U. PORTO

Identifying Novel Genetic Causes for Hereditary Myopathies: from Conventional Approaches to Next-Generation Sequencing

Jorge Oliveira

D
2019

Jorge Manuel Santos Marques Oliveira

IDENTIFYING NOVEL GENETIC CAUSES FOR HEREDITARY MYOPATHIES:

FROM CONVENTIONAL APPROACHES TO NEXT-GENERATION SEQUENCING

Tese de Candidatura ao grau de Doutor em Ciências Biomédicas submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Orientador – Prof. Doutor Mário Sousa
Categoria – Professor Catedrático
Afiliação – Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

Coorientador – Prof. Doutor José Carlos Machado
Categoria – Professor Associado
Afiliação – Faculdade de Medicina da Universidade do Porto

Coorientador – Doutora Rosário Santos
Categoria – Assessor Principal
Afiliação – Centro Hospitalar Universitário do Porto

ACKNOWLEDGMENTS

In a more personal perspective, my special thanks to my parents, parents-in-law, remaining family and friends, for all the incentives, support, and, above all, patience over such a demanding period of my life.

Finally, for my wife Marta and my children Ana Jorge, Carolina, Guilherme and Duarte, to whom I would like to dedicate this work, I have no words to express my deepest gratitude for all your sacrifices. I am sorry for not being as supportive and available as I should have been. Also, for not spending more quality time as you all deserve. Sincerely hope that, in the future, you can understand the path I've chosen to undertake during these years.

LIST OF PUBLICATIONS

Ao abrigo do disposto do n.º 2, alínea a) do artigo 31.º do Decreto-Lei n.º 115/2013 de 7 de agosto, constam nesta tese os seguintes artigos já publicados:

- **Oliveira J**, Oliveira ME, Kress W, Taipa R, Pires MM, Hilbert P, Baxter P, Santos M, Buermans H, den Dunnen JT, Santos R. (2013). Expanding the MTM1 mutational spectrum: novel variants including the first multi-exonic duplication and development of a locus-specific database. European Journal of Human Genetics, 21(5):540-549.

- **Oliveira J,** Gonçalves A, Oliveira ME, Fineza I, Pavanello RC, Vainzof M, Bronze-da-Rocha E, Santos R, Sousa M. (2014). Reviewing large LAMA2 deletions and duplications in congenital muscular dystrophy patients. Journal of Neuromuscular Diseases, 1(2):169-179.

- **Oliveira J**, Negrão L, Fineza I, Taipa R, Melo-Pires M, Fortuna AM, Gonçalves AR, Froufe H, Egas C, Santos R, Sousa M. (2015). New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing. Journal of Human Genetics, 60(6):305-312.

- **Oliveira J,** Gonçalves A, Taipa R, Melo-Pires M, Oliveira ME, Costa JL, Machado JC, Medeiros E, Coelho T, Santos M, Santos R, Sousa M. (2016). New massive parallel sequencing approach improves the genetic characterization of congenital myopathies. Journal of Human Genetics, 61(6):497-505.

- **Oliveira J,** Martins M, Leite RP, Sousa M, Santos R. (2017). The new neuromuscular disease related with defects in the ASC-1 complex: report of a second case confirms ASCC1 involvement. Clinical Genetics, 92(4):434-439.

- Gonçalves A*, **Oliveira J***, Coelho T, Taipa R, Melo-Pires M, Sousa M, Santos R. (2017). Exonization of an intronic LINE-1 element causing Becker Muscular Dystrophy as a novel mutational mechanism in Dystrophin gene. Genes (Basel), 8(10). pii: E253. *- Equally contributing authors.

- Samões R*, **Oliveira J***, Taipa R, Coelho T, Cardoso M, Gonçalves A, Santos R, Melo Pires M, Santos M. (2017). RYR1-related myopathies: clinical, histopathologic and genetic heterogeneity among 17 patients from a Portuguese tertiary centre. Journal of Neuromuscular Diseases, 4(1):67-76. *- Equally contributing authors.

- **Oliveira J,** Pereira R, Santos R, Sousa M. (2017). Homozygosity mapping using Whole-Exome Sequencing: A Valuable Approach for Pathogenic Variant Identification in Genetic Diseases. Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 3: BIOINFORMATICS, 210-216.

- **Oliveira J,** Gruber A, Cardoso M, Taipa R, Fineza I, Gonçalves A, Laner A, Winder TL, Schroeder J, Rath J, Oliveira ME, Vieira E, Sousa AP, Vieira JP, Lourenço T, Almendra L, Negrão L, Santos M, Melo-Pires M, Coelho T, den Dunnen JT, Santos R, Sousa M. (2018). LAMA2 gene mutation update: toward a comprehensive picture of the laminin-α2 variome and its related phenotypes. Human Mutation, 39:1314-1337.

Em cumprimento das disposições do referido Decreto-Lei e das recomendações da comissão de ética da Universidade do Porto, o candidato declara que efetuou trabalho experimental, recolha e tratamento dados, tendo elaborado os referidos artigos com a colaboração ativa dos respetivos co-autores.

ABSTRACT

This academic thesis targeted the identification of novel genetic causes in a vast group of diseases primarily affecting the skeletal muscle, and collectively known as hereditary myopathies (HM). HM traditional classification subdivides further into: congenital muscular diseases [congenital myopathies (CM) and congenital muscular dystrophies], progressive muscular dystrophies (MD) [Duchenne/Becker MD (D/BMD), limb-girdle MD (LGMD), Emery-Dreifuss MD and facioscapulohumeral MD], myofibrillar and distal myopathies. Although distinguishable from a clinical and histopathological standpoint, there is considerable genetic overlap between these groups of diseases. This overlap, their vast clinical heterogeneity, and the potential for novel disease-causing collectively explain how a large number of clinical cases still remain genetically unsolved. Besides the overarching objective of expanding HM genetic etiology, this work was also geared towards obtaining refined genotype-phenotype correlations, expand the phenotypic spectrum of HM, and develop and implement new sequencing strategies to facilitate its genetic characterization.

The author implemented the genetic analysis of congenital muscle diseases at a national basis over a time-span of 13 years totalizing the analysis of 246 patients. The first part of the work consisted of the reassessment of the genetic data gathered through conventional approaches: Sanger sequencing complemented by multiplex ligation-dependent probe amplification (MLPA). Overall, 99 distinct pathogenic variants over thirteen different genes were identified in 112 patients (106 families). Variants in laminin-α2 (*LAMA2*), myotubularin (*MTM1*) and ryanodine receptor 1 (*RYR1*) genes accounted for more than three-quarters of the genetically characterized patients. Based on a systematic screening by MLPA, novel large deletions and duplications in both *MTM1* and *LAMA2 loci* were identified for the first time ever.

To assist variant collection on a global scale and to obtain further insights about the impact of novel variants and new genotype-phenotype correlations, two publicly available locus-specific databases, were developed for *LAMA2* (http://tiny.cc/LOVD_LAMA2) and *MTM1* (http://tiny.cc/LOVD_MTM1), using the Leiden Open Variation Database (LOVD) software and updated as part of this work. The analysis of variant data in both datasets showed a non-random distribution of missense variants in laminin-α2 and myotubularin, as they clustered in specific domains of the protein with relevant biological function.

The phenotype associated with *LAMA2* defects was progressively expanded to include milder forms with later disease-onset. Most of these phenotypes were correlated with the presence of missense variants, such as the one previously reported

(p.Thr821Pro) in two unrelated Portuguese patients. To expand present knowledge about this variant, additional patients were described. Variant screening showed a statistically significant frequency difference (58-fold increase) in MD patients when compared to a control group.

This research also comprised the development and application of different next-generation sequencing (NGS) approaches, namely gene panels, whole-exome sequencing (WES) and whole-genome sequencing (WGS) to address the genetic etiology of HM. Besides highly increasing the likelihood of identifying novel genetic defects in HM patients, these novel approaches were effective strategies to cope with the genetic and clinical heterogeneity seen in these diseases. To evaluate the impact of novel variants and mutational mechanisms genetic studies were complemented with gene expression analysis. From the diversity of bioinformatic algorithms applied in this work, one effective strategy was the detection of runs of homozygosity (ROH) using WES data.

In the present work, a new NGS gene panel was developed to analyze 20 genes, in simultaneously, known to be defective in CM. A total of 12 genetically unsolved patients were carefully selected and their DNA sequenced using the Ion Torrent NGS platform. Pathogenic variants were identified in *ACTA1*, *NEB*, *RYR1* and *TTN* genes in eight patients, demonstrating the effectiveness and utility of this strategy. Using a commercial gene panel available for neurological diseases (in four cases), an additional variant in the *MATR3* gene was identified in a patient with an adult-onset distal myopathy and rimmed vacuoles.

Further increasing the resolving power, WES was performed in eight unrelated patients (a total of 12 individuals: six singletons and two trios). In five of such cases, pathogenic or likely pathogenic variants were identified in *ASCC1*, *CCDC78*, *CHKB*, *KLHL41,* and *LAMA2* genes. This task provided the first-ever patient to be entirely solved by WES in Portugal (http://tiny.cc/jhg_2015_20). As these genes were previously reported as causing genetic diseases in a very limited number of patients (typically less than a dozen), this new data generated also contributed to expand the mutational and clinical spectra of these disorders. An outstanding contribution of this research was the confirmation of a new disease-causing gene (*ASCC1*) associated with a severe neuromuscular disease.

Finally, out of 28 genetically unsolved cases of dystrophinopathies, two patients with BMD phenotype and a strong level of clinical suspicion were selected for further research. Although thoroughly studied by several routine techniques the pathogenic variant in *DMD* failed to be identified. Following *DMD* transcript analysis different bioinformatic and NGS strategies were used to further investigate these cases. In one

of the patients, by employing low-coverage WGS a large (~8 Mb) genomic inversion involving part of the *DMD* gene was identified, leading to the loss of the C-terminal region of dystrophin. In the second case, a novel mutational mechanism in the *DMD* gene was identified, caused by the partial exonization of a new LINE-1 element deeply inserted in intron 51 of *DMD*. The documentation of such extremely rare events brought to discussion the importance of mobile genomic elements such as LINE-1 as a potential cause of genetic diseases and from a cell biology perspective their importance as a genomic evolutionary driving force.

Globally, the clinical heterogeneity and the diversity of genetic profiles of the patients here presented provided clear indications that the gene-by-gene (stepwise) analysis through conventional genomic approaches is no longer the most efficient strategy. Although muscle biopsy is key for diagnoses, a targeted genetic analysis can be effective as a first-tier approach, avoiding an invasive procedure in some cases. However, in the case of CM, an integrated analysis of the patient's muscle biopsy with the results from the NGS-enabled gene panel was shown to be of extreme clinical utility. As for patients with muscle dystrophies, it would be more effective to use more comprehensive NGS strategies considering the increasing genetic overlap between the congenital and the later disease-onset forms of MD. This overlap is likely to be one of the major explanations for the significant number of HM patients that remained genetically uncharacterized for several years. The best example were the MD patients with *LAMA2* defects found in this work, that ultimately contributed to include this gene as an additional cause for LGMD (https://omim.org/entry/618138/).

The collated genetic data also contributed towards refined genotype-phenotype correlations in *MTM1*- and *LAMA2*-related diseases, particularly for missense variants. The identification of such patterns has strong clinical implications, mainly for disease prognostic purposes. This research also contributed with additional insights about the possible pathophysiological mechanisms in the *ASCC1*-related disease and its phenotypic delineation. Its initial classification as spinal muscular atrophy with bone fractures was arguably, and additional data could justify its inclusion in CM, or even in both groups.

Besides promoting pioneering work this research also laid the ground foundations for new sequencing capacities, further genomic research and the possibility of improved diagnostics at Centro Hospitalar Universitário do Porto.

RESUMO

Esta tese académica teve como principal objetivo a identificação de novas causas genéticas para um vasto grupo de doenças que afetam o músculo esquelético, coletivamente designadas como miopatias hereditárias (MH). Na sua classificação tradicional estas miopatias são subdivididas em: doenças musculares congénitas [nomeadamente as miopatias congénitas (MC) e as distrofias musculares congénitas], distrofias musculares (DM) progressivas [nomeadamente a DM de Duchenne / Becker (DMD/ DMB), DM das cinturas (LGMD), DM de Emery-Dreifuss e DM facioscapuloumeral], miopatias miofibrilares e distais. Apesar de se diferenciaram do ponto de vista clínico e histopatológico, existe considerável sobreposição genética entre os diferentes grupos. Para além do objetivo de expandir a etiologia genética das MH, este trabalho pretendeu refinar as correlações genótipo-fenótipo, expandir o espectro fenotípico das MH, bem como o desenvolvimento e implementação de novas abordagens de sequenciação por forma a facilitar a sua caracterização genética.

O autor implementou a análise genética das doenças musculares congénitas a nível nacional, que num período de 13 anos totalizou 246 casos estudados. A primeira parte do trabalho consistiu na descrição dos dados genéticos recolhidos nestes estudos, primariamente recorrendo a abordagens convencionais: sequenciação de Sanger complementada com a técnica de MLPA ("multiplex ligation-dependent probe amplification"). No total, foram detetadas 99 variantes patogénicas distintas (em 13 genes) identificadas em 112 doentes (106 famílias). Variantes nos genes laminina-α2 (*LAMA2*), miotubularina 1 (*MTM1*) e receptor 1 da rianodina (*RYR1*) foram responsáveis por mais de três quartos dos doentes geneticamente caracterizados. O rastreio sistemático realizado em doentes sem diagnóstico pela técnica de MLPA, permitiu identificar pela primeira vez novos tipos de variantes patogénicas, mais concretamente grandes duplicações e deleções, nos genes *MTM1* e *LAMA2*.

Por forma adjuvar a identificação e classificação de novas variantes numa perspetiva internacional, bem como obter novos dados sobre o impacto de novas variantes e sobre a correlação genótipo-fenótipo, duas bases de dados específicas de *locus* (LSDB). Estas estão disponíveis publicamente, e foram implementadas para os genes *LAMA2* (http://tiny.cc/LOVD_LAMA2) e *MTM1* (http://tiny.cc/LOVD_MTM1), recorrendo ao software Leiden Open Variation Database (LOVD) e atualizados como parte deste trabalho. A análise dos dados nestes repositórios mostrou uma distribuição não aleatória das variantes do "tipo missense" na laminina-α2 e na miotubularina 1, dado que se agrupam em domínios específicos da proteína com função biológica relevante. O fenótipo associado aos defeitos no gene *LAMA2* foi expandido nos últimos anos por forma a incluir formas de início mais tardio da doença e com quadro clínico mais ligeiro e/ou lentamente progressiva. A maioria desses

fenótipos está correlacionada com a presença de variantes do tipo "missense", como a alteração p.Thr821Pro anteriormente descrita em dois doentes portugueses. Por forma a expandir o conhecimento relativamente a esta alteração, foram caracterizados mais doentes com DM com esta variante patogénica. A realização de um rastreio sistemático desta variante em dois grupos, mostrou uma diferença estatisticamente significativa (aumento de 58 vezes) entre a frequência nos doentes com DM e no grupo controle.

Este trabalho prosseguiu com desenvolvimento e a aplicação de diferentes abordagens de sequenciação de nova geração (NGS), mais concretamente, painéis de genes, sequenciação total do exoma (WES) e sequenciação total do genoma (WGS). Além de aumentar a probabilidade de identificar novos defeitos genéticos em pacientes com MH, essas novas abordagens revelaram-se eficientes para abordar a heterogeneidade genética e clínica observada nas MH. Para avaliar o impacto de novas variantes e respetivos mecanismos mutacionais, os estudos genéticos efetuados a nível genómico foram complementados com a análise da expressão génica. A partir da aplicação de diferentes algoritmos bioinformáticos, foi utilizada uma estratégia eficaz para o estudo de doentes com consanguinidade parental, recorrendo à deteção de regiões de homozigotia (ROH) com os dados gerados de WES.

O novo painel de NGS desenvolvido permitiu analisar em simultâneo 20 genes previamente associados às MC. No total 12 doentes foram criteriosamente selecionados, a partir do grupo de doentes sem diagnóstico, e sequenciados com recurso ao painel de gene recorrendo à plataforma Ion Torrent. Foram identificadas variantes patogénicas em oito doentes, nos genes *ACTA1*, *NEB*, *RYR1* e *TTN* demonstrando a eficácia e a utilidade da referida estratégia. Através do uso um painel de genes envolvidos em doenças neurológicas, foi identificada uma variante adicional no gene *MATR3* num doente com miopatia distal de início adulto e com vacúolos bordejados na biópsia muscular.

A análise do exoma foi realizada em oito pacientes não relacionados (um total de 12 indivíduos: seis exomas individuais e dois trios). Em cinco destes casos, variantes patogénicas ou provavelmente patogénicas foram identificadas nos genes *ASCC1*, *CCDC78*, *CHKB*, *KLHL41* e *LAMA2*. Esta tarefa do projeto permitiu obter o primeiro diagnóstico genético recorrendo à WES inteiramente realizado em Portugal (http://tiny.cc/jhg_2015_20). Como a maioria das variantes patogénicas foram identificadas em genes associados a um número muito limitado de doentes (menos de duas dezenas de casos), os novos dados obtidos também contribuíram para expandir os espectros clínicos e mutacionais destas doenças. Uma contribuição notável foi a confirmação do envolvimento de um novo gene, que codifica para a subunidade 1 do complexo cointegrador 1 activador de sinal (*ASCC1*), como causa de uma nova doença neuromuscular congénita.

Finalmente, a partir de 28 casos de distrofinopatias geneticamente não esclarecidos, foram selecionados dois doentes, com uma forte suspeita clínica de DMB, para prosseguir

investigação. Apesar de estudados durante vários anos por diversas técnicas de rotina, a variante patogénica no gene *DMD* não tinha sido identificada. A primeira abordagem, consistiu na análise ao nível dos transcritos de *DMD*, seguindo-se diferentes estratégias bioinformáticas e de NGS. Num dos doentes, recorrendo a WGS de baixa cobertura foi identificada uma grande inversão genómica (~8 Mb) envolvendo parte do gene *DMD*, que originou a perda da região C-terminal da distrofina em virtude de um efeito posicional. No segundo caso, identificou-se um novo mecanismo mutacional no gene *DMD*, causado pela exonização parcial de um novo elemento LINE-1 que se inseriu no intrão 51 deste gene. Este tipo de elementos para além de potenciais causadores de doenças genéticas são, numa perspetiva da biologia celular, uma importante força motriz para a evolução do genoma humano.

A heterogeneidade clínica e a diversidade de perfis genéticos dos doentes apresentados neste trabalho sugerem que análise gene-a-gene já não se configura, atualmente, como a estratégia de diagnóstico mais eficiente. Embora a realização da biópsia muscular seja fundamental para o diagnóstico, a análise genética pode ser eficaz como uma abordagem de primeira linha, evitando um procedimento invasivo em alguns casos. No entanto, no caso das MC, a análise integrada da biópsia muscular com a caracterização genética com o painel de NGS desenvolvido revelou-se de extrema utilidade. Relativamente aos pacientes com distrofias musculares a aplicação de estratégias de NGS mais abrangentes será mais indicada, dada a crescente sobreposição entre as formas congénitas e as progressivas. Esta sobreposição será uma das principais explicações para o número significativo de doentes ainda não caracterizados geneticamente. Este aspeto foi também evidente nos doentes com defeitos no gene *LAMA2* caracterizados neste trabalho e que contribuíram para a inclusão deste *locus* como uma causa adicional de LGMD (https://omim.org/entry/618138/). Globalmente, os dados genéticos obtidos nesta investigação contribuíram para determinar novas correlações genótipo-fenótipo nas doenças relacionadas com os genes *MTM1* e *LAMA2*. A identificação destes padrões mutacionais são importantes para prognóstico clínico e desenvolvimento de futuras terapias. Esta investigação também contribuiu com novos dados relativos à doença neuromuscular relacionada com o gene *ASCC1*, permitindo mais detalhes fenotípicos e sobre a sua possível patofisiologia. Apesar da sua classificação inicial como atrofia muscular espinhal com fraturas ósseas existem algumas evidências para que esta entidade possa ser incluída também no subgrupo das MC.

Para além de promover este trabalho pioneiro no Centro Hospitalar Universitário do Porto, esta investigação serviu como base para a disponibilização de novos recursos de sequenciação e investigação genómica, bem como a possibilidade de obtenção de diagnósticos mais precisos neste grupo de doenças primárias do músculo esquelético.

ABBREVIATIONS

ATP	adenosine triphosphate	IGV	integrative genomic viewer
AD	autosomal dominant	IHC	immunohistochemistry
AR	autosomal recessive	INDELs	insertions or deletions
BAM	binary alignment map	LGMD	limb-girdle muscular dystrophies
BM	Bethlem myopathy	LOVD	Leiden open variation database
BMD	Becker muscular dystrophy	LSDB	locus-specific databases
bp	base pairs	MD	muscular dystrophy
CCD	central core disease	MDC1A	congenital muscular dystrophy type 1A
CFTD	congenital fiber-type disproportion	MDDG	muscle-dystrophy dystroglycanopathy
CK	creatine kinase	MEB	muscle-eye-brain disease
CM	congenital myopathy	MFM	myofibrillar myopathies
CMD	congenital muscular dystrophies	MHS	malignant hyperthermia susceptibility
CMOS	complementary metal–oxide–semiconductor	MLPA	multiplex ligation-dependent probe amplification
CNM	centronuclear myopathy	MmD	multiminicore disease
CNS	central nervous system	MNV	multi-nucleotide variants
CNVs	copy number variants	MPS	massively parallel sequencing
CRISPR	clustered regularly interspaced short palindromic repeats	MRI	magnetic resonance imaging
		mRNA	messenger RNA
CT	computerized tomography	NADH-TR	nicotinamide adenine dinucleotide tetrazolium reductase
DAPC	dystrophin-associated protein complex		
DGC	dystrophin-glycoprotein complex	NEM	nemaline myopathy
DHPRs	dihydropyridine receptors	NGS	next-generation sequencing
DMD	Duchenne muscular dystrophy	NIPT	non-invasive prenatal tests
DMYO	distal myopathies	NMJ	neuromuscular junction
DNA	deoxyribonucleic acid	nNOS	neuronal nitric oxide synthase
dNTP	deoxynucleotide triphosphate	OLC	overlap-layout-consensus
ECC	excitation-contraction coupling	PCR	polymerase chain reaction
ECM	extracellular matrix	PGM	personal genome machine
EDMD	Emery-Dreifuss muscular dystrophy	PMD	progressive muscular dystrophy
EMG	electromyography	QC	quality control
FCMD	Fukuyama congenital muscular dystrophies	RNA	ribonucleic acid
		rRNA	ribosomal RNA
FP-PCR	fluorescent primer PCR	SAM	sequence alignment map
FSHD	facioscapulohumeral muscular dystrophy	SBS	sequencing-by-synthesis
		SERCA	sarcoplasmic reticulum calcium ATPase
gDNA	genomic DNA	SMRT	single-molecule real-time
GVDs	genetic variants databases	SNV	single nucleotide variants
HM	hereditary myopathies	SPRI	solid-phase reversible immobilization
HMERF	hereditary myopathy with early respiratory failure	SR	sarcoplasmic reticulum
		SV	structural variants
HPO	human phenotype ontology	TP-PCR	triplet repeat primed PCR

T-tubules	transverse tubules	WGA	whole-genome amplification
TVC	torrent variant caller	WGS	whole-genome sequencing
uBAM	unaligned binary alignment map	WMC	white matter changes
UCMD	Ullrich congenital muscular dystrophy	WWS	Walker-Warburg syndrome
VCF	variant calling format	XLR	X chromosome-linked recessive
WES	whole-exome sequencing	ZMW	zero-mode waveguide

Amino acids:

A- Ala	Alanine
C- Cys	Cysteine
D- Asp	Aspartic acid
E- Glu	Glutamic acid
F - Phe	Phenylalanine
G- Gly	Glycine
H- His	Histidine
I- Ile	Isoleucine
K- Lys	Lysine
L- Leu	Leucine
M- Met	Methionine
N- Asn	Asparagine
P- Pro	Proline
Q- Gln	Glutamine
R- Arg	Arginine
S- Ser	Serine
T- Thr	Threonine
V- Val	Valine
W- Trp	Tryptophane
Y- Tyr	Tyrosine
X- Ter	STOP codon

Nucleotides:

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracil

TABLE OF CONTENTS

Chapter 1

GENERAL INTRODUCTION

Content

1.1 THE SKELETAL MUSCLE

It is common-sense to associate the capacity to generate strength and movement with skeletal muscle. However, the current scientific knowledge gathered from cell biology and physiology extended its role much further. There is a very broad involvement of skeletal muscle over the human body, providing shape, stability to the skeleton and joints. Also ensuring basic bodily functions such as ingestion, excretion or egestion under voluntary control. Skeletal muscles also constitute an important protective layer over the body vital organs against external trauma or aggression. There is an additional function in maintaining homeostasis through the generation of heat via muscle contractions. Also, as it is being increasingly recognized, skeletal muscle constitutes an important metabolic site, as the largest insulin-sensitive body tissue and the primary site for insulin-stimulated glucose utilization (Stump et al., 2006). Therefore, it is not surprising that skeletal muscle represents one of the most paradigmatic examples of a dynamic and plastic tissue, within all those that compose the human body.

In humans, skeletal muscle accounts for approximately 40% of the total body weight, contains as much as 50 to 75% of all proteins and contributes towards 30 to 50% of whole-body protein turnover (Frontera & Ochala, 2015). As further detailed in this chapter, the remarkable and unmatched protein enrichment found in skeletal muscle is entirely supported at the cellular and genomic level, where particularly large genes encode the large and complex proteins. Such a unique organization of the human genome should be kept in mind when revisiting not only the normal structure of the skeletal muscle but also the pathological changes found within a wide variety of disorders affecting skeletal muscles.

The fundamental and basic cellular unit of the skeletal muscle is the myofiber (Figure 1.1.1). The muscular system contains other important components: contractile apparatus (aka, the sarcomere), connective tissue or extracellular matrix (ECM), mitochondria (in high abundance to provide energy), nerves and blood capillaries to supply the muscle (Mahdy, 2019). ECM represents as much as 10% of the skeletal muscle weight and plays a major role in force transmission, maintenance, and repair of muscle fibers following injury. ECM also constitutes the backbone structure that maintains myofibers, blood vessels, and nerves (Mahdy, 2019). Basically, the ECM is distributed in three different layers recognizable in skeletal muscle. The individual myofibers are enfolded by an endomysium sheath of connective tissue, groups of myofibers (fascicles) are wrapped by the perimysium, and finally, these are also sheathed in a dense irregular layer of connective tissue called epimysium (reviewed by McNally & Pytel, 2007). In the very end of the myofiber, the myotendinous junctions are established, consisting of a specialized connection to bony insertion that can withstand a considerable amount of stress forces (McNally & Pytel, 2007).

At the subcellular level myofibers are highly differentiated, the cytoplasm of an individual mature myofiber is highly specialized and organized being known as sarcoplasm, that contains several chains of the contractile apparatus, known as sarcomeres, located along the run of the myofiber. As for the plasma membrane of muscle is called the sarcolemma, whereas the specialized smooth endoplasmic reticulum, being involved in storage and release of calcium ions (Ca^{2+}), is known as the sarcoplasmic reticulum (SR).
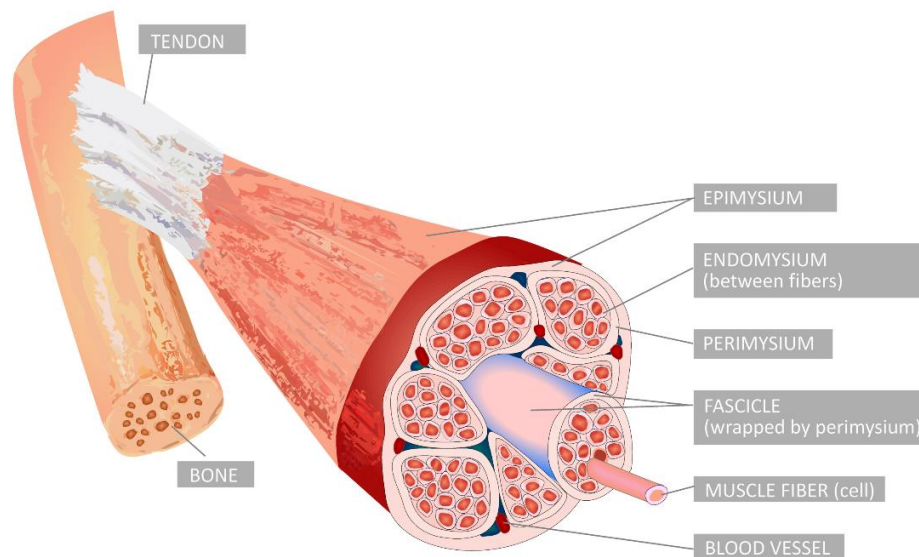
Figure 1.1.1- Skeletal muscle illustration depicting its structure and the association with tendon and bone. The extracellular matrix in skeletal muscles is arranged in three distinct layers: the endomysium (between myofibers), perimysium (around each fascicle) and epimysium (the outermost layer, surrounds the entire muscle).

During embryogenesis, the structures precursors of muscles - known as the somites - derive from the paraxial mesoderm (starting from day 20 after conception), which later segments into structures on either side of the neural tube and notochord (Buckingham, 2003). Proceeding through the embryo development, the somites differentiate into the dermomyotome and finally establishing the myotome. At the cellular level, myoblasts undergo frequent divisions, during embryogenesis and merge establishing a multinucleated, syncytial and elongated muscle fiber or the myotube, and displaying an undifferentiated contractile apparatus (Hill, 2018). This process is called myogenesis, can also occur after birth in the so-called postnatal growth and during regeneration after damage as a result of the activation of quiescent muscle stem cells (satellite cells) (Blais, 2015). At this stage during development, the nuclei of myotubes are still centrally located in the muscle fiber. As the synthesis of the myofilaments/myofibrils occurs, the nuclei are gradually shifted towards the periphery of the myofiber and with fully differentiated sarcomeres. Also, during

development, two populations of myofibers can be recognized: first-formed myofibers act as a structural framework, upon which myoblasts proliferate and are fused in a linear sequence; and secondary myofibers, a later population of cells that form surrounding the primary (Hill, 2018).

An essential structure for the skeletal muscle function is the sarcomere, constituting the basic contractile unit (Figure 1.1.2). Within sarcomere two proteins (myofilaments) are the most abundant, actin and myosin, comprising most of the protein content within a single myofiber (Frontera & Ochala, 2015). Actin is a globular protein (G-actin) which polymerizes into double helical strands (F-actin). This polymerization of actin involves splitting ATP and binding of ADP, comprising nearly 90% of the total ADP in muscle (Jones et al., 2004). The thin filaments, besides actin, are composed of tropomyosin and the three different (C, I and T) troponin molecules (Jones et al., 2004). Those proteins constitute a regulatory complex, which is the calcium-dependent troponin complex, having important roles in the activation process that ultimately leads to myofilament sliding and the generation force (Frontera & Ochala, 2015). Myosin, the main molecular motor, is a hexamer consisting of two heavy chains and four light chains, all composing the thick filaments. The amino-terminal region of myosin assumes a globular nature and the two pairs of light chains bind at the junction between the globular head and the carboxy-terminal rod domain. The headpiece contains the enzymatic active site required to perform both the ATP hydrolysis and the actin-binding (McNally et al., 2006).
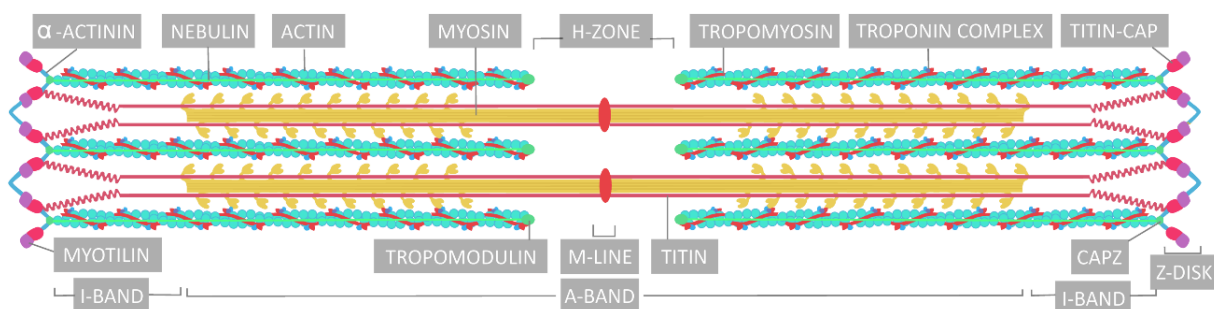
Figure 1.1.2- The basic structure of the sarcomere, depicting the thin and thick filaments, along with the main regions Z-disk, I- and A-bands. Adapted from Rahimov & Kunkel, 2013.

Z-disks define the lateral edges of the sarcomeres and cross-link the protruding ends of actin-based thin filaments from adjacent sarcomeres via α-actinin protein (Henderson et al., 2017). This sarcomeric region act as an anchor site for the N-terminal domain of titin and nebulin/nebulette filamentary systems, making it indispensable for contractile force transmission (Henderson et al., 2017). Titin is a gigantic protein (composed by over 38,000

amino acids) that helps to stabilize and align the thick filaments. It acts as a molecular spring, connecting a wide variety of molecular patterners, being essential to the passive mechanical properties of the myofilaments (Freundt & Linke, 2018). Nebulin, in its turn, its integrated within the thin filaments where it regulates actin-myosin interactions and regulates calcium uptake to the SR (Tskhovrebova & Trinick, 2017; Henderson et al., 2017). For muscle contraction, and according to the sliding filaments theory, actin and myosin filaments must overlap, sliding one over the other, causing the sarcomere to shorten or to increase the length (known as eccentric contraction). Such process is triggered by the release of calcium ions which enables the process of ATP hydrolysis and consequently, the molecular interaction of myosin with actin (McNally et al., 2006).

Excitation-contraction coupling (ECC), the fundamental mechanism for muscle function, is explainable by a cascade of events, occurring from the generation of the action potential in the skeletal muscle fibers to the beginning of muscle tension (Calderón et al., 2014). Two main processes are required for the generation of force: the transmission of a stimulus from the nerve to the triad followed by the rapid release of calcium from the SR cisternae and the interaction between actin and myosin that forms cross-bridges (Frontera & Ochala, 2015). The molecular events during ECC have been characterized in further detail:

i) the activation of skeletal muscle at the neuromuscular junction (NMJ), by the nerve impulse, induces a membrane depolarization that propagates along the plasma membrane to invaginations called "transverse tubules" (T-tubules, Figure 1.1.3) (Marty & Fauré, 2016);

ii) depolarization in this structure activate L-type calcium channel dihydropyridine receptors (DHPRs);

iii) the voltage sensor subunits of the DHPRs stimulates the opening of the SR calcium channel and ryanodine receptors type 1 (RYR1) allowing an influx of calcium. This ECC process is performed in a specific location, where the T-tubule is near the terminal cisternae of the SR that stores calcium - the triad - which is localized specifically at the transition of A-I bands in mammalian skeletal muscle;

iv) the opening of the ryanodine receptors in the terminal cisternae of the SR is promoted by the calcium flow, through this opening a significant amount of calcium is transferred to the sarcoplasm;

v) the released calcium binds to troponin C protein located at the actin thin myofilament, triggering a series of molecular events that displace the tropomyosin which was blocking the active site of the actin filament. This is part of the cross-bridge cycle, the molecular basis for the sliding filamin theory that explains muscle contraction. Subsequently, the exposure of the active-site on actin allows the binding of the myosin head molecule with actin. ATP and the ATPase also located in the myosin head, facilitate the detachment of myosin from actin (a cross-bridge formed in a previous contraction) and the formation of a new cross-bridge. The

speed of each muscle contraction is proportional to the rate at which myosin hydrolyzes ATP. Muscle relaxation occurs after the closing of RYR1, and calcium uptake into the SR by the sarcoplasmic reticulum calcium ATPase (SERCA) (Marty & Fauré, 2016).
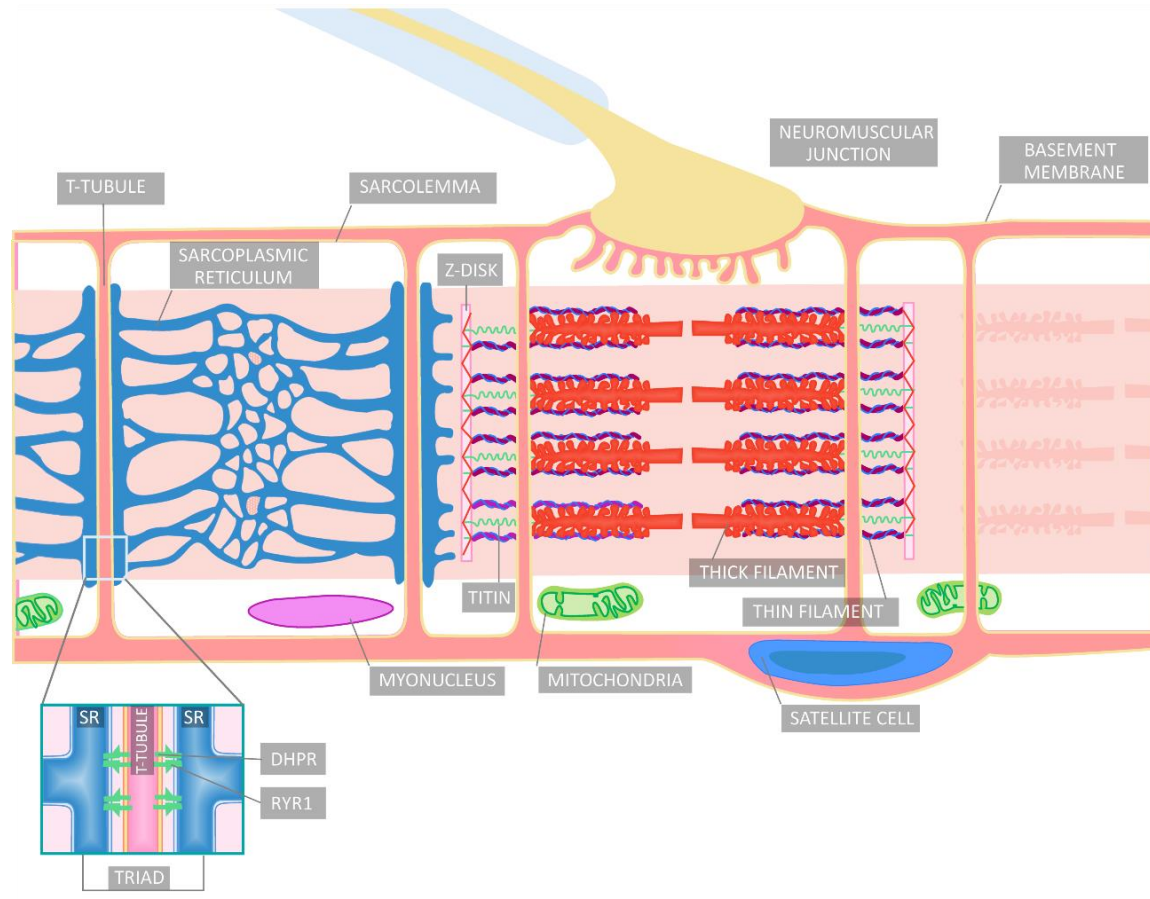
Figure 1.1.3- Schematic representation of a myofiber, highlighting its unique intracellular configuration, more specifically the transverse tubules (T-tubules) and the sarcoplasmic reticulum (SR) systems, which establish the triad in specific sites. Adapted from Ravenscroft et al., 2015.

1.2 HEREDITARY MYOPATHIES: PHENOTYPES, GENES AND PATHOPHYSIOLOGY

Hereditary myopathies (HM), also known as primary muscle disorders, are a vast group of heterogeneous and genetically-determined diseases, having as a common denominator the pathological changes in the myofiber structure, function or its capacity to regenerate. In this section HM will be presented following its traditional classification, further subdivided according to following dichotomies: congenital *versus* progressive; dystrophic *versus* myopathic; proximal *versus* distal. According to this classification, the major groups are: i) congenital muscular diseases, more specifically: congenital myopathies and congenital muscular dystrophies; ii) progressive muscular dystrophies: Duchenne/Becker muscular dystrophies, limb-girdle muscular dystrophies, Emery-Dreifuss muscular dystrophies and facioscapulohumeral muscular dystrophies; and finally, iii) myofibrillar and distal myopathies.

The diagnosis of HM involves three closely related areas of expertise: clinical (neurology and/or pediatric neurology), the neuropathology and the molecular genetics (Laing, 2012). Clinically, several signs and symptoms are usually important for diagnosis: age of onset, severity and disease progression, pattern and distribution of muscle weakness. The involvement of the systems and organs, namely the central nervous system (CNS), heart or respiratory function is also evaluated. Serum levels of creatine kinase (CK) are usually highly elevated in muscular dystrophies, and other enzymes such as alanine aminotransferase, aspartate aminotransferase, and aldolase. In addition, other invasive [muscle/nerve biopsy and electromyography (EMG)] and non-invasive techniques [namely computerized tomography (CT) scan and magnetic resonance imaging (MRI) of muscle and/or brain] are necessary to complete the diagnostic workup. Muscle histopathology is of extreme importance to distinguish, at a higher level, a myogenic condition (primary defect in muscle) from other NMD conditions of neurogenic or myasthenic origin. It is also important to classify HM into different subtypes such as muscular dystrophies or structural (congenital) myopathies. With the advent of the molecular era, specific antibodies were developed to detect proteins involved in these diseases. Moreover, changes in the expression profile of certain muscle proteins (e.g. those from the dystrophin-associated protein complex) are routinely assessed by immunohistochemistry (IHC) or western-blot using specific antibodies. These can be used to identify the disease etiology and to steer the genetic analysis. However, for most of the proteins involved in HM, there are no specific antibodies available to perform IHC. Alongside genetic discovery, outstanding progress in the understanding of pathophysiological mechanisms has been made during the past two decades (Schorling et al., 2017). This knowledge also had implication in the classification the HM, being progressively designated according to the genetic entity (e.g. *dystrophinopathies*) or even the main pathophysiological mechanism involved (e.g. *sarcomeropathies* or *triadopathies*).

1.2.1 CONGENITAL MYOPATHIES

The term congenital myopathy (CM) was initially adopted by Magee and Shy (1956) to describe a wide diversity of muscle diseases with congenital onset, that share characteristic histopathological and ultrastructural changes in muscle biopsy, and absence of dystrophic features (Schorling et al., 2017). Thus, CM comprises a heterogeneous group of diseases from a clinical and genetic perspective (Table 1.2.1).

Clinical aspects of CM

The initial symptoms of a baby with CM include hypotonia, feeding difficulties, muscle weakness, and even respiratory distress in some patients, and are typically noticed around the neonatal period (Gonorazky et al., 2018). However, it should be stressed that later onset, typically within the first two years of life or even diagnosed during adulthood in rarer cases, has been reported in some phenotypes of CM (Jungbluth & Voermans, 2016). Thus, the clinical spectrum of CM extends from very severe neonatal forms with congenital arthrogryposis and hypomobility to milder childhood-onset subtypes usually non-progressive or slow-progression of muscle weakness. This vast clinical variability does not associate exclusively with the *locus* involved, but rather with the nature or type of variants and the underlying pathophysiological mechanism involved.

At the severest end of this spectrum, the first symptoms are generally more exacerbated, and may consist of polyhydramnios and reduced fetal movements leading to the development of arthrogryposis and clubfoot (Cassandrini et al., 2017). Severe muscle hypotonia is often present at birth and/or during the first months of life (floppy baby syndrome), together with muscle weakness, feeding difficulties and respiratory insufficiency (Gonorazky et al., 2018). Dysmorphic features, secondary to muscle weakness, are also characteristic in CM patients: with high arched palate, micrognathia and *pectus carinatum* or *excavatum*, the latter resulting from chest muscles weakness (Cassandrini et al., 2017). Other nonspecific clinical features of some CM subtypes include ocular muscles weakness (ophthalmoplegia or ophthalmoparesis), ptosis and even strabismus; and bulbar involvement (Mah & Joseph, 2016). As the child grows older, motor milestones are considerably delayed or even not attained, while hypotonia and muscle weakness remains relatively stable. Bone and joint deformities may appear at this stage, such as contractures, lordosis, scoliosis or even rigid spine (Cassandrini et al., 2017).

Table 1.2.1- Genes and phenotypes associated with congenital myopathies.

Subtype (phenotype OMIM)	Gene	Nr. Exons	Chromosomal localization	Protein	Inheritance pattern	Reference (phenotype-gene association)
I. Nemaline myopathy (NEM)						
NEM pathophysiological mechanism I: Contractile dysfunction of the thin filament (Z-disk) due to misassembly or instability of its components.						
NEM1 (#609284)	*TPM3*	13	1q21.2	Tropomyosin α3-chain	AD, AR	Laing et al., 1995
NEM2 (#6256030)	*NEB*	183	2q22	Nebulin	AR	Pelin et al., 1999
NEM3 (#161800)	*ACTA1*	7	1q42.1	Actin, alpha, skeletal muscle	AD, AR	Nowak et al., 1999
NEM5 (#605355)	*TNNT1*	14	19q13.4	Troponin T, slow skeletal muscle	AR	Johnston et al., 2000
NEM4 (#609285)	*TPM2*	9	9p13	Tropomyosin β-chain	AD	Donner et al., 2002
NEM11 (#617336)	*MYPN*	24	10q21.3	Myopalladin	AR	Miyatake et al., 2017
NEM pathophysiological mechanism II: Impaired myofibrillar assembly and maintenance/dynamics, dysregulation of thin filament proteins breakdown (loss or reduced poly-ubiquitination).						
NEM7 (#610687)	*CFL2*	4	14q13.1	Cofilin 2	AR	Agrawal et al., 2007
NEM6 (#609273)	*KBTBD13*	1	15q22.31	Kelch repeat and BTB (POZ) domain containing protein 13	AD	Sambuughin et al., 2010
NEM8 (#615348)	*KLHL40*	6	3p22.1	Kelch-like family member 40	AR	Ravenscroft et al., 2013
NEM9 (#615731)	*KLHL41*	6	2q31.1	Kelch-like family member 41	AR	Gupta et al., 2013
NEM10 (#616165)	*LMOD3*	4	3p14.1	Leiomodin 3	AR	Yuen et al., 2014
NEM?, Klippel-Feil syndrome 4 (#616549)	*MYO18B*	44	22q12.1	Myosin XVIIIB	AR	Malfatti et al., 2015
Other genes involved in NEM: *RYR1*						
II. Centronuclear myopathy (CNM)						
CNM pathophysiological mechanism: Dysregulation of membrane traffic and remodeling, altering the formation and maintenance of the excitation-contraction coupling (ECC) apparatus, through the disturbance the structure of the T-tubule (tubulogenesis) and the terminal sarcoplasmic reticulum.						
Myotubular myopathy, X-linked (#310400)	*MTM1*	15	Xq28	Myotubularin	XLR	Laporte et al., 1996
CNM1 (#160150)	*DNM2*	21	19p13.2	Dynamin 2	AD	Bitoun et al., 2005
CNM2 (#255200)	*BIN1*	19	2q14.3	Bridging integrator 1 (Amphiphysin 2)	AR	Nicot et al., 2007

Table 1.2.1- Genes and phenotypes related to congenital myopathies (continues).

CNM4 (#614408)	*CCDC78*	14	16p13.3	Coiled-Coil Domain-Containing protein 78	AD	Majczencko et al., 2012
CNM5 (#615959)	*SPEG*	41	2q35	SPEG complex locus	AR	Agrawal et al., 2014
CNM4 (#614807)	*MAP3K20*	12	2q31.1	Mitogen-activated protein kinase kinase kinase 20	AR	Vasli et al., 2017
Other genes involved in CNM: *RYR1* and *TTN*.						
III. Central core disease (CCD) / multi(mini)core disease (MmD)						
CCD/ MmD pathophysiological mechanism I: Leaky RYR1 channel activity, channel instability or reduced Ca^{2+} conductance. Changes in the excitation-contraction coupling and/or triadic assembly (loss of the link between DHPR and RYR1).						
CCD (#117000); minicore (255320)	*RYR1*	106	19q13.2	Ryanodine receptor 1	AD, AR	Quane et al., 1993
CCD/ MmD pathophysiological mechanism II: Impairment of the oxidative stress regulation. Increased levels of basal oxidative activity and sensitivity to oxidant exposure in SEPN1-MmD patients. SEPN1 possibly regulates EC coupling indirectly, either by a modulation of the Ca^{2+} reuptake through SERCA channels or through secondary RYR1 changes.						
Myopathy, CFTD (#602771)	*SELENON*	13	1p36.11	Selenoprotein N	AR	Ferreiro et al., 2002
CCD/ MmD pathophysiological mechanism III: Sarcomere structural changes due to defects in Titin. Cardiac and skeletal muscle are both compromised.						
Salih myopathy (#611705), minicore myopathy	*TTN*	364	2q31.2	Titin	AR	Ceyhan-Birsoy et al., 2013
CCD/ MmD pathophysiological mechanism IV: Satellite cell depletion, reduced myofiber diameter and lack of PAX+ nuclei in patients with MEGF10 deficiency.						
EMARDD (#614399)	*MEGF10*	26	5q23.2	Multiple epidermal growth factor-like domains 10	AR	Logan et al., 2011
other genes involved in CCD/MmD: *MYH7.*						
IV. Congenital fiber-type disproportion (CFTD)						
CFTD pathophysiological mechanism: Very heterogeneous, its pathophysiology is mostly shared among the other subtypes of CM.						
Myopathy, myosin storage (#608358)	*MYH7*	40	14q12	Myosin, Heavy chain 7, cardiac muscle, beta	AD	Ortolano et al., 2011
Genes involved in CFTD: *ACTA1, SEPN1, RYR1, TPM2* and *TPM3.*						

Table 1.2.1- Genes and phenotypes related to congenital myopathies (continues).

V. Other congenital myopathies						
Sarcotubular myopathy:						
Phenotype not in OMIM	*TRIM32*	4	9q33.2	Tripartite motif-containing 32	AR	Schoser et al., 2005
Lethal congenital myopathy:						
Compton-North type (#612540)	*CNTN1*	29	12q12	Contactin 1	AR	Compton et al., 2008
Congenital skeletal myopathy and fatal cardiomyopathy:						
Phenotype not in OMIM	*MYBPC3*	35	11p11.2	Cardiac myosin binding protein-C	AD	Tajsharghi et al., 2010
Bailey-Bloch congenital myopathy:						
Bailey-Bloch type (#255995)	*STAC3*	12	12q13.3	SH3 and cysteine-rich domains 3	AR	Horstick et al., 2013
Tubular-aggregate myopathy:						
TAM1 (#160565)	*STIM1*	13	11p15.4	Stromal interaction molecule 1	AD	Böhm et al., 2013a
TAM2 (#615883)	*ORAI1*	2	12q24.31	ORAI calcium release-activated calcium modulator 1	AD	Nesin et al., 2014
Congenital myopathy with ophthalmoplegia related to CACNA1S:						
Phenotype not in OMIM	*CACNA1S*	44	1q32	Calcium channel, voltage-dependent, L type, alpha 1S subunit	AR	Hunter et al., 2015
Congenital myopathy with Moebius and Robin sequence (Carey-Fineman-Ziter syndrome):						
Carey-Fineman-Ziter syndrome (#254940)	*MYMK*	5	9q34.2	Myomaker	AR	Di Gioia et al., 2017
Congenital myopathy with excess of muscle spindles:						
Congenital myopathy with excess of muscle spindles (#218040)	*HRAS*	7	11p15.5	V-Ha-RAS Harvey Rat Sarcoma Viral	AD	Quélin et al., 2017

Footnote: AD, autosomal dominant; AR- autosomal recessive; Nr- number; data used to generate this table was obtained from the Gene Table of Neuromuscular Disorders (Bonne & Rivier, 2018), Online Mendelian Inheritance in Man (OMIM, 2018) database and cited literature.

Cardiomyopathy is also present in some CM subtypes, especially those originated from molecular defects in proteins simultaneously expressed in skeletal and cardiac muscles. In contrast with other congenital muscle diseases, usually there is no CNS involvement in CM, consequently, the cognitive function of these patients is usually fully preserved. These clinical features are helpful clues towards the classification and the diagnostic workup of CM.

Muscle pathology found in CM

The classification of CM is further established based on characteristic histological features (Figure 1.2.1) found in muscle biopsies of these patients.

i) Nemaline myopathy (NEM), characterized by the presence of rod-like structures, called nemaline bodies, in variable number and size that stain dark blue with Gomori trichrome (Romero & Clarke, 2013). Nemaline bodies are formed by the abnormal aggregation of proteins within muscle thin filaments (Ramirez-Martinez et al., 2017).

ii) Centronuclear myopathy (CNM), nuclei are misplaced assuming a central position in the myofiber, their frequency and characteristics vary along the type of muscles and the genetic etiology (Jungbluth et al., 2008).

iii) Core myopathy: central core disease (CCD) and Multi(mini)core disease (MmD), histochemical appearance typically consisting of focally reduced oxidative enzyme activity, which corresponds to myofibrillar changes on ultrastructural examination. In simplistic terms, such areas reflect dimmed mitochondrial activity and might appear in different forms, thus commonly known as central cores or minicores (Schorling et al., 2017). CCD is characterized by larger centrally located well-demarcated cores that run along the fiber axis and visible through longitudinal sections (Jungbluth et al., 2007a). Whereas MmD is defined by the presence of multiple small cores of less well-defined appearance and more-limited length along the fiber axis (Jungbluth et al., 2007b).

iv) There are additional variants of the three main groups detailed above, including mixture types such as those combining central core with nemalinic features (core-rod myopathy). Nonetheless, in most of these cases, type 1 fibers are smaller in size (hypotrophic) and constitute the predominant type of muscle fibers seen (type 1 fiber predominance). When these type 1 hypotrophy and predominance is marked and in the absence of cores, nemaline bodies, or centralized nuclei, this clinicopathologic entity is denominated as congenital fiber-type disproportion (CFTD) and is recognized as the fourth main subtype of CM (Clarke et al., 2010).
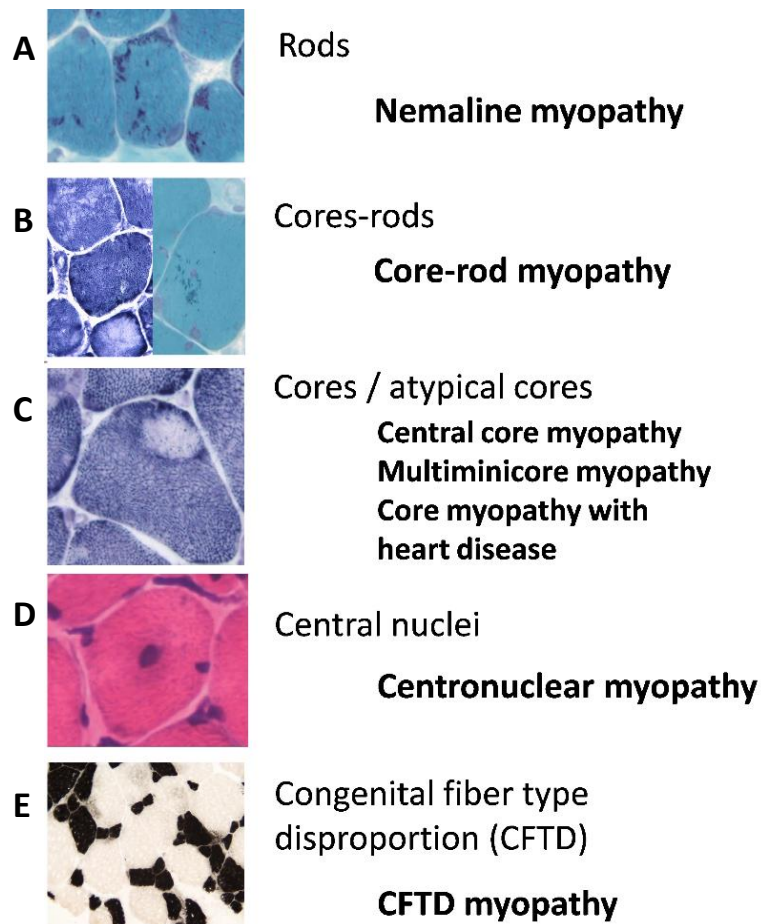
Figure 1.2.1- Major histopathological changes found in muscle biopsies of patients with CM. A) Presence of rod-like (dark blue) structures in Gomori Trichrome Stain, known as nemaline bodies. B) The mixed appearance of muscle biopsy with rods (right) and core-like structures (left) consisting of focally reduced oxidative enzyme activity in NADH-TR staining. C) Isolated large core (NADH-TR staining). D) Centralized nuclei visualized by hematoxylin & eosin staining. E) Congenital fiber-type disproportion (CFTD), type 1 fibers smaller are shown in dark, and type 2 is larger and pale (ATPase pH 4.3 staining). Pictures kindly provided by Dr. Ricardo Taipa and Prof. Melo Pires (Centro Hospitalar Universitário do Porto).

Genetic etiology of CM

The last decade brought an outstanding explosion of gene discovery in CM, now totalizing at least 32 different genetic causes for this disease (Table 1.2.1). This augmented understanding of the genetic basis of CM brought the knowledge that such defects are associated with a wide variety of clinical phenotypes and a broad spectrum of histopathologic findings on muscle biopsy. In addition, genetic defects in different *loci* can share the same histopathologic features which contribute towards an extreme genetic heterogeneity in CM. Moreover, within the list of genes with a higher number of exons in the human genome, three genes are linked to CM (*TTN*, *NEB,* and *RYR1*).

As for the *locus* known to be involved in NEM, the majority encode for proteins that are involved in the assembly and maintenance of the sarcomere, such as the kelch-like family

members or are components of the thin filament of the sarcomere, such as α-actin (*ACTA1* gene) or nebulin (*NEB*). In particular, these two genes are considered the most frequent causes of NEM (up to, respectively, 25% and 50% of genetically established cases) (Schorling et al., 2017). In this subtype, most genetic defects are associated with an AR inheritance, but autosomal dominant (essentially *de novo*) cases are also frequent in *ACTA1*-related cases. Pathogenic variants in *NEB* have been identified across a wide diversity of NEM clinical phenotypes, being the most frequent of congenital type (Lehtokari et al., 2006; Romero & Clarke, 2013). Due to its considerable dimension and rather repetitive structure, the *NEB* gene has placed difficulties in both research and diagnostic testing (Donner et al., 2004). Variants in *ACTA1* gene are found in 25–30% of NEM patients (Ryan et al., 2001), but in higher frequency (nearly half) in patients with severe congenital weakness, typically heterozygous dominant changes and altering single amino acids (reviewed by Romero & Clarke, 2013). In terms of recessive pathogenic variants are predominantly functionally null mutations (such as nonsense or frameshift) or certain missense variants. Somatic and gonadal mosaicism was also reported for *ACTA1* variants (Miyatake et al., 2014; Seidahmed et al., 2016). The remaining genes are extremely rare causes of NEM (Table 1.2.1), some even limited to single reports/families (Jungbluth et al., 2018).

As for CNM, the most frequent genetic cause is undoubtedly related with myotubularin gene (*MTM1*) defects, having a recessive inheritance linked to the X chromosome (XLR) and also known as myotubular myopathy (XLMTM). Its estimated incidence, 1/50,000 male births, makes it a very rare condition (Jungbluth et al., 2008). More than three hundred *MTM1* variants have been reported collectively worldwide and typically distributed throughout the entire gene´s coding sequence and with a small subset of recurrent variants. Previous attempts to establish genotype-phenotype correlations studies have been limited considering the more private nature of many *MTM1* variants. As a tendency, truncating variants usually give rise to a more severe phenotype, while non-truncating variants outside the most relevant myotubularin's domain, such as the catalytical, are reported to originate milder disease presentations (Jungbluth & Gautel, 2014). Over the years there was an increasing number of reports of manifesting females due to skewed X chromosome inactivation (Biancalana et al., 2017). Ryanodine receptor 1 gene (*RYR1*) it is likely the second most frequent cause of CNM (Wilmshurst et al., 2010; Amburgey et al., 2013), being typically AR, it shows some overlapping clinical features with other forms of recessively inherited *RYR1*-related myopathy (Jungbluth et al., 2018). Other autosomal recessive (AR) forms of CNM have been associated with pathogenic variants in amphiphysin (*BIN1*) and titin (*TTN*) genes. Additional AR forms of CNM are linked to SPEG complex locus (*SPEG*) and Mitogen-activated protein kinase kinase kinase 20 gene (*MAP3K20*) with causative variants identified in a small number of families. While AD disease-causing variants were found in dynamin 2 (*DNM2*) and

in the coiled-coil domain-containing protein 78 gene (*CCDC78*), the latter found defective in only a single case report (Majczenko et al., 2018).

Core myopathy is one of the more heterogeneous subtypes of CM, only in terms of pathophysiology but also in terms of the genetics' etiology. Here, a defective *RYR1* is one of the most frequent causes of core myopathy, namely associated with CCD and MmD phenotypes. CCD having a typical AD inheritance was one of the first genetically established cause of CM (Quane et al., 1993; Zhang et al., 1993). Clinically, dominantly inherited *RYR1*-related myopathy is commonly associated with early onset, presenting as neonatal hypotonia and/or delay in motor acquisition (Sewry & Wallgren-Pettersson, 2017). There is considerable clinical variability, from very severe cases with prenatal onset to milder forms only diagnosed during adulthood. Typically, weakness is often axial and can involve facial muscles, further complications such as hip dislocation and joint laxity are common in *RYR1*-related core myopathy (Sewry & Wallgren-Pettersson, 2017). This predominant axial involvement was identified as an important clinical clue towards CCD (Rocha et al., 2014). However, other features are non-specific such as joint laxity, showing some degree of clinical overlap with other neuromuscular disorders, namely collagen VI-related myopathies (Sewry & Wallgren-Pettersson, 2017). In some cases, CCD phenotype is found in conjunction with malignant hyperthermia susceptibility (MHS) but it can also present as a separate clinical entity, where patients do not necessarily display overt muscle weakness. Considering MmD, patients have variable clinical features, and the majority harbor recessive variants in *RYR1* and selenoprotein N (*SELENON*) gene. *RYR1*- and *SELENON*-related MmD are characterized by marked weakness, early spinal rigidity, scoliosis and respiratory impairment (Cassandrini et al., 2017). Thus, there is some degree of similarity between these two recessively inherited MmD, but in the *RYR1*-related form, extraocular muscle involvement (ophthalmoplegia or ophthalmoparesis) is a very common feature and, with only limited exceptions, do not acquire severe respiratory problems (Jungbluth et al., 2018). Defects in myosin heavy chain 7 (*MYH7*, AD transmission) and multiple epidermal growth factor-like domains 10 (*MEGF10*, with AR inheritance) genes are less frequently implicated in MmD. Within the spectrum of titinopathy (*TTN*-related clinical entity), patients can display muscle biopsies with cores, core-like structures, or minicores along with other dystrophic features (Jungbluth et al., 2018). The presence of scoliosis, rigid spine, mainly distal contractures, and cardiomyopathy can occur in this *TTN*-related form (Chaveau et al., 2014a). *MYH7*-related CM presents in childhood with proximal or distal muscle weakness slowly progressive and a variable degree of cardiorespiratory impairment with onset later in life (Cullup et al., 2012). Myopathies related with *MEGF10* deficiency, display a broad clinical spectrum, ranging from a congenital myopathy characterized by proximal and generalized muscle weakness, respiratory difficulties, joint contractures, and scoliosis (designated early-onset

myopathy with areflexia, respiratory distress and dysphagia, EMARDD), to later-onset cases (from childhood to adulthood) with minicores in muscle biopsies (Takayama et al., 2016).

As for CFTD, six genes have been associated with this presentation, namely *TPM3*, *RYR1*, *ACTA1, TMP2, SELENON* and *MYH7* (DeChene et al., 2013). All have been implicated in other histological presentations of CM. The most frequently implicated *locus* in CFTD is *TPM3* gene which codes for α-tropomyosin slow, one of the isoforms of tropomyosin expressed in skeletal muscle (but also in cardiac) (Clarke, 2011a). Although pathogenic variants in *TPM3* were firstly reported as causing NEM, it appears to be predominantly associated with CFTD instead. Heterozygous missense variants are found in most cases, giving rise to mild to moderate weakness keeping independent ambulation into adulthood (Clarke, 2011a). Additional clinical features include the development of scoliosis, prominent involvement of axial and ankle dorsiflexor muscles and mild ptosis. Respiratory involvement, initially nocturnal hypoventilation, is also present on older or more severe cases (Clarke, 2011a). Further to the clinical features already detailed for *RYR1*-related myopathy, CFTD cases present a broad clinical phenotype and disease severity, varying from the neonatal period with respiratory distress to less overt difficulties climbing stairs in teenagers. Still, there is prominent hypotonia and axial muscles weakness, myopathic facial appearance and respiratory involvement (Clarke, 2011a). Ophthalmoparesis is also an important diagnostic clue towards *RYR1*-related cases since it is not observed in the other causes of CFTD with different genetic etiology. The remain four *loci*: *SELENON*, *ACTA1*, *TPM2,* and *MYH7*, have been associated with a very limited number of CFTD cases reported so far and may be part of a large spectrum. The overall data, together with the inexistence of a specific genetic marker and distinctive clinical symptoms in CFTD, suggest it probably represents a spectrum of a disease process and is a diagnosis of exclusion (Kissiedu & Prayson, 2016).

Besides the genes known to be causative of CM, classified in the four major groups already outlined before, other nine *loci* were reported: *CACNA1S*, *CNTN1*, *HRAS*, *MYMK*, *MYBPC3*, *ORAI1*, *STAC3*, *STIM1,* and *TRIM32*. From these, an emerging group, with several new cases of *STIM1*-related and *ORAI1*-related, has been reported caused by heterozygous variants giving rise to tubular aggregate myopathy (TAM) (Böhm et al., 2013a). It is a slowly progressive myopathy with varying degrees of extraocular muscle weakness, exertional myalgia and calf hypertrophy. The TAM phenotype relates with defective calcium homeostasis since STIM1 is a reticular calcium sensor, and ORAI1 a plasma membrane channel for calcium (Böhm et al., 2013a). Strikingly, these molecular defects also give rise to Stormorken syndrome, a broader phenotype involving not only muscle weakness but also variable features such as miosis, thrombocytopenia, hyposplenism, ichthyosis, dyslexia, and short stature (Böhm & Laporte, 2018).

Main pathophysiological mechanisms in CM

Currently, there is a broad knowledge concerning the pathophysiology of CM, mostly derived from animal models, genetic and pathophysiological studies in human patients.

As for NEM two main mechanisms have been identified. Since the majority of the molecular defects comprise components or modifiers of the actin thin filament, the disease is mainly caused by a dysfunction of thin filaments (Gonorazky et al., 2018). The muscle contractibility is compromised as the actin filaments are either not properly formed and/or its interaction with the myosin thick filaments is disrupted (Ravenscroft et al., 2015). More recently, an additional alternative mechanism in NEM was unveiled, related to molecular defects in proteins that are not components of the thin filament machinery. This list comprises six different proteins, where three: *KLHL40*, *KLHL41*, and *KBTBD13*, share Kelch motifs. One possible hypothesis is that these proteins are involved in myofibrillar assembly and maintenance/dynamics, dysregulation of thin filament proteins breakdown (loss or reduced poly-ubiquitination) (Gupta & Beggs, 2014). This hypothesis was raised from the recent observations that: i) KLHL40 can bind to LMOD3 and nebulin directly, promoting the stability of LMOD3 by blocking its degradation (Garg et al., 2014; Gonorazky et al., 2018); ii) KLHL41 stabilizes the sarcomere and maintains muscle function by acting as a molecular chaperone (Ramirez-Martinez et al., 2017). A similar mechanism for protein stabilization was proposed to be likely involved in the other Kelch-containing proteins as well (Ramirez-Martinez et al., 2017).

Concerning the CNM pathophysiological mechanisms, they remained elusive for several years, given the identification of very different genetic defects leading to a somewhat similar histopathological presentation. Dowling and collaborators (2014) suggested that the presence of triad abnormalities is the overarching explanation for CNM pathophysiology. Although the mechanism underlying each individual gene/protein is distinct, they all converge in the presence of altered ECC and structural changes in the triad (Figure 1.1.3). These changes were identified in different CNM genetic entities both in animal models and from patients' muscle biopsies (Dowling et al., 2009; Durieux et al., 2010; Gibbs et al., 2014; Smith et al., 2014). These alterations in the triad can be attributed to different molecular events. Myotubularin is a lipidic phosphatase that mainly regulates the levels of the phosphoinositides PI3P and PI3,5P2 (Dowling et al., 2014). The link between the lack of phosphatase activity and the changes in the triad structure is not very clear, despite myotubularin being localized in the triad. Regarding BIN1, it was characterized even before its association with CNM as a modulator of T-tubule biogenesis, and BAR (Bin1 Amphiphysin RVS167) domain-containing protein regulates and modifies membrane curvature (Jungbluth et al., 2008). Dynamin-2, a GTPase playing a role in membrane scission events, it is known to be involved in the trafficking of triad proteins possibly by its interaction with BIN1 (Marty &

Fauré, 2016). Recent findings showed that this actin-dependent trafficking impairment also contributes to the pathological mechanism in dynamin-2-associated CNM (González-Jamett et al., 2017). Ultimately, all gathered data suggests that *DNM2* disease-causing variants might affect the triad's structure. The identification of pathogenic variants in *RYR1* as causing CNM provided further consistent evidence towards the connection between triad abnormalities and this disease (Dowling et al., 2014). These phenotypes associated with RYR1, a protein exclusively functioning and localized within the triad, suggest that the disturbance of this structure may be enough to generate central nucleation (Dowling et al., 2014). Until now there is no link between titin and the triad besides the commonly shared CNM histopathology. Taking into account the massive size of this protein and the multiple protein-protein interactions with other molecular partners, one feasible explanation is that CNM-related pathogenic variants in *TTN* would interfere with or alter other proteins that more directly regulate the triad (Dowling et al., 2014). More recently, it has been proposed that the pathogenicity of titin is multidimensional, since it is likely to involve several mechanisms, namely those implicated in ECC, RYR1 recruitment to the triad mediated by calpain 3, and obscurin-mediated interactions between the T-tubules, the SR and the sarcomere (Jungbluth et al., 2018).

The pathophysiology of CCD/MmD is diverse; four main mechanisms are highlighted in Table 1.2.1. One involved is related to RYR1 defects. As depicted, this channel as a pivotal role in calcium release (homeostasis) during the process of ECC, thus this constitutes the primary pathomechanism (Gonorazky et al., 2018). Functional studies, using a variety of cellular and animal models, have shown the effects of *RYR1* pathogenic variants (reviewed by Marty & Fauré, 2016; Jungbluth et al., 2018). Apart from the malignant hyperthermia-associated RYR1 variants that lead to excessive calcium release (hypersensitivity channel) and lower RYR1 activation thresholds, the CCD dominantly inherited variants cause completely different effects. Namely the depletion of calcium in the SR that contributes to increased levels in the cytosol (the so-called 'leaky channel' hypothesis) and disturbed ECC mechanism (the 'excitation-contraction uncoupling' hypothesis) (Jungbluth et al., 2018). This is explainable by the modified RYR1–DHPR interaction, reduced RYR1 expression and/or stability, and impaired calcium release due to mutations in the channel pore (Gonorazky et al., 2018). In contrast, recessive variants may cause mislocalized RYR1 channels or its reduced expression, yet the pathogenesis of recessive *RYR1* mutations has been poorly characterized (Zhou et al., 2013; Ravenscroft et al., 2014). The presence of a hypomorphic (disrupted) *RYR1* allele is not enough, by itself, to induce muscle weakness, or a damaging reduction in the amount of protein, as demonstrated in ryr1 knockout mouse model (Cacheux et al., 2015; Marty & Fauré, 2016). Nonetheless, such allele it is found in association with a second point mutation in recessive forms of *RYR1*-related CM (Marty & Fauré, 2016). The

tetrameric organization of RYR1 and its interaction with the DHPR, adds further intricacy towards the characterization of the variant's effect (Marty & Fauré, 2016). The marked reduction of the amount of the RYR1 protein found in some patients has led to the assumption that MmD-related *RYR1* variants might have an effect on protein expression and stability as much as its function (Jungbluth, 2007a). Pathways other than ECC have been associated with loss of RYR1 function, namely aberrant oxidative stress suggesting that its regulation could be mediated by calcium homeostasis (Dowling et al., 2012; Gonorazky et al., 2018). In fact, the second mechanism in MmD involves the impairment of the oxidative stress regulation, primarily due to molecular defects in selenoprotein N (Jungbluth et al., 2011). In *SELENON*-MmD patients, the basal levels of oxidative activity are increased leading to sensitivity towards oxidative exposure. Furthermore, there is at least another process that could be implicated with selenoprotein N. It relates with the possibility of ECC modulation (Arbogast et al., 2009), either by modulation of calcium reuptake through SERCA channels or through a secondary change in RYR1 (Gonorazky et al., 2008). Since this protein has been specifically implicated in redox regulation during myogenesis, the main question to unfold is why the disease only manifests postnatally. This role is also shared by the protein encoded by *MEGF10* gene, that gives rise to a rare form of MmD when mutated (Jungbluth et al., 2018). In patients with MEGF10 deficiency, another pathophysiological mechanism was uncovered. MEGF10 protein is primarily expressed in the satellite cells of the skeletal muscle, a mononucleated cell type positioned between the plasma and the basement membranes (Takayama et al., 2016). Upon muscle injury, these satellite cells proliferate and differentiate to enable the regeneration of new muscle fibers (Takayama et al., 2016). This process is mediated through the inhibition of the expression of the myogenic factors MyoD, myogenin and myosin heavy chain (Takayama et al., 2016). Thus, in patients with MEGF10 defects displaying a MmD histopathology, there is poor muscle regeneration due to satellite cells depletion. The last mechanism within CCD/ MmD highlighted here, relates with alterations within the sarcomere itself due to titin mutations. Pathogenic variants in genes encoding sarcomeric proteins in skeletal muscle lead to a variety of pathological changes. It should be emphasized that pathogenic variants in genes of the thick filament, such as *MYH7*, are a cause of cores (non-functioning compacted sarcomeres) (Hwang & Sykes, 2015). The link between titin and cores is probably related to the regulation of thick filament length and stability which depends upon titin and it is critical for maintaining muscle health (Tonino et al., 2017). In the case of minicore *TTN*-related disease, the sarcomeres are disorganized and out of proper order, presenting as internal nuclei and/or cores in muscle pathology (Ceyhan-Birsoy et al., 2013; Chauveau et al., 2014).

1.2.2 CONGENITAL MUSCULAR DYSTROPHIES

The second group of congenital muscle diseases comprises the congenital muscular dystrophies (CMD), collectively gathered based on the presence of a distinctive pathological signature found in the patients' muscle biopsies (Mercuri & Muntoni, 2013). From the early times of clinical research, the definition of muscular dystrophy (MD), was established to describe a genetic myopathy with a range of muscle histologic changes different as those presented for CM (Shieh et al., 2013). Typically, MD pathology includes necrotic and regenerating fibers, myofiber size variation, and abnormal internalization of muscle nuclei (Figure 1.2.2).
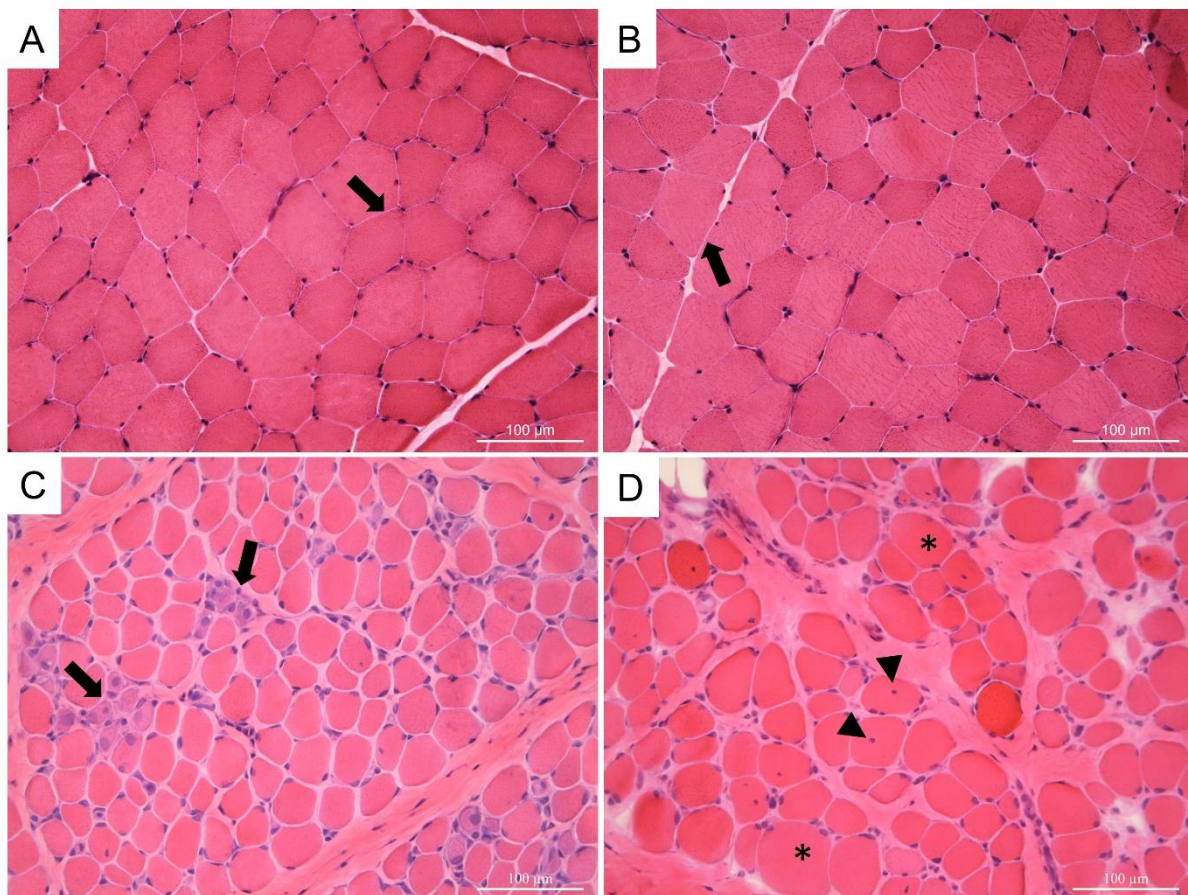
Figure 1.2.2- Examples of morphologic patterns in muscle biopsies (hematoxylin & eosin staining). (A & B) Normal muscle histology, with normal (thin) endomysial (arrow in A) and perimysial (arrow in B). C & D) Muscle biopsies from Duchenne muscular dystrophy patients, depicting characteristic dystrophic features: increased fiber size variation; myocyte hypertrophy (asterisks); endomysial and perimysial fibrosis; frequent central nuclei (arrowheads); necrosis, myophagocytosis, and regeneration foci (arrows). Pictures kindly provided by Dr. Ricardo Taipa and Prof. Melo Pires (Centro Hospitalar Universitário do Porto).

Successive steps of degeneration and regeneration of muscle fibers eventually result in necrosis and ultimately replacement of muscle with fatty and fibrosis. Consequently, CMD are a highly heterogeneous group of conditions, and as similarly with CM, have onset just after birth or during the first years of life, and are clinically characterized by hypotonia and muscle weakness (Mercuri & Muntoni, 2013). There is also variable involvement of other organs such as eyes, heart, and CNS. Table 1.2.2 displays the currently known phenotypes and genes involved in this group of diseases.

Table 1.2.2- Genes and phenotypes associated with congenital muscular dystrophies.

Subtype (phenotype OMIM)	Gene	Nr. Exons	Chromosomal localization	Protein	Inheritance pattern	Reference (phenotype-gene association)
I. Congenital muscular dystrophy (CMD) with merosin deficiency (MDC1A)						
MDC1A / LAMA2-related MD pathophysiological mechanism: Defective laminin-2 structure leads to the destabilized sarcolemma, due to the loss of the link between DAPC (glycosylated α-dystroglycan and α7-β1 integrin) and the extracellular matrix. Mitochondrial apoptosis also appears to contribute to disease progression.						
MD, LAMA2 deficiency (#607855)	*LAMA2*	65	6q22.33	Laminin-alpha 2	AR	Helbling-Leclerc et al., 1995
II. Ullrich CMD & Bethlem myopathy (UCMD-BM)						
UCMD-BM pathophysiological mechanism: Loss of connection between the basement membrane and the interstitial matrix, resulting in the reduced anchorage to the extracellular matrix. Dysfunction of mitochondria and sarcoplasmic reticulum, and of defective autophagy was identified in a mouse model for collagen VI.						
UCMD 1 (#158810), BM 1 (#254090)	*COL6A1*	35	21q22.3	Collagen, type VI, alpha-1	AR, AD	Jöbsis et al., 1996
UCMD 1 (#158810), BM 1 (#254090)	*COL6A2*	28	21q22.3	Collagen, type VI, alpha-2	AR, AD	Camacho Vanegas et al., 2001
UCMD 1 (#158810), BM 1 (#254090)	*COL6A3*	44	2q37.3	Collagen, type VI, alpha-3	AR, AD	Speer et al., 1996
UCMD 2 (#616470), BM 2 (#616471)	*COL12A1*	66	6q13-q14	Collagen, type XII, alpha-1	AR, AD	Hicks et al., 2014; Zou et al., 2014
III. Congenital muscle dystrophies - dystroglycanopathies (CMD-DG)						
CMD-DG pathophysiological mechanism I: Loss of the link between dystrophin-glycoprotein complex (DGC) and the basal lamina due to primary dystroglycan changes.						
CMD-DG type A, 9 (#609308)	*DAG1*	6	3p21	Dystroglycan 1	AR	Geis et al., 2013
CMD-DG pathophysiological mechanism II: Loss of the link between DGC and the basal lamina due to α-dystroglycan hypoglycosylation. Abnormal basement membrane formation. Reduced α-dystroglycan binding to laminin, neurexin, and agrin.						
MD-DG 4 type A (#253800) & B (#613152)	*FKTN*	12	9q31.2	Fukutin	AR	Kobayashi et al., 1998

Table 1.2.2- Genes and phenotypes associated with congenital muscular dystrophies (continues).

MD-DG 3 type A (#253280) & B (#613151)	*POMGNT1*	22	1p34.1	O-linked mannose beta1,2-N-acetylglucosaminyltransferase	AR	Yoshida et al., 2001
MD-DG 1 type A (#236670) & B (#613155)	*POMT1*	20	9q34.13	Protein-O-mannosyltransferase 1	AR	Beltrán-Valero de Bernabé et al., 2002
MD-DG 6 type A (#613154) & B (#608840)	*LARGE1*	16	22q12.3	Like-glycosyltransferase	AR	Longman et al., 2003
MD-DG 5 type A (#613153) & B (#606612)	*FKRP*	4	19q13.32	Fukutin-related protein	AR	Beltrán-Valero de Bernabé et al., 2004
MD-DG 2 type A (#613150) & B (#613156)	*POMT2*	21	14q24.3	Protein-O-mannosyltransferase 2	AR	van Reeuwijk et al., 2005
MD-DG 7 type A (#614643)	*ISPD*	10	7p21.2	Isoprenoid synthase domain containing	AR	Willer et al., 2012
MD-DG 8 type A (#614830)	*POMGNT2*	2	3p22.1	protein O-linked mannose N-acetylglucosaminyltransferase 2	AR	Manzini et al., 2012
MD-DG 10 type A (#615041)	*RXYLT1*	6	12q14.2	Ribitol xylosyltransferase 1	AR	Vuillaumier-Barrot et al., 2012
MD-DG 11 type A (#615181)	*B3GALNT2*	12	1q42.3	Beta-1,3-N-acetylgalacto-saminyltransferase 2	AR	Stevens et al., 2013
MD-DG 12 type A (#615249)	*POMK*	5	8p11.21	Protein-O-mannose kinase	AR	Jae et al., 2013
MD-DG 13 type A (#615287)	*B4GAT1*	2	11q13.2	Beta-1,4-Glucuronyltransferase 1	AR	Buysse et al., 2013
Phenotype not in OMIM	*TRAPPC11*	31	4q35.1	Trafficking protein particle complex 11	AR	Liang et al., 2015
CMD-DG pathophysiological mechanism III: Dolichol-P-Mannose pathway affected leading to combined O-mannosylation and N-glycosylation defects, overlapping with congenital disorders of glycosylation (CDG).						
CDG, type Im (#610768)	*DOLK*	1	9q34.11	Dolichol kinase	AR	Lefeber et al., 2011
CDG, type Iu (#615042)	*DPM2*	4	9q34.13	Dolichyl-phosphate mannosyltransferase 2, catalytic subunit	AR	Barone et al., 2012
MD-DG 14 type A (#615350), type B (#615351)	*GMPPB*	8	3p21.31	GDP-mannose pyrophosphorylase B	AR	Carss et al., 2013
CDG, type Ie (#608799)	*DPM1*	10	20q13.13	Dolichyl-phosphate mannosyltransferase 1, catalytic subunit	AR	Yang et al., 2013

Table 1.2.2- Genes and phenotypes associated with congenital muscular dystrophies (continues).

IV. Other congenital muscular dystrophies						
Congenital muscular dystrophy due to integrin defects:						
CMD due to ITGA7 deficiency (#613204)	*ITGA7*	25	12q13	Integrin alpha 7 precursor	AR	Hayashi et al., 1998
Rigid spine syndrome:						
MD, rigid spine, 1 (#602771)	*SELENON*	13	1p36.11	Selenoprotein N1	AR	Moghadaszadeh et al., 2001
Rigid spine syndrome due to FHL1 defects:						
Myopathy, X-linked, postural atrophy (#300696)	*FHL1*	8	Xq26.3	Four and a half LIM domain 1	XLR	Shalaby et al., 2008
Congenital muscular dystrophy due to *LMNA* defects:						
CMD (#613205)	*LMNA*	13	1q22	Lamin A/C	AD	Quijano-Roy et al., 2008
Congenital muscular dystrophy due to dynamin 2 defects:						
Phenotype not in OMIM	*DNM2*	21	19p13.2	Dynamin 2	AD	Bitoun et al., 2009
Congenital muscular dystrophy due to telethonin defects:						
Phenotype not in OMIM	*TCAP*	2	17q12	Telethonin	AR	Ferreiro et al., 2011
Congenital muscular dystrophy with mitochondrial structural abnormalities (megaconial type):						
CMD megaconial type (#602541)	*CHKB*	11	22q13	Choline kinase beta	AR	Mitsuhashi et al., 2011b
Congenital muscular dystrophy with rigid spine due to ACTA1 defects:						
Phenotype not in OMIM	*ACTA1*	7	1q42.13	Alpha actin, skeletal muscle	AR	O'Grady et al., 2015
Congenital muscular dystrophy, Davignon-Chauveau type:						
CMD, Davignon-Chauveau type (#617066)	*TRIP4*	13	15q22.31	Thyroid hormone receptor interactor 4	AR	Davignon et al., 2016
Congenital muscular dystrophy with cataracts and intellectual disability (ID):						
CMD, with cataracts and ID (#617404)	*INPP5K*	14	17p13.3	Inositol Polyphosphate-5-Phosphatase K	AR	Osborn et al., 2017; Wiessner et al., 2017

Footnote: AD, autosomal dominant; AR- autosomal recessive; Nr- number; XLR- recessive linked to X chromosome; data used to generate this table was obtained from the Gene Table of Neuromuscular Disorders (Bonne & Rivier, 2018), OMIM (2018) and cited literature.

CMD due to laminin-α2 defects

The main subtype of CMD in Europe, congenital muscular dystrophy type 1A (MDC1A), is caused by defects in laminin-α2 one of the three subunits comprising the laminin-211. This laminin (Figure 1.2.3) is an extracellular protein that establishes a link between the ECM and the dystrophin-associated protein complex (DAPC), through the glycosylated α-dystroglycan and with the α7β1 integrin (Schorling et al., 2017). The genetic cause of MDC1A is related to pathogenic variants in the *LAMA2* gene that encodes for laminin-α2, identified more than two decades ago (Helbling-Leclerc et al., 1995). The classical phenotype of MDC1A is typically a severe form of CMD, with onset during the newborn period or within the first six months of life. Its initial presentation consists of marked hypotonia and muscle weakness, with facial muscle involvement and serum CK is typically elevated (Schorling et al., 2017). As the disease progresses, motor development milestones are not properly attained, as MDC1A patients are often unable to stand or even to sit without support. Patients may also present distal contractures that may bear a resemblance to arthrogryposis (Schorling et al., 2017). Feeding difficulties and respiratory distress may also be present secondary due to muscle weakness. Although the cognitive development is not usually affected, nearly all MDC1A patients show white matter changes (WMC) identifiable, beyond the first 6 months of life, by brain MRI (Mercuri et al., 2001). A very residual percentage of patients showed brain structural defects, intellectual disability and/or seizures (Jones et al., 2001). The amount of this protein being expressed can be routinely evaluated in muscle or skin biopsies of patients by IHC using different antibodies, or by performing western-blot in muscle protein extracts. The classical MDC1A phenotype correlates with complete laminin-α2 deficiency, whereas, milder phenotypes are associated with a partial (incomplete) reduction of the amount of protein. Having later disease onset and less severe muscular weakness, the pattern of affected muscles in these milder cases could suggest a limb-girdle MD (LGMD) (Schorling et al., 2017). The characterization of MDC1A has been greatly expanded with the increasing number of patients reported with *LAMA2* pathogenic variants (Oliveira et al., 2008). As for the main pathophysiological mechanism in MDC1A, it is consensual that the link between the ECM and the DAPC is lost, leading to a destabilized sarcolemma and ultimately to a dystrophic process in muscle.

Collagen muscle disorders: Bethlem myopathy and Ullrich CMD

Collagen VI is a major component of the muscle ECM where it interacts with a wide range of molecules, namely other collagens, proteoglycans, matrilins, fibronectin, von Willebrand A-domain related protein and microfibril-associated glycoprotein 1 (Lamandé & Bateman, 2017); it is intimately related to the basement membrane of several tissues, namely skin, kidney, nerves, blood vessels and especially skeletal muscle (Lamandé & Bateman, 2017).

Thus, the disturbance of cell-matrix interactions is believed to be an important feature of these disorders (Mercuri & Muntoni, 2013). The main collagen VI isoforms are composed of three distinct chains (α1, α2, and α3) encoded by different genes *COL6A1* and *COL6A2* (*loci* in 21q22.3) and *COL6A3* (2q37). Pathogenic dominant or recessive variants in each of these genes can cause two types of muscle disorders: Ullrich congenital muscular dystrophy (UCMD) towards the more severe end of the phenotype, and a mild to moderate phenotype Bethlem myopathy (BM) (Bushby et al., 2014). More recently, pathogenic variants in the *COL12A1* gene (*locus* at 6q13-q14), a member of the fibril-associated collagens with interrupted triple helices group and encoding the alpha 1 chain of collagen XII, was associated with a phenotype resembling the classical BM (Hicks et al., 2014; Zhou et al., 2014).
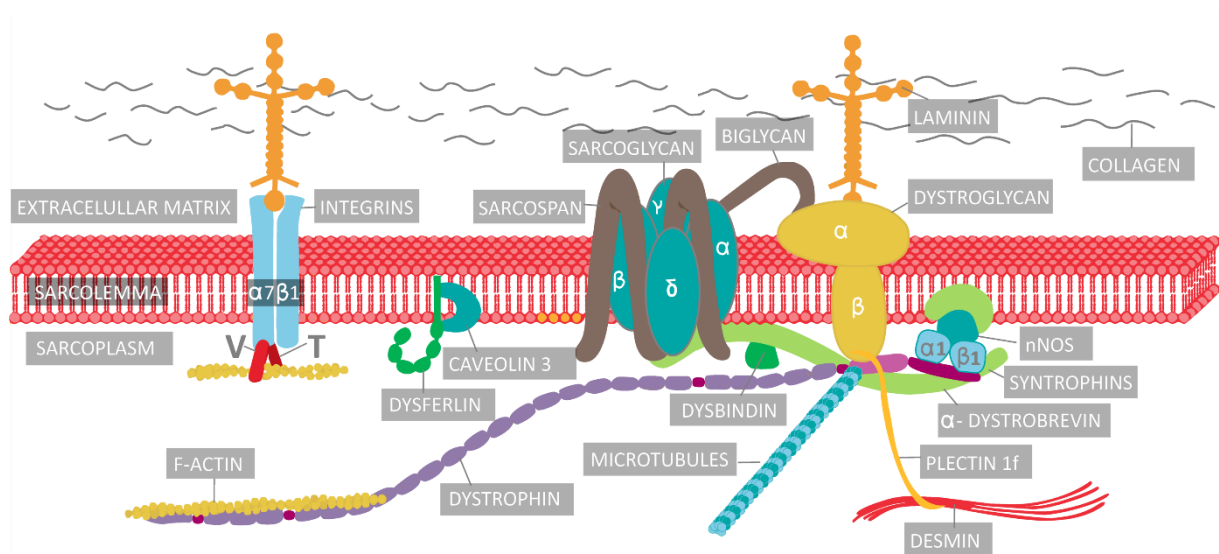
Figure 1.2.3- Schematic representation of the sarcolemma and its major components, highlighting the connection from the inner cytoskeletal proteins to the extracellular matrix. Collagen fibers are bundled with sarcolemma proteins, mainly the laminin-211 (α2β1γ1), which in its turn is either connected to the glycosylated branches of α-dystroglycan or integrin α7β1 or. Both dystroglycan and sarcoglycans act as a hub, connecting a variety of proteins composing the dystrophin-glycoprotein complex (DGC). Besides syntrophin, α-dystrobrevin, and sarcospan, one central element of such complex is dystrophin. The muscle isoform of dystrophin consists of an N-terminal actin-binding domain, a central rod-like domain with 24 spectrin-like triple helical coiled coils, and a cysteine-rich C-terminus that allows the assembly to the sarcolemma proteins. Dystrophin also binds F-actin filaments through its N-terminal domain. Dystrophin as a central role in assembling and maintaining the link between cytoskeletal actin and the extracellular matrix. Adapted from Rahimov & Kunkel, 2013.

UCMD typically presents during the neonatal period as with hypotonia and muscle weakness, failure to thrive, congenital hip dislocation and contractures including torticollis, kyphoscoliosis, and proximal joints (Fu & Xiong, 2017). Extreme distal joint laxity in

conjunction with contractures and skeletal deformities were identified in the proximal joints in nearly half of these patients (Bönnemann, 2011). An important diagnostic feature of UCMD are a range of skin abnormalities that include: soft velvet skin on the soles and palms, hyperkeratosis pilaris on the extensor limb surfaces and abnormal development of hypertrophic scar and keloid formation (Bushby et al., 2011). There is also a variable delay in motor development; some affected children will acquire independent ambulation, while others never walk. Further, the majority UCMD patients that achieve ambulation will lose this ability later during the first or early second decade of life due to increased muscle weakness and contractures (Fu & Xiong, 2017). Intelligence and cardiac function are normally spared in this condition. In contrast, respiratory insufficiency can occur, in the majority of patients, during the first or second decade of life, with diaphragmatic weakness sometimes being out-of-proportion to the skeletal muscle weakness, requiring non-invasive ventilatory support (Bushby et al., 2011). Nearly, all UCMD patients will develop a rigid spine, often associated with scoliosis, and a selective pattern of muscle involvement identifiable by MRI (Mercuri et al., 2010; Mercuri & Muntoni, 2013). In terms of CK levels is often normal or only mildly elevated. Muscle biopsy can show a variable pathology ranging from unspecific myopathic changes to dystrophic features with abundant adipose tissue deposition (Lamandé & Bateman, 2017). In the case of UCMD muscle biopsy frequently display dystrophic features (Bushby et al., 2011). A more specific pathological change in UCMD is the marked deficiency or absence of collagen VI, but this deficiency is often subtle and affecting more specifically the expression in the basal lamina but not its interstitial deposition (Lamandé & Bateman, 2017).

In contrast BM, the first phenotype reported associated with collagen VI (Jöbsis et al., 1996), is a mild progressive disorder characterized by proximal muscle weakness and flexion contractures that mainly affecting the fingers, wrists, elbows and/or ankles (Bushby et al., 2011). Early onset contractures found in BM tend to resolve spontaneously (Bönnemann et al., 2011). BM can also present congenital symptoms but has typically onset during childhood, and in some cases, later onset can even occur during adulthood (Fu & Xiong, 2017). The normal life expectancy is nearly normal, but these patients can require aids for ambulation later in life as the disease progresses (Lamandé & Bateman, 2017). Respiratory involvement is also less pronounced in the BM, although weakness may progress in adulthood (Fu & Xiong, 2017). As newer contractures develop, usually by the end of the first decade of life, these are often stable or might progress creating a degree of disability, such as hand restricted function due to the flexion contractures (Bönnemann et al., 2011). As compared with UCMD, rigid spine is also less pronounced in BM. Concerning, muscle weakness is slowly progressive but more evident beyond the third and fourth decades, making about two-thirds of patients over the age of 60 years dependent of ambulation

support (Bönnemann et al., 2011). Between the two extreme ends of the collagen VI-related disease spectrum, are a range of intermediate phenotypes, being more severe than classical BM but milder as compared with classical UCMD. These patients present with significant muscle weakness in early childhood and mixture of the other two phenotypes, namely distal laxity of the interphalangeal joints found in UCMD, while at the same time showing early Bethlem-like contractures of the long finger flexors like in BM (Bönnemann et al., 2011). As previously mentioned, these disorders are not genetically distinct since they are either caused by dominant or recessive pathogenic variants in collagen VI genes. The phenotypical outcome is much influenced by the way these variants affect proteins. Genetic analysis of the three collagen VI genes detected pathogenic variants in a range of 56-79% of cases, depending on the phenotype, being the highest value within the group of patients with UCMD (Bushby et al., 2011). The proportion of collagen VI-related cases attributed to defects either in *COL6A1* and *COL6A2* genes are quite equivalent around 40%. Whereas, pathogenic variants in *COL6A3* are less much frequent found in about 20% of cases (Bushby et al., 2011). There are several reports of intra-familial phenotypic variability which could consist of the influence of other genes presumably playing a role in disease modulation (Lamandé & Bateman, 2017).

Dystroglycanopathies (CMD-related)

Dystroglycanopathies are a group of diseases where, one of the main functions of dystroglycan, the connection to laminin-α2 is greatly compromised. Within this group, three different pathophysiological mechanisms were identified (Table 1.2.2). The first comprises defects in dystroglycan itself, having an essential role of anchoring diverse extracellular proteins to the cytoskeleton (Geis et al., 2013). It is encoded by the *DAG1* gene and cleaved by post-translational modifications into two subunits: α-dystroglycan a highly glycosylated extracellular, being associated with the transmembrane β-dystroglycan (Muntoni et al., 2011). The α-dystroglycan is subject to O-glycosylation originating several oligosaccharides branches that are crucial for its normal binding function to the ECM (Muntoni et al., 2011). Defects in this glycosylation process constitute the main cause of dystroglycanopathy, being associated with pathogenic variants in genes encoding for enzymes (Figure 1.2.4) involved in the O-glycosylation pathway (Sparks, 2012). More recently, dolichol-P-Mannose biosynthesis (Figure 1.2.4), primarily associated with the congenital disorders of glycosylation, was also been implicated in the group of these diseases as it leads to combined O-mannosylation and N-glycosylation defects (Barone et al., 2012). In CMD patients with defective O-glycosylation of α-dystroglycan, further to skeletal muscle involvement, there is also CNS and ocular involvement in the more severe phenotypes. The main explanation is that α-dystroglycan is also essential for neuronal migration, axon guidance, neuromuscular junction formation in the

nervous system and in the optic tissue (Sparks et al., 2012). The clinical variability within dystroglycanopathies is broad, involving severe forms of CMD namely Walker-Warburg syndrome (WWS), muscle-eye-brain disease (MEB), and Fukuyama CMD (FCMD), and non-syndromic CMD variants, as well as numerous subtypes of LGMD. The heterogeneity surrounding dystroglycanopathies includes both the genetic and biochemical aspects.
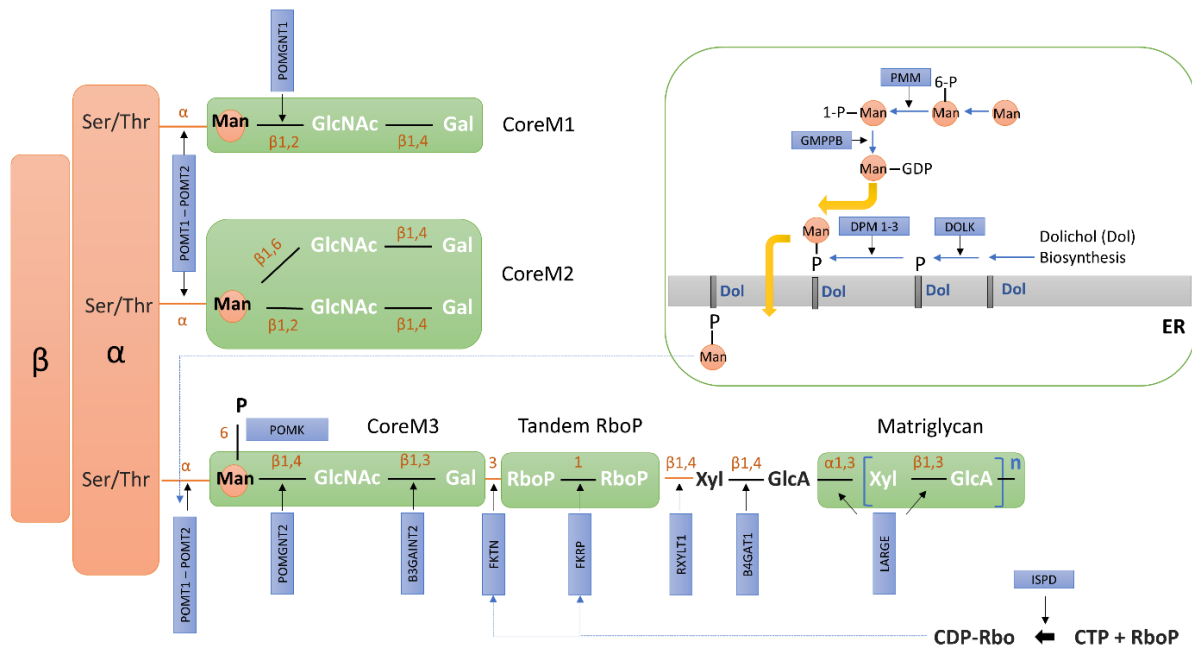
Figure 1.2.4- Overview of the sugar chain structures in α-dystroglycan and their respective o-glycosylation pathway. Adapted from Kanagawa & Toda, 2017. α-dystroglycanopathies genes products are indicated in blue boxes. Box in top-right corner highlights the dolichol and mannose metabolic pathway leading to Dol-P-Man, the substrate of mannosyltransferases. ER- endoplasmic reticulum; RboP- ribitol 5-phosphate; GlcA- glucuronic acid; Xyl- xylose; GalNAc- N-acetylgalactosamine; GlcNAc- N-acetylglucosamine; Man- mannose; Dol-P-Man- dicholphosphate mannose. Adapted from Bouchet-Séraphin et al., 2015.

WWS is within the most severe end of the dystroglycanopathy spectrum. Disease onset usually occurs during early fetal development presenting with severe brain structural abnormalities. These changes are detectable prenatally by ultrasound and are collectively known as cobblestone complex, a combination of CNS that include type II lissencephaly, agenesis of the corpus callosum, cerebellar hypoplasia, and enlarged ventricles (Taniguchi-Ikeda et al., 2016). Thus, the diagnostic hallmark in WWS is the combination of severe brain malformations, muscular dystrophy, and structural eye abnormalities, particularly microphthalmia and can include cataracts (Sparks, 2012). As there are extreme muscle weakness and CNS defects, these patients fail to attain any motor milestone and life expectancy is typically less than 3 years of age, mainly due to complications such as

pneumonia or heart failure (Sparks, 2012; Taniguchi-Ikeda et al., 2016). Recessive pathogenic variants in one of the following genes can give rise to WWS: *POMT1*, *POMT2*, *FKTN*, *FKRP*, *LARGE*, *ISPD*, *B3GALNT2*, *B4GAT1*, *POMK*, *DAG1*, *POMGNT2* and *RXYLT1* (Taniguchi-Ikeda et al., 2016).

Positioned as the second severest phenotype within dystroglycanopathy, MEB is mainly characterized by a muscular dystrophy with a distinctive predominant ocular involvement (retinal defects, optic atrophy, glaucoma, juvenile cataracts, congenital myopia and nystagmus) and structural brain defects extending from microcephaly to changes within the Cobblestone complex, usually milder than those found in WWS (Sparks, 2012; Taniguchi-Ikeda et al., 2016). In terms of genetics, at least five genes (*POMGNT1*, *POMT1*, *POMT2*, *FKTN,* and *GMPPB*) have been implicated in MEB.

Finally, FCMD caused by pathogenic variants in the *FKTN* gene was identified initially in Japan, where it represents the most frequent CMD phenotype (Toda et al., 2005). Variants in this *locus* have been progressively identified worldwide (Yis et al., 2011; Costa et al., 2013). In a similar way as in WWS or MEB, FCMD is characterized by severe muscle and brain involvement, however, ocular features (such as myopia and retinal dysplasia) are less frequent (Toda et al., 2005). Structural brain changes including cobblestone lissencephaly, cerebellar cysts, and hypomyelination, found in such patients causing seizures and moderate to severe intellectual disability (Taniguchi-Ikeda et al., 2016). Further features include a myopathic facies and progressive cardiac disease in patients older than age ten years, the vast majority of patients are unable to walk, and they usually die by the age of 20 years (Saito, 2012).

Due to this overlying distribution of genotypes and phenotypes, the categorization of these diseases has become challenging over the years. Thus, OMIM adopted a more simplified classification of MD-dystroglycanopathy (MDDG) as proposed by Godfrey and collaborators (Godfrey et al., 2011). This includes three wide phenotypic groups: i) type A comprises the considered classical phenotype, within the more severe end, with eye and brain anomalies of congenital onset such as WWS, WWS-like, MEB and FCMD-like; ii) type B encompasses CMD phenotypes without optic symptoms but with cerebellar involvement, with or without intellectual disability; and iii) type C being the mildest, including LGMD phenotypes (Table 1.2.3, section 1.2.3) with or without intellectual disability (Godfrey et al., 2011). Within this classification, a number is added to differentiate the *locus*. Three genes listed in Table 1.2.2 (*DOLK*, *DPM2,* and *DPM1*) do not follow this classification and were initially associated with the congenital disorders of glycosylation as they encode for enzymes involved in different steps of the dolichol-P-Mannose biosynthesis pathway (Figure 1.2.4).

Main pathophysiological mechanisms in CMD

As delineated in this section, the defects in collagens (VI and XII), laminin-α2 and primary or secondary changes in α-dystroglycan, give rise to an MD typically with onset during the first two years of life. The data collected in CMD patients and the research performed in animal models are in favor of the ECM disruption as the main pathological mechanism in these diseases. ECM of skeletal muscle is a complex network composed of collagen, proteoglycans and glycosaminoglycans, glycoproteins and matricellular proteins serving for the attachment and as a signaling scaffold for myofibers (Shieh, 2013). There is a growing body of evidence suggesting that ECM is pivotal to the muscle's normal function, providing the ability to adapt to changes in the microenvironment, and constitutes a biological reservoir of stem cells (Gillies & Lieber, 2011). The concept of a static ECM was completely reshaped with the identification of matrix remodeling enzymes that are responsible for the ECM turnover. This turnover is required in different events such as cell migration, myotube formation, and reorganization of the matrix during muscle adaptation (Gillies & Lieber, 2011). Given the early onset nature of these diseases, it is conceivable that the pathological consequences of ECM disruption could initiate, prematurely even during the fusion of myoblasts that occurs during embryogenesis.

Other CMD subtypes

Further to those already mentioned the genetic defects causing CMD phenotypes, includes eleven additional *loci* with very distinct clinical features and high heterogeneity in term of the pathological mechanisms involved. Seven are commonly shared with other groups of hereditary myopathies, specifically CM (*ACTA1*, *DNM2*, *FHL1,* and *SELENON* genes), EDMD/LGMD (*LMNA* gene) and LGMD (*TCAP* and *TRAPC11* genes).

One of the first identified causes of CMD is related with the integrin α7 subunit (encoded by *ITGA7* gene mainly expressed in skeletal and cardiac muscle) deficiency, despite its extreme rarity, since with only three cases have been reported so far (Hayashi et al., 1998). Integrins are a vast family of transmembrane cell surface receptors participating in a wide variety of cell interactions. Through these interactions with ECM proteins and counterreceptors on other cells, integrins play an important role in cell migration, in cell-cell interactions, and cell shape (Burkin & Kaufman, 1999). The predominant integrin in skeletal muscle that binds laminin-211 and laminin-221 is the α7β1 isoform a heterodimeric transmembrane glycoprotein (Burkin & Kaufman, 1999). Changes in α7β1 integrin are known to be associated with neuromuscular diseases, such as upregulation in DMD patients thought to be a compensation mechanism and primary α7 subunit deficiency in CMD. The phenotype of this subtype is variable, patients suffered from a mild CMD combined with delayed motor milestones (Fu & Xiong, 2017).

Abnormalities in the nuclear envelope proteins lamins A and C, both encoded by the *LMNA* gene, is an additional cause of CMD (Quijano-Roy et al., 2008). These proteins are important structural components of the nuclear lamina, a network underlying the inner nuclear membrane which contributes to determine the nuclear shape and size (Quijano-Roy et al., 2008). In LMNA-CMD phenotype, initially considered as part of a disease spectrum of EDMD, weakness is usually identified during the first 6 months of life, sometimes including a brief phase of more rapid progression during the first 2 years of age with loss of early motor milestones (Mercuri et al., 2004; Fu & Xiong, 2017). One distinctive clinical feature during this stage is the "dropped head" sign, caused by the weakness of axial and neck muscles (flexors and extensors) (Karaoglu et al., 2017). The weakness involving proximal upper limbs and distal lower limbs as, however, a slower progression and facial muscles are usually spared (Bonne & Quijano-Roy, 2013). Like Emery-Dreifuss phenotype, there is cardiac involvement (atrial arrhythmogenic cardiomyopathy with conduction block and ventricular tachyarrhythmias) and contractures manifesting mainly in the lower limbs and spinal rigidity (Fu & Xiong, 2017). Nocturnal respiratory insufficiency may occur early as muscle weakness progresses (Mercuri & Muntoni, 2013).

Megaconial CMD has a completely different pathological mechanism as compared with other CMD subtypes, and only a very limited number of cases have been reported so far. As highlighted by the term "megaconial", this CMD subtype is mainly characterized by the presence of enlarged mitochondria, located predominantly at the periphery of muscle fibers of these patients (Nishino et al., 1998a). It was later shown to be caused by pathogenic variants in the choline kinase beta (*CHKB*) gene which encodes an enzyme that performs the first step of the Kennedy pathway (phosphatidylcholine biosynthesis) (Mitsuhashi et al., 2011.). The lack of this enzyme leads to defects in mitochondrial dynamics such as fusion and fission events, important control mechanisms that regulate the morphology and function of this organelle (Wortmann et al., 2015). The main clinical picture in this disease is characterized by the presence of early-onset muscle weakness and delayed motor skills, the presence of moderate to severe intellectual disability but without any brain changes visible through by MRI and cardiomyopathy (Mistuhashi & Nishino, 2011).

The remaining three genes were identified in the last few years, further expanding the causes of CMD. One relates with defects in the inositol polyphosphate-5-phosphatase K (*INPP5K* gene), giving rise to a variety of different symptoms that comprise a MD with a reduction in α-dystroglycan glycosylation (overlapping with dystroglycanopathies), short stature, intellectual disability, and cataracts (Osborn et al., 2017; Wiessner et al., 2017). Another gene, *TRAPPC11*, encodes for a protein that is a member of the trafficking protein complex 11 involved in endoplasmic reticulum-to-Golgi trafficking (Bögershausen et al., 2013. It was initially associated with an LGMD and later as causing a more severe form that,

giving the age of onset, should be included within the CMD spectra (Liang et al., 2015). In terms of the *TRAPPC11*-related phenotype besides the muscle disease, it includes fatty liver pathology (steatosis) and infantile-onset cataract (Liang et al., 2015). Interestingly, recent reports also linked the defects of this membrane trafficking protein to dystroglycanopathy (α-dystroglycan hypoglycosylation) (Matalonga et al., 2017; Larson et al., 2018).

Finally, defects in *TRIP4* encoding one subunit of the Activating Signal Cointegrator-1 (ASC-1), a poorly characterized transcription coactivator was associated with a congenital muscle disease (Davignon et al., 2016). This clinical entity presents as a neonatal-onset muscle weakness with respiratory failure, skin abnormalities and joint hyperlaxity without contractures (Davignon et al., 2016). Patient's muscle pathology showed an atypical mixed association of multiminicores, caps and dystrophic features (Davignon et al., 2016). Although yet poorly characterized, defects in ASC-1 were identified as a novel pathophysiological mechanism, due to the dysregulation of late myogenic differentiation causing impaired myotube growth (Davignon et al., 2016).

1.2.3 PROGRESSIVE MUSCULAR DYSTROPHIES

In contrast with CMD, the progressive muscular dystrophy (PMD) referrers to a vast group of MDs that typically manifest beyond the first two years of life. These patients have apparent normal motor milestones during infancy, but beyond that period develop loss of muscle function, affecting ambulation, posture, cardiac and respiratory functions, while those with later onset may be milder affected and associated with only slight weakness or fatigue (Chelly & Desguerre, 2013). The typical muscle pathology in MDs, similarly as in CMD, consists of necrosis following regeneration, the proliferation of connective tissue, and increased adipose tissue (Figure 1.2.2). During necrosis, there is lysis of the sarcolemma followed by the Z-disk disruption which causes fibers hypertrophy and rounding of its contours (Malfatti & Romero, 2017). An increase of connective tissue in endomysial and perimysial areas is also common in MD muscle pathology presentation. The etiology of MD, giving rise to muscle weakness and wasting, are heterogeneous and include a wide diversity of genetic pathways dysfunction and genes encoding proteins for components of the plasma membrane, the ECM, sarcomere, and the nuclear membrane. In further detail, the main pathological axis within MDs includes the disruption of the cytoskeleton–ECM connection, aberrant glycosylation of α-dystroglycan (shared with CMD), impaired cell signaling at the sarcolemma, cytoskeletal defects in muscular dystrophy, defective sarcolemma repair and disintegration of muscle sarcomeres (Rahimov & Kunkel, 2013).

Dystrophinopathies

This designation collectively gathers Duchenne MD (DMD) and Becker MD (BMD) two muscle diseases caused by defects in the same gene (Table 1.2.3).

DMD was one of the firstly recognized neuromuscular diseases, more than 160 years ago, through the reports of Edward Meryon and Duchenne de Boulogne (Tyler, 2003). This early disease awareness was probably related with the fact that it is one of the most frequent neuromuscular disorders, with an overall reported incidence ranging from 1:3,802 to 1:6,291 male births (Flanigan, 2014). From the clinical perspective DMD patients present first symptoms between the age of 2 and 5 years, as delayed motor milestones, progressive proximal muscle weakness (positive Gowers' sign), unsteady gait becoming lordotic and waddling, tendency to walk on tiptoes, muscle pain and fatigability, frequent falls, difficulties in running properly and climbing stairs (Malfatti & Romero, 2017). As the disease progresses additional clinical signs may include difficulties rising stairs, running or jumping, Achilles tendon contractures, calf and tongue hypertrophy (Flanigan, 2014). Coupled with the disease's mechanism cognitive function can be affected in a subset of DMD patients, with language development delay, intellectual disability, increased risk of autism, attention deficit

and hyperactivity disorder (Flanigan, 2014). As the majority of the other PMD, patients with dystrophinopathies have elevated (a 10 to 50-fold increase) values of CK. Given that this CK rise can even precede the first clinical symptoms of DMD, it is recommended the assessment of this enzyme as part of the routine screening of all infants with motor delay (Noritz & Murphy, 2013), or even in some countries, as part of the new-born screening program (Gatheridge et al., 2016; Scheuerbrandt, 2018). As muscle weakness progress, patients decline to wheelchair dependence, which in the absence of oral steroids administration occurs by the age of 12 years (Flanigan, 2014). After losing ambulation these patients develop progressive respiratory insufficiency, scoliosis, and cardiomyopathy (Domingos et al., 2017).

In comparison with DMD, Becker muscle dystrophy (BMD), first clearly described by Emil Becker in 1955, is a rarer (incidence is about one-fifth for that in DMD) and milder allelic form of dystrophinopathy (Domingos et al., 2017) BMD is classified by a later-onset muscle weakness and loss of ambulation after the age of 16 years. There is also an intermediate phenotype in severity between DMD and BMD, these cases become wheelchair-bounded between the ages of 13 and 16 years. However, the age at loss of ambulation varies from 16 years to the 6th decade of life, with mean age is in the fourth decade (Chelly & Desguerre, 2013). Thus, BMD presents very a wide range of clinical phenotypes extending from patients that lose the ability to walk towards late teens to other cases experiencing only mild weakness, cramps, and increased CK levels (Chelly & Desguerre, 2013). For most BMD patients, the natural history of the disease is similar to that of DMD, although having later disease onset and the progression rate is slower (Chelly & Desguerre, 2013). In opposition to muscle involvement being less severe than in DMD, progressive dilated cardiomyopathy in BMD is frequent being a major cause of mortality (Domingos et al., 2017). Echocardiographic studies in patients with subclinical or benign BMD (mean age of 18 years), showed cardiac involvement in up to 70% of patients (Domingos et al., 2017). Some BMD patients have cognitive issues as well, mainly behavioral problems and attention deficits and less frequently intellectual disability and autism (Young et al., 2008).

Both DMD and BMD are caused by pathogenic variants in the dystrophin gene (*DMD*), located in the short arm of chromosome X (Xp21), thus explaining why this condition affects almost exclusively males (X-linked recessive inheritance). Female carriers, however, may exhibit DMD/BMD-like phenotypes (including dilated cardiomyopathy) due to skewed inactivation of the X chromosome. That is when inactivation occurs preferentially in the chromosome containing the normal *DMD* gene, leading to the expression of the mutated dystrophin gene (Iyadurai & Kissel, 2016). The *DMD* gene is, in terms of its total size (2.4 Mb), the largest gene in the human genome. Most of its size is mainly due to several introns of massive size, and the coding region, subdivided into 79 exons, is only about 0.6% of this

sequence (Iyadurai & Kissel, 2016). Dystrophin, the *DMD* product, is a protein of 3,685 amino acid and 427 kDa in skeletal muscle. *DMD* originates, through seven independent promoters, the direct transcription of seven transcripts/isoforms that are expressed in different tissues. The full-length isoform (Dp427) is expressed in striated, smooth, and cardiac muscles, but also in the brain; shorter isoforms (Dp260, 140, 116 and 71) are also expressed in the brain, peripheral nerve, and retina (Domingos et al., 2017). In skeletal muscle, dystrophin connects the contractile apparatus with the ECM through the sarcolemma (Iyadurai & Kissel, 2016).

The considerable large size of *DMD* is the explanation for the high mutation rate found in dystrophinopathies, is estimated that 25–33% of all DMD cases are explainable by *de novo* mutational events (Santos et al., 2014). However, in a mother of a D/BMD testing negative for a pathogenic variant in *DMD* gene, there is still the possibility of having a subsequently affected son due to germline mosaicism, thus caution should be used in genetic counseling of such cases. The most common variants found in the *DMD* gene are deletions or duplications of one or more exons, comprising a total of nearly 80% of D/BMD patients (Bladen et al., 2015). As for large deletions, although they can occur almost anywhere along the gene, two main clusters have been identified: a 'minor hotspot' spanning exons 2 to 20, and a 'major hotspot' comprising exons 45 to 53 (Santos et al., 2014). The remaining pathogenic changes (~20%) are essentially sub-exonic or sub-intronic, that include nonsense, splice-site variants, frameshifts (small deletions or insertions), mid-intronic variants that activate cryptic splice sites and very rare missense changes, and even complex rearrangements involving *DMD* locus (Santos et al., 2014).

In terms of genotype-phenotype correlations, DMD is usually caused by pathogenic variants that disrupt the reading frame leading to the complete loss of dystrophin expression, whereas BMD phenotypes are generally associated with in-frame deletions or insertions, which are translated in a mutated yet partially-functional protein. This correlation is known as the 'reading-frame rule' and has been demonstrated in about 91% of cases (Aartsma-Rus et al., 2006). However, several exceptions to this rule have been reported, more commonly found in BMD patients where their frequency depends on the type of genetic defect involved (Aartsma-Rus et al., 2006; Santos et al., 2014).

Traditionally the diagnosis of dystrophinopathies depended upon the use of muscle biopsy. Besides a variable dystrophic pattern found in these patients, the use of specific antibodies directed against different dystrophin domains is of great diagnostic values. The majority of DMD patients have a complete absence of dystrophin labeling in sarcolemma, however, in some cases, it is possible to have some fibers (known as revertant fibers), showing normal or almost normal intensity that is expressing some degree of dystrophin (Malfatti & Romero, 2017). By contrast, the muscles of BMD patients show a reduced, fainted

or uneven labeling of muscle fibers. This pattern is explainable by the type of effect of the *DMD* variant involved. The reading frame is mostly disrupted or containing a premature termination codon in DMD cases or roughly maintained in BMD, where inframe deletions or duplications are mostly found.

For the molecular genetics diagnosis of dystrophinopathies, standard routine techniques include multiplex ligation-probe amplification (MLPA) or chromosome microarray (array-CGH) that are used to detect large deletions and duplications. These approaches replaced, almost by completely, Southern-Blot and multiplex PCR approaches that were used in the initial period of molecular studies. For the identification of the remaining types of pathogenic variants, that is smaller/point variants, *DMD* gene sequencing is usually performed as a second-tier level, being complemented in some cases with studies at the cDNA level (Santos et al., 2014).

The pathophysiology underlying dystrophinopathies have been subject to extended analysis over the past two decades, and the gathered knowledge propelled the development of new tailored therapeutic approaches. It is known, since the early biochemistry studies to characterize dystrophin, that this protein plays an important role in connecting the cytoskeleton to the ECM, stabilizing the muscle fiber, especially during muscle contraction. Dystrophin is an intracellular rod-shaped protein (Figure 1.2.3) localized to the inner surface of the sarcolemma being particularly abundant at costameres and sites of cell-cell contact (Rahimov & Kunkel, 2013; Giraud et al., 2015). As reviewed by Constantin (2014), dystrophin is organized in four main regions: i) The N-terminal domain contains a pair of calponin homology (CH) modules binding F-actin; ii) a large central rod domain of 24 homologous triple helical repeats and four hinge domains, which could as a shock absorber and providing flexibility to the protein; iii) the third section is composed of a WW domain, a short conserved region in different proteins, folding as a stable, triple-stranded beta-sheet. This domain is known to bind proteins with proline, and/or phosphoserine-phosphothreonine containing motifs. It is frequently associated with other domains typical for proteins in signal transduction processes. This WW domain is followed by a cysteine-rich domain with two EF-hand motifs and two ZZ modules that bind to calmodulin through a calcium-dependent way; iv) and finally the C-terminal domain interacts with multiple proteins to assemble the DAPC (Rahimov & Kunkel, 2013). This complex is composed of several proteins, namely dystroglycan, sarcoglycans, dystrobrevins and syntrophin (Figure 1.2.3). The link between DAPC and dystrophin, commonly referred to as the dystrophin-glycoprotein complex (DGC), provided the molecular basis for the contraction-induced sarcolemma injury model underlying dystrophinopathies pathogenesis (Constantin, 2014). In addition, DAPC is also now considered a putative cellular signaling complex by conferring the scaffold for numerous signaling proteins and through the association with neuronal nitric oxide synthase (nNOS)

(Constantin, 2014). Dystrophin deficiency was shown to destabilize DAPC, consequently, muscle cells become more susceptible to contraction-induced membrane damage and ultimately there is an increase in membrane fragility (Chelly & Desguerre, 2013; Domingos et al., 2017). Consequently, this fragility was thought to have implications in calcium homeostasis, a critical aspect of muscle function, and it was hypothesized that increased intracellular calcium could account for muscle necrosis through the enhancement of calcium-activated proteases (Domingos et al., 2017). The influx of calcium and the abnormal release of cytoplasmic content to the extracellular space also causes the activation of the immune system with an associated inflammatory response and muscle fiber necrosis (Domingos et al., 2017). Inflammation is thought to be a major driver of the progressive fibrous tissue deposition that mainly characterizes DMD (Domingos et al., 2017). Additional mechanisms are likely involved, such as deregulation of cell signaling that could lead to cellular and tissue changes such as oxidative damage and increased apoptosis (Chelly & Desguerre, 2013).

An extensive list of therapeutic strategies for DMD has been investigated or are currently under clinical trials (reviewed by Leigh et al., 2018). The first group comprises approaches that are applied regardless of the type of the DMD pathogenic variant involved:

i) Reducing inflammation is the primary target of the classical approach to treat DMD patients, resorting to glucocorticoids (prednisone or deflazacort). Also aiming to reduce inflammatory and fibrotic changes, new drugs as Vamorolone (VBP15) are being tested to confirm that it can inhibit more specifically the NF-κB pathways with fewer side effects (Leigh et al., 2018).

ii) Antioxidants, as one of the disease's mechanism, is related to the overproduction of ROS and ultimately leading to membrane permeability and protein turnover. One antioxidant that inhibits lipid's peroxidation, Idebenone, has been effective to improve respiratory function in phase III clinical trial (Buyse et al., 2015; Leigh et al., 2018).

iii) Utrophin modulation, using SMT C1100 molecule to increase utrophin expression, in a way to compensate the dystrophin deficiency and ameliorate disease symptoms (Tinsley et al., 1993).

iv) Myostatin inhibitors have been demonstrated to increase considerably muscle mass and strength, decreased serum CK concentration, and improved muscle histology in the mdx mouse a model for DMD (Bogdanovich et al., 2002). Despite previous tests with MYO-029 an antibody against myostatin failed, a new generation of anti-myostatin drugs such as PF-06252616 and adnectin are being evaluated on clinical trials (Leigh et al., 2018).

v) Stem cell therapy, studies are currently underway some with promising results, namely the isolation and transplantation of muscle satellite cells, mesenchymal stem cell therapy, concerning clinical trial myoblast transfer was not successful but further optimization is still the preclinical phase (Brandsema & Darras, 2015).

vi) Gene replacement, systemically delivery of an Adeno-associated virus (AAV) vector resulted in successful delivery of DMD to affected muscles of dystrophin-deficient mdx mice (Gregorevic et al., 2006). As the large size of the *DMD* gene causes extreme difficulties for vector packaging, a micro-version of dystrophin was used. Clinical trials are underway, but despite progress in the field, cellular immunity issues may be an obstacle to successful dystrophin gene therapy for DMD (Brandsema & Darras, 2015).

The second group of therapeutic approaches aims to restore the reading frame or correct the native dystrophin by correcting/surpassing the genetic defects found in patients. This makes the knowledge about their genotype pivotal for treatment, here are some examples:

i) Stop codon readthrough drugs, a strategy to suppress termination codons, thus having only applicability to pathogenic variants in *DMD* that cause premature translation arrest (nonsense variants, ~10-15% of cases). The initial pieces of evidence were provided by clinical research using an aminoglycoside – gentamycin. This antibiotic induces the misreading of such codons in RNA and allows alternative amino acids to be incorporated in the polypeptide being synthesized (Leigh et al., 2018). Translarna (ataluren, PTC124) is a newer, orally administered nonantibiotic drug that also promotes ribosomal read-through of nonsense variants (Brandsema & Darras, 2015). Phase III clinical trials results have shown promising, making its approval by European Medicines Agency, although being considered inconclusive by the American Food and Drugs Administration (FDA) (Leigh et al., 2018).

ii) Exon skipping using antisense oligonucleotides and morpholinos was conceptualized as a feasible approach to restore the *DMD* gene reading frame in DMD patients. Since pathogenic variants around exon 51 is a mutational hotspot in *DMD*, two drugs Eteplirsen (Exondys 51) and Drisapersen, have been developed to promote this exon skipping specifically (Leigh et al., 2018). After some promising results in clinical trials, Eteplirsen received in 2016 FDA authorization under accelerated approval. A recent systematic review and meta-analysis suggested that all data gathered so far does not clearly support the effectiveness of such compounds to treat DMD. Further, and despite some potential utility when used at a specific dose, significant side effects were reported with Drisapersen usage (Shimizu-Motohashi et al., 2018).

iii) Gene editing comprises technologies that allow genetic material to be added, removed, or altered in specific locations of the genome. One recently introduced and widely used is the CRISPR/Cas9, that is clustered regularly interspaced short palindromic repeats and CRISPR-associated protein 9, considered a fast, inexpensive, and relatively accurate method for genomic editing (Cong et al., 2013). This technique allows the correction of a broad spectrum of genome modifications, thus having a high potential for gene therapy of several genetic conditions (Kuruvilla et al., 2018). CRISPR/Cas9 system is composed of two main components, the first being the Cas9, an endonuclease enzyme functioning to edit

DNA, and the second a guide RNA that is made to match the target (Wang et al., 2017). Genome editing with the CRISPR/Cas9 system has been recently reported effective to delete exons of the gene in dystrophic mice leading to an improvement of muscle expression and function (Long et al., 2016). More recently, Amoasii and co-workers (2018) tested the same technique in a dog model for DMD, which displays many disease features of the human counterpart. Using both intramuscular and systemic delivery of the gene editing components to promote exon 51 skipping, the research showed a variable increase in the protein levels of dystrophin in skeletal and cardiac muscle. Restoration of dystrophin expression was accompanied by improved muscle histology (Amoasii et al., 2018). One major drawback of this technique is the possibility of off-target genomic changes.

Table 1.2.3- Genes and phenotypes associated with progressive muscular dystrophies.

Phenotypes (phenotype OMIM)	Gene	Nr. Exons	Chromosomal localization	Protein	Inheritance pattern	Reference (phenotype-gene association)
I. Duchenne/Becker Muscular Dystrophies (D/BMD)						
D/BMD pathophysiological mechanism: The connection between the cytoskeleton and extracellular matrix is compromised by defects in dystrophin. Thus, muscle fibers are not stabilized during muscle contraction, becoming more susceptible to contraction-induced membrane damage.						
Duchenne / Becker muscular dystrophy (#310200, #300376)	*DMD*	79	Xp21.2	Dystrophin	XLR	Koenig et al., 1988
II. Limb-girdle muscular dystrophies (LGMD) [1]						
LGMD pathophysiological mechanism I: Disruption of the cytoskeleton-extracellular matrix connection. Defects in one of the components of the subcomplex of integral proteins sarcoglycans (SGC) compromises the necessary mechanical support to the dystrophin-associated protein complex, leading to the reduction or complete loss of the other SGC subunits, destabilizing the entire protein complex which consequently weakens the sarcolemma.						
LGMD R3 [LGMD2D] (#608099)	*SGCA*	10	17q21	Alpha-sarcoglycan	AR	Roberds et al., 1994
LGMD R5 [LGMD2C] (#253700)	*SGCG*	8	13q12	Gamma-sarcoglycan	AR	Noguchi et al., 1995
LGMD R4 [LGMD2E] (#604286)	*SGCB*	6	4q12	Beta-sarcoglycan	AR	Bönnemann et al., 1995
LGMD R6 [LGMD2F] (#601287)	*SGCD*	9	5q33-q34	Delta-sarcoglycan	AR	Nigro et al., 1996
LGMD pathophysiological mechanism II: Defective calpain-3 cannot cleave filamin C consequently altering its interaction with subunits of the sarcoglycan complex. Pathogenic variants in calpain-3 also compromise sarcomere turnover and remodeling.						
LGMD R4, LGMD D1 [LGMD2A, LGMD1I] (#253600, #618129)	*CAPN3*	24	15q15.1-q21.1	Calpain 3	AR, AD	Richard et al., 1995; Vissing et al., 2016

Table 1.2.3- Genes and phenotypes associated with progressive muscular dystrophies (continues).

LGMD pathophysiological mechanism III: Disintegration of muscle sarcomeres.						
(a) LGMD1A (#159000)	*MYOT*	11	5q31	Myotilin	AD	Hauser et al., 2000
LGMD R4 [LGMD2H] (#254110)	*TRIM32*	2	9q33.2	Tripartite motif-containing 32	AR	Frosk et al., 2002
LGMD R7 [LGMD2G] (#601954)	*TCAP*	2	17q12	Telethonin	AR	Moreira et al., 2000
LGMD R10 [LGMD2J] (#608807)	*TTN*	363	2q31	Titin	AR	Hackman et al., 2002
(a) [LGMD2R] (#615325)	*DES*	9	2q35	Desmin	AD, AR	Hedberg et al., 2012; Cetin et al., 2013
LGMD R17 [LGMD2Q] (#613723)	*PLEC*	33	8q24.3	Plectin	AR	Gundesli et al., 2010; Fattahi et al., 2015
LGMD pathophysiological mechanism IV: The link between dystrophin-associated glycoprotein complex (DGC) and the basal lamina is compromised due to α-dystroglycan hypoglycosylation or other glycosylation defects.						
LGMD R9 [LGMD2I or MD-dystroglycanopathy type C, 5] (#607155)	*FKRP*	4	19q13.32	Fukutin-related protein	AR	Brockington et al., 2001
LGMD R11 [LGMD2K or MD-dystroglycanopathy type C, 1] (#609308)	*POMT1*	20	9q34.13	Protein-O-mannosyltransferase 1	AR	Balci et al., 2005
LGMD R13 [LGMD2M or MD-dystroglycanopathy type C, 4] (#611588)	*FKTN*	12	9q31.2	Fukutin	AR	Godfrey et al., 2006; Murakami et al., 2006
LGMD R14 [LGMD2N or MD-dystroglycanopathy type C, 2] (#613158)	*POMT2*	21	14q24.3	Protein-O-mannosyltransferase 2	AR	Biancheri et al., 2007
LGMD R15 [LGMD2O or MD-dystroglycanopathy type C, 3] (#613157)	*POMGNT1*	22	1p34.1	O-linked mannose beta1,2-N-acetylglucosaminyltransferase	AR	Clement et al., 2008
(a) [Congenital disorder of glycosylation (CDG), type Io] (#612937)	*DPM3*	2	1q22	Dolichyl-phosphate mannosyltransferase polypeptide 3	AR	Lefeber et al., 2009
LGMD R16 [LGMD2P or MD-dystroglycanopathy type C, 9] (#613818)	*DAG1*	6	3p21	Dystroglycan	AR	Hara et al., 2011
LGMD R18 [LGMD2S] (#615356)	*TRAPPC11*	31	4q35.1	trafficking protein particle complex 11	AR	Bögershausen et al., 2013

Table 1.2.3- Genes and phenotypes associated with progressive muscular dystrophies (continues).

LGMD R19 [LGMD2T or MD-dystroglycanopathy type C, 14] (#615352)	*GMPPB*	8	3p21.31	GDP-mannose pyrophosphoryl ase B	AR	Carss et al., 2013
LGMD R20 [LGMD2U, MD-dystroglycanopathy type C, 7] (#616052)	*ISPD*	10	7p21.2	Isoprenoid synthase domain containing	AR	Tasca et al., 2013
(a) MD-dystroglycanopathy type C, 12 (#615247)	*POMK*	5	8p11.21	Protein-O-mannose kinase	AR	Jae et al., 2014
LGMD R24 [MD-dystroglycanopathy type C, 8] (#618135)	*POMGNT2*	2	3p22.1	protein O-linked mannose N-acetylglucosam inyltransferase 2	AR	Endo et al., 2015
LGMD R21 [LGMD2Z] (#617232)	*POGLUT1*	11	3q13.33	Protein O-Glucosyltransfe rase 1	AR	Servián-Morilla et al., 2016
LGMD pathophysiological mechanism V: Defective sarcolemma repair or disruption of caveolae formation.						
LGMD R2 (LGMD2B) (#253601)	*DYSF*	56	2p12-14	Dysferlin	AR	Bashir et al., 1998; Liu et al., 1998
(a) [LGMD1C] (#607801)	*CAV3*	3	3p25	Caveolin 3	AD, AR	Minetti et al., 1998
LGMD R12 [LGMD2L] (#611307)	*ANO5*	22	11p14-12	Anoctamin 5	AR	Bolduc et al., 2010
LGMD pathophysiological mechanism VI: Defective nuclear membrane or nuclear import proteins.						
(a) [LGMD1B] (#159001)	*LMNA*	12	1q22	Lamin A/C	AD	Muchir et al., 2000
LGMD D2 [LGMD1F] (#608423)	*TNPO3*	23	7q32.1-q32.2	Transportin 3	AD	Melià et al., 2013; Torella et al., 2013
LGMD pathophysiological mechanism VII: Chaperone protein that binds BAG3 and HSP70, protects proteins from irreversible aggregation. Mutated protein has limited capacity to prevent protein aggregation.						
LGMD D1 [LGMD1E] (#603511)	*DNAJB6*	10	7q36	HSP-40 homologue, subfamily B, number 6	AD	Sarparanta et al., 2012
LGMD autosomal dominant with progressive fingers and toes flexion limitation:						
LGMD D3 [LGMD1G] (#609115)	*HNRNPDL*	8	4q21	Heterogeneous nuclear ribonucleoprote in D-like	AD	Vieira et al., 2014
Other LGMD phenotypes not included in the newest classification system (only one family identified):						
LGMD2Y (#617072)	*TOR1AIP1*	10	1q25.2	Torsin A interacting protein 1	AR	Kayman-Kurekci et al., 2014

Table 1.2.3- Genes and phenotypes associated with progressive muscular dystrophies (continues).

LGMD2W (#616827)	*LIMS2*	11	2q14.3	LIM and senescent cell antigen-like domains 2	AR	Chardon et al., 2015
LGMD2X (#616812)	*BVES*	8	6q21	Blood vessel epicardial substance	AR	Schindler et al., 2016

III. Emery-Dreifuss muscular dystrophies (EDMD)

EDMD pathophysiological mechanism: Enhanced nuclear fragility due to defective membrane proteins may affect mechanically stressed tissues such as cardiac and skeletal muscles. Further, the altered nuclear membrane could interfere with satellite cell function and thereby skeletal muscle regeneration.

EDMD1 (#310300)	*EMD*	6	Xq28	Emerin	XLR	Bione et al., 1994
EDMD2 & 3 (#181350, #616516)	*LMNA*	12	1q22	Lamin A/C	AD, AR	Bonne et al., 1999
EDMD4 (#612998)	*SYNE1*	146	6q25	Spectrin repeat containing, nuclear envelope 1 (nesprin 1)	AD	Zhang et al., 2007
EDMD5 (#612999)	*SYNE2*	116	14q23.2	Spectrin repeat containing, nuclear envelope 2 (nesprin 2)	AD	Zhang et al., 2007
EDMD6 (#300696)	*FHL1*	7	Xq26.3	Four and a half LIM domain 1	XLR	Gueneau et al., 2009
EDMD7 (#614302)	*TMEM43*	12	3p25.1	Transmembrane protein 43	AD	Liang et al., 2011

IV. Facioscapulohumeral muscular dystrophy (FSHD)

FSHD pathophysiological mechanism: Transcriptional derepression of DUX4 in skeletal muscle through the activation of a canonical polyadenylation signal. This causes a gain-of-function attributable to the stabilized DUX4 transcripts, ultimately contributing to the pathogenesis of this disease.

Facioscapulohumeral muscular dystrophy 1 (#158900)	*DUX4*	3	4q35	Double homeobox 4	AD	Wijmenga et al., 1992
Facioscapulohumeral muscular dystrophy 2, digenic (#1158901)	*SMCHD1*	48	18p11.32	Structural maintenance of chromosomes flexible hinge domain containing 1	Digenic	Lemmers et al., 2012

Footnote: AD, autosomal dominant; AR- autosomal recessive; XLR- recessive linked to X chromosome; data used to generate this table was obtained from the GeneTable of Neuromuscular Disorders (Bonne & Rivier, 2018), OMIM (2018) and cited literature. The LGMD nomenclature based on a recently reformed classification proposed by Straub and collaborators (2018) is indicated between square brackets. According to this same classification, *LAMA2*, *COL6A1*, *COL6A2,* and *COL6A3* genes and respective phenotypes are now included in LGMD and (a) indicated genes now excluded from LGMD list based in the newest nomenclature (Straub et al., 2018).

Limb-girdle muscular dystrophies

Limb-girdle muscular dystrophies (LGMD) is a vast and heterogeneous group of diseases, mainly characterized by a specific pattern of muscle weakness and wasting primarily affecting the pelvic and shoulder girdle (Domingos et al., 2017). This heterogeneity is not exclusively clinical, where additional symptoms can help to ascertain the different subtypes, but essentially derives from the diverse genetic etiology found in LGMDs. Overall, disease onset ranges from childhood (beyond two years of life) to adulthood (up to the 4th decade of life). There are variable degrees of CK from nearly normal values to extremely elevated values. In term of muscle pathologic findings, there is also high variability ranging from nonspecific myopathic changes to dystrophic features (Liewluck & Milone, 2018). Its hereditary nature was long-recognized just from the early classification of the disease (Walton & Nattrass, 1954) and led to further classification in AR and AD forms, being the former the most frequent inheritance pattern in LGMDs.

According to the most recent internationally-recognized classification system proposed by Straub and collaborators (2018), LGMD "*is a genetically inherited condition that primarily affects skeletal muscle leading to progressive, predominantly proximal muscle weakness at presentation caused by a loss of muscle fibers. To be considered a form of limb-girdle muscular dystrophy the condition must be described in at least two unrelated families with affected individuals achieving independent walking, must have an elevated serum creatine kinase activity, must demonstrate degenerative changes on muscle imaging over the course of the disease, and have dystrophic changes on muscle histology, ultimately leading to end-stage pathology for the most affected muscles*" (Straub et al., 2018). As it is foreseeable in Table 1.2.3, this new classification of LGMD is genetically-driven (gene and inheritance pattern). Since the first identification of a defective gene nearly a quarter of a century ago - alpha-sarcoglycan gene (*SGCA*) (Roberds et al., 1994) - the number of LGMD-related genes has grown immensely.

The frequencies among the different subtypes are variable within the different countries or ethnical groups. Calpainopathy is considered one of the most frequent LGMD subtypes in most European countries and in the United States (Liewluck & Milone, 2018). However, there are some exceptions, namely in Denmark where *FKRP*-related is predominant (Sveen et al., 2006), whereas in Portugal gamma-sarcoglycanopathy is the most prevalent form of LGMD (31% of total genetically confirmed LGMD patients) followed by dysferlinopathy (24%) (Vieira et al., 2012).

Following the initial identification of the *SGCA* gene (causing LGMD R3) the remaining genes that encode for the other components of the sarcoglycan complex, i.e. gamma (*SGCG*), beta (*SCGB*) and delta (*SGCD*), were also subsequently associated to LGMDs (LGMD R5, R4, and R6 respectively). These four heavily glycosylated proteins are the fabric

of a larger sarcomeric structure known as DGC (McNally & Pytel, 2007). The sarcoglycan complex has a pivotal role in maintaining sarcolemmal integrity and the necessary mechanical support to dystrophin-associated protein complex (DAPC). Most of the patients with sarcoglycanopathies display a phenotype similar to DMD, while others with a milder presentation, may mimic BMD, or even display exercise-induced myalgias and rhabdomyolysis with myoglobinuria mainly identified by multi-gene sequencing panels (Tarnopolsky et al., 2015; Liewluck & Milone, 2018). This phenotypic overlap with dystrophinopathies also includes cardiorespiratory involvement and loss of ambulation during adolescence (Liewluck & Milone, 2018). Patients with milder phenotype may remain independent locomotion throughout adulthood (Kirschner & Lochmuller, 2011). The presence of sarcoglycans in the muscle sarcolemma is detectable by IHC using specific antibodies.

Calpainopathy, one of the most frequent subtypes of LGMDs, shares some of the typical, clinical features including symmetric and progressive weakness of the girdle muscles. However, it presents some degree of clinical variability (three main phenotypes) and the age at onset varies from two to 40 years of age (Angelini & Fanin, 2017). Scapular winging is also a frequent unspecific clinical sign of calpainopathy. It is essentially an AR disease caused by pathogenic variants in *CAPN3* gene, which product – calpain-3 – is a calcium-activated neutral protease highly expressed in skeletal muscle, involved in the mechanism of sarcomere remodeling (Fanin & Angelini, 2015). Additional studies have also shown that filamin C may be a substrate for calpain-3, functioning to regulate its protein-protein interaction with the sarcoglycans (Guyon et al., 2003). Although being widely established as an AR, the inheritance pattern of calpainopathies was recently expanded to include autosomal dominance (Vissing et al., 2016; Martinez-Thompson et al., 2018). These patients shared the same heterozygous variant in *CAPN3* (c.643_663del) and were reported having a milder phenotype (myalgia, back pain or hyperlordosis) as compared with LGMD2A. The AD calpainopathy has onset around the 4th decade of life (Vissing et al., 2016). CK values were normal to 1,800 U/L and muscle biopsies were normal to only mildly myopathic changes, and calpain-3 expression was reduced (Martinez-Thompson et al., 2018). Western-blot analysis of muscle biopsies using a calpain-3 antibody is a highly specific assay for diagnosing calpainopathies (Fanin & Angelini, 2015). Not only it may demonstrate calpain-3 protein reduction or absence but, in some patients, additional further information about its function through the autolytic activity assay (Fanin et al., 2009).

The next group of LGMD phenotypes are caused by defects in the sarcomere components, more specifically of the Z-disk an attachment region for thin filaments and delineates the sarcomeric boundaries (Figure 1.1.2). As further introduced in 1.2.4 section, genetic defects affecting structural components of the Z-disk, or other proteins that contribute towards the maintenance of its integrity, typically cause myofibrillar myopathies (Liewluck &

Milone, 2018). It may seem somewhat contradictory their inclusion in LGMD, however, some patients with pathogenic variants in myotilin (*MYOT*), telethonin (*TCAP*), titin (*TTN*), desmin (*DES*) a plectin (*PLEC*) genes do not display the full histopathology picture of myofibrillar myopathies (Liewluck & Milone, 2018). Based on this newest classification *DES* and *MYOT* were even formally excluded from LGMD. The clinical presentation of such subtypes is variable, involving not only proximal girdle muscles but tend to affect also distal muscles (those in hands, forearms, lower legs, and feet). Cardiac symptoms are also reported in nearly all subtypes with exception of *PLEC*-related diseases (Domingos et al., 2017). Early respiratory involvement has been associated with defects in desmin. An additional clinical clue is facial weakness which may be present in *MYOT*- and *DES*-related phenotypes (Thompson & Straub, 2016). TRIM32 is a ubiquitin ligase and not a direct sarcomere component, was included here as it binds to myosin and ubiquitinates actin and participates in sarcomere recycling (Kudryashova et al., 2005).

Considerations about the genetic diversity and the pathophysiology of dystroglycanopathies have been highlighted in section 1.2.2. The majority of genes associated with the dystroglycanopathies are associated with both allelic disorders CMD and LGMD, as part of the same genetic entity. Within the list of twelve genes shown in Table 1.2.3, only two (*DPM3* and *POGLUT1*) have been found exclusively associated with LGMD and were not reported as a cause of CMD yet. Muscle pseudohypertrophy, cardiac involvement, and most distinctively a wide range of cognitive impairment are commonly found clinical features among LGMD patients with dystroglycanopathy (Liewluck & Milone, 2018). One of the genes listed, *GMPPB*, is commonly shared with an emerging group of congenital myasthenic syndromes secondary to hypoglycosylation of proteins in the neuromuscular junction (Carss et al., 2013). Another phenotype of LGMD (type R21) involves defects in O-glucosyltransferase 1 (*POGLUT1*) gene (Servián-Morilla et al., 2016). Although patients with variants in this gene showed a mild hypoglycosylation of α-dystroglycan, this feature is not the primary disease mechanism. The key pathomechanism in this form of LGMD is the Notch-dependent loss of muscle satellite cells that impairs muscle regeneration and causing to muscular dystrophy (Servián-Morilla et al., 2016).

The considerable size of individual muscle fibers and the mechanical stress forces that they are subject during the normal process of muscle contraction renders the fragile sarcolemma susceptible to micro injuries (wear-and-tear). In healthy muscle, the lesions caused by such events by a repair pathway that is Ca^{2+}-dependent (Laval & Bushby, 2004). This dynamic repair process is mediated by the fusion of membrane-bound intracellular vesicles with the sarcolemma. However, its repair is dependent on the size of the lesions is and in the interaction of multiple proteins residing both on the sarcolemma and in the vesicles complexes, being tightly coordinated in a spatial-temporal way (Barthélémy et al.,

2018). Defects in such repair process cause muscle degeneration and ultimately leads to muscle dystrophy phenotype. The first mutated gene identified in this group was dysferlin (*DYSF*), that encodes for a 230-kD transmembrane protein with six intracellular C2 domains, a major component of the Ca^{2+}-dependent sarcolemma repair pathway (Bashir et al., 1998; Liu et al., 1998). Three distinct phenotypes have been associated with pathogenic variants in the *DYSF* gene: LGMD 2R, Miyoshi myopathy, and distal anterior compartment myopathy (Illa et al., 2001), collectively known as dysferlinopathies. LGMD 2R, a frequent subtype among LGMDs, has typical onset during late teens to early adulthood, presenting as muscle pain, increased fatigue and difficulties in running, very high CK levels and generally a slow disease progression (Santos et al., 2010). As dysferlin promotes the intracellular vesicles fusion among them and with the sarcolemma at the injury site, a defective protein will fail to perform this activity, leading to the characteristic accumulation of subsarcolemmal vesicles in affected muscles (Bansal et al., 2003). Dysferlin also interacts with multiple proteins, one of such is caveolin-3 a major constituent of the plasma membrane caveolae, that is small invaginations of the plasma membrane (Matsuda et al., 2001). Pathogenic variants in *CAV3* gene are responsible for rippling muscle disease (percussion-induced repetitive contractures, formerly known as LGMD1C), with such peculiar clinical sign would prompt for this genetic entity (Bruno et al., 2012). An additional genetic cause of LGMDs with defective membrane repair, are disease-causing variants in anoctamin 5 (*ANO5*) gene (Bolduc et al., 2010). These give rise to a slowly progressive LGMD (LGMD R12), although there is a wide spectrum of clinical manifestations, including hyperCKemia, and exercise-induced myalgia (Penttila et al., 2012; Schessl et al., 2012). Given the phenotypic overlap with dysferlin-associated MD, it was suggested that *ANO5* could be involved in repair skeletal muscle membrane upon injury. Anoctamin 5 is a putative calcium-activated chloride channel, but its exact function is not yet fully known (Wicklund & Kissel, 2014). Muscle biopsies of patients with *ANO5* mutations displayed multifocal sarcolemma lesions, but unlike in typical dysferlinopathies, no accumulation of subsarcolemmal vesicle was observed (Bolduc et al., 2010). Thus, making unlikely that it acts through the dysferlin-dependent repair pathway previously mentioned. One possible explanation is that increased Ca^{2+} concentration within a damaged myofiber might activate ANO5 residing on intracellular vesicles. This activation allows the transport of chloride into the vesicles to modify their conformation and facilitate their recruitment to the injury site (Bolduc et al., 2010). This may reflect the scramblases activity found in anoctamin family, which can modify phospholipids orientation, that are normally placed in the inner leaflet of the membrane, by exposing them to the extracellular surface (Benarroch, 2017).

The last LGMD-related pathophysiological mechanism highlighted here comprises a defective nuclear membrane or nuclear import proteins. Collectively known as nuclear

envelopathies, they encompass a very genetically heterogeneous group of human disorders due to mutations in genes encoding a variety of nuclear membrane proteins (Worman & Dauer, 2014). Concerning muscle diseases, these envelopathies go beyond LGMD and are included in CMD and especially in Emery-Dreifuss muscular dystrophy (EDMD) as depicted in the next topic. Concerning LGMDs, it can be caused essentially by pathogenic variants in transportin 3 gene (*TNPO3*, LGMD 2D) (Melià et al., 2013; Torella et al., 2013). Transportin 3 has been implicated in the nuclear import of a wide range of Ser/Arg-rich domain-containing proteins (including splicing factors) involved in mRNA metabolism (Maertens et al., 2014). This nuclear importer was also surprisingly associated with HIV-1 infection, through a functional interaction with viral capsid proteins (Bhargava et al., 2018). The remaining genes, both encoding nuclear envelope proteins, lamin A/C (*LMNA*) and lamina-associated polypeptide 1B (*TOR1AIP1*) were previously included in LGMD (1B and 2Y, respectively) but were meanwhile excluded due to different motives (Straub et al., 2018). *LMNA* as it is essentially associated to a phenotype of Emery-Dreifuss muscular dystrophy and *TOR1AIP1* because only one family has been found so far with such genetic defects (Kayman-Kurekci et al., 2014; Straub et al., 2018).

Emery-Dreifuss muscular dystrophy

Emery-Dreifuss muscular dystrophy (EDMD) is mainly characterized by a core of clinical features that include: i) early childhood onset of joint contractures; ii) muscle weakness and wasting, slowly progressing from a humeroperoneal distribution, which latter extends to the scapular and pelvic girdle muscles; iii) cardiac involvement that may manifest as an array of features such as palpitations, presyncope and syncope, and heart failure; and iv) myopathic features, mainly dystrophic features, on muscle biopsy (Bonne et al., 2015). As in other muscle diseases there are also wide inter- and intrafamilial variability (Bonne et al., 2015). Most EDMD patients are normal at birth and during their first years of life, and contractures often developing in the second decade of life, affecting the elbows or ankles (Puckelwartz & McNally, 2011). Unlike other MDs, the early elbow contractures limit the use of arm strength, constituting a diagnostically distinctive feature (Puckelwartz & McNally, 2011). Contractures can also affect the posterior cervical muscles, limiting neck flexion and eventually, the spine forward flexion may also become limited (Puckelwartz & McNally, 2011). EDMD is a genetically heterogeneous condition until now six different genes have been identified as a cause of EDMD, all encoding proteins that are associated with the inner and outer membranes of the nucleus, comprising the nuclear envelope (Table 1.2.3). The first genetically determined EDMD was related to defects in the emerin gene (*EMD*) (Bione et al., 1994). Emerin and other LEM (Lap2, emerin, Man1)-domain proteins are integral membrane polypeptides localized in the inner membrane (Berk et al., 2013). These LEM-domain

proteins bind directly to lamins (nuclear intermediate filament) and to barrier-to-autointegration factor (BAF) forming a major component of NE-associated nucleoskeletal structure known as the nuclear lamina (Berk et al., 2013). *EMD* gene is located in Xq28 and associated with an XLR disease inheritance. Pathogenic variants in *EMD* are the second most frequent of EDMD, just following defects in the gene encoding lamin A/ C (*LMNA*) the leading cause of this disease. The nuclear lamina is a multi-protein lattice composed of A- and B-type lamins and their associated proteins, associates not only with the inner nuclear membrane proteins but also with heterochromatin, providing links among the genome, nucleoskeleton, and cytoskeleton (Brady et al., 2018). Probably due to its intricated role, *LMNA* gene is a classic example of pleiotropism being associated, when defective, with ten different diseases that affect not only skeletal muscle but also cardiac muscle, peripheral nerve, or fat tissue (Ho & Hegele, 2018). Most EDMD patients in LMNA-related are either sporadic due to *de novo* mutations or with AD inheritance in familial cases, but also, much less frequently, AR (Bonne et al., 1999).

Later in 2007, Zhang and collaborators identified AD pathogenic variants in nesprins genes (*SYNE1* and *SYNE2*) in EDMD patients. These genes encode for proteins of the spectrin-repeat family, that link the inner nuclear membrane to cytoskeletal proteins to mediate nuclear envelope localization and integrity (Zhang et al., 2007). Nesprin was shown to be important to stabilize emerin and laminin, being among the strongest emerin-binding partners (Zhang et al., 2007). This was further corroborated by in vitro experiments with mutated emerin (with either missense or frameshift variants), revealing altered binding affinities to nesprin-1 and -2 when compared to the wild-type, implicating the emerin–nesprin binding defects in EDMD (Wheeler et al., 2007; Zhang et al., 2007). Besides the role of nesprins as a scaffold for the nucleoskeleton composed of spectrins, actin, and lamins, disruption of the integrity of this structure can also result in chromatin reorganization, with concomitant changes in gene expression and nuclear disorganization (Zhang et al., 2007). However, the number of cases associated with nesprins are still very limited. Missense variants found in *SYNE1* and *SYNE2* affected highly conserved evolutionarily residues, located within the lamin and emerin binding domains (Zhang et al., 2007). An additional cause of EDMD was associated with the four-and-a-half LIM domain 1 (*FHL1*) *locus* in (Gueneau et al., 2009). The gene originates multiple protein isoforms that were not initially established as involved with the nuclear envelope, making the pathophysiology mechanism unclear and constituting an exception among the genetic defects involved in EDMD. To add further complexity, three other myopathic phenotypes have been correlated with pathogenic variants in *FHL1*: X-linked myopathy with postural muscle atrophy and generalized hypertrophy, X-linked dominant scapuloperoneal myopathy, and the reducing body myopathy (Cowling et al., 2011). A possible explanation for the different expression of phenotype is the

localization of variants within the gene, being more distally located in EDMD (exons 5 to 8) (Gueneau et al., 2009). More recently, a specific isoform, FHL1B, was identified and determined to be a nuclear envelope protein, just like lamin A/C and emerin (Ziat et al., 2016).

The sixth genetic cause of EDMD is *TMEM43*, encoding a protein, also known as LUMA, found within the inner nuclear membrane being associated with the linker of the nucleoskeleton and cytoskeleton complex (Liang et al., 2011). Pathogenic variants in this gene are associated with an AD and clinical data is still limited. Similarly, the remaining EDMD-related genes, it was initially reported as a cause of isolated arrhythmogenic right ventricular cardiomyopathy (Merner et al., 2008). Additional research provided the topological and functional characterization of LUMA, depicting a large hydrophilic and four transmembrane domains. Pathogenic variants causing EDMD are specifically clustered in this hydrophilic domain. LUMA structurally interacts with lamins and emerin being involved in the organization of the nuclear membrane (Bengtsson & Otto, 2008). Overall, two main hypotheses have been drawn to explain how defects in genes encoding ubiquitously expressed nuclear envelope proteins give rise to the muscle-specific changes seen in EDMD. The first considers the disruption of the inner nuclear membrane complexes and the nuclear lamina, causing changes in gene expression because of the disorganization in nuclear heterochromatin (Zhang et al., 2007). The second proposes that the mechanical strength of the nucleus is entirely compromised by the weakened when the nuclear lamina is, contributing leading to structural and adaptive signaling defects in tissues, such as skeletal and cardiac muscle, that are subject to intense mechanical stress (Zhang et al., 2007).

Facioscapulohumeral muscular dystrophy

Facioscapulohumeral muscular dystrophy (FSHD) is widely recognized as one of the most common forms of MD, with an estimated prevalence of approximately 1:15,000–1:20,000 (Tawil et al., 2015). Despite a distinctive pattern of muscle involvement which would increase its diagnostic awareness, there is a wide spectrum of disease severity and the involvement of a peculiar mutational mechanism, entailing specific strategies for its detection. Thus, FSHD prevalence is probably underestimated as it may be under-diagnosed. Accumulated clinical data and findings arising with the advent of molecular genetics demonstrated wide phenotypic expressions associated with FSHD, ranging from nearly asymptomatic individuals to more severe phenotypes where independent locomotion is lost (Nikolic et al., 2016). The classical FSHD phenotype, in accordance with the initial description made by Louis Landouzy and Joseph Dejerine around 1884, is characterized by an adult disease onset progressive muscle weakness and atrophy mainly affecting the facial, scapular and humeral

muscles (Nikolic et al., 2016). Within the full FSHD spectrum disease onset varies from infancy to late adulthood, with most cases starting between late teens to the early 20s (Tawil, 2018). The disease onset of FSHD has a particular pattern of muscle weakness starting in the face and the muscles around scapular girdle and then, as the disease progresses, towards upper arms, trunk, and lower-extremities muscles (Tawil, 2018). Asymmetric involvement is another diagnostically important clinical feature, which is not exclusive to FSHD but is rarely encountered in other subtypes of MD (Sacconi et al., 2015). Besides the skeletal muscle being affected, there are extra-muscular manifestations namely sensorineural hearing loss, retinal telangiectasia, respiratory compromise, cardiomyopathy and dysphagia (Sacconi et al., 2015). CNS involvement is usually not described associated to the FSHD clinical picture, except for early-onset cases reported as having intellectual disability and epilepsy (Sacconi et al., 2015). The diagnosis of FSHD is initial mainly clinical since muscle biopsy is often not very helpful as different aspects may be revealed, including muscular dystrophy, an inflammatory pattern like myositis, or even normal findings (Chelly & Desguerre, 2013). The molecular genetic basis of FSHD is quite atypical. The inheritance pattern in the majority of FSHD patients is autosomal dominant, with as many as one third the cases explainable by *de novo* mutations (Tawil, 2018). A small of a subset of patients (~5%) have FSHD type 2, a clinically undistinguishable subtype of the disease, but with a more complex Inheritance involving two *loci* (digenic) (Tawil et al., 2015). The genetic cause of FSHD was mapped to chromosome 4q35 (Wijmenga et al., 1991) and later associated with the loss of a repetitive sequence subtelomeric macrosatellite repeat known as D4Z4 having a total size of 3.3Kb (Tawil et al., 2015). The double homeobox 4 (*DUX4*) gene was identified within this repeat (Ding et al., 1998). *DUX4* is a transcription factor which is normally expressed only in germline cells or in somatic cells during early development (Sacconi et al., 2015). Thus, in normal somatic cells, *DUX4* is not expressed as the repeats are confined in a closed chromatin conformation (heterochromatin) (Hewitt et al., 1994). Normal individuals in the population have 11 to 100 of D4Z4 repeats, whereas in FSHD patients there are only 1-10 of such repeats, so there is a contraction of the number of repeats. These contractions are only pathogenic when located in a permissive genetic background – the A allele - one of two present in 4q35 regions (4aA or 4qB) (Tawil et al., 2015). Thus, the pathogenesis is associated with this contraction in the number of repeats, which leads to a more open chromatin structure, allowing gene expression to occur within these repeats (Tawil, 2018). FSHD type 1 is caused by monoallelic contractions in 'A' background that contains a polyadenylation site sequence which is critical for stabilizing *DUX4* mRNA (Lemmers et al., 2010). In contrast, type 2 FSHD patients do not harbor D4Z4 contractions but are associated with pathogenic variants in another *locus* (*SMCHD1)* in chromosome 18, that originates the phenotype in conjunction with a "permissive" 4qA allele,

consistent a digenic inheritance (Lemmers et al., 2012). *SMCHD1* is known to be one of the genes required for hypomethylation and repression of *DUX4* (Sacconi et al., 2015). So, the accepted model is that both FSHD1 and FSHD2 have an identical molecular basis: the abnormal expression of the *DUX4* gene in skeletal muscle. This event influences other genes that are normally expressed only in germline cells, which in turn may alternatively use promoters for other genes causing the production of non-physiological transcripts, long non-coding RNAs, or antisense transcripts (Sacconi et al., 2015). These changes in the genetic expression profile are likely to be involved in the downstream cascade of cellular events leading to muscle degeneration, muscle regeneration, and activation of an immune response in FSHD patients (Sacconi et al., 2015; Tawil et al., 2015). Genotype-phenotype correlation studies depicted an inverted correlation between the number of D4Z4 repeats and the severity of FSHD phenotype, compatible with the observation that alleles with very reduced size (1–3 D4Z4 repeats) are within with the most severe spectrum of the disease (Nikolic et al., 2016). The molecular diagnosis of FSHD can be technically challenging, as it may involve different molecular approaches ranging from: i) Southern-blot, where DNA is digested with EcoRI and BlnI enzymes and resolved on a linear gel electrophoresis followed by detection using a specific probe (p13E-11) to estimate the number of D4Z4 repeats; ii) haplotyping and pulsed-field gel electrophoresis are also used to differentiate between 4qA and 4qB alleles; iii) hypomethylation screening plus *SMCHD1* sequencing to detect sequencing variants involved in FSHD2 (Lemmers et al., 2012; Larsen et al., 2015).

1.2.4 MYOFIBRILLAR AND DISTAL MYOPATHIES

Despite being classified as two distinct groups, myofibrillar myopathies and distal myopathies, share some common aspects such as later disease onset, predominantly distal muscle weakness and myopathic changes in muscle biopsy ranging from myofibrillar changes to autophagic vacuolar pathology (rimmed vacuoles).

Myofibrillar myopathies

The unifying characteristic of myofibrillar myopathies (MFM) is the presence of distinctive histopathology of abnormal protein aggregations and a significant myofibrillar disintegration in muscle biopsies of these patients (reviewed by Palmio & Udd, 2016). Myofibril breakdown initiates in the proximity of the Z-disk leading to its streaming, and the accumulation of degraded filamentous material in various patterns around nuclei and under the sarcolemma (Fichna et al., 2018). This material consists of a variety of proteins such as desmin, α-B-crystallin, dystrophin, myotilin, sarcoglycans, neural cell adhesion molecule, plectin, gelsolin, ubiquitin or filamin C and TAR DNA-binding protein 43 (TDP-43) (Selcen, 2011; Béhin et al., 2015). The characteristic MFM histological abnormalities see on muscle histology comprise: i) amorphous granular or hyaline deposition visible on trichrome-staining; ii) altered oxidative enzyme activities translated as core-like or faded-out fibers; presence of rimmed and non-rimmed vacuoles, (myotilinopathies and zaspopathies); and more unspecific myopathic features such as fiber-size variability (Batonnet-Pichone et al., 2017). Other features found in myofibrillar pathology are also shared by other muscle diseases, including LGMD, distal myopathy, scapuloperoneal syndrome or rigid spine syndrome (Selcen, 2011). MFM mainly comprises AD forms, but AR and XLR have been also associated, with onset during adulthood and variable muscle weakness manifestation, from the most typical distal presentation, but also proximal, proximo-distal or even generalized (Fichna et al., 2018). As for CK levels in MFM patients, it can be mild to moderate elevated, but normal values can also be found (Palmio & Udd, 2016). Instead of the progressive muscle weakness as the initial clinical sign of MFM, in some cases, the cardiomyopathy may precede the skeletal muscle presentation. Nonetheless, MFM patients can also develop cardiomyopathy and respiratory insufficiency later in life as the disease progresses (Palmio & Udd, 2016). EMG studies in MFM, reveal mostly myopathic features and abnormal electrical irritability often with myotonic discharges (Selcen, 2011). A small subset of patients has a combination of myopathic and neurogenic motor unit potentials or slow nerve conduction velocities, making peripheral neuropathy an associated feature (Selcen, 2011). Table 1.2.4 lists the currently known genetic causes of MFM, which includes 11 different *loci*. However, there is no absolute consensus towards as other authors propose a different assortment of a *locus*

involved in MFM (Kley et al., 2016). Despite this, the universally accepted genes encode for proteins that either located in or are closely associated with the Z-disk, which is important in maintaining the sarcomere structural integrity. This group includes myotilin, titin, LIM domain binding 3 (ZASP), filamin C, BCL2-associated athanogene 3, desmin and α-B-crystallin. MFM patients with *TTN* variants in particular generally present dyspnea in the earliest stage of disease, which lead the designation of hereditary myopathy with early respiratory failure (HMERF) (Pfeffer et al., 2012). The awareness of this clinical entity has grown considerably over the last few years, as several cases have been identified worldwide including by our group. Notwithstanding the complexity of analyzing such a large gene as *TTN*, the arrangement of cytoplasmic bodies in a necklace fashion is suggestive of HMERF (Uruha et al., 2015), and there is also a mutational hot-spot in exon 344 that encodes a disease-associated domain (Hedberg et al., 2014). Further pathophysiological mechanisms in MFM, not directly involved with this sarcomeric region, have also been proposed (Table 1.2.4). One consists of chaperones proteins associated to the Z-disk, such as α-B-crystallin a member of the small heat shock protein family, that in muscle localizes in the Z-disk (Fichna et al., 2018). It was proposed to enable proper folding of the muscle proteins such as actin and desmin filaments, allowing normal myofibril organization and preventing abnormal aggregation (Fichna et al., 2018). Genetic defects in the oxidoreductase *PYROXD1* gene were linked to an early-onset myopathy with distinctive histopathology (frequent internalized nuclei, myofibrillar disorganization, desmin-positive inclusions, and thickened Z-band), but more importantly adding the altered redox regulation to the portfolio of the primary causes of muscle disease (O'Grady et al., 2016).

Distal myopathies

Distal myopathies (DMYO) are a group of diseases in which muscle weakness is selectively or predominantly distal (Milone & Liewluck, 2018). Accordingly, in DMYO the upper limbs, lower limbs or both are primarily affected, resulting in a suggestive pattern of muscle weakness and atrophy. As the disease follow its natural course proximal musculature can be affected as well, but even in a more advanced state, the weakness is usually more prominent in the feet and/or hands (Palmio & Udd, 2016). The disease onset typically occurs during adulthood, but in some cases during childhood. Also, a wide variety of muscle pathological changes have been documented in DMYO, from nonspecific myopathic or dystrophic features to rimmed or non-rimmed vacuoles often without significant protein aggregation (Milone & Liewluck, 2018). The levels of CK are normal to mildly elevated, with exception of few subtypes. Genetic research in DMYO, essentially over the last decade, allowed the identification of 16 different *loci* (Table 1.2.4).

Table 1.2.4- Genes and phenotypes associated with myofibrillar and distal myopathies.

Subtype (phenotype OMIM)	Gene	Nr. Exons	Chromosomal localization	Protein	Inheritance pattern	Reference (phenotype-gene association)
I. Myofibrillar myopathy (MFM), childhood to adult-onset slowly progressive myopathy with proximal-distal muscle weakness.						
MFM pathophysiological mechanism I: Contractile dysfunction of the thin filament (Z-disk) due to the instability of its components.						
Myopathy, myofibrillar, 1 (#601419)	*DES*	9	2q35	Desmin	AD, AR	Goldfarb et al., 1998
Hereditary myopathy with early respiratory failure (#603689)	*TTN*	363	2q31	Titin	AD	Lange et al., 2005; Pfeffer et al., 2012
Myopathy, myofibrillar, 3 (#609200)	*MYOT*	11	5q31	Myotilin	AD	Selcen & Engel, 2004
Myopathy, myofibrillar, 4 (#609452)	*LDB3*	14	10q22	LIM domain binding 3	AD	Selcen & Engel, 2005
Myopathy, myofibrillar, 5 (#609524)	*FLNC*	48	7q32	Filamin C	AD	Vorgerd et al., 2005
Myopathy, myofibrillar, 6 (#612954)	*BAG3*	4	10q25.2-q26.2	BCL2-associated athanogene 3	AD	Selcen et al., 2009
Myopathy, myofibrillar, 7 (#617114)	*KY*	11	3q22.2	Kyphoscoliosis peptidase	AD	Straussberg et al., 2016
Phenotype not in OMIM	*ACTA1*	7	1q42.13	Actin, α-skeletal muscle	AD	Selcen, 2015; Liewluck et al., 2017
MFM pathophysiological mechanism II: Chaperone protein of desmin & actin. Possible through a dominant negative effect affects its function in protein quality control and degradation.						
Myopathy, myofibrillar, 2 (#608810)	*CRYAB*	3	11q22.3-q23.1	Crystallin, alpha B	AD	Selcen & Engel, 2003
MFM pathophysiological mechanism III: Link between myofibrils and sarcolemma, nucleus & mitochondria is lost. Structural proteins that maintain the integrity of sarcomeres (and consequently myofibrils) are compromised.						
Reducing body myopathy, X-linked 1b (#300718)	*FHL1*	8	Xq26.3	Four and a half LIM domain 1	XLD/XLR	Schessl et al., 2008
MFM pathophysiological mechanism IV: Poorly characterized, altered redox regulation possibly affected.						
Myopathy, myofibrillar, 8 (#617258)	*PYROXD1*	12	12p12.1	Pyridine nucleotide-disulfide oxidoreductase domain 1	AR	O'Grady et al., 2016

Table 1.2.4- Genes and phenotypes associated with myofibrillar and distal myopathies (continues).

II. Distal myopathies (DMYO): Slowly progressive myopathy with adult or late adult-onset, symptoms starting in distal muscle, mainly in feet or hands.						
DMYO pathophysiological mechanism I: Contractile dysfunction of the thin filament (Z-disk) due to the instability of its components.						
Tibial muscular dystrophy, tardive (#600334)	*TTN*	364	2q31	Titin	AD	Hackman et al., 2002
Distal myopathy with a myotilin defect (phenotype not in OMIM)	*MYOT*	11	5q31	Myotilin	AD	Pénisson-Besnier et al., 2006
Distal myopathy with nebulin defect (phenotype not in OMIM)	*NEB*	183	2q22	Nebulin	AR	Wallgren-Pettersson et al., 2007
Late-onset distal myopathy (Markesbery-Griggs, a phenotype not in OMIM)	*LDB3*	14	10q22	LIM domain binding 3	AD	Griggs et al., 2007
Myopathy, distal, 4 (#614065)	*FLNC*	48	7q32	Filamin C	AD	Duff et al., 2011
DMYO pathophysiological mechanism II: Defective sarcolemma repair or disruption of caveolae formation.						
Miyoshi muscular dystrophy 1 (#254130)	*DYSF*	56	2p12-14	Dysferlin	AR	Liu et al., 1998
Myopathy, distal, Tateyama type (#614321)	*CAV3*	3	3p25	Caveolin 3	AD	Tateyama et al., 2002
Miyoshi muscular dystrophy 3 (#613319)	*ANO5*	22	11p14-12	Anoctamin 5	AR	Bolduc et al., 2010
DMYO pathophysiological mechanism III: The enzyme that catalyzes the first two steps of N-acetylneuraminic (sialic) acid synthesis is defective. Abnormal O-glycan sialylation of muscle proteins.						
Nonaka myopathy (#605820)	*GNE*	12	9p13.3	UDP-N-acetylglucosamine-2- epimerase/N-acetylmannosamine kinase	AR	Eisenberg et al., 2001
DMYO pathophysiological mechanism IV: Disrupts ability of myosin tail to establish a normal coiled-coil structure, leading to an altered structure of thick filament.						
Laing distal myopathy (#160500)	*MYH7*	40	14q12	Myosin, Heavy chain 7, cardiac muscle, beta	AD	Meredith et al., 2004

.

Table 1.2.4- Genes and phenotypes associated with myofibrillar and distal myopathies (continues).

DMYO pathophysiological mechanism V: Dysregulation of RNA metabolism, cytoplasmic mislocalization of RNA binding proteins. Dysfunction in stress granule dynamics of RBPs and increased tendency to form aggregate.					
Distal myopathy with matrin 3 defects (phenotype not in OMIM)	*MATR3*	18	5q31	Matrin 3	AD	Senderek et al., 2009
Welander distal myopathy (#604454)	*TIA1*	13	2p13	Cytotoxic granule associated RNA binding protein	AD	Hackman et al., 2013
DMYO pathophysiological mechanism VI: Increased levels of ubiquitinated cell proteins (sensitization to proteasome stress). Impaired degradation of proteins by endoplasmic reticulum-associated degradation & formation of aggregates promoted. Impaired protein aggregate clearance.					
Distal myopathy with valosin-containing protein defects (phenotype not in OMIM)	*VCP*	17	9p13-p12	Valosin-containing protein	AD	Palmio et al., 2011
Other distal myopathies:					
Distal myopathy with dynamin 2 defects (phenotype not in OMIM)	*DNM2*	21	19p13.2	Dynamin 2	AD	Fischer et al., 2006
Distal myopathy with kelch-like 9 defects (phenotype not in OMIM)	*KLHL9*	1	9p21.3	Kelch-like 9	AD	Cirak et al., 2010
Myopathy, distal, 5 (#617030)	*ADSSL1*	13	14q32.33	Adenylosuccinate synthase-like 1	AR	Park et al., 2016

Footnote: AD, autosomal dominant; AR- autosomal recessive; Nr.- number; XLR- recessive, or XLD- dominant, linked to X chromosome; data used to generate this table was obtained from the Gene Table of Neuromuscular Disorders (Bonne & Rivier, 2018), OMIM (2018) and cited literature.

However, as there is some degree of overlap with other myopathies (MFM and LGMD) some authors adopted broader classifications that include up to 20 or 24 different *loci* (Palmio & Udd, 2016; Milone & Liewluck, 2018). Most of the genetic entities within DMYO (n=12) have an autosomal dominant inheritance pattern or sporadic if associated with *de novo* occurring variants, being the remaining cases AR and only one linked with the X-chromosome. The most striking aspect is the diversity of pathophysiological mechanisms involved in DMYO, which are in favor of multiple genetic defects manifesting as predominant distal muscle weakness. The most extensive group, that includes five different molecular members, is related to contractile dysfunction of the thin filament (z-disk) due to the instability of its

components. In such group, variants in nebulin are the only giving rise to an early (childhood) onset phenotype and associated with an autosomal inheritance pattern (Wallgren-Pettersson et al., 2007). In *NEB*-related DMYO muscle weakness is predominant in lower legs and fingers extensors and also neck flexors (Wallgren-Pettersson et al., 2007). As for the tibial muscular dystrophy (*TTN*), distal myopathy with a myotilin defect (*MYOT*) and Markesbery-Griggs myopathy (*LDB3*) all sharing the fact that disease onset occurs during late adulthood (late adult-onset) and rimmed vacuoles are relatively frequent in the muscle biopsy of these patients. Whereas in the phenotype associated with defects in filamin-C with an early adult-onset, rimmed vacuoles are not among the findings in muscle pathology. Another muscle contraction dysfunction in DMYO is associated with defects in one myosin heavy chain gene (*MYH7*), encoding the main isoform in slow/ β-cardiac and the predominant in slow-twitch type I fibers (Feinstein-Linial et al., 2016). Being the major motor protein in muscle, myosin has a unique structure consisting of an amino-terminal motor/head domain and a carboxy-terminal tail or rod domain. The later allows the myosin assembly into the thick filament. Besides causing a CM and cardiomyopathy, pathogenic variants in *MYH7* also give rise to Laing distal myopathy (Meredith et al., 2004). Such DMYO phenotype as early onset (often during childhood) and displays muscle fiber-size variation (with type fibers I hypotrophy), expression of both slow and fast myosin and signs of mild necrosis and regeneration in muscle (Feinstein-Linial et al., 2016). An additional link between DMYO and CM (CNM in particular) is associated with defects in dynamin 2 (*DNM2*) gene. The disease presents during adulthood, with distal upper and lower limbs weakness, and normal CK levels (Fischer et al., 2006). The second main mechanism causing DMYO is associated with a defective sarcolemma's repair process or the disruption of caveolae formation. Here three genes, all commonly shared with LGMD, have been identified: *DYSF* (Miyoshi muscular dystrophy 1; Liu et al., 1998), *ANO5* (Miyoshi muscular dystrophy 3; Bolduc et al., 2010) and *CAV3* (Tateyama distal myopathy; Tateyama et al., 2002). Patients with pathogenic variants in these genes have very high CK levels and scattered necrotic fibers in muscle biopsies (Palmio & Udd, 2016). One further mechanism causing DMYO is related with defects in genes that encode for RNA binding proteins (RBP), namely *MATR3* (Senderek et al., 2009) and *TIA1* (Hackman et al., 2013). Such proteins were identified as a potential component of the cytoplasmic stress granule proteome (Mensch et al., 2018). Molecular defects cause dysregulation of RNA metabolism and dysfunction in stress granule dynamics of RBP and higher predisposition to form aggregates (Zhao et al., 2018). In both, *MATR3*- and *TIA*-related DMYO there is late adult-onset and rimmed vacuoles are also often found. Pathogenic variants in valosin-containing protein (*VCP*) gene are associated to a wide spectrum of neurological diseases, namely inclusion body myopathy with Paget disease of bone and/or frontotemporal dementia, familial amyotrophic lateral sclerosis and a distal

myopathy phenotype (Boland-Freitas et al., 2016). In the myopathic presentation, there is also late disease-onset with rimmed vacuolar muscle pathology, and a pivotal diagnostic clue is the occurrence, in some cases, of late-occurring dementia (Boland-Freitas et al., 2016). One additional cause of DMYO is related to changes in ubiquitin-dependent protein degradation, due to defects in kelch-like 9 (*KLHL9*) gene (Cirak et al., 2010). Patients with pathogenic variants in *KLHL9* present with slowly progressive muscle weakness and atrophy with onset between childhood to adolescence, involving primarily the anterior tibial muscles followed by atrophy of intrinsic hand muscles (Cirak et al., 2010). Serum creatinine kinase levels were mildly increased, and non-specific changes are found in muscle biopsies of these patients (Cirak et al., 2010). The pathophysiological diversity among DMYO also includes two distinct enzymatic deficiencies. The first is related to UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase deficiency, caused by pathogenic variants in the *GNE* gene (Eisenberg et al., 2001). The activity of this enzyme is rate-limiting for the biosynthesis of sialic acid, thus essential for sialylation of a variety of glycoproteins (Eisenberg et al., 2001). The clinical suspicion of GNE-related myopathy (also known as Nonaka myopathy) should be raised in young adults presenting with bilateral foot drop, with histopathologic findings on muscle biopsies include fiber-size variation and rimmed vacuoles (Carrillo et al., 2018). Finally, alterations in adenylosuccinate synthase-like, a muscle-specific synthase that catalyzes the initial step of the conversion of inosine monophosphate to adenosine monophosphate (Sun et al., 2005), can also give rise to an adolescent-onset DMYO (Park et al., 2016). Typically, the clinical picture includes distal and facial muscle weakness which progresses to quadriceps during the third decade of life, also rimmed vacuoles in muscle, and mild elevation of CK levels.

1.3 NEXT-GENERATION SEQUENCING IN CLINICAL GENETICS

1.3.1 GENETIC ANALYSIS IN THE PRE-NEXT-GENERATION SEQUENCING ERA

The impact of genetics in human health has long been recognized since the first identification of a genetic disorder, the Down syndrome, nearly sixty years ago and through the analysis of the human karyotype (Lejeune et al., 1959). Since then the field of clinical genetics has grown immensely. The main purposes of genetics in a health care setting consists in providing the most accurate and definitive diagnoses, predicting disease outcome (prognosis), assist the choice of the best standards of care and therapeutic interventions available. In addition, it enables genetics prevention actions through carrier screening, preimplantation or prenatal genetic diagnosis, allowing the best reproductive options to those affected families. During these last four decades considerable technological developments in DNA (and RNA) sequencing were generated The pioneer works from Paul Berg (1972), Frederick Sanger (1975) and Walter Gilbert (1977), that jointly won the Nobel Prize in Chemistry in 1980 (Kolata et al., 1980), made possible the accomplishment of several progresses in the field which contributed to the development of Sanger sequencing. The next additional breakthrough was the identification of the entire map of the Human Genome. To tackle just enormous task an International Consortium - Human Genome Project - was established. The first draft of the human genome was reported in February 2001 (Venter et al., 2001) and its more definitive map was published two years later. Empowered by this enormous international endeavor, there were considerable developments in sequencing technology, namely through automation and the computational resources made available. These advances unlocked a new avenue for clinical genetics, by providing the possibility to study a wide variety of monogenic diseases through the screening of variants even at a single-base resolution. Since the first commercially available automated sequencer, the Prism 373 from Applied Biosystems (ABI), several sequencing systems have been successfully launched with increased throughput, automation, and easier operation. Even today, Sanger sequencing is still used to confirm a suspected diagnosis, especially in genetically homogeneous diseases. However, when diseases involve large *loci* or in genetic heterogeneous conditions, where dozens of genes need to be studied, the analysis through conventional sequencing is no longer the most efficient approach.

The portfolio of molecular techniques available for genetic analysis is not confined to genomic sequencing. Several complementing approaches are necessary to uncover the full diversity of mutational events and mutation types associated with human genetic diseases. In summary, these could be classified as qualitative and quantitative approaches. To detect and

determine dynamic expansions (or contractions) of repetitive elements, Southern-blot or PCR-based techniques, such as fluorescent primer PCR (FP-PCR) or triplet repeat primed (TP-PCR) are often required. As for detecting large genomic copy number variants (CNVs), such as large heterozygous deletions or duplications, quantitative approaches such as qPCR (real-time or quantitative PCR), MLPA or microarrays (such as array-CGH or SNP array) can be used to accurately determine the number of copies present in one or more genomic segments. Genetic studies at the RNA level are useful to determine the effect of splice-site affecting variants or even to detect deep intronic variants. As for large genomic rearrangement, their characterization can be further delineated (to determine their breakpoints) by long-range PCR and Southern-blot techniques.

1.3.2 NEXT-GENERATION SEQUENCING TECHNOLOGY

As mentioned, over the last two decades decisive developments in nanotechnology and informatics, contributed to new sequencing methods, aiming to complement and eventually even replace Sanger sequencing. This technology is collectively referred to as next-generation sequencing (NGS) or massively parallel sequencing (MPS), often an umbrella to designate a wide diversity of approaches. With this technology it is possible to generate a massive amount of data per instrument run, in a faster and cost-effective way, streaming parallel analysis of several genes. Currently, NGS is being made available by four main brands: Illumina, Ion Torrent (Thermo Fischer Scientific), PacBio and Oxford Nanopore Technologies. All presenting completely different strategies towards the same problem, that is sequencing data massification. For simplicity, although a not fully consensual classification among literature, one can consider that second-generation sequencing is based on massive parallel and clonal (PCR) amplification of molecules; whereas, the third-generation sequencing rely on single-molecule sequencing without a prior clonal amplification step (Schadt et al., 2010; Niedringhaus et al., 2011).

Second-generation sequencing platforms

Second-generation platforms belong to the group of cyclic-array sequencing technologies (reviewed by Shendure et al., 2008). These have in common that PCR products (resulting from clonal expansion) end up spatially clustered, either to a single location on a planar substrate (e.g. by bridge PCR in Illumina system), or to the surface of micron-scale beads (e.g. by emulsion PCR in Ion Torrent), which can be recovered and arrayed subsequently. The sequencing process consists of sequencing-by-synthesis (SBS), alternating cycles of enzyme-driven biochemistry and imaging-based data acquisition (Illumina) or flows of the

different nucleotides and detection of pH fluctuation due to the release of protons when incorporation of nucleotides occurs (Ion Torrent). Other second-generation platforms included the SOLiD platform (Life technologies), which sequencing methodology is based on sequential ligation with dye-labeled oligonucleotides (Pandey et al., 2008); and the Polonator G.007 (Dover/Harvard) is an automated, genome sequencing machine that is based on polony sequencing technology (Edwards et al., 2008).

The first commercially available NGS platform was the GS20, launched in 2005 by 454 Life Sciences (later acquired by Roche). This platform was based on an emulsion method for DNA amplification and an instrument for SBS using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes (Margulies et al., 2005). In 2006 emerged the first Solexa (later acquired by Illumina) sequencer, the Genome Analyzer that revolutionized DNA sequencing, reaching gigabases per run (Illumina, 2018a). The basic process of Illumina sequencing is based on capture (on a solid surface) of individual molecules, followed by bridge PCR allowing its amplification into small clusters of identical molecules. Illumina adopted the concept of SBS, in which during each sequencing cycle, a single labeled deoxynucleotide triphosphate (dNTP) is added to the template on the flow cell. The labeled nucleotide serves as a "reversible terminator". These fluorescently labeled reversible terminators make that one single base incorporation on each molecule temporarily stoping the reaction. Then, the fluorescent dye is identified through laser excitation and imaging, then enzymatically cleaved to allow the next round of incorporation. To minimize incorporation bias, the possible four reversible terminator-bound dNTPs (A, C, T, G) are present. Base calls, the process of converting the raw data captured by the sequencing machine into sequence reads, are made directly from signal intensity measurements during each cycle, which results in highly accurate base-by-base sequencing (Bentley et al., 2008). Although highly accurate and robust, this technology presents some limitations. For example, during the sequencing process, some dyes may lose activity by isomerizing while being in an excited state and may result in sequencing errors. One concerns with the partial overlap between emission spectra of the fluorophores (particularly evident in A and C nucleotides and also as for G and T pair), consequently their separation is not completely effective by optical filters which limit the base calling on the Illumina platform (Fuller et al., 2009; Kircher et al., 2011). The second problematic aspect occurs during the sequencing process with the defective extension of the fluorescent signal known as phasing and pre-phasing. Phasing is caused by incomplete removal of the 3' terminators and fluorophores, as well as sequences in the cluster that missed an incorporation cycle. Whereas, pre-phasing derives from the incorporation of nucleotides without an effective 3'-blocking step (Kircher et al., 2011). Over many cycles, these errors will accumulate and decrease the overall signal to noise ratio per

single cluster, causing a decrease in quality towards the ends of the reads (Kircher et al., 2011).

Contrasting with Illumina and 454 sequencers, Ion Torrent developed a new sequencing concept that does not employ optical signals, known as semi-conductor sequencing. The method of DNA sequencing in Ion Torrent is based on pH changes caused by the release of hydrogen ions (H^+) during the polymerization of DNA. The DNA sequencing is done a massively parallel sensor array on a complementary metal–oxide–semiconductor (CMOS) chip (Rothberg et al., 2011). This chip contains millions of wells that individually harbor a single sequencing reaction and allows the detection of chemical changes that will be converted into a digital signal (Rothberg et al., 2011). After library and template preparation, each bead containing multiple copies of a DNA fragment is randomly deposited into a well. Just before sequencing, the chip wells are filled with a reaction solution that includes DNA polymerase and the nucleotides (dNTPs). There are different flows of each different dNTP at a time (4 flows correspond to a cycle). Every time that the polymerase incorporates the complementary nucleotide into the elongating DNA strand, a proton is released leading to a shift in the pH of the solution, which is detected by an ion sensor incorporated on the chip. This sensor converts this pH fluctuation into a voltage signal thus confirming that the nucleotide flooded into the chip was in fact incorporated. This process occurs in all wells simultaneously allowing the massive parallel sequencing (Rothberg al., 2011). To transform raw voltages into base calls, i.e. sequence reads, signal-processing software converts this data into measurements of incorporation in each well for each successive nucleotide flow using a physical model. The model considers diffusion rates, buffering effects and polymerase rates to determine the sequence for which the expected incorporation signal best matches the observed incorporation signal (Rothberg et al., 2011). Once the most likely sequence is found, any remaining difference between expected and observed signals influences the assignment of quality values. Ion Torrent sequencer requires a Torrent Server (Torrent Suite Software) to perform the base calling and alignment. There are also limitations concerning Ion Torrent technology. The best well-known issue is related to homopolymeric template stretches. During the sequencing of such specific templates, multiple incorporations of the same base on each strand will occur, generating the release of a higher concentration of H^+ in a single flow. The increased shift in pH generates a greater incorporation signal, indicating to the system that more than one nucleotide was incorporated in that flow. However, for longer homopolymers there are sensitivity issues making their quantification inaccurate, that is the number the number of repetitive nucleotides is incorrectly established, constituting as the major source of sequencing errors in this platform (Merriman et al., 2012; Quail et al., 2012).

In 2012, Quail and co-workers, compared three Illumina platforms (HiSeq, GAIIx, and MiSeq) with the Ion Torrent Personal Genome Machine (PGM). Illumina showed lower raw error rate and read length and higher yield, accuracy and higher running time. However, over the last years, several improvements in Ion Torrent algorithms have been proposed to correct sequencing errors for PGM data and more recently reports are suggestive of improvements in Ion Torrent technologies, namely with the release of the newer Ion Torrent S5 XL and the new Hi-Q chemistry (Pereira et al., 2016; Song et al., 2017), with an accuracy at least comparable with Illumina MiSeq platform (Shin et al., 2017).

Third-generation sequencers

The 3rd-generation NGS technology brought the possibility of circumventing common and transversal limitations of PCR, such as nucleotide misincorporation by a polymerase, chimera formation and allelic drop-outs (preferential amplification of one allele) causing an artificial homozygosity call. The platform Helicos Genetic Analysis System was considered the first commercial 3rd-generation sequencer. To address the principle of single-molecule fluorescent sequencing, in that system the DNA was simply sheared, tailed with poly A, and hybridized to a flow cell surface containing oligo-T nucleotides. Then, the flow cell is loaded onto the HeliScope™ Sequencing System. A laser illuminates the surface of the flow cell and images of all labeled templates are collected by a charge-coupled device camera. A DNA polymerase (with a fluorescent label nucleotide) is added and the sequencing reaction begins by the incorporation of the labeled nucleotide. A series of incorporation, image and wash steps continues will all nucleotides. Every strand is thus, uniquely and independently sequenced (Thompson et al., 2010).

In 2010, Pacific Biosystems developed an instrument (PacBio RS II) capable of sequencing individual DNA molecules in real-time, introducing the concept of single-molecule real-time (SMRT) sequencing (McCarthy, 2010). Individual DNA polymerases are attached to the Zero-Mode Waveguide (ZMW) wells, which are nanoholes where a single molecule of the DNA polymerase enzyme can be directly placed (McCarthy, 2010). A single DNA molecule is used as a template for the incorporation of fluorescently labeled nucleotides through polymerase activity. Each base has a different fluorescent dye, thereby emitting a signal out of the ZMW (McCarthy, 2010). A detector reads the fluorescent signal and names the base according to the detected signal. Once the base is added, the fluorescent tag is cleaved by the polymerase (McCarthy, 2010). Given the circular nature of the SMRTbell DNA template, the polymerase sequences the same DNA molecule several times through multiple passes. A circular consensus sequencing method is used in PacBio technology, that enables long reads with high accuracy at the single-molecule level. The machine PacBio RS II, and the recently introduced Sequel system can sequence single DNA molecules in real-time without

means of amplification. Besides avoiding amplification step, SMRT has other advantages, namely long read lengths (with average read lengths up to 30 kb), enabling *de novo* assembly and direct detection of haplotypes; high consensus accuracy; simultaneous epigenetic characterization (direct detection of DNA base modifications at one-base resolution) (Schadt et al., 2010; Nakano et al., 2017).

In 2014, a new 3rd-generation sequencer was released under the name MinION, from Oxford Nanopore Technology. This sequencer offers a direct, electrical detection of single DNA molecules. Nanopore sequencing uses electrophoresis to transport an unknown sample through a small orifice. An ionic current pass through nanopores and the current is changed as the bases pass through the pore in different combinations. The information about the change in current is used to identify those molecules and perform the sequencing (Jain et al., 2015). Although obeyed to the promise of longer read lengths (> 800 Kb), it generates higher sequencing error rates (with accuracies ranging from 65 to 88%) as compared with PacBio and NGS sequencers (Mikheyev & Tin, 2014; Rhoads & Au, 2015). Longer reads are critical to span repetitive elements and other complex sequences to generate high-quality, contiguous assemblies, as well as, solving the problematic repetitive sequences, transposable elements, segmental duplications, and telomeric/centromeric regions that are difficult address with short reads (Goodwin et al., 2015). Given its importance, efforts are been made to improve the quality of the reads and reduce the error rate of 3rd-generation sequencers (Goodwin et al., 2015).

1.3.3 NGS WORKFLOW

Prior to implementing an NGS approach, it is crucial to clearly define the biological question to be answered. Choosing the right sample preparation workflow and the sequencing technology itself is highly dependent upon this definition. The typical scenario in human clinical genetics and in the context of rare diseases consists in the analysis of specific DNA segments to identify for pathogenic variants that can explain a particular phenotype/disease. As already mentioned, a wide variety of genetic defects should be considered. These include single nucleotide variants (SNV), small insertions or deletions (INDELs) and CNVs. In these cases, DNA is used as template, and the target could consist in a single *locus*, but often a set of genes (gene panels), or to investigate the entire coding regions of the genome and its boundaries through whole-exome sequencing (WES) and even the entire genome (whole-genome sequencing, WGS). For certain diseases or more complex cases, it is also important to study changes in gene expression and transcriptome sequence analysis (RNA analysis), collectively known as RNA-seq (Wang et al., 2009). NGS also allows more complex studies, such as the study of protein DNA interactions, known as

ChIP-Seq (Park, 2009), well as, the analysis of DNA patterns of methylation called bisulfite-seq (Li & Tollefsbol, 2011). The subject of NGS applications will be further detailed in section 1.3.5. Regardless of the technology or application, the first step consists of library preparation. An NGS library is defined as a collection of DNA fragments that represent regions of interest for study, which can be either the entire human genome or a small target region. In most NGS experiments using 2nd-generation sequencers, the preparation of a library starts with the fragmentation of nucleic acids. The obtained fragments are then selected according to the required size, which is determined by the sequencing read-length, a parameter which is quite variable across the different platforms. Alternatively, fragments with the desired size from the target region can be generated by highly multiplexed PCR reactions which are called amplicon-based sequencing. The next step consists in binding specific adapters to the fragments or PCR product, using a DNA ligase. Again, these adapters are platform-specific. These short oligonucleotides are necessary to allow subsequent enrichment and latter parallel clonal amplification by a modified PCR (emulsion PCR in a bead or bridge PCR in the surface of the flow cell). For fragmentation-derived libraries, it is possible to proceed without further step if WGS is intended, whereas if the objective is to perform WES or to sequence a set of genes (gene panel) it is necessary to perform targeted enrichment of genomic DNA regions. This enrichment is performed by capture using a solution-based hybridization. It consists of a pool of biotinylated oligonucleotide probes targeting the desired regions that are added to a library of adapter-ligated fragments. The hybridized probes are then captured by streptavidin magnetic beads.

Regardless of the type of NGS library suitable sensitivity and specificity should be ensured. Sensitivity is essentially expressed by bias, which can be defined by a systematic data distortion due to the experimental design (Head et al., 2014). The main objective is to keep bias as low as possible so that all fragments are equally represented in the library. A reduced number of target regions usually means higher bias and lower complexity. In terms of library complexity, the ideal situation is a highly complex library that could reflect with high fidelity the original diversity of the source material (Head et al., 2014). Specificity can be affected by non-random errors that can be introduced during the amplification/enrichment steps or during capture. Certain genomic regions are more prone to enrichment problems due to differences in GC/AT content and existence of repetitive regions, which makes hard to develop a highly sensitive and specific library (Aird et al., 2011). Currently, numerous kits for preparing libraries from DNA or RNA are commercially available for the multiple NGS applications and for a broad range of starting material, from pico- to micrograms of nucleic acids.

Samples requirements

The extraction of nucleic acids is the very first step in all NGS workflows. Although it is often neglected or underscored, it is of utmost importance as the quality of the library greatly depends on the conditions of the starting sample. The extraction method should be selected carefully based on the biological material available and the type of downstream application. The technique must guarantee the minimum presence of contaminants and the maximum quality given that sample quality is crucial for NGS success. Another important aspect, especially for fragmentation-derived libraries, is the integrity of the nucleic acid. Although automated nucleic acid extraction methods provide high throughput, reproducibility, and yield, it is still recommended to perform quality control (QC) to check quantity, purity, and integrity of each sample prior to its processing. The basic starting sample to construct an NGS library is a nucleic acid. These samples, often genomic DNA (gDNA) or RNA, must have high purity and integrity and enough concentration for the sequencing reaction. To determine the integrity of gDNA a first-level and inexpensive QC assessment consist in evaluating the sample by gel electrophoresis, a high molecular weight intact band should be seen without any signs of smearing (a continuum of lower size bands). Regarding RNA this inspection is often more demanding, and it is necessary to determine the RNA integrity number (RIN) especially in quantitative or more comprehensive approaches (Schroeder et al., 2006). The RIN is estimated using an algorithm applied to electrophoretic RNA measurements usually obtained through capillary gel electrophoresis integrated with an automated detection system. To calculate this value several parameters including the ratio of the areas of 28S and 18S ribosomal RNA (rRNA) peaks are used. The RIN ranges from 1 to 10, the higher values correspond to the best level of integrity (Schroeder et al., 2006). Typically, values over 8 are considered suitable to prepare an NGS library. Contaminants such as nucleic acids, proteins or chemicals can interfere with library preparation. Sample purity can be assessed following nucleic acid extraction and throughout the library preparation workflow using UV/Visible spectrophotometry. For DNA and RNA samples the relative abundance of proteins can be assessed by determining the A260/A280 ratio. This ratio should be between 1.8-2.0 for DNA or slightly higher for RNA (2.0-2.2). The A260/A230 ratio can be useful to detect organic compounds contaminates, and ideally must be higher than 2.0 for DNA and in RNA higher than 1.5. To perform these quantifications two of the most used equipment are the fluorometer Qubit and the spectrophotometer NanoDrop (both from Thermo Fisher Scientific). The choice between Qubit and NanoDrop is controversial. Haque and co-workers stated that DNA quantification by spectrophotometry is less biased than fluorometers (Haque et al., 2003). Others consider Qubit more accurate than NanoDrop, which may overestimate the amount of DNA (O'Neill et al., 2011; Kapp et al., 2015). Simbolo and co-workers suggested that the correct QC of DNA samples for NGS may require the

combination of both (Simbolo et al., 2013). They demonstrated that although Qubit shows high accuracy than NanoDrop, it does not enable the detection of impurities, which can affect the overall NGS process. To quantify DNA for amplicon-based libraries, especially those derived from multiplexed reactions (up to 20,000 different targets in a single reaction) qPCR is highly recommended.

DNA fragmentation

Libraries prepared with DNA fragments are commonly used in WGS or *de novo* sequencing (sequencing without considering the reference genome) applications. The first step consists of performing controlled fragmentation of nucleic acids. Fragmentation is typically done either by physical or enzymatic methods. Physical methods include: i) acoustic shearing, performed in dedicated apparatus (such as Covaris, Woburn, MA) that generate focused bursts of ultrasonic acoustic energy to break nucleic acids in smaller fragments ranging from 100-5,000 bp (Head et al., 2014); ii) sonication, through devices (e.g. Bioruptor, Denville, NJ) that uses ultrasonic waves whose vibrations produce gaseous cavitation in the liquid that can break molecules by resonance vibrations (Knierim et al., 2011), small volumes of DNA can be sheared to 150-1000 bp in length; iii) hydrodynamic shear, (e.g. Hydroshear from Digilab, Marlborough, MA) in which molecules are forced to cross through a point-sink such as a small orifice or the bore of a small-diameter tube at high velocity, this causes hydrodynamic shear stress leading to the fragmentation of the molecules (Joneja & Huang et al., 2009). Other alternatives are the g-TUBEs (Covaris), that only requires a benchtop centrifuge, being useful for sequencing applications requiring longer DNA fragments (6 kb – 20 kb range). As for enzymatic methods, it includes digestion by DNase I or dsDNA Fragmentase (New England Biolabs, Ipswich, MA) and other restriction endonucleases. Knierim and co-workers, compared both enzymatic and physical fragmentation methods and found similar yields. Generally, enzymatic fragmentation was shown to be more consistent, but not so efficient when compared to physical shear methods, due to bias and for the detecting of insertions/deletions (Knierim et al., 2011). The choice between physical or enzymatic method should also consider additional external factors, such as lab facilities and experimental design (Knierim et al., 2011).

Binding adapters

After the starting material has been sheared, specific adaptors are connected to the obtained fragments. The adaptors are important since they define the boundaries of the region of interest (known begins and ends), which depending on the methods used can be random sequences, allowing their parallel processing. To ligate Illumina platform adaptors (P5 and P7), the fragment ends are blunted and 5´ phosphorylated. Then, the 3´ ends are

adenylated with a single base, which allows the hybridization of the 3' thymine overhang of sequencing adaptors and binding using a DNA ligase. In Ion Torrent, is slightly different, fragments are blunted-end and two different adaptor sequences (A and P1) are connected. During the adapter ligation reaction, a suitable adapter:fragment ratio must be achieved, in order to avoid either a weak ligation or the formation of dimers (Head et al., 2014; van Dijk et al., 2014). If during an experiment or sequencing run it is necessary to combine several different samples, distinct barcodes are required for each library besides the platform-specific adaptors. Barcodes allow the independent processing of different samples or targets (up to 384 distinct reactions) in the same run. These specific sequences allow to informatically demultiplex the generated reads. An alternative strategy was developed that combines fragmentation and adaptor ligation in a single step, thus making the process simpler, faster and requiring a reduced amount of sample. The process is known as tagmentation and it is based on transposon-based technology (Adey et al., 2010). This allows the enzymatic fragmentation and simultaneous incorporation of tags of known sequence at the ends of the fragments, allowing the addition of sequencing adaptors by PCR. There are commercially available transposon-based library preparation kits for both the Illumina and Ion Torrent platforms: the Nextera DNA Sample Preparation Kit and the MuSeek Library Preparation Kit, respectively.

Size-selection

Upon nucleic acid fragmentation, the fragments are select according to the desired library size. This is limited either by the type of NGS instrument and by the specific sequencing application. Short-read sequencers, such as Illumina and Ion Torrent, present best results when DNA libraries contain shorter fragments of similar sizes. Illumina sequences are up to 600 bp (2x300) in length (MiSeq, Illumina), while the Ion Torrent platform allows a maximum fragment size of about 400 bp. Optimal library size is also limited by the sequencing application. In general, for WGS the longer is the fragment the better, while for WES shorter reads are still suitable since most exons are under 200 bp in length (Sakharkar et al., 2004). Size fractionation of randomly sheared DNA often constitutes a laborious process and a true bottleneck in NGS applications. Different fragment size selection methods are available. Most are still done by agarose gel electrophoresis, since comparing with other methods are more precise and cheaper. As an example, E-Gel electrophoresis system (Thermo Fisher Scientific) allow this size selection by simply pipetting-out the band of interest from the gel. More automated systems such as the BluePippin (Sage Science) comprises a computerized instrument which combines electrophoresis with a fluorescence-based DNA detection unit. DNA fragments are detected and selected according to a user-programmed size range, allowing DNA fractions to be electro-eluted. The solid-phase reversible immobilization (SPRI)

paramagnetic beads with reversible binding affinity for DNA fragments (Agencourt AMPure XP, Beckman Coulter) are also widely used to select the size of the desired fragments. By adjusting the ratio between beads solution and the DNA sample, it is possible to keep DNA in the supernatant or bounded to beads (Deangelis et al., 1995). An alternative method, not as commonly used as compared with those previously mentioned, is the use of spin columns, also through commercial kits, such as GeneJET NGS Cleanup Kit (Thermo Fisher Scientific) and Select-a-Size DNA Clean & Concentrator (Zymo Research). Between the different library preparation steps, rounds of amplification and purification steps might be required to obtain a suitable amount or concentration of the library.

Target enrichment

Target enrichment libraries are prepared when only a specific segment of the genome is required to be sequenced. It can be a single locus, a gene panel or even the entire exome. The methods to generate these libraries can be divided into hybridization-based capture and amplicon-based methods (Mertes et al., 2011; Samorodnitsky et al., 2015). In hybrid capture methods, the prepared DNA fragments (fragmentation-derived library) are hybridized specifically to DNA or RNA probes that are complementary to the regions of interest, either in a solution or on a solid support, allowing to be physically captured and isolate the fragments from the regions of interest (Mertes et al., 2011). This could be done by different methods on solid support built upon the knowledge of microarray research (Albert et al., 2007; Hodges et al., 2007) or using biotinylated oligonucleotide probes in solution (Gnirke et al., 2009), to physically capture and isolate the sequences of interest. An increasing number of protocols and vendors are offering out of the box solutions for hybrid capture, meaning, the researcher does not need to perform its development, but rather choosing between a pre-set targeted enrichment region (e.g. whole-exome or a panel of genes) or specify their own custom enrichment region. Example vendors would include Agilent (SureSelect product), NimbleGen (SeqCap EZ product), Illumina (TruSeq DNA and Nextera Rapid Capture), Flexgen and microarray. The main steps comprise probe hybridization, capture hybrids on magnetic beads, and amplification of captured libraries. Haloplex is another commercial method from Agilent and can be applied in both platforms. Basically, the DNA is digested by restriction enzymes and hybridized with HaloPlex probe library, resulting in DNA circulation and incorporation of indexes primer cassette and sequencing motifs, forming DNA-probe hybrids. Then, biotinylating of probe DNA, using magnetic beads, allows the hybrid capture of target regions, which are amplified by PCR to target enrich the sample.

Amplicon-based methods are based on the design of synthetic oligonucleotides (primers), with complementary sequences to the flanking regions of the target DNA. Both Ion Torrent and Illumina have their own commercial kits available to amplicon generation, known as

Ampliseq and TruSeq (custom amplicon library), respectively. For the Ion Torrent method, the first step consists of a multiplex PCR to generate a library of DNA fragments. Depending on the application, it can contain up to 24,000 primer pairs in a single reaction. Then primers are partially digested and the specific platform adapters are ligated. As for the TruSeq assay, it allows the simultaneous capture of multiple targets of interest and to sequence thousands of amplicons.

Library quality-control and quantification

The next and final step in library preparation workflow is quantification and QC assessment. Several methods are currently available and should be chosen according to the type of library and sequencing platform. With the purpose to quantify spectrophotometrically (NanoDrop), fluorometric (Qubit) and qPCR methods have been reported (Robin et al, 2016). One drawback the spectrophotometric and fluorometric methods is their lower sensitivity. Thus, one main reason why NGS protocols often require higher library concentrations is to ensure that they are within the quantification method's detection range. The main problem is that to achieve this, an additional library amplification step is required, posing additional quality problems, distortion of molecules proportions (with the eventual loss of rarer molecules) and over-amplification of shorter fragments. To perform QC, it is necessary to inspect the profile of the libraries' fragments. Automated capillary electrophoresis separation systems that incorporate instrument optics to detect laser-induced fluorescent signal [e.g. Bioanalyser or TapStation (Agilent) or LabChip GX Touch 24 (PerkinElmer)] are very effective, in addition to quantification, to perform this profile inspection.

Clonal amplification and template preparation

Independently of the starting material used to prepare NGS library, most of the 2nd-generation sequencers require several thousands of input molecules to generate a detectable sign and to produce a readable (higher quality) sequence. This requirement makes clonal amplification a mandatory step in such NGS workflow, it is performed either by cluster generation (Illumina platforms) or by micro-emulsion PCR (Ion Torrent platform). In Ion Torrent, the adaptors are used to capture single molecules of the template into polyacrylamide micro-beads by primer hybridization (Ion Torrent, 2018a). These special beads were specifically developed and placed under the commercial name Ion Sphere™ particles. Subsequently, library fragments are clonally amplified by emulsion PCR. Here, beads with single strands previously fused by primer hybridization, are incorporated into a carefully controlled emulsion, in which each droplet constitutes a microreactor containing DNA template, primer, and reagents for PCR. Thus, library fragments are amplified and following amplification, each bead is coated with clonally amplified molecules, constituting

the target library (Ion Torrent, 2018a). As during emulsion PCR template, molecules are dispersed into numerous compartments, it is possible to avoid common PCR amplification problems, namely, template competition and minimizing the chances of recombination, and the formation of chimeric DNA (i.e. incomplete PCR products that serve as primers to amplify other fragments) (Kanagal-Shamanna, 2016). After the emulsion PCR, the Ion Sphere particles, covered with several copies of the templates, are loaded into the Ion chip wells by a short centrifugation step. The chip is placed on the sequencer machine to perform the sequencing itself (Ion Torrent, 2018a). This method, besides avoiding the mentioned chimeric amplicons, has more advantages, namely there is no DNA fragmentation thus reduce the risk of loss start material or introduce DNA damage by the fragmentation process (Knierim et al., 2011), it is efficient with a low amount of sample and the overall process of library preparation is complete in few hours and with little hands-on time, while others approaches make take days. The Illumina overall process of clonal amplification is performed in a completely different way known as cluster generation by bridge PCR (Illumina, 2018b). DNA fragments or amplicon are denatured into single-stranded molecules. These fragments are added into a flow cell with a surface with sequences corresponding to adapter sequences that allows their random binding. To perform bridge amplification unlabelled nucleotides and DNA polymerase enzyme are blended with the solid-phase fragments. As a result, double-stranded bridges are obtained stabilized on the surface (Illumina, 2018b). The process of denaturation is repeated generating single-stranded templates and bridge PCR is repeated until dense clusters of dsDNA are generated. Here each channel signal derived from the fluorescently labeled dNTP's is intensity enough to be detected by the system. Thus, bridge PCR is used to generate clusters of amplified sequencing library fragments on a solid support (Illumina, 2018b).

1.3.4 NGS BIOINFORMATICS

Along with all this new laboratory techniques, there was a considerable development in NGS (bio)informatics, which could only take place greatly due the increasing computational capacities (hardware), as well as algorithms and applications (software) to assist all the required steps: from the raw data processing to more detailed data analysis and interpretation of variants in a clinical context. Typically, NGS bioinformatics is subdivided in the primary, secondary and tertiary analysis.

Primary analysis

The primary data analysis consists of the detection and analysis of raw data (signal analysis), targeting the generation of legible sequencing reads (base calling) and scoring base quality. Typical outputs from this primary analysis are FASTQ file (Illumina) or unmapped binary alignment map (uBAM) file (Ion Torrent).

In Ion Torrent platform primary analysis is performed in the Ion Torrent Suite Software (Ion Torrent, 2018b). It starts with signal processing, in which the signal of nucleotide incorporation is detected by the sensor at the bottom of chip cell and converted to voltage and outputted from the sequencer to the server as a raw voltage data (DAT) file. For each nucleotide flow, one acquisition file is generated that contains the raw signal measurement in each chip well for the given nucleotide flow. During the analysis pipeline, these raw signal measurements are converted into incorporation measures (WELLS) file (Ion Torrent, 2018b). Base calling is the final step of primary analysis and is performed by a base-caller module. Its objective is to determine the most likely base sequence, the best match for the incorporation signal stored in WELLS file. The mathematical models behind this base-calling are complex (including heuristic and generative approaches) and comprising three sub-steps, namely: key sequence-based normalization, iterative normalization, and phase correction (Ion Torrent, 2018b). Such a procedure is required to address some of the errors occurring during the SBS process. Besides signal droop (DR), two other major errors, collectively known as called phase errors, may lead to false negative or positive calls. One is an incomplete extension (IE), where the flowed nucleotide did not attach to an available position in the template. The other is the carry forward (CF) problem typically resulting from a nucleotide binding where it was not supposed to. The parameters (DR, IE, and CF) are used to determine the 'most likely' base sequence in the well by the Solver algorithm (Ion Torrent, 2018b).

As for the Illumina platform, the principle for signal detection relies on fluorescence. Therefore, the base calling is apparently much simpler and is performed directly from fluorescent signal intensity measurements resulting from the incorporated nucleotides during each cycle. Illumina claims that their SBS technology delivers the highest percentage of error-free reads. The latest versions of their chemistry have been reoptimized to enable accurate base calling in difficult genomic regions such as GC rich, homopolymers and of repetitive nature (Illumina, 2018c). Further, the dNTPs have been chemically modified to contain a reversible blocking group that acts a temporary terminator for DNA polymerization. After each dNTP incorporation, the fluorescent dye is the image is processed to identify the corresponding base and then enzymatically cleaved-off to allow incorporation of the next one (Illumina, 2018c). However, a single flow cell often contains billions of DNA clusters tightly and randomly packed into a very small area. Such physical proximity could lead to crosstalk

events between neighboring DNA clusters. As fluorophores attached to each base produce light emissions, there can be some degree of interference between the nucleotide signals, that can overlap with the optimal emissions of the fluorophores of the surrounding clusters. Thus, although the base calling is simpler than in the Ion Torrent platform, the image processing step is quite complex. The overall process requires aligning each image to the template of cluster position on the flow cell, image extraction to assign an intensity value for each DNA cluster, followed by intensity correction. Besides this crosstalk correction, other issues are addressed through base-calling algorithms such phasing (failures in nucleotide incorporation), fading (or signal decay) and T accumulation (thymine fluorophores are not always efficiently removed after each iteration, causing a build-up of the signal along the sequencing run). Some of the initial base-callers for the Illumina platform were Alta-Cyclic (Erlich et al., 2008) and Bustard (Kao et al., 2009). Currently, there are multiple other base-callers differing in the statistical and computational methodologies used to infer the correct base (Ledergerber & Dessimoz, 2011).

As errors occur both in Ion Torrent and Illumina sequencing, these are expressed in quality scores of the base call based using Phred score, a logarithmic error probability. Thus, a Phred score of 10 (Q10) refers to a base with a 1 in 10 probabilities of being incorrect or accuracy of 90.0%, and as for Q30 means 1 in 1000 of the probability of an incorrect base or 99.9% accuracy (Ewing et al., 1998). The quality of the raw sequences is critical for the overall success of NGS analysis, thereby several bioinformatic tools were developed to evaluate the quality of raw data, such as NG QC toolkit (Patel & Jain, 2012), QC-Chain (Zhou et al., 2013) and FastQC (Andrews et al., 2016). FastQC is one of the most popular, generating a report containing well-structured and graphical information about different parameters of the read quality. The Phred score is also useful to filter and trimming sequences. Additional trimming is performed at the ends of each read to remove adapters' sequences. Overall the trimming step, although reducing the overall number and the length of reads, it raises quality to acceptable levels. Several tools were developed to perform trimming, namely with Illumina data, such as BTrim (Kong, 2011), leeHom (Renaud et al., 2014), AdapterRemoval (Lindgreen, 2012) and Trimmonatic (Bolger et al., 2014). The choice of this algorithm is highly dependent on the dataset, downstream analysis and parameters used (Del Fabbro et al., 2013). To further analyze the sequences, a demultiplexing process is required which consists of separating the sequencing reads into separate files according to the barcodes used for each sample.

Secondary Analysis

The next step of the NGS data analysis pipeline is a secondary analysis which includes the reads alignment against the reference human genome (typically hg19 or hg38) and

variants calling. To map sequencing reads two different alternatives can be followed: their alignment against a reference genome, or *de novo* assembly that involves mapping from scratch without resourcing to a reference genome (Flicek & Birney, 2009). For most NGS applications mapping against a reference sequence is usually the first choice. As for *de novo* assembly, it is mostly confined to more specific projects, especially targeting to correct inaccuracies in the reference genome and to improve the identification of structural variants (SVs) and other complex rearrangements. Sequence alignment is a classic problem addressed by bioinformatics, especially when dealing with the existence of reads derived from repetitive elements, which at an extreme condition the algorithm must choose from which repeat copy the read belongs to. In such a context it is impossible to make high-confidence calls, the decision must be taken either by the user or software through a heuristic approach. Further, sequencing errors or discrepancies between the sequenced data and the reference genome also cause misalignment problems. Another difficulty is to establish a threshold between what is a real variation and a misalignment. A mapping algorithm will try to locate a location in the reference sequence that matches the read, tolerating a certain number of mismatches to allow subsequence variation detection. Among the commonly used methods to perform short reads alignments we highlight Burrows-Wheeler Aligners (BWA) and Bowtie mostly used for Illumina; whereas for Ion Torrent, the Torrent Mapping Alignment Program (TMAP) is the recommended alignment software as it was specifically optimized for this platform. BWA uses the Burrows-Wheeler transform algorithm, initially developed to prepare data for compression techniques such as bzip2, it is a fast and efficient aligner performing very well for both short and long reads (Li & Durbin, 2009; Li & Durbin, 2010). Bowtie (now Bowtie 2) is faster than BWA for some types of alignment (Langmead et al., 2009), but it may compromise sensitivity and accuracy. Bowtie is also used to align reads derived from RNA sequencing experiments. As for TMAP it comprises two stages: i) initial mapping (using Smith-Waterman or Needleman-Wunch algorithms) allowing reads to be roughly aligned against the reference genome; ii) alignment refinement, as particular positions of the read are realigned to the corresponding position in the reference. This alignment refinement is designed to compensate for specific systematic biases of the Ion Torrent sequencing process (such as homopolymer alignment and phasing errors with low INDEL scores) (Homer, 2012).

De novo assemblers are mostly based on graphs theory and can be categorized into three major groups: i) the Overlap-Layout-Consensus (OLC) method rely on an overlap graph, ii) the de Bruijn graph (BDG, also known as Eurelian) methods using some form of K-mer graph, and iii) the greedy graph algorithms that can use either OLC or DBG (Miller et al., 2010).

The reads alignments are typically stored in binary alignment map (BAM) or in sequence alignment map (SAM) files. BAM and SAM formats are designed to contain the same information, namely read sequence, base quality scores, the location of alignments, differences relative to reference sequence and mapping quality scores (Li et al., 2009). The main difference is that SAM being a text file is much easier to process by conventional text-based processing programs, whereas the BAM format provides a binary version of the same data (Li et al., 2009). These alignments files can be inspected using visualization software such as Integrative Genomic Viewer (IGV) (Thorvaldsdóttir et al., 2012) or GenomeBrowse (GoldenHelix Bozeman, MT).

Post-alignment processing is recommended for Illumina data prior to perform the variant call. Its objective is to increase the variant call accuracy and quality of the downstream process, by reducing base call and alignment artifacts (Tian et al., 2016). In general terms, it consists of filtering (removal) of duplicates reads, intensive local realignment (mostly near INDELs), and base quality score recalibration (Robinson et al., 2017). SAMtools (Li et al., 2009), Genome Analysis Toolkit (GATK) (McKenna et al., 2010) and Picard (http://broadinstitute.github.io/picard/) are some of the bioinformatic tools used to perform this post-alignment processing. Since variant calling algorithms assume that, in the case of fragmentation-based libraries, all reads are independent, removal of PCR duplicates and non-unique alignments (i.e., reads with more than one optimal alignment) is critical. This step can be performed using Picard tools (MarkDuplicates). If not removed, a given fragment will be considered as a different read, increasing the number of incorrect variant calls and leading to an incorrect coverage and/or zygosity assessment (Robinson et al., 2017). This PCR duplicates removal is usually not performed in reads obtained from amplicon-based libraries. Reads spanning INDELs impose further processing. Given the fact that each read is independently aligned to the reference genome when an INDEL is part of the read, there is a higher change for alignment mismatches. The realigner tool firstly determines suspicious intervals requiring a realignment due to the presence of INDELs, next the realigner runs over those intervals combining shreds of evidence to generate a consensus score to support the presence of the INDELs (Robinson et al., 2017). IndelRealigner from the GATK suite can be used to run this step. As mentioned, the confidence of the base call is given by the Phred-scaled quality score, that is generated by the sequencer machine and represents the raw quality score. However, this score may be influenced by multiple factors namely the sequencing platform and the sequence composition, and not reflecting the base-calling error rate (Nielsen et al., 2011). Consequently, is necessary to recalibrate this score to improve variant calling accuracy. BaseRecalibrator from the GATK suite is one of the most commonly used tools.

The variant call step has the main objective of identifying variants using the post-processed BAM file. Several tools are available for variant calling, some identify variants based on the number of high confidence base calls that disagree with the reference genome position of interest. Other, use Bayesian, likelihood, or machine learning statistical methods that use factor parameters, such as base and mapping quality scores, to identify variant differences. SAMtools, GATK, and Freebayes belong to the latter group and are among the most widely used toolkits for Illumina data (Hwang et al., 2015). Ion Torrent has its own variant caller known as the Torrent Variant Caller (TVC). Running as a plugin in the Ion Torrent server, TVC calls single-nucleotide variants (SNVs), multi-nucleotide variants (MNVs), INDELs in a sample across a reference or within a targeted subset of that reference. Several parameters can be customized, but often predefined configurations (Germ-Line vs Somatic, High vs Low stringency) can be used depending on the type of experiment performed. Most of these tools use the SAM/BAM format as input and generate a variant calling format (VCF) file as their output (Danecek et al., 2011). The VCF format is a standard format file, currently in version 4.2, developed by the large sequencing projects such as the 1000 genomes project. VCF is basically a text file contains meta-information lines, a header line, followed by data lines each containing information chromosomal position, the reference base, the identified alternative base or bases. The format also contains genotype information on samples for each position (https://samtools.github.io/hts-specs/VCFv4.2.pdf). VCFtools provide the possibility to easily manipulate VCF files, e.g., merge multiple files or extracting SNPs from specific regions (Danecek et al., 2011).

Genetic variations can occur in the human genome ranging from SNV and INDELS to more complex (submicroscopic) SVs (Feuk et al., 2006). These SVs include both large insertions/duplications and deletions (also known as copy number variants, CNVs) and large inversions. Although SVs can be detected by traditional techniques (as outlined in 1.3.1) its identification by NGS gained further interest over recent years. Incorporating the calling of such SVs would increase the diagnostic yield of these NGS approaches, overcoming some of the limitations present in other methods and with the potential to eventually replace them. Reflecting this growing tendency, several bioinformatic tools have been developed to detect CNVs from NGS data. These can be classified into five different major groups according to type algorithms/strategies used: paired-end mapping, split read, read depth, *de novo* genome assembly, and combinatorial approach (Zhao et al., 2013). Paired-end mapping algorithm compares the average insert size between the actually sequenced read-pairs and that expected based on the reference genome. Discordant reads maps may indicate either the presence of a deletion or insertion. The paired-end mapping methods can also efficiently identify mobile element insertions, insertions (smaller than the average insert size of the genome library inversions), and tandem duplications (Zhao et al., 2013). A limitation of

paired-end mapping is the inability to detect CNVs in low-complexity regions with segmental duplications (Korbel et al., 2009). BreakDancer (Chen et al., 2009) and PEMer (Korbel et al., 2009) are two examples of tools for paired-end mapping strategy. Concerning split read methods, they derive from the concept that due to an SV only one of the reads of the pair is correctly aligned to the reference genome, while the other one fails to map or only partially aligns (Zhao et al., 2013). The latter reads also have the potential to provide the accurate breakpoints of the underlying genetic defect at the base pair level. The partially mapped reads are spliced into multiple fragments that are aligned to the reference genome independently (Zhao et al., 2013). This further remapping step provides the exact starting and ending positions of the insertion/deletion events. Some example of tools using the split read method is Pindel (Ye et al., 2018) and SLOPE (Abel et al., 2010). As for the read depth methodology, it assumes that is possible to establish a correlation between the number of copies and the read coverage depth. The study design, as for normalization or data comparison, can be based either in single samples, paired cases, control samples, or a large dataset (population) of samples. In general terms, these algorithms present for main steps: mapping, normalization, estimation of copy number, and segmentation (Zhao et al., 2013). First, reads are aligned, and coverage is estimated across an individual genome in predefined intervals. Next, the CNV-calling algorithm must perform a normalization in terms of the number of reads to compensate potential biases, for instance, due to GC content or repetitive sequences. Having a normalized read depth, it is possible to obtain a rough estimation of copy numbers (gains or losses) across genomic segments. As for the final step, segmentation, regions with similar copy number are merged to detect abnormal copy number (Magi et al., 2012; Zhao et al., 2013). In comparison with pair-end and split reads methods, read depth approaches have the advantage to estimate the number copies of CNVs and a tendency to obtain better performances with larger CNVs (Zhao et al., 2013).

Detection of CNV mainly relies on WGS data since includes non-coding regions which are known to encompass a significant percentage of SVs (Zarrei et al., 2015). WES has emerged as a more cost-effective alternative to WGS and the interest in detecting CNV from WES data has grown considerably. However, since only a small fraction of the human genome is sequenced by WES it is not able to detect the complete spectrum of CNVs. Also, the lower uniformity of WES as compared with WGS may reduce its sensitivity to detect CNVs. Usually, WES generates higher depth for targeted regions as compared with WGS. Consequently, most of the tools developed for CNVs detection using WES data, have depth-based calling algorithms implemented and require multiple samples or matched case-control samples as input (Zhao et al., 2013). Ion Torrent also developed a proprietary algorithm, as part of the Ion Reporter software, for the detection of CNVs in NGS data derived from amplicon-based libraries (Ion Torrent, 2018c).

Tertiary analysis

The third main step of the NGS analysis pipeline addresses the important issue of "making sense" or data interpretation, that is finding, in human clinical genetics context, the fundamental link between variant data and the phenotype observed in a patient. The tertiary analysis starts with variant annotation, which adds a further layer of information to predict the functional impact of all variants previously founded in variant calling step. Variant annotation is followed by variant filtering, prioritization, and data visualization tools. These analytical steps can be performed by a combination of a wide variety of software, which should be in a constant update to include the recent scientific findings, requiring constant support and further improvements by the developers. The variant annotation is a key initial step for the analysis of sequencing variants. As mentioned before, the output of the variant calling is a VCF file. Each line in such file contains raw information about a variant, such as genomic position, reference, and alternate bases, but no data about its biological consequences. Variant annotation offers such biological context for the all variants found. Given the massive amount of NGS data, data annotation is performed automatically. Several tools are currently available, and each uses different methodologies and databases for variant annotation. Most of the tools can perform both the annotation of SNVs and the annotation of INDELs, whereas annotation of SVs or CNVs are more complex and are not performed by all methods (Scherer et al., 2007). One basic step in the annotation is to provide the variant's context. That is in which gene the variant is located, its position within the gene and the impact of the variation (missense, nonsense, synonymous, stop-loss, etc). Additional annotation might integrate data from other algorithms such as SIFT (Ng et., 2003), PolyPhen-2 (Adzhubei et al., 2010), CADD (Kircher et al., 2014) and Condel (González-Pérez et al., 2011), that computes the consequence scores for each variant based on various different parameters, like the degree of conservation of amino acid residues, sequence homology, evolutionary conservation, protein structure or statistical prediction based on known mutations. Additional annotation can resource to population or disease variants databases (ClinVar or HGMD), where information about its frequency or clinical association is retrieved. Among the extensive list of annotation tools, the most widely used are ANNOVAR (Yang et al., 2015), variant effect predictor (VEP) (McLaren et al., 2016), snpEff (Cingolani et al., 2012) and SeattleSeq (Ng et al., 2009). ANNOVAR is a command-line tool that can identify SNPs, INDELs, and CNVs. It annotates the functional effects of variants with respect to genes and other genomic elements and compares variants to existing variation databases. ANNOVAR can also evaluate and filter-out subsets of variants that are not reported in public databases, which is important especially when dealing with rare variants causing Mendelian diseases (Wang et al., 2010). Like ANNOVAR, VEP from Ensembl (EMBL-EBI) can provide genomic

annotation for numerous species. However, in contrast with ANNOVAR, that requires software installation and experienced users, VEP has a user-friendly interface through a dedicated web-based genome browser, although can have programmatic access via a standalone Perl script or a REST API. A wider range of input file formats are supported, and it can annotate SNPs, INDELs, CNVs or SV. VEP searches the Ensembl Core database and determines where in the genomic structure the variant falls and depending on that gives a consequence prediction. SnpEff is another widely used annotation tool, standalone or integrated with other tools commonly used in sequencing data analysis pipelines such as Galaxy, GATK and GKNO projects support. In contrast with VEP and ANNOVAR, it does not annotate CNVs but has the capability to annotate non-coding regions. It can perform annotation for multiple variants being faster than VEP (Cingolani et al., 2012).

The variant annotation may seem like a simple and straightforward process; however, it can be very complex considering the genetic organization intricacy. In theory, the exonic regions of the genome are transcribed into RNA, which in its turn is translated into a protein. Making that one gene would originate only one transcript and ultimately a single protein. However, such concept (one gene-one enzyme hypothesis) is completely out-dated, as the genetic organization and its machinery are much more complex. Due to a process is known as alternative splicing, from the same gene, several transcripts and thus different in proteins can be produced. Alternative splicing is the major mechanism for the enrichment of transcriptome and proteome diversity (Keren et al., 2010). While imperative to explain genetic diversity, when considering annotation this is a major setback, depending on the transcript choice the biological information and implications concerning the variant can be very different. Additional blurriness concerning annotation tools is caused by the existence of a diversity of databases and reference genomes datasets, which are not completely consistent and overlapping in terms of content. The most frequently used are Ensembl (http://www.ensembl.org), RefSeq (http://www.ncbi.nlm.nih.gov/RefSeq/) and UCSC genome browser (http://genome.ucsc.edu), that contain such reference datasets and additional genetic information for several species. These databases also contain a compilation about the different set of transcripts that were observed for each gene and are used for variant annotation. Each database has its own particularities and thus depending on the database used for annotation the outcome may turn different. For instance, if for a given *locus* one of the possible transcripts has an intron retained, while in the others have not, a variant located in such region will be considered as located in the coding sequencing in only one of the isoforms. In order to minimize the problem of multiple transcripts, the collaborative consensus coding sequence (CCDS) project was developed. This project aims to catalog identical protein annotations both on human and mouse reference genomes with stable identifiers and to uniformize its representation on the different databases (Pruitt et al., 2009).

The use of different annotations tools also introduces more variability to NGS data. For instance, ANNOVAR by default uses 1 Kb window to define upstream and downstream regions (Wang et al., 2010), while SnpEff and VEP use 5 kb (Cingolani et al., 2012; McLaren et al., 2016). This makes that the classification of variant different even though the same transcript was used. McCarthy and co-workers found significant differences for VEP and ANNOVAR annotations of the same transcript (McCarthy et al., 2014). Besides the problems related to multiple transcripts and annotation tools, there are also problems with overlapping genes, i.e. more than one gene in the same genomic position. There is still no complete/definitive solution to deal with these limitations, thus results from variant annotation should be analyzed with respect to the research context problem and, if possible, resorting to multiple sources.

After annotation of a VCF file from WES, the total number of variants may range between 30,000 to 50,000. To make clinical sense of such a large number of variants and to identify the disease-causing variant(s), some filtering strategies are required. Although quality control was performed in previous steps, several false-positive variants are still present. Consequently, when starting the third-level of NGS analysis, is highly recommended to, based on quality parameters or previous knowledge of artifacts, reduce the number of false-positive calls and variant call errors. Parameters such as the total number of independent reads and the percentage of reads showing the variant, as well as, the homopolymer length (particularly for Ion Torrent, with stretches longer than 5 bases being suspicious) are examples of filters that could be applied. The user should define the threshold based on observed data and research question but, relatively to the first parameter, less than 10 independent reads are typically rejected since it is likely due to sequencing bias or the low coverage itself.

One of the most commonly used NGS filters is the population frequency filter, in which variants with allele frequency above a certain threshold (1% in most studies) are excluded. Minor allele frequency (MAF), one of the metrics used to filter based on allele frequency, can sort variants in three groups: rare variants (MAF < 0.5), low frequent variants (MAF between 0.5–5%) and common variants (MAF > 5%) (International HapMap 3 Consortium et al., 2010). Populational databases, that includes data from thousands of individuals from several populations represent a powerful source of variant information about the global patterns of human genetic variation. It also helps, not only to better identify diseases alleles but also are important to understand the populational origins, migrations, relationships, admixture and changes in population size, which could be useful to understand some diseases pattern (Stoneking & Krause, 2011). For instance, the population frequency filter threshold should be set according to the expected prevalence/incidence of the disease. Population history, geographical events (e.g. genetic bottleneck), and cultural factors (e.g. marriage between

close biological relatives) contribute to high incidence of diseases in certain populations and thus, in this populations, the allele frequency needs to be adjusted (Bittles, 2001). Population databases such 1000 genome project (Siva, 2008), Exome Aggregation Consortium (ExAC) (Lek et al., 2016), and the Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/) are the most widely used databases. Nonetheless, this filter has also limitations and could cause erroneous exclusion, which is difficult to overcome. For instance, as carriers of recessive disorders carriers do not show any signs of the disease, the frequency of damaging alleles in populational variant databases can be higher than the established threshold. In-house variant databases are important to assist variant filtering, namely, to help to understand the frequency of variants in a study population and to identify systematic errors of an in-house system.

In kindreds with a recognizable inheritance pattern, it is advisable to perform family inheritance-based model filtering. These are especially useful if more than one patient of such families is available for study, as it would greatly reduce the number of variants to be thoroughly analyzed. For instance, for AD diseases the ideal situation would be testing at least three patients from a different generation and select only variants in heterozygosity. If a pedigree indicates a likely x-linked disease, variants located in the X chromosome are select and the remaining variants in other chromosomes are not primarily inspected. As for AR diseases, with more than one affect sibling, it would be important to study as many patients as possible. Two models should be considered, either homozygous variants in patients that are found in heterozygosity in both parents or two heterozygous variants in the same gene with distinct parental origins. For sporadic cases, the trio analysis can constitute as extremely useful to reduce the analytical burden. In such a context, heterozygous variants found only in the patient and not present in both parents would indicate a *de novo* origin. Even in non-related cases with very homogeneous phenotypes, such as those typically syndromic, it is possible to use an overlap-based strategy (Gilissen et al., 2012), assuming that the same gene or even the same variant is shared among all the patients. An additional filter, useful when many variants persist after applying others, is based on the predicted impact of variants (functional filter). In some pipelines intronic or synonymous variants, based on the assumption that they likely to be benign (non-disease associated). Nonetheless, care should be taken since numerous intronic and apparent synonymous variants, have been implicated in human diseases (Sauna et al., 2011; Cartegni et al., 2002). Thus, a functional filter is applied in which the variants are prioritized based on its genomic location (exonic or splice-sites). Additional information for filtering missense variants includes evolutionary conservation, predicted effect on protein structure, function or interactions. To enable such filtering, the scores annotated in the VCF generated by algorithms to evaluate missense variants (PolyPhen, SIFT and CADD for instance) are extremely useful. The same applies to

variants that might have an effect over splicing, as prediction algorithms are being incorporated in VCF annotation, such as the Human Splice finder (Desmet et al., 2009) in VarAFT (Desvignes et al., 2018).

Although functional annotation adds an important layer of information for filtering, the fundamental question to be answered in the context of gene discovery is the identification of the disease-causing gene (MacArthur et al. 2012; Lelieveld et al., 2016). To address this complex question, a new generation of tools are being developed, that instead of merely excluding information perform variants ranking, thus allowing their prioritization. Different approaches have been proposed. The first such PHIVE explores the similarity between human diseases phenotype and those derived from knockout experiments in animal model organisms (Robinson et al. 2014). Other algorithms attempt to solve this problem by an entirely different way, through the computation of a deleteriousness score (also known as burden score) for each gene, based on how intolerant genes are to normal variation and using data from population variation databases (Eilbeck et al., 2017). It was proposed that human disease genes are much more intolerant to variants than non-disease associated genes (Khurana et al. 2013; Petrovski et al. 2013). The Human Phenotype Ontology (HPO) enables the hierarchical sorting by disease names and clinical features (symptoms) for describing medical conditions. Based on these descriptions, HPO can also provide an association between symptoms and known disease genes. Several tools attempt to use these phenotype descriptions to generate a ranking of potential candidates in variant prioritization. As an example, some attempt to simplify analysis in a clinical context, such as the phenotypic interpretation of exomes (PhenIX, Zemojtel et al., 2014) that only reports genes previously associated with genetic diseases. While others can also be used to identify novel genes, such as Phevor (Singleton et al., 2014) that uses data gathered in other related ontologies, Gene Ontology (GO) for example, to suggest novel gene-disease associations.

1.3.5 NGS APPLICATIONS IN CLINICAL GENETICS

With the advances of the sequencing technology, multiple applications become available, being extremely useful not only for genetic research but also in a clinical setting aiming to establish a definitive genetic diagnosis.

Gene panels, WES and WGS

The most common NGS approaches, consists of (multi) gene panels, that is a set of regions associated with heterogeneous genetic diseases. Their size is extremely variable, ranging from smaller panels with only a handful of genes to larger solutions comprising all

the known *loci* associated to genetic disease: the clinical exome or Mendeliome (Lee et al., 2014; Xue et al., 2015). In comparison with the other comprehensive genomic approaches, gene panels are much easier to handle as their cost, size of the data files, and the total number of variants detected are often much smaller. A possible drawback is a difficulty in defining the appropriate list of genes for a given disease/phenotype, owing to the remarkable heterogeneity of clinical features in some conditions. Although with the clear risk of losing information, gene panels undoubtedly facilitate the analysis due to a significant reduction of the number of variants obtained. Hence, gene panels have two key advantages: i) faster concerning data processing and data analysis, and ii) lower costs comparing with WES or WGS (Xue et al., 2015). These two comprehensive genomic approaches, that have become accessible to genetics laboratories during the last years, have generated much discussion about which is the most effective/feasible for genetic research. WGS allows a complete examination of all genomic regions, thus SNVs, INDELs, and SVs both in coding and non-coding regions of the genome, whereas WES targets coding regions and flanking intronic sequences, lacking non-coding regions and regulatory regions. Although the average coverage is higher in WES as compared to WGS, the uniformity is much effective in the latter. This means that the likelihood of having a region not covered is much lower in WGS. It is explainable by the hybridization efficiency of capture probes or amplification (in the case of amplicon-based approaches) performed in WES, which results in target regions (such as those with high GC content) that have little or no coverage. WGS is beyond a doubt the more powerful approach having the potential to efficiently detect all types of genomic variants (Belkadi et al., 2015; Meienberg et al., 2016). However, it comes with the drawback of higher sequencing costs and greater informatics resources, both in terms of hardware (data processing and storage) and software (to handle the data), along with longer turnaround time and bioinformatics expertise to deal with more vast and complex data analysis. Consequently, WGS is mainly accessible to large research centers or sequencing services providers. WES is less expensive than WGS and data analysis requires less computational resources. Furthermore, 85% of the known disease-causing variants are located within the coding sequences those being effectively targeted by WES (Antonarakis et al., 2000). Thus, WES has emerged as a powerful approach for researchers elucidating genetic variants underlying human diseases.

Single-cell sequencing

Single-cell sequencing enables the capability of examining the genome of an individual cell. Single-cell sequencing makes possible to evaluate the roles of genetic mosaicism in normal physiology and disease and can be applied in a variety of fields such as metagenomics or cancer. For instance, in cancer genomics, it may be helpful in the

determination of the contributions of intra-tumor genetic heterogeneity in cancer development or its response to treatment (Gawad et al., 2016). Single-cell sequencing involves isolating a single cell, cell lysis, and nucleic acid extraction, whole genome amplification (WGA), constructing sequencing libraries and then sequencing the DNA (Yilmaz & Singh, 2012). Both Ion Torrent and Illumina having commercially available methods to perform this analysis. Notwithstanding its helpfulness, several challenges should be considered namely sample loss or degradation, sample contamination either by other cell types or by other environmental contaminants, which have critical effects on the quality of sequencing. Besides, as the amount of starting material is very low (picogram level) to achieve satisfactory coverage, extensive amplification is needed. This step may contribute to errors and bias, preferential allele amplification and allele dropout which could contribute to a lower quality of sequencing data (Findlay et al., 1995). Novel variant-calling algorithms would overcome some of these current technical limitations (Eberwine et al., 2014; Gawad et al., 2016). Great expectations come from the 3rd-generation sequencers with the promise to overcome some of these technical limitations, by performing single-molecule sequencing without a previous amplification step. However, the still high sequencing error rates and costs associated with such platforms still render them as actually suitable alternatives.

NGS non-invasive prenatal test

Nearly two decades ago, Dennis Lo and collaborators identified the presence of free-fetal DNA in maternal plasma and serum (Lo et al., 1997). The presence of this cell-free DNA opened a new chapter in terms of non-invasive approaches for prenatal testing. These non-invasive prenatal tests (NIPT) are safer and can be applied much sooner than the traditional approaches, consequently, several efforts have been made to translate them into clinical routine. NIPT allows testing for fetal aneuploidies and even the diagnosis of specific monogenic diseases. NGS has already been shown useful to perform NIPT (Oepkes et al., 2016). Two main technical approaches can be applied: namely low-coverage WGS (i.e. massive parallel shotgun sequencing) and targeted resequencing (Norwitz & Levy, 2013). For instance, in WGS to detect aneuploidies, the entire cell-free DNA is fragmented in a controlled way and sequenced. Reads are aligned to the reference genome and mapped to a specific chromosome. Then, the number of reads observed in the specific location of a chromosome is normalized and compared to reference samples. Aneuploidies are detected by the differences between the number reads are present. This implies sequencing millions of DNA fragments to guarantee that chromosome reads differences are statistically significant. A very low amount of fetal DNA (fetal fraction <4%) and maternal mosaicism are some challenges of applying NGS to perform NIPT, and a major source of a false positive or false negative. The existence of a vanishing twin, which can result in assisted reproductive

techniques, are also recognized as a source of false positive results (Grömminger et al., 2014)

Transcriptome analysis by NGS (RNA sequencing)

By definition, transcriptome is the complete set of transcripts expressed in a cell. The analysis of such molecules is important to understand the expression profile of the genome, to identify changes in such profiles in different cells and tissues, or to determine the effect of genetic defects which cannot be inferred from genomic data. In clinical research, transcriptome analysis assumes that some genetic variants may have detrimental effects on gene expression, namely through changes on splicing patterns. Gene expression variation may also occur due to changes in physiological, metabolic and environmental states of a particular cell. The characterization of such modifications, under a specific developmental or physiological condition, could even be useful in the characterization or identification of disease-causing genes. As in its genomic counterpart, the development of NGS methods applied to transcriptome analysis open new avenues for research, essentially through the method designated as RNA sequencing (Wang et al., 2009). It presents several advantages over the traditional assays to study transcriptome (e.g. microarray hybridization-based approaches). RNA sequencing requires a very low amount of starting material which is helpful in cases where biological samples are limited. Moreover, it does not require prior knowledge about the transcriptomic composition of an organism, being capable of identifying novel transcripts and to distinguish different isoforms of the same gene. RNA sequencing has higher sensitivity (even for genes expressed at very low levels) and better dynamic range for quantifying gene expression (Wang et al., 2009). Still, a few drawbacks and limitations of RNA sequencing should be pinpointed. It is especially sensitive to sample degradation and contaminants as the other techniques using RNA as a template. The RNA sample should be cleared of ribosomal RNA which can reach high abundance in some tissue (typically constituting up to 90% of total RNA) (Conesa et al., 2016). Otherwise, this ribosomal material would overrun messenger RNA (mRNA) molecules which are the primary target of this approach. For eukaryotic samples, this may require enrichment for mRNA using poly(A) selection or to deplete rRNA (O'Neil et al., 2013; Zhao et al., 2014). The sequencing depth of coverage is another critical aspect of RNA sequencing. As a general rule of thumb, higher coverage leads to more precise sequencing data. However, there is no full consensus towards the minimum number of mapped reads as it is very dependent with experiment objective, making very difficult comparisons between experiments and data analysis. The number of technical replicates that should be included in such experiments and type of tissues to be analyzed are additional aspects to consider depending on the biological question to be addressed. It depends on both the amount of technical variability of the

technical procedure and also the biological variability of the system under observation. Nevertheless, RNA sequencing can provide unprecedented benefits to clinical research (Byron et al., 2016). For example, multigene mRNA profiling assays are being used to measure the expression of several different genes with broad clinical applications. Again, cancer genetics is among the areas where RNA-sequencing can have applicability, given the importance of the detection of oncogenic gene fusions and of other genetic biomarkers (Wang et al., 2015).

NGS applications to study epigenetics

Epigenetics refers to reversible biochemical modifications of DNA and associated proteins, that do not alter the DNA sequence itself but alters gene expression and its function. Different epigenetic modifications have been identified over DNA (cytosine methylation and hydroxymethylation) or histone proteins (lysine acetylation, ubiquitination, and sumoylation; lysine and arginine methylation; serine and threonine phosphorylation) (Belzil et al., 2016). Although still in an early stage, there is a growing interest in DNA methylation patterns identified by NGS, being applied to unveil pathophysiological mechanisms and diseases biomarkers in different fields such as cancer, neurodegenerative and psychiatric disorders (Zhang et al., 2010). New protocols were developed to improve its detection through NGS. Namely, the PacBio system was used for differential methylation status of CpG islands, with or without bisulfite conversion, through single-molecule real-time (Yang et al., 2015; Flusberg et al., 2010). Chromatin immunoprecipitation coupled with short-tag sequencing (ChIP-seq) was developed to identify genome-wide DNA binding sites proteins including transcription factors. Such protocols use specific antibodies to immunoprecipitate DNA-bound protein which after suitable purification are sequenced (Mirabella et al., 2016). ChIP-seq can provide valuable insights into gene regulation events involved in various diseases and biological pathways, such as those typically connected with development and cancer progression.

Chapter 2

AIMS AND OUTLINE OF THE THESIS

Contents

2.1 Aims and objectives

2.2 Outline of the thesis

2.3 Patients and methods overview

2.1 AIMS AND OBJECTIVES

The overwhelming developments that occurred in genetic sequencing technologies have considerably reduced the gap between basic research and applied clinical applications. From here, derives an entirely new research perspective that strives to develop, optimize and apply new methodologies to achieve difficult genetic diagnoses, which otherwise would ultimately turn into a genetic odyssey. The knowledge gathered through such research can also be translated back to the clinics at different levels with a primary goal of improving the diagnostic workups for these diseases. This PhD thesis places into perspective the interconnected technological and scientific challenges which clearly remain to be addressed concerning the vast group of hereditary myopathies.

Conceptually, five broad research questions were initially drawn:

I) Considering the unique nature of congenital muscles diseases, especially in terms of its genetic and clinical heterogeneity, can this group of diseases fit as a model to improve the diagnostic workup of hereditary myopathies (*senso lato*)?

II) As for the most frequent subtypes of congenital muscle diseases (*LAMA2*-, *MTM1*- and *RYR1*-related diseases) could further insights about genotype-phenotype correlations and their mutational spectra be obtained?

III) How can diagnostic algorithms be improved, specifically through the development and application of new genotyping strategies, for example, next-generation sequencing?

IV) What sort of molecular defects give rise to the myopathies still uncharacterized from a genetic perspective?

V) Could the new genetic data gathered provide further understanding concerning the pathophysiological mechanisms underlying myopathies or even towards fundamentals in cell biology?

To address the research questions outlined above, the aims of this work are:

I) Systematically reassess the genetic data gathered from a large cohort of patients with congenital muscle diseases (from 2004 to 2017).

II) Apply additional conventional genotyping methodologies (besides Sanger sequencing), such as multiplex ligation-dependent probe amplification, to improve the mutation detection rates in congenital muscle diseases and as an attempt to identify new variants.

III) Establish novel and comprehensive locus-specific databases (LSDB) to display *LAMA2* and *MTM1* variants, as important resources made available for the scientific community to collect unreported variant and to refine genotype-phenotype correlations in these genetic entities.

IV) Develop new methods based on NGS technology to facilitate the identification of novel genetic defects causing hereditary myopathies, thus as an efficient and improved strategy to address the genetic heterogeneity seen in these diseases. One primary objective consisted in the setup of a new targeted resequencing approach based on NGS to analyze several genes in simultaneous, known to be defective in congenital myopathies.

V) Apply NGS gene panels, whole-exome sequencing and whole-genome sequencing technology to study patients with hereditary myopathies aiming to identify novel genetic causes.

VI) Integrate bioinformatic, transcriptomic and proteomic analysis, to further characterize novel variants and disease mechanisms.

2.2 OUTLINE OF THE THESIS

The experimental and research work essentially performed over the past five years is presented in this thesis, is structured in five main chapters. Besides the general introduction and the present chapter, it includes further two (chapters number three and four) with novel scientific findings. It was a clear intention to organize these two chapters with original work composed by standalone articles. These articles were subject to peer-review and the majority published in international journals of the field of human and clinical genetics or devoted to neuromuscular diseases.

Chapter three addresses the genetic characterization of a large group of patients with congenital muscle diseases studied by the author over a time-period of thirteen years. In this chapter, the so-called conventional approaches, such as Sanger sequencing and MLPA, were primarily used to characterize this patient cohort. To assist variant collection and draw further insights about the impact of novel variants and genotype-phenotype correlations, two publicly accessible databases, used as the central point to collect sequence variants and curated by the author, are also reported.

In chapter four, the genetic heterogeneity of different subtypes of hereditary myopathies: congenital myopathies and progressive muscular dystrophies (limb-girdle and dystrophinopathies), were revisited through a wide range of NGS approaches. The focus of this chapter was to pursuit novel genetic causes for these diseases, especially those defects which could not be identified using classical approaches in a straightforward manner.

Finally, chapter number five comprises the concluding remarks and future perspectives. It provides an overarching outlook of the main outcomes and conclusions that were drawn from the results attained in this work. The future research path is also delineated in this last chapter.

2.3 PATIENTS AND METHODS OVERVIEW

2.3.1 GENETIC ANALYSIS IN A COHORT OF CONGENITAL MUSCLE DISEASE PATIENTS BY CONVENTIONAL APPROACHES (CHAPTER 3)

This section derives from the genetic characterization attempt of 246 patients with clinical suspicion of congenital muscle diseases (CMD and CM), studied over a period of 13 years (2004 - 2017). An overview of the work performed in chapter 3 is depicted in Figure 2.3.1.
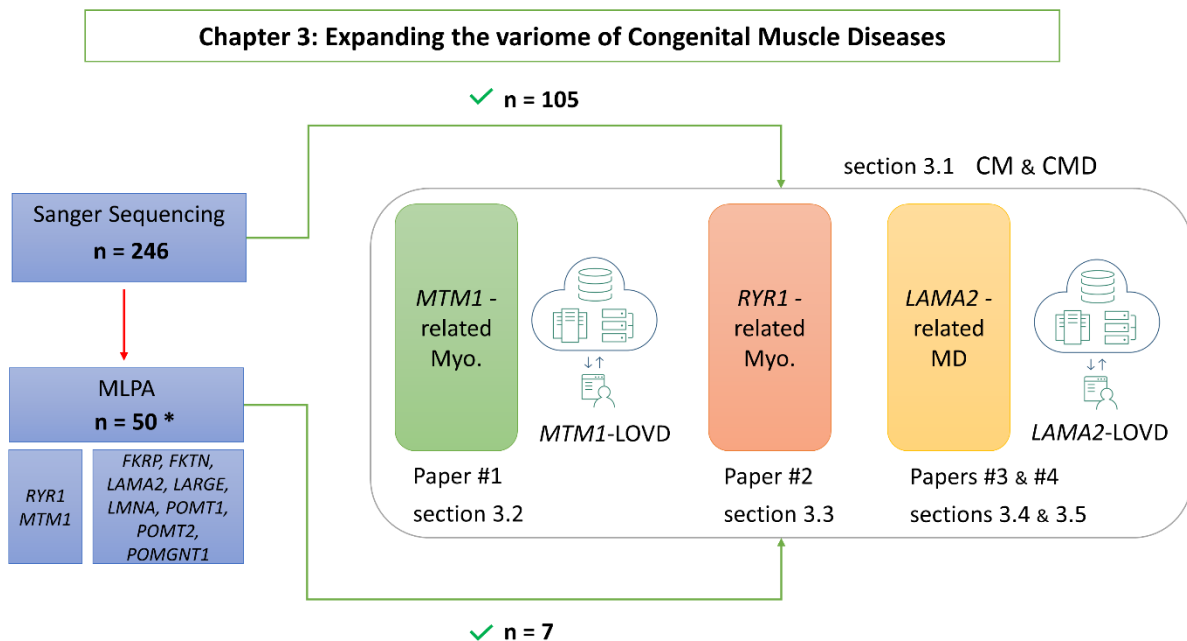
Figure 2.3.1- Chapter 3 experimental outline. MD- muscular dystrophies; Myo- myopathies; *- selected from the 141 unsolved cases.

As explored in further detail in the General Introduction, these diseases are individually very rare and throughout these years their genetic causes have been progressively established with the identification of new disease *loci*. During this time period, the author has implemented on a nationwide basis, the genetic analysis by conventional sequencing (Sanger sequencing) of twenty genes related to congenital muscle diseases (Table 2.3.1). These genetic studies were gradually introduced into routine genetic diagnostics and were complemented with MLPA analysis, a technique which is not currently available for all *loci*. Therefore, the data collectively presented wherein derives from a rather heterogeneous patient cohort, which also includes individuals from different geographical origins besides Portugal (namely Belgium, Brazil, Iran, Spain, Switzerland, and the United Kingdom).

In general terms, Sanger sequencing provides the highest quality data in terms of sequence for a specific region of interest but does not allows a quantitative analysis (i.e. the number of copies present) of the genomic region interrogated. As a result, large heterozygous gene deletions or duplications involving one or more exons are virtually undetectable through genomic sequencing. The frequency of these type of mutations differs considerably on the different loci, depending on their size and the relative abundance of region-specific repetitive sequences, such as low copy repeats (i.e. segmental duplications), that may contribute to the creation of recombination hotspots (Gu et al., 2008). Duchenne/ Becker muscular dystrophy is one of the best examples of a neuromuscular disease where most of the cases (approximately two thirds) are explained by large deletions or duplications (Santos et al., 2015). The introduction of the MLPA technique has significantly simplified and improved the detection of this type of mutations (Schouten et al., 2002). However, for most of the congenital muscle diseases, the frequency of large deletion and duplications and its foreseeable impact remained elusive. A systematic reassessment of large deletions and duplications resorted to MLPA technique. In four CMD patients with only one heterozygous pathogenic variant identified after completing *LAMA2* gene and compatible clinical features, the P391 and P392 MLPA kits (MRC-Holland, Amsterdam) were used in the search for a second variant. Additionally, in 28 patients with a CMD phenotype remaining genetically unsolved, were subject to MLPA analysis using six different commercial kits (MRC-Holland): P061 probe mix for *POMGNT1* and *POMT1*; P391 and P392 for *LAMA2*; P326 for *LARGE*, *FKTN* and *POMT2*; P116 for *FKRP*; and P048 for *LMNA*. Also, in 18 patients with a congenital myopathy (CNM or CCD subtypes), *MTM1* (P309 probe mix) and *RYR1* (P281 and P282) genes were also screened, respectively, for copy number variants. To further characterize large deletions or duplications detected by MLPA, an extensive set of molecular techniques such as genomic sequencing, Southern-blot, long-range PCR, cDNA analysis and even NGS, were used as detailed in the articles' material and methods.

Table 2.3.1- Genes analyzed by conventional sequencing.

Gene Symbol	Full name	Cytogenetic location	cDNA reference sequence	# exons	Protein reference sequence	# a.a.	Disease / disease group	Inherit. pattern
ACTA1	actin, alpha 1, skeletal muscle	1q42.13	NM_001100.3	7	NP_001091.1	377	CM & CMD	AD, AR
B4GAT1	beta-1,4-glucuronyltransferase 1	11q13.2	NM_006876.2	2	NP_006867.1	415	CMD	AR
BIN1	bridging integrator 1	2q14.3	NM_139343.2	19	NP_647593.1	593	CM	AR
COL6A1	collagen type VI alpha 1 chain	21q22.3	NM_001848.2	35	NP_001839.2	1028	CMD; (Bethlem myopathy)	AD, AR
COL6A2	collagen type VI alpha 2 chain	21q22.3	NM_001849.3	28	NP_001840.3	1018	CMD; (Bethlem myopathy)	AD, AR
COL6A3	collagen type VI alpha 3 chain	2q37.3	NM_004369.3	44	NP_004360.2	3177	CMD; (Bethlem myopathy)	AD, AR
DNM2	dynamin 2	19p13.2	NM_001190716.1	21	NP_001177645.1	869	CM & CMD	AD, AR
FKRP	fukutin related protein	19q13.32	NM_024301.4	4	NP_077277.1	495	CMD; (progressive MD)	AR
FKTN	fukutin	9q31.2	NM_001351496.1	12	NP_001338425.1	461	CMD	AR
ISPD	isoprenoid synthase domain containing	7p21.2	NM_001101426.3	10	NP_001094896.1	451	CMD	AR
LAMA2	laminin subunit alpha 2	6q22.33	NM_000426.3	65	NP_000417.2	3122	CMD; (progressive MD)	AR
LARGE1	LARGE xylosyl- and glucuronyltransferase 1	22q12.3	NM_004737.5	16	NP_004728.1	756	CMD	AR
LMNA	lamin A/C	1q22	NM_170707.3	12	NP_733821.1	664	CMD; (progressive MD)	AD, AR
MTM1	myotubularin 1	Xq28	NM_000252.3	15	NP_000243.1	603	CM	XLR
POMGNT1	protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)	1p34.1	NM_017739.3	22	NP_060209.3	660	CMD; (progressive MD)	AR
POMGNT2	protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-)	3p22.1	NM_032806.5	2	NP_116195.2	580	CMD	AR
POMT1	protein O-mannosyltransferase 1	9q34.13	NM_007171.3	20	NP_009102.3	747	CMD; (progressive MD)	AR
POMT2	protein O-mannosyltransferase 2	14q24.3	NM_013382.5	21	NP_037514.2	750	CMD; (progressive MD)	AR
RYR1	ryanodine receptor 1	19q13.2	NM_000540.2	106	NP_000531.2	5038	CM; (MD features in some cases)	AD, AR
SELENON	selenoprotein N	1p36.11	NM_020451.2	13	NP_996809.1	590	CM & CMD	AR

Footnote: #- number; a.a.-aminoacids; AD- autosomal dominant; AR- autosomal recessive; Inherit. - inheritance; XLR- linked to chromosome X recessive.

Additionally, in collaboration with Leiden University Medical Center (LUMC, Prof. Johan den Dunnen) novel resources to report and maintain sequence variants were established. Specifically, two internationally recognized locus-specific databases (LSDB) have been developed and/or maintained by the author:

i) **MTM1**, https://databases.lovd.nl/shared/genes/MTM1, or: http://tiny.cc/LOVD_MTM1

ii) **LAMA2**, https://databases.lovd.nl/shared/genes/LAMA2, or: http://tiny.cc/LOVD_LAMA2

To support its development, the Leiden Open (source) Variation Database (LOVD) software developed and stored by LUMC team was used in this task. Further to the variants collected directly from this research, additional sequence variants reported in publications and from scientific communications were also retrieved reviewed and included in these databases. This task also provided the opportunity to reinforce collaborations with other groups international and collect a significant number of novel sequence variants.

2.3.2 CHARACTERIZATION OF HEREDITARY MYOPATHIES USING NGS TECHNOLOGY AND BIOINFORMATICS (CHAPTER 4)

This second part of the thesis aimed to characterize a total of 22 patients with distinct subtypes of hereditary myopathies, which remained undiagnosed after being subject to a considerable number of previous genetic studies (Table 2.3.2). To address the genetic heterogeneity of these diseases, the author resorted to different NGS approaches to study these patients. Figure 2.3.2 summarizes the methodologies used in chapter 4.

The first task consisted in the development of a new NGS gene panel using semiconductor sequencing technology (Rothberg et al., 2011) using Ion AmpliSeq Designer (https://www.ampliseq.com/). This strategy was mainly intended to deal with the vast clinical and genetic heterogeneity seen in CM and to expand the list of genes sequenced, including some large *loci* such as nebulin (*NEB*) and titin (*TTN*).

Table 2.3.2- Data overview from the twenty-two patients studied by NGS approaches.

Patient #	Gender	Age *	Main clinical features	Muscle histology	I. pattern Fam. HX	Previous genetic studies	Research strategy	Results
I. Congenital myopathy with neonatal onset or during infancy								
1	F	31yrs	Neonatal hypotonia. Proximal tetraparesis, axial muscle weakness, and scoliosis. Ophthalmoparesis and bulbar weakness. Never walked.	Muscle biopsy was only performed in an affected brother.	AR (affected brother: case 2)	NEG: *MTM1*, *DNM2* genes	CM NGS gene panel	Chapter 4.1, paper #5
2	M	37yrs	Neonatal hypotonia. Proximal tetraparesis, axial muscle weakness, and scoliosis. Ophthalmoparesis and bulbar weakness. Never walked.	Myopathic features: biopsies showed considerable fiber size variability and frequent fibers with abnormal centrally placed nuclei. majority of atrophic and hypotrophic fibers were mainly of type 1 (evocative of congenital fiber-type size disproportion).	AR (affected sister: case 1)	not performed	Screening of *RYR1* variants performed by Sanger sequencing	Chapter 4.1, paper #5
3	F	27yrs	Neonatal hypotonia. Proximal tetraparesis, axial muscle weakness and scoliosis. Ophthalmoparesis and bulbar weakness.	Myopathic features: considerable fiber size variability and frequent fibers with abnormal centrally placed nuclei.	Sporadic	NEG: *RYR1* (hotspot) gene	CM NGS gene panel	Chapter 4.1, paper #5
4	M	78yrs	Severe and early bilateral foot drop. Later with milder facial, cervical and proximal upper limb weakness.	Myopathic features: fiber size variability, with atrophic fibers. On Gomori trichrome staining alfa-actinin aggregates (rods) were clearly seen.	AR (affected brother)	NEG: *CAPN3*, *DYSF,* and *TTN* genes	CM NGS gene panel	Chapter 4.1, paper #5

I. Congenital myopathy with neonatal onset or during infancy (continues)

Patient #	Gen der	Age *	Main clinical features	Muscle histology	I. pattern Fam. HX	Previous genetic studies	Research strategy	Results
5	M	20yrs	Early respiratory muscle involvement. Facial and bulbar weakness; proximal tetraparesis and later distal involvement. Axial muscle weakness. Respiratory insufficiency. Currently walks without assistance.	Myopathic features: fiber size variability, with atrophic fibers, corresponding to the majority to type 1 fibers. On the Gomori trichrome stain, there were areas with structures resembling rods.	AR (affected deceased brother)	NEG: *SMN1* gene	CM NGS gene panel	Chapter 4.1, paper #5
6	M	21yrs	Facial weakness and proximal tetraparesis (limb-girdle weakness, mostly scapular weakness). Delayed walking during infancy.	Myopathic features: fiber size variability and multiple small minicores seen on oxidative stains and semi-thin sections.	Sporadic	NEG: *RYR1* (hotspot) gene	CM NGS gene panel	Chapter 4.1, paper #5
7	M	15yrs	Congenital onset (bilateral clubfoot). Neonatal hypotony. Facial diparesis and bulbar weakness. Predominantly distal tetraparesis. Respiratory insufficiency.	Myopathic features: fiber size variability, with severely atrophic fibers, corresponding to the majority to type 1. Typical nemaline rods were clearly seen in Gomori trichrome staining.	Affected relative (a distant cousin)	NEG: *SMN1* and *DMPK* genes	CM NGS gene panel	Chapter 4.1, paper #5
8	M	30yrs	Muscle weakness, rigid spine, scoliosis, joint hyperlaxity, and contractures.	Myopathic features: type I fiber predominance and fiber size disproportion (type 1 smaller).	Sporadic	NEG: *COL6A1*, *COL6A2*, *COL6A3*, *RYR1* and *SEPN1* genes	CM NGS gene panel	No candidate variant

I. Congenital myopathy with neonatal onset or during infancy (continues)

| Patient #	Gender	Age *	Main clinical features	Muscle histology	I. pattern Fam. HX	Previous genetic studies	Research strategy	Results
9	M	22yrs	Muscle weakness and arthrogryposis.	Myopathic features: type I fiber predominance.	Sporadic	NEG: *RYR1* gene	CM NGS gene panel	No candidate variant
10	M	Deceased with one month	Neonatal hypotonia and respiratory difficulties. Electromyography with myopathic features.	Myopathic features, fiber size variability. Centrally placed nuclei, suggestive of centronuclear myopathy.	Sporadic	NEG: *MTM1* gene. Normal karyotype.	CM NGS gene panel	No candidate variant
11	F	2yrs	Reduced fetal movements and polyhydramnios. Neonatal hypotonia and respiratory difficulties. Severe muscle weakness: distal, proximal and bulbar.	Muscle biopsy was only performed in deceased brother: type I fiber predominance, the presence of rods suggestive of nemaline myopathy.	AR (affected brother)	NEG: *NEB* (pathogenic variants panel) gene.	CM NGS gene panel; WES (singleton)	Chapter 5.1
12	F	Deceased at the neonatal period	Severe neonatal hypotonia, lack of spontaneous movements (including respiratory). Myopathic appearance, arthrogryposis, bilateral femoral fractures, and thin gracile ribs.	Limited information as this study was performed in the context of an autopsy, atrophic muscle fibers reported	AR (affected siblings)	NEG: *SMN1, DMPK, ACTA1, MTM1, TNNT1, TPM2* and *TPM3*. Normal karyotype.	WES (singleton)	Chapter 4.3, Paper #7
13	M	11yrs	Facial asymmetry (6 months of age). Started walking at 12 months but with frequent falls. Axial weakness mainly in neck flexors. Partial facial paresis, scapular winging weakness, and atrophy (8yrs). Distal muscle weakness in lower limbs.	Myopathic features, fiber lesions compatible with myofibrillar myopathy (mainly desmin accumulation)	Sporadic	NEG: *FHL1* and *MYH7*.	CM NGS gene panel	Chapter 5.1

II. Myopathy with onset from early childhood to adolescence								
Patient #	Gen der	Age *	Main clinical features	Muscle histology	I. pattern Fam. HX	Previous genetic studies	Research strategy	Results
14	F	44yrs	Progressive muscle weakness (limb-girdle muscular dystrophy). Delayed motor development (delayed walking, difficulties in climbing stairs and running) and elevated CK levels (average ~850 U/l). Electromyography (7 yrs) myopathic. Currently, unable to rise from a chair, arms do not rise to shoulder level and walk with braces.	Dystrophic features (muscular dystrophy).	Sporadic	NEG: *ANO5, CAPN3, LARGE, FKRP, FKTN, POMGNT1, POMT1* and *POMT2* genes.	WES (Trio)	Chapter 4.2, paper #6
15	F	14yrs	Progressive muscle weakness (limb-girdle muscular dystrophy). Delayed motor development, lumbar lordosis, and elevated CK levels (1300 U/l).	Dystrophic features (muscular dystrophy).	Sporadic	NEG*: CAPN3, DMPK, FKRP* and *LMNA* genes.	WES (singleton)	Chapter 3.5, paper #4
16	F	25yrs	Progressive muscle weakness and atrophy affecting distinct muscle groups (ocular extrinsic, facial, pharyngeal, respiratory, cervical, axial, upper and lower limbs. Elevated CK levels (4x to 5x normal values).	Muscles changes compatible with a vacuolar myopathy (frequent rimmed vacuoles).	Sporadic	NEG: *GNE* gene.	WES (Trio)	Under research

III. Myopathy with onset during adulthood								
Patient #	**Gen der**	**Age ***	**Main clinical features**	**Muscle histology**	**I. pattern Fam. HX**	**Previous genetic studies**	**Research strategy**	**Results**
17	F	45yrs	Progressive muscle weakness (limb-girdle muscular dystrophy).	Dystrophic features (muscular dystrophy); occasional central nuclei in atrophic fibers (being this predominantly type 1 fiber).	Sporadic	*CAPN3* (one heterozygous variant); increased expression of *CAPN3* at the mRNA level.	Neurological NGS gene panel; WES (singleton)	Chapter 5.1
18	M	51yrs	Muscle weakness in lower limbs, with distal predominance; elevated CK levels (1500 U/l).	Dystrophic features (muscular dystrophy), IHC irregular staining for dysferlin.	Sporadic	NEG: *ANO5, CAPN3, CAV3, DYSF, FKRP, TCAP,* and *TTN* genes.	Neurological NGS gene panel; WES (singleton); overlap approach	Under research
19	M	55yrs	Muscle weakness in lower limbs, pelvic girdle, and quadriceps; elevated CK levels (2000 U/l).	Dystrophic features (muscular dystrophy), IHC irregular staining for dysferlin.	Sporadic	NEG: *ANO5, CAPN3, DYSF, FKRP,* and *TTN* genes.	Neurological NGS gene panel; WES (singleton); overlap approach	Under research
20	F	46yrs	Predominantly distal muscle weakness (more severe in lower limbs) with onset during the 4th decade of life. Elevated CK levels (2x normal values).	Myopathic features, fiber size variability. Replacement of muscle fibers by fat. Presence of rimmed vacuoles.	AD	NEG: *DYSF, CAPN3, FKRP* genes.	Neurological NGS gene panel	Chapter 5.1

IV. Dystrophinopathies								
Patient #	Gen der	Age *	Main clinical features	Muscle histology	I. pattern Fam. HX	Previous genetic studies	Research strategy	Results
21	M	20yrs	Delayed motor development (noticed at 2 yrs of age) and elevated CK levels. Mild intellectual disability, with poor language skills and reduced social interaction. Limb-girdle and scapular muscle atrophy, especially in deltoids. Dysmorphic facial features, with an ogival palate. No cardiac or respiratory involvement.	Dystrophic features (muscular dystrophy), IHC irregular staining for dystrophin	Sporadic	NEG: *DMD by* MLPA & Sanger sequencing (gDNA); FKRP, SGCA, SGCB, SGCG, and SGCD genes; triplet expansion in *FMR1* and *AFF2*, subtelomeric CNVs; normal karyotype	*DMD* transcripts analysis; Western-blot; WGS; bioinformatics	Chapter 5.1
22	M	51 yrs	Progressive proximal weakness of lower limbs with onset at 13 yrs of age, electromyography showing myopathic signs, and high CK levels.	Dystrophic features (muscular dystrophy), IHC irregular staining for dystrophin	Sporadic	NEG: *DMD* by MLPA, Sanger sequencing (gDNA)	*DMD* transcripts & bioinformatic analysis; 3rd-generation sequencing	Chapter 4.4, paper #8

Footnote: #- number; *- at June 2018; Fam. Hx- Family history; Heteroz.- Heterozygous; I. pattern- Inheritance pattern; yrs- years.

A total of 12 patients with CM without a genetic diagnosis after conventional studies (chapter 3) were studied using this strategy based on an NGS gene panel. Most of these sequencing experiments were performed in an Ion Personal Genome Machine™ (PGM™) System (Thermo Fisher Scientific, Waltham, MA, USA).
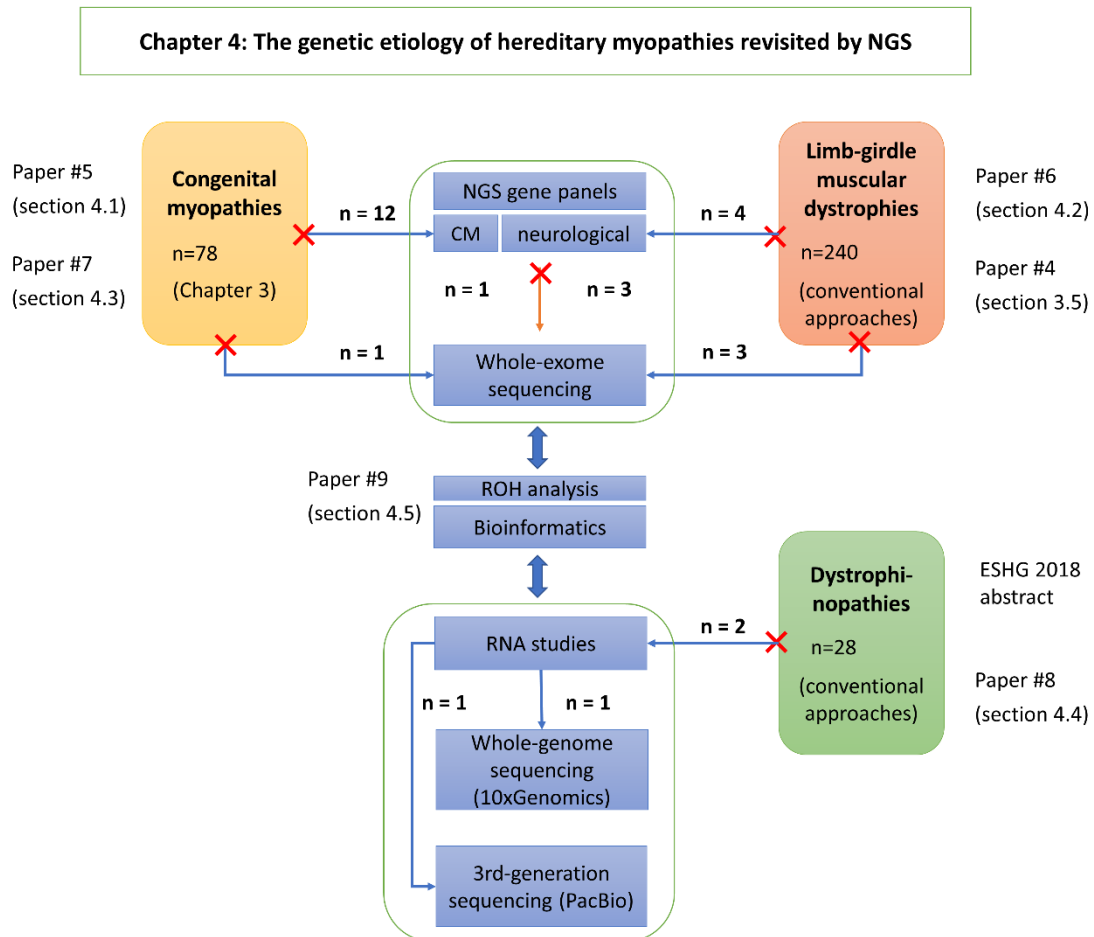
Figure 2.3.2- Experimental outline of Chapter 4.

A second NGS gene panel, made available for neurological diseases - Ion AmpliSeq™ Neurological Research Panel (Thermo Fisher Scientific, Waltham, MA, USA) was applied using the same technology. It was evaluated and tested in four patients with progressive MD with onset during adulthood. This panel was developed as a rapid and cost-effective method to identify pathogenic variants associated with inherited neurological disorders. It covers an extensive list of 751 genes that are known to harbor pathogenic variants affecting the functioning of the brain and of the nervous system. In addition to target all coding regions, 101 non-coding disease-causing *loci* and 14 regions that were reported to contain repeat-

expansion in different genes are targeted by this gene panel (Appendix I.1). These studies were performed in an Ion S5 sequencer (Thermo Fisher Scientific, Waltham, MA, USA).

To deal with extreme genetic heterogeneity and as an attempt to identify novel genetic causes for hereditary myopathies, WES analysis was performed. For exome capture, Ion AmpliSeq™ Exome Panel was used and sequenced in an Ion Proton System (Thermo Fisher Scientific, Waltham, MA, USA). This large-scale NGS analysis may contribute to broadening considerably the phenotypic spectrum of hereditary myopathies and even, in a most optimistic perspective, the identification of novel disease-causing genes. However, this objective by itself is extremely challenging and requires the development of novel strategies for data analysis, validation of variants pathogenicity thus obtaining a consistent association between genotype and phenotype. In this subtopic, a total of eight patients were analyzed by WES, being in six a singleton-exome and in two cases the trio analysis (exomes of the patient and both parents) was performed. Thus, a total of 12 exomes were sequenced in this work. Half of the patients enrolled in this task were primarily subject to an NGS gene panel (CM or neurological as outlined above) which turned out to be negative/inconclusive. In the other four patients, WES was performed as a first-level research approach. The underlying reasons towards this decision consisted in the presence of atypical phenotypes which would not fit in any known NGS gene panel or when clinical presentations were suggestive towards the application of more than one gene panel (congenital *versus* progressive MD for instance).

Analysis of NGS data, specially WES-derived, required several bioinformatic tools to address different analytical steps such as for reads alignment, variants' calling, annotation, and filtering, but also for the applicability of different disease inheritance models (Gilissen et al., 2012). Application and optimization of bioinformatic algorithms and pipelines were a central aspect of chapter 4, making decisive contributes towards the identification and clarifying the relevance of genetic defects some of considerable unique nature. One particularly pertinent strategy included homozygosity mapping or the detection of long runs of homozygosity (ROH). This approach was applied in this work for gene discovery and for disease-causing variants identification in recessive conditions. ROH was traditionally performed resorting to SNP arrays, WES data, however, has been increasingly used for this analysis having the additional advantageous possibility of variant identification.

Finally, also in chapter 4 two cases of dystrophinopathies were thoroughly investigated by several techniques including RNA studies (*DMD* transcripts analysis). These patients were selected, from a subset of twenty-eight genetically unsolved cases, upon their high level of clinical suspicion, namely a phenotype suggestive of BMD and irregular dystrophin staining in immunohistochemistry performed in muscle biopsy. The routine studies by MLPA and Sanger sequencing have failed to identify a pathogenic variant in *DMD*. Being the transcriptomic data suggestive of the gene's involvement, bioinformatic strategies were used followed by: i) whole-

genome sequencing (Chromium™ Genome Solution, 10x Genomics, Pleasanton, CA, United States) (Elyanow et al., 2017); or ii) single molecule real-time sequencing (PacBio RS II System, Pacific Biosciences, CA, United States) (Rhoads & Au, 2015). The combination of bioinformatics, transcriptomic and genomic approaches were pivotal for providing further clarification of the mutational events causing disease.

Chapter 3

EXPANDING THE VARIOME OF CONGENITAL MUSCLE DISEASES

Contents

3.1 Congenital muscle diseases: genetic profiles in a large patient cohort

This section recapitulates the genetic analysis of 246 patients with congenital muscle diseases (CMD and CM) studied over a period of 13 years (from 2004 to 2017). As explored in further detail in the General Introduction, these diseases are individually very rare and only throughout these years their genetic causes have been progressively established, through the identification of new disease *loci*. During this time-period I have implemented in a nationwide basis, the genetic analysis by conventional sequencing (Sanger sequencing) of several genes related with congenital muscle diseases. These were gradually introduced into genetics routine diagnostics and complemented with *MLPA* analysis, a technique which is not available for all loci. Therefore, the data collectively presented wherein derives from a rather heterogeneous patient cohort, which includes individuals from different geographical origins.

Overall, pathogenic variants were identified in 112 patients (106 families) out of the 246 analyzed, leading to an overall pathogenic detection rate of 45.5%. This lower detection rate reflects the challenges concerning the genetic analysis of these diseases. Table 3.1.1 summarizes the overall results obtained for these 112 patients. The list of genes with pathogenic variants comprises thirteen different *loci*. Variants in three genes (*LAMA2*, *MTM1* and *RYR1*) account for more than three quarters of all cases in our cohort. These findings are concordant with previous publications (Maggi et al., 2013; Sframeli et al., 2017). As for *LAMA2*-related diseases, ten patients with a late-onset phenotype - defined as having onset beyond the first two years - were also identified. Whereas variants in genes associated with dystroglycanopathies were only found in 7 patients (6.3% of all characterized patients), which is clearly lower as compared with other publications (Sframeli et al., 2017). In seven genes (*B4GAT1*, *BIN1*, *ISPD*, *LARGE1, POMGNT2*, *POMT1* and *SELENON*) no pathogenic variants have been identified. It was interesting to note that the *SELENON* gene is included in this list since it has been reported by other international groups as an important cause of CM and CMD with rigid spine (Maggi et al., 2013; Sframeli et al., 2017). The scarcity of Portuguese patients with SELENON-related myopathies is corroborated by a review of the literature, where only one of such patients has been reported so far (Ferreiro et al., 2002). Future epidemiological research might provide an explanation for this observation.

As shown in Table 3.1.2 the majority of the variants have been identified in *LAMA2* (37.4%) *RYR1* (25.3%) and *MTM1* (12.1%) genes. In terms of their distribution by type, missense variants are the more represented category (43.4%). It is followed by nonsense variants, small out-of-frame (OF) deletions or duplications and changes affecting splicing, all with a total of fifteen variants each. Eight variants (~8% of total unique variants) were classified as large deletions or duplications. Whereas, only three are small deletions or duplications which maintain the reading frame.

Table 3.1.1- Overall results obtained from 112 genetically characterized patients with congenital muscle diseases.

Gene Symbol	Disease / disease group	Number of patients with (pathogenic) disease-causing variants	Frequency within genetically characterized patients (%, total n=112)	Data & references
ACTA1	CM	2	1.8%	Appendix II.1 (Table II.1.1)
COL6A1	CMD	5	4.5%	Appendix II.1 (Table II.1.1)
COL6A2	CMD	5	4.5%	Appendix II.1 (Table II.1.1)
COL6A3	CMD	1	0.9%	Appendix II.1 (Table II.1.1)
DNM2	CM & CMD	1	0.9%	Appendix II.1 (Table II.1.1)
FKRP	CMD	1	0.9%	Oliveira et al., 2016b [1] & Appendix II.1 (Table II.1.1)
FKTN	CMD	1	0.9%	Costa et al., 2013
LAMA2	CMD (+ progressive MD*)	52 (+10*)	46.4%	Oliveira et al., 2008b; Marques et al., 2014; Papers in sections 3.4 and 3.5
LMNA	CMD	3	2.7%	Appendix II.1 (Table II.1.1)
MTM1	CM	12	10.7%	Paper in section 3.2; Appendix III.1 (Table III.1) (n=7)
POMGNT1	CMD	3	2.7%	Oliveira et al., 2008a; Oliveira et al., 2016b [1] & Appendix II.1 (Table II.1)
POMT2	CMD	2	1.8%	Oliveira et al., 2016b [1] & appendix II.1 (Table II.1.1)
RYR1	CM & CMD	24	21.4%	Duarte et al., 2011; paper in section 3.3; Appendix II.1 (Table II.1.1) (n=11)

Footnote: *- A total of ten cases with *LAMA2*-related MD with latter onset (non-congenital) were also characterized; [1]- Published as an abstract in the European Human Genetics Conference (2016).

It is interesting to note that the majority of variants classified as being missense were found in *RYR1*, whereas those of nonsense or small OF deletions/duplication type were identified in *LAMA2* gene. As for the large deletions or duplications were also predominantly reported in *LAMA2.*

Table 3.1.2- Overview of variants classified according to type and zygosity.

	Type of variant							Zygosity	
	Total	MS	NS	Small OF del/dup	SPL	Small IF del/dup	Large del/dup	Het.	Hom. [Hem.]
ACTA1	2	2	0	0	0	0	0	2	0
COL6A1	6	4	0	0	2	0	0	6	0
COL6A2	4	0	1 (3)	0	2	1	0	4	2
COL6A3	2	2	0	0	0	0	0	2	0
DNM2	1	1	0	0	0	0	0	1	0
FKRP	2	1	0	1	0	0	0	2	0
FKTN	1	0	0	0	0	0	1	0	1
LAMA2	37	3	10 (32)	11 (18)	7 (14)	1	5 (15)	62	21
LMNA	3	2	0	0	1	0	0	3	0
MTM1	12	6	3	0	1	0	2	0	[12]
POMGNT1	3	1	0	0	2 (3)	0	0	2	2
POMT2	1	1	0	0	0	0	0	0	1
RYR1	25	20 (34*)	1	3	0	1	0	39	0
Total	99	43	15	15	15	3	8	123	27 [12]

Footnote: Between brackets are indicated the total number of variants; *- three missense variants in the same allele (n=2), and two missense variants in the same allele (n=2); del/dup- deletion or duplication; IF- in-frame; MS- missense; NS- nonsense; OF- out-of-frame; SPL- splicing affecting variant; Het.- heterozygous; Hom.- homozygous; Hem.- hemizygous.

In terms of zygosity most variants were found in heterozygosity. It is interesting to note the overall high allelic diversity found in the 112 patients included in this study, where a total of 99 distinct variants were identified, which translates to an average of 0.88 variants per case. Additional data is shown in Appendix II.1 (Table II.1.1) which includes a total of 48 different variants identified in 40 patients, where 23 variants have not been previously reported in the literature. Finally, considering the genetic entities with higher prevalence within this cohort, the work was in a privileged position to provide further important relevant insights into the genotypes and phenotypes of these diseases. Therefore, genetic and clinical data is presented in more detail in the following sections of this chapter.

3.2 EXPANDING THE *MTM1* MUTATIONAL SPECTRUM: NOVEL VARIANTS INCLUDING THE FIRST MULTI-EXONIC DUPLICATION AND DEVELOPMENT OF A LOCUS-SPECIFIC DATABASE

OLIVEIRA J, OLIVEIRA ME, KRESS W, TAIPA R, PIRES MM, HILBERT P, BAXTER P, SANTOS M, BUERMANS H, DEN DUNNEN JT, SANTOS R. (2013). EUROPEAN JOURNAL OF HUMAN GENETICS, 21(5):540-549.

3.2.1 ABSTRACT

Myotubular myopathy (MIM#310400), the X-linked form of Centronuclear myopathy (CNM) is mainly characterized by neonatal hypotonia and inability to maintain unassisted respiration. The *MTM1* gene, responsible for this disease, encodes myotubularin– a lipidic phosphatase involved in vesicle trafficking regulation and maturation. Recently, it was shown that myotubularin interacts with desmin, being a major regulator of intermediate filaments. We report the development of a locus-specific database for *MTM1* using the Leiden Open Variation database software (http://www.lovd.nl/MTM1), with data collated for 474 mutations identified in 472 patients (by June 2012). Among the entries are a total of 25 new mutations, including a large deletion encompassing introns 2–15. During database implementation it was noticed that no large duplications had been reported. We tested a group of eight uncharacterized CNM patients for this specific type of mutation, by multiplex ligation-dependent probe amplification (MLPA) analysis. A large duplication spanning exons 1–5 was identified in a boy with a mild phenotype, with results pointing toward possible somatic mosaicism. Further characterization revealed that this duplication causes an in-frame deletion at the mRNA level (r.343_444del). Results obtained with a next generation sequencing approach suggested that the duplication extends into the neighboring *MAMLD1* gene and subsequent cDNA analysis detected the presence of a *MTM1/MAMLD1* fusion transcript. A complex rearrangement involving the duplication of exon 10 has since been reported, with detection also enabled by MLPA analysis. It is thus conceivable that large duplications in *MTM1* may account for a number of CNM cases that have remained genetically unresolved.

3.2.2 INTRODUCTION

Congenital myopathies are a heterogeneous group of diseases, generally characterized by muscle weakness, and with onset at birth or during infancy. These myopathies have characteristic histological hallmarks in muscle biopsy allowing the differential classification in distinct entities: centronuclear myopathy (CNM), core myopathy (centralcore and minicore diseases) and nemaline rod myopathy (Romero, 2010; Nance et al., 2012). In CNM, the most prominent histopathological features include hypotrophy of type 1 fibers and a high frequency of centrally located nuclei with perinuclear halos lacking myofilaments and occupied by mitochondrial and glycogen aggregates.[1] Several genes are reported to be associated with CNM; these include *MTM1* in the X-linked form (Laporte et al., 2010; Biancalana et al., 2012) *DNM2* and *MTMR14* in the autosomal dominant forms (Bitoun et al., 2005; Tosch et al., 2006; Nicot et al., 2007; Biancalana et al., 2012), *BIN1* and *RYR1* associated with the autossomal recessive forms (Wilmshurst et al., 2010; Bevilacqua et al., 2011; Biancalana et al., 2012).

X-linked myotubular myopathy (XLMTM; MIM 310400) has a prevalence of approximately 1/50,000 males and is characterized by severe hypotonia present at birth and inability to maintain sustained spontaneous respiration (Jungbluth et al., 2008).

Different authors have proposed that patients be classified according to their phenotype, as: i) *severe* - characteristic facial features, markedly delayed motor milestones and requiring prolonged ventilatory support (>12 hours); ii) *moderate* – more rapid acquirement of motor milestones and independent respiration for >12 hours per day; iii) *mild* – motor milestones slightly delayed and independent spontaneous respiratory function achieved after the neonatal period (Herman et al., 1999; McEntagart et al., 2002). Carrier females are usually asymptomatic, however there are several records of manifesting heterozygotes due to skewed X chromosome inactivation (Tanner et al., 1999; Hammans et al., 2000; Sutton et al., 2001; Jungbluth et al., 2003; Schara et al., 2003; Pénisson-Besnier et al., 2007).

The *MTM1* gene (in Xq28) is composed of 15 exons and has an open reading frame of 1.8 kb encoding the myotubularin protein. Structurally, myotubularins are constituted by four characteristic domains: the PTP (protein tyrosine phosphatase), the predicted GRAM (glucosyltransferases, Rab-like GTPases activators and myotubularins), the RID (Rac-induced recruitment domain) and SID (SET-interaction domain). Functionally, myotubularin is a phosphatase acting specifically on PtdIns3P and PtdIns(3,5)P2, two phosphoinositides (PIs). PIs participate in the regulation of various cellular mechanisms by direct binding to PI-binding domains of effector proteins (that control membrane/vesicular trafficking) and subsequent recruitment/activation of these at specific membrane sites. PtdIns3P and PtdIns(3,5)P2 have a direct role in the endosomal-lysosomal pathway (Robinson et al.,

2006). The PTP-catalytic domain of myotubularins is responsible for the phosphoester hydrolysis of the 3-phosphate of PIs. This hydrolysis involves two residues of cysteine and arginine located on a Cys-X5-Arg motif, characteristic for the PTP domain (Laporte et al., 2001). The loss of phosphatase activity or the production of truncated proteins as a result of *MTM1* mutations could lead to abnormal dephosphorylation of PtdIns3P/PtdIns(3,5)P2 and subsequent abnormal trafficking of the effector proteins of the endosomal-lysosomal pathway (Pénisson-Besnier et al., 2007). Similar results were observed with mutations located in the GRAM domain of myotubularin. This domain of about 70 aminoacids is responsible for PtdIns(3,5)P2 binding (Tsujita et al., 2004). Recently mitochondrial homoeostasis in muscle fibres and regulation of the desmin cytoskeletal system was attributed to mytobularin. It was experimentally demonstrated that myotubularin interacts with desmin, and that this complex is disrupted by specific mutations in the *MTM1* gene (Hnia et al., 2011).

It is recognized that, despite the genetic advances in this field and the large number of cases reported (Laporte et al., 2000), a significant number of CNM cases remain genetically unresolved. This may be explained either by the involvement of further gene *loci* or by the presence of mutations in known genes, that are not detectable by routine techniques. During the development of a LSDB (Locus-specific database) for the *MTM1* gene we noticed that no large duplications (involving one or more exons) had been reported for this gene. This observation led us to investigate the possibility of their occurrence in molecularly unresolved CNM patients. Accordingly, we report the first multi-exonic duplication in *MTM1* (exons 1 to 5) detected in a male patient with a mild XLMTM phenotype. In line with this finding, a complex rearrangement involving the duplication of exon 10 was recently published (Trump et al., 2012).

3.2.3 MATERIAL AND METHODS

MTM1-LOVD database development

The *MTM1* mutation database was implemented using LOVD v2.0 software (Fokkema et al., 2011), and is integrated in Leiden Muscular Dystrophy pages (http://www.dmd.nl/). Currently (by 29th June 2012), this large database installation displays a total of 141 genes with 70512 variants (9265 unique) described in 55775 individuals. The *MTM1* database is subdivided into two main tables, for variant data (n=20) and for patient/clinical items (n=22), the latter being shared among the different LSDBs in the Leiden Muscular Dystrophy pages (example of an entry in Appendix II.2 Figure II.2.1). Data were retrieved from peer-reviewed literature and new variants were directly submitted by different sources (laboratories and clinicians). The curators' tasks included confirmation of information concerning the sequence variants, especially with regard to their descriptions following the recommendations of the

Human Genome Variation Society (HGVS) (den Dunnen et al., 2000), using the cDNA reference sequence NM_000252.2. Classification of the clinical phenotype (severe, moderate or mild) was added to the patient's entry when adequate information was reported or submitted.

New *MTM1* variants

Variants submitted to MTM1-LOVD that had not been described previously, are presented here. These were reported from different centers (Belgium, Germany and Portugal) and submitted to MTM1-LOVD. Information regarding the patient's phenotype and the mutation origins was also collected. Bioinformatic analysis of sequence variants, in particular missense and splicing mutations, was performed using Polyphen v2 (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al., 2010) and the Human Splicing Finder v2.3 (http://www.umd.be/HSF/) (Desmet et al., 2009). Phylogenetic conservation analysis of the affected residues and population screening of variants were also performed.

Studies performed in patient 1

Clinical description. Patient 1 (P1) is a 7 year old boy born to a non-consanguineous couple. He has a healthy younger sister. Prenatal manifestations included oligohydramnios and signs of premature birth. Delivery was at 35 weeks by caesarean section, with an Apgar score of 8/9, weight 2465g, height 46cm and head circumference 33cm. During the neonatal period the only clinically relevant sign was facial paresis. There were no feeding or respiratory difficulties. During the first month his paediatrician noticed that he was hypotonic and started global stimulation. He started to walk at 21 months. At the age of two he was referred to a paediatric neurology clinic. He was shy and had poor language skills. There was limitation in the abduction of the right eye, facial diparesis with the left side more exacerbated, with a lagophtalmos. Proximal tetraparesis was detected; the patient could not raise his arms completely and was unable to stand up from the floor without bilateral help. He had a waddling gait and could not run. CK levels were normal (36 U/L). Biopsy was performed at age 3. Over the last 3 years muscle weakness progressed and there has been a gradual loss of motor skills.

Multiplex ligation-dependent probe amplification (MLPA) analysis. Screening for duplications and deletions in *MTM1* was performed by the MLPA technique using the P309-A1 Probe Set (MRC-Holland, Amsterdam). This contains 16 probes for the *MTM1* gene, 7 probes for the *MTMR1* gene, 3 probes located on Xq28 (*DKC1* and *FLNA* genes) and 11 reference probes for distinct regions on the X chromosome (Appendix II.2 Table II.2.1). gDNA samples of P1 and four healthy controls (150 ng each) were used in the procedure.

Products were separated by capillary electrophoresis on an ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, CA). Data analysis was conducted using GeneMarker® software (SoftGenetics LLC, State College, PA). Population normalization method was selected and data were plotted using probe ratio.

Southern blotting and hybridization. gDNA samples from P1, his mother and controls were digested with EcoRI (New England Biolabs, Beverly, MA) and resolved on a 0.8% agarose gel. DNA fragments were vacuum-transferred to a nylon membrane using a saline method. A cDNA probe recognizing *MTM1* exons 2 to 7 was prepared using DIG (digoxigenin) DNA Labeling Kit (Roche Applied Science, Indianapolis, IN) and incubated overnight using the Easy Hyb Buffer (Roche Applied Science, Indianapolis, IN). The membrane was prepared with DIG Wash and Block Buffer Set (Roche Applied Science, Indianapolis, IN), incubated with Anti-DIG-AP conjugate (Roche Applied Science, Indianapolis, IN), and the DIG-labeled probe detected with ready-to-use CDP-Star (Roche Applied Science, Indianapolis, IN).

cDNA analysis. Total RNA was extracted from muscle samples of P1 and controls using the PerfectPure RNA Fibrous Tissue Kit (5 PRIME, Hamburg) and converted to cDNA using the High Capacity RNA-to-cDNA Kit (Applied Biosystems, Foster City, CA). *MTM1* transcripts were subjected to PCR amplification of the region(s) encompassing exons 2 to 7, using the specific primers (2F: 5'-TCCAGGATGGCTTCTGCATC-3' and 7R: 5'-CAAGCCCTGCCTCCTGTATTC-3'). For the detection of the *MTM1/MAMLD1* fusion transcript total RNA was extracted from blood using the PerfectPure RNA Blood Kit (5 PRIME, Hamburg). After cDNA conversion, PCR amplification was performed using the primer mentioned above for exon 2 of *MTM1* and a reverse primer for exon 5 of *MAMLD1* (cMAMLD1-5R: 5'-AGTCTGGCCTGAGTGTGAGAGG-3'). PCR products were purified, sequenced using BigDye® Terminator Cycle Sequencing Kit v1.1 (Applied Biosystems, Foster City, CA) and resolved on an ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, CA).

Next Generation Sequencing. gDNA was Covaris (S-series) sheared to an average size of 300 to 400 bp. An Illumina sequencing library was prepared using the NEBNext kit (New England Biolabs, Beverly, MA) without modifications and using Illumina Truseq adapters. No library pre-amplification was performed. Sequencing data was generated on an Illumina Hiseq2000 (Illumina, San Diego, CA) using standard instrument settings for flowcell clustering and paired-end sequencing of 2x100 bp reads on a flowcell with v3 reagents. Fastq files were generated using the Illumina Casava v1.8 pipeline. Stampy (v1.0.13) was used to align the sequence data to the human genome (Hg19) using standard parameters.

The total number of aligned reads per 10,000 bp bins across the whole genome was calculated using the Bedtools package (Quinlan et al., 2010). Bins overlapping more than 1% with simple-repeat regions were excluded from further processing. For sample comparison of tag coverage in the *MTM1* locus, the bin counts were normalized relative to total number of reads aligned to the chromosome X. For visualization purposes, BED tracks were generated for the patient, control and the ratio patient/control, and were uploaded to the UCSC genome browser (Kent et al., 2002).

Studies performed in patient 2

Clinical description. Patient 2 (P2) was the first child of healthy parents, with no consanguinity or family history of note. After delivery he was very hypotonic and required immediate respiratory support with CPAP (continuous positive airway pressure) and naso jujenal feeding. At two months of age he underwent tracheal endoscopy and aryepiglottoplasty after which he needed ventilatory support. After extubation he was transferred for intensive care and was diagnosed with congenital nystagmus and pyloric stenosis. At three months he was referred for a neurology opinion. He had a frog like posture, absent antigravity movement in the upper limbs, with some in the lower limbs, and areflexia. At pyloromyotomy a quadriceps muscle biopsy was taken which confirmed the diagnosis of centronuclear myopathy. From 5 months respiratory function improved and by 7 months he was off all respiratory support. He began to lift each leg off the cot for 1-2 minutes and the nystagmus disappeared, with normal eye movements, but he had persisting marked neck weakness. Following an episode of lung collapse he began to show restriction of lateral and vertical eye movements. He was able to lift his arms against gravity and bring his hands together as well as to hold a rattle, and was able to lift his legs momentarily against gravity. He had a marked scoliosis, a bell shaped chest and there was early hip and knee flexion contractures. Placed prone he was unable to turn his head or push himself up on his arms. At 8 months he developed bronchiolitis and pulmonary collapse as well as a brief cardiac arrest. Following this he was re-intubated and given ventilatory support. After recovery he was much less active and lost antigravity limb movement and eye movements. At the age of 10 months he became unwell again and died.

Molecular analysis. Initially gDNA was PCR amplified for subsequent sequencing. Since no products were obtained for exons 3 to 14, a large hemizygous *MTM1* deletion was suspected. Deletion was confirmed by MLPA analysis, as described for P1. The breakpoints of the large deletion encompassing exons 3 to 14 were determined by long-range PCR. Amplification of gDNA was performed using the BIO-X-ACT™ Long DNA Polymerase Kit (Bioline, Taunton, MA) and primers complementary to introns 2 and 14 of *MTM1* (2iF: 5'-

GAAAGGTTGCTGAAGGACATACTG-3' and 14iR: 5'-GCCTTGGGTATGAATGCTGG-3').
PCR products were resolved on a 2% agarose gel and purified with QIAquick Gel Extraction
Kit (QIAGEN, Valencia, CA), followed by sequencing.

3.2.4 Results

MTM1-LOVD

A total of 496 variants are currently (June 2012) listed in the *MTM1* LSDB. Besides non-pathogenic variants (n=18) or sequence changes with unknown significance (n=4), this database includes 474 *MTM1* mutations identified in 472 XLMTM patients. The mutation profile of the *MTM1* gene, according to LOVD-MTM1 data, is summarized in Table 3.2.1 and Figure 3.2.1.

Large deletions encompassing one or more *MTM1* exons account for 6.3% of mutations reported. The eighteen different large deletions reported to date (identified in 30 cases) are detailed in Figure 3.2.1B. Small duplications were identified in 6.5% of total mutations, all but one case predictably creating PTC. Mutations that consist in a deletion associated with sequence insertion (delins) were identified in 7 cases (1.5%), whereas simple insertions were reported in 6 patients (1.3%). Only two cases (0.4%) involving large (exonic) duplications were reported to date. One patient was reported as having a complex rearrangement that involves the duplication of exon 10 (Trump et al., 2012). Patient P1 presented in detail in this work is the only recorded case with a multi-exonic duplication of *MTM1*. A total of 55 cases were not reported in the literature and were submitted directly to the MTM1-LOVD.
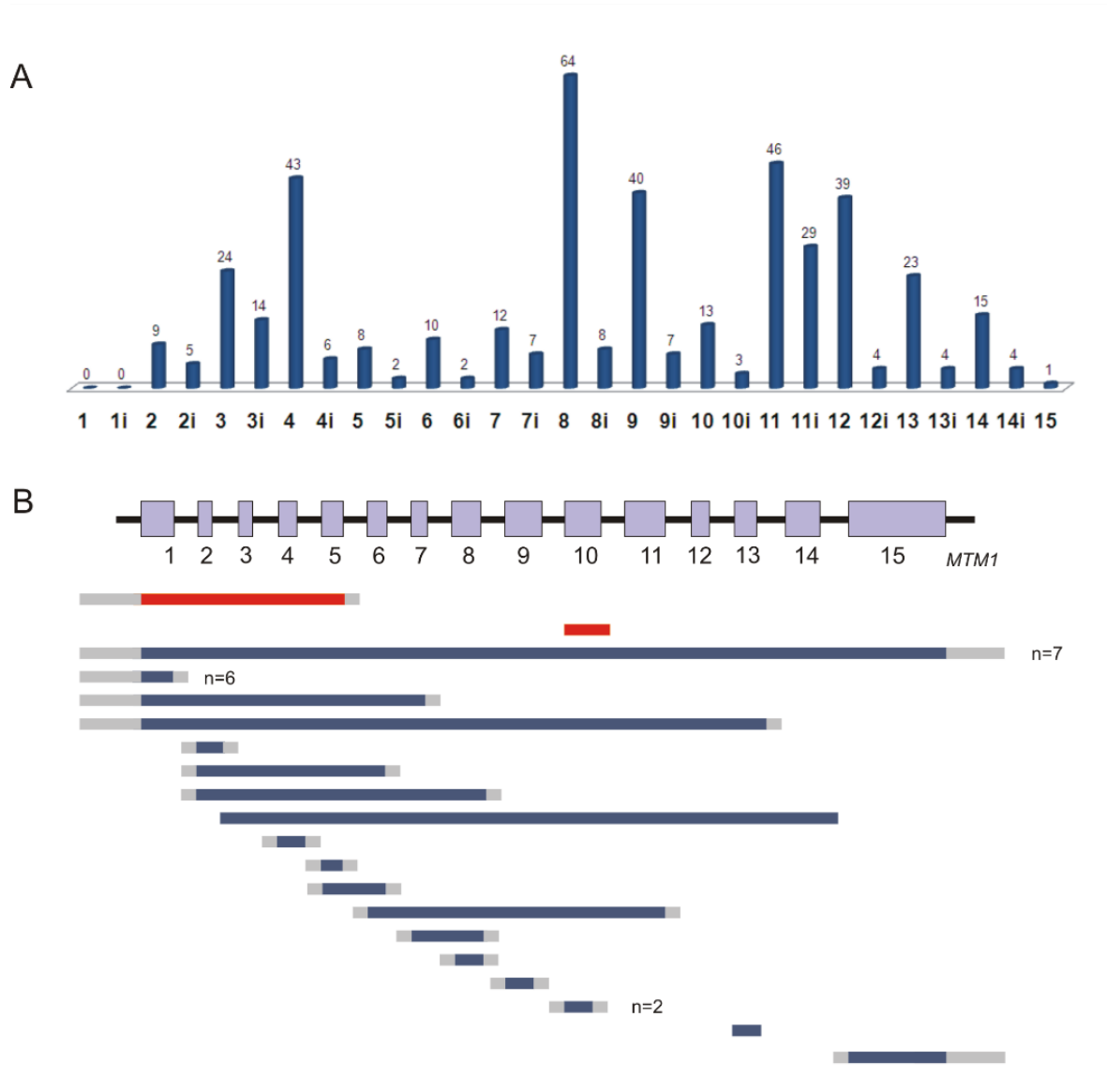
Figure 3.2.1- Mutations registered in MTM1-LOVD as distributed along the *MTM1* gene. (A) Total number of small mutations is shown for each exon and intron of *MTM1*. (B) Large deletions and duplications identified in *MTM1*. Solid bars represent regions involved (blue - deletions; red - duplications); grey bars represent undelineated breakpoints. In cases with more than one independent report the total number (n) of patients is indicated.

Table 3.2.1- Overview of mutations in *MTM1*-LOVD

DNA	RNA[a]				Protein[b]							Total
								del dup Ins delins				
Mutation type	r.(0?)	del dup ins subs	r.(spl?)	r.(?)	p.0?	Missense	Nonsense	IF	OF	MP	p.(?)	**Total**
Substitutions	-	26	62	238	5	145	91	35	8	3	39	326 [152]
Deletions	-	2	7	63	-	-	3	10	52	-	7	72 [50]
Duplications	-	-	1	30	-	-	7	-	23	-	1	31 [23]
Insertions	-	1	-	5	-	-	-	1	5	-	-	6 [6]
Deletion /insertions	-	-	2	5	-	-	1	-	4	-	2	7 [7]
Large deletions*	15	2	-	13	18	-	-	10	1	-	1	30 [18]
Large duplications*	-	2	-	-	-	-	-	2	-	-	-	2 [2]
												474 [258]

Footnote: Data extracted from MTM1-LOVD (29/06/2012). (a) - Experimentally tested at RNA level (del, dup, ins, subs) or predicted (remaining columns). (b) - Predicted protein changes inferred from DNA or RNA sequence. Numbers between square brackets indicate the total count of unique (different) mutations. r.(0?) - predictably no expression at RNA level; del - deletion; dup - duplication; ins - insertion; subs - substitution; r.(spl?) - predicted to affect splicing; r.(?) - RNA effect unknown; p.0? - predictably no protein production; delins - deletion and insertion; IF - in-frame; OF - out-of-frame; MP - multiple polypeptides; p.(?) - effect at protein level unknown. (*) - involving one or more exons.

New *MTM1* variants

The twenty-five novel point mutations identified in CNM male patients are described in Table 3.2.2. Among these, eleven are single nucleotide substitutions located within exonic sequences: c.2T>A, c.32C>A, c.323G>A, c.469G>T, c.595C>A, c.637C>T, c.659G>C, c.1241T>C, c.1247A>G, c.1318C>T and c.1600T>C. With the exception of c.2T>A that predictably affects the initiation codon of myotubularin (p.0?), the majority of these substitutions are predicted to be missense mutations (p.Gly108Asp, p.Pro199Thr, p.Leu213Phe, p.Arg220Thr, p.Phe414Ser, p.His416Arg and p.Trp534Arg). The remaining three are nonsense mutations (p.Ser11*, p.Glu157* and p.Gln440*). A condensed view of all data corroborating the pathogenicity of missense changes is presented in Appendix II.2 Table II.2.2. Briefly, missense variants were considered pathogenic when affecting phylogenetically conserved residues and/or were not detected in ethnically matched control chromosomes. Three single nucleotide substitutions are located in intronic donor splice consensus sequences and predictably affect splicing: c.231+1G>T, c.342+5G>A and c.867+1G>A. In one of these changes (c.342+5G>A) cDNA studies demonstrated that it promotes exon 4 skipping (r.232_342del) leading to an in-frame deletion at the protein level (p.Ser79_Asp115del).

In addition, four deletions (c.1088_1089del, c.1328_1331del, c.1509_1510del and c.1519_1522del), two duplications (c.509_528dup and c.596dup) and two deletion/insertion mutations (c.765_767delinsGG and c.1319_1321delinsTA) were also detected. All of these variants are predicted, at the RNA and protein level, to be frame-shift mutations that generate premature termination codons. An insertion of 376bp in exon 13 (c.1388_1389ins376, GenBank JQ403527) was detected in a patient with a severe phenotype. Using the Repbase database (Jurka et al., 2005), this insertion was identified as an *AluYa5* sequence. Further studies revealed that this alteration affects exon 13 splicing (Appendix II.2, Figure II.2.1), predictably resulting in an in-frame deletion of 38 aminoacids (p.Phe452_Gln489del) located in the SID region of myotubularin. The remaining two novel mutations are the large deletion and duplication presented below in more detail.

Table 3.2.2- Novel *MTM1* mutations submitted to the *MTM1*-LOVD

Patient ID in MTM1 LOVD	DNA mutation	Gene location	Type of mutation	cDNA effect	Protein effect	Origin of mutation	Phenotype	Geographic origin
18914	c.=/(?_-76)_342+?dup	Exons 1-5	Large duplication	r.[=, 343_444del]	p.Asp115_Leu148del	*de novo,* somatic	Mild	Portugal
10959	c.63+834_1645-2104del	Intron 2-15	Large deletion	r.(?)	p.(Thr22_Gln548del)	Inherited (MC)	Severe	United Kingdom
24703	c.2T>A	Exon 2	Affects initiation codon	r.(?)	p.0?	Inherited (de novo, in mother)	Severe (at birth)	Germany
21000	c.32C>A	Exon 2	Nonsense	r.(?)	p.(Ser11*)	*de novo*	Severe	Germany
24706	c.231+1G>T	Intron 4	Donor splice site disruption (a)	r.(spl?)	p.(?)	Inherited (MGC)	Severe (died at 10 months)	Germany
20980	c.323G>A	Exon 5	Missense	r.(?)	p.(Gly108Asp) (b)	*de novo*	Mild (24 years old, still walking)	Germany
18861	c.342+5G>A	Intron 5	Donor splice site disruption	r.232_342del	p.Ser79_Asp115del	Inherited (MC)	Severe	Germany
20960	c.469G>T	Exon 7	Nonsense	r.(?)	p.(Glu157*)	Inherited (MC)	Severe	United Kingdom
20961	c.509_528dup	Exon 7	Out-of-frame duplication	r.(?)	p.(Gly177Trpfs*16)	Unknown	Severe (long-term survivor)	Germany
20933	c.595C>A	Exon 8	Missense	r.(?)	p.(Pro199Thr) (b)	Inherited (MC)	Unknown (neonatal death)	Portugal
24711	c.596dup	Exon 8	Out-of-frame duplication	r.(?)	p.(Ala200Cysfs*12)	Inherited (MC)	Severe	Germany
24702	c.637C>T	Exon 8	Missense	r.(?)	p.(Leu213Phe) (b)	Inherited (MC)	Mild (diagnosed at 24 years of age)	Germany

Table 3.2.2- Novel *MTM1* mutations submitted to the MTM1-LOVD (continues)

Patient ID in MTM1 LOVD	DNA mutation	Gene location	Type of mutation	cDNA effect	Protein effect	Origin of mutation	Phenotype	Geographic origin
24707	c.659G>C	Exon 8	Missense	r.(?)	p.(Arg220Thr) (b)	Unknown	Severe (at birth)	Germany / Albania
20951	c.765_767delinsGG	Exon 9	Out-of-frame deletion	r.(?)	p.(Asp256Valfs*28)	Inherited (MC)	Severe	Turkey
11124	c.867+1G>A	Intron 9	Donor splice site disruption (a)	r.(spl?)	p.(?)	Inherited (MC)	Unknown	Portugal
16734	c.1088_1089del	Exon 11	Out-of-frame deletion	r.(?)	p.(Lys363Serfs*14)	Inherited (MC)	Severe	Germany
11170	c.1241T>C	Exon 11	Missense	r.(?)	p.(Phe414Ser) (b)	Unknown	Unknown	Portugal
12852	c.1247A>G	Exon 11	Missense	r.(?)	p.(His416Arg) (b)	Unknown	Severe	Germany
20944	c.1318C>T	Exon 12	Nonsense	r.(?)	p.(Gln440*)	Inherited (MC)	Severe	Germany
20955	c.1319_1321delinsTA	Exon 12	Out-of-frame deletion/insertion	r.(?)	p.(Gln440Leufs*24)	Inherited (MC)	Severe	Germany
24712	c.1328_1331del	Exon 12	Out-of-frame deletion	r.(?)	p.(Asp443Valfs*20)	Unknown	Severe	Germany
14208	c.1388_1389ins376, GenBank JQ403527	Exon 13	Exon skipping	r.1354_1467del	p.Phe452_Gln489del	Inherited (de novo, in mother)	Severe	Turkey
20938	c.1509_1510del	Exon 14	Out-of-frame deletion	r.(?)	p.(Asn503Lysfs*2)	Unknown	Severe	United Kingdom
20952	c.1519_1522del	Exon 14	Out-of-frame deletion	r.(?)	p.(Glu507Asnfs*28)	Inherited (MC)	Severe	Germany
24705	c.1600T>C	Exon 14	Missense	r.(?)	p.(Trp534Arg) (b)	Inherited (MGC)	Severe (at birth)	Germany

Footnote: ID - Patient identification in database; r.(spl?) - RNA not analysed; predicted to affect splicing; (a) - predicted to affect splicing by bioinformatic analysis; (b) - pathogenicity assessment of missense variants in Appendix II.2 Table II.2.2; r.(?) - RNA not studied; (MC) - mother carrier; (MGC) - maternal grandmother carrier. Variants are described according to the reference sequence NM_000252.2, using HGVS nomenclature guidelines.

Multi-exonic duplication in patient 1

Patients included in this work were initially studied by routine *MTM1* gene analysis. From our molecularly unresolved CNM patient cohort, 6 males and 2 females were selected for quantitative studies by MLPA analysis of *MTM1*, based on clinical and histological criteria that were compatible with CNM. Of these eight candidates, a single male patient (P1) was found to be positive, with a multi-exonic duplication in *MTM1*. A second male patient was subsequently found to carry a *DNM2* mutation and the remainder are as yet uncharacterized.

Muscle biopsy of P1 showed fibre size variation with atrophy, numerous central nuclei and endomysium fibrosis (Figure 3.2.2A). In some areas muscle was better preserved, with scanty central nuclei and atrophic fibres (Figure 3.2.2B). There was central dark staining with NADH-TR, SDH and PAS, reflecting aggregation of mitochondria and glycogen (Figures 3.2.2C and D). "Necklace"-fibres were easily identified with routine histological stains (Figure 3.2.2E), PAS (Figure 3.2.2D) and histoenzimological stains. ATPase histoenzymology showed the presence of only type 1 fibres.

Following the detection of a large duplication spanning exons 1 to 5 of the *MTM1* gene (Figure 3.2.3A), the 3' breakpoint of the duplication was narrowed down by Southern blot analysis (Figure 3.2.3B), which revealed the absence of a ~10 kb band corresponding to exons 5 and 6 of the gene. cDNA studies were carried out on reminiscent muscle tissue, in order to further characterize the rearrangement and evaluate its impact at the mRNA level. This analysis revealed a residual amount of normal transcript and a predominant mutant isoform lacking exon 6 (r.343_444del) (Figure 3.2.3C). It is predictable that this abnormal transcript will originate an in-frame deleted myotubularin lacking 34 aminoacids of the GRAM domain.

In order to gain more insight into the structural rearrangement at the *MTM1* locus, we performed a whole genome low coverage analysis by next generation sequencing. After alignment, the total number of reads aligned per 10kb interval was calculated for the patient and an unrelated normal control sample. The ratio of aligned reads in the 10kb bins between patient and control is indicative for amplification (ratio >1) or deletion (ratio <1) events. Multiple 10kb bins in the 5' region of the *MTM1* locus showed amplification (Figure 3.2.3D). Interestingly, the amplification appeared to extend into the 3' region of the neighbouring *MAMLD1* gene. An in depth analysis of the orientation of the aligned read pairs as well as several long-range PCR experiments failed to provide further information regarding the breakpoints of the structural rearrangement. Considering the NGS results, we postulated that the duplication could originate a *MTM1/MAMLD1* fusion transcript. In order to test this possibility, additional expression experiments were performed using RNA obtained from whole blood. Results revealed that a *MTM1/MAMLD1* fusion transcript was indeed generated

(Figure 3.2.3E). Based on these results, a schematic representation of the duplication is presented (Figure 3.2.3F).
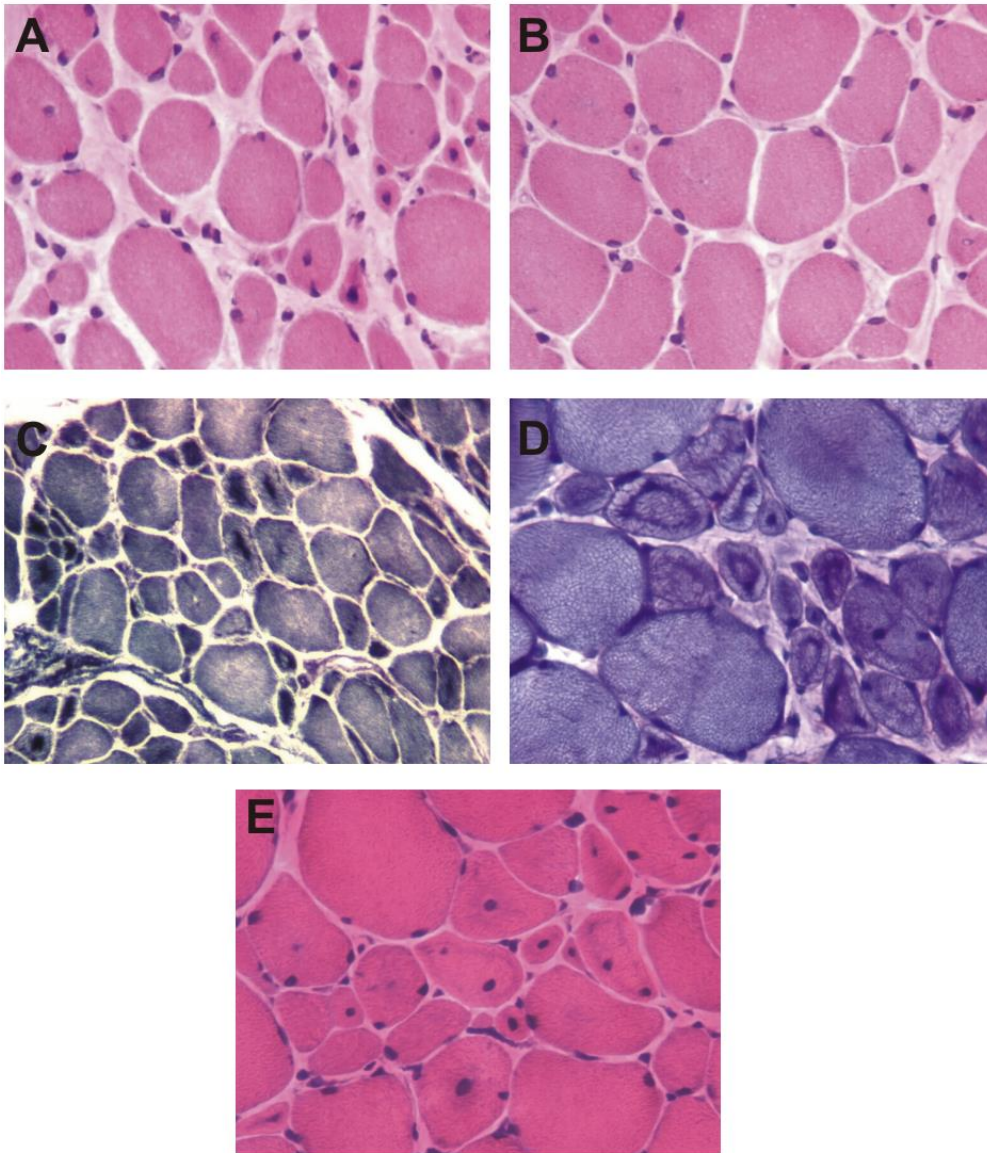
Figure 3.2.2- Histological features of patient 1. Histological analysis of the left deltoid muscle biopsy performed using staining with hematoxylin and eosin (A, B and E), histochemical reactions for NADH-TR (C), and PAS (D). Abnormal fibre size variability is seen on all stainings. The key feature of the pathology is the central nuclei (more evident on H&E image A and E). The dark central staining areas accompanying the central nuclei are another striking feature (C and D). The "necklace fibres", recently described as a histopathological marker of MTM1, can be best appreciated on PAS staining (D) and faintly on H&E (E).
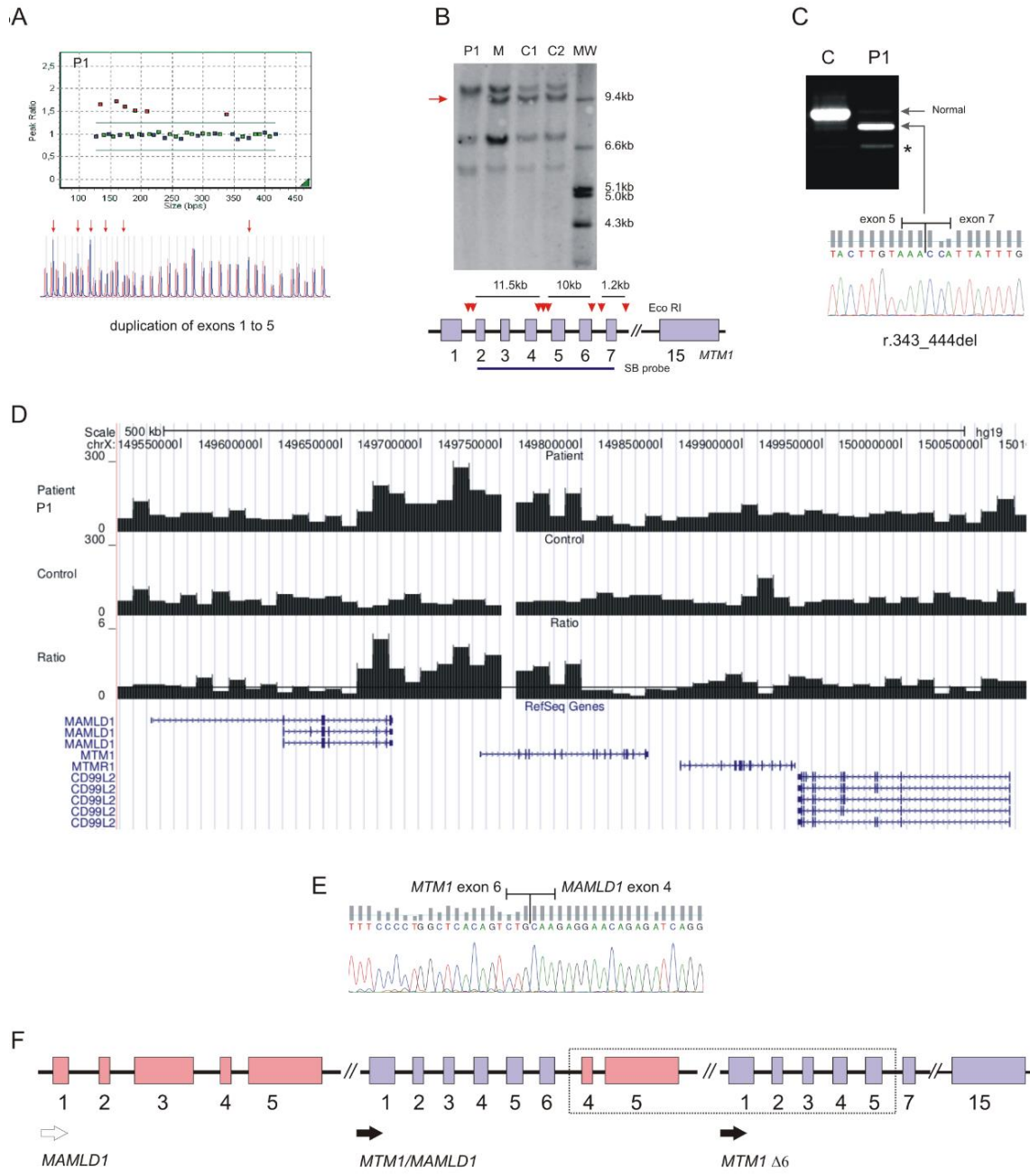
Figure 3.2.3- Large genomic duplication detected in patient 1. (A) MLPA analysis in patient 1 (P1) shows duplication of exons 1 to 5 of the *MTM1* gene. Red arrows indicate the duplicated probes. (B) Southern blot technique using a cDNA probe spanning exons 2 to 7 (SB probe); red arrow-heads represent EcoRI restriction sites; red arrow highlight the absence of a band in the patient, corresponding to exons 5 and 6. (C) RT-PCR analysis of mRNA obtained from muscle, revealing a mutated isoform lacking exon 6; (*) - smaller faint band corresponding to an alternative splicing product (Δ exon 5 and 6, ENSEMBL transcript ID: ENST00000424519). (D) Low coverage NGS analysis of a 500kb interval within the X chromosome. Histograms correspond to the total number of reads aligned per 10kb interval for the patient (top), an unrelated normal control sample (middle) and ratio between patient and control (bottom) where the line corresponds to ratio value of 1. (E) Partial sequence of the fusion transcript, showing *MTM1* exon 6 adjacent to *MAMLD1* exon 4. (F) Schematic representation of duplicated region (dashed rectangle); resulting transcripts are indicated by black arrows - observed (filled) and predicted (unfilled). P1 - patient; C/ C1/ C2 - controls; M - patient's mother; MW - molecular weight marker.

Large deletion in Patient 2

During routine *MTM1* gDNA sequencing in P2, no symmetrical PCR amplification was obtained for the majority of exons (3 to 14). This led us to suspect the presence of a large intragenic deletion, which was subsequently confirmed by MLPA (Appendix II.2, Figure II.2.3). This mutation was further characterized by long-range PCR, in order to predict its impact at the protein level and to facilitate carrier screening and prenatal diagnosis. Sequencing of the junction fragment obtained by long-range PCR revealed breakpoints located in introns 2 and 14 (~75 kb sized deletion) and enabled us to determine the full description of the mutation as c.63+834_1645-2104del (Appendix II.2, Figure II.2.3). At the protein level, if efficient translation occurred, this would correspond to the loss of approximately 87% of primary protein sequence including all of the functionally relevant myotubularin domains.

3.2.5 Discussion

MTM1-LOVD

Providing evidence for pathogenicity of variants is costly and time-consuming, so diagnostic laboratories rely on previous reports of variants detected in patients. For *MTM1* however, such data has been dispersed in the literature, in various formats. These difficulties prompted us to develop an LSDB for the *MTM1* gene. We also engaged in this task because the scientific community has expressed the importance and the need for establishing a dedicated XLMTM database (mutational and clinical) (Laporte et al., 2000). The implementation and curation of this LSDB followed guidelines reported elsewhere (Celli et al., 2011; Vihinen et al., 2011), and is registered with the HGVS (LSDB list, http://www.hgvs.org/dblist/glsdb.html).

The main objective for this freely accessible database is the description of *MTM1* sequence variants with phenotypic impact, collected from different sources. These data ultimately will allow further insights with respect to the mutational spectrum of the *MTM1* gene and epidemiology, abetting the molecular diagnosis and research in this field. Currently >11% of the entries included in this database have not been previously described in the literature. This means that this database contributes to the visibility of new *MTM1* mutations within the public domain that would otherwise have a limited possibility to be released as single or individual case reports in the literature.

Database content analysis

Although large genomic deletions contributed to narrowing down the XLMTM candidate region and ultimately the identification of the *MTM1* gene (Dahl et al., 1995), large intragenic deletions appear to be a relatively rare cause of myotubular myopathy (6.3% of all cases) compared to point mutations. In fact, these deletions are generally identified in single case reports, with the exception of exon 1 and full *MTM1* gene deletions. The majority are associated with a severe disease outcome, even in deletions that are predictably in-frame. However, it should be noted that besides the large deletion reported in this work (P2) only a few mutations, namely the deletion of exon 9 and the deletion of exon 13, have been studied in enough detail to allow a more accurate prediction of their impact at the protein level (Tsai et al., 2005; Jungbluth et al., 2008).

The largest proportion (93.3%) of pathogenic sequence variants described to date in *MTM1* is that comprising small mutations. Since the last published *MTM1* mutation update (Laporte et al., 2000), there has been a disproportional increase of missense and splicing mutations reported in *MTM1*. This might be attributed to the recent identification of additional XLMTM cases with a milder phenotype, often associated with these types of sequence variants (Biancalana et al., 2003; Yu et al., 2003; Hoffjan et al., 2006; Bevilacqua et al., 2009). Analysis of the distribution of point mutations reveals a higher incidence in exons 8, 11, 9, 4 and 12 (in decreasing order of frequency), and this should be taken into consideration when conducting *MTM1* gene mutation screening. Altogether, mutations in these exons account for almost half of all reported XLMTM cases.

A significant number of *MTM1* point mutations coincide with the hypermutable CpG dinucleotides that are amenable to methylation-mediated deamination. This mechanism could explain the recurrence, hence higher frequency, of some variants (Table 3.2.3). In fact six mutations (c.109C>T, c.141_144delAGAA, c.205C>T, c.1261-10A>G, c.1261C>T and c.1262G>A) were each detected in nine or more patients, accounting for approximately 24% of total cases in the MTM1-LOVD database. The most frequent variant in the MTM1-LOVD is the splicing mutation c.1261-10A>G that activates a cryptic splice site, promoting the inclusion of 9bp in the open reading frame. This cryptic splice site is preferentially used in *MTM1* transcripts detected in skeletal muscle (de Gouvon et al., 1997; Nishino et al., 1998b).

Table 3.2.3- Frequent mutations in MTM1-LOVD.

Number of DB entries (n)	Gene Region	Mutation (DNA)	Mutation (protein)	Sequence context	Possible cause
12	Exon 3	c.109C>T	p.Arg37*	TCCT<u>C</u>GACT	CpG
13	Exon 4	c.141_144delAGAA	p.Glu48Leufs*24	CAA<u>AGAA</u>GT	slippage
10	Exon 4	c.205C>T	p.Arg69Cys	TTAT<u>C</u>GTCT	CpG
9	Exon 8	c.614C>T	p.Pro205Leu	GTTC<u>C</u>GTAT	CpG
13	Exon 9	c.721C>T	p.Arg241Cys	TGTG<u>C</u>GTTG	CpG
28	Intron 11	c.1261-10A>G	p.Ser420_Arg421insPheIleGln	ATCA<u>A</u>TTTA	unknown
9	Exon 12	c.1261C>T	p.Arg421*	TCAG<u>C</u>GAAT	CpG
18	Exon 12	c.1262G>A	p.Arg421Gln	CAGC<u>G</u>AATA	CpG

Footnote: Description of mutations with 9 or more independent entries. DB - database; CpG - dinucleotide mutational hotspot. Nucleotides affected by the mutation are underlined. Variants are described according to the reference sequence NM_000252.2, using HGVS nomenclature guidelines.

The change is predicted to include 3 aminoacids in the core of the protein-tyrosine phosphatase (active) site. It was identified in several patients from different ethnic origins (US, Japan and Europe) and associated with a severe disease outcome in male patients. The cause of this recurrent mutational event is still unknown.

In a previous study (McEntagart et al., 2002), the analysis of a large number of patients revealed a statistically significant association between non-truncating mutations and the mild phenotype, as opposed to the intermediate/severe phenotype associated with mutations that give rise to PTC. We performed a similar analysis of all data available in MTM1-LOVD with the aim of obtaining further correlation between genotype and phenotype. In line with previous reports, it was evident that truncating mutations are predominantly associated with a severe phenotype. A total of 188 male patients were reported to have mutations that originate PTC. Clinical classification was available in 146 of these (77.7%). The majority (n=139, 95.2%) were reported as severe. Only four patients were reported as mild or mild/moderate, and the PTC inducing mutations were c.836delC, c.1558C>T (n=2) and c.1792delC. Besides these, there are three patients with a moderate phenotype, one associated with a splicing mutation (c.137-7T>G) and the other two with nonsense mutations which had also been found in patients with a severe disease outcome. Overall, the registries

show a clear bias towards the severest end of the disease, which may reflect either a specific biological pattern related to defects in myotubularin or the patient selection criteria used to conduct *MTM1* gene analysis.

Since the majority of non-truncating mutations are of the missense type (~80%) we centered our analysis on these variants and their localization on the protein (Table 3.2.4).

Table 3.2.4- Analysis of *MTM1*-LOVD missense mutations.

Protein domain	Residues involved (% of total protein)	Missense mutations						Nonsense mutations
		n (% of total)	Phenotype reported (b)					n (% of total)
			Severe	Mild/ Moderate	Inconsistent	Unknown		
GRAM	aa 34-149 (19.1%)	9 (12.3%)	2 (22.2%)	5 (55.6%)	0	2 (22.2%)		5 (15.6%)
RID (a)	aa 162-265 (17.1%)	28 (38.4%)	15 (53.6%)	7 (25.0%)	3 (10.7%)	3 (10.7%)		6 (18.8%)
PTP	aa 274-434 (26.5%)	28 (38.4%)	18 (64.3%)	5 (17.9%)	1 (3.5%)	4 (14.3%)		3 (9.4%)
SID	aa 435-486 (8.5%)	5 (6.8%)	3 (60.0%)	1 (20.0%)	0	1 (20.0%)		5 (15.6%)
other	(28.8%)	3 (4.1%)	1 (33.3%)	2 (66.7%)	0	0		13 (40.6%)
Total		73	39	22	4	10		32

Footnote: GRAM - glucosyltransferase, Rab-like GTPase activators and myotubularins; RID - Rac-induced recruitment domain; PTP - protein tyrosine phosphatase; SID - SET-interacting domain; aa - aminoacid position within myotubularin protein (reference sequence NP_000243.1). (a) - this domain partially overlaps desmin binding region. (b) - only phenotypes reported in male patients were considered for this analysis; the mild and moderate phenotypes were combined since several reports do not distinguish the two clinical classifications.

Nonsense mutations were also included for comparative purposes. Most missense mutations (96%) are located in myotubularin regions with known function (representing 71% of the total protein sequence), whereas only 59% of nonsense mutations are located in these domains. This suggests that missense changes are not randomly scattered in the protein. Moreover, 85% of mutations associated with a severe phenotype are located in the PTP (catalytic) domain and the RID/ desmin-binding region. In terms of its relative proportion, mild/moderate

missense mutations are more frequent in the GRAM domain and in the remaining regions of myotubularin. However, it should be noted that there is some degree of phenotypic variability; four missense mutations were classified as "inconsistent" in terms of the reported clinical severity. The presence of more than one sequence variant in a patient can hamper attempts at genotype/phenotype correlations. In fact two independent publications reported male patients with double *MTM1* mutations. In the first case a patient classified as having a severe phenotype (died at 6 weeks of age) two missense changes were identified (p.Asp431Asn and p.Asp433Asn) (Laporte et al., 1997). The second patient was reported with severe neonatal hypotonia requiring assisted ventilation due to respiratory failure. Mutational analysis revealed the presence of two *MTM1* changes that create PTC: c.109C>T (p.Arg37*) and c.386_387insA (p.Ser129Argfs*7) (Tachi et al., 2001).

Novel multi-exonic duplication

We describe two patients with large rearrangements in the *MTM1* gene, one of which (P1) carries the first multiexonic duplication described to date. The clinical presentation of this patient is milder than the classical form of XLMTM - although he had delayed motor development skills and a progressive tetraparesis, at age 7 he remains ambulant and there is no report of respiratory impairment. Analysis of the patient's muscle biopsy revealed variation of fibre size and typical central nuclei which lead to the diagnosis of CNM. Additionally, "necklace" fibres were identified. These structures, characterized by a basophilic ring deposit following the contour of the cell in which the nuclei are aligned, were initially described as a particular feature of older CNM cases (Romero, 2010). It is now evident that "necklace" fibres can also be found in early onset cases with a milder phenotype. Using a variety of experimental methods, we have shown that the duplication encompassing exons 1 to 5 in *MTM1* has an unexpected effect at the mRNA level, resulting in an in-frame deletion of the sequence corresponding to exon 6 (r.343_444del). At the gDNA level, using a genome wide low coverage analysis of the patient's genome we were able to confirm the amplification within the *MTM1* gene and obtained evidence that it extended into the *MAMLD1* gene, known to be associated with hypospadias (Ogata et al., 2008). However, unlike intragenic duplications, it is foreseeable that because this duplication affects only the 3' end of the *MAMLD1* gene, a normal copy would be left intact, hence the duplication should have little or no phenotypic consequence, thereby explaining why no genital abnormalities were found in the patient. MLPA analysis had shown that the average ratio of duplicated *MTM1* probes (1.6) was significantly lower than expected for a male (approximately 2.0). This, together with the presence of a residual amount of normal transcript and the finding that the patient's mother was not a carrier, provides strong evidence for somatic mosaicism in the patient – the

first such case described to date – which might explain the patient's relatively mild phenotype in relation to what could be predicted from the genotype.

Standard routine methods for XLMTM molecular diagnosis, such as genomic analysis of *MTM1* by Sanger sequencing, do not allow the identification of duplications and may also fail to detect mosaics. The only other description of a duplication involving a complex *MTM1* rearrangement concerns a male infant with characteristic clinical and histopathologic findings of XLMTM (Trump et al., 2012). The patient had severe neonatal hypotonia and respiratory insufficiency. Other XLMTM compatible findings included absence of deep tendon reflexes, cryptorchidism, and elongated fingers and toes. There was a progressive respiratory deterioration culminating in death at 1 month of age, due to respiratory failure. It is conceivable that there may be a considerable number of CNM cases with *MTM1* duplications, and it is therefore advisable to perform MLPA analysis in all CNM cases, after excluding point mutations.

Future perspectives

It would be important to gather more *MTM1* mutational data that are presently dispersed among other non-public databases (such as UMD-MTM1 and the Cardiff database for X-linked myotubular myopathy) (Jungbluth et al., 2009) as well as in private registries of diagnostic laboratories around the world. According to the meeting report of the 6th ENMC workshop on centronuclear (myotubular) myopathies, efforts are currently being made to develop a patient registry for CNM with the support of TREAT-NMD and the Myotubular Trust (Jungbluth et al., 2009). Until now the harmonized clinical items to be included in this registry are not yet in the public domain. After full release of LOVD software v3.0 the clinical items to be included will follow these international recommendations for the *MTM1* registry, ultimately allowing data integration which at this point in time is a recognized difficulty.

3.3 *RYR1*-RELATED MYOPATHIES: CLINICAL, HISTOPATHOLOGIC AND GENETIC HETEROGENEITY AMONG 17 PATIENTS FROM A PORTUGUESE TERTIARY CENTRE

SAMÕES R.*, OLIVEIRA J.*, TAIPA R., COELHO T., CARDOSO M., GONÇALVES A., SANTOS R., MELO-PIRES M., SANTOS M. (2017). JOURNAL OF NEUROMUSCULAR DISEASES, 4(1):67-76.

* THESE AUTHORS CONTRIBUTED EQUALLY TO THIS WORK HAS FOLLOWS: RS- COLLECTED CLINICAL AND NEUROPATHOLOGY DATA, DRAFTED THE MANUSCRIPT. JO- PERFORMED GENETIC EXPERIMENTS AND INTERPRETED GENETIC VARIANT DATA, MADE SIGNIFICANT CONTRIBUTIONS IN THE PAPER.

3.3.1 ABSTRACT

BACKGROUND: Pathogenic variants in ryanodine receptor type 1 (*RYR1*) gene are an important cause of congenital myopathy. The clinical, histopathologic and genetic spectrum is wide.

OBJECTIVE: Review a group of the patients diagnosed with ryanodinopathy in a tertiary centre from North Portugal, as an attempt to define some phenotypical patterns that may help guiding future diagnosis.

METHODS: Patients were identified from the database of the reference centre for Neuromuscular Disorders in North Portugal. Their data (clinical, histological and genetic) was retrospectively accessed.

RESULTS: Seventeen *RYR1*-related patients (including 4 familial cases) were identified. They were divided in groups according to three distinctive clinical characteristics: extraocular muscle (EOM) weakness (N=6), disproportionate axial muscle weakness (N=2) and joint laxity (N=5). The fourth phenotype includes patients with mild tetraparesis and no distinctive clinical features (N=4). Four different histopathological patterns were found: centronuclear (N=5), central core (N=4), type 1 fibres predominance (N=4) and congenital fibre type disproportion (N=1) myopathies. Each index case, except two patients, had a different *RYR1* variant. Four new genetic variants were identified. All centronuclear myopathies were associated with autosomal recessive inheritance and EOM weakness. All central core myopathies were caused by pathogenic variants in hotspot 3 with autosomal dominant inheritance. Three genetic variants were reported to be associated to malignant hyperthermia susceptibility.

CONCLUSIONS: Distinctive clinical features were recognized as diagnostically relevant: extraocular muscle weakness (and centronuclear pattern on muscle biopsy), severe axial weakness disproportionate to the ambulatory state and mild tetraparesis associated with (proximal) joint laxity. There was a striking genetic heterogeneity, including four new *RYR1* variants.

3.3.2 INTRODUCTION

Pathogenic variants in the ryanodine receptor type 1 (*RYR1*) gene have been identified as the most common cause of congenital myopathies, in several large cohorts (Amburgey et al., 2011; Colombo et al., 2015). This gene codifies the ryanodine receptor, a calcium release channel of the skeletal muscle sarcoplasmic reticulum with a crucial role in excitation–contraction coupling. *RYR1* is a large gene (106 exons) but it has three mutational hotspot regions: region 1 (exons 2-17, amino acid residues 35–614), region 2 (exons 39-46, amino acid residues 2,163–2,458) and region 3 (exons 85-104, amino acid residues 3,916–4,942) (Treves et al., 2005).

The best characterized entity associated to *RYR1* is the dominantly inherited central core disease with malignant hyperthermia susceptibility, a potentially life-threatening pharmacogenetic reaction to volatile anesthetics and/or depolarising muscle relaxants (Robinson et al., 2006). However, in recent years, an increasing range of additional histopathological and clinical variants have been associated with *RYR1* variants, including with recessive inheritance (Treves et al., 2008).

We aimed to review the clinical, histopathologic and genetic characteristics of the patients diagnosed with ryanodinopathy followed in our tertiary centre. In face of the great variability described in literature, we attempted to define some key features of ryanodinopathies that could help future diagnosis.

3.3.3 MATERIAL AND METHODS

Centro Hospitalar do Porto – Hospital de Santo António is the reference centre for Neuromuscular Disorders in North Portugal. Adult and paediatric patients with the genetic diagnosis of ryanodinopathy were identified from the Centre database. Their clinical records and muscle biopsies were retrospectively reviewed. As muscle magnetic resonance imaging is still not available in our centre in the routine clinical practice, we do not have such data. This retrospective series of cases is in accord with the ethical standards of the Committee on Human Experimentation of our institution and in accord with the Helsinki Declaration of 1975.

3.3.4 RESULTS

Seventeen patients with ryanodine myopathy, including four families (FAM), were identified and included. Table 3.3.1 summarises the clinical, histopathological and genetic data of the patients (P1 to P17).

Clinical manifestations

At the time when data were retrospectively collected, the patients' mean age was 22.3 years old (range 5 months-51 yo). Twelve patients (70.6%) were female. The disease onset was reported by most patients or their parents as having occurred congenitally or during the first year of life; only two patients noticed the first symptoms between 2 years and school age. All patients had a myopathic syndrome with thin appearance of muscle bulk, along with other specific features:

- Extraocular muscle (EOM) weakness (Figure 3.3.1, Panel A): This was present in six patients (including family FAM.I) and associated with different severities of tetraparesis. Patients 1 and 2 (two siblings aged 28 and 34yo) constitute our most disabled patients; they have never gained the ability to walk; they also have facial, bulbar, respiratory and spinal muscles weakness and were submitted to surgery in the second decade of life due to a rapid evolving scoliosis. Patient 3 (aged 24yo) has an intermediate presentation in terms of severity, as she has a slight tetraparesis but with a significant weakness of the axial muscles. She also had early scoliosis surgery and needs non-invasive ventilation. Patient 4 (aged 3yo) presented with a mild motor delay as she only acquired the ability to walk at the age of 24 months. Patients 5 (aged 37yo) and 6

(aged 33yo) have proximal tetraparesis with no other associated features apart from facial muscle weakness, also shared by the other patients.

- Severe axial muscle weakness (Figure 3.3.1, panel D): Patients 7 (FAM.II) and 8 (unrelated, aged 17 and 20yo) had rapidly progressive scoliosis that led to surgery at the age of 16 and 14yo, respectively. The second one has also been using non-invasive ventilation since that time. They also have mild facial muscle weakness. Both patients maintain unassisted walking and none has bulbar muscle involvement.

- Joint laxity (Figure 3.3.1, panel F): In two families, FAM.III and FAM.IV, the coexistence of weakness and laxity was the main key to the diagnosis. In the first kindred, the index case was patient 9, a 4 years-old girl with normal development and a mild to moderate proximal tetraparesis, mainly affecting the lower limbs, associated with an evident laxity in axial and limbs joints (shoulders and elbows) noticed at the age of two. Patient 10, her mother (aged 20yo), has recurrent patellar dislocations and the same pattern of laxity but with less weakness. This worsened transiently during her second pregnancy. She gave birth to our youngest patient, patient 11, who was born with tetraparesis (mainly upper limbs), laxity, facial and transitory bulbar involvement, needing to be fed by a nasogastric tube in the neonatal period. She progressively improved in months and at present, aged 12 months, she achieved to sit unsupported.

Table 3.3.1- Clinical, histopathological and genetic findings of patients included in this work.

Patients	Distinctive clinical features	Additional clinical features					Main histopathological findings	*RYR1* DNA/ protein variants (all heterozygous)	Variant type	Exon	Protein domain / hotspot	Variant Reference	Disease inheritance
		AW	F	B	SS	NIV							
P1, P2 (FAM.I)	Extraocular muscles weakness	N	Y	Y	15, 13	*, 25	Central nuclei	c.12010C>T / p.Gln4004*	nonsense	87	HD3	[4]	AR
								c.14643G>A / p.Met4881Ile	missense	101	HD3, Triadin	[4]	
P3		3.5	Y	N	15	17		c.3800C>G / p.Pro1267Arg	missense	28	-	[4]	Sporadic (AR, compound heterozygosity)
								c.9157C>T / p.Arg3053*	nonsense	61	apoCaM	[4]	
P4		1	Y	N	N	N		c.12063_12064dup / p.Met4022Thrfs*4	frameshift dup.	88	HD3	New [b]	Sporadic (AR?, parents not tested)
								[c.7027G>A / p.Gly2343Ser (spl.?); c.13672C>T / p.Arg4558Trp]	missense / spl.? missense	43 94	HD2 HD3, S100A1	[5]	
P5		1	Y	N	N	N		c.1901delC / p.Thr634Lysfs*32	frameshift del.	17	-	New	Sporadic (AR, compound heterozygosity)
								[c.7027G>A / p.Gly2343Ser (spl.?); c.13672C>T / p.Arg4558Trp]	missense / spl.? missense	43 94	HD2 HD3, S100A1	[5]	

| Patient	Clinical						Histopathology	Mutation	Type	Exon	Domain	Reference	Inheritance
P6		3	Y	N	N	N	Congenital fibre type disproportion	c.1342A>T / p.Ile448Phe	missense	13	HD1	New [c]	Sporadic (AR, compound heterozygosity)
								c.2348C>T / p.Ser783Leu	missense	19	SPRY	New	
P7 (FAM.II)	Severe axial muscles weakness	1	Y	N	16	N	Type 1 fibre predominance	c.14677C>T / p.Arg4893Trp	missense	102	HD3	[6]	AD (niece of P14)
P8		2	Y	N	14	14		c.7843C>T / p.Arg2615Cys [d]	missense	49	DHPR	[3]	Sporadic (variant inherited from healthy father)
P9, P10, P11 (FAM.III)	Joint laxity	2,1, N	[a]	[a]	N	N	Type 1 fibre predominance (P9)	c.14693T>C / p.Ile4898Thr [MHS]	missense	102	HD3	[6]	AD
P12, P13 (FAM.IV)		1	N	N	N	N	Central core (P12)	c.14656T>C / p.Phe4886Leu	missense	102	HD3	New [3] [e]	AD
P14 (FAM.II)		1	Y	N	N	N	Central core	c.14677C>T / p.Arg4893Trp	missense	102	HD3	[2]	AD (uncle of P7)
P15	Mild tetraparesis only ± facial weakness	1	Y	N	N	N	Type 1 fibre predominance	[c.4711A>G / p.Ile1571Val; c.10097G>A / p.Arg3366His; c.11798G>A/ p.Tyr3933Cys][MHS]	missense	33 67 86	- - HD3	[7], [8]	Sporadic (AR?, parents not tested)
								c.14537C>T / p.Ala4846Val	missense	101	HD3	[9]	

P16		1	N	N	N	N	Central core	c.14582G>A / p.Arg4861His[MHS]	missense	101	HD3	[5]	AD (other family members affected)
P17		1	N	N	N	N		c.13910C>T / p.Thr4637Ile	missense	95	HD3	[10]	Sporadic (*de novo*)

Footnote: Additional clinical features: AW – ability to walk (age in years); F – Facial weakness; B – Bulbar weakness; SS – Scoliosis surgery (age in years); NIV – Non-invasive ventilation (age in years); N – no, Y – yes, * - Patient refuses; FAM. – Family; HD1/2/3- hotspot domain 1/2/3; Triadin- Triadin interaction domain; apoCaM - apocalmodulin binding site; S100A1 - S100A1 domain; SPRY - Domain in SPla and the RYanodine Receptor; DHPR - dihydropyridine receptor association domain; [a] – Only P11 had bulbar involvement and facial palsy in the neonatal period; [b] – variant listed in the Leiden Muscular Dystrophy LOVD database; [c] – variant also found in another unpublished patient from our cohort with central core disease; [d] – variant of uncertain clinical significance; [e] - Aminoacid change previously reported; AR – autosomal recessive; AD – autosomal dominant; Del. – deletion; Dup. – duplication; MHS - Documented risk for malignant hyperthermia susceptibility; P1, P2 and P3 were previously reported in Oliveira et al., 2016 [1] and P7, P8 and P15 previously presented in Rocha et al., 2014 [2]; [3]- Klein et al., 2012; [4]- Abath-Neto OL, 2014; [5]- Monnier et al., 2001; [6]- Lynch et al., 1999; [7]- Duarte et al., 2011; [8]- Kraeva et al., 2015; [9]- Gambelli et al., 2007; [10]- Davis et al., 2003.

In FAM.IV, the daughter (P12, aged 9yo) was also the index case, with moderate tetraparesis, elbow laxity and past history of congenital hip dislocation. Her mother (P13, aged 38 yo) was found to have also past history of congenital hip dislocation and easy fatigue. On examination, a mild tetraparesis was revealed. So, apart from patient 11 in her first month of life, no patient in this group had facial, bulbar or axial muscle involvement.
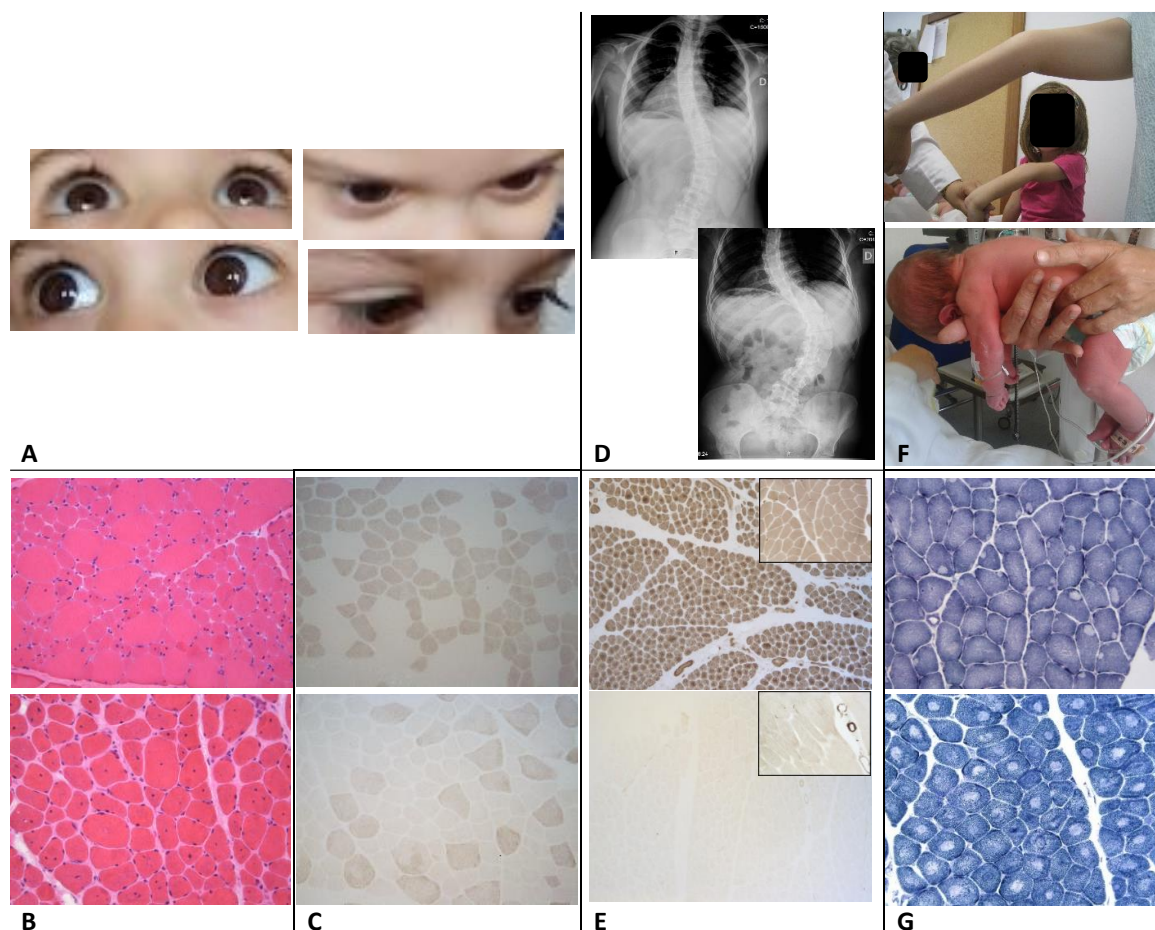
Figure 3.3.1- Clinical and histopathological features of some patients.

Panel A (permission was granted by the patient) – Extraocular muscle weakness in patient 4: upper left – looking upwards, upper right – looking downwards, lower left – looking rightwards, lower right – looking leftwards;

Panel B –Centronuclear myopathy, H&E patients 4 (top) and 5 (bottom), magnification, 200x.

Panel C – Myopathy with congenital fibre type disproportion as main feature, patient 6; ATPase 4.35 (top) and 9.4 (bottom), magnification 100x.

Panel D – Spinal X-ray showing worsening of scoliosis in patient 7, five years apart: at 11 years (top) and 16 years (bottom);

Panel E – Myopathy with type 1 fibres predominance in patient 9, ATPase 4.35 (top) and 9.4 (bottom), magnification, 100x.

Figure 3.3.1- (continues).

Panel F – Family 3: patients 9 and 10 showing hyperlaxity at elbows and wrists (top) and patient 11 as a floppy newborn with nasogastric tube (bottom).

Panel G – Centralcore myopathy in patients12 (top) and 16 (bottom), NADH, magnification 100xanalysis of the left deltoid muscle biopsy performed using staining with hematoxylin and eosin (A, B and E), histochemical reactions for NADH-TR (C), and PAS (D). Abnormal fibre size variability is seen on all stainings. The key feature of the pathology is the central nuclei (more evident on H&E image A and E). The dark central staining areas accompanying the central nuclei are another striking feature (C and D). The "necklace fibres", recently described as a histopathological marker of MTM1, can be best appreciated on PAS staining (D) and faintly on H&E (E).

The fourth phenotype includes four patients (P14 to P17) with a mild tetraparesis and no distinctive clinical feature. The two patients with later reported age of onset belong to this group (P14 and P15). They also have facial muscles weakness but no other additional feature. Patient 14 belongs to FAM.II and is a maternal uncle of patient 7 (different phenotype). Patient 17 presented with neonatal arthrogryposis that required several surgical procedures. He was only able to walk at the age of 6 and became fully functional. None of the patients has cardiac involvement. Eight patients (47.1%) are not aware of any family history suggestive of myopathy.

Histopathology

Fourteen patients (82.4%) had a muscle biopsy, at a mean age of 10.9 (1-29) yo. The examined muscle (deltoid) was abnormal in every patient, with four histopathologic myopathic patterns found (Figure 1, panels B to G): centronuclear (N=5), centralcore (N=4), type 1 fibres predominance (T1FP; N=4) and congenital fibre type disproportion (CFTD; N=1). All patients with EOM weakness had the centronuclear pattern, except P6 who showed CFTD. From the four centralcore patterns identified, three belong to patients with isolated mild tetraparesis with or without facial weakness. The remaining belongs to patient 12, with the laxity component. The group of patients with T1FP was more heterogeneous in terms of clinical presentation but both patients with severe axial muscles involvement having this histopathologic pattern. From the other two T1FP patients, one was the youngest patient with the laxity (P11) and the other was one of the less severely affected patients (P15).

In the extraocular muscles weakness phenotype, the majority of cases showed severe myopathic changes. In the other phenotypes, the numbers are too small to draw conclusions (Table 3.3.2).

Table 3.3.2- Correlation between clinical and histopathological severity histopathological

Patient	Clinical phenotype	Clinical severity score	Hystopathological pattern	Fibre-size variation	Atrophic fibres	Fibrosis	Necrosis
1	Extraocular muscles weakness	4+2=6	Central nuclei	++	++	0	0
2		4+3=7		NA			
3		2+3=5		+++	+++	0	0
4		1+0=1		+++	+++	0	0
5		2+0=2		+++	+++	0	0
6		1+0=1	CFTD	++	++	0	0
7	Severe axial muscles weakness	2+0=2	Type 1 fibre predominance	+	+	0	0
8		2+3=5		+++	+++	+	0
9	Joint laxity	2+0=2		+	+	0	0
10		1+0=1	NP				
11		1*+1=2	NP				
12		3+0=3	Central core	+	+	0	0
13		2+0=2	NP				
14	Mild tetraparesis only ± facial weakness	0+0=0	Central core	NA			
15		2+2=4	T1FP	+	+	0	0
16		2+0=2	Central core	++	++	0	0
17		2+2=4		NA			

Footnote: CFTD – congenital fibre type disproportion, T1FP – type 1 fibres predominance, NA – not available for review, NP – not performed, + - slight, ++ - moderate, +++ - severe, * Still in the first year of life. Clinical severity score (adapted from Amburgey, 2013): Sum of ambulatory (0: No Impairment, 1: Delay Motor Development, 2: Able to Walk Independently but with limitations, 3: Able to Walk with assistance, 4: Wheelchair dependent) and respiratory ratings (0: No Respiratory Impairment, 1: Respiratory Impairment as Neonate Only, 2: Some Respiratory Impairment without support, 3: Vent, CPAP, or BiPap at night, 4: Vent, CPAP, or BiPap during the day).

Genetics

High genetic heterogeneity was found in this cohort of patients. Except for patients 4 and 5, each index case had different pathogenic variants in heterozygosity, the majority being of the missense type. Five pathogenic variants had not previously published in the literature (P4, P5, P6 and P12). As detailed in Table 1, variants found in compound heterozygosity were more dispersed along the *RYR1* gene. Cases with autosomal recessive (AR) inheritance patterns were more frequent in the group of patients with EOM involvement (P1 to P6). In addition, all the centronuclear histopathologic presentations were also found in this group. In the remaining patients, there was a predominance of pathogenic variants in hotspot 3 (exons 93 to 106) along with autosomal dominant (AD) inheritance patterns. All the central core histopathologic patterns were found in this last group. The variant c.7843C>T (p.Arg2615Cys) identified in patient 8 has a paternal origin; considering that the patient's father is asymptomatic, its clinical significance remains unclear.

Malignant hyperthermia events were not reported in any of the patients or in their family members. However, three variants were previously associated with malignant hyperthermia susceptibility, by in vitro or functional contracture tests. They were identified in the first family with laxity (FAM.III), in one of the least disabled patients with EOM involvement (P15) and in one of the four patients with mild tetraparesis and facial weakness only (P16).

3.3.5 DISCUSSION

We report a rather heterogeneous group of 17 patients (including four families) with *RYR1*-related myopathies.

Clinically, all patients have muscle weakness and thin appearance of muscle bulk. We divided them in major groups according to three distinctive features: EOM weakness (N=6), severe axial muscle weakness (N=2) and marked joint laxity (N=5). The fourth phenotype had mild tetraparesis with or without facial weakness and no clinical distinctive feature (N=4).

There was no significant association between EOM weakness and clinical severity, as the patients who presented with this feature were quite heterogeneous in this respect (considering ambulation, respiratory and bulbar involvement). This has also been previously reported (Lyfenko et al., 2007).

There is clear disproportion between axial weakness and the ambulatory state in patients 3, 7 and 8 who were submitted to scoliosis surgery early in life while maintaining ambulation so far. This has been described by Colombo et al in a large published series of congenital myopathies, 44.4% of whom with the genetic diagnosis of *RYR1* related myopathy (Colombo et al., 2015). They also report that in *RYR1*-related myopathies, feeding difficulties can be transitory, as occurred with patient 11. Of interest, four out of six patients with EOM weakness also have axial weakness, with three of them having been submitted to early surgery. Therefore, the combination of these two patterns of muscle weakness may thus point towards the diagnosis of ryanodinopathy.

Concerning the third phenotype that we describe, the association of *RYR1* myopathy and laxity has been previously reported (Donkervoort et al., 2015). In our patients, laxity was mostly observed in proximal joints like shoulder, elbow, hip and knee, which may also help in the distinction from other myopathies with prominent joint hypermobility (e.g. collagen VI-related dystrophies). In this regard, muscle biopsy can also guide genetic testing.

In terms of histopathology, among the 14 patients who had muscle biopsy, four different myopathic patterns were found, previously described as being associated with *RYR1* myopathies: centronuclear (N=5), central core (N=4), T1FP (N=4) and CFTD (N=1).

In our cohort, all centronuclear myopathies were associated with EOM weakness. Wilmshurst and collegues identified *RYR1* mutations in 17 patients from a cohort of 24 patients with unresolved centronuclear myopathy. All patients except one had EOM weakness [8]. This association has also been described by other authors (Amburgey et al., 2013; Jungbluth et al., 2014)

Three of the patients with the T1FP pattern were already published by our group, emphasising the absence of central cores. The authors suggest that T1FP and involvement of axial muscles may be an important element to consider *RYR1* as a candidate gene (Rocha et al., 2011). In Maggi and colleagues' series of 66 patients with a congenital myopathy, the identification of *RYR1* mutations in 4 patients with isolated type 1 uniformity or predominance suggested that *RYR1* mutations are a relatively common cause of these histological findings (Maggi et al., 2013). Sato et al published ten unrelated Japanese patients diagnosed with congenital neuromuscular disease with uniform type 1, four of whom (40%) had *RYR1* mutations. These patients had milder clinical features compared with those without *RYR1* mutations (Sato et al., 2008). Our findings do not support such a clear association between T1FP and milder clinical involvement in *RYR1* patients because half of them had severe axial involvement (P7 and P8).

Clarke et al identified *RYR1* mutations in four out of seven families with typical CFTD in whom other known genetic causes had been excluded. The authors suggested that *RYR1* is a relatively common cause of this clinico-pathological pattern. They also noted an association between the presence of *RYR1* mutations and ophthalmoplegia, proposing that it may be the most specific clinical indicator of *RYR1* mutations in the setting of CFTD as this is not reported in the other known causes (Clarke et al., 2010). In agreement, our only CFTD case had EOM weakness.

The classic description of a *RYR1* congenital myopathy was a core myopathy but this paradigm has changed. Most of the previously cited papers that describe non-core histologic presentations propose that cores may evolve overtime, in a dynamic model. None of our patients have repeated the muscle biopsy and the mean age at muscle biopsy was in fact lower in central core (7.7yo) than in non-central core presentations (11.1 yo), as opposed to what would be expected based on the possible development of cores with age increasing. The small number of patients may contribute to this finding.

Concerning the genetic profiles found in our patient cohort there is a high genetic heterogeneity and variability. The majority of cases had different *RYR1* variants including four described for the first time in this work.

It is not infrequent to find variant of uncertain significance (VUS) during these studies. Particularly those of missense type pose additional interpretation difficulties and as consequence the clinical utility of these studies is sometimes limited. According to the most

recent guidelines for sequence variants interpretation from the American College of Medical Genetics and Genomics, several lines of evidences should be considered when performing variant pathogenicity assessment (Richards et al., 2015).

As an example, the 2348C>T (p.Ser783Leu) variant if interpreted individually would be classified as VUS (class 3), in spite of the bioinformatics analysis suggesting its deleterious effect and absence from population variant databases. However, when interpreted together with the c.1342A>T (p.Ile448Phe) variant (detected in another patient of our laboratory cohort and considered likely pathogenic) would be in favour of being causative for *RYR1*-related myopathy. Nonetheless, it would be important to further characterize the functional properties of these variants as previously demonstrated in the literature (Lyfenko et al., 2007; Goonasekera et al., 2007).

We attempted to correlate the sequence variants distribution found in this cohort and their impact on the structure and/or activity of ryanodine receptor type 1 using the information available in previous reports.

It is well known that the majority of patients with autosomal dominant central core disease have *RYR1* pathogenic variants that are clustered in the hydrophobic COOH-terminal region (hotspot domain 3) of this receptor. Zhou et al published a series of 28 patients in which the dominant RYR1 variants were mainly found in the C-terminal and only in the central domain of RYR1. These authors also stated that most patients with dominant C-terminal variants had the classical central core phenotype, characterized by variable degrees of proximal weakness and absence of significant bulbar, respiratory or extraocular involvement (Zhou et al., 2007). We have also found these milder phenotypes in our patients with AD inheritance, having variants located in hotspot 3 and central cores in their muscle biopsies. It is interesting to note that the majority of pathogenic variants found in our AD patients are within a small stretch of 38 residues (4861-4898) of the protein. Since the high (near-atomic) resolution structure of RYR1 has been recently described, it is now evident that these residues correspond to critical segments of the ion-conducting pathway: the luminal loop, the pore helix and the selectivity filter (Yan et al., 2015). In particular, the p.Arg4893Trp variant (found in P7 and P14) may affect one of the potential residues that constitute the inner cation-binding site. These structures were proposed to be a hotspot for disease-related mutations (Yan et al., 2015). Some variants in this region have been further characterized, namely the p.Ile4898Thr variant found in family FAM.III, which was found to create a dominant-negative suppression of RYR1 channel calcium ion permeation (Loy et al., 2011; Amburgey et al., 2013).

The genetic variants associated with autosomal recessive inheritance and other histopathological patterns (e.g. centronuclear) are distributed along the entire coding sequence (Wu et al., 2006; Treves et al., 2008; Amburgey et al., 2013). Our series is

concordant with this description. Among 26 patients with *RYR1* pathogenic variants, Maggi and colleagues (2013) identified compound heterozygous variants consistent with recessive inheritance distributed throughout the *RYR1* coding sequence in 15 patients (57.7%) (Maggi et al., 2013). Amburgey and collaborators attempted to establish genotype-phenotype correlations in a large cohort of recessive *RYR1*-related cases (Amburgey et al., 2013). Our data is in accordance with the main observations of this work. Hypomorphic mutations (nonsense and frame-shifting, expected to reduce *RYR1* expression) were identified in AR non-central core myopathies (P1 to P5) and in association with more severe clinical phenotypes. In our cohort, all patients with AR inheritance pattern and at least one hypomorphic allele had EOM weakness and centronuclear myopathy. This was previously described by other previous works (Zhou et al., 2007; Klein et al., 2012; Amburgey et al., 2013; Maggi et al., 2013; Snoeck et al., 2015). Regarding malignant hyperthermia, we identified three genetic variants (five patients) with proven susceptibility. In concordance with previous reports, this susceptibility may exist even in the context of a mild and/or non centralcore *RYR1* myopathy (Rueffert et al., 2009).

In conclusion, there is a great variability in clinical presentation, genetics and muscle pathology in patients with ryanodinopathies but some clinical features may be diagnostically relevant: extraocular muscle weakness (with centronuclear pattern on muscle biopsy), severe axial weakness disproportionate to the ambulatory state and mild tetraparesis associated with (proximal) joint laxity.

3.4 REVIEWING LARGE *LAMA2* DELETIONS AND DUPLICATIONS IN CONGENITAL MUSCULAR DYSTROPHY PATIENTS

OLIVEIRA J, GONÇALVES A, OLIVEIRA ME, FINEZA I, PAVANELLO RC, VAINZOF M, BRONZE-DA-ROCHA E, SANTOS R, SOUSA M. (2014). JOURNAL OF NEUROMUSCULAR DISEASES, 1(2):169-179.

3.4.1 ABSTRACT

BACKGROUND: Congenital muscular dystrophy (CMD) type 1A (MDC1A) is caused by recessive mutations in laminin-α2 (*LAMA2*) gene. Laminin-211, a heterotrimeric glycoprotein that contains the α2 chain, is crucial for muscle stability establishing a bond between the sarcolemma and the extracellular matrix. More than 215 mutations are listed in the locus specific database (LSDB) for *LAMA2* gene (May 2014).

OBJECTIVE: A limited number of large deletions/duplications have been reported in *LAMA2*. Our main objective was the identification of additional large rearrangements in *LAMA2* found in CMD patients and a systematic review of cases in the literature and LSDB.

METHODS: In four of the fifty-two patients studied over the last 10 years, only one heterozygous mutation was identified, after sequencing and screening for a frequent *LAMA2* deletion. Initial screening of large mutations was performed by multiplex ligation-dependent probe application (MLPA). Further characterization implied several techniques: long-range PCR, cDNA and Southern-blot analysis.

RESULTS: Three novel large deletions in *LAMA2* and the first pathogenic large duplication were successfully identified, allowing a definitive molecular diagnosis, carrier screening and prenatal diagnosis. A total of fifteen deletions and two duplications previously reported were also reviewed. Two possible mutational "hotspots" for deletions may exist, the first encompassing exons 3 and 4 and second in the 3' region (exons 56 to 65) of *LAMA2*.

CONCLUSIONS: Our findings show that this type of mutation is fairly frequent (18.4% of mutated alleles) and is underestimated in the literature. It is important to include the screening of large deletions/duplications as part of the genetic diagnosis strategy.

3.4.2 INTRODUCTION

LAMA2-related dystrophy (*LAMA2*-RD) collectively gathers two distinct clinical entities: the classical phenotype with congenital onset known as MDC1A, and a milder limb-girdle type muscular dystrophy with onset during childhood (late-onset or "ambulant" *LAMA2*-RD) (Bönnemann et al., 2014; Quijano-Roy et al., 2014). As foreseeable by this designation, these entities are caused by recessive mutations in *LAMA2* gene located on chromosome 6q22-23 and spanning 65 exons (Helbling-Leclerc et al., 1995; Zhang et al., 1996). This gene codes for the α2 chain of laminin-211, an extracellular glycoprotein expressed in the basal membrane of striated muscles, peripheral nerves, brain and trophoblast (Leivo et al., 1988; Campbell et al., 1995; Villanova et al., 1997). The interaction of laminin-211 with cell-surface receptors such as α-dystroglycan and integrin (mainly α7β1 in adult skeletal muscle) explains its major relevance in the overall extracellular architecture, integrity and cell adhesion (Holmberg et al., 2013).

MDC1A represents the most frequent form of CMD in western countries, accounting for 30 to 50% of cases (Allamand et al., 2002). Typical clinical features includes severe hypotonia associated with muscle weakness manifesting at birth or during early infancy, proximal joint contractures, elevated creatine kinase (CK) levels, cerebral white matter abnormalities and delayed motor milestones with affected children usually not achieving independent ambulation (Tomé et al., 1994; Philpot et al., 1995; Scessl et al., 2006). Feeding problems and respiratory insufficiency are commonly reported complications often requiring gastrostomy and/or artificial ventilator support (Philpot et al., 1999; Bönnemann et al., 2011).

Other features such as cardiac involvement, a sensory and motor demyelinating neuropathy, epilepsy and mental retardation have been also documented in some forms of *LAMA2*-RD (Deodato et al., 2002; Di Muzio et al., 2003; Geranmayeh et al., 2010; Carboni et al., 2011; Marques et al., 2014).

An important diagnostic aspect is that skeletal muscle biopsies from these patients have changes in laminin-α2 expression detected by immunohistochemical (IHC) analysis (Tomé et al., 1994). However, there is a recent report of a muscular dystrophy patient with apparently normal laminin-α2 IHC expression and having mutations in *LAMA2* gene (Kevelam et al., 2014).

Milder *LAMA2*-RD cases have been reported in the past few years expanding the phenotypic spectrum of the disease (Jones et al., 2001). These patients have slower disease progression and acquire independent locomotion, and are usually associated with a partial expression of laminin-α2 (Hayashi et al., 1997).

More than 375 distinct sequence variants (215 of them with known clinical relevance) have been reported in the LSBD for *LAMA2* gene (http://www.dmd.nl/LAMA2, data accessed

May 2014). Pathogenic changes include small deletions/insertions (34.9%), nonsense mutations (25.1%), changes affecting splicing (25.6%) and also missense substitutions (12.1%). In spite of the relevant amount of mutational data available there is still a limited number (2.3%) of large deletions and duplications reported in this gene. The initial suspicion of a large *LAMA2* deletion (which was predicted to include exons 23 to 56) was identified by protein truncation test in the work of Pegoraro and collaborators (Pegoraro et al., 1998). The first fully characterized large deletion in *LAMA2*, corresponds to an out-of-frame deletion of exon 56 (c.7750-1713_7899-2154del), which has been proven to be one of the most frequent pathogenic variants detected in Portuguese MDC1A patients (Oliveira et al., 2008).

One of the main objectives of this work was to describe additional novel pathogenic deletions and duplications associated with the *LAMA2* gene, identified in our cohort of CMD patients. Moreover, a systematic review of all cases with large deletions/duplications reported in the literature and mutation databases is presented. Our findings showed that this type of mutation is fairly frequent and is underestimated in the literature, reinforcing the importance to screen large deletions/duplications in LAMA2 gene as part of the genetic diagnosis strategy.

3.4.3 MATERIAL AND METHODS

Patients

Over the last 10 years (2004-2014) our group performed genetic studies in 94 CMD patients. Mutations in genes related with CMD were identified in 68% (n=64) of these patients. The majority of these patients have *LAMA2* mutations (n=52) and were referred for molecular studies due to changes in muscle laminin-α2 detected by IHC analysis and/or compatible white matter anomalies detected by magnetic resonance imaging (MRI). In four patients of this cohort only one heterozygous mutation was detected upon *LAMA2* genomic sequencing (described in Table 3.4.1). In these patients we conducted screening of large deletions and duplications in the *LAMA2* gene. This research was approved by the ethics committee from Hospital Centre of Porto (CHP).

LAMA2 gene analysis

Genomic DNA (gDNA) was obtained from peripheral blood using the salting-out method (Miller et al., 1988). *LAMA2* gene sequencing was done according to (Oliveira et al., 2008), which comprised all coding and adjacent intronic sequences of *LAMA2*. Variants were described according to the Human Genome Variation Society (HGVS) guidelines for mutation nomenclature (version 2.0) (de Dunnen et al., 2000) and using the cDNA reference sequence with accession number NM_000426.3.

MLPA analysis

Screening for deletions and duplications in *LAMA2* gene was performed by multiplex ligation-dependent probe application (MLPA) technique using two sets of probe mixes (P391-A1 and P392-A1) from MRC-Holland (Amsterdam, the Netherlands). These probe mixes contain one probe for each exon of the gene with the exception of exons 18, 44 and 48. Two probes are present for exon 1, 2, 4 and 65 and three probes for exon 56. Also, probemix P391 contains 9 reference probes and P392 contains 8 reference probes detecting different genomic regions. For the MLPA procedure, 150 ng gDNA was used for each patient and normal control samples. Amplification products were subsequently separated by capillary electrophoresis on an ABI 3130xl genetic analyser (Applied Biosystems, Foster City, CA). Data analysis was conducted using GeneMarker Software V1.5 (SoftGenetics LLC, State College, PA). Population normalization method was selected and data was plotted using probe ratio.

Southern Blot

gDNA samples from patient P3, respective parents and normal controls were digested with BglI (NewEngland Biolabs, Beverly, MA), resolved on a 0.7% agarose gel and vacuum transferred to a GeneScreen Plus membrane (Perkin Elmer, Waltham, MA) using a saline method. A cDNA probe recognizing exons 2-4 was prepared using digoxigenin (DIG) DNA Labeling Kit (Roche Applied Science, Indianapolis, IN, USA) and incubated overnight using the Easy Hyb Buffer (Roche Applied Science). The membrane was washed at 60°C in 1xSSC (Saline-Sodium Citrate)/0.1% SDS (Sodium Dodecyl Sulfate) (Sigma-Aldrich, St. Louis, MO) and twice in 0.5xSSC/0.1% SDS for 15 min each. Subsequently, the membrane was prepared with DIG Wash and Block Buffer Set (Roche Applied Science), incubated with Anti-DIG-AP conjugate (Roche Applied Science), and the DIG-labeled probe detected with ready-to-use CDP-Star (Roche Applied Science).

LAMA2 cDNA analysis

cDNA studies were carried out in patient P2. Total RNA was extracted from patient and control muscle biopsy samples using the PerfectPure RNA FibrousTissue kit (5 PRIME, Germany) and converted to cDNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City CA).

Table 3.4.1- Clinical data of CMD patients with novel large deletions and duplications in *LAMA2*.

Patient	Sex	Age *	Age of onset	Clinical presentation	Highest CK (IU/l)	Pattern/ progression of weakness	Best motor achievement	Contractures	Central nervous system involvement/ seizures	Magnetic resonance imaging	Laminin-α2 in muscle
P1 (P19 in [28])	M	20 yr	At birth	Generalized hypotonia and areflexia	3264	Muscular weakness with axial and proximal predominance and scoliosis	Independent ambulation	Elbows and ankles	No cognitive delay and no seizures	White matter changes and no gyral abnormalities	nd
P2 (P24 in [28])	M	3 yr	At birth	Hypotonia and feeding problems	1770	Muscular weakness with proximal predominance and hip congenital luxation	Cephalic control and assisted trunk control	Knees	No cognitive delay and no seizures	White matter changes and no gyral abnormalities	Total absence
P3	M	2 yr	4 mo	Hypotonia and muscular weakness	7644	Muscular weakness with axial and proximal UL predominance	Cephalic control and assisted trunk control	Knees, ankles and rigid spine	No cognitive delay and no seizures	White matter changes	nd
P4	M	8 mo	At birth	Generalized hypotonia	4400	Muscular weakness with proximal UL predominance	Cephalic and trunk control	Discrete equinus	No cognitive delay and no seizures	nd	Partial deficiency

Footnote: CK- creatine phosphokinase; M- male; mo- months; nd- not determined; P- patient; UL- upper limbs; yr- years; *- age at last clinical follow-up.

LAMA2 transcripts were subjected to PCR amplification using specific primers for the region corresponding to exons 27-32 (27F- 5′AAATTTCATGCGACAAAGCAGG; 32R-5′GCTTGCAGCCGTCACACTTC). Resulting PCR products were purified and sequenced as described before.

Long-range PCR

The deletion breakpoints encompassing exon 3 (patient P3) and also exon 17 (patient P1) were determined by amplification of gDNA using the BIO-X-ACT Long DNA Polymerase kit (Bioline, Taunton, MA). Specific primers were designed for each case; complementary to intron 2 and exon 4, for the deletion of exon 3 (2i-F:5′ACAAAGCCTGATGGAGGGAAAC; 4-R:5′AAAGCGTTAGGCACTCCGTGTC) and complementary to regions in exons 16 and 18,for the deletion of exon 17 (16F:5′-TTGGTCATGCGGAGTCCTG; 18R:5′ TGGCACGTTGGGCTAAAGC). Resolved PCR fragments were purified using the QIAquick® Gel Extraction Kit (Qiagen, Valencia, CA), and subsequently sequenced.

3.4.4 Results

Novel large deletions and duplications

We previously reported a large deletion encompassing exon 56 of the *LAMA2* gene which is relatively frequent in our laboratory patient cohort (present in 23% of our cases) [28]. This initial study suggested that it is clinically relevant to screen this type of mutations in CMD patients, and indirectly drove the development of a MLPA commercial kit for *LAMA2* gene (P391-A1 and P392-A1 from MRC-Holland). As part of this work we initially assessed the effectiveness of the MLPA kit using previously genotyped DNA samples from our patient cohort. Patients presenting homozygous and heterozygous deletions encompassing *LAMA2* exon 56 were tested. MLPA technique successfully detected this mutation in homo- and heterozygous states (Appendix II.3, Figure II.3.1).

Four patients are presented in detail in this work; P1 and P2 were previously described in the literature as P19 and P24, respectively in (Oliveira et al., 2008) while two additional patients (P3 and P4) were more recently referred for *LAMA2* gene analysis. All of these patients had incomplete molecular characterization: only one heterozygous mutation was detected by genomic sequencing and thus the molecular defect in the other *LAMA2* allele remained unknown. MLPA technique was performed in DNA samples from these patients to tentatively identify the second pathogenic mutation in LAMA2.

Patient P1 presents a congenital muscular dystrophy (neonatal onset) and remained ambulatory until the age of 17 years. In this patient a heterozygous codon deletion

(c.1798_1800del, p.Gly600del) was initially detected in *LAMA2*. MLPA analysis for this patient revealed reduced hybridization of one probe corresponding to exon 17 (Figure 3.4.1A), compatible with the presence of a heterozygous deletion involving this exon. By reviewing genomic sequencing data we excluded a potential sequence change which could compromise the affinity of the MLPA probe. A long-range PCR experiment was performed using primers designed to bind regions presumably not involved in the deletion (exons 16 and 18). Upon amplification, two PCR products were detected in the patient, whereas the experimental control had a single band (data not shown). Sequencing across the deletion breakpoint revealed that part of intron 16 is joined to intron 17, corresponding to the loss of ~5.3Kb that spans exon 17 (Figure 3.4.1B). This novel large mutation, c.2322+259_2450+2037del, predictably causes frame-shifting. DNA samples from the patient's parents were unavailable for study.

Patient P1

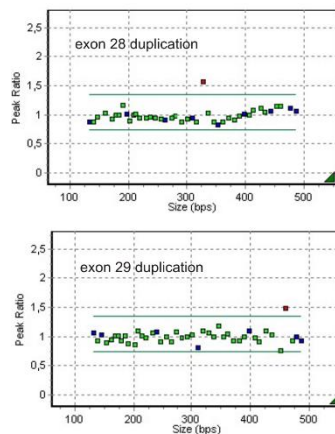**1A-** MLPA analysis
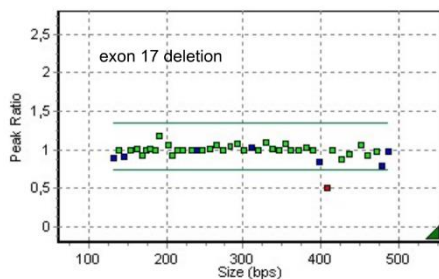
B- Breakpoint sequencing
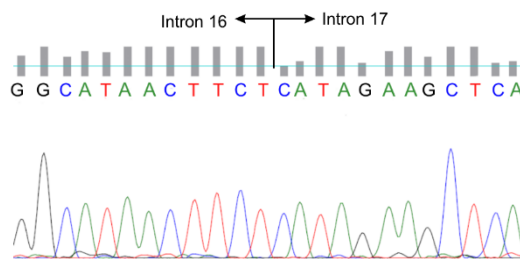
Figure 3.4.1- Results obtained for patient P1. (A) MLPA technique showing reduction of one probe (0.5 peak ratio) corresponding to exon 17 of the *LAMA2* gene. (B) Sequencing electropherogram of a fragment (obtained from long-range PCR) that enabled the identification of the deletion breakpoint, corresponding to the loss of ~5.3 Kb.

Patient P2 was referred for molecular study at 3 years of age having a typical MDC1A phenotype with absence of laminin-α2 in IHC analysis of muscle. The heterozygous nonsense mutation c.3085C>T (p.Arg1029*) was the only pathogenic sequence change detected by sequencing the *LAMA2* gene. A large heterozygous duplication encompassing exons 28 and 29 was identified in patient P2 by MLPA (Figure 3.4.2A). cDNA studies performed in the patient revealed a normal transcript together with other abnormal PCR products (Figure 3.4.2B). These include one out-of-frame transcript resulting from the contiguous duplication of exons 28 and 29 (Figure 3.4.2C).
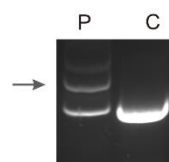
Patient P3 with an MDC1A phenotype has a novel nonsense mutation in exon 4 (c.497G>A; p.Trp166*) detected in a heterozygous and apparently homozygous state, depending on the primer-pair used to study this region (Figure 3.4.3A). These ambiguous results led us to suspect a possible deletion comprising at least part of intron 3. The application of MLPA confirmed this assumption, since a reduced amplification signal was observed for the exon 3-specific probe (Figure 3.4.3B). For further characterization, and since no RNA was available for study, we performed Southern-blotting and hybridization using a cDNA probe that recognizes exons 1 to 4. This experiment suggested that the genomic deletion originates a new fragment of approximately 6 Kb in the patient, that is absent in the control (Figure 3.4.3C). To delineate the deletion endpoints several primers were tested to perform a deletion-specific PCR. A 42931 bp deletion combined with the insertion of three nucleotides was identified, annotated as c.284-4685_397-146delinsATA (Figure 3.4.3D). Compound heterozygosity for these two mutations was confirmed by the analysis of patient's parents.
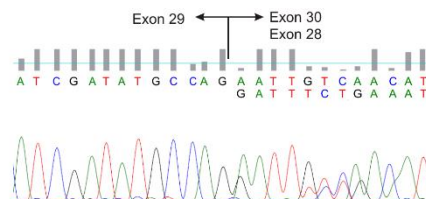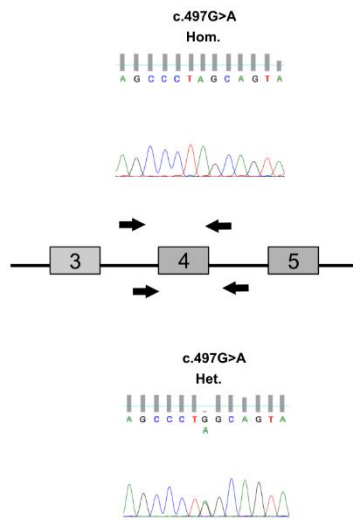

Figure 3.4.2- Duplication identified in patient P2. (A) Increased signal for MLPA probes (peak ratios around 1.5) recognizing exon 28 (upper panel) and exon 29 (lower panel). (B) cDNA analysis of *LAMA2* transcripts revealed the presence of abnormal PCR products in the patient that were not detected in the control. (C) Sequencing electropherogram of the cDNA PCR product (indicated by arrow) reveals the adjacent duplication of exons 28 and 29. C-control; P- patient.
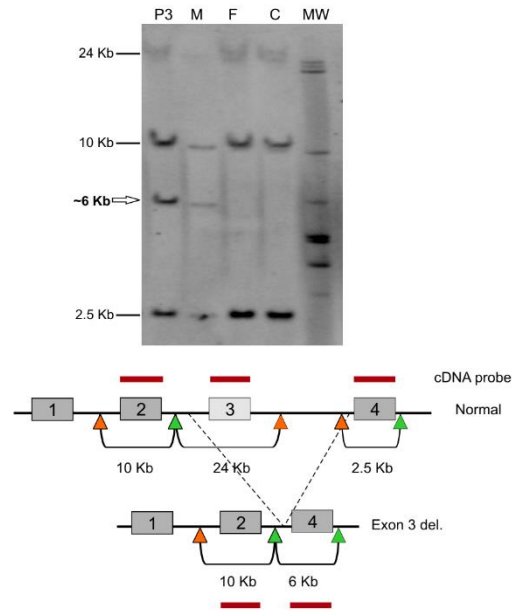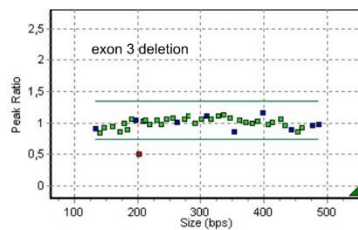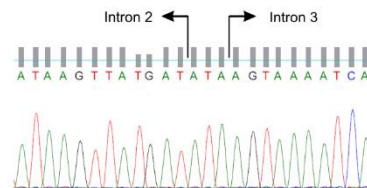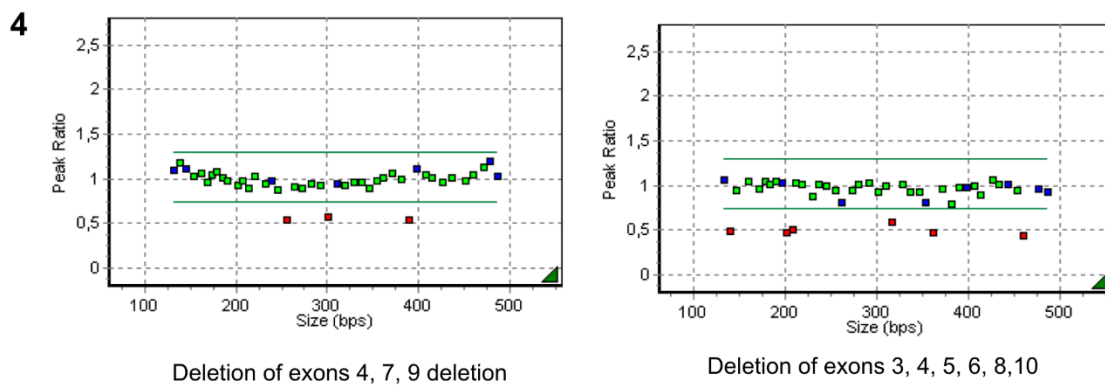
Patient P3

Figure 3.4.3- Characterization of a heterozygous deletion in patient P3. (A) The initial suspicion of a heterozygous deletion encompassing intron 3 derived from the genomic sequencing data. The nonsense mutation c.497G>A located in exon 4 was detected both in heterozygosity and homozygosity depending upon the primers used. (B) Confirmation by MLPA, with reduced amplification signal for the exon 3 probe. (C) Southern-blot followed by hybridization using a cDNA probe that recognizes exons 1 to 4, revealing a 6 Kb fragment in the patient and in his mother, but not in the father nor the control. (D) Deletion breakpoint identified by deletion-specific PCR followed by sequencing. Sequencing electropherogram revealed a 42.9 Kb deletion combined with the insertion of three nucleotides. C- control; F- father; M- mother; MW- molecular weight marker; P3- patient 3.

Lastly, patient P4 has also an MDC1A phenotype but partial laminin-α2 absence in muscle. Besides one heterozygous 8 bp duplication in exon 13 (c.1854_1861dup) previously reported in the literature [9], we were able to identify a large heterozygous deletion involving exons 3 to 10 by MLPA (Figure 3.4.4). This deletion is predicated to be out-of-frame. A more extensive characterization was not possible since no RNA sample was available for study. Compound heterozygosity was verified since each parent carried a different mutation.

Patient P4

Deletion of exons 4, 7, 9 deletion

Deletion of exons 3, 4, 5, 6, 8,10

Figure 3.4.4- MLPA results for patient 4. Several *LAMA2* probes with decreased peak ratios (~0.5), corresponding to exons 3 to 10, compatible with a large heterozygous deletion.

Novel point mutations in LAMA2

Fifty-two patients with *LAMA2* mutations have been characterized until now: 26 of which were previously reported in 2008, and another two more recently in a publication describing their atypical phenotype associated with novel missense mutations (Marques et al., 2014). In all cases mutations have been identified in both disease alleles. A list of 10 novel mutations is shown in Appendix II.3, Table II.3.1. These include four nonsense mutations (c.3520C>T, c.5263A>T, c.6501C>G e c.6979G>T), four changes affecting splice sites (c.396+1G>T, c.2450+4A>G, c.6708-1G>T and c.8988+1G>A), one single nucleotide duplication (c.2350dupT) and a missense mutation (c.3235T>G, p.Cys1079Gly).

Reviewing large deletions and duplications in LAMA2

As part of this work we reviewed all large deletions and duplications reported in the literature or in the locus specific database for the *LAMA2* gene (http://www.lovd.nl/LAMA2, information last accessed in May 2014). In addition to the mutations presented in this publication, 12 different large deletions and a single duplication were previously described (Table 3.4.2).

LAMA2 large deletions (n=15, reported in 35 patients) are apparently dispersed throughout the gene, but two possible mutational "hotspots" may exist: i) one includes exons 3 and 4 (5 different deletions) and ii) in the 3' region of the gene (from exons 56 to 65). The largest *LAMA2* deletion, encompassing exons 23 to 56, was detected by the protein truncation test (Pegoraro et al., 1998) and comprehends more than half of the gene's coding regions. At least six different deletions affect single exons and were confirmed by a second technique in order to exclude false positive results. The majority of deletions are predicted to cause frame-shifting (out-of-frame deletions). Still, only two cases were further characterized at the cDNA level which limits the accuracy of this data. Considering deletions predicted to be in-frame, all except one (Piluso et al., 2011) were reported in combination with a second truncating mutation (causing a frame-shift or a nonsense mutation). All of these patients have typical MDC1A phenotypes, except the patient with the deletion of exons 41 to 48, which remains ambulant at the age of 10 years (patient #6, reported by Xiong et al., 2014). Six patients have been described with large homozygous deletions which may be explained by a higher frequency of a particular mutation within the population (deletion of exon 56 in Portugal and exon 4 deletion in Chinese patients) or due to consanguinity in individual sporadic cases. Patients with homozygous deletions are usually detected by genomic sequencing, since the affected regions will fail to amplify during PCR. Until now, no patients have been reported with compound heterozygosity between two different large rearrangements.

LAMA2 duplications are even rarer mutational events; besides our report of a novel duplication encompassing exons 28 and 29 detected in patient P2, only one other heterozygous in-frame duplication involving exons 5 to 12 has been documented in a patient presenting muscular dystrophy (Xiong et al., 2014). However, these authors did not identify a second mutation in the patient, which might have explained an autosomal recessive *LAMA2*-RD.

The frequency of large deletions and duplications in *LAMA2* may be estimated based on the two largest patient cohorts reported and that employed quantitative techniques: 18/104 alleles from our patient cohort of 52 patients and 17/86 alleles from the recent work of Xiong and collaborators that studied 43 patients (Xiong et al., 2014). The overall frequency of these mutations is thus around 18.4% (35/190 alleles).

Table 3.4.2- Review of large deletions and duplications reported in the *LAMA2* gene.

Affected gene regions (exons)	Nr. of affected exons	Mutation description		Impact on reading frame (prediction)	Nr. of patients reported	Zygosity // other mutation type	Phenotype	References
		gDNA	RNA					
Deletions								
2-3	2	c.113-?_396+?del	r.(del)	OF	2	het. (n=2) // splice-site mutation	MDC1A; MDC1A (ambulant at 5yr)	Xiong et al., 2014
3	1	c.284-4685_397-146delinsATA	r.(del)	OF	1	het. // nonsense mutation	MDC1A	P3, this paper
3-4	2	c.284-?_639+?del	r.(del)	OF	1	hom.	MDC1A	Xiong et al., 2014
3-10	8	c.284-?_1467+?del	r.(del)	OF	1	het. // 8 bp OF duplication	MDC1A	P4, this paper
4	1	c.397-?_639+?del	r.397_639del	IF	5	hom. (n=2), het. (n=3) // several	MDC1A	Xiong et al., 2014
5	1	c.640-?_819+?del	r.(del)	IF	2	het. (n=2) // 2 bp OF deletion	MDC1A	Xiong et al., 2014
10-12	3	c.1307-?_1782+?del	r.(del)	OF	1	het. // 4 bp OF duplication	MDC1A	Xiong et al., 2014
13-37	15	c.1783-19594_5445+1681del	r.(del)	IF	1	het. // unknown	myopathy	Piluso et al., 2011 Db
17	1	c.2322+259_2450+2037 del	r.(del)	OF	1	het. // 3 bp IF deletion	MDC1A (ambulant at 17yr)	P1, this paper P19, Oliveira et al., 2008

Table 3.4.2- Review of large deletions and duplications reported in the *LAMA2* gene (continues).

Affected gene regions (exons)	Nr. of affected exons	Mutation description		Impact on reading frame (prediction)	Nr. of patients reported	Zygosity // other mutation type	Phenotype	References
		gDNA	RNA					
Deletions (continues)								
23-56	34	c.(?_3175)_(7898_?)del	r.(del)	OF	1	het. // 1 bp OF deletion	MDC1A	Pegoraro et al., 1998
41-48	8	c.5866-?_6867+?del	r.(del)	IF	1	het. // 4 bp OF duplication	MDC1A (ambulant at 10yr)	Xiong et al., 2014
56	1	c.7750-1713_7899-2154del	r.7750_7898del	OF	14	hom. (n=2), het. (n=12) // several	MDC1A; LAMA2-related MD	Oliveira et al., 2008, Db
57-3UTR	9	c.7899-?_(*219_?)del	r.(?)	?	2	hom. (n=2)	MDC1A	Db
59-63	5	c.8245-?_8988+?del	r.(del)	IF	1	het. // nonsense mutation	MDC1A	Xiong et al., 2014
63	1	c.8858-?_8988+?del	r.(del)	OF	1	het. // nonsense mutation	MDC1A	Xiong et al., 2014
Duplications								
5-12	8	c.640-?_1782+?dup	r.(dup)	IF	1	het. // unknown	MD	Piluso et al., 2011 Db
28-29	2	c.4059-?_4311+?dup	r.4059_4311dup	OF	1	het. // nonsense mutation	MDC1A	P2, this paper P24, Oliveira et al., 2008

Footnote: Mutations described according to HGVS nomenclature using cDNA reference sequence with accession number NM_000426.3. bp- base pairs; Db- *locus*-specific mutation database for *LAMA2* gene (http://www.lovd.nl/LAMA2); het.- heterozygous; hom.- homozygous; IF- in-frame; OF- out-of-frame; MD- muscular dystrophy; Nr.- number; yr- years.

3.4.5 DISCUSSION

This work describes the detailed genetic characterization of four patients with compatible features with a MDC1A phenotype, but with only one heterozygous pathogenic sequence variant detected upon complete *LAMA2* sequencing. Three novel large deletions in the *LAMA2* gene and the first pathogenic large duplication were successfully identified in this group of patients, allowing a definitive molecular diagnosis, carrier screening and prenatal diagnosis. Characterization of these mutations implied the use of a variety of techniques such as long-range PCR, cDNA and Southern-blot analysis. These methods are not generally used in the routine genetic diagnosis of this disease, but are essential to obtain accurate genotype-phenotype correlations.

Up to now reports of large deletions and duplications in *LAMA2* are very rare; only three publications have referred this type of mutation (Oliveira et al., 2008; Piluso et al., 2011; Xiong et al., 2014). The work of the Italian group included a more heterogeneous patient cohort and a broader technical approach. An array-based comparative genomic hybridization (array-CGH) developed to screen genes implicated in neuromuscular diseases, enabled the identification of several novel copy number variants (CNVs) including two present in the *LAMA2* gene (Piluso et al., 2011). Based on the data from a total of 95 fully genotyped *LAMA2*-RD patients, from two large cohorts, we estimate that the frequency of large deletions and duplications in *LAMA2* may be as high as 18.4%. Considering this relatively high frequency, it is important to include screening techniques such as MLPA or array-CGH in the molecular diagnostic work-up. Here, laboratories should consider the variety of equipments required, running costs and sensitivity of these two approaches to screen this type of rearrangement. The presence of a single heterozygous large deletion or duplication, especially when in-frame, should be carefully evaluated. It is conceivable that the presence of a non-pathogenic CNV in a CMD patient may not necessarily explain the clinical phenotype.

Readers should also be aware that genomic sequencing is the technique with the highest sensitivity to screen for *LAMA2* mutations (>80%), especially in CMD cases with laminin-α2 deficiency. Our current strategy for *LAMA2* genetic analysis is sub-divided in three tiers: i) the first level comprising a selected number of exons (namely: 3, 13, 22, 27, 33, 36, 54, 58, and 61) corresponding to the genomic regions where the majority of point mutations in our population are located, together with the screening of exon 56 deletion; ii) the second tier includes the remaining *LAMA2* exons; and finally iii) MLPA analysis (two panels). Until now, seventeen patients have been analyzed in this manner. In 35% of patients both mutated alleles were identified using tier 1, and in 82% at least one heterozygous mutation was detected. We consider feasible in our population to screen these

LAMA2 regions in patients with compatible features of CMD (such as white matter changes in brain MRI), even before performing a muscle biopsy.

Large deletions and duplications detected in CMD patients are not confined to LAMA2 gene; in fact we have recently reported a patient with a Fukuyama CMD caused by a multi-exonic duplication in *FKTN* (fukutin) (Costa et al., 2013). There are additional reports of other pathogenic CNVs in CMD genes, such as: *ISPD* (isoprenoid synthase domain containing) (Czeschik et al., 2013), *LARGE* (like-glycosyltransferase) (van Reeuwijk et al., 2007; Clarke et al., 2011) and *POMGNT1* [protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)] (Saredi et al., 2012).

New mutation screening methods are currently being developed based on next-generation sequencing (NGS) technology, which will contribute to establish the genetic causes of hereditary myopathies that remain unsolved. However, prior to its application, it is important to exclude large deletions and duplications as a cause of these diseases. Bioinformatic pipelines for NGS usually do not incorporate algorithms that enable their automatic detection, but we have previously shown that, when properly applied, this technology can help delineate large genomic rearrangements (Oliveira et al., 2013).

In summary, we have reassessed the impact of large deletions and duplications in *LAMA2*-RD and emphasize the importance of including screening for these rearrangements as part of the diagnostic strategy, especially in patients where a single heterozygous mutation has been detected.

3.5 *LAMA2* GENE MUTATION UPDATE: TOWARD A MORE COMPREHENSIVE PICTURE OF THE LAMININ-α2 VARIOME AND ITS RELATED PHENOTYPES

Oliveira J, Gruber A, Cardoso M, Taipa R, Fineza I, Gonçalves A, Laner A, Winder TL, Schroeder J, Rath J, Oliveira M. E., Vieira E, Sousa AP, Vieira JP, Lourenço T, Almendra L, Negrão L, Santos M, Melo-Pires M, Coelho T, den Dunnen JT, Santos R, Sousa M.(2018). Human Mutation, 39(10):1314-1337.

3.5.1 ABSTRACT

Congenital muscular dystrophy type 1A (MDC1A) is one of the main subtypes of early-onset muscle disease, caused by disease-associated variants in the laminin-α2 (*LAMA2*) gene. MDC1A usually presents as a severe neonatal hypotonia and failure to thrive. Muscle weakness compromises normal motor development, leading to the inability to sit unsupported or to walk independently. The phenotype associated with *LAMA2* defects has been expanded to include milder and atypical cases, being now collectively known as *LAMA2*-related muscular dystrophies (*LAMA2*-MD). Through an international multicenter collaborative effort, 61 new *LAMA2* disease-associated variants were identified in 86 patients, representing the largest number of patients and new disease-causing variants in a single report. The collaborative variant collection was supported by the LOVD-powered *LAMA2* gene variant database (http://www.LOVD.nl/LAMA2), updated as part of this work. As of December 2017, the database contains 486 unique *LAMA2* variants (309 disease-associated), obtained from direct submissions and literature reports. Database content was systematically reviewed and further insights concerning *LAMA2*-MD are presented. We focused on the impact of missense changes, especially the c.2461A>C (p.Thr821Pro) variant and its association with late-onset *LAMA2*-MD. Finally, we report diagnostically challenging cases, highlighting the relevance of modern genetic analysis in the characterization of clinically heterogeneous muscle diseases.

3.5.2 BACKGROUND

Laminin-211 is a heterotrimetric cruciform-shaped complex that establishes a stable link between the sarcolemma of muscle fibers and the extracellular matrix, being a major component of the extrasynaptic skeletal muscle basement membrane (BM) (Durbeej, 2015). The -211 classification derives from the three specific chains (α2, β1 and γ1) that compose this specific laminin form (Aumailley et al., 2005). Laminin-211 binds to the glycosylated residues of α-dystroglycan (α-DG) and also self-assembles (polymerizes) into networks through its N-terminal domain (Yurchenco, 2015). This supramolecular network connects to collagen IV and to perlecan (heparan sulfate proteoglycan) through nidogens cross-linking (Jones et al., 2000). Laminin-211 expression is not confined to skeletal muscle but has also been shown to be expressed in a variety of other tissues, more importantly in peripheral nerve (Schwann cells) and in brain (Yurchenco, 2015). Post-translational changes have been reported in laminin-211 components. More specifically, laminin-α2 chain was found to undergo cleavage at residue 2580 under specific conditions to generate an N-terminal 300 kDa peptide and a C-terminal 80 kDa peptide which are subsequently connected through a noncovalent process (Durbeej, 2015).

Disease associated (pathogenic) variants located in the gene that codes for the α2 chain (*LAMA2*; MIM# 156225) of laminin-211, give rise to a group of diseases collectively designated as *LAMA2*-related muscular dystrophy (*LAMA2*-MD). *LAMA2* maps to chromosome 6q22.33 and spans over 260 kb. It comprises 65 exons and codes for a protein with a molecular mass of approximately 390 kDa (Zhang et al., 1996). The majority of patients with *LAMA2* mutations have a congenital muscular dystrophy (CMD) phenotype classified as type 1A (MDC1A; MIM# 607855). The classical phenotype manifests as neonatal hypotonia or muscle weakness during the first months of life and reduced spontaneous movements (Helbling-Leclerc et al., 1995). As muscle weakness persists during development, it compromises the achievement of normal motor milestones (no cephalic control or inability to sit unsupported) and frequently gives rise to failure to thrive. Other manifestations such as gastroesophageal reflux, aspiration, recurrent chest infections, and even respiratory failure were reported in MDC1A (Jones et al., 2001). Facial muscle weakness, ophthalmoparesis and macroglossia are also features present in these patients but are often beyond early childhood (Quijano-Roy et al., 2012). Other relevant clinical hallmarks of MDC1A include elevated creatine kinase (CK) levels and dystrophic changes (necrosis and regeneration of fibers, chronic inflammation and fibrosis) recognizable in muscle biopsies of these patients (Tomé et al., 1994). Diagnostically important features are the complete absence of laminin-α2 staining evaluated by immunohistochemistry (IHC) performed in muscle or in skin biopsies (Sewry et al. 1996) using specific antibodies, and

typical white matter changes (WMC) in brain detectable by magnetic resonance imaging (MRI) (Lamer et al., 1998). WMC are related with alterations in the brain's water content, due to modifications in the maturation and/or function of the blood-brain barrier, and are detectable after the first six months to one year of life (Menezes et al., 2014). Besides WMC, brain structural defects have been reported in patients with laminin deficiency, in an estimated ~4% of *LAMA2*-MD cases (Jones et al., 2001). In some initial studies, performed before *LAMA2* genotyping was available, this association was based solely on laminin staining by IHC (Sunada et al., 1995; Pini et al., 1996; Brett et al., 1998; Martinello et al., 1998; Tsao et al., 1998; Philpot et al., 1999). It is plausible that any dystroglycanopathy could account for the partial laminin deficiency observed in some patients, explaining the diversity of structural brain defects reported. It is nonetheless consensual that primary laminin-alpha2 deficiency can contribute to structural abnormalities in the cerebral cortex during fetal development. Malformations found in patients with *LAMA2* disease-causing variants includes: i) cortical dysplasia (Mercuri et al., 1999), ii) changes within the lissencephaly spectrum, namely agyria or pachygyria (Geranmayeh et al 2010), and iii) polymicrogyria (Vigliano et al., 2009).

In a subset of MDC1A cases there is partial laminin-α2 deficiency (reduced / irregular laminin-α2 staining in IHC) which translates into a CMD with a slower disease progression (Oliveira et al., 2008; Geranmayeh et al., 2010). There is some degree of correlation between independent ambulation and IHC status of laminin-α2. The majority of MDC1A patients that do not acquire independent locomotion have complete laminin-α2 deficiency on muscle biopsy, whereas in the majority of cases that are able to walk independently a partial laminin-α2 deficiency has been documented (Geranmayeh et al., 2010).

Further to MDC1A, "milder" *LAMA2*-related phenotypes have been increasingly reported over the past few years. These late-onset *LAMA2*-MD patients are mainly characterized by proximal muscle weakness with onset during childhood, delayed motor milestones, achievement of independent ambulation, and persistently elevated CK levels (Gavassini et al., 2011). Some reports classified these patients as a subtype of limb-girdle muscular dystrophy (LGMD). Patients included in this group may also show muscle hypertrophy, rigid spine syndrome and pronounced joint contractures which are often more evident in the elbows. In addition to cardiac involvement in a limited number of cases, these clinical features are evocative of Emery-Dreifuss muscular dystrophy (EDMD) (Nelson et al., 2015). It should be emphasized that patients with late-onset *LAMA2*-MD still manifest typical brain WMC, but IHC labeling of laminin-α2 in muscle biopsy may show only very subtle changes.

As laminin-α2 is also expressed in Schwann cells, there is a range of clinical features related with peripheral nerve involvement in *LAMA2*-MD patients. In a particular series of MDC1A patients, the majority had decreased motor nerve conduction, suggesting that

peripheral demyelinating neuropathy is a disease feature (Shorer et al., 1995). Later it was also shown that laminin-α2 related neuropathic abnormalities also included sensory nerves (Quijano-Roy et al., 2004). More importantly, in a milder case of *LAMA2*-MD there was evidence of a myelinogenesis disorder, leading to the assumption that the neuropathy in laminin-α2 deficient cases is actually dysmyelinating (Di Muzio et al., 2003). These changes are more evident in milder *LAMA2*-MD patients (Mora et al., 1996; Deodato et al., 2002; Chan et al., 2014), whereas as in MDC1A presentations the more severe muscle involvement probably masks the subtle neuropathic features of the disease.

In terms of the mutation spectrum of the *LAMA2* gene, four independent studies described cohorts with more than twenty patients (Pegoraro et al., 1998; Oliveira et al., 2008; Geranmayeh et al., 2010; Xiong et al., 2015). The most frequent reported genotypes include variants that create premature termination codons (PTC) in both disease alleles, and are associated with complete deficiency of laminin-α2 in muscle biopsy as well as a MDC1A phenotype. In contrast, missense variants are present in a smaller number of cases and usually correlate with partial laminin-α2 deficiency giving rise to milder phenotypes. The asymmetrical proportion between truncating and non-truncating variants, explains the higher prevalence of MDC1A as compared with other emerging *LAMA2*-related phenotypes.

A relatively high frequency (18.4% of disease causing variants) of large deletions and duplications in *LAMA2* was also reported. Variants of this sort are detectable by multiplex ligation-dependent probe amplification (MLPA) or array comparative genomic hybridization (array-CGH) (Oliveira et al., 2014).

The *LAMA2* locus-specific database (LSDB), which we initiated in 2002, was continuously updated and used to assist the collection of new variants as reported here. Of the 486 unique variants registered to date (December 2017), a total of 61 novel disease associated variants detected in 86 patients are reported for the first time. Database content is systematically presented and further insights into the genotypes and phenotypes of *LAMA2*-MD are presented.

3.5.3 DEVELOPMENT AND UPDATE OF *LAMA2* LSDB

As part of the work we report the development of a comprehensive database for *LAMA2* variants, an important resource made available for the scientific community since 2002. The LOVD software (Fokkema et al., 2011) was used to store genetic and clinical data, allowing an off-the-shelf LSDB deployment in accordance with international guidelines for the curation and creation of these databases (Celli et al., 2012; Vihinen et al., 2012). The LSDB content was updated and migrated to LOVD version 3.0, being completely redesigned in terms of its

database architecture. Variant data was collected from publications accessed by the curators (64%) or through direct database submissions (36%). Currently (by December 2017), the *LAMA2*-LOVD contains a total of 1186 of entries (486 unique) identified in a total of 748 individuals. Based on disease impact, these entries comprise: 816 disease associated variants (309 unique), 317 benign (141 unique) and 53 variants of unknown clinical significance (VUS, 38 unique).

3.5.4 DESCRIPTION OF NOVEL *LAMA2* VARIANTS

A total of 61 novel disease-associated or likely associated variants were identified in the *LAMA2* gene (Table 3.5.1), representing more than 20% (61/309) of the total disease variants currently listed in the *LAMA2* LSDB. Variant interpretation followed the standards and guidelines for the classification of sequence variants, proposed by the American College of Medical Genetics and Genomics (ACMG) (Richards et al., 2015). The LOVD *LAMA2* database gives two classification, a Functional classification (column Effect) and a Clinical classification (column ClassClinical). The functional classification indicates the consequences of the variant for the function of the gene/protein (e.g. affects function), the clinical classification the consequences for the individual carrying the variant (e.g. ACMG:5, disease-associated, autosomal recessive (pathogenic)). The summary conclusion of the curators for specific variants, based on all individual observations of the variants, is given in a SUMMARY record. All unpublished variants collected and/or classified in the course of this project can be retrieved from the database using the following link: https://databases.lovd.nl/shared/references/DOI:10.1002/humu.23599.

Variants were identified by different international groups (material and methods in Appendix II.4 Data II.4.1), which reflects by the diversity of the patients' geographical origins (eleven distinct nationalities). Most variants are predicted to be truncating, 20 nonsense type and 23 small frame-shift variants (16 deletions and 7 duplications). In addition, this list includes a significant number of variants affecting canonical splice-sites (n=13), the majority located in donor sites (+1 and +2 positions). Due to the inability to obtain proper biological samples or study limitations it was mostly not possible to evaluate their impact at the mRNA level. Thus, the impact of these splice-site variants was evaluated with bioinformatic tools (see section 3.1), which for all variants indicated unequivocal deleterious effects. One fully characterized was c.819+2T>C, located in the donor splice-site of intron 5. Analysis by RT-PCR followed by sequencing, showed the presence of aberrant transcripts (details in the 'Diagnostic Relevance' section and Appendix II.4, Figures II.4.1 and II.4.2).

Table 3.5.1- Novel pathogenic variants identified in *LAMA2* gene listed in the locus-specific database.

Exon/ Intron	DNA Variant (NM_000426.3)	Interpretation [a]	DNA Variant (NC_000006.11) hg19	RNA Variant	Predicted effect on Protein	External variant databases	Number of entries in LSDB	Patient-ID in LSBD	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/ fibroblasts	Zygosity / 2nd variant / orientation (cis, trans or unknown)	Interpretation of the second variant
1	c.47del	Pathogenic	g.129204437delG	r.(?)	p.(Gly16Alafs*29)	-	1	102376	M	United States	MDC1A	-	Het. / c.2T>C / trans	Pathogenic
1	c.94C>T	Likely pathogenic	g.129204484C>T	r.(?)	p.(Gln32*)	-	1	102361	F	Canada	MDC1A	Deficiency	Het. / c.8245-2A>G / unknown	Likely pathogenic
2	c.164delA	Likely pathogenic	g.129371114delA	r.(?)	p.(Asn55Metfs*16)	-	1	102378	M	Canada	CMD	Deficiency	Het. / ? / unknown	No second pathogenic variant found
2i	c.283+2del	Likely pathogenic	g.129371235delT	r.(spl?)	p.(?)	-	1	102463	F	United States	MDC1A	Deficiency	Het. / c.1609-41_1609-7inv / unknown	VUS [1]
3i	c.396+1G>T	Pathogenic	g.129381042G>T	r.(spl?)	p.(?)	ClinVar (RCV000316746.1); dbSNP (rs770617208); gnomAD (0.0024%)	6	102366	F	United States	MDC1A	-	Het. / c.498G>A / unknown	Pathogenic
								102732	F	Mexico	MDC1A	-	Het. / c.5116C>T / unknown	Pathogenic
								102386	M	United States	MDC1A	-	Hom./ n.a./ n.a.	n.a.
								131976	M	United States	MDC1A	-	Het. / c.6501C>A / unknown	Likely pathogenic
								102478	F	United States	MDC1A	-	Het. / c.6690C>A / unknown	Pathogenic
								36041	M	Lybia	Unknown	-	Het. / c.8586T>G / trans	Pathogenic
4i	c.639+2T>A	Likely pathogenic	g.129419562T>A	r.spl?	p.(?)	-	1	102476	F	United States	MDC1A	-	Het. / c.2049_2050del / trans	Pathogenic

Exon	cDNA	Classification	Genomic	RNA	Protein	Database	n	ID	Sex	Country	Phenotype	Deficiency	Allele	Classification
5i	c.819+2T>C	Pathogenic	g.129465227T>C	r.[640_819del;640_1027del]	p.[Ile214_Arg273del; Ile214Hisfs*22]	-	3	102735 [2]	M	Portugal	Late-onset LAMA2-related MD	Partial deficiency	Het. / c.3976C>T / unknown	Pathogenic
								102736 [2]	F	Portugal	Late-onset LAMA2-related MD	-	Het. / c.3976C>T / unknown	Pathogenic
								103207 [3]	F	Portugal	Late-onset LAMA2-related MD	-	Het. / c.1854_1861dup / trans	Pathogenic
7	c.939_940del	Pathogenic	g.129470153_129470154del	r.(?)	p.(Cys314Trpfs*3)	gnomAD (0.0028%)	7	102373	F	United States	MDC1A	-	Het. / c.7732C>T / unknown	Pathogenic
								102382 [2]	F	United States	MDC1A	Partial deficiency	Het. / c.5562+5G>C / unknown	Pathogenic
								132007 [2]	M	United States	MDC1A		Het. / c.5562+5G>C / unknown	Pathogenic
								102396	F	United States	MDC1A	-	Hom. / n.a. / n.a	n.a.
								132008	F	United States	MDC1A	-	Het./ c.7658delC / unknown	Pathogenic
								132009	M	United States	MDC1A	-	Hom./ n.a./ n.a.	n.a.
								102655	M	United States	MDC1A	Deficiency	Het. / c.7732C>T / unknown	Pathogenic
7	c.991A>T	Likely pathogenic	g.129470205A>T	r.(?)	p.(Arg331*)	gnomAD (0.00041%)	1	102467	M	United States	MDC1A	-	Het. / c.5325dupA / unknown	Pathogenic
12	c.1762delG	Pathogenic	g.129513978delG	r.(?)	p.(Ala588Leufs*11)	Clinvar (RCV000171527.1); dbSNP (rs786205654)	7	102328	F	Saudi Arabia	MDC1A	-	Hom. / n.a. / n.a	n.a.
								102349	F	Unknown	MDC1A	Deficiency	Hom. / n.a. / n.a	n.a.
								132010	M	Saudi Arabia	MDC1A	-	Hom. / n.a. / n.a	n.a.

Exon	cDNA	Classification	Genomic	RNA	Protein	Reference	N	ID	Sex	Country	Phenotype	Expression	2nd allele	2nd classification
								132011	M	Saudi Arabia	MDC1A	-	Het. / c.1303C>T/ trans	Pathogenic
								132012 [4]	M	Saudi Arabia	MDC1A	-	Hom./ n.a./ n.a.	n.a.
								132013 [4]	M	Saudi Arabia	MDC1A	-	Hom./ n.a./ n.a.	n.a.
								102363	F	Saudi Arabia	MDC1A	-	Hom. / n.a. / n.a	n.a.
13	c.1823_1824 del	Likely pathogenic	g.129571297_129571298del	r.(?)	p.(Tyr608*)	dbSNP (rs754600708); gnomAD (0.00041%)	1	102471	F	United States	MDC1A	Deficiency	Hom. / n.a. / n.a	n.a.
14	c.2017G>T	Likely pathogenic	g.129573361G>T	r.(?)	p.(Glu673*)	-	1	102460	F	United States	MDC1A	Deficiency	Het. / c.2023_2024del / unknown	Likely pathogenic
14	c.2023_2024 del	Likely pathogenic	g.129573367_129573368del	r.(?)	p.(Met675Aspfs*29)	gnomAD (0.00041%)	1						Het. / c.2017G>T / unknown	Likely pathogenic
17	c.2350dup	Pathogenic	g.129591796dup	r.(?)	p.(Tyr784Leufs*3)	ClinVar (RCV000486406.1)	1	103206	F	Spain	MDC1A	-	Het. / c.4692_4695dup / unknown	Pathogenic
17	c.2383G>T	Likely pathogenic	g.129591829G>T	r.(?)	p.(Glu795*)	dbSNP (rs149896793); ESP (0.01%); gnomAD (0.00041%)	1	102397	F	United States	MDC1A	-	Het. / c.4761dupT / unknown	Likely pathogenic
17i	c.2450+4A>G	Likely pathogenic	g.129591900A>G	r.(spl?)	p.(?)	-	2	103191	F	Portugal	MDC1A	Partial deficiency	Het. / c.8244+1G>A / unknown	Pathogenic
								103972	M	Portugal	Late-onset LAMA2-related MD	-	Het. / c.7750-1713_7899-2154del / unknown	Pathogenic
18i	c.2538-1G>A	Likely pathogenic	g.129608991G>A	r.spl?	p.(?)	-	2	102547	F	United States	MDC1A	Deficiency	Het. / c.3735+2T>A / unknown	Likely pathogenic
								132014	F	United States	MD	-	Het. / ? / unknown	No 2nd pathogenic variant found

21	c.2875C>T	Pathogenic	g.129618848C>T	r.(?)	p.(Gln959*)	-	1	102661	F	Saudi Arabia	MDC1A	-	Hom. / n.a. / n.a	n.a.
23	c.3338_3345 dup	Likely pathogenic	g.129634169_129634176dup	r.(?)	p.(Thr1116Glnfs*26)	-	1	102385	M	United States	MDC1A	-	Het. / c.6207C>A / unknown	Likely pathogenic
23	c.3372dup	Likely pathogenic	g.129634203dup	r.(?)	p.(Cys1125Metfs*4)	-	1	103970	U [6]	Portugal	LGMD/ EDMD [6]	-	Het. / c.2461A>C / unknown	Pathogenic
24	c.3472A>T	Likely pathogenic	g.129635860A>T	r.(?)	p.(Lys1158*)	-	1	102324	F	United States	Unknown	-	Het. / ? / unknown	No second pathogenic variant found
24	c.3520C>T	Pathogenic	g.129635908C>T	r.(?)	p.(Gln1174*)	-	1	103127	F	Spain	MDC1A	-	Het. / c.3976C>T / trans	Pathogenic
25	c.3560_3569 del	Pathogenic	g.129636625_129636634del	r.(?)	p.(Thr1187Metfs*9)	-	1	102486	F	Canada	MDC1A	Deficiency	Hom. / n.a. / n.a	n.a.
25i	c.3735+2T>A	Likely pathogenic	g.129636802T>A	r.spl?	p.(?)	-	1	102547	F	United States	MDC1A	Deficiency	Het. / c.2538-1G>A / unknown	Likely pathogenic
26	c.3829C>T	Likely pathogenic	g.129637000C>T	r.(?)	p.(Arg1277*)	-	1	102383	M	Canada	MDC1A	Deficiency	Het. / c.4654G>A / trans	VUS [1]
27	c.4002T>G	Pathogenic	g.129637260T>G	r.(?)	p.(Tyr1334*)	-	1	102535	F	United States	MDC1A	-	Het. / c.7658delC / unknown	Pathogenic
27	c.4049del	Pathogenic	g.129637307delG	r.(?)	p.(Arg1350Hisfs*12)	-	2	102663	M	United States	MDC1A	Deficiency	Het. / c.2049_2050del / trans	Pathogenic
								102401	F	United States	MDC1A	-	Het. / c.5562+5G>C / trans	Pathogenic
29	c.4261C>T	Pathogenic	g.129649507C>T	r.(?)	p.(Gln1421*)	-	1	102662 [4]	M	United States	MDC1A	-	Het. / c.283+1G>A / trans	Pathogenic
30	c.4348C>T	Pathogenic	g.129663524C>T	r.(?)	p.(Arg1450*)	ClinVar (RCV000171401.1); dbSNP (rs200923373); gnomAD (0.0012%)	1	102364	M	United States	MDC1A	Deficiency	Het. / c.2049_2050del / trans	Pathogenic

33	c.4761dup	Likely pathogenic	g.129687407dupT	r.(?)	p.(Arg1588Serfs*20)	-	1	102397	F	United States	MDC1A	-	Het. / c.2383G>T / unknown	Likely Pathogenic
34	c.4941del	Likely pathogenic	g.129691117delG	r.(?)	p.(Met1647Ilefs*5)	-	1	102353	M	United States	Father of affected child (carrier study)	-	Het. / n.a. / n.a.	n.a.
35	c.5050G>T	Pathogenic	g.129704357G>T	r.(?)	p.(Glu1684*)	ClinVar (RCV000078775.3; RCV000177827.2): dbSNP (rs201632009)	1	131883	M	Italy	MDC1A	Deficiency	Het. / c.2901C>A / trans	Pathogenic
35i	c.5072-1454_5154delinsAGATTGCC	Likely pathogenic	g.129711182_129712718delinsAGATTGCC	r.spl?	p.(?)	-	1	102381	M	United States	MDC1A	-	Hom. / n.a. / n.a	n.a.
36	c.5132del	Pathogenic	g.129712696delA	r.(?)	p.(Glu1711Glyfs*14)	-	1	102658	M	United States	MDC1A	-	Het. / c.363C>A / trans	Pathogenic
36	c.5134_5153del	Pathogenic	g.129712698_129712717del	r.(?)	p.(Arg1712Glufs*4)	-	1	111376	M	Turkey	MDC1A	Deficiency	Hom. / n.a. / n.a	n.a.
36	c.5182del	Likely pathogenic	g.129712746delC	r.(?)	p.(Leu1728*)	-	1	102358	F	United States	MDC1A	Deficiency	Hom. / n.a. / n.a	n.a.
37	c.5259del	Pathogenic	g.129714214delA	r.(?)	p.(Val1754*)	-	1	102469	F	Israel	MDC1A	Deficiency	Het. / c.7147C>T / trans	Pathogenic
37	c.5263A>T	Pathogenic	g.129714218A>T	r.(?)	p.(Lys1755*)	-	1	103189	M	Iran	MDC1A	Deficiency	Het. / c.6501C>G/ unknown	Pathogenic
41	c.5914C>T	Pathogenic	g.129748945C>T	r.(?)	p.(Gln1972*)	ClinVar (RCV000078782.3, RCV000178452.1); dbSNP (rs398123378)	4	102365	M	United States	Unknown	-	Hom. / n.a. / n.a	n.a.
								102384	M	United States	MDC1A	Deficiency	Hom. / n.a. / n.a	n.a.
								132015 [4]	F	United States	Unknown	-	Het. / ? / unknown	No second pathogenic variant found

								102461	F	United States	MDC1A	-	Hom. / n.a. / n.a	n.a.
42	c.5998delA	Likely pathogenic	g.129759820delA	r.(?)	p.(Thr2000Profs*3)	-	1	102362	M	United States	MDC1A	Deficiency	Het. / c.7147C>T / unknown	Pathogenic
43	c.6207C>A	Likely pathogenic	g.129762082C>A	r.(?)	p.(Tyr2069*)	dbSNP (rs143343647); ESP (0.02%)	1	102385	M	United States	MDC1A	-	Het. / c.3338_3345dup / unknown	Likely pathogenic
43	c.6266del	Pathogenic	g.129762141delA	r.(?)	p.(Asn2089Thrfs*14)	-	1	102371	F	Mexico	MDC1A	Deficiency	Het. / c.2962C>T / trans	Pathogenic
45i	c.6429+1G>T	Likely pathogenic	g.129766967G>T	r.spl?	p.(?)	gnomAD (0.0032%)	2	102400	M	United States	MDC1A	Deficiency	Het. / c.2901C>A / unknown	Pathogenic
								132016	M	United States	Unknown	Partial deficiency	Hom./ n.a./ n.a.	n.a.
46	c.6501C>G	Pathogenic	g.129774204C>G	r.(?)	p.(Tyr2167*)	-	1	103189	M	Iran	MDC1A	Deficiency	Het. / c.5263A>T/ unknown	Pathogenic
47	c.6588dup	Likely pathogenic	g.129775314dupT	r.(?)	p.(Ile2197Tyrfs*5)	gnomAD (0.00041%)	1	102379	F	United States	MDC1A	Partial deficiency	Het. / c.7571A>T/ unknown	VUS [1]
49	c.6979G>T	Pathogenic	g.129781456G>T	r.(?)	p.(Gly2327*)	-	1	103192	M	Iran	MDC1A	-	Hom. / n.a. / n.a	n.a.
51	c.7297C>T	Likely pathogenic	g.129786431C>T	r.(?)	p.(Gln2433*)	-	1	102334	F	United States	MDC1A	Deficiency	Het. / c.35T>G / unknown	Pathogenic
54	c.7491del	Likely pathogenic	g.129799877delA	r.(?)	p.(Asp2498Ilefs*49)	-	1	102480	M	United States	MDC1A	Deficiency	Het. / c.8244+3_8244+6del / unknown	Likely pathogenic
54i	c.7572+1G>A	Pathogenic	g.129799959G>A	r.spl?	p.(?)	-	2	111379	F	Germany	Late-onset LAMA2-related MD	-	Het. / c.245A>T / unknown	VUS [1]
								102548	F	United States	MDC1A	Deficiency	Het. / c.2T>C / trans	Pathogenic
56i	c.7898+2T>G	Likely pathogenic	g.129807769T>G	r.spl?	p.(?)	-	1	111374	M	Turkey	MDC1A	Deficiency	Hom. / n.a. / n.a	n.a.
56i_65_	c.7898+732_*39282del	Likely pathogenic	g.129808499_129876774del	r.(?)	p.(?)	-	1	102475 [5]	M	Saudi Arabia	MDC1A	-	Hom. / n.a. / n.a	n.a.

58i	c.8244+1G>C	Likely pathogenic	g.129813629G>C	r.spl?	p.(?)	-	1	102380	M	United States	MDC1A	-	Hom. / n.a. / n.a	n.a.
58i	c.8244+2dup	Likely pathogenic	g.129813630dup	r.spl?	p.(?)	-	1	111380	F	Saudi Arabia	MDC1A	-	Hom. / n.a. / n.a	n.a.
59i	c.8357+1G>A	Pathogenic	g.129823917G>A	r.spl?	p.(?)	-	1	102391	M	United States	MDC1A	Deficiency	Het. / c.2049_2050del / trans	Pathogenic
61	c.8556_8558del	Likely pathogenic	g.129826353_129826355del	r.(?)	p.(Ile2852del)	ClinVar (RCV000078805.4); dbSNP rs398123389	1	102737	M	Portugal	MDC1A	Partial deficiency	Het. / c.5234+1G>A / unknown	Pathogenic
61	c.8586T>G	Pathogenic	g.129826383T>G	r.(?)	p.(Tyr2862*)	-	1	36041	M	Lybia	Unknown	-	Het. / c.396+1G>A / trans	Likely pathogenic
61	c.8669dup	Pathogenic	g.129826466dupT	r.(?)	p.(Leu2890Phefs*16)	-	3	102395	M	United States	MDC1A	Deficiency	Het. / c.2370T>A / trans	VUS [1]
								102472	M	United States	MDC1A	Deficiency	Het. / c.2049_2050del / trans	Pathogenic
								102369	M	United States	MDC1A	Partial deficiency	Het. / ? / unknown	No second pathogenic variant found
63	c.8947C>T	Pathogenic	g.129833597C>T	r.(?)	p.(Gln2983*)	-	1	111373	M	Turkey	MDC1A	Deficiency	Het. / c.6955C>T / trans	Pathogenic
64	c.9095dup	Likely pathogenic	g.129835624dupA	r.(?)	p.(Ile3033Aspfs*6)	-	2	102474	M	United States	MDC1A	Partial deficiency	Het. / c.5562+5G>C / unknown	Pathogenic
								132025 [4]	M	United States	Unknown	-	Het./ c.4860G>A / unkown	VUS

Footnote: CMD- congenital muscular dystrophy; F- female; Het.- heterozygous; Hom.- homozygous; ID#- identification number; IHC- immunohistochemistry; M-male; MD- muscular dystrophy; MDC1A- congenital muscular dystrophy type 1A; n.a.- not applicable; VUS- variant of unknown significance. [1]- Variant listed in table 2; [2]- Siblings; Variants were detected by Sanger sequencing, except for: [3]- Whole-exome sequencing (patient ID# 103207), [4]- NGS gene panel (patients' ID#s: 102662, 132012, 132013, 132015, 132025), and [5]- Array-CGH (patient ID# 102475), more details are available in Appendix II.4 Figure II.4.3; [6]- Variant identified through an anonymized screening performed in genetically uncharacterized LGMD/EDMD patients; [a]- According to the ACMG guidelines. References sequences used to describe variants: M_000426.3 and NC_000006.11.

In addition to the most prevalent type of variants already stated, the remainder include: i) two missense variants (one of which might also have an effect on splicing), ii) one in-frame (IF) codon deletion, iii) one deletion-insertion variant, and iv) a large deletion encompassing exons 57 to 65. This large deletion was detected in a homozygous patient with a MDC1A phenotype by array-CGH technique (Appendix II.4, Figure II.4.3). The 61 new variants were identified in 87 patients (85 families) and in one obligate carrier. In terms of genotypes, a total of 25 patients had homozygous variants. In the remaining 57 cases compound heterozygous variants were found: 52 classified as pathogenic and 5 VUS. From this cohort, only five cases (5.7%) had incomplete genotyping as only one mutated allele was identified. Four variants were detected in more than three unrelated patients: c.939_940del (n=7), c.1762del (n=7), c.396+1G>T (n=6), and c.5914C>T (n=4). This is explainable by study inclusion criteria and the higher frequency of variants identified in patients of specific populations or ethnic groups still underrepresented in the literature (c.1762del in patients from Saudi Arabia, and the others were found in patients with Hispanic ancestry). Finally, concerning the clinical presentation, the majority of patients were classified as MDC1A, whereas only seven had onset beyond infancy, and achieved independent locomotion. This particular phenotype ("late-onset" *LAMA2*-MD) was seen in patients with splicing variants or missense substitutions, presumably non-truncating alleles, and with partial *LAMA2* deficiency documented in some of the cases.

Bioinformatic analysis of novel LAMA2 variants

The novel *LAMA2* variants, especially those of the missense type and/or predicted to affect splicing, were further assessed resorting to bioinformatic prediction tools. A total of fourteen variants fitting into these categories are shown in Table 3.5.2. With the exception of homozygotes, all variants are heterozygous and found in combination with a second change known to be disease-associated or likely disease-associated. Since experimental evidence could not be obtained, bioinformatic analysis were pivotal to attempt their classification in terms of pathogenicity.

Listed in Appendix II.4 Data II.4.3 are the *in silico* tools used to evaluate variants, more specifically tolerance predictors and splicing predictors. Considering the extensive list of tools available to evaluate missense variants, we sought to determine which could be most efficient in the case of *LAMA2* variants. The performance measures for binary classifiers, as described by Niroula & Vihinen (2016), were calculated for nine tolerance predictors algorithms. Two control sets of *LAMA2* variants consistently classified in LOVD database as either pathogenic (n=29) or benign (n=22) (http://tiny.cc/LAMA2_2018) were used to perform these calculations. The tools with best performance for our purpose, based on the accuracy and Matthews correlation coefficient (MCC), were MutPred2 (Pejaver et al., 2017),

PolyPhen-2 (HumVar) (Adzhubei et al., 2010), SIFT (Kumar et al., 2009) and UMD-Predictor (Salgado et al., 2016) (Appendix II.4, Table II.4.1). These algorithms were subsequently applied to evaluate the new missense variants reported in this work (Table 3.5.2).

The majority of variants listed in Table 3.5.2 were inferred as being of the missense type (n=8). Five variants were consistently classified as deleterious by all four tolerance predictors used, and two variants were classified as deleterious by three out of the four algorithms. The other missense variant was considered deleterious by half of the tolerance predictors.

In four variants, also inferred to be of the missense type, a dual effect was predicted as they could also influence the splicing mechanism. In these, the majority of algorithms tested to evaluate missense variants consistently pointed towards intolerance (all except MutPred2 in three and SIFT in one of the variants) and also indicated an effect on splicing by all tools used, except GeneSplicer in two of the variants.

From the two variants remaining in Table 3.5.2, one is an apparently synonymous substitution predicted to create a new acceptor splice-site by three distinct algorithms (all except GeneSplicer) and the other is a large intronic inversion that was predicted to disrupt the canonical acceptor splice-site.

3.5.5 BIOLOGICAL RELEVANCE: CONTENT ANALYSIS OF THE LAMA2 LSDB

An overview of disease associated variants found in the *LAMA2* gene is shown in Figures 3.5.1 and 3.5.2. These may be subdivided as: 59.6% SNV (n=184 unique; 496 in total), 24.9% small deletions (n=77; 214), 8.7% small insertions (n=27; 62), 6.2% large deletions or duplications (n=19; 42), and two deletion/insertions (0.6%). In terms of their foreseeable impact, the most frequent group is that of variants that cause a premature termination code (PTC). These include nonsense (n=79) and out-of-frame (OF) changes (65 deletions, 23 duplications). A total of 79 variants were predicted or experimentally demonstrated to affect splicing. The first and last two conserved nucleotides of introns concentrate the vast majority of splicing variants. It should be highlighted that both PTC-inducing and splice-site variants are widespread throughout the gene with no clear 'mutational' hotspots. In terms of distribution throughout the gene, missense variants (n=40, 13% of total disease associated variants) do not follow a similar pattern; they seem to cluster in specific regions of laminin-α2 (Figure 3.5.1). The first group (n=11, 27.5% of missense variants) affect residues located in domain VI corresponding to the N-terminal part of the laminin-α2. A possible explanation is that missense variants located this region have a detrimental effect on laminin-211 function through the disruption of protein folding and loss of the ability for polymerization into

Table 3.5.2- New *LAMA2* missense changes and other variants possibly affecting splicing based on bioinformatic prediction tools.

Exon/ Intron	DNA Variant (NM_000426 .3)	Interpre tation [a]	DNA Variant (NC_000006 .11) hg19	RNA Vari ant	Predicted effect on Protein	External variant database s	Bioinformatic assessment [b]	Patient- ID in LSDB	Gen der	Geogr aphic origin	Phenoty pe	IHC for LAMA2 in muscle/ skin	Zygosity / second variant / orientation (cis, trans or unknown)	Interp retatio n of the secon d varian t
1	c.112G>A	VUS	g.12920450 2G>A	r.(?) ^ r.(sp l?)	p.(Gly38Ser) ^ p.(?)	-	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2) -Effect on splicing predicted in 4/4 of tools tested	102536	M	Saudi Arabia	MDC1A	Deficien cy	Hom. / n.a. / n.a.	n.a.
2	c.245A>T	VUS	g.12937119 5A>T	r.(?)	p.(Gln82Leu)	-	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2)	111379	F	Germa ny	Late- onset LAMA2- related MD	-	Het. / c.7572+1G> A / unknown	Patho genic
3	c.437C>T	VUS	g.12941935 8C>T	r.(?)	p.(Ser146Ph e)	ClinVar (RCV0005 05761.1); dbSNP (rs143680 577)	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102470	M	United States	MDC1A	-	Het. / c.6617del / unknown	Patho genic
5	c.745C>T	VUS	g.12946515 1C>T	r.(?)	p.(Arg249Cy s)	dbSNP (rs376437 110); ESP (0.01%); gnomAD (0.0014 %)	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102368	M	United States	MDC1A	-	Het. / c.9101_910 4dup/ unknown	Patho genic
5	c.818G>A	VUS	g.12946522 4G>A	r.(?)	p.(Arg273Ly s)	ClinVar (RCV0003 94734.1)	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2)	102398	M	United States	MDC1A	-	Het. / c.3976C>T / unknown	Patho genic

#	cDNA	Class.	Genomic	RNA	Protein		Database	Prediction	ID	Sex	Country	Disease	Laminin	Other allele	Class.
10	c.1326T>G	VUS	g.129498870T>G	r.(?)	p.(Cys442Trp)	-		Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102360	M	United States	MDC1A	Deficiency	Het. / c.3976C>T / unknown	Pathogenic
									102549	M	United States	MDC1A	-	Het. / c.7658delC / unknown	Pathogenic
11i	c.1609-41_1609-7inv	VUS	g.129513784_129513818inv	r.(spl?)	p.(?)	-		-Effect on splicing predicted in 4/4 of tools tested	102463	F	United States	MDC1A	Deficiency	Het. / c.283+2delT / unknown	Likely Pathogenic
17	c.2370T>A	VUS	g.129591816T>A	r.(?)^r.(spl?)	p.(790=)^p.(?)	-		Possibly affects function: -Effect on splicing predicted in 3/4 of tools tested (all except NNSplice)	102395	M	United States	MDC1A	Deficiency	Het. / c.8669dupT / trans	Pathogenic
23	c.3235T>G	Pathogenic	g.129634066T>G	r.(?)	p.(Cys1079Gly)	-		Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102726	F	Portugal	Late-onset LAMA2-related MD	Normal	Het. / c.7750-1713_7899-2154del / unknown	Pathogenic
									102718	M	Portugal	Late-onset LAMA2-related MD	Partial deficiency	Het. / c.3085C>T / trans	Pathogenic
31	c.4523G>A	Pathogenic	g.129670529G>A	r.(?)^r.(spl?)	p.(Arg1508Lys)^p.(?)		ClinVar (RCV000483171.1); dbSNP (rs770084568); gnomAD (0.00041%)	Possibly affects function: -Missense variant not tolerated in 2/4 predictors (UMD and PolyPhen-2) -Effect on splicing predicted in 4/4 of tools tested	102457 [1]	F	United States	MDC1A	-	Het. / c.2049_2050del / trans	Pathogenic
									102484 [1]	F	United States	MDC1A	-	Het. / c.2049_2050del / trans	Pathogenic
									102656	M	United States	MDC1A	-	Hom. / n.a. / n.a	n.a.
32	c.4654G>A	VUS	g.129674439G>A	r.(?)	p.(Ala1552Thr)		dbSNP (rs771891309) gnomAD (0.0012%)	-Missense variant not tolerated in 2/4 predictors (UMD and PolyPhen-2)	102383	M	Canada	MDC1A	-	Het. / c.3829C>T / trans	Likely Pathogenic
46	c.6548T>G	VUS	g.129774251T>G	r.(?)	p.(Leu2183Arg)	-		Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102399	M	Iran	CMD	-	Hom. / n.a. / n.a.	n.a.

47	c.6707G>A	VUS	g.12977543 3G>A	r.(?) ^ r.(sp l?)	p.(Arg2236L ys)^p.(?)	-	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used -Effect on splicing predicted in 3/4 of tools tested (all except GeneSplicer)	103971	U [1]	Portug al	LGMD/ EDMD [2]	-	Het. / c.2461A>C / unknown	Patho genic
54	c.7571A>T	VUS	g.12979995 7A>T	r.(?) ^ r.(sp l?)	p.(Glu2524V al)^p.(?)	-	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2) -Effect on splicing predicted in 3/4 of tools tested (all except GeneSplicer)	102379	F	United States	MDC1A	Partial deficienc y	Het. / c.6588dupT / unknown	Likely pathog enic

Footnote: CMD- congenital muscular dystrophy; F- female; Het.- heterozygous; Hom.- homozygous; ID- identification; IHC- immunohistochemistry; M- male; MD- muscular dystrophy; MDC1A- congenital muscular dystrophy type 1A; n.a.- not applicable; U- unknown; VUS- variant of unknown (clinical) significance; [1]- Siblings; [2]- Variants identified through an anonymized screening performed in genetically uncharacterized LGMD/ EDMD patients; [a]- According to the ACMG guidelines. [b]- More detailed information available in Appendix II.4 Table II.4.2. References sequences used to describe variants: NM_000426.3 and NC_000006.11.

supramolecular networks, that occurs through a cooperative self-assembly process of laminin-211 (Durbeej, 2015; Yurchenco, 2015). A subset of these missense substitutions (namely p.Tyr138His, p.Gln167Pro, p.Leu243Pro and p.Gly284Arg) are located on the presumed polymerization face near a patch containing the sequence P-L-E-N-G-E, corresponding to residues 208–213 of laminin-α2 (Yurchenco, 2015). These changes were identified in patients with late-onset *LAMA2*-MD with moderately reduced protein levels.
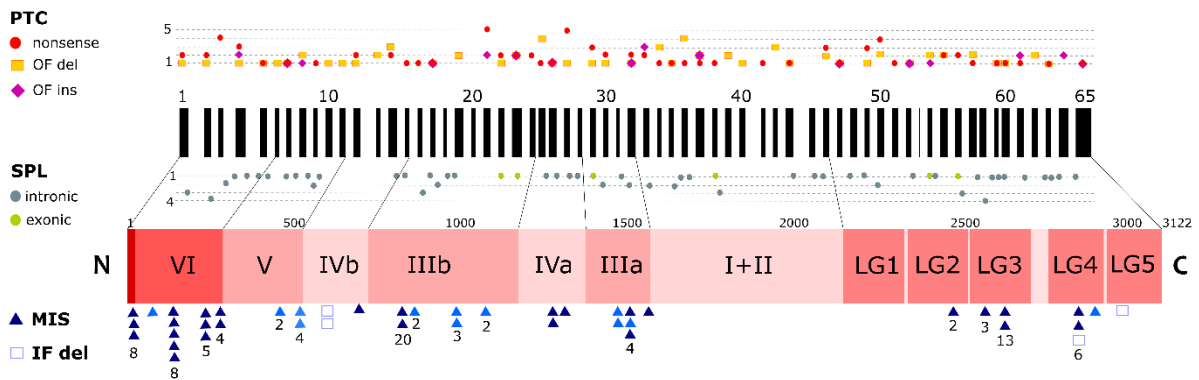
Figure 3.5.1- Point variants recorded in the *LAMA2*-LOVD. Top layer unveils the number of unique variants that originate premature termination codons (PTC): nonsense, out-of-frame (OF) deletions (DEL) or insertions (INS), per *LAMA2* exon (black rectangles). Middle layer shows splice-site variants (SPL), also indicating the number of unique variants per region: intronic (grey) or exonic (light green). Laminin-211 protein domains: I to VI, and Laminin G-like (LG) are shown in light pink to red boxes, from the C-terminal (C) to N-terminal region (N). Bottom layer displays missense (MIS) changes or single codon in-frame (IF) deletions. Light blue triangles indicate substitution of a cysteine. Variants are clustered per exonic region, and numbers bellow each symbol indicate the total number of changes.

The next cluster consists of missense variants (n=10, 25% of total) that specifically alter cysteine residues, located in one of the three EGF-like repeats (domains V, IIIb and IIIa), known to establish disulfide bridges. Here, the solenoid like structure conveyed by these rigid rod-like structures is probably modified in a way that alters the integrity of the connection between the sarcolemma and extracellular matrix mediated by laminin-α2. The last group of missense variants affects residues located in the C-terminal region of the protein that contains a tandem of five laminin G-like (LG) domains - LG1-5. A total of seven disease associated missense variants (17.5% of all missense) are in LG2, LG3 or LG4 domains. LG4 and LG5 domains mediate the binding of laminin-α2 to the O-linked carbohydrate chains of α-dystroglycan, whereas LG2 and LG3 bind to integrin α7/β1. Rare missense variants and in-frame deletions pose a problem for genetic data interpretation. Four IF codon deletions have been reported so far and, since there are no functional analysis strategies currently available

for laminin-α2 variants, their impact remains unclear. A second aspect to be considered, as highlighted before, is that some changes predicted to be missense may instead have an effect on mRNA splicing. In addition to the application of bioinformatics tools used to assess the pathogenicity of missense changes (such as those mentioned in *Bioinformatic analysis of novel LAMA2 variants*) it is advisable to consider if their location coincides with the potential hotspots outlined here.

Since our previous assessment (Oliveira et al., 2014) only two novel large deletions have been reported (Ding et al., 2016; Bhowmik et al., 2016), totaling seventeen deletions and two duplications (Figure 3.5.2). There are two apparent mutational "hotspots" for large deletions, the first region includes exons 3 and 4, and the second is in the 3' end of *LAMA2* gene (exons 56 to 65).

Figure 3.5.2- Large deletions and duplications listed in the *LAMA2*-LOVD. Large duplications (DUP) are shown in the top layer as yellow rectangles encompassing affected gene regions, and deletions (DEL) are shown in the bottom part of the picture as red rectangles. Black rectangles represent the *LAMA2* gene exons. Grey boxes indicate undetermined breakpoints and numbers the total of entries in the database (otherwise only one entry is present).

Considering the distribution of disease associated variants, exons 14, 21, 22, 26, 27, 36, 38 and 56 contain over 25 variant entries. In contrast, seven exons (namely 20, 28, 44, 45, 48, 53 and 58) have no disease-causing variants reported so far. Fourteen disease associated variants are among the most prevalent in the *LAMA2*-LOVD database, with at least 10 independent entries each (Table 3.5.3). The most frequent across different ethnical backgrounds are: c.2049_2050del (p.Arg683Serfs*21), c.3085C>T (p.Arg1029*) and c.3976C>T (p.Arg1326*). Interestingly, these variants are also represented in population variant databases such as gnomAD (ExAC), found in heterozygosity with frequencies

ranging from 0.012% to 0.001%. Other variants such as c.1854_1861dup (p.Leu621Hisfs*7), seem to be population- or ethnic group-specific, exhibiting a relatively high frequency (0.23%) within control alleles from the "Latino" population (gnomAD).

3.5.6 CLINICAL RELEVANCE: THE EXPANDING DISEASE SPECTRUM OF *LAMA2*-RELATED MD

Genotype-phenotype correlations

The severest end of the spectrum of *LAMA2*-related MD - MDC1A - corresponds to a neonatal onset disease that gives rise to hypotonia and compromised normal motor development. In *LAMA2*-related MD the locomotion attainment has been considered an important clinical measure of disease severity. In a series of 26 MDC1A patients only two had acquired independent locomotion (Oliveira et al., 2008). Interestingly, all patients that harbored variants inducing PTC in both disease alleles were unable to achieve independent walking. In contrast, the two patients that were able to walk had a missense or a single codon deletion in one of the disease genes. LSDB content and other studies reported in the literature (Geranmayeh et al., 2010) further corroborated our findings. However, there are exceptions to this rule; for example, a patient with a homozygous nonsense variant (p.Arg1549*) was able to reach ambulation and even climb stairs (Geranmayeh et al., 2010). This particular variant was reported in association with partial deficiency of laminin-α2 in several unrelated patients (Pegoraro et al., 1998; Di Blasi et al., 2000; Geranmayeh et al., 2010). Here, an explanation for this discrepancy is the fact that the variant is located within exon 32 that undergoes alternative splicing (Pegoraro et al., 2000). The exon removal leads to an IF deletion at the mRNA level, thereby restoring the reading frame from the PTC created by the nonsense variant.

Geranmayeh and co-workers (2010) provided further genotype-phenotype correlations with prognostic clinical implications. Statistically significant differences were identified between patients with complete deficiency and those with partial deficiency of laminin-α2. Patients with absence of laminin-α2 had earlier onset (P = 0.0073), lack of independent ambulation (P = 0.0215), and were more prone to requiring artificial feeding (P = 0.0099) or respiratory support (P = 0.0354) (Geranmayeh et al., 2010). Within MDC1A, there is a subset of patients with early onset phenotype but a "milder" disease progression and with partial laminin-α2 deficiency. This partial deficiency is often associated with missense variants, IF deletions and splicing variants (leaky or inducing IF exon-skipping) (Allamand & Guicheney, 2002; Quijano-Roy et al., 2012). One of the earliest such cases reported had a homozygous variant (p.Cys996Arg) that affects domain IIb of laminin-α2 (Nissinen et al., 1996).

Table 3.5.3- List of the most frequent pathogenic variants in the *LAMA2* LSDB (variants with 10 or more entries in LOVD).

Exon/ Intron	DNA Variant (NM_000426.3)	DNA Variant (hg19)	RNA Variant	Predicted effect on Protein	Number of independent entries in LOVD	Geographic origin of patients (LOVD)	gnomAD / ExAC data: population, nr alleles/ total alleles (Frequency)
13	c.1854_1861dup	g.129571328_129571335dup	r.1854_1861dup	p.Leu621Hisfs*7	12	France, Portugal, Brazil, Spain	Latino: 2/838 (0.23%)
14	c.2049_2050del	g.129573393_129573394del	r.2049_2050del	p.Arg683Serfs*21	42	Several countries	All populations except Ashkenazi Jewish: 34/277,008 (0.012%)
18	c.2461A>C	g.129601216A>C	r.2461a>c	p.Thr821Pro	18	Portugal	-
22	c.3085C>T	g.129621928C>T	r.(?)	p.(Arg1029*)	24	Portugal, Spain, United States	Latino: 1/34,418 (0.003%); European (Non-Finnish): 1/126,676 (0.001%)
26i	c.3924+2T>C	g.129637097T>C	r.3736_3924del	p.Leu1246_Glu1308del	33	Saudi Arabia, Sudan, United States	-
27	c.3976C>T	g.129637234C>T	r.(?)	p.(Arg1326*)	24	Portugal, Sweden, United States, Spain, Denmark	European (Non-Finnish): 14/126,524 (0.001%); European (Finnish): 1/25,788 (0.004%)
32	c.4645C>T	g.129674430C>T	r.[4645c>u, 4580_4717del]	p.[Arg1549*, Cys1527_Val1572del]	10	Australia, Italy, United States	South Asian: 2/30,780 (0.006%)
36i	c.5234+1G>A	g.129712799G>A	r.5072_5234del	p.Val1765Serfs*21	10	Portugal, Canada, United States	Latino: 1/34,376 (0.003%); European (Non-Finnish): 2/126,266 (0.002%)
38	c.5476C>T	g.129722399C>T	r.(?)	p.(Arg1826*)	11	China, Saudi Arabia, United Kingdom	East Asian: 2/17,240 (0.012%); European (Non-Finnish): 5/111,398 (0.0045%)

38i	c.5562+5G>C	g.129722490G>C	r.[5446_5562del, 5562_5563ins5562+1_5562+11{5562+5g>c}]	p.[Lys1816_Asp1854del, Tyr1855Valfs*24]	14	United Kingdom, United States	European (Non-Finnish): 7/125,864 (0.0056%); European (Finnish): 1/25,408 (0.0039%)
46	c.6488delA	g.129774191delA	r.(?)	p.(Lys2163Argfs*12)	15	Qatar, Saudi Arabia, United States	-
55	c.7732C>T	g.129802567C>T	r.(?)	p.(Arg2578*)	12	China, Denmark, Mexico, Russian Federation, United States	Latino: 3/34,380 (0.0087%); European (Non-Finnish): 10/126,598 (0.0079%); East Asian: 1/18,834 (0.0053%); South Asian: 1/30,782 (0.0032%)
55i_56i	c.7750-1713_7899-2154del	g.129805906_129810892del	r.7750_7898del	p.Ala2584Hisfs*8	17	Portugal	-
58i	c.8244+1G>A	g.129813629G>A	r.8076_8244del	p.Pro2693Valfs*12	12	Germany, Portugal, Tunisia, United States	European (Non-Finnish): 1/111,114 (0.0009%)

Footnote: nr- number.

Despite the general consistency between phenotype, the type of variant and the IHC status, some exceptions have been documented in the literature. These include patients with complete laminin-α2 deficiency and missense variants that achieved independent locomotion (Geranmayeh et al., 2010), although this could be attributed to IHC sensitivity issues. Intrafamilial clinical variability has also been reported, such as that found among patients from one large Kenyan kindred of Asian ancestry. Here, patients shared the same genotype (homozygous missense variant located in the G-domain of laminin-α2) but locomotion was not achieved in all cases (Geranmayeh et al., 2010).

As previously mentioned, a very small fraction of *LAMA2*-MD patients have brain structural defects which are frequently associated with intellectual disability (ID) and/or refractory seizures (Vigliano et al., 2009, Geranmayeh et al., 2010). However, there are also reports of patients, with these structural defects who, apparently have no seizures or ID. The opposite also holds true in the case of seizures (and to a lesser extent ID) since they have been reported in patients without cerebral structural changes. Based on the reassessment of data available in the LOVD and reported in the literature, no association was found between epilepsy, cognitive function or brain anomalies, and a particular set of *LAMA2* genotypes/variants. The variants found in these cases are diverse in terms of their impact, ranging from those causing PTC to missense changes, and are apparently dispersed with no obvious hotspot along the gene. Furthermore, phenotypical discrepancies have been found in patients sharing with same genotype. For example, two siblings reported by Di Blasi et al. (2001) and case #2 from Nelson et al. (2015) share the same genotype (an homozygous nonsense variant p.Arg744*), but cortical polymicrogyria and lissencephaly were only reported in the latter patient. It is conceivable that other genetic factors besides *LAMA2* variants are contributing to these phenotypes.

Over the last few years there has been a significant increase in reports of late-onset *LAMA2*-related MD patients (Gavassini et al., 2011; Rajakulendran et al., 2011; Kevelam et al., 2014; Marques et al., 2014; Nelson et al., 2015; Lokken et al., 2015; Ding et al., 2016; Harris et al, 2017; Kim et al., 2017). Most of these patients have heterozygous or homozygous missense or splice variants. Their clinical presentation is also variable but often overlapping with a childhood-onset LGMD, consisting of proximal muscle weakness and delayed motor milestones, but in all cases achieving independent ambulation. Rigid spine syndrome with joint contractures has been also reported in some patients (Nelson et al., 2015).

Additional cases of late-onset LAMA2-related MD sharing the p.Thr821Pro variant

Phenotypic variability in *LAMA2*-related MD has been clearly underestimated so far, with only a limited number of patients with this later-onset phenotype reported in the literature. As for establishing further genotype-phenotype correlations, the cases are still relatively scarce and there is a vast diversity of genetic defects and/or genotypes which makes it difficult to stratify patients into homogeneous groups.

To address some of these limitations and resorting to our large *LAMA*-related MD patient cohort, the clinical and genetic characterization of six additional patients with a late-onset phenotype from four unrelated families is reported (Table 3.5.4). They all share the same missense variant: p.Thr821Pro. In five cases the genotype was similar in that, besides this missense substitution, the second allele was a truncating variant: c.7750-1713_7899-2154del (p.Ala2584Hisfs*8) in patients P1 and P2, c.3976C>T (p.Arg1326*) in P3 and P4, and c.1854_1861dup (p.Leu621Hisfs*7) in P5. The sixth patient (P6) represents the first documented case with a homozygous p.Thr821Pro missense variant. Most of these patients were only diagnosed during adulthood, which reflects the diagnostic difficulties concerning non-MDC1A cases. All have a very mild muscle weakness (as compared with typical MDC1A) with lower limb weakness resulting in gait disturbances. In the oldest patient (P6) this weakness culminated in loss of ambulation during the sixth decade of life. In four patients brain MRI was performed (P1, P2, P3 and P6), revealing WMC like those usually found in *LAMA2*-related MD (Figure 3.5.3 A-C). These findings were pivotal for conducting *LAMA2* gene analysis in three of the cases. Patient P6, who developed dementia over the last 2 years, also had hypothalamus and pons alterations (data not shown). Five patients were subjected to a muscle biopsy. These showed myopathic or dystrophic features (Figure 3.5.3 D-F), and IHC analysis for laminin-α2 revealed apparently normal labeling (n=3, Figure 3. 5.3 G-I) or partial deficiency (n=1, data not shown).

Table 3.5.4- Additional *LAMA2*-related muscular dystrophy patients sharing the p.Thr821Pro variant.

Family Patient #	Genotype	Predicted effect on Protein	Age (gender)	Age of first symptoms	Phenotype at onset	Pattern of muscle weakness; other clinical features	Cardiac involvement?	Contractures?	Independent locomotion? (age)	Loss of ambulation? (age)	CK levels (U/l)	Muscle biopsy	Laminin-α2 IHC [1]	Brain changes (MRI)
F.I-P1	c.2461A>C + c.7750-1713_7899-2154del	p.Thr821Pro + p.Ala2584Hisfs*8	42 yrs (F)	Third decade	Migraine-like headaches for 8 mo. Complaints of limb weakness and walking difficulties	Mild generalized muscular atrophy and tetraparesis (4+/5 grade), feet dorsiflexion (4/5). Paresis of trunk and neck flexion, cannot do sit-ups or lift head while in supine position.	N	N	Y	N	n.p.	Moderate myopathic changes, with discrete endomysial fibrosis	Normal	WMC
F.I-P2			50 yrs (M)	Fourth decade	Running difficulties and progressive lower limb weakness	Proximal paresis (4/5 grade) in upper limbs, distal and proximal paresis in lower limbs. Paresis of trunk and neck flexion (grade 2 and 4+, respectively). Lordosis and myopathic gait with slight steppage.	N	N	Y	N	n.p.	Myopathic changes with fiber necrosis, also "ragged red fibers" and large number of COX negative fibers.	Normal	WMC
F.II-P3	c.2461A>C + c.3976C>T	p.Thr821Pro + p.Arg1326*	15 yrs (F)	18 Mo	Running difficulties and stairs	Proximal paresis (2/5 grade) in upper limbs and in lower limbs (3/5 grade), neck flexion (grade 2/5).	N	Y	Y (15 Mo)	N	839	Dystrophic changes	Normal	WMC
F.II-P4			11 yrs (F)	5 yrs	Facial fatigue	Proximal paresis (2/5 grade) in upper limbs and in lower limbs (4/5 grade), neck flexion (grade 1/5).	N	N	Y (14 Mo)	N	2466	n.p.	n.p.	n.p.

F.III-P5	c.2461A>C + c.1854_1861dup	p.Thr821Pro + p.Leu621Hisfs*7	33 yrs (F)	Childhood	Running difficulties Difficulty in getting out of a bed	Proximal paresis (4-3/5 grade) in lower limbs. Paresis of trunk and neck flexion (grade 3/5). Lordosis and myopathic gait.	N	N	Y	N	~2000	Dystrophic changes	Partial deficiency	n.p.
F.IV-P6	c.2461A>C (hom.)	p.Thr821Pro	71 yrs (M)	Childhood	Gait impairment	LGMD initially suspected. Proximal tetraparesis (grade 4/5, 4-/5 in lower limbs). Moderate intellectual disability (last 2 yrs).	Y [2]	N	Y	Y (last 2 years)	579	Moderate dystrophic changes associated to angulated atrophic fibers and nuclear clumps.	Normal	WMC, thalamus and pons involvement

Footnote: F- female; hom.- Homozygous; LGMD- Limb-girdle muscular dystrophy; M- male; MD- muscular dystrophy; mo- months; N- no; n.p.- not performed; U- unknown; WMC- white matter changes; Y- yes; yrs- years. [1]- Clone: Mer3/22B2 (Leica Biosystems, Newcastle upon Tyne, United Kingdom); [2]- Left ventricle hypokinesia of unknown cause, normal ejection fraction.
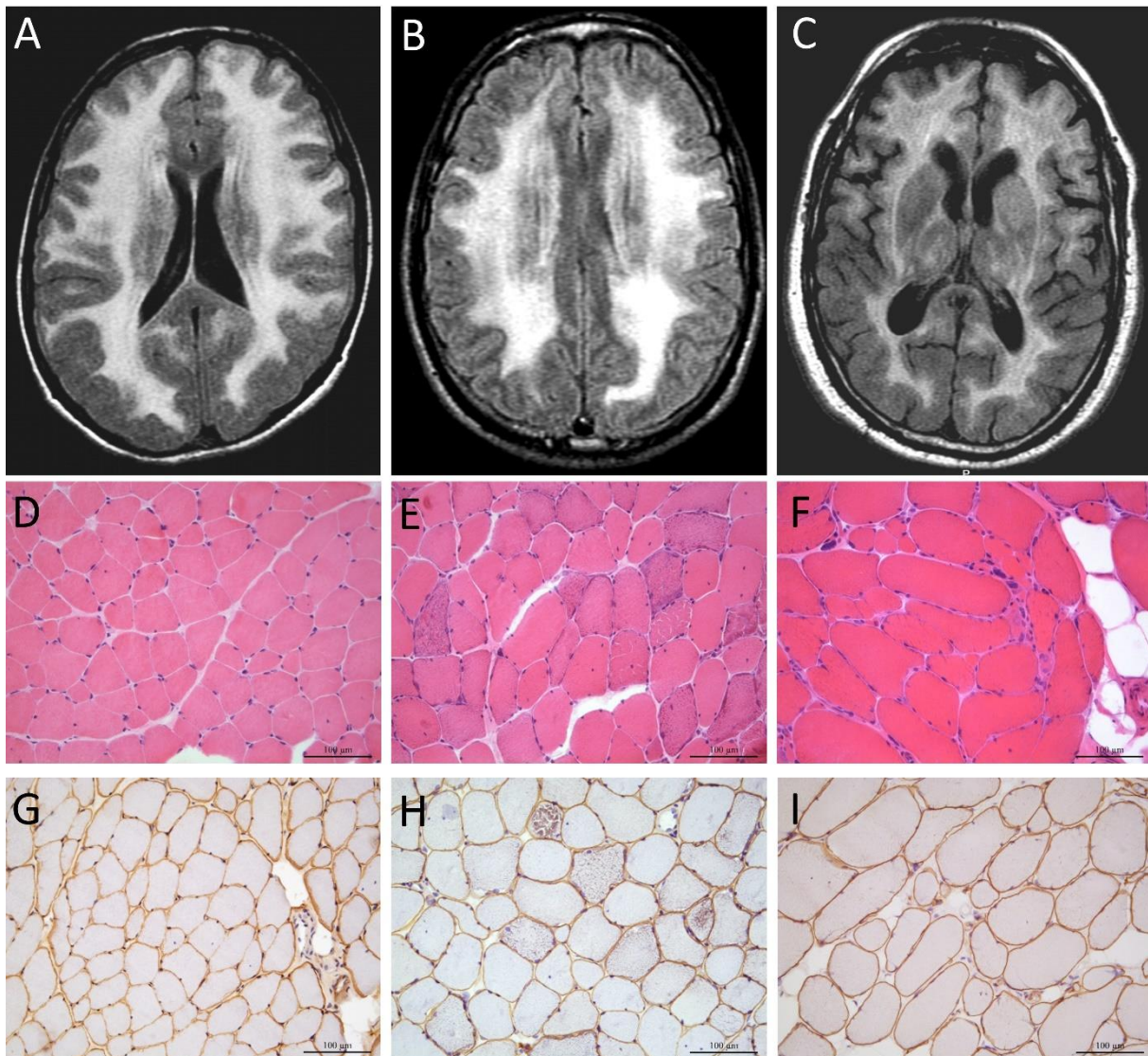
Figure 3.5.3- Brain MRI and muscle neuropathology results. Patient P1: A) brain MRI (FLAIR) shows typical white matter changes (WMC) with normal structural cerebral cortex changes; D) moderate myopathic changes with discrete endomysial fibrosis in hematoxylin eosin (H&E) staining, and G) normal immunohistochemistry (IHC) for laminin-α2. Patient P2: B) brain MRI (FLAIR) with WMC and normal cerebral cortex; E) myopathic changes with necrotic fibers and several "ragged red fibers" (H&E), and H) normal IHC for laminin-α2. Patient P6: C) WMC in brain MRI (FLAIR); F) dystrophic changes (necrotic fibers under myophagocytosis, fiber splitting and hypersegmentation, and fat substitution) and mild neurogenic features (atrophic angulated fibers and nuclear clumps), and I) normal IHC for laminin-α2.

Prevalence of p.Trp821Pro variant in a genetically uncharacterized MD patient cohort

The missense variant p.Trp821Pro is one of the most frequent genetic causes of late-onset *LAMA2*-MD in a population-specific (Portuguese) patient cohort. This missense substitution was initially identified in two patients with atypical presentations (Marques et al.,

2014), prior to the six patients described above. A further three patients, from two unrelated families with Portuguese ancestry, have also been reported by other groups: one patient from Canada (with #102482 in *LAMA2*-LOVD), and two brothers studied in France (Nelson et al., 2015). Thus far, all patients are reported to have milder muscle weakness and the majority were initially classified as possible LGMD or EDMD.

To evaluate if this missense variant could account for additional uncharacterized cases, we screened an irreversibly anonymised group of 239 myopathic Portuguese patients with clinical presentation is compatible with LGMD or EDMD. Variant screening was performed by restriction fragment length analysis (RFLA, Appendix II.4 Data II.4.4), since the c.2461A>C change creates a new restriction site for *Hpy*CH4III (Appendix II.4, Figure II.4.4). Positive samples were confirmed by Sanger sequencing.

A total of seven patients carried this missense substitution (2.9% of the cohort), three of which were homozygotes and four were heterozygotes (2% of all disease alleles). To further ascertain the genotype of the four patients carrying the c.2461A>C variant in heterozygosity, the entire coding sequence of the *LAMA2* gene was sequenced. In all patients an additional heterozygous variant was detected. Three were previously identified in other (MDC1A) patients: c.4739dup (p.Leu1581Profs*5) (Oliveira et al., 2008), c.3372dup (p.Cys1125Metfs*4) (patient #103970 in Table 3.5.1) and a missense variant c.32T>C (p.Leu11Pro) listed in ClinVar (RCV000157587.1) as being disease associated. The fourth variant, c.6707G>A, is also new and was interpreted as a VUS; it predictably gives rise to a missense change (p.Arg2236Lys) and/or may have an effect on splicing (r.spl?) (Table 3.5.2, patient #103971).

Since the c.2461A>C variant was not listed in variant population databases, its prevalence was estimated in control individuals using the afore-mentioned RFLA screening strategy. For this study, we randomly selected and irreversibly anonymised 1,100 out of a total of 11,000 samples previously analyzed in the laboratory. These were residual samples from genetic studies for diseases unrelated with neuromuscular disorders that are performed on a nationwide basis. The c.2461A>C variant was identified in one of these samples, in heterozygosity. Its allelic frequency in the general population was estimated as 0.0452% (1/2,200) which explains the relatively high prevalence of this variant among Portuguese patients with *LAMA2*-MD.

Overall, the presented data reinforces that it is diagnostically important to consider *LAMA2* gene involvement not only in CMD patients, but also as a possible cause of MD with onset beyond childhood and even in adulthood. The association between *LAMA2* and this later onset phenotype was not fully established, considering the limited number of cases reported so far. Nonetheless, it is advisable to include *LAMA2* in the list of candidate genes for MDs (LGMD or EDMD). The p.Trp821Pro missense variant constitutes an interesting genotype-

phenotype linker, as it may give rise to different phenotypes depending on the variant found in the second allele (Figure 3.5.4).
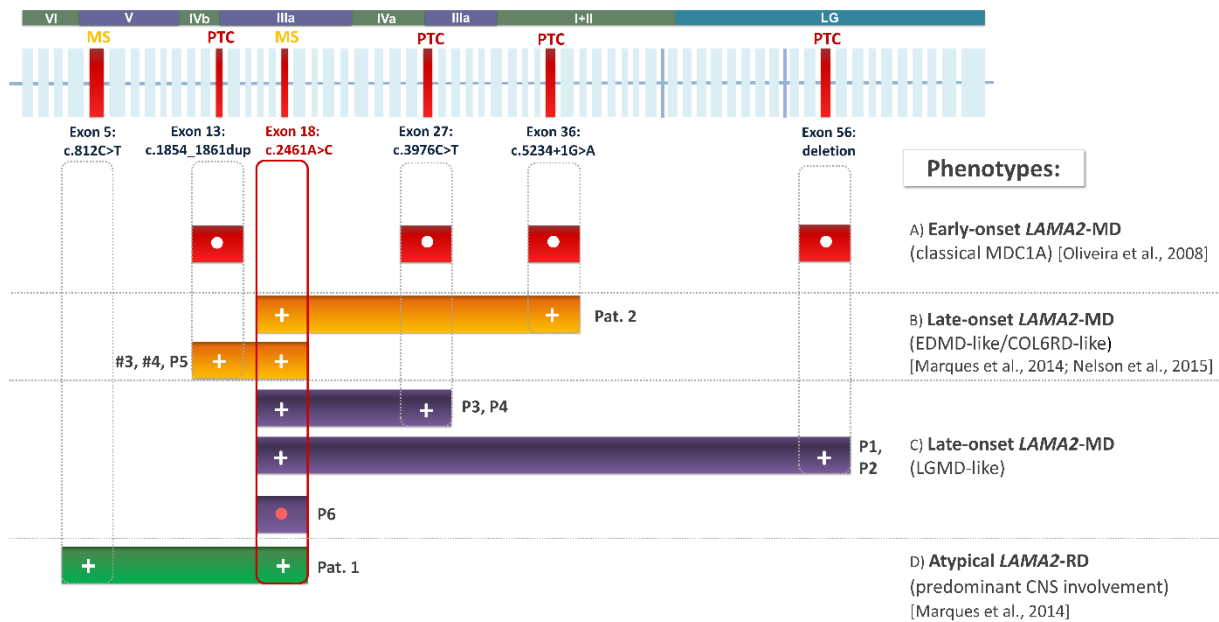
Figure 3.5.3- Different phenotypes in *LAMA2*-MD found in association with the c.2461A>C (p.Trp821Pro) variant. A) Early-onset (classical MDC1A): no independent ambulation; muscle biopsy shows dystrophic features and no labelling for laminin-α2 in IHC. Several patients reported in (Oliveira et al., 2008). B) Late-onset (EDMD/COL6-RD-like): rigid spine syndrome; cardiac involvement in some patients; walking difficulties; dystrophic features in MD, normal and irregular laminin α2 staining in IHC. Pat.2- patient 2 in (Marques et al., 2014); cases #3, #4 in (Nelson et al., 2015); P5 (this work). C) Late-onset (LGMD-like): slow progression; dystrophic features, normal and irregular laminin-α2 staining in IHC, walking difficulties later in life; P1-4, P6 (this work). D) Atypical *LAMA2*-RD: predominant CNS involvement, (occipital agyria, WMC, epilepsy); increased variability of muscle fiber diameter and irregular laminin-α2 staining in IHC. Pat.1- patient 1 in (Marques et al., 2014). Genotype-phenotype correlations suggest that the classical MDC1A presentation is explainable by variants causing premature termination codons (PTC) in both disease alleles. While late-onset *LAMA2*-MD are more likely to be associated with missense (MS) substitutions. •- Homozygous; +- Heterozygous.

3.5.7 DIAGNOSTIC RELEVANCE

Molecular defects in the *LAMA2* gene are the main genetic cause (~30%) of CMDs in most countries, except for Japan where Fukuyama-type CMD has the highest prevalence, due to a frequent founder mutation in the *FKTN* gene (Kobayashi et al., 1998). Besides the clinical examination, the clinical diagnostic workup of CMDs conventionally relies upon performing a muscle biopsy (Bönnemann et al., 2014). In addition to standard staining

methods, muscle pathology analysis includes a panel of antibodies for IHC against proteins involved in MD (laminin-α2, sarcoglycans, dystrophin and dysferlin). Three different commercial antibodies are currently available for laminin-α2 IHC studies: clone 5H2 detects the 80 kDa protein (C-terminal region), clone Mer3/22B2 detects the 300 kDa product (N-terminal region) and clone 4H8-2 clone which also recognizes the N-terminal domain. The diagnostic sensitivity of IHC is extremely high for typical MDC1A cases, where complete deficiency would be detectable regardless of the antibody used for analysis. The milder *LAMA2*-MD cases are more challenging as often only a partial deficiency is often documented. Moreover, depending on the underlying molecular defects, this IHC deficiency may not be consistent for the different antibodies (Cohn et al., 1998). N-terminal antibodies usually have higher sensitivity for cases with partial laminin-α2 deficiency, as there was a relatively intact labelling with the antibody for the 80 kDa fragment, when compared with that using the other antibodies (Cohn et al., 1998). It is therefore advisable to include at least two different antibodies against laminin-α2 in order to increase IHC sensitivity. In a small fraction of CMD patients there is also irregular labelling or partial laminin-α2 deficiency. There is some degree of genetic heterogeneity among these patients, depending on whether it is a primary or a secondary deficiency. To distinguish between these two possibilities, antibodies against glycosylated residues of α-DG and laminin-α4/5 may be effective. If changes are detected in α-DG, this would indicate a defective glycosylation pathway and involvement of other *loci*. On the other hand, normal α-DG labeling and overexpression of laminin-α4/5 (a compensatory gene expression mechanism) is suggestive of a primary laminin-α2 deficiency.

Brain MRI performed beyond the first 6 to 12 months is also an important diagnostic resource for CMDs. As previously mentioned all *LAMA2*-related MD patients have brain WMC, consisting in bilateral hyperintensity signal on T2-weighted and FLAIR MRI, in periventricular areas and subcortical cerebral hemisphere (Quijano-Roy et al., 2012). These findings alone should be an indication to perform *LAMA2* genetic testing. As demonstrated by this work and previously suggested by Gavassini and collaborators (2011), it is diagnostically relevant to perform brain MRI in uncharacterized LGMD patients. This could be performed even during adulthood, as these typical brain changes will persist throughout life. Brain MRI is especially relevant for "atypical" or mild MD cases where IHC for laminin-α2 has a lower diagnostic yield.

Considering the size and number of exons in the *LAMA2* gene, its genetic analysis has been simplified, more than a decade ago, with the introduction of automatized sequencers (fragment analyzers) for Sanger sequencing and the use of universal-tailed primers. Gene sequencing is undoubtedly the approach with the highest sensitivity for *LAMA2* analysis, detecting approximately 80% of disease associated variants. Based on the variant data collected, there is a significant frequency (~18%) of large deletions and duplications. The

genetic study should therefore be complemented with other molecular techniques such as MLPA or array-CGH.

Data available in the LSDB and population-specific cohorts can help to optimize the *LAMA2* genetic analysis. This was exemplified in a Portuguese CMD patient cohort where a 3-tier genetic test was proposed (Oliveira et al., 2014): i) sequencing a small set of selected exons where the majority of point mutations are located (based on a specific population or ethnical group variant data); ii) sequencing the remaining *LAMA2* exons; and iii) MLPA analysis or array-CGH.

The introduction of next-generation sequencing technology (NGS, or massive parallel sequencing) has remodeled genetic analysis strategies, especially in genetically heterogeneous conditions such as the MDs. Distinct NGS applications such as gene panels or whole-exome sequencing (WES) can be extremely useful to address diagnostically difficult cases. In fact, novel cases with milder *LAMA2*-related phenotypes recently reported in the literature have been solved resorting to NGS (Ding et al., 2016; Dean et al., 2017; Harris et al, 2017; Kim et al., 2017).

The impact of NGS technology is also reflected in five patients described in this work: (ID#s: 102662, 132012, 132013, 132015, 132025 in Table 3.5.1) whose disease associated variants listed were identified by NGS gene panels. One further patient (ID# 103207 in Table 3.5.1) demonstrates the utility of NGS to address genetic and clinical heterogeneity. This is a patient with a LGMD phenotype and ID, who has remained without genetic characterization for several years. The patient, currently 14 years of age, had delayed motor development (started walking at 31 months of age), lumbar lordosis, and elevated CK levels (~1300 U/l). Muscle biopsy performed at 6 years of age (in another clinical center where she was initially followed) revealed dystrophic features and normal IHC results for dystrophin and sarcoglycans. Genetic analysis of *FKRP*, *CAPN3*, *LMNA* and *DMPK* genes were negative. The patient was studied by WES as previously reported in a similar research (Oliveira et al., 2017). As a first approach WES data analysis was restricted to a set of genes known to be associated with muscle diseases (Appendix II.4 Data II.4.5). Within the list of filtered-in variants, two heterozygous variants were identified in *LAMA2* (Appendix II.4, Figure II.4.5). The first was the c.1854_1861dup variant, previously reported as disease associated in several MDC1A patients, and the second was a novel splicing variant c.819+2T>C located in the donor splice-site of intron 5 (Table 3.5.1). To further characterize the effect of this splice-site variant, a muscle fragment available from patient ID# 102735 (shares the same variant) was used (Appendix II.4 Data II.4.2). *LAMA2* transcript analysis by RT-PCR showed the presence of multiple aberrant products that, upon sequencing, were attributed to multiple skipping events involving exons 5 to 7 (Table 3.5.1, Appendix II.4 Figures II.4.1 and II.4.2). Study of the patient's parents confirmed compound heterozygosity, as each progenitor

carried a different *LAMA2* variant. Brain MRI performed after WES analysis, revealed WMC but not configuring the typical pattern found in MDC1A cases. In this patient, axial T2 and FLAIR revealed small focal white matter hyperintensities in the subcortical part of brain, more specifically in the frontal, temporal-anterior, parahyppocampus and insula regions with sparing of the internal capsule and corpus callosum (data not shown).

Finally, as demonstrated by five cases listed in Table 3.5.1 (ID#s 102324, 102369, 102378, 132014 and 132015), a small percentage of patients were found to have only one heterozygous *LAMA2* disease-causing variant. This could be attributed to deeply placed intronic variants affecting splicing or variants located in the gene's promotor region, both of which are not covered by conventional sequencing, gene panels or even WES. Here, a more comprehensive NGS approach such as whole-genome sequencing and/or RNA sequencing (RNA-seq) may ultimately provide a final answer to such cases with incomplete genotyping.

3.5.8 FUTURE PROSPECTS

Although there is an increasing recognition of the involvement of *LAMA2* disease-associated variants in the genetic etiology of muscular dystrophies, the incidence is probably still underestimated. To improve the diagnosis of these cases, it is necessary to include both brain MRI and to evaluate the expression of laminin-α2 in muscle by IHC. These two approaches are often not considered in the clinical workup of patients with (non-congenital) myopathies. NGS can also contribute towards the identification of further cases. However, the interpretation of variants from such studies often leads to their classification as VUS which considerably limits the clinical utility of these genetic data.

As for further research, it is necessary not only to continue to document clinical data and *LAMA2* variants to obtain further genotype-phenotype correlations, but also to develop strategies for functional analysis and validation of new variants, especially those predictably of the missense type. This task may be complex, as variants might affect several key aspects of the laminin-211 life-cycle: i) post-translational modification, ii) protein translocation and secretion process, iii) interaction with membrane-specific receptors, and iv) variety of molecular partners in the BM, possibly some yet to be identified. One strategy for functional analysis would imply obtaining a biological sample from the patient by an invasive procedure (e.g. muscle or skin biopsy), expanding cells through *in vitro* culture, and performing protein-protein interaction studies, such as pull-down assays using a battery of different bait-proteins known to interact with laminin-α2. Failure to detect a particular interaction would indicate a deleterious effect. To enable such studies, further research should primarily focus in a comprehensive search for domain-specific interactions, which could be accomplished by

high-throughput proteomics analyses. An assay for those variants specifically affecting domains involved in laminin polymerization has been reported (Cheng et al., 1997; Hussain et al., 2011). Basically, a mixture of wild-type with the mutated form of laminin would show a failure in establishing normal polymerization levels. Here, the limiting step would be generating and purifying sufficient amounts of proteins to conduct these *in vitro* studies.

As laminin-α2 is not confined to muscle or brain cells, in a transgenic mouse model with deficient laminin-α2 it was shown that the loss of this protein caused disruption of the apical ectoplasmic specialization-blood-testis barrier, and leading to male infertility (Häger et al., 2005). The laminin-α2 in testis was further implicated in the regulation of an axis that functionally links the BM to the blood-testis barrier of Sertoli cells (Gao et al., 2017). Considering that human infertility has not been linked to laminin-α2, it would relevant to evaluate male reproductive issues in late-onset *LAMA2*-MD patients.

One of the most important aspects concerning *LAMA2*-MD is the development of a suitable treatment for this condition. Several approaches have been proposed, developed and tested in laminin-α2 deficient mice and zebrafish models (reviewed by Wood et al., 2014; Durbeej et al., 2015). One particularly effective approach targets extracellular matrix modulation as way to ameliorate MDC1A. Here, strategies aim to improve muscle viability, through the augmentation of residual functionality within the cellular system, such as up-regulation of other laminins (α4 or α1) and integrin-α7 (Wood et al., 2014). However, with laminin-411 there are some limitations for BM repair, since this laminin only forms a trimeric structure, lacking capacity to further self-polymerize into superstructures such as those derived from laminins-211 or -111. Overall there are some hurdles towards its applicability, namely the large size of laminins which make its delivery to target locations extremely challenging. An effective way to address this problem is to use shorter engineered proteins, such as the chimeric laminin/nidogen protein or mini-agrin, shown to be effective in a *LAMA2*-MD mouse model (dy^W^/dy^W^) (McKee et al., 2017; Reinhard et al., 2017). Probably in a near future, we will witness a new generation of laminin-binding proteins that, depending on the underlying genetic defects, are able to replace defective domains of laminin and promote the assembly of a stable and fully functional BM.

Chapter 4

Contents

4.1 New massive parallel sequencing approach improves the genetic characterization of congenital myopathies

4.2 New splicing mutation in the choline kinase beta (*CHKB*) gene causing a muscular dystrophy detected by whole-exome sequencing

4.3 The new neuromuscular disease related with defects in the ASC-1 complex: report of a second case confirms *ASCC1* involvement

4.4 Exonization of an intronic LINE-1 element causing Becker Muscular Dystrophy as a novel mutational mechanism in Dystrophin gene

4.5 Homozygosity mapping using Whole-Exome Sequencing: A Valuable Approach for Pathogenic Variant Identification in Genetic Diseases

4.1 New massive parallel sequencing approach improves the genetic characterization of congenital myopathies

Oliveira, J., Gonçalves, A., Taipa, R., Melo-Pires, M., Oliveira, M. E., Costa, J. L., Machado, J. C., Medeiros, E., Coelho, T., Santos, M., Santos, R., Sousa, M. (2016). Journal of Human Genetics. 61(6):497-505.

4.1.1 Abstract

Congenital myopathies (CM) are a heterogeneous group of muscle diseases characterized by hypotonia, delayed motor skills and muscle weakness with onset during the first years of life. The diagnostic workup of CM is highly dependent on the interpretation of the muscle histology, where typical pathognomonic findings are suggestive of a CM but are not necessarily gene-specific. Over 20 loci have been linked to these myopathies including three exceptionally large genes (*TTN*, *NEB* and *RYR1*) which are a challenge for molecular diagnosis. We developed a new approach using massive parallel sequencing technology (MPS) to simultaneously analyze 20 genes linked to CM. Assay design was based on the Ion AmpliSeq strategy and sequencing runs were performed on an Ion PGM system. A total of 12 patients were analyzed in this study. Among the 2534 variants detected, 14 pathogenic mutations were successfully identified, in the *DNM2*, *NEB*, *RYR1*, *SEPN1* and *TTN* genes. Most of these had not been documented and/or fully characterized, hereby contributing to expand the CM mutational spectrum. The utility of this approach was demonstrated by the identification of mutations in 70% of the patients included in this study, which is relevant for CM especially considering its wide phenotypic and genetic heterogeneity.

4.1.2 Introduction

Congenital myopathies (CM) are a highly heterogeneous and continuously expanding, group of muscle diseases with an estimated incidence of around 6 per 100 000 live births (Jungbluth et al., 2014). On the severest end of the disease spectrum, CM are characterized by muscle weakness and hypotonia with neonatal onset or during the first years of life, which often give rise to respiratory and feeding difficulties (Oliveira et al., 2013; Jungbluth et al., 2014). Typical features may include weakness of facial and bulbar muscles, and also involvement of extra-ocular muscles (ophthalmoparesis or ophthalmoplegia) (North et al., 2014). Typical forms of CM may also present during childhood as motor development delay and/or waddling gait (Jungbluth et al., 2014). At the other end of the scale, it is now clear that some patients with subtle congenital muscle weakness might remain undiagnosed until adulthood, when muscle strength deteriorates or respiratory insufficiency settles in (Jungbluth et al., 2014; Snoeck et al., 2015).

The classification and the diagnostic workup of CM is highly reliant upon the identification of distinct structural abnormalities detected in muscle biopsies (Romero et al., 2013; North et al., 2014). Based on pathognomonic findings CM are subdivided into four major groups: i) CM with rods in nemaline myopathy (NM), ii) CM with cores, that includes central core disease (CCD), multiminicore disease (MmD) and multicore myopathy, iii) CM with central nuclei (centronuclear myopathy, CNM) and iv) congenital fiber-type size disproportion (CFTD) (Romero et al., 2013). Other less common subtypes such cap or core-rod myopathies, which might represent overlapping myopathological entities, have been also reported (Romero et al., 2013; Piteau et al., 2014). Although, these findings are of extreme diagnostic value, in some patients the muscle biopsy may display only unspecific myopathic features such as type 1 fiber predominance/uniformity (Rocha et al., 2014).

More than 20 different genes are currently known to be associated with CM (http://musclegenetable.fr/, August 2015). The majority of these loci encode basic structural components of the sarcomere or proteins involved in calcium homeostasis, both crucial for normal muscle structure and function (Jungbluth et al., 2014). Other genetic defects give rise to abnormal triad structure due to aberrant tubulogenesis and/or abnormal membrane recycling (Dowling et al., 2014). Among these loci there are three particularly large genes: Titin (*TTN*) with 363 exons, Nebulin (*NEB*) with 183 and the Ryanodine receptor 1 (*RYR1*) with 106. Consequently, conventional Sanger sequencing of these genes is extremely laborious and costly. These aspects might explain the lack of thorough studies for large genes such as *TTN* and the limited number of patients reported in the literature (Chaveau et al., 2014). The genetic study workup is complex considering that the same clinical entity can be caused by mutations in different genes, as may the same defective gene give rise to distinct myopathies. *RYR1-*

related myopathies are among the best examples. Nearly all subtypes of CM (CCD, CNM, CFTD and core-rod myopathy) have been reported as linked to *RYR1* mutations (North et al., 2014; Snoeck et al., 2015). In addition, although muscle histology is important for the diagnostic workup of CM, pathognomonic findings (such as rods in nemaline myopathy) are not necessarily gene-specific (Wallgreen-Pettersson et al., 2011). The considerable genetic and clinical heterogeneity is reflected by the large number of genetic studies reported to be inconclusive. Essentially owing to these challenging diagnostic and technical difficulties, a significant proportion of CM patients are still genetically unsolved (Oliveira et al., 2013).

With the advent of the novel massive parallel sequencing (MPS) technologies, we have developed a new targeted resequencing approach, which allows the simultaneous analysis of 20 genes implicated in CM. Besides contributing to expand the mutational spectrum of CM, our study enabled the identification of mutations in seven CM patients (out of ten that were prospectively studied) thereby demonstrating its clinical utility.

4.1.3 MATERIAL AND METHODS

Patients

A total of twelve patients were included in this study. Two patients with mutations previously identified by conventional sequencing were used for assay validation. The first patient used as a positive control (C1), was diagnosed with a centronuclear myopathy due to a heterozygous mutation (c.1393C>T, p.Arg465Trp) in the *DNM2* gene. The second patient (C2) has a mild core myopathy with adulthood onset, caused by two heterozygous *RYR1* missense mutations (patient 1 in Duarte et al., 2011).

For the prospective part of this study, ten genetically uncharacterized patients (P1-P10) were selected based on features compatible with CM, namely: i) early disease onset (neonatal or up to early childhood), and ii) muscle histology suggestive of CM or with structural defects. Eight of these patients had structural changes on muscle biopsy: three with central nuclei (P1B, P2 and P8), three with rods (P3, P5 and P7); and two with minicores (P4 and P6). The remaining two patients (P9 and P10) had unspecific myopathic changes seen on muscle biopsy, mainly type 1 fiber predominance. Parents and other family members, when available, were studied to demonstrate compound heterozygosity in autosomal recessive cases and/or to ascertain the parental origin of the mutations. This research work was approved by the ethics committee of Centro Hospitalar do Porto.

Muscle histology analysis

The routine diagnostic workup consisted of an open biopsy (deltoid, quadriceps or gastrocnemius muscle) and histological [hematoxylin and eosin stain (H&E), periodic acid–

Schiff (PAS) and gomori thrichrome (GT)] and histochemical [reduced nicotinamide adenine dinucleotide (NADH), succinic dehydrogenase (SDH), cytochrome oxidase (COX) and adenosine triphosphatase (ATPase)] studies performed on frozen material. Semi-thin sections at 1μm from resin embedded muscle were also routinely used.

AmpliSeq assay design

The first task consisted in the development of a targeted resequencing approach based on semiconductor technology to simultaneous analyze 20 genes linked to CM, namely: *ACTA1, BIN1, CFL2, CNTN1, DNM2, KBTBD13, KLHL40, MEGF10, MTM1, MYBPC3, MYH2, MYH7, NEB, RYR1, SEPN1, TNNT1, TPM2, TPM3, TRIM32 and TTN* (Appendix III.1 Figure III.1.1). Assay design was performed using the Ion AmpliSeq Designer software (Life Technologies; Foster City, California, USA) pipeline 2.2.1. Based on a list of representative transcripts for these loci (Appendix III.1 Table III.1.1), target regions were selected from the University of California Santa Cruz, California, USA (UCSC) Genome Browser (http://genome.ucsc.edu/index.html). Targets included exonic regions (with 50 bp into flanking introns) and untranslated regions (UTR). The obtained custom Ampliseq assay covered 92.6% of these regions comprising a total of ~320 thousand base pairs (bp) and consisted of 2077 different amplicons (ranging from 64 to 239 bp, 174 bp on average), divided into two independent primer pools for multiplex polymerase chain reaction (PCR).

Library and template preparation

Sample quality of patient genomic DNA (gDNA) was evaluated by gel electrophoresis and quantified using Qubit dsDNA HS Assay Kit (Life Technologies). A total of 10 ng of gDNA were used in each multiplex PCR reaction. Library amplifications, digestion, bar-coding and purification was performed according to the Ion AmpliSeq Library Kit version 2.0 (Life Technologies) instructions. Library quantification was performed using the Qubit dsDNA HS Assay. All libraries were diluted to the same concentration and pooled to ensure an equal representation of the different samples.

The diluted and combined libraries were subjected to amplification by emulsion PCR using Ion Personal Genome Machine (PGM) Template OT 2 200 Kit (Life Technologies) and prepared on an Ion OneTouch 2 Instrument (Life Technologies) according to the manufacture's protocol. Enrichment of template Ion Sphere particles was performed using the Ion OneTouch 2 enrichment system (Life Technologies).

Semiconductor sequencing

Sequencing was carried out on a PGM system platform based on semiconductor technology.13 Two independent experiments were performed, using the Ion 316 and 318 chips, allowing 4 or 8 bar-coded samples respectively. The Ion Sequencing Kit v2.0 (Life Technologies) was used to perform sequencing runs, following the manufacture's recommended protocols. Torrent Server version 4.0.2 was used to obtain the basic run metrics and to generate Binary Alignment Map (BAM) and Variant Caller Format (VCF) files. Torrent Variant Caller was set for germline - low stringency - calls.

Data analysis and interpretation

Variant annotation and filtering steps were performed using Ion Reporter v4.2 software (http://ionreporter.lifetechnologies.com/) (Life Technologies). VCF files from each patient were added to Ion Reporter and were annotated using the "annotate variants single sample" workflow incorporated in this software. A filter chain was created to restrict the number of variants, which consisted in the removal of single nucleotide polymorphisms (SNP) present on the UCSC common SNPs list (those with at least 1% minor allele frequency). Alamut Visual v2.4 software (Interactive Biosoftware, Rouen, France) was used to assist variant analysis and interpretation. The software also incorporates algorithms to evaluate the impact of missense mutations and the effect on splicing, and allows the simultaneous visualization of VCF and BAM files. GenomeBrowse v2.0.0 (Golden Helix Inc., Bozeman, MT) was used for visual inspection of BAM files.

Variant validation and expression analysis

Clinically relevant variants (classified as pathogenic) were confirmed by a new PCR using in-house designed primers (sequences in Appendix III.1 Table III.1.2) and the resulting amplicons sequenced by the Sanger method.

Variants possibly affecting splicing were evaluated by expression studies at the mRNA level. Total RNA was obtained from cryopreserved muscle samples (P4 and P7) using the PerfectPure RNA Fibrous Tissue kit (5 PRIME, Hilden, Germany) or from whole blood (patient P3) extracted with the PerfectPure RNA Blood Kit (5 PRIME). After conversion to cDNA with the High Capacity cDNA Reverse Transcription kit (Life Technologies) these samples were amplified by PCR using specific primers annealing to different regions of *TTN* and *NEB* transcripts (primers sequences in Appendix III.1 Table III.1.2). PCR products were purified with Illustra ExoStar (GE Healthcare, Little Chalfont, UK). A new PCR was prepared with BigDye Terminator kit V3.1 chemistry (Life Technologies). Sequencing reactions were run on an ABI 3130xl genetic analyser (Life Technologies).

4.1.4 RESULTS

Development of a new sequencing approach for CM

The initial task was the development of a new gene panel based on MPS, designed using the Ion AmpliSeq software. An overall coverage of 95% was obtained for the genomic coding sequence (CDS) which was considered suitable for this research. The CDS coverage of each locus ranged between 77 and 100% with the majority of the genes (n=15) having a coverage above 90% (Appendix III.1 Table III.1.1). For the remaining loci, all except the *NEB* gene could be explained by the presence of high GC content regions, having higher impact in smaller genes. These regions are amenable to filling in by Sanger sequencing. In the case of *NEB* the lower coverage (85%) is explained by the existence of a triplicated region (three almost identical sequence blocks) encompassing exons 82-89, 90-97 and 98-105. Since the majority of MPS approaches are based on the alignment of short sequences, the obtained reads and variants will not be mapped with precision.

The sequencing runs performed on the total of 12 samples generated 6.5 million sequence reads, with an average coverage depth of 257x (Appendix III.1 Table III.1.3) and with 97.0% of the target regions successfully covered. A total of 2535 sequence variants were called: 2348 single nucleotide variants (SNV) and 187 insertion/deletions (INDEL), corresponding to an average of 211.4 variants per individual (Appendix III.1 Table III.1.3).

The high number of variants obtained in this work required the development of a computational filtering strategy to facilitate the analytic process and to concentrate on the variants that most likely correlated with the patients' phenotypes. SNV listed as common SNPs in the UCSC database were excluded from further analysis. As a result, 86% of the detected variants were filtered-out, meaning that an average of ~29 variants per patient remained to be analyzed in detail. Data interpretation implied evaluating: i) the variant predictable impact, ii) the frequency in publically available genetic variant databases [dbSNP and exome variant server (EVS)] and iii) the data available in locus specific databases (Leiden Muscular Dystrophy Pages).

A total of 14 pathogenic variants listed in Table 4.1.1 were detected in 9 patients whose clinical data is shown in Table 4.1.2. The three missense mutations previously identified by conventional sequencing in the two positive controls (C1 and C2) were successfully detected. The remaining 11 mutations were identified in 7 patients of the prospective study; these were located in *RYR1* (n=4), *NEB* (n=4) and *TTN* (n=3). One additional heterozygous mutation in the selenoprotein N1 gene (*SEPN1*) was also identified. In this case, and so as to exclude the presence of a second pathogenic variant, this gene was completely resequenced by the Sanger method. In one of the patients (P2) with a novel heterozygous nonsense mutation in

RYR1, gene analysis was complemented by conventional sequencing, leading to the identification of a second heterozygous mutation located in a region not covered by the assay (part of exon 101).

In this work, a total of 67 variants (the majority intronic or located in UTRs) were found to be unclassified variants (UV); these included both ultra-rare sequence changes (reported in genetic databases with a frequency bellow 0.1%) and novel variants. Finally, 43 known polymorphisms were also identified, reported in the EVS database with frequencies higher than 0.1% and thus not listed in the USCS common SNP table.

Difficulties and potential pitfalls

During the validation of the MPS gene panel for CM, the zygosity of one of the mutations (chr19:g.38986918C>G, c.6612C>G) in positive control C2 was not correctly established. After reviewing Sanger sequencing data and the MPS alignment, a rare SNP (c.6549-51C>T) was identified in the intronic region where the respective MPS forward primer anneals. This variant prevented the amplification of the normal allele and caused a PCR allele dropout, thereby explaining the apparent genotype discrepancy (Appendix III.1 Figure III.1.2).

It is noteworthy that during the optimization of the analytical process, an intronic mutation (chr2:g.152417626_152417632del) was initially filtered-out. The reference SNP (rs) cluster identification of a known variant was attributed to this 7 bp deletion during the annotation performed by Ion Reporter (Appendix III.1 Figure III.1.3). By default the Ion Reporter algorithm associates variants to the rs number based on the genomic coordinates involved rather than the specific variant itself. As a consequence, if a deletion or duplication coincides with the genomic position of a known SNP, the software attributes the same rs number. Considering this potential pitfall, all filtered-out INDEL variants were manually rechecked. About 9% of the total numbers of variants called were classified as false positives (FP). These FP variants resulted from: i) sequencing artifacts in homopolymeric regions (consistently found in several samples tested), ii) variants in shorter reads due to mis-priming events (similar to those reported by McCall et al., 2014) and usually showing a biased proportion of mutated/normal reads) and to a lesser extent iii) variants located in regions with low coverage depth.

Table 4.1.1- Pathogenic sequence variants detected by the massive parallel sequencing gene panel.

Patient Id./ Gender	Gene / Accession Number	Gene location (exon/ intron)	DNA change Genomic cDNA [zygosity]	RNA change	Protein	Mutation type	Mutation origin	Family history	Frequency in ESP	Mutation reference
C1 M	*DNM2* NM_00100536 0.2	11	Chr19:g.10909219C>T c.1393C>T [heterozygous]	r.(?)	p.Arg465Trp	Missense	*de novo*? (unaffected parents were not tested)	Sporadic	n.p.	Bitoun et al., 2007
C2 F	*RYR1* NM_000540.2	40	Chr19:g.38986918C>G c.6612C>G [heterozygous] (*)	r.(?)	p.His2204Gln	Missense	Inherited (parent #1)	Affected brother (same genotype)	n.p.	Duarte et al., 2011
		98	Chr19:g.39068613G>A c.14228G>A [heterozygous]	r.(?)	p.Gly4743Asp	Missense	Inherited (parent #2)		n.p.	Duarte et al., 2011
P1 F	*RYR1* NM_000540.2	87	Chr19:g.39034513C>T NM_000540.2:c.12010C>T [heterozygous]	r.(?)	p.Gln4004*	Nonsense	Inherited (maternal)	Affected brother (P1B, with the (same genotype)	n.p.	New
		101	c.14643G>A (**) [heterozygous]	r.(?)	p.Met4881Ile	Missense	Inherited (paternal)		n.p.	New
P2 F	*RYR1* NM_000540.2	61	Chr19:g.39002235C>T c.9157C>T [heterozygous]	r.(?)	p.Arg3053*	Nonsense	Inherited? (paternal inferred)	Sporadic	n.p.	New
		28	Chr19:g.38964051C>G c.3800C>G [heterozygous]	r.(?)	p.Pro1267Arg	Missense	Inherited (maternal)		1/13006	New
P3 M	*NEB* NM_00127120 8.1	129	Chr2:g.152408252C>T c.19944G>A [homozygous]	r.[19944g>a; 19944_19445ins199 44+1_19944+120, =]	p.Asn6649_Ile 8560delins27	Affects splicing	Inherited? (untested parents)	Affected brother (not studied)	n.p.	Lehtokari et al., 2014/ new (splicing effect)

P4 M	TTN NM_001267550.2	Intron 273	Chr2:g.179473930C>T c.52102+5G>A [heterozygous]	r.52034_52102del	p.Gly17345_Leu17367del	Affects splicing	Inherited (maternal)	Affected sib (same genotype, deceased in the neonatal period)	n.p.	New
		326	Chr2:g.179434749_179434750del c.76109_76110del [heterozygous]	r.(?)	p.Ile25370Argfs*6	Frameshift deletion	Inherited (paternal)		n.p.	New
P5 M	NEB NM_001271208.1	171	Chr2:g.152354789_152354792dup c.24294_24297dup [homozygous]	r.(?)	p.Glu8100Serfs*5	Frameshift duplication	Inherited? (consanguineous parents)	Affected sister (not studied, deceased)	n.p.	Pelin et al., 2002; Lehtokari et al., 2006; Lehtokari et al., 2014
P6 M	TTN NM_001267550.2	304	Chr2:g.179454772_179454773insGG c.61679_61680insCC [heterozygous]	r.(?)	p.Ser20561Leufs*17	Frameshift insertion	de novo	Sporadic	n.p.	New
P7 M	NEB NM_001271208.1	Intron 122	Chr2:g.152417626_152417632del c.19102-4_19102-10del [heterozygous]	r.[19102_19206del, 19101_19102ins19101+1_19102-1, =]	p.[Val6368_Ser6402del, Val6368_Ile8560delins25]	Affects splicing	Inherited? (maternal inferred)	Affected relative (distant cousin) (shares p.Arg7026* mutation)	n.p.	Lehtokari et al., 2014/ new (splicing effect)
		140	Chr2:g.152394412G>A c.21076C>T [heterozygous]	r.(?)	p.Arg7026*	Nonsense	Inherited (paternal inferred)		n.p.	Lehtokari et al., 2014
	SEPN1 NM_020451.2	10	Chr1:g.26139280T>G c.1384T>G [heterozygous]	r.(?)	p.Scy462Gly	Selenocystein codon (TGA) loss	?		n.p.	Ferreiro et al., 2002

Footnote: C- control; ESP- NHLBI GO Exome Sequencing Project, data accessed through the exome variant server (http://evs.gs.washington.edu/EVS/); Id- identification; n.p.- variant not present in the database; P- patient; (*)- zygosity incorrectly determined due to allele dropout (see results section); (**)- mutation not detected by gene panel, region not included in the assay.

Table 4.1.2- Clinical data of the patients presented in this work.

Patient / Gender	Age (years)	Onset	Hypotony	Muscle weakness	CK (IU/L)	Ambulant?	Facial involvement?	Cardiac involvement?	Respiratory insufficiency?	Scoliosis?
C1 M	20 yr	early childhood	N	Ophthalmoparesis; tetraparesis; weakness predominantly in lower limbs.	47	Y	Y	N	N	N
C2 F (*)	43 yr (**)	adulthood	N	Only myalgia.	456	Y	N	na	N	na
P1 F	28 yr	neonatal	Y	Ophthalmoparesis; bulbar weakness; proximal tetraparesis; axial muscle weakness.	np	N (never walked)	Y	N	Y (refused NIV)	Y (corrected by surgery 13 yr)
P1B M	34 yr	neonatal	Y	Ophthalmoparesis; bulbar weakness; proximal tetraparesis; axial muscle weakness.	18	N (never walked)	Y	N	Y (refused NIV)	Y (corrected by surgery 15 yr)
P2 F	23 yr	neonatal	Y	Ophthalmoparesis; bulbar weakness; proximal tetraparesis; axial muscle weakness.	5	Y (3.5 yr)	Y	N	Y (NIV since 16 yr of age)	Y (corrected by surgery 12 yr)
P3 M	75 yr	early childhood	N	Severe and early bilateral foot drop; later with milder facial, cervical and proximal upper limb weakness.	Norm.	Y	Y	N	N	N

Table 4.1.2- Clinical data of the patients presented in this work (continues).

Patient / Gender	Age (years)	Onset	Hypotony	Muscle weakness	CK (IU/L)	Ambulant?	Facial involvement?	Cardiac involvement?	Respiratory insufficiency?	Scoliosis?
P4 M	Deceased (at 10 yr of age)	congenital (arthrogryposis)	Y	Facial and bulbar weakness; severe tetraparesis; axial muscle weakness.	Norm.	N (never walked)	Y	Y (dilated cardiomyopathy)	Y (IV during neonatal period)	Y (not corrected)
P5 M	16 yr	infancy	N	Early respiratory muscle involvement; facial and bulbar weakness; proximal tetraparesis and later distal involvement; axial muscle weakness.	137	Y	Y	N	Y (NIV since 4 yr of age)	Y (corrected by surgery 15 yr)
P6 M	17 yr	infancy	N	Facial weakness; proximal tetraparesis.	54	Y	Y	N	N	N
P7 M	12 yr	congenital (bilateral club foot)	Y	Facial diparesis and bulbar weakness; tetraparesis predominantly distal.	np	Y (2 yr)	Y	N	Y (NIV since 8 yr of age)	N

Footnote: M- male; F- female; Y- yes; N- no; NIV- non-invasive ventilation; IV- invasive ventilation; yr-years; na- not assessed; norm.- within normal range; np- not performed; (*)- reported in Duarte et al., 2011; (**)- age at the time of publication.

RYR1-related myopathies

Three patients (P1, P1B and P2) from two families were identified has having an autosomal recessive congenital myopathy due to defects in the ryanodine receptor 1 gene (Figure 4.4.1). All three patients had marked hypotonia and respiratory distress during the neonatal period but there was some degree of clinical variability. The two brothers from the first family are both wheelchair dependent (P1 never walked) and have a severe tetraparesis, whereas patient P2 achieved independent locomotion at 3.5 years of age and currently (at 28 years of age) walks without support. The patients' muscle biopsies showed considerable fiber size variability and frequent fibers with abnormal centrally placed nuclei - pathognomonic features for a central nuclear myopathy (Figure 4.1.1A).
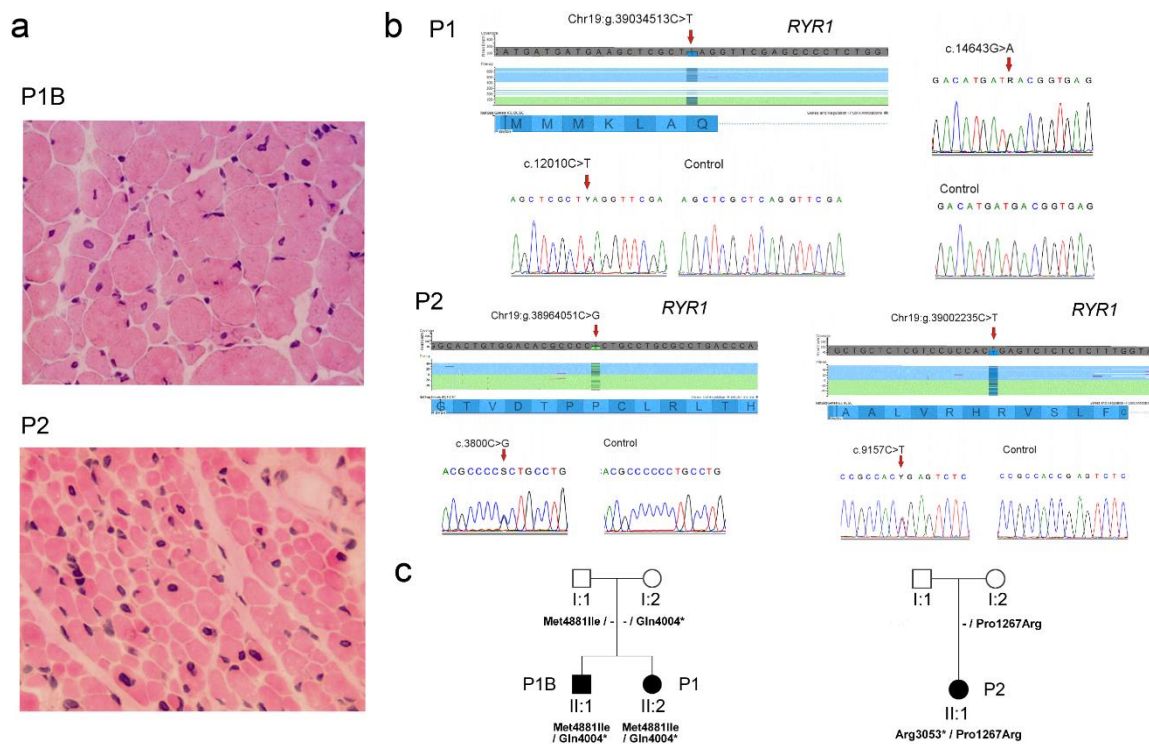
Figure 4.1.1- *RYR1*-related centronuclear myopathy. (A) Patients P1B and P2 histology of deltoid muscle biopsies (H&E with 400x magnification). In both there is abnormal fiber size variability, with atrophic fibers and central nuclei. In patient P1B the majority of atrophic and hypotrophic fibers were type 1 fibers resembling congenital fiber size type disproportion (not shown). (B) Heterozygous *RYR1* pathogenic variants identified in patient P1: i) nonsense mutation c.12010C>T (p.Gln4004*) was identified by MPS (top part) and confirmed by Sanger sequencing (bottom); ii) missense mutation c.14643G>A p.(Met4881Ile). Mutations identified in patient P2: i) nonsense mutation c.9157C>T (p.Arg3053*) and ii) missense mutation c.3800C>G (p.Pro1267Arg). (C) Patients' family pedigrees suggesting an autosomal inheritance pattern.

Four novel mutations were identified in *RYR1* (Figure 4.1.1B); these comprised two nonsense mutations (p.Gln4004* and p.Arg3053*) and two missense mutations (p.Met4881Ile and p.Pro1267Arg) affecting highly conserved residues. Bioinformatic analysis resorting to two distinct algorithms corroborated the pathogenic impact of these missense variants. The genotype was concordant between all three cases, since there is compound heterozygosity of a missense and a nonsense mutation (Figure 4.1.1C).

NEB-related myopathies

Patients P3, P5 and P7 were seen to have rods in their muscle biopsy, evocative of a nemaline myopathy (Figure 4.1.2A). Interestingly, in the first biopsy of P5 there were features compatible with congenital fiber type size disproportion, with a focal area showing structures resembling rods. A second biopsy performed 2 years later showed larger areas and number of the typical rods. In terms of their clinical presentation, phenotypes were rather dissimilar, varying from a very mild distal myopathy diagnosed during adulthood to a congenital form of nemaline myopathy (NM) with bilateral club foot and requiring non-invasive ventilation (NIV). The first patient (P3) is a 75 year old man with a slowly progressive distal myopathy. His motor and cognitive development was described as normal. After his sixties there were complaints of proximal upper limb and cervical weakness. There is no respiratory or cardiac involvement. All the variants identified in the *NEB* gene of this patient were detected in homozygosity including one mutation (c.19944G>A) located in the last base of exon 129 (Figure 4.1.2B). Since no known parental consanguinity was reported, the possibility of a large deletion spanning the entire *NEB* gene was excluded by the analysis of the amplicon's normalized coverage depth (Appendix III.1 Figure III.1.4). The c.19944G>A mutation was previously reported as pathogenic and possibly affecting splicing (Lehtokari et al., 2014), which was corroborated by our bioinformatic analysis. However, until now, its effect had not been experimentally proven. This mutation causes the disruption of the canonical splice-site (naturally-occurring), and a cryptic splice-site located within intron 129 is used instead (Figure 4.1.2C). The outcome of this splice-site change is a partial intronic retention (120 bp) culminating in a premature termination codon and a shortened nebulin protein (p.Asn6649_Ile8560delins27). The patient has a brother with similar muscle complaints but he was unavailable for study; no other family members were known to be affected. The second NEB mutation identified in this work (c.24294_24297dup) was detected in patient P5 (Figure 4.1.2B), probably in homozygosity since his parents are first degree cousins. This frameshift duplication was previously found in heterozygosity in patients with NM (Pelin et al., 2002; Lehtokari et al., 2006; Lehtokari et al., 2014). The neonatal period of patient P5 was uneventful with only initial feeding difficulties reported. There was significant respiratory involvement

starting at 2 years of age. Two years later the respiratory difficulties progressed, requiring NIV. Corrective surgery for scoliosis was performed at 15 years of age. He currently has a proximal tetraparesis and walks without assistance. The patient had an older brother (deceased and without genetic studies) with a similar phenotype.
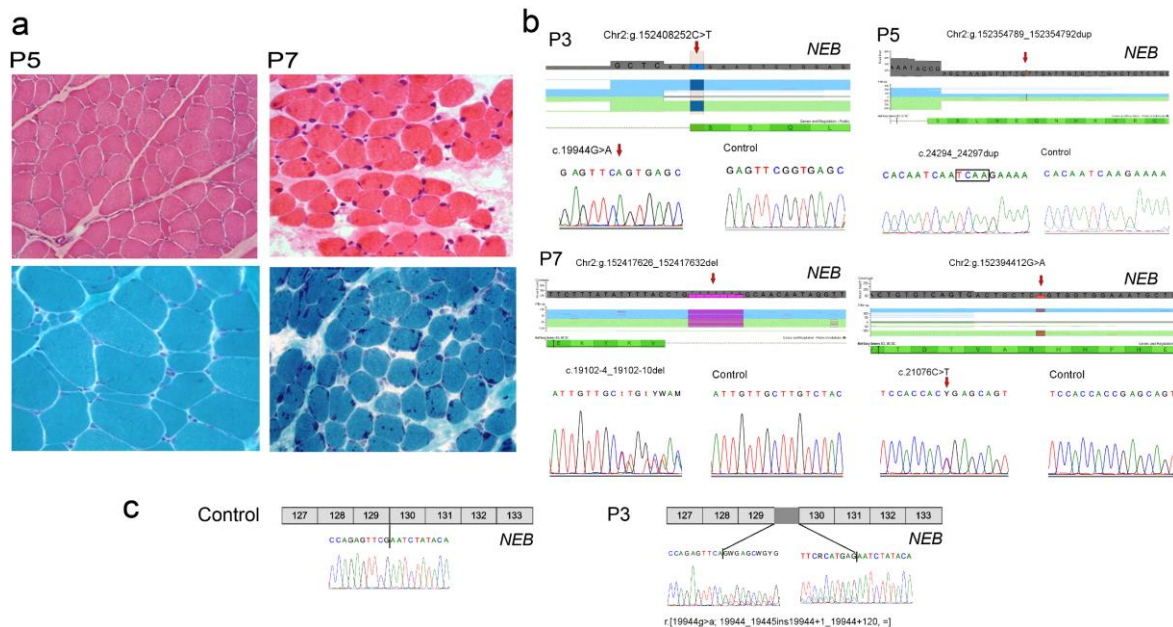
Figure 4.1.2- *NEB*-related nemaline myopathy. (A) Histological features of deltoid muscle biopsies of patients P5 and P7 (H&E with 200x magnification in P5 and 400x in P7; GT with 400x magnification). In patient P5 there was fiber size variability, with severely atrophic fibers, corresponding the majority to type 1 fibers. On the GT stain, there was a small area with some structures resembling rods. In P7 case, the typical rods were clearly seen on the GT stain. (B) Homozygous *NEB* mutations identified in patient P3 (c.19944G>A) and P5 (c.24294_24297dup), identified by MPS (top part) and confirmed by Sanger sequencing. Two heterozygous NEB mutations identified in patient P7: i) an intronic deletion of 7 bp (c.19102-4_19102-10del) affecting the acceptor splice-site and ii) a nonsense mutation c.21076C>T (p.Arg7026*). (C) Characterization of the effect of the c.19944G>A mutation at the mRNA level. This mutation abolishes the normal donor splice-site, shifting towards the use of an alternative (cryptic) splice-site in intron 129, which predictably originates a premature stop codon and a shorter polypeptide. centronuclear myopathy.

Patient P7 had the severest NM presentation of the three cases included in this work. He had a congenital disease onset, presenting at birth with hypotonia and predominantly distal muscle weakness with bilateral club foot. The patient required NIV since the age of 8. He has two heterozygous mutations in the *NEB* gene (Figure 4.1.2B). Although his parents were not studied, these mutations are likely to have distinct origins, since only one (p.Arg7026*) is in common with a distant cousin also affected with NM. This nonsense mutation has been

214

previously identified in a third Portuguese patient with a (*NEB*-related) nemaline myopathy (Lehtokari et al., 2014). Surprisingly, this patient is reported to carry the same nonsense mutation as found in our patient P7 and the splicing mutation (c.19944G>A) identified in our patient P3. She was described as having a typical NM, which is milder than patient P7 but more severe than P3. The other *NEB* mutation found in patient P7, a 7 bp deletion located at the end of intron 122 (c.19102-4_19102-10del), was also previously described (Lehtokari et al., 2014) but not evaluated at the mRNA level. We demonstrated that it affects the *NEB* splicing process (Appendix III.1 Figure III.1.5). Two abnormal transcripts were identified in muscle-derived total RNA: one is consequential of the exon 123 skipping (r.19102_19206del) and the other is the total retention of intron 122 (r.19101_19102ins19101+1_19102-1) in mature mRNA, predictably originating a shorter polypeptide. Patient P7 is also a carrier of a pathogenic variant (c.1384T>G, p.Scy462Gly) in the *SEPN1* gene. Although it may not influence the patient's phenotype, this variant should be considered for genetic counseling purposes in this family.

TTN-related myopathies

Mutations in the *TTN* gene were identified in two unrelated cases (P4 and P6) (Figure 4.1.3). Patient P4 had a severe neonatal hypotonia with ophthalmoparesis, facial and bulbar weakness, dilated cardiomyopathy (cause of death at 10 years of age), requiring invasive ventilation and a feeding tube. The youngest of his two brothers had a similar phenotype and died during the neonatal period. P4's muscle biopsy showed myopathic features and multiple minicores (Figure 4.1.3A). Two novel heterozygous mutations in *TTN* were identified in patient P4. One of the mutations is a 2 bp deletion (c.76109_76110del, Figure 4.1.3B) of paternal origin, which presumably gives rise to a premature termination codon and a shorter polypeptide (p.Ile25370Argfs*6). The second mutation, maternally inherited, is an intronic SNV located near the splice site (c.52102+5G>A). As predicted by bioinformatic analysis, muscle RNA studies revealed that this variant is disease-causing: the canonical splice-site is abolished and a cryptic splice-site located within exon 273 is alternatively used, originating an in-frame deletion of 69 bp (Figure 4.1.3C). The analysis of the other family members demonstrated the autosomal recessive inheritance pattern (Figure 4.1.3D).

Patient P6 had a milder phenotype, presenting with delayed walking during infancy. He had facial and limb-girdle weakness, mostly scapular weakness. Muscle biopsy showed myopathic changes and multicores (Figure 4.1.3A). The patient is currently 17 years old and shows no respiratory or cardiac involvement. The genetic study revealed a *de novo* heterozygous mutation: a 2 bp insertion in exon 304 (c.61679_61680insCC) (Figure 4.1.3B), that predictably gives rise to a premature stop codon in the A-band of Titin (p.Ser20561Leufs*17).
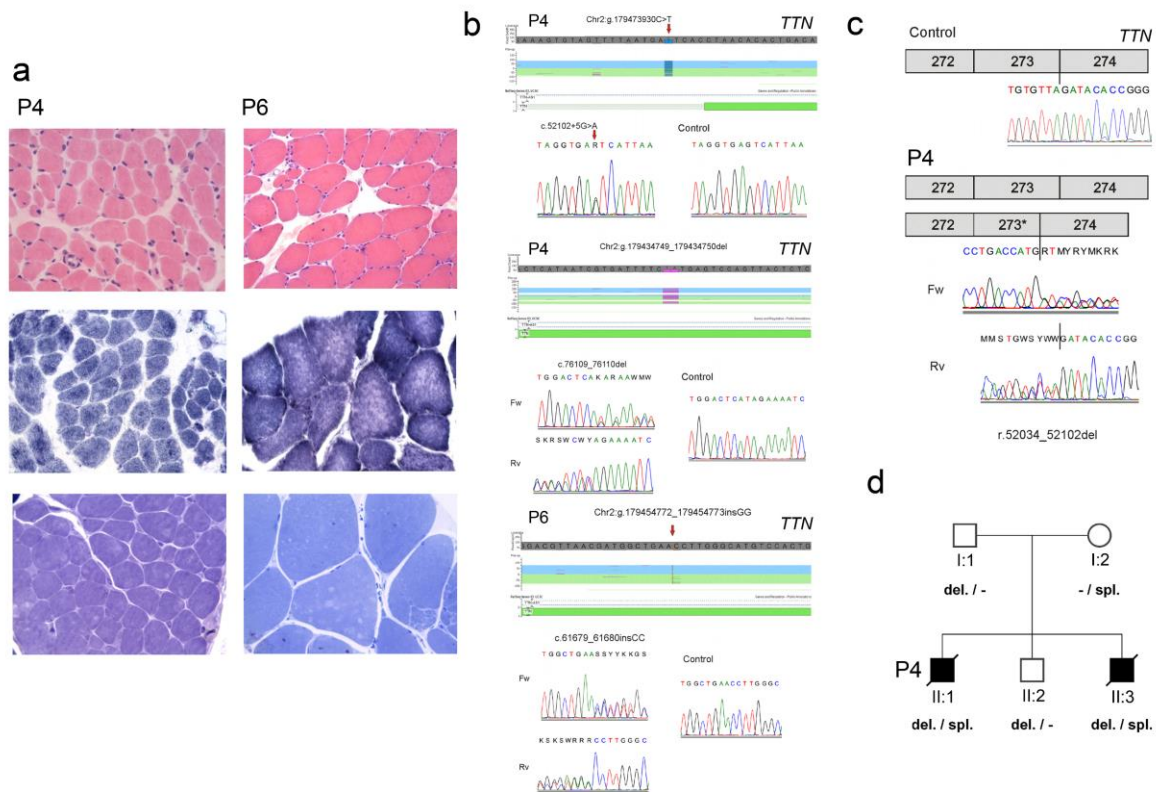
Figure 4.1.3- *TTN*-related myopathies. (A) Histological features of patients P4 and P6 deltoid muscle biopsies (H&E with 400x magnification in P4 and 200x in P6, NADH with 400x magnification and toluidine blue semi-thin sections with 630x magnification in P4 and 400x in P6). In both cases there was fiber size variability and multiple small minicores seen on oxidative stains and semi-thin sections. (B) Heterozygous *TTN* mutations identified in patient P4: i) splice-site mutation in intron 279 (c.52102+5G>A) and ii) an out-of-frame deletion (c.76109_76110del) in exon 326. In patient P6 a heterozygous insertion of two nucleotides was detected (c.61679_61680insCC), which leads to a shift on the *TTN* reading frame (p.Ser20561Leufs*17). (C) The c.52102+5G>A mutation was further characterized at the mRNA level, where it was shown that it abolishes the donor splice-site and a cryptic splice-site in exon 273 is used in alternative, leading to an in-frame deletion (r.52034_52102del). (D) Pedigree of patient P4 family is compatible with an autosomal recessive inheritance pattern, where the two affected sibs share the genotype. Fw- forward; Rv- reverse.

Additionally, a novel silent variant was identified in exon 358 (c.105648G>A), but its effect at the mRNA level has yet to be determined.

4.1.5 DISCUSSION

We report the development of a novel and efficient method for mutation detection based on MPS and its application in the genetic characterization of patients with CM.

Several centers reported low positivity rates in the molecular characterization of CM patients. These numbers may be seen in light of the fact that these studies were restricted to a relatively small number of candidate genes that were known, until recently (North et al., 2014). Additionally, although some clinical and histological features allow some diagnostic guidance to the genetic study, in a subset of cases neither can orientate towards a (single) specific gene. Since the pathognomonic morphological abnormalities may show progressive or age-related changes (Romero et al., 2013), the diagnostic value of the muscle biopsy depends when it is performed. As a final point, the severest end of the CM disease spectrum includes cases with pre-natal onset and neonatal death, where the scarcity of clinical symptoms and the limited number of additional tests often cut-short the diagnostic workup.

The introduction of new sequencing technologies (next-generation sequencing or MPS) circumvented some of these limitations, enabling genetic diagnosis of diseases which had been considered extremely difficult to study by traditional methods. Different strategies based on MPS have been proposed for the genetic characterization of hereditary myopathies. These range from gene panels with a limited number of loci aiming to obtain high coverage depth (Kondo et al., 2012; Ankala et al., 2015) to larger panels targeting several loci known to be implicated in myopathies (Savarese et al., 2014) or even those with a broader scope, such as that covering all known neuromuscular disorders (Valsi et al., 2012). To address the genetic heterogeneity of these diseases, several successful applications of whole-exome or -genome sequencing were reported in the literature (Böhm et al., 2013; Bögershausen et al., 2014; Couthouis et al., 2014: Oliveira et al., 2015). Irrespective of the approach taken, the histopathological findings should be correlated with the genetic data that is generated, as these can help narrow down the genetic studies considerably (Uruha et al., 2015).

Our strategy consisted in the development of a well delimited gene panel for CM which in the present cohort led to an overall mutation detection rate of 70%. This is relevant considering that the assay does not include all the genomic regions, nor does it detect the presence of large genomic rearrangements. In patient P6 only one heterozygous frameshift mutation was identified. This was a *de novo* mutation (not detected in the patient's parents and paternity was confirmed). *TTN*-related myopathies are currently an expanding group of diseases, seen to have a wide clinical presentation that ranges from congenital to late adulthood onset, and with autosomal recessive or dominant inheritance (Chauveau et al., 2014). The histological data of patient P6 showed the presence of multicores in NADH staining and by electron microscopy

217

areas of sarcomeric disruption with irregularities and smearing of the Z line and disorganization of the myofibrils (data not shown). Additional experimental evidence is being gathered to clarify the impact of the apparently silent variant (c.105648G>A) and to rule out or confirm the presence of an additional *TTN* mutation, not detectable by the methods used.

About 25% of the variants detected in this group of patients required further studies (in particular, expression analysis at mRNA level). Although bioinformatic analysis was highly suggestive of their pathogenic nature, their impact was ultimately demonstrated by these additional studies. The best example was the study of patient P3, where *NEB* gene analysis revealed the presence of an apparently silent homozygous variant (c.19944G>A), described in previous reports as likely to be deleterious (Pelin et al., 2002). Given that the patient's biopsy was performed several years ago and there were no muscle fragments to obtain RNA, the possibility of using blood-derived transcripts was evaluated. In fact, we detected suitable expression of nebulin in whole blood nucleated cells, which enabled the study of the c.19944G>A variant and its effect on splicing. This mutation abolishes the normal splice-site consensus sequence leading to the use of an alternative (cryptic) splice in the adjacent intron. It is predictable that this partial intronic retention creates a premature termination codon and thus a shorter nebulin protein. The presence of a residual amount of normal transcripts, possibly due to a leaky mutational effect, explains the milder phenotype in this patient.

Even though the application of this gene panel was very successful, some technological limitations should be mentioned. First, one variant in a positive control was incorrectly genotyped due to an allele drop-out during PCR, owing to the presence of a rare SNP in the patient. Such events are not unusual in conventional sequencing. In an attempt to avoid this pitfall, the bioinformatic process for primer design used in the Ampliseq strategy takes into account positions of known SNPs. However, de novo, ultra-rare or population-specific variants will not be considered in the pipeline. This false-positive result had no practical implication in our study, because all the variants identified by MPS were confirmed by Sanger sequencing. Our main concern is the scenario where the allele drop-out prevents the amplification of the mutated allele, generating a false-negative result. Data analysis is one of the biggest hurdles in this technology, given the different possibilities of error that may lead to false-positive or -negative results (Robasky et al., 2014). As shown here, besides the development of the MPS gene panel it was necessary to deploy a strategy to deal with the large number of variants by using specific variant filters. During this process, one of the mutations detected in this work was initially filtered-out due to incorrect annotation. It is highly recommended, especially in large variant datasets, that different sources of variant annotation be used so as to ensure to allow accuracy and consistency. Variant annotation is one additional layer of variability in MPS (McCarthy et al., 2014) which very often remains disregarded.

As a future perspective, the intention is to expand this gene panel to include loci that have recently been linked to CM, namely: *KLHL41*, *LMOD3* and *SPEG* (Gupta et al., 2013: Agrawal et al., 2014; Yuen et al., 2014). It would also be relevant to perform whole-exome sequencing in patients that remain genetically unsolved after screening with this gene panel.

In conclusion, the wide phenotypic heterogeneity and huge size of the candidate genes makes diagnosis of CM complex, costly and labor-intensive. We have developed and demonstrated the clinical utility of a new MPS gene panel for the screening of these patients. In seven of the ten undiagnosed cases, pathogenic variants were identified, most of which contributed towards widening both the mutational and the phenotypic spectra of CM.

4.2 New splicing mutation in Choline Kinase Beta (*CHKB*) gene causing a muscular dystrophy detected by whole-exome sequencing

Oliveira, J., Negrão, L., Fineza, I., Taipa, R., Melo-Pires, M., Fortuna, A. M., Gonçalves, A. R., Froufe, H., Egas, C., Santos, R., Sousa, M. (2015). Journal of Human Genetics, 60(6):305-312.

4.2.1 Abstract

Muscular dystrophies (MD) are a group of hereditary muscle disorders that include two particularly heterogeneous subgroups: limb-girdle MD and congenital MD, linked to 52 different genes (7 common to both subgroups). Massive parallel sequencing technology may avoid the usual stepwise gene-by-gene analysis. We report whole exome sequencing (WES) analysis of a patient with childhood-onset progressive MD, also presenting mental retardation and dilated cardiomyopathy. Conventional sequencing had excluded eight candidate genes. WES of the trio (patient and parents) was performed using the Ion Proton sequencing system. Data analysis resorted to filtering steps using GEMINI software revealed a novel silent variant in the choline kinase beta (*CHKB*) gene. Inspection of sequence alignments ultimately identified the causal variant (*CHKB*:c.1031+3G>C). This splice site mutation was confirmed by Sanger sequencing and its effect was further evaluated by gene expression analysis. On reassessment of the muscle biopsy, typical abnormal mitochondrial oxidative changes were observed. Mutations in *CHKB* have been shown to cause phosphatidylcholine deficiency in myofibers, causing a rare form of CMD (only 21 patients reported). Notwithstanding interpretative difficulties that need to be overcome prior to the integration of WES in the diagnostic workflow, this work corroborates its utility in solving cases from highly heterogeneous groups of diseases, where conventional diagnostic approaches fail to provide a definitive diagnosis.

4.2.2 INTRODUCTION

Muscular dystrophies (MD) are a set of highly heterogeneous muscle diseases sharing distinctive dystrophic features on muscle histology, namely degenerative fibers and fibrosis, variable myofiber size, and abnormally internalized nuclei (Shieh et al., 2013). MD are classically subdivided into five major groups: Congenital MD (CMD), Duchenne/Becker MD, Emery-Dreifuss MD, Facioscapulohumeral MD and Limb-girdle MD (LGMD). Two of these groups, CMD and LGMD, are particularly heterogeneous from both a clinical and a genetic perspective. Until now 31 different loci have been reported for the LGMDs and 28 genes in CMD (Nigro et al., 2014; Bönnemann et al., 2014). Notably, at least 7 genes are shared between these two groups and this list is increasing over time. Concerning CMD, mutations in the laminin-α2 (*LAMA2*) gene are a major cause of the disease, corresponding to approximately 58% of patients referred for genetic studies in this disease group (Oliveira et al., 2014). In patients with inconclusive imunohistochemistry (IHC) studies, the diagnostic workup may be more complex and is usually steered by other muscle biopsy features, clinical evaluation, electromyogram, brain and muscle imaging. Specific clinical signs are important to consider, namely the distribution and progression of muscle weakness, the presence of brain structural changes (with or without mental retardation) and other features such the presence of contractures, rigid spine and hyperlaxidity (Bönnemann et al., 2014). All these features have been used to classify CMD into further subtypes and to establish robust diagnostic algorithms. Besides laminin-α2 deficiency or changes for α-dystroglycan, both detected by IHC, other clinically relevant histological findings (such as structural changes) are rare.

Nishino and collaborators (1998) reported a subtype of CMD with mitochondrial structural abnormalities, presumably of autosomal recessive inheritance. The four patients described shared some clinical features with other CMD forms but had marked mental retardation without structural brain changes (Nishino et al., 1998). Disease progression was slow and some patients developed dilated cardiomyopathy. Creatine kinase (CK) levels were mildly elevated. Muscle phenotype included dystrophic features and a striking finding was observed with succinate dehydrogenase (SDH) and cytochrome c oxidase staining, revealing reduction of mitochondrial content in the center of the myofiber and enlarged mitochondria at the periphery, assuming a "megaconial" appearance (Nishino et al., 1998). This entity was later demonstrated to be caused by mutations in the choline kinase beta gene (*CHKB*), leading to loss of activity for this enzyme in rostrocaudal muscular dystrophy mice6 and in humans in a very limited number of patients reported to date (Mitsuhashi et al., 2011a; Gutiérrez Ríos et al., 2012; Quinlivan et al., 2013; Castro-Gago et al., 2014; Cabrera-Serrano et al., 2014). CHKB together with two isoforms of choline kinase alfa (CHKA) phosphorylates choline into phosphocholine, the initial step of the Kennedy's pathway. This step is essential in the

biosynthesis of phosphatidylcholine - an important component of the phospholipidic membrane layer of eukaryotes - and also a key precursor for many signaling molecules with relevant cellular functions such as cell growth signaling and proliferation (Aoyama et al., 2004). The disease mechanism has been attributed to the decreased levels of phosphatidylcholine in mitochondria, leading to its dysfunction and subsequently mitophagy and clearance in skeletal muscle (Mitsuhashi et al., 2011b; Mitsuhashi et al., 2011c). Immunoprecipitation studies performed in mouse liver suggested that chk activity was mainly due to heterodimeric CHKA/CHKB (60%), although some of the activity was due to homodimeric forms (Sher et al., 2006). However, in human skeletal muscle CHKA is not expressed, therefore the CHKB isoform is essential to sustain normal phosphatidylcholine levels (Mitsuhashi et al., 2011a).

The development and application of new sequencing technologies, frequently coined next-generation sequencing (NGS), is currently accelerating research in the field of clinical genetics, essentially given its capacity to generate and analyze more genomic sequences at a lower cost and at a faster pace (Buermans & den Dunnen, 2014). Using this technology it is possible to analyze several specific genes simultaneously, comprising a disease panel, or at the limit it may include all the exonic regions of the human genome - whole exome sequencing (WES) or even the entire genome (individual genome sequencing). Whole exome or genome sequencing may also contribute to the identification of new genes associated with genetic diseases. Nevertheless, its introduction into routine diagnostics still faces some challenges, particularly considering the amount of data to be analyzed, and the possibility of obtaining false positive and negative results should not be neglected.

In this work we report the use of WES in a muscular dystrophy patient, which contributed to the identification of a novel splicing mutation in the *CHKB* gene. Our study corroborates the potential of WES to identify genetic defects involved in rare neuromuscular diseases and illustrates the difficulties that may be encountered using this technology, especially regarding data analysis.

4.2.3 MATERIAL AND METHODS

Patient

The patient is the first child of a non-consanguineous healthy couple; pregnancy and neonatal period were uneventful. She was first observed by neuropediatrics at the age of 4 years because of motor problems: independent walking at 2.5 years, progressive lower limb weakness with difficulties in climbing stairs and running. She is presently unable to rise from a chair, the arms reach do not rise to shoulder level and she walks with braces. The CK values,

ascertained over a period of thirty years, were slightly elevated, ranging from 299 to 1857 U/L (average 849 U/l; normal < 200 U/l). Electromyography performed at age 7 showed a myopathic pattern without evidence of fibrillation potentials or myotonic discharges. She also presented behavioral abnormalities and cognitive deficit, which prevented her from attaining the basic academic goals. At the end of the first decade dilated cardiomyopathy was diagnosed, adequately controlled with medication (digoxin and furosemide). Standard genetic screening tests due to delay in psychomotor development (including speech delay), performed at age 22, were all normal. The patient had no signs of ichthyosis. Informed consent was obtained and the research was approved by the ethics committee from the Centro Hospitalar do Porto.

Sanger Sequencing

Eight genes known to be defective in muscular dystrophies were previously analyzed by conventional (Sanger) sequencing, namely: *ANO5, CAPN3, LARGE, FKRP, FKTN, POMGNT1, POMT1* and *POMT2*. Genomic sequencing included all coding regions and adjacent intronic sequences. Results were analyzed using SeqScape software V2.5 (Life Technologies). Variants were described according to the Human Genome Variation Society (HGVS) guidelines for mutation nomenclature (version 2.0) (den Dunnen & Antonarakis et al., 2001). Large heterozygous deletions and duplications were also screened in several genes using the multiplex ligation-dependent probe amplification (MLPA) technique. Different commercial kits from MRC-Holland (Amsterdam, the Netherlands) were used*: SGCA, SGCB, SGCD, SGCG* and *FKRP* (P116); *FKTN, LARGE* and *POMT2* (P326); *POMT1* and *POMGNT1* (P061). Genomic DNA (150 ng) from the patient and at least three normal controls were used in the procedure. MLPA products were resolved by capillary electrophoresis on an ABI 3130xl genetic analyser (Life Technologies). Data analysis was conducted using GeneMarker Software V1.5 (SoftGenetics LLC, State College, PA). After using the population normalization method data was plotted using the probe ratio option.

Exome sequencing

Seventy five nanograms of high quality DNA from each individual (patient and both parents) were amplified in twelve primer pools of 200 bp amplicons with the Ion AmpliSeq™ Exome Library Preparation kit (Life Technologies). Samples were barcoded with the Ion Xpress Barcode Adapter (Life Technologies) in order to pool two exomes per chip. The libraries were evaluated for quality with High Sensitivity DNA Kit in Bioanalyzer (Agilent Technologies, Santa Clara, CA) and quantified using Ion Library Quantitation Kit (Life Technologies). Library fragments were clonally amplified by emulsion PCR using the Ion PI Template OT2 200 kit v2 and the Ion OneTouch 2 System (Life Technologies), and the positive Ion Sphere Particles

enriched in the Ion OneTouch ES machine (Life Technologies). Finally, the positive spheres were loaded in Ion PI chip v2 and sequenced in the Ion Proton System (Life Technologies). All procedures were carried out according to the manufacturer's instructions.

Exome mapping and variant calling

The data generated from the three exomes (patient and parents) was processed with the Torrent Suite Software 4.1 (Life Technologies). Reads were mapped against the human reference genome hg19 using Torrent Mapping Alignment Program (TMAP) version 4.0.6 (Life Technologies). Variant calling was performed by running Torrent Variant Caller plugin version 4.0, using the optimized parameters for exome sequencing recommended for AmpliSeq sequencing (Life Technologies). The variant call format (VCF) files from the trio were combined using vcftools version 0.1.9.0. (Danecek et al., 2011) and all variants were annotated and prioritized with GEMINI (Paila et al., 2013). Variant filtering was performed using a list of 108 candidate genes known to be implicated in hereditary myopathies (Appendix III.2 Table III.2.1). Each gene was screened for variants matching the recessive disease model, being: i) homozygous in the patient and parents carrying the same variant in heterozygosity ("autosomal recessive") or ii) two heterozygous variants in the patient (compound heterozygosity, "comp_hets" function) and each having a distinct parental origin. Candidate variants were manually checked on the Binary Alignment Map (BAM) files through Integrative Genomics Viewer (IGV) version 2.3 (Thorvaldsdóttir et al., 2013) and GenomeBrowse V2.0.0 (Golden Helix Inc., Bozeman, MT).

Confirmation of variants by Sanger sequencing

To verify the candidate variant detected in the *CHKB* gene of the three individuals, specific primers were designed (9F- 5'GTGGAGTCTGGAAGGAATGGC; 9R- 5'TTTAACTTCTCCCCCACTGTCC) to amplify the region of interest by PCR. Resulting PCR products were purified using Illustra ExoStar (GE Healthcare, Little Chalfont, UK) and subjected to a new cycling reaction prepared with BigDye Terminator kit V3.1 chemistry (Life Technologies). Sequencing reactions were analyzed in an ABI 3130xl genetic analyser (Life Technologies).

Gene expression analysis

Total RNA was obtained from peripheral blood (patient and a control) using the PerfectPure RNA Blood Kit (5 PRIME, Germany) and from muscle (only available from a control) with the PerfectPure RNA Fibrous Tissue kit (5 PRIME). RNA was converted into cDNA using the High Capacity cDNA Reverse Transcription kit (Life Technologies). For the analysis of *CHKB* transcripts PCR amplification was performed using primers annealing to the regions

corresponding to exons 7 to 11 (6/7F-5'TCCAGGAAGGGAACATCTTG; 11R-5'GGTGGAGTCAGGATGAGGAG; based on reference sequence NM_005198.4). As an experimental control, the 3' region of the transcripts for protein-O-mannosyltransferase 1 (*POMT1*) was simultaneously amplified, using the following primers: 18F-5'CGGCGAAGAAATGTCCATGAC and 3'UTR-R 5'GAGCTTTTCAATGAGACCCCC (based on reference sequence with accession number NM_007171.3). Resulting PCR products were resolved in a 2% agarose gel; visible bands were excised and purified using PureLink™ Quick Gel Extraction kit (Invitrogen by Life Technologies). Products were sequenced as described above.

Muscle histology

Two muscle biopsies were performed, at 9 and 20 years of age. In the first biopsy histochemical studies were not performed and paraffin sections showed non-specific changes. The second biopsy revealed striking variation in fiber size, with type 1 and type 2 fiber atrophy, necrotic fibers, several basophilic fibers and endomysial fibrosis. IHC showed normal labeling for sarcoglycans, dystrophin and laminin-α2. SDH and nicotinamide adenine dinucleotide (NADH) staining was uneven in central areas of the myofiber. The conclusion at that time (1995) was that these findings pointed towards a possible muscular dystrophy.

4.2.4 RESULTS

Preceding molecular genetic analysis

The patient was followed by neuropediatrics and later neurology, with the clinical suspicion of LGMD or possibly a dystroglycanopathy. Eight candidate genes were initially analyzed by Sanger sequencing as part of the routine genetic diagnostic workup. Several variants were detected in these studies, including a novel silent heterozygous variant (NM_024301.4:c.474C>G) in the fukutin related protein (*FKRP*) gene, which prompted further characterization at the cDNA level. Since no splicing defects were identified, this unclassified variant was considered non-pathogenic. These studies were complemented by screening for large genomic deletions or duplications at several *loci* by MLPA, but again no causal mutations were detected. These continuous molecular studies were inconclusive and unable to identify the genetic defect responsible for the patient's phenotype. The high number of candidate genes that were screened and the detailed clinical characterization consistent with a genetically determined MD, led us to consider WES as an appropriate alternative approach to identify the underlying defect.

Exome sequencing metrics

Overall quality parameters of the exome trio are summarized in Table 4.2.1. On average each exome generated a total of 37 million sequence reads, 99% of which successfully aligned against the human reference genome. Accordingly, the exome's target regions were on average covered 105.6 times and 93% of these had a minimum coverage depth of 20-fold. These values are relatively high and above the threshold considered appropriate for this study. Approximately 83% of all sequenced bases had a Phred quality score above Q20 (i.e. a 99% accuracy of the base call). The size of the total captured regions was ~58 Mbp. For each sequenced exome one BAM and VCF file were generated. An overview of the variants called for each exome is described in Table 4.2.2. WES identified 48 847 single nucleotide variants (SNV) in the index case and 48,870 and 47 878 SNV in each parent. More than 2 300 insertions/deletions were also detected in each individual exome. The patient's exome contains 1 820 variants not previously reported in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/).

Table 4.2.1- Exome sequencing quality metrics.

Individual	Patient	Mother	Father
Total Reads (number)	41 776 885	38 365 995	31 441 741
Aligned Reads passed filtering (number)	41 271 480	37 922 459	31 102 979
Aligned Reads passed filtering (%)	99	99	99
Mean Coverage of Target Region (x)	117.3	109.5	89.96
Bases with Q > 20 (%)	83.02	83.37	83.60
Total captured regions size (Mb)	57.74	57.74	57.74
Captured regions with coverage >20 (%)	93.19	93.42	91.17

Footnote: Sequencing and mapping metrics are presented per individual.

Table 4.2.2- Exome variant metrics.

Individual	Patient	Mother	Father
Total number of SNVs	48 847	48 870	47 878
Total number of indels	2 473	2 375	2 332
Homozygous variants	19 510	19 231	18 958
Heterozygous variants	31 810	32 014	31 252
Novel variants (*)	1 820	1 749	1 610

Footnote: Variants (with quality ≥ 50) are presented per individual. Variant calling files from each sequenced individual were combined and analyzed with GEMINI. (*) – not listed in dbSNP (no "rs" number attributed).

227

Exome analysis and variant filtering

To deal with the large amount of variants identified by WES (>50 000 variants in each individual exome) the analysis approach was based on the patient's clinical phenotype and the most probable inheritance pattern for the disease. The first filtering strategy focused on the known genetic causes for hereditary myopathies (including muscular dystrophies), where a total of 108 candidate genes were selected (Appendix III.2 Table III.2.1), based on the information available in the Muscle Gene Table website (http://musclegenetable.fr/, accessed April 2014). The GEMINI database framework was used to annotate and filter the large number of sequence variants listed in the three generated VCF files. This software has different built-in analysis tools that are useful for identifying de novo mutations, but also to filter variants meeting autosomal recessive or autosomal dominant inheritance patterns in family-based studies (Paila et al., 2013). Since no other relatives of our patient were reported as being affected with a muscle disease, the autosomal recessive model was used, which in GEMINI is sub-divided in two different functions: "comp_hets" and "autosomal_recessive". Considering that parental consanguinity was not reported, the first tool used was "comp_hets". This function is suitable to identify potential compound heterozygote genotypes. A list of compound heterozygous variants in the selected 108 genes was generated for the patient (Appendix III.2 Table III.2.2). A total of 29 heterozygous variants in 11 different genes were identified, 24 are considered polymorphisms (frequency higher than 1.0%) and five had not been documented in dbSNP. After manually checking the reads alignments on those genomic coordinates (for the trio and by comparing with a control exome) three variants were considered sequencing artifacts. One heterozygous missense variant (NM_004369.3:c.1927C>A, p.Leu643Ile) was detected in the α3 chain of the collagen VI gene (COL6A3), but it was seen to have been inherited from the father, who was homozygous for this change. A second variant (NM_001267550.1:c.33513_33515dup) located in the titin gene (*TTN*), predictably resulting in the insertion of a new codon, was considered non-pathogenic since it was previously reported in the Exome Variant Server (http://evs.gs.washington.edu/EVS/) with an overall frequency of 1.3%.

As no suitable candidates were found, a second filtering approach ("autosomal_recessive") was used. This identifies canonical recessive mutations by generating a list of homozygous variants present in the patient and inherited from each parent, heterozygotes for those changes. In this analysis fewer variants were identified (n=14) present in a total of 8 different genes (one variant each in *CHKB*, *CNTN1*, *DNM2*, *ETFDH*, *ITGA7*, *NEB* and *TIA1* and 7 variants in *SYNE1*) (Appendix III.2 Table III.2.2). Thirteen variants were already listed in publically available sequence variant databases (frequency higher than 1.0%) and one was a novel sequence change located in the CHKB gene (NM_005198.4:c.843T>C).

Variant data interpretation

The novel homozygous variant c.843T>C located in exon 8 of the *CHKB* gene (Appendix III.2 Figure III.2.1) was evaluated by bioinformatic tools and prompted a detailed analysis of the alignments encompassing this gene. An additional novel variant was detected (NM_005198.4:c.1031+3G>C), located in the donor splice site of intron 9 (Figure 4.2.1A and Table 4.2.3). In this region the read depth was around 40-fold and only reverse sequences were successfully aligned (Appendix III.2 Figure III.2.2). It should be noted that the c.1031+3G>C variant was listed as heterozygous in the VCF files of all three exomes, thereby explaining why it was filtered-out with GEMINI's "autosomal_recessive" function. Conventional (Sanger) sequencing confirmed it to be homozygous in the patient and heterozygous in both parents (Figure 4.2.1B). In addition, this variant was not found in homozygosity in the patient's clinically unaffected sister (data not shown).

Table 4.2.2- Data details for the main candidate variant.

Individual	Patient	Mother	Father
Chromosome	22	22	22
Position	51 018 153	51 018 153	51 018 153
Gene Name	*CHKB*	*CHKB*	*CHKB*
Depth	47	39	34
Reference Allele	C	C	C
Number of reads with reference allele [freq.]	2 [0.04]	16 [0.41]	14 [0.41]
Alternate Allele	G	G	G
Number of reads with alternate allele [freq.]	45 [0.96]	23 [0.59]	20 [0.49]
Mutation Type	Splice site	Splice site	Splice site
Refseq accession number	NM_005198.4	NM_005198.4	NM_005198.4
Mutation DNA	c.1031+3G>C	c.1031+3G>C	c.1031+3G>C
Confirmed by Sanger Sequencing	Yes	Yes	Yes

Footnote: Data derived from manual inspection of the BAM file and correction of artifacts for the candidate position in each of the parents and the patient.

Taking into account the possibility of other false negatives in the remaining candidate loci, all heterozygous variants were reassessed. The genotype quality (GQ) score of 90 was determined as a suitable cut-off value for this reassessment (Appendix III.2 Figure III.2.3). Eleven sequence variants (including the c.1031+3G>C change in *CHKB*) were retrieved that had been called as heterozygous in all three samples and having a GQ score <90 in the patient. With the exception of the *CHKB* variant, all were polymorphisms and thus not considered as possible candidates.

The location of the c.1031+3G>C variant, at the exon-intron junction, lead us to suspect a possible effect on the splicing process; approximately 95% of mammalian transcripts have an A or a G nucleotide in the +3 position of the donor splice site (Shapiro et al., 1987). This, together with the clinical presentation, led us to consider c.1031+3G>C as a candidate variant accountable for the disease.
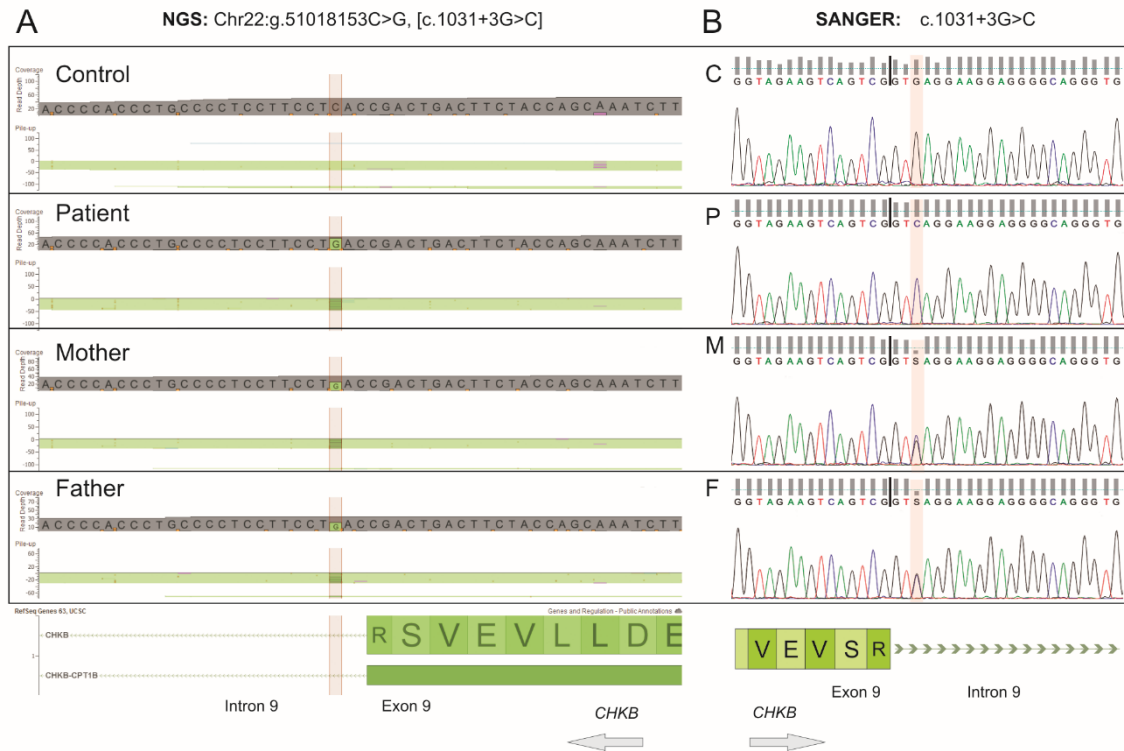
Figure 4.2.1- Novel splice mutation (c.1031+3G>C) identified in the *CHKB* gene. (A) Sequencing results from NGS showing this variant in the patient and in both parents. For this region of the gene only reversely orientated reads were generated. (B) Although annotated in the VCF file as being present in heterozygosity, Sanger sequencing confirmed this variant to be homozygous in the patient (P); the patient's mother (M) and father (F) are both heterozygous carriers. cDNA reference sequence: NM_005198.4.

CHKB gene expression analysis

Further evidence supporting the pathogenicity of the c.1031+3G>C variant was obtained through bioinformatic analysis. Five different algorithms were used to evaluate splicing: SpliceSiteFinder-Like, MaxEntScan (Yeo et al., 2004), NNSPLICE (Reese et al., 1997), GeneSplicer (Pertea et al., 2001) and Human Splicing Finder (Desmet et al., 2009), incorporated in Alamut Visual V2.4 software (Interactive Biosoftware, Rouen, France). In all algorithms there was a reduction of probability scores and particularly in two (NNSPLICE and GeneSplicer) the results were below the value set as a minimum threshold for splicing

(Appendix III.2 Figure III.2.4). This effect was experimentally determined by the analysis of C*HKB* transcripts. cDNA derived from peripheral blood was used, given that no muscle specimens of the patient were available for study. CHKB is ubiquitously expressed and enzymatic activity for this kinase was previously reported in different tissues (Aoyama et al., 2002). Moreover, in control individuals there were no differences in *CHKB* expression between cDNA samples derived from blood and those derived from muscle (Figure 4.2.2A).
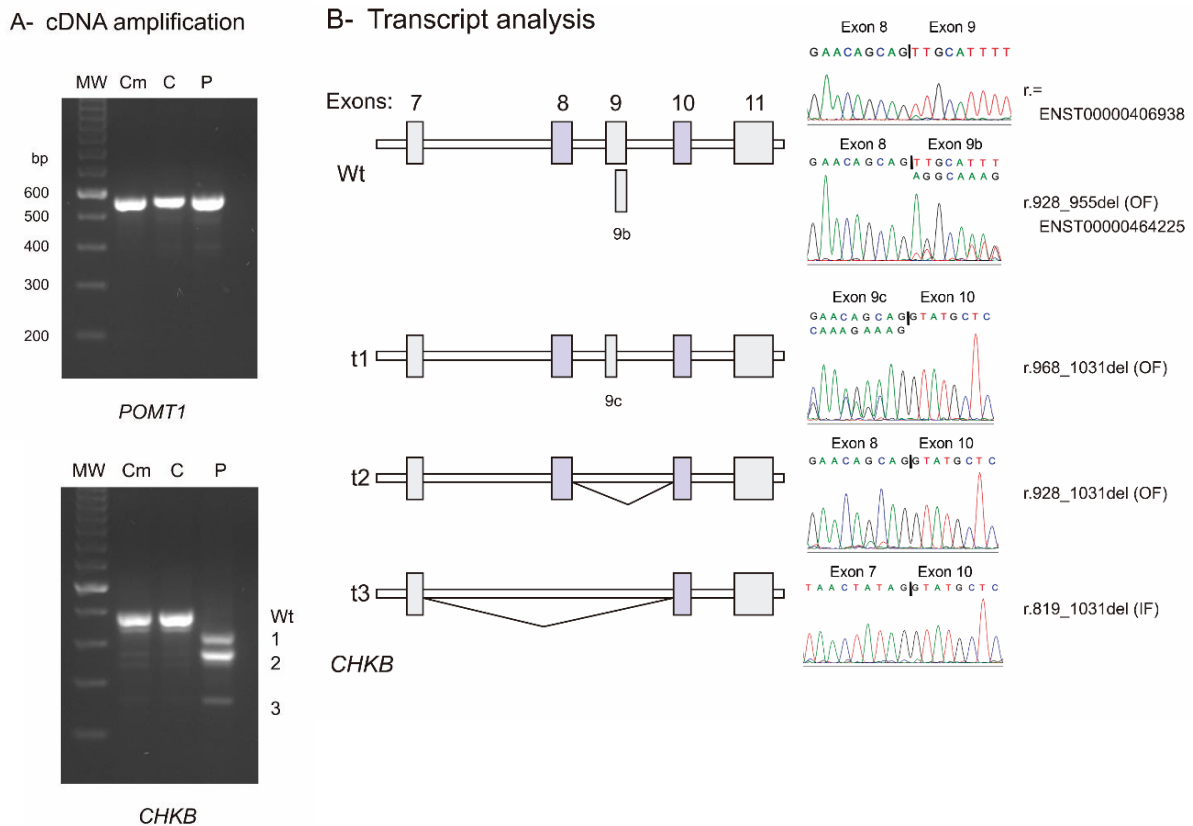
Figure 4.2.2– *CHKB* gene expression analysis. (A) RT-PCR experiments using total RNA obtained from peripheral blood (C- control; P- patient) and from muscle (available only for the control sample, Cm). Amplification of *POMT1* transcripts was performed to show successful amplification in the different samples. Normal *CHKB* transcripts (Wt) were seen in the control sample, whereas in the patient only three aberrant transcripts (1-3) were present. (B) Resulting PCR products were sequenced for further transcript characterization. Besides the normal full-length transcript, there is also an alternatively spliced isoform (9b). Three aberrant transcripts were identified in the patient: t1) resulting from the use of a cryptic splice site in exon 9, causing an out-of-frame (OF) deletion; t2) full exon 9 skipping causing a frame-shift; and t3) skipping of exons 8 and 9 leading to corresponds to an in-frame (IF) deletion.

C- Impact at protein level (*in-silico* prediction)
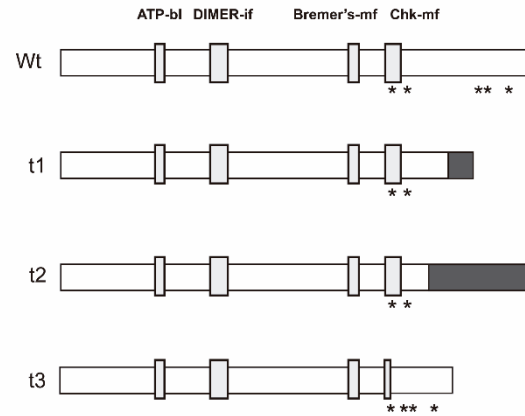
D- Human choline kinase 3D structure

CHKB   CHKA
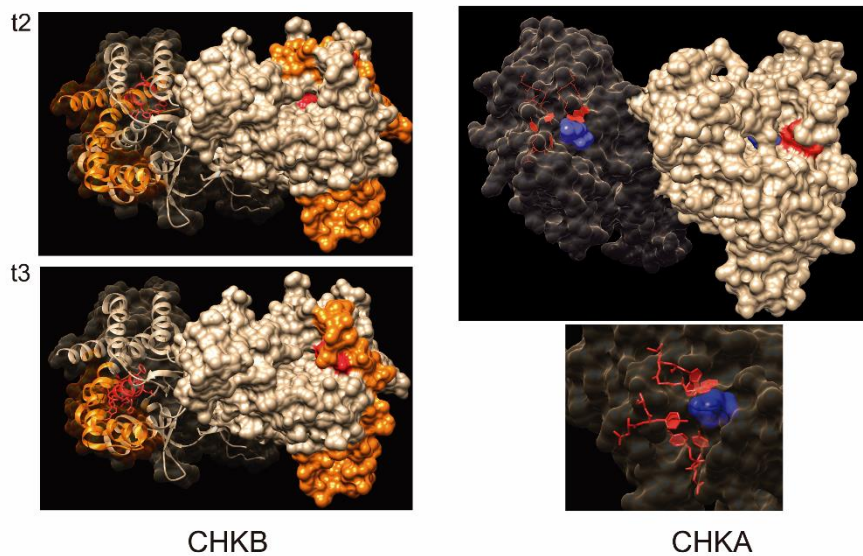
Figure 4.2.2- *CHKB* gene expression analysis (continues). (C) Bioinformatic analysis of aberrant transcripts. Grey rectangles draw attention to relevant CHKB domains: i) ATP-binding loop (bl); ii) dimer-interface (if); iii) Bremer's-motif (mf); and iv) Chk-mf.26 Asterisks highlight important aromatic residues involved in stabilization of the positively charged amine of choline. Black regions represent residues predictably altered as a result of the frame-shift. (D) Crystallographic three dimensional (3D) structures of the human choline kinase beta (CHKB) with Protein Data Bank (PDB): 2IG7 and also the α isoform (CHKA) with PDB: 2CKQ, showing its interaction with the subtract (choline). Both structures were visualized using Chimera software V1.9.27 In one of the chains of the CHK homodimers, transparency was increased to show relevant secondary structures and/ or residues. The regions affected by the deletions identified in transcripts t2 and t3 are depicted in orange. Red residues indicate aromatic aminoacids known to interact with choline, which is shown in blue.

232

In the patient there was a drastic effect on splicing; no normal expression for *CHKB* was detected since the expected normal transcript (with 468 bp, as seen to be present in control samples) was not observed. Instead, she presented at least three aberrant transcripts not detected in the controls. These were further characterized by sequencing (Figure 4.2.2B). All aberrant transcripts were due to splicing defects related with the mutation-derived weakening of the donor splice site. The first transcript (r.968_1031del) corresponds to a 64 bp out-of-frame deletion, where a cryptic splice site located in the middle of exon 9 is used as an alternative. The other two transcripts originate by exon 9 skipping (r.928_1031del) leading to a 104 bp out-of-frame deletion and by exon 8 and 9 skipping (r.819_1031del) leading to a 213 bp in-frame deletion. Should these mutated transcripts be translated, they predictably result in the loss of critical portions of the enzyme (Figure 4.2.2C and D).

The crystallographic structures are available in the Protein Data Bank (PDB) for the human CHKB (PDB: 2IG7) and the α isoform CHKA (PDB: 2CKQ) (Malito et al., 2006). Based on these structures it is predictable that in the mutated polypeptides derived from the transcripts 1 and 2, there is loss of some aromatic residues known to stabilize the positively charged amine of choline; the interaction with this substrate would thus be affected (Figure 4.2.2c and d). In addition, the formation of the choline binding pocket site itself is probably also compromised. The mutated polypeptide with an in-frame deletion derived from transcript 3 (Figure 4.2.2c and d), is predicted to abolish part of the choline kinase motif located in the C-terminal region of CHKB, also involved in the binding of the choline moiety (Malito et al., 2006).

Muscle histology

The results of WES analysis prompted the revision of the patient's muscle biopsy. Oxidative enzyme staining (SDH) revealed not only large pale central areas but also abnormal oxidative activity, particularly towards the periphery of the myofiber (Figure 4.2.3). Electron microscopy shows the presence of slightly enlarged mitochondria (Appendix III.2 Figure III.2.5).

4.2.5 Discussion

We describe the identification of a novel intronic splicing mutation in the CHKB gene, resorting to the innovative analysis of an exome trio (patient and respective parents). Mutations in this locus were previously reported as associated with an extremely rare form of CMD, with only 21 patients accounted for in the literature (Mitsuhashi et al., 2011a; Gutiérrez Ríos et al., 2012; Quinlivan et al., 2013; Castro-Gago et al., 2014; Cabrera-Serrano et al., 2014) and at least three additional cases presented in conference proceedings (Behin et al., 2013; Nascimento et al., 2013). Defects in CHKB have been shown to cause phosphatidylcholine deficiency in muscle cells (Mitsuhashi et al., 2011a). The c.1031+3G>C variant located in the

donor splice site near exon 9, was considered pathogenic upon *CHKB* expression analysis. It was experimentally demonstrated that in the patient reported herein, there are no normal full length CHKB transcripts, and only aberrant transcripts were identified. If these transcripts escape nonsense mediated decay they will probably fail to code for a functional CHKB enzyme. Our experimental findings are similar to those observed in a Turkish CMD patient with a different mutation (c.1031+1G>A) affecting the same donor splice site (Mitsuhashi et al., 2011b).
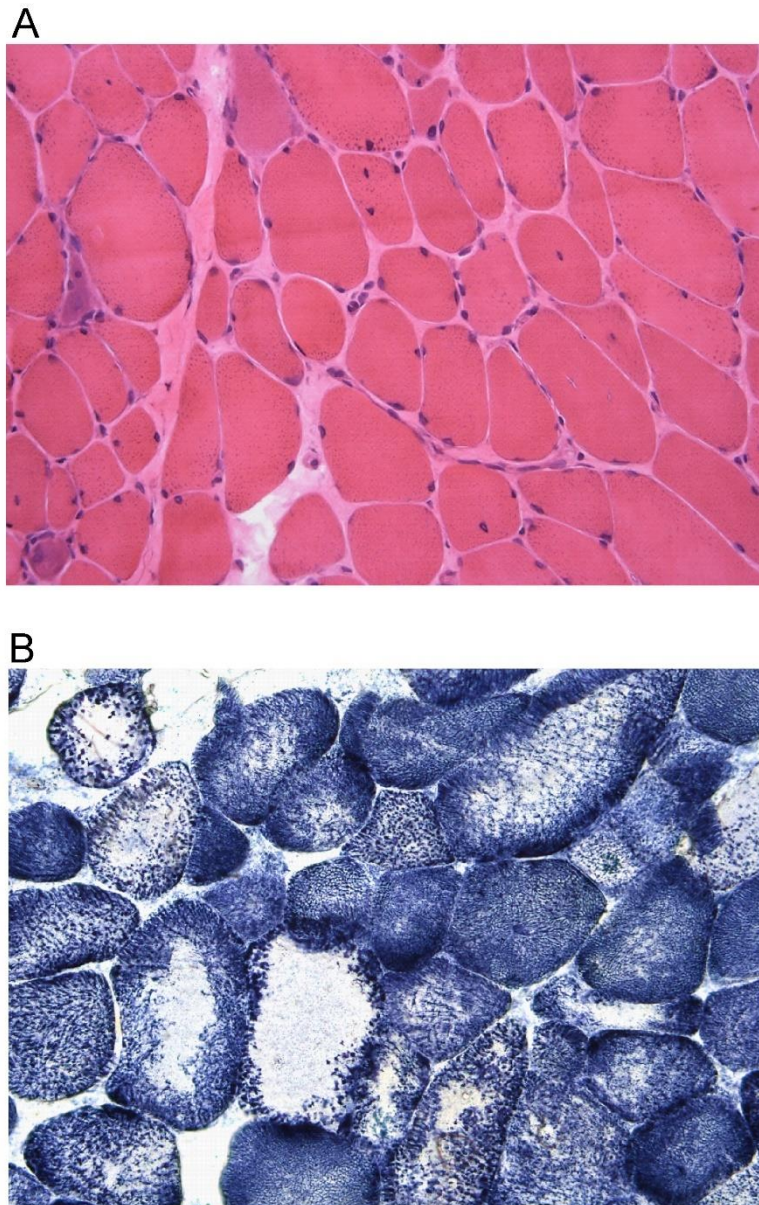
Figure 4.2.3- Muscle biopsy showing: a) variation in fiber size, basophilic fibers and endomysial fibrosis (HE, magnification 200x) and b) large pale central areas with large mitochondria towards the periphery, seen with oxidative enzyme stain (SDH, magnification 200x).

The major clinical features found in the patient (progressive muscle weakness, mental retardation and dilated cardiomyopathy) are in agreement with the phenotype previously described for this entity (OMIM #602541). In the first clinical report of fifteen patients (Mitsuhashi et al., 2011b) findings included mildly to moderately elevated CK levels, floppiness during the neonatal period (in 9/14 patients), mental retardation (15/15), cardiomyopathy (6/14) and independent walking (11/15). Muscle histology was systematically dystrophic and mitochondrial enlargement was observed in all patients. An isolated case with a homozygous nonsense mutation in CHKB was identified in a child with a similar clinical and muscle phenotype (Gutiérrez Ríos et al., 2012). The phenotypic spectrum linked to CHKB deficiency has been recently expanded with additional reports, including a female patient with a milder phenotype resembling LGMD (Quinlivan et al., 2013) and the oldest known patient (50 years old) having a progressive "limb-girdle" myopathy (Behin et al., 2013). Both patients presented missense mutations (two compound heterozygotes and one homozygote in the second report). Muscle histology showed large mitochondria located at the periphery and central areas in the myofiber, devoid of activity. In fact, it should be emphasized that the majority of patients described in the literature harboring *CHKB* mutations, were initially identified by the existence of this specific histological hallmark: the presence of enlarged mitochondria in the periphery of the myofibers and the absence of this organelle in the center of the cells (Nishino et al., 1998). These enlarged mitochondria were correlated with the reduced levels of phosphatidylcholine in the mitochondrial membrane. This seems to be the consequence of a compensatory mechanism resulting from the depletion of functionally compromised mitochondria by autophagy (Mitsuhashi & Nishino, 2011; Mitsuhashi & Nishino, 2013). A common aspect, especially in the atypical cases, is that several muscle biopsies needed to be performed. In the oldest patient reported to date these mitochondrial changes were more evident in the last biopsy (from a total of three) which was performed at the age of 49 years (Behin et al., 2013). This pathognomonic finding (megaconial) is highly specific for the defective gene; however, their detection requires awareness and high quality histological preparations. Therefore, when muscular dystrophy (congenital or progressive) presents in association with mental retardation and cardiac involvement, the muscle biopsy should be carefully reassessed and the CHKB gene considered a possible candidate.

As previously referred, NGS technology is improving the capabilities to perform molecular genetic diagnostics. The most rapidly growing application of this technology is the development of NGS gene panels (targeted resequencing), which can help deal with high genetic and clinical heterogeneity found in several diseases, such as neuromuscular disorders (Valencia et al., 2013). By performing targeted resequencing it is possible to obtain higher redundancy and thus more reliable variant data. This is achievable because the extension of these assays (covered regions) is usually quite smaller when compared to other (broader) NGS approaches

such as WES or whole genome sequencing. Accordingly, NGS targeted resequencing will rapidly replace conventional sequencing for routine genetic diagnosis. However, there is still lack of consensus as to which genes should be included (or not) in such disease-specific gene panels and each laboratory implementing these techniques has been defining their own sets of genes. Another important aspect is that the application of NGS gene panels in some diseases such as hereditary myopathies, will perpetuate the long-established clinical classification, which is currently being challenged by the increasing reports of the same gene being involved in several different subtypes. The ryanodine receptor 1 (*RYR1*) gene is an excellent example of this paradigm, considering its association with different diseases, ranging from congenital myopathies and muscular dystrophies to malignant hyperthermia susceptibility, as well as by presenting dissimilar muscle phenotypes: myopathic with or without structural changes (such as central cores, minicores or centrally placed nuclei), dystrophic signs or apparently normal histology but with compromised function (defective intracellular calcium homeostasis causing hypermetabolic response when exposed to anesthetic agents) (Brislin et al., 2013).

In contrast to NGS targeted resequencing, WES has been proposed as an important research tool for gene discovery or further expanding the genetic causes of a particular disease (Ku et al., 2012). Probably in the near future WES will be used as a first tier diagnostic tool for several medical specialties, allowing the identification of genetic defects linked to monogenic or even complex multigenic disorders. Our initial analysis strategy was orientated towards the genes known to be implicated in hereditary myopathies and, among the attempted approaches, an autosomal recessive model was assumed. It is interesting to note that although the mutation in the *CHKB* gene was successfully identified by WES, its zygosity was not called correctly in the VCF file used for automated analysis. A different unreported variant in the CHKB gene drew our attention to this locus and enabled us to identify the causative mutation. At this stage, considering the novelty aspect of the work, there is no indication if this is a common pitfall of WES data analysis, and the overall impact on filtering strategies is as yet unknown. To ensure that no additional false negative results were obtained upon filtering, a strategy based on the GQ score (threshold <90) was used. Presently WES should be seen as a screening technique, and it does not provide the answer to all aspects of clinical molecular genetic testing. As shown here, several other techniques and biological samples are very often required to clarify the impact on phenotype of novel sequence variants. Moreover, different molecular techniques are required for the detection of more complex mutations such as repeat expansions or structural (copy number) variants that are not readily identified by NGS (Vasli et al., 2013). Another WES analysis constraint is the occurrence of false positives. Some of these sequence artifacts are not easily flagged, so they may be interpreted by the software as

true variants. Even though they are later excluded by conventional sequencing, they can certainly create extra burden on the analysis.

In conclusion, we present a case resolved by WES which enabled improvement of patient follow-up and management and genetic counseling of family members. A novel CHKB gene mutation linked to a rare form of MD was identified by WES, where previously the conventional diagnostic approaches had failed to provide a definitive diagnosis.

4.3 THE NEW NEUROMUSCULAR DISEASE RELATED WITH DEFECTS IN THE ASC-1 COMPLEX: REPORT OF A SECOND CASE CONFIRMS *ASCC1* INVOLVEMENT.

4.3.1 ABSTRACT

Next-generation sequencing technology aided the identification of the underlying genetic cause in a female newborn with a severe neuromuscular disorder. The patient presented generalized hypotonia, congenital bone fractures, lack of spontaneous movements and poor respiratory effort. She died within the first days of life. Karyotyping and screening for several genes related with neuromuscular diseases all tested negative. A male sibling was subsequently born with the same clinical presentation. Whole-exome sequencing was performed with variant filtering assuming a recessive disease model. Analysis focused on genes known to be related firstly with congenital myopathies, extended to muscle diseases and finally to other neuromuscular disorders. No disease-causing variants were identified. A similar disorder was described in patients with recessive variants in two genes: *TRIP4* (three families) and *ASCC1* (one family), both encoding subunits of the nuclear activating signal cointegrator 1 (ASC-1) complex. Our patient was also found to have a homozygous frameshift variant (c.157dupG, p.Glu53Glyfs*19) in *ASCC1*, thereby representing the second known case. This confirms *ASCC1* involvement in a severe neuromuscular disease lying within the spinal muscular atrophy or primary muscle disease spectra.

4.3.2 INTRODUCTION

Perinatal manifesting neuromuscular diseases include a wide variety of clinical entities such as fetal akinesia/hypokinesia, arthrogryposis multiplex congenita, spinal muscular atrophy (SMA) type I, congenital myopathies and muscular dystrophies, and congenital myasthenic syndromes (Todd et al., 2015; Ravenscroft et al., 2016). There is some degree of clinical overlap between these diseases and also wide genetic heterogeneity, as reflected by the vast number of genes known to be involved and the diversity of mutation types. Genes such as *TTN*, *NEB* and *RYR1* are exceptionally large, adding complexity to the genetic studies. In addition to the clinical evaluation, muscle biopsy, brain magnetic resonance imaging and electromyogram are necessary steps in the diagnostic algorithm, in order to guide the genetic study. However, these exams may not be possible in patients at the severest end of the disease spectrum with a lethal outcome. Such limitations contribute towards the high number of cases that remain elusive in terms of diagnosis. Massive parallel sequencing technology, commonly referred to as next-generation sequencing, now offers novel diagnostic possibilities for such cases. Current genetic research is applying this sequencing technology, especially with applications such as gene panels and whole-exome sequencing (WES), in an effort to solve complex muscle diseases, identifying new genetic causes and expanding mutational profiles (Oliveira et al., 2015; Oliveira et al., 2016).

Here we report one such research study that identified the second known case originating from genetic defects in *ASCC1*, thereby confirming its involvement in a severe congenital neuromuscular disease with bone fractures.

4.3.3 MATERIAL AND METHODS

Human Subjects

This work was approved by the ethics committee of Centro Hospitalar do Porto. Written informed consent was obtained from the participants and legal tutors. Besides the parental samples, biological samples and detailed phenotype information was only available from one of the patients (II.2, Figure 4.3.1A).

Routine genetic tests

The screening of homozygous *SMN1* deletions was performed by PCR of exons 7 and 8 followed by restriction enzyme digestion to differentiate *SMN1* from *SMN2*. Pathogenic expansions in *DMPK* were excluded by PCR and southern-blot. Analysis of *ACTA1, MTM1, TNNT1, TPM2* and *TPM3* was performed by Sanger Sequencing.

Whole-exome sequencing

Whole-exome sequencing (WES) was performed using the Ion AmpliSeq Exome kit (Thermofisher Scientific) as previously described (Oliveira et al., 2015). This strategy uses 12 primer pools (293,903 primer pairs in total) to amplify target regions covering >97% of the consensus coding sequence annotation. Sequencing was performed in the Ion Proton System (Thermofisher Scientific). Reads were aligned in the Ion Torrent server against the human genome assembly hg19. Sequence variants were called using the Torrent Variant Caller plugin version 4.4. Variant filtering resorted to Ion Reporter software and in-house adapted LOVD databases (version v.3.0) that store rare sequence variants and sequencing artifacts detected in our studies. Candidate variants were manually inspected on the Binary Alignment Map (BAM) file through GenomeBrowse V2.1.1 (GoldenHelix Inc.).

ASCC1 targeted variant analysis

Conventional (Sanger) sequencing was carried out in the patient and in both parents to confirm the variant identified by WES in the *ASCC1* locus. Exon 3 was amplified using PCR oligonucleotides: 5'-CTTCCCAGGGTTCCAAGACTG-3' (forward) and 5'-CCACTTCCCGCTGAGTTTCC-3' (reverse). Amplicons were subjected to a further PCR using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermofisher Scientific) and the resulting products were sequenced on a 3130xl Genetic Analyzer (Applied Biosystems). The reference sequence with accession number NM_001198799.2 was used to describe variants according to HGVS nomenclature.

4.3.4 Results

A healthy apparently non-consanguineous Portuguese couple referred to genetic counselling had one late intrauterine fetal death, of unknown cause, at 32 weeks of gestation. This first pregnancy was followed by two live births - one female and one male (Figure 4.3.1A). Both newborns had severe neonatal hypotonia, lack of spontaneous movements (including respiratory apparatus) and died within a few days of life. Detailed clinical and anatomopathological data was available only for the female patient (II.2). She had a "myopathic" appearance, with tent-shaped mouth, microretrognathia and arthrogryposis (Figure 4.3.1B-C), bilateral femoral fractures (Figure 4.3.1D) and thin, gracile ribs (Figure 4.3.1E). Reports on the histological analysis of skeletal muscle, performed in the context of autopsy, describe atrophic muscle fibers (data not shown).
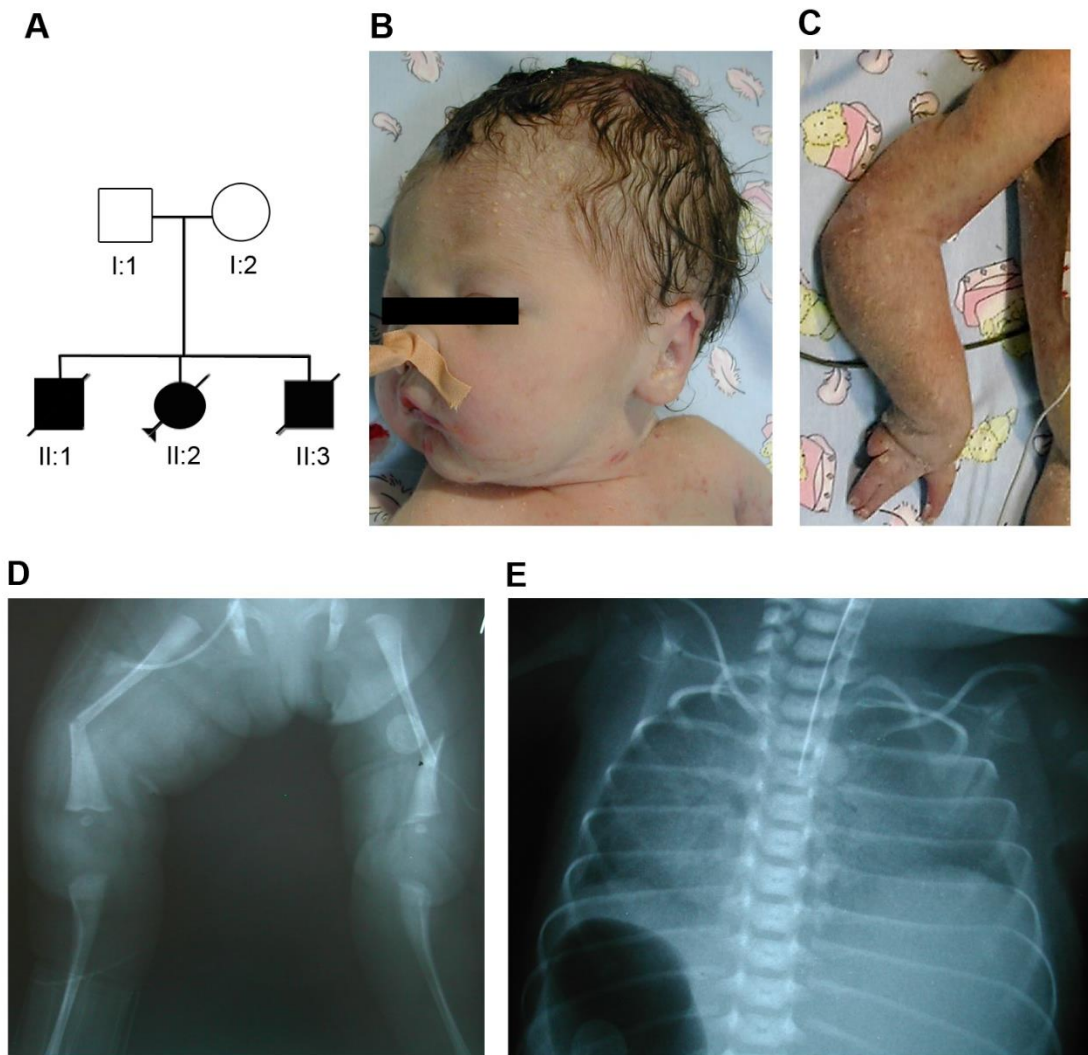
Figure 4.3.1- Family pedigree and patient's main clinical findings. (A) Family pedigree compatible with an autosomal recessive inheritance pattern. (B) Myopathic appearance, including high arched palate, tent-shaped mouth and microretrognathia. (C) Hand showing signs of arthrogryposis. (D) Bilateral congenital femoral fractures. (E) Chest radiograph showing thin, gracile ribs.

Diagnostic workup tests included karyotyping and mutation screening for several genes implicated in different monogenic neuromuscular diseases, namely SMA (MIM: 25330) (*SMN1*, MIM: 600354), myotonic dystrophy (MIM: 160900) (*DMPK*, MIM: 605377) and congenital myopathy genes (*ACTA1*, MIM: 102610; *MTM1*, MIM: 310400; *TNNT1*, MIM: 191041; *TPM2*, MIM: 190990; and *TPM3*, MIM: 191031). These all tested negative.

WES performed on the genomic DNA of patient II.2 targeted a total of 57.7 Mb of target regions. The mean read depth (target base coverage) of these captured regions was 145.6 (fold), with 94.51% of these regions covered by at least 20 reads (metrics determined with

Torrent Suite variant caller plugin). A total of 52,830 variants were called, including 49,681 single and multiple nucleotide variants (SNVs and MNPs) and 3,149 insertion/deletions (indels). For data analysis, variant filtering assumed an autosomal recessive disease model: homozygous variants or two heterozygous variants in the same locus. Variants were excluded if the minor allele frequency was above 0.1% in the population variant databases [dbSNP, 1000 genome project, exome sequencing project (ESP) and exome aggregation consortium (ExAC)]. This high stringency was adopted on the premise that the disease incidence/prevalence would be extremely low. The primary strategy focused on genes known to be related with congenital myopathies (n=24) extended to those linked to muscle diseases (n=146, Appendix III.3 Table III.3.1) and finally to all genes known to be involved in neuromuscular diseases (n=407). No plausible disease-causing variants were identified in any of these sets, and exome analysis proceeded with the interpretation of variants in the remaining *loci*. The extensive list of candidate variants included a homozygous single base pair duplication located in the coding region of *ASCC1* (Figure 4.3.2A). Meanwhile, Knierim and collaborators reported four families with a severe prenatal SMA with respiratory distress and congenital bone fractures (Knierim et al., 2016). All cases showed an autosomal recessive inheritance pattern and were said to be caused by loss-of-function variants in two different genes (*ASCC1* and *TRIP4*), both encoding subunits of the nuclear activating signal cointegrator 1 (ASC-1) transcriptional complex (Knierim et al., 2016). These cases shared several similar features with our patient II.2 (Table 4.3.1). The variant reported in *ASCC1* (NM_001198799.2:c.157dupG, p.Glu53Glyfs*19), identified in the single family of Turkish origin, was exactly the same as that found in our patient. Our WES results were confirmed by Sanger sequencing of *ASCC1* exon 3 (Figure 4.3.2B). Likewise, this region was interrogated in the patient's parents, revealing that both are heterozygotes for the NM_001198799.2:c.157dupG variant (Figure 4.3.2B). No specimens were available for analysis of the patient's siblings (II.1 and II.3).

4.3.5 DISCUSSION

ASC-1 is a complex composed of four proteins encoded by different genes: thyroid hormone receptor interactor 4 (*TRIP4*, MIM: 604501), subunit 1 (*ASCC1*, MIM: 614215), subunit 2 (*ASCC2*, MIM: 614216) and subunit 3 (*ASCC3*, MIM: 614217) of the ASC-1 complex (Jung et al., 2002; Knierim et al., 2016). This poorly understood complex is known to interact with nuclear receptors and other transcription factors (Kim et al., 1999). More recently, based on the domains or motifs found in ASC-1 subunits, it was suggested to be a ribonucleoprotein complex involved in transcriptional coactivation of a wide range of genes and in RNA processing (Knierim et al., 2016).

Although an *ASCC1* variant had been picked up as a possible candidate in our study, information on related diseases or phenotypes was scarce.
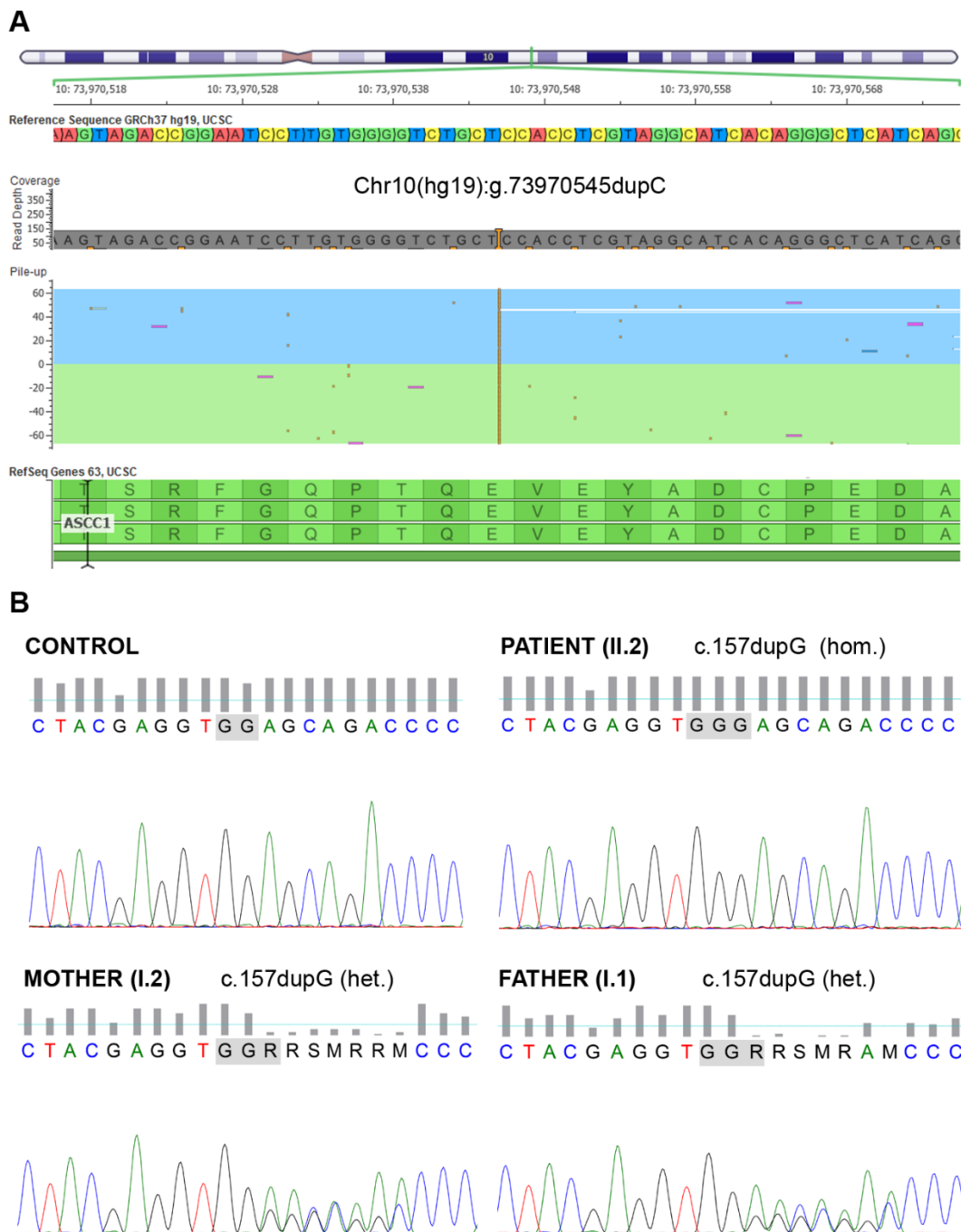
Figure 4.3.2- Identification of NM_001198799.2:c.157dupG variant in ASCC1 gene. (A) Visualization of the frame-shift variant located in ASCC1 in the BAM file obtained from whole-exome sequencing data. (B) Confirmation of variant by Sanger sequencing in the patient (homozygosity) and in both parents (heterozygosity). Family pedigree and patient's main clinical findings.

Table 4.3.1- Clinical features of the patient reported in this work and comparison with published cases with ASC-1 related neuromuscular diseases.

	Family 1 Davignon et al.[14]	Family A Knierim et al.[5]	Family B Knierim et al.[5]	Family C Knierim et al.[5]	Family D Knierim et al.[5]	This report
Number of patients (n)	4	2	2	1	2	1+1[a]
Country	France	Kosovo	Kosovo	Albania	Turkey	Portugal
Locus (cDNA ref. seq.)	*TRIP4* (NM_016213.4)				*ASCC1* (NM_001198799.2)	
Pathogenic Variant	c.890G>A (p.Trp297*) Homo.	c.760C>T (p.Arg254*) Homo.	c.832C>T (p.Arg278*) Homo.	c.760C>T (p.Arg254*) + c.832C>T (p.Arg278*)	c.157dupG (p.Glu53Glyfs*19) Homo.	
Reduced/ absent fetal movements	U	Y	Y	U	Y	Y
Poly/ hydramnios	Y (n=1)	N	N	U	Y	N
Oligohydramnios	N	Y	Y	U	N	N
Premature delivery (<37 w)	U	Y	Y	U	Y	N
Neonatal hypotonia	Y (mainly axial)	Y	Y	Y	Y	Y
Neonatal respiratory distress	Y (n=3)	Y	Y	Y	Y	Y
Congenital bone fractures	N	Y	Y (n=1)	U	Y	Y
Joint contractures	Y (n=2) (mild/mod.) after 14yrs of age	Y (arthrogryposis)	Y (arthrogryposis)	Y (arthrogryposis)	Y (arthrogryposis)	Y (mild arthrogryposis)
High-arched palate	Y (n=1)	Y (n=1)	Y (n=1)	U	N	Y

Table 4.3.1- Clinical features of the patient reported in this work and comparison with published cases with ASC-1 related neuromuscular diseases (continues).

	Family 1 Davignon et al.[14]	Family A Knierim et al.[5]	Family B Knierim et al.[5]	Family C Knierim et al.[5]	Family D Knierim et al.[5]	This report
Micro/ retrognathia	U	Y	Y	U	N	Y
Muscle weakness and atrophy	Y	Y	Y	U	Y	Y
Delayed motor milestones	Y	U	U	U	U	NA
Best motor achievement	Walk (n=2); sit (n=2)	U	U	U	U	NA
Cardiomyopathy	N	Y (n=1)	Y (n=1)	U	N	N
Skin changes	Y (hyperelasticity, dryness, follicular hyperkerastosis)	U	U	U	U	N
Brain imaging	Normal (n=1) (CT scan)	U	U	U	Abnormal cortical gyration (MRI)	N (transfontanel ultrasonography)
Skeletal muscle histology	Fiber size variation. Frequent minicores and cap lesions, few fibers with central nuclei. (n=3)	Fiber size variation and atrophy (n=1). Type 1 fiber grouping.	Fiber size variation and atrophy (n=2). Type 1 fiber grouping.	U	Fiber size variation and atrophy (n=1). Type 1 fiber grouping.	Fiber atrophy (limited analysis, in the context of autopsy)

Footnote: Homo- homozygous; Y- yes; N- no; U- unknown; NP- not performed; NA- not applicable; mod- moderate; a- Patient's sibling not studied (similar phenotype with a neonatal death and fractures); CT- Computed tomography; MRI- Magnetic resonance imaging; W- weeks.

The first disease association reported was a germline *ASCC1* missense variant predisposing to a form of cancer (Barrett esophagus and esophageal adenocarcinoma) (Orloff et al., 2011; van Nistelrooij et al., 2014). The ASC-1 complex was also proposed to be an inhibitor of the nuclear factor kappa B (NF-κB). This transcription factor plays an important role in the development and the function of the immune system (Hayden et al., 2011) and, when inadequately active, is known to be a cause of chronic inflammatory disorders and cancer. In particular, a heterozygous *ASCC1* nonsense variant detected in patients with rheumatoid arthritis was shown to abrogate the NF-κB-inhibiting capacity of *ASCC1*, thereby possibly acting as a risk factor for rheumatoid arthritis and modulating the disease outcome (Torices et al., 2015).

Recently, two independent groups reported a combined total of four families with *TRIP4*-related neuromuscular disease (Knierim et al., 2016; Davignon et al., 2016) and a further family with an *ASCC1*-related form (Knierim et al., 2016). The pathophysiological mechanisms of these newly identified clinical entities arising from pathogenic variants in *ASCC1* and *TRIP4* remain elusive. Ascc1 and trip4 zebrafish morphants have been used to study the ASC-1-related neuromuscular phenotypes (Knierim et al., 2016). In these disease models, where gene expression knockdown is mediated by antisense morpholino oligonucleotides, there was a severe disturbance in neuromotor unit development. More specifically, the axonal outgrowth, neuromuscular junction density and the myotome were found to be compromised (Knierim et al., 2016). However, the second report of additional patients with *TRIP4* mutations, sharing clinical features with the previous patients but without bone fractures, added further speculation on the underlying disease mechanism(s) (Davignon et al., 2016). Ex-vivo and in-vitro experiments performed on murine myoblastic C2C12 cell lines and on muscle primary cultures obtained from patients with *TRIP4* variants, led Davignon and colleagues to propose that ASC-1 might be a key player in muscle development, having a role in late myogenesis and/or myotube growth (Davignon et al., 2016).

Our patient shows clinical features strikingly similar to those of the siblings described by Knierim and co-workers (Knierim et al., 2016), namely: a neonatal (most likely pre-natal) onset, severe hypotonia, lack of spontaneous movements, contractures and congenital bone fractures.

Congenital fractures of long bones have been reported in some perinatal-manifesting neuromuscular diseases, such as SMA and primary myopathies (Ryan et al., 2001; Grotto et al., 2016). These fractures are rare and are attributed to fetal akinesia or hypokinesia, where the decrease of mechanical use affects bone development (Rodríguez et al., 1988). As an isolated clinical finding, bone fractures are a nonspecific clinical feature. However, in conjunction with other signs, such as muscle histopathological markers, it may provide some

diagnostic guidance. One such example is *KLHL40*-related myopathy, where muscle histology enables classification as nemaline myopathy and bone fractures (reported in approximately half of the patients) provides a high likelihood of *KLHL40* involvement (Ravenscroft et al., 2013).

It is noteworthy that, besides sharing a similar phenotype, the variant identified in our patient is exactly the same as that found in the Turkish family reported by Knierim and collaborators (Knierim et al., 2016). There is no known ethnic connection between the two cases. Our patient's parents are both Portuguese and originating from the same village. Homozygosity mapping using WES data of our patient, revealed only two runs of homozygosity, with 1.8 and 0.2 Mb (in chromosomes 11 and 19, respectively) and neither region contained the disease-causing gene. This is not in agreement with high consanguinity, and it is unlikely that the couple shares a recent a common ancestor.

In conclusion, we report the second known case with a pathogenic variant in *ASCC1*. This independent description provides the confirmation that homozygous truncating variants in *ASCC1* give rise to a severe neuromuscular disease, possibly within the SMA or primary muscle diseases spectra. As for future research, it will be important to identify additional ASC-1-related patients, especially those with milder disease presentations, so as this should enable further characterization of this new clinical entity.

4.4 EXONIZATION OF AN INTRONIC LINE-1 ELEMENT CAUSING BECKER MUSCULAR DYSTROPHY AS A NOVEL MUTATIONAL MECHANISM IN DYSTROPHIN GENE

GONÇALVES A*, OLIVEIRA J*, COELHO T, TAIPA R, MELO-PIRES M, SOUSA M, SANTOS R. (2017). GENES (BASEL), 8(10). PII: E253.

* THESE AUTHORS CONTRIBUTED EQUALLY TO THIS WORK.

4.4.1 ABSTRACT

A broad mutational spectrum in the dystrophin (*DMD*) gene, from large deletions/duplications to point mutations, causes Duchenne/Becker muscular dystrophy (D/BMD). Comprehensive genotyping is particularly relevant considering the mutation-centered therapies for dystrophinopathies. We report the genetic characterization of a patient with disease onset at age 13 years, elevated creatine kinase levels and reduced dystrophin labeling, where multiplex ligation-dependent probe amplification (MLPA) and genomic sequencing failed to detect pathogenic variants. Bioinformatic, transcriptomic (real time PCR, RT-PCR), and genomic approaches (Southern blot, long-range PCR, and single molecule real-time sequencing) were used to characterize the mutation. An aberrant transcript was identified, containing a 103-nucleotide insertion between exons 51 and 52, with no similarity with the *DMD* gene. This corresponded to the partial exonization of a long interspersed nuclear element (LINE-1), disrupting the open reading frame. Further characterization identified a complete LINE-1 (~6 kb with typical hallmarks) deeply inserted in intron 51. Haplotyping and segregation analysis demonstrated that the mutation had a de novo origin. Besides underscoring the importance of mRNA studies in genetically unsolved cases, this is the first report of a disease-causing fully intronic LINE-1 element in *DMD*, adding to the diversity of mutational events that give rise to D/BMD.

4.4.2 Introduction

Duchenne or Becker muscular dystrophies (D/BMD), caused by pathogenic variants in the Dystrophin (DMD) gene, are among the most common inherited diseases of muscle, with an estimated prevalence of ~1/3800 live male births (Mendell et al., 2012). A broad mutational spectrum for D/BMD has been thoroughly described in the literature, ranging from large multi-exonic deletions/duplications to smaller single nucleotide variants (Flanigan, 2014). More complex and rarer DMD mutations, such as large rearrangements and gene disruption mediated by retrotransposition activity, have also been reported (Solyom et al., 2012; Ishmukhametova et al., 2013). The genetic heterogeneity, size, and complexity of the DMD gene demands expertise in a vast number of molecular techniques, besides the routinely used multiplex ligation-dependent probe amplification (MLPA) and genomic sequencing. Since dystrophinopathies are now amenable to therapy, the genetic characterization of these patients has gained relevance beyond clinical follow-up and genetic counselling purposes.

We previously reported the characterization of 308 dystrophinopathy patients, from 284 unrelated families, leading to the identification of 175 distinct mutations (Santos et al., 2014). This 91% positivity rate (284 of 312 families) was achieved in a cohort with strict inclusion criteria. Since then, and considering all referrals with clinical suspicion of D/BMD, over 100 cases remain unsolved at the genetic level.

This report describes a wide combination of genetic studies performed on a patient presenting a mild BMD phenotype, where a unique mutational event involving the insertion of a long interspersed nuclear element 1 (LINE-1) was identified.

4.4.3 Material and methods

Patient Samples

Formal written informed consent for publication of this case report was obtained from the patient and other family members whose data is presented. The study was conducted in accordance with the Declaration of Helsinki, and with approval of the institutional (CHP) ethics committee (Code: 336-13(196-DEFI/285-CES); date of approval: 11 December 2013).

RNA Studies

Total RNA extracted from patient and control muscle samples with PerfectPure RNA Fibrous Tissue kit (5 PRIME) was converted to cDNA using the High Capacity RNA-to-cDNA Kit (Thermo Fisher Scientific, Waltham, MA, USA). DMD transcripts were amplified by PCR

covering the entire coding region. Amplicons were purified with Illustra ExoStar (GE Healthcare, Little Chalfont, UK) and sequenced using BigDyeTM Terminator Cycle Sequencing Kit V3.1 (Thermo Fisher Scientific). Reference sequence for variant description: NM_004006.2.

Bioinformatics

Genome similarity sequence search was conducted using the Basic Local Alignment Search Tool (BLASTN v2.2.32) (Johnson et al., 2008). Analysis of repetitive elements was performed using CENSOR (Kohany et al., 2006), RepeatMasker (http://www.repeatmasker.org/) and L1Xplorer (Penzkofer et al., 2005). Potential acceptor splice-sites and branchpoints were assessed using different algorithms available in Alamut Visual software (v2.8, Interactive Biosoftware, Rouen, France) (Appendix III.4 Figure III.4.1).

LINE-1 Characterization

In order to identify the 5′ insertion site, primers were designed against five candidate target regions within intron 51 (Appendix III.4 Data III.4.1). For the 3′ insertion site, a forward primer was designed against a conserved region of the LINE-1 3′UTR (L1-F: AAATTAGGTATTGATGGGACGTATT) and a reverse primer within intron 51 (51int-R: GAGAAGATGACAGTTAAATCAAAGC) (Appendix III.4 Data III.4.1). Resultant amplicons were sequenced as described above. LINE-1 was genotyped by single molecule real-time sequencing (PacBio RS II system, Pacific Biosciences, San Francisco, CA, USA) using custom DNA libraries (Appendix III.4 Data III.4.1). FASTA/Q files were mapped against a LINE-1 reference and consensus sequence was obtained from BAM files using Samtools mpileup command (Appendix III.4 Data III.4.1). Sequence artifacts and ambiguous sites were clarified via long-range PCR followed by Sanger sequencing.

4.4.4 Results

We report the genetic characterization a 50-year-old male patient with clinical features compatible with BMD, namely, onset at 13 years of age with progressive proximal weakness of lower limbs, electromyography showing myopathic signs, and high creatine kinase levels. The patient was initially referred for DMD molecular testing in 2001, but multiplex PCR and Southern blot failed to detect pathogenic variants. The case was re-evaluated in the context of genetic counselling of the patient's daughter. A muscle biopsy performed in the patient revealed dystrophic features and irregular staining for Dystrophin (Figure 4.4.1).
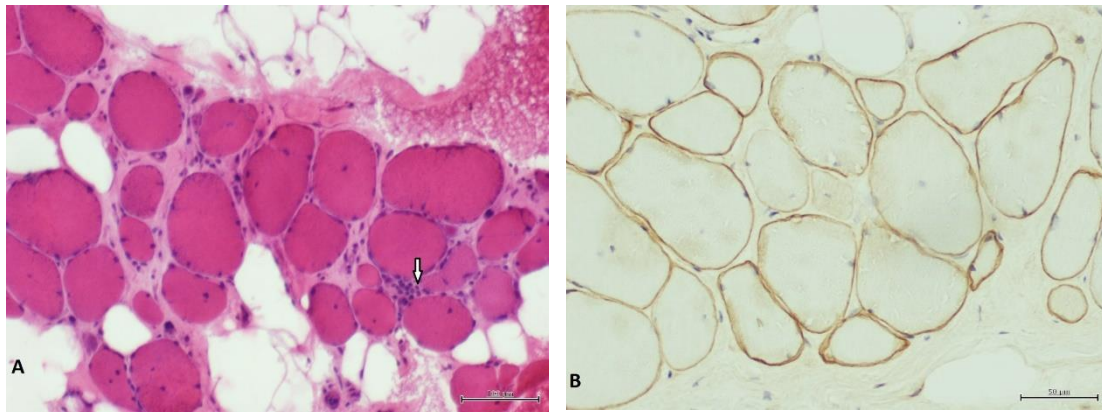
Figure 4.4.1- Patient's muscle biopsy. (A) Hematoxylin and eosin stain showing severe fibrosis and fat substitution. Arrow indicates a necrotic fiber. Scale bar corresponds to 100 μm. (B) Dys2 antibody showing irregular and faint staining of dystrophin. Scale bar corresponds to 50 μm.

Given that large deletions/duplications and point mutations were not detected by current routine genetic studies (MLPA and DMD genomic sequencing), complete DMD cDNA analysis of the muscle specimen was performed. An insertion of 103 nucleotides (not traceable in DMD) was identified between exons 51 and 52 (r.7542_7543ins(103), Figure 4.42A), predictably shifting the DMD open reading frame (ORF). Besides the predominant mutated transcript, a residual amount of wild-type transcript was detectable (Figure 4.4.2A). To identify the origin of the mutated sequence, a BLASTN query was performed against the human nucleotide collection. Identity of 95.1% (98/103 base pairs (bp)) was retrieved against two human LINE-1 sequences: L1.21 and L1.14 (GenBank accession numbers U93570 and U93566, respectively) (Appendix III.4 Data III.4.2). Comparative analysis also showed high similarity with other human genomic sequences containing LINE-1 elements. Further confirmation was obtained using CENSOR software where 98/103 bp had 100% similarity with the consensus sequence of the human LINE-1 element (Appendix III.4 Figure III.4.2).

A strategy was delineated to identify the genomic insertion site in the DMD gene. The first five nucleotides (AATTC), having no correspondence to the LINE-1 consensus sequence, were presumed to belong to intron 51 of DMD. This ~44 Kb intron was scanned for the sequence AG/AATTC (AG being the canonical dinucleotide for acceptor splice-sites); a total of eight such sequences were found. Composite splice-site analysis narrowed this down to five potential sites—those presenting high splice-site scores and suitable (cryptic) branch-points located nearby (Appendix III.4 Figure III.4.1).
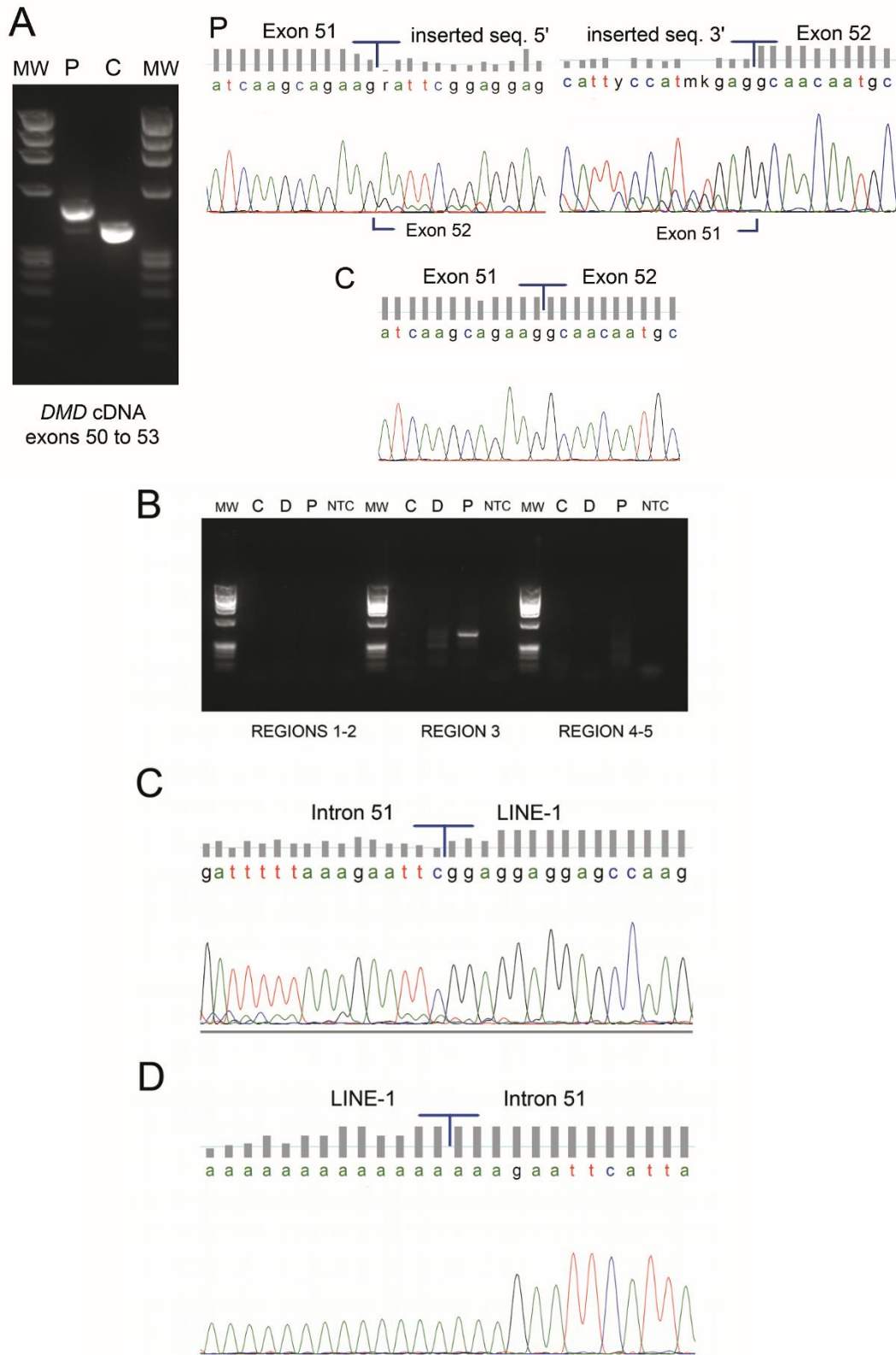
Figure 4.4.2- (A) cDNA analysis of DMD transcript revealed an abnormal PCR product with higher molecular weight in the patient (P) which was not detectable in a control sample (C). Sequencing electropherogram shows an insertion of 103 nucleotides between exons 51 and 52. A residual wild-type transcript is present in the patient sample (faint PCR band).

Figure 4.4.2- (continues). (B) PCR amplification of five candidate regions to identify the long interspersed nuclear element (LINE-1) insertion site in intron 51 of DMD (C—control, D—control with a deletion of intron 51, P—patient, MW—molecular weight marker, NTC—no template control). A single amplicon was obtained in the patient sample for the candidate region 3. (C) Upon sequencing, the exact location of the LINE-1 in intron 51 was mapped to lie between positions c.7542+8951 and c.7542+8952 (NM_004006.2). (D) The 3′ insertion site was confirmed through a specific PCR followed by sequencing; the LINE-1 poly-A tail is also seen.

To identify the LINE-1 5′ insertion site by PCR, three forward oligonucleotides were designed to encompass these five regions of interest and a single reverse oligonucleotide annealing to the known inserted LINE-1 sequence (detected by cDNA analysis). PCR experiments and subsequent sequencing showed that the LINE-1 was inserted at position NM_004006.2:c.7542+8951_c.7542+8952 of intron 51 (Figure 4.4.2B and C). The 3′ end was then identified, containing a poly-A tail and a stretch of 9 bp (AAAGAATTC) consistent with a flanking target site duplication (TSD) (Figure 4.4.2D). Southern blot and hybridization was performed to estimate the size of insertion. Results revealed a ~6 Kb size increase, thus corresponding to a complete or almost complete LINE-1 element (Appendix III.4 Figure III.4.3).

The patient's daughter was seen to be a carrier of the LINE-1 insertion mutation. Additional family members (the patient's healthy brother and two sisters presenting at-risk haplotypes) were also screened for the mutation using a LINE-1-specific PCR (Figure 4.4.3). Only the patient´s daughter tested positive, suggesting a de novo event.

Further genotyping confirmed that a full-length LINE-1 was present (sequence available in Appendix III.4 Data III.4.2 and submitted to GenBank, accession number MF421743). L1Xplorer and RepeatMasker tools classified the element as a member of the L1HS subfamily, as it had all the typical hallmarks of these retrotransposons: a 5′-untranslated region (UTR), two non-overlapping ORFs (ORF1 and ORF2), a short 3′UTR and a poly-A tail (Figure 4.4.4).
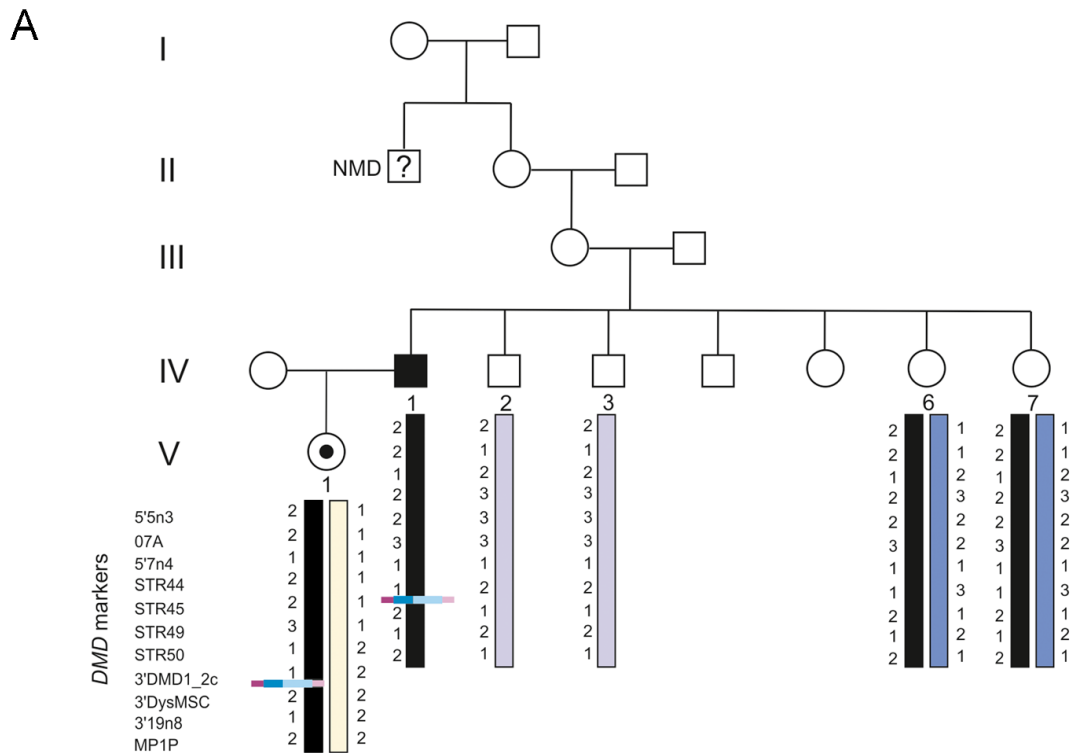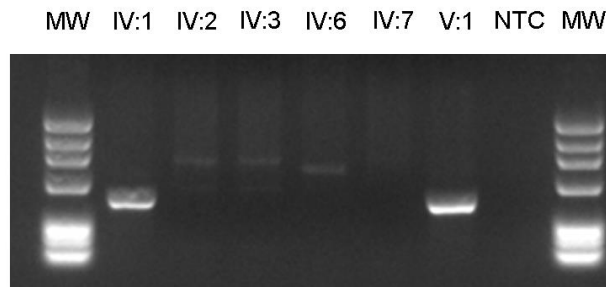
Figure 4.4.3- (A) Patient's family tree highlighting the segregation of the LINE-1 insertion and *DMD* haplotyping. This insertion is present in the patient (IV:1) and his daughter (V:1), and not detectable in the patient's sisters (IV:6 and IV:7 both carriers of the same at-risk haplotype). Interestingly, this family was initially thought to have an X-linked transmission, since one of the patient's maternal great-uncles (deceased) was suspected to have a neuromuscular disease (NMD). (B) LINE-1 specific PCR used to screen for additional carriers.

4.4.5 DISCUSSION

LINE-1s are the most abundant type of retrotransposable elements, accounting for nearly 17% of the human genome (Kim et al., 2012). Typically, they are ~6 kb in length and exhibit characteristic components (Figure 4.4.4). ORF1 encodes an RNA-binding protein while

ORF2 encodes a protein with endonuclease and reverse transcriptase activity. Although the transcriptional mechanism of LINE-1 is not fully understood, it has been proposed to involve target-site primed reverse transcription. The cDNA originated by this process recombines with genomic DNA, giving rise to characteristic signatures: a 7–20 bp direct repeat of the endonuclease target flanking the inserted LINE-1 (TSD) (Hancks & Kazazian, 2016). Only 80–100 LINE-1s in the human genome (0.1% of total) are believed to be capable of active retrotransposition (Hancks & Kazazian, 2016).
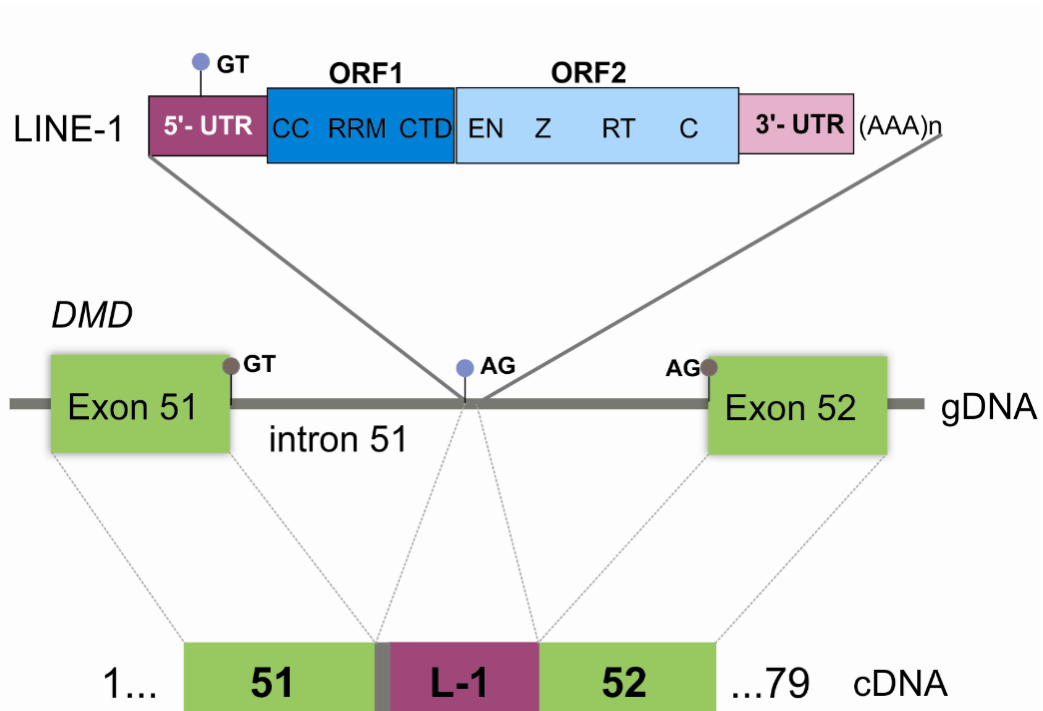
Figure 4.4.4- Graphical representation of the pathogenic LINE-1 insertion. At the genomic level (gDNA), the integration site is located in intron 51 of DMD. LINE-1 is displayed showing its characteristic features: 5′UTR, open reading frame (ORF) 1 (CC—coiled coiled domain, RRM—RNA recognition motif, CTD—C-terminal domain), ORF2 (EN—endonuclease, Z domain, RT—reverse transcriptase, C—cysteine-rich), and 3′ UTR with a poly-A tail. At the cDNA level, the retention of five base pairs from intron 51 (grey box), and the partial exonization of LINE-1 (L-1) (dark pink box) is explainable by the recognition of cryptic splice-sites (light blue circles) located in intron 51 (acceptor splice-site) and in the LINE-1 sequence itself (donor splice-site). Gray circles correspond to the canonical splice-sites.

Despite their importance in evolution and genome diversity, the insertion of a LINE-1 within a gene could have a deleterious effect, giving rise to disease. To date, only 30 such insertions have been reported, the majority located in exonic regions and causing frame-

shifts or exon skipping (Narita et al., 1993). In contrast, intronic LINE-1 insertions are rarely reported.

Regarding the DMD gene, only five pathogenic insertions have been described. Exonic disruptions, giving rise to a DMD phenotype, have been reported twice in exon 44 (Narita et al., 1993; Musova et al., 2006), and also in exons 48 and 67 (Holmes et al., 1994; Awano et al., 2010). A further pathogenic insertion was detected in two unrelated Japanese families with X-linked dilated cardiomyopathy, where a 5′-truncated form of a LINE-1 was integrated in the *DMD* 5′UTR, thought to affect the transcription or the stability of muscle transcripts (Yoshida et al., 1998). A different repetitive element (Alu-like) was also reported to cause dilated cardiomyopathy, activating a cryptic acceptor splicing site in intron 11 of DMD (Ferlini et al., 1998). The mutational event in our patient is completely distinct in two aspects: it is a deep-intronic insertion and a full LINE-1 sequence is present. This LINE-1 was classified as a member of the L1HS subfamily, responsible for the majority of the documented LINE-1 retrotransposition events. Our results showed its partial exonization at the cDNA level, due to the recognition of a cryptic 3′ splice-site located in intron 51 and a 5′ splice-site within this element (Figure 4.4.4). This presumably gives rise to a truncated polypeptide (p.Ala2515Asnfs*21). The presence of a residual wild-type transcript explains the patient's milder dystrophinopathy (BMD phenotype). This LINE-1 sequence has 100% identity with a LINE-1 element located in chromosome 2 (GenBank accession number AC216112), which could constitute its original source. It has a near-complete identity (only one bp difference) with another pathogenic LINE-1 (GenBank accession number AF149422) inserted in the hemoglobin-beta locus; a seemingly active retrotransposable element of the human genome (Kimberland et al., 1999).

Intronic LINE-1 insertions causing exonization have only been described in three other cases: chronic granulomatous disease (*CYBB* gene) (Meischl et al., 2000), Chanarin-Dorfman syndrome (*ABHD5* gene) (Samuelov et al., 2011), and familial retinoblastoma (RB1 gene) (Rodriguez-Martin et al., 2016). The rarity of LINE-1-mediated pathogenic insertions described in the literature and in variant databases is attributable mostly to their low activity throughout the genome and to the technical difficulties in their detection (especially the full-length insertions). With massive parallel sequencing technology, there are also considerable limitations, particularly through short reads sequencing, and here, tailored bioinformatics analysis tools and strategies are required (Tica et al., 2016). In the case of intronic LINE-1 insertions, detection may be hampered by the intron's length and the fact that it mainly affects transcriptional events (e.g., intronic retentions or exonization events). One possibility is therefore to conduct mRNA studies in yet uncharacterized B/DMD patients, as previously suggested (Santos et al., 2014; Tuffery-Giraud et al., 2017).

It is known that repetitive transposable elements such as short interspersed nuclear elements (e.g. Alu sequences) or LINE sequences are frequent in DMD intronic regions. However, besides being the underlying cause of some gross rearrangements, they may also influence gene expression by mediating alternative splicing. One can thus speculate that this would ultimately interfere with the efficacy of RNA-based therapies, especially those designed to restore the reading frame. To our knowledge, this is the first report of a deep-intronic insertion of a LINE-1 element in the DMD gene shown to cause disease. Besides its scientific relevance, while expanding the mutational mechanisms underlying B/DMD, this finding also reinforces the need to develop comprehensive to identify LINE-1 insertion profiles in the human genome.

4.5 Homozygosity mapping using whole-exome sequencing: a valuable approach for pathogenic variant identification in genetic diseases

4.5.1 Abstract

In the human genome, there are homozygous regions presenting as sizeable stretches, or 'runs' of homozygosity (ROH). The length of these ROH is dependent on the degree of shared parental ancestry, being longer in individuals descending from consanguineous marriages or those from isolated populations. Homozygosity mapping is a powerful tool in clinical genetics. It relies on the assumption that, due to identity-by-descent, individuals affected by a recessive disease are likely to have homozygous markers surrounding the disease locus. Consequently, the analysis of ROH shared by affected individuals in the same kindred often helps to identify the disease-causing gene. However, scanning the entire genome for blocks of homozygosity, especially in sporadic cases, is not a straight-forward task. Whole-exome sequencing (WES) has been shown to be an effective approach for finding pathogenic variants, particularly in highly heterogeneous genetic diseases. Nevertheless, the huge amount of data, especially variants of unknown clinical significance, and the presence of false-positives due to sequencing artifacts, makes WES analysis complex. This paper briefly reviews the different algorithms and bioinformatics tools available for ROH identification. We emphasize the importance of performing ROH analysis using WES data as an effective way to improve diagnostic yield.

4.5.2 INTRODUCTION

Mendelian diseases are caused by pathogenic variants in genes that follow the biological inheritance laws originally proposed by Gregor Mendel. The disease-causing gene may be in an autosome or in a sex-chromosome, and may be dominant or recessive. Individually, these diseases are considered to be rare, but collectively they occur at a high rate, with an estimated 7.9 million children being born annually with a serious birth defect of genetic origin (Christianson et al., 2006).

Sanger sequencing has been the gold standard in molecular diagnostics of Mendelian diseases and is still the first choice to confirm a suspected diagnosis, enabling accurate genetic counselling. However, for diseases with genetic heterogeneity, such as hereditary myopathies or primary ciliary dyskinesia, gene-by-gene Sanger sequencing is not the most cost-effective or efficient approach (Oliveira et al., 2015; Pereira et al., 2015). Technological advances over the past decade has led to the development of high-throughput sequencing platforms, revolutionizing the sequencing capabilities and boosting the use of this so called "next-generation sequencing" (NGS) both in research and in clinical diagnostic settings. With this technology, the human genome can be completely sequenced, allowing the simultaneous analysis of multiple genes. The application of NGS in Mendelian diseases focuses firstly on exonic regions of DNA, as the majority of disease-causing mutations are found in exons or in the flanking intronic regions (Ng et al., 2010).

Marriage between close biological relatives increases the probability of the offspring inheriting two deleterious copies of a recessive gene. Thus, children from consanguineous couples have a higher incidence of autosomal recessive disorders (Bittles, 2001). In addition, as alleles are parts of haplotypes, not only will the affected descendant have two identical copies of the ancestral allele, but the surrounding DNA segment (haplotype) will also be homozygous. Thus, the child will be homozygous for that segment, the so-called runs of homozygosity (ROH) (McQuillan et al., 2008). These homozygous segments that are identical by descent (IBD) are generally longer in consanguinity cases. Nevertheless, even in the absence of known recent inbreeding, ROH can be detected in geographically isolated populations and historical bottlenecks (Pemberton et al., 2012). The ROH length is dependent on the degree of shared parental ancestry and its age. Recent inbreeding events/parental consanguinity tend to have longer ROH (measuring tens of Mb) since there are fewer recombination events interrupting the segments that are IBD. Conversely, older ROH are generally much shorter because the homozygous stretches have been split down by repeated meiosis over the generations, with the exception of genomic regions where the recombination rates are lower (McQuillan et al., 2008).

4.5.3 HOMOZYGOSITY MAPPING

Homozygosity mapping (also known as autozygosity mapping), consists in the identification of homozygous regions in the genome. This is a powerful strategy to associate new genes to diseases (Alkuraya 2010; Goodship et al., 2000). As mentioned, affected individuals are likely to have two IBD alleles at markers located in the vicinity of the disease locus and thus will be homozygous for these markers. This method relies on the search for ROH that are shared by affected individuals in the same family. However, scanning the genome for blocks of homozygosity, although simple and efficient, requires sophisticated techniques such as the use of numerous microsatellite markers or high-density single nucleotide polymorphism (SNP) genotyping. For most autozygosity mapping projects, multipoint linkage analyses under a recessive disease model have been performed, with software such as GENEHUNTER (Kruglyak et al., 1996), SIMWALK2 (Sobel et al., 2002), MERLIN (Abecasis et al., 2002) or ALLEGRO (Gudbjartsson et al., 2000).

In general, haplotypes are inspected manually for homozygous regions that are shared by all affected individuals and can be inferred to be IBD if genotypes from the parents or other close relatives are available. But, in practical terms, conventional parametric methods of multipoint linkage analysis for large datasets in complex consanguineous families are often difficult, because of the time and computational power required, since homozygous blocks among affected individuals tend to be large (mean 4.4 Mb) and contain dozens or hundreds of genes. Moreover, genotyping errors of SNP array platforms and poorly represented chromosomal regions also limit the potential of this technology. If ROH are incorrectly mapped by the introduction of erroneous heterozygous genotypes, the analysis of the causative gene will be compromised.

4.5.4 NEXT-GENERATION SEQUENCING

NGS, and in particular whole-exome sequencing (WES), prompted great advances in the study of genetic diseases and gene discovery (Boycott et al. 2013; Ng et al., 2010). However, this technology has still some limitations (Sirmaci et al., 2012). Firstly, some deleterious variations may be in non-coding regions, which cannot be detected by WES. Genetic and phenotypic heterogeneity in affected individuals makes exome sequencing difficult to interpret. Sequencing errors related to poor capture efficiency, mechanical and analytical errors, as well as misalignment of repetitive regions, lead to erroneous results. These hamper the analysis and impose the need to validate candidate causative variants by Sanger sequencing. Moreover, WES analysis imposes the need of applying filtering strategies. This step is critical as it will limit data analysis and consequently influence the

results. For instance, almost half of the variants may be excluded for being synonymous. Despite the fact that these are usually not considered deleterious, numerous synonymous mutations have been implicated in human diseases (Sauna & Kimchi-Sarfaty, 2011). In addition, during WES analysis a frequency filter is usually set to exclude variants with minor allele frequency above a certain threshold (above 1% in most cases). This threshold should be set in accordance with the expected prevalence of the disease. Even so, considering that in recessive disorders carriers do not show any signs of the disease, the frequency of damaging alleles in populational variant databases can still be higher than the established threshold. This would lead to the erroneous exclusion of this variant during filtering.

Homozygosity Mapping using WES Data

Recent studies have clearly demonstrated the power and the effectiveness of applying homozygosity mapping to WES data, in an attempt to identify causative genes for Mendelian disorders (Gillespie et al., 2014; Shamseldin et al., 2015). This approach has the advantage of unraveling the causal variant irrespective of the gene involved. The homozygosity map allows narrowing down of the target data sets; examination at the base-pair level then enables identification of candidate causative variants.

This approach would start by mapping WES reads against a reference genome (human hg19 in our examples). Data derived from whole-genome and RNA sequencing can also be used. This step is usually performed through a bioinformatic pipeline that generates SAM/ BAM (Sequence /Binary Alignment Map) and, at the end of the process, a variant call format (VCF) file. These files are then used as input for homozygosity mapping analysis. The position and zygosity of the obtained sequence variants can be used to retrieve/infer ROH regions. As all bioinformatic approaches, there is an error rate associated that is difficult to estimate, since it is highly dependent on the WES metrics and the sequencing platform used.

Table 4.5.1 lists some of the tools available to perform ROH. For instance, the web-based tool HomozygosityMapper (Seelow & Schuelke 2012), allows users to interactively analyze NGS data for homozygosity mapping. Furthermore, PLINK (Purcell et al., 2007) and GERMLINE (Gusev et al., 2008), originally developed for the analysis of SNP array data, are tools based on sliding-window algorithms. In a sliding window analysis, the statistics are calculated for a small frame of the data. The window incrementally advances across the region of interest and, at each new position, the reported statistics are calculated. In this way, chromosomes are scanned by moving a window of a fixed size along their entire length and variation in genetic markers across the region of interest can be measured. This type of analysis reveals how variation patterns change across a surveyed genomic segment (Srinuandee & Satirapod 2015).

Table 4.5.1- List of bioinformatic tools developed for ROH detection and their main features.

Software	OS	UI	Algorithm	Main features / experimental design	Input data files	ROH size range (Mb)	Other features	Ref.
Homozygosity-Mapper	Unix/Linux (web server [a])	GUI	Detection of homozygous blocks of selectable length	Perform autozygosity mapping from SNP arrays and NGS data	VCF files and SNP geno-types	> 1.5	•Independent of parameters like family structure/allele frequencies. •Robust against genotyping errors. •Integrated with GeneDistiller (candidate gene search engine).	Seelow et al., 2009
PLINK	Unix/Linux, Mac OS, Win.	CLI & GUI	Sliding-window	WGAS analysis tool set. Estimation and use of IBD in the context of population-based studies	BED, PED, and FAM files	[0.5 - 1.5] > 1.5	•Makes a variety of standard association tests. •Maps disease loci that contain multiple rare variants in a population-based linkage analysis. •Integrated with Haploview.	Purcell et al., 2007
GERMLINE	Unix/Linux	CLI	Sliding-window	Designed for genome-wide discovery of IBD segments shared within large populations (SNP arrays)	PED, MAP and Hap-map files	> 1.5	•Overcomes the computational barrier of pairwise analysis and can scale the analysis linearly with the sample size.	Gusev et al., 2008

Table 4.5.1- List of bioinformatic tools developed for ROH detection and their main features (continues).

HomSI	Unix/Linux, Mac OS, Win.	GUI	Sliding-window	Identify ROH in consangui-neous families from NGS data	VCF files	> 1.5	• Takes into account the distribution of the variants within genomic coordinates. • Reported to be consistent with data derived from SNP microarrays.	Gormez et al., 2014
H³M²	Unix/Linux	CLI	Heterogeneous hidden Markov model	Analyse medium and short ROH obtained from WES data	BAM files	< 0.5 [0.5-1.5] > 1.5	• Reported to be more accurate than GERMLINE and PLINK, especially in the detection of short and medium ROHs.	Magi et al., 2014
Agile-Genotyper and Agile-VariantMapper	Win.	GUI	User-controllable visualization of homozygous regions	ROH analysis WES data and SNP genotyping file data	SAM; tab-delimited text files	> 1.5	• AgileVariantMapper uses the genotypes of all positions found to be polymorphic. • AgileGenotyper deduces genotypes from positions previously found to be polymorphic in the 1000 Genomes Project data set.	Carr et al., 2013

Footnote: [a]-http://www.homozygositymapper.org; CLI- command-line interface; GUI- graphical user interface; NGS- Next-generation sequencing; OS- operative system; Ref.- references; ROH- runs of homozygosity; SNP- single nucleotide polymorphism; UI- user interface; VCF- variant call format; WES- whole-exome sequencing; WGAS- whole-genome association studies; Win.- Windows.

EXome-HOMozygosity is an example in which a sliding-window algorithm (PLINK) is applied for WES-based ROH detection (Pippucci et al., 2011). However, the sliding-windows approaches cannot be used easily with short/medium ROH sizes. In order to solve this issue, Magi et al proposed a new algorithm, H3M2, that is capable of detecting smaller ROH (Magi et al., 2014). AgileGenotyper (Carr et al., 2013) and HomSI (Gormez et al., 2014) are other tools that can be used for the graphical visualization of ROHs.

Case Studies

The following examples are shown to elucidate the relevance and limitations of WES-based ROH. Two patients listed in Table 4.5.2 have recessive conditions caused by homozygous pathogenic variants identified by WES. There was no indication of parental consanguinity, and patient P2 had a sibling affected by the same condition. In both cases, retrospective analysis was based on genome-wide homozygosity mapping performed using HomozygosityMapper algorithm.

Table 4.5.2- Cases selected to illustrate the use of homozygosity mapping using WES data in patients with rare genetic diseases.

Patient	Phenotype	Genotype	Reference
P1, male adult	Infertility due to total sperm immotility. Clinical features compatible with Kartagener syndrome. No known parental consanguinity.	Homozygous pathogenic variant in *CCDC103* gene: Chr17(h19):g.42978470G>C	Pereira et al., 2015.
P2, female neonate	Severe neuromuscular disease, congenital hypotonia, respiratory distress and bone fractures. No known parental consanguinity.	Homozygous pathogenic variant in *ASCC1* gene: Chr10(hg19):g.73970545dupC	Oliveira et al., 2017

Data obtained from patient P1 revealed the presence of two long stretches of homozygous SNPs in chromosomes 1 and 17 (with 19 and 17 Mb, respectively) (Figure 4.5.1, top panel). The affected gene (*CCDC103*) is located in one of these regions, specifically in chromosome 17. In this example the proposed approach would be suitable to generate a considerably shorter list of candidate loci, and consequently reduce the number of variants to be evaluated in terms of their pathogenicity. In the second example the same strategy was applied in a patient with a rare neuromuscular disease. Results for patient P2 revealed two smaller ROH (1.8 and 0.2 Mb) in chromosomes 11 and 19 respectively (Figure 4.5.1, bottom panel). However, the *ASCC1* gene, carrying the disease-causing variant, is located in chromosome 10. In this last example, and considering the algorithm used, the

analysis based on the assumption of a homozygous variant located in a ROH would clearly fail. The causal gene would not be included in the list of candidate genes, most likely due to the size of the actual ROH where the genetic defect is located. In the first example, the long ROH can be attributed to consanguinity (although it was not formally confirmed), while in the second case, a distant common ancestor could explain the presence of a rare pathogenic variant in a smaller ROH tract.
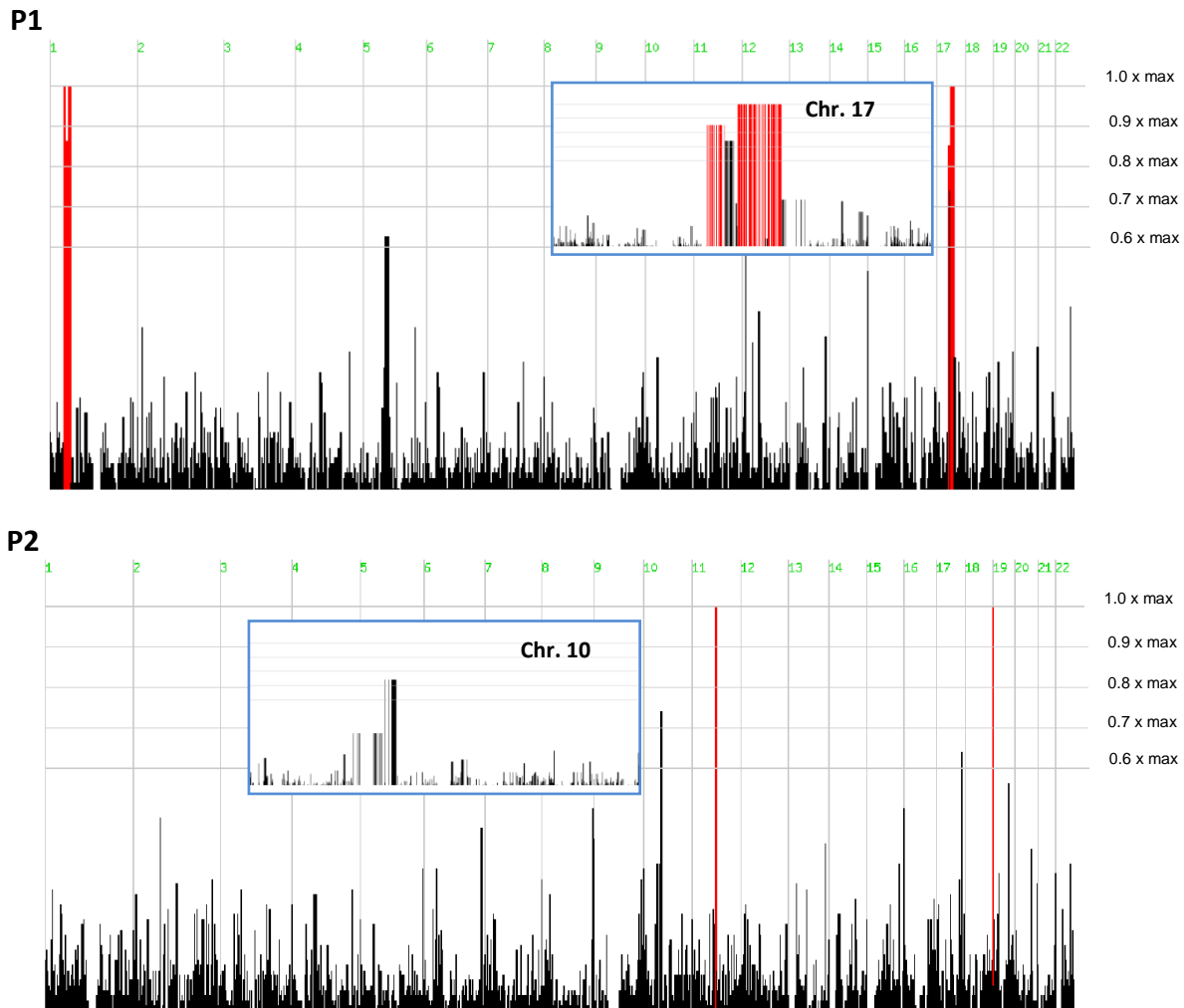
Figure 4.5.1- Genome-wide homozygosity mapping using WES and the HomozygosityMapper software. Results are shown for patients P1 and P2. Longer ROH are shown in red color. Inserted blue boxes show the genomic regions in more detail where the disease-related gene is located.

4.5.5 CONCLUSIONS

The aim of this position paper is to revisit homozygosity mapping as an important tool for clinical genetics in cases where a recessive disease is suspected. We have reviewed the

proposed algorithms and available bioinformatic tools designed for ROH detection based on WES data. Selection of appropriate algorithms should mainly consider specific features of the case under study, such as the genetic context, the ROH size and the number of relatives affected by the same condition.

Monogenic disorders have been studied by classical approaches aiming to unravel several disease-causing genes. The presence of numerous genes in a candidate genomic region was a limiting factor, considering the costs and the time required, to screen mutations by Sanger sequencing. Another limitation with the "traditional" approaches is more evident when they are used to study trios (only the parents and their affected child) or larger families with only one or two affected members. The genotype data extracted in these familial contexts generally remain statistically insufficient for classical analytical approaches (Jorde, 2000).

As compared with conventional homozygosity mapping that uses known SNPs, WES has the added advantage of allowing the identification of the actual disease-causing variant. Therefore, instead of using two different procedures, one for identifying candidate loci and the other to identify the genetic defect itself at the nucleotide level, both can be performed in a single step by WES. Nonetheless, there are still limitations and further bioinformatic developments are required. Considering the examples presented, there are sensitivity issues especially if the genetic defect is in a small ROH. Finally, we consider that it would be useful to develop a bioinformatic tool that combines variant filtering and homozygosity mapping (Appendix IV.5 Figure IV.5.1), which currently need to be performed separately.

Chapter 5

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Contents

5.1 Concluding Remarks

5.2 Future Perspectives

5.1 CONCLUDING REMARKS

5.1.1 GENETIC HETEROGENEITY AND ALLELIC DIVERSITY IN CONGENITAL MUSCLE DISEASES

This research contributed to the characterization of the genetic profiles of a cohort of patients with congenital muscle diseases. The results shown in this thesis **(Chapter 3.1)** are suggestive towards a considerable mutational heterogeneity, particularly for CM and CMD. This heterogeneity can be measured in two different dimensions: i) genetic heterogeneity and ii) allelic diversity. Regarding genetic heterogeneity, this is related to variants in different genes that translate into similar phenotypes. This is widely recognized considering the extended lists of *loci* causing disease in CM (n=32) and CMD (n=34). Further implications can be drawn as the different genes have distinct inheritance patterns. Thus, this heterogeneity can present as problematic for diagnostics purposes, genetic counseling and carrier screening. Congenital myopathies are among the most challenging subgroup of HM, greatly in part due to the inability to use gene-specific diagnostic biomarkers. As highlighted in Chapter 1.2, the diagnosis of CM usually comprises other auxiliary approaches such as EMG and muscle biopsy (eventually brain MRI for differential diagnosis), but none of these can provide guidance towards a specific *locus* and much less a specific genetic variant. Allelic diversity is defined as the presence of distinct disease-causing (pathogenic) variants within the same gene which can give rise to highly similar or diverse phenotypes. Targeted analysis of a small subset of specific variants in such a context of high allelic diversity, may turn into a failure to detect the genetic cause of the diseases. This parameter depends on the gene's size, mutational rate and even a sequence context which could explain the existence of potential mutational hotspots. For the patient cohort studied in this work an average of 0.88 unique variants per patient was obtained, which demonstrates a high allelic heterogeneity.

Within all the analyzed genes, *RYR1* is probably the best representative of this heterogeneity (**Chapter 3.3**). Most non-related patients with *RYR1*-related diseases have distinct variants although they are predominantly of the missense type. During *RYR1* analysis, it is relatively common to find variants of uncertain significance, which poses additional interpretation difficulties and limits the clinical utility of such findings. Moreover, genetic defects in *RYR1* may be associated with an autosomal dominant or recessive inheritance.

In spite of this genetic heterogeneity, for the most prevalent genetic causes in terms of number variants/cases, namely those associated with *MTM1* and *LAMA2*, I was able to identify potential mutational hotspots or at least gene regions with higher tendency to accumulate disease-causing variants (**Chapters 3.2 and 3.5**). Concerning *MTM1*, and

despite its relatively small size in terms of the number of exons (n=15), the analysis of the distribution of pathogenic variants revealed a higher incidence in exons 8, 11, 9, 4 and 12. Overall, nearly half of all XLMTM cases in the *MTM1*-LOVD have variants in these exons. In fact, a subset of six variants (each detected in at least nine patients) accounted for approximately one-quarter of the total cases in the database. Most of these variants coincide with the hypermutable CpG dinucleotides that are prone to methylation-mediated deamination, a mechanism that could explain their recurrence (Cooper et al., 2010), and higher frequency especially in cases where variants have arisen *de novo.* As for the *LAMA2* gene, a group of fourteen disease-associated variants are amongst the most prevalent, three of which were found across different ethnical backgrounds. In fact, these variants were represented in population variant databases such as gnomAD/ExAC, with frequencies as high as 0.012%. One additional frequent variant - c.1854_1861dup - that was found present in MDC1A Portuguese patients, seems to be population-specific (Latino subgroup), and has a relatively high frequency (0.23%) in gnomAD. This frequency is consistent with the autosomal recessive nature of *LAMA2*-MD and that this disorder is the most frequent subtype of congenital muscle diseases.

5.1.2 EXPANDING THE MUTATIONAL SPECTRUM OF CONGENITAL MUSCLE DISEASES

In collaboration with international groups, 130 novel pathogenic variants identified in 14 genes and scattered over the CM and CMD, were reported in this work. Besides this additional contribution towards its variome, novel insights were provided towards the mutational spectrum of these diseases. Large duplications in *MTM1* (**Chapter 3.2**) and *LAMA2* (**Chapter 3.4**) *loci* were identified for the first time ever, and novel large heterozygous deletions were also characterized. This type of variants is virtually undetectable by conventional sequencing which explains their refractory nature. Large deletions and duplications detected in CMD and CM patients are not restricted to these two genes; in fact, our group has also reported a patient with a Fukuyama CMD caused by a multi-exonic duplication in the *FKTN* gene (Costa et al., 2013). Further reports of other pathogenic CNVs in other related genes include *COL6A1* and *COL6A2* (Foley et al., 2011); *ISPD* (Czeschik et al., 2013); *LARGE1* (van Reeuwijk et al., 2007; Clarke et al., 2011); *NEB* (Kiiski et al., 2013); *POMGNT1* (Saredi et al., 2012); *RYR1* (Remiche et al., 2015) and *TPM3* (Kiiski et al., 2015). As shown the characterization of these genetic defects requires the use of a wide variety of techniques such as long-range PCR, cDNA and Southern-blot analysis. Based on the data reviewed from two large patient cohorts (95 fully genotyped *LAMA2*-MD patients) the estimation of the allelic frequency of large deletions and duplications in *LAMA2* was attempted. Considering that its frequency was determined to be as high as 18.4%, it is

necessary to expand routine genetic diagnosis methods to effectively detect such variants. Our research certainly increased the international awareness towards their existence, bridging towards diagnostic implications especially for *LAMA2*.

5.1.3 WIDENING THE CLINICAL SPECTRUM OF HEREDITARY MYOPATHIES

Large deletions and duplications in the *LAMA2* gene are not particularly associated with phenotypic variability, meaning that from a clinical perspective they are indistinguishable from other variants types. However, this was not the case for the first multi-exonic duplication described in *MTM1* **(Chapter 3.2)**. The clinical presentation of our patient is atypical in the sense that it is less severe than the classical form of XLMTM. He showed delayed motor development skills and muscle weakness, but at seven years of age, he maintained independent ambulation and no respiratory involvement. The genomic and transcriptomic data suggested the presence of mosaicism in this case, as determined by quantitative data from the duplication and a residual amount of normal transcript still present in the muscle. Thus, the explanation for the patient's relatively mild phenotype does not lay upon the type of variant nor his genotype, but in the presence of somatic mosaicism of embryonic origin (**Chapter 3.2**). It involves the presence of two or more populations of cells, with dissimilar genotypes in the same individual. This event reflects high *de novo* mutation rates, adding another layer of phenotypic variability to genetic diseases. Together with sensitivity issues in variants' detection, it is most likely that mosaicism might explain a small subset of patients that persist with no genetic characterization. During this research, a second patient was identified as being mosaic for a missense variant in the *ACTA1* gene (patient #13 in Table 5.1.1 and Appendix IV.1), detected using the NGS gene panel for CM. Again, neither the clinical presentation nor muscle histology was highly suggestive of the involvement of this specific gene. Only one publication described an additional *ACTA1* variant with histopathological features consistent with myofibrillar myopathy (Selcen, 2015). Thus, this gene should be considered in patients with congenital hypotonia associated with muscle weakness and features of myofibrillar myopathy. The topic of somatic mosaicism concerning muscle diseases is not new. It is one of the possible explanations for clinical variability including intrafamilial, as observed in mildly affected parents with a severely affected offspring. It was also a plausible explanation of incomplete penetrance along two consecutive generations. An additional patient with a mosaic variant in *ACTA1* has been also reported in the literature (Laing et al., 2009), and several other HM cases were associated with mosaic variants in *COL6A* genes (Donkervoort et al., 2015), *DMD* (van Essen et al., 1992), *LMNA* (Makri et al., 2009), *RYR1* (Marks et al., 2018) and *TPM2* (Tasca et al., 2012).

The phenotype of *LAMA2*-MD was expanded over the last decade with the description of a limited number of patients with late-onset phenotype. Previously, we have characterized a novel missense variant (c.2461A>C, p.Thr821Pro) detected in two unrelated patients, found in association with a later onset *LAMA2*-MD (Marques et al., 2014). It was interesting to note that further three patients, from two unrelated families with Portuguese ancestry, have also been reported by other groups from Canada (patient #102482 in *LAMA2*-LOVD) and France (Nelson et al., 2015). To expand the knowledge about this variant, we have characterized from a clinical and genetic perspective six additional patients with late-onset phenotype from four unrelated families all sharing this missense change (**Chapter 3.5**). The phenotype associated with this variant translates into mild muscle weakness, when compared to the classical MDC1A phenotype, and gait disturbance. Only the oldest patient, homozygous for this variant, has lost independent ambulation during the sixth decade of life. Based on these independent findings I have postulated that this variant could have a relatively high frequency in the Portuguese population and could account for a subset of patients with LGMD or EDMD phenotypes. Accordingly, its frequency in controls and uncharacterized patients was 0.05% and 2.9% respectively. This difference was, using Fisher's exact test, deemed extremely statistically significant (P<0.0001). Relevant conclusions could be drafted from these results. The first is that it is diagnostically important to consider *LAMA2* gene involvement not only in patients with congenital phenotypes but also as a possible cause of MD with onset beyond childhood and even in adulthood. The second is that *LAMA2* defects and in particular the p.Thr821Pro variant could account for a significant number of cases in our country. Consequently, it should be emphasized that the *LAMA2* gene should be included in the list of candidate *loci* for progressive MD. This has been recognized in the newest LGMD classification that now includes *LAMA2* locus. Finally, this variant has a 58-fold difference in frequency comparing both groups (controls vs patients). This constitutes, according to the American College of Medical Genetics guidelines for sequence variant interpretation, an important evidence towards its pathogenicity: PS4 (Pathogenic / Strong evidence 4), "the prevalence of the variant in affected individuals is significantly increased compared with the prevalence in control" (Richards et al., 2015).

The progressive overlap between congenital and non-congenital forms in terms of their genetic etiology has been greatly expanded with the introduction of NGS technology, especially through more extensive approaches such as WES or WGS. Two patients analyzed by WES in the present work are illustrative of this paradigm (**Chapters 3.5 and 4.2**). In both cases, delayed motor development (walking), muscle weakness and elevated CK levels were the initial symptoms. Interestingly, there was also an association between muscular dystrophy and intellectual disability. Extensive diagnostic workup failed to identify the genetic conditions underlying these phenotypes. Similarly, and although distinct genetic

causes (*LAMA2* and *CHKB)* have been identified, the milder phenotype found in the two patients could be attributed to the presence of variants affecting donor splice sites.

The study depicted in **Chapter 4.2** reports the characterization of an LGMD patient (#14) with *CHKB* gene defects. *CHKB* codes for the kinase choline beta, an enzyme involved in the first step of the Kennedy (CDP-choline) pathway. Changes in the activity of this isoform were previously demonstrated to cause phosphatidylcholine deficiency particularly in muscle cells (Mitsuhashi et al., 2011a). Variants in this *locus* were initially reported as associated with a form of CMD with large mitochondria (megaconial).

Table 5.1.1- Additional still unpublished results generated by this research.

Patient #	Gender	Age	Research strategy	Outcome / results summary
11	F	2yrs	CM NGS gene panel; WES (singleton)	Two novel heterozygous variants in *KLHL41* gene: NM_006063.2:c.667C>T (p.Arg223Cys) Large deletion involving exon 2 and part of exon 3 (Appendix IV.2)
13	M	11yrs	CM NGS gene panel (tested in two different DNA samples obtained from blood and muscle tissue)	A novel mosaic variant in *ACTA1* gene: NM_001100.3:c.461T>C (p.Val154Ala) (Appendix IV.1)
17	F	45yrs	Neurological NGS gene panel; WES (singleton)	Novel heterozygous variant in *CCDC78* gene: NM_001031737.2:c.60+1G>C (r.spl?)
20	F	46yrs	Neurological NGS gene panel	Disease-causing variant in *MATR3* gene: NM_018834.5:c.254C>G (p.Ser85Cys) (Appendix IV.3)
21	M	20yrs	*DMD* transcripts analysis; Western-blot; WGS; bioinformatics	Novel large genomic inversion disrupting *DMD* gene (Appendix IV.4); (Oliveira et al., 2018b).

Footnote: F- female, M- male, yrs- years.

Only a limited number of cases (less than 30) have been reported so far making this clinical entity very rare and still poorly characterized. Our contribution was to provide an additional case towards this global knowledge. Moreover, the phenotype initially classified as congenital should be reviewed as it may comprise progressive MD. The c.1031+3G>C variant found in homozygosity promotes changes in the normal splicing mechanism reshaping the transcript profile of *CHKB*. The milder presentation could be explained by the presence of at least residual functional choline kinase activity, as inferred by *in-silico* analysis. Typical muscle histology displays the concomitant presence of dystrophic features and large mitochondria located at the periphery and central areas in the myofiber and devoid of activity. Thus, most patients with CHKB defects were initially identified through the presence of enlarged mitochondria in the periphery of the myofibers and the absence of this organelle in the center of the cells and were considered to be pathognomonic (Nishino et al., 1998). These augmented mitochondria correlated with the reduced levels of phosphatidylcholine in the membrane. This appears to be the consequence of a compensatory mechanism resulting from the depletion of functionally compromised mitochondria by autophagy (Mitsuhashi & Nishino, 2011; Mitsuhashi & Nishino, 2013). A common aspect in the atypical cases was that several muscle biopsies needed to be performed. It seems that it is not always straightforward to identify or establish the significance of these somewhat subtle pathological changes. The pathognomonic finding, the enlarged mitochondria, although presenting high specificity, appears to be a marker with low sensitivity particularly in the milder forms of the disease. To obtain further insight into the epidemiology of this disease, the *CHKB* gene was analyzed in eight additional patients presenting both MD and intellectual disability. Patients were selected from our cohort, but no pathogenic variants were found (Oliveira et al., 2016). I have just recently performed the genetic characterized of one additional Portuguese female patient, with the typical triad: muscle dystrophy, cardiomyopathy, and intellectual disability. She has the same splice variant found in patient #14, in compound heterozygosity with a novel frameshift deletion (data not shown). This reinforces that this subtype of CMD is most likely underestimated.

5.1.4 FURTHER GENOTYPE-PHENOTYPE CORRELATIONS IN *MTM1*- AND *LAMA2*-RELATED DISEASES

Considering the valuable development of LSDBs for *MTM1* and *LAMA2* (section 5.1.5), it was possible to collect new data (clinical and genetic) for these two *loci*. Even for rare diseases, as the size of the data increases, it is possible to refine genotype-phenotype correlations. I have particularly focused on missense variants, as these also pose further difficulties in their clinical interpretation. In the absence of functional tests, distinguishing

between disease-causing and non-disease missense variants requires the use of several criteria including bioinformatic analysis.

For the data collected in *MTM1*-LOVD, it was possible to corroborate that truncating variants (nonsense or out-of-frame deletions or duplications) are predominantly associated with a severe phenotype, based on a large number of male patients reported with such variants. I've determined that most missense mutations are in myotubularin regions with known function, suggesting that these changes are not randomly scattered in the protein, especially when compared with the distribution of nonsense variants. Moreover, most missense variants associated with a severe phenotype are located in two specific regions: the protein tyrosine phosphatase (PTP) which constitutes the catalytic domain; and the Rac-induced recruitment domain (RID). In terms of its relative proportion, mild/moderate missense mutations are more frequent in the Rab-like GTPase Activator and Myotubularin (GRAM) domain, followed by the other regions of myotubularin. These patterns had not been previously delineated in the literature.

As for *LAMA2* missense variants listed in the *LAMA2*-LOVD, they also followed a non-random distribution. They cluster in specific regions of the laminin-α2 chain: a first group are located in domain VI corresponding to the N-terminal domain; a second group consists of missense variants that specifically alter cysteine residues, located in one of the three EGF-like repeats (domains V, IIIb, and IIIa) known to establish disulfide bridges; a third group is affected in residues located in the C-terminal region of the protein which contains a tandem of five laminin G-like (LG) domains. In addition to the identification of these clusters, possible explanations for their pathogenicity were also proposed. For the first group, the missense variants located in domain VI may disrupt laminin-211 protein folding, leading to the loss of the ability for polymerization into supramolecular networks. The second group, affecting cysteine residues, may compromise the solenoid-like structure conveyed by EGF-like repeats, altering the integrity of the connection between the sarcolemma and extracellular matrix mediated by laminin-α2. Finally, the disease associated missense variants in the LG domains, may compromise the binding of laminin-α2 to the *O*-linked carbohydrate chains of α-dystroglycan, and more specifically the connection to integrin α7/β1 (LG2 and LG3 domains).

5.1.5 DEVELOPMENT AND CURATION OF *LOCUS*-SPECIFIC DATABASES

Since the beginning of the molecular genetics age, researchers wondered how to convey to others the variant data that was being obtained, how it could be shared and updated alternatively to the static display in a scientific publication. Still in the pre-genomic age and with the development of world-wide-web in the early 90s (consisting in websites and mostly

stationary HTML pages accessible through a web browser), the first genetic variants databases (GVDs) began to flourish. These databases suffered significant expansion in diversity, purposes, and number. The first example of GVDs, were publicly acessible population sequence databases (e.g. dbSNP, Exome Variant Server, 1000 Genome Project and gnomAD/ExAC), a second group of locus-specific databases (LSDBs), and finally, disease databases (such as Clinvar, HGMD, and OMIM). The latter presented limiting aspects especially in terms of mutational coverage and/or representativeness, but still contributed to accelerate genetic diagnostics. With the development of specific software and tools, such as the Leiden Open (source) Variation Database (LOVD), the creation and maintenance of LSDB was facilitated (Fokkema et al., 2005; Fokkema et al., 2011). As sequencing technologies evolved, more and more data has been generated, reaching the status of exponentially growing sequencing capacities, through the so-called massive parallel sequencing. For a significant number of variants, we are still unable to ascertain their true nature/impact, and they are consequently cataloged as unclassified variants or VUS. Additionally, for rare pathogenic variants, it is often difficult to determine the specific corresponding phenotype. All these aspects certainly place LSDBs in a center stage, especially to assist the classification of new data. Some authors stated that LSDBs should perhaps be known as genotype-phenotype databases, mainly due to the fact as they could shed light on the relationship between underlying genetic variants and their disease manifestations (Dalgleish, 2016). My involvement in this initiative started as I progressively obtained more expertise in the curation of LSDBs and perceived that there was a real need for their creation (Oliveira et al., 2008; Oliveira et al., 2010). As part of this thesis, I have developed and reported two LDSBs, internationally recognized as such, for *LAMA2* and *MTM1* genes. The necessary steps to perform the development and curation of such databases are described in the literature (Celli et al., 2011). Broadly, it comprises four stages: pre-development, initiation, curation and maintenance. At the time of writing, these databases contain i) a total of 1923 (553 distinct) variants in the case of *LAMA2*-LOVD; and ii) 526 (290 distinct) variants in the *MTM1* LSDB. Overall, these figures represent a large effort from the curators' perspective, especially since this task was not subjected to specific funding or financial support. Although the work concerning these databases cannot be directly and globally measured in terms of scientific productivity, their updates were reported in two separate publications (**Chapters 3.2 and 3.5**). Besides the considerable number of new variants released into the public domain, the novel genotype-phenotype insights are summarized in section 5.1.4. Finally, I must emphasize that the newly developed LSDBs are extremely useful resources in clinical genetics both for research and diagnostic purposes. They are also key players in the global effort to establish a reference dataset of human sequence variants in *LAMA2* and *MTM1* genes.

5.1.6 APPLYING NGS TECHNOLOGY TO INVESTIGATE GENETICALLY UNSOLVED CASES OF HEREDITARY MYOPATHIES

During the last decade, a new chapter in human genetics has been initiated, due to the development of high-throughput sequencing methodologies (NGS or MPS) capable of analyzing an entire genome at the individual scale. An important part of my research consisted of developing and applying NGS technology to identify new genetic causes for hereditary myopathies, with the overarching objective of incorporating this technology into future diagnosis. Thus, several different approaches comprising gene panels, WES, WGS, and single molecule real-time sequencing were tentatively used depending on the context of the research.

The first task consisted in the development of a tailored NGS gene panel specific for CM (**Chapter 4.1**). I've chosen to start with this specific subgroup of muscle diseases, due to their lower diagnostic yield based on conventional approaches, which have been observed not only by our group but are recurrently reported in the literature. Several factors can justify this low sensitivity, that recapitulates the challenges also present in other muscle diseases. In CMs no specific antibodies are available to perform IHC in muscle biopsy, in contrast to LGMDs where such an approach is useful to guide the genetic study. The extent of this guidance is extremely variable in CM, depending on the type of structural changes found in the myofiber. It may be less useful when there is only predominance of type 1 fibers, which would indicate a possible CM but nearly all subtypes could be still accounted for. Some pathognomonic histological features allow more specific guidance in a subset of CM cases, but none can provide guidance towards a (single) specific gene. The second factor explaining the difficulties in CM are related to expanding genetic etiology. Since the initial design of this NGS gene panel (over a 5-year period) several new CM-related genes have been identified: *KLHL41*, *LMOD3*, *MAP3K20*, *MYPN* and *SPEG* (Gupta et al., 2013; Agrawal et al., 2014; Yuen et al., 2014; Miyatake et al., 2017; Vasli et al., 2017). As a final factor, the severest end of the CM disease spectrum includes cases with prenatal onset and neonatal death, where the scarcity of clinical symptoms and the limited number of additional auxiliary tests often cut-short the diagnostic workup. Our strategy, that consisted of applying a new gene panel for CM, led to an overall pathogenic variant detection rate of 70% (7 out of 10) in a well-selected group of patients (**Chapter 4.1**). This was quite relevant considering that the assay does not include all the genomic (specially intronic) regions, nor does it detect all possible large genomic rearrangements in these *loci*. Furthermore, seven novel variants in *RYR1* and *TTN* genes were identified, and the effect on the splicing mechanisms was determined for variants in the *NEB* gene (**Chapter 4.1**). In fact, nearly one-quarter of all variants detected in this group of patients required further studies, namely expression

analysis at the mRNA level. Despite the wide phenotypic heterogeneity and the huge size of the candidate genes - which makes the diagnosis of CM complex, costly and labor-intensive - a novel integrated NGS approach for the genetic diagnosis of CM patients was proposed.

Next, I applied a large NGS gene panel that is commercially available for neurological disorders that include several genes (n=751) known to harbor variants that affect the functioning of the brain and the nervous system. Thus, this gene panel covers other diseases beyond the group of hereditary myopathies. A total of four cases with LGMD-type phenotypes were analyzed using this strategy. A previously reported variant in *MATR3* (c.254C>G, p.Ser85Cys) was found in one of these patients (#20 in Table 5.1.1, Appendix IV.3). This female patient presents adult-onset distal muscle weakness, frequent rimmed vacuoles in the muscle, and a family history of an autosomal dominant inheritance. Matrin-3 related distal myopathy is a very rare disease, having been reported in only 10 families worldwide - the first family in France was just recently reported by Barp et al., 2018 - all sharing the same missense variant. Clinically, the condition manifests as an adult-onset distal weakness in lower and upper limbs, frequently associated with dysphagia, dysphonia, vocal cord and pharyngeal weakness. The description of this new family in Portugal emphasizes its worldwide dispersion and highlights the relevance of NGS to identify such rare conditions.

While NGS gene panels were proven to be useful to address the genetic variability within myopathies, they enclose intrinsic limitations, especially when dealing with extreme genetic heterogeneity. If a set of genes are chosen to be included in the panel, their applicability will perpetuate the established clinical classification, e.g. the division between congenital vs non-congenital forms. Also, in terms of genetic research NGS panels have limited utility for the identification of new disease genes. WES in its turn has been proposed as an important research tool for gene discovery, further expanding the genetic etiology of hereditary conditions (Ku et al., 2012). In this thesis, exome sequencing was performed in eight unrelated cases (a total of 12 individuals: six singletons and two trios). In four of these, pathogenic or likely pathogenic variants were identified for *ASCC1* (**Chapter 4.3**), *CCDC78* (patient in #17, **Table 5.1.1**), *CHKB* (**Chapter 4.1**) and *LAMA2* (**Chapter 3.5**) genes. Interestingly, most of these variants affected donor splice-sites, and all except the c.60+1G>C substitution, have been further characterized at the mRNA level. An additional novel heterozygous missense variant (p.Arg223Cys) in the *KLHL41* gene was detected in a CM patient with nemaline myopathy (patient in #11, **Table 5.1.1**). As the *KLHL41*-related disease is autosomal recessive, its involvement in the phenotype of the patient could not be confirmed just with the application of exome sequencing. However, further investigations carried-out just recently identified a large genomic deletion by CytoScan XON array (details in Appendix IV.2). The bioinformatic analysis of exome sequencing to identify CNVs was not

successful in this case, thus it is necessary to screen such alterations using more sensitive approaches.

It is noteworthy that most of the variants were identified in genes related to diseases with only a very limited number of patients reported in the literature so far. Thus, these new cases have contributed to the expansion of the clinical spectrum of the disorders as detailed in section 5.1.3. Overall, the application of WES was successful in at least 50% (4/8) of the cases included in this work. This reinforces the growing body of evidence suggesting the clinical utility of WES, inclusive as a first-tier diagnostic tool for several medical specialties, allowing the identification of genetic defects linked to monogenic or even complex multigenic disorders.

In collaboration with LUMC, I have applied different NGS approaches such as Single Molecule Real-Time (SMRT) sequencing (PacBio) and 10x GemCode Technology (10x Genomics), both part of the 3rd-generation sequencing technologies. As detailed in **Chapter 1.3**, this newer technology, is a major improvement as compared with the so-called short sequencers (2nd generation). Here, it is possible to generate reads which can reach more than 30 Kbp in length. These longer reads are primarily presented as useful for *de novo* assemblies of microbial, plant or animal genomes. But concerning clinical genetics, this 3rd-generation sequencing technology can be also useful for resequencing, especially to identify structural variants and phasing variants across large regions of the human genome. In addition, several researchers have been applying these longer reads to fill-in the gaps present in the human reference genome, especially those consisting of repetitive elements and complex regions, that are refractory to scrutiny by other methods and sequencing technologies. This 3rd-generation technology was applied in this work and was presented in **Chapter 4.4**, where it was useful to retrieve the sequence of a LINE-1, that was newly inserted in the *DMD* gene and caused a Becker muscular dystrophy (more details in section 5.1.8). Also, an additional patient with a similar phenotype but with cognitive involvement (#21 in Table 5.1.1, Appendix IV.4, Oliveira et al., 2018b), was characterized resorting to low-coverage WGS using the 10x GemCode Technology (10xGenomics Chromium system). Routine diagnostic techniques, such as MLPA and conventional sequencing, failed to detect pathogenic variants in *DMD*. Its involvement was highly suspected considering that the patient's dystrophic muscle showed irregular labeling for dystrophin. By combining transcriptomic and genomic approaches it was possible to narrow the genetic defect to *DMD*, probably involving exons 75 to 79. More detailed data analysis of WGS data, through visual BAM file inspection, showed a possible breakpoint within intron 74 of DMD. Some reads had homology with a region located upstream of the peroxiredoxin 4 (*PRDX4*) gene located at Xp22.11. Correspondingly, some reads in that gene's location showed homology with intron 74 *DMD* sequences in an inverted order. Further experiments confirmed the presence of a

large (~8 Mb) genomic inversion. Abnormal *DMD* transcripts were subsequently identified, including some containing segments from the region upstream of *PRDX4*. The patient's phenotype is thereby explained by this *DMD* inversion, leading to the loss of the C-terminal region of dystrophin. Besides expanding the *DMD* mutational spectrum, these results emphasize the importance of WGS in clinical genetics, having the potential to detect a wide variety of mutation types in an "all-in-one" approach. However, applying WGS in routine diagnosis is still challenging, especially due to the demanding informatic resources required and the vast number of VUS often found.

5.1.7 BIOINFORMATIC ANALYSIS OF NGS DATA

One challenging aspect for the use of NGS, especially for larger applications like WES or WGS, is its demand for bioinformatics. The success or failure in research, and especially diagnostics, is much dependent upon the correct choice of algorithms and inheritance disease models. As part of this research, several bioinformatic tools and pipelines have been used or established to deal with NGS data or to facilitate the identification of the specific genetic alterations. Moreover, this novel sequencing technology is prone to errors and artifacts, which bioinformatics can certainly help to reduce.

During this research, I have concentrated significant efforts in optimizing and evaluating several open-source and commercial software. One of the aims of this work was to associate novel genes to genetic diseases. Usually, the identification of such an association can be accomplished using different strategies, much relying on special kindreds with multiple affected members or non-related patients that share very similar phenotypes. As for the intra-familial studies, they are either suggestive of an autosomal dominant inheritance pattern with several affected individuals in different generations, or pedigrees suggesting an x-linked inheritance, or alternatively families with high consanguinity rates. In such a context, it is possible, by means of indirect approaches and resorting to genetic markers, to narrow down the region of the genome where the genetic defect causing a disease lays upon, or even, using NGS approaches (WES or WGS), identify straightaway the disease-causing variant. Although research-appealing, these families are very scarce, and none have been accessible to the researchers during this project. At least in hereditary myopathies, our experience indicates that many still unsolved cases are sporadic, meaning no family history of the disease exists. In such cases, at least when an autosomal recessive disease is suspected, the application of homozygosity mapping (HomMap) should be considered, not only for research but also in routine diagnostics.

The currently available algorithms and bioinformatic tools designed for ROH detection from array-SNP and WES data were reviewed (**Chapter 4.5**), highlighting their main

features, advantages, and limitations. Based on this assessment, selection of appropriate algorithms for HomMap should mainly consider the technology used to generate the data, but also specific features of the patient/family under study, such as genetic context, ROH average size and the number of relatives affected by the same condition. The presence of consanguinity has its particularities: not only the disease inheritance model is self-evident (even in sporadic cases) but concentrating the analysis of variants in longer ROH regions reduces considerably the analytical burden. However, ROH analysis should not be confined to consanguineous cases as it may be useful in cases with unknown parental consanguinity, especially if parents are from the same remote/confined village, where mutation founder effects or inbreeding events might be an issue. The work exemplified in this thesis, using our own data and the software HomozygosityMapper (Seelow & Schuelke, 2012), shows how the analysis of ROH regions can be useful for variant filtering and accelerating the identification of clinically relevant variants.

Further analysis, still preliminary, has been performed using our WES data. By correlating the highest homozygosity score and the total size of ROH from HomozygosityMapper and applying the k-means clustering algorithm, we were able to generate two different clusters (Oliveira et al., 2018a). One cluster assembles from exomes with a low number (and size) of ROH, where heterozygous variants should be primarily inspected, and the other (broader) cluster aggregates the exomes with larger ROH and a higher number of such regions. If our assumption is correct, even in sporadic cases and based on where they are clustered, it will be possible to infer about the inheritance disease model.

5.1.8 CONFIRMING *ASCC1* AS A NOVEL DISEASE-CAUSING GENE

One of the most relevant contributions of this research was providing the confirmation of a new gene-disease association. From the early stages of human molecular genetics, one of the primary goals of research is the identification of novel disease-causing genes. With the sequencing of the human genome and later with the development of NGS technology, such reports have increased exponentially. Concerning the diseases that primarily affect skeletal muscles, considerable advances in gene discovery have been also accomplished during the last decade. In **Chapter 4.3** a patient with a rare and severe form of severe neuromuscular disease was reported. The parents, a healthy apparently non-consanguineous Portuguese couple, had two other affected pregnancies which ultimately lead to the loss of the babies. The phenotype consisted of generalized hypotonia, congenital fractures of the long bones, lack of spontaneous movements and poor respiratory effort, leading to death within the first days of life. Although this case was subject to an extensive diagnostic effort, the genetic cause of the disease remained unclear. Thus, this family was included in this research

project and primarily studied by WES. Since only a DNA sample from the proband was available, our strategy consisted of performing exome sequencing followed by filtering rare variants in this case. The still extensive list of variants included a homozygous duplication located in the coding region of the activating signal cointegrator 1 (ASC-1) complex, subunit 1 (*ASCC1*) gene causing a shift in the open reading frame. At that time little was known about this *locus.* Its product takes part of the ASC-1 complex, a tetramer composed of three other proteins encoded by different genes: subunit 2 (*ASCC2*) and subunit 3 (*ASCC3*) of the ASC-1 complex and the thyroid hormone receptor interactor 4 (*TRIP4*) (Jung et al., 2002; Knierim et al., 2016). This complex was shown to interact with nuclear receptors and other transcription factors (Kim et al., 1999). *ASCC1* was initially disease-associated with cancer, where a germline missense variant was shown to predispose for Barrett esophagus and esophageal adenocarcinoma (Orloff et al., 2011; van Nistelrooij et al., 2014). Further research has placed the ASC-1 complex as a repressor of the nuclear factor kappa B (NF-κB). This factor, although playing an important role in the immune system (Hayden et al., 2011), when improperly activated can give rise to inflammatory diseases and cancer. Interestingly during the analysis of our patient, four families with severe neuromuscular disease, termed as a prenatal SMA with respiratory distress and congenital bone fractures, were reported (Knierim et al., 2016). These autosomal recessive pedigrees were shown to be caused by loss-of-function variants in two different genes, *TRIP4* locus in three families and one concerning *ASCC1* gene. Our patient shows clinical features strikingly similar to the Turkish siblings described by Knierim and co-workers and all shared the same variant in *ASCC1*. The third report of additional patients with *TRIP4* mutations shared clinical features with the previous patients but without bone fractures (Davignon et al., 2016). This last paper added further speculation on the underlying disease mechanism, which further addressed in section 5.2.4 of this thesis. Although the *ASCC1* discovery as a novel disease-causing gene, has been outrun by the excellent publication by Knierim and collaborators, our case - the second known ever - was important as it provided an essential final proof of *ASCC1* involvement in a severe congenital neuromuscular disease.

5.1.9 CHARACTERIZATION OF A NOVEL MUTATIONAL MECHANISM IN THE *DMD* GENE

In **Chapter 4.4**, the genetic characterization of a patient with a Becker muscular dystrophy phenotype is reported. Again, the patient remained several years without a definitive diagnosis, as routine methods for *DMD* gene analysis (MLPA and genomic sequencing) were unsuccessful to identify pathogenic variants. It should be noted that *DMD* is the largest gene of the human genome and, mainly for this reason, it is particularly prone to genomic rearrangements, often intragenic multi-exonic deletions or duplications (~80% of total cases)

(Bladen et al., 2015). In the remaining patients, smaller (point) variants are accounted for, whereas complex genomic rearrangements encompassing *DMD* are much rarer. Further research in this patient, using multiple strategies, allowed the identification of a completely novel mutational mechanism involving the partial exonization of a newly inserted LINE-1 element in intron 51 of *DMD*. Its identification brings to the discussion a subject with almost three decades from the initial observation by Haig Kazazian Jr. and colleagues. They investigated a patient with hemophilia A caused by an exonic LINE-1 insertion in the factor VIII (F8) gene (Kazazian et al. 1988). LINE-1s are one of the frequent transposable elements of the human genome (and the only active and autonomous), approximately 0.5 million LINE-1s totalize 17% of the genomic sequence. In its active form, the LINE-1 complete sequence is about 6 Kbp in length and contains several conserved elements: a 5′- and 3′-UTR, encodes two proteins by ORF1 and ORF2, separated by an inter-ORF spacer with 63 bp and ending in a long poly-A tail (Hancks & Kazazian, 2016). In summary, active LINE-1 can be "copy & pasted" through transcription, translation and cDNA sequence integration into novel genomic locations. Besides this normal retransposition activity, an additional 3′-transduction has also been reported, by which LINE-1 contributes further to genomic evolution, by shuffling protein-coding sequences throughout the genome (Hancks & Kazazian, 2016). Only under 100 LINE-1s are believed to be capable of retrotransposition activity in an individual's genome (Hancks & Kazazian, 2016). After the initial report in hemophilia A, LINE-1-mediated retrotransposition and disease-causing insertions have been progressively reported in the literature reaching more than 100 cases so far (Suarez et al., 2018). Concerning the *DMD* gene, five pathogenic insertions have been described. However, the mutational event characterized in our patient is completely distinct. It is a deep-intronic insertion of a full LINE-1 sequence (member of the L1HS subfamily). Just recently with the scientific inputs of Prof. Haig Kazazian, it was possible to track down its original location in the reference genome at 4p15.31 (a 100% match). Previous research by Brouha et al. (2003) showed that this element was modestly active, with only 7% of activity relative to a very hot LINE-1. It has a population gene frequency of 0.46, meaning that there are many filled sites and many empty ones, for this specific LINE-1, in the genomes of the general population (Haig Kazazian, personal communication). Such an insertion leads to its partial exonization at the cDNA level, due to the recognition of a cryptic 3′ splice-site located in intron 51, and a 5′ splice-site within this element. Consequently, the known mutational mechanisms giving rise to DMD/B phenotypes especially those affecting splicing (Tuffery-Giraud et al., 2017), have been expanded with this additional report of a deep-intronic insertion of a LINE-1 element in the *DMD* gene.

5.1.10 FINAL CONCLUSIONS AND RESEARCH QUESTIONS RESTATEMENT

"As for the most frequent subtypes of congenital muscle diseases (LAMA2-, MTM1- and RYR1- related diseases) could further insights about genotype-phenotype correlations and their mutational spectra be obtained?"

The genetic data obtained from 112 patients with congenital muscle diseases characterized in this work demonstrated that the most prevalent genetic entities are related to defects in *LAMA2*, *RYR1*, and *MTM1*. Given the large number of variants collected in this cohort and together with the data from the LSDB, refined genotype-phenotype correlations were established. Male patients with missense variants in the *MTM1* gene can either present a milder or a severe phenotype. Our analysis showed that missense changes associated with a severe phenotype are most likely localized in two specific regions of myotubularin with known biological function. These variants distribution pattern, not identified before, is particularly important for disease prognosis and for patient management. As for *LAMA2* missense variants, they also follow a non-random distribution and could be typified into three different groups. In addition to the identification of such clusters, it was possible to propose the pathomechanisms likely to be involved in these variants and related phenotypes.

"Considering the unique nature of congenital muscles diseases, especially in terms of its genetic and clinical heterogeneity, can this group of diseases fit as a model to improve the diagnostic workup of hereditary myopathies (senso lato)?"

Based on the clinical heterogeneity and the diversity of genetic profiles of CM and CMD patients, there was a clear indication that the gene-by-gene stepwise analysis conducted through conventional sequencing is no longer the most efficient strategy. In the presence of a patient with a muscle disease, with onset just after birth or during the first months of life, variants in *LAMA2*, *RYR1*, and *MTM1* should be considered as the most likely genetic causes involved. If the CK levels are highly elevated and/or brain white matter changes are part of the phenotype, a set of frequent variants or all *LAMA2* exons can be effectively screened as a first-tier approach, prior to submitting the patient to an invasive muscle biopsy. Muscle biopsies have been paramount since early historical characterization of skeletal muscle diseases and, although several factors should be considered in clinical applications, they are still an important step of the diagnostic workup. Considering patients with normal or only slightly elevated CK levels, a muscle biopsy could be considered, as it is most informative for CM patients, followed by the analysis of a suitable gene panel or even a specific gene if enough evidence is provided from muscle histology, inheritance pattern, or more subtle clinical features. Muscle biopsies also constitute an important (and for several

genes the only) source of biological material to proceed with the characterization of novel variants with unknown clinical significance at RNA or protein levels.

"How can diagnostic algorithms be improved, specifically through the development and application of new genotyping strategies, for example, next-generation sequencing?"

Considering its genetics and muscle pathology CM constitute a unique disease group within HM. Therefore, it is difficult to make further extrapolations towards the remaining groups of muscle diseases. CM unique characteristics also justified the decision of designing a dedicated NGS gene panel for diagnostic purposes such as the one described in this work. Nonetheless, this gene panel will require periodic adjustments concerning its composition based on the novel genetic data reported by the scientific community. As for patients with dystrophic features in muscle biopsy, it would be more effective to use a more comprehensive gene panel, considering the considerable and increasing overlap between congenital and progressive (later onset forms). Such an NGS panel should include at least the genes known to be defective in LGMD, EDMD and CMD phenotypes.

"What sort of molecular defects give rise to the myopathies still uncharacterized from a genetic perspective?"

The clinical overlap also explains a relatively significant fraction of HM patients that remained genetically uncharacterized for several years. In these patients, the genetic etiology was not related to yet uncharacterized genes, but rather "unexpected" *loci* considering the traditional classification of HM. The most illustrative example of such a paradigm was the LGMD patients with *LAMA2* defects found in this work, which ultimately lead to its inclusion as a novel cause for this entity (https://omim.org/entry/618138/). The second group relates with genetic alterations in expected genes, but not detected by routine methods, such as CNVs, deep intronic variants or even more complex variations such as those found in the two BMD cases presented. Finally, there is also the possibility of the involvement of genes previously unassociated with human diseases, as shown with the confirmation of *ASCC1* gene involvement.

"Could the new genetic data gathered provide further understanding concerning the pathophysiological mechanisms underlying myopathies or even towards fundamentals in cell biology?"

Our report contributed with further details about the phenotype and additional insights towards the pathophysiological mechanisms in *ASCC1*-related disease. Its initial

classification as spinal muscular atrophy was arguably, and our data could justify its inclusion in the CM group. Just recently the identification of six additional patients from three families further corroborated this possibility (Böhm et al., 2018). The patients included in this report were subject to a muscle biopsy, further documenting the presence of sarcomeric changes and enlarged Z-bands (Böhm et al., 2018), which could constitute a histopathological hallmark for *ASCC1*-related diseases. Furthermore, this research contributed to the characterization of a novel mutation mechanism related to LINE-1-mediated retrotransposition activity in the *DMD* gene. The observation of such novel events brought to discussion, in the context of cell biology research, the importance of mobile genomic elements such as LINE-1, not only as potential causes of genetic diseases but also as driving-forces behind genome evolution.

5.2 FUTURE PERSPECTIVES

5.2.1 PROCEED WITH THE GENETIC CHARACTERIZATION OF UNSOLVED CASES

From the cohort included in this research seven patients (#8, 9, 10, 16, 18 and 19) have yet to be fully characterized. We expect to continue their characterization through the application of WES (patients #8, 9 and 10) and WGS (patient #16). Finally, concerning the last two unrelated patients #18 and 19, these were preliminarily analyzed by WES. As their phenotype was somewhat similar, including reduced/irregular IHC staining for dysferlin in muscle biopsy, an overlapping variant filtering strategy was used to potentially identify variants shared by both cases. A novel missense variant was identified in a *locus* previously reported as a candidate gene for muscular dystrophy and cardiomyopathy or possibly constitutes as muscular dystrophy phenotype modifier. Further experiments at the transcriptomic and protein level should be carried-out to functionally characterize this variant and to evaluate its potential impact over the phenotype.

5.2.2 INTEGRATION OF NGS IN ROUTINE DIAGNOSTICS

The research work applied different sequencing strategies, which marked a clear transition from the so-called conventional (Sanger-based) sequencing to the more recent NGS technologies. As part of my work as a clinical laboratory geneticist and the clinical context where I am currently integrated, I was particularly interested in deploying some of these NGS approaches into routine diagnostics. Overall, the results shown in the thesis corroborate the importance of applying this technology to genetically heterogeneous diseases such as hereditary myopathies. However, to implement this transition to NGS some relevant aspects should be carefully addressed: which I have integrated into the following categories: i) adjusting technical workflow and bioinformatics; ii) acquiring new core competencies and skills for the laboratory personnel; iii) reducing intrinsic and technological limitations and pitfalls of NGS. Regardless of the technology used, to introduce NGS it is necessary to completely change the way nucleic acids are processed and analyzed, as compared with more conventional sequencing strategies, so that in the final step of the process disease-causing variants can be effectively identified. As detailed in Chapter 1.3, the initial step consists in preparing a library of DNA fragments usually diluted at very low concentrations (within the picomolar range) for subsequent sequencing. Thus, additional new laboratory machinery (besides the sequencer itself) are necessary to perform controlled DNA fragmentation, library quantification, and its QC. From the informatics perspective, the reader must necessarily be aware of the large amount of data being generated from these

sequencers, even from the smaller benchtop apparatus. It is necessary to increase data storage capacity and to perform convenient backups. Also, although much of the primary bioinformatics analysis can be performed in the sequencer's server, it is essential to delineate a custom pipeline to convey secondary and tertiary data analysis. This will likely be translated into the need for additional computational resources. The second aspect concerns laboratory staff training and acquiring of competencies to work with NGS data. As much of the equipment, software, and data analysis strategies are completely different, there is a steep learning curve in NGS. Also, there should be less rotativity between the different work tasks as it may increase overall productivity and quality. As many steps of the workflow are now automatized and even robotized, the traditional paradigm has changed considerably, meaning that fewer technicians are required in the "wet" laboratory and more staff should be involved in other tasks such as data analysis and reporting. Despite the obvious advantageous application of NGS technology, there are important limitations and pitfalls to keep in mind. One relates to the error rate which is considerably higher than conventional sequencing, particularly in Ion Torrent which is known to underperform in homopolymeric regions (Bragg et al., 2013). In the current state of the art, this technology has other potential pitfalls which were thoroughly detailed in **Chapters 4.1** and **4.2**. All these aspects can give rise to false positive and negative results causing analysis constraints. For false positives, these can occur as sequence artifacts that are not easily flagged being interpreted by the software as true variants during variant calling step. Even though they are later excluded by conventional sequencing, they can certainly create an extra burden on the analysis, especially when involving manual inspection of reads alignments using BAM files. False-positives usually there are no major implications for diagnostics, because all the variants, especially the novel ones, should always be confirmed by a second technique (e.g. Sanger sequencing). As for false negative results, there are several critical aspects to consider particularly when performing NGS for diagnostics. At present, there are regions of the genome that are poorly covered by NGS, especially in experiments using 2nd generation NGS technology. Often, these regions have high GC content, extremely repetitive sequences, or locate in genomic segments have almost identical copies in other locations of the genome (pseudogenes). Moreover, the detection of more complex mutations such as repeat expansions, SV or CNVs, are usually undetected by NGS. For that reason, it is highly recommended, when resorting to gene panels, the use of conventional sequencing and other techniques to compensate for the potential failure of NGS in such difficult regions. Thus, a "filling" panel by Sanger sequencing should be implemented along with the NGS gene panels. Other issues giving rise to false-negative results derive from the bioinformatic analysis, especially the annotation and variant filtering steps. Thus, I would recommend validating different sources of variant annotation and software to test for consistency and

ensure higher accuracy. As one of the main conclusions of this work, several techniques that use different biological samples are needed to clarify the impact of novel sequence variants on the phenotype. Therefore, it is unlikely that NGS alone will clarify the genetic complexity present in some cases of hereditary myopathies.

5.2.3 FURTHER ADVANCES IN GENETIC VARIANT DATABASES

Genetic variant databases have been striving for the last two decades. From a conceptual perspective, different types of variant databases were proposed, ranging from the so-called central databases to the more gene-centered repositories (LSDBs), all having advantages and limitations when compared with each other. These databases have gained further relevance with an increasing amount of data being generated with the 2nd generation sequencers. In fact, this newer sequencing technology has introduced an interesting uniqueness: not only do they generate a considerable number of new variants but the knowledge concerning their frequency and impact over the phenotype can be of extremely utility for NGS data analysis. Thus, NGS creates a necessity but can also provide an answer especially if variant data can be shared to the public domain.

In addition to the work of curating the two LSDBs reported in the thesis, I intend to further develop LSDBs and GVDs. Being already involved in the management and curation of *RYR1* gene LSDB (http://tiny.cc/LOVD_RYR1), I will continue its development into a more matured state. To ensure data integration and increase variant data accessibility it will be necessary to develop algorithms based on artificial intelligence (AI) more specifically deep learning algorithms, to automatize the feed-in of genetic variant databases. These can be used, in theory, to scan publications, web pages, and other public databases, allowing the collection of variants into a common repository. As in many other areas, such an AI system could constitute a huge breakthrough, assisting many of the tasks around LSDB curation that currently require human intervention and a considerable amount of manual work.

5.2.4 CLARIFYING THE PATHOPHYSIOLOGICAL MECHANISMS UNDERLYING ASC-1 RELATED DISEASES

As previously mentioned, we have identified the second worldwide case of a new entity of neuromuscular disease caused by truncating variants in the *ASCC1* gene. The scarcity of cases and the reduced knowledge about its molecular basis certainly provide us with a fertile ground for further research. There are also apparently contradictory findings between *TRIP4* and *ASCC1*-related phenotypes and their pathology. In part, these contradictions might be explained by the strategies and animal models used to mimic human diseases. Knierim and

collaborators (2016), created *ascc1* and *trip4* zebrafish mutants transiently generated using antisense morpholino oligonucleotides at the expression level (knock-down). Using this model, a severe disturbance in the development of the neuromotor unit, more specifically, the axonal outgrowth, neuromuscular junction density, and the myotome were found to be compromised (Knierim et al., 2016). In opposition, ex-vivo and in-vitro studies performed in myoblastic C2C12 murine cell lines and using muscle cultures obtained from patients with *TRIP4* variants, suggested that ASC-1 might have a different role, being involved in muscle development (late myogenesis and/or myotube growth) (Davignon et al., 2016). Thus, these reports although describing patients sharing clinical features, added further speculation on the underlying disease mechanism(s), the former being a motor neuron disease/neuropathy and the second a primary myopathy. It is also possible that this complex might be important in regulating a wide variety of molecular targets, and that both mechanisms can become compromised at different levels depending on the ASC-1 subunit primarily affected. As a small amount of muscle tissue from the deceased patient is still preserved in paraffin, it is theoretically possible to extract messenger RNA to perform a whole-transcriptomic analysis. From the analysis of genome-wide differential RNA expression (compared with suitable controls), further insights into biological pathways and molecular mechanisms underlying this disease could be obtained. Also, the identification of additional ASC-1-related patients might further characterize this new clinical entity.

5.2.5 DEVELOPMENT OF A MOUSE MODEL FOR THE HUMAN *LAMA2* P.THR821PRO VARIANT

During this research additional Portuguese cases with limb-girdle muscular dystrophies have been characterized. Interestingly, one recurrent genetic cause was the *LAMA2* missense variant - p.Thr821Pro – so far found in seventeen patients with Portuguese ancestry. In homozygosity, this variant causes a mild muscular dystrophy phenotype, but it still manifests typical brain white matter changes. The pathophysiology mechanism of this variant has not been determined especially in the CNS. Our main goal is to study its effect in a mouse model from a multisystemic perspective and to evaluate the possibility of correcting this phenotype. Thus, further understanding will be obtained how this specific missense substitution can lead to a defective laminin-α2 protein and the underlying mechanisms leading to muscular dystrophy and cerebral white matter changes. Although several transgenic mice have been generated as a disease model for MDC1A (Sunada et al., 1994; Miyagoe et al., 1997; Kuang et al., 1998; Besse et al., 2003), for milder muscular phenotypes that retain typical central nervous system involvement, no suitable animal models have been developed until now at present. Primarily, to study the impact of the p.Thr821Pro variant, a

knock-in mouse model could be effectively generated through the CRISPR/Cas9 system (Sander & Joung, 2014) or using a recent alternative based on CRISPR/Cpf1 (Zetsche et al., 2015). Morphologic, histological and immunohistochemistry analysis could be performed in different mutant mouse tissues, including skeletal muscle and brain. Based on this model it would be also possible to perform pull-down assays (or co-immunoprecipitation) using samples from brain tissue and compare them with control mice to eventually identify novel protein-protein interactions. This project/task could potentially expand our current knowledge about the impact of this genetic defect, especially on the brain-blood barrier and the connection between the sarcolemma and extracellular matrix.

5.2.6 EVALUATING MALE REPRODUCTIVE ISSUES DUE TO LAMININ-α2 DEFICIENCY

It is known that laminin-α2 expression is not confined to muscle and brain cells. Homozygous mice models for *LAMA2* deficiency are usually infertile likely due to abnormal testicular basal membrane and impaired testis maturation (Häger et al., 2005; Gao et al., 2017). The laminin α2 present in testis basal membrane was implicated in the regulation of the axis that functionally links the testis basal membrane and Sertoli cell blood-testis barrier (Gao et al., 2017). However, the effect of laminin-α2 deficiency over human fertility has not been studied or characterized so far. This could be accomplished through two different approaches. The first would be an assessment of male reproductive issues in milder *LAMA2*-MD patients that have reached adulthood. A second would require a knock-in mouse model for a missense variant as proposed in section 5.2.5. Male fertility would be evaluated by the analysis of mouse sperm concentration, motility and morphology, and by the existence of offspring. Finally, we could also evaluate if this *LAMA2*-related infertility can be effective restored. To evaluate this, spermatogonial stem cells could be isolated from knock-in mice and corrected back to wild-type genotype using genome-editing techniques, followed by autotransplantation on mouse testis. This research can potentially lead to a new line of research concerning male infertility.

5.2.7 FINAL CONSIDERATIONS

The present thesis summarizes an extensive effort to bring new genetic discovery in HM. it is clear that we are on the brink of yet another revolution in the research and diagnostics of human diseases, driven by the advent of NGS-enabled technology. Besides advancements in AI-machine learning and computational infrastructures, continuous translational genetic research is needed to establish the association between genetic data and the diseases/phenotypes related to skeletal muscle.

Chapter 6

REFERENCES

REFERENCES

Aartsma-Rus, A., Deutekom, J. C., Fokkema, I. F., Ommen, G. B., & Dunnen, J. T. (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: An overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle & Nerve*, *34(*2), 135-144. *doi:* 10.1002/mus.20586.

Abath-Neto, O. L. (2014). Estudo clínico, histológico e molecular da miopatia centronuclear (Doctoral dissertation). Retrieved from http://www.teses.usp.br/teses/disponiveis/ 5/5138/tde-13012015-113703/publico/OsorioLopesAbathNeto.pdf. *doi:* 10.11606/T.5. 2014.tde- 13012015-113703.

Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. (2002). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, *30*(1), 97-101. *doi*: 10.1038/ng786.

Abel, H. J., Duncavage, E. J., Becker, N., Armstrong, J. R., Magrini, V. J., & Pfeifer, J. D. (2010). SLOPE: A quick and accurate method for locating non-SNP structural variation from targeted next-generation sequence data. *Bioinformatics*, *26*(21), 2684-2688. *doi:* 10.1093/bioinformatics/btq528.

Adey, A., Morrison, H. G., Asan, Xun, X., Kitzman, J. O., & Turner, E. H., et al. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology*, *11*(12). *doi:* 10.1186/gb-2010-11-12-r119.

Adzhubei, I. A., Schmidt S., Peshkin, L., Ramensky, V. E., Gerasimova, A., & Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), 248-249. *doi:* 10.1038/nmeth0410-248.

Agrawal, P., Greenleaf, R., Tomczak, K., Lehtokari, V., Wallgren-Pettersson, C., & Wallefeld, W., et al. (2007). Nemaline Myopathy with Minicores Caused by Mutation of the CFL2 Gene Encoding the Skeletal Muscle Actin–Binding Protein, Cofilin-2. *The American Journal of Human Genetics*, *80*(1), 162-167. *doi:* 10.1086/510402.

Agrawal, P. B, Pierson, C. R., Joshi, M., Liu, X., Ravenscroft, G., & Moghadaszadeh, B., et al. (2014). SPEG interacts with myotubularin, and its deficiency causes centronuclear myopathy with dilated cardiomyopathy. *American Journal of Human Genetics*, *95*(2), 218-226. *doi*: 10.1016/j.ajhg.2014.07.004

Aird, D., Ross, M. G., Chen, W., Danielsson, M., Fennell, T., & Russ, C., et al. (2011). Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology*, *12*(2). *doi:* 10.1186/gb-2011-12-2-r18.

Albert, T. J., Molla, M. N., Muzny, D. M., Nazareth, L., Wheeler, D., & Song, X., **et al.** (2007). Direct selection of human genomic loci by microarray hybridization. *Nature Methods*, *4*(11), 903-905. *doi:* 10.1038/nmeth1111.

Allamand, V. & Guicheney, P. (2002). Merosin-deficient congenital muscular dystrophy, autosomal recessive (MDC1A, MIM#156225, LAMA2 gene coding for alpha2 chain of laminin). *European Journal of Human Genetics, 10*(2), 91-94. *doi*: 10.1038/sj.ejhg.5200743.

Alkuraya, F. S. (2010). Homozygosity mapping: One more tool in the clinical geneticist's toolbox. *Genetics in Medicine, 12*(4), 236-239. *doi*: 10.1097/GIM.0b013e3181ceb95d.

Amburgey, K., McNamara, N., Bennett, L. R., McCormick, M. E., Acsadi, G., & Dowling, J. J. (2011). Prevalence of congenital myopathies in a representative pediatric united states population. *Annals of neurology, 70*(4), 662-665. *doi:* 10.1002/ana.22510.

Amburgey, K., Bailey, A., Hwang, J.H., Tarnopolsky, M.A., Bonnemann, C.G., & Medne, L., et al. (2013). Genotype-phenotype correlations in recessive RYR1-related myopathies. *Orphanet journal of rare diseases*, 8, 117. *doi:* 10.1186/1750-1172-8-117.

Amoasii, L., Hildyard, J. C., Li, H., Sanchez-Ortiz, E., Mireault, A., & Caballero, D., et al. (2018). Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science, 362*(6410), 86-91. *doi:*1 0.1126/science.aau1549.

Andrews, S. (2016). FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved October 1, 2018, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Angelini, C. & Fanin, M. (2017). Calpainopathy. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1313/.

Ankala, A., da Silva, C., Gualandi, F., Ferlini, A., Bean, L. J., & Collins, C., et al. (2015). A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Annals of Neurology, 77*(2), 206-214. *doi:* 10.1002/ana.24303.

Antonarakis, S. E., Krawczak, M., & Cooper, D. N. (2000). Disease-causing mutations in the human genome. *European Journal of Pediatrics, 159*(S3). *doi:* 10.1007/pl00014395.

Aoyama, C., Ohtani, A. & Ishidate, K. (2002). Expression and characterization of the active molecular forms of choline/ethanolamine kinase-alpha and -beta in mouse tissues, including carbon tetrachloride-induced liver. *The Biochemical Journal, 363*(pt3), 777-784*. doi:* 10.1042/bj3630777.

Aoyama, C., Liao, H. & Ishidate, K. (2014). Structure and function of choline kinase isoforms in mammalian cells. *Progress in Lipid Research, 43*(3), 266-281. *doi*: 10.1016/j.plipres.2003.12.001.

Arbogast, S., Beuvin, M., Fraysse, B., Zhou, H., Muntoni, F., & Ferreiro, A. (2009). Oxidative stress in SEPN1-related myopathy: From pathophysiology to treatment. *Annals of Neurology, 65*(6), 677-686. *doi:*10.1002/ana.21644.

Aumailley, M., Bruckner-Tuderman, L., Carter, W. G., Deutzmann, R., Edgar, D., & Ekblom, P., et al. (2005). A simplified laminin nomenclature. *Matrix Biology*, *24*(5), 326-332. *doi:* 10.1016/j.matbio.2005.05.006.

Awano, H., Malueka, R. G., Yagi, M., Okizuka, Y., Takeshima, Y., & Matsuo, M. (2010). Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA. *Journal of Human Genetics. 55*(12), 785-790. *doi:* 10.1038/jhg.2010.111.

Balci, B., Uyanik, G., Dincer, P., Gross, C., Willer, T., & Talim, B., et al. (2005). An autosomal recessive limb girdle muscular dystrophy (LGMD2) with mild mental retardation is allelic to Walker–Warburg syndrome (WWS) caused by a mutation in the POMT1 gene. *Neuromuscular Disorders, 15*(4), 271-275. *doi*: 10.1016/j.nmd.2005.01.013.

Bansal, D., Miyake, K., Vogel, S. S., Groh, S., Chen, C., & Williamson, R., et al. (2003). Defective membrane repair in dysferlin-deficient muscular dystrophy. *Nature, 423*(6936), 168-172. *doi:* 10.1038/nature01573.

Barthélémy, F., Defour, A., Lévy, N., Krahn, M., & Bartoli, M. (2018). Muscle Cells Fix Breaches by Orchestrating a Membrane Repair Ballet. *Journal of Neuromuscular Diseases, 5*(1), 21-28. *doi:*10.3233/jnd-170251.

Barone, R., Aiello, C., Race, V., Morava, E., Foulquier, F., & Riemersma, M., et al. (2012). DPM2-CDG: A muscular dystrophy-dystroglycanopathy syndrome with severe epilepsy. *Annals of Neurology, 72*(4), 550-558. *doi:* 10.1002/ana.23632.

Barp, A., Malfatti, E., Metay, C., Jobic, V., Carlier, R. Y. & Laforet, P. (2018). The first French case of MATR3-related distal myopathy: Clinical, radiological and histopathological characterization. *Reviews Neurology (Paris), 174*(10), 752-755. *doi:* 10.1016/j.neurol.2017.08.004.

Bashir, R., Britton, S., Strachan, T., Keers, S., Vafiadaki, E., & Lako, M., et al. (1998). A gene related to Caenorhabditis elegans spermatogenesis factor fer-1is mutated in limb-girdle muscular dystrophy type 2B. *Nature Genetics, 20*(1), 37-42. *doi:* 10.1038/1689.

Batonnet-Pichon, S., Behin, A., Cabet, E., Delort, F., Vicart, P., & Lilienbaum, A. (2017). Myofibrillar Myopathies: New Perspectives from Animal Models to Potential Therapeutic Approaches. *Journal of Neuromuscular Diseases, 4*(1), 1-15. *doi:* 10.3233/jnd-160203.

Béhin, A., Laforêt, P., Malfatti, E., Pellegrini, N., Hayashi, Y., & Carlier, R. Y. et al. (2013). P.15.10 Megaconial myopathy presenting as a progressive limb-girdle myopathy. *Neuromuscular Disorders*, 23 (9-10), 821. *doi*: 10.1016/j.nmd.2013.06.635.

Béhin, A., Salort-Campana, E., Wahbi, K., Richard, P., Carlier, R., & Carlier, P., et al. (2015). Myofibrillar myopathies: State of the art, present and future challenges. *Revue Neurologique, 171*(10), 715-729. *doi:* 10.1016/j.neurol.2015.06.002.

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., & Antipenko, A., et al. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences of USA*, *112*(17), 5473-5478. *doi:* 10.1073/pnas.1418631112.

Beltrán-Valero de Bernabé, D., Currier, S., Steinbrecher, A., Celli, J., Beusekom, E. V., & Zwaag, B. V., et al. (2002). Mutations in the O-Mannosyltransferase Gene POMT1 Give Rise to the Severe Neuronal Migration Disorder Walker-Warburg Syndrome. *The American Journal of Human Genetics, 71*(5), 1033-1043. doi: 10.1086/342975.

Beltrán-Valero de Bernabé, D., Voit, T., Longman, C., Steinbrecher, A., Straub, V., Yuva, Y., et al. (2004). Mutations in the FKRP gene can cause muscle-eye-brain disease and Walker-Warburg syndrome. *Journal of Medical Genetics, 41*(5). doi:10.1136/jmg.2003.013870.

Belzil, V. V., Katzman, R. B., & Petrucelli, L. (2016). ALS and FTD: An epigenetic perspective. Acta Neuropathologica, *132*(4), 487-502. *doi:* 10.1007/s00401-016-1587-4.

Benarroch, E. E. (2017). Anoctamins (TMEM16 proteins). *Neurology, 89*(7), 722-729. *doi:* 10.1212/wnl.0000000000004246.

Bengtsson, L., & Otto, H. (2008). LUMA interacts with emerin and influences its distribution at the inner nuclear membrane. *Journal of Cell Science, 121*(4), 536-548. *doi:* 10.1242/jcs.019281.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., & Brown, C. G., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*(7218), 53-59. *doi:*10.1038/nature07517.

Berk, J. M., Tifft, K. E., & Wilson, K. L. (2013). The nuclear envelope LEM-domain protein emerin. *Nucleus, 4*(4), 298-314. *doi:* 10.4161/nucl.25751.

Besse, S., Allamand, V., Vilquin, J. T., Li, Z., Poirier, C., & Vignier, N., et al. (2003). Spontaneous muscular dystrophy caused by a retrotransposal insertion in the mouse laminin alpha2 chain gene. *Neuromuscular Disorders*,*13*(3), 216-222. *doi:* 10.1016/s0960-8966(02)00278-x.

Bevilacqua, J. A., Bitoun, M., Biancalana, V., Oldfors, A., Stoltenburg, G., & Claeys, K. G. et al. (2009). "Necklace" fibers, a new histological marker of late-onset MTM1-

related centronuclear myopathy. *Acta neuropathologica*, *117*(3), 283-291. *doi:* 10.1007/s00401-008-0472-1.

Bevilacqua, J. A., Monnier, N., Bitoun, M., Eymard, B., Ferreiro, A., & Monges, S., et al. (2011). Recessive RYR1 mutations cause unusual congenital myopathy with prominent nuclear internalization and large areas of myofibrillar disorganization. *Neuropathology and Applied Neurobiology*, *37*(3), 271-284. *doi:* 10.1111/j.1365-2990.2010.01149.x.

Bhargava, A., Lahaye, X., & Manel, N. (2018). Let me in: Control of HIV nuclear entry at the nuclear envelope. *Cytokine & Growth Factor Reviews, 40*, 59-67. *doi:* 10.1016/j.cytogfr.2018.02.006.

Bhowmik, A. D., Dalal, A. B., Matta, D., Sundaram, C. & Aggarwal, S. (2016). Targeted Next Generation Sequencing Identifies a Novel Deletion in LAMA2 Gene in a Merosin Deficient Congenital Muscular Dystrophy Patient. *Indian Journal of Pediatrics*, 83(4), 354-355. *doi*: 10.1007/s12098-015-1822-3

Biancalana, V., Caron, O., Gallati, S., Baas, F., Kress, W., & Novelli, G. et al. (2003). Characterisation of mutations in 77 patients with X-linked myotubular myopathy, including a family with a very mild phenotype. *Human Genetics*, *112*(2), 135-142. *doi:* 10.1007/s00439-002-0869-1.

Biancalana, V., Beggs, A.H., Das, S., Jungbluth, H., Kress, W., & Nishino, I., et al. (2012). Clinical utility gene card for: Centronuclear and myotubular myopathies. *European Journal of Human Genetics*, *20*(10). *doi:* 10.1038/ejhg.2012.91.

Biancalana, V., Scheidecker, S., Miguet, M., Laquerrière, A., Romero, N. B., & Stojkovic, T., et al. (2017). Affected female carriers of MTM1 mutations display a wide spectrum of clinical and pathological involvement: Delineating diagnostic clues. *Acta Neuropathologica, 134*(6), 889-904. *doi:* 10.1007/s00401-017-1748-0.

Biancheri, R., Falace, A., Tessa, A., Pedemonte, M., Scapolan, S., & Cassandrini, D., et al. (2007). POMT2 gene mutation in limb-girdle muscular dystrophy with inflammatory changes. *Biochemical and Biophysical Research Communications, 363*(4), 1033-1037. *doi:* 10.1016/j.bbrc.2007.09.066.

Bione, S., Maestrini, E., Rivella, S., Mancini, M., Regis, S., Romeo, G., & Toniolo, D. (1994). Identification of a novel X-linked gene responsible for Emery-Dreifuss muscular dystrophy. *Nature Genetics, 8*(4), 323-327. *doi:* 10.1038/ng1294-323.

Bitoun, M., Maugenre, S., Jeannet, P. Y., Lacène, E., Ferrer, X., & Laforêt, P., et al. (2005). Mutations in dynamin 2 cause dominant centronuclear myopathy. *Nature Genetics*, *37*(11), 1207-1209. *doi:* 10.1038/ng1657.

Bitoun, M., Bevilacqua, J. A., Prudhon, B., Maugenre, S., Taratuto, A. L., & Monges, S., et al. (2007). Dynamin 2 mutations cause sporadic centronuclear myopathy with neonatal onset. *Annals of Neurology, 62*(6), 666-670. *doi:* 10.1002/ana.21235.

Bitoun, M., Durieux, A., Prudhon, B., Bevilacqua, J. A., Herledan, A., & Sakanyan, V., et al. (2009). Dynamin 2 mutations associated with human diseases impair clathrin-mediated receptor endocytosis. *Human Mutation, 30*(10), 1419-1427. *doi:* 10.1002/humu.21086.

Bittles, A. H. (2001). Consanguinity and its relevance to clinical genetics. *Clinical genetics, 60*(2), 89-98. *doi:* 10.1034/j.1399-0004.2001.600201.x.

Bladen, C. L., Salgado, D., Monges, S., Foncuberta, M. E., Kekou, K., & Kosma, K., et al. (2015). The TREAT-NMD DMD Global Database: Analysis of More than 7,000 Duchenne Muscular Dystrophy Mutations. *Human Mutation, 36*(4), 395-402. *doi:* 10.1002/humu.22758.

Blais, A. (2015). Myogenesis in the genomics era. *Journal of Molecular Biology, 427*(11), 2023-2038. *doi:* 10.1016/j.jmb.2015.02.009.

Bogdanovich, S., Krag, T. O., Barton, E. R., Morris, L. D., Whittemore, L., Ahima, R. S., & Khurana, T. S. (2002). Functional improvement of dystrophic muscle by myostatin blockade. *Nature, 420*(6914), 418-421. *doi:* 10.1038/nature01154.

Bögershausen, N., Shahrzad, N., Chong, J. X., von Kleist-Retzow, J. C., Stanga, D., & Li, Y., et al. (2013). Recessive TRAPPC11 mutations cause a disease spectrum of limb girdle muscular dystrophy and myopathy with movement disorder and intellectual disability. *American Journal of Human Genetics*, 93(1), 181-190. *doi:* 10.1016/j.ajhg.2013.05.028.

Böhm, J., Chevessier, F., De Paula, A., Koch, C., Attarian, S., & Feger, C. et al. (2013a). Constitutive Activation of the Calcium Sensor STIM1 Causes Tubular-Aggregate Myopathy. *The American Journal of Human Genetics, 92*(2), 271-278. *doi:* 10.1016/j.ajhg.2012.12.007.

Böhm, J., Vasli, N., Malfatti, E., Le Gras, S., Feger, C., & Jost, B., et al. (2013b). An integrated diagnosis strategy for congenital myopathies. *PLoS One, 8*(6), e67527. *doi:* 10.1371/journal.pone.0067527.

Böhm, J., & Laporte, J. (2018). La myopathie à agrégats tubulaires et le syndrome de Stormorken. *Médecine/sciences, 34*, 26-31. *doi:* 10.1051/medsci/201834s208.

Boland-Freitas, R., Graham, J., Davis, M., Geevasinga, N., Vucic, S., & Ng, K. (2016). Late-onset distal myopathy of the upper limbs due to p.Ile151Val mutation in the valosin-containing protein. *Muscle & Nerve, 54*(1), 165-166. *doi:* 10.1002/mus.25073.

Bolduc, V., Marlow, G., Boycott, K. M., Saleki, K., Inoue, H., & Kroon, J., et al. (2010). Recessive Mutations in the Putative Calcium-Activated Chloride Channel Anoctamin

5 Cause Proximal LGMD2L and Distal MMD3 Muscular Dystrophies. *The American Journal of Human Genetics, 86*(2), 213-221. *doi:* 10.1016/j.ajhg.2009.12.013.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. *doi:* 10.1093/bioinformatics/btu170.

Bonne, G., Barletta, M. R., Varnous, S., Bécane, H., Hammouda, E., & Merlini, L., et al. (1999). Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nature Genetics, 21*(3), 285-288. *doi:* 10.1038/6799.

Bonne, G., & Quijano-Roy, S. (2013). Emery–Dreifuss muscular dystrophy, laminopathies, and other nuclear envelopathies. *Handbook of Clinical Neurology Pediatric Neurology Part III,* 1367-1376. *doi:* 10.1016/b978-0-444-59565-2.00007-1.

Bonne, G., Leturcq, F., & Ben Yaou, R. (2015). Emery-Dreifuss Muscular Dystrophy. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1436/.

Bonne, G., & Rivier, F. (2018). GeneTable of Neuromuscular Disorders. Retrieved December 1, 2018, from http://www.musclegenetable.fr/.

Bönnemann, C. G., Modi, R., Noguchi, S., Mizuno, Y., Yoshida, M., & Gussoni, E., et al. (1995). β–sarcoglycan (A3b) mutations cause autosomal recessive muscular dystrophy with loss of the sarcoglycan complex. *Nature Genetics, 11*(3), 266-273. *doi:* 10.1038/ng1195-266.

Bönnemann, C. G., Rutkowski, A., Mercuri, E., Muntoni, F., & CMD Outcomes Consortium. (2011). 173rd ENMC International Workshop: congenital muscular dystrophy outcome measures 5-7 March 2010, Naarden, The Netherlands. *Neuromuscular Disorders*, *21*(7), 513-522. *doi:* 10.1016/j.nmd.2011.04.004.

Bönnemann, C. G., Wang, C. H., Quijano-Roy, S., Deconinck, N., Bertini, E., & Ferreiro, A., et al. (2014). Diagnostic approach to the congenital muscular dystrophies. *Neuromuscular Disorders*, *24*(4), 289-311. *doi:* 10.1016/j.nmd.2013.12.011.

Bouchet-Séraphin, C., Vuillaumier-Barrot, S., & Seta, N. (2015). Dystroglycanopathies: About Numerous Genes Involved in Glycosylation of One Single Glycoprotein. *Journal of Neuromuscular Diseases*, *2*(1), 27-38. *doi:* 10.3233/JND-140047.

Boycott, K. M., Vanstone, M. R., Bulman, D. E., & MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews Genetics*, *14*(10), 681-691. *doi:* 10.1038/nrg3555.

Brady, G. F., Kwan, R., Cunha, J. B., Elenbaas, J. S., & Omary, M. B. (2018). Lamins and Lamin-Associated Proteins in Gastrointestinal Health and Disease. *Gastroenterology, 154*(6). *doi:* 10.1053/j.gastro.2018.03.026.

Bragg, L. M., Stone, G., Butler, M. K., Hugenholtz, P., & Tyson, G. W. (2013). Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. *PLoS Computational Biology*, *9*(4), e1003031. *doi:* 10.1371/journal.pcbi.1003031.

Brandsema, J., & Darras, B. (2015). Dystrophinopathies. *Seminars in Neurology, 35*(04), 369-384. *doi:* 10.1055/s-0035-1558982.

Brett, F. M., Costigan, D., Farrell, M. A., Heaphy, P., Thornton, J., & King, M. D. (1998). Merosin-deficient congenital muscular dystrophy and cortical dysplasia. *European Journal of Paediatric Neurology*, *2*(2), 77-82. *doi:* 10.1016/S1090-3798(98)80045-7.

Brislin, R. P. & Theroux, M. C. (2013). Core myopathies and malignant hyperthermia susceptibility: a review. *Paediatric Anaesthesia*, *23*(9), 834-841. doi: 10.1111/pan.12175.

Brockington, M., Yuva, Y., Prandini, P., Brown, S. C., Torelli, S., & Benson, M. A., et al. (2001). Mutations in the fukutin-related protein gene (FKRP) identify limb girdle muscular dystrophy 2I as a milder allelic variant of congenital muscular dystrophy MDC1C. *Human Molecular Genetics, 10*(25), 2851-2859. *doi:* 10.1093/hmg/10.25.2851.

Brouha, B., Schustak, J., Badge, R. M, Lutz-Prigge, S., Farley, A. H., & Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of USA*, *100*(9), 5280-5285. *doi:* 10.1073/pnas.0831042100.

Bruno, C., Sotgia, F., Gazzerro, E., Minetti, C. & Lisanti, M. P. (2015). Caveolinopathies. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1385/.

Buckingham, M. (2003). How the community effect orchestrates muscle differentiation. *Bioessays*, *25*(1):13-16. *doi:* 10.1002/bies.10221.

Buermans, H. P. & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta*, *1842*(10), 1932-1941. *doi:* 10.1016/j.bbadis.2014.06.015.

Burkin, D. J., & Kaufman, S. J. (1999). The α7β1 integrin in muscle development and disease. *Cell and Tissue Research, 296*(1), 183-190. *doi:* 10.1007/s004410051279.

Bushby, K. M., Collins, J., & Hicks, D. (2013). Collagen Type VI Myopathies. Advances in Experimental Medicine and Biology Progress in Heritable Soft Connective Tissue Diseases, 185-199. *doi:* 10.1007/978-94-007-7893-1_12.

Buyse, G. M., Voit, T., Schara, U., Straathof, C. S., Dangelo, M. G., & Bernert, G., et al (2015). Efficacy of idebenone on respiratory function in patients with Duchenne muscular dystrophy not using glucocorticoids (DELOS): A double-blind randomised placebo-controlled phase 3 trial. *The Lancet, 385*(9979), 1748-1757. *doi:* 10.1016/s0140-6736(15)60025-3.

Buysse, K., Riemersma, M., Powell, G., Reeuwijk, J. V., Chitayat, D., & Roscioli, T., et al. (2013). Missense mutations in β-1,3-N-acetylglucosaminyltransferase 1 (B3GNT1) cause Walker–Warburg syndrome. *Human Molecular Genetics, 22*(9), 1746-1754. *doi:* 10.1093/hmg/ddt021.

Byron, S. A., Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., & Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nature Reviews Genetics*, *17*(5), 257-271. *doi:* 10.1038/nrg.2016.10.

Cabrera-Serrano, M., Junckerstorff, R. C., Atkinson, V., Sivadorai, P., Allcock, R. J., & Lamont, P. et al. (2015). Novel CHKB mutation expands the megaconial muscular dystrophy phenotype. *Muscle & Nerve*, *51*(1), 140-143. *doi:* 10.1002/mus.24446

Cacheux, M., Blum, A., Sébastien, M., Wozny, A. S., Brocard, J., & Mamchaoui, K., et al. (2015). Functional Characterization of a Central Core Disease RyR1 Mutation (p.Y4864H) Associated with Quantitative Defect in RyR1 Protein. *Journal of Neuromuscular Diseases, 2*(4), 421-432. *doi:* 10.3233/jnd-150073.

Calderón, J. C., Bolaños, P. & Caputo, C. The excitation-contraction coupling mechanism in skeletal muscle. *Biophysic Reviews*, *6*(1), 133-160. *doi:* 10.1007/s12551-013-0135-x.

Campbell, K. P. (1995). Three muscular dystrophies: loss of cytoskeleton-extracellular matrix linkage. Cell, 80(5), 675-679. doi: 10.1016/0092-8674(95)90344-5.

Carboni, N., Marrosu, G., Porcu, M., Mateddu, A., Solla, E., & Cocco, E., et al. (2011). Dilated cardiomyopathy with conduction defects in a patient with partial merosin deficiency due to mutations in the laminin-α2-chain gene: a chance association or a novel phenotype? *Muscle & Nerve*, *44*(5), 826-828. *doi:* 10.1002/mus.22228.

Carr, I. M., Bhaskar, S., O'Sullivan, J., Aldahmesh, M. A., Shamseldin, H. E., & Markham, A. F., et al. (2013). Autozygosity Mapping with Exome Sequence Data. *Human Mutation*, *34*(1), 50-56. *doi:* 10.1002/humu.22220.

Carrillo, N., Malicdan, M. C., & Huizing, M. (2018). GNE Myopathy: Etiology, Diagnosis, and Therapeutic Challenges. *Neurotherapeutics, 15*(4), 900-914. *doi:* 10.1007/s13311-018-0671-y.

Carss, K., Stevens, E., Foley, A., Cirak, S., Riemersma, M., & Torelli, S., et al. (2013). Mutations in GDP-Mannose Pyrophosphorylase B Cause Congenital and Limb-Girdle Muscular Dystrophies Associated with Hypoglycosylation of α-Dystroglycan. *The American Journal of Human Genetics, 93*(1), 29-41. *doi:* 10.1016/j.ajhg.2013.05.009.

Cartegni, L., Chew, S. L., & Krainer, A. R. (2002). Listening to silence and understanding nonsense: Exonic mutations that affect splicing. *Nature Reviews Genetics, 3*(4), 285-298. *doi:* 10.1038/nrg775.

Cassandrini, D., Trovato, R., Rubegni, A., Lenzi, S., Fiorillo, C., & Baldacci, J., et al. (2017). Congenital myopathies: clinical phenotypes and new diagnostic tools. *Italian Journal of Pediatrics, 43*(1), 101. *doi:* 10.1186/s13052-017-0419-z.

Castro-Gago, M., Dacruz-Alvarez, D., Pintos-Martínez, E., Beiras-Iglesias, A., Delmiro, A., & Arenas, J. et al. (2015). Exome sequencing identifies a CHKB mutation in Spanish patient with Megaconial Congenital Muscular Dystrophy and mtDNA depletion. *European Journal of Paediatric Neurology, 18*(6), 796-800. *doi:* 10.1016/j.ejpn.2014.06.005.

Celli, J., Dalgleish, R., Vihinen, M., Taschner, P. E. & den Dunnen, J. T. (2012). Curating gene variant databases (LSDBs): toward a universal standard. *Human Mutation, 33*(2), 291-297. *doi:* 10.1002/humu.21626.

Cetin, N., Balci-Hayta, B., Gundesli, H., Korkusuz, P., Purali, N., & Talim, B., et al. (2013). A novel desmin mutation leading to autosomal recessive limb-girdle muscular dystrophy: Distinct histopathological outcomes compared with desminopathies. *Journal of Medical Genetics, 50*(7), 437-443. *doi:* 10.1136/jmedgenet-2012-101487.

Ceyhan-Birsoy, O., Agrawal, P., Hidalgo, C., Schmitz-Abe, K., DeChene, E., & Swanson, L., et al. (2013). Recessive truncating titin gene, TTN, mutations presenting as centronuclear myopathy. *Neurology, 81*(14), 1205-1214. *doi:* 10.1212/WNL.0b013e3182a6ca62.

Chan, S. H., Foley, A. R., Phadke, R., Mathew, A. A., Pitt, M., Sewry, C. & Muntoni, F. (2014). Limb girdle muscular dystrophy due to LAMA2 mutations: diagnostic difficulties due to associated peripheral neuropathy. *Neuromuscular Disorders, 24*(8), 677-683. *doi:* 10.1016/j.nmd.2014.05.008.

Chardon, J. W., Smith, A., Woulfe, J., Pena, E., Rakhra, K., & Dennie, C., et al. (2015). LIMS2 mutations are associated with a novel muscular dystrophy, severe cardiomyopathy and triangular tongues. *Clinical Genetics, 88*(6), 558-564. doi:10.1111/cge.12561.

Chauveau, C., Bonnemann, C. G., Julien, C., Kho, A. L., Marks, H., & Talim, B., et al. (2014a). Recessive TTN truncating mutations define novel forms of core myopathy

with heart disease. *Human Molecular Genetics, 23*(4), 980-991. doi:10.1093/hmg/ddt494.

Chauveau, C., Rowell, J. & Ferreiro, A. (2014b). A rising titan: TTN review and mutation update. *Human Mutation*, *35*(9), 1046-1059. *doi:* 10.1002/humu.22611.

Chelly, J., & Desguerre, I. (2013). Progressive muscular dystrophies. *Handbook of Clinical Neurology Pediatric Neurology Part III,* 1343-1366. *doi:* 10.1016/b978-0-444-59565-2.00006-x

Chen, K., Wallis, J. W., Mclellan, M. D., Larson, D. E., Kalicki, J. M., & Pohl, C. S., et al. (2009). BreakDancer: An algorithm for high-resolution mapping of genomic structural variation. *Nature Methods*, *6*(9), 677-681. *doi:* 10.1038/nmeth.1363.

Cheng, Y. S., Champliaud, M. F., Burgeson, R. E., Marinkovich, M. P., & Yurchenco, P. D. (1997). Self-assembly of laminin isoforms. *The Journal of Biological Chemistry*, 272(50), 31525-31532.

Christianson, A., Howson, C.P. & Modell, B. (2006). Global report on birth defects: the hidden toll of dying and disabled children, New York. Available at: http://www.marchofdimes.org/materials/global-report-on-birth-defects-the-hidden-toll-of-dying-and-disabled-children-full-report.pdf (Accessed November 15, 2016).

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., & Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly, 6*(2), 80-92. *doi:* 10.4161/fly.19695.

Cirak, S., Deimling, F. V., Sachdev, S., Errington, W. J., Herrmann, R., & Bönnemann, C., et al. (2010). Kelch-like homologue 9 mutation is associated with an early onset autosomal dominant distal myopathy. *Brain, 133*(7), 2123-2135. *doi:* 10.1093/brain/awq108.

Clarke, N. F., Waddell, L. B., Cooper, S. T., Perry, M., Smith, R. L., & Kornberg, A. J, et al. (2010). Recessive mutations in RYR1 are a common cause of congenital fiber type disproportion. *Human mutation*, *31*(7), E1544-1550. *doi:* 10.1002/humu.21278.

Clarke, N. F. (2011a). Congenital Fiber-Type Disproportion. *Seminars in Pediatric Neurology, 18*(4), 264-271. *doi:* 10.1016/j.spen.2011.10.008.

Clarke, N. F., Maugenre, S., Vandebrouck, A., Urtizberea, J. A., Willer, T., & Peat, R. A, et al. (2011b). Congenital muscular dystrophy type 1D (MDC1D) due to a large intragenic insertion/deletion, involving intron 10 of the LARGE gene. *European Journal of Human Genetics*, *19*(4), 452-427. *doi:* 10.1038/ejhg.2010.212.

Clement, E., Mercuri, E., Godfrey, C., Smith, J., Robb, S., & Kinali, M., et al. (2008). Brain involvement in muscular dystrophies with defective dystroglycan glycosylation. *Annals of Neurology, 64*(5), 573-582. *doi:* 10.1002/ana.21482.

Cohn, R. D., Herrmann, R., Sorokin, L., Wewer, U. M., & Voit, T. (1998). Laminin alpha2 chain-deficient congenital muscular dystrophy: variable epitope expression in severe and mild cases. *Neurology*, *51*(1), 94-100. *doi:* 10.1212/WNL.51.1.94.

Colombo, I., Scoto, M., Manzur, A. Y., Robb, S. A., Maggi, L., & Gowda, V., et al. (2015). Congenital myopathies: Natural history of a large pediatric cohort. *Neurology*, *84*(1), 28-35. *doi:* 10.1212/WNL.0000000000001110.

Compton, A., Albrecht, D., Seto, J., Cooper, S., Ilkovski, B., & Jones, K. et al. (2008). Mutations in Contactin-1, a Neural Adhesion and Neuromuscular Junction Protein, Cause a Familial Form of Lethal Congenital Myopathy. *The American Journal of Human Genetics*, *83*(6), 714-724. *doi:* 10.1016/j.ajhg.2008.10.022.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., & Mcpherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology, 17*(1). *doi:* 10.1186/s13059-016-0881-8.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., & Habib, N., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science, 339*(6121), 819-823. *doi:* 10.1126/science.1231143.

Constantin, B. (2014). Dystrophin complex functions as a scaffold for signalling proteins. *Biochimica Et Biophysica Acta (BBA) - Biomembranes, 1838*(2), 635-642. *doi:* 10.1016/j.bbamem.2013.08.023.

Cooper, D. N., Mort, M., Stenson, P. D., Ball, E. V. & Chuzhanova, N. A. (2010). Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Human Genomics*, 4(6), 406-410. *doi:* 10.1186/1479-7364-4-6-406.

Costa, C., Oliveira, J., Gonçalves, A., Santos, R., Bronze-da-Rocha, E., & Rebelo, O., et al. (2013). A Portuguese case of Fukuyama congenital muscular dystrophy caused by a multi-exonic duplication in the fukutin gene. *Neuromuscular Disorders*, *23*(7), 557-561. *doi:* 10.1016/j.nmd.2013.03.005.

Couthouis, J., Raphael, A. R., Siskind, C., Findlay, A. R., Buenrostro, J. D., & Greenleaf, W. J., et al. (2014). Exome sequencing identifies a DNAJB6 mutation in a family with dominantly-inherited limb-girdle muscular dystrophy. *Neuromuscular Disorders*, *24*(5), 431-435. *doi:* 10.1016/j.nmd.2014.01.014.

Cowling, B. S., Cottle, D. L., Wilding, B. R., D'Arcy, C. E., Mitchell, C. A., & Mcgrath, M. J. (2011). Four and a half LIM protein 1 gene mutations cause four distinct human myopathies: A comprehensive review of the clinical, histological and pathological features. *Neuromuscular Disorders, 21*(4), 237-251. *doi:* 10.1016/j.nmd.2011.01.001.

Cullup, T., Lamont, P., Cirak, S., Damian, M., Wallefeld, W., & Gooding, R., et al. (2012). Mutations in MYH7 cause Multi-minicore Disease (MmD) with variable cardiac involvement. *Neuromuscular Disorders, 22*(12), 1096-1104. *doi:* 10.1016/j.nmd.2012.06.007.

Czeschik, J. C., Hehr, U., Hartmann, B., Lüdecke, H. J., Rosenbaum, T., Schweiger, B. & Wieczorek, D. (2013). 160 kb deletion in ISPD unmasking a recessive mutation in a patient with Walker-Warburg syndrome. *European Journal of Medical Genetics*, *56*(12), 689-694. *doi:* 10.1016/j.ejmg.2013.09.014.

Dahl, N., Hu, L. J., Chery, M., Fardeau, M., Gilgenkrantz, S., & Nivelon-Chevallier, A., et al. (1995). Myotubular myopathy in a girl with a deletion at Xq27-q28 and unbalanced X inactivation assigns the MTM1 gene to a 600-kb region. *American Journal of Human Genetics*, *56*(5), 1108-1115.

Dalgleish, R. (2016). LSDBs and How They Have Evolved. *Human Mutation*, 37(6), 532-539. *doi:* 10.1002/humu.22979.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., & DePristo, M. A. et al. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. *doi:* 10.1093/bioinformatics/btr330.

Davignon, L., Chauveau, C., Julien, C., Dill, C., Duband-Goulet, I., & Cabet, E., et al. (2016). The transcription coactivator ASC-1 is a regulator of skeletal myogenesis, and its deficiency causes a novel form of congenital muscle disease. *Human Molecular Genetics*, 25(8), 1559-1573. *doi:* 10.1093/hmg/ddw033.

Davis, M. R., Haan, E., Jungbluth, H., Sewry, C., North, K., & Muntoni, F., et al. (2003). Principal mutation hotspot for central core disease and related myopathies in the C-terminal transmembrane region of the RYR1 gene. *Neuromuscular disorders*, *13*(2), 151-157. *doi:* 10.1016/S0960-8966(02)00218-3.

de Gouyon, B. M., Zhao, W., Laporte, J., Mandel, J. L., Metzenberg, A., & Herman, G. E. (1997). Characterization of mutations in the myotubularin gene in twenty six patients with X-linked myotubular myopathy. *Human Molecular Genetics*, *6*(9), 1499-1504. *doi:* 10.1093/hmg/6.9.1499.

Del Fabbro, C., Scalabrin, S., Morgante, M., & Giorgi, F. M. (2013). An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS ONE*, *8*(12). *doi:* 10.1371/journal.pone.0085024.

den Dunnen, J. T. & Antonarakis, S. E. (2000). Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Human Mutation*, *15*(1), 7-12. *doi:* 10.1002/(SICI)1098-1004(200001)15:1<7::AID-HUMU4>3.0.CO;2-N.

den Dunnen, J. T. & Antonarakis, S. E. (2001). Nomenclature for the description of human sequence variations. *Human Genetics*, *109*(1), 121-124. *doi:* 10.1007/s004390100505.

Dean, M., Rashid, S., Kupsky, W., Moore, S. A. & Jiang, H. (2017). Child Neurology: LAMA2 muscular dystrophy without contractures. *Neurology*, *88*(21), e199-e203. doi: 10.1212/WNL.0000000000003958.

Deangelis, M. M., Wang, D. G., & Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research*, *23*(22), 4742-4743. *doi:* 10.1093/nar/23.22.4742.

DeChene, E. T., Kang, P. B., & Beggs, A. H. (2013). Congenital Fiber-Type Disproportion. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1259/.

Deodato, F., Sabatelli, M., Ricci, E., Mercuri, E., Muntoni, F., & Sewry, C., et al. (2002). Hypermyelinating neuropathy, mental retardation and epilepsy in a case of merosin deficiency. *Neuromuscular Disorders*, 12(4), 392-398. *doi:* 10.1016/S0960-8966(01)00312-1.

Desmet, F. O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., & Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, *37*(9), e67. *doi:* 10.1093/nar/gkp215.

Desvignes, J., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Béroud, C., & Salgado, D. (2018). VarAFT: A variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Research*, *46*(W1). *doi:* 10.1093/nar/gky471.

Di Blasi, C., Mora, M., Pareyson, D., Farina, L., Sghirlanzoni, A., & Vignier, N., et al. (2000). Partial laminin alpha2 chain deficiency in a patient with myopathy resembling inclusion body myositis. *Annals of Neurology*, *47*(6), 811-816. *doi:* 10.1002/1531-8249(200006)47:6<811::AID-ANA16>3.0.CO;2-9.

Di Blasi, C., He Y, Morandi, L., Cornelio, F., Guicheney, P., & Mora, M. (2001). Mild muscular dystrophy due to a nonsense mutation in the LAMA2 gene resulting in exon skipping. *Brain*, *124*(Pt 4), 698-704. *doi:* 10.1093/brain/124.4.698.

Di Gioia, S., Connors, S., Matsunami, N., Cannavino, J., Rose, M., & Gilette, N. et al. (2017). A defect in myoblast fusion underlies Carey-Fineman-Ziter syndrome. *Nature Communications*, *8*, 16077. *doi:* 10.1038/ncomms16077.

Di Muzio, A., De Angelis, M. V., Di Fulvio, P., Ratti, A., Pizzuti, A., & Stuppia, L., et al. (2003). Dysmyelinating sensory-motor neuropathy in merosin-deficient congenital muscular dystrophy. *Muscle & Nerve*, *27*(4), 500-506. doi: 10.1002/mus.10326.

Dijk, E. L., Jaszczyszyn, Y., & Thermes, C. (2014). Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research*, *322*(1), 12-20. *doi:* 10.1016/j.yexcr.2014.01.008.

Ding, H., Beckers, M., Plaisance, S., Marynen, P., Collen, D., & Belayew, A. (1998). Characterization of a Double Homeodomain Protein (DUX1) Encoded by a cDNA Homologous to 3.3 Kb Dispersed Repeated Elements. *Human Molecular Genetics, 7*(11), 1681-1694. *doi:* 10.1093/hmg/7.11.1681.

Ding, J., Zhao, D., Du, R., Zhang, Y., Yang, H., & Liu, J., et al. (2016). Clinical and molecular genetic analysis of a family with late-onset LAMA2-related muscular dystrophy. *Brain & Development*, *38*(2), 242-249. doi: 10.1016/j.braindev.2015.08.005.

Domingos, J., Sarkozy, A., Scoto, M., & Muntoni, F. (2017). Dystrophinopathies and Limb-Girdle Muscular Dystrophies. *Neuropediatrics, 48*(04), 262-272. *doi:* 10.1055/s-0037-1601860.

Donkervoort, S., Bonnemann, C. G., Loeys, B., Jungbluth, H., & Voermans, N. C. (2015). The neuromuscular differential diagnosis of joint hypermobility. *American journal of medical genetics Part C, Seminars in medical genetics*, *169C*(1):23-42. *doi:* 10.1002/ajmg.c.31433.

Donner, K., Ollikainen, M., Ridanpää, M., Christen, H., Goebel, H., & de Visser, M., et al. (2002). Mutations in the β-tropomyosin (TPM2) gene – a rare cause of nemaline myopathy. *Neuromuscular Disorders*, *12*(2), 151-158. *doi:* 10.1016/S0960-8966(01)00252-8.

Donner, K., Sandbacka, M., Lehtokari, V., Wallgren-Pettersson, C., & Pelin, K. (2004). Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts. *European Journal of Human Genetics, 12*(9), 744-751. *doi:* 10.1038/sj.ejhg.5201242.

Dowling, J. J., Vreede, A. P., Low, S. E., Gibbs, E. M., Kuwada, J. Y., Bonnemann, C. G., & Feldman, E. L. (2009). Loss of Myotubularin Function Results in T-Tubule Disorganization in Zebrafish and Human Myotubular Myopathy. *PLoS Genetics, 5*(2). *doi:* 10.1371/journal.pgen.1000372.

Dowling, J. J., Arbogast, S., Hur, J., Nelson, D. D., Mcevoy, A., & Waugh, T., et al. (2012). Oxidative stress and successful antioxidant treatment in models of RYR1-related myopathy. *Brain, 135*(4), 1115-1127. *doi:* 10.1093/brain/aws036.

Dowling, J. J., Lawlor, M. W. & Dirksen, R. T. (2014). Triadopathies: an emerging class of skeletal muscle diseases. *Neurotherapeutics*, *11*(4), 773-85. *doi:* 10.1007/s13311-014-0300-3.

Duarte, S. T., Oliveira, J., Santos, R., Pereira, P., Barroso, C., & Conceição, I., et al. (2011). Dominant and recessive RYR1 mutations in adults with core lesions and mild muscle symptoms. *Muscle & Nerve*, *44*(1), 102-108. *doi:* 10.1002/mus.22009.

Duff, R., Tay, V., Hackman, P., Ravenscroft, G., Mclean, C., & Kennedy, P., et al. (2011). Mutations in the N-terminal Actin-Binding Domain of Filamin C Cause a Distal Myopathy. *The American Journal of Human Genetics, 88*(6), 729-740. *doi:* 10.1016/j.ajhg.2011.04.021.

Durbeej, M. (2015). Laminin-α2 Chain-Deficient Congenital Muscular Dystrophy: Pathophysiology and Development of Treatment. *Current Topic in Membranes*, *76*, 31-60. *doi:* 10.1016/bs.ctm.2015.05.002.

Durieux, A., Vignaud, A., Prudhon, B., Viou, M. T., Beuvin, M., & Vassilopoulos, S., et al. (2010). A centronuclear myopathy-dynamin 2 mutation impairs skeletal muscle structure and function in mice. *Human Molecular Genetics, 19*(24), 4820-4836. *doi:* 10.1093/hmg/ddq413.

Eberwine, J., Sul, J., Bartfai, T., & Kim, J. (2014). The promise of single-cell sequencing. *Nature Methods, 11*(1), 25-27. *doi:* 10.1038/nmeth.2769.

Edwards, J. S. (2008) Polony Sequencing: History, Technology, and Applications. In: Janitz, M., editors. *Next Generation Genome Sequencing,* 57–76. *doi:* 10.1002/9783527625130.

Eilbeck, K., Quinlan, A., & Yandell, M. (2017). Settling the score: Variant prioritization and Mendelian disease. *Nature Reviews Genetics*, *18*(10), 599-612. *doi:* 10.1038/nrg.2017.52.

Eisenberg, I., Avidan, N., Potikha, T., Hochner, H., Chen, M., & Olender, T., et al. (2001). The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. *Nature Genetics, 29*(1), 83-87. *doi:* 10.1038/ng718.

Elyanow, R., Wu, H. T. & Raphael, B. J. (2017). Identifying structural variants using linked-read sequencing data. *Bioinformatics. doi:* 10.1093/bioinformatics/btx712.

Endo, Y., Dong, M., Noguchi, S., Ogawa, M., Hayashi, Y. K., & Kuru, S., et al. (2015). Milder forms of muscular dystrophy associated with POMGNT2 mutations. *Neurology Genetics, 1*(4). *doi:* 10.1212/nxg.0000000000000033.

Erlich, Y., Mitra, P. P., Delabastide, M., Mccombie, W. R., & Hannon, G. J. (2008). Alta-Cyclic: A self-optimizing base caller for next-generation sequencing. *Nature Methods*, *5*(8), 679-682. *doi:* 10.1038/nmeth.1230.

Ewing, B. & Green, P. (1998). Base-Calling of Automated Sequencer Traces UsingPhred.II. Error Probabilities. *Genome Research*, *8*(3), 186-194. *doi:* 10.1101/gr.8.3.186.

Fanin, M., Nascimbeni, A. C., Tasca, E., & Angelini, C. (2009). How to tackle the diagnosis of limb-girdle muscular dystrophy 2A. *European Journal of Human Genetics, 17*(5), 598-603. *doi:* 10.1038/ejhg.2008.193.

Fanin, M. & Angelini, C. (2015). Protein and genetic diagnosis of limb girdle muscular dystrophy type 2A: The yield and the pitfalls. *Muscle & Nerve, 52*(2), 163-173. *doi:* 10.1002/mus.24682.

Feinstein-Linial, M., Buvoli, M., Buvoli, A., Sadeh, M., Dabby, R., & Straussberg, R., et al. (2016). Two novel MYH7 proline substitutions cause Laing Distal Myopathy-like phenotypes with variable expressivity and neck extensor contracture. *BMC Medical Genetics, 17*(1). *doi:* 10.1186/s12881-016-0315-1

Ferlini, A. & Muntoni, F. (1998). The 5′ Region of intron 11 of the dystrophin gene contains target sequences for mobile elements and three overlapping ORFs. *Biochemical and Biophysical Research Communications*, *242*(2), 401–406. *doi:* 10.1006/bbrc.1997.7976.

Ferreiro, A., Quijano-Roy, S., Pichereau, C., Moghadaszadeh, B., Goemans, N., & Bönnemann, C., et al. (2002). Mutations of the selenoprotein N gene, which is implicated in rigid spine muscular dystrophy, cause the classical phenotype of multiminicore disease: reassessing the nosology of early-onset myopathies. *The American Journal of Human Genetics*, 71(4), 739-749. Ddoi: 10.1086/342719.

Ferreiro, A., Mezmezian, M., Olivé, M., Herlicoviez, D., Fardeau, M., Richard, P., & Romero, N. B. (2011). Telethonin-deficiency initially presenting as a congenital muscular dystrophy. *Neuromuscular Disorders, 21*(6), 433-438. *doi:* 10.1016/j.nmd.2011.03.005.

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. Nature Reviews Genetics, 7(2), 85-97. *doi:* 10.1038/nrg1767.

Fichna, J. P., Maruszak, A., & Żekanowski, C. (2018). Myofibrillar myopathy in the genomic context. *Journal of Applied Genetics, 59*(4), 431-439. *doi:* 10.1007/s13353-018-0463-4.

Findlay, I., Ray, P., Quirke, P., Rutherford, A., & Lilford, R. (1995). Allelic drop-out and preferential amplification in single cells and human blastomeres: Implications for preimplantation diagnosis of sex and cystic fibrosis. *Human Reproduction*, *10*(6), 1609-1618. *doi:* 10.1093/humrep/10.6.1609.

Fischer, D., Herasse, M., Bitoun, M., Barragán-Campos, H. M., Chiras, J., & Laforêt, P., et al. (2006). Characterization of the muscle involvement in dynamin 2-related centronuclear myopathy. *Brain, 129*(6), 1463-1469. *doi:* 10.1093/brain/awl071.

Flanigan, K. M. (2014). Duchenne and Becker muscular dystrophies. *Neurological Clinics*, *32*(3), 671-688. *doi:* 10.1016/j.ncl.2014.05.002.

Flicek, P., & Birney, E. (2009). Sense from sequence reads: Methods for alignment and assembly. *Nature Methods*, *6*(11). *doi:* 10.1038/nmeth.1376.

Fokkema, I. F., den Dunnen, J. T. & Taschner, P. E. (2005). LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. *Human Mutation*, *26*(2):63-68. *doi:*10.1002/humu.20201.

Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F. & den Dunnen, J. T. (2011). LOVD v.2.0: the next generation in gene variant databases. *Human Mutation*, 32(5), 557-563. *doi:* 10.1002/humu.21438.

Foley, A. R., Hu, Y., Zou, Y., Yang, M., Medne, L., & Leach, M., et al. (2011). Large genomic deletions: a novel cause of Ullrich congenital muscular dystrophy. *Annals of Neurology*, *69*(1), 206-211. *doi:* 10.1002/ana.22283.

Freundt, J. K. & Linke, W. A. (2018). Titin as a force generating muscle protein under regulatory control. *Journal of Applied Physiology.* [Epub ahead of print]. *doi:* 10.1152/japplphysiol.00865.2018.

Frontera, W. R. & Ochala, J. (2015). Skeletal muscle: a brief review of structure and function. *Calcified Tissue International*, *96*(3), 183-195. *doi:* 10.1007/s00223-014-9915-y.

Frosk, P., Weiler, T., Nylen, E., Sudha, T., Greenberg, C. R., & Morgan, K., et al. (2002). Limb-Girdle Muscular Dystrophy Type 2H Associated with Mutation in TRIM32, a Putative E3-Ubiquitin–Ligase Gene. *The American Journal of Human Genetics, 70*(3), 663-672. *doi:* 10.1086/339083.

Fu, X. & Xiong, H. (2017). Genetic and Clinical Advances of Congenital Muscular Dystrophy. *Chinese Medical Journal, 130*(21), 2624. *doi:* 10.4103/0366-6999.217091.

Fuller, C. W., Middendorf, L. R., Benner, S. A., Church, G. M., Harris, T., & Huang, X., et al. (2009). The challenges of sequencing by synthesis. *Nature Biotechnology*, *27*(11), 1013-1023. *doi:*10.1038/nbt.1585.

Gambelli, S., Malandrini, A., Berti, G., Gaudiano, C., Zicari, E., & Brunori, P., et al. (2007). Inheritance of a novel RYR1 mutation in a family with myotonic dystrophy type 1. *Clinical genetics*, *71*(1), 93-94. *doi:* 10.1111/j.1399-0004.2006.00725.x.

Gao, Y., Mruk, D., Chen, H., Lui, W. Y., Lee, W. M. & Cheng, C. Y. (2017). Regulation of the blood-testis barrier by a local axis in the testis: role of laminin α2 in the basement membrane. *FASEB Journal*, *31*(2), 584-597. *doi:* 10.1096/fj.201600870R.

Garg, A., O'Rourke, J., Long, C., Doering, J., Ravenscroft, G., & Bezprozvannaya, S., et al. (2014). KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy. *Journal of Clinical Investigation, 124*(8), 3529-3539. *doi:*10.1172/jci74994.

Gatheridge, M. A., Kwon, J. M., Mendell, J. M., Scheuerbrandt, G., Moat, S. J., & Eyskens, F., et al. (2016). Identifying Non–Duchenne Muscular Dystrophy–Positive and False Negative Results in Prior Duchenne Muscular Dystrophy Newborn Screening Programs. *JAMA Neurology, 73*(1), 111. *doi:* 10.1001/jamaneurol.2015.3537.

Gavassini, B. F., Carboni, N., Nielsen, J. E., Danielsen, E. R., Thomsen, C., & Svenstrup, K., et al. (2011). Clinical and molecular characterization of limb-girdle muscular dystrophy due to LAMA2 mutations. *Muscle & Nerve, 44*(5), 703-709. doi: 10.1002/mus.22132.

Gawad, C., Koh, W., & Quake, S. R. (2016). Single-cell genome sequencing: Current state of the science. *Nature Reviews Genetics, 17*(3), 175-188. doi: 10.1038/nrg.2015.16.

Geis, T., Marquard, K., Rödl, T., Reihle, C., Schirmer, S., & Kalle, T. V., et al. (2013). Homozygous dystroglycan mutation associated with a novel muscle–eye–brain disease-like phenotype with multicystic leucodystrophy. *Neurogenetics, 14*(3-4), 205-213. *doi:* 10.1007/s10048-013-0374-9.

Geranmayeh, F., Clement, E., Feng, L. H., Sewry, C., Pagan, J., & Mein, R., et al. (2010). Genotype-phenotype correlation in a large population of muscular dystrophy patients with LAMA2 mutations. *Neuromuscular Disorders, 20*(4), 241-250. *doi:* 10.1016/j.nmd.2010.02.001.

Gibbs, E. M., Davidson, A. E., Telfer, W. R., Feldman, E. L., & Dowling, J. J. (2013). The myopathy-causing mutation DNM2-S619L leads to defective tubulation in vitro and in developing zebrafish. *Disease Models & Mechanisms, 7*(1), 157-161. *doi:* 10.1242/dmm.012286.

Gilissen, C., Hoischen, A., Brunner, H. G., & Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics, 20*(5), 490-497. *doi:* 10.1038/ejhg.2011.258.

Gillespie, R. L., Lloyd, I. C. & Black, G. C. M. (2014). The Use of Autozygosity Mapping and Next-Generation Sequencing in Understanding Anterior Segment Defects Caused by an Abnormal Development of the Lens. *Human Heredity, 77*(1–4), 118-137. *doi:* 10.1159/000362599.

Gillies, A. R., & Lieber, R. L. (2011). Structure and function of the skeletal muscle extracellular matrix. *Muscle & Nerve. doi:* 10.1002/mus.22094.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., Leproust, E. M., & Brockman, W., et al. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology, 27*(2), 182-189. *doi:* 10.1038/nbt.1523.

Godfrey, C., Escolar, D., Brockington, M., Clement, E. M., Mein, R., & Jimenez-Mallebrera, C., et al. (2006). Fukutin gene mutations in steroid-responsive limb girdle muscular dystrophy. *Annals of Neurology, 60*(5), 603-610. *doi:* 10.1002/ana.21006.

Godfrey, C., Foley, A. R., Clement, E., & Muntoni, F. (2011). Dystroglycanopathies: Coming into focus. *Current Opinion in Genetics & Development, 21*(3), 278-285. *doi:* 10.1016/j.gde.2011.02.001.

Goldfarb, L. G., Park, K., Cervenáková, L., Gorokhova, S., Lee, H., & Vasconcelos, O., et al. (1998). Missense mutations in desmin associated with familial cardiac and skeletal myopathy. *Nature Genetics, 19*(4), 402-403. *doi:* 10.1038/1300.

Gonorazky, H. D., Bönnemann, C. G., & Dowling, J. J. (2018). The genetics of congenital myopathies. *Handbook of Clinical Neurology, 148*, 549-564. *doi:* 10.1016/B978-0-444-64076-5.00036-3.

González-Jamett, A. M., Baez-Matus, X., Olivares, M. J., Hinostroza, F., Guerra-Fernández, M. J., & Vasquez-Navarrete, J., et al. (2017). Dynamin-2 mutations linked to Centronuclear Myopathy impair actin-dependent trafficking in muscle cells. *Scientific Reports, 7*(1). doi:10.1038/s41598-017-04418-w.

González-Pérez, A., & López-Bigas, N. (2011). Improving the Assessment of the Outcome of Nonsynonymous SNVs with a Consensus Deleteriousness Score, Condel. *The American Journal of Human Genetics, 88*(4), 440-449. *doi:* 10.1016/j.ajhg.2011.03.004.

Goodship, J., Gill, H., Carter, J., Jackson, A., Splitt, M., & Wright, M. (2000). Report Autozygosity Mapping of a Seckel Syndrome Locus to Chromosome 3q22.1-q24. *American Journal of Human Genetics, 67*(2), 498-503. *doi:* 10.1086/303023.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., & McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Research, 25*(11), 1750-1756. *doi:* 10.1101/gr.191395.115.

Goonasekera, S. A, Beard, N. A, Groom, L., Kimura, T., Lyfenko, A. D., & Rosenfeld, A., et al. (2007). Triadin binding to the C-terminal luminal loop of the ryanodine receptor is important for skeletal muscle excitation contraction coupling. *The Journal of general physiology, 130*(4), 365-378. *doi:* 10.1085/jgp.200709790.

Gormez, Z., Bakir-Gungor, B. & Sagiroglu, M. S. (2014). HomSI: a homozygous stretch identifier from next-generation sequencing data. *Bioinformatics, 30*(3), 445-447. doi: 10.1093/bioinformatics/btt686.

Gregorevic, P., Allen, J. M., Minami, E., Blankinship, M. J., Haraguchi, M., & Meuse, L., et al. (2006). RAAV6-microdystrophin preserves muscle function and extends

lifespan in severely dystrophic mice. *Nature Medicine, 12*(7), 787-789. *doi:* 10.1038/nm1439.

Griggs, R., Vihola, A., Hackman, P., Talvinen, K., Haravuori, H., & Faulkner, G., et al. (2007). Zaspopathy in a large classic late-onset distal myopathy family. *Brain, 130*(6), 1477-1484. *doi:* 10.1093/brain/awm006.

Grömminger, S., Yagmur, E., Erkan, S., Nagy, S., Schöck, U., & Bonnet, J., et al. (2014). Fetal Aneuploidy Detection by Cell-Free DNA Sequencing for Multiple Pregnancies and Quality Issues with Vanishing Twins. *Journal of Clinical Medicine*, *3*(3), 679-692. *doi:* 10.3390/jcm3030679.

Grotto, S., Cuisset, J. M., Marret, S., Drunat, S., Faure, P., & Audebert-Bellanger, S. et al. (2016). Type 0 Spinal Muscular Atrophy: Further Delineation of Prenatal and Postnatal Features in 16 Patients. Journal of Neuromuscular Diseases, 3(4), 487-495. *doi:* 10.3233/JND-160177.

Gu, W., Zhang, F., & Lupski, J. R. (2008). Mechanisms for human genomic rearrangements. *PathoGenetics, 1*(1), 4. *doi:* 10.1186/1755-8417-1-4.

Gudbjartsson, D. F., Jonasson, K., Frigge, M. L. & Kong, A. (2000). Allegro, a new computer program for multipoint linkage analysis. *Nature Genetics*, *25*(1), 12–13. *doi:* 10.1038/75514.

Gueneau, L., Bertrand, A. T., Jais, J., Salih, M. A., Stojkovic, T., & Wehnert, M., et al. (2009). Mutations of the FHL1 Gene Cause Emery-Dreifuss Muscular Dystrophy. *The American Journal of Human Genetics, 85*(3), 338-353. *doi:* 10.1016/j.ajhg.2009.07.015.

Gundesli, H., Talim, B., Korkusuz, P., Balci-Hayta, B., Cirak, S., & Akarsu, N. A., et al. (2010). Mutation in Exon 1f of PLEC, Leading to Disruption of Plectin Isoform 1f, Causes Autosomal-Recessive Limb-Girdle Muscular Dystrophy. *The American Journal of Human Genetics, 87*(6), 834-841. *doi:* 10.1016/j.ajhg.2010.10.017.

Gupta, V. A., Ravenscroft, G., Shaheen, R., Todd, E. J., Swanson, L. C., & Shiina, M., et al. (2013). Identification of KLHL41 Mutations Implicates BTB-Kelch-Mediated Ubiquitination as an Alternate Pathway to Myofibrillar Disruption in Nemaline Myopathy*. American Journal of Human Genetics*, *93*(6), 1108-1117. *doi:* 10.1016/j.ajhg.2013.10.020.

Gupta, V. A., & Beggs, A. H. (2014). Kelch proteins: Emerging roles in skeletal muscle development and diseases. *Skeletal Muscle, 4*(1), 11. *doi:*10.1186/2044-5040-4-11.

Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., & Breslow, J. L., et al. (2008). Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, *19*(2), 318-326. *doi:* 10.1101/gr.081398.108.

Gutiérrez Ríos, P., Kalra, A. A., Wilson, J. D., Tanji, K., Akman, H. O., & Area Gómez, E. et al. (2012). Congenital megaconial myopathy due to a novel defect in the choline kinase Beta gene. *Archives of Neurology, 69*(5), 657-661. *doi:* 10.1001/archneurol.2011.2333.

Guyon, J. R., Kudryashova, E., Potts, A., Dalkilic, I., Brosius, M. A., & Thompson, T. G., et al. (2003). Calpain 3 cleaves filamin C and regulates its ability to interact with gamma- and delta-sarcoglycans. *Muscle & Nerve, 28*(4), 472-483. *doi:* 10.1002/mus.10465.

Hackman, P., Vihola, A., Haravuori, H., Marchand, S., Sarparanta, J., & Seze, J. D., et al. (2002). Tibial Muscular Dystrophy Is a Titinopathy Caused by Mutations in TTN, the Gene Encoding the Giant Skeletal-Muscle Protein Titin. *The American Journal of Human Genetics, 71*(3), 492-500. *doi:* 10.1086/342380.

Hackman, P., Sarparanta, J., Lehtinen, S., Vihola, A., Evilä, A., & Jonson, P. H., et al. (2013). Welander distal myopathy is caused by a mutation in the RNA-binding protein TIA1. *Annals of Neurology, 73*(4), 500-509. *doi:* 10.1002/ana.23831.

Häger, M., Gawlik, K., Nyström, A., Sasaki, T., & Durbeej, M. (2005). Laminin {alpha}1 chain corrects male infertility caused by absence of laminin {alpha}2 chain. *The American Journal of Pathology, 167*(3), 823-833. *doi*: 10.1016/S0002-9440(10)62054-8.

Hammans, S. R., Robinson, D. O., Moutou, C., Kennedy, C. R., Dennis, N. R., & Hughes, P. J., et al. (2000). A clinical and genetic study of a manifesting heterozygote with X-linked myotubular myopathy. *Neuromuscular Disorders, 10*(2), 133-137. *doi:* 10.1016/S0960-8966(99)00073-5.

Hancks, D. C. & Kazazian, H. H., Jr. (2016). Roles for retrotransposon insertions in human disease. *Mobile DNA, 7*, 9. *doi:* 10.1186/s13100-016-0065-9.

Haque, K. A., Pfeiffer, R. M., Beerman, M. B., Struewing, J. P., Chanock, S. J., & Bergen, A. W. (2003). *BMC Biotechnology, 3*(1), 20. *doi:* 10.1186/1472-6750-3-20.

Hara, Y., Balci-Hayta, B., Yoshida-Moriguchi, T., Kanagawa, M., Bernabé, D. B., & Gündeşli, H., et al. (2011). A Dystroglycan Mutation Associated with Limb-Girdle Muscular Dystrophy. *New England Journal of Medicine, 364*(10), 939-946. *doi:*10.1056/nejmoa1006939.

Harris, E., McEntagart, M., Topf, A., Lochmüller, H., Bushby, K., & Sewry, C., et al. (2017). Clinical and neuroimaging findings in two brothers with limb girdle muscular dystrophy due to LAMA2 mutations. *Neuromuscular Disorders, 27*(2), 170-174. *doi:* 10.1016/j.nmd.2016.10.009.

Hauser, M. A., Horrigan, S. K., Salmikangas, P., Torian, U. M., Viles, K. D., & Dancel, R., et al. (2000). Myotilin is mutated in limb girdle muscular dystrophy 1A. *Human Molecular Genetics, 9*(14), 2141-2147. *doi:* 10.1093/hmg/9.14.2141.

Hayashi, Y. K., Ishihara, T., Domen, K., Hori, H., & Arahata, K. (1997). A benign allelic form of laminin a2 chain deficient muscular dystrophy. *Lancet, 349*(9059), 1147. *doi:* 10.1016/S0140-6736(05)63023-1.

Hayashi, Y. K., Chou, F., Engvall, E., Ogawa, M., Matsuda, C., & Hirabayashi, S., et al. (1998). Mutations in the integrin α7 gene cause congenital myopathy. *Nature Genetics, 19*(1), 94-97. *doi:* 10.1038/ng0598-94.

Hayden, M. S., & Ghosh, S. (2011). NF-κB in immunobiology. *Cell Research, 21*(2), 223-244. *doi:* 10.1038/cr.2011.13.

Head, S. R., Komori, H. K., Lamere, S. A., Whisenant, T., Nieuwerburgh, F. V., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: Overviews and challenges. *BioTechniques, 56*(2). *doi:* 10.2144/000114133.

Helbling-Leclerc, A., Zhang, X., Topaloglu, H., Cruaud, C., Tesson, F., & Weissenbach, J., et al. (1995). Mutations in the laminin alpha 2-chain gene (LAMA2) cause merosin-deficient congenital muscular dystrophy. *Nature Genetics, 11*(2), 216-218. *doi:* 10.1038/ng1095-216.

Hedberg, C., Melberg, A., Kuhl, A., Jenne, D., & Oldfors, A. (2012). Autosomal dominant myofibrillar myopathy with arrhythmogenic right ventricular cardiomyopathy 7 is caused by a DES mutation. *European Journal of Human Genetics, 20*(9), 984-985. *doi:* 10.1038/ejhg.2012.39.

Hedberg, C., Melberg, A., Dahlbom, K., & Oldfors, A. (2014). Hereditary myopathy with early respiratory failure is caused by mutations in the titin FN3 119 domain. *Brain, 137*(4). *doi:* 10.1093/brain/awt305.

Henderson, C. A., Gomez, C. G., Novak, S. M., Mi-mi, L., & Gregorio, C. C. (2017). Overview of the Muscle Cytoskeleton. *Comprehensive Physiology*, 7(3): 891-944. *doi:* 10.1002/cphy.c160033.

Herman, G. E., Finegold, M., Zhao, W., de Gouyon, B., & Metzenberg, A. (1999). Medical complications in long-term survivors with X-linked myotubular myopathy. *The Journal of Pediatrics*, 134(2), 206-214. *doi:* 10.1016/S0022-3476(99)70417-8.

Hewitt, J. E., Lyle, R., Clark, L. N., Valleley, E. M., Wright, T. J., & Wijmenga, C., et al. (1994). Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystropothhy. *Human Molecular Genetics, 3*(8), 1287-1295. *doi:* 10.1093/hmg/3.8.1287.

Hill, M. A. (2018). Embryology Musculoskeletal System - Muscle Development. Retrieved September 18, 2018 from https://embryology.med.unsw.edu.au/ embryology/index.php/Musculoskeletal_System_- _Muscle_Development

Hicks, D., Farsani, G. T., Laval, S., Collins, J., Sarkozy, A., & Martoni, E., et al. (2013). Mutations in the collagen XII gene define a new form of extracellular matrix-related myopathy. *Human Molecular Genetics, 23*(9), 2353-2363. *doi:* 10.1093/hmg/ddt637.

Hnia, K., Tronchère, H., Tomczak, K. K., Amoasii, L., Schultz, P., & Beggs, A. H, et al. (2011). Myotubularin controls desmin intermediate filament architecture and mitochondrial dynamics in human and mouse skeletal muscle. *The Journal of Clinical Investigation, 121*(1), 70-85. *doi:* 10.1172/JCI44021.

Ho, R., & Hegele, R. A. (2018). Complex effects of laminopathy mutations on nuclear structure and function. *Clinical Genetics. doi:* 10.1111/cge.13455.

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., & Smith, S. W., et al. (2007). Genome-wide in situ exon capture for selective resequencing. Nature Genetics, 39(12), 1522-1527. doi: 10.1038/ng.2007.42.

Hoffjan, S., Thiels, C., Vorgerd, M., Neuen-Jacob, E., Epplen, J. T. & Kress, W. (2006). Extreme phenotypic variability in a German family with X-linked myotubular myopathy associated with E404K mutation in MTM1. *Neuromuscular Disorders, 16*(11), 749-753. *doi:* 10.1016/j.nmd.2006.07.020.

Holmberg, J., & Durbeej, M. (2013). Laminin-211 in skeletal muscle function. *Cell Adhesion & Migration, 7*(1), 111-121. *doi:* 10.4161/cam.22618.

Holmes, S. E., Dombroski, B. A., Krebs, C. M., Boehm, C. D. & Kazazian, H. H., Jr. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nature Genetics, 7*(2): 143-148. *doi:* 10.1038/ng0694-143.

Homer, N. (2012). TMAP: The torrent mapping program. Retrieved October 1, 2018, from https://github.com/iontorrent/TMAP/blob/master/doc/tmap-book.pdf.

Horstick, E., Linsley, J., Dowling, J., Hauser, M., McDonald, K., & Ashley-Koch, A. et al. (2013). Stac3 is a component of the excitation–contraction coupling machinery and mutated in Native American myopathy. Nature Communications, 4(1), 1952. *doi:* 10.1038/ncomms2952.

Hussain, S. A., Carafoli, F., & Hohenester, E. (2011). Determinants of laminin polymerization revealed by the structure of the α5 chain amino-terminal region. *EMBO Reports, 12*(3), 276-282. *doi:* 10.1038/embor.2011.

Hunter, J., Ahearn, M., Balak, C., Liang, W., Kurdoglu, A., & Corneveaux, J. et al. (2015). Novel pathogenic variants and genes for myopathies identified by whole

exome sequencing. *Molecular Genetics & Genomic Medicine*, *3*(4), 283-301. *doi:* 10.1002/mgg3.142.

Hwang, P. M., & Sykes, B. D. (2015). Targeting the sarcomere to correct muscle function. *Nature Reviews Drug Discovery, 14*(5), 313-328. *doi:* 10.1038/nrd4554.

Hwang, S., Kim, E., Lee, I., & Marcotte, E. M. (2015). Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific Reports*, *5*(1). *doi:* 10.1038/srep17875.

Illa, I., Serrano-Munuera, C., Gallardo, E., Lasa, A., Rojas-García, R., & Palmer, J., et al. (2001). Distal anterior compartment myopathy: A dysferlin mutation causing a new muscular dystrophy phenotype. *Annals of Neurology, 49*(1), 130-134. *doi:*10.1002/1531-8249(200101)49:13.3.co;2-s.

Illumina. (2018a). History of Illumina Sequencing. Retrieved October 1, 2018, from https://www.illumina.com/science/technology/next-generation-sequencing/illumina-sequencing-history.html.

Illumina. (2018b). An introduction to next-generation sequencing technology. Retrieved October 1, 2018, from https://www.illumina.com/documents/products/ illumina_sequencing_introduction.pdf.

Illumina. (2018c). Illumina Sequencing Technology. Technology Spotlight: Illumina Sequencing. Retrieved October 1, 2018, from https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencin g.pdf.

International HapMap 3 Consortium, Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., & Gibbs, R. A., et al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52-58. *doi:* 10.1038/nature09298.

Ion Torrent. (2018a). Ion Torrent - Amplicon sequence. Application Note. Retrieved October 1, 2018, from https://docplayer.net/23705099-Ion-torrent-amplicon-sequencing.html.

Ion Torrent. (2018b). Torrent Suite Software Analysis Pipeline. Technical note. Retrieved October 1, 2018, from http://coolgenes.cahe.wsu.edu/ion-docs/Technical-Note---Analysis-Pipeline_6455567.html.

Ion Torrent. (2018c). Ion torrent CNV detection by Ion semiconductor sequencing. Retrieved October 1, 2018, from https://assets.thermofisher.com/TFS-Assets/LSG/brochures/CNV-Detection-by-Ion.pdf.

Ishmukhametova, A., Chen, J.M., Bernard, R., de Massy, B., Baudat, F., Boyer, A., Méchin, D., Thorel, D., Chabrol, B., & Vincent, M.C., et al. (2013). Dissecting the structure and mechanism of a complex duplication-triplication rearrangement in the DMD gene. *Human Mutation*, *34*(8), 1080-1084. *doi:* 10.1002/humu.22353.

Iyadurai, S. J., & Kissel, J. T. (2016). The Limb-Girdle Muscular Dystrophies and the Dystrophinopathies. *CONTINUUM: Lifelong Learning in Neurology, 22*(6), 1954-1977. *doi:* 10.1212/con.0000000000000406.

Jackson, D. A., Symonst, R. H., & Berg, P. (1972) Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of Escherichia coli. *Proceedings of the National Academy of Scinces of the United States of America*, 69, 2904–2909.

Jae, L. T., Raaben, M., Riemersma, M., Beusekom, E. V., Blomen, V. A., & Velds, A., et al. (2013). Deciphering the Glycosylome of Dystroglycanopathies Using Haploid Screens for Lassa Virus Entry. *Science, 340*(6131), 479-483. *doi:* 10.1126/science.1233675.

Jain, M., Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., & Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *Nature Methods*, *12*(4), 351-356. *doi:* 10.1038/nmeth.3290.

Jobsis, G. J., Bolhuis, P. A., Boers, J. M., Baas, F., Wolterman, R. A., Hensels, G. W., & Visser, M. D. (1996). Genetic localization of Bethlem myopathy. *Neurology, 46*(3), 779-782. *doi:* 10.1212/wnl.46.3.779.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. & Madden, T. L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Research, 36, W5-9. *doi:* 10.1093/nar/gkn201.

Johnston, J., Kelley, R., Crawford, T., Morton, D., Agarwala, R., & Koch, T., et al. (2000). A Novel Nemaline Myopathy in the Amish Caused by a Mutation in Troponin T1. *The American Journal of Human Genetics*, *67*(4), 814-821. *doi:* 10.1086/303089.

Joneja, A., & Huang, X. (2009). A device for automated hydrodynamic shearing of genomic DNA. *BioTechniques*, *46*(7), 553-556. *doi:* 10.2144/000113123.

Jones, J. C., Dehart, G. W., Gonzales, M., & Goldfinger, L. E. (2000). Laminins: an overview. *Microscopy Research and Technique*, *51*(3), 211-213. *doi:* 10.1002/1097-0029(20001101)51:3<211::AID-JEMT1>3.0.CO;2-P.

Jones, K. J., Morgan, G., Johnston, H., Tobias, V., Ouvrier, R. A., & Wilkinson, I., et al. (2001). The expanding phenotype of laminin alpha2 chain (merosin) abnormalities: case series and review. Journal of Medics Genetics, 38(10), 649-657. *doi:* 10.1136/jmg.38.10.649.

Jones, D., Round, J., & de Haan, A. (2004). Chapter 1 - Structure of the muscle fibre, Editors: Jones, D., Round, J., de Haan, A. *Skeletal Muscle from Molecules to Movement, Churchill Livingstone*, 1-8. *doi:* 10.1016/B978-0-443-07427-1.50005-4.

Jorde, L. B. (2000). Linkage disequilibrium and the search for complex disease genes. *Genome research*, *10*(10), 1435-1444. *doi:* 10.1101/gr.144500.

Jung, D. J., Sung, H. S., Goo, Y. W., Lee, H. M., Park, O. K., Jung, S. Y., Lim, J., & Kim, H. J., et al. (2002). Novel transcription coactivator complex containing activating signal cointegrator 1. *Molecular and Cellular Biology*, *22*(14), 5203-5211. *doi:* 10.1128/MCB.22.14.5203-5211.2002.

Jungbluth, H., Sewry, C. A., Buj-Bello, A., Kristiansen, M., Ørstavik, K. H., & Kelsey, A., et al. (2003). Early and severe presentation of X-linked myotubular myopathy in a girl with skewed X-inactivation. *Neuromuscular Disorders*, *13*(1), 55-59. *doi:* 10.1016/S0960-8966(02)00194-3.

Jungbluth, H. (2007a). Central core disease. *Orphanet Journal of Rare Diseases*, *2*(1), 25. *doi:* 10.1186/1750-1172-2-25.

Jungbluth, H. (2007b). Multi-minicore Disease. *Orphanet Journal of Rare Diseases*, *2*(1), 31. *doi:* 10.1186/1750-1172-2-31.

Jungbluth, H., Wallgren-Pettersson, C., & Laporte, J. (2008). Centronuclear (myotubular) myopathy. *Orphanet Journal of Rare Diseases*, *3*, 26. *doi:* 10.1186/1750-1172-3-26.

Jungbluth, H., Wallgren-Pettersson, C., Laporte, J. F., & Centronuclear (Myotubular) Myopathy Consortium. (2009). 164th ENMC International workshop: 6th workshop on centronuclear (myotubular) myopathies, 16-18th January 2009, Naarden, The Netherlands. *Neuromuscular Disorders*, *19*(10), 721-729. *doi:* 10.1016/j.nmd.2009.06.373.

Jungbluth, H., Sewry, C. A., & Muntoni, F. (2011). Core Myopathies. *Seminars in Pediatric Neurology, 18*(4), 239-249. *doi:* 10.1016/j.spen.2011.10.005.

Jungbluth, H., & Wallgren-Pettersson, C. (2014). Congenital (Structural) Myopathies. in Emery and Rimoin's Principles and Practice of Medical Genetics 6th edn (eds Rimoin, D. L., Pyeritz, R. E., Korf, B.) chapter 127 pp 1–51 (Academic Press, 2014).

Jungbluth, H., & Gautel, M. (2014). Pathogenic mechanisms in centronuclear myopathies. *Frontiers in aging neuroscience*, *6*, 339. *doi:* 10.3389/fnagi.2014.00339.

Jungbluth, H., & Voermans, N., C. (2016). Congenital myopathies: not only a paediatric topic. *Current Opinion in Neurology*, *29*(5):642-650. *doi:* 10.1097/WCO.0000000000000372.

Jungbluth, H., Treves, S., Zorzato, F., Sarkozy, A., Ochala, J., & Sewry, C., et al. (2018). Congenital myopathies: Disorders of excitation–contraction coupling and muscle contraction. *Nature Reviews Neurology, 14*(3), 151-167. *doi:* 10.1038/nrneurol.2017.191.

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O. & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, *110*(1-4), 462-467. *doi:* 10.1159/000084979.

Kanagal-Shamanna, R. (2016). Emulsion PCR: Techniques and Applications. *Clinical Applications of PCR Methods in Molecular Biology*, 33-42. *doi:* 10.1007/978-1-4939-3360-0_4.

Kanagawa, M., & Toda, T. (2017). Muscular Dystrophy with Ribitol-Phosphate Deficiency: A Novel Post-Translational Mechanism in Dystroglycanopathy. *Journal of Neuromuscular Diseases, 4*(4), 259-267. *doi:* 10.3233/jnd-170255.

Kao, W., Stevens, K., & Song, Y. S. (2009). BayesCall: A model-based base-calling algorithm for high-throughput short-read sequencing. *Genome Research*, *19*(10), 1884-1895. *doi:* 10.1101/gr.095299.109.

Kapp, J. R., Diss, T., Spicer, J., Gandy, M., Schrijver, I., & Jennings, L. J., et al. (2015). Variation in pre-PCR processing of FFPE samples leads to discrepancies in BRAF and EGFR mutation detection: A diagnostic RING trial. *Journal of Clinical Pathology*, *68*(2), 111-118. *doi:* 10.1136/jclinpath-2014-202644.

Karaoglu, P., Quizon, N., Pergande, M., Wang, H., Polat, A. I., & Ersen, A., et al. (2017). Dropped head congenital muscular dystrophy caused by de novo mutations in LMNA. *Brain and Development, 39*(4), 361-364. *doi:* 10.1016/j.braindev.2016.11.002.

Kayman-Kurekci, G., Talim, B., Korkusuz, P., Sayar, N., Sarioglu, T., & Oncel, I., et al. (2014). Mutation in TOR1AIP1 encoding LAP1B in a form of muscular dystrophy: A novel gene related to nuclear envelopathies. *Neuromuscular Disorders, 24*(7), 624-633. *doi:* 10.1016/j.nmd.2014.04.007.

Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G. & Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature*, *332*(6160), 164-166. *doi:* 10.1038/332164a0.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H. & Zahler, A. M. (2002). The human genome browser at UCSC. *Genome Research*, *12*(6), 996-1006. *doi:* 10.1101/gr.229102.

Keren, H., Lev-Maor, G., & Ast, G. (2010). Alternative splicing and evolution: Diversification, exon definition and function. *Nature Reviews Genetics*, *11*(5), 345-355. *doi:* 10.1038/nrg2776.

Kevelam, S. H., van Engelen, B. G., van Berkel, C. G., Küsters, B. & van der Knaap, M. S. (2014). LAMA2 mutations in adult-onset muscular dystrophy with leukoencephalopathy. *Muscle & Nerve*, *49*(4), 616-617. *doi:* 10.1002/mus.24147.

Khurana, E., Fu, Y., Chen, J., & Gerstein, M. (2013). Interpretation of Genomic Variants Using a Unified Biological Network Approach. *PLoS Computational Biology*, *9*(3). *doi:* 10.1371/journal.pcbi.1002886.

Kiiski, K., Laari, L., Lehtokari, V. L., Lunkka-Hytönen, M., Angelini, C., & Petty, R., et al. (2013). Targeted array comparative genomic hybridization--a new diagnostic tool for the detection of large copy number variations in nemaline myopathy-causing genes. *Neuromuscular Disorders*, *23*(1), 56-65. *doi:* 10.1016/ j.nmd.2012.07.007.

Kiiski, K., Lehtokari, V. L., Manzur, A. Y., Sewry, C., Zaharieva, I., & Muntoni, F., et al. (2015). A Large Deletion Affecting TPM3, Causing Severe Nemaline Myopathy. *Journal of Neuromuscular Disorders*, 2(4), 433-438. *doi*: 10.3233/JND-150107.

Kim, H. J., Yi, J. Y., Sung, H. S., Moore, D. D., Jhun, B. H., Lee, Y. C., & Lee, J. W. (1999). Activating signal cointegrator 1, a novel transcription coactivator of nuclear receptors, and its cytosolic localization under conditions of serum deprivation. Molecular and Cellular Biology, 19(9), 6323-6332. *doi:* 10.1128/MCB.19.9.6323.

Kim, M. W., Jang, D. H., Kang, J., Lee, S., Joo, S. Y., & Jang J. H., et al. (2017). Novel Mutation (c.8725T>C) in Two Siblings With Late-Onset LAMA2-Related Muscular Dystrophy. *Annals of Laboratory Medicine*, 37(4), 359-361. *doi:* 10.3343/alm.2017.37.4.359.

Kim, Y. J., Lee, J. & Han, K. (2012). Transposable Elements: No More 'Junk DNA'. *Genomics & Informatics*, *10*(4), 226-233. doi: 10.5808/GI.2012.10.4.226.

Kimberland, M.L., Divoky, V., Prchal, J., Schwahn, U., Berger, W. & Kazazian, H.H., Jr. (1999). Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Human Molecular Genetics, *8*(8), 1557-1560. *doi:* 10.1093/hmg/8.8.1557.

Kircher, M., Heyn, P., & Kelso, J. (2011). Addressing challenges in the production and analysis of illumina sequencing data. *BMC Genomics*, *12*(1). *doi:*10.1186/1471-2164-12-382. *doi:* 10.1186/1471-2164-12-382.

Kircher, M., Witten, D. M., Jain, P., Oroak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310-315. *doi:* 10.1038/ng.2892.

Kirschner, J., & Lochmüller, H. (2011). Sarcoglycanopathies. *Handbook of Clinical Neurology Muscular Dystrophies,* 41-46. *doi:* 10.1016/b978-0-08-045031-5.00003-7.

Kissiedu, J., & Prayson, R. A. (2016). Congenital fiber type disproportion. *Journal of Clinical Neuroscience, 26*, 136-137. *doi:* 10.1016/j.jocn.2015.08.029**.**

Klein, A., Lillis, S., Munteanu, I., Scoto, M., Zhou, H., & Quinlivan, R., et al. (2012). Clinical and genetic findings in a large cohort of patients with ryanodine receptor 1

gene-associated myopathies. *Human mutation*, *33*(6), 981-988. *doi*: 10.1002/humu.22056.

Kley, R. A., Olivé, M., & Schröder, R. (2016). New aspects of myofibrillar myopathies. *Current Opinion in Neurology, 29*(5), 628-634. *doi:* 10.1097/wco.0000000000000357.

Knierim, E., Lucke, B., Schwarz, J. M., Schuelke, M., & Seelow, D. (2011). Systematic Comparison of Three Methods for Fragmentation of Long-Range PCR Products for Next Generation Sequencing. *PLoS ONE*, *6*(11). *doi:* 10.1371/journal.pone.0028240.

Knierim, E., Hirata, H., Wolf, N. I., Morales-Gonzalez, S., Schottmann, G., & Tanaka, Y., et al. (2016). Mutations in Subunits of the Activating Signal Cointegrator 1 Complex Are Associated with Prenatal Spinal Muscular Atrophy and Congenital Bone Fractures. *American Journal of Human Genetics*, *98*(3), 473-489. *doi:* 10.1016/j.ajhg.2016.01.006.

Kobayashi, K., Nakahori, Y., Miyake, M., Matsumura, K., Kondo-Iida, E., & Nomura, Y., et al. (1998). An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. *Nature, 394*(6691), 388-392. *doi:*10.1038/28653.

Koenig, M., Monaco, A., & Kunkel, L. (1988). The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein. *Cell, 53*(2), 219-228. *doi:* 10.1016/0092-8674(88)90383-2.

Kohany, O., Gentles, A. J., Hankus, L., & Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics*, *7*, 474. *doi:* 10.1186/1471-2105-7-474.

Kolata, G. B. (1980). The 1980 Nobel Prize in Chemistry. *Science*, *210*(4472), 887-889. *doi:* 10.1126/science.7001629.

Kondo, E., Nishimura, T., Kosho, T., Inaba, Y., Mitsuhashi, S., & Ishida, T., et al. (2012). Recessive RYR1 mutations in a patient with severe congenital nemaline myopathy with ophthalomoplegia identified through massively parallel sequencing. *American Journal of Medical Genetics Part A*, *158A*(4), 772-778. *doi:* 10.1002/ajmg.a.35243.

Kong, Y. (2011). Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics*, *98*(2), 152-153. *doi:* 10.1016/j.ygeno.2011.05.009.

Kobayashi, K., Nakahori, Y., Mizuno, K., Miyake, M., Kumagai, T., & Honma, A., et al. (1998). Founder-haplotype analysis in Fukuyama-type congenital muscular dystrophy (FCMD). *Human Genetics*, *103*(3), 323-327. *doi:* 10.1007/s004390050824.

Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., & Simons, J. F., et al. (2007). Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. *Science*, *318*(5849), 420-426. *doi:* 10.1126/science.1149504.

Kruglyak, L., Daly, M. J., Reeve-Daly, M. P., & Lander, E. S. (1996). Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach. Am. J. Hum. Genet, 58, pp.1347–1363.

Kraeva, N., Heytens, L., Jungbluth, H., Treves, S., Voermans, N., & Kamsteeg, E., et al. (2015). Compound RYR1 heterozygosity resulting in a complex phenotype of malignant hyperthermia susceptibility and a core myopathy. *Neuromuscular disorders*, *25*(7), 567-576. *doi:* 10.1016/j.nmd.2015.04.007.

Ku, C. S., Cooper, D. N., Polychronakos, C., Naidoo, N., Wu, M., & Soong R. (2012). Exome sequencing: dual role as a discovery and diagnostic tool. *Annals of Neurology*, *71*(1), 5-14. *doi:* 10.1002/ana.22647.

Kuang, W., Xu, H., Vachon, P. H., Liu, L., Loechel, F., & Wewer, U. M., et al. (1998). Merosin-deficient congenital muscular dystrophy. Partial genetic correction in two mouse models. Journal of Clinical Investigation, *102*(4):844-852. *doi*: 10.1172/JCI3705.

Kudryashova, E., Kudryashov, D., Kramerova, I., & Spencer, M. J. (2005). Trim32 is a Ubiquitin Ligase Mutated in Limb Girdle Muscular Dystrophy Type 2H that Binds to Skeletal Muscle Myosin and Ubiquitinates Actin. *Journal of Molecular Biology, 354*(2), 413-424. *doi:* 10.1016/j.jmb.2005.09.068.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073-1081. *doi:* 10.1038/nprot.2009.86.

Kuruvilla, J., Sasmita, A. O., & Ling, A. P. (2018). Therapeutic potential of combined viral transduction and CRISPR/Cas9 gene editing in treating neurodegenerative diseases. *Neurological Sciences, 39*(11), 1827-1835. *doi:* 10.1007/s10072-018-3521-0.

Laing, N. G., Wilton, S. D., Akkari, P. A., Dorosz, S., Boundy, K., & Kneebone, C., et al. (1995). A mutation in the alpha tropomyosin gene TPM3 associated with autosomal dominant nemaline myopathy NEM1. *Nature Genetics*, *10*(2), 249. *doi:* 10.1038/ng0695-249.

Laing, N.G., Dye, D.E., Wallgren-Pettersson, C., Richard, G., Monnier, N., & Lillis, S., et al. (2009). Mutations and polymorphisms of the skeletal muscle alpha-actin gene (ACTA1). *Human Mutation*, *30*(9), 1267-1277. *doi:* 10.1002/humu.21059.

Laing, N. G. (2012). Genetics of neuromuscular disorders. *Critical Reviews in Clinical Laboratory Sciences*, *49*(2), 33-48. *doi:* 10.3109/10408363.2012.658906.

Lamandé, S. R., & Bateman, J. F. (2018). Collagen VI disorders: Insights on form and function in the extracellular matrix and beyond. *Matrix Biology, 71-72*, 348-367. *doi:* 10.1016/j.matbio.2017.12.008.

Lamer, S., Carlier, R. Y., Pinard, J. M., Mompoint, D., Bagard, C., & Burdairon, E., et al. (1998). Congenital muscular dystrophy: use of brain MR imaging findings to predict merosin deficiency. *Radiology*, 206(3), 811-816. *doi:* 10.1148/radiology.206.3.9494506.

Lange, S., Xiang, F., Yakovenko, A., Vihola, A., Hackman, P., & Rostkova, E., et al. (2005). The Kinase Domain of Titin Controls Muscle Gene Expression and Protein Turnover. *Science, 308*(5728), 1599-1603. *doi:* 10.1126/science.1110463.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). *doi:* 10.1186/gb-2009-10-3-r25.

Laporte, J., Hu, L. J., Kretz, C., Mandel, J. L., Kioschis, P., & Coy, J. F., et al. (1996). A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast. *Nature Genetics, 13*(2), 175-182. *doi*: 10.1038/ng0696-175.

Laporte, J., Guiraud-Chaumeil, C., Vincent, M. C., Mandel, J. L., Tanner, S. M., & Liechti-Gallati, S., et al. (1997). Mutations in the MTM1 gene implicated in X-linked myotubular myopathy. ENMC International Consortium on Myotubular Myopathy. European Neuro-Muscular Center. *Human Molecular Genetics*, *6*(9), 1505-1511. *doi:* 10.1093/hmg/6.9.1505.

Laporte J., Biancalana, V., Tanner, S. M., Kress, W., Schneider, V., & Wallgren-Pettersson, C., et al. (2000). MTM1 mutations in X-linked myotubular myopathy. *Human Mutation*, *15*(5), 393-409. *doi*: 10.1002/(SICI)1098-1004(200005)15:5<393::AID-HUMU1>3.0.CO;2-R.

Laporte, J., Blondeau, F., Buj-Bello, A., & Mandel, J. L. (2001). The myotubularin family: from genetic disease to phosphoinositide metabolism. *Trends in Genetics*, *17*(4), 221-228. *doi:* 10.1016/S0168-9525(01)02245-4.

Larsen, M., Rost, S., Hajj, N. E., Ferbert, A., Deschauer, M., & Walter, M. C., et al. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. *European Journal of Human Genetics, 23*(6), 808-816. *doi:* 10.1038/ejhg.2014.191.

Larson, A. A., Baker, P. R., Milev, M. P., Press, C. A., Sokol, R. J., & Cox, M. O., et al. (2018). TRAPPC11 and GOSR2 mutations associate with hypoglycosylation of α-dystroglycan and muscular dystrophy. *Skeletal Muscle, 8*(1). *doi:* 10.1186/s13395-018-0163-0.

Laval, S. H., & Bushby, K. M. (2004). Limb-girdle muscular dystrophies - from genetics to molecular pathology. *Neuropathology and Applied Neurobiology, 30*(2), 91-105. *doi:* 10.1111/j.1365-2990.2004.00555.x.

Ledergerber, C., & Dessimoz, C. (2011). Base-calling for next-generation sequencing platforms. *Briefings in Bioinformatics*, *12*(5), 489-497. *doi:* 10.1093/bib/bbq077.

Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., & Quintero-Rivera, F., et al. (2014). Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA*, *312*(18), 1880. *doi:* 10.1001/jama.2014.14604.

Lefeber, D. J., Schönberger, J., Morava, E., Guillard, M., Huyben, K. M., & Verrijp, K., et al. (2009). Deficiency of Dol-P-Man Synthase Subunit DPM3 Bridges the Congenital Disorders of Glycosylation with the Dystroglycanopathies. *The American Journal of Human Genetics, 85*(1), 76-86. *doi:* 10.1016/j.ajhg.2009.06.006.

Lefeber, D. J., Brouwer, A. P., Morava, E., Riemersma, M., Schuurs-Hoeijmakers, J. H., & Absmanner, B., et al. (2011). Autosomal Recessive Dilated Cardiomyopathy due to DOLK Mutations Results from Abnormal Dystroglycan O-Mannosylation. *PLoS Genetics, 7*(12). *doi:* 10.1371/journal.pgen.1002427.

Lehtokari, V. L., Pelin, K., Sandbacka, M., Ranta, S., Donner, K., & Muntoni, F., et al. (2006). Identification of 45 novel mutations in the nebulin gene associated with autosomal recessive nemaline myopathy. *Human Mutation*, *27*(9), 946-956. *doi*: 10.1002/humu.20370.

Lehtokari, V. L., Kiiski, K., Sandaradura, S. A., Laporte, J., Repo, P., & Frey, J. A., et al. (2014). Mutation update: the spectra of nebulin variants and associated myopathies. *Human Mutation*, *35*(12), 1418-1426. *doi:* 10.1002/humu.22693.

Leigh, F., Ferlini, A., Biggar, D., Bushby, K., Finkel, R., Morgenroth, L. P., & Wagner, K. R. (2018). Neurology Care, Diagnostics, and Emerging Therapies of the Patient with Duchenne Muscular Dystrophy. Pediatrics, 142(Supplement 2). *doi:* 10.1542/peds.2018-0333c.

Leivo, I., & Engvall, E. (1998). Merosin, a protein specific for basement membranes of Schwann cells, striated muscle, and trophoblast, is expressed late in nerve and muscle development. *Proceedings of the National Academy of Scinces of the United States of America*, *85*(5), 1544-1548. *doi:* 10.1073/pnas.85.5.1544.

Lejeune, J., Gautier, M., & Turpin, R. (1959). [Study of somatic chromosomes from 9 mongoloid children]. Comptes rendus de l'Académie des Sciences, *248*(11), 1721-1722.

Lelieveld, S. H., Veltman, J. A., & Gilissen, C. (2016). Novel bioinformatic developments for exome sequencing. Human Genetics, 135(6), 603-614. doi: 10.1007/s00439-016-1658-6.

Lemmers, R. J., Vliet, P. J., Klooster, R., Sacconi, S., Camano, P., & Dauwerse, J. G., et al. (2010). A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy. *Science, 329*(5999), 1650-1653. *doi:* 10.1126/science.1189044

Lemmers, R. J., Tawil, R., Petek, L. M., Balog, J., Block, G. J., & Santen, G. W., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nature Genetics, 44*(12), 1370-1374. *doi:* 10.1038/ng.2454.

Leventer, R. J., Guerrini, R., & Dobyns, W. B. (2008). Malformations of cortical development and epilepsy. *Dialogues in Clinical Neurosciences*, *10*(1), 47-62. *doi:* 10.1101/cshperspect.a022392.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078-2079. doi: 10.1093/bioinformatics/btp352.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*(5), 589-595. *doi:* 10.1093/bioinformatics/btp698.

Li, Y., & Tollefsbol, T. O. (2011). DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis. *Methods in Molecular Biology Epigenetics Protocols*, 11-21. doi: 10.1007/978-1-61779-316-5_2.

Liang, W., Mitsuhashi, H., Keduka, E., Nonaka, I., Noguchi, S., Nishino, I., & Hayashi, Y. K. (2011). TMEM43 mutations in emery-dreifuss muscular dystrophy-related myopathy. *Annals of Neurology, 69*(6), 1005-1013. *doi:* 10.1002/ana.22338.

Liang, W., Zhu, W., Mitsuhashi, S., Noguchi, S., Sacher, M., & Ogawa, M., et al. (2015). Congenital muscular dystrophy with fatty liver and infantile-onset cataract caused by TRAPPC11 mutations: Broadening of the phenotype. *Skeletal Muscle, 5*(1). *doi:* 10.1186/s13395-015-0056-4.

Liewluck, T., Sorenson, E. J., Walkiewicz, M. A., Rumilla, K. M., & Milone, M. (2017). Autosomal dominant distal myopathy due to a novel ACTA1 mutation. *Neuromuscular Disorders, 27*(8), 742-746. *doi:* 10.1016/j.nmd.2017.05.003.

Lindgreen, S. (2012). AdapterRemoval: Easy Cleaning of Next Generation Sequencing Reads. *BMC Research Notes*, 5(1), 337. *doi:* 10.1186/1756-0500-5-337.

Liu, J., Aoki, M., Illa, I., Wu, C., Fardeau, M., & Angelini, C., et al. (1998). Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nature Genetics, 20*(1), 31-36. *doi:* 10.1038/1682.

Lo, Y. M., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W., & Wainscoat, J. S. (1997). Presence of fetal DNA in maternal plasma and serum. *The Lancet, 350*(9076), 485-487. *doi:* 10.1016/s0140-6736(97)02174-0.

Logan, C., Lucke, B., Pottinger, C., Abdelhamed, Z., Parry, D., & Szymanska, K. et al. (2011). Mutations in MEGF10, a regulator of satellite cell myogenesis, cause early onset myopathy, areflexia, respiratory distress and dysphagia (EMARDD). *Nature Genetics, 43*(12): 1189-1192. *doi:* 10.1038/ng.995.

Løkken, N., Born, A.P., Duno, M., & Vissing, J. (2015). LAMA2-related myopathy: Frequency among congenital and limb-girdle muscular dystrophies. *Muscle & Nerve, 52*(4), 547-553. *doi*: 10.1002/mus.24588.

Long, C., Amoasii, L., Mireault, A. A., Mcanally, J. R., Li, H., & Sanchez-Ortiz, E., et al. (2016). Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science, 351*(6271), 400-403. *doi:* 10.1126/science.aad5725.

Longman, C., Brockington, M., Torelli, S., Jimenez-Mallebrera, C., Kennedy, C., & Khalil, N., et al. (2003). Mutations in the human LARGE gene cause MDC1D, a novel form of congenital muscular dystrophy with severe mental retardation and abnormal glycosylation of alpha-dystroglycan. *Human Molecular Genetics, 12*(21), 2853-2861. *doi:* 10.1093/hmg/ddg307.

Loy, R. E., Orynbayev, M., Xu, L., Andronache, Z., Apostol, S., & Zvaritch, E., et al. (2011). Muscle weakness in Ryr1I4895T/WT knock-in mice as a result of reduced ryanodine receptor Ca2+ ion permeation and release from the sarcoplasmic reticulum. *The Journal of general physiology, 137*(1), 43-57. *doi:* 10.1085/jgp.201010523.

Lyfenko, A. D., Ducreux, S., Wang, Y., Xu, L., Zorzato, F., & Ferreiro, A., et al. (2007). Two central core disease (CCD) deletions in the C-terminal region of RYR1 alter muscle excitation-contraction (EC) coupling by distinct mechanisms. *Human mutation, 28*(1), 61-68. *doi:* 10.1002/humu.20409.

Lynch, P. J., Tong, J., Lehane, M., Mallet, A., Giblin, L., & Heffron, J. J., et al. (1999). A mutation in the transmembrane/luminal domain of the ryanodine receptor is associated with abnormal Ca2+ release channel function and severe central core disease. *Proceedings of the National Academy of Sciences of the United States of America, 96*(7), 4164-4169. *doi:* 10.1002/humu.20409.

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., & Walter, K., et al. (2012). A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science, 335*(6070), 823-828. *doi:* 10.1126/science.1215040.

Maertens, G. N., Cook, N. J., Wang, W., Hare, S., Gupta, S. S., & Oztop, I., et al. (2014). Structural basis for nuclear import of splicing factors by human Transportin 3. *Proceedings of the National Academy of Sciences, 111*(7), 2728-2733. *doi:* 10.1073/pnas.1320755111.

Magee, K. R., & Shy, G. M. (1956). A new congenital non-progressive myopathy. *Brain*, *79*(4), 610-621.

Maggi, L., Scoto, M., Cirak, S., Robb, S. A., Klein, A., & Lillis, S., et al. (2013). Congenital myopathies--clinical features and frequency of individual subtypes diagnosed over a 5-year period in the United Kingdom. *Neuromuscular disorders*, *23*(3), 195-205. *doi:* 10.1016/j.nmd.2013.01.004.

Magi, A., Tattini, L., Palombo, F., Benelli, M., Gialluisi, A., & Giusti, B., et al. (2014). H3M2: detection of runs of homozygosity from whole-exome sequencing data. *Bioinformatics*, *30*(20), 2852-2859. *doi:* 10.1093/bioinformatics/btu401.

Mah, J. K., & Joseph, J. T. (2016). An Overview of Congenital Myopathies. *Continuum (Minneap Minn)*, *22*(6), 1932-1953. *doi:* 10.1212/CON.0000000000000404.

Mahdy MAA. (2019). Skeletal muscle fibrosis: an overview. *Cell Tissue Research*. [Epub ahead of print]. *doi:* 10.1007/s00441-018-2955-2.

Majczenko, K., Davidson, A., Camelo-Piragua, S., Agrawal, P., Manfready, R., & Li, X., et al. (2012). Dominant Mutation of CCDC78 in a Unique Congenital Myopathy with Prominent Internal Nuclei and Atypical Cores. *The American Journal of Human Genetics*, *91*(2): 365-371. *doi*: 10.1016/j.ajhg.2012.06.012.

Makri, S., Clarke, N. F., Richard, P., Maugenre, S., Demay, L., & Bonne, G., et al. (2009). Germinal mosaicism for LMNA mimics autosomal recessive congenital muscular dystrophy. *Neuromuscular Disorders*, *19*(1):26-28. *doi:* 10.1016/ j.nmd.2008.09.016

Malfatti, E., Böhm, J., Lacène, E., Beuvin, M., Guy Brochier, & Romero, N. et al. (2015). A Premature Stop Codon in MYO18B is Associated with Severe Nemaline Myopathy with Cardiomyopathy. *Journal of Neuromuscular Diseases*, *2*(3), 219-227. *doi:* 10.3233/JND-150085.

Malfatti, E., & Romero, N. B. (2017). Diseases of the skeletal muscle. *Handbook of Clinical Neurology Neuropathology,145*, 429-451. *doi:* 10.1016/b978-0-12-802395-2.00030-4.

Malito, E., Sekulic, N., Too, W. C., Konrad, M., & Lavie, A. (2006). Elucidation of human choline kinase crystal structures in complex with the products ADP or phosphocholine. Journal of Molecular Biology, 364(2), 136-151. *doi:* 10.1016/j.jmb.2006.08.084.

Manzini, M., Tambunan, D., Hill, R., Yu, T., Maynard, T., & Heinzen, E., et al. (2012). Exome Sequencing and Functional Validation in Zebrafish Identify GTDC2 Mutations

as a Cause of Walker-Warburg Syndrome. *The American Journal of Human Genetics, 91*(3), 541-547. *doi:* 10.1016/j.ajhg.2012.07.009.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., & Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature, 437*(7057), 376-380. *doi:* 10.1038/nature03959.

Marks, S., van Ruitenbeek, E., Fallon, P., Johns, P., Phadke, R., & Mein, R., et al. (2018). Parental mosaicism in RYR1-related Central Core Disease. *Neuromuscular Disorders, 28*(5), 422-426. *doi:* 10.1016/j.nmd.2018.02.011.

Marques, J., Duarte, S. T., Costa, S., Jacinto, S., Oliveira, J., & Oliveira, M. E., et al. (2014). Atypical phenotype in two patients with LAMA2 mutations. *Neuromuscular Disorders, 24*(5), 419-424. *doi:* 10.1016/j.jmb.2006.08.084.

Martinez-Thompson, J. M., Niu, Z., Tracy, J. A., Moore, S. A., Swenson, A., Wieben, E. D., & Milone, M. (2018). Autosomal dominant calpainopathy due to heterozygous CAPN3 C.643_663del21. *Muscle & Nerve, 57*(4), 679-683. *doi:* 10.1002/mus.25970

Marty, I., & Fauré, J. (2016). Excitation-Contraction Coupling Alterations in Myopathies. *Journal Neuromuscular Disorders, 3*(4), 443-453. *doi:* 10.3233/JND-160172.

Martinello, F., Angelini, C., & Trevisan, C. P. (1998). Congenital muscular dystrophy with partial merosin deficiency and late onset epilepsy. *European Neurology, 40*(1), 37-45. *doi:* 10.1159/000007954.

Matalonga, L., Bravo, M., Serra-Peinado, C., García-Pelegrí, E., Ugarteburu, O., & Vidal, S., et al. (2016). Mutations in TRAPPC11 are associated with a congenital disorder of glycosylation. *Human Mutation, 38*(2), 148-151. *doi:* 10.1002/humu.23145.

Matsuda, C., Hayashi, Y. K., Ogawa, M., Aoki, M., Murayama, K., & Nishino, I., et al. (2001). The sarcolemmal proteins dysferlin and caveolin-3 interact in skeletal muscle. *Human Molecular Genetics, 10*(17), 1761-1766. *doi:* 10.1093/hmg/10.17.1761.

Maxam, A. M., & Gilbert, W. (1972). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of USA, 74*(2):560-564.

McCall, C. M., Mosier, S., Thiess, M., Debeljak, M., Pallavajjala, A., & Beierl, K., et al. (2014). False positives in multiplex PCR-based next-generation sequencing have unique signatures. *The Journal of Molecular Diagnosis, 16*(5), 541-549. *doi:* 10.1016/j.jmoldx.2014.06.001.

McCarthy, A. (2010). Third Generation DNA Sequencing: Pacific Biosciences Single Molecule Real Time Technology. *Chemistry & Biology, 17*(7), 675-676. *doi:* 10.1016/j.chembiol.2010.07.004.

McCarthy, D. J., Humburg, P., Kanapin, A., Rivas, M. A., Gaulton, K., Cazier, J., & Donnelly, P. (2014). Choice of transcripts and software has a large effect on variant annotation. *Genome Medicine,* 6(3), 26. *doi:* 10.1186/gm543.

McEntagart, M., Parsons, G., Buj-Bello, A., Biancalana, V., Fenton, I., & Little, M., et al. (2002). Genotype-phenotype correlations in X-linked myotubular myopathy. *Neuromuscular Disorders*, *12*(10), 939-946. *doi*: 10.1016/S0960-8966(02)00153-0.

McKee, K. K., Crosson, S. C., Meinen, S., Reinhard, J. R., Rüegg, M. A., & Yurchenco, P. D. (2017). Chimeric protein repair of laminin polymerization ameliorates muscular dystrophy phenotype. *The Journal of Clinical Investigation*, *127*(3), 1075-1089. *doi:* 10.1172/JCI90854.

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., & Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. *doi:* 10.1101/gr.107524.110.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., & Thormann, A., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biology,* 17(1). *doi:* 10.1186/s13059-016-0974-4.

McNally, E. M., Lapidos, K. A., & Wheeler, M. T. (2006). Skeletal Muscle Structure and Function, Editor(s): Runge, M. S., Patterson, C. *Principles of Molecular Medicine*. *Humana Press. doi:* 10.1007/978-1-59259-963-9.

McNally, E. M., & Pytel, P. (2007). Muscle diseases: the muscular dystrophies. *Annual Reviews of Pathology*, 2, 87-109. *doi:* 10.1146/annurev.pathol.2.010506.091936.

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., & Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *American journal of human genetics*, *83*(3), 359-372. *doi*: 10.1016/j.ajhg.2008.08.007.

Meienberg, J., Bruggmann, R., Oexle, K., & Matyas, G. (2016). Clinical sequencing: Is WGS the better WES? *Human Genetics*, *135*(3), 359-362. *doi:* 10.1007/s00439-015-1631-9.

Meischl, C., Boer, M., Ahlin, A., & Roos, D. (2000). A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. European Journal of Human Genetics, *8*(9), 697-703. *doi:* 10.1038/sj.ejhg.5200523.

Melià, M. J., Kubota, A., Ortolano, S., Vílchez, J. J., Gámez, J., & Tanji, K., et al. (2013). Limb-girdle muscular dystrophy 1F is caused by a microdeletion in the transportin 3 gene. *Brain, 136*(5), 1508-1517. *doi:* 10.1093/brain/awt074.

Mendell, J. R., Shilling, C., Leslie, N. D., Flanigan, K. M., al-Dahhak, R., & Gastier-Foster, J. et al. (2012). Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Annals of Neurology*, *71*(3), 304-313. *doi:* 10.1002/ana.23528.

Menezes, M. J., McClenahan, F. K., Leiton, C. V., Aranmolate, A., Shan, X., & Colognato, H. (2014). The extracellular matrix protein laminin α2 regulates the

maturation and function of the blood-brain barrier. *The Journal of Neuroscience*, *34*(46), 15260-15280. *doi:* 10.1523/JNEUROSCI.3678-13.2014.

Mensch, A., Meinhardt, B., Bley, N., Hüttelmaier, S., Schneider, I., & Stoltenburg-Didinger, G., et al. (2018). The p.S85C-mutation in MATR3 impairs stress granule formation in Matrin-3 myopathy. *Experimental Neurology, 306*, 222-231. *doi:* 10.1016/j.expneurol.2018.05.012.

Mercuri, E., Gruter-Andrew, J., Philpot, J., Sewry, C., Counsell, S., & Henderson, S., et al. (1999). Cognitive abilities in children with congenital muscular dystrophy: correlation with brain MRI and merosin status. *Neuromuscular Disorders, 9*(6-7), 383-387.

Mercuri, E., Rutherford, M., Vile, C. D., Counsell, S., Sewry, C., & Brown, S., et al. (2001). Early white matter changes on brain magnetic resonance imaging in a newborn affected by merosin-deficient congenital muscular dystrophy. *Neuromuscular Disorders, 11*(3), 297-299. *doi:* 10.1016/s0960-8966(00)00190-5.

Mercuri, E., Poppe, M., Quinlivan, R., Messina, S., Kinali, M., & Demay, L., et al. (2004). Extreme Variability of Phenotype in Patients With an Identical Missense Mutation in the Lamin A/C Gene. *Archives of Neurology, 61*(5), 690. *doi:* 10.1001/archneur.61.5.690.

Mercuri, E., Clements, E., Offiah, A., Pichiecchio, A., Vasco, G., & Bianco, F, et al. (2010). Muscle magnetic resonance imaging involvement in muscular dystrophies with rigidity of the spine. *Annals of Neurology, 67*(2), 201-208. *doi:* 10.1002/ana.21846.

Mercuri, E., & Muntoni, F. (2013). Muscular dystrophies. *The Lancet, 381*(9869), 845-860. *doi:* 10.1016/s0140-6736(12)61897-2.

Meredith, C., Herrmann, R., Parry, C., Liyanage, K., Dye, D. E., & Durling, H. J., et al. (2004). Mutations in the Slow Skeletal Muscle Fiber Myosin Heavy Chain Gene (MYH7) Cause Laing Early-Onset Distal Myopathy (MPD1). *The American Journal of Human Genetics, 75*(4), 703-708. *doi:* 10.1086/424760.

Merner, N. D., Hodgkinson, K. A., Haywood, A. F., Connors, S., French, V. M., & Drenckhahn, J., et al. (2008). Arrhythmogenic Right Ventricular Cardiomyopathy Type 5 Is a Fully Penetrant, Lethal Arrhythmic Disorder Caused by a Missense Mutation in the TMEM43 Gene. *The American Journal of Human Genetics, 82*(4), 809-821. *doi:* 10.1016/j.ajhg.2008.01.010.

Merriman, B., Team, I. T., & Rothberg, J. M. (2012). Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis, 33*(23), 3397-3417. doi: 10.1002/elps.201200424.

Mertes, F., Elsharawy, A., Sauer, S., J. M. L. M. Van Helvoort, Zaag, P. J., & Franke, A., et al. (2011). Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics*, *10*(6), 374-386. *doi:* 10.1093/bfgp/elr033.

Messina, S., Bruno, C., Moroni, I., Pegoraro, E., D'Amico, A., & Biancheri, R., et al. (2010). Congenital muscular dystrophies with cognitive impairment. A population study. *Neurology, 75*(10), 898-903. *doi:* 10.1212/WNL.0b013e3181f11dd5.

Mikheyev, A. S., & Tin, M. M. (2014). A first look at the Oxford Nanopore MinION sequencer. Molecular Ecology Resources, 14(6), 1097-1102. doi: 10.1111/1755-0998.12324.

Miller, S. A., Dykes, D. D. & Polesky H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research, *16*(3), 1215.

Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, *95*(6), 315-327. *doi:* 10.1016/j.ygeno.2010.03.001.

Milone, M., & Liewluck, T. (2018). The unfolding spectrum of inherited distal myopathies. *Muscle & Nerve. doi:* 10.1002/mus.26332.

Minetti, C., Sotgia, F., Bruno, C., Scartezzini, P., Broda, P., & Bado, M., et al. (1998). Mutations in the caveolin-3 gene cause autosomal dominant limb-girdle muscular dystrophy. *Nature Genetics, 18*(4), 365-368. doi:10.1038/ng0498-365.

Mirabella, A. C., Foster, B. M., & Bartke, T. (2016). Chromatin deregulation in disease. *Chromosoma*, *125*(1), 75-93. *doi:* 10.1007/s00412-015-0530-0.

Mitsuhashi, S., Hatakeyama, H., Karahashi, M., Koumura, T., Nonaka, I., & Hayashi, Y.K., et al. (2011a). Muscle choline kinase beta defect causes mitochondrial dysfunction and increased mitophagy. *Human Molecular Genetics*, *20*(19), 3841-3851. *doi*: 10.1093/hmg/ddr305.

Mitsuhashi, S., Ohkuma, A., Talim, B., Karahashi, M., Koumura, T., & Aoyama, C., et al. (2011b). A congenital muscular dystrophy with mitochondrial structural abnormalities caused by defective de novo phosphatidylcholine biosynthesis. *American Journal of Human Genetics*, *88*(6), 845-851. *doi:* 10.1016/j.ajhg.2011.05.010.

Mitsuhashi, S., & Nishino, I. (2011c). Phospholipid synthetic defect and mitophagy in muscle disease. *Autophagy*, *7*(12), 1559-1561. *doi*: 10.4161/auto.7.12.17925.

Mitsuhashi, S., & Nishino, I. (2013). Megaconial congenital muscular dystrophy due to loss-of-function mutations in choline kinase β. *Current Opinion in Neurology*, *26*(5), 536-543. *doi:* 10.1097/WCO.0b013e328364c82d.

Miyagoe, Y., Hanaoka, K., Nonaka, I., Hayasaka, M., Nabeshima, Y., & Arahata, K., et al. (1997). Laminin alpha2 chain-null mutant mice by targeted disruption of the Lama2

gene: a new model of merosin (laminin 2)-deficient congenital muscular dystrophy. *FEBS Letter*, *415*(1):33-39. *doi*: 10.1016/S0014-5793(97)01007-7.

Miyatake, S., Koshimizu, E., Hayashi, Y. K., Miya, K., Shiina, M., & Nakashima, M., et al. (2014). Deep sequencing detects very-low-grade somatic mosaicism in the unaffected mother of siblings with nemaline myopathy. *Neuromuscular Disorders, 24*(7), 642-647. *doi:* 10.1016/j.nmd.2014.04.002.

Miyatake, S., Mitsuhashi, S., Hayashi, Y. K., Purevjav, E., Nishikawa, A., & Koshimizu, E. et al. (2017). Biallelic Mutations in MYPN, Encoding Myopalladin, Are Associated with Childhood-Onset, Slowly Progressive Nemaline Myopathy*. American Journal of Human Genetics*, *100*(1), 169-178. *doi:* 10.1016/j.ajhg.2016.11.017.

Moghadaszadeh, B., Petit, N., Jaillard, C., Brockington, M., Roy, S. Q., Merlini, L., et al. (2001). Mutations in SEPN1 cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome. *Nature Genetics, 29*(1), 17-18. *doi:* 10.1038/ng713.

Monnier, N., Romero, N. B., Lerale, J., Landrieu, P., Nivoche, Y., & Fardeau, M., et al. (2001). Familial and sporadic forms of central core disease are associated with mutations in the C-terminal domain of the skeletal muscle ryanodine receptor. *Human molecular genetics*, *10*(22), 2581-2592. *doi*: 10.1093/hmg/10.22.2581.

Mora, M., Moroni, I., Uziel, G., di Blasi, C., Barresi, R., & Farina, L., et al. (1996). Mild clinical phenotype in a 12-year-old boy with partial merosin deficiency and central and peripheral nervous system abnormalities. Neuromuscular Disorders, 6(5), 377-381. *doi*: 10.1016/0960-8966(96)00359-8.

Moreira, E. S., Wiltshire, T. J., Faulkner, G., Nilforoushan, A., Vainzof, M., & Suzuki, O. T., et al. (2000). Limb-girdle muscular dystrophy type 2G is caused by mutations in the gene encoding the sarcomeric protein telethonin. *Nature Genetics, 24*(2), 163-166. *doi:* 10.1038/72822.

Muchir, A., Bonne, G., van der Kooi, A. J., van Meegen, M., Baas, F., & Bolhuis, P. A., et al. (2000). Identification of mutations in the gene encoding lamins A/C in autosomal dominant limb girdle muscular dystrophy with atrioventricular conduction disturbances (LGMD1B). *Human Molecular Genetics, 9*(9), 1453-1459. *doi:* 10.1093/hmg/9.9.1453.

Muntoni, F., Torelli, S., Wells, D. J., & Brown, S. C. (2011). Muscular dystrophies due to glycosylation defects. *Current Opinion in Neurology, 24*(5), 437-442. *doi:* 10.1097/wco.0b013e32834a95e3.

Murakami, T., Hayashi, Y. K., Noguchi, S., Ogawa, M., Nonaka, I., & Tanabe, Y., et al. (2006). Fukutin gene mutations cause dilated cardiomyopathy with minimal muscle weakness. *Annals of Neurology, 60*(5), 597-602. *doi:* 10.1002/ana.20973.

Musova, Z., Hedvicakova, P., Mohrmann, M., Tesarova, M., Krepelova, A., & Zeman, J., et al. (2006). A novel insertion of a rearranged L1 element in exon 44 of the dystrophin gene: Further evidence for possible bias in retroposon integration. *Biochemical and Biophysical Research Communications*, 347(1), 145-149. *doi*: 10.1016/j.bbrc.2006.06.071.

Nakano, K., Shiroma, A., Shimoji, M., Tamotsu, H., Ashimine, N., & Ohki, S., et al. (2017). Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area. *Human Cell*, 30(3), 149-161. *doi:* 10.1007/s13577-017-0168-8.

Nance, J. R., Dowling, J. J., Gibbs, E. M., & Bönnemann, C. G. (2012). Congenital Myopathies: An Update. *Current Neurology and Neuroscience Reports*, *12*(2), 165-174. *doi:* 10.1007/s11910-012-0255-x.

Naom, I., D'Alessandro, M., Sewry, C. A., Philpot, J., Manzur, A. Y., & Dubowitz, V., et al. (1998). Laminin a2-chain gene mutations in two siblings presenting with limb-girdle muscular dystrophy. *Neuromuscular Disorders*, *8*(7): 495-501. *doi:* 10.1016/S0960-8966(98)00065-0.

Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., & Minami, R., et al. (1993). Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *The Journal of Clinical Investigation*, *91*(5), 1862-1867. *doi:* 10.1172/JCI116402.

Nascimento, A., Jou, C., Ortez, C., Hayashi, Y.K., Nishino, I., & Olivé, M., et al. (2013). P.15.11 Megaconial congenital muscular dystrophy in two children with mutations in the CHKB Gene. *Neuromuscular Disorders*, *23*(9-10), 822. *doi:* 10.1016/j.nmd.2013.06.636.

Nelson, I., Stojkovic, T., Allamand, V., Leturcq, F., Bécane, H. M., & Babuty, D., et al. (2015). Laminin α2 Deficiency-Related Muscular Dystrophy Mimicking Emery-Dreifuss and Collagen VI related Diseases. *Journal of Neuromuscular Diseases*, *2*(3), 229-240. *doi:* 10.3233/JND-150093.

Nesin, V., Wiley, G., Kousi, M., Ong, E., Lehmann, T., & Nicholl, D. et al. (2014). Activating mutations in STIM1 and ORAI1 cause overlapping syndromes of tubular myopathy and congenital miosis. *Proceedings of the National Academy of Sciences of USA*, *111*(11), 4197-4202. *doi:* 10.1073/pnas.1312520111.

Ng, P. C. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812-3814. *doi:* 10.1093/nar/gkg509.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., & Lee, C., et al. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, *461*(7261), 272-276. *doi:* 10.1038/nature08250.

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., & Dent, K. M., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*, 42(1), 30-35. *doi:* 10.1038/ng.499.

Nicot, A. S., Toussaint, A., Tosch, V., Kretz, C., Wallgren-Pettersson, C., & Iwarsson, E., et al. (2007). Mutations in amphiphysin 2 (BIN1) disrupt interaction with dynamin 2 and cause autosomal recessive centronuclear myopathy. *Nature Genetics*, *39*(9), 1134-1139. *doi*: 10.1038/ng2086.

Niedringhaus, T. P., Milanova, D., Kerby, M. B., Snyder, M. P., & Barron, A. E. (2011). Landscape of Next-Generation Sequencing Technologies. *Analytical Chemistry*, *83*(12), 4327-4341. *doi:*10.1021/ac2010857.

Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12(6), 443-451. *doi:* 10.1038/nrg2986.

Nigro, V., & Savarese, M. (2014). Genetic basis of limb-girdle muscular dystrophies: the 2014 update. *Acta Myologica*, *33*(1), 1-12.

Niroula, A., & Vihinen, M. (2016). Variation Interpretation Predictors: Principles, Types, Performance, and Choice. Human Mutation, 37(6), 579-597. doi: 10.1002/humu.22987.

Nigro, V., Moreira, E. D., Piluso, G., Vainzof, M., Belsito, A., & Politano, L., et al. (1996). Autosomal recessive limbgirdle muscular dystrophy, LGMD2F, is caused by a mutation in the δ–sarcoglycan gene. *Nature Genetics, 14*(2), 195-198. *doi:* 10.1038/ng1096-195.

Nikolic, A., Ricci, G., Sera, F., Bucci, E., Govi, M., & Mele, F., et al. (2016). Clinical expression of facioscapulohumeral muscular dystrophy in carriers of 1–3 D4Z4 reduced alleles: Experience of the FSHD Italian National Registry. *BMJ Open, 6*(1). doi:10.1136/bmjopen-2015-007798

Nishino, I., Kobayashi, O., Goto, Y., Kurihara, M., Kumagai, K., & Fujita, T., et al. (1998a). A new congenital muscular dystrophy with mitochondrial structural abnormalities. *Muscle & Nerve*, 21(1), 40-47. *doi:* 10.1002/(SICI)1097-4598(199801)21:1<40::AID-MUS6>3.0.CO;2-G.

Nishino, I., Minami, N., Kobayashi, O., Ikezawa, M., Goto, Y., & Arahata, K., et al. (1998b). MTM1 gene mutations in Japanese patients with the severe infantile form of myotubular myopathy. *Neuromuscular Disorders*, 8(7), 453-458. *doi:* 10.1016/S0960-8966(98)00075-3.

Nissinen, M., Helbling-Leclerc, A., Zhang, X., Evangelista, T., Topaloglu, H., & Cruaud C, et al. (1996). Substitution of a conserved cysteine-996 in a cysteine-rich motif of the laminin alpha2-chain in congenital muscular dystrophy with partial deficiency of the protein. *American Journal of Human Genetics*, *58*(6), 1177-1784.

Noguchi, S., Mcnally, E. M., Othmane, K. B., Hagiwara, Y., Mizuno, Y., & Yoshida, M., et al. (1995). Mutations in the Dystrophin-Associated Protein gamma-Sarcoglycan in Chromosome 13 Muscular Dystrophy. *Science, 270*(5237), 819-822. *doi:* 10.1126/science.270.5237.819.

Noritz, G. H., & Murphy, N. A. (2013). Motor Delays: Early Identification and Evaluation. *Pediatrics, 131*(6). *doi:* 10.1542/peds.2013-1056.

Norwitz, E. R., & Levy, B. (2013). Noninvasive prenatal testing: the future is now. *Reviews Obstetrics & Gynecology*, *6*(2):48-62. *doi:* 10.3909/riog0201.

North, K. N., Wang, C. H., Clarke, N., Deconinck, N., Bertini, E., & Ferreiro, A., et al. (2014). Approach to the diagnosis of congenital myopathies. *Neuromuscular Disorders*, *24*(2), 97-116. *doi:* 10.1016/j.nmd.2013.11.003.

Nowak, K., Wattanasirichaigoon, D., Goebel, H., Wilce, M., Pelin, K., & Donner, et al. (1999). Mutations in the skeletal muscle α-actin gene in patients with actin myopathy and nemaline myopathy. *Nature Genetics*, *23*(2), 208-212. *doi:* 10.1038/13837.

O'Grady, G. L., Best, H. A., Oates, E. C., Kaur, S., Charlton, A., & Brammah, S., et al. (2014). Recessive ACTA1 variant causes congenital muscular dystrophy with rigid spine. *European Journal of Human Genetics, 23*(6), 883-886. *doi:* 10.1038/ejhg.2014.169.

O'Grady, G. L., Best, H. A., Sztal, T. E., Schartner, V., Sanjuan-Vazquez, M., & Donkervoort, S., et al. (2016). Variants in the Oxidoreductase PYROXD1 Cause Early-Onset Myopathy with Internalized Nuclei and Myofibrillar Disorganization. *The American Journal of Human Genetics, 99*(5), 1086-1105. *doi:* 10.1016/j.ajhg.2016.09.005.

O'Neil, D., Glowatz, H., & Schlumpberger, M. (2013). Ribosomal RNA Depletion for Efficient Use of RNA-Seq Capacity. *Current Protocols in Molecular Biology*, Chapter 4:Unit 4.19. *doi*: 10.1002/0471142727.mb0419s103.

O'Neill, M., Mcpartlin, J., Arthure, K., Riedel, S., & Mcmillan, N. (2011). Comparison of the TLDA with the Nanodrop and the reference Qubit system. *Journal of Physics: Conference Series*, 307, 012047. *doi:* 10.1088/1742-6596/307/1/012047.

Oepkes, D., Page-Christiaens, G. C., Bax, C. J., Bekker, M. N., Bilardo, C. M., & Boon, E. M., et al. (2016). Trial by Dutch laboratories for evaluation of non-invasive prenatal testing. Part I-clinical impact. *Prenatal Diagnosis*, *36*(12), 1083-1090. *doi:* 10.1002/pd.4945.

Ogata, T., Wada, Y., & Fukami, M. (2008). MAMLD1 (CXorf6): a new gene for hypospadias. Sexual Development, 2(4-5), 244-250. *doi:* 10.1159/000152040.

Oliveira, J., Soares-Silva, I., Fokkema, I., Gonçalves, A., Cabral, A., & Diogo, L., et al. (2008a). Novel synonymous substitution in POMGNT1 promotes exon skipping in a patient with congenital muscular dystrophy. Journal of Human Genetics, 53(6), 565-572. *doi:* 10.1007/s10038-008-0263-5.

Oliveira, J., Santos, R., Soares-Silva, I., Jorge, P., Vieira, E., & Oliveira, M. E., et al. (2008b). LAMA2 gene analysis in a cohort of 26 congenital muscular dystrophy patients. *Clinical Genetics*, *74*(6), 502-512. *doi:* 10.1111/j.1399-0004.2008.01068.x.

Oliveira, J., Dias, C., Redeker, E., Costa, E., Silva, J., & Reis Lima, M., et al. (2010). Development of NIPBL locus-specific database using LOVD: from novel mutations to further genotype-phenotype correlations in Cornelia de Lange Syndrome. *Human Mutation*, *31*(11), 1216-1222. *doi:* 10.1002/humu.21352.

Oliveira, J., Oliveira, M. E., Kress, W., Taipa, R., Pires, M. M., & Hilbert, P., et al. (2013). Expanding the MTM1 mutational spectrum: novel variants including the first multi-exonic duplication and development of a locus-specific database. *European Journal of Human Genetics, 21*(5), 540-549. *doi:* 10.1038/ejhg.2012.201.

Oliveira, J., Gonçalves, A., Oliveira, M. E., Fineza, I., Pavanello, R. C., & Vainzof, M., et al. (2014). Reviewing Large LAMA2 Deletions and Duplications in Congenital Muscular Dystrophy Patients. *Journal of Neuromuscular Diseases*, *1*(2), 169-179. *doi:* 10.3233/JND-140031.

Oliveira, J., Negrão, L., Fineza, I., Taipa, R., Melo-Pires, M., & Fortuna, A. M., et al. (2015). New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing. *Journal of Human Genetics*, *60*(6), 305-312. *doi:* 10.1038/jhg.2015.20.

Oliveira, J., Gonçalves, A., Taipa, R., Melo-Pires, M., Oliveira, M. E., & Costa, J. L., et al. (2016a). New massive parallel sequencing approach improves the genetic characterization of congenital myopathies. *Journal of Human Genetics*, *61*(6), 497-505. *doi:* 10.1038/jhg.2016.2.

Oliveira, J., Janssens, S., Gonçalves, A., Vieira, E., Oliveira, M. E., & Moreno, T., et al. (2016b). Genetic and clinical heterogeneity of dystroglycanopathies: challenging for next-generation sequencing? Proceedings of the European Society of Human Genetics Conference 2016, Barcelona, Spain. page. 214. Retrieved from: https://www.eshg.org/fileadmin/www.eshg.org/conferences/2016/downloads/ESHG2016_Abstracts_final.pdf.

Oliveira, J., Negrão, L., Fineza, I., Taipa, R., Melo-Pires, M., & Gonçalves, A. R., et al. (2016c). P16 – Beta choline kinase deficiency: a rare cause of muscular dystrophy,

cardiomyopathy and intellectual disability. UMIB Summit 2015. *Medicine*, *95*(10), e02371. *doi:* 10.1097/MD.0000000000002371.

Oliveira, J., Martins, M., Pinto Leite, R., Sousa, M., & Santos, R. (2017). The new neuromuscular disease related with defects in the ASC-1 complex: report of a second case confirms ASCC1 involvement. *Clinical Genetics*, *92*(4), 434-439. *doi:* 10.1111/cge.12997.

Oliveira, J., Pereira, R., Santos, R., & Sousa, M. (2018a). Evaluating Runs of Homozygosity in Exome Sequencing Data - Utility in Disease Inheritance Model Selection and Variant Filtering. In: Peixoto N., Silveira M., Ali H., Maciel C., van den Broek E. (eds) Biomedical Engineering Systems and Technologies. BIOSTEC 2017. *Communications in Computer and Information Science, vol 881*. Springer, Cham. *doi*: 10.1007/978-3-319-94806-5_15.

Oliveira J., Gonçalves, A., Ariyurek, Y., den Dunnen, J. T., Sousa, M., & Santos, R. (2018b). Whole-genome sequencing detects a large genomic inversion disrupting the *DMD* gene in a Becker muscular dystrophy patient. Proceedings of the European Society of Human Genetics Conference 2018, Milan, Italy.

OMIM. (2018). Online Mendelian Inheritance in Man. Retrieved December 1, 2018, from https://omim.org/.

Orloff, M., Peterson, C., He, X., Ganapathi, S., Heald, B., & Yang, Y. R., et al. (2011). Germline mutations in MSR1, ASCC1, and CTHRC1 in patients with Barrett esophagus and esophageal adenocarcinoma. *JAMA*, *306*(4), 410-419. *doi:* 10.1001/jama.2011.1029.

Ortolano, S., Tarrío, R., Blanco-Arias, P., Teijeira, S., Rodríguez-Trelles, F., & García-Murias, M. et al. (2011). A novel MYH7 mutation links congenital fiber type disproportion and myosin storage myopathy. *Neuromuscular Disorders*, *21*(4), 254-262. *doi:* 10.1016/j.nmd.2010.12.011.

Osborn, D. P., Pond, H. L., Mazaheri, N., Dejardin, J., Munn, C. J., & Mushref, K., et al. (2017). Mutations in INPP5K Cause a Form of Congenital Muscular Dystrophy Overlapping Marinesco-Sjögren Syndrome and Dystroglycanopathy. *The American Journal of Human Genetics, 100*(3), 537-545. *doi:* 10.1016/j.ajhg.2017.01.019.

Paila, U., Chapman, B. A., Kirchner, R. & Quinlan, A. R. (2013). GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Computational Biology*, *9*(7), e1003153. *doi:* 10.1371/journal.pcbi.1003153.

Palmio, J., Sandell, S., Suominen, T., Penttilä, S., Raheem, O., & Hackman, P., et al. (2011). Distinct distal myopathy phenotype caused by VCP gene mutation in a Finnish family. *Neuromuscular Disorders, 21*(8), 551-555. *doi:* 10.1016/j.nmd.2011.05.008.

Palmio, J., & Udd, B. (2016). Myofibrillar and distal myopathies. *Revue Neurologique, 172*(10), 587-593. *doi:* 10.1016/j.neurol.2016.07.019.

Pandey, V., Nutter, R. C., & Prediger, E. (2008). Applied biosystems SOLID system: ligation-based sequencing. In: Janitz M, editor. Next Generation Genome Sequencing: Towards Personalized Medicine. Germany: Wiley-VCH, Weinheim, 431-444.

Park, P. J. (2009). ChIP–seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*, *10*(10), 669-680. *doi:* 10.1038/nrg2641.

Park, H. J., Hong, Y. B., Choi, Y., Lee, J., Kim, E. J., & Lee, J., et al. (2016). ADSSL 1 mutation relevant to autosomal recessive adolescent onset distal myopathy. *Annals of Neurology, 79*(2), 231-243. *doi:* 10.1002/ana.24550.

Patel, R. K., & Jain, M. (2012). NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. *PLoS ONE*, *7*(2). *doi:* 10.1371/journal.pone.0030619.

Pegoraro, E., Marks, H., Garcia, C. A., Crawford, T., Mancias, P., & Connolly, A. M., et al. (1998). Laminin alpha2 muscular dystrophy: genotype/phenotype studies of 22 patients. *Neurology*, *51*(1), 101-110.

Pegoraro, E., Fanin, M., Trevisan, C. P., Angelini, C., Hoffman, E. P. (2000). A novel laminin alpha2 isoform in severe laminin alpha2 deficient congenital muscular dystrophy. *Neurology*, *55*(8), 1128–1134. *doi:* 10.1212/WNL.55.8.1128.

Pejaver, V., Mooney, S. D., & Radivojac, P. (2017). Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. *Human Mutation*, *38*(9), 1092-1108. *doi:* 10.1002/humu.23258.

Pelin, K., Hilpelä, P., Donner, K., Sewry, C., Akkari, P. A., & Wilton, S. D, et al. (1999). Mutations in the nebulin gene associated with autosomal recessive nemaline myopathy. *Proceedings of the National Academy of Sciences of USA*, *96*(5), 2305-2310. *doi:* 10.1073/pnas.96.5.2305.

Pelin, K., Donner, K., Holmberg, M., Jungbluth, H., Muntoni, F., & Wallgren-Pettersson, C. (2002). Nebulin mutations in autosomal recessive nemaline myopathy: an update. *Neuromuscular Disorders*, *12*(7-8), 680-686.

Pemberton, T. J., Absher, D., Feldman, M. W., Myers, R. M., Rosenberg, N. A., & Li, J. Z. (2012). Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics*, *91*(2), 275-292. *doi:* 10.1016/j.ajhg.2012.06.014.

Pénisson-Besnier, I., Talvinen, K., Dumez, C., Vihola, A., Dubas, F., & Fardeau, M., et al. (2006). Myotilinopathy in a family with late onset myopathy. *Neuromuscular Disorders, 16*(7), 427-431. *doi:* 10.1016/j.nmd.2006.04.009.

Pénisson-Besnier, I., Biancalana, V., Reynier, P., Cossée, M., & Dubas, F. (2007). Diagnosis of myotubular myopathy in the oldest known manifesting female carrier: a clinical and genetic study. *Neuromuscular Disorders*, *17*(2), 180-185. doi: 10.1016/ j.nmd.2006.10.008.

Penttilä, S., Palmio, J., & Udd, B. (2012). Anoctaminopathy. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK114459/.

Penzkofer, T., Dandekar, T., & Zemojtel, T. (2005). L1Base: from functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Research*, *33*, 498-500. doi: 10.1093/nar/gki044.

Pereira, R., Oliveira, J., Ferraz, L., Barros, A., Santos, R., & Sousa, M. (2015). Mutation analysis in patients with total sperm immotility. *Journal of assisted reproduction and genetics*, *32*(6), 893-902. *doi:* 10.1007/s10815-015-0474-6.

Pereira, F. L., Soares, S. C., Dorella, F. A., Leal, C. A., & Figueiredo, H. C. (2016). Evaluating the efficacy of the new Ion PGM Hi-Q Sequencing Kit applied to bacterial genomes. Genomics, 107(5), 189-198. doi: 10.1016/j.ygeno.2016.03.004.

Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Research*, *29*, 1185-1190.

Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S., & Goldstein, D. B. (2013). Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *PLoS Genetics*, *9*(8). *doi:* 10.1371/journal.pgen.1003709.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S, Greenblatt, D. M, & Meng, E. C., et al. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605-1612. *doi:* 10.1002/jcc.20084.

Pfeffer, G., Elliott, H. R., Griffin, H., Barresi, R., Miller, J., & Marsh, J., et al. (2012). Titin mutation segregates with hereditary myopathy with early respiratory failure. *Brain, 135*(6), 1695-1713. *doi:* 10.1093/brain/aws102.

Philpot, J., Sewry, C., Pennock, J., & Dubowitz, V. (1995). Clinical phenotype in congenital muscular dystrophy: correlation with expression of merosin in skeletal muscle. *Neuromuscular Disorders*, *5*(4), 301-305.

Philpot, J., Cowan, F., Pennock, J., Sewry, C., Dubowitz, V., & Bydder, G., et al. (1999a). Merosin-deficient congenital muscular dystrophy: the spectrum of brain involvement on magnetic resonance imaging. *Neuromuscular Disorders*, *9*(2), 81-85.

Philpot, J., Bagnall, A., King, C., Dubowitz, V., & Muntoni, F. (1999b). Feeding problems in merosin deficient congenital muscular dystrophy. *Archives of Disease in Childhood*, *80*(6), 542-547.

Piluso, G., Dionisi, M., Del Vecchio Blanco, F., Torella, A., Aurino, S., & Savarese, M., et al. (2011). Motor chip: a comparative genomic hybridization microarray for copy-number mutations in 245 neuromuscular disorders. *Clinical Chemistry*, *57*(11), 1584-1596. doi: 10.1373/clinchem.2011.168898.

Pini, A., Merlini, L., Tomé, F. M., Chevallay, M., & Gobbi, G. (1996). Merosin-negative congenital muscular dystrophy, occipital epilepsy with periodic spasms and focal cortical dysplasia. Report of three Italian cases in two families. *Brain & Development*, *18*(4), 316-322.

Pippucci, T., Benelli, M., Magi, A., Martelli, P. L., Magini, P., & Torricelli, F., et al. (2011). EX-HOM (EXome HOMozygosity): A Proof of Principle. *Human Heredity*, *72*(1), 45-53. *doi:* 10.1159/000330164.

Piteau, S. J., Rossiter, J. P., Smith, R. G., & MacKenzie, J. J. (2014). Congenital myopathy with cap-like structures and nemaline rods: case report and literature review. *Pediatric Neurology*, *51*(2), 192-197. *doi:* 10.1016/j.pediatrneurol.2014.04.002.

Pruitt, K. D., Harrow, J., Harte, R. A., Wallin, C., Diekhans, M., & Maglott, D. R., et al. (2009). The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Research*, *19(*7), 1316-1323. *doi:* 10.1101/gr.080531.108.

Puckelwartz, M., & Mcnally, E. M. (2011). Emery–Dreifuss muscular dystrophy. *Handbook of Clinical Neurology Muscular Dystrophies,* 155-166. *doi:* 10.1016/b978-0-08-045031-5.00012-8.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., & Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, *81*(3), 559–575. *doi:* 10.1086/519795.

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., & Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: Comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. BMC Genomics, 13(1), 341. *doi:* 10.1186/1471-2164-13-341.

Quane, K., Healy, J., Keating, K., Manning, B., Couch, F., & Palmucci, L., et al. (1993). Mutations in the ryanodine receptor gene in central core disease and malignant hyperthermia. *Nature Genetics*, *5*(1), 51-55. *doi:* 10.1038/ng0993-51

Quélin, C., Loget, P., Rozel, C., D'Hervé, D., Fradin, M., & Demurger, F. et al. (2017). Fetal costello syndrome with neuromuscular spindles excess and p.Gly12Val HRAS

mutation. *European Journal of Medical Genetics*, *60*(7), 395-398. *doi:* 10.1016/j.ejmg.2017.03.014.

Quijano-Roy, S., Renault, F., Romero, N., Guicheney, P., Fardeau, M., & Estournet, B. (2004). EMG and nerve conduction studies in children with congenital muscular dystrophy. *Muscle & Nerve, 29*(2), 292-299. *doi:* 10.1002/mus.10544.

Quijano-Roy, S., Mbieleu, B., Bönnemann, C. G., Jeannet, P., Colomer, J., & Clarke, N. F., et al. (2008). De novo LMNA mutations cause a new form of congenital muscular dystrophy. *Annals of Neurology, 64*(2), 177-186. *doi:* 10.1002/ana.21417.

Quijano-Roy, S., Sparks, S., & Rutkowski, A. (2012). LAMA2-Related Muscular Dystrophy. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: http://www.ncbi.nlm.nih.gov/books/NBK97333/.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841-842. *doi*: 10.1093/bioinformatics/btq033.

Quinlivan, R., Mitsuahashi, S., Sewry, C., Cirak, S., Aoyama, C., & Mooore, D. et al. (2013). Muscular dystrophy with large mitochondria associated with mutations in the CHKB gene in three British patients: extending the clinical and pathological phenotype. *Neuromuscular Disorders*, *23*(7), 549-556. *doi*: 10.1016/j.nmd.2013.04.002.

Rahimov, F., & Kunkel, L. M. (2013). The cell biology of disease: cellular and molecular mechanisms underlying muscular dystrophy. *Journal of Cell Biology*, *201*(4), 499-510. *doi:* 10.1083/jcb.201212142.

Rajakulendran, S., Parton, M., Holton, J. L., & Hanna, M. G. (2011). Clinical and pathological heterogeneity in late-onset partial merosin deficiency. Muscle & Nerve, *44*(4), 590-593. doi: 10.1002/mus.22196.

Ramirez-Martinez, A., Cenik, B., Bezprozvannaya, S., Chen, B., Bassel-Duby, R., Liu, N., & Olson, E. (2017). KLHL41 stabilizes skeletal muscle sarcomeres by nonproteolytic ubiquitination. *Elife*, *6*. *doi:* 10.7554/elife.26439.

Ravenscroft, G., Miyatake, S., Lehtokari, V. L., Todd, E. J., Vornanen, P., & Yau, K. S., et al. (2013). Mutations in KLHL40 are a frequent cause of severe autosomal-recessive nemaline myopathy. *American Journal of Human Genetics*, *93*(1), 6-18. *doi:* 10.1016/ j.ajhg.2013.05.004.

Ravenscroft, G., Laing, N. G., & Bönnemann, C. G. (2015). Pathophysiological concepts in the congenital myopathies: blurring the boundaries, sharpening the focus. *Brain*, *138*(Pt 2), 246-268. *doi:* 10.1093/brain/awu368.

Ravenscroft, G., Davis, M. R., Lamont, P., Forrest, A., & Laing, N. G. (2016). New era in genetics of early-onset muscle disease: Breakthroughs and challenges. *Seminars in Cell & Development Biology*, *64*,160-170. *doi*: 10.1016/j.semcdb.2016.08.002.

Reese, M.G., Eeckman, F.H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. *Journal of Computational Biology*, *4*(3), 311-323. doi: 10.1089/cmb.1997.4.311.

Reinhard, J. R., Lin, S., McKee, K. K., Meinen, S., Crosson, S. C., & Sury, M., et al. (2017). Linker proteins restore basement membrane and correct LAMA2-related muscular dystrophy in mice. *Science Translational Medicine*, *9*(396), pii: eaal4649. *doi:* 10.1126/ scitranslmed.aal4649.

Remiche, G., Kadhim, H., Abramowicz, M., Mavroudakis, N., & Monnier, N., Lunardi J. (2015). A novel large deletion in the RYR1 gene in a Belgian family with late-onset and recessive core myopathy. *Neuromuscular Disorders*, *25*(5), 397-402. *doi*: 10.1016/j.nmd.2015.01.016.

Renaud, G., Stenzel, U., & Kelso, J. (2014). LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Research*, *42*(18). *doi:* 10.1093/nar/gku699.

RepeatMasker. Available online: http://www.repeatmasker.org (accessed on 3 March 2017).

Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, Proteomics & Bioinformatics*, *13*(5), 278-289. *doi:* 10.1016/j.gpb.2015.08.002.

Richard, I., Broux, O., Allamand, V., Fougerousse, F., Chiannilkulchai, N., & Bourg, N., et al. (1995). Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. *Cell, 81*(1), 27-40. *doi:* 10.1016/0092-8674(95)90368-2.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., & Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405-424. doi: 10.1038/gim.2015.30.

Robasky, K., Lewis, N. E., & Church, G. M. (2014). The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics*, *15*(1), 56-62. *doi:* 10.1038/nrg3655.

Roberds, S. L., Leturcq, F., Allamand, V., Piccolo, F., Jeanpierre, M., & Anderson, R. D., et al. (1994). Missense mutations in the adhalin gene linked to autosomal recessive muscular dystrophy. *Cell, 78*(4), 625-633. *doi:* 10.1016/0092-8674(94)90527-4.

Robin, J. D., Ludlow, A. T., Laranger, R., Wright, W. E., & Shay, J. W. (2016). Comparison of DNA Quantification Methods for Next Generation Sequencing. Scientific Reports, 6(1). doi: 10.1038/srep24067.

Robinson, F. L., & Dixon, J. E. (2006). Myotubularin phosphatases: policing 3-phosphoinositides. Trends in Cell Biology, *16*(8), 403-412. *doi:* 10.1016/j.tcb.2006.06.001.

Robinson, P. N., Kohler, S., Oellrich, A., Wang, K., Mungall, C. J., & Lewis, S. E., et al. (2014). Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Research*, *24*(2), 340-348. *doi:* 10.1101/gr.160325.113.

Robinson, P. N., Piro, R. M., & Jäger, M. (2017). Chapter 10- Postprocessing the Alignment. Computational exome and genome analysis. Boca Raton, FL: CRC Press.

Robinson, R., Carpenter, D., Shaw, M. A., Halsall, J., & Hopkins, P. (2006). Mutations in RYR1 in malignant hyperthermia and central core disease. *Human mutation*, *27*(10), 977-989. doi: 10.1002/humu.20356.

Rocha, J., Taipa, R., Melo Pires, M., Oliveira, J., Santos, R., & Santos, M. (2014). Ryanodine myopathies without central cores--clinical, histopathologic, and genetic description of three cases. *Pediatric neurology*, *51*(2), 275-278. *doi:* 10.1016/j.pediatrneurol. 2014.04.024.

Rodríguez, J. I., Garcia-Alix, A., Palacios, J., & Paniagua, R. (1988). Changes in the long bones due to fetal immobility caused by neuromuscular disease. A radiographic and histological study. *The Journal of Bone and Joint Surgery American volume*, *70*(7), 1052-1060.

Rodríguez-Martín, C., Cidre, F., Fernández-Teijeiro, A., Gómez-Mariano, G., de la Vega, L., & Ramos, P., et al. (2016). Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene. *Journal of Human Genetics*, *61*(5), 463–466. *doi:* 10.1038/jhg.2015.173.

Romero, N. B., Lehtokari, V. L., Quijano-Roy, S., Monnier, N., Claeys, K. G., & Carlier, R. Y., et al. (2009). Core-rod myopathy caused by mutations in the nebulin gene. *Neurology*, *73*(14), 1159-1161. *doi:* 10.1212/WNL.0b013e3181bacf45.

Romero, N. B. (2010). Centronuclear myopathies: a widening concept. *Neuromuscular Disorders*, *20*(4), 223-228. *doi:* 10.1016/j.nmd.2010.01.014.

Romero, N. B., & Clarke, N. F. (2013). Congenital myopathies. *Handbook of Clinical Neurology*, 113, 1321-1336. *doi:* 10.1016/B978-0-444-59565-2.00004-6.

Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., & Davey, M., et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348-352. *doi:* 10.1038/nature10242.

Rueffert, H., Wehner, M., Ogunlade, V., Meinecke, C. & Schober R. (2009). Mild clinical and histopathological features in patients who carry the frequent and causative

malignant hyperthermia RyR1 mutation p.Thr2206Met. *Clinical neuropathology*, *28*(6), 409-416. *doi:* 10.2379/NPX08167.

Ryan, M. M., Schnell, C., Strickland, C. D., Shield, L. K., Morgan, G., & Iannaccone, S. T, et al. (2001). Nemaline myopathy: a clinical study of 143 cases. *Annals of Neurology*, *50*(3), 312-320. *doi:* 10.1002/ana.1080.

Sacconi, S., Salviati, L., & Desnuelle, C. (2015). Facioscapulohumeral muscular dystrophy. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1852*(4), 607-614. *doi:* 10.1016/j.bbadis.2014.05.021.

Saito, K. (2012). Fukuyama Congenital Muscular Dystrophy. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1206/.

Sakharkar, M. K., Chow, V. T., & Kangueane, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biology*, *4*(4):387-93.

Salgado, D., Desvignes, J. P., Rai, G., Blanchard, A., Miltgen, M., Pinard, A., et al. (2016). UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of any Human cDNA Substitution. *Human Mutation*, *37*(5):439-446. *doi:* 10.1002/humu.22965.

Sambuughin, N., Yau, K., Olivé, M., Duff, R., Bayarsaikhan, M., & Lu, S., et al. (2010). Dominant Mutations in KBTBD13, a Member of the BTB/Kelch Family, Cause Nemaline Myopathy with Cores. *The American Journal of Human Genetics*, *87*(6), 842-847. *doi:* 10.1016/j.ajhg.2010.10.020.

Samorodnitsky, E., Jewell, B. M., Hagopian, R., Miya, J., Wing, M. R., & Lyon, E., et al. (2015). Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. Human Mutation, 36(9), 903-914. doi: 10.1002/humu.22825.

Samuelov, L., Fuchs-Telem, D., Sarig, O., & Sprecher, E. (2011). An exceptional mutational event leading to Chanarin-Dorfman syndrome in a large consanguineous family. *The British Journal of Dermatology*, *164*(6), 1390-1392. *doi:* 10.1111/j.1365-2133.2011.10252.x.

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, *32*(4), 347-55. *doi:* 10.1038/nbt.2842.

Sanger, F., & Coulson, A. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, *94*(3), 441-448. *doi:*10.1016/0022-2836(75)90213-2.

Santos, R., Oliveira, J., Vieira, E., Coelho, T., Carneiro, A. L., & Evangelista, T., et al. (2010). Private dysferlin exon skipping mutation (c.5492GA) with a founder effect

reveals further alternative splicing involving exons 49–51. *Journal of Human Genetics, 55*(8), 546-549. *doi:* 10.1038/jhg.2010.60.

Santos, R., Gonçalves, A., Oliveira, J., Vieira, E., Vieira, J. P., & Evangelista, T., et al. (2014). New variants, challenges and pitfalls in DMD genotyping: Implications in diagnosis, prognosis and therapy. *Journal of Human Genetics*, *59*(8), 454–464. *doi:* 10.1038/jhg.2014.54.

Saredi, S., Ardissone, A., Ruggieri, A., Mottarelli, E., Farina, L., & Rinaldi, R., et al. Novel POMGNT1 point mutations and intragenic rearrangements associated with muscle-eye-brain disease. *Journal of the Neurological Sciences*, *318*(1-2), 45-50. *doi:* 10.1038/jhg.2014.54.

Sarparanta, J., Jonson, P. H., Golzio, C., Sandell, S., Luque, H., & Screen, M., et al. (2012). Mutations affecting the cytoplasmic functions of the co-chaperone DNAJB6 cause limb-girdle muscular dystrophy. *Nature Genetics, 44*(4), 450-455. *doi:* 10.1038/ng.1103.

Sato, I., Wu, S., Ibarra, M. C., Hayashi, Y. K., Fujita, H., & Tojo, M., et al. (2008). Congenital neuromuscular disease with uniform type 1 fiber and RYR1 mutation. *Neurology*, *70*(2), 114-122. *doi:* 10.1212/01.wnl.0000269792.63927.86.

Sauna, Z. E., & Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics*, 12(10), 683-691. doi: 10.1038/nrg3051.

Savarese, M., Di Fruscio, G., Mutarelli, M., Torella, A., Magri, F., & Santorelli F. M., et al. (2014). MotorPlex provides accurate variant detection across large muscle genes both in single myopathic patients and in pools of DNA samples. *Acta Neuropathological Communications*, 2, 100. *doi:* 10.1186/s40478-014-0100-3.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, *19*, 227-230. *doi:*10.1093/hmg/ddq416.

Schara, U., Kress, W., Tücke, J., & Mortier, W. (2003). X-linked myotubular myopathy in a female infant caused by a new MTM1 gene mutation. *Neurology*, *60*(8), 1363-1365.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135-1145. *doi:*10.1038/nbt1486.

Scherer, S. W., Lee, C., Birney, E., Altshuler, D. M., Eichler, E. E., & Carter, N. P., et al. (2007). Challenges and standards in integrating surveys of structural variation. *Nature Genetics*, 39(7s). *doi:* 10.1038/ng2093.

Schessl, J., Zou, Y., & Bönnemann, C. G. (2006). Congenital muscular dystrophies and the extracellular matrix. *Seminars in Pediatric Neuroly*, *13*(2), 80-89. *doi*: 10.1016/j.spen.2006.06.003.

Schessl, J., Zou, Y., Mcgrath, M. J., Cowling, B. S., Maiti, B., & Chin, S. S., et al. (2008). Proteomic identification of FHL1 as the protein mutated in human reducing body myopathy. *Journal of Clinical Investigation. doi:* 10.1172/jci34450.

Schessl, J., Kress, W., & Schoser, B. (2012). Novel ANO5 mutations causing hyper-CKemia, limb girdle muscular weakness and miyoshi type of muscular dystrophy. *Muscle & Nerve, 45*(5), 740-742. *doi:* 10.1002/mus.23281.

Scheuerbrandt, G. (2018). Screening for Duchenne muscular dystrophy in Germany, 1977-2011: A personal story. *Muscle & Nerve, 57*(2), 185-188. *doi:* 10.1002/mus.25979.

Schindler, R. F., Scotton, C., Zhang, J., Passarelli, C., Ortiz-Bonnin, B., & Simrick, S., et al. (2016). POPDC1S201F causes muscular dystrophy and arrhythmia by affecting protein trafficking. *Journal of Clinical Investigation, 126*(1), 239-253. *doi:* 10.1172/jci79562.

Schorling, D. C., Kirschner, J., & Bönnemann, C. G. (2017). Congenital Muscular Dystrophies and Myopathies: An Overview and Update. *Neuropediatrics, 48*(4):247-261. *doi:* 10.1055/s-0037-1604154.

Schoser, B., Frosk, P., Engel, A., Klutzny, U., Lochmüller, H., & Wrogemann, K. (2005). Commonality of TRIM32 mutation in causing sarcotubular myopathy and LGMD2H. *Annals of Neurology, 57*(4), 591-595. *doi:* 10.1002/ana.20441.

Schouten, J. P., McElgunn, C. J., Waaijer, R., Zwijnenburg, D., Diepvens, F., & Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Research, 30*(12). *doi:* 10.1093/nar/gnf056.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., et al. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology, 7*(1), 3. *doi:* 10.1186/1471-2199-7-3.

Seelow, D., Schuelke, M., Hildebrandt, F., & Nürnberg, P. (2009). HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic acids research, 37*(Web Server issue), W593-599. *doi:* 10.1093/nar/gkp369.

Seelow, D., & Schuelke, M., (2012). HomozygosityMapper2012—bridging the gap between homozygosity mapping and deep sequencing. *Nucleic acids research, 40*(Web Server issue), W516-520. *doi:* 10.1093/nar/gks487.

Seidahmed, M. Z., Salih, M. A., Abdelbasit, O. B., Alassiri, A. H., Hussein, K. A., & Miqdad, A., et al. (2016). Gonadal mosaicism forACTA1gene masquerading as autosomal recessive nemaline myopathy. *American Journal of Medical Genetics Part A, 170*(8), 2219-2221. *doi:*10.1002/ajmg.a.37768.

Selcen, D., & Engel, A. G. (2004). Mutations in myotilin cause myofibrillar myopathy. *Neurology, 62*(8), 1363-1371. *doi:* 10.1212/01.wnl.0000123576.74801.75

Selcen, D., & Engel, A. G. (2003). Myofibrillar myopathy caused by novel dominant negative alpha b-crystallin mutations. *Annals of Neurology, 54*(6), 804-810. *doi:* 10.1002/ana.10767.

Selcen, D., & Engel, A. G. (2005). Mutations in ZASP define a novel form of muscular dystrophy in humans. *Annals of Neurology, 57*(2), 269-276. *doi:* 10.1002/ana.20376.

Selcen, D., Muntoni, F., Burton, B. K., Pegoraro, E., Sewry, C., Bite, A. V., & Engel, A. G. (2009). Mutation in BAG3 causes severe dominant childhood muscular dystrophy. *Annals of Neurology, 65*(1), 83-89. *doi:* 10.1002/ana.21553.

Selcen, D. (2011). Myofibrillar myopathies. *Neuromuscular Disorders, 21*(3), 161-171. *doi:* 10.1016/j.nmd.2010.12.007.

Selcen D. (2015). Severe congenital actin related myopathy with myofibrillar myopathy features. *Neuromuscular Disorders*, *25*(6), 488-492. *doi:* 10.1016/j.nmd.2015.04.002.

Senderek, J., Garvey, S. M., Krieger, M., Guergueltcheva, V., Urtizberea, A., & Roos, A., et al. (2009). Autosomal-Dominant Distal Myopathy Associated with a Recurrent Missense Mutation in the Gene Encoding the Nuclear Matrix Protein, Matrin 3. *The American Journal of Human Genetics, 84*(4), 511-518. *doi:* 10.1016/j.ajhg.2009.03.006.

Servián-Morilla, E., Takeuchi, H., Lee, T. V., Clarimon, J., Mavillard, F., & Area-Gómez, E., et al. (2016). A POGLUT1 mutation causes a muscular dystrophy with reduced Notch signaling and satellite cell loss. *EMBO Molecular Medicine, 8*(11), 1289-1309. *doi:*10.15252/emmm.201505815.

Sewry, C. A., Philpot, J., Sorokin, L. M., Wilson, L.A., Naom, I., & Goodwin, F., et al. (1996). Diagnosis of merosin (laminin-2) deficient congenital muscular dystrophy by skin biopsy. *Lancet*, *347*(9001), 582-584.

Sewry, C. A., & Wallgren-Pettersson, C. (2017). Myopathology in congenital myopathies. *Neuropathology and Applied Neurobiology, 43*(1), 5-23. *doi:*10.1111/nan.12369.

Sframeli, M., Sarkozy, A., Bertoli, M., Astrea, G., Hudson, J., & Scoto, M., et al. (2017). Congenital muscular dystrophies in the UK population: Clinical and molecular spectrum of a large cohort diagnosed over a 12-year period. *Neuromuscular Disorders*, *27*(9), 793-803. *doi:* 10.1016/j.nmd.2017.06.008.

Shalaby, S., Hayashi, Y. K., Goto, K., Ogawa, M., Nonaka, I., Noguchi, S., & Nishino, I. (2008). Rigid spine syndrome caused by a novel mutation in four-and-a-half LIM domain 1 gene (FHL1). *Neuromuscular Disorders, 18*(12), 959-961. *doi:* 10.1016/j.nmd.2008.09.012.

Shamseldin, H. E., Tulbah, M., Kurdi, W., Nemer, M., Alsahan, N., & Al Mardawi, E., et al. (2015). Identification of embryonic lethal genes in humans by autozygosity

mapping and exome sequencing in consanguineous families. *Genome Biology*, *16*(1), 116. *doi*: 10.1186/s13059-015-0681-6.

Shapiro, M. B., & Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research*, 15(*17*), 7155-7174.

Sher, R. B., Aoyama, C., Huebsch, K. A., Ji, S., Kerner, J., & Yang, Y., et al. (2016). A rostrocaudal muscular dystrophy caused by a defect in choline kinase beta, the first enzyme in phosphatidylcholine biosynthesis. *The Journal of Biological Chemistry*, *281*(8), 4938-4948. *doi:* 10.1074/jbc.M512578200.

Shieh, P. B. (2013). Muscular dystrophies and other genetic myopathies. *Neurologic Clinics*, *31*(4), 1009-1029. *doi:* 10.1016/j.ncl.2013.04.004.

Shimizu-Motohashi, Y., Murakami, T., Kimura, E., Komaki, H., & Watanabe, N. (2018). Exon skipping for Duchenne muscular dystrophy: A systematic review and meta-analysis. *Orphanet Journal of Rare Diseases, 13*(1). *doi:* 10.1186/s13023-018-0834-2.

Shin, S., Kim, Y., Oh, S. C., Yu, N., Lee, S., Choi, J. R., & Lee, K. (2017). Validation and optimization of the Ion Torrent S5 XL sequencer and Oncomine workflow for BRCA1 and BRCA2 genetic testing. *Oncotarget*, 8(21). *doi:* 10.18632/oncotarget.16799.

Shorer, Z., Philpot, J., Muntoni, F., Sewry, C., & Dubowitz, V. (1995). Demyelinating peripheral neuropathy in merosin-deficient congenital muscular dystrophy. *Journal of Child Neurology*, *10*(6), 472-475. *doi:* 10.1177/088307389501000610.

Simbolo, M., Gottardi, M., Corbo, V., Fassan, M., Mafficini, A., Malpeli, G., et al. (2013). DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLoS ONE*, *8*(6). *doi:* 10.1371/journal.pone.0062692.

Singleton, M., Guthery, S., Voelkerding, K., Chen, K., Kennedy, B., Margraf, R., et al. (2014). Phevor Combines Multiple Biomedical Ontologies for Accurate Identification of Disease-Causing Alleles in Single Individuals and Small Nuclear Families. *The American Journal of Human Genetics*, *94*(4), 599-610. *doi:* 10.1016/j.ajhg.2014.03.010.

Sirmaci, A., Edwards, Y. J., Akay, H., & Tekin, M. et al. (2012). Challenges in Whole Exome Sequencing: An Example from Hereditary Deafness. *PLoS One*, *7*(2), e32000. *doi:* 10.1371/journal.pone.0032000.

Siva, N. (2008). 1000 Genomes project. *Nature Biotechnology*, *26*(3), 256-256. *doi:* 10.1038/nbt0308-256b.

Smith, L. L., Gupta, V. A., & Beggs, A. H. (2014). Bridging integrator 1 (Bin1) deficiency in zebrafish results in centronuclear myopathy. *Human Molecular Genetics, 23*(13), 3566-3578. *doi:* 10.1093/hmg/ddu067.

Snoeck, M., van Engelen, B. G., Küsters, B., Lammens, M., Meijer, R., & Molenaar, J. P., et al. (2015). RYR1-related myopathies: a wide spectrum of phenotypes throughout life. *European Journal of Neurology*, *22*(7), 1094-1112. *doi:* 10.1111/ene.12713.

Sobel, E., Papp, J.C., & Lange, K. (2002). Detection and Integration of Genotyping Errors in Statistical Genetics. *American Journal of Human Genetics*, *70*(2), 496–508. *doi:* 10.1086/338920.

Solyom, S., Ewing, A. D., Hancks, D. C., Takeshima, Y., Awano, H., & Matsuo, M., et al. (2012). Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. *Human Mutation*, *33*(2), 369-371. *doi:* 10.1002/humu.21663.

Song, L., Huang, W., Kang, J., Huang, Y., Ren, H., & Ding, K. (2017). Comparison of error correction algorithms for Ion Torrent PGM data: Application to hepatitis B virus. *Scientific Reports*, *7*(1). *doi:* 10.1038/s41598-017-08139-y.

Sparks, S. (2012). Congenital protein hypoglycosylation diseases. *The Application of Clinical Genetics,* 5, 43-54. *doi:* 10.2147/tacg.s18673.

Speer, M. C., Tandan, R., Rao, P. N., Fries, T., Stajich, J. M., & Bolhuis, P. A., et al. (1996). Evidence for locus heterogeneity in the Bethlem myopathy and linkage to 2q37. *Human Molecular Genetics, 5*(7), 1043-1046. *doi:* 10.1093/hmg/5.7.1043.

Srinuandee, P., & Satirapod, C. (2015). Use of genetic algorithm and sliding windows for optimising ambiguity fixing rate in GPS kinematic positioning mode. *Survey Review*, *47*(340), 1–6. *doi:* 10.1179/1752270614Y.0000000088.

Stevens, E., Carss, K., Cirak, S., Foley, A., Torelli, S., & Willer, T., et al. (2013). Mutations in B3GALNT2 Cause Congenital Muscular Dystrophy and Hypoglycosylation of α-Dystroglycan. *The American Journal of Human Genetics, 92*(3), 354-365. *doi:* 10.1016/j.ajhg.2013.01.016.

Stoneking, M., & Krause, J. (2011). Learning about human population history from ancient and modern genomes. *Nature Reviews Genetics*, *12*(9), 603-614. *doi:* 10.1038/nrg3029.

Straub, V., Murphy, A., & Udd, B. (2018). 229th ENMC international workshop: Limb girdle muscular dystrophies – Nomenclature and reformed classification Naarden, the Netherlands, 17–19 March 2017. *Neuromuscular Disorders, 28*(8), 702-710. *doi:* 10.1016/j.nmd.2018.05.007.

Straussberg, R., Schottmann, G., Sadeh, M., Gill, E., Seifert, F., Halevy, A., et al. (2016). Kyphoscoliosis peptidase (KY) mutation causes a novel congenital myopathy with core targetoid defects. *Acta Neuropathologica, 132*(3), 475-478. *doi:* 10.1007/s00401-016-1602-9.

Stump, C. S., Henriksen, E. J., Wei, Y., & Sowers, J. R. (2006). The metabolic syndrome: role of skeletal muscle metabolism. *Annals of Medicine*, *38*(6), 389-402. *doi:* 10.1080/07853890600888413.

Suarez, N. A., Macia, A., & Muotri, A. R. (2018). LINE-1 retrotransposons in healthy and diseased human brain. *Development Neurobiology*, *78*(5), 434-455. doi: 10.1002/dneu.22567.

Sun, H., Li, N., Wang, X., Chen, T., Shi, L., Zhang, L., et al. (2005). Molecular cloning and characterization of a novel muscle adenylosuccinate synthetase, AdSSL1, from human bone marrow stromal cells. *Molecular and Cellular Biochemistry, 269*(1), 85-94. *doi:* 10.1007/s11010-005-2539-9.

Sunada, Y., Edgar, T. S., Lotz, B. P., Rust, R. S., & Campbell, K. P. (1995). Merosin-negative congenital muscular dystrophy associated with extensive brain abnormalities. *Neurology*, *45*(11), 2084-2089.

Sunada, Y., Bernier, S. M., Kozak, C. A., Yamada, Y., & Campbell, K. P. (1994). Deficiency of merosin in dystrophic dy mice and genetic linkage of laminin M chain gene to dy locus. *Journal of Biological Chemistry*, *269*(19), 13729-13732.

Sutton, I. J., Winer, J. B., Norman, A. N., Liechti-Gallati, S., & MacDonald, F. (2001). Limb girdle and facial weakness in female carriers of X-linked myotubular myopathy mutations. *Neurology*, *57*(5), 900-902. *doi:* 10.1212/WNL.57.5.900.

Sveen, M., Schwartz, M., & Vissing, J. (2006). High prevalence and phenotype-genotype correlations of limb girdle muscular dystrophy type 2I in Denmark. *Annals of Neurology, 59*(5), 808-815. *doi:* 10.1002/ana.20824.

Tachi, N., Kozuka, N., Chiba, S., Miyaji, M., & Watanabe, I. (2001). A double mutation in a patient with X-linked myotubular myopathy. *Pediatric Neurology*, *24*(4), 297-299.

Tajsharghi, H., Leren, T., Abdul-Hussein, S., Tulinius, M., Brunvand, L., Dahl, H., & Oldfors, A. (2009). Unexpected myopathy associated with a mutation in MYBPC3 and misplacement of the cardiac myosin binding protein C. *Journal of Medical Genetics*, *47*(8): 575-577. *doi*: 10.1136/jmg.2009.072710.

Takayama, K., Mitsuhashi, S., Shin, J., Tanaka, R., Fujii, T., & Tsuburaya, R., et al. (2016). Japanese multiple epidermal growth factor 10 (MEGF10) myopathy with novel mutations: A phenotype–genotype correlation. *Neuromuscular Disorders, 26*(9), 604-609. *doi:* 10.1016/j.nmd.2016.06.005.

Tan, E., Topaloglu, H., Sewry, C., Zorlu, Y., Naom, I., & Erdem, S., et al. (1997). Late onset muscular dystrophy with cerebral white matter changes due to partial merosin deficiency. *Neuromuscular Disorders*, *7*(2), 85-89.

Taniguchi-Ikeda, M., Morioka, I., Iijima, K., & Toda, T. (2016). Mechanistic aspects of the formation of α-dystroglycan and therapeutic research for the treatment of α-

dystroglycanopathy: A review. *Molecular Aspects of Medicine, 51*, 115-124. *doi:* 10.1016/j.mam.2016.07.003.

Tanner, S. M., Orstavik, K. H., Kristiansen, M., Lev, D., Lerman-Sagie, T., & Sadeh, M., et al. (1999). Skewed X-inactivation in a manifesting carrier of X-linked myotubular myopathy and in her non-manifesting carrier mother. *Human Genetics, 104*(3), 249-253. *doi:* 10.1007/s004390050943.

Tarnopolsky, M., Hoffman, E., Giri, M., Shoffner, J., & Brady, L. (2015). Alpha-sarcoglycanopathy presenting as exercise intolerance and rhabdomyolysis in two adults. *Neuromuscular Disorders, 25*(12), 952-954. *doi:* 10.1016/j.nmd.2015.09.010.

Tasca, G., Fattori, F., Ricci, E., Monforte, M., Rizzo, V., & Mercuri, E., et al. (2013a). Somatic mosaicism in TPM2-related myopathy with nemaline rods and cap structures. *Acta Neuropathologica, 125*(1):169-171. *doi:* 10.1007/s00401-012-1049-6.

Tasca, G., Moro, F., Aiello, C., Cassandrini, D., Fiorillo, C., & Bertini, E., et al. (2013b). Limb-girdle muscular dystrophy with α-dystroglycan deficiency and mutations in the ISPD gene. *Neurology, 80*(10), 963-965. *doi:* 10.1212/wnl.0b013e3182840cbc.

Tateyama, M., Aoki, M., Nishino, I., Hayashi, Y., Sekiguchi, S., & Shiga, Y., et al. (2002). Mutation in the caveolin-3 gene causes a peculiar form of distal myopathy. *Neurology, 58*(2), 323-325. doi:10.1212/wnl.58.2.323.

Tawil, R., Kissel, J. T., Heatwole, C., Pandya, S., Gronseth, G., & Benatar, M. (2015). Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy. *Neurology, 85*(4), 357-364. *doi:* 10.1212/wnl.0000000000001783.

Tawil, R. (2018). Facioscapulohumeral muscular dystrophy. *Neurogenetics, Part II Handbook of Clinical Neurology,* 541-548. *doi:*10.1016/b978-0-444-64076-5.00035-1.

Thompson, J. F., & Steinmann, K. E. (2010). Single Molecule Sequencing with a HeliScope Genetic Analysis System. *Current Protocols in Molecular Biology*. Chapter 7: Unit 7.10. *doi:* 10.1002/0471142727.mb0710s92.

Thompson, R., & Straub, V. (2016). Limb-girdle muscular dystrophies — international collaborations for translational research. *Nature Reviews Neurology, 12*(5), 294-309. *doi:* 10.1038/nrneurol.2016.35.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2012). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics, 14*(2), 178-192. *doi:* 10.1093/bib/bbs017.

Tian, S., Yan, H., Kalmbach, M., & Slager, S. L. (2016). Impact of post-alignment processing in variant discovery from whole exome data. *BMC Bioinformatics*, 17(1). *doi:* 10.1186/s12859-016-1279-z.

Tica, J., Lee, E., Untergasser, A., Meiers, S., Garfield, D. A., & Gokcumen, O., et al. (2016). Next-generation sequencing-based detection of germline L1-mediated transductions. *BMC Genomics*, *17*, 342. *doi:* 10.1186/s12864-016-2670-x.

Tinsley, J. M., & Davies, K. E. (1993). Utrophin: A potential replacement for dystrophin? *Neuromuscular Disorders, 3*(5-6), 537-539. *doi:* 10.1016/0960-8966(93)90111-v.

Toda, T., Chiyonobu, T., Xiong, H., Tachikawa, M., Kobayashi, K., & Manya, H., et al. (2005). Fukutin and alpha-dystroglycanopathies. *Acta Myologica*, *24*(2), 60-63.

Todd, E. J., Yau K. S., Ong, R., Slee, J., McGillivray, G., & Barnett, C.P., et al. (2015). Next generation sequencing in a large cohort of patients presenting with neuromuscular disease before or at birth. *Orphanet Journal of Rare Diseases*, *10*, 148. *doi*: 10.1186/s13023-015-0364-0.

Tomé, F. M., Evangelista, T., Leclerc, A., Sunada, Y., Manole, E., & Estournet, B., et al. (1994). Congenital muscular dystrophy with merosin deficiency. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, 317(4), 351-357.

Tonino, P., Kiss, B., Strom, J., Methawasin, M., Smith, J. E., & Kolb, J., et al. (2017). The giant protein titin regulates the length of the striated muscle thick filament. *Nature Communications, 8*(1), 1041. *doi*:10.1038/s41467-017-01144-9.

Torella, A., Fanin, M., Mutarelli, M., Peterle, E., Blanco, F. D., & Rispoli, R., et al. (2013). Next-Generation Sequencing Identifies Transportin 3 as the Causative Gene for LGMD1F. *PLoS ONE, 8*(5). *doi:* 10.1371/journal.pone.0063536.

Torices, S., Alvarez-Rodríguez, L., Grande, L., Varela, I., Muñoz, P., & Pascual, D., et al. (2015). A Truncated Variant of ASCC1, a Novel Inhibitor of NF-κB, Is Associated with Disease Severity in Patients with Rheumatoid Arthritis. *Journal of Immunology*, *195*(11), 5415-5420. *doi:* 10.4049/jimmunol.150153.

Tosch, V., Rohde, H. M., Tronchère, H., Zanoteli, E., Monroy, N., & Kretz, C., et al. (2006). A novel PtdIns3P and PtdIns(3,5)P2 phosphatase with an inactivating variant in centronuclear myopathy. *Human Molecular Genetics*, *15*(21), 3098-3106. *doi*: 10.1093/hmg/ddl250.

Treves S, Anderson AA, Ducreux S, Divet A, Bleunven C, & Grasso C, et al. (2005). Ryanodine receptor 1 mutations, dysregulation of calcium homeostasis and neuromuscular disorders. *Neuromuscular disorders*, *15*(9-10), 577-587. *doi*: 10.1016/j.nmd.2005.06.008.

Treves, S., Jungbluth, H., Muntoni, F., & Zorzato, F. (2008). Congenital muscle disorders with cores: the ryanodine receptor calcium channel paradigm. *Current opinion in pharmacology*, *8*(3), 319-326. *doi:* 10.1016/j.coph.2008.01.005.

Trump, N., Cullup, T., Verheij, J. B., Manzur, A., Muntoni, F., & Abbs, S. X-linked myotubular myopathy due to a complex rearrangement involving exon 10 of the

myotubularin (MTM1) gene. *Neuromuscular Disorders*, *22*(5), 384-388. *doi*: 10.1016/j.nmd.2011.11.004.

Tsai, T. C., Horinouchi, H., Noguchi, S., Minami, N., Murayama, K., & Hayashi, Y. K., et al. (2005). Characterization of MTM1 mutations in 31 Japanese families with myotubular myopathy, including a patient carrying 240 kb deletion in Xq28 without male hypogenitalism. *Neuromuscul Disorders*, *15*(3), 245-252. *doi:* 10.1016/j.nmd.2004.12.005.

Tsao, C. Y., Mendell, J. R., Rusin, J. & Luquette, M. (1998). Congenital muscular dystrophy with complete laminin-alpha2-deficiency, cortical dysplasia, and cerebral white-matter changes in children. *Journal of Child Neurology*, *13*(6), 253-256. *doi*: 10.1177/088307389801300602.

Tskhovrebova, L., & Trinick, J. (2017). Titin and Nebulin in Thick and Thin Filament Length Regulation. *Subcellular Biochemistry*, *82*, 285-318. *doi:* 10.1007/978-3-319-49674-0_10.

Tsujita, K., Itoh, T., Ijuin, T., Yamamoto, A., Shisheva, A., & Laporte, J., et al. (2004). Myotubularin regulates the function of the late endosome through the gram domain-phosphatidylinositol 3,5-bisphosphate interaction. *The Journal of Biological Chemistry*, *279*(14), 13817-13824. *doi:* 10.1074/jbc.M312294200.

Tuffery-Giraud, S., Miro, J., Koenig, M., & Claustres, M. (2017). Normal and altered pre-mRNA processing in the DMD gene. *Human Genetics*, *136*(9), 1155-1172. *doi*:10.1007/s00439-017-1820-9.

Tyler, K. L. (2003). Origins and early descriptions of "Duchenne muscular dystrophy". *Muscle & Nerve, 28*(4), 402-422. *doi:* 10.1002/mus.10435.

Uruha, A., Hayashi, Y. K., Oya, Y., Mori-Yoshimura, M., Kanai, M., & Murata, M., et al. (2015). Necklace cytoplasmic bodies in hereditary myopathy with early respiratory failure. *Journal of Neurology, Neurosurgery, and Psychiatry*, *86*(5), 483-489. *doi:* 10.1136/jnnp-2014-309009.

Valencia, C.A., Ankala, A., Rhodenizer, D., Bhide, S., Littlejohn, M. R., & Keong, L. M., et al. (2013). Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS One*, *8*(1), e53083. *doi:* 10.1371/journal.pone.0053083.

van Essen, A. J., Abbs, S., Baiget, M., Bakker, E., Boileau, C., & van Broeckhoven, C., et al. (1992). Parental origin and germline mosaicism of deletions and duplications of the dystrophin gene: a European study. *Human Genetics*, *88*(3), 249-257.

van Nistelrooij, A. M., Dinjens, W. N., Wagner, A., Spaander, M. C., van Lanschot, J. J., & Wijnhoven, B.P. (2014). Hereditary Factors in Esophageal Adenocarcinoma. *Gastrointestinal Tumors*, *1*(2), 93-98. *doi:* 10.1159/000362575.

van Reeuwijk, J., Janssen, M., van den Elzen, C., Beltran-Valero de Bernabé, D., Sabatelli, P., & Merlini, L., et al. (2005). POMT2 mutations cause alpha-dystroglycan hypoglycosylation and Walker-Warburg syndrome. *Journal of Medical Genetics, 42*(12), 907-912. *doi:* 10.1136/jmg.2005.031963.

van Reeuwijk, J., Grewal, P. K., Salih, M. A., Beltrán-Valero de Bernabé, D., McLaughlan, J. M., & Michielse, C. B., et al. (2007). Intragenic deletion in the LARGE gene causes Walker-Warburg syndrome. *Human Genetics*, *121*(6), 685-690. *doi:* 10.1007/s00439-007-0362-y.

Vanegas, O. C., Bertini, E., Zhang, R., Petrini, S., Minosse, C., & Sabatelli, P., et al. (2001). Ullrich scleroatonic muscular dystrophy is caused by recessive mutations in collagen type VI. *Proceedings of the National Academy of Sciences, 98*(13), 7516-7521. *doi:* 10.1073/pnas.121027598.

Vasli, N., Böhm, J., Le Gras, S., Muller, J., Pizot, C., & Jost, B., et al. (2012). Next generation sequencing for molecular diagnosis of neuromuscular diseases. *Acta Neuropathologica*, *124*(2), 273-283. *doi:* 10.1007/s00401-012-0982-8.

Vasli, N., & Laporte, J. (2013). Impacts of massively parallel sequencing for genetic diagnosis of neuromuscular disorders. *Acta Neuropathologica*, *125*(2), 173-185. *doi:* 10.1007/s00401-012-1072-7.

Vasli, N., Harris, E., Karamchandani, J., Bareke, E., Majewski, J., & Romero, N. B., et al. (2017). Recessive mutations in the kinase ZAK cause a congenital myopathy with fibre type disproportion. *Brain*, *140*(1), 37-48. *doi:* 10.1093/brain/aww257.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., & Sutton, G. G., et al. (2001). The Sequence of the Human Genome. *Science*, *291*(5507), 1304-1351. *doi:* 10.1126/science.1058040.

Vieira, E., Gonçalves, A., Maia, N., Oliveira, M. E., Oliveira, J., Bronze-da-Rocha, E., & Santos, R. (2012). Distribution of limb girdle muscular dystrophy subtypes among Portuguese patients. Oral presentation at the 2012 Portuguese Society of Human Genetics. Porto. Portugal. Retrieved from: http://repositorio.insa.pt/bitstream/10400.18/1224/1/LGMD_abstract_EV-2012.pdf.

Vieira, N. M., Naslavsky, M. S., Licinio, L., Kok, F., Schlesinger, D., & Vainzof, M., et al. (2014). A defect in the RNA-processing protein HNRPDL causes limb-girdle muscular dystrophy 1G (LGMD1G). *Human Molecular Genetics, 23*(15), 4103-4110. *doi:* 10.1093/hmg/ddu127.

Vigliano, P., Dassi, P., Di Blasi, C., Mora, M., & Jarre, L. (2009). LAMA2 stop-codon mutation: merosin-deficient congenital muscular dystrophy with occipital polymicrogyria, epilepsy and psychomotor regression. *European Journal of Paediatric Neurology*, *13*(1), 72-76. *doi:* 10.1016/j.ejpn.2008.01.010.

Vissing, J., Barresi, R., Witting, N., Ghelue, M. V., Gammelgaard, L., & Bindoff, L. A., et al. (2016). A heterozygous 21-bp deletion inCAPN3causes dominantly inherited limb girdle muscular dystrophy. *Brain, 139*(8), 2154-2163. *doi:* 10.1093/brain/aww133.

Vihinen, M., den Dunnen, J. T., Dalgleish, R. & Cotton, R. G. (2012). Guidelines for establishing locus specific databases. *Human Mutation*, *33*(2), 298-305. *doi*: 10.1002/humu.21646.

Villanova, M., Malandrini, A., Sabatelli, P., Sewry, C. A., Toti, P., & Torelli, S., et al. (1997). Localization of laminin alpha 2 chain in normal human central nervous system: an immunofluorescence and ultrastructural study. *Acta Neuropathologica*, *94*(6), 567-571. *doi:* 10.1007/s004010050751.

Vorgerd, M., Ven, P. F., Bruchertseifer, V., Löwe, T., Kley, R. A., & Schröder, R., et al. (2005). A Mutation in the Dimerization Domain of Filamin C Causes a Novel Type of Autosomal Dominant Myofibrillar Myopathy. *The American Journal of Human Genetics, 77*(2), 297-304. *doi:* 10.1086/431959.

Vuillaumier-Barrot, S., Bouchet-Séraphin, C., Chelbi, M., Devisme, L., Quentin, S., & Gazal, S., et al. (2012). Identification of Mutations in TMEM5 and ISPD as a Cause of Severe Cobblestone Lissencephaly. *The American Journal of Human Genetics, 91*(6), 1135-1143. *doi:* 10.1016/j.ajhg.2012.10.009.

Wallgren-Pettersson, C., Lehtokari, V., Kalimo, H., Paetau, A., Nuutinen, E., & Hackman, P., et al. (2007). Distal myopathy caused by homozygous missense mutations in the nebulin gene. *Brain, 130*(6), 1465-1476. doi:10.1093/brain/awm094.

Wallgren-Pettersson, C., Sewry, C. A., Nowak, K. J. & Laing, N. G. (2011). Nemaline myopathies. *Seminars in Pediatric Neurology*, *18*(4), 230-238. *doi:* 10.1016/j.spen.2011.10.004.

Walton, J. N., & Nattrass, F. J. (1954). On the classification, natural history and treatment of the myopathies. *Brain*, 77(2):169-231.

Wang, J., Wu, P., Shi, Z., Xu, Y., & Liu, Z. (2017). The AAV-mediated and RNA-guided CRISPR/Cas9 system for gene therapy of DMD and BMD. *Brain and Development, 39*(7), 547-556. *doi:* 10.1016/j.braindev.2017.03.024.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16). *doi:* 10.1093/nar/gkq603.

Wang, Y., Wu, N., Liu, J., Wu, Z., & Dong, D. (2015). FusionCancer: A database of cancer fusion genes derived from RNA-seq data. *Diagnostic Pathology*, *10*(1). *doi:* 10.1186/s13000-015-0310-4.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57-63. *doi:* 10.1038/nrg2484.

Wheeler, M. A., Davies, J. D., Zhang, Q., Emerson, L. J., Hunt, J., Shanahan, C. M., & Ellis, J. A. (2007). Distinct functional domains in nesprin-1α and nesprin-2β bind directly to emerin and both interactions are disrupted in X-linked Emery–Dreifuss muscular dystrophy. *Experimental Cell Research, 313*(13), 2845-2857. *doi:* 10.1016/j.yexcr.2007.03.025.

Wiessner, M., Roos, A., Munn, C. J., Viswanathan, R., Whyte, T., & Cox, D., et al. (2017). Mutations in INPP5K, Encoding a Phosphoinositide 5-Phosphatase, Cause Congenital Muscular Dystrophy with Cataracts and Mild Cognitive Impairment. *The American Journal of Human Genetics, 100*(3), 523-536. *doi:* 10.1016/j.ajhg.2017.01.024.

Wicklund, M. P., & Kissel, J. T. (2014). The Limb-Girdle Muscular Dystrophies. *Neurologic Clinics, 32*(3), 729-749. *doi:* 10.1016/j.ncl.2014.04.005.

Wijmenga, C., Padberg, G. W., Moerer, P., Wiegant, J., Liem, L., & Brouwer, O. F., et al. (1991). Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics, 9*(4), 570-575. *doi:* 10.1016/0888-7543(91)90348-i.

Wijmenga, C., Hewitt, J. E., Sandkuijl, L. A., Clark, L. N., Wright, T. J., & Dauwerse, H. G., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nature Genetics, 2*(1), 26-30. *doi:* 10.1038/ng0992-26.

Willer, T., Lee, H., Lommel, M., Yoshida-Moriguchi, T., Bernabe, D. B., & Venzke, D., et al. (2012). ISPD loss-of-function mutations disrupt dystroglycan O-mannosylation and cause Walker-Warburg syndrome. *Nature Genetics, 44*(5), 575-580. *doi:* 10.1038/ng.2252.

Wilmshurst, J. M., Lillis, S., Zhou, H., Pillay, K., Henderson, H., & Kress, W., et al. (2010). RYR1 mutations are a common cause of congenital myopathies with central nuclei. *Annals of neurology, 68*(5), 717-276. *doi*: 10.1002/ana.22119.

Wood, A. J. & Currie, P. D. (2014). Analysing regenerative potential in zebrafish models of congenital muscular dystrophy. *The International Journal of Biochemistry & Cell Biology, 56*, 30-37. *doi:* 10.1016/j.biocel.2014.10.021.

Worman, H. J., & Dauer, W. T. (2014). The Nuclear Envelope: An Intriguing Focal Point for Neurogenetic Disease. *Neurotherapeutics, 11*(4), 764-772. *doi:* 10.1007/s13311-014-0296-8.

Wortmann, S. B., Koolen, D. A., Smeitink, J. A., Heuvel, L. V., & Rodenburg, R. J. (2015). Whole exome sequencing of suspected mitochondrial patients in clinical practice. *Journal of Inherited Metabolic Disease, 38*(3), 437-443. *doi:* 10.1007/s10545-015-9823-y.

Wu, S., Ibarra, M. C., Malicdan, M. C., Murayama, K., Ichihara, Y., & Kikuchi, H., et al. (2006). Central core disease is due to RYR1 mutations in more than 90% of patients. *Brain: a journal of neurology*, *129*(Pt 6), 1470-1480. *doi:* 10.1093/brain/awl077.

Xiong, H., Tan, D., Wang, S., Song, S., Yang, H., & Gao, K., et al. (2015). Genotype/phenotype analysis in Chinese laminin-α2 deficient congenital muscular dystrophy patients. *Clinical Genetics*, *87*(3), 233-243. *doi:* 10.1111/cge.12366.

Xue, Y., Ankala, A., Wilcox, W. R., & Hegde, M. R. (2015). Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: Single-gene, gene panel, or exome/genome sequencing. *Genetics in Medicine*, *17*(6), 444-451. *doi:* 10.1038/gim.2014.122.

Yan, Z., Bai, X. C., Yan, C., Wu, J., Li, Z., & Xie, T., et al. (2015). Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution. *Nature*, *517*(7532), 50-55. *doi*: 10.1038/nature14063.

Yang, A. C., Ng, B. G., Moore, S. A., Rush, J., Waechter, C. J., & Raymond, K. M., et al. (2013). Congenital disorder of glycosylation due to DPM1 mutations presenting with dystroglycanopathy-type congenital muscular dystrophy. *Molecular Genetics and Metabolism, 110*(3), 345-351. *doi:* 10.1016/j.ymgme.2013.06.016.

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols*, *10*(10), 1556-1566. *doi:* 10.1038/nprot.2015.105.

Yang, Y., Sebra, R., Pullman, B. S., Qiao, W., Peter, I., & Desnick, R. J., et al. (2015). Quantitative and multiplexed DNA methylation analysis using long-read single-molecule real-time bisulfite sequencing (SMRT-BS). *BMC Genomics*, *16*(1). *doi:* 10.1186/s12864-015-1572-7.

Ye, K., Guo, L., Yang, X., Lamijer, E., Raine, K., & Ning, Z. (2018). Split-Read Indel and Structural Variant Calling Using PINDEL. *Methods in Molecular Biology Copy Number Variants*, 95-105. *doi:* 10.1007/978-1-4939-8666-8_7.

Yeo, G. & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology*, *11*(2-3), 377-394. *doi:* 10.1089/1066527041410418.

Yilmaz, S., & Singh, A. K. (2012). Single cell genome sequencing. *Current Opinion in Biotechnology*, 23(3), 437-443. *doi:* 10.1016/j.copbio.2011.11.018.

Yis, U., Uyanik, G., Heck, P. B., Smitka, M., Nobel, H., & Ebinger, F., et al. (2011). Fukutin mutations in non-Japanese patients with congenital muscular dystrophy: Less severe mutations predominate in patients with a non-Walker-Warburg phenotype. *Neuromuscular Disorders, 21*(1), 20-30. *doi:* 10.1016/j.nmd.2010.08.007.

Yoshida, K., Nakamura, A., Yazaki, M., Ikeda, S., & Takeda, S. (1998). Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. *Human Molecular Genetics*, *7*(7), 1129-1132.

Yoshida, A., Kobayashi, K., Manya, H., Taniguchi, K., Kano, H., & Mizuno, M., et al. (2001). Muscular Dystrophy and Neuronal Migration Disorder Caused by Mutations in a Glycosyltransferase, POMGnT1. *Developmental Cell, 1*(5), 717-724. *doi:* 10.1016/s1534-5807(01)00070-3.

Young, H. K., Barton, B. A., Waisbren, S., Dale, L. P., Ryan, M. M., Webster, R. I., & North, K. N. (2007). Cognitive and Psychological Profile of Males with Becker Muscular Dystrophy. *Journal of Child Neurology, 23*(2), 155-162. *doi:* 10.1177/0883073807307975.

Yu, S., Manson, J., White, S., Bourne, A., Waddy, H., & Davis, M., et al. (2003). X-linked myotubular myopathy in a family with three adult survivors. Clinical Genetics, 64(2), 148-152. *doi:* 10.1034/j.1399-0004.2003.00118.x.

Yuen, M., Sandaradura, S. A., Dowling, J. J., Kostyukova, A. S., Moroz, N., & Quinlan, K. G., et al. (2014). Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *The Journal of Clinical Investigations*, *124*(1), 4693-4708. *doi*: 10.1172/ JCI80057.

Yurchenco, P. D. (2015). Integrating Activities of Laminins that Drive Basement Membrane Assembly and Function. *Current Topic in Membranes*, 76, 1-30. doi: 10.1016/bs.ctm.2015.05.001.

Zemojtel, T., Köhler, S., Mackenroth, L., Jäger, M., Hecht, J., & Krawitz, P., et al. (2014). Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Science Translational Medicine, 6(252). doi: 10.1126/scitranslmed.3009262.

Zhang, Y., Chen, H. S., Khanna, V. K., Leon, S. D., Phillips, M. S., & Schappert, K., et al. (1993). A mutation in the human ryanodine receptor gene associated with central core disease. *Nature Genetics, 5*(1), 46-50. *doi:*10.1038/ng0993-46.

Zhang, X., Vuolteenaho, R., & Tryggvason, K. (1996). Structure of the human laminin alpha2-chain gene (LAMA2), which is affected in congenital muscular dystrophy. *The Journal of Biological Chemistry*, *271*(44), 27664-27669.

Zhang, Q., Bethmann, C., Worth, N. F., Davies, J. D., Wasner, C., & Feuer, A., et al. (2007). Nesprin-1 and -2 are involved in the pathogenesis of Emery–Dreifuss muscular dystrophy and are critical for nuclear envelope integrity. *Human Molecular Genetics, 16*(23), 2816-2833. *doi:* 10.1093/hmg/ddm238.

Zhang, Y., & Jeltsch, A. (2010). The Application of Next Generation Sequencing in DNA Methylation Analysis. *Genes*, *1*(1), 85-101. *doi:* 10.3390/genes1010085.

Zhao, M., Wang, Q., Wang, Q., Jia, P., & Zhao, Z. (2013). Computational tools for copy number variation (CNV) detection using next-generation sequencing data: Features and perspectives. *BMC Bioinformatics*, *14*(Suppl 11). *doi:* 10.1186/1471-2105-14-s11-s1.

Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D., & Perou, C. M. (2014). Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*, *15*(1), 419. *doi:* 10.1186/1471-2164-15-419.

Zhao, M., Kim, J. R., van Bruggen, R., & Park, J. (2018). RNA-Binding Proteins in Amyotrophic Lateral Sclerosis. *Molecules and Cells, 41*(9), 818-829. *doi:* 10.14348/molcells.2018.0243.

Zhou, H., Jungbluth, H., Sewry, C. A., Feng, L., Bertini, E., & Bushby, K., et al. (2007). Molecular mechanisms and phenotypic variation in RYR1-related congenital myopathies*. Brain: a journal of neurology*, *130*(Pt 8), 2024-2036. doi: 10.1093/brain/awm096.

Zhou, H., Rokach, O., Feng, L., Munteanu, I., Mamchaoui, K., & Wilmshurst, J. M., et al. (2013). RyR1 Deficiency in Congenital Myopathies Disrupts Excitation-Contraction Coupling. *Human Mutation, 34*(7), 986-996. *doi:* 10.1002/humu.22326.

Zhou, Q., Su, X., Wang, A., Xu, J., & Ning, K. (2013). QC-Chain: Fast and Holistic Quality Control Method for Next-Generation Sequencing Data. *PLoS ONE*, 8(4). *doi:* 10.1371/journal.pone.0060234.

Ziat, E., Mamchaoui, K., Beuvin, M., Nelson, I., Azibani, F., & Spuler, S., et al. (2016). FHL1B Interacts with Lamin A/C and Emerin at the Nuclear Lamina and is Misregulated in Emery-Dreifuss Muscular Dystrophy. *Journal of Neuromuscular Diseases, 3*(4), 497-510. *doi:* 10.3233/jnd-160169.

Zou, Y., Zwolanek, D., Izu, Y., Gandhy, S., Schreiber, G., & Brockmann, K., et al. (2013). Recessive and dominant mutations in COL12A1 cause a novel EDS/myopathy overlap syndrome in humans and mice. *Human Molecular Genetics, 23*(9), 2339-2352. *doi:* 10.1093/hmg/ddt627.

Appendices

Appendix I.1

Contents

Table I.1.1 - List of genes included in the Ion AmpliSeq Neurological Research Panel.

Gene Symbol	Full gene name	Locus
AAAS	Aladin	12q13
AARS	Alanyl-tRNA synthetase	16q22
AARS2	Alanyl-tRNA synthetase 2	6p21.1
ABAT	4-aminobutyrate aminotransferase	16p13.3
ABCD1	ATP-binding cassette, subfamily D, member 1	Xq28
ABHD12	Abhydrolase domain-containing protein 12	20p11.21
ACAD9	Acyl-CoA dehydrogenase family, member 9	3q26
ACO2	Aconitase, mitochondrial	22q13.2
ACSL4	Acyl-CoA synthetase long-chain family member 4	Xq22.3
ACTA1	Actin, alpha-1, skeletal muscle	1q42.1
ACTB	Actin, beta	7p22-p12
ACTG1	Actin, gamma-1	17q25.3
ACY1	Aminoacylase-1	3p21.1
ADAR	Adenosine deaminase, RNA-specific	1q21.3
ADGRG1	Adhesion G protein-coupled receptor G1	16q13
ADK	Adenosine kinase	10q11-q24
ADRA2B	Adrenergic, alpha-2B-, receptor	2q11.2
ADSL	Adenylosuccinate lyase	22q13.1
AFF2	AF4/FMR2 family, member 2 (fragile site, X-linked, E)	Xq28
AFG3L2	ATPase family gene 3-like 2	18p11
AGRN	Agrin	1pter-p32
AGTR2	Angiotensin receptor 2	Xq23
AHI1	Abelson helper integration site 1	6q23.3
AIMP1	ARS-interacting multifunctional protein 1	4q24
ALDH5A1	Succinic semialdehyde dehydrogenase	6p22
ALDH7A1	Aldehyde dehydrogenase 7 family, member A1 (antiquitin 1)	5q31
ALS2	Alsin	2q33
AMPD1	Adenosine monophosphate deaminase-1, muscle	1p21-p13
ANG	Angiogenin	14q11.2
ANO10	Anoctamin 10	3p22.1
ANO3	Anoctamin 3	11p14
ANO5	Anoctamin 5	11p14.3
AP1S2	Adaptor-related protein complex 1, sigma-2 subunit	Xp22
AP4B1	Adaptor-related protein complex 4, beta-1 subunit	1p13.2
AP4E1	Adaptor-related protein complex 4, epsilon-1 subunit	15q21.2
AP4M1	Adaptor-related protein complex 4, mu-1 subunit	7q22.1
AP4S1	Adaptor-related protein complex 4, sigma-1 subunit	14q12
AP5Z1	Adaptor-related protein complex 5, zeta-1 subunit	7p22.1
APOB	Apolipoprotein B (including Ag(x) antigen)	2p24
APTX	Aprataxin	9p13.3

AR	Androgen receptor (dihydrotestosterone receptor)	Xq11-q12
ARFGEF2	ADP-ribosylation factor guanine nucleotide-exchange factor 2, brefeldin A-inhibited	20q13.13
ARHGEF6	Rho guanine nucleotide exchange factor-6	Xq26
ARHGEF9	Rho guanine nucleotide exchange factor 9	Xq22.1
ARID1A	AT rich interactive domain 1A, SWI-like	1p35.3
ARID1B	AT-rich interaction domain-containing protein 1B	6q25.1
ARL13B	ADP-ribosylation factor-like 13B	3q11.2
ARL14EP	Chromosome 11 open reading frame 46	11p14.1-p13
ARNT2	Aryl hydrocarbon receptor nuclear translocator 2	15q24
ARSA	Arylsulfatase A	22q13.31-qter
ARX	Aristaless-related homeobox, X-linked	Xp22.13
ASCC3	Activating signal cointegrator 1 complex, subunit 3	6q16.3
ASCL1	Achaete-scute complex, Drosophila, homolog-like 1	12q22-q23
ASPM	Abnormal spindle-like, microcephaly-associated	1q31
ATCAY	Caytaxin	19p13.3
ATL1	Atlastin GTPase 1	14q11-q21
ATN1	Atrophin 1	12p13.31
ATP13A2	ATPase, type 13A2	1p36
ATP1A2	ATPase, Na+K+ transporting, alpha-2 polypeptide	1q21-q23
ATP1A3	ATPase, Na+K+ transporting, alpha-3 polypeptide	19q12-q13.2
ATP2A1	ATPase, Ca++ transporting, fast-twitch, 1	16p12
ATP2B3	ATPase, Ca++ transporting, plasma membrane, 3	Xq28
ATP5F1E	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit	20q13.3
ATP6AP2	ATPase, H+ transporting, lysosomal, accessory protein 2	Xp11.4
ATP7A	ATPase, Cu++ transporting, alpha polypeptide	Xq12-q13
ATPAF2	ATP synthase, mitochondrial F1 complex, assembly factor 2	17p11.2
ATR	Ataxia-telangiectasia and Rad3-related (FRAP-related protein-1)	3q22-q24
ATRX	ATR-X gene (helicase 2; X-linked nuclear protein)	Xq13
ATXN1	Ataxin-1	6p23
ATXN10	Ataxin 10	22q13
ATXN2	Ataxin-2	12q24
ATXN3	Ataxin-3 (josephin)	14q32.12
ATXN7	Ataxin 7	3p14.1
ATXN8OS	Ataxin 8 opposite strand	13q21
AUTS2	KIAA0442 gene	7q11.2
B3GALNT2	Beta-1,3-N-acetylgalactosaminyltransferase 2	1q42.3
BAG3	BCL2-associated athanogene 3	10q25.2-q26.2
BCS1L	bcs1, S. cerevisiae, homolog-like	2q33
BDNF	Brain-derived neurotrophic factor	11p13

BIN1	Box-dependent MYC-interacting protein-1 (amphiphysin-like)	2q14
BRAT1	BRCA1-associated ATM activator 1	7p22.3
BRWD3	Bromodomain-and WD repeat-containing protein 3	Xq13
BSCL2	Seipin	11q13
BUB1B	Budding uninhibited by benzimidazoles 1, S. cerevisiae, homolog of, beta	15q15
C12orf57	Chromosome 12 open reading frame 57	12p13
C12orf65	Chromosome 12 open reading frame 65	12q24.31
C19orf12	Chromosome 19 open reading frame 12	19q12
C9orf72	Chromosome 9 open reading frame 72	9p21
CA8	Carbonic anhydrase VIII	8q11-q12
CACNA1A	Calcium channel, voltage-dependent, P/Q type, alpha 1A subunit	19p13
CACNA1G	Calcium channel, voltage-dependent, T type, alpha-1G subunit	17q22
CACNB4	Calcium channel, voltage-dependent, beta 4 subunit	2q22-q23
CACNG2	Calcium channel, voltage-dependent, gamma-2 subunit (stargazin)	22q13.1
CAMTA1	Calmodulin-binding transcription activator 1	1p36
CAPN10	Calpain-10	2q37.3
CAPN3	Calpain, large polypeptide L3	15q15.1-q21.1
CASK	Calcium/calmodulin-dependent serine protein kinase	Xp11.4
CASP2	Caspase 2, apoptosis-related cysteine protease (neural precursor cell expressed, developmentally down-regulated 2)	7q35
CAV3	Caveolin-3	3p25
CC2D1A	Coiled-coil and C2 domain-containing 1A	19p13.12
CCDC78	Coiled-coil domain-containing protein 78	16p13.3
CCDC88C	Coiled-coil domain-containing protein 88C	14q32.11
CCNA2	Cyclin A	4q27
CCT5	chaperonin containing TCP1, subunit 5 (epsilon)	5p15.2
CDC6	Cell division cycle 6, S. cerevisiae, homolog of	17q21.2
CDH15	Cadherin-15, M-cadherin (myotubule)	16q24.3
CDK5RAP2	CDK5 regulatory subunit-associated protein 2	9q33.3
CDKL5	Cyclin-dependent kinase-like 5 (serine/threonine protein kinase 9)	Xp22
CDON	Cell adhesion molecule-related/downregulated by oncogenes	11q24.2
CENPJ	Centromeric protein J	13q12.2
CEP135	Centrosomal protein, 135kD	4q12
CEP152	Centrosomal protein, 152kD	15q21.1
CEP41	Centrosomal protein, 41kD	7q32
CEP57	Centrosomal protein 57kD	11q21
CEP63	Centrosomal protein, 63kD	3q22.2

CFL2	Cofilin 2, muscle	14q12
CHAT	Choline acetyltransferase	10q11.2
CHKB	Choline kinase, beta	22q13
CHMP1A	CHMP family, member 1A	16q24.3
CHMP2B	CHMP family, member 2B	3p11.2
CHRNA1	Cholinergic receptor, nicotinic, alpha polypeptide-1, muscle	2q24-q32
CHRNA2	Cholinergic receptor, nicotinic, alpha polypeptide-2	8p21
CHRNA4	Cholinergic receptor, nicotinic, alpha polypeptide-4	20q13.2-q13.3
CHRNB1	Cholinergic receptor, nicotinic, beta polypeptide-1, muscle	17p12-p11
CHRNB2	Cholinergic receptor, nicotinic, beta polypeptide-2	1q21
CHRND	Cholinergic receptor, nicotinic, delta polypeptide	2q33-q34
CHRNE	Cholinergic receptor, nicotinic, epsilon polypeptide	17p13-p12
CHRNG	Cholinergic receptor, nicotinic, gamma polypeptide	2q33-q34
CLCN1	Chloride channel-1, skeletal muscle	7q35
CLCN2	Chloride channel-2	3q26
CLN3	Battenin	16p12.1
CLN5	CLN5 gene	13q22.3
CLN6	CLN6 gene	15q21-q23
CLN8	CLN8 gene	8p23
CNBP	CCHC-type zinc finger nucleic acid-binding protein	3q21.3
CNKSR1	Connector enhancer of KSR 1	1p36.11
CNTN1	Contactin 1	12q11-q12
CNTNAP2	Contactin-associated protein-like 2	7q35-q36
COA5	Cytochrome c oxidase assembly factor 5	2q11.2
COG6	Component of oligomeric golgi complex 6	13q13.3
COL4A2	Collagen IV, alpha-2 polypeptide	13q34
COL6A1	Collagen VI, alpha-1 polypeptide	21q22.3
COL6A2	Collagen VI, alpha-2 polypeptide	21q22.3
COL6A3	Collagen VI, alpha-3 polypeptide	2q37
COLQ	Collagenic tail of endplate acetylcholinesterase	3p25
COQ2	CoQ2, S. cerevisiae, homolog of (parahydroxybenzoate-polyprenyltransferase, mitochondrial)	4q21-q22
COQ5	Coq5, S. cerevisiae, homolog of	12q24.31
COQ6	Coq6, S. cerevisiae, homolog of	14q24.3
COQ8A	AARF domain-containing kinase 3	1q42.2
COX10	Cytochrome c oxidase, subunit X	17p12-p11.2
COX14	Cytochrome c oxidase assembly protein COX14	12q13.12
COX15	Cytochrome c oxidase, subunit 15	10q24
COX6B1	Cytochrome c oxidase, subunit VIb polypeptide 1 (ubiquitous)	19q13.1
CP	Ceruloplasmin	3q23-q24
CPA6	Carboxypeptidase A6	8q13

CPLANE1	Ciliogenesis and planar polarity effector 1	5p13.2
CRADD	Caspase and RIP adaptor with death domain	12q21.33-q23.1
CRBN	Cereblon	3p26.2
CRPPA	Isoprenoid synthase domain-containing protein	7p21.2
CRYAB	Crystallin, alpha B	11q22.3-q23.1
CSF1R	Colony-stimulating factor-1 receptor; oncogene FMS (McDonough feline sarcoma)	5q32
CSNK1D	Casein kinase-1, delta	17q25
CSTB	Cystatin B (stefin B)	21q22.3
CTNNB1	Catenin (cadherin-associated protein), beta 1, 88kD	3p22.1
CTSD	Cathepsin D (lysosomal aspartyl protease)	11p15.5
CUL4B	Cullin-4B	Xq23
CYP27A1	Cytochrome P450, subfamily XXVIIA, polypeptide 1 (sterol 27-hydroxylase)	2q33-qter
CYP2U1	Cytochrome P450, family 2, subfamily U, polypeptide 1	4q25
CYP7B1	Cytochrome P450, subfamily VIIB (oxysterol 7-alpha-hydroxylase), polypeptide 1	8q21.3
DAG1	Dystrophin-associated glycoprotein-1	3p21
DARS2	Aspartyl-tRNA synthetase 2	1q25.1
DCC	Deleted in colorectal carcinoma	18q21.3
DCTN1	Dynactin 1 (p150, glued, Drosophila, homolog of)	2p13
DCX	Doublecortin	Xq22.3-q23
DDHD1	DDHD domain-containing protein 1	14q22.1
DDHD2	DDHD domain-containing protein 2	8p11.23
DES	Desmin	2q35
DHTKD1	Dehydrogenase E1 and transketolase domains-containing protein 1	10p14
DIP2B	Disco-interacting protein 2, Drosophila, homolog of, B	12q13.12
DIS3L2	Dis3 mitotic control, S. cerevisiae, homolog-like 2	2q37.2
DLG3	Discs large, Drosophila, homolog of, 3	Xq13.1
DMD	Dystrophin	Xp21.2
DMPK	Dystrophia myotonica-protein kinase	19q13.2-q13.3
DNA2	DNA replication helicase 2, yeast, homolog of	10q21.3-q22.1
DNAJB2	DnaJ, E. coli, homolog of, subfamily B, member 2 (heat-shock protein, DNAJ-like 1)	2q35
DNAJB6	DNAJ/HSP40 homolog, subfamily B, member 6	7q36.3
DNAJC5	DNAJ/HSP40 homolog, subfamily C, member 5	20q13.33
DNM2	Dynamin-2	19p13.2
DNMT1	DNA methyltransferase 1	19p13.3-p13.2
DOK7	Downstream of tyrosine kinase 7	4p16.2
DPAGT1	Dolichyl-phosphate N-acetylglucosamine phosphotransferase	11q23.3

DPM2	Dolichyl-phosphate mannosyltransferase 2, regulatory subunit	9q34.11
DPYD	Dihydropyrimidine dehydrogenase	1p22
DRD2	Dopamine receptor D2	11q23.1
DRD4	Dopamine receptor D4	11p15.5
DRD5	Dopamine receptor D5	4p16.1-p15.3
DST	Dystonin (bullous pemphigoid antigen 1)	6p12-p11
DUX4	Double homeo box protein 4	4q35
DYNC1H1	Dynein, cytoplasmic-1, heavy chain-1	14q32
DYRK1A	Dual specificity tyrosine-(Y)-phosphorylation regulated kinase-1A (\'minibrain\', Drosophila, homolog of)	21q22.1
DYSF	Dysferlin	2p13.3-p13.1
EARS2	Glutamyl-tRNA synthetase 2	16p13.1-p11.2
EEF1B2	Eukaryotic translation elongation factor-1, beta-2	2q33-q34
EFHC1	EF hand domain (C-terminal)-containing 1	6p12-p11
EGR2	KROX-20, Drosophila, homolog of (early growth response-2)	10q21.1-q22.1
EIF2B1	Eukaryotic translation initiation factor 2B, subunit 1	12q24.31
EIF2B2	Eukaryotic translation initiation factor 2B, subunit 2	14q24
EIF2B3	Eukaryotic translation initiation factor 2B, subunit 3	1p34.1
EIF2B4	Eukaryotic translation initiation factor 2B, subunit 4	2p23.3
EIF2B5	Eukaryotic translation initiation factor 2B, subunit 5	3q27
EIF4G1	Eukaryotic translation initiation factor 4 gamma, 1	3q27
ELP1	Inhibitor of kappa light polypeptide gene enhancer in B cells, kinase complex-associated protein	9q31
ELP2	Elongator acetyltransferase complex, subunit 2	18q12.2
EMD	Emerin	Xq28
EMX2	Empty spiracles, Drosophila, homolog of, 2	10q26.1
ENTPD1	Ectonucleoside triphosphate diphosphohydrolase 1 (CD39 antigen)	10q24
EPB41L1	Erythrocyte membrane protein band 4.1-like 1	20q11.2-q12
EPM2A	Laforin	6q24
ERBB3	Transformation gene ERBB-3	12q13
ERLIN2	Endoplasmic reticulum lipid raft-associated protein 2	8p11.2
EXOSC3	Exosome component 3	9p13.2
FA2H	Fatty acid 2-hydroxylase	16q23
FAM126A	Hyccin	7p15.3
FARS2	Phenylalanyl-tRNA synthetase 2, mitochondrial	6p25.1
FASN	Fatty acid synthase	17q25
FASTKD2	FAST kinase domains 2	2q33.3
FBXO7	F-box only protein 7	22q12-q13
FGD1	FYVE, RhoGEF, and PH domain-containing protein 1	Xp11.21
FGD4	FYVE, RhoGEF, and PH domain-containing protein 4	12p11.2
FGF14	Fibroblast growth factor-14	13q34

FHL1	Four-and-a-half LIM domains 1	Xq27.2
FIG4	Fig4, S. cerevisiae, homolog of	6q21
FKRP	Fukutin-related protein	19q13.3
FKTN	Fukutin	9q31
FLNA	Filamin A, alpha (actin-binding protein-280)	Xq28
FLNC	Filamin C (actin-binding protein-280)	7q32
FLVCR2	Feline leukemia virus subgroup C receptor 2	14q24.3
FMR1	FMR1 gene	Xq27.3
FOLR1	Folate receptor-1, adult	11q13.3-q13.5
FOXG1	Forkhead box G1B	14q13
FOXP1	Forkhead box P1	3p14.1
FOXP2	Forkhead box P2	7q31
FOXRED1	FAD-dependent oxidoreductase domain-containing protein 1	11q24.2
FRY	Furry, Drosophila, homolog of	13q13.1
FTL	Ferritin, light chain	19q13.3-q13.4
FTSJ1	FTSJ homolog 1	Xp11.23
FUS	Fusion, derived from 12-16 translocation, malignant liposarcoma	16p11.2
FUZ	Fuzzy, Drosophila, homolog of	19q13
FXN	Frataxin	9q13
GABRA1	Gamma-aminobutyric acid (GABA) A receptor, alpha-1	5q34-q35
GABRB3	Gamma-aminobutyric acid (GABA) A receptor, beta-3	15q11.2-q12
GABRD	Gamma-aminobutyric acid (GABA) A receptor, delta	1p36.3
GABRG2	Gamma-aminobutyric acid (GABA) A receptor, gamma-2	5q34
GAD1	Glutamate decarboxylase-1, brain, 67kD	2q31
GALC	Galactosylceramidase	14q31
GAMT	Guanidinoacetate methyltransferase	19p13.3
GAN	Gigaxonin	16q24.1
GARS	Glycyl-tRNA synthetase	7p15
GATAD2B	GATA zinc finger domain-containing protein 2B	1q23.1
GATM	L-arginine:glycine amidinotransferase	15q21.1
GBA2	Glucosidase, beta, acid 2	9p13.3
GDAP1	Ganglioside-induced differentiation-associated protein 1	8q21.11
GDI1	GDP dissociation inhibitor 1	Xq28
GFAP	Glial fibrillary acidic protein	17q21
GFER	Growth factor, erv1, S. cerevisiae, homolog of (augmenter of liver regeneration)	16p13.3-p13.12
GFPT1	Glutamine-fructose-6-phosphate transaminase	2p13
GIGYF2	GRB10-interacting GYF protein 2	2q37.1
GJB1	Gap junction protein, beta-1, 32kD (connexin 32)	Xq13.1
GJC2	Gap junction protein, gamma-2 (47kD)	1q42.13
GLI2	GLI-Kruppel family member GLI2 (oncogene GLI2)	2q14

GLRA1	Glycine receptor, alpha-1 polypeptide	5q32
GLRB	Glycine receptor, beta subunit	4q31.3
GNAL	Guanine nucleotide-binding protein, alpha-subunit, olfactory type	18p11.21
GNB4	Guanine nucleotide-binding protein, beta-4	3q26.3
GNE	UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase	9p13.3
GON4L	GON4-like protein	1q22
GOSR2	Golgi snap receptor complex member 2	17q21
GPHN	Gephyrin	14q24
GRIA3	Glutamate receptor, ionotropic, AMPA 3	Xq25-q26
GRIK2	Glutamate receptor, ionotropic, kainate 2	6q21
GRIN1	Glutamate receptor, ionotropic, N-methyl D-aspartate 1	9q34.3
GRIN2A	Glutamate receptor, ionotropic, N-methyl D-aspartate 2A	16p13
GRIN2B	Glutamate receptor, ionotropic, N-methyl D-aspartate 2B	12p12
GRM1	Glutamate receptor, metabotropic, 1	6q24
GRN	Granulin	17q21.32
HCFC1	Host cell factor C1 (VP16-accessory protein)	Xq28
HDAC1	Histone deacetylase-1	1p34.1
HDAC4	Histone deacetylase 4	2q37.2
HEPACAM	Hepatocyte cell adhesion molecule	11q24
HEXA	Hexosaminidase A, alpha polypeptide	15q23-q24
HINT1	Histidine triad nucleotide-binding protein 1 (protein kinase C inhibitor 1)	5q31.2
HIST1H4B	Histone 1, H4b	6p21.3
HIST3H3	Histone 3, H3	1q42
HK1	Hexokinase-1	10q22
HOXB1	Homeo box-B1	17q21-q22
HOXD10	Homeo box-D10	2q31-q32
HSD17B10	17-beta-hydroxysteroid dehydrogenase X	Xp11.2
HSD17B4	Hydroxysteroid (17-beta) dehydrogenase 4	5q23.1
HSPB1	Heat-shock 27kD protein-1	7q11.23
HSPB3	Heat-shock 27kD protein 3	5q11.2
HSPB8	Heat-shock 22-kD protein 8	12q24
HSPD1	Heat-shock 60kD protein 1	2q33.1
HTRA2	HTRA serine peptidase 2	2p12
HTT	Huntingtin	4p16.3
HUWE1	HECT, UBA, and WWE domains-containing protein 1	Xp11.2
HYLS1	HYLS1 gene	11q24.2
IAPP	Islet amyloid polypeptide (diabetes-associated peptide; amylin)	12p12.3-p12.1
IER3IP1	Immediate-early response 3-interacting protein 1	18q21.1

IGBP1	Immunoglobulin-binding protein 1	Xq13.1-q13.3
IGHMBP2	Immunoglobulin mu binding protein 2	11q13.2-q13.4
IL1RAPL1	Il-1 receptor accessory protein-like 1	Xp22.1-p21.3
INF2	Inverted formin 2	14q32.33
INPP4A	Inositol polyphosphate-4-phosphatase, type I, 107kD	2q11.2
INPP5E	Inositol polyphosphate-5-phosphatase, 72kD	9q34.3
IQSEC2	IQ motif- and Sec7 domain-containing protein 2	Xp11.22
ISCU	Iron-sulfur cluster scaffold, E. coli, homolog of	12q24.1
ITGA7	Integrin, alpha-7	12q13
ITM2B	Integral membrane protein 2B (BRI gene)	13q14
ITPR1	Inositol 1,4,5-triphosphate receptor, type 1	3p26.1
JPH3	Junctophilin 3	16q24.3
KANK1	KN motif- and ankyrin repeat domain-containing protein 1	9p24.3
KARS	Lysyl-tRNA synthetase	16q22.2-q22.3
KBTBD13	Kelch repeat and BTB/POZ domains-containing protein 13	15q22.31
KCNA1	Potassium voltage-gated channel, shaker-related subfamily, member 1	12p13
KCNC3	Potassium voltage-gated channel, Shaw-related subfamily, member 3	19q13.3-q13.4
KCNJ10	Potassium inwardly-rectifying channel, subfamily J, member 10	1q23.2
KCNMA1	Potassium large conductance calcium-activated channel, subfamily M, alpha member 1 (slowpoke, Drosophila, homolog of)	10q22.3
KCNQ2	Potassium voltage-gated channel, KQT-like subfamily, member 2	20q13.3
KCNQ3	Potassium voltage-gated channel, KQT-like subfamily, member 3	8q24
KCNT1	Potassium channel, subfamily T, member 1	9q34.3
KCTD7	Potassium channel tetramerization domain containing 7	7q11.21
KDM5A	Lysine(K)-specific demethylase 5A	12p11
KDM5C	Lysine-specific demethylase 5C (Jumonji, AT-rich interactive domain 1C)	Xp11.22-p11.21
KDM6B	Lysine-specific demethylase 6B	17p13.1
KIF1A	Kinesin family member 1A	2q37
KIF1B	Kinesin family member 1B	1p36.2
KIF5A	Kinesin family member 5A	12q13
KIF7	Kinesin family member 7	15q26.1
KIRREL3	Kin of IRRE-like 3	11q24.2
KNL1	Kinetochore scaffold 1	15q15.1
L1CAM	L1 cell adhesion molecule	Xq28
LAMA1	Laminin, alpha-1	18p11.31

LAMA2	Laminin, alpha-2 (merosin)	6q22-q23
LAMB1	Laminin, beta-1	7q31.1-q31.3
LAMC3	Laminin, gamma-3	9q33-q34
LARGE1	Acetylglucosaminyltransferase-like protein	22q12.3-q13.1
LDB3	LIM domain binding 3	10q22.2-q23.3
LGI1	Leucine-rich gene, glioma-inactivated, 1	10q24
LINS1	Lines, Drosophila, homolog of, 1	15q26
LITAF	LPS-induced TNFA factor	16p13.3-p12
LMNA	Lamin A/C	1q21.2
LMNB1	Lamin B1	5q23.2
LRPPRC	Leucine-rich PPR motif-containing protein	2p21
LRRK2	Leucine-rich repeat kinase 2 (dardarin)	12q12
LRSAM1	Leucine-rich repeat- and sterile alpha motif-containing 1	9q33.3-q34.11
MAGT1	Magnesium transporter 1	Xq13.1-q13.2
MAN1B1	Mannosidase, alpha, class 1B member 1	9q34.3
MAOA	Monoamine oxidase A	Xp11.23
MAPK10	Mitogen-activated protein kinase 10	4q21.3
MAPT	Microtubule-associated protein tau	17q21.1
MATR3	Matrin 3	5q31.2
MBD5	Methyl-CpG-binding domain protein 5	2q23.1
MCPH1	Microcephalin	8p23
MECP2	Methyl-CpG-binding protein-2	Xq28
MED12	Mediator of RNA polymerase II transcription, subunit 12, S. cerevisiae, homolog of	Xq13
MED13L	Mediator complex subunit 13-like	12q24
MED17	Mediator complex subunit 17	11q21
MED23	Mediator complex subunit 23	6q23.2
MED25	Mediator of RNA polymerase II transcription, subunit 25, S. cerevisiae, homolog of	19q13.3
MEF2C	MADS box transcription enhancer factor 2, polypeptide C (myocyte enhancer factor 2C)	5q14
MEGF10	Multiple epidermal growth factor-like domains 10	5q23.2
MFF	Mitochondrial fission factor	2q36.3
MFSD8	Major facilitator superfamily domain-containing protein 8	4q28.1-q28.2
MLC1	MLC1 gene	22q13.33
MMD2	Monocyte-to-macrophage differentiation-associated protein 2	7p22.1
MOG	Myelin-oligodendrocyte glycoprotein	6p21.3
MPC1	Brain protein 44-like	6q27
MPDZ	Multiple PDZ domain protein	9p23
MPP3	Membrane protein, palmitoylated-3 (MAGUK p55 subfamily member 3)	17q12-q21

MPZ	Myelin protein zero	1q22
MR1	Major histocompatibility complex, class I-related	1q25.3
MRE11	Meiotic recombination 11, S. cerevisiae, homolog A of	11q21
MRI1	Methylthioribose-1-phosphate isomerase, S. cerevisiae, homolog of	19p13.2
MRPL3	Mitochondrial ribosomal protein L3	3q22.1
MSTN	Growth differentiation factor-8 (myostatin)	2q32.2
MTFMT	Mitochondrial methionyl-tRNA formyltransferase	15q22.31
MTM1	Myotubularin	Xq28
MTMR2	Myotubularin-related protein 2	11q22
MTPAP	Mitochondrial poly(A) polymerase	10p11.23
MUSK	Receptor tyrosine kinase MuSK	9q31.3-q32
MYBPC1	Myosin-binding protein C, slow type	12q23.2
MYF6	Myogenic factor-6	12q21
MYH2	Myosin, heavy polypeptide-2, skeletal muscle, adult	17p13.1
MYH3	Myosin, heavy polypeptide-3, skeletal muscle, embryonic	17p13.1
MYH7	Myosin, heavy polypeptide-7, cardiac muscle, beta	14q12
MYH8	Myosin, heavy polypeptide-8, skeletal muscle, perinatal	17p13.1
MYOT	Myotilin (titin immunoglobulin domain protein)	5q31
NAGA	Acetylgalactosaminidase, alpha-N- (alpha-galactosidase B)	22q11
NBN	Nibrin	8q21
NDE1	nudE neurodevelopment protein 1	16p13.1
NDRG1	N-myc downstream-regulated gene 1	8q24.3
NDST1	N-deacetylase/N-sulfotransferase (heparan sulfate-N-deacetylase/N-sulfotransferase)	5q32-q33.3
NDUFA1	NADH-ubiquinone oxidoreductase subunit A1	Xq24
NDUFA10	NADH-ubiquinone oxidoreductase subunit A10	2q37.3
NDUFA11	NADH-ubiquinone oxidoreductase subunit A11	19p13.3
NDUFA12	NADH-ubiquinone oxidoreductase subunit A12	12q22
NDUFA2	NADH-ubiquinone oxidoreductase subunit A2	5q31.2
NDUFA7	NADH-ubiquinone oxidoreductase subunit A7	19p13.2
NDUFA9	NADH-ubiquinone oxidoreductase subunit A9	12p
NDUFAB1	NADH-ubiquinone oxidoreductase subunit AB1	16p12.3-p12.1
NDUFAF1	NADH-ubiquinone oxidoreductase complex assembly factor 1	15q13.3
NDUFAF2	NADH-ubiquinone oxidoreductase complex assembly factor 2	5q12.1
NDUFAF3	NADH-ubiquinone oxidoreductase complex assembly factor 3	3p21.31
NDUFAF4	NADH-ubiquinone oxidoreductase complex assembly factor 4	6q16.1
NDUFAF5	NADH-ubiquinone oxidoreductase complex assembly factor 5	20p12.1

NDUFAF6	NADH-ubiquinone oxidoreductase complex assembly factor 6	8q22.1
NDUFB3	NADH-ubiquinone oxidoreductase subunit B3	2q31.3
NDUFS1	NADH-ubiquinone oxidoreductase core subunit S1	2q33-q34
NDUFS2	NADH-ubiquinone oxidoreductase core subunit S2	1q23
NDUFS3	NADH-ubiquinone oxidoreductase core subunit S3	11p11.11
NDUFS4	NADH-ubiquinone oxidoreductase subunit S4	5q11.1
NDUFS5	NADH-ubiquinone oxidoreductase subunit S5	1p34.2-p33
NDUFS6	NADH-ubiquinone oxidoreductase subunit S6	5pter-p15.33
NDUFS7	NADH-ubiquinone oxidoreductase core subunit S7	19p13
NDUFS8	NADH-ubiquinone oxidoreductase core subunit S8	11q13
NDUFV1	NADH-ubiquinone oxidoreductase core subunit V1	11q13
NDUFV2	NADH-ubiquinone oxidoreductase core subunit V2	18p11.31-p11.2
NEB	Nebulin	2q22
NEFL	Neurofilament, light polypeptide	8p21
NFU1	NFU1 iron-sulfur cluster scaffold, S. cerevisiae, homolog of	2p15-p13
NGF	Nerve growth factor, beta	1p13.1
NHEJ1	Nonhomologous end-joining factor 1	2q35
NHLRC1	NHL repeat-containing 1 gene (malin)	6p22.3
NIN	Ninein	14q22.1
NIPA1	Nonimprinted gene in Prader-Willi syndrome/Angelman syndrome chromosome region 1	15q11.1
NKX2-1	NK2 homeobox 1	14q13
NLGN3	Neuroligin 3	Xq13
NLGN4X	Neuroligin 4	Xp22.33
NOL3	Nucleolar protein 3	16q21-q23
NOP56	Nop56, S. cerevisiae, homolog of	20p13
NPC1	NPC1 gene	18q11-q12
NPC2	Epididymal secretory protein HE1	14q24.3
NRXN1	Neurexin 1	2p16.3
NTRK1	Neurotrophic tyrosine kinase, receptor, type 1	1q21-q22
NUBPL	Nucleotide-binding protein-like protein	14q12
NUP62	Nucleoporin, 62-kD	19q13.33
OCLN	Occludin	5q13.1
OCRL	Phosphatidylinositol polyphosphate 5-phosphatase (OCRL gene)	Xq26.1
OPHN1	Oligophrenin-1	Xq12
OPTN	Optineurin	10p15-p14
ORC1	Origin recognition complex, subunit 1, S. cerevisiae, homolog	1p32
ORC6	Origin recognition complex, subunit 6, S. cerevisiae, homolog of	16q12
PABPN1	Poly(A)-binding protein, nuclear 1	14q11.2-q13

PACS1	Phosphofurin acidic cluster sorting protein 1	11q13.1
PAFAH1B1	Platelet-activating factor acetylhydrolase, isoform 1B, alpha subunit	17p13.3
PAK3	p21-activated kinase-3	Xq23
PANK2	Pantothenate kinase 2	20p13-p12.3
PARK7	Oncogene DJ-1	1p36
PARP1	Poly(ADP-ribose) polymerase 1	1q42
PC	Pyruvate carboxylase	11q13.4-q13.5
PCDH19	Protocadherin 19	Xq22
PDGFRB	Platelet-derived growth factor receptor, beta polypeptide	5q32
PDHA1	Pyruvate dehydrogenase, E1-alpha polypeptide-1	Xp22.2-p22.1
PDHB	Pyruvate dehydrogenase, E1 beta polypeptide	3p13-q23
PDHX	Pyruvate dehydrogenase complex, lipoyl-containing component X	11p13
PDSS2	Prenyl diphosphate synthase, subunit 2	6q21
PDX1	Pancreas/duodenum homeobox protein 1	13q12.1
PDYN	Prodynorphin	20p13
PDZD7	PDZ domain-containing 7	10q24.3
PECR	Peroxisomal trans-2-enoyl-CoA reductase	2q35
PEX1	Peroxisome biogenesis factor-1	7q21-q22
PEX12	Peroxisome biogenesis factor 12	17q12
PEX14	Peroxisome biogenesis factor 14	1p36.2
PEX16	Peroxisome biogenesis factor 16	11p12-p11.2
PEX2	Peroxisome biogenesis factor 2	8q21.1
PEX26	Peroxisome biogenesis factor 26	22q11.21
PEX3	Peroxisomal biogenesis factor-3	6q24.2
PEX5	Peroxisome biogenesis factor 5	12p13.3
PEX6	Peroxisomal biogenesis factor 6 (peroxisomal AAA-type ATPase 1)	6p21.1
PFN1	Profilin-1	17p13.3
PHC1	Polyhomeotic-like 1	12p13.31
PHF8	PHD finger protein 8	Xp11.2
PHGDH	Phosphoglycerate dehydrogenase	1p12
PHOX2B	Paired mesoderm homeo box 2B	4p12
PHYH	Phytanoyl-CoA hydroxylase	10pter-p11.2
PIGN	Phosphatidylinositol glycan, class N	18q21.33
PIK3R2	Phosphatidylinositol 3-kinase, regulatory subunit 2	19q13.2-q13.4
PIK3R5	Phosphatidylinositol 3-kinase, regulatory subunit 5	17p13.1
PINK1	PTEN-induced putative kinase 1	1p36
PIP5K1C	Phosphatidylinositol-4-phosphate 5-kinase, type I, gamma	19p13.3
PLA2G6	Phospholipase A2, group VI	22q13.1
PLCB1	Phospholipase C, beta-1	20p12

PLEC	Plectin 1	8q24
PLEKHG5	Pleckstrin homology domain- and RhoGEF domain-containing protein G5	1p36
PLP1	Proteolipid protein 1	Xq22
PMM2	Phosphomannomutase 2	16p13.3-p13.2
PMP22	Peripheral myelin protein-22	17p11.2
PNKD	Myofibrillogenesis regulator 1	2q35
PNKP	Polynucleotide kinase 3' phosphatase	19q13.4
PNPLA2	Patatin-like phospholipase domain-containing protein 2	11p15.5
PNPLA6	Patatin-like phospholipase domain-containing protein 6	19p13.3
PNPO	Pyridoxamine 5'-phosphate oxidase	17q21.32
POLG	Polymerase (DNA directed), gamma	15q25
POLR3A	Polymerase III, RNA, subunit A	10q22.3
POLR3B	Polymerase III, RNA, subunit B	12q23.3
POMGNT1	Protein 0-mannose beta-1,2-N-acetylglucosaminyltransferase	1p34-p33
POMGNT2	Protein O-mannose beta-1,2-N-acetylglucosaminyltransferase 2	3p22.1
POMT1	Protein O-mannosyltransferase 1	9q34.1
POMT2	Putative protein O-mannosyltransferase 2	14q24.3
PPP2R2B	Protein phosphatase 2, regulatory subunit B, beta	5q31-q33
PPT1	Palmitoyl-protein thioesterase 1	1p32
PQBP1	Polyglutamine-binding protein 1	Xp11.23
PRICKLE1	Prickle-like 1	12q12
PRICKLE2	Prickle-like 2	3p14
PRKCG	Protein kinase C, gamma polypeptide	19q13.4
PRKN	Parkin	6q25.2-q27
PRKRA	Protein kinase, interferon-inducible double-stranded RNA-dependent activator	2q31.3
PRMT9	Protein arginine methyltransferase 9	4q31.23
PRNP	Prion protein (p27-30)	20pter-p12
PRPS1	Phosphoribosyl pyrophosphate synthetase-1	Xq22-q24
PRRT2	Proline-rich transmembrane protein 2	16p11.2
PRSS12	Protease, serine, 12	4q25-q26
PRX	Periaxin	19q13.1-q13.2
PSAP	Prosaposin (sphingolipid activator protein-1)	10q22.1
PSEN1	Presenilin 1	14q24.3
PSEN2	Presenilin 2	1q31-q42
PTCH1	Patched, Drosophila, homolog of	9q22.3
PTEN	Phosphatase and tensin homolog (mutated in multiple advanced cancers 1)	10q23.31
RAB18	Ras-associated protein RAB18	10p12.1
RAB39B	Ras-associated protein RAB39B	Xq28
RAB7A	Ras-associated protein RAB7	3q21

RABL6	RAB-like protein 6	9q34.3
RAD50	RAD50, S. cerevisiae, homolog of	5q31
RAI1	Retinoic acid-induced gene 1	17p11.2
RALGDS	ral guanine nucleotide dissociation stimulator	9q34
RAPSN	Receptor-associated protein of the synapse, 43kD	11p11.2-p11.1
RARS2	Arginyl-tRNA synthetase 2	6q16.1
RBBP8	Retinoblastoma-binding protein 8	18q11.2
REEP1	Receptor expression-enhancing protein 1	2p11.2
RELN	Reelin	7q22
RETREG1	Family with sequence similarity 134, member B	5p15.1
RGS7	Regulator of G protein signaling 7	1q43
RMND1	Required for meiotic nuclear division 1, S. cerevisiae, homolog of	6q25
RNASEH2A	Ribonuclease H2, large subunit	19p13.13
RNASEH2B	Ribonuclease H2, subunit B	13q14.1
RNASEH2C	Ribonuclease H2, subunit C	11q13.2
RNASET2	Ribonuclease T2	6q27
RNF170	RING finger protein 170	8p11.2
RNU4ATAC	RNA, U4, small nuclear, AT-AC form	2q14.2
ROGDI	Rogdi, Drosophila, homolog of	16p13.3
RPGRIP1L	RPGRIP1-like	16q12.2
RPS6KA3	Ribosomal protein S6 kinase, 90kD, polypeptide 3	Xp22.2-p22.1
RTN2	Reticulon-2	19q13.3
RTTN	Rotatin	18q22.2
RXYLT1	Ribitol xylosyltransferase 1	12q14.2
RYR1	Ryanodine receptor-1, skeletal	19q13.1
SACS	Sacsin	13q12
SAMHD1	SAM domain- and HD domain-containing protein 1	20q11.2
SBF2	SET binding factor 2 (myotubularin-related 13)	11p15
SC5D	Sterol C5-desaturase-like	11q23.3
SCAPER	S-phase cyclin A-associated protein in the endoplasmic reticulum	15q24
SCARB2	Scavenger receptor class B, member 2	4q13-q21
SCN1A	Sodium channel, voltage-gated, type I, alpha polypeptide	2q24
SCN1B	Sodium channel, voltage-gated, type I, beta polypeptide	19q13.1
SCN2A	Sodium channel, voltage-gated, type II, alpha subunit	2q23-q24.3
SCN4A	Sodium channel, voltage-gated, type IV, alpha polypeptide	17q23.1-q25.3
SCN8A	Sodium channel, voltage gated, type VIII, alpha polypeptide	12q13
SCN9A	Sodium channel, voltage-gated, type IX, alpha subunit	2q24
SCO1	SCO1 cytochrome c oxidase assembly protein	17p13-p12
SCO2	SCO2 cytochrome c oxidase assembly protein	22q13

SDHA	Succinate dehydrogenase complex, subunit A, flavoprotein	5p15
SDHAF1	Succinate dehydrogenase complex assembly factor 1	19q12-q13.2
SELENON	Selenoprotein N	1p36-p35
SEPSECS	O-phosphoserine tRNA-selenocysteine tRNA synthase	4p15.2
SEPT9	Septin 9	17q25
SERPINI1	Protease inhibitor 12	3q26
SETX	Senataxin	9q34
SGCA	Sarcoglycan, alpha (50kD dystrophin-associated glycoprotein; adhalin)	17q12-q21.33
SGCB	Sarcoglycan, beta (43kD dystrophin-associated glycoprotein)	4q12
SGCD	Sarcoglycan, delta (35kD dystrophin-associated glycoprotein)	5q33
SGCE	Sarcoglycan, epsilon	7q21
SGCG	Sarcoglycan, gamma (35kD dystrophin-associated glycoprotein)	13q12
SH3TC2	SH3 domain and tetratricopeptide repeat domain 2	5q32
SHANK2	SH3 and multiple ankyrin repeat domains 2	11q13.3-q13.4
SHANK3	SH3 and multiple ankyrin repeat domains 3	22q13.3
SHH	Sonic hedgehog	7q36
SHROOM4	Shroom family member 4	Xp11.2
SIGMAR1	Sigma nonopioid intracellular receptor 1	9p13
SIL1	Sil1, S. cerevisiae, homolog of	5q31
SIX3	Sine oculis homeo box, Drosophila, homolog of, 3	2p21
SLC12A6	Solute carrier family 12 (potassium/chloride transporters), member 6	15q13-q14
SLC1A3	Solute carrier family 1 (glial high affinity glutamate transporter), member 3	5p13
SLC20A2	Solute carrier family 20, phosphate transporter, member 2 (murine leukemia virus, amphotropic, receptor for)	8p11.21
SLC25A19	Solute carrier family 25 (mitochondrial deoxynucleotide carrier), member 19	17q25.3
SLC25A22	Solute carrier family 25 (mitochondrial carrier, glutamate), member 22	11p15.5
SLC2A1	Solute carrier family 2 (facilitated glucose transporter), member 1	1p34.2
SLC30A10	Solute carrier family 30 (zinc transporter), member 10	1q41
SLC31A1	Solute carrier family 31 (copper transporter), member 1	9q31-q32
SLC33A1	Solute carrier family 33 (acetyl-CoA transporter), member 1	3q25.31
SLC35A2	Solute carrier family 35 (UDP-galactose transporter), member 2	Xp11.23-p11.22

SLC52A2	Solute carrier family 52, riboflavin transporter, member 2	8q24.3
SLC52A3	Solute carrier family 52, riboflavin transporter, member 3	20p13
SLC5A7	Solute carrier family 5 (choline transporter), member 7	2q12.3
SLC6A3	Solute carrier family 6 (neurotransmitter transporter, dopamine), member 3	5p15.3
SLC6A5	Solute carrier family 6 (neurotransmitter transporter, glycine), member 5	11p15.2-p15.1
SLC6A8	Solute carrier family 6 (neurotransmitter transporter, creatine), member 8	Xq28
SLC9A6	Solute carrier family 9 (sodium/hydrogen exchanger), member 6	Xq26.3
SLITRK1	SLIT- and NTRK-like family, member 1	13q31
SMARCA2	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2	9p24-p23
SMARCA4	SWI/SNF-related, matrix-associated, actin-dependent regulator of chromatin, subfamily A, member 4	19p13.2
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1	22q11
SMCHD1	Structural maintenance of chromosomes flexible hinge domain-containing protein 1	18p11.32
SMN1	Survival of motor neuron 1, telomeric	5q12.2-q13.3
SMS	Spermine synthase	Xp22.1
SNAP29	Synaptosomal-associated protein, 29kD	22q11.2
SNCA	Synuclein, alpha (non A4 component of amyloid precursor)	4q21
SNCB	Synuclein, beta	5q35
SNIP1	SMAD nuclear interacting protein 1	1p34.3
SOBP	Sine oculis-binding protein, Drosophila, homolog of	6q21
SOD1	Superoxide dismutase-1, soluble	21q22.1
SPART	Spartin	13q12.3
SPAST	Spastin	2p22-p21
SPG11	Spastascin	15q21.1
SPG21	Acidic cluster protein, 33kD, (maspardin)	15q21-q22
SPG7	Paraplegin	16q24.3
SPR	Sepiapterin reductase	2p14-p12
SPTAN1	Spectrin, alpha, nonerythrocytic-1 (alpha-fodrin)	9q33-q34
SPTBN2	Spectrin, beta, nonerythrocytic, 2	11q13
SPTLC1	Serine palmitoyltransferase, long-chain base subunit 1	9q22.1-q22.3
SPTLC2	Serine palmitoyltransferase, long-chain base subunit 2	14q24.3-q31
SRPX2	SUSHI repeat-containing protein, X-linked, 2	Xq21.33-q23
ST3GAL3	ST3 beta-galactoside alpha-2,3-sialyltransferase 3	1p34.1
ST3GAL5	Sialyltransferase 9	2p11.2
STAMBP	STAM binding protein	2p13.1

STIL	SCL/TAL1-interrupting locus	1p33
STIM1	Stromal interaction molecule 1	11p15.5
STRADA	STE20-related kinase adaptor alpha	17q23.3
STXBP1	Syntaxin-binding protein 1	9q34.1
SUCLG1	Succinate-CoA ligase, alpha subunit	2p11.2
SUOX	Sulfite oxidase	Chr.12
SURF1	Surfeit-1	9q34
SYN1	Synapsin I	Xp11.4-p11.2
SYNE1	Spectrin repeat-containing nuclear envelope protein 1 (nesprin 1)	6q25
SYNE2	Spectrin repeat-containing nuclear envelope protein 2 (nesprin 2)	14q23
SYNGAP1	Synaptic Ras GTPase activating protein 1	6p21.3
SYP	Synaptophysin	Xp11.23-p11.22
SYT14	Synaptotagmin 14	1q32.2
TACO1	Translational activator of mitochondrially encoded cytochrome c oxidase subunit I	17q22-q24.2
TAF1	TAF1 RNA polymerase II, TATA box-binding protein-associated factor, 250kD	Xq13
TAF2	TAF2 RAN polymerase II, TATA box-binding protein-associated factor, 150kD	8q24.12
TARDBP	TAR DNA-binding protein	1p36.2
TBC1D24	TBC1 domain family, member 24	16p13.3
TBP	TATA box binding protein	6q27
TCAP	Telethonin	17q12
TCTN1	Tectonic family, member 1	12q24.1
TDGF1	Teratocarcinoma-derived growth factor-1	3p23-p21
TDP1	Tyrosyl-DNA phosphodiesterase 1	14q31-q32
TECPR2	Tectonin beta-propeller repeat-containing protein 2	14q32.31
TECR	Trans-2,3-enoyl-CoA reductase	19p13.12
TFG	TRK-fused gene	3q11-q12
TGIF1	TG-interacting factor 1	18p11.3
TGM6	Transglutaminase 6	20p13-p12.2
TH	Tyrosine hydroxylase	11p15.5
THAP1	THAP domain-containing protein 1	8p11.21
TIA1	TIA1 cytotoxic granule-associated RNA-binding protein	2p13
TK2	Thymidine kinase, mitochondrial	16q22
TMEM135	Transmembrane protein 135	11q14.2
TMEM138	Transmembrane protein 138	11q12.2
TMEM216	Transmembrane protein 216	11q12.2
TMEM237	Transmembrane protein 237	2q33.2
TMEM67	Transmembrane protein 67 (meckelin)	8q21.13-q22.1
TMEM70	Transmembrane protein 70	8q21.11

TNNI2	Troponin I, fast-twitch skeletal muscle isoform	11p15.5
TNNT1	Troponin-T1, skeletal, slow	19q13.4
TNNT3	Troponin-T3, skeletal, fast	11p15.5
TOR1A	Torsin A	9q34
TPK1	Thiamine pyrophosphokinase	7q34
TPM2	Tropomyosin-2, beta	9p13.2-p13.1
TPM3	Tropomyosin 3	1q22-q23
TPP1	Tripeptidyl peptidase 1	11p15.5
TRAPPC9	Trafficking protein particle complex 9	8q24.3
TREM2	Triggering receptor expressed on myeloid cells 2	6p21.2
TREX1	3' repair exonuclease 1	3p21.3-p21.2
TRIM32	Tripartite-motif-containing protein 32	9q31-q34.1
TRMT1	tRNA methyltransferase 1, S. cerevisiae, homolog of	19p13.3
TRPA1	Transient receptor potential cation channel, subfamily A, member 1 (ankyrin-like protein with transmembrane domains 1)	8q13
TRPV4	Transient receptor potential cation channel, subfamily V, member 4 (vanilloid receptor-related osmotically activated channel)	12q24.1
TSC1	Hamartin (tuberous sclerosis 1 gene)	9q34
TSC2	Tuberin (tuberous sclerosis 2 gene)	16p13.3
TSEN2	tRNA splicing endonuclease 2, S. cerevisiae, homolog of	3p25.1
TSEN34	tRNA splicing endonuclease 34, S. cerevisiae, homolog of	19q13.4
TSEN54	tRNA splicing endonuclease 54, S. cerevisiae, homolog of	17q25.1
TSFM	Ts translation elongation factor, mitochondrial	12q13-q14
TSPAN7	Tetraspanin 7	Xq11
TTBK2	Tau tubulin kinase 2	15q15.2
TTC19	Tetratricopeptide repeat domain 19	17p12
TTI2	TELO2-interacting protein 2	8p12
TTN	Titin	2q31
TTPA	Tocopherol, alpha, transfer protein	8q13.1-q13.3
TUBA1A	Tubulin, alpha-1A	12q12-q14
TUBA8	Tubulin, alpha 8	22q11
TUBB2B	Tubulin, beta-2B	6p25.2
TUBB3	Tubulin, beta-3	16q24.3
TUBB4A	Tubulin, beta-4A	19p13.3
TUBGCP6	Tubulin-gamma complex-associated protein 6	22q13.33
TUFM	Tu translation elongation factor, mitochondrial	16p11.2
TUSC3	Tumor suppressor candidate 3	8p22
TYMP	Thymidine phosphorylase	22q13.32-qter
TYROBP	TYRO protein tyrosine kinase-binding protein	19q13.1
UBA1	Ubiquitin-like modifier-activating enzyme 1	Xp11.23
UBE2A	Ubiquitin-conjugating enzyme E2A (RAD6 homolog)	Xq24

UBE3A	Ubiquitin protein ligase E3A	15q11-q13
UBQLN2	Ubiquilin 2	Xp11.23-p11.1
UBR7	Ubiquitin protein ligase E3 component N-recognin 7	14q32.12
UPF3B	UPF3 regulator of nonsense transcripts, yeast, homolog of, B	Xq25-q26
UQCRB	Ubiquinol-cytochrome c reductase binding protein	8q22
UQCRC2	Ubiquinol-cytochrome c reductase core protein II	16p12
UQCRQ	Ubiquinol-cytochrome c reductase, complex III subunit VII, 9.5kD	5q31.1
UROC1	Urocanase domain-containing protein 1	3q21.3
VANGL1	Vang-like 1	1p13
VAPB	Vesicle-associated membrane protein-associated protein B	20q13.3
VCP	Valosin-containing protein	9p13-p12
VLDLR	Very low density lipoprotein receptor	9p24
VMA21	Vma21, S. cerevisiae, homolog of	Xq28
VPS13A	Vacuolar protein sorting 13A (chorein)	9q21
VPS33B	Vacuolar protein sorting 33, yeast, homolog of, B	15q26.1
VPS35	Vacuolar protein sorting 35, yeast, homolog of	16q11.2
VPS37A	Vacuolar protein sorting 37A	8p23-p21
VRK1	Vaccinia-related kinase-1	14q32
WASHC5	WASH complex, subunit 5	8q24.13
WDR45	WD repeat-containing protein 45	Xp11.23
WDR45B	WD repeat domain 45B	17q25.3
WDR62	WD repeat-containing protein 62	19q13.12
WDR81	WD repeat-containing protein 81	17p13.3
WNK1	WNK lysine deficient protein kinase 1	12p13
YARS	Tyrosyl-tRNA synthetase	1p35
ZBTB40	Zinc finger- and BTB domain-containing protein 40	1p36
ZCCHC8	Zinc finger CCHC domain-containing protein 8	12q24.31
ZDHHC15	Zinc finger DHHC domain-containing protein 15	Xq13.3
ZDHHC9	Zinc finger DHHC domain-containing protein 9	Xq26.1
ZEB2	Zinc finger E box-binding homeobox 2	2q22
ZFYVE26	Zinc finger FYVE domain-containing protein 26	14q24.1
ZFYVE27	Zinc finger FYVE domain-containing protein 27	10q24.2
ZIC2	ZIC family, member 2	13q32
ZNF335	Zinc finger protein 335	20q11.2-q13.1
ZNF41	Zinc finger protein-41	Xp22.1-cen
ZNF526	Zinc finger protein 526	19q13.2
ZNF592	Zinc finger protein 592	15q25.3
ZNF674	Zinc finger protein 674	Xp11
ZNF711	Zinc finger protein-711	Xq21.1-q21.3
ZNF81	Zinc finger protein-81 (HFZ20)	Xp22.1-p11

Appendix II.1

Contents

Table II.1.1- Additional pathogenic variants identified in patients with congenital muscular dystrophies or congenital myopathies.

Gene	DNA Variant (hg19)	Exon/ Intron	DNA Variant	RNA effect	Predicted effect on Protein	Gen der	Pheno type	Zygosity / 2nd variant / orientatio n [a]	Reference
ACTA1	g.229567260G>A	7	NM_001100.3: c.1120C>T	r.(?)	p.(Arg374Cys)	M	NEM	Het. / na / de novo	Laing et al., 2009
ACTA1	g.229568041G>A	4	NM_001100.3: c.592C>T	r.(?)	p.(Arg198Cys)	F	NEM	Het. / na / de novo	Laing et al., 2009
MTM1	g.149767061G>T	4	NM_000252.2: c.142G>T	r.(?)	p.(Glu48*)	M	XLMTM	Hemi. / c.0 / maternal	Tanner et al., 1999
MTM1	g.149809779A>G	8	NM_000252.2: c.566A>G	r.(?)	p.(Asn189Ser)	M	XLMTM	Hemi. / c.0 / maternal	Laporte et al., 1996
MTM1	g.149809806A>G	8	NM_000252.2: c.593A>G	r.(?)	p.(Tyr198Cys)	M	XLMTM	Hemi. / c.0 / maternal	Novel
MTM1	g.149828138G>A	12	NM_000252.2: c.1262G>A	r.(?)	p.(Arg421Gln)	M	XLMTM	Hemi. / c.0 / de novo	Gonyon et al., 1997
MTM1	g.149826468G>T	11	NM_000252.2: c.1228G>T	r.(?)	p.(Glu410*)	M	XLMTM	Hemi. / c.0 / de novo	Novel
MTM1	g.149764968C>T	3	NM_000252.2: c.70C>T	r.(?)	p.(Arg24*)	M	XLMTM	Hemi. / c.0 / maternal	Laporte et al., 1997
MTM1	g.149826487A>G	11	NM_000252.2: c.1247A>G	r.(?)	p.(His416Arg)	M	XLMTM	Hemi. / c.0/ maternal	Oliveira et al., 2013
COL6A1	g.47409531G>A	10	NM_001848.2: c.868G>A	r.(?)	p.(Gly290Arg)	M	UCMD	Het./ na / unknown	Lampe et al., 2005
COL6A1	g.47409043G>A	9	NM_001848.2: c.850G>A	r.(?)	p.(Gly284Arg)	F	UCMD	Het./ na / unknown.	Lampe et al., 2005
COL6A1	g.47409540G>A	10	NM_001848.2: c.877G>A	r.(?)	p.(Gly293Arg)	M	UCMD/ BM?	Het. / na / de novo	Briñas et al., 2010
COL6A1	g.47409665G>C	10i	NM_001848.2: c.904-1G>C	r.[903_904ins903+1_904-1; 904-1g>c]	p.Gly302Valfs*83	M	BM	Het. / c.1921G>A / unknown.	Novel
COL6A1	g.47421265G>A	30	NM_001848.2: c.1921G>A	r.(?)	p.(Val641Ile)	M	BM	Het. / c.904-1G>C/ unknown	Novel
COL6A1	g.47409881C>T	11i	NM_001848.2: c.930+189C>T	r.930_931ins930+115_930+186	p.Lys310_Gly311ins24	M	BM	Het. / na / unknown	Cummings et al., 2017
COL6A2	g.47537839C>T	12	NM_001849.3: c.1105C>T	r.(?)	p.(Gln369*)	M	UCMD	Hom. / c.1105C>T/ na	Novel
COL6A2	g.47537839C>T	12	NM_001849.3: c.1105C>T	r.(?)	p.(Gln369*)	F	UCMD	Hom. / c.1105C>T/ na	Novel
COL6A2	g.47535800_47535811del	6	NM_001849.3: c.816_827del	r.(?)	p.(Asn272_Pro276delinsLys)	F	UCMD	Het. / na / de novo	Novel
COL6A2	g.47537839C>T	12	NM_001849.3: c.1105C>T	r.(?)	p.(Gln369*)	F	UCMD	Het. / c.1970-9G>A / unknown.	Novel
COL6A2	g.47545690G>A	25i	NM_001849.3: c.1970-9G>A	r.1789_1790ins1790-7_1790-1	p.Gly657Alafs*18	F	UCMD	Het. / c.1105C>T / unknown	Martoni et al., 2009
COL6A2	g.47537368_47537377del	11i	NM_001849.3: c.1053+1_1053+10del	r.1000_1053del	p.Lys335_Gly352del	F	BM	Het. / na / de novo.	Novel
COL6A3	g.238287870T>C	6	NM_004369.3: c.1906A>G	r.(?)	p.(Asn636Asp)	M	UCMD/ BM?	Het. / c.6308A>G /unknown	Novel rs111379417
COL6A3	g.238268006T>C	18	NM_004369.3: c.6308A>G	r.(?)	p.(Lys2103Arg)	M	UCMD/ BM?	Het. / c.1906A>G /unknown	Novel
LMNA	g.156084803A>G	1	NM_170707.3: c.94A>G	r.(?)	p.(Lys32Glu)	M	CMD	Het. / na / unknown	Monges et al., 2011
LMNA	g.156106706_156106715del	7i_8	NM_170707.3: c.1381-6_1384del	r.1381_1386del.	p.Asp461_Gln462del)	M	CMD	Het. / na / unknown	Novel
LMNA	g.156084800G>A	1	NM_170707.3: c.91G>A	r.(?)	p.(Glu31Lys)	M	CMD	Het. / na / unknown	Menezes et al., 2012
POMGNT1	g.46660532G>A	7	NM_017739.3: c.636C>T	r.535_652del	p.Asp179Valfs*23)	F	CMD	Hom. / c.636C>T / na	Oliveira et al., 2008

Gene	Genomic	Exon	Transcript (c.)	r.	p.	Sex	Disease	Zygosity / 2nd allele / origin[a]	Reference
POMGNT1	g.46660532G>A	7	NM_017739.3: c.636C>T	r.535_652 del	p.Asp179Valfs*2 3)	M	MEB	Het. / c.1342G>C / unknown	Oliveira et al., 2008
POMGNT1	g.46658051C>G	16	NM_017739.3: c.1342G>C	r.(?)	p.(Gly448Arg)	M	MEB	Het. / c.636C>T / unknown	Novel rs38683401 4
POMGNT1	g.46661483C>T	6	NM_017739.3: c.534G>A	r.421_534 del)	p.Gly141_Lys17 8del	F	CMD	Hom. / c.534G>A / na	Novel
POMT2	g.77745107T>C	19	NM_013382.5: c.1997A>G	r.(?)	p.(Tyr666Cys)	F	CMD	Hom. / c.1997A>G / na	Godfrey et al., 2007
DNM2	g.10909219C>T	11	NM_001005360.2: c.1393C>T	r.(?)	p.(Arg465Trp)	M	CNM	Het. / na / unknown.	Bitoun et al., 2005
FKRP	g.47259533C>A	4	NM_024301.4: c.826C>A	r.(?)	p.(Leu276Ile)	F	CMD/ LGMD?	Het. / c.1234dup / trans	de Paula et al., 2003
FKRP	g.47259941dup	4	NM_024301.4: c.1234dup	r.(?)	p.(His412Profs*5 2)	F	CMD/ LGMD?	Het. / c.826C>A / trans	Novel
RYR1	g.38991283G>A	46	NM_000540.2: c.7361G>A	r.(?)	p.(Arg2454His)	M	CCD	Het. / c.14468C>T / trans	Barone et al., 1999
RYR1	g.39070725C>T	100	NM_000540.2: c.14468C>T	r.(?)	p.(Thr4823Met)	M	CCD	Het. / c.7361G>A / trans	Negrão et al., 2008
RYR1	g.39075614G>A	102	NM_000540.2: c.14678G>A	r.(?)	p.(Arg4893Gln)	F	CCD	Het. / na / unknown	Davis et al., 2003
RYR1	g.38943556A>T	13	NM_000540.2: c.1342A>T	r.(?)	p.(Ile448Phe)	F	CCD	Het. / na / unknown	Novel
RYR1	g.38989859del	43	NM_000540.2: c.7003del	r.(?)	p.(Leu2335Cysfs *95)	F	CM	Het. / c.7027G>A; c.13672C>T / unknown	Novel
RYR1	g.38989883G>A	43	NM_000540.2: c.7027G>A	r.(?)^ r.spl?	p.(Gly2343Ser)^ p.?	F	CM	Het. / c.7027G>A; c.13672C>T / unknown	Novel rs53659696 9
RYR1	g.39061259C>T	94	NM_000540.2: c.13672C>T	r.(?)	p.(Arg4558Trp)	F	CM	Het. / c.7003del; c.7027G>A / unknown	Novel rs77174160 6
RYR1	g.39009932G>A	67	NM_000540.2: c.10097G>A	r.(?)	p.(Arg3366His)	F	CCD	Het. / c.11798A> G / cis	Duarte et al., 2011
RYR1	g.38931454G>A	2	NM_000540.2: c.115G>A	r.(?)	p.(Glu39Lys)	F	CCD	Het. / na / unknown	Novel rs53920127 6
RYR1	g.39034191A>G	86	NM_000540.2: c.11798A>G	r.(?)	p.(Tyr3933Cys)	F	CCD	Het. / c.10097G> A / cis	Gillies et al., 2008
RYR1	g.39070684_3907 0686del	100	NM_000540.2: c.14427_14429del	r.(?)	p.(Phe4810del)	M	CCD	Het. / na / unknown	Novel
RYR1	g.38959747G>A	26	NM_000540.2: c.3523G>A	r.(?)	p.(Glu1175Lys)	M	CM	Het. / c.4837C>T / trans	Novel rs76974443 8
RYR1	g.38974059C>T	33	NM_000540.2: c.4837C>T	r.(?)	p.(Gln1613*)	M	CM	Het. / c.3523G>A/ trans	Novel
RYR1	g.39062864A>C	95	NM_000540.2: c.13952A>C	r.(?)	p.(His4651Pro)	F	CCD	Het. / na / unknown	Davis et al., 2003
RYR1	g.39052093A>G	90	NM_000540.2: c.12623A>G	r.(?)	p.(Gln4208Arg)	M	CM	Het. / na / maternal	Novel
RYR1	g.38931461T>C	2	NM_000540.2: c.122T>C	r.(?)	p.(Phe41Ser)	M	CM	Het. / c.13673G> A trans	Klein et al., 2011
RYR1	g.39061260G>A	94	NM_000540.2: c.13673G>A	r.(?)	p.(Arg4558Gln)	M	CM	Het. / c.122T>C trans	Kossugue et al., 2007

Footnote: BM- Bethlem myopathy; CCD- central core disease; CM- congenital myopathy; CMD- congenital muscular dystrophy; F- female; Hem.- hemizygous; Het.- heterozygous; Hom.- Homozygous; LGMD- Limb-girdle muscular dystrophy; M- male; MD- muscular dystrophy; MEB- muscle-eye-brain muscular dystrophy; na- not applicable; NEM- nemaline myopathy; UCMD- Ullrich congenital muscular dystrophy; XLMTM- X-linked myotubular myopathy. [a]- cis, trans, unknown or origin (maternal or de novo).

Appendix II.2

Contents

A

Patient data (#0012336)	
Patient ID	███████
Disease	myopathy, myotubular, type 1, X-linked (XL-MTM1)
Phenotype additional	Mild
Reference	Portugal:Porto
Remarks	–
# reported	1
Geographic origin	Portugal
Ethnic origin	–
Gender	F
Inheritance	isolated (sporadic)
Consanguinity	–
CK level	–
MR-IQ	normal
FVC	–
Age diagnosis	30y
Age onset	4–6y
Phenotype onset	Frequent falls
Motor_ability	Walks (only in flat surfaces)
Wheelchair_dependent	–
Age at death	>32y
Protein data	–
Remarks (non public)	Slowly progressive proximal weakness (especially in lower limbs); severe difficulty rising up from the floor and climbing stairs; moderate difficulty rising up from sitting. Positive Gowers sign.

B

Variant data	
Allele	Unknown
Reported pathogenicity	Pathogenic
Concluded pathogenicity	Unknown
Exon	12
DNA change	c.1262G>A (View in UCSC Genome Browser, Ensembl)
Var_pub_as	–
RNA change	r.(?)
Protein change	p.(Arg421Gln)
DB-ID	MTM1_00174
Variant remarks	conserved residue (C. elegans and/or S. cerevisiae R); functional effect (disruption of ligand binding site); de novo, in patient
PolyPhen	1.000
Genet_ori	de novo
Reference	–
Template	DNA
Technique	SEQ
Frequency	–
RE-site	–
DNA_g	–

Figure II.2.1 - Mutational and clinical items in MTM1-LOVD database, example of an entry. Each database entry is subdivided into two tables: patient/clinical items (A) shared among the different LSDBs in the Leiden Muscular Dystrophy pages, and variant data (B).

Table II.2.1 - *MTM1* probes in P309-A1 MLPA kit.

Probe name	Product length (bp)	Probe location	Ligation site (a)	Partial sequence (b)
12646-L14952	189	Intron 1	c.-11+81_ -11+82	GAGGAGGAAGTT-GGGACTCCTTG
12330-L13333	338	Intron 1	c.-11+529_ -11+530	GTTTCTGTGTCA-TACACGAGAGGT
10741-L12837	208	Exon 2	c.-2_-1	TAGAGTTTCCAG-GATGGCTTCTGC
12647-L11339	135	Exon 3	c.83_84	AGATGGAGTCAA-TCGAGATCTCAC
10744-L11340	158	Exon 4	c.178_179	ATGGCCCCATTA-AGGGAAGAGTTT
12327-L13330	171	Exon 5	c.271_272	TGGGTGTGATCT-CGAGAATTGAAA
12650-L13723	274	Exon 6	c.396_397	AGCAGAAGAGAT-ATGTTTGAGATC
12328-L13331	256	Intron 7	c.528+27_ 528+28	CTATTGTCTGGT-ATGTGATGAACC
10748-L12835	177	Exon 8	c.655_656	GAGTTGCAACTT-TTAGGTCCCGAA
10749-L11345	327	Exon 9	c.720_721	ACGGTCATTGTG-CGTTGCAGTCAG
10750-L11346	213	Exon 10	c.906_907	GATGCATATCAT-AACGCCGAACTT
12648-L11347	139	Exon 11	c.1106_1107	TTCAGGGAAGAG-TTCAGTGCTTGT
12331-L13334	365	Exon 12	c.1302_1303	GATGCTGACCGT-TCTCCTATTTTT
12649-L13332	301	Intron 13	c.1467+127_ 1467+128	AGGAATTACATT-AAAGGTATGCAT
10754-L11350	387	Exon 14	c.1616_1617	GAATTACTACAT-TAGATGGAACCC
10755-L11351	310	Exon 15	c.1703_1704	CGACGAATACAT-AAAGCGGCTTGA

(a)- cDNA nucleotide position using reference sequence NM_000252.2; (b)- 24 nucleotides neighboring the ligation site.

The information in this table originally derived from the SALSA MLPA KIT P309-A1 MTM1 documentation (description version 01; 28-01-2009). We included exclusively the information of the probes that recognize *MTM1* sequences; in addition this kit also contains 11 reference probes for the X chromosome, 8 probes for *MTMR1* and 4 probes for other genes located in Xq28 (*IDS*, *FLNA* and *DKC1*).

Table II.2.2 - Pathogenicity assessment of new missense variants.

Gene location	DNA mutation	Protein effect	Grantham distance (a)	PolyPhen V2 score (b)	Phylogenetic conservation (c)	Human Splice Finder (d)	Population screening (e)	Remarks	Conclusion
Exon 5	c.323G>A	p.Gly108Asp	90	0.998	9/10 (*y*)	-	-	-	**Probably pathogenic**
Exon 8	c.595C>A	p.Pro199Thr	38	0.995	10/10	-	0/180	other reported mutations affecting the same residue	**Pathogenic**
Exon 8	c.637C>T	p.Leu213Phe	22	0.523	9/10 (*ce*)	-	-	-	**Probably pathogenic**
Exon 8	c.659G>C	p.Arg220Thr	71	1.000	10/10	-	-	-	**Probably pathogenic**
Exon 11	c.1241T>C	p.Phe414Ser	155	0.998	10/10	-	0/140	-	**Pathogenic**
Exon 11	c.1247A>G	p.His416Arg	29	1.000	10/10	NDSS (75.1%)	-	disruption of ligand binding site	**Probably pathogenic**
Exon 14	**c.1600T>C**	**p.Trp534Arg**	**101**	**1.000**	**10/10**	**DSS: +2%**	**-**	**-**	**Probably pathogenic**

Pathogenicity assessment was performed with the aid of the commercial software Alamut version 2.1 (Interactive Biosoftware, Rouen, France). Results corroborating pathogenicity are underlined. Variants described according to the reference sequence NM_000252.2, using HGVS nomenclature guidelines.

(a) - Grantham's distance (Grantham, 1974), compares wild type and mutated aminoacids considering physical and chemical parameters (volume, weight, polarity, and carbon-composition); range of values: 0-215 (higher value indicates larger difference).

(b) - PolyPhen version 2 (Adzhubei *et al.*, 2010). This software attributes a score to the impact of substitutions on the structure and function of protein; range of values: 0-1.000 (higher value more likely to be pathogenic).

(c) - Phylogenetic conservation analysis: performed using an alignment of myotubularin's orthologues from 10 different species, 10/10 means conserved residue in all species; 9/10 means conserved except in one specie: yeast (*y*) or *C. elegans* (*ce*).

(d) - Impact of variants on splice-sites scores using the Human Splice Finder (Desmet *et al.*, 2009); NDSS - new donor splice site; DSS - donor splice site.

(e) - Population screening performed in ethnically matched controls, N/Nt - indicates the number of control chromosomes with variant / total number of chromosomes tested.

References:

Grantham, R. Amino acid difference formula to help explain protein evolution. Science. 1974; 185:862-4.

Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4):248-9.

Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009; 37(9):e67.

Appendix II.2
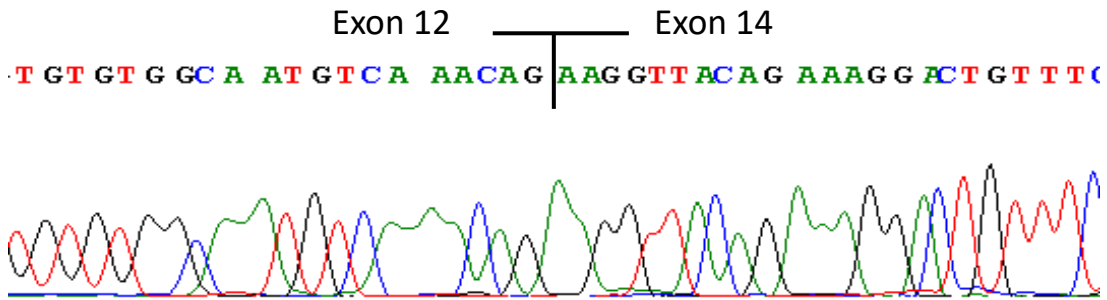
Figure II.2.2 - cDNA studies performed in a XLMTM patient with an AluYa5 insertion. Partial *MTM1* cDNA sequencing result showing skipping of exon 13 in the patient.

Material and Methods

Total RNA was extracted from the patient's affected muscle and reverse transcribed with random primers. Amplification was performed using a forward primer located in exon 12 (G1099A) 5' AAAACCACACCGATGCTGAC 3' and a reverse primer located in exon 14 (G1099B) 5' TCAGTGACCATAAAGAAACAGTCC 3'. PCR products were purified on Microcon YM-100 (Millipore) and sequenced on an ABI 3130 genetic analyzer (Applied Biosystems) with primers used for the PCR amplification using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems).
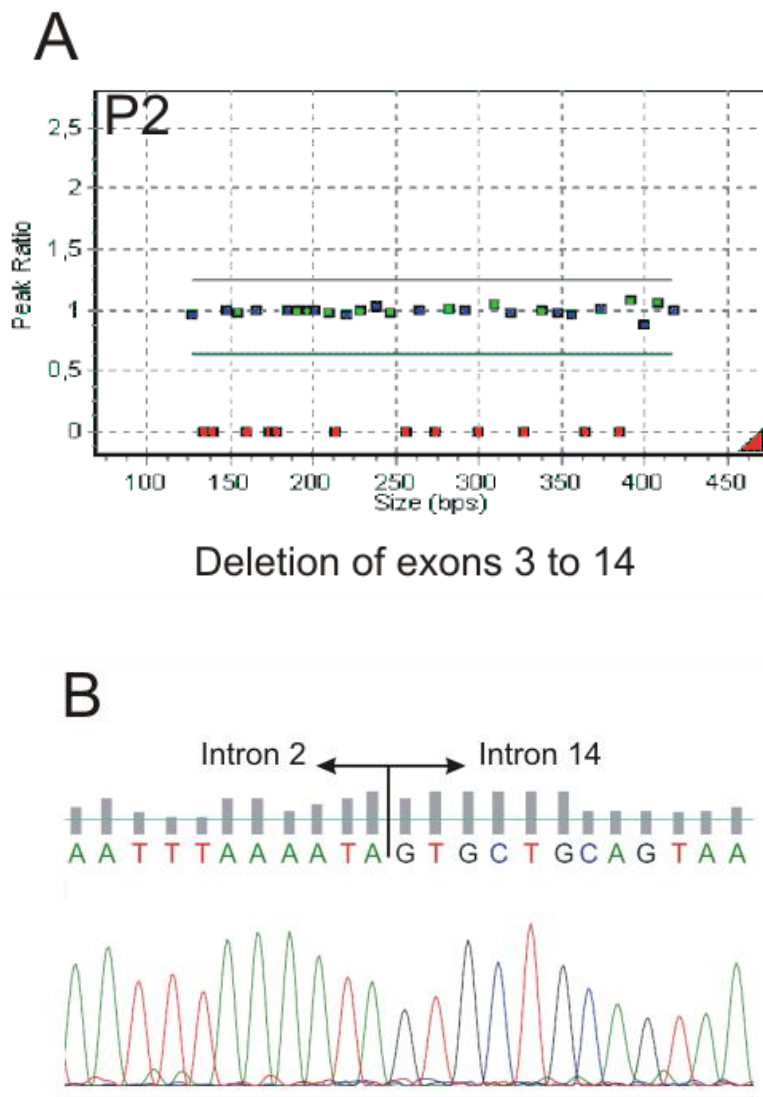
Figure II.2.3 - Multi-exonic *MTM1* deletion identified in patient P2. (A) Deletion of exons 3 to 14 of *MTM1* in patient 2 (P2) confirmed by MLPA technique. Red squares correspond to the deleted regions (peak ratios are equal to 0). (B) Electropherogram of the junction fragment (introns 2 and 14), initially obtained by long-range PCR.

Appendix II.3

Contents

Table II.3.1 - List of ten novel *LAMA2* variants

Mutations [a,b]	Gene location	Type of mutation	Protein impact (prediction)	Bioinformatic analysis	Zygosity / Number of patients
c.396+1G>T	Intron 3	splicing	p.?	HSF: WT=77.1; MUT=50.26 (variation: -34.81%)	Het./n=1
c.2450+4A>G	Intron 17	splicing	p.?	HSF: WT=87.44; MUT=79.1 (variation: -9.54%)	Het./n=1
c.2350dupT	Exon 17	small duplication	p.Tyr784Leufs *3	n.a.	Het./n=1
c.3235T>G	Exon 23	missense	p.Cys1079Gly	PolyPhen-2: 1.000 (probably damaging) SIFT: 0.01 (damaging) MutationTaster: disease causing (prob. 0.999) Grantham distance†: 159 [0-215] EVS: not reported	Het./n=2
c.3520C>T	Exon 24	nonsense	p.Gln1174*	n.a.	Het./n=1
c.5263A>T	Exon 37	nonsense	p.Lys1755*	n.a.	Het./n=1
c.6501C>G	Exon 46	nonsense	p.Tyr2167*	n.a.	Het./n=1
c.6708-1G>T	Intron 47	splicing	p.?	HSF : WT=77.89; MUT=48.95 (variation: -37,16%)	Het./n=1
c.6979G>T	Exon 49	nonsense	p.Gly2327*	n.a.	Hom./n=1
c.8988+1G>A	Intron 63	splicing	p.?	HSF : WT=97.89; MUT=71.05 (variation: -27.41%)	Het./n=1

Legend: a- mutations described according to HGVS nomenclature; b- cDNA reference sequence with accession number NM_000426.3. Het.- heterozygous; Hom.- Homozygous; MUT- Mutation ; n.a.- not applicable; WT- Wild type. EVS- Exome Variant Server (http://evs.gs.washington.edu/EVS/); HSF- Human Splicing Finder (http://www.umd.be/HSF/); MutationTaster (http://www.mutationtaster.org/); PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/); SIFT (http://sift.jcvi.org/).

References :

†- Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974; 185(4154): 862-4.

Exon 56 deletion (homozygous)

Exon 56 deletion (heterozygous)
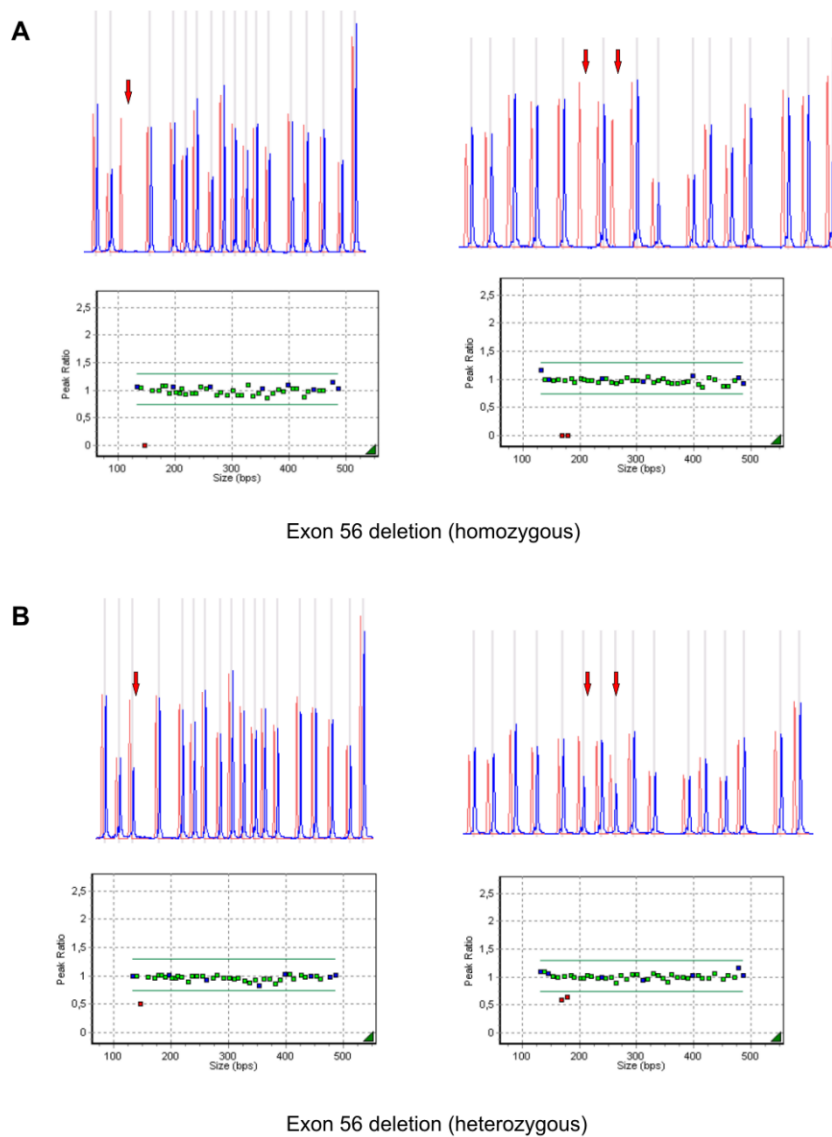
Figure II.3.1 – (A) Homozygous and (B) heterozygous exon 56 deletions detected by MLPA technique.

Appendix II.4

Contents

Data II.4.1 - Methods used for *LAMA2* genotyping

Group I- Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto Magalhães, Centro Hospitalar do Porto, Porto, Portugal.

LAMA2 gene analysis was performed by Sanger sequencing and Multiplex Ligation-dependent Probe Amplification (MLPA) as we previously reported in the literature (Oliveira et al., 2008; Oliveira et al., 2014).

MLPA analysis - Screening for deletions and duplications in *LAMA2* gene was performed using two sets of probe mixes (P391-A1 and P392-A1) from MRC-Holland (Amsterdam, the Netherlands). These probe mixes contain at least one probe for each exon of the gene except for exons 18, 44 and 48. A total of 150 ng of gDNA was used for each patient and normal controls samples. Amplification products were subsequently separated by capillary electrophoresis on an ABI 3130xl genetic analyser (Applied Biosystems, Foster City, CA). Data analysis was conducted using GeneMarker Software V1.5 (SoftGenetics LLC, State College, PA). Population normalization method was selected, and data was plotted using probe ratio.

Sanger sequencing - All 65 exons of *LAMA2* were amplified by Polymerase Chain Reaction (PCR) using intronic M13-tailed primers. Amplicons were purified using Illustra® ExoProStar 1-Step Kit (GE Healthcare, Buckinghamshire, UK) and sequenced with M13 universal primers and BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA). Products were resolved on an ABI3130xl Genetic Analyzer (Applied Biosystems). Sequencing analysis was aided by SeqScape V2.5 software (Thermo Fisher Scientific).

Group II- PreventionGenetics, Marshfield, Wisconsin, USA.

Sanger sequencing – DNA was extracted from the patient specimen using the 5 Prime ArchivePure DNA Blood Kit. PCR was used to amplify all exons of the *LAMA2* gene. After purification of the PCR products, cycle sequencing is carried out using the ABI Big Dye Terminator v.3.0 kit. PCR products are resolved by electrophoresis on an ABI 3730xl capillary sequencer.

Next Generation Sequencing (NGS) of *LAMA2* via single gene or NGS gene panels - A combination of NGS and Sanger sequencing technologies is used to cover the full coding region of the *LAMA2* gene plus ~20 bases of non-coding DNA flanking each exon. For NGS, patient DNA corresponding to these regions is captured using an optimized set of DNA hybridization probes. Captured DNA is sequenced using Illumina's Reversible Dye Terminator (RDT) platform (Illumina, San Diego, CA, USA). Regions with insufficient coverage by NGS are covered by Sanger sequencing. All pathogenic, likely pathogenic, variants of uncertain significance, and suspect NGS variant calls are confirmed by Sanger sequencing.

PreventionGenetics' gene-centric custom designed array Comparative Genomic Hybridization (aCGH) - PreventionGenetics' aCGH is designed with very high-density probe coverage within each gene. The average probe spacing within each exon is 10 bp with a total of 141,549 overlapping probes covering all targeted exons, while the average probe spacing within each intron, 5`UTR and 3`UTR is 25 bp with a total of 408,546 overlapping probes covering all targeted introns, 5`UTR and 3`UTR. DNA was extracted from the patient specimen using the 5 Prime ArchivePure DNA Blood Kit. Equal amounts of genomic DNA from the patient and a gender matched reference sample are amplified and labeled with Cy3 and Cy5 dyes, respectively. Each purified labeled product is combined with the same amount of purified labeled product of its matched reference sample along with a unique sample tracking control, to prevent any sample cross contamination. The combined sample is allowed to hybridize for at least 40 hours at 42ºC. Arrays are then washed and scanned immediately with 2.5 µM resolution.

Group III- MGZ Center of Medical Genetics, Munich, Germany.

Sanger sequencing - After DNA extraction from EDTA blood, standard PCR of all coding exons of the LAMA2 gene was performed. PCR was purified using Millipores 96-plates and Seq96-plates. Direct sequencing was done with ABI PRISM 3100 Avant (Big Dye 1.1). All coding exons and the flanking regions (minimum 30 bp 5´ and 3´ of each exon) of the *LAMA2* gene

were analyzed by direct sequencing. For sequence analysis the Mutation surveyor 3.1 (SoftGenetics) software was used.

NGS - DNA was obtained from peripheral blood of all patients. For panel enrichment approximately 1.5 µg genomic DNA was required. Targeted enrichment was performed with a Custom SureSelectXT Kit (Agilent), which targets the complete coding sequences of 1800 genes associated with rare diseases. Sequencing was carried out on an Illumina NextSeq 500 instrument as 150 bp paired-end runs with V2 chemistry. Reads were aligned to the human reference genome (GRCh37/hg19) using BWA (v 0.7.8-r455) with standard parameters. Duplicate reads and reads that did not map unambiguously were removed. The percentage of reads overlapping targeted regions and coverage statistics of targeted regions were calculated using Shell scripts. Calling of single-nucleotide variants and small insertions and deletions (INDELs) for the 1800 genes associated to rare disease was conducted using SAMtools (v1.1). We used the following parameters: a maximum read depth of 10000 (parameter -d), a maximum per sample depth of 10000 for INDEL calling (parameter -L), adjustment of mapping quality (parameter -C) and recalculation of per-Base Alignment Quality (parameter -E). Additionally, we required putative SNVs to fulfill the following criteria: a minimum of 20% of reads showing the variant base and the variant base is indicated by reads coming from different strands. For INDELs we required that at least 15% of reads covering this position indicate the INDEL. Variant annotation was performed with snpEff (v 4.0e) and Alamut-Batch (v 1.3.1) based on the RefSeq database. Only variants (SNVs/small INDELs) in the coding region and the flanking intronic regions (±15 bp) were evaluated).

Data II.4.2 - Characterization of the *LAMA2:* c.819+2T>C variant at the RNA level

Total RNA was extracted from a fragment of muscle biopsy (patient #102735 and a control sample) using the PerfectPure RNA Fibrous Tissue kit (5 PRIME, Germany). RNA was converted into cDNA using the High Capacity cDNA Reverse Transcription kit (Thermo Fisher Scientific, Waltham, MA, USA). For the analysis of *LAMA2* transcripts, cDNA PCR amplification was performed using Ranger DNA polymerase (Bioline, Tauton, MA, USA) and with primers binding to the regions of exons 4 and 11:

Forward: 5'- GGTGTTCCAGATCGCGTATG-3'

Reverse: 5'-AACAGAAACACTCATCGCAGCC-3'

PCR products were visualized in a 2% agarose gel (Supp. Figure S1). After PCR purification with Illustra® ExoProStar 1-Step Kit (GE Healthcare, Buckinghamshire, UK) sequencing was performed with BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA), and forward or reverse primer in separate reactions. Sanger sequencing products were resolved and detected in a 3130xl sequencer (Applied Biosystems, Thermo Fisher Scientific). As compared with control sample two smaller PCR products were identified in the patient. These bands correspond to aberrant transcripts due to skipping of exon 5 (in-frame deletion) and of exons 5 to 7 (out-of-frame deletion) (Supp. Figure S2).
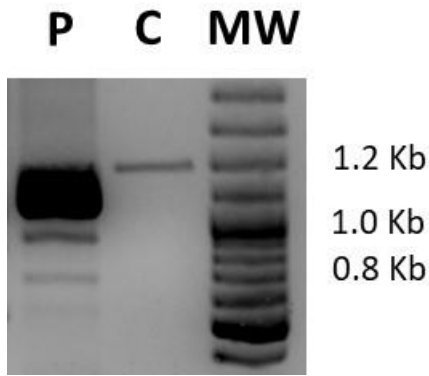
Figure II.4.1 – Evaluation of PCR products obtained from the amplification of cDNA samples (P‑ patient ID# 102735; C‑ control), corresponding of regions corresponding to exons 4 to 11 of LAMA2. MW‑ molecular weight marker.
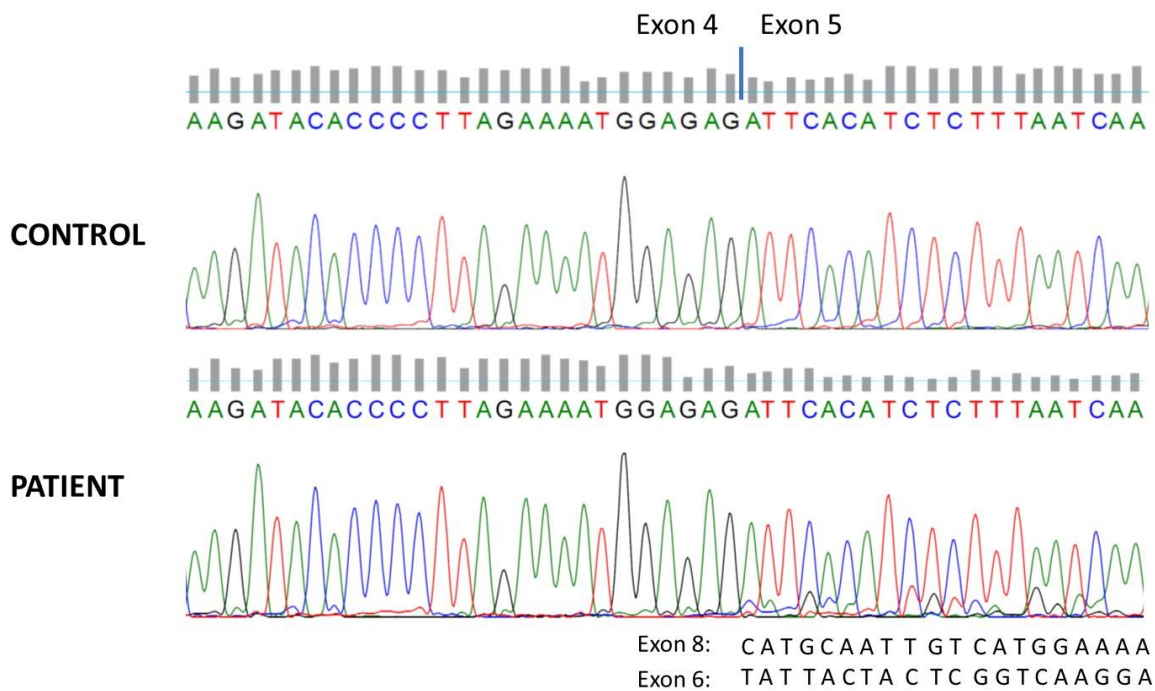
Figure II.4.2 – Sanger sequencing electropherogram of cDNA amplification products for *LAMA2* transcripts from Control and Patient ID# 102735 as in Figure II.4.1. As compared with control, patient's electropherogram revealed additional peaks, with lower intensity, corresponding to skipping of exon 5 and also of exons 5 to 7.
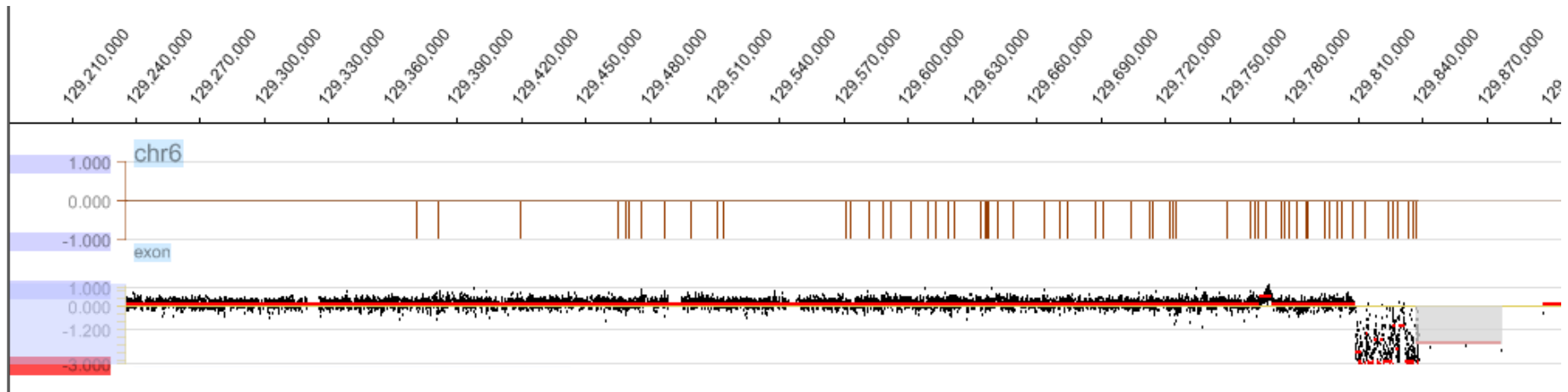
Figure II.4.3 - Large homozygous deletion in *LAMA2* gene detected in a MDC1A patient by array-CGH.

Magnification of *LAMA2 locus* of patient #102475 showing reduced signal in array's probes located in the 3' end (right side). It corresponds to a deletion of positions chr6(hg19):129,808,499 to chr6(hg19):129,876,774 encompassing exons 56 to 65.

Data II.4.3 - Bioinformatic tools used to assess missense variants and splicing

Splicing predictors used to evaluate variants

Human Splice Finder

Website:	http://umd.be/HSF3/
Version:	3.1
Reference:	[Desmet et al., 2009]

MaxEntScan

Website:	http://genes.mit.edu/burgelab/maxent/ Xmaxentscan_scoreseq.html
Version:	not available
Reference:	[Yeo et al., 2004]

NNSPLICE

Used in Alamut Visual Software v2.11
(Interactive Biosoftware, Rouen, France)

Reference: [Reese et al., 1997]

GeneSplicer

Used in Alamut Visual Software v2.11
(Interactive Biosoftware, Rouen, France)

Reference: [Pertea et al., 2001]

Tolerance predictors used to classify missense variants

Combined Annotation Dependent Depletion (CADD):

Website:	http://cadd.gs.washington.edu/
Version:	1.3
Reference:	[Kircher et al., 2014]

Mutation Assessor 3:

Website:	http://mutationassessor.org/r3/
Version:	release 3
Reference:	[Reva et al., 2011]

MutationTaster 2:

       Website: http://www.mutationtaster.org/

       Version: build NCBI 37 / Ensembl 69

       Reference: [Schwarz et al., 2014]

MutPred 2:

       Website: http://mutpred.mutdb.org/

       Version: not available

       Reference: [Pejaver et al., 2017]

Polymorphism Phenotyping v2 (PolyPhen-2):

       Website: http://genetics.bwh.harvard.edu/pph2/

       Version: 2.2.2 (build r394)

       Reference: [Adzhubei et al., 2010]

PON-P2:

       Website: http://structure.bmc.lu.se/PON-P2/

       Version: not available

       Reference: [Niroula et al., 2015]

SIFT:

       Website: http://sift.jcvi.org/

       Version: JCVI-SIFT v1.03

       Reference: [Kumar et al., 2009]

Protein Variation Effect Analyzer (PROVEAN):

       Website: http://provean.jcvi.org/index.php

       Version: 1.1.3

       Reference: [Choi et al., 2012]

UMD-Predictor:

       Website: http://umd-predictor.eu/

       Version: not available

       Reference: [Salgado et al., 2016]

References:

- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248-9.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012;7(10):e46688.
- Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.
- Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310-315.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.
- Niroula A, Urolagin S, Vihinen M. PON-P2: prediction method for fast and reliable identification of harmful variants. PLoS One. 2015 Feb 3;10(2):e0117380.
- Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H, Mort M, Cooper DN, Sebat J, Iakoucheva LM, Mooney SD, Radivojac P. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. bioRxiv 134981; doi: https://doi.org/10.1101/134981.
- Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. 2001;29(5):1185-90.
- Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie.
- Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011 Sep 1;39(17):e118.
- Salgado D, Desvignes JP, Rai G, Blanchard A, Miltgen M, Pinard A, Lévy N, Collod-Béroud G, Béroud C. UMD-Predictor: A High-Throughput Sequencing Compliant System for Pathogenicity Prediction of any Human cDNA Substitution. Hum Mutat. 2016 May;37(5):439-46.
- Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014 Apr;11(4):361-2.
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2-3):377-94.

Table II.4.1 - Tolerance predictors method's performance assessment to predict the impact of *LAMA2* missense variants. Performance measures determined as detailed in http://tiny.cc/LAMA2_2018 and as suggested by Vihinen & Niroula (2016).

Prediction tool	Range / cut-offs used	Performance measures for binary classifiers					
		PPV	NPV	Sensitivity	Specificity	Accuracy	MCC
CADD	PHRED score: > 20.0 deleterious	0.64	1.00	1.00	0.27	0.69	0.38
Mutation Assessor 3	Functional (high, medium), predicted non-functional (low, neutral)	0.68	0.71	0.86	0.45	0.69	0.32
MutationTaster 2	Disease causing vs Polymorphism; range of the probability score: 0 to 1	0.72	0.92	0.97	0.50	0.76	0.47
MutPred2	Probability score: > 0.50 pathogenic	0.76	1.00	1.00	0.59	0.82	0.56
	Probability score: > 0.68 pathogenic (10% FPR)	0.86	0.82	0.86	0.82	0.84	0.56
	Probability score: > 0.80 pathogenic (5% FPR)	0.88	0.73	0.76	0.86	0.80	0.52
PolyPhen-2 (HumVar)	Benign, possibly damaging (<20% FPR) or probably damaging (<10% FPR)	0.71	1.00	1.00	0.45	0.76	0.49
PON-P2 (*)	Pathogenic: score > 0.5), neutral: score < 0.5) or unknown (SD >c0.05)	0.79	0.83	0.95	0.50	0.80	0.48
PROVEAN	Score: <-2.5 deleterious, >-2.5 neutral	0.71	1.00	0.97	0.45	0.75	0.44
SIFT	Score: > 0.05 tolerated, < 0.05 damaging	0.74	1.00	1.00	0.55	0.80	0.54
UMD-Predictor	UMD score: <50 polymorphism, [50; 64] probable poly., [65; 74] probably path., >74 pathogenic; score range: 0 to 100	0.74	1.00	1.00	0.55	0.80	0.54

Footnote: NPV- negative predictive value, Path.- pathogenic, Poly.- polymorphism, PPV- positive predictive value, MCC- Matthews correlation coefficient, FPR- false positive rate. (*)- As a total of 21 variants (31%) were classified as unknown by PON-P2 these were not included to calculate performance measure.

Table II.4.2 - Bioinformatic evaluation of variants of missense type and/or predicted to affect splicing

DNA Variant (NM_000426.3)	RNA Variant	Predicted effect on Protein	Tolerance predictors				Splicing predictors			
			MutPred2	PolyPhen-2	SIFT	UMD-predictor	HSF	MaxEnt Scan	NNSplice	GeneSplicer
c.112G>A	r.(?)^ r.(spl?)	p.(Gly38Ser)^ p.(?)	Non-pathogenic (score: 0.176)	Probably damaging (score: 0.998)	Damaging (score: 0.049)	Pathogenic (score: 100)	WT donor SS: 75.34 Mut: 64.76	WT donor SS: 6.62 Mut: -2.3	WT donor SS: 0.88 Mut: 0.02	WT donor SS: 8.68 Mut: 2.82
c.245A>T	r.(?)	p.(Gln82Leu)	Non-pathogenic (score: 0.626)	Probably damaging (score: 0.994)	Damaging (score: 0.002)	Pathogenic (score: 90)	NEP	NEP	NEP	NEP
c.437C>T	r.(?)	p.(Ser146Phe)	Pathogenic (score: 0.901)	Probably damaging (score: 0.999)	Damaging (score: 0.001)	Pathogenic (score: 93)	NEP	NEP	NEP	NEP
c.745C>T	r.(?)	p.(Arg249Cys)	Pathogenic (score: 0.822)	Probably damaging (score: 0.999)	Damaging (score: 0)	Pathogenic (score: 99)	NEP	NEP	NEP	NEP
c.818G>A	r.(?)	p.(Arg273Lys)	Non-pathogenic (score: 0.488)	Probably damaging (score: 0.992)	Damaging (score: 0)	Probably pathogenic (score: 72)	NEP	NEP	NEP	NEP
c.1326T>G	r.(?)	p.(Cys442Trp)	Pathogenic (score: 0.940)	Probably damaging (score: 0.999)	Damaging (score: 0)	Pathogenic (score: 100)	NEP	NEP	NEP	NEP
c.1609-41_1609-7inv	r.(spl?)	p.(?)	NA	NA	NA	NA	WT acceptor SS: 81.5 Mut: 68.41	WT acceptor SS: 6.92 Mut: 1.55	WT acceptor SS: 0.60 Mut: 0	WT acceptor SS: 8.36 Mut: 0
c.2370T>A	r.(?)^ r.(spl?)	p.(790=)^p.(?)	NA	NA	NA	NA	New acceptor SS: 88.37	New acceptor SS: 7.52	NEP	New acceptor SS: 3.39
c.3235T>G	r.(?)	p.(Cys1079Gly)	Pathogenic (score: 0.901)	Probably damaging (score: 0.999)	Damaging (score: 0)	Pathogenic (score: 100)	NEP	NEP	NEP	NEP
c.4523G>A	r.(?)^r.(spl?)	p.(Arg1508Lys)^p.(?)	Non-pathogenic (score: 0.442)	Probably damaging (score: 0.892)	Tolerated (score: 0.078)	Pathogenic (score: 100)	WT donor SS: 75.89 Mut: 65.32	WT donor SS: 7.87 Mut: -0.65	WT donor SS: 0.84 Mut: 0	WT donor SS: 3.76 Mut: 0
c.4654G>A	r.(?)	p.(Ala1552Thr)	Non-pathogenic (score: 0.205)	Possibly damaging (score: 0.543)	Tolerated (score: 0.070)	Pathogenic (score: 78)	NEP	NEP	NEP	NEP
c.6548T>G	r.(?)	p.(Leu2183Arg)	Pathogenic (score: 0.927)	Probably damaging (score: 1)	Damaging (score: 0)	Pathogenic (score: 93)	NEP	NEP	NEP	NEP
c.6707G>A	r.(?)^ r.(spl?)	p.(Arg2236Lys)^p.(?)	Pathogenic (score: 0.864)	Probably damaging (score: 0.996)	Damaging (score: 0)	Pathogenic (score: 100)	WT donor SS: 84.86 Mut: 74.28	WT donor SS: 9.24 Mut: 2.93	WT donor SS: 0.98 Mut: 0.14	NEP
c.7571A>T	r.(?)^ r.(spl?)	p.(Glu2524Val)^p.(?)	Non-pathogenic (score: 0.489)	Probably damaging (score: 0.974)	Damaging (score: 0.001)	Pathogenic (score: 100)	WT donor SS: 84.88 Mut: 80.13	WT donor SS: 6.36 Mut: 2.44	WT donor SS: 0.60 Mut: 0.01	NEP

Footnote: Mut- Mutated; NA- not applicable; NEP- no effect predicted; SS- splice-site; WT- wild-type.

Data II.4.4 - Screening of c.2461A>C variant by restriction fragment length analysis.

The screening of the c.2461A>C variant was performed by polymerase chain reaction (PCR) of exon 18 of the *LAMA2* gene, followed by restriction fragment length analysis (RFLA) using *Hpy*CH4III restriction enzyme. As this substitution adds a new restriction site for this enzyme, specific primers were designed to encompass this region and another restriction site present in the "wild-type" sequence as follows:

Forward:	5'-TGAGAATGACCAGCCTGTACT-3'
Reverse:	5'-CCCAAGGTAGAAATCATTTTGCA-3'

PCR experiments were performed in standard conditions using a Master Mix (Promega, Madison, WI, USA). In each batch a normal control and two positive controls (one homozygous for the c.2461A>C variant and another heterozygous) were included. The obtained PCR products were digested with enzyme *Hpy*CH4III (overnight, 37°C). The generated fragments were analyzed by electrophoresis using an agarose gel (2% TAE) and the band pattern compared with the respective controls (Supp. Figure S4).
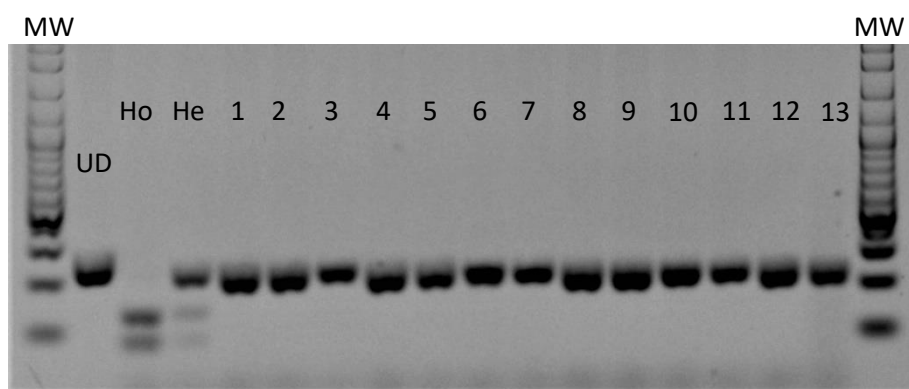
Figure II.4.4 - Representative image of the variant c.2461A>C screening by RFLA, showing the typical band pattern. MM- molecular marker, UD- undigested sample, HO- control sample homozygous for the c.2461A>C variant, HE- heterozygous control. The numbers (1-13) represent different tested samples.

Appendix II.4

Data II.4.5 - Genetic analysis of patient #103207 (#15) with a muscular dystrophy and intellectual disability by whole-exome sequencing

Whole-exome sequencing (WES) was performed in patient #103207 using the Ion AmpliSeq Exome kit (Thermofisher Scientific, Waltham, MA, USA) as described in previous reports (Oliveira et al., 2015; Oliveira et al., 2017). Exome sequencing generated a total of ~50,000,000 mapped reads, which translates in a mean read depth (target base coverage) of 162.4 (fold), with 95.8% of these regions covered by at least 20 reads. A total of 52,182 variants were identified, comprising 49,253 single and multiple nucleotide variants, and 2,929 insertion/deletions (indels). Variants with frequency above 1% in the population variant databases (dbSNP, 1000 genome project, exome sequencing project and exome aggregation consortium) were excluded. As a first line strategy we focused on genes known to be related with muscle diseases (n=146, Oliveira et al., 2017). Within the list of candidates, two heterozygous variants in *LAMA2* were identified: one pathogenic frameshift duplication (NM_000426.3:c.1854_1861dup) and one novel likely pathogenic variant affecting a canonical donor splice site (NM_000426.3:c.819+2T>C) (Supp. Figure S5 A and B respectively). Variants were confirmed by Sanger sequencing as detailed above.

References:

- Oliveira J, Negrão L, Fineza I, Taipa R, Melo-Pires M, Fortuna AM, Gonçalves AR, Froufe H, Egas C, Santos R, Sousa M. New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing. J Hum Genet. 2015 Jun;60(6):305-12.

- Oliveira J, Martins M, Pinto Leite R, Sousa M, Santos R. The new neuromuscular disease related with defects in the ASC-1 complex: report of a second case confirms ASCC1 involvement. Clin Genet. 2017 Oct;92(4):434-439.
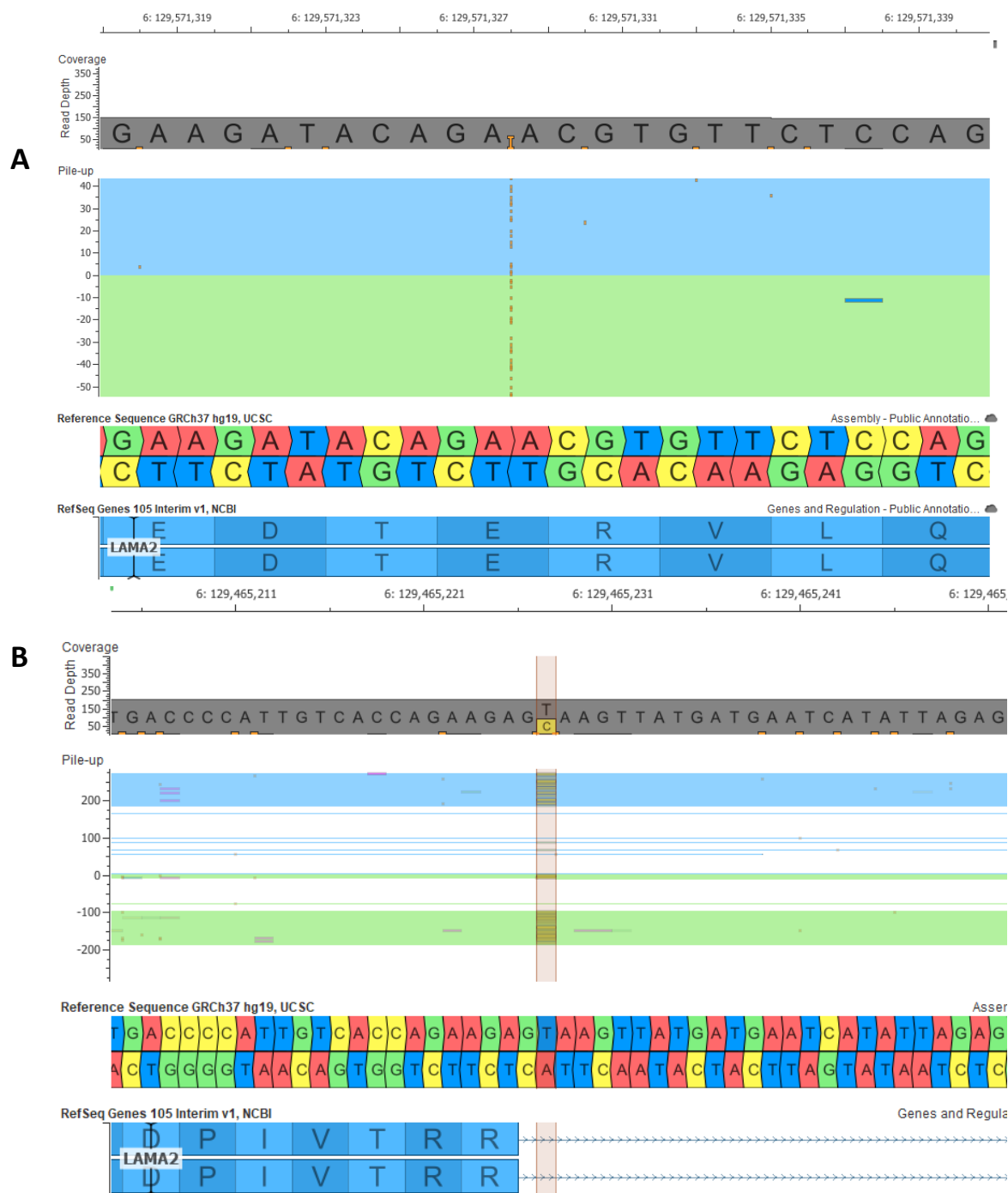
Figure II.4.5 - Visual inspection of two heterozygous *LAMA2* variants in the WES BAM file from patient #103207. (A) NC_000006.11:g.129571328_129571335dup, identified as an insertion of the sequence ACGTGTTC between positions 6:129571327 and 6:129571328, corresponds to exon 13. (B) NC_000006.11:g.129465227T>C changing the second base (+2 position) of intron 5.

Appendix III.1

Contents

Figure III.1.1 - Genes included in the massive parallel sequencing approach panel known to be implicated in congenital myopathies. The main pathognomonic and diagnostic features of CM subtypes are highlighted. The *loci* not analyzed in this panel are indicated between brackets.

Table III.1.1 - *Loci* included in the massive parallel sequencing gene panel for congenital myopathies**.**

Gene symbol	Number of exons (n)	Locus	Gene (official full name)	cDNA reference sequences accession numbers	Size of target regions (bp)	Target regions covered (design)	Target regions covered (>20x read depth) (*)	Average coverage depth (fold) (*)	CDS size (bp)	CDS covered by design
ACTA1	7	1q42.1	Alpha actin, skeletal muscle	NM_001100.3	2192	88%	74%	255	1134	77%
BIN1	19	2q14	Amphiphysin	NM_139343.2 NM_139346.2	4714	98%	92%	374	1826	100%
CFL2	4	14q12	Cofilin 2 (muscle)	NM_138638.7	3911	85%	79%	282	501	89%
CNTN1	23	12q11-q12	Contactin-1	NM_001843.3	8060	91%	88%	361	3057	91%
DNM2	22	19p13.2	Dynamin 2	NM_001005360.2 NM_001005361.2	6005	96%	95%	405	2613	99%
KBTBD13	1	15q22.31	Kelch repeat and BTB (POZ) domain containing 13	NM_001101362.2	3223	100%	86%	321	1377	100%
KLHL40	6	3p22.1	kelch-like family member 40	NM_152393.3	3019	99%	99%	464	1866	100%
MEGF10	26	5q23.2	Multiple EGF-like-domains 10	NM_032446.2	10828	94%	94%	398	3423	100%
MTM1	15	Xq28	Myotubularin	NM_000252.2	4936	92%	92%	234	1812	100%
MYBPC3	35	11p11.2	Cardiac myosin binding protein-C	NM_000256.3	7875	93%	90%	346	3825	96%
MYH2	42	17p13.1	Myosin, heavy polypeptide 2, skeletal muscle	NM_017534.5 NM_001100112.1	10251	94%	94%	412	5826	95%
MYH7	40	14q12	Myosin, heavy polypeptide 7, cardiac muscle, beta	NM_000257.2	10030	97%	94%	417	5808	98%
NEB	183	2q22	Nebulin	NM_001164507.1 NM_001164508.1	44402	85%	84%	390	25683	85%
RYR1	106	19q13.1	Ryanodine receptor 1 (skeletal)	NM_000540.2	25991	93%	90%	422	15117	93%
SEPN1	13	1p36.13	Selenoprotein N1	NM_020451.2	5632	83%	79%	447	1773	83%
TNNT1	14	19q13.4	Slow troponin T	NM_003283.5	2448	94%	69%	406	837	99%
TPM2	11	9p13	Tropomyosin 2 (beta)	NM_003289.3 NM_213674.1	3124	96%	96%	409	855	87%
TPM3	10	1q21.2	Tropomyosin 3	NM_152263.5 NM_001043352.1	12482	76%	76%	387	855	100%
TRIM32	2	9q33.2	Tripartite motif-containing 32	NM_012210.3	5981	100%	100%	329	1962	100%
TTN	363	2q31	Titin	NM_001267550.2	145523	96%	94%	340	107976	97%

bp- base pair; CDS- coding sequences; (*)- calculated using the amplicons' metrics generated in the first sequencing run.

Table III.1.2 - Primers used for Sanger sequencing, to confirm the presence of the pathogenic variants detected by massive parallel sequencing or to determine their impact at the mRNA level.

Gene / Accession Number	Variant	Region of the gene studied	Primer Sequence (5´-3´)
NEB NM_001271208.1	c.19102-4_19102-10del	Intron 122	F – ACAGACACGCCCCTCTATGTGAC R – GATTGAAGGTCCAGGAAAAAACC
		Exons 121-125 (cDNA)	F – CATGCCAAGAAATCATACGACC R – TTCGGATCATCGTCAACACTTC
	c.19944G>A	Exon 129	F – GGGCCATCTCTTCTGTTTCTCA R – CATGCCCATTATGTTGGTTTGA
		Exons 127-132 (cDNA)	F – AGGATGACCTCAACTGGCTGA R – TCAGTTTGTGGTACAATTCCCG
	c.21076C>T	Exon 140	F – TGTGTTGGGCTCTACAACTCACC R – TCTCCCAAGTTGAAATTCACTCAC
	c.24294_24297dup	Exon 171	F – CAGTGCTGCAGAGATTAGTTCAAAC R - AGAAGCAAAGAAGGGAAATGGG
RYR1 NM_000540.2	c.3800C>G	Exon 28	F - <u>TGTAAAACGACGGCCAGT</u>AAGCATTGAAGTTGAGCTGCC R - <u>CAGGAAACAGCTATGACC</u>GCAGGTTTTCGTAGTCAGGGTC
	c.9157C>T	Exon 61	F - <u>TGTAAAACGACGGCCAGT</u>TCGTCTCCTTGGCCTCCTC R - <u>CAGGAAACAGCTATGACC</u>TCCATGTCTCCAGGTCCAGG
	c.12010C>T	Exon 87	F - <u>TGTAAAACGACGGCCAGT</u>GATCCCTGATCCCTTCTCGG R - <u>CAGGAAACAGCTATGACC</u>GGGCAAGACTTGGAAATGGAC
	c.14643G>A	Exon 101	F - <u>TGTAAAACGACGGCCAGT</u>GTAGAGCCACAGGGACTGAACC R - <u>CAGGAAACAGCTATGACC</u>GACTCAAGTAATCGTCCCGCC
TTN NM_001267550.2	c.52102+5G>A	Intron 273	F – AACCATTTGTCCTGCCTCTGAC R - TGACATGATTTGTGGTTTTCCC
		Exons 272-274 (cDNA)	F – AGAGATGCAATGAACACCTGGTAC R - TCTGAGTCGGGTTTTGTCTTGTC
	c.61679_61680insCC	Exon 304	F – ACCGTGAGAGTAGGCCAAAC R - TTCATTGTTGGCTAAAAGGTT
	c.76109_76110del	Exon 326	F – TGGTGTTGGTGAACCTCTTG R - TTGGTTTGCTCCAAGAAAGG
SEPN1 NM_020451.2	c.1384T>G	Exon 10	F - <u>TGTAAAACGACGGCCAGT</u>GCCGCTTTGATGATGGCTT R - <u>CAGGAAACAGCTATGACC</u>TCAGCACAGAAACCCCCAA

M13 primer sequences are underlined; F- forward; R- reverse.

Table III.1.3 - Sequencing and variant metrics.

EXP	ID	Total reads (n)	On target (%)	Mean coverage of target regions (x)	Bases with Q>20 (%)	Variants (n)			Filtered-in variants (n)			Pathogenic variants			Rare polymorphisms (n)			Unclassifed variants (n)				False positives (n)		
						Total	SNV	INDEL	Total	SNV	INDEL	Total	SNV	INDEL	Total	SNV	INDEL	Total	Mis.	Silent	Intr.	Total	SNV	INDEL
1	C1	769694	98.5	368.3	98.3	246	231	15	28	25	3	1	1	0	6	5	1	7	4	1	2	14	12	2
	C2	704271	98.4	335.7	97.9	236	220	16	29	25	4	2	2	0	5	4	1	8	2	1	5	14	11	3
	P1	767199	98.3	356.9	97.9	157	144	13	20	17	3	1	1	0	4	4	0	4	1	1	2	11	9	2
	P3	855888	98.4	396.7	98.1	286	271	14	25	21	4	1	1	0	4	4	0	5	1	0	4	15	12	3
2	P2	395952	96.2	188.0	90.5	195	182	13	33	28	5	2	2	0	2	0	0	9	1	0	8	20	16	4
	P4	435835	95.2	204.4	91.0	178	161	17	32	23	9	2	1	1	1	0	0	6	3	0	3	23	15	8
	P5	439667	96.2	210.4	91.0	185	174	11	32	26	6	1	0	1	1	0	0	4	0	1	3	26	21	5
	P6	443067	96.6	211.9	91.0	214	196	17	38	32	6	1	0	1	7	6	1	8	2	1	5	22	18	4
	P7	443489	96.9	212.9	91.0	220	198	22	28	20	8	3	2	1	1	0	0	2	0	0	2	22	15	7
	P8	453240	96.5	216.2	91.1	210	189	21	30	20	10	0	0	0	4	0	0	5	1	0	4	21	14	7
	P9	387413	96.4	185.4	91.3	193	177	16	36	28	8	0	0	0	6	0	0	7	4	1	2	23	15	8
	P10	422698	96.5	201.8	91.2	217	205	12	22	18	4	0	0	0	2	0	0	2	1	0	1	18	14	4
					Total:	2537			353			14	10	4	43	23	3	67	20	6	41	229	172	57

EXP- experiment; ID- sample identification; INDEL- Insertions/deletions; Intr.- intron; Mis.- Missense; SNV- single nucleotide variants.

A

B

C

Ampliseq forward primer:

5' TGGTGGCTATGGCCCTCT 3'
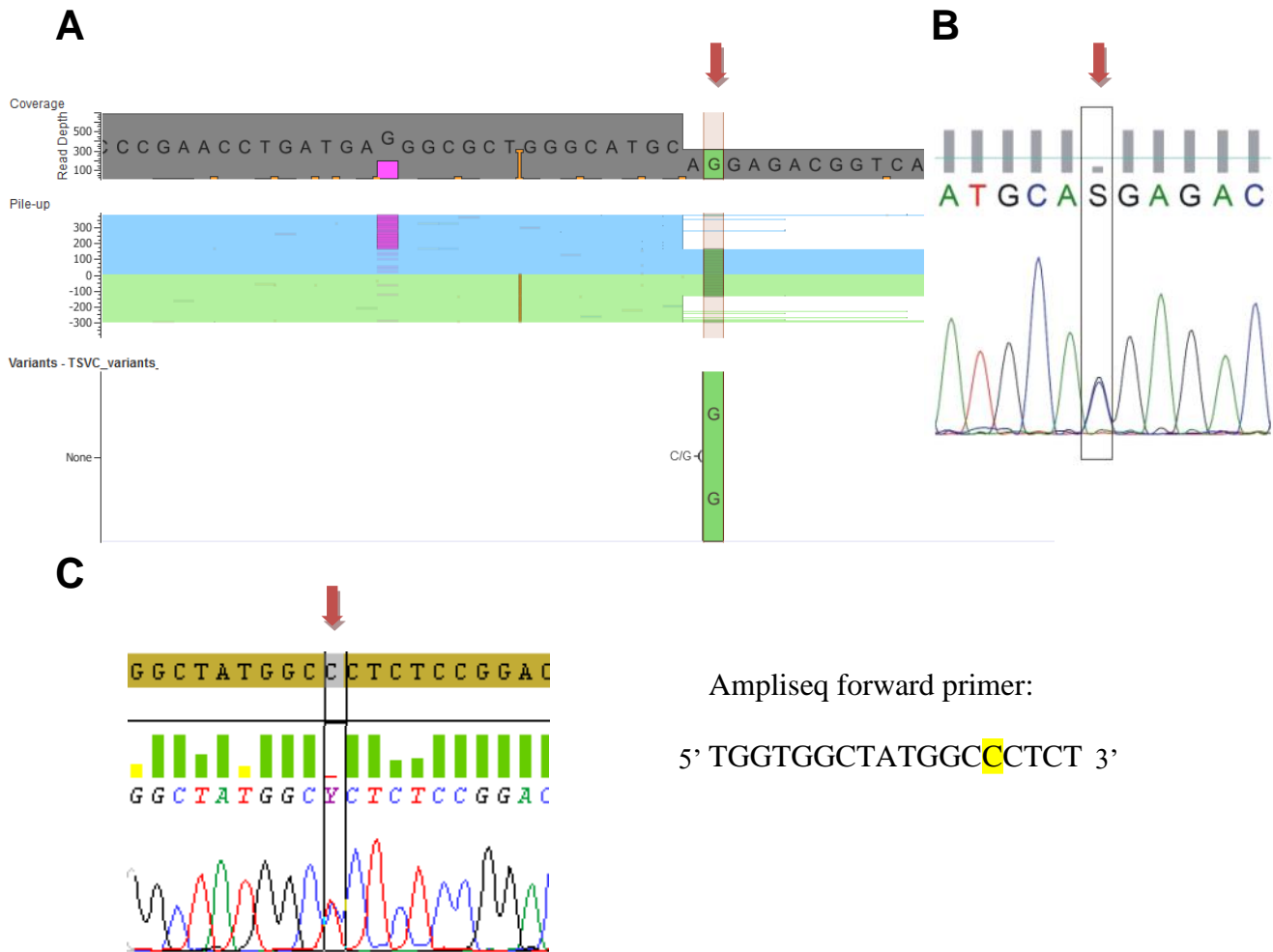
Figure III.1.2 - Examples of potential pitfall found during the application of the massive parallel sequencing (MPS) gene panel for congenital myopathies: the zygosity of c.6612C>G variant was incorrectly established. (A) Visualization of BAM (top) and VCF (bottom) files for the positive control C2. Arrow indicates the position of the genomic coordinate chr19:38986918, where the mutation is shown as being homozygous; all sequencing reads have a G instead of a C. (B) Sanger sequencing electropherogram showing the c.6612C>G mutation in heterozygosity described in a previous publication (Duarte et al., 2011). (C) On the left is the sequencing electropherogram of the region recognized by the MPS forward primer (sequence shown on the right). A rare heterozygous SNP (c.6549-51C>T, highlighted) with a reported frequency of 1/13005, prevents the amplification of the normal allele and causes a PCR allele dropout, thereby explaining the apparent genotype discrepancy.

A

Genotype	Ref	Type	Genes	dbSNP	DGV	MAF	EMAF	AMAF	GMAF	UCSC
G/G	GC	INDEL	BIN1							
AT/A	AT	INDEL	NEB	rs369188503 ...(2)		0.436	0.3824	0.3919	0.4523	YES
GTAGACAA/G	GTAGACAA	INDEL	NEB	rs145127681		0.011				YES

B
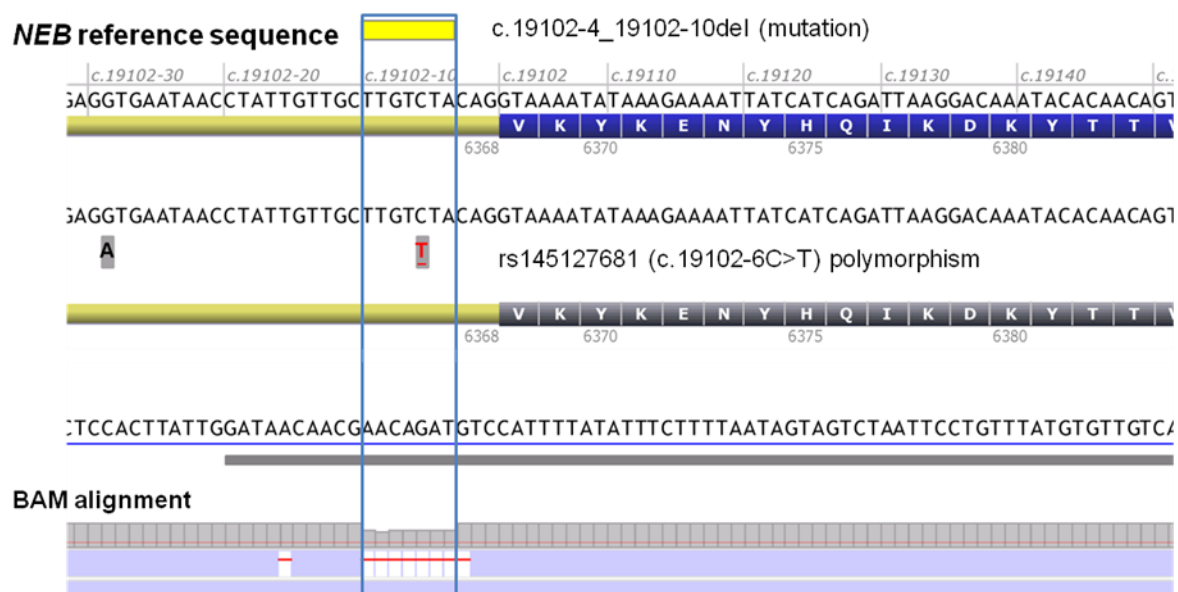
Figure III.1.3 - Example of a potential pitfall found during the application of the massive parallel sequencing (MPS) gene panel for congenital myopathies: an intronic mutation was initially filtered-out during data analysis. (A) Ion reporter caption showing the chr2:g.152417626_152417632del (c.19102-10_19102-4del) variant (blue box) in patient P7, annotated with the rs number "rs1415127681" – an SNP in the UCSC database, with a minor allele frequency of 0.011. (B) Partial *NEB* gene reference sequence (end of intron 122 and initial part of exon123), known SNPs and BAM alignment visualization in Alamut software. As shown in the midlle section, rs145127681 is actually associated with a known SNP (c.19102-6C>T) and not with the intronic deletion (c.19102-10_19102-4del) as suggested by Ion Reporter annotation, as shown in (A). The genomic coordinates of this 7 bp deletion encompass the SNP at c.19102-6. The incorrect annotation in Ion Reporter software filtered-out this pathogenic intronic deletion, which affects the acceptor splite of intron 122.
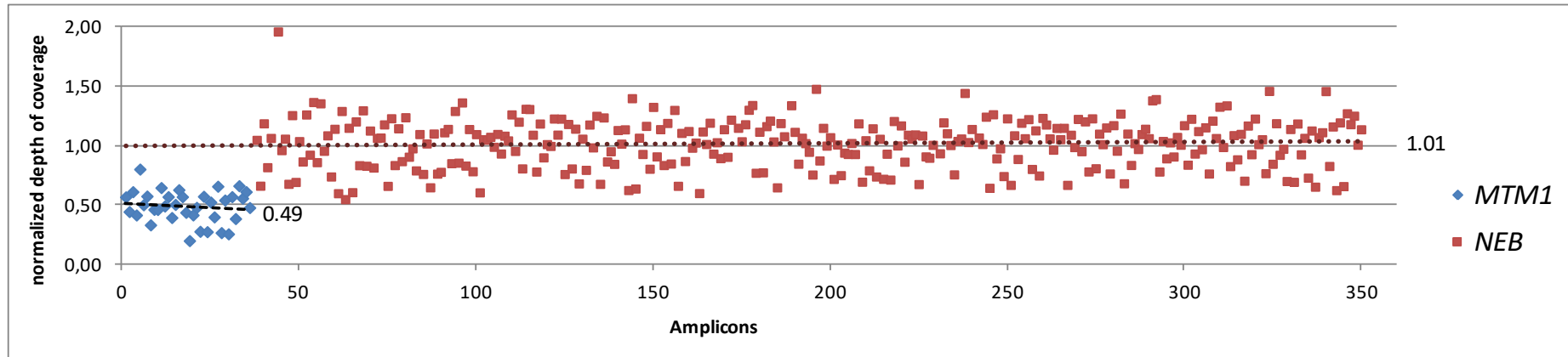
Figure III.1.4 - Dosage analysis of *NEB* and *MTM1* genes in P3, using data derived from the massive parallel sequencing approach. The analysis was done to exclude the presence of a large deletion spanning the entire *NEB* gene in patient P3. This figure shows the ratio of normalized (depth) coverage for *NEB* and *MTM1* amplicons in patient P3. Although the library preparation based on multiplex PCR may introduce some bias, this data analysis suggests that patient P3 has two copies of the *NEB* gene (average ratio of 1.01) and one copy of *MTM1* gene (average ratio of 0.49) which is located on the X chromosome. The analysis was performed as follows: 1) amplicon intra-normalization, which consisted in dividing the coverage of individual amplicons by the sum of all amplicons obtained for the patient; 2) the value obtained in this manner was divided by the average of intra-normalized values from three individuals analyzed in the same experiment. Considering the *MTM1* amplicons used to demonstrate the feasibility of this approach, the normalization was performed using only female individuals as references.

A

B
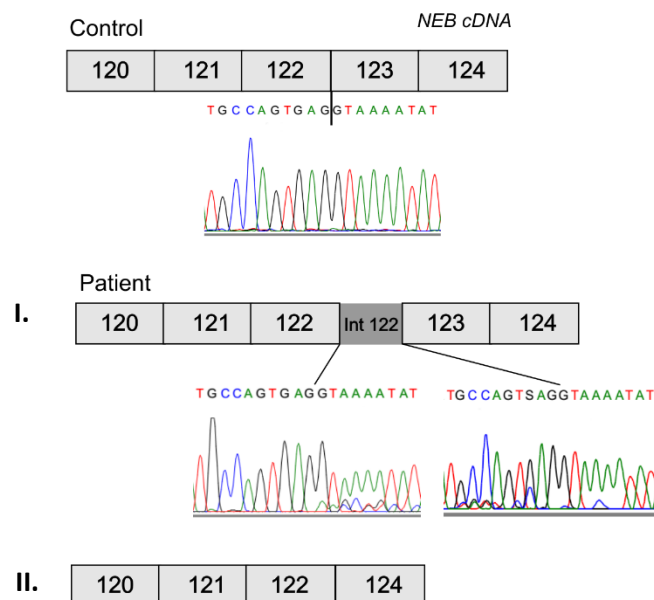
Figure III.1.5 - Characterization of the c.19102-10_19102-4del mutation at the mRNA level. (A) Bioinformatic analysis using Alamut Visual v2.4 software (Interactive Biosoftware, Rouen, France), shows that the deletion of 7 nucleotides near the acceptor splice-site (intron 122 of *NEB*) causes a significant reduction of its probabilistic value, as assessed by different algorithms (MaxEnt: -68.6%, NNSPLICE: -71.6%, HSF: -2.9%). (B) cDNA analysis proved the effect of this variant on splicing. Besides the wild-type transcript, two aberrant transcripts were identified in patient P7, as a result of the small intronic deletion. These transcripts correspond to the total retention of intron 122 (I.) and the skipping of exon 123 (II.).

Appendix III.2

Contents

Table III.2.1 - List of 108 candidate genes implicated in hereditary myopathies.

Gene Symbol	Full gene name	Locus
ABHD5	Abhydrolase domain containing 5	3p25.3-p24.3
ACADVL	Acyl-Coenzyme A dehydrogenase, very long chain	17p13
ACTA1	Alpha actin, skeletal muscle	1q42.1
ACVR1	Activin A receptor, type II-like kinase 2	2q23-q24
AGL	Amylo-1,6-glucosidase, 4-alpha-glucanotransferase	1p21
ALG13	UDP-N-acetylglucosami-nyltransferase	Xq23
ANO5	Anoctamin 5	11p14-12
B3GALNT2	Beta-1,3-N-acetylgalacto-saminyltransferase 2	1q42.3
B3GNT1	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyl-transferase 1	11q13.2
BAG3	BCL2-associated athanogene 3	10q25.2-q26.2
BIN1	Amphiphysin	2q14
CAPN3	Calpain 3	15q15.1-q21.1
CAV3	Caveolin 3	3p25
CFL2	Cofilin 2 (muscle)	14q12
CHKB	Choline kinase beta	22q13
CNTN1	Contactin-1	12q11-q12
COL6A1	Alpha 1 type VI collagen	21q22.3
COL6A2	Alpha 2 type VI collagen	21q22.3
COL6A3	Alpha 3 type VI collagen	2q37
CPT2	Carnitine palmitoyltransferase II	1p32
CRYAB	Crystallin, alpha B	11q22.3-q23.1
DAG1	Dystroglycan1	3p21
DES	Desmin	2q35
DMD	Dystrophin	Xp21.2
DNAJB6	HSP-40 homologue, subfamily B, number 6	7q36
DNM2	Dynamin 2	19p13.2
DPM1	Dolichyl-phosphate mannosyltransferase 1, catalytic subunit	20q13.13

DPM2	Dolichyl-phosphate mannosyltransferase polypeptide 2, regulatory subunit	9q34.13
DPM3	Dolichyl-phosphate mannosyltransferase polypeptide 3	1q22
DUX4	Double homeobox 4	4q35
DYSF	Dysferlin	2p12-14
EMD	Emerin	Xq28
ENO3	Enolase 3, beta muscle specific	17pter-p11
ETFA	Electron-transfer-flavoprotein, alpha polypeptide	15q23-q25
ETFB	Electron-transfer-flavoprotein, beta polypeptide	19q13.3-q13.4
ETFDH	Electron-transferring-flavoprotein dehydrogenase	4q32-q35
FHL1	Four and a half LIM domain 1	Xq26.3
FKRP	Fukutin-related protein	19q13.32
FKTN	Fukutin	9q31-q33
FLNC	Filamin C, gamma (actin-binding protein - 280)	7q32
GAA	Acid alpha-glucosidase preproprotein	17q25.2-q25.3
GBE1	Glucan (1,4-alpha-), branching enzyme 1	3p12
GMPPB	GDP-mannose pyrophosphorylase B	3p21.31
GNE	UDP-N-acetylglucosamine-2- epimerase/N-acetylmannosamine kinase	9p13.3
GYG1	Glycogenin 1	3q24
GYS1	glycogen synthase 1 (muscle)	19q13.3
ISCU	Iron-sulfur cluster scaffold homolog (E. coli)	12q24.1
ISPD	Isoprenoid synthase domain containing	7p21.2
ITGA7	Integrin alpha 7 precursor	12q13
KBTBD13	Kelch repeat and BTB (POZ) domain containing 13	15q22.31
KLHL40	Kelch-like family member 40	2p22.1
KLHL9	Kelch-like homologue 9	9p21.2-p22.3
LAMA2	Laminin alpha 2 chain of merosin	6q22-q23
LAMP2	Lysosomal-associated membrane protein 2 precursor	Xq24
LARGE	Like-glycosyltransferase	22q12.3-q13.1
LDB3	LIM domain binding 3	10q22
LDHA	Lactate dehydrogenase A	11p15.4
LMNA	Lamin A/C	1q22

LPIN1	Lipin 1 (phosphatidic acid phosphatase 1)	2p25.1
MATR3	Matrin 3	5q31
MEGF10	Multiple EGF-like-domains 10	5q23.2
MSTN	Myostatin, muscular hypertrophy	2q32
MTM1	Myotubularin	Xq28
MYBPC3	Cardiac myosin binding protein-C	11p11.2
MYH2	Myosin, heavy polypeptide 2, skeletal muscle	17p13.1
MYH7	Myosin, heavy polypeptide 7, cardiac muscle, beta	14q12
MYOT	Myotilin	5q31
NEB	Nebulin	2q22
PABPN1	Poly(A) binding protein, nuclear 1	14q11.2-q13
PFKM	Phosphofructokinase, muscle	12q13.3
PGAM2	Phosphoglycerate mutase 2 (muscle)	7p13-p12
PGK1	Phosphoglycerate kinase 1	Xq13
PGM1	Phosphoglucomutase 1	1p31
PHKA1	Phosphorylase b kinase, alpha submit	Xq13
PLEC	Plectin 1, intermediate filament binding protein 500kDa	8q24
PNPLA2	Adipose triglyceride lipase (desnutrin)	1p15.5
POMGNT1	O-linked mannose beta1,2-N-acetylglucosaminyltransferase 1	1p34.1
POMGNT2	O-linked mannose beta1,2-N-acetylglucosaminyltransferase 2	3p22.1
POMK	Protein-O-mannose kinase	8p11.21
POMT1	Protein-O-mannosyltransferase 1	9q34.1
POMT2	Protein-O-mannosyltransferase 2	14q24.3
PRKAG2	Protein kinase, AMP-activated, gamma 2 non-catalytic subunit	7q31
PTPLA	Protein tyrosine phosphatase-like (3-Hydroxyacyl-CoA dehydratase	10p12.33
PTRF	Polymerase I and transcript release factor	17q21-q23
PYGM	Glycogen phosphorylase	11q12-q13.2
RYR1	Ryanodine receptor 1 (skeletal)	19q13.1
SEPN1	Selenoprotein N1	1p36.13
SGCA	Alpha sarcoglycan	17q21

SGCB	Beta sarcoglycan	4q12
SGCD	Delta-sarcoglycan	5q33-q34
SGCG	Gamma sarcoglycan	13q12
SLC22A5	Solute carrier family 22 member 5	5q31
SLC25A20	Carnitine-acylcarnitine translocase	3p21.31
SMCHD1	Structural maintenance of chromosomes flexible hinge domain containing 1	18p11.32
SYNE1	Spectrin repeat containing, nuclear envelope 1 (nesprin 1)	6q25
SYNE2	Spectrin repeat containing, nuclear envelope 2 (nesprin 2)	14q23.2
TCAP	Telethonin	17q12
TIA1	Cytotoxic granuleassociated RNA binding protein	2p13
TMEM43	Transmembrane protein 43	3p25.1
TMEM5	Transmembrane protein 5	12q14.2
TNNT1	Slow troponin T	19q13.4
TNPO3	Transportin 3	7q32.1-q32.2
TPM2	Tropomyosin 2 (beta)	9p13
TPM3	Tropomyosin 3	1q21.2
TRAPPC11	trafficking protein particle complex 11	4q35.1
TRIM32	Tripartite motif-containing 32	9q33.2
TTN	Titin	2q31
VCP	Valosin-containing protein	9p13-p12
POMGNT2	O-linked mannose beta1,2-N-acetylglucosaminyltransferase 2	3p22.1
POMK	Protein-O-mannose kinase	8p11.21
POMT1	Protein-O-mannosyltransferase 1	9q34.1
POMT2	Protein-O-mannosyltransferase 2	14q24.3
PRKAG2	Protein kinase, AMP-activated, gamma 2 non-catalytic subunit	7q31
PTPLA	Protein tyrosine phosphatase-like (3-Hydroxyacyl-CoA dehydratase	10p12.33
PTRF	Polymerase I and transcript release factor	17q21-q23
PYGM	Glycogen phosphorylase	11q12-q13.2
RYR1	Ryanodine receptor 1 (skeletal)	19q13.1
SEPN1	Selenoprotein N1	1p36.13

Appendix III.2

SGCA	Alpha sarcoglycan	17q21
SGCB	Beta sarcoglycan	4q12
SGCD	Delta-sarcoglycan	5q33-q34
SGCG	Gamma sarcoglycan	13q12
SLC22A5	Solute carrier family 22 member 5	5q31
SLC25A20	Carnitine-acylcarnitine translocase	3p21.31
SMCHD1	Structural maintenance of chromosomes flexible hinge domain containing 1	18p11.32
SYNE1	Spectrin repeat containing, nuclear envelope 1 (nesprin 1)	6q25
SYNE2	Spectrin repeat containing, nuclear envelope 2 (nesprin 2)	14q23.2
TCAP	Telethonin	17q12
TIA1	Cytotoxic granuleassociated RNA binding protein	2p13
TMEM43	Transmembrane protein 43	3p25.1
TMEM5	Transmembrane protein 5	12q14.2
TNNT1	Slow troponin T	19q13.4
TNPO3	Transportin 3	7q32.1-q32.2
TPM2	Tropomyosin 2 (beta)	9p13
TPM3	Tropomyosin 3	1q21.2
TRAPPC11	trafficking protein particle complex 11	4q35.1
TRIM32	Tripartite motif-containing 32	9q33.2
TTN	Titin	2q31
VCP	Valosin-containing protein	9p13-p12

Table III.2.2 - Variants filtered-in by GEMINI

Variants filtered-in by GEMINI ("comp_hets" function)

Chromosome	Start position	End position	Reference	Alteration	Quality score	dbSNP id	Gene	Frequency in ESP/ dbSNP (*)	Genotypes (Patient/Mother/Father)	Variant description (HGVS nomenclature)		Gene location
chr1	26131653	26131654	G	A	501.44	rs7349185	SEPN1	0.167 (*)	G/A,A/A,G/A	NM_020451.2:c.425G>A	p.Cys142Tyr	exonic
chr1	26140572	26140573	C	A	190.18	rs2294228	SEPN1	0.368	C/A,A/A,C/A	NM_020451.2:c.1506C>A	p.Asn502Lys	exonic
chr1	100316588	100316589	A	G	165.69	rs2307130	AGL	0.416	A/G,A/G,./.	NM_000028.2:c.-10A>G	p.=	5'UTR/exonic
chr1	100358102	100358103	C	T	249.45	rs3753494	AGL	0.139	C/T,./.,T/T	NM_000028.2:c.3199C>T	p.Pro1067Ser	exonic
chr2	152352842	152352843	C	G	300.85	rs7575451	NEB	0.371	C/G,C/G,G/G	NM_001271208.1:c.24538G>C	p.Ala8180Pro	exonic
chr2	152404900	152404901	G	A	213.38	rs35707762	NEB	0.009	G/A,./.,G/A	NM_001271208.1:c.20078C>T	p.Thr6693Ile	exonic
chr2	152422075	152422076	C	G	489.19	rs2288210	NEB	0.350	C/G,C/G,G/G	NM_001271208.1:c.18305G>C	p.Arg6102Thr	exonic
chr2	152476027	152476028	C	G	275.88	rs10172023	NEB	0.269	C/G,C/G,C/G	NM_001271208.1:c.10809G>C	p.Trp3603Cys	exonic
chr2	179432184	179432185	A	G	445.5	rs12463674	TTN	0.226	A/G,A/G,./.	NM_001267550.2:c.78674T>C	p.Ile26225Thr	exonic
chr2	179544684	179544685	C	CTCT	283.88	None	TTN	0.013	C/CTCT,C/CTCT,./.	NM_001267550.2:c.33513_33515dup	p.Glu11172dup	exonic
chr2	179545858	179545859	C	T	355.85	rs36051007	TTN	0.226	C/T,C/T,./.	NM_001267550.2:c.33287G>A	p.Arg11096His	exonic
chr2	179554304	179554305	C	T	167.79	rs2244492	TTN	0.433	C/T,C/T,./.	NM_001267550.2:c.31864G>A	p.Gly10622Arg	exonic
chr2	238244862	238244878	TGCAGCAGCAGCAGC	TGCAGCAGCAGCG,T	218.78	rs35879189	COL6A3	0.095	TGCAGCAGCAGCAGCG/TGCAG	NM_004369.3:c.8877_8879del	p.Ala2960del	exonic
chr2	238287848	238287849	G	T	1071.0	None	COL6A3	None	G/T,./.,T/T	NM_004369.3:c.1927C>A	p.Leu643Ile	exonic
chr3	81643166	81643167	T	C	331.55	rs2172397	GBE1	0.021	T/C,C/C,C/C	NM_000158.3:c.1000G>A	p.Val334Ile	exonic
chr3	81692084	81692085	C	T	405.7	rs28763902	GBE1	0.002	C/T,C/T,./.	NM_000158.3:c.839G>A	p.Gly280Asp	exonic
chr5	126732426	126732427	G	A	38.03	rs3812054	MEGF10	0.108	G/A,./.,G/A	NM_032446.2:c.616G>A	p.Val206Ile	exonic
chr5	126791281	126791282	G	A	284.05	rs17164935	MEGF10	0.144	G/A,./.,G/A	NM_032446.2:c.3215G>A	p.Arg1072Lys	exonic
chr9	134385435	134385436	A	G	762.58	rs2296949	POMT1	0.121	A/G,G/G,A/G	NM_007171.3:c.752G>A	p.Arg251Gln	exonic
chr9	134386780	134386781	G	A	62.14	rs4740164	POMT1	0.038	G/A,./.,G/A	NM_007171.3:c.979G>A	p.Val327Ile	exonic
chr15	65369394	65369395	C	T	286.09	rs2919358	KBTBD13	0.396	C/T,C/T,./.	NM_001101362.2:c.242C>T	p.Ala81Val	exonic
chr15	65369946	65369947	G	A	55.07	rs146917406	KBTBD13	0.001	G/A,G/A,./.	NM_001101362.2:c.794G>A	p.Gly265Asp	exonic
chr17	4856375	4856376	A	G	58.28	rs238238	ENO3	0.331	A/G,./.,A/G	NM_001976.4:c.212A>G	p.Asn71Ser	exonic
chr17	4856579	4856580	T	C	588.74	rs238239	ENO3	0.450	T/C,./.,T/C	NM_001976.4:c.254T>C	p.Val85Ala	exonic
chr21	47545767	47545768	G	A	424.13	rs1042917	COL6A2	0.413	G/A,G/A,G/A	NM_001849.3:c.2039G>A	p.Arg680His	exonic
chr21	47552208	47552209	G	A	275.12	rs35548026	COL6A2	0.081	G/A,./.,G/A	NM_001849.3:c.2803G>A	p.Gly935Arg	exonic

Variants excluded (sequencing artifact)

Chromosome	Start position	End position	Reference	Alteration	Quality score	dbSNP id	Gene	Frequency ESP	Genotypes (Patient/Mother/Father)
chr1	26140658	26140659	A	C	19.44	None	SEPN1	None	A/C,./.,./.
chr2	179611942	179611943	A	AG	52.86	None	TTN	None	A/AG,./.,./.
chr21	47549160	47549163	AGG	AG,AGA	198.0	None	COL6A2	None	AG/AGA,./.,./.

Variants filtered-in by GEMINI ("autossomal_recessive" function)

Chromosome	Start position	End position	Reference	Alteration	Quality score	dbSNP id	Gene	Frequency in ESP/ dbSNP (*)	Genotypes (Patient/Mother/Father)	Variant description	(HGVS nomenclature)	Gene location
chr6	152665260	152665261	C	A	320.9	rs4645434	SYNE1	0.433	A/A,C/A,C/A	NM_182961.3:c.12180G>T	p.Glu4060Asp	exonic
chr6	152671474	152671475	A	C	443.67	rs9478320	SYNE1	0.404	C/C,A/C,A/C	NM_182961.3:c.11734-5T>G	-	intronic
chr6	152673146	152673147	G	A	340.19	rs6908392	SYNE1	0.422	A/A,A/G,A/G	NM_182961.3:c.11580+15C>T	-	intronic
chr6	152675853	152675854	A	G	489.63	rs9397102	SYNE1	0.419	G/G,A/G,A/G	NM_182961.3:c.10866T>C	p.= (p.Ser3622Ser)	exonic
chr6	152683412	152683413	G	T	929.3	rs4407724	SYNE1	0.352	T/T,G/T,G/T	NM_182961.3:c.10191C>A	p.= (p.Gly3397Gly)	exonic
chr6	152690558	152690559	G	A	242.48	rs1873176	SYNE1	0.463	A/A,A/G,A/G	NM_182961.3:c.9651+47C>T	-	intronic
chr6	152694183	152694184	T	C	920.02	rs6913579	SYNE1	0.471	C/C,T/C,T/C	NM_182961.3:c.9495A>G	p.= (p.Glu3165Glu)	exonic
chr4	159620034	159620035	A	C	204.66	rs7679753	ETFDH	0.285 (*)	C/C,A/C,A/C	NM_004453.3:c.973-104A>C	-	intronic
chr2	152500448	152500449	C	G	386.61	rs13013209	NEB	0.350	G/G,C/G,C/G	NM_001271208.1:c.7839G>C	p.Lys2613Asn	exonic
chr2	70456585	70456586	C	T	354.68	rs2706768	TIA1	0.398 (*)	T/T,C/T,C/T	NM_022173.2:c.223-136G>A	-	intronic
chr19	10916683	10916684	C	T	750.15	rs2287029	DNM2	0.129	T/T,C/T,C/T	NM_001005360.2:c.1545+41C>T	-	intronic
chr12	41421641	41421642	A	G	318.14	rs12367345	CNTN1	0.121	G/G,A/G,A/G	NM_001843.3:c.2711-17A>G	-	intronic
chr12	56088205	56088206	T	C	768.56	rs7953669	ITGA7	0.436	C/C,T/C,T/C	NM_001144996.1:c.2369+21A>G	-	intronic
chr22	51018486	51018487	A	G	1000.36	None	CHKB	None	G/G,A/G,A/G	NM_005198.4:c.843T>C	p.= (p.Phe281Phe)	exonic

NGS: Chr22:g.51018487A>G, [c.843T>C]
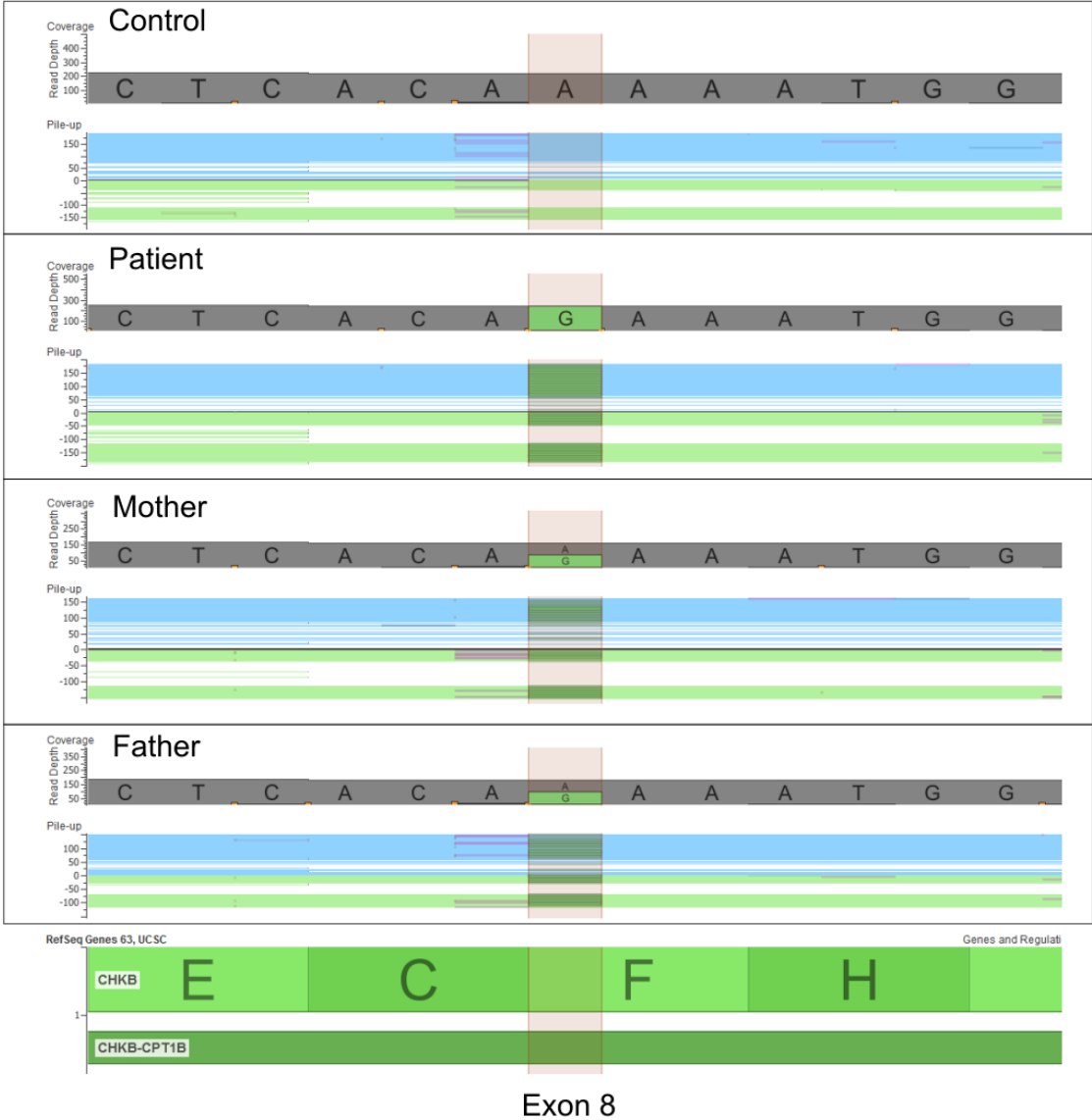
Exon 8

Figure III.2.1 - Partial alignment obtained for exon 8 of the *CHKB* gene, showing a novel silent variant (c.843T>C) homozygous in the patient and heterozygous in the parents. This sequence variant was detected in reads from both orientations (forward and reverse) as indicated by the blue and light green graphics. cDNA reference sequence: NM_005198.4.
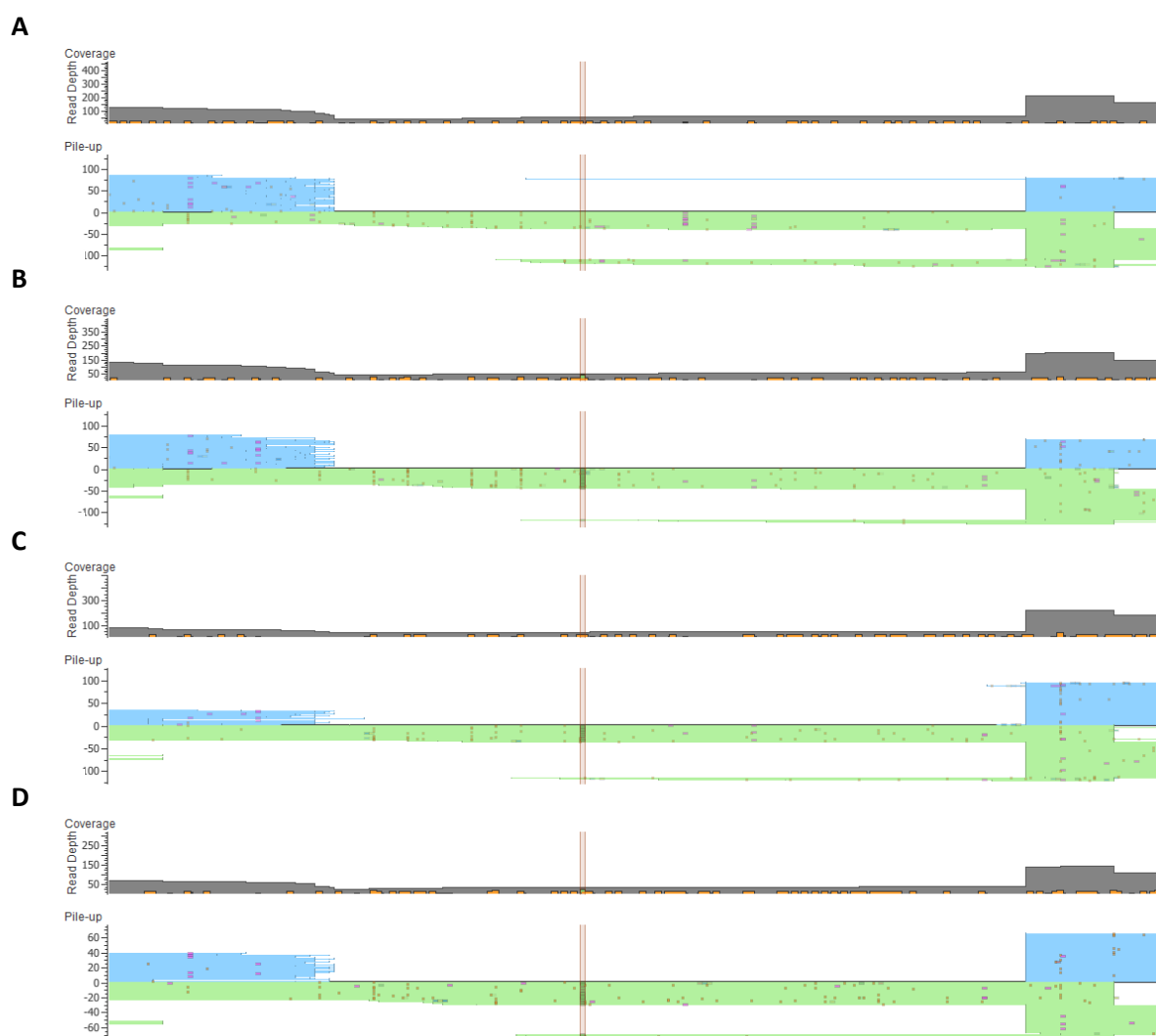
Figure III.2.2 - Reads' alignment obtained for intron 9/exon 9 of the *CHKB* gene in: (A) control; (B) patient; patient's (C) mother and (D) father. In location of the c.1031+3G>C variant (highlighted by vertical bars) only reverse sequences (light green graphics) were obtained. The forward sequences do not reach the full length of the amplicon in all samples.
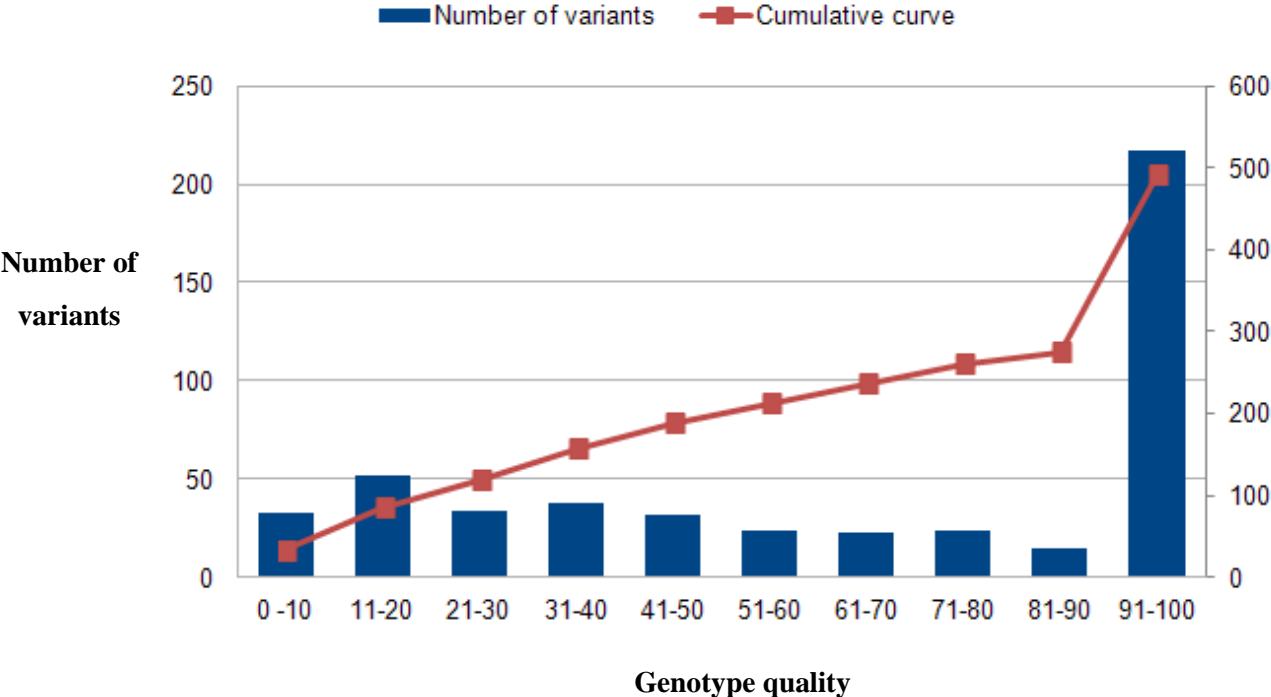
Figure III.2.3 - Number of heterozygous variants and respective cumulative curve plotted against genotype quality (GQ) class intervals. These variants are located in the 108 loci implicated in hereditary myopathies. The GQ value is a Phred-scaled value, thus higher values have a higher likelihood of true genotype calling. Data stratification shows a clear shift between GQ intervals 81-90 and 91-100.
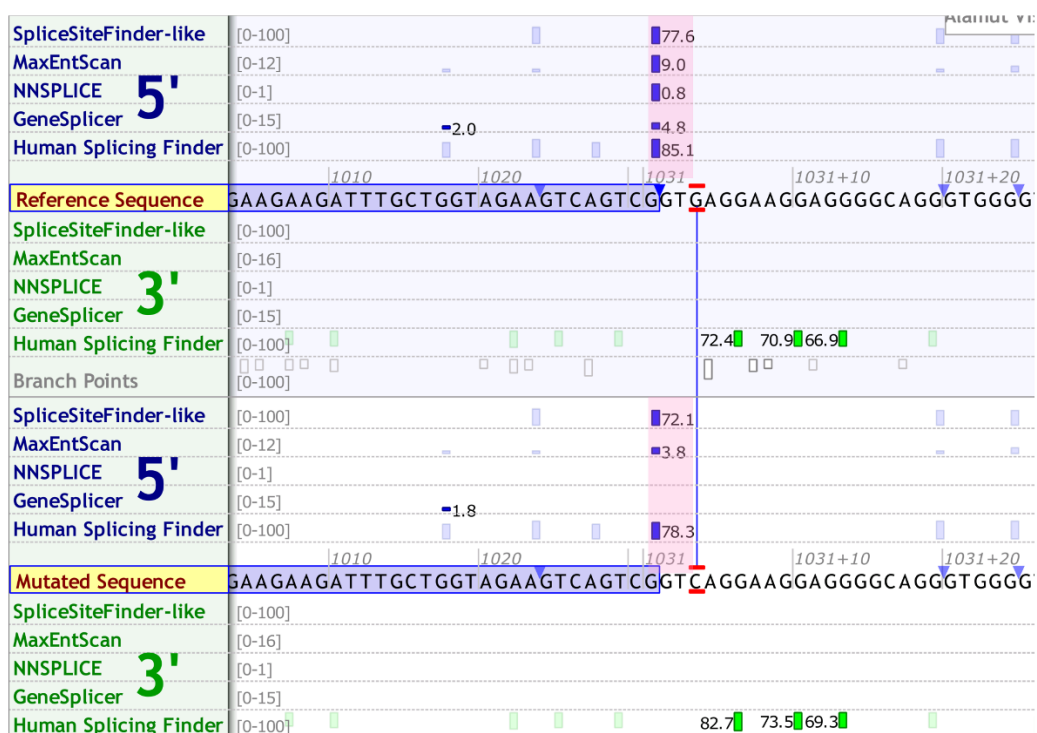
Figure III.2.4 - Bioinformatic analysis of the c.1031+3G>C variant using different splicing prediction tools incorporated in Alamut Visual V2.4 software (Interactive Biosoftware, Rouen, France).

Figure III.2.5 - Muscle fiber surrounded by extensive fibrous tissue. Around the nucleus several slightly enlarged mitochondria are seen. Electron microscopy at 4000x magnification.

Appendix III.3

Contents

Table III.3.1 - List of genes related with muscle diseases.

HGNC Gene Symbol:	Official Full Name	Location
ACADVL	acyl-CoA dehydrogenase, very long chain	17p13.1
ACTA1 *	actin, alpha 1, skeletal muscle	1q42.13
AGL	amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase	1p21.2
AGRN	agrin	1p36.33
ALG13	ALG13, UDP-N-acetylglucosaminyltransferase subunit	Xq23
ALG2	ALG2, alpha-1,3/1,6-mannosyltransferase	9q22.33
ANO5	anoctamin 5	11p14.3
ATP2A1	ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 1	16p11.2
B3GALNT2	beta-1,3-N-acetylgalactosaminyltransferase 2	1q42.3
B4GAT1 (B3GNT1)	beta-1,4-glucuronyltransferase 1	11q13.2
BAG3	BCL2 associated athanogene 3	10q26.11
BICD2	BICD cargo adaptor 2	9q22.31
BIN1 *	bridging integrator 1	2q14.3
CACNA1S	calcium voltage-gated channel subunit alpha1 S	1q32.1
CAPN3	calpain 3	15q15.1
CAV3	caveolin 3	3p25.3
CCDC78 *	coiled-coil domain containing 78	16p13.3
CFL2 *	cofilin 2	14q13.1
CHAT	choline O-acetyltransferase	10q11.23
CHKB	choline kinase beta	22q13.33
CHRNA1	cholinergic receptor nicotinic alpha 1 subunit	2q31.1
CHRNB1	cholinergic receptor nicotinic beta 1 subunit	17p13.1
CHRND	cholinergic receptor nicotinic delta subunit	2q37.1
CHRNE	cholinergic receptor nicotinic epsilon subunit	17p13.2
CLCN1	chloride voltage-gated channel 1	7q34
CNTN1 *	contactin 1	12q12
COL12A1	collagen type XII alpha 1 chain	6q14.1
COL6A1	collagen type VI alpha 1 chain	21q22.3
COL6A2	collagen type VI alpha 2 chain	21q22.3
COL6A3	collagen type VI alpha 3 chain	2q37.3
COLQ	collagen like tail subunit of asymmetric acetylcholinesterase	3p25.1
CPT2	carnitine palmitoyltransferase 2	1p32.3
CRYAB	crystallin alpha B	11q23.1
DAG1	dystroglycan 1	3p21.31
DES	desmin	2q35
DMD	dystrophin	Xp21.1
DNAJB6	DnaJ heat shock protein family (Hsp40) member B6	7q36.3
DNM2 *	dynamin 2	19p13.2

DOK7	docking protein 7	4p16.3
DPAGT1	dolichyl-phosphate N-acetylglucosaminephosphotransferase 1	11q23.3
DPM1	dolichyl-phosphate mannosyltransferase subunit 1, catalytic	20q13.13
DPM2	dolichyl-phosphate mannosyltransferase subunit 2, regulatory	9q34.11
DPM3	dolichyl-phosphate mannosyltransferase subunit 3	1q22
DYNC1H1	dynein cytoplasmic 1 heavy chain 1	14q32.31
DYSF	dysferlin	2p13.2
EMD	emerin	Xq28
ENO3	enolase 3	17p13.2
ERBB3	erb-b2 receptor tyrosine kinase 3	12q13.2
ETFA	electron transfer flavoprotein alpha subunit	15q24.3
ETFB	electron transfer flavoprotein beta subunit	19q13.41
ETFDH	electron transfer flavoprotein dehydrogenase	4q32.1
FHL1	four and a half LIM domains 1	Xq26.3
FKRP	fukutin related protein	19q13.32
FKTN	fukutin	9q31.2
FLNC	filamin C	7q32.1
GAA	glucosidase alpha, acid	17q25.3
GBE1	1,4-alpha-glucan branching enzyme 1	3p12.2
GFPT1	glutamine--fructose-6-phosphate transaminase 1	2p13.3
GMPPB	GDP-mannose pyrophosphorylase B	3p21.31
GNE	glucosamine (UDP-N-acetyl)-2-epimerase/ N-acetylmannosamine kinase	9p13.3
GYG1	glycogenin 1	3q24
GYS1	glycogen synthase 1	19q13.33
HNRNPDL	heterogeneous nuclear ribonucleoprotein D like	4q21.22
HRAS	HRas proto-oncogene, GTPase	11p15.5
HSPG2	heparan sulfate proteoglycan 2	1p36.12
IGHMBP2	immunoglobulin mu binding protein 2	11q13.3
ISCU	iron-sulfur cluster assembly enzyme	12q23.3
ISPD	isoprenoid synthase domain containing	7p21.2
ITGA7	integrin subunit alpha 7	12q13.2
KBTBD13 *	kelch repeat and BTB domain containing 13	15q22.31
KLHL40 *	kelch like family member 40	3p22.1
KLHL41 *	kelch like family member 41	2q31.1
KLHL9	kelch like family member 9	9p21.3
LAMA2	laminin subunit alpha 2	6q22.33
LAMP2	lysosomal associated membrane protein 2	Xq24
LARGE1 (LARGE)	LARGE xylosyl- and glucuronyltransferase 1	22q12.3
LDB3	LIM domain binding 3	10q23.2

LDHA	lactate dehydrogenase A	11p15.1
LMNA	lamin A/C	1q22
LMOD3 *	leiomodin 3	3p14.1
LPIN1	lipin 1	2p25.1
MATR3	matrin 3	5q31.2
MEGF10 *	multiple EGF like domains 10	5q23.2
MSTN	myostatin	2q32.2
MTM1 *	myotubularin 1	Xq28
MUSK	muscle associated receptor tyrosine kinase	9q31.3
MYBPC3 *	myosin binding protein C, cardiac	11p11.2
MYF6	myogenic factor 6	12q21.31
MYH2 *	myosin heavy chain 2	17p13.1
MYH3	myosin heavy chain 3	17p13.1
MYH7 *	myosin heavy chain 7	14q11.2
MYOT	myotilin	5q31.2
NEB *	nebulin	2q23.3
ORAI1	ORAI calcium release-activated calcium modulator 1	12q24.31
PABPN1	poly(A) binding protein nuclear 1	14q11.2
PFKM	phosphofructokinase, muscle	12q13.11
PGAM2	phosphoglycerate mutase 2	7p13
PGK1	phosphoglycerate kinase 1	Xq21.1
PGM1	phosphoglucomutase 1	1p31.3
PHKA1	phosphorylase kinase regulatory subunit alpha 1	Xq13.1
PIP5K1C	phosphatidylinositol-4-phosphate 5-kinase type 1 gamma	19p13.3
PLEC	plectin	8q24.3
PNPLA2	patatin like phospholipase domain containing 2	11p15.5
POMGNT1	protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)	1p34.1
POMGNT2	protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-)	3p22.1
POMK	protein-O-mannose kinase	8p11.21
POMT1	protein O-mannosyltransferase 1	9q34.13
POMT2	protein O-mannosyltransferase 2	14q24.3
PRKAG2	protein kinase AMP-activated non-catalytic subunit gamma 2	7q36.1
PTRF	polymerase I and transcript release factor	17q21.2
PYGM	phosphorylase, glycogen, muscle	11q13.1
RAPSN	receptor associated protein of the synapse	11p11.2
RBCK1	RANBP2-type and C3HC4-type zinc finger containing 1	20p13
RYR1 *	ryanodine receptor 1	19q13.2
SCN4A	sodium voltage-gated channel alpha subunit 4	17q23.3

SELENON (SEPN1)*	selenoprotein N	1p36.11
SGCA	sarcoglycan alpha	17q21.33
SGCB	sarcoglycan beta	4q12
SGCD	sarcoglycan delta	5q33.2
SGCG	sarcoglycan gamma	13q12.12
SIL1	SIL1 nucleotide exchange factor	5q31.2
SLC22A5	solute carrier family 22 member 5	5q31.1
SLC25A20	solute carrier family 25 member 20	3p21.31
SMCHD1	structural maintenance of chromosomes flexible hinge domain containing 1	18p11.32
SPEG *	SPEG complex locus	2q35
STIM1	stromal interaction molecule 1	11p15.4
SYNE1	spectrin repeat containing nuclear envelope protein 1	6q25.2
SYNE2	spectrin repeat containing nuclear envelope protein 2	14q23.2
TAZ	tafazzin	Xq28
TCAP	titin-cap	17q12
TIA1	TIA1 cytotoxic granule associated RNA binding protein	2p13.3
TMEM43	transmembrane protein 43	3p25.1
TMEM5	transmembrane protein 5	12q14.2
TNNI2	troponin I2, fast skeletal type	11p15.5
TNNT1 *	troponin T1, slow skeletal type	19q13.42
TNPO3	transportin 3	7q32.1
TOR1AIP1	torsin 1A interacting protein 1	1q25.2
TPM2 *	tropomyosin 2 (beta)	9p13.3
TPM3 *	tropomyosin 3	1q21.3
TRAPPC11	trafficking protein particle complex 11	4q35.1
TRIM32 *	tripartite motif containing 32	9q33.1
TRPV4	transient receptor potential cation channel subfamily V member 4	12q24.11
TTN *	titin	2q31.2
UBA1	ubiquitin like modifier activating enzyme 1	Xp11.3
VCP	valosin containing protein	9p13.3
VMA21	VMA21 vacuolar H+-ATPase homolog (S. cerevisiae)	Xq28

Footnote: This list of candidate genes used in analysis of WES data (version of 2015), includes those that cause primary myopathies - muscular dystrophies, congenital muscular dystrophies, congenital myopathies (identified with asterisks) and other myopathies. Genes related with metabolic myopathies and congenital myasthenic syndromes were also included as they can mimic primary muscle diseases. Data was derived from the Gene Table of Neuromuscular disorders (http://musclegenetable.fr/) and OMIM - Online Mendelian Inheritance in Man (http://omim.org/). Some gene symbols were altered during the course of this study; former symbols (used during analysis) are indicated between parentheses. HGNC- HUGO gene nomenclature committee.

Appendix III.4

Contents

REGION 1 – Potential acceptor splice-site

REGION 1 – Potential branch-points

REGION 2 – Potential acceptor splice-site

REGION 2 – Potential branch-points

REGION 3 – Potential acceptor splice-site

REGION 3 – Potential branch-points
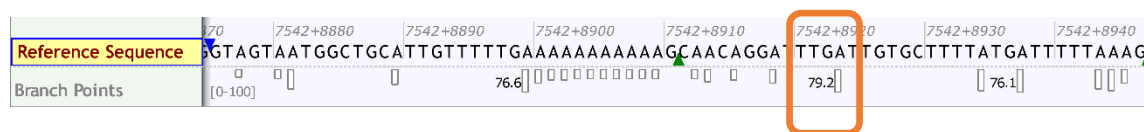
REGION 4 – Potential acceptor splice-site

REGION 4 – Potential branch-points

REGION 5 – Potential acceptor splice-site
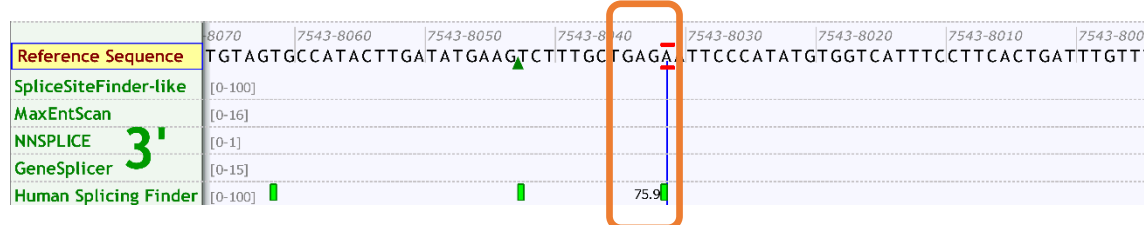
REGION 5 – Potential branch-points
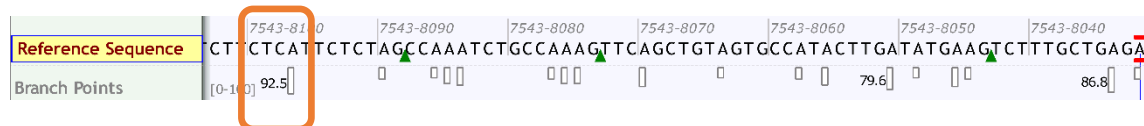
Figure III.4.1 - Bioinformatic splice-site and branch-point analysis. To identify candidate regions for the LINE-1 insertion, different algorithms were used to evaluate potential acceptor splice sites and branch points within intron 51, namely the SpliceSiteFinder-Like, MaxEntScan (Yeo et al., 2004), NNSPLICE (Reese et al., 1997), GeneSplicer (Pertea et al., 2001) and Human Splice Finder (Desmet et al., 2009), integrated in Alamut Visual software (v2.8, Interactive Biosoftware, Rouen, France).

Appendix III.4

References:

Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.

Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. 2001;29(5):1185-90.

Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Comput Biol. 1997;4(3):311-23.

Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2-3):377-94.

Data III.4.1 - Additional material and methods for the LINE-1 characterization.

I. Primers used to identify the LINE-1 5' insertion site in *DMD*.

    a) Schematic showing primers' location:

    b)
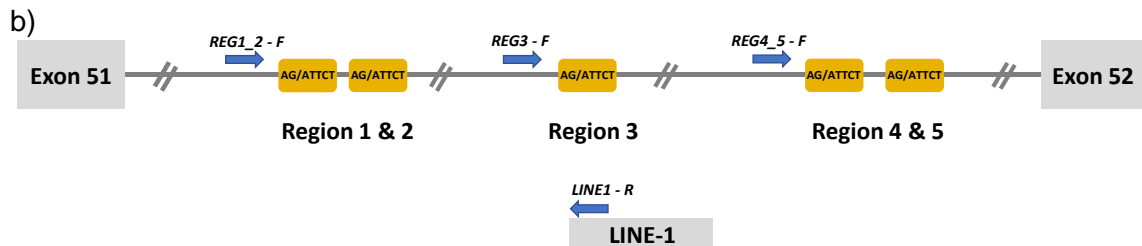
    c) Primers sequences:

        REG1_2-F: TCAAGCAAAACTGATCCAGG

        REG3-F: GCAGTCATGGCCACAAAAG

        REG4_5-F: ACAAAATCTATGGAGGCACAGG

        LINE1-R: TGGAAATGCAGAAATCACCC

II. Sequences of custom probes used for isolating and enriching target region by hybridization capture of DNA libraries for (PacBio) single molecule real time sequencing.

    > Probe nr1 (NM_004006.2:c.7542+8207_7542+8516)

    AGATGAGGCAGTGATCACACATCACTATTAGTAAAAGTGTTTCTGTACCTGTATCCACACT
    TTTATGTATATGGTTACTTATGTTAAAGTGATACATATTATATAAAATTAACGTATACATTAA
    GTAGATATTTTAATAGTCTGTAATTAAATACTACTAGTATTTTCTTTCCTCCTTCAAGTGCTT
    ACTTTTGATACCTCGAGTTACAGTGTCATAAAGATTCTTTAGAAATATATTGACTGTCTTTT
    AAGAGCTTTTGATACAATACTGAGTTTACATTCATCTGTTATTTATTGAACACTTGCTGGTG
    AAAGGCATC

    Primers: F: AGATGAGGCAGTGATCACACA R: GATGCCTTTCACCAGCAAGT

Appendix III.4

> Probe nr2 (NM_004006.2:c.7542+9241_7542+9589)

AATGTAACCCCTCCCCCATATCAAGTTAATCTATGTTCAACTCCAGAATTATTTTTGAACAC
TCAAACTAGAAATTAAAAAAAATTAAATCCATGAAGACGATTTTTGCCAAAAGCATATAGAT
AAATTGAGTTGATTCTATACTTAAGAAAGTGGAGAGGAGAGAGTAATTTGGAGAGAGTAA
TTTACTCTTAATCCCATATTTTTTCCCTAAATGTGAAAGAAGTAGATTGTAGTGAGAGGGA
AAATAACCTGTAGCAACTTCATTGAGGCTAAGCTTTCTGTCATGTTATATTATACGAAAGT
AATGAAATGCTTCCACAGATAGAATCAGAAGTCCCCTCTGAGA

PRIMERS: F: AATGTAACCCCTCCCCCATA R: TCTCAGAGGGGACTTCTGATTC

III. Custom bioinformatic pipeline for single molecule real-time sequencing analysis

a) FASTA/Q files were mapped against a LINE-1 reference sequence (ID: L1HS from http://www.girinst.org/repbase/) using BLASR alignment software ("-clipping soft - minMatch 7 -minPctSimilarity 90") (Chaisson et al., 2012).

b) BAM files were generated from SAM files with Samtools (v0.1.19; "view -b -S") (Li et al., 2009).

c) Consensus sequence was called from BAM files using Samtools mpileup command ("-uf" option), bcftools ("view -cg" option) and vcfutils.pl (vcf2fq command).

References:

Chaisson MJ, Tesler G. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinformatics. 2012;13:238.

Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078-2079.

Figure III.4.2 - Homology search bioinformatics. The search for the origin of the 103 bp sequence identified in the patient's DMD transcripts resorted to the Basic Local Alignment Search Tool (BLAST v2.2.32, http://blast.ncbi.nlm.nih.gov/) at the nucleotide level.

A)

B)

C)

D)
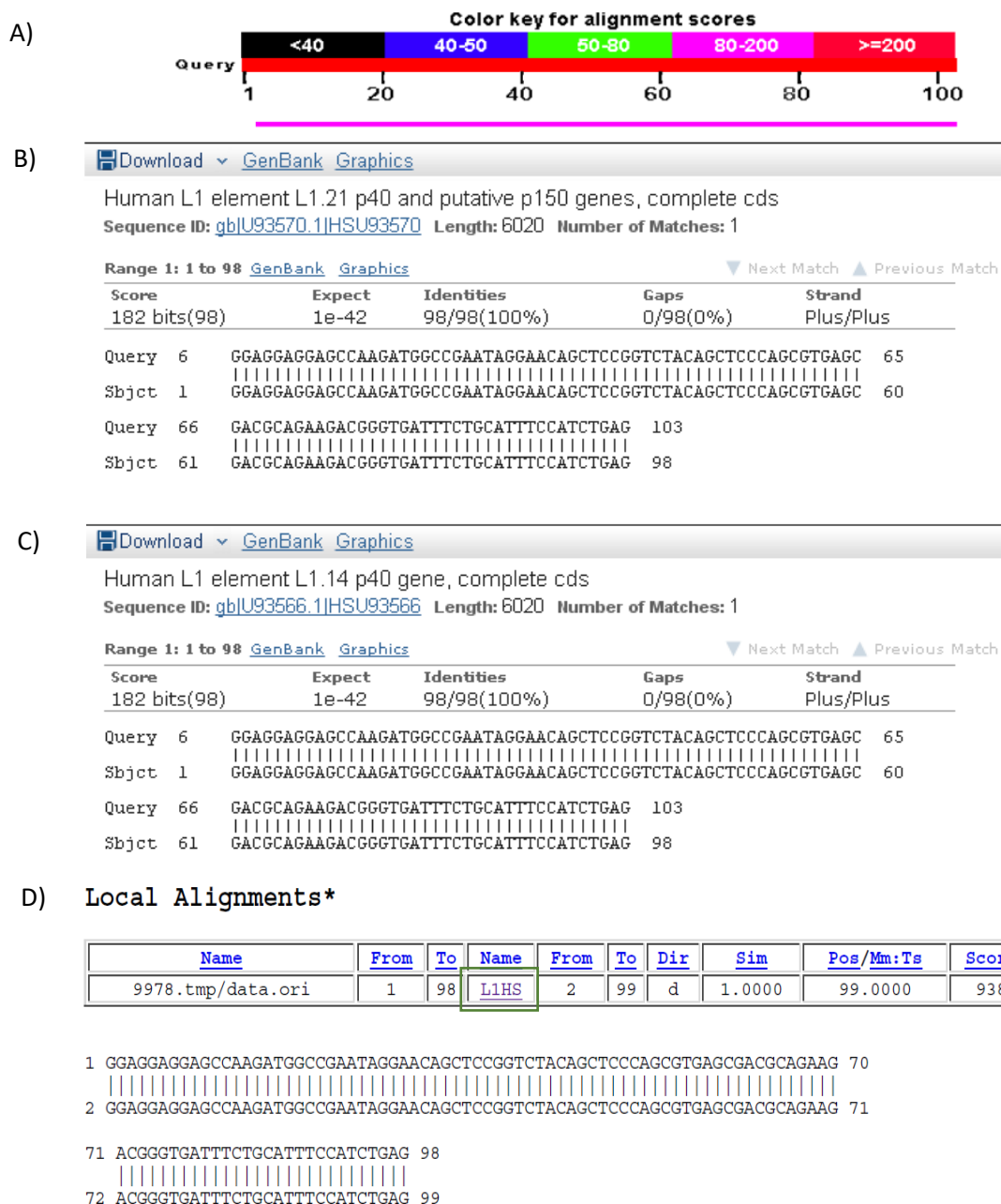
Legend: A) BLAST results showing that a 95.1% identity (98/103) was obtained against two human LINE-1 sequences: B) L1.21 (sequence ID: U93570) and C) L1.14 (sequence ID: U93566). D) To confirm the nature of this sequence, analysis of repetitive elements was performed using CENSOR software (http://www.girinst.org/censor/). Censor results showing 100% similarity with a stretch of 98 nucleotides belonging to the consensus sequence of the L1Hs family.
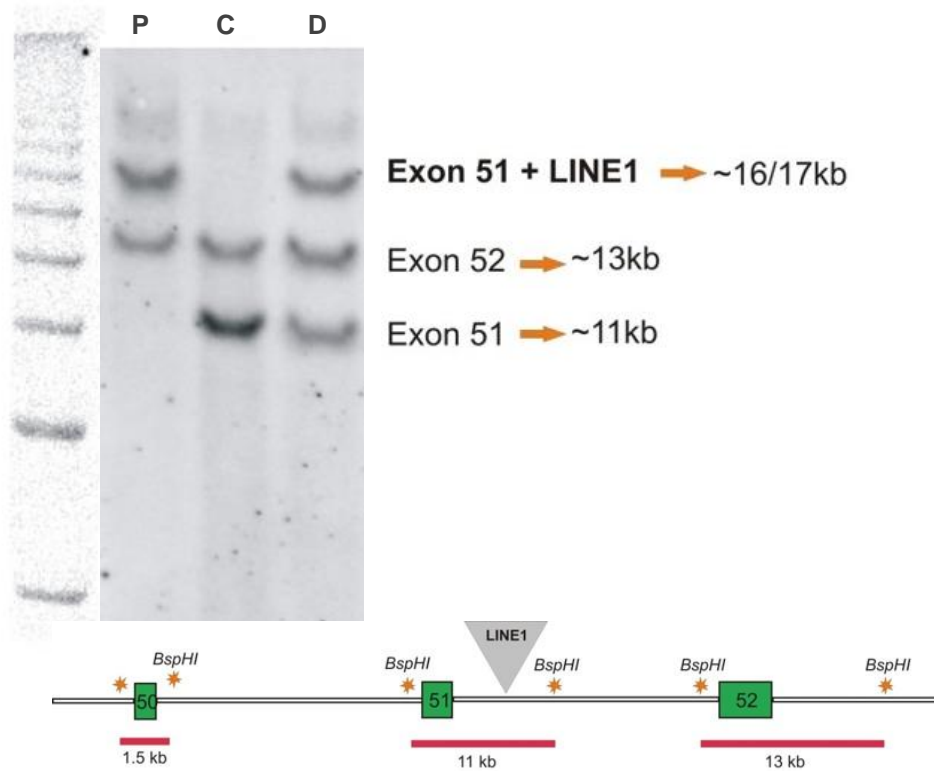
Figure III.4.3 - Southern blot and hybridization analysis using a cDNA probe recognizing exons 50 to 52. A larger fragment (increment of ~ 6 Kb) was identified in the patient (P) and his daughter (D), and not detected in the control sample (C).

Material & Methods:

Genomic DNA (gDNA) samples from patient, daughter, as well as male normal control, were digested with BspHI (NewEngland Biolabs, Beverly, MA), resolved on a 0.7% agarose gel and vacuum transferred to a GeneScreen Plus membrane (Perkin Elmer, Waltham, MA) using an saline method. A custom cDNA probe recognizing exons 50-52 was prepared using the digoxigenin (DIG) DNA Labeling Kit (Roche Applied Science, Indianapolis, IN). The membrane was incubated with this probe overnight using the Easy Hyb Buffer (Roche Applied Science). Subsequently, the membrane was washed at 60ºC in 1xSSC (saline-sodium citrate) /0.1% SDS (Sodium dodecyl sulfate) and 0.5xSSC/0.1% SDS, and prepared with DIG Wash and Block Buffer Set (Roche Applied Science, Indianapolis, IN). Incubation was performed with Anti-DIG-AP conjugate (Roche Applied Science, Indianapolis, IN), and the DIG-labeled probe was detected with ready-to-use CDP-Star (Roche Applied Science, Indianapolis, IN).

Data III.4.2 – Full LINE-1 sequence

> LINE-1_DMD_int51

```
   1 ggaggaggag ccaagatggc cgaataggaa cagctccggt ctacagctcc cagcgtgagc
  61 gacgcagaag acgggtgatt tctgcatttc catctgaggt accgggttca tctcaatagg
 121 gagtgccaga cagtgggcgc aggccagtgt gtgtgcgcac cgtgcgcaag ccgaagcagg
 181 gcgaggcatt gcctcacctg ggaagcgcaa ggggtcaggg agttcccttt ccgagtcaaa
 241 gaaaggggtg acggacgtac ctggaaaatc gggtcactcc cacccgaata ttgcgctttt
 301 cagaccggct taagaaacgg cgcaccacga gactatatcc cacacctggc tcagagggtc
 361 ctacgcccac ggaatctcgc tgattgctag cacagcagtc tgagatcaaa ctgcaaggcg
 421 gcaacgaggc tgggggaggg gcgcccgcca ttgcccaggc ttgcttaggt aaacaaagca
 481 gccgggaagc tcgaactggg tggagcccac cacagctcaa ggaggcctgc ctgcctctgt
 541 aggctccacc tctgggggca gggcacagac aaacaaaaag acagcagtaa cctctgcaga
 601 cttaagtgtc cctgtctgac agctttgaag agagcagtgg ttctcccagc acgcagctgg
 661 agatctgaga acgggcagac tgcctcctca agtgggtccc tgacccctga cccccgagca
 721 gcctaactgg gaggcacccc ccagcagggg cacactgaca cctcacacgg cagggtattc
 781 caacagacct gcagctgagg gtcctgtctg ttagaaggaa aactaacaac cagaaaggac
 841 atctacaccg aaaacccatc tgtacatcac catcatcaaa gaccaaaagt agataaaacc
 901 acaaagatgg ggaaaaaaca gaacagaaaa actggaaact ctaaaacgca gagcgcctct
 961 cctcctccaa aggaacgcag ttcctcacta gcaacagaac aaagctggat ggagaatgat
1021 tttgacgagc tgagagaaga aggcttcaga cgatcaaatt actctgagct acgggaggac
1081 attcaaacca aaggcaaaga agttgaaaac tttgaaaaaa atttagaaga atgtataact
1141 agaataacca atacagagaa gtgcttaaag gagctgatgg agctgaaaac caaggctcga
1201 gaactacgtg aagaatgcag aagcctcagg agccgatgcg atcaactgga agaaagggta
1261 tcagcaatgg aagatgaaat gaatgaaatg aagcgagaag ggaagtttag agaaaaaaga
1321 ataaaaagaa atgagcaaag cctccaagaa atatgggact atgtgaaaag accaaatcta
1381 cgtctgattg gtgtacctga aagtgatgtg gagaatggaa ccaagttgga aaacactctg
1441 caggatatta tccaggagaa cttccccaat ctagcaaggc aggccaacgt tcagattcag
1501 gaaatacaga gaacgccaca agatactcc tcgagaagag caactccaag acacatactt
1561 gtcagattca ccaaagttga aatgaaggaa aaaatgttaa gggcagccag agagaaaggt
1621 caggttaccc tcaaaggaaa gcccatcaga ctaacagcgg atctctcggc agaaacccta
1681 caagccagaa gagagtgggg gccaatattc aacattctta agaaaagaa ttttcaaccc
1741 agaatttcat atccagccaa actaagcttc ataagtgaag gagaaataaa atactttata
1801 gacaagcaaa tgctgagaga ttttgtcacc accaggcctg ccctaaaaga gctcctgaag
1861 gaatcactaa acatggaaag gaacaaccgg taccagccgc tgcaaaatca tgccaaaatg
1921 taaagacatc gagactagga agaaactgca tcaactaatg agcaaaatca ccagctaaca
```

```
1981 tcataatgac aggatcaaat tcacacataa caatattaac tttaaatata aatggactaa
2041 attctgcaat taaaagacac agactggcaa gttggataaa gagtcaagac ccatcagtgt
2101 gctgtattca ggaaacccat ctcacgtgca gagacacaca taggctcaaa ataaaaggat
2161 ggaggaagat ctaccaagca aatggaaaac aaaaaaaggc aggggttgca atcctagtct
2221 ctgataaaac agacttcaaa ccaacgaaga tcaaaagaga caaagaaggc cattacataa
2281 tggtaaaggg atcaattcaa caagaggagc taactatcct aaatatttat gcacccaata
2341 caggagcacc cagattcata aagcaagtcc tcagtgagct acaaagagac ttagactccc
2401 acacattaat aatgggagac tttaacaccc cactgtcaac attagacaga tcaacgagac
2461 agaaagtcaa caaggatacc caggaattga actcagctct gcaccaagca gacctaatag
2521 acatctacag aactctccac cccaaatcaa cagaatatac attttttca gcaccacacc
2581 acacctattc caaaattgac cacatagttg gaagtaaagc tctcctcagc aaatgtaaaa
2641 gaacagaaat tataacaaac tatctctcag accacagtgc aatcaaacta gaactcagga
2701 ttaagaatct cactcaaagc cgctcaacta catggaaact gaacaacctg ctcctgaatg
2761 actactgggt acataacgaa atgaaggcag aaataaagat gttctttgaa accaacgaga
2821 acaaagacac cacataccag aatctctggg acgcattcaa agcagtgtgt agagggaaat
2881 ttatagcact aaatgcctac aagagaaagc aggaaagatc caaaattgac accctaacat
2941 cacaattaaa agaactagaa aagcaagagc aaacacattc aaaagctagc agaaggcaag
3001 aaataactaa aatcagagca gaactgaagg aaatagagac acaaaaaacc cttcaaaaaa
3061 tcaatgaatc caggagctgg tttttgaaa ggatcaacaa aattgataga ccgctagcaa
3121 gactaataaa gaaaaaaga gagaagaatc aaatagacac aataaaaaat gataaagggg
3181 atatcaccac cgatcctaca gaaatacaaa ctaccatcag agaatactac aaacacctct
3241 atgcaaataa actagaaaat ctagaagaaa tggatacatt cctcgacaca tacactctcc
3301 caagactaaa ccaggaagaa gttgaatctc tgaatagacc aataacaggc tctgaaattg
3361 tggcaataat caatagttta ccaaccaaaa agagtccagg accagatgga ttcacagccg
3421 aattctacca gaggtacaag gaggaactgg taccattcct tctgaaacta ttccaatcaa
3481 tagaaaaaga gggaatcctc cctaactcat tttatgaggc cagcatcatt ctgataccaa
3541 agccgggcag agacacaacc aaaaaagaga attttagacc aatatccttg atgaacattg
3601 atgcaaaaat cctcaataaa atactggcaa accgaatcca gcagcacatc aaaaagctta
3661 tccaccatga tcaagtgggc ttcatccctg ggatgcaagg ctggttcaat atacgcaaat
3721 caataaatgt aatccagcat ataaacagag ccaaagacaa aaaccacatg attatctcaa
3781 tagatgcaga aaaagccttt gacaaaattc aacaacccтt catgctaaaa actctcaata
3841 aattaggtat tgatgggacg tatttcaaaa taataagagc tatctatgac aaacccacag
3901 ccaatatcat actgaatggg caaaaactgg aagcattccc tttgaaaact ggcacaagac
3961 agggatgccc tctctcaccg ctcctattca acatagtgtt ggaagttctg gccagggcaa
4021 tcaggcagga gaaggaaata aagggtattc aattaggaaa agaggaagtc aaattgtccc
4081 tgtttgcaga cgacatgatt gtttatctag aaaaccccat catctcagcc caaaatctcc
4141 ttaagctgat aagcaacttc agcaaagtct caggatacaa aatcaatgta caaaaatcac
```

456

```
4201 aagcattctt atacaccaac aacagacaaa cagagagcca aatcatgggt gaactcccat
4261 tcacaattgc ttcaaagaga ataaaatact taggaatcca acttacaagg gatgtgaagg
4321 acctcttcaa ggagaactac aaaccactgc tcaaggaaat aaaagaggac acaaacaaat
4381 ggaagaacat tccatgctca tgggtaggaa gaatcaatat cgtgaaaatg gccatactgc
4441 ccaaggtaat ttacagattc aatgccatcc ccatcaagct accaatgact ttcttcacag
4501 aattggaaaa aactacttta aagttcatat ggaaccaaaa aagagcccgc attgccaagt
4561 caatcctaag ccaaaagaac aaagctggag gcatcacact acctgacttc aaactatact
4621 acaaggctac agtaaccaaa acagcatggt actggtacca aaacagagat atagatcaat
4681 ggaacagaac agagccctca gaaataatgc cgcatatcta caactatctg atctttgaca
4741 aacctgagaa aaacaagcaa tggggaaagg attccctatt taataaatgg tgctgggaaa
4801 actggctagc catatgtaga aagctgaaac tggatccctt ccttacacct tatacaaaaa
4861 tcaattcaag atggattaaa gacttaaatg ttagacctaa aaccataaaa accctagagg
4921 aaaacctagg cattaccatt caggacatag gcgtgggcaa ggacttcatg tccaaaacac
4981 caaaagcaat ggcaacaaaa accaaaattg acaaatggga tctaattaaa ctaaagagct
5041 tttgcacagc aaaagaaact accatcagag tgaacaggca acctacaaca tgggagaaaa
5101 tttttgcaac ctactcatct gacaaagggc taatatccag aatctacaat gaactcaaac
5161 aaatttacaa gaaaaaaaca aacaaccca tcaaaaagtg ggcgaaggac atgaacagac
5221 acttctcaaa agaagacatt tatgcagcca aaaaacacat gaaaaaatgc tcatcatcac
5281 tggccatcag agaaatgcaa atcaaaacca ctatgagata tcatctcaca ccagttagaa
5341 tggcaatcat taaaaagtca ggaaacaaca ggtgctggag aggatgtgga gaaataggaa
5401 cactttaca ctgttggtgg gactgtaaac tagttcaacc attgtggaag tcagtgtggc
5461 gattcctcag ggatctagaa ctagaaatac catttgaccc agccatccca ttactgggta
5521 tatacccaaa ggactataaa tcatgctgct ataaagacac atgcacacgt atgtttattg
5581 cggcactatt cacaatagca aagacttgga accaacccaa atgtccaaca atgatagtct
5641 ggattaagaa aatgtggcac atatacacca tggaatacta tgcagccata aaaaatgatg
5701 agttcatatc ctttgtaggg acatggatga aattggaaac catcattctc agtaaactat
5761 cgcaagaaca aaaaaccaaa cactgcatat tctcactcat aggtgggaat tgaacaatga
5821 gatcacatgg acacaggaag gggaatatca cactctgggg actgtggtgg ggtcgggggga
5881 ggggggaggg atagcattgg gagatatacc taatgctaga tgacacatta gtgggtgcag
5941 tgcaccagca tggcacatgt atacatatgt aactaacctg cacaatgtgc acatgtaccc
6001 taaaactgag agtataataa aaaaaaaaaa ataaaaaata aaaaaaaaaa gaaaaaaaaa
6061 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa
```

Appendix III.5

Contents

Figure III.5.1 - Integration of homozygosity mapping in the analysis workflow of whole-exome sequencing

Appendix IV.1

Contents
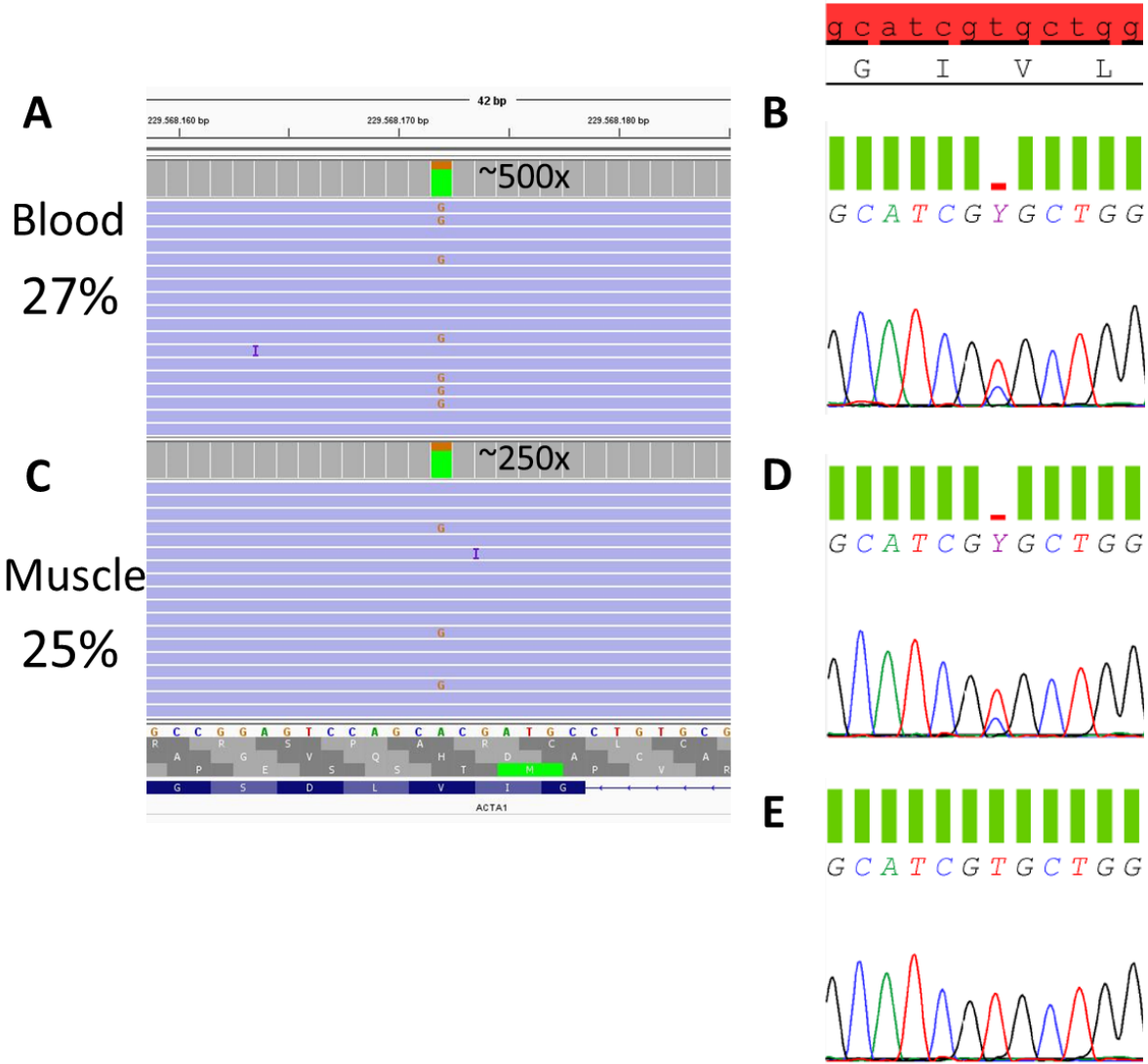
Figure IV.1.1 – Novel *ACTA1* variant (NM_001100.3:c.461T>C, p.Val154Ala) found in apparent mosaicism in patient #13. The screening of variants was initially performed by the (high coverage) NGS gene panel for congenital myopathies in a DNA sample obtained from blood (A) and confirmed by Sanger sequencing (B). This analysis was subsequently confirmed in an additional DNA sample extracted from the patient's muscle tissue (C and D). A control sequencing electropherogram is shown in (E). NGS analysis allowed to estimate the representativity of the variants: ~25% of reads in muscle and 27% in blood DNA.

Appendix IV.2

Contents

A

B

Figure IV.2.1 – (A) Family pedigree of patient #11 (red arrow highlights this proband). Symptoms start shortly after birth, with severe tetraparesis, hypotonia, bulbar and facial weakness. Respiratory muscles' weakness requiring invasive ventilation. She also had a high arched palate and clubfoot. Patient's sibling died at 2 months of age with a similar severe phenotype of a congenital neuromuscular disease. (B) Analysis of a muscle biopsy performed in the patient's sibling showed frequent rod-like inclusions (nemaline bodies) in the Gomori Trichrome stain sarcoplasm suggestive of a nemaline myopathy.
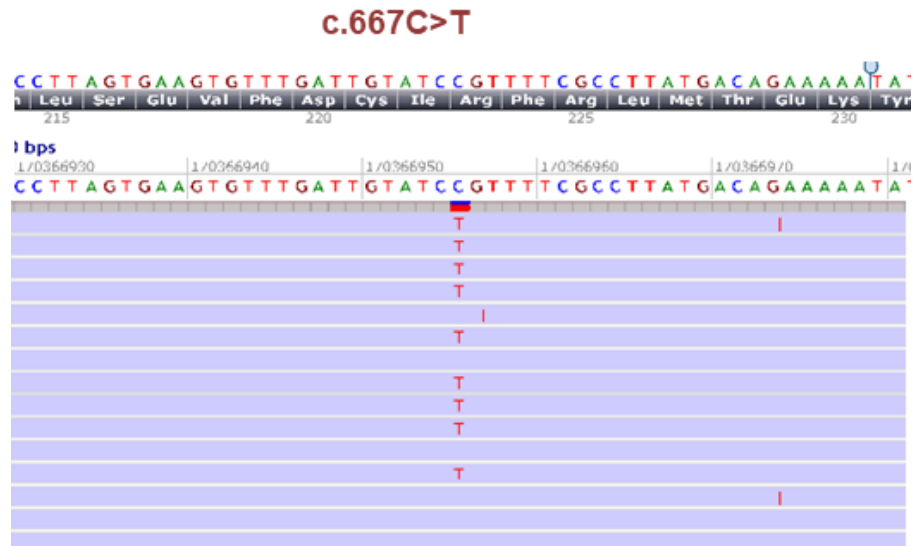
Figure IV.2.2 – From the whole-exome sequencing (WES) data obtained for patient #11 it was not possible to identify a genetic cause consistent both with the phenotype and an autosomal recessive inheritance (the most likely to be involved in this family). However, a single heterozygous missense variant (NM_006063.2:c.667C>T, p.Arg223Cys) in *KLHL41,* not previously reported in the literature, was identified in the patient.

Figure IV.2.3 – In an attempt to identify an additional variant and to complement WES analysis, an exon level microarray (CytoScan XON, Thermo Fisher Scientific) was performed. Results showed a large heterozygous deletion in *KLHL41*, affecting exon 2 and 3 (the later apparently only partially).

c.1110+1836_1276delinsACAC

AG AA CG G TAT G CC A G G ATA A A A A A CA CA AT G G A A C G A A G TA A A A A A A C T
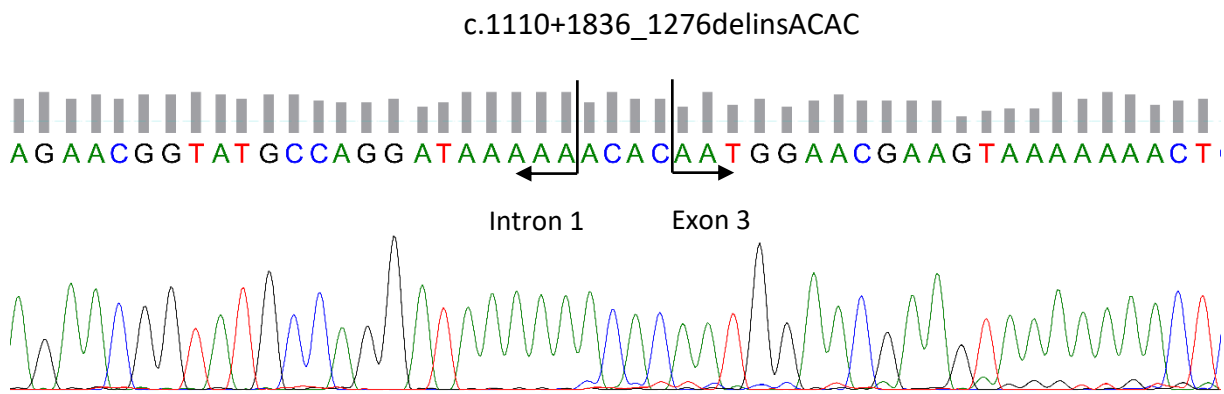
Intron 1   Exon 3

Figure IV.2.4 – Long-range PCR followed by sequencing showed genomic deletion breakpoints, involving from part of intron 1 (c.1110+1836) to the first eight base pairs of exon 3 (c.1276) and insertion of four bases (ACAC). This deletion totalizes 2.1 Kbp.

Figure IV.2.5 – Further experiments at the mRNA level revealed the absence of exons 2 and 3 (r.1111_1376del) in some *KLHL41* transcripts, while others have the c.667C>T substitutions. In addition, family studies confirmed compound heterozygosity for the two variants and co-segregation with the disease (both variants are present in the siblings).

Data IV.2.1 – Abstract from the work presented in the 22nd meeting of the Portuguese Society of Human Genetics (2018).

Title: Extending genetic analysis of hereditary myopathies beyond next-generation targeted and exome sequencing

Authors: Ana Gonçalves; Manuela Santos; Ana Fortuna; Teresa Coelho; Ricardo Taipa; Manuel Melo-Pires; Johan T. Den Dunnen; Mário Sousa; Rosário Santos; Jorge Oliveira

Background

Next-generation sequencing applications have significantly improved the research and diagnostic yield of genetically heterogeneous conditions. A subset of patients remains unsolved even upon WES analysis. We describe 3 complex cases of hereditary myopathies.

Methods

P1: new-born; severe nemaline myopathy; brother with similar phenotype. A combination of WES, exon-level microarray (XON), RNA and genomic breakpoint sequencing was done.

P2 and P3: suspected Becker Muscular Dystrophy. No pathogenic variants in *DMD* were identified by conventional testing. In P2, bioinformatic, transcriptomic and genomic approaches (SB, long-range PCR and single-molecule real-time sequencing) were applied. In P3, RT-PCR, cDNA-MLPA, low-coverage whole genome sequencing (LC-WGS) and breakpoint sequencing approaches were used.

Results

P1: WES failed to explain an autosomal-recessive condition. XON array showed a large heterozygous deletion in *KLHL41*, where WES had identified a single heterozygous missense variant. This novel deletion was confirmed by RNA and genomic breakpoint sequencing. Family studies confirmed compound heterozygosity for the two variants and co-segregation with the disease.

P2: An aberrant transcript was identified, containing a 103nt insertion between *DMD* exons 51 and 52, with no similarity to the gene. This corresponded to the partial exonization of a LINE-1 sequence. Further characterization identified the deep-intronic insertion of a complete LINE-1.

P3: *DMD* cDNA studies disclosed the absence of exons 75-79. Automated structural variant calling from LC-WGS was inconclusive, but binary alignment map file inspection showed a putative breakpoint within intron 74, as some reads had homology with a region upstream of *PRDX4* (Xp22.11). Breakpoint sequencing showed a ~8Mb inversion comprising part of *DMD* and upstream of *PRDX4*.

Conclusions

This work revealed unexpected and hitherto unknown mutational events underlying some myopathies. Together with solid bioinformatics, WGS complemented by transcriptome analysis has the potential to detect the majority of mutation types.

Funding

Appendix IV.3

Contents

Figure IV.3.1 – Pedigree of patient #20's (III:3) family depicting a typical autosomal dominant inheritance pattern.

Total reads: 2,441,340

Bases in target regions	3,566,796
Average base coverage depth	113.3
Uniformity of base coverage	94.97%
Target base coverage at 20x	95.55%
Target bases with no strand bias	95.65%
Percent end-to-end reads	81.96%
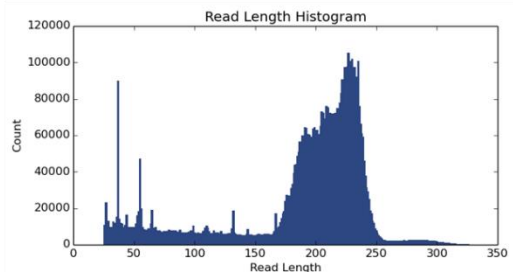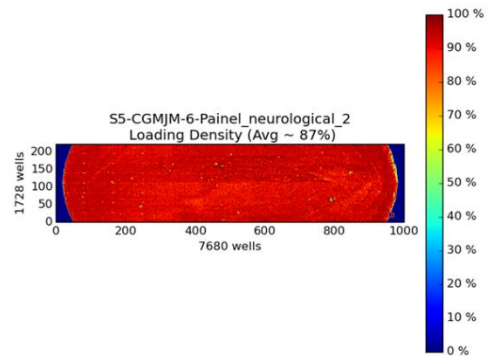
Figure IV.3.2 – Details of the neurology gene panel sequencing run performed in patient #20 and its respective metrics.
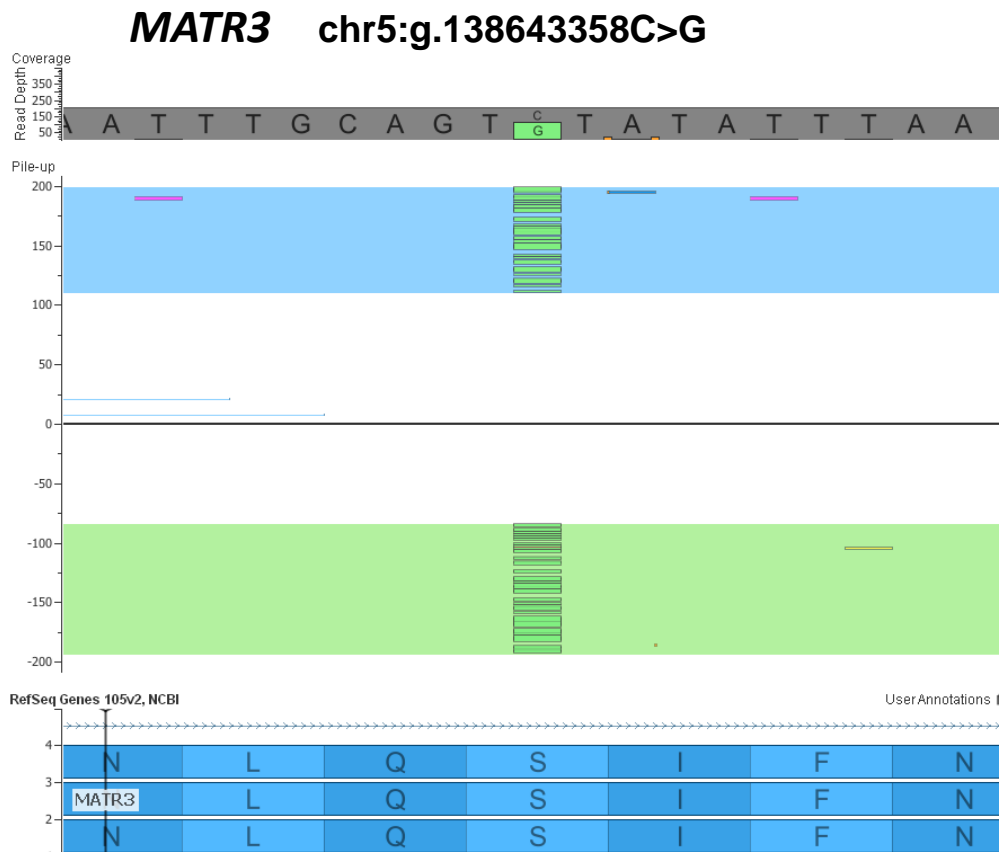
Figure IV.3.3 – Visualization of the patient #20 BAM file showing the NM_199189.2:c.254C>G heterozygous variant found in *MATR3* gene, which predictably leads to a missense substitution (p.Ser85Cys).
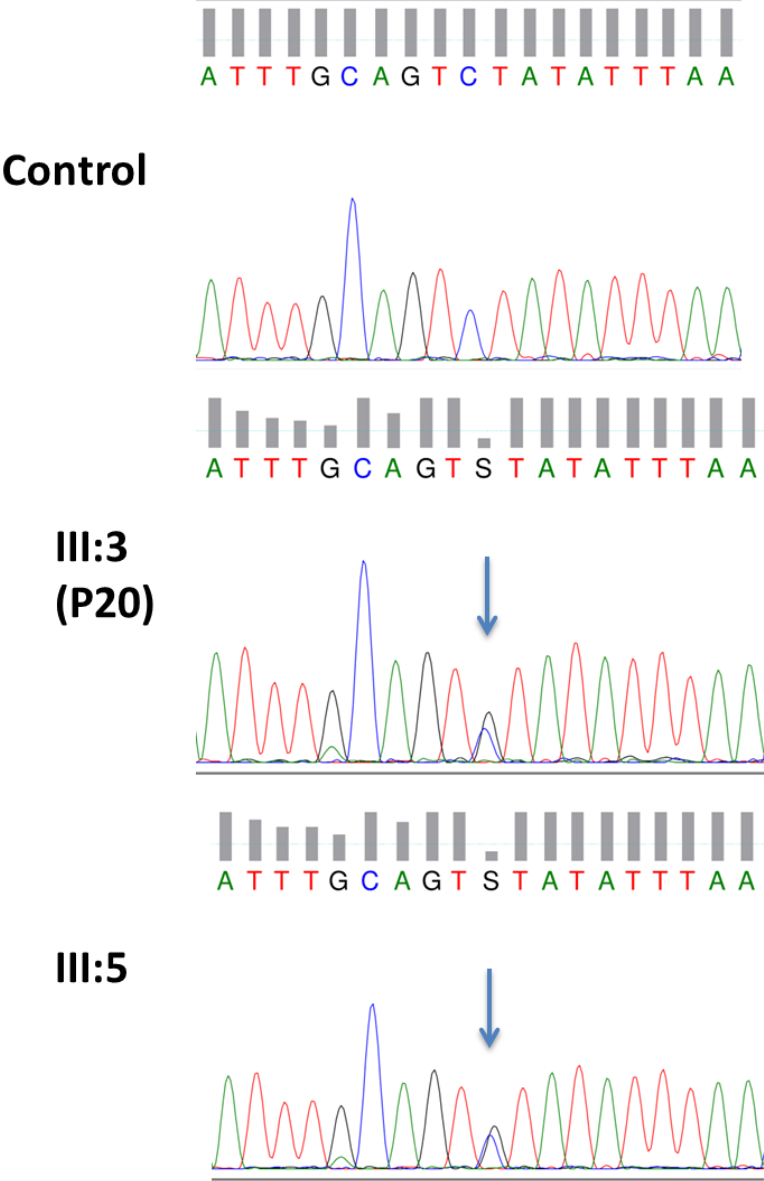
Figure IV.3.4 – Partial sequencing electropherogram of exon 5 of *MATR3* gene, in a control, patient III:3 (patient #20) and III:5, confirms the presence of the NM_199189.2:c.254C>G variant in the patients (arrows).

Appendix IV.4

Contents

1999
Electromyography: myopathic
IHC in muscle biopsy:
    Dys1,3 irregular, Dys2 absent
    SGs & Merosin normal

DMD **Multiplex PCR** (18 exons)
& Southern Blot (6 probes):
No deletions
2000

2005
DMD **MLPA:**
No deletions/duplications

FMR1 & AFF2: negative
Karyotype: 46 XY
Subtelomeric MLPA: Normal
2008

2009
DMD **sequencing:**
No pathogenic variants

2nd IHC muscle biopsy:
Dys irregular
α-DG Normal
2010

2012
Multiplex Western Blot:
DYS1 reduced;
DYS2,3 absent

FKRP, SGCB, SGCG and *SGCD*
sequencing: negative
SGCA: heterozygous variant of
unknown clinical significance (VUS)
2012

PREVIOUS DIAGNOSTIC STUDIES

Figure IV.4.1 – Previous diagnostic studies performed in patient #21.
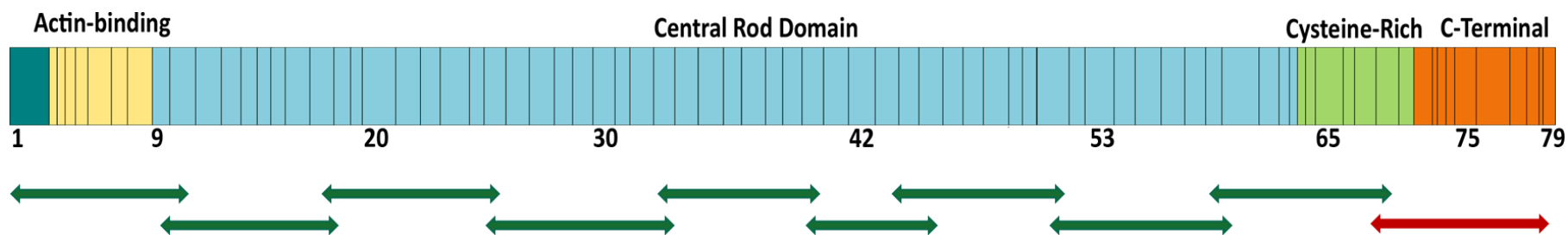
Figure IV.4.2 –*DMD* transcripts (cDNA) were amplified using 10 overlapping fragments covering the entire sequence; one amplicon (red arrow) corresponding to exons 67 to 79 failed to amplify.
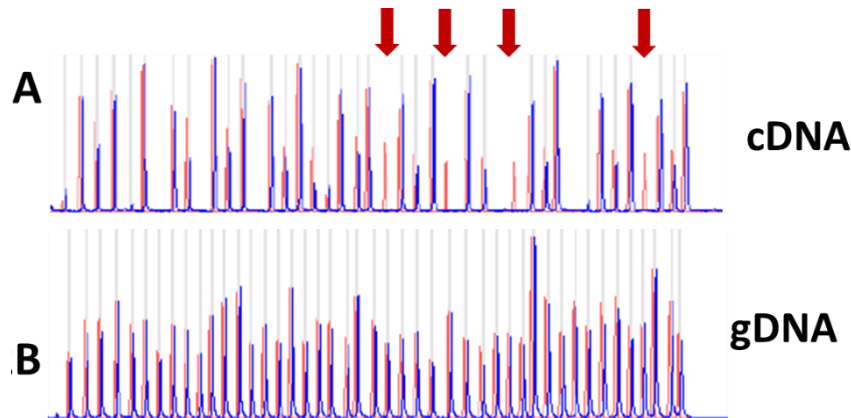
Figure IV.4.3 – Multiplex ligation-dependent probe amplification (MLPA) was performed in the patient #21's cDNA sample revealed the absence of signal from probes detecting the sequence corresponding to exons 75, 76, 77 and 79 of *DMD* (A), whereas at the genomic level no deletion was detected (B).
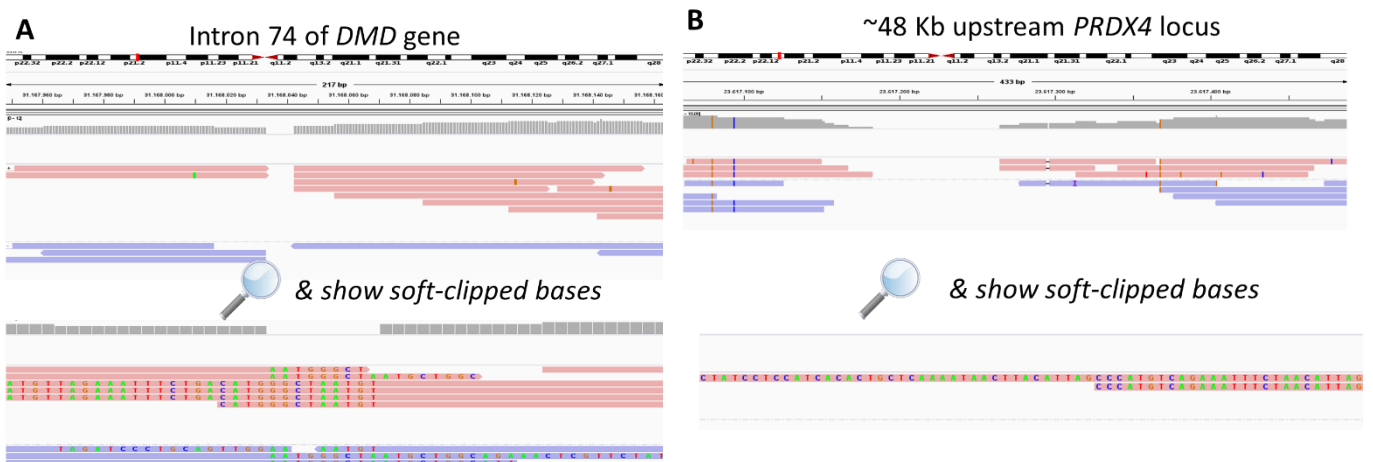
Figure IV.4.4 – Low-coverage whole-genome sequencing (WGS) was performed using the Chromium system (10x Genomics). Structural variants automated calling from WGS data using Long Ranger software (10x Genomics) was inconclusive. Visual BAM file inspection using IGV showed a possible breakpoint within intron 74 of *DMD* (A). Some reads contained soft-clipped bases with homology with a region 48 Kb upstream of the *PRDX4* gene (Xp22.1). Correspondingly, some partially aligned reads in *PRDX4* showed homology with inverted intron 74 *DMD* sequences (B).
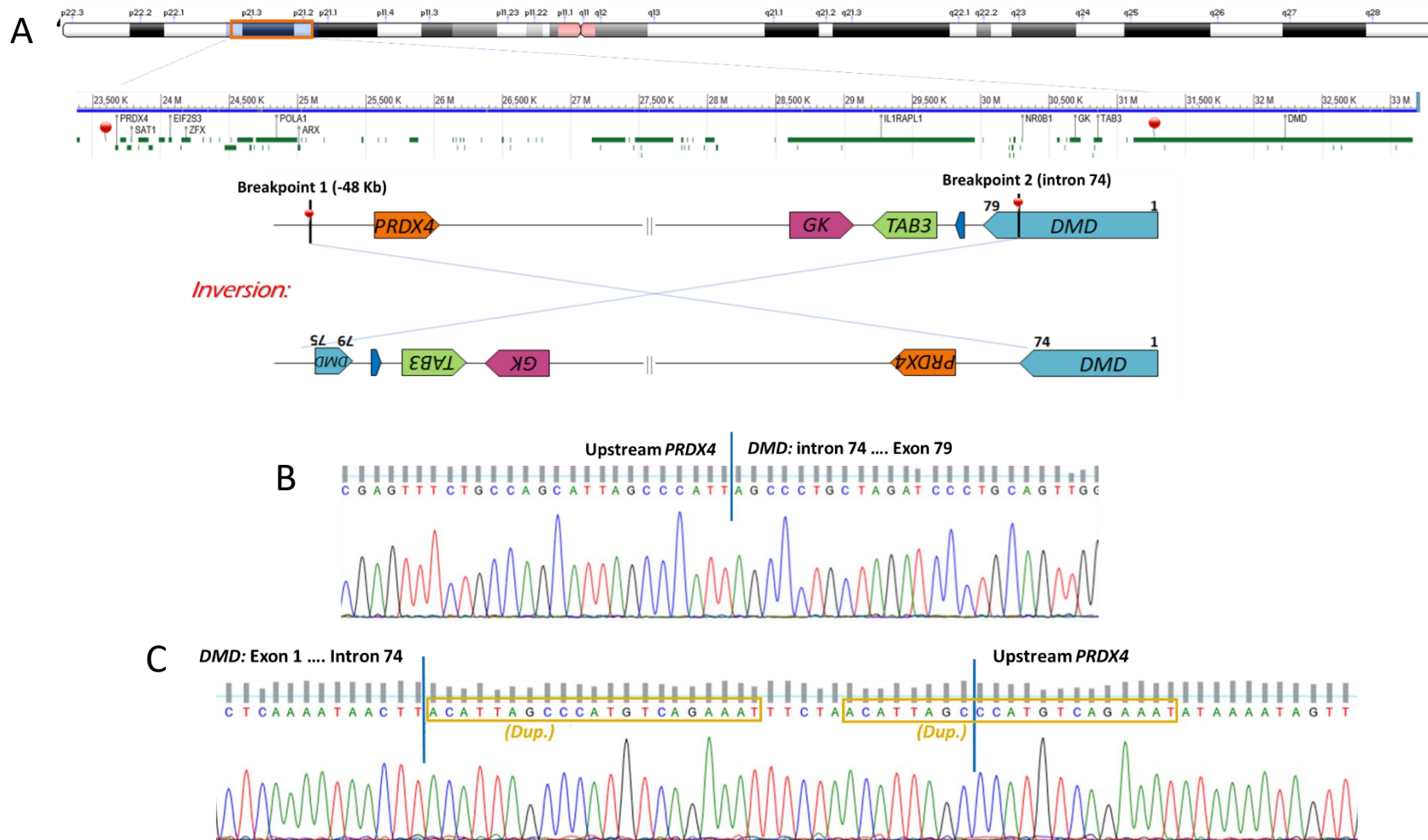
Figure IV.4.5 – Schematic representation of the inversion at the genomic level (A). The inversion involves the region Xp22.1 to Xp21.2 (~8 Mb),

comprising 74 genes. Breakpoints are localized 48 kb upstream of *PRDX4* (breakpoint 1) and within intron 74 of *DMD* (breakpoint 2). Breakpoints

were confirmed by Sanger sequencing (B and C). In the second breakpoint (C) the inversion originated a 21 bp duplication (dup.) and the inclusion

of a short sequence with no homology to the reference genome. There was a loss of 9 bp in intron 74 of *DMD* and a loss of 3 bp in the upstream

region of *PRDX4*.
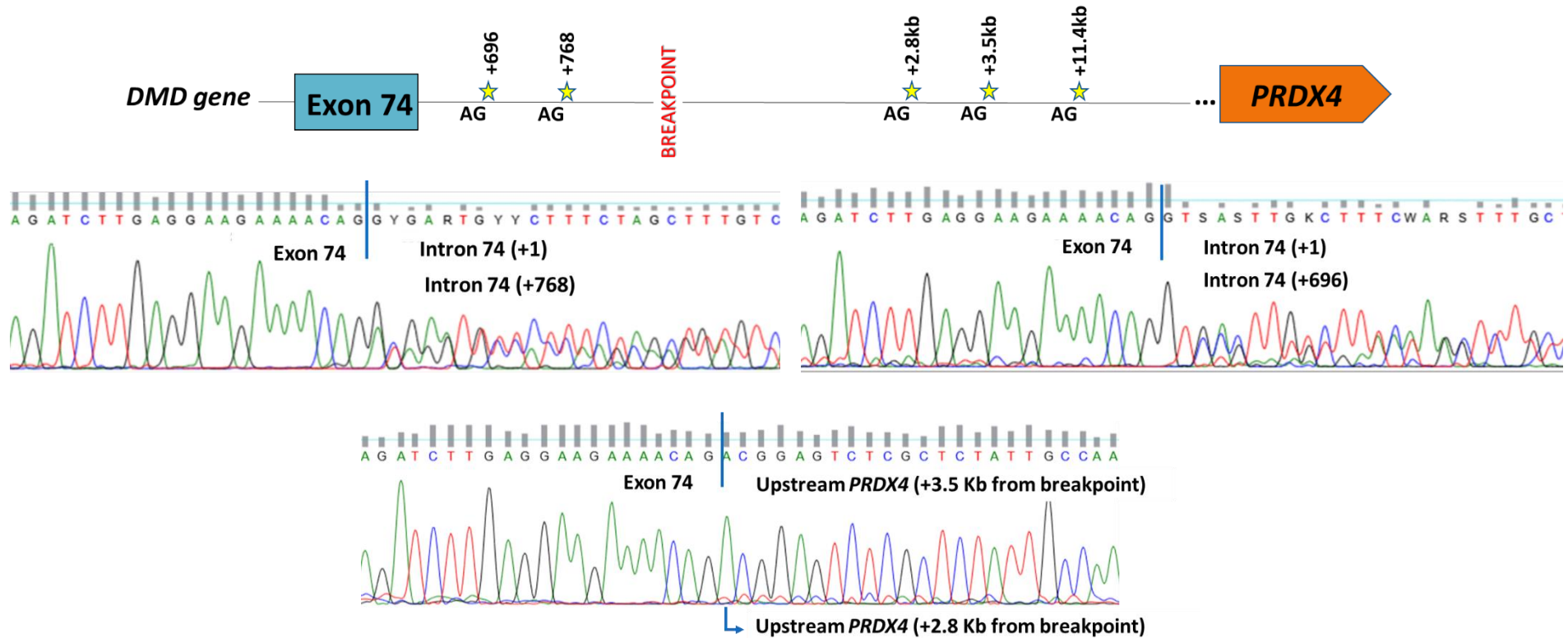
Figure IV.4.6 – Several transcripts were detected in the patient, representing the use of cryptic splice-sites located in *DMD* (intron 74) as well as in the region upstream of *PRDX4* ("chimeric" transcripts).

Data IV.4.1 – Abstract from the work presented in the 2018 European Human Genetics conference

Title: Whole-genome sequencing detects a large genomic inversion disrupting the *DMD* gene in a Becker muscular dystrophy patient

Authors: J. Oliveira , A. Gonçalves , Y. Ariyurek , J. T. den Dunnen , M. Sousa , R. Santos ;

Introduction

Duchenne and Becker muscular dystrophies (D/BMD) are caused by pathogenic variants in the dystrophin (DMD)

gene. Being the largest locus of the human genome (2.3Mb, Xp21.1), DMD is particularly prone to genomic rearrangements, often intragenic multi-exonic deletions or duplications (~70% of cases). Single nucleotide variants are identified in most of the remaining D/BMD cases, whereas complex genomic rearrangements encompassing DMD are rarer.

Materials and Methods

Routine techniques failed to detect pathogenic DMD variants in a BMD patient with progressive muscle weakness, mild intellectual disability and dystrophic muscle showing irregular dystrophin labelling. DMD transcript analysis (RT-PCR and cDNA-MLPA) was performed, as well as low-coverage whole-genome sequencing (WGS) using the 10xGenomics Chromium system.

Results

RNA analysis showed the absence of exons 75-79. While automated structural variant calling from WGS was inconclusive, visual BAM Ile inspection showed a possible breakpoint within intron 74 of DMD. Some reads had homology with a region located upstream of the PRDX4 gene (Xp22.11). Similarly, some reads in that location showed homology with inverted intron 74 DMD sequences. Breakpoint PCR and Sanger sequencing confirmed the presence of a ~8Mb inversion. Abnormal DMD transcripts were subsequently identified, some of which contained segments from the region upstream of PRDX4.

Conclusions

The patient's phenotype is explainable by this DMD inversion with concomitant loss of the C-terminal region of dystrophin. Besides expanding the DMD mutational spectrum, this report reinforces the importance of WGS in clinical genetics, having the potential to detect a wide variety of mutation types.

CHP grant reference 336-13(196-DEFI/285-CES).

Appendix V

Contents

ARTICLE

Expanding the *MTM1* mutational spectrum: novel variants including the first multi-exonic duplication and development of a locus-specific database

Jorge Oliveira[1,8], Márcia E Oliveira[1,8], Wolfram Kress[2], Ricardo Taipa[3], Manuel Melo Pires[3], Pascale Hilbert[4], Peter Baxter[5], Manuela Santos[6], Henk Buermans[7], Johan T den Dunnen[7] and Rosário Santos*,[1]

Myotubular myopathy (MIM#310400), the X-linked form of Centronuclear myopathy (CNM) is mainly characterized by neonatal hypotonia and inability to maintain unassisted respiration. The *MTM1* gene, responsible for this disease, encodes myotubularin – a lipidic phosphatase involved in vesicle trafficking regulation and maturation. Recently, it was shown that myotubularin interacts with desmin, being a major regulator of intermediate filaments. We report the development of a locus-specific database for *MTM1* using the Leiden Open Variation database software (http://www.lovd.nl/MTM1), with data collated for 474 mutations identified in 472 patients (by June 2012). Among the entries are a total of 25 new mutations, including a large deletion encompassing introns 2–15. During database implementation it was noticed that no large duplications had been reported. We tested a group of eight uncharacterized CNM patients for this specific type of mutation, by multiple ligation-dependent probe amplification (MLPA) analysis. A large duplication spanning exons 1–5 was identified in a boy with a mild phenotype, with results pointing toward possible somatic mosaicism. Further characterization revealed that this duplication causes an in-frame deletion at the mRNA level (r.343_444del). Results obtained with a next generation sequencing approach suggested that the duplication extends into the neighboring *MAMLD1* gene and subsequent cDNA analysis detected the presence of a *MTM1/MAMLD1* fusion transcript. A complex rearrangement involving the duplication of exon 10 has since been reported, with detection also enabled by MLPA analysis. It is thus conceivable that large duplications in *MTM1* may account for a number of CNM cases that have remained genetically unresolved.
European Journal of Human Genetics (2013) **21**, 540–549; doi:10.1038/ejhg.2012.201; published online 12 September 2012

Keywords: locus-specific database; *MTM1*; novel mutations

INTRODUCTION

Congenital myopathies are a heterogeneous group of diseases, generally characterized by muscle weakness, and with onset at birth or during infancy. These myopathies have characteristic histological hallmarks in muscle biopsy allowing the differential classification in distinct entities: centronuclear myopathy (CNM), core myopathy (centralcore and minicore diseases) and nemaline rod myopathy.[1,2] In CNM, the most prominent histopathological features include hypotrophy of type 1 fibers and a high frequency of centrally located nuclei with perinuclear halos lacking myofilaments and occupied by mitochondrial and glycogen aggregates.[1] Several genes are reported to be associated with CNM; these include *MTM1* in the X-linked form,[3,4] *DNM2* and *MTMR14* in the autosomal dominant forms,[4–6] *BIN1* and *RYR1* associated with the autosomal recessive forms.[4,7–9]

X-linked myotubular myopathy (XLMTM; MIM 310400) has a prevalence of approximately 1/50 000 males and is characterized by severe hypotonia present at birth and inability to maintain sustained

spontaneous respiration.[10] Different authors have proposed that patients be classified according to their phenotype, as: (i) *severe* – characteristic facial features, markedly delayed motor milestones and requiring prolonged ventilatory support (>12 h); (ii) *moderate* – more rapid acquirement of motor milestones and independent respiration for >12 h per day; (iii) *mild* – motor milestones slightly delayed and independent spontaneous respiratory function achieved after the neonatal period.[11,12] Carrier females are usually asymptomatic, however there are several records of manifesting heterozygotes due to skewed X chromosome inactivation.[13–18]

The *MTM1* gene (in Xq28) is composed of 15 exons and has an open reading frame of 1.8 kb encoding the myotubularin protein. Structurally, myotubularins are constituted by four characteristic domains: the protein tyrosine phosphatase (PTP), the predicted GRAM (glucosyltransferases, Rab-like GTPases activators and myotubularins), the RID (Rac-induced recruitment domain) and SID (SET-interaction domain). Functionally, myotubularin is a

[1]Unidade de Investigação e Desenvolvimento, Departamento de Genética, Centro de Genética Médica Dr Jacinto Magalhães, Instituto Nacional de Saúde Dr Ricardo Jorge, IP, Porto, Portugal; [2]Institute of Human Genetics, University of Wuerzburg, Wuerzburg, Germany; [3]Unidade de Neuropatologia, Centro Hospitalar do Porto, Porto, Portugal; [4]Département de Biologie Moléculaire, Institut de Pathologie et de Génétique ASBL, Gosselies, Belgium; [5]Department of Paediatric Neurology, Sheffield Children's Hospital, Sheffield, UK; [6]Consulta de Neuromusculares, Centro Hospitalar do Porto, Porto, Portugal; [7]Leiden Genome Technology Center, Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
*Correspondence: Dr R Santos, Unidade de Investigação e Desenvolvimento, Departamento de Genética, Centro de Genética Médica Dr Jacinto Magalhães, Instituto Nacional de Saúde Dr Ricardo Jorge, IP, Praça Pedro Nunes, 88, 4099-028 Porto, Portugal. Tel: +351 22 607 0330; Fax: +351 22 607 0399;
E-mail: rosario.santos@insa.min-saude.pt
[8]These authors contributed equally to this work.
Received 18 April 2012; revised 17 July 2012; accepted 8 August 2012; published online 12 September 2012

phosphatase acting specifically on PtdIns3P and PtdIns(3,5)P2, two phosphoinositides (PIs). PIs participate in the regulation of various cellular mechanisms by direct binding to PI-binding domains of effector proteins (that control membrane/vesicular trafficking) and subsequent recruitment/activation of these at specific membrane sites. PtdIns3P and PtdIns(3,5)P2 have a direct role in the endosomal-lysosomal pathway.[19] The PTP-catalytic domain of myotubularins is responsible for the phosphoester hydrolysis of the 3-phosphate of PIs. This hydrolysis involves two residues of cysteine and arginine located on a Cys-X5-Arg motif, characteristic for the PTP domain.[20] The loss of phosphatase activity or the production of truncated proteins as a result of *MTM1* mutations could lead to abnormal dephosphorylation of PtdIns3P/PtdIns(3,5)P2 and subsequent abnormal trafficking of the effector proteins of the endosomal-lysosomal pathway.[19] Similar results were observed with mutations located in the GRAM domain of myotubularin. This domain of about 70 amino acids is responsible for PtdIns(3,5)P2 binding.[21] Recently mitochondrial homeostasis in muscle fibers and regulation of the desmin cytoskeletal system was attributed to mytobularin. It was experimentally demonstrated that myotubularin interacts with desmin, and that this complex is disrupted by specific mutations in the *MTM1* gene.[22]

It is recognized that, despite the genetic advances in this field and the large number of cases reported,[23] a significant number of CNM cases remain genetically unresolved. This may be explained either by the involvement of further gene *loci* or by the presence of mutations in known genes, that are not detectable by routine techniques. During the development of a locus-specific database (LSDB) for the *MTM1* gene we noticed that no large duplications (involving one or more exons) had been reported for this gene. This observation led us to investigate the possibility of their occurrence in molecularly unresolved CNM patients. Accordingly, we report the first multi-exonic duplication in *MTM1* (exons 1–5) detected in a male patient with a mild XLMTM phenotype. In line with this finding, a complex rearrangement involving the duplication of exon 10 was recently published.[24]

MATERIALS AND METHODS
MTM1-LOVD database development
The *MTM1* mutation database was implemented using LOVD v2.0 software,[25] and is integrated in Leiden Muscular Dystrophy pages (http://www.dmd.nl/). Currently (by 29th June 2012), this large database installation displays a total of 141 genes with 70 512 variants (9265 unique) described in 55 775 individuals. The *MTM1* database is subdivided into two main tables, for variant data ($n = 20$) and for patient/clinical items ($n = 22$), the latter being shared among the different LSDBs in the Leiden Muscular Dystrophy pages (example of an entry in Supplementary Figure S1). Data were retrieved from peer-reviewed literature and new variants were directly submitted by different sources (laboratories and clinicians). The curators' tasks included confirmation of information concerning the sequence variants, especially with regard to their descriptions following the recommendations of the Human Genome Variation Society (HGVS),[26] using the cDNA reference sequence NM_000252.2. Classification of the clinical phenotype (severe, moderate or mild) was added to the patient's entry when adequate information was reported or submitted.

New MTM1 variants
Variants submitted to MTM1-LOVD that had not been described previously, are presented here. These were reported from different centers (Belgium, Germany and Portugal) and submitted to MTM1-LOVD. Information regarding the patient's phenotype and the mutation origins was also collected. Bioinformatic analysis of sequence variants, in particular missense and splicing mutations, was performed using Polyphen v2 (http://genetics.bwh.harvard.edu/pph2/)[27] and the Human Splicing Finder v2.3 (http://www.umd.be/HSF/).[28]

Phylogenetic conservation analysis of the affected residues and population screening of variants were also performed.

Studies performed in patient 1
Clinical description. Patient 1 (P1) is a 7-year-old boy born to a non-consanguineous couple. He has a healthy younger sister. Prenatal manifestations included oligohydramnios and signs of premature birth. Delivery was at 35 weeks by cesarean section, with an Apgar score of 8/9, weight 2465 g, height 46 cm and head circumference 33 cm. During the neonatal period the only clinically relevant sign was facial paresis. There were no feeding or respiratory difficulties. During the first month his pediatrician noticed that he was hypotonic and started global stimulation. He started to walk at 21 months. At the age of two he was referred to a pediatric neurology clinic. He was shy and had poor language skills. There was limitation in the abduction of the right eye, facial diparesis with the left side more exacerbated, with a lagophtalmos. Proximal tetraparesis was detected; the patient could not raise his arms completely and was unable to stand up from the floor without bilateral help. He had a waddling gait and could not run. CK levels were normal (36 U/l). Biopsy was performed at age 3. Over the last 3 years muscle weakness progressed and there has been a gradual loss of motor skills.

Multiple ligation-dependent probe amplification (MLPA) analysis. Screening for duplications and deletions in *MTM1* was performed by the MLPA technique using the P309-A1 Probe Set (MRC-Holland, Amsterdam, The Netherlands). This contains 16 probes for the *MTM1* gene, 7 probes for the *MTMR1* gene, 3 probes located on Xq28 (*DKC1* and *FLNA* genes) and 11 reference probes for distinct regions on the X chromosome (Supplementary Table S2). gDNA samples of P1 and four healthy controls (150 ng each) were used in the procedure. Products were separated by capillary electrophoresis on an ABI 3130xl genetic analyzer (Applied Biosystems, Foster City, CA, USA). Data analysis was conducted using GeneMarker software (SoftGenetics LLC, State College, PA, USA). Population normalization method was selected and data were plotted using probe ratio.

Southern blotting and hybridization. gDNA samples from P1, his mother and controls were digested with *Eco*RI (New England Biolabs, Beverly, MA, USA) and resolved on a 0.8% agarose gel. DNA fragments were vacuum-transferred to a nylon membrane using a saline method. A cDNA probe recognizing *MTM1* exons 2–7 was prepared using digoxigenin (DIG) DNA Labeling Kit (Roche Applied Science, Indianapolis, IN, USA) and incubated overnight using the Easy Hyb Buffer (Roche Applied Science). The membrane was prepared with DIG Wash and Block Buffer Set (Roche Applied Science), incubated with Anti-DIG-AP conjugate (Roche Applied Science), and the DIG-labeled probe detected with ready-to-use CDP-Star (Roche Applied Science).

cDNA analysis. Total RNA was extracted from muscle samples of P1 and controls using the PerfectPure RNA Fibrous Tissue Kit (5 PRIME, Hamburg, Germany) and converted to cDNA using the High Capacity RNA-to-cDNA Kit (Applied Biosystems). *MTM1* transcripts were subjected to PCR amplification of the region(s) encompassing exons 2–7, using the specific primers (2F: 5′-TC CAGGATGGCTTCTGCATC-3′ and 7R: 5′-CAAGCCCTGCCTCCTGTATTC-3′). For the detection of the *MTM1/MAMLD1* fusion transcript total RNA was extracted from blood using the PerfectPure RNA Blood Kit (5 PRIME). After cDNA conversion, PCR amplification was performed using the primer mentioned above for exon 2 of *MTM1* and a reverse primer for exon 5 of *MAMLD1* (cMAMLD1-5R: 5′-AGTCTGGCCTGAGTGTGAGAGG-3′). PCR products were purified, sequenced using BigDye Terminator Cycle Sequencing Kit v1.1 (Applied Biosystems) and resolved on an ABI 3130xl genetic analyzer (Applied Biosystems).

Next generation sequencing. gDNA was Covaris (S-series) sheared to an average size of 300 to 400 bp. An Illumina sequencing library was prepared using the NEBNext kit (New England Biolabs) without modifications and using Illumina Truseq adapters. No library pre-amplification was performed. Sequencing data was generated on an Illumina Hiseq2000 (Illumina, San Diego, CA, USA) using standard instrument settings for flowcell clustering and paired-end sequencing of 2 × 100 bp reads on a flowcell with v3 reagents. Fastq

Table 1 Overview of mutations in MTM1-LOVD

| DNA | RNA[a] | | | | | Protein[b] | | del/dup/ins/delins | | | | |
Mutation type	r.(0?)	del/dup/ins/subs	r.(spl?)	r.(?)	p.0?	Missense	Nonsense	IF	OF	MP	p.(?)	Total
Substitutions	—	26	62	238	5	145	91	35	8	3	39	326 (152)
Deletions	—	2	7	63	—	—	3	10	52	—	7	72 (50)
Duplications	—	—	1	30	—	—	7	—	23	—	1	31 (23)
Insertions	—	1	—	5	—	—	—	1	5	—	—	6 (6)
Deletion/insertions	—	—	2	5	—	—	1	—	4	—	2	7 (7)
Large deletions[c]	15	2	—	13	18	—	—	10	1	—	1	30 (18)
Large duplications[c]	—	2	—	—	—	—	—	2	—	—	—	2 (2)
												474 (258)

Abbreviations: r.(0?), predictably no expression at RNA level; del, deletion; dup, duplication; ins, insertion; subs, substitution; r.(spl?), predicted to affect splicing; r.(?), RNA effect unknown; p.0?, predictably no protein production; delins, deletion and insertion; IF, in-frame; OF, out-of-frame; MP, multiple polypeptides; p.(?), effect at protein level unknown.
Data extracted from MTM1-LOVD (29/06/2012). Numbers between brackets indicate the total count of unique (different) mutations.
[a]Experimentally tested at RNA level (del, dup, ins, subs) or predicted (remaining columns).
[b]Predicted protein changes inferred from DNA or RNA sequence.
[c]Involving one or more exons.

files were generated using the Illumina Casava v1.8 pipeline. Stampy (v1.0.13) was used to align the sequence data to the human genome (Hg19) using standard parameters. The total number of aligned reads per 10 000 bp bins across the whole-genome was calculated using the Bedtools package.[29] Bins overlapping more than 1% with simple-repeat regions were excluded from further processing. For sample comparison of tag coverage in the *MTM1* locus, the bin counts were normalized relative to total number of reads aligned to the chromosome X. For visualization purposes, BED tracks were generated for the patient, control and the ratio patient/control, and were uploaded to the UCSC genome browser.[30]

Studies performed in patient 2

Clinical description. Patient 2 (P2) was the first child of healthy parents, with no consanguinity or family history of note. After delivery he was very hypotonic and required immediate respiratory support with continuous positive airway pressure and naso jujenal feeding. At 2 months of age he underwent tracheal endoscopy and aryepiglottoplasty after which he needed ventilatory support. After extubation he was transferred for intensive care and was diagnosed with congenital nystagmus and pyloric stenosis. At 3 months he was referred for a neurology opinion. He had a frog like posture, absent antigravity movement in the upper limbs, with some in the lower limbs and areflexia. At pyloromyotomy a quadriceps muscle biopsy was taken, which confirmed the diagnosis of CNM. From 5 months respiratory function improved and by 7 months he was off all respiratory support. He began to lift each leg off the cot for 1–2 min and the nystagmus disappeared, with normal eye movements, but he had persisting marked neck weakness. Following an episode of lung collapse he began to show restriction of lateral and vertical eye movements. He was able to lift his arms against gravity and bring his hands together, as well as to hold a rattle, and was able to lift his legs momentarily against gravity. He had a marked scoliosis, a bell shaped chest and there was early hip and knee flexion contractures. Placed prone he was unable to turn his head or push himself up on his arms. At 8 months he developed bronchiolitis and pulmonary collapse, as well as a brief cardiac arrest. Following this he was re-intubated and given ventilatory support. After recovery he was much less active and lost antigravity limb movement and eye movements. At the age of 10 months he became unwell again and died.

Molecular analysis. Initially gDNA was PCR amplified for subsequent sequencing. As no products were obtained for exons 3–14, a large hemizygous *MTM1* deletion was suspected. Deletion was confirmed by MLPA analysis, as described for P1. The breakpoints of the large deletion encompassing exons 3–14 were determined by long-range PCR. Amplification of gDNA was performed using the BIO-X-ACT Long DNA Polymerase Kit (Bioline, Taunton, MA, USA) and primers complementary to introns 2 and

14 of *MTM1* (2iF: 5′-GAAAGGTTGCTGAAGGACATACTG-3′ and 14iR: 5′-GCCTTGGGTATGAATGCTGG-3′). PCR products were resolved on a 2% agarose gel and purified with QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA, USA), followed by sequencing.

RESULTS

MTM1-LOVD

A total of 496 variants are currently (June 2012) listed in the *MTM1* LSDB. Besides non-pathogenic variants ($n = 18$) or sequence changes with unknown significance ($n = 4$), this database includes 474 *MTM1* mutations identified in 472 XLMTM patients. The mutation profile of the *MTM1* gene, according to LOVD-MTM1 data, is summarized in Table 1 and Figure 1. The most represented group of mutations is single-nucleotide substitutions (68.8% of total mutations), which can be subdivided as: missense ($n = 145$), nonsense ($n = 91$), those affecting splicing ($n = 85$) or translation ($n = 5$). Deletions were identified in 102 patients. These include 'small' deletions reported in 72 patients (15.2% of total mutations) the majority of which (76.4%) predictably induce premature termination codons (PTC). Large deletions encompassing one or more *MTM1* exons account for 6.3% of mutations reported. The eighteen different large deletions reported to date (identified in 30 cases) are detailed in Figure 1b. Small duplications were identified in 6.5% of total mutations, all but one case predictably creating PTC. Mutations that consist in a deletion associated with sequence insertion (delins) were identified in seven cases (1.5%), whereas simple insertions were reported in six patients (1.3%). Only two cases (0.4%) involving large (exonic) duplications were reported to date. One patient was reported as having a complex rearrangement that involves the duplication of exon 10.[24] Patient P1 presented in detail in this work is the only recorded case with a multi-exonic duplication of *MTM1*. A total of 55 cases were not reported in the literature and were submitted directly to the MTM1-LOVD.

New MTM1 variants

The 25 novel point mutations identified in CNM male patients are described in Table 2. Among these, 11 are single-nucleotide substitutions located within exonic sequences: c.2T>A, c.32C>A, c.323G>A, c.469G>T, c.595C>A, c.637C>T, c.659G>C, c.1241T>C, c.1247A>G, c.1318C>T and c.1600T>C. With the exception of c.2T>A that predictably affects the initiation codon of
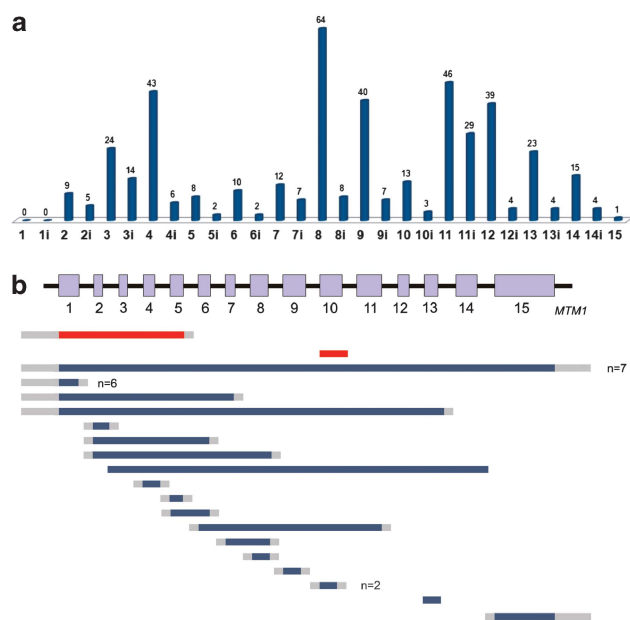
Figure 1 Mutations registered in MTM1-LOVD as distributed along the *MTM1* gene. (**a**) Total number of small mutations is shown for each exon and intron of *MTM1*. (**b**) Large deletions and duplications identified in *MTM1*. Solid bars represent regions involved (blue – deletions; red – duplications); gray bars represent undelineated breakpoints. In cases with more than one independent report the total number (*n*) of patients is indicated.

molecularly unresolved CNM patient cohort, six males and two females were selected for quantitative studies by MLPA analysis of *MTM1*, based on clinical and histological criteria that were compatible with CNM. Of these eight candidates, a single male patient (P1) was found to be positive, with a multi-exonic duplication in *MTM1*. A second male patient was subsequently found to carry a *DNM2* mutation and the remainder are as yet uncharacterized.

Muscle biopsy of P1 showed fiber size variation with atrophy, numerous central nuclei and endomysium fibrosis (Figure 2a). In some areas muscle was better preserved, with scanty central nuclei and atrophic fibers (Figure 2b). There was central dark staining with NADH-TR, SDH and PAS, reflecting aggregation of mitochondria and glycogen (Figures 2c and d). 'Necklace'-fibers were easily identified with routine histological stains (Figure 2e), PAS (Figure 2d) and histoenzimological stains. ATPase histoenzymology showed the presence of only type 1 fibers.

Following the detection of a large duplication spanning exons 1–5 of the *MTM1* gene (Figure 3a), the 3'-breakpoint of the duplication was narrowed down by Southern blot analysis (Figure 3b), which revealed the absence of a ~10 kb band corresponding to exons 5 and 6 of the gene. cDNA studies were carried out on reminiscent muscle tissue, in order to further characterize the rearrangement and evaluate its impact at the mRNA level. This analysis revealed a residual amount of normal transcript and a predominant mutant isoform lacking exon 6 (r.343_444del) (Figure 3c). It is predictable that this abnormal transcript will originate an in-frame deleted myotubularin lacking 34 amino acids of the GRAM domain.

In order to gain more insight into the structural rearrangement at the *MTM1* locus, we performed a whole-genome low coverage analysis by next generation sequencing. After alignment, the total number of reads aligned per 10 kb interval was calculated for the patient and an unrelated normal control sample. The ratio of aligned reads in the 10 kb bins between patient and control is indicative for amplification (ratio >1) or deletion (ratio <1) events. Multiple 10 kb bins in the 5'-region of the *MTM1* locus showed amplification (Figure 3d). Interestingly, the amplification appeared to extend into the 3'-region of the neighboring *MAMLD1* gene. An in depth analysis of the orientation of the aligned read pairs, as well as several long-range PCR experiments, failed to provide further information regarding the breakpoints of the structural rearrangement. Considering the NGS results, we postulated that the duplication could originate a *MTM1/MAMLD1* fusion transcript. In order to test this possibility, additional expression experiments were performed using RNA obtained from whole blood. Results revealed that a *MTM1/MAMLD1* fusion transcript was indeed generated (Figure 3e). Based on these results, a schematic representation of the duplication is presented (Figure 3f).

Large deletion in patient 2. During routine *MTM1* gDNA sequencing in P2, no symmetrical PCR amplification was obtained for the majority of exons (3–14). This led us to suspect the presence of a large intragenic deletion, which was subsequently confirmed by MLPA (Supplementary Figure S5a). This mutation was further characterized by long-range PCR, in order to predict its impact at the protein level and to facilitate carrier screening and prenatal diagnosis. Sequencing of the junction fragment obtained by long-range PCR revealed breakpoints located in introns 2 and 14 (~75 kb sized deletion) and enabled us to determine the full description of the mutation as c.63 + 834_1645-2104del (Supplementary Figure S5b). At the protein level, if efficient translation occurred, this would correspond to the

myotubularin (p.0?), the majority of these substitutions are predicted to be missense mutations (p.Gly108Asp, p.Pro199Thr, p.Leu213Phe, p.Arg220Thr, p.Phe414Ser, p.His416Arg and p.Trp534Arg). The remaining three are nonsense mutations (p.Ser11*, p.Glu157* and p.Gln440*). A condensed view of all data corroborating the pathogenicity of missense changes is presented in Supplementary Information (Supplementary Table S3). Briefly, missense variants were considered pathogenic when affecting phylogenetically conserved residues and/or were not detected in ethnically matched control chromosomes. Three single-nucleotide substitutions are located in intronic donor splice consensus sequences and predictably affect splicing: c.231 + 1G > T, c.342 + 5G > A and c.867 + 1G > A. In one of these changes (c.342 + 5G > A) cDNA studies demonstrated that it promotes exon 4 skipping (r.232_342del) leading to an in-frame deletion at the protein level (p.Ser79_Asp115del). In addition, four deletions (c.1088_1089del, c.1328_1331del, c.1509_1510del and c.1519_1522del), two duplications (c.509_528dup and c.596dup) and two deletion/insertion mutations (c.765_767delinsGG and c.1319_1321delinsTA) were also detected. All of these variants are predicted, at the RNA and protein level, to be frame-shift mutations that generate PTC. An insertion of 376 bp in exon 13 (c.1388_1389ins376, GenBank JQ403527) was detected in a patient with a severe phenotype. Using the Repbase database,[31] this insertion was identified as an *AluYa5* sequence. Further studies revealed that this alteration affects exon 13 splicing (Supplementary Figure S4), predictably resulting in an in-frame deletion of 38 amino acids (p.Phe452_Gln489del) located in the SID region of myotubularin. The remaining two novel mutations are the large deletion and duplication presented below in more detail.

Multi-exonic duplication in patient 1. Patients included in this work were initially studied by routine *MTM1* gene analysis. From our

Table 2 Novel *MTM1* mutations submitted to the MTM1-LOVD

Patient ID	DNA mutation	Gene location	Type of mutation	cDNA effect	Protein effect	Origin of mutation	Phenotype	Geographic origin
18914	c.=/(?_−76)_342 + ?dup	Exons 1–5	Large duplication	r.[= , 343_444del]	p.Asp115_Leu148del	*De novo*, somatic	Mild	Portugal
10959	c.63 + 834_1645-2104del	Intron 2–15	Large deletion	r.(?)	p.(Thr22_Gln548del)	Inherited (MC)	Severe	United Kingdom
24703	c.2T>A	Exon 2	Affects initiation codon	r.(?)	p.0?	Inherited (*de novo*, in mother)	Severe (at birth)	Germany
21000	c.32C>A	Exon 2	Nonsense	r.(?)	p.(Ser11*)	*de novo*	Severe	Germany
24706	c.231+1G>T	Intron 4	Donor splice site disruption[a]	r.(spl?)	p.(?)	Inherited (MGC)	Severe (died at 10 months)	Germany
20980	c.323G>A	Exon 5	Missense	r.(?)	p.(Gly108Asp)[b]	*de novo*	Mild (24 years old, still walking)	Germany
18861	c.342+5G>A	Intron 5	Donor splice site disruption	r.232_342del	p.Ser79_Asp115del	Inherited (MC)	Severe	United Kingdom
20960	c.469G>T	Exon 7	Nonsense	r.(?)	p.(Glu157*)	Inherited (MC)	Severe	Germany
20961	c.509_528dup	Exon 7	Out-of-frame duplication	r.(?)	p.(Gly177Trpfs*16)	Unknown	Severe (long-term survivor)	Portugal
20933	c.595C>A	Exon 8	Missense	r.(?)	p.(Pro199Thr)[b]	Inherited (MC)	Unknown (neonatal death)	Germany
24711	c.596dup	Exon 8	Out-of-frame duplication	r.(?)	p.(Ala200Cysfs*12)	Inherited (MC)	Severe	Germany/Albania
24702	c.637C>T	Exon 8	Missense	r.(?)	p.(Leu213Phe)[b]	Inherited (MC)	Mild (diagnosed at 24 years of age)	Turkey
24707	c.659G>C	Exon 8	Missense	r.(?)	p.(Arg220Thr)[b]	Unknown	Severe (at birth)	Portugal
20951	c.765_767delinsGG	Exon 9	Out-of-frame deletion	r.(?)	p.(Asp256Valfs*28)	Inherited (MC)	Severe	Germany
11124	c.867+1G>A	Intron 9	Donor splice site disruption[a]	r.(spl?)	p.(?)	Inherited (MC)	Unknown	Portugal
16734	c.1088_1089del	Exon 11	Out-of-frame deletion	r.(?)	p.(Lys363Serfs*14)	Inherited (MC)	Severe	Germany
11170	c.1241T>C	Exon 11	Missense	r.(?)	p.(Phe414Ser)[b]	Unknown	Unknown	Germany
12852	c.1247A>G	Exon 11	Missense	r.(?)	p.(His416Arg)[b]	Unknown	Severe	Germany
20944	c.1318C>T	Exon 12	Nonsense	r.(?)	p.(Gln440*)	Inherited (MC)	Severe	Germany
20955	c.1319_1321delinsTA	Exon 12	Out-of-frame deletion/insertion	r.(?)	p.(Gln440Leufs*24)	Inherited (MC)	Severe	Turkey
24712	c.1328_1331del	Exon 12	Out-of-frame deletion	r.(?)	p.(Asp443Valfs*20)	Unknown	Severe	United Kingdom
14208	c.1388_1389ins376, GenBank JQ403527	Exon 13	In-frame exon skipping	r.1354_1467del	p.Phe452_Gln489del	Inherited (*de novo*, in mother)	Severe	Germany
20938	c.1509_1510del	Exon 14	Out-of-frame deletion	r.(?)	p.(Asn503Lysfs*2)	Unknown	Severe	Germany
20952	c.1519_1522del	Exon 14	Out-of-frame deletion	r.(?)	p.(Glu507Asnfs*28)	Inherited (MC)	Severe (at birth)	
24705	c.1600T>C	Exon 14	Missense	r.(?)	p.(Trp534Arg)[b]	Inherited (MGC)		

Abbreviations: ID, patient identification in database; MC, mother carrier; MGC, maternal grandmother carrier.
Variants are described according to the reference sequence NM_000252.2, using HGVS nomenclature guidelines.
[a]Predicted to affect splicing by bioinformatic analysis.
[b]Pathogenicity assessment of missense variants in Supplementary Information (S3).
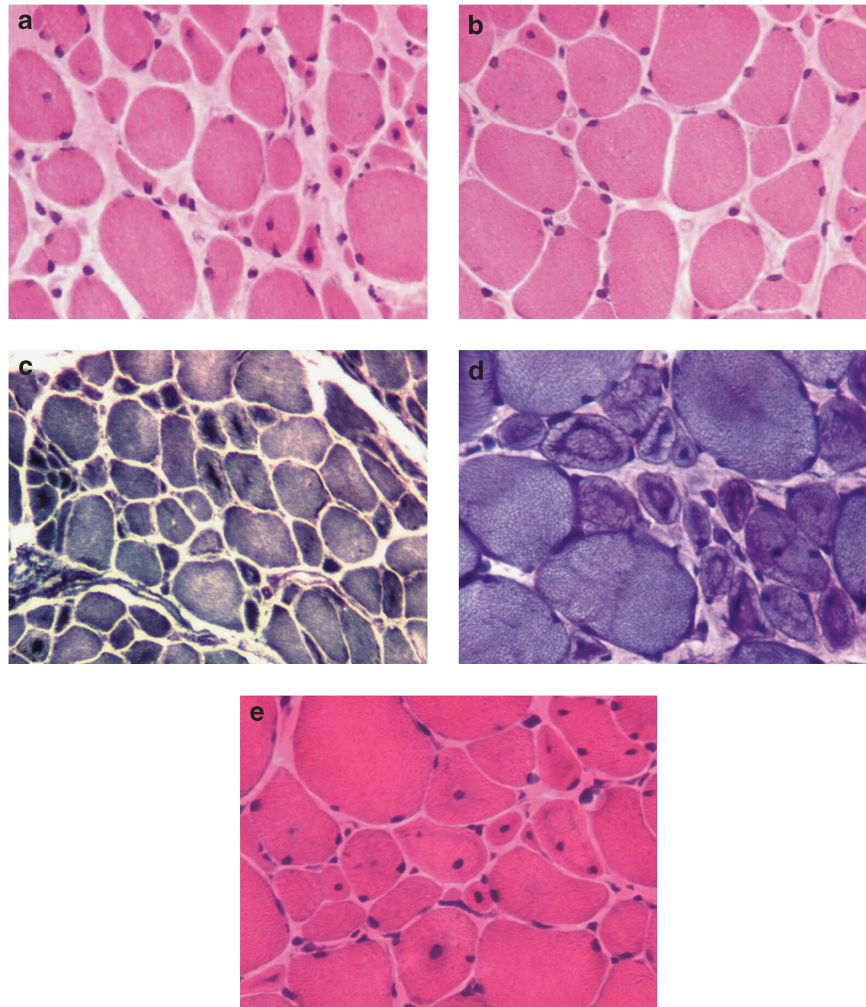
Figure 2 Histological features of patient 1. Histological analysis of the left deltoid muscle biopsy performed using staining with hematoxylin and eosin (**a**, **b** and **e**), histochemical reactions for NADH-TR (**c**), and PAS (**d**). Abnormal fiber size variability is seen on all stainings. The key feature of the pathology is the central nuclei (more evident on H&E image **a** and **e**). The dark central staining areas accompanying the central nuclei are another striking feature (**c** and **d**). The 'necklace fibers', recently described as a histopathological marker of *MTM1*, can be best appreciated on PAS staining (**d**) and faintly on H&E (**e**).

loss of ∼87% of primary protein sequence including all of the functionally relevant myotubularin domains.

DISCUSSION
MTM1-LOVD
Providing evidence for pathogenicity of variants is costly and time-consuming, so diagnostic laboratories rely on previous reports of variants detected in patients. For *MTM1* however, such data has been dispersed in the literature, in various formats. These difficulties prompted us to develop an LSDB for the *MTM1* gene. We also engaged in this task because the scientific community has expressed the importance and the need for establishing a dedicated XLMTM database (mutational and clinical).[23] The implementation and curation of this LSDB followed guidelines reported elsewhere,[32,33] and is registered with the HGVS (LSDB list, http://www.hgvs.org/dblist/glsdb.html).

The main objective for this freely accessible database is the description of *MTM1* sequence variants with phenotypic impact, collected from different sources. These data ultimately will allow further insights with respect to the mutational spectrum of the *MTM1*

gene and epidemiology, abetting the molecular diagnosis and research in this field. Currently >11% of the entries included in this database have not been previously described in the literature. This means that this database contributes to the visibility of new *MTM1* mutations within the public domain that would otherwise have a limited possibility to be released as single or individual case reports in the literature.

Database content analysis
Although large genomic deletions contributed to narrowing down the XLMTM candidate region and ultimately the identification of the *MTM1* gene,[34] large intragenic deletions appear to be a relatively rare cause of myotubular myopathy (6.3% of all cases) compared with point mutations. In fact, these deletions are generally identified in single case reports, with the exception of exon 1 and full *MTM1* gene deletions. The majority are associated with a severe disease outcome, even in deletions that are predictably in-frame. However, it should be noted that besides the large deletion reported in this work (P2) only a few mutations, namely the deletion of exon 9 and the deletion of
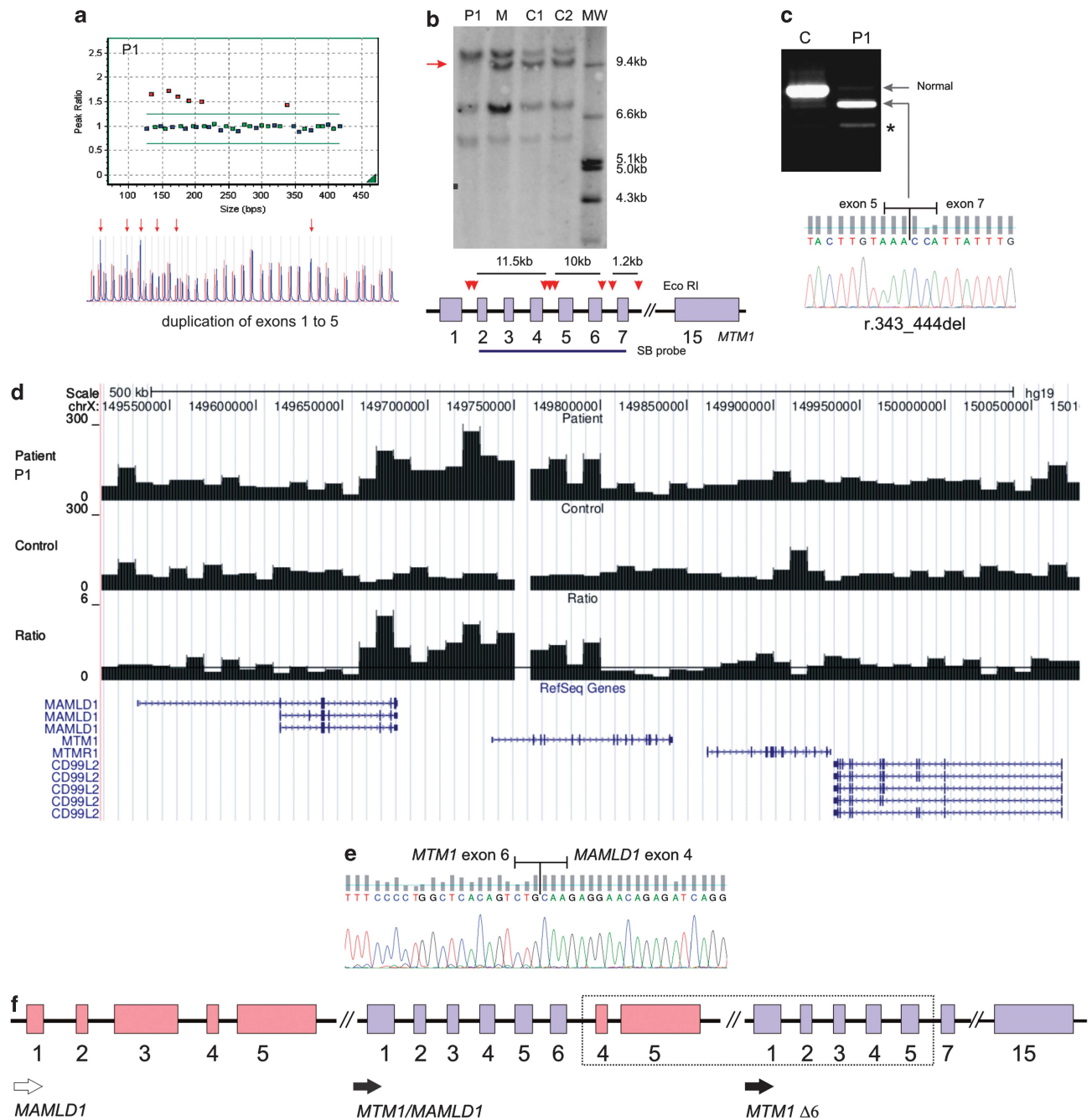
Figure 3 Large genomic duplication detected in patient 1. (**a**) MLPA analysis in patient 1 (P1) shows duplication of exons 1–5 of the *MTM1* gene. Red arrows indicate the duplicated probes. (**b**) Southern blot technique using a cDNA probe spanning exons 2–7 (SB probe); red arrow-heads represent *EcoR*I restriction sites; red arrow highlight the absence of a band in the patient, corresponding to exons 5 and 6. (**c**) RT-PCR analysis of mRNA obtained from muscle, revealing a mutated isoform lacking exon 6; (*) – smaller faint band corresponding to an alternative splicing product (Δ exon 5 and 6, ENSEMBL transcript ID: ENST00000424519). (**d**) Low coverage NGS analysis of a 500 kb interval within the X chromosome. Histograms correspond to the total number of reads aligned per 10 kb interval for the patient (top), an unrelated normal control sample (middle) and ratio between the patient and control (bottom) where the line corresponds to ratio value of 1. (**e**) Partial sequence of the fusion transcript, showing *MTM1* exon 6 adjacent to *MAMLD1* exon 4. (**f**) Schematic representation of duplicated region (dashed rectangle); resulting transcripts are indicated by black arrows – observed (filled) and predicted (unfilled). P1, patient; C/C1/C2, controls; M, patient's mother; MW, molecular weight marker.

exon 13, have been studied in enough detail to allow a more accurate prediction of their impact at the protein level.[10,35]

The largest proportion (93.3%) of pathogenic sequence variants described to date in *MTM1* is that comprising small mutations. Since the last published *MTM1* mutation update,[23] there has been a disproportional increase of missense and splicing mutations reported in *MTM1*. This might be attributed to the recent identification of additional XLMTM cases with a milder phenotype, often associated with these types of sequence variants.[36–39] Analysis of the distribution of point mutations reveals a higher incidence in exons

Table 3 Frequent mutations in MTM1-LOVD

Number of DB entries (n)	Gene region	Mutation (DNA)	Mutation (protein)	Sequence context	Possible cause
12	Exon 3	c.109C>T	p.Arg37*	TCCTCGACT	CpG
13	Exon 4	c.141_144delAGAA	p.Glu48Leufs*24	CAAAGAAGT	Slippage
10	Exon 4	c.205C>T	p.Arg69Cys	TTATCGTCT	CpG
9	Exon 8	c.614C>T	p.Pro205Leu	GTTCCGTAT	CpG
13	Exon 9	c.721C>T	p.Arg241Cys	TGTGCGTTG	CpG
28	Intron 11	c.1261-10A>G	p.Ser420_Arg421insPheIleGln	ATCAATTTA	Unknown
9	Exon 12	c.1261C>T	p.Arg421*	TCAGCGAAT	CpG
18	Exon 12	c.1262G>A	p.Arg421Gln	CAGCGAATA	CpG

Abbreviations: DB, database; CpG, dinucleotide mutational hotspot.
Description of mutations with nine or more independent entries. Nucleotides affected by the mutation are underlined. Variants are described according to the reference sequence NM_000252.2, using HGVS nomenclature guidelines.

Table 4 Analysis of MTM1-LOVD missense mutations

			Missense mutations				Nonsense mutations
	Residues involved			Phenotype reported[a]			
Protein domain	(% of total protein)	n (% of total)	Severe	Mild/Moderate	Inconsistent	Unknown	n (% of total)
GRAM	aa 34–149 (19.1%)	9 (12.3%)	2 (22.2%)	5 (55.6%)	0	2 (22.2%)	5 (15.6%)
RID[b]	aa 162–265 (17.1%)	28 (38.4%)	15 (53.6%)	7 (25.0%)	3 (10.7%)	3 (10.7%)	6 (18.8%)
PTP	aa 274–434 (26.5%)	28 (38.4%)	18 (64.3%)	5 (17.9%)	1 (3.5%)	4 (14.3%)	3 (9.4%)
SID	aa 435–486 (8.5%)	5 (6.8%)	3 (60.0%)	1 (20.0%)	0	1 (20.0%)	5 (15.6%)
Other	(28.8%)	3 (4.1%)	1 (33.3%)	2 (66.7%)	0	0	13 (40.6%)
	Total	73	39	21	4	9	32

Abbreviations: aa, amino acid position within myotubularin protein (reference sequence NP_000243.1); GRAM, glucosyltransferase, Rab-like GTPase activators and myotubularins; PTP, protein tyrosine phosphatase; RID, Rac-induced recruitment domain; SID, SET-interacting domain.
[a]Only phenotypes reported in male patients were considered for this analysis; the mild and moderate phenotypes were combined as several reports do not distinguish the two clinical classifications.
[b]This domain partially overlaps desmin-binding region.

8, 11, 9, 4 and 12 (in decreasing order of frequency), and this should be taken into consideration when conducting *MTM1* gene mutation screening. Altogether, mutations in these exons account for almost half of all reported XLMTM cases.

A significant number of *MTM1* point mutations coincide with the hypermutable CpG dinucleotides that are amenable to methylation-mediated deamination. This mechanism could explain the recurrence, hence higher frequency, of some variants (Table 3). In fact six mutations (c.109C>T, c.141_144delAGAA, c.205C>T, c.1261-10A>G, c.1261C>T and c.1262G>A) were each detected in nine or more patients, accounting for ∼24% of total cases in the MTM1-LOVD database. The most frequent variant in the MTM1-LOVD is the splicing mutation c.1261-10A>G that activates a cryptic splice site, promoting the inclusion of 9 bp in the open reading frame. This cryptic splice site is preferentially used in *MTM1* transcripts detected in skeletal muscle.[40,41] The change is predicted to include three amino acids in the core of the PTP (active) site. It was identified in several patients from different ethnic origins (US, Japan and Europe) and associated with a severe disease outcome in male patients. The cause of this recurrent mutational event is still unknown.

In a previous study,[12] the analysis of a large number of patients revealed a statistically significant association between non-truncating mutations and the mild phenotype, as opposed to the intermediate/severe phenotype associated with mutations that give rise to PTC. We performed a similar analysis of all data available in MTM1-LOVD with the aim of obtaining further correlation between genotype and phenotype. In line with previous reports, it was evident that

truncating mutations are predominantly associated with a severe phenotype. A total of 188 male patients were reported to have mutations that originate PTC. Clinical classification was available in 146 of these (77.7%). The majority (*n* = 139, 95.2%) were reported as severe. Only four patients were reported as mild or mild/moderate, and the PTC inducing mutations were c.836delC, c.1558C>T (*n* = 2) and c.1792delC. Besides these, there are three patients with a moderate phenotype, one associated with a splicing mutation (c.137-7T>G) and the other two with nonsense mutations which had also been found in patients with a severe disease outcome. Overall, the registries show a clear bias towards the severest end of the disease, which may reflect either a specific biological pattern related to defects in myotubularin or the patient selection criteria used to conduct *MTM1* gene analysis.

As the majority of non-truncating mutations are of the missense type (∼80%) we centered our analysis on these variants and their localization on the protein (Table 4). Nonsense mutations were also included for comparative purposes. Most missense mutations (96%) are located in myotubularin regions with known function (representing 71% of the total protein sequence), whereas only 59% of nonsense mutations are located in these domains. This suggests that missense changes are not randomly scattered in the protein. Moreover, 85% of mutations associated with a severe phenotype are located in the PTP (catalytic) domain and the RID/desmin-binding region. In terms of its relative proportion, mild/moderate missense mutations are more frequent in the GRAM domain and in the remaining regions of myotubularin. However, it should be noted that there is some degree

548

of phenotypic variability; four missense mutations were classified as 'inconsistent' in terms of the reported clinical severity. The presence of more than one sequence variant in a patient can hamper attempts at genotype/phenotype correlations. In fact two independent publications reported male patients with double *MTM1* mutations. In the first case a patient classified as having a severe phenotype (died at 6 weeks of age) two missense changes were identified (p.Asp431Asn and p.Asp433Asn).[42] The second patient was reported with severe neonatal hypotonia requiring assisted ventilation due to respiratory failure. Mutational analysis revealed the presence of two *MTM1* changes that create PTC: c.109C>T (p.Arg37*) and c.386_387insA (p.Ser129Argfs*7).[43]

Novel multi-exonic duplication. We describe two patients with large rearrangements in the *MTM1* gene, one of which (P1) carries the first multi-exonic duplication described to date. The clinical presentation of this patient is milder than the classical form of XLMTM – although he had delayed motor development skills and a progressive tetraparesis, at age 7 he remains ambulant and there is no report of respiratory impairment. Analysis of the patient's muscle biopsy revealed variation of fiber size and typical central nuclei, which lead to the diagnosis of CNM. Additionally, 'necklace' fibers were identified. These structures, characterized by a basophilic ring deposit following the contour of the cell in which the nuclei are aligned, were initially described as a particular feature of older CNM cases.[1] It is now evident that 'necklace' fibers can also be found in early onset cases with a milder phenotype. Using a variety of experimental methods, we have shown that the duplication encompassing exons 1–5 in *MTM1* has an unexpected effect at the mRNA level, resulting in an in-frame deletion of the sequence corresponding to exon 6 (r.343_444del). At the gDNA level, using a genome wide low coverage analysis of the patient's genome we were able to confirm the amplification within the *MTM1* gene and obtained evidence that it extended into the *MAMLD1* gene, known to be associated with hypospadias.[44] However, unlike intragenic duplications, it is foreseeable that because this duplication affects only the 3′-end of the *MAMLD1* gene, a normal copy would be left intact, hence the duplication should have little or no phenotypic consequence, thereby explaining why no genital abnormalities were found in the patient.

MLPA analysis had shown that the average ratio of duplicated *MTM1* probes (1.6) was significantly lower than expected for a male (~2.0). This, together with the presence of a residual amount of normal transcript and the finding that the patient's mother was not a carrier, provides strong evidence for somatic mosaicism in the patient – the first such case described to date – which might explain the patient's relatively mild phenotype in relation to what could be predicted from the genotype.

Standard routine methods for XLMTM molecular diagnosis, such as genomic analysis of *MTM1* by Sanger sequencing, do not allow the identification of duplications and may also fail to detect mosaics. The only other description of a duplication involving a complex *MTM1* rearrangement concerns a male infant with characteristic clinical and histopathologic findings of XLMTM.[24] The patient had severe neonatal hypotonia and respiratory insufficiency. Other XLMTM compatible findings included the absence of deep tendon reflexes, cryptorchidism and elongated fingers and toes. There was a progressive respiratory deterioration culminating in death at 1 month of age, due to respiratory failure. It is conceivable that there may be a considerable number of CNM cases with *MTM1* duplications, and it is therefore advisable to perform MLPA analysis in all CNM cases, after excluding point mutations.

Future perspectives

It would be important to gather more *MTM1* mutational data that are presently dispersed among other non-public databases (such as UMD-MTM1 and the Cardiff database for XLMTM),[45] as well as in private registries of diagnostic laboratories around the world. According to the meeting report of the 6th ENMC workshop on centronuclear (myotubular) myopathies, efforts are currently being made to develop a patient registry for CNM with the support of TREAT-NMD and the Myotubular Trust.[45] Until now the harmonized clinical items to be included in this registry are not yet in the public domain. After full release of LOVD software v3.0 the clinical items to be included will follow these international recommendations for the *MTM1* registry, ultimately allowing the data integration, which at this point in time is a recognized difficulty.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Romero NB: Centronuclear myopathies: a widening concept. *Neuromuscul Disord* 2010; **20**: 223–228.
2 Nance JR, Dowling JJ, Gibbs EM, Bönnemann CG: Congenital myopathies: an update. *Curr Neurol Neurosci Rep* 2012; **12**: 165–174.
3 Laporte J, Hu LJ, Kretz C et al: A gene mutated in X-linked myotubular myopathy defines a new putative tyrosine phosphatase family conserved in yeast. *Nat Genet* 1996; **13**: 175–182.
4 Biancalana V, Beggs AH, Das S et al: Clinical utility gene card for: Centronuclear and myotubular myopathies. *Eur J Hum Genet* 2012; doi:10.1038/ejhg.2012.91.
5 Bitoun M, Maugenre S, Jeannet PY et al: Mutations in dynamin 2 cause dominant centronuclear myopathy. *Nat Genet* 2005; **37**: 1207–1209.
6 Tosch V, Rohde HM, Tronchère H et al: A novel PtdIns3P and PtdIns(3,5)P2 phosphatase with an inactivating variant in centronuclear myopathy. *Hum Mol Genet* 2006; **15**: 3098–3106.
7 Nicot AS, Toussaint A, Tosch V et al: Mutations in amphiphysin 2 (BIN1) disrupt interaction with dynamin 2 and cause autosomal recessive centronuclear myopathy. *Nat Genet* 2007; **39**: 1134–1139.
8 Wilmshurst JM, Lillis S, Zhou H et al: RYR1 mutations are a common cause of congenital myopathies with central nuclei. *Ann Neurol* 2010; **68**: 717–726.
9 Bevilacqua JA, Monnier N, Bitoun M et al: Recessive RYR1 mutations cause unusual congenital myopathy with prominent nuclear internalization and large areas of myofibrillar disorganization. *Neuropathol Appl Neurobiol* 2011; **37**: 271–284.
10 Jungbluth H, Wallgren-Pettersson C, Laporte J: Centronuclear (myotubular) myopathy. *Orphanet J Rare Dis* 2008; **3**: 26.
11 Herman GE, Finegold M, Zhao W, de Gouyon B, Metzenberg A: Medical complications in long-term survivors with X-linked myotubular myopathy. *J Pediatr* 1999; **134**: 206–214.
12 McEntagart M, Parsons G, Buj-Bello A et al: Genotype-phenotype correlations in X-linked myotubular myopathy. *Neuromuscul Disord* 2002; **12**: 939–946.
13 Tanner SM, Orstavik KH, Kristiansen M et al: Skewed X-inactivation in a manifesting carrier of X-linked myotubular myopathy and in her non-manifesting carrier mother. *Hum Genet* 1999; **104**: 249–253.
14 Hammans SR, Robinson DO, Moutou C et al: A clinical and genetic study of a manifesting heterozygote with X-linked myotubular myopathy. *Neuromuscul Disord* 2000; **10**: 133–137.
15 Sutton IJ, Winer JB, Norman AN, Liechti-Gallati S, MacDonald F: Limb girdle and facial weakness in female carriers of X-linked myotubular myopathy mutations. *Neurology* 2001; **57**: 900–902.
16 Jungbluth H, Sewry CA, Buj-Bello A et al: Early and severe presentation of X-linked myotubular myopathy in a girl with skewed X-inactivation. *Neuromuscul Disord* 2003; **13**: 55–59.
17 Schara U, Kress W, Tücke J, Mortier W: X-linked myotubular myopathy in a female infant caused by a new MTM1 gene mutation. *Neurology* 2003; **60**: 1363–1365.
18 Pénisson-Besnier I, Biancalana V, Reynier P, Cossée M, Dubas F: Diagnosis of myotubular myopathy in the oldest known manifesting female carrier: a clinical and genetic study. *Neuromuscul Disord* 2007; **17**: 180–185.
19 Robinson FL, Dixon JE: Myotubularin phosphatases: policing 3-phosphoinositides. *Trends Cell Biol* 2006; **16**: 403–412.
20 Laporte J, Blondeau F, Buj-Bello A, Mandel JL: The myotubularin family: from genetic disease to phosphoinositide metabolism. *Trends Genet* 2001; **17**: 221–228.

21 Tsujita K, Itoh T, Ijuin T *et al*: Myotubularin regulates the function of the late endosome through the gram domain-phosphatidylinositol 3,5-bisphosphate interaction. *J Biol Chem* 2004; **279**: 13817–13824.

22 Hnia K, Tronchère H, Tomczak KK *et al*: Myotubularin controls desmin intermediate filament architecture and mitochondrial dynamics in human and mouse skeletal muscle. *J Clin Invest* 2011; **121**: 70–85.

23 Laporte J, Biancalana V, Tanner SM *et al*: Mtm1 Mutations in X-linked myotubular myopathy. *Hum Mutat* 2000; **15**: 393–409.

24 Trump N, Cullup T, Muntoni F, Verheij J, Jungbluth H: X-linked myotubular myopathy due to a complex rearrangement involving exon 10 of the myotubularin (MTM1) gene. *Neuromuscul Disord* 2012; **22**: 384–388.

25 Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT: LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat* 2011; **32**: 557–563.

26 den Dunnen JT, Antonarakis SE: Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat* 2000; **15**: 7–12.

27 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.

28 Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C: Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**: e67.

29 Quinlan AR, Hall IM: BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; **26**: 841–842.

30 Kent WJ, Sugnet CW, Furey TS *et al*: The human genome browser at UCSC. *Genome Res* 2002; **12**: 996–1006.

31 Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 2005; **110**: 462–467.

32 Celli J, Dalgleish R, Vihinen M, Taschner PE, den Dunnen JT: Curating gene variant databases (LSDBs): Toward a universal standard. *Hum Mutat* 2011; **33**: 291–297.

33 Vihinen M, den Dunnen JT, Dalgleish R, Cotton RG: Guidelines for establishing locus specific databases. *Hum Mutat* 2011; **33**: 298–305.

34 Dahl N, Hu LJ, Chery M *et al*: Myotubular myopathy in a girl with a deletion at Xq27-q28 and unbalanced X inactivation assigns the MTM1 gene to a 600-kb region. *Am J Hum Genet* 1995; **56**: 1108–1115.

35 Tsai TC, Horinouchi H, Noguchi S *et al*: Characterization of MTM1 mutations in 31 Japanese families with myotubular myopathy, including a patient carrying 240 kb deletion in Xq28 without male hypogenitalism. *Neuromuscul Disord* 2005; **15**: 245–252.

36 Biancalana V, Caron O, Gallati S *et al*: Characterisation of mutations in 77 patients with X-linked myotubular myopathy, including a family with a very mild phenotype. *Hum Genet* 2003; **112**: 135–142.

37 Yu S, Manson J, White S *et al*: X-linked myotubular myopathy in a family with three adult survivors. *Clin Genet* 2003; **64**: 148–152.

38 Hoffjan S, Thiels C, Vorgerd M, Neuen-Jacob E, Epplen JT, Kress W: Extreme phenotypic variability in a German family with X-linked myotubular myopathy associated with E404K mutation in MTM1. *Neuromuscul Disord* 2006; **16**: 749–753.

39 Bevilacqua JA, Bitoun M, Biancalana V *et al*: 'Necklace' fibers, a new histological marker of late-onset MTM1-related centronuclear myopathy. *Acta Neuropathol* 2009; **117**: 283–291.

40 de Gouyon BM, Zhao W, Laporte J, Mandel JL, Metzenberg A, Herman GE: Characterization of mutations in the myotubularin gene in twenty six patients with X-linked myotubular myopathy. *Hum Mol Genet* 1997; **6**: 1499–1504.

41 Nishino I, Minami N, Kobayashi O *et al*: MTM1 gene mutations in Japanese patients with the severe infantile form of myotubular myopathy. *Neuromuscul Disord* 1998; **8**: 453–458.

42 Laporte J, Guiraud-Chaumeil C, Vincent MC *et al*: Mutations in the MTM1 gene implicated in X-linked myotubular myopathy. ENMC International Consortium on Myotubular Myopathy. European Neuro-Muscular Center. *Hum Mol Genet* 1997; **6**: 1505–1511.

43 Tachi N, Kozuka N, Chiba S, Miyaji M, Watanabe I: A double mutation in a patient with X-linked myotubular myopathy. *Pediatr Neurol* 2001; **24**: 297–299.

44 Ogata T, Wada Y, Fukami M: MAMLD1 (CXorf6): a new gene for hypospadias. *Sex Dev* 2008; **2**: 244–250.

45 Jungbluth H, Wallgren-Pettersson C, Laporte JF: Centronuclear (Myotubular) Myopathy Consortium. 164th ENMC International workshop: 6th workshop on centronuclear (myotubular) myopathies, 16-18th January 2009, Naarden, The Netherlands. *Neuromuscul Disord* 2009; **19**: 721–729.

Research Report

RYR1-Related Myopathies: Clinical, Histopathologic and Genetic Heterogeneity Among 17 Patients from a Portuguese Tertiary Centre

Raquel Samões[a,1,*], Jorge Oliveira[b,c,1], Ricardo Taipa[d], Teresa Coelho[e], Márcio Cardoso[e], Ana Gonçalves[b,c], Rosário Santos[b,c,f], Manuel Melo Pires[d] and Manuela Santos[g]

[a]*Department of Neurology, Centro Hospitalar do Porto, Porto, Portugal*
[b]*Unidade de Genética Molecular, Centro de Genética Médica, Centro Hospitalar do Porto, Porto, Portugal*
[c]*Unidade Multidisciplinar de Investigação Biomédica (UMIB), Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal*
[d]*Neuropathology Unit, Centro Hospitalar do Porto, Porto, Portugal*
[e]*Department of Neurophysiology and Neuromuscular Disorders Outpatient Clinic, Centro Hospitalar do Porto, Porto, Portugal*
[f]*UCIBIO/REQUIMTE, Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal*
[g]*Neuromuscular Disorders Outpatient Clinic and Department of Neuropaediatrics, Centro Hospitalar do Porto, Porto, Portugal*

Abstract.
Background: Pathogenic variants in ryanodine receptor type 1 (*RYR1*) gene are an important cause of congenital myopathy. The clinical, histopathologic and genetic spectrum is wide.
Objective: Review a group of the patients diagnosed with ryanodinopathy in a tertiary centre from North Portugal, as an attempt to define some phenotypical patterns that may help guiding future diagnosis.
Methods: Patients were identified from the database of the reference centre for Neuromuscular Disorders in North Portugal. Their data (clinical, histological and genetic) was retrospectively accessed.
Results: Seventeen *RYR1-related* patients (including 4 familial cases) were identified. They were divided in groups according to three distinctive clinical characteristics: extraocular muscle (EOM) weakness ($N = 6$), disproportionate axial muscle weakness ($N = 2$) and joint laxity ($N = 5$). The fourth phenotype includes patients with mild tetraparesis and no distinctive clinical features ($N = 4$). Four different histopathological patterns were found: centronuclear ($N = 5$), central core ($N = 4$), type 1 fibres predominance ($N = 4$) and congenital fibre type disproportion ($N = 1$) myopathies. Each index case, except two patients, had a different *RYR1* variant. Four new genetic variants were identified. All centronuclear myopathies were associated with autosomal recessive inheritance and EOM weakness. All central core myopathies were caused by pathogenic variants in hotspot 3 with autosomal dominant inheritance. Three genetic variants were reported to be associated to malignant hyperthermia susceptibility.

---

[1]These authors contributed equally to this work.

*Correspondence to: Raquel Samões, Centro Hospitalar do Porto, Serviço de Neurologia, Largo do Prof. Abel Salazar, 4099-001 Porto, Portugal. Tel.: +351967429395; E-mail: araquelfsamoes@hotmail.com.

Conclusions: Distinctive clinical features were recognized as diagnostically relevant: extraocular muscle weakness (and centronuclear pattern on muscle biopsy), severe axial weakness disproportionate to the ambulatory state and mild tetraparesis associated with (proximal) joint laxity. There was a striking genetic heterogeneity, including four new RYR1 variants.

Keywords: Central core, centronuclear, congenital myopathies, joint laxity, malignant hyperthermia, myopathy, ryanodine receptor 1, RYR1

INTRODUCTION

Pathogenic variants in the ryanodine receptor type 1 (*RYR1*) gene have been identified as the most common cause of congenital myopathies, in several large cohorts [1, 2]. This gene codifies the ryanodine receptor, a calcium release channel of the skeletal muscle sarcoplasmic reticulum with a crucial role in excitation–contraction coupling. *RYR1* is a large gene (106 exons) but it has three mutational hotspot regions: region 1 (exons 2–17, amino acid residues 35–614), region 2 (exons 39–46, amino acid residues 2,163–2,458) and region 3 (exons 85–104, amino acid residues 3,916–4,942) [3].

The best characterized entity associated to *RYR1* is the dominantly inherited central core disease with malignant hyperthermia susceptibility, a potentially life-threatening pharmacogenetic reaction to volatile anesthetics and/or depolarising muscle relaxants [4]. However, in recent years, an increasing range of additional histopathological and clinical variants have been associated with *RYR1* variants, including with recessive inheritance [5, 6].

We aimed to review the clinical, histopathologic and genetic characteristics of the patients diagnosed with ryanodinopathy followed in our tertiary centre. In face of the great variability described in literature, we attempted to define some key features of ryanodinopathies that could help future diagnosis.

MATERIALS AND METHODS

Centro Hospitalar do Porto – Hospital de Santo António is the reference centre for Neuromuscular Disorders in North Portugal.

Adult and paediatric patients with the genetic diagnosis of ryanodinopathy were identified from the Centre database. Their clinical records and muscle biopsies were retrospectively reviewed. As muscle magnetic resonance imaging is still not available in our centre in the routine clinical practice, we do not have such data.

This retrospective series of cases is in accord with the ethical standards of the Committee on Human Experimentation of our institution and in accord with the Helsinki Declaration of 1975.

RESULTS

Seventeen patients with ryanodine myopathy, including four families (FAM), were identified and included. Table 1 summarises the clinical, histopathological and genetic data of the patients (P1 to P17).

Clinical manifestations

At the time when data were retrospectively collected, the patients' mean age was 22.3 years old (range 5 months-51 yo). Twelve patients (70.6%) were female.

The disease onset was reported by most patients or their parents as having occurred congenitally or during the first year of life; only two patients noticed the first symptoms between 2yo and school age.

All patients had a myopathic syndrome with thin appearance of muscle bulk, along with other specific features:

- Extraocular muscle (EOM) weakness (Fig. 1, Panel A): This was present in six patients (including family FAM.I) and associated with different severities of tetraparesis. Patients 1 and 2 (two siblings aged 28 and 34yo) constitute our most disabled patients; they have never gained the ability to walk; they also have facial, bulbar, respiratory and spinal muscles weakness and were submitted to surgery in the second decade of life due to a rapid evolving scoliosis. Patient 3 (aged 24yo) has an intermediate presentation in terms of severity, as she has a slight tetraparesis but with a significant weakness of the axial muscles. She also had early scoliosis surgery and needs non-invasive ventilation. Patient 4 (aged 3yo) presented with a mild motor delay as she only acquired the ability to walk at the age of 24 months. Patients 5 (aged 37yo) and 6 (aged 33yo) have proximal tetraparesis with no other associated features apart

Table 1
Clinical, histopathological and genetic findings of patients included in this work

Patients	Distinctive clinical features	Additional clinical features					Main histopathological findings	RYR1 DNA/ protein variants (all heterozygous)	Variant type	Exon	Protein domain / hotspot	Variant Reference	Disease inheritance
		AW	F	B	SS	NIV							
P1, P2 (FAM.I)		N	Y	Y	15, 13	*, 25	Central nuclei	c.12010C>T / p.Gln4004*	nonsense	87	HD3	[24]	AR
P3		3.5	Y	N	15	17		c.14643G>A / p.Met4881Ile	missense	101	HD3, Triadin	[24]	Sporadic (AR, compound heterozygosity)
								c.3800C>G / p.Pro1267Arg	missense	28	–	[24]	
P4	Extraocular muscles weakness	1	Y	N	N	N		c.9157C>T / p.Arg3053*	nonsense	61	apoCaM	[24]	Sporadic (AR?, parents not tested)
								c.12063_12064dup / p.Met4022Thrfs*4	frameshift dup.	88	HD3	New [b]	
P5		1	Y	N	N	N		[c.7027G>A / p.Gly2343Ser (spl.?);	missense / spl.?	43	HD2	[25]	Sporadic (AR, compound heterozygosity)
								c.13672C>T / p.Arg4558Trp]	missense	94	HD3, S100A1	New	
								c.1901delC / p.Thr634Lysfs*32	frameshift del.	17	–	[25]	
P6		3	Y	N	N	N	Congenital fibre type disproportion	[c.7027G>A / p.Gly2343Ser (spl.?);	missense / spl.?	43	HD2		Sporadic (AR, compound heterozygosity)
								c.13672C>T / p.Arg4558Trp]	missense	94	HD3, S100A1		
								c.1342A>T / p.Ile448Phe	missense	13	HD1	New [c]	
P7 (FAM.II)	Severe axial muscles weakness	1	Y	N	16	N	Type 1 fibre predominance	c.2348C>T / p.Ser783Leu	missense	19	SPRY	New	AD (niece of P14)
								c.14677C>T / p.Arg4893Trp	missense	102	HD3	[26]	
P8		2	Y	N	14	14		c.7843C>T / p.Arg2615Cys [d]	missense	49	DHPR	[11]	Sporadic (variant inherited from healthy father)

(Continued)

Table 1
(Continued)

Patients	Distinctive clinical features	Additional clinical features					Main histopathological findings	RYR1 DNA/ protein variants (all heterozygous)	Variant type	Exon	Protein domain / hotspot	Variant Reference	Disease inheritance
		AW	F	B	SS	NIV							
P9, P10, P11 (FAM.III)	Joint laxity	2,1,N	[a]	[a]	N	N	Type 1 fibre predominance (P9)	c.14693T>C / p.Ile4898Thr[MHS]	missense	102	HD3	[27]	AD
P12, P13 (FAM.IV)		1	N	N	N	N	Central core (P12)	c.14656T>C / p.Phe4886Leu	missense	102	HD3	New [6] [e]	AD
P14 (FAM.II)	Mild tetraparesis only ± facial weakness	1	Y	N	N	N	Central core	c.14677C>T / p.Arg4893Trp	missense	102	HD3	[11]	AD (uncle of P7)
P15	Type 1 fibre predominance	1	Y	N	N	N	Type 1 fibre predominance	[c.4711A>G / p.Ile1571Val;	missense	33	–	[28],	Sporadic
								c.10097G>A / p.Arg3366His;		67	–	[29]	(AR?, parents not tested)
								c.11798G>A/ p.Tyr3933Cys][MHS]		86	HD3		
								c.14537C>T / p.Ala4846Val	missense	101	HD3	[30]	
P16		1	N	N	N	N	Central core	c.14582G>A / p.Arg4861His[MHS]	missense	101	HD3	[26]	AD (other family members affected)
P17		1	N	N	N	N		c.13910C>T / p.Thr4637Ile	missense	95	HD3	[31]	Sporadic (de novo)

Additional clinical features: AW – ability to walk (age in years); F – Facial weakness; B – Bulbar weakness; SS – Scoliosis surgery (age in years); NIV – Non-invasive ventilation (age in years); N – no, Y – yes, * - Patient refuses; FAM. – Family; HD1/2/3- hotspot domain 1/2/3; Triadin- Triadin interaction domain; apoCaM - apocalmodulin binding site; S100A1 - S100A1 domain; SPRY - Domain in SP1a and the RYanodine Receptor; DHPR - dihydropyridine receptor association domain; [a] – Only P11 had bulbar involvement and facial palsy in the neonatal period; [b] – variant listed in the Leiden Muscular Dystrophy LOVD database; [c] – variant also found in another unpublished patient from our cohort with central core disease; [d] – variant of uncertain clinical significance; [e] - Aminoacid change previously reported; AR – autosomal recessive; AD – autosomal dominant; Del. – deletion; Dup. – duplication; MHS - Documented risk for malignant hyperthermia susceptibility; P1, P2 and P3 were previously reported in Oliveira et al., 2016 [24] and P7, P8 and P15 previously presented in Rocha et al., 2014 [11].
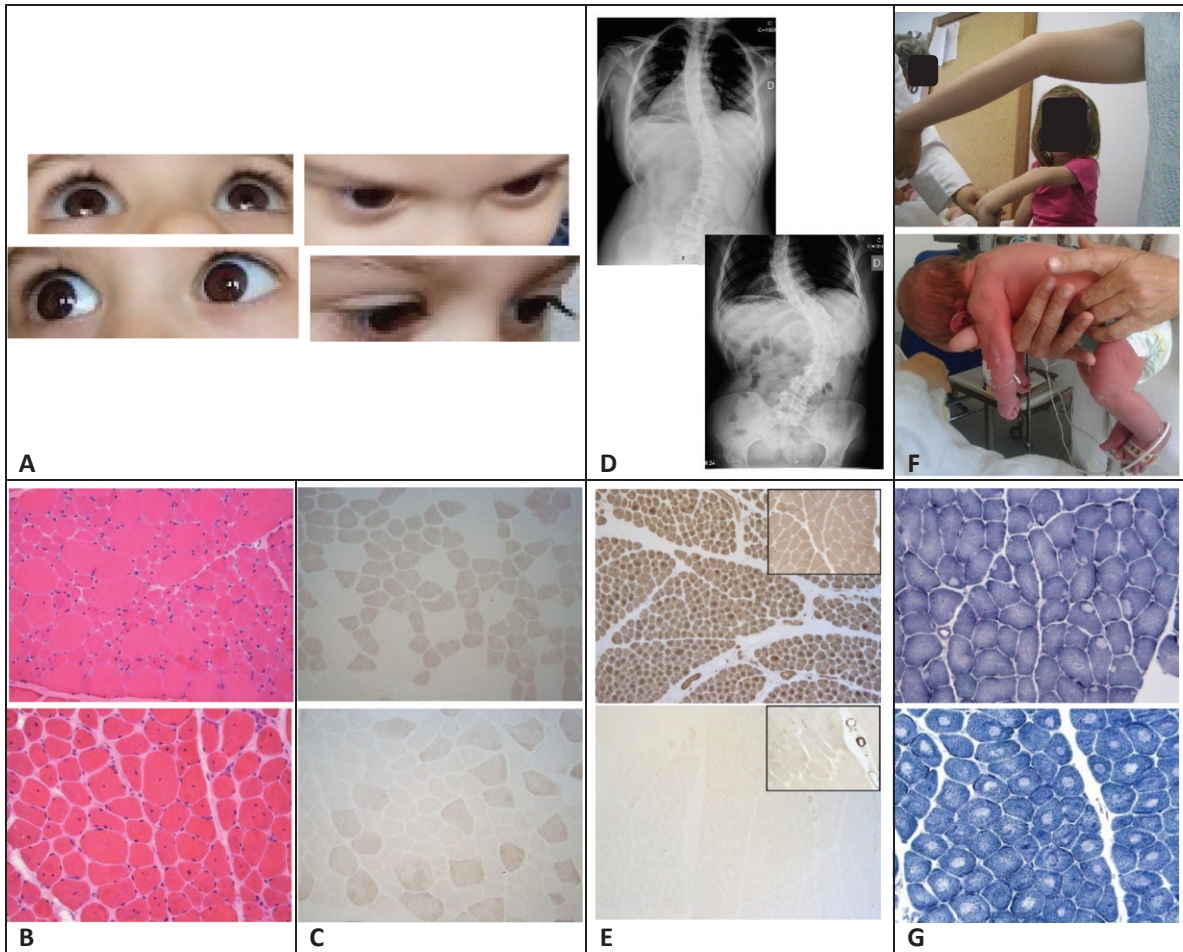
Fig. 1. Clinical and histopathological features of some patients (Color figures in the online PDF version). Panel A (permission was granted by the patient) – Extraocular muscle weakness in patient 4: upper left – looking upwards, upper right – looking downwards, lower left – looking rightwards, lower right – looking leftwards;. Panel B – Centronuclear myopathy, H&E patients 4 (top) and 5 (bottom), magnification, 200x. Panel C – Myopathy with congenital fibre type disproportion as main feature, patient 6; ATPase 4.35 (top) and 9.4 (bottom), magnification 100x. Panel D – Spinal X-ray showing worsening of scoliosis in patient 7, five years apart: at 11yo (top) and 16yo (bottom);. Panel E – Myopathy with type 1 fibres predominance in patient 9, ATPase 4.35 (top) and 9.4 (bottom), magnification, 100x. Panel F – Family 3: patients 9 and 10 showing hyperlaxity at elbows and wrists (top) and patient 11 as a floppy newborn with nasogastric tube (bottom). Panel G – Centralcore myopathy in patients12 (top) and 16 (bottom), NADH, magnification 100x.

from facial muscle weakness, also shared by the other patients.

- Severe axial muscle weakness (Fig. 1, panel D): Patients 7 (FAM.II) and 8 (unrelated, aged 17 and 20yo) had rapidly progressive scoliosis that led to surgery at the age of 16 and 14yo, respectively. The second one has also been using non-invasive ventilation since that time. They also have mild facial muscle weakness. Both patients maintain unassisted walking and none has bulbar muscle involvement.
- Joint laxity (Fig. 1, panel F): In two families, FAM.III and FAM.IV, the coexistence of

weakness and laxity was the main key to the diagnosis. In the first kindred, the index case was patient 9, a 4yo girl with normal development and a mild to moderate proximal tetraparesis, mainly affecting the lower limbs, associated with an evident laxity in axial and limbs joints (shoulders and elbows) noticed at the age of two. Patient 10, her mother (aged 20yo), has recurrent patellar dislocations and the same pattern of laxity but with less weakness. This worsened transiently during her second pregnancy. She gave birth to our youngest patient, patient 11, who was born with tetraparesis (mainly

upper limbs), laxity, facial and transitory bulbar involvement, needing to be fed by a nasogastric tube in the neonatal period. She progressively improved in months and at present, aged 12 months, she achieved to sit unsupported. In FAM.IV, the daughter (P12, aged 9yo) was also the index case, with moderate tetraparesis, elbow laxity and past history of congenital hip dislocation. Her mother (P13, aged 38 yo) was found to have also past history of congenital hip dislocation and easy fatigue. On examination, a mild tetraparesis was revealed. So, apart from patient 11 in her first month of life, no patient in this group had facial, bulbar or axial muscle involvement.

The fourth phenotype includes four patients (P14 to P17) with a mild tetraparesis and no distinctive clinical feature. The two patients with later reported age of onset belong to this group (P14 and P15). They also have facial muscles weakness but no other additional feature. Patient 14 belongs to FAM.II and is a maternal uncle of patient 7 (different phenotype). Patient 17 presented with neonatal arthrogryposis that required several surgical procedures. He was only able to walk at the age of 6 and became fully functional.

None of the patients has cardiac involvement.

Eight patients (47.1%) are not aware of any family history suggestive of myopathy.

Histopathology

Fourteen patients (82.4%) had a muscle biopsy, at a mean age of 10.9 (1–29) yo. The examined muscle (deltoid) was abnormal in every patient, with four histopathologic myopathic patterns found (Fig. 1, panels B to G): centronuclear ($N = 5$), centralcore ($N = 4$), type 1 fibres predominance (T1FP; $N = 4$) and congenital fibre type disproportion (CFTD; $N = 1$). All patients with EOM weakness had the centronuclear pattern, except P6 who showed CFTD. From the four centralcore patterns identified, three belong to patients with isolated mild tetraparesis with or without facial weakness. The remaining belongs to patient 12, with the laxity component. The group of patients with T1FP was more heterogeneous in terms of clinical presentation but both patients with severe axial muscles involvement having this histopathologic pattern. From the other two T1FP patients, one was the youngest patient with the laxity (P11) and the other was one of the less severely affected patients (P15).

In the extraocular muscles weakness phenotype, the majority of cases showed severe myopathic changes. In the other phenotypes, the numbers are too small to drawing conclusions (Table 2).

Genetics

High genetic heterogeneity was found in this cohort of patients. Except for patients 4 and 5, each index case had different pathogenic variants in heterozygosity, the majority being of the missense type. Five pathogenic variants had not previously published in the literature (P4, P5, P6 and P12). As detailed in Table 1, variants found in compound heterozygosity were more dispersed along the *RYR1* gene. Cases with autosomal recessive (AR) inheritance patterns were more frequent in the group of patients with EOM involvement (P1 to P6). In addition, all the centronuclear histopathologic presentations were also found in this group. In the remaining patients, there was a predominance of pathogenic variants in hotspot 3 (exons 93 to 106) along with autosomal dominant (AD) inheritance patterns. All the central core histopathologic patterns were found in this last group. The variant c.7843 C>T (p.Arg2615Cys) identified in patient 8 has a paternal origin; considering that the patient's father is asymptomatic, its clinical significance remains unclear.

Malignant hyperthermia events were not reported in any of the patients or in their family members. However, three variants were previously associated with malignant hyperthermia susceptibility, by *in vitro* or functional contracture tests. They were identified in the first family with laxity (FAM.III), in one of the least disabled patients with EOM involvement (P15) and in one of the four patients with mild tetraparesis and facial weakness only (P16).

DISCUSSION

We report a rather heterogeneous group of 17 patients (including four families) with RYR1-related myopathies.

Clinically, all patients have muscle weakness and thin appearance of muscle bulk. We divided them in major groups according to three distinctive features: EOM weakness ($N = 6$), severe axial muscle weakness ($N = 2$) and marked joint laxity ($N = 5$). The fourth phenotype had mild tetraparesis with or without facial weakness and no clinical distinctive feature ($N = 4$).

Table 2
Correlation between clinical and histopathological severity. CFTD – congenital fibre type disproportion

Patient	Clinical phenotype	Clinical severity score	Hystopathological pattern	Fibre-size variation	Atrophic fibres	Fibrosis	Necrosis
1	Extraocular	4+2=6	Central nuclei	+++	+++	+	0
2	muscles weakness	4+3=7			NA		
3		2+3=5		+++	+++	0	0
4		1+0=1		+++	+++	0	0
5		2+0=2		+++	+++	0	0
6		1+0=1	CFTD	+++	+	0	0
7	Severe axial	2+0=2	Type 1 fibre predominance	+	+	0	0
8	muscles weakness	2+3=5		+++	+++	+	0
9	Joint laxity	2+0=2		+	+	0	0
10		1+0=1			NP		
11		1*+1=2			NP		
12		3+0=3	Central core	+	+	0	0
13		2+0=2			NP		
14	Mild tetraparesis	0+0=0	Central core		NP		
15	only ± facial	2+2=4	T1FP	+	+	0	0
16	weakness	2+0=2	Central core	+	+	0	0
17		2+2=4			NP		

T1FP – type 1 fibres predominance, NA – not available for review, NP – not performed,+- slight,++- moderate,+++- severe, * Still in the first year of life. Clinical severity score (adapted from Amburgey, 2013): Sum of ambulatory. (0: No Impairment, 1: Delay Motor Development, 2: Able to Walk Independently but with limitations, 3: Able to Walk with assistance, 4: Wheelchair dependent) and respiratory ratings (0: No Respiratory Impairment, 1: Respiratory Impairment as Neonate Only, 2: Some Respiratory Impairment without support, 3: Vent, CPAP, or BiPap at night, 4: Vent, CPAP, or BiPap during the day)

There was no significant association between EOM weakness and clinical severity, as the patients who presented with this feature were quite heterogeneous in this respect (considering ambulation, respiratory and bulbar involvement). This has also been previously reported [16].

There is clear disproportion between axial weakness and the ambulatory state in patients 3, 7 and 8 who were submitted to scoliosis surgery early in life while maintaining ambulation so far. This has been described by Colombo et al. in a large published series of congenital myopathies, 44.4% of whom with the genetic diagnosis of RYR1 related myopathy [2]. They also report that in RYR1-related myopathies, feeding difficulties can be transitory, as occurred with patient 11. Of interest, four out of six patients with EOM weakness also have axial weakness, with three of them having been submitted to early surgery. Therefore, the combination of these two patterns of muscle weakness may thus point towards the diagnosis of ryanodinopathy. Concerning the third phenotype that we describe, the association of RYR1 myopathy and laxity has been previously reported [7]. In our patients, laxity was mostly observed in proximal joints like shoulder, elbow, hip and knee, which may also help in the distinction from other myopathies with prominent joint hypermobility (e.g. collagen VI-related dystrophies).

In this regard, muscle biopsy can also guide genetic testing.

In terms of histopathology, among the 14 patients who had muscle biopsy, four different myopathic patterns were found, previously described as being associated with *RYR1* myopathies: centronuclear ($N = 5$), central core ($N = 4$), T1FP ($N = 4$) and CFTD ($N = 1$).

In our cohort, all centronuclear myopathies were associated with EOM weakness. Wilmshurst and collegues identified RYR1 mutations in 17 patients from a cohort of 24 patients with unresolved centronuclear myopathy. All patients except one had EOM weakness [8]. This association has also been described by other authors [9, 10].

Three of the patients with the T1FP pattern were already published by our group, emphasising the absence of central cores. The authors suggest that T1FP and involvement of axial muscles may be an important element to consider RYR1 as a candidate gene [11]. In Maggi and colleagues' series of 66 patients with a congenital myopathy, the identification of RYR1 mutations in 4 patients with isolated type 1 uniformity or predominance suggested that RYR1 mutations are a relatively common cause of these histological findings [12]. Sato et al. published ten unrelated Japanese patients diagnosed with congenital neuromuscular disease with uniform type 1,

four of whom (40%) had RYR1 mutations. These patients had milder clinical features compared with those without RYR1 mutations [13]. Our findings do not support such a clear association between T1FP and milder clinical involvement in RyR1 patients because half of them had severe axial involvement (P7 and P8).

Clarke et al. identified RYR1 mutations in four out of seven families with typical CFTD in whom other known genetic causes had been excluded. The authors suggested that RYR1 is a relatively common cause of this clinico-pathological pattern. They also noted an association between the presence of RYR1 mutations and ophthalmoplegia, proposing that it may be the most specific clinical indicator of RYR1 mutations in the setting of CFTD as this is not reported in the other known causes [14]. In agreement, our only CFTD case had EOM weakness.

The classic description of a *RYR1* congenital myopathy was a core myopathy but this paradigm has changed. Most of the previously cited papers that describe non-core histologic presentations propose that cores may evolve overtime, in a dynamic model. None of our patients have repeated the muscle biopsy and the mean age at muscle biopsy was in fact lower in central core (7.7yo) than in non-central core presentations (11.1 yo), as opposed to what would be expected based on the possible development of cores with age increasing. The small number of patients may contribute to this finding.

Concerning the genetic profiles found in our patient cohort there is a high genetic heterogeneity and variability. The majority of cases had different *RYR1* variants including four described for the first time in this work.

It is not infrequent to find variant of uncertain significance (VUS) during these studies. Particularly those of missense type pose additional interpretation difficulties and as consequence the clinical utility of these studies is sometimes limited. According to the most recent guidelines for sequence variants interpretation from the American College of Medical Genetics and Genomics, several lines of evidences should be considered when performing variant pathogenicity assessment [15].

As an example, the 2348 C>T (p.Ser783Leu) variant if interpreted individually would be classified as VUS (class 3), in spite of the bioinformatics analysis suggesting its deleterious effect and absence from population variant databases. However, when interpreted together with the c.1342A>T (p.Ile448Phe) variant (detected in another patient of our laboratory cohort and considered likely pathogenic) would be in favour of being causative for RYR1-related myopathy. Nonetheless, it would be important to further characterize the functional properties of these variants as previously demonstrated in the literature [16, 17].

We attempted to correlate the sequence variants distribution found in this cohort and their impact on the structure and/or activity of ryanodine receptor type 1 using the information available in previous reports.

It is well known that the majority of patients with autosomal dominant central core disease have *RYR1* pathogenic variants that are clustered in the hydrophobic COOH-terminal region (hotspot domain 3) of this receptor. Zhou et al. published a series of 28 patients in which the dominant *RYR1* variants were mainly found in the C-terminal and only in the central domain of *RYR1*. These authors also stated that most patients with dominant C-terminal variants had the classical central core phenotype, characterized by variable degrees of proximal weakness and absence of significant bulbar, respiratory or extraocular involvement [18]. We have also found these milder phenotypes in our patients with AD inheritance, having variants located in hotspot 3 and central cores in their muscle biopsies. It is interesting to note that the majority of pathogenic variants found in our AD patients are within a small stretch of 38 residues (4861–4898) of the protein. Since the high (near-atomic) resolution structure of RyR1 has been recently described, it is now evident that these residues correspond to critical segments of the ion-conducting pathway: the luminal loop, the pore helix and the selectivity filter [19]. In particular, the p.Arg4893Trp variant (found in P7 and P14) may affect one of the potential residues that constitute the inner cation-binding site. These structures were proposed to be a hotspot for disease-related mutations [9, 19]. Some variants in this region have been further characterized, namely the p.Ile4898Thr variant found in family FAM.III, which was found to create a dominant-negative suppression of RYR1 channel calcium ion permeation [20].

The genetic variants associated with autosomal recessive inheritance and other histopathological patterns (e.g. centronuclear) are distributed along the entire coding sequence [5, 9, 21]. Our series is concordant with this description. Among 26 patients with *RYR1* pathogenic variants, Maggi and colleagues (2013) identified compound heterozygous variants consistent with recessive inheritance

distributed throughout the *RYR1* coding sequence in 15 patients (57.7%) [12]. Amburgey and collaborators attempted to establish genotype-phenotype correlations in a large cohort of recessive RYR1-related cases [9]. Our data is in accordance with the main observations of this work. Hypomorphic mutations (nonsense and frame-shifting, expected to reduce RyR1 expression) were identified in AR non-central core myopathies (P1 to P5) and in association with more severe clinical phenotypes. In our cohort, all patients with AR inheritance pattern and at least one hypomorphic allele had EOM weakness and centronuclear myopathy. This was previously described by other previous works [6, 9, 12, 18, 22].

Regarding malignant hyperthermia, we identified three genetic variants (five patients) with proven susceptibility. In concordance with previous reports, this susceptibility may exist even in the context of a mild and/or non centralcore RYR1 myopathy [23].

In conclusion, there is a great variability in clinical presentation, genetics and muscle pathology in patients with ryanodinopathies but some clinical features may be diagnostically relevant: extraocular muscle weakness (with centronuclear pattern on muscle biopsy), severe axial weakness disproportionate to the ambulatory state and mild tetraparesis associated with (proximal) joint laxity.

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

REFERENCES

[1] Amburgey K, McNamara N, Bennett LR, McCormick ME, Acsadi G, Dowling JJ. Prevalence of congenital myopathies in a representative pediatric united states population. Annals of Neurology. 2011;70(4):662-5.

[2] Colombo I, Scoto M, Manzur AY, Robb SA, Maggi L, Gowda V, et al. Congenital myopathies: Natural history of a large pediatric cohort. Neurology. 2015;84(1):28-35.

[3] Treves S, Anderson AA, Ducreux S, Divet A, Bleunven C, Grasso C, et al. Ryanodine receptor 1 mutations, dysregulation of calcium homeostasis and neuromuscular disorders. Neuromuscular Disorders : NMD. 2005;15(9-10):577-87.

[4] Robinson R, Carpenter D, Shaw MA, Halsall J, Hopkins P. Mutations in RYR1 in malignant hyperthermia and central core disease. Human Mutation. 2006;27(10):977-89.

[5] Treves S, Jungbluth H, Muntoni F, Zorzato F. Congenital muscle disorders with cores: The ryanodine receptor calcium channel paradigm. Current Opinion in Pharmacology. 2008;8(3):319-26.

[6] Klein A, Lillis S, Munteanu I, Scoto M, Zhou H, Quinlivan R, et al. Clinical and genetic findings in a large cohort of patients with ryanodine receptor 1 gene-associated myopathies. Human Mutation. 2012;33(6):981-8.

[7] Donkervoort S, Bonnemann CG, Loeys B, Jungbluth H, Voermans NC. The neuromuscular differential diagnosis of joint hypermobility. American Journal of Medical Genetics Part C, Seminars in Medical Genetics. 2015;169c(1):23-42.

[8] Wilmshurst JM, Lillis S, Zhou H, Pillay K, Henderson H, Kress W, et al. RYR1 mutations are a common cause of congenital myopathies with central nuclei. Annals of Neurology. 2010;68(5):717-26.

[9] Amburgey K, Bailey A, Hwang JH, Tarnopolsky MA, Bonnemann CG, Medne L, et al. Genotype-phenotype correlations in recessive RYR1-related myopathies. Orphanet Journal of Rare Diseases. 2013;8:117.

[10] Jungbluth H, Gautel M. Pathogenic mechanisms in centronuclear myopathies. Frontiers in Aging Neuroscience. 2014;6:339.

[11] Rocha J, Taipa R, Melo Pires M, Oliveira J, Santos R, Santos M. Ryanodine myopathies without central cores–clinical, histopathologic, and genetic description of three cases. Pediatric Neurology. 2014;51(2):275-8.

[12] Maggi L, Scoto M, Cirak S, Robb SA, Klein A, Lillis S, et al. Congenital myopathies–clinical features and frequency of individual subtypes diagnosed over a 5-year period in the United Kingdom. Neuromuscular Disorders: NMD. 2013;23(3):195-205.

[13] Sato I, Wu S, Ibarra MC, Hayashi YK, Fujita H, Tojo M, et al. Congenital neuromuscular disease with uniform type 1 fiber and RYR1 mutation. Neurology. 2008;70(2):114-22.

[14] Clarke NF, Waddell LB, Cooper ST, Perry M, Smith RL, Kornberg AJ, et al. Recessive mutations in RYR1 are a common cause of congenital fiber type disproportion. Human Mutation. 2010;31(7):E1544-50.

[15] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. 2015;17(5):405-24.

[16] Lyfenko AD, Ducreux S, Wang Y, Xu L, Zorzato F, Ferreiro A, et al. Two central core disease (CCD) deletions in the C-terminal region of RYR1 alter muscle excitation-contraction (EC) coupling by distinct mechanisms. Human Mutation. 2007;28(1):61-8.

[17] Goonasekera SA, Beard NA, Groom L, Kimura T, Lyfenko AD, Rosenfeld A, et al. Triadin binding to the C-terminal luminal loop of the ryanodine receptor is important for skeletal muscle excitation contraction coupling. The Journal of General Physiology. 2007;130(4):365-78.

[18] Zhou H, Jungbluth H, Sewry CA, Feng L, Bertini E, Bushby K, et al. Molecular mechanisms and phenotypic variation in RYR1-related congenital myopathies. Brain: A Journal of Neurology. 2007;130(Pt 8):2024-36.

[19] Yan Z, Bai XC, Yan C, Wu J, Li Z, Xie T, et al. Structure of the rabbit ryanodine receptor RyR1 at near-atomic resolution. Nature. 2015;517(7532):50-5.

[20] Loy RE, Orynbayev M, Xu L, Andronache Z, Apostol S, Zvaritch E, et al. Muscle weakness in Ryr1I4895T/WT knock-in mice as a result of reduced ryanodine receptor Ca2+ion permeation and release from the sarcoplasmic reticulum. The Journal of General Physiology. 2011;137(1):43-57.

[21] Wu S, Ibarra MC, Malicdan MC, Murayama K, Ichihara Y, Kikuchi H, et al. Central core disease is due to RYR1 mutations in more than 90% of patients. Brain: A Journal of Neurology. 2006;129(Pt 6):1470-80.

[22] Snoeck M, van Engelen BG, Kusters B, Lammens M, Meijer R, Molenaar JP, et al. RYR1-related myopathies: A wide spectrum of phenotypes throughout life. European Journal of Neurology. 2015;22(7):1094-112.

[23] Rueffert H, Wehner M, Ogunlade V, Meinecke C, Schober R. Mild clinical and histopathological features in patients who carry the frequent and causative malignant hyperthermia RyR1 mutation p.Thr2206Met. Clinical Neuropathology. 2009;28(6):409-16.

[24] Oliveira J, Goncalves A, Taipa R, Melo-Pires M, Oliveira ME, Costa JL, et al. New massive parallel sequencing approach improves the genetic characterization of congenital myopathies. Journal of Human Genetics. 2016;61(6): 497-505.

[25] Abath-Neto OL. Estudo clínico, histológico e molecular da miopatia centronuclear [Dissertation]2014.

[26] Monnier N, Romero NB, Lerale J, Landrieu P, Nivoche Y, Fardeau M, et al. Familial and sporadic forms of central core disease are associated with mutations in the C-terminal domain of the skeletal muscle ryanodine receptor. Human Molecular Genetics. 2001;10(22):2581-92.

[27] Lynch PJ, Tong J, Lehane M, Mallet A, Giblin L, Heffron JJ, et al. A mutation in the transmembrane/luminal domain of the ryanodine receptor is associated with abnormal Ca2+ release channel function and severe central core disease. Proceedings of the National Academy of Sciences of the United States of America. 1999;96(7):4164-9.

[28] Duarte ST, Oliveira J, Santos R, Pereira P, Barroso C, Conceicao I, et al. Dominant and recessive RYR1 mutations in adults with core lesions and mild muscle symptoms. Muscle & Nerve. 2011;44(1):102-8.

[29] Kraeva N, Heytens L, Jungbluth H, Treves S, Voermans N, Kamsteeg E, et al. Compound RYR1 heterozygosity resulting in a complex phenotype of malignant hyperthermia susceptibility and a core myopathy. Neuromuscular Disorders: NMD. 2015;25(7):567-76.

[30] Gambelli S, Malandrini A, Berti G, Gaudiano C, Zicari E, Brunori P, et al. Inheritance of a novel RYR1 mutation in a family with myotonic dystrophy type 1. Clinical Genetics. 2007;71(1):93-4.

[31] Davis MR, Haan E, Jungbluth H, Sewry C, North K, Muntoni F, et al. Principal mutation hotspot for central core disease and related myopathies in the C-terminal transmembrane region of the RYR1 gene. Neuromuscular Disorders: NMD. 2003;13(2):151-7.

Research Report

Reviewing Large LAMA2 Deletions and Duplications in Congenital Muscular Dystrophy Patients

Jorge Oliveira[a,e,f], Ana Gonçalves[a,f], Márcia E. Oliveira[a,f], Isabel Fineza[b], Rita C.M. Pavanello[c],
Mariz Vainzof[c], Elsa Bronze-da-Rocha[d], Rosário Santos[a,d,f,1,*] and Mário Sousa[e,f,1]

[a] *Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto de Magalhães, Centro Hospitalar do Porto, Porto, Portugal*
[b] *Unidade de Neuropediatria, Centro de Desenvolvimento da Criança Luís Borges, Hospital Pediátrico de Coimbra, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal*
[c] *Centro de Estudos do Genoma Humano e Células Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brasil*
[d] *Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal*
[e] *Departamento de Microscopia, Laboratório de Biologia Celular, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal*
[f] *Unidade Multidisciplinar de Investigação Biomédica-UMIB, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal*

Abstract.
Background: Congenital muscular dystrophy (CMD) type 1A (MDC1A) is caused by recessive mutations in laminin-α2 (*LAMA2*) gene. Laminin-211, a heterotrimeric glycoprotein that contains the α2 chain, is crucial for muscle stability establishing a bond between the sarcolemma and the extracellular matrix. More than 215 mutations are listed in the *locus* specific database (LSDB) for *LAMA2* gene (May 2014).
Objective: A limited number of large deletions/duplications have been reported in *LAMA2*. Our main objective was the identification of additional large rearrangements in *LAMA2* found in CMD patients and a systematic review of cases in the literature and LSDB.
Methods: In four of the fifty-two patients studied over the last 10 years, only one heterozygous mutation was identified, after sequencing and screening for a frequent *LAMA2* deletion. Initial screening of large mutations was performed by multiplex ligation-dependent probe application (MLPA). Further characterization implied several techniques: long-range PCR, cDNA and Southern-blot analysis.
Results: Three novel large deletions in *LAMA2* and the first pathogenic large duplication were successfully identified, allowing a definitive molecular diagnosis, carrier screening and prenatal diagnosis. A total of fifteen deletions and two duplications previously reported were also reviewed. Two possible mutational "hotspots" for deletions may exist, the first encompassing exons 3 and 4 and second in the 3' region (exons 56 to 65) of *LAMA2*.

---

[1]Authors contributed equally to this work.

*Correspondence to: Rosário Santos, Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto de Magalhães, Centro Hospitalar do Porto, Praça Pedro Nunes, 88, 4099-028 Porto, Portugal. Tel.: +351 226 070 304; Fax: +351 226 070 399; E-mail: rosario.santos@chporto.min-saude.pt.

Conclusions: Our findings show that this type of mutation is fairly frequent (18.4% of mutated alleles) and is underestimated in the literature. It is important to include the screening of large deletions/duplications as part of the genetic diagnosis strategy.

Keywords: LAMA2, congenital muscular dystrophy, large deletion, large duplication, review

INTRODUCTION

LAMA2-related dystrophy (LAMA2-RD) collectively gathers two distinct clinical entities: the classical phenotype with congenital onset known as MDC1A, and a milder limb-girdle type muscular dystrophy with onset during childhood (late-onset or "ambulant" LAMA2-RD) [reviewed by 1, 2]. As foreseeable by this designation, these entities are caused by recessive mutations in *LAMA2* gene located on chromosome 6q22-23 and spanning 65 exons [3, 4]. This gene codes for the α2 chain of laminin-211, an extracellular glycoprotein expressed in the basal membrane of striated muscles, peripheral nerves, brain and trophoblast [5–7]. The interaction of laminin-211 with cell-surface receptors such as α-dystroglycan and integrin (mainly α7β1 in adult skeletal muscle) explains its major relevance in the overall extracellular architecture, integrity and cell adhesion [reviewed by 8].

MDC1A represents the most frequent form of CMD in western countries, accounting for 30 to 50% of cases [9]. Typical clinical features includes severe hypotonia associated with muscle weakness manifesting at birth or during early infancy, proximal joint contractures, elevated creatine kinase (CK) levels, cerebral white matter abnormalities and delayed motor milestones with affected children usually not achieving independent ambulation [10–12]. Feeding problems and respiratory insufficiency are commonly reported complications often requiring gastrostomy and/or artificial ventilator support [13, 14].

Other features such as cardiac involvement, a sensory and motor demyelinating neuropathy, epilepsy and mental retardation have been also documented in some forms of LAMA2-RD [15–19].

An important diagnostic aspect is that skeletal muscle biopsies from these patients have changes in laminin-α2 expression detected by immunohistochemical (IHC) analysis [10]. However, there is a recent report of a muscular dystrophy patient with apparently normal laminin-α2 IHC expression and having mutations in *LAMA2* gene [20].

Milder LAMA2-RD cases have been reported in the past few years expanding the phenotypic spectrum of the disease [21]. These patients have slower disease progression and acquire independent locomotion,

and are usually associated with a partial expression of laminin-α2 [22–26].

More than 375 distinct sequence variants (215 of them with known clinical relevance) have been reported in the LSBD for *LAMA2* gene (http://www.dmd.nl/LAMA2, data accessed May 2014). Pathogenic changes include small deletions/insertions (34.9%), nonsense mutations (25.1%), changes affecting splicing (25.6%) and also missense substitutions (12.1%). In spite of the relevant amount of mutational data available there is still a limited number (2.3%) of large deletions and duplications reported in this gene. The initial suspicion of a large *LAMA2* deletion (which was predicted to include exons 23 to 56) was identified by protein truncation test in the work of Pegoraro and collaborators [27]. The first fully characterized large deletion in *LAMA2*, corresponds to an out-of-frame deletion of exon 56 (c.7750-1713_7899-2154del), which has been proven to be one of the most frequent pathogenic variants detected in Portuguese MDC1A patients [28].

One of the main objectives of this work was to describe additional novel pathogenic deletions and duplications associated with the *LAMA2* gene, identified in our cohort of CMD patients. Moreover, a systematic review of all cases with large deletions/duplications reported in the literature and mutation databases is presented. Our findings showed that this type of mutation is fairly frequent and is underestimated in the literature, reinforcing the importance to screen large deletions/duplications in *LAMA2* gene as part of the genetic diagnosis strategy.

MATERIAL AND METHODS

Patients

Over the last 10 years (2004-2014) our group performed genetic studies in 94 CMD patients. Mutations in genes related with CMD were identified in 68% (n = 64) of these patients. The majority of these patients have *LAMA2* mutations (n = 52) and were referred for molecular studies due to changes in muscle laminin-α2 detected by IHC analysis and/or compatible white matter anomalies detected by magnetic resonance imaging (MRI). In four patients of this cohort only

Table 1
Clinical data of CMD patients with novel large deletions and duplications in *LAMA2*

Patient	Sex	Age*	Age of onset	Clinical presentation	Highest CK (IU/l)	Pattern/ progression of weakness	Best motor achievement	Contractures	Central nervous system involvement/ seizures	Magnetic resonance imaging	Laminin-α2 in muscle
P1 (P19 in [28])	M	20 yr	At birth	Generalized hypotonia and areflexia	3264	Muscular weakness with axial and proximal predominance and scoliosis	Independent ambulation	Elbows and ankles	No cognitive delay and no seizures	White matter changes and no gyral abnormalities	nd
P2 (P24 in [28])	M	3 yr	At birth	Hypotonia and feeding problems	1770	Muscular weakness with proximal predominance and hip congenital luxation	Cephalic control and assisted trunk control	Knees	No cognitive delay and no seizures	White matter changes and no gyral abnormalities	Total absence
P3	M	2 yr	4 mo	Hypotonia and muscular weakness	7644	Muscular weakness with axial and proximal UL predominance	Cephalic control and assisted trunk control	Knees, ankles and rigid spine	No cognitive delay and no seizures	White matter changes	nd
P4	M	8 mo	At birth	Generalized hypotonia	4400	Muscular weakness with proximal UL predominance	Cephalic and trunk control	Discrete equinus	No cognitive delay and no seizures	nd	Partial deficiency

CK–creatine phosphokinase; M– male; mo– months; nd– not determined; P– patient; UL– upper limbs; yr– years; *– age at last clinical follow-up.

one heterozygous mutation was detected upon *LAMA2* genomic sequencing (described in Table 1). In these patients we conducted screening of large deletions and duplications in the *LAMA2* gene. This research was approved by the ethics committee from Hospital Centre of Porto (CHP).

LAMA2 gene analysis

Genomic DNA (gDNA) was obtained from peripheral blood using the salting-out method [29]. *LAMA2* gene sequencing was done according to [28], which comprised all coding and adjacent intronic sequences of *LAMA*2. Variants were described according to the Human Genome Variation Society (HGVS) guidelines for mutation nomenclature (version 2.0) [30] and using the cDNA reference sequence with accession number NM_000426.3.

MLPA analysis

Screening for deletions and duplications in *LAMA2* gene was performed by multiplex ligation-dependent probe application (MLPA) technique using two sets of probe mixes (P391-A1 and P392-A1) from MRC-Holland (Amsterdam, the Netherlands). These probe mixes contain one probe for each exon of the gene with the exception of exons 18, 44 and 48. Two probes are present for exon 1, 2, 4 and 65 and three probes for exon 56. Also, probemix P391 contains 9 reference probes and P392 contains 8 reference probes detecting different genomic regions. For the MLPA procedure, 150 ng gDNA was used for each patient and normal control samples. Amplification products were subsequently separated by capillary electrophoresis on an ABI 3130xl genetic analyser (Applied Biosystems, Foster City, CA). Data analysis was conducted using GeneMarker Software V1.5 (SoftGenetics LLC, State College, PA). Population normalization method was selected and data was plotted using probe ratio.

Southern blot

gDNA samples from patient P3, respective parents and normal controls were digested with *BglI* (NewEngland Biolabs, Beverly, MA), resolved on a 0.7% agarose gel and vacuum transferred to a GeneScreen Plus membrane (Perkin Elmer, Waltham, MA) using a saline method. A cDNA probe recognizing exons 2–4 was prepared using digoxigenin (DIG) DNA Labeling Kit (Roche Applied Science, Indianapolis, IN, USA) and incubated overnight using the Easy Hyb Buffer (Roche Applied Science). The membrane was washed

at 60°C in 1xSSC (Saline-Sodium Citrate)/0.1% SDS (Sodium Dodecyl Sulfate) (Sigma-Aldrich, St. Louis, MO) and twice in 0.5xSSC/0.1% SDS for 15 min each. Subsequently, the membrane was prepared with DIG Wash and Block Buffer Set (Roche Applied Science), incubated with Anti-DIG-AP conjugate (Roche Applied Science), and the DIG-labeled probe detected with ready-to-use CDP-Star (Roche Applied Science).

LAMA2 cDNA analysis

cDNA studies were carried out in patient P2. Total RNA was extracted from patient and control muscle biopsy samples using the PerfectPure RNA Fibrous Tissue kit (5 PRIME, Germany) and converted to cDNA using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA). *LAMA2* transcripts were subjected to PCR amplification using specific primers for the region corresponding to exons 27–32 (27F- 5′AAATTTCATGCGACAAAGCAGG; 32R-5′GCTTGCAGCCGTCACACTTC). Resulting PCR products were purified and sequenced as described before.

Long-range PCR

The deletion breakpoints encompassing exon 3 (patient P3) and also exon 17 (patient P1) were determined by amplification of gDNA using the BIO-X-ACT Long DNA Polymerase kit (Bioline, Taunton, MA). Specific primers were designed for each case; complementary to intron 2 and exon 4, for the deletion of exon 3 (2i-F:5′ACAAAGCCTGATGGAGG GAAAC; 4-R:5′AAAGCGTTAGGCACTCCGTGTC) and complementary to regions in exons 16 and 18, for the deletion of exon 17 (16F:5′-TTGGTCATG CGGAGTCCTG; 18R:5′ TGGCACGTTGGGC-TAAAGC). Resolved PCR fragments were purified using the QIAquick® Gel Extraction Kit (Qiagen, Valencia, CA), and subsequently sequenced.

RESULTS

Novel large deletions and duplications

We previously reported a large deletion encompassing exon 56 of the *LAMA2* gene which is relatively frequent in our laboratory patient cohort (present in 23% of our cases) [28]. This initial study suggested that it is clinically relevant to screen this type of mutations in CMD patients, and indirectly drove the development of a MLPA commercial kit for *LAMA2* gene (P391-A1

and P392-A1 from MRC-Holland). As part of this work we initially assessed the effectiveness of the MLPA kit using previously genotyped DNA samples from our patient cohort. Patients presenting homozygous and heterozygous deletions encompassing *LAMA2* exon 56 were tested. MLPA technique successfully detected this mutation in homo- and heterozygous states (Supplementary data I).

Four patients are presented in detail in this work; P1 and P2 were previously described in the literature as P19 and P24, respectively in [28] while two additional patients (P3 and P4) were more recently referred for *LAMA2* gene analysis. All of these patients had incomplete molecular characterization: only one heterozygous mutation was detected by genomic sequencing and thus the molecular defect in the other *LAMA2* allele remained unknown. MLPA technique was performed in DNA samples from these patients to tentatively identify the second pathogenic mutation in *LAMA2*.

Patient P1 presents a congenital muscular dystrophy (neonatal onset) and remained ambulatory until the age of 17 years. In this patient a heterozygous codon deletion (c.1798_1800del, p.Gly600del) was initially detected in *LAMA2*. MLPA analysis for this patient revealed reduced hybridization of one probe corresponding to exon 17 (Fig. 1A), compatible with the presence of a heterozygous deletion involving this exon. By reviewing genomic sequencing data we excluded a potential sequence change which could compromise the affinity of the MLPA probe. A long-range PCR experiment was performed using primers designed to bind regions presumably not involved in the deletion (exons 16 and 18). Upon amplification, two PCR products were detected in the patient, whereas the experimental control had a single band (data not shown). Sequencing across the deletion breakpoint revealed that part of intron 16 is joined to intron 17, corresponding to the loss of ~5.3Kb that spans exon 17 (Fig. 1B). This novel large mutation, c.2322+259_2450+2037del, predictably causes frameshifting. DNA samples from the patient's parents were unavailable for study.

Patient P2 was referred for molecular study at 3 years of age having a typical MDC1A phenotype with absence of laminin-α2 in IHC analysis of muscle. The heterozygous nonsense mutation c.3085C>T (p.Arg1029*) was the only pathogenic sequence change detected by sequencing the *LAMA2* gene. A large heterozygous duplication encompassing exons 28 and 29 was identified in patient P2 by MLPA (Fig. 2A). cDNA studies performed in the

Patient P1

A- MLPA analysis
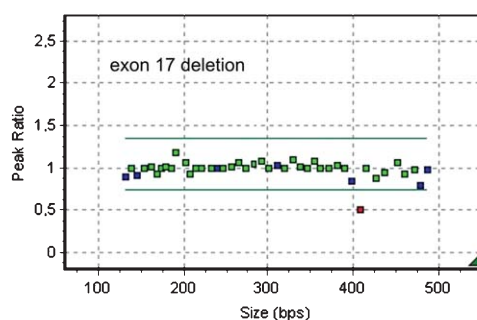
B- Breakpoint sequencing
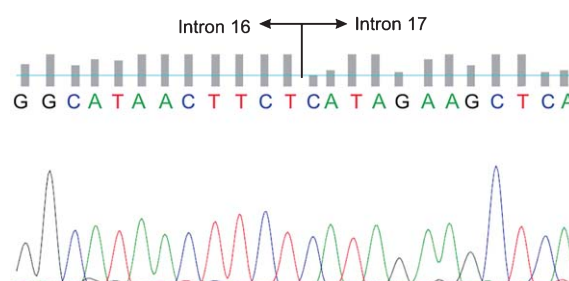
Fig. 1. Results obtained for patient P1. (A) MLPA technique showing reduction of one probe (0.5 peak ratio) corresponding to exon 17 of the *LAMA2* gene. (B) Sequencing electropherogram of a fragment (obtained from long-range PCR) that enabled the identification of the deletion breakpoint, corresponding to the loss of ~5.3 Kb.

patient revealed a normal transcript together with other abnormal PCR products (Fig. 2B). These include one out-of-frame transcript resulting from the contiguous duplication of exons 28 and 29 (Fig. 2C).

Patient P3 with an MDC1A phenotype has a novel nonsense mutation in exon 4 (c.497G>A, p.Trp166*) detected in a heterozygous and apparently homozygous state, depending on the primer-pair used to study this region (Fig. 3A). These ambiguous results led us to suspect a possible deletion comprising at least part of intron 3. The application of MLPA confirmed this assumption, since a reduced amplification signal was observed for the exon 3-specific probe (Fig. 3B). For further characterization, and since no RNA was available for study, we performed Southern-blotting and hybridization using a cDNA probe that recognizes exons 1 to 4. This experiment suggested that the genomic deletion originates a new fragment of approximately 6 Kb in the patient, that is absent in the control (Fig. 3C). To delineate the deletion endpoints several primers were tested to perform a deletion-specific PCR. A 42931 bp deletion combined

Patient P2

A- MLPA analysis

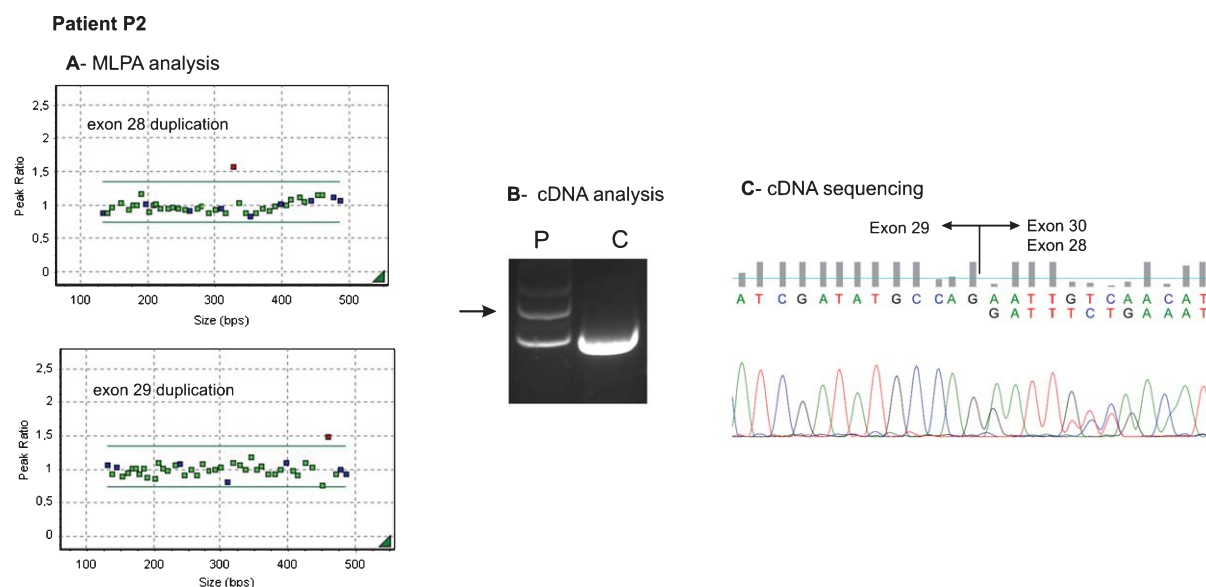

B- cDNA analysis

C- cDNA sequencing

Fig. 2. Duplication identified in patient P2. (A) Increased signal for MLPA probes (peak ratios around 1.5) recognizing exon 28 (upper panel) and exon 29 (lower panel). (B) cDNA analysis of *LAMA2* transcripts revealed the presence of abnormal PCR products in the patient that were not detected in the control. (C) Sequencing electropherogram of the cDNA PCR product (indicated by arrow) reveals the adjacent duplication of exons 28 and 29. C- control; P- patient.

with the insertion of three nucleotides was identified, annotated as c.284-4685_397-146delinsATA (Fig. 3D). Compound heterozygosity for these two mutations was confirmed by the analysis of patient's parents.

Lastly, patient P4 has also an MDC1A phenotype but partial laminin-α2 absence in muscle. Besides one heterozygous 8 bp duplication in exon 13 (c.1854_1861dup) previously reported in the literature [9], we were able to identify a large heterozygous deletion involving exons 3 to 10 by MLPA (Fig. 4). This deletion is predicated to be out-of-frame. A more extensive characterization was not possible since no RNA sample was available for study. Compound heterozygosity was verified since each parent carried a different mutation.

Novel point mutations in LAMA2

Fifty-two patients with *LAMA2* mutations have been characterized until now: 26 of which were previously reported in 2008, and another two more recently in a publication describing their atypical phenotype associated with novel missense mutations [19]. In all cases mutations have been identified in both disease alleles. A list of 10 novel mutations is shown in supplementary data II. These include four nonsense mutations (c.3520C>T, c.5263A>T, c.6501C>G e c.6979G>T), four changes affecting splice sites (c.396+1G>T, c.2450+4A>G, c.6708-1G>T and c.8988+1G>A), one

single nucleotide duplication (c.2350dupT) and a missense mutation (c.3235T>G, p.Cys1079Gly).

Reviewing large deletions and duplications in LAMA2

As part of this work we reviewed all large deletions and duplications reported in the literature or in the *locus* specific database for the *LAMA2* gene (http://www.lovd.nl/LAMA2, information last accessed in May 2014). In addition to the mutations presented in this publication, 12 different large deletions and a single duplication were previously described (Table 2).

LAMA2 large deletions (*n* = 15, reported in 35 patients) are apparently dispersed throughout the gene, but two possible mutational "hotspots" may exist: i) one includes exons 3 and 4 (5 different deletions) and ii) in the 3' region of the gene (from exons 56 to 65). The largest *LAMA2* deletion, encompassing exons 23 to 56, was detected by the protein truncation test [27] and comprehends more than half of the gene's coding regions. At least six different deletions affect single exons and were confirmed by a second technique in order to exclude false positive results. The majority of deletions are predicted to cause frameshifting (out-of-frame deletions). Still, only two cases were further characterized at the cDNA level which limits the accuracy of this data. Considering deletions

Patient P3

A- gDNA sequencing

c.497G>A
Hom.

A G C C C T A G C A G T A

c.497G>A
Het.

A G C C C T G G C A G T A
A

C- Southern Blot

B- MLPA analysis

D- Breakpoint sequencing

Intron 2 ← → Intron 3
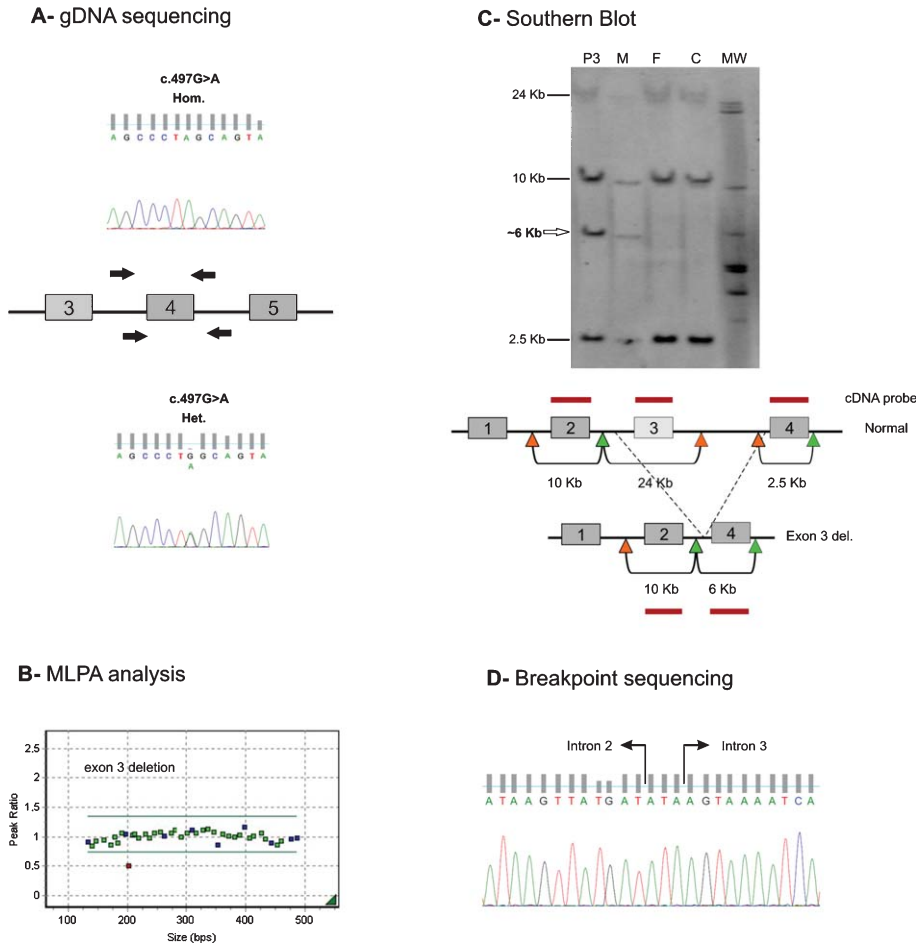
A T A A G T T A T G A T A T A A G T A A A A T C A

Fig. 3. Characterization of a heterozygous deletion in patient P3. (A) The initial suspicion of a heterozygous deletion encompassing intron 3 derived from the genomic sequencing data. The nonsense mutation c.497G>A located in exon 4 was detected both in heterozygosity and homozygosity depending upon the primers used. (B) Confirmation by MLPA, with reduced amplification signal for the exon 3 probe. (C) Southern-blot followed by hybridization using a cDNA probe that recognizes exons 1 to 4, revealing a 6 Kb fragment in the patient and in his mother, but not in the father nor the control. (D) Deletion breakpoint identified by deletion-specific PCR followed by sequencing. Sequencing electropherogram revealed a 42.9 Kb deletion combined with the insertion of three nucleotides. C- control; F- father; M- mother; MW- molecular weight marker; P3- patient 3.

predicted to be in-frame, all except one [31] were reported in combination with a second truncating mutation (causing a frame-shift or a nonsense mutation). All of these patients have typical MDC1A phenotypes, except the patient with the deletion of exons 41 to 48, which remains ambulant at the age of 10 years (patient #6, reported by [32]). Six patients have been described with large homozygous deletions which may be explained by a higher frequency of a particular mutation within the population (deletion of exon 56 in Portugal and exon 4 deletion in Chinese patients) or due to consanguinity in individual sporadic cases. Patients with homozygous deletions are usually detected by

genomic sequencing, since the affected regions will fail to amplify during PCR. Until now, no patients have been reported with compound heterozygosity between two different large rearrangements.

LAMA2 duplications are even rarer mutational events; besides our report of a novel duplication encompassing exons 28 and 29 detected in patient P2, only one other heterozygous in-frame duplication involving exons 5 to 12 has been documented in a patient presenting muscular dystrophy [31]. However, these authors did not identify a second mutation in the patient, which might have explained an autosomal recessive LAMA2-RD.
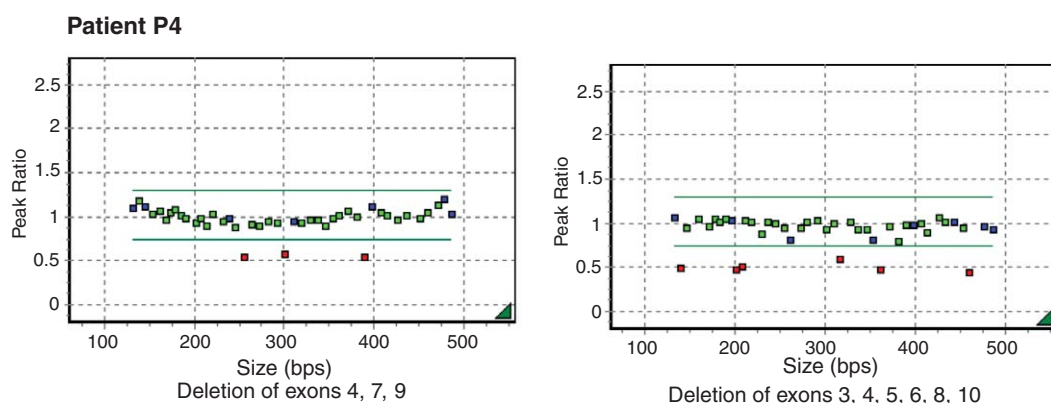
Fig. 4. MLPA results for patient 4. Several *LAMA2* probes with decreased peak ratios (0.5), corresponding to exons 3 to 10, compatible with a large heterozygous deletion.

The frequency of large deletions and duplications in *LAMA2* may be estimated based on the two largest patient cohorts reported and that employed quantitative techniques: 18/104 alleles from our patient cohort of 52 patients and 17/86 alleles from the recent work of Xiong and collaborators that studied 43 patients [32]. The overall frequency of these mutations is thus around 18.4% (35/190 alleles).

DISCUSSION

This work describes the detailed genetic characterization of four patients with compatible features with a MDC1A phenotype, but with only one heterozygous pathogenic sequence variant detected upon complete *LAMA2* sequencing. Three novel large deletions in the *LAMA2* gene and the first pathogenic large duplication were successfully identified in this group of patients, allowing a definitive molecular diagnosis, carrier screening and prenatal diagnosis. Characterization of these mutations implied the use of a variety of techniques such as long-range PCR, cDNA and Southern-blot analysis. These methods are not generally used in the routine genetic diagnosis of this disease, but are essential to obtain accurate genotype-phenotype correlations.

Up to now reports of large deletions and duplications in *LAMA2* are very rare; only three publications have referred this type of mutation [28, 31, 32]. The work of the Italian group included a more heterogeneous patient cohort and a broader technical approach. An array-based comparative genomic hybridization (array-CGH) developed to screen genes implicated in neuromuscular diseases, enabled the identification of several novel copy number variants (CNVs) including two present in the *LAMA2* gene [31]. Based on the data from a total of 95 fully genotyped LAMA2-RD patients, from two large cohorts, we estimate that the frequency of large deletions and duplications in *LAMA2* may be as high as 18.4%. Considering this relatively high frequency, it is important to include screening techniques such as MLPA or array-CGH in the molecular diagnostic work-up. Here, laboratories should consider the variety of equipments required, running costs and sensitivity of these two approaches to screen this type of rearrangement. The presence of a single heterozygous large deletion or duplication, especially when in-frame, should be carefully evaluated. It is conceivable that the presence of a non-pathogenic CNV in a CMD patient may not necessarily explain the clinical phenotype.

Readers should also be aware that genomic sequencing is the technique with the highest sensitivity to screen for *LAMA2* mutations (>80%), especially in CMD cases with laminin-α2 deficiency. Our current strategy for *LAMA2* genetic analysis is sub-divided in three tiers: i) the first level comprising a selected number of exons (namely: 3, 13, 22, 27, 33, 36, 54, 58, and 61) corresponding to the genomic regions where the majority of point mutations in our population are located, together with the screening of exon 56 deletion; ii) the second tier includes the remaining *LAMA2* exons; and finally iii) MLPA analysis (two panels). Until now, seventeen patients have been analyzed in this manner. In 35% of patients both mutated alleles were identified using tier 1, and in 82% at least one heterozygous mutation was detected. We consider feasible in our population to screen these *LAMA2* regions in patients with compatible features of CMD (such as white matter changes in brain MRI), even before performing a muscle biopsy.

Table 2
Review of large deletions and duplications reported in the *LAMA2* gene

Affected gene regions (exons)	Nr. of affected exons	Mutation description		Impact on reading frame (prediction)	Nr. of patients reported	Zygosity // other mutation type	Phenotype	References
		gDNA	RNA					
Deletions								
2–3	2	c.113-?_396+?del	r.(del)	OF	2	het. (n=2) // splice-site mutation	MDC1A; MDC1A (ambulant at 5yr)	[32]
3	1	c.284-4685_397-146delinsATA	r.(del)	OF	1	het. // nonsense mutation	MDC1A	P3, this paper
3–4	2	c.284-?_639+?del	r.(del)	OF	1	hom.	MDC1A	[32]
3–10	8	c.284-?_1467+?del	r.(del)	OF	1	het. // 8 bp OF duplication	MDC1A	P4, this paper
4	1	c.397-?_639+?del	r.397_639del	IF	5	hom. (n=2), het. (n=3) // several	MDC1A	[32]
5	1	c.640-?_819+?del	r.(del)	IF	2	het. (n=2) // 2 bp OF deletion	MDC1A	[32]
10–12	3	c.1307-?_1782+?del	r.(del)	OF	1	het. // 4 bp OF duplication	MDC1A	[32]
13–37	15	c.1783-19594_5445+1681del	r.(del)	IF	1	het. // unknown	myopathy	[31] Db
17	1	c.2322+259_2450+2037del	r.(del)	OF	1	het. // 3 bp IF deletion	MDC1A (ambulant at 17yr)	P1, this paper P19, [28]
23–56	34	c.(?_3175)_(7898_?)del	r.(del)	OF	1	het. // 1 bp OF deletion	MDC1A	[27]
41–48	8	c.5866-?_6867+?del	r.(del)	IF	1	het. // 4 bp OF duplication	MDC1A	[32]
56	1	c.7750-1713_7899-2154del	r.7750_7898del	OF	14	hom. (n=2), het. (n=12) // several	MDC1A; LAMA2-related MD	[28] Db
57–3UTR	9	c.7899-?_(*219_?)del	r.(?)	?	2	hom. (n=2)	MDC1A	Db
59–63	5	c.8245-?_8988+?del	r.(del)	IF	1	het. // nonsense mutation	MDC1A	[32]
63	1	c.8858-?_8988+?del	r.(del)	OF	1	het. // nonsense mutation	MDC1A	[32]
Duplications								
5–12	8	c.640-?_1782+?dup	r.(dup)	IF	1	het. // unknown	muscular dystrophy	[31] Db
28–29	2	c.4059-?_4311+?dup	r.4059_4311dup	OF	1	het. // nonsense mutation	MDC1A	P2, this paper P24, [28]

Mutations described according to HGVS nomenclature using cDNA reference sequence with accession number NM_000426.3. bp- base pairs; Db- *locus*-specific mutation database for *LAMA2* gene (http://www.lovd.nl/LAMA2); het.- heterozygous; hom.- homozygous; IF- in-frame; OF- out-of-frame; MD- muscular dystrophy; Nr.- number; yr- years.

Large deletions and duplications detected in CMD patients are not confined to *LAMA2* gene; in fact we have recently reported a patient with a Fukuyama CMD caused by a multi-exonic duplication in *FKTN* (fukutin) [33]. There are additional reports of other pathogenic CNVs in CMD genes, such as: *ISPD* (isoprenoid synthase domain containing) [34], *LARGE* (like-glycosyltransferase) [35, 36] and *POMGNT1* [protein O-linked mannose N-acetylglucosaminyltransferase 1 (beta 1,2-)] [37].

New mutation screening methods are currently being developed based on next-generation sequencing (NGS) technology, which will contribute to establish the genetic causes of hereditary myopathies that remain unsolved. However, prior to its application, it is important to exclude large deletions and duplications as a cause of these diseases. Bioinformatic pipelines for NGS usually do not incorporate algorithms that enable their automatic detection, but we have previously shown that, when properly applied, this technology can help delineate large genomic rearrangements [38].

In summary, we have reassessed the impact of large deletions and duplications in LAMA2-RD and emphasize the importance of including screening for these rearrangements as part of the diagnostic strategy, especially in patients where a single heterozygous mutation has been detected.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

REFERENCES

[1] Bönnemannm, C. G., Wang, C. H., Quijano-Roy, S., Deconinck, N., Bertini, E., Ferreiro, A., Muntoni, F., Sewry, C., Béroud, C., Mathews, K. D., Moore, S. A., Bellini, J., Rutkowski, A., North, K. N.; Members of International Standard of Care Committee for Congenital Muscular Dystrophies. Diagnostic approach to the congenital muscular dystrophies. Neuromuscul Disord. 2014; 24(4): 289-311.

[2] Quijano-Roy, S., Sparks, S., Rutkowski, A. LAMA2-Related Muscular Dystrophy. In: Pagon RA, Adam MP, Ardinger HH, Bird TD, Dolan CR, Fong CT, Smith RJH, Stephens K, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2014. Available from: http://www.ncbi.nlm.nih.gov/books/NBK97333/

[3] Helbling-Leclerc, A., Zhang, X., Topaloglu, H., Cruaud, C., Tesson, F., Weissenbach, J., Tomé, F. M., Schwartz, K., Fardeau, M., Tryggvason, K., Guicheney, P. Mutations in the laminin alpha 2-chain gene (LAMA2) cause merosin-deficient congenital muscular dystrophy. Nat Genet. 1995; 11(2): 216-218.

[4] Zhang, X., Vuolteenaho, R., Tryggvason, K. Structure of the human laminin alpha2-chain gene (LAMA2), which is affected in congenital muscular dystrophy. J Biol Chem. 1996; 271(44): 27664-27669.

[5] Leivo, I., Engvall, E. Merosin, a protein specific for basement membranes of Schwann cells, striated muscle, and trophoblast, is expressed late in nerve and muscle development. Proc Natl Acad Sci U S A. 1988; 85(5): 1544-1548.

[6] Campbell, K. P. Three muscular dystrophies: Loss of cytoskeleton-extracellular matrix linkage. Cell. 1995; 80(5): 675-679.

[7] Villanova, M., Malandrini, A., Sabatelli, P., Sewry, C. A., Toti, P., Torelli, S., Six, J., Scarfó, G., Palma, L., Muntoni, F., Squarzoni, S., Tosi, P., Maraldi, N. M., Guazzi, G. C. Localization of laminin alpha 2 chain in normal human central nervous system: An immunofluorescence and ultrastructural study. Acta Neuropathol (Berl). 1997; 94(6): 567-571.

[8] Holmberg, J., Durbeej, M. Laminin-211 in skeletal muscle function. Cell Adh Migr. 2013; 7(1): 111-121.

[9] Allamand, V., Guicheney, P. Merosin-deficient congenital muscular dystrophy, autosomal recessive (MDC1A, MIM#156225, LAMA2 gene coding for alpha2 chain of laminin). Eur J Hum Genet. 2002; 10(2): 91-94.

[10] Tomé, F. M., Evangelista, T., Leclerc, A., Sunada, Y., Manole, E., Estournet, B., Barois, A., Campbell, K. P., Fardeau, M. Congenital muscular dystrophy with merosin deficiency. C R Acad Sci III. 1994; 317(4): 351-357.

[11] Philpot, J., Sewry, C., Pennock, J., Dubowitz, V. Clinical phenotype in congenital muscular dystrophy: Correlation with expression of merosin in skeletal muscle. Neuromuscul Disord. 1995; 5(4): 301-305.

[12] Schessl, J., Zou, Y., Bönnemann, C. G. Congenital muscular dystrophies and the extracellular matrix. Semin Pediatr Neurol. 2006; 13(2): 80-89.

[13] Philpot, J., Bagnall, A., King, C., Dubowitz, V., Muntoni, F. Feeding problems in merosin deficient congenital muscular dystrophy. Arch Dis Child. 1999; 80(6): 542-547.

[14] Bönnemann, C. G., Rutkowski, A., Mercuri, E., Muntoni, F. 173rd ENMC International Workshop: Congenital muscular dystrophy outcome measures 5-7 March 2010, Naarden, The Netherlands. Neuromuscul Disord. 2011; 21(7): 513-522.

[15] Deodato, F., Sabatelli, M., Ricci, E., Mercuri, E., Muntoni, F., Sewry, C., Naom, I., Tonali, P., Guzzetta, F. Hypermyelinating neuropathy, mental retardation and epilepsy in a case of merosin deficiency. Neuromuscul Disord. 2002; 12(4): 392-398.

[16] Di Muzio, A., De Angelis, M. V., Di Fulvio, P., Ratti, A., Pizzuti, A., Stuppia, L., Gambi, D., Uncini, A. Dysmyelinating sensory–motor neuropathy in merosin-deficient congenital muscular dystrophy. Muscle Nerve. 2003; 27(4): 500-506.

[17] Geranmayeh, F., Clement, E., Feng, L. H., Sewry, C., Pagan, J., Mein, R., Abbs, S., Brueton, L., Childs, A. M., Jungbluth, H., De Goede, C. G., Lynch, B., Lin, J. P., Chow, G.,

Sousa, C. D., O'Mahony, O., Majumdar, A., Straub, V., Bushby, K., Muntoni, F. Genotype-phenotype correlation in a large population of muscular dystrophy patients with LAMA2 mutations. Neuromuscul Disord. 2010; 20(4): 241-250.

[18] Carboni, N., Marrosu, G., Porcu, M., Mateddu, A., Solla, E., Cocco, E., Maioli, M. A., Oppo, V., Piras, R., Marrosu, M. G. Dilated cardiomyopathy with conduction defects in a patient with partial merosin deficiency due to mutations in the laminin-α2-chain gene: A chance association or a novel phenotype? Muscle Nerve. 2011; 44(5): 826-828.

[19] Marques, J., Duarte, S. T., Costa, S., Jacinto, S., Oliveira, J., Oliveira, M. E., Santos, R., Bronze-da-Rocha, E., Silvestre, A. R., Calado, E., Evangelista, T. Atypical phenotype in two patients with LAMA2 mutations. Neuromuscul Disord. 2014; 24(5): 419-424.

[20] Kevelam, S. H., van Engelen, B. G., van Berkel, C. G., Küsters, B., van der Knaap, M. S. LAMA2 mutations in adult-onset muscular dystrophy with leukoencephalopathy. Muscle Nerve. 2014; 49(4): 616-617.

[21] Jones, K. J., Morgan, G., Johnston, H., Tobias, V., Ouvrier, R. A., Wilkinson, I., North, K. N. The expanding phenotype of laminin alpha2 chain (merosin) abnormalities: Case series and review. J Med Genet. 2001; 38(10): 649-657.

[22] Hayashi, Y. K., Ishihara, T., Domen, K., Hori, H., Arahata, K. A benign allelic form of laminin a2 chain deficient muscular dystrophy. Lancet. 1997; 349(9059): 1147.

[23] Naom, I., D'Alessandro, M., Sewry, C. A., Philpot, J., Manzur, A. Y., Dubowitz, V., Muntoni, F. Laminin a2-chain gene mutations in two siblings presenting with limb-girdle muscular dystrophy. Neuromuscul Disord. 1998; 8(7): 495-501.

[24] Tan, E., Topaloglu, H., Sewry, C., Zorlu, Y., Naom, I., Erdem, S., D'Alessandro, M., Muntoni, F., Dubowitz, V. Late onset muscular dystrophy with cerebral white matter changes due to partial merosin deficiency. Neuromuscul Disord. 1997; 7(2): 85-89.

[25] Di Blasi, C., He, Y., Morandi, L., Cornelio, F., Guicheney, P., Mora, M. Mild muscular dystrophy due to a nonsense mutation in the LAMA2 gene resulting in exon skipping. Brain. 2001; 124(Pt4): 698-704.

[26] Gavassini, B. F., Carboni, N., Nielsen, J. E., Danielsen, E. R., Thomsen, C., Svenstrup, K., Bello, L., Maioli, M. A., Marrosu, G., Ticca, A. F., Mura, M., Marrosu, M. G., Soraru, G., Angelini, C., Vissing, J., Pegoraro, E. Clinical and molecular characterization of limb-girdle muscular dystrophy due to LAMA2 mutations. Muscle Nerve. 2011; 44(5): 703-709.

[27] Pegoraro, E., Marks, H., Garcia, C. A., Crawford, T., Mancias, P., Connolly, A. M., Fanin, M., Martinello, F., Trevisan, C. P., Angelini, C., Stella, A., Scavina, M., Munk, R. L., Servidei, S., Bönnemann, C. C., Bertorini, T., Acsadi, G., Thompson, C. E., Gagnon, D., Hoganson, G., Carver, V., Zimmerman, R. A., Hoffman, E. P. Laminin alpha2 muscular dystrophy: Genotype/phenotype studies of 22 patients. Neurology. 1998; 51(1): 101-110.

[28] Oliveira, J., Santos, R., Soares-Silva, I., Jorge, P., Vieira, E., Oliveira, M. E., Moreira, A., Coelho, T., Ferreira, J. C., Fonseca, M. J., Barbosa, C., Prats, J., Aríztegui, M. L., Martins, M. L., Moreno, T., Heinimann, K., Barbot, C., Pascual-Pascual, S. I., Cabral, A., Fineza, I., Santos, M., Bronze-da-Rocha, E.

LAMA2 gene analysis in a cohort of 26 congenital muscular dystrophy patients. Clin Genet. 2008; 74(6): 502-512.

[29] Miller, S. A., Dykes, D. D., Polesky, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 1988; 16(3): 1215.

[30] den Dunnen, J. T., Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. Hum Mutat. 2000; 15(1): 7-12.

[31] Piluso, G., Dionisi, M., Del Vecchio Blanco, F., Torella, A., Aurino, S., Savarese, M., Giugliano, T., Bertini, E., Terracciano, A., Vainzof, M., Criscuolo, C., Politano, L., Casali, C., Santorelli, F. M., Nigro, V. Motor chip: A comparative genomic hybridization microarray for copy-number mutations in 245 neuromuscular disorders. Clin Chem. 2011; 57(11): 1584-1596.

[32] Xiong, H., Tan, D., Wang, S., Song, S., Yang, H., Gao, K., Liu, A., Jiao, H., Mao, B., Ding, J., Chang, X., Wang, J., Wu, Y., Yuan, Y., Jiang, Y., Zhang, F., Wu, H., Wu, X. Genotype/phenotype analysis in Chinese laminin-α2 deficient congenital muscular dystrophy patients. Clin Genet. 2014; doi: 10.1111/cge.12366 [Epub ahead of print].

[33] Costa, C., Oliveira, J., Gonçalves, A., Santos, R., Bronze-da-Rocha, E., Rebelo, O., Pais, R. P., Fineza, I. A Portuguese case of Fukuyama congenital muscular dystrophy caused by a multi-exonic duplication in the fukutin gene. Neuromuscul Disord. 2013; 23(7): 557-561.

[34] Czeschik, J. C., Hehr, U., Hartmann, B., Lüdecke, H. J., Rosenbaum, T., Schweiger, B., Wieczorek, D. 160 kb deletion in ISPD unmasking a recessive mutation in a patient with Walker-Warburg syndrome. Eur J Med Genet. 2013; 56(12): 689-694.

[35] van Reeuwijk, J., Grewal, P. K., Salih, M. A., Beltrán-Valero de Bernabé, D., McLaughlan, J. M., Michielse, C. B., Herrmann, R., Hewitt, J. E., Steinbrecher, A., Seidahmed, M. Z., Shaheed, M. M., Abomelha, A., Brunner, H. G., van Bokhoven, H., Voit, T. Intragenic deletion in the LARGE gene causes Walker-Warburg syndrome. Hum Genet. 2007; 121(6): 685-690.

[36] Clarke, N. F., Maugenre, S., Vandebrouck, A., Urtizberea, J. A., Willer, T., Peat, R. A., Gray, F., Bouchet, C., Manya, H., Vuillaumier-Barrot, S., Endo, T., Chouery, E., Campbell, K. P., Mégarbané, A., Guicheney, P. Congenital muscular dystrophy type 1D (MDC1D) due to a large intragenic insertion/deletion, involving intron 10 of the LARGE gene. Eur J Hum Genet. 2011; 19(4): 452-457.

[37] Saredi, S., Ardissone, A., Ruggieri, A., Mottarelli, E., Farina, L., Rinaldi, R., Silvestri, E., Gandioli, C., D'Arrigo, S., Salerno, F., Morandi, L., Grammatico, P., Pantaleoni, C., Moroni, I., Mora, M. Novel POMGNT1 point mutations and intragenic rearrangements associated with muscle-eye-brain disease. J Neurol Sci. 2012; 318(1-2): 45-50.

[38] Oliveira, J., Oliveira, M. E., Kress, W., Taipa, R., Pires, M. M., Hilbert, P., Baxter, P., Santos, M., Buermans, H., den Dunnen, J. T., Santos, R. Expanding the MTM1 mutational spectrum: Novel variants including the first multi-exonic duplication and development of a locus-specific database. Eur J Hum Genet. 2013; 21(5): 540-549.

WILEY | HGVS HUMAN GENOME VARIATION SOCIETY

LAMA2 gene mutation update: Toward a more comprehensive picture of the laminin-α2 variome and its related phenotypes

Jorge Oliveira[1,2] (iD) | Angela Gruber[3] | Márcio Cardoso[4] | Ricardo Taipa[5] |
Isabel Fineza[6] | Ana Gonçalves[1,2] | Andreas Laner[7] | Thomas L. Winder[8] |
Jocelyn Schroeder[3] | Julie Rath[3] | Márcia E. Oliveira[1,2] | Emília Vieira[1,2] |
Ana Paula Sousa[4] | José Pedro Vieira[9] | Teresa Lourenço[10] | Luciano Almendra[11] |
Luís Negrão[11] | Manuela Santos[12] | Manuel Melo-Pires[5] | Teresa Coelho[4] |
Johan T. den Dunnen[13] | Rosário Santos[1,2,14]* | Mário Sousa[2,15,16]*

[1]Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto Magalhães, Centro Hospitalar do Porto, Porto, Portugal

[2]Unidade Multidisciplinar de Investigação Biomédica (UMIB), Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal

[3]PreventionGenetics, Marshfield, Wisconsin

[4]Consulta de Doenças Neuromusculares e Serviço de Neurofisiologia, Departamento de Neurociências, Centro Hospitalar do Porto, Porto, Portugal

[5]Unidade de Neuropatologia, Centro Hospitalar do Porto, Porto, Portugal

[6]Unidade de Neuropediatria, Centro de Desenvolvimento da Criança Luís Borges, Hospital Pediátrico de Coimbra, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal

[7]MGZ - Center of Medical Genetics, Munich, Germany

[8]Invitae Corporation, San Francisco, California

[9]Serviço de Neurologia, Hospital de Dona Estefânia, Centro Hospitalar de Lisboa Central, Lisboa, Portugal

[10]Serviço de Genética Médica, Hospital de Dona Estefânia, Centro Hospitalar de Lisboa Central, Lisboa, Portugal

[11]Consulta de Doenças Neuromusculares, Hospitais da Universidade de Coimbra, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal

[12]Consulta de Doenças Neuromusculares e Serviço de Neuropediatria, Centro Hospitalar do Porto, Porto, Portugal

[13]Departments of Human Genetics and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands

[14]UCIBIO/REQUIMTE, Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal

[15]Departamento de Microscopia, Laboratório de Biologia Celular, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal

[16]Centro de Genética da Reprodução Prof. Alberto Barros, Porto, Portugal

Correspondence
Jorge Oliveira, Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto de Magalhães, Centro Hospitalar do Porto, Praça Pedro Nunes, 88, 4099-028 Porto, Portugal.
Email: jorge.oliveira@chporto.min-saude.pt

*Rosário Santos and Mário Sousa contributed equally to this work.

Abstract

Congenital muscular dystrophy type 1A (MDC1A) is one of the main subtypes of early-onset muscle disease, caused by disease-associated variants in the laminin-α2 (*LAMA2*) gene. MDC1A usually presents as a severe neonatal hypotonia and failure to thrive. Muscle weakness compromises normal motor development, leading to the inability to sit unsupported or to walk independently. The phenotype associated with *LAMA2* defects has been expanded to include milder and atypical cases, being now collectively known as *LAMA2*-related muscular dystrophies (*LAMA2*-MD). Through an international multicenter collaborative effort, 61 new *LAMA2* disease-associated variants were identified in 86 patients, representing the largest number of patients and new disease-causing variants in a single report. The collaborative variant collection was supported by the LOVD-powered *LAMA2* gene variant database (https://www.LOVD.nl/LAMA2), updated as part of this work. As of December 2017, the database contains 486 unique *LAMA2* variants (309 disease-associated), obtained from direct submissions and literature reports. Database content was systematically reviewed and further insights concerning *LAMA2*-MD are presented. We focus on the impact of missense changes, especially the c.2461A > C (p.Thr821Pro) variant and its association with late-onset *LAMA2*-MD. Finally, we report diagnostically challenging cases, highlighting the

1 | BACKGROUND

Laminin-211 is a heterotrimetric cruciform-shaped complex that establishes a stable link between the sarcolemma of muscle fibers and the extracellular matrix, being a major component of the extrasynaptic skeletal muscle basement membrane (BM; Durbeej, 2015). The -211 classification derives from the three specific chains (α2, β1, and γ1), which compose this specific laminin form (Aumailley et al., 2005). Laminin-211 binds to the glycosylated residues of α-dystroglycan (α-DG) and also self-assembles (polymerizes) into networks through its N-terminal domain (Yurchenco, 2015). This supramolecular network connects to collagen IV and to perlecan (heparan sulfate proteoglycan) through nidogens cross-linking (Jones, Dehart, Gonzales, & Goldfinger, 2000). Laminin-211 expression is not confined to skeletal muscle but has also been shown to be expressed in a variety of other tissues, more importantly in peripheral nerve (Schwann cells) and in brain (Yurchenco, 2015). Posttranslational changes have been reported in laminin-211 components. More specifically, laminin-α2 chain was found to undergo cleavage at residue 2580 under specific conditions to generate an N-terminal 300 kDa peptide and a C-terminal 80 kDa peptide, which are subsequently connected through a noncovalent process (Durbeej, 2015).

Disease-associated (pathogenic) variants located in the gene that codes for the α2 chain (*LAMA2*; MIM# 156225) of laminin-211, give rise to a group of diseases collectively designated as *LAMA2*-related muscular dystrophy (*LAMA2*-MD). *LAMA2* maps to chromosome 6q22.33 and spans over 260 kb. It comprises 65 exons and codes for a protein with a molecular mass of approximately 390 kDa (Zhang, Vuolteenaho, & Tryggvason, 1996). The majority of patients with *LAMA2* mutations have a congenital muscular dystrophy (CMD) phenotype classified as type 1A (MDC1A; MIM# 607855). The classical phenotype manifests as neonatal hypotonia or muscle weakness during the first months of life and reduced spontaneous movements (Helbling-Leclerc et al., 1995). As muscle weakness persists during development, it compromises the achievement of normal motor milestones (no cephalic control or inability to sit unsupported) and frequently gives rise to failure to thrive. Other manifestations such as gastroesophageal reflux, aspiration, recurrent chest infections, and even respiratory failure were reported in MDC1A (Jones et al., 2001). Facial muscle weakness, ophthalmoparesis, and macroglossia are also features present in these patients but are often beyond early childhood (Quijano-Roy, Sparks, & Rutkowski, 2012). Other relevant clinical hallmarks of MDC1A include elevated creatine kinase (CK) levels and dystrophic changes (necrosis and regeneration of fibers, chronic inflammation, and fibrosis) recognizable in muscle biopsies of these patients (Tomé et al., 1994). Diagnostically important features are the complete absence of laminin-α2 staining evaluated by

immunohistochemistry (IHC) performed in muscle or in skin biopsies (Sewry et al., 1996) using specific antibodies, and typical white matter changes (WMC) in brain detectable by magnetic resonance imaging (MRI; Lamer et al., 1998). WMC are related with alterations in the brain's water content, due to modifications in the maturation and/or function of the blood–brain barrier, and are detectable after the first six months to one year of life (Menezes et al., 2014). Besides WMC, brain structural defects have been reported in patients with laminin deficiency, in an estimated ∼4% of *LAMA2*-MD cases (Jones et al., 2001). In some initial studies, performed before *LAMA2* genotyping was available, this association was based solely on laminin staining by IHC (Brett et al., 1998; Martinello, Angelini, & Trevisan, 1998; Philpot et al., 1999; Pini, Merlini, Tomé, Chevallay, & Gobbi, 1996; Sunada, Edgar, Lotz, Rust, & Campbell, 1995; Tsao, Mendell, Rusin, & Luquette, 1998). It is plausible that any dystroglycanopathy could account for the partial laminin deficiency observed in some patients, explaining the diversity of structural brain defects reported. It is nonetheless consensual that primary laminin-α2 deficiency can contribute to structural abnormalities in the cerebral cortex during fetal development. Malformations found in patients with *LAMA2* disease-causing variants includes: (a) cortical dysplasia (Mercuri et al., 1999), (b) changes within the lissencephaly spectrum, namely agyria or pachygyria (Geranmayeh et al., 2010), and (c) polymicrogyria (Vigliano, Dassi, Di Blasi, Mora, & Jarre, 2009).

In a subset of MDC1A cases there is partial laminin-α2 deficiency (reduced/irregular laminin-α2 staining in IHC), which translates into a CMD with a slower disease progression (Geranmayeh et al., 2010; Oliveira et al., 2008). There is some degree of correlation between independent ambulation and IHC status of laminin-α2. The majority of MDC1A patients that do not acquire independent locomotion have complete laminin-α2 deficiency on muscle biopsy, whereas in the majority of cases that are able to walk independently a partial laminin-α2 deficiency has been documented (Geranmayeh et al., 2010).

Further to MDC1A, "milder" *LAMA2*-related phenotypes have been increasingly reported over the past few years. These late-onset *LAMA2*-MD patients are mainly characterized by proximal muscle weakness with onset during childhood, delayed motor milestones, achievement of independent ambulation, and persistently elevated CK levels (Gavassini et al., 2011). Some reports classified these patients as a subtype of limb-girdle muscular dystrophy (LGMD). Patients included in this group may also show muscle hypertrophy, rigid spine syndrome, and pronounced joint contractures which are often more evident in the elbows. In addition to cardiac involvement in a limited number of cases, these clinical features are evocative of Emery–Dreifuss muscular dystrophy (EDMD; Nelson et al., 2015). It should be emphasized that patients with late-onset *LAMA2*-MD still manifest typical brain

WMC, but IHC labeling of laminin-α2 in muscle biopsy may show only very subtle changes.

As laminin-α2 is also expressed in Schwann cells, there is a range of clinical features related with peripheral nerve involvement in *LAMA2*-MD patients. In a particular series of MDC1A patients, the majority had decreased motor nerve conduction, suggesting that peripheral demyelinating neuropathy is a disease feature (Shorer, Philpot, Muntoni, Sewry, & Dubowitz, 1995). Later it was also shown that laminin-α2 related neuropathic abnormalities also included sensory nerves (Quijano-Roy et al., 2004). More importantly, in a milder case of *LAMA2*-MD there was evidence of a myelinogenesis disorder, leading to the assumption that the neuropathy in laminin-α2 deficient cases is actually dysmyelinating (Di Muzio et al., 2003). These changes are more evident in milder *LAMA2*-MD patients (Chan et al., 2014; Deodato et al., 2002; Mora et al., 1996), whereas as in MDC1A presentations the more severe muscle involvement probably masks the subtle neuropathic features of the disease.

In terms of the mutation spectrum of the *LAMA2* gene, four independent studies described cohorts with more than twenty patients (Geranmayeh et al., 2010; Oliveira et al., 2008; Pegoraro et al., 1998; Xiong et al., 2015). The most frequent reported genotypes include variants that create premature termination codons (PTC) in both disease alleles, and are associated with complete deficiency of laminin-α2 in muscle biopsy as well as an MDC1A phenotype. In contrast, missense variants are present in a smaller number of cases and usually correlate with partial laminin-α2 deficiency giving rise to milder phenotypes. The asymmetrical proportion between truncating and non-truncating variants, explains the higher prevalence of MDC1A as compared with other emerging *LAMA2*-related phenotypes.

A relatively high frequency (18.4% of disease-causing variants) of large deletions and duplications in *LAMA2* was also reported. Variants of this sort are detectable by multiplex ligation-dependent probe amplification (MLPA) or array comparative genomic hybridization (array-CGH; Oliveira et al., 2014).

The *LAMA2* locus-specific database (LSDB), which we initiated in 2002, was continuously updated and used to assist the collection of new variants as reported here. Of the 486 unique variants registered to date (December 2017), a total of 61 novel disease-associated variants detected in 86 patients are reported for the first time. Database content is systematically presented and further insights into the genotypes and phenotypes of *LAMA2*-MD are presented.

2 | DEVELOPMENT AND UPDATE OF *LAMA2* LSDB

As part of the work we report the development of a comprehensive database for *LAMA2* variants, an important resource made available for the scientific community since 2002. The LOVD software (Fokkema et al., 2011) was used to store genetic and clinical data, allowing an off-the-shelf LSDB deployment in accordance with international guidelines for the curation and creation of these databases (Celli, Dalgleish, Vihinen, Taschner, & den Dunnen, 2012; Vihinen, den Dunnen, Dalgleish, & Cotton, 2012). The LSDB content was updated and migrated to LOVD version 3.0, being completely redesigned in terms of its database architecture.

Variant data was collected from publications accessed by the curators (64%) or through direct database submissions (36%). Currently (by December 2017), the *LAMA2*-LOVD contains a total of 1,186 of entries (486 unique) identified in a total of 748 individuals. Based on disease impact, these entries comprise: 816 disease-associated variants (309 unique), 317 benign (141 unique), and 53 variants of unknown clinical significance (VUS, 38 unique).

3 | DESCRIPTION OF NOVEL *LAMA2* VARIANTS

A total of 61 novel disease-associated or likely associated variants were identified in the *LAMA2* gene (Table 1), representing more than 20% (61/309) of the total disease variants currently listed in the *LAMA2* LSDB. Variant interpretation followed the standards and guidelines for the classification of sequence variants, proposed by the American College of Medical Genetics and Genomics (ACMG; Richards et al., 2015). The LOVD *LAMA2* database gives two classifications, a Functional classification (column Effect) and a Clinical classification (column ClassClinical). The functional classification indicates the consequences of the variant for the function of the gene/protein (e.g., affects function), the clinical classification the consequences for the individual carrying the variant (e.g., ACMG:5, disease-associated, autosomal recessive [pathogenic]). The summary conclusion of the curators for specific variants, based on all individual observations of the variants, is given in a SUMMARY record. All unpublished variants collected and/or classified in the course of this project can be retrieved from the database using the following link: https://databases.lovd.nl/shared/references/DOI:10.1002/humu.23599.

Variants were identified by different international groups (material and methods in Supporting Information I), which reflects by the diversity of the patients' geographical origins (11 distinct nationalities). Most variants are predicted to be truncating, 20 nonsense type and 23 small frame-shift variants (16 deletions and seven duplications). In addition, this list includes a significant number of variants affecting canonical splice-sites ($n = 13$), the majority located in donor sites (+1 and +2 positions). Due to the inability to obtain proper biological samples or study limitations it was mostly not possible to evaluate their impact at the mRNA level. Thus, the impact of these splice-site variants was evaluated with bioinformatic tools (see Section 3.1), which for all variants indicated unequivocal deleterious effects. One fully characterized was c.819+2T > C, located in the donor splice-site of intron 5. Analysis by RT-PCR followed by sequencing, showed the presence of aberrant transcripts (details in Section 6 and Supporting Information II Figures S1 and S2). In addition to the most prevalent type of variants already stated, the remainder include: (a) two missense variants (one of which might also have an effect on splicing), (b) one in-frame (IF) codon deletion, (c) one deletion-insertion variant, and (d) a large deletion encompassing exons 57 to 65. This large deletion was detected in a homozygous patient with an MDC1A phenotype by array-CGH technique (Supporting Information Figure S3). The 61 new variants were

TABLE 1 Novel pathogenic variants identified in *LAMA2* gene listed in the locus-specific database

Exon/Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Number of entries in LSDB	Patient-ID in LSBD	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/fibroblasts	Zygosity/2nd variant/orientation (cis, trans, or unknown)	Interpretation of the second variant
1	c.47del	Pathogenic	g.129204437delG	r.(?)	p.(Gly16Alafs*29)	–	1	102376	M	United States	MDC1A	–	Het./c.2T > C / trans	Pathogenic
1	c.94C > T	Likely pathogenic	g.129204484C > T	r.(?)	p.(Gln32*)	–	1	102361	F	Canada	MDC1A	Deficiency	Het./c.8245-2A > G / unknown	Likely pathogenic
2	c.164del	Likely pathogenic	g.129371114delA	r.(?)	p.(Asn55Metfs*16)	–	1	102378	M	Canada	CMD	Deficiency	Het./ ? / unknown	No second pathogenic variant found
2i	c.283+2del	Likely pathogenic	g.129371235delT	r.(spl?)	p.(?)	–	1	102463	F	United States	MDC1A	Deficiency	Het./ c.1609-41_1609-7inv / unknown	VUS [1]
3i	c.396+1G > T	Pathogenic	g.129381042G > T	r.(spl?)	p.(?)	ClinVar (RCV000316746.1); dbSNP (rs777061 7208); gnomAD (0.0024%)	6	102366	F	United States	MDC1A	–	Het./ c.498G > A / unknown	Pathogenic
								102732	F	Mexico	MDC1A	–	Het./ c.5116C > T / unknown	Pathogenic
								102386	M	United States	MDC1A	–	Hom/ n.a./ n.a.	n.a.
								131976	M	United States	MDC1A	–	Het./ c.6501C > A / unknown	Likely pathogenic
								102478	F	United States	MDC1A	–	Het./ c.6690C > A / unknown	Pathogenic
								36041	M	Lybia	Unknown	–	Het./ c.8586T > G / trans	Pathogenic
4i	c.639+2T > A	Likely pathogenic	g.129419562T > A	r.spl?	p.(?)	–	1	102476	F	United States	MDC1A	–	Het./ c.2049_2050 del / trans	Pathogenic
5i	c.819+2T > C	Pathogenic	g.129465227T > C	r.[640_819 del;640 _1027del]	p.[Ile214_Arg273del; Ile214 Hisfs*22]	–	3	102735 [2]	M	Portugal	Late-onset LAMA2-related MD	Partial deficiency	Het./ c.3976C > T / unknown	Pathogenic

(Continues)

TABLE 1 (Continued)

Exon/ Intron	DNA variant (NM_ 000426.3) [a]	Interpretation [a]	RNA variant	Predicted effect on protein	External variant databases	Number of entries in LSDB	Patient-ID in LSDB	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/ fibroblasts	Zygosity/ 2nd variant/ orientation (cis, trans, or unknown)	Interpretation of the second variant
							102736 [2]	F	Portugal	Late-onset LAMA2-related MD	–	Het./ c.3976C > T / unknown	Pathogenic
							103207 [3]	F	Portugal	Late-onset LAMA2-related MD	–	Het./ c.1854_1861 dup / trans	Pathogenic
7	c.939_940del	Pathogenic	r.(?)	p.(Cys314 Trpfs*3)	gnomAD (0.0028%)	7	102373	F	United States	MDC1A	–	Het./ c.7732C > T / unknown	Pathogenic
							102382 [2]	F	United States	MDC1A	Partial deficiency	Het./ c.5562+5G> C / unknown	Pathogenic
							132007 [2]	M	United States	MDC1A	–	Het./ c.5562+5G> C / unknown	Pathogenic
							102396	F	United States	MDC1A	–	Hom./ n.a. / n.a.	n.a.
							132008	F	United States	MDC1A	–	Het./ c.7658delC / unknown	Pathogenic
							132009	M	United States	MDC1A	–	Hom./ n.a. / n.a.	n.a.
							102655	M	United States	MDC1A	Deficiency	Het./ c.7732C > T / unknown	Pathogenic
7	c.991A > T	Likely pathogenic	r.(?)	p.(Arg331*)	gnomAD (0.00041%)	1	102467	M	United States	MDC1A	–	Het./ c.5325dupA / unknown	Pathogenic
12	c.1762del	Pathogenic	r.(?)	p.(Ala588 Leufs*11)	Clinvar (RCV000 171527.1); dbSNP (rs78620 5654)	7	102328	F	Saudi Arabia	MDC1A	–	Hom./ n.a. / n.a.	n.a.
							102349	F	Unknown	MDC1A	Deficiency	Hom./ n.a. / n.a.	n.a.
							132010	M	Saudi Arabia	MDC1A	–	Hom./ n.a. / n.a.	n.a.
							132011	M	Saudi Arabia	MDC1A	–	Het./ c.1303C > T/ trans	Pathogenic

(Continues)

TABLE 1 (Continued)

Exon/Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Number of entries in LSDB	Patient-ID in LSBD	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/fibroblasts	Zygosity/2nd variant/orientation (cis, trans, or unknown)	Interpretation of the second variant
								132012 [4]	M	Saudi Arabia	MDC1A	–	Hom./n.a./n.a.	n.a.
								132013 [4]	M	Saudi Arabia	MDC1A	–	Hom./n.a./n.a.	n.a.
								102363	F	Saudi Arabia	MDC1A	–	Hom./n.a./n.a.	n.a.
13	c.1823_1824del	Likely pathogenic	g.129571297_129571298del	r.(?)	p.(Tyr608*)	dbSNP (rs754600708); gnomAD (0.00041%)	1	102471	F	United States	MDC1A	Deficiency	Hom./n.a./n.a.	n.a.
14	c.2017G>T	Likely pathogenic	g.129573361G>T	r.(?)	p.(Glu673*)	–	1	102460	F	United States	MDC1A	Deficiency	Het./c.2023_2024del/unknown	Likely pathogenic
14	c.2023_2024del	Likely pathogenic	g.129573367_129573368del	r.(?)	p.(Met675Aspfs*29)	gnomAD (0.00041%)	1						Het./c.2017G>T/unknown	Likely pathogenic
17	c.2350dup	Pathogenic	g.129591796dup	r.(?)	p.(Tyr784Leufs*3)	ClinVar (RCV000486406.1)	1	103206	F	Spain	MDC1A	–	Het./c.4692_4695dup/unknown	Pathogenic
17	c.2383G>T	Likely pathogenic	g.129591829G>T	r.(?)	p.(Glu795*)	dbSNP (rs149896793); ESP (0.01%); gnomAD (0.00041%)	1	102397	F	United States	MDC1A	–	Het./c.4761dupT/unknown	Likely pathogenic
17i	c.2450+4A>G	Likely pathogenic	g.129591900A>G	r.(spl?)	p.(?)	–	2	103191	F	Portugal	MDC1A	Partial deficiency	Het./c.8244+1G>A/unknown	Pathogenic
								103972	M	Portugal	Late-onset LAMA2-related MD	–	Het./c.7750-1713_7899-2154del/unknown	Pathogenic
18i	c.2538-1G>A	Likely pathogenic	g.129608991G>A	r.spl?	p.(?)	–	2	102547	F	United States	MDC1A	Deficiency	Het./c.3735+2T>A/unknown	Likely pathogenic
								132014	F	United States	MD	–	Het./?/unknown	No second pathogenic variant found
21	c.2875C>T	Pathogenic	g.129618848C>T	r.(?)	p.(Gln959*)	–	1	102661	F	Saudi Arabia	MDC1A	–	Hom./n.a./n.a.	n.a.

(Continues)

TABLE 1 (Continued)

Exon/Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Number of entries in LSDB	Patient-ID in LSDB	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/fibroblasts	Zygosity/2nd variant/orientation (cis, trans, or unknown)	Interpretation of the second variant
23	c.3338_3345dup	Likely pathogenic	g.129634169_129634176dup	r.(?)	p.(Thr1116 Glnfs*26)	–	1	102385	M	United States	MDC1A	–	Het./ c.6207C >A / unknown	Likely pathogenic
23	c.3372dup	Likely pathogenic	g.129634203dup	r.(?)	p.(Cys1125 Metfs*4)		1	103970	U [6]	Portugal	LGMD/ EDMD [6]	–	Het./ c.2461A > C / unknown	Pathogenic
24	c.3472A > T	Likely pathogenic	g.129635860A > T	r.(?)	p.(Lys1158*)	–	1	102324	F	United States	Unknown	–	Het./ ? / unknown	No second pathogenic variant found
24	c.3520C > T	Pathogenic	g.129635908C > T	r.(?)	p.(Gln1174*)	–	1	103127	F	Spain	MDC1A	–	Het./ c.3976C > T / trans	Pathogenic
25	c.3560_3569del	Pathogenic	g.129636625_129636634del	r.(?)	p.(Thr1187 Metfs*9)	–	1	102486	F	Canada	MDC1A	Deficiency	Hom. / n.a. / n.a.	n.a.
25i	c.3735 +2T > A	Likely pathogenic	g.129636802T > A	r.spl?	p.(?)		1	102547	F	United States	MDC1A	Deficiency	Het./ c.2538-1G > A / unknown	Likely pathogenic
26	c.3829C > T	Likely pathogenic	g.129637000C > T	r.(?)	p.(Arg1277*)	–	1	102383	M	Canada	MDC1A	Deficiency	Het./ c.4654G > A / trans	VUS [1]
27	c.4002T > G	Pathogenic	g.129637260T > G	r.(?)	p.(Tyr1334*)	–	1	102535	F	United States	MDC1A	–	Het./ c.7658delC / unknown	Pathogenic
27	c.4049del	Pathogenic	g.129637307delG	r.(?)	p.(Arg1350 Hisfs*12)	–	2	102663	M	United States	MDC1A	Deficiency	c.2049_2050 del / trans	Pathogenic
								102401	F	United States	MDC1A	–	Het./ c.5562+5G > C / trans	Pathogenic
29	c.4261C > T	Pathogenic	g.129649507C > T	r.(?)	p.(Gln1421*)	–	1	102662 [4]	M	United States	MDC1A	–	Het./ c.283+1G > A / trans	Pathogenic
30	c.4348C > T	Pathogenic	g.129663524C > T	r.(?)	p.(Arg1450*)	ClinVar (RCV000171401.1); dbSNP (rs200923373); gnomAD (0.0012%)	1	102364	M	United States	MDC1A	Deficiency	Het./ c.2049_2050 del / trans	Pathogenic
33	c.4761dupT	Likely pathogenic	g.129687407dupT	r.(?)	p.(Arg1588 Serfs*20)	–	1	102397	F	United States	MDC1A	–	Het./ c.2383G > T / unknown	Likely Pathogenic

(Continues)

TABLE 1 (Continued)

Exon/Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Number of entries in LSDB	Patient-ID in LSDB	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/fibroblasts	Zygosity/2nd variant/orientation (cis, trans, or unknown)	Interpretation of the second variant
34	c.4941del	Likely pathogenic	g.129691117delG	r.(?)	p.(Met1647 Ilefs*5)	–	1	102353	M	United States	Father of affected child (carrier study)	–	Het./n.a./n.a.	n.a.
35	c.5050G>T	Pathogenic	g.129704357G>T	r.(?)	p.(Glu1684*)	ClinVar (RCV0000 78775.3; RCV000 177827.2): dbSNP (rs2016 32009)	1	131883	M	Italy	MDC1A	Deficiency	Het./ c.2901C>A / trans	Pathogenic
35i	c.5072-1454_5154 delinsAGA TTGCC	Likely pathogenic	g.129711182_ 129712718 delins AGATTGCC	r.spl?	p.(?)	–	1	102381	M	United States	MDC1A	–	Hom./n.a./n.a.	n.a.
36	c.5132del	Pathogenic	g.129712696delA	r.(?)	p.(Glu1711 Glyfs*14)	–	1	102658	M	United States	MDC1A	–	Het./ c.363C>A/ trans	Pathogenic
36	c.5134_5153del	Pathogenic	g.129712698_ 129712717del	r.(?)	p.(Arg1712 Glufs*4)	–	1	111376	M	Turkey	MDC1A	Deficiency	Hom./n.a./n.a.	n.a.
36	c.5182del	Likely pathogenic	g.129712746delC	r.(?)	p.(Leu1728*)	–	1	102358	F	United States	MDC1A	Deficiency	Hom./n.a./n.a.	n.a.
37	c.5259del	Pathogenic	g.129714214delA	r.(?)	p.(Val1754*)	–	1	102469	F	Israel	MDC1A	Deficiency	Het./ c.7147C>T / trans	Pathogenic
37	c.5263A>T	Pathogenic	g.129714218A>T	r.(?)	p.(Lys1755*)	–	1	103189	M	Iran	MDC1A	Deficiency	Het./ c.6501C>G/ unknown	Pathogenic
41	c.5914C>T	Pathogenic	g.129748945C>T	r.(?)	p.(Gln1972*)	ClinVar (RCV00 0078782.3, RCV0001 78452.1); dbSNP (rs39 8123378)	4	102365	M	United States	Unknown	–	Hom./n.a./n.a.	n.a.
								102384	M	United States	MDC1A	Deficiency	Hom./n.a./n.a.	n.a.
								132015 [4]	F	United States	Unknown	–	Het./?/ unknown	No second pathogenic variant found

(Continues)

TABLE 1 (Continued)

Exon/ Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Number of entries in LSDB	Patient-ID in LSDB	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/ fibroblasts	Zygosity/ 2nd variant/ orientation (cis, trans, or unknown)	Interpretation of the second variant
								102461	F	United States	MDC1A	–	Hom./ n.a. / n.a.	n.a.
42	c.5998del	Likely pathogenic	g.129759820delA	r.(?)	p.(Thr2000 Profs*3)	–	1	102362	M	United States	MDC1A	Deficiency	Het./ c.7147C>T / unknown	Pathogenic
43	c.6207C>A	Likely pathogenic	g.129762082C>A	r.(?)	p.(Tyr2069*)	dbSNP (rs1433 43647); ESP (0.02%)	1	102385	M	United States	MDC1A	–	Het./ c.3338 _3345dup / unknown	Likely pathogenic
43	c.6266del	Pathogenic	g.129762141delA	r.(?)	p.(Asn2089 Thrfs*14)	–	1	102371	F	Mexico	MDC1A	Deficiency	Het./ c.2962C>T / trans	Pathogenic
45i	c.6429+ 1G>T	Likely pathogenic	g.129766967G>T	r.spl?	p.(?)	gnomAD (0.0032%)	2	102400	M	United States	MDC1A	Deficiency	Het./ c.2901C>A / unknown	Pathogenic
								132016	M	United States	Unknown	Partial deficiency	Hom./ n.a./ n.a.	n.a.
46	c.6501C>G	Pathogenic	g.129774204C>G	r.(?)	p.(Tyr2167*)	–	1	103189	M	Iran	MDC1A	Deficiency	Het./ c.5263A>T / unknown	Pathogenic
47	c.6588dup	Likely pathogenic	g.129775314dupT	r.(?)	p.(Ile2197 Tyrfs*5)	gnomAD (0.00041%)	1	102379	F	United States	MDC1A	Partial deficiency	Het./ c.7571A>T / unknown	VUS [1]
49	c.6979G>T	Pathogenic	g.129781456G>T	r.(?)	p.(Gly2327*)	–	1	103192	M	Iran	MDC1A	–	Hom./ n.a./ n.a.	n.a.
51	c.7297C>T	Likely pathogenic	g.129786431C>T	r.(?)	p.(Gln2433*)	–	1	102334	F	United States	MDC1A	Deficiency	Het./ c.35T>G / unknown	Pathogenic
54	c.7491del	Likely pathogenic	g.129799877delA	r.(?)	p.(Asp2498 Ilefs*49)	–	1	102480	M	United States	MDC1A	Deficiency	Het./ c.8244 +3_8244+6 del/ unknown	Likely pathogenic
54i	c.7572+ 1G>A	Pathogenic	g.129799959G>A	r.spl?	p.(?)	–	2	111379	F	Germany	Late-onset LAMA2-related MD	–	Het./ c.245A>T / unknown	VUS [1]
								102548	F	United States	MDC1A	Deficiency	Het./ c.2T>C / trans	Pathogenic
56i	c.7898+ 2T>G	Likely pathogenic	g.129807769T>G	r.spl?	p.(?)	–	1	111374	M	Turkey	MDC1A	Deficiency	Hom./ n.a./ n.a.	n.a.
56i_ 65_	c.7898+732 _*39282del	Likely pathogenic	g.129808499_ 129876774del	r.(?)	p.(?)	–	1	102475 [5]	M	Saudi Arabia	MDC1A	–	Hom./ n.a./ n.a.	n.a.

(Continues)

TABLE 1　(Continued)

Exon/ Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Number of entries in LSDB	Patient-ID in LSBD	Gender	Geographic origin	Phenotype	IHC for laminin-α2 in muscle/ fibroblasts	Zygosity/ 2nd variant/ orientation (cis, trans, or unknown)	Interpretation of the second variant
58i	c.8244+1G>C	Likely pathogenic	g.129813629G>C	r.spl?	p.(?)	–	1	102380	M	United States	MDC1A	–	Hom./n.a./n.a.	n.a.
58i	c.8244+2dup	Likely pathogenic	g.129813630dup	r.spl?	p.(?)	–	1	111380	F	Saudi Arabia	MDC1A	–	Hom./n.a./n.a.	n.a.
59i	c.8357+1G>A	Pathogenic	g.129823917G>A	r.spl?	p.(?)	–	1	102391	M	United States	MDC1A	Deficiency	Het./c.2049_2050del/trans	Pathogenic
61	c.8556_8558del	Likely pathogenic	g.129826353_129826355del	r.(?)	p.(Ile2852del)	ClinVar (RCV00078805.4); dbSNP rs398123389	1	102737	M	Portugal	MDC1A	Partial deficiency	Het./c.5234+1G>A/unknown	Pathogenic
61	c.8586T>G	Pathogenic	g.129826383T>G	r.(?)	p.(Tyr2862*)	–	1	36041	M	Lybia	Unknown	–	Het./c.396+1G>A/trans	Likely pathogenic
61	c.8669dup	Pathogenic	g.129826466dupT	r.(?)	p.(Leu2890Phefs*16)	–	3	102395	M	United States	MDC1A	Deficiency	Het./c.2370T>A/trans	VUS[1]
								102472	M	United States	MDC1A	Deficiency	Het./c.2049_2050del/trans	Pathogenic
								102369	M	United States	MDC1A	Partial deficiency	Het./?/unknown	No second pathogenic variant found
63	c.8947C>T	Pathogenic	g.129833597C>T	r.(?)	p.(Gln2983*)	–	1	111373	M	Turkey	MDC1A	Deficiency	Het./c.6955C>T/trans	Pathogenic
64	c.9095dup	Likely pathogenic	g.129835624dupA	r.(?)	p.(Ile3033Aspfs*6)	–	2	102474	M	United States	MDC1A	Partial deficiency	Het./c.5562+5G>C/unknown	Pathogenic
								132025 [4]	M	United States	Unknown	–	Het./c.4860G>A/unknown	VUS

Notes. CMD: congenital muscular dystrophy; F: female; Het.: heterozygous; Hom..: homozygous; ID#: identification number; IHC: immunohistochemistry; M: male; MD: muscular dystrophy; MDC1A: congenital muscular dystrophy type 1A; n.a.: not applicable; VUS: variant of unknown significance. [1]: Variant listed in Table 2; [2]: Siblings. Variants were detected by Sanger sequencing, except for: [3]: Whole-exome sequencing (patient ID# 103207), [4]: NGS gene panel (patients' ID#s: 102662, 132012, 132013, 132015, 132025), and [5]: Array-CGH (patient ID# 102475), more details are available in Supporting Information I; [6]: Variant identified through an anonymized screening performed in genetically uncharacterized LGMD/EDMD patients; [a]: According to the ACMG guidelines. References sequences used to describe variants: M_000426.3 and NC_000006.11.

identified in 87 patients (85 families) and in one obligate carrier. In terms of genotypes, a total of 25 patients had homozygous variants. In the remaining 57 cases compound heterozygous variants were found: 52 classified as pathogenic and five VUS. From this cohort, only five cases (5.7%) had incomplete genotyping as only one mutated allele was identified. Four variants were detected in more than three unrelated patients: c.939_940del (n = 7), c.1762del (n = 7), c.396+1G > T (n = 6), and c.5914C > T (n = 4). This is explainable by study inclusion criteria and the higher frequency of variants identified in patients of specific populations or ethnic groups still underrepresented in the literature (c.1762del in patients from Saudi Arabia, and the others were found in patients with Hispanic ancestry). Finally, concerning the clinical presentation, the majority of patients were classified as MDC1A, whereas only seven had onset beyond infancy, and achieved independent locomotion. This particular phenotype ("late-onset" LAMA2-MD) was seen in patients with splicing variants or missense substitutions, presumably non-truncating alleles, and with partial LAMA2 deficiency documented in some of the cases.

3.1 | Bioinformatic analysis of novel LAMA2 variants

The novel LAMA2 variants, especially those of the missense type and/or predicted to affect splicing, were further assessed resorting to bioinformatic prediction tools. A total of 14 variants fitting into these categories are shown in Table 2. With the exception of homozygotes, all variants are heterozygous and found in combination with a second change known to be disease-associated or likely disease-associated. Since experimental evidence could not be obtained, bioinformatic analysis was pivotal to attempt their classification in terms of pathogenicity.

Listed in Supporting Information III are the in silico tools used to evaluate variants, more specifically tolerance predictors and splicing predictors. Considering the extensive list of tools available to evaluate missense variants, we sought to determine which could be most efficient in the case of LAMA2 variants. The performance measures for binary classifiers, as described by Niroula and Vihinen (2016), were calculated for nine tolerance predictors algorithms. Two control sets of LAMA2 variants consistently classified in LOVD database as either pathogenic (n = 29) or benign (n = 22) were used to perform these calculations (Supporting Table). The tools with best performance for our purpose, based on the accuracy and Matthews correlation coefficient, were MutPred2 (Pejavar, Mooney, & Radivojac, 2017), PolyPhen-2 (HumVar; Adzhubei et al., 2010), SIFT (Kumar, Henikoff, & Ng, 2009), and UMD-Predictor (Salgado et al., 2016; Supporting Information Table S1). These algorithms were subsequently applied to evaluate the new missense variants reported in this work (Table 2).

The majority of variants listed in Table 2 were inferred as being of the missense type (n = 8). Five variants were consistently classified as deleterious by all four tolerance predictors used, and two variants were classified as deleterious by three out of the four algorithms. The other missense variant was considered deleterious by half of the tolerance predictors.

In four variants, also inferred to be of the missense type, a dual effect was predicted as they could also influence the splicing mecha-

nism. In these, the majority of algorithms tested to evaluate missense variants consistently pointed toward intolerance (all except MutPred2 in three and SIFT in one of the variants) and also indicated an effect on splicing by all tools used, except GeneSplicer in two of the variants.

From the two variants remaining in Table 2, one is an apparently synonymous substitution predicted to create a new acceptor splice-site by three distinct algorithms (all except GeneSplicer) and the other is a large intronic inversion that was predicted to disrupt the canonical acceptor splice-site.

4 | BIOLOGICAL RELEVANCE: CONTENT ANALYSIS OF THE LAMA2 LSDB

An overview of disease-associated variants found in the LAMA2 gene is shown in Figures 1 and 2. These may be subdivided as: 59.6% single nucleotide variants (SNV) (n = 184 unique; 496 in total), 24.9% small deletions (n = 77; 214), 8.7% small insertions (n = 27; 62), 6.2% large deletions or duplications (n = 19; 42), and two deletion/insertions (0.6%). In terms of their foreseeable impact, the most frequent group is that of variants that cause a PTC. These include nonsense (n = 79) and out-of-frame changes (65 deletions, 23 duplications). A total of 79 variants were predicted or experimentally demonstrated to affect splicing. The first and last two conserved nucleotides of introns concentrate the vast majority of splicing variants. It should be highlighted that both PTC-inducing and splice-site variants are widespread throughout the gene with no clear "mutational" hotspots. In terms of distribution throughout the gene, missense variants (n = 40, 13% of total disease-associated variants) do not follow a similar pattern; they seem to cluster in specific regions of laminin-α2 (Figure 1). The first group (n = 11, 27.5% of missense variants) affect residues located in domain VI corresponding to the N-terminal part of the laminin-α2. A possible explanation is that missense variants located in this region have a detrimental effect on laminin-211 function through the disruption of protein folding and loss of the ability for polymerization into supramolecular networks, that occurs through a cooperative self-assembly process of laminin-211 (Durbeej, 2015; Yurchenco, 2015). A subset of these missense substitutions (namely p.Tyr138His, p.Gln167Pro, p.Leu243Pro, and p.Gly284Arg) are located on the presumed polymerization face near a patch containing the sequence P-L-E-N-G-E, corresponding to residues 208–213 of laminin-α2 (Yurchenco, 2015). These changes were identified in patients with late-onset LAMA2-MD with moderately reduced protein levels.

The next cluster consists of missense variants (n = 10, 25% of total) that specifically alter cysteine residues, located in one of the three EGF-like repeats (domains V, IIIb, and IIIa), known to establish disulfide bridges. Here, the solenoid-like structure conveyed by these rigid rod-like structures is probably modified in a way that alters the integrity of the connection between the sarcolemma and extracellular matrix mediated by laminin-α2. The last group of missense variants affects residues located in the C-terminal region of the protein that contains a tandem of five laminin G-like (LG) domains—LG1–5. A total of seven disease-associated missense variants (17.5% of all missense) are in LG2, LG3, or LG4 domains. LG4 and LG5 domains mediate

TABLE 2 New *LAMA2* missense changes and other variants possibly affecting splicing based on bioinformatic prediction tools

Exon/ Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Bioinformatic assessment [b]	Patient-ID in LSDB	Gender	Geographic origin	Phenotype	IHC for LAMA2 in muscle/skin	Zygosity/second variant/orientation (cis, trans, or unknown)	Interpretation of the second variant
1	c.112G > A	VUS	g.129204502G > A	r.(?)^ r.(spl?)	p.(Gly38Ser)^ p.(?)	–	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2) -Effect on splicing predicted in 4/4 of tools tested	102536	M	Saudi Arabia	MDC1A	Deficiency	Hom. / n.a. / n.a.	n.a.
2	c.245A > T	VUS	g.129371195A > T	r.(?)	p.(Gln82Leu)	–	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2)	111379	F	Germany	Late-onset LAMA2-related MD	–	Het. / c.7572+1G: / unknown	Pathogenic
3	c.437C > T	VUS	g.129419358C > T	r.(?)	p.(Ser146Phe)	ClinVar (RCV000505761.1); dbSNP (rs143680577)	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102470	M	United States	MDC1A	–	Het. / c.6617del / unknown	Pathogenic
5	c.745C > T	VUS	g.129465151C > T	r.(?)	p.(Arg249Cys)	dbSNP (rs376437110); ESP (0.01%); gnomAD (0.0014 %)	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102368	M	United States	MDC1A	–	Het. / c.9101_910. unknown	Pathogenic
5	c.818G > A	VUS	g.129465224G > A	r.(?)	p.(Arg273Lys)	ClinVar (RCV000394734.1)	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2)	102398	M	United States	MDC1A	–	Het. / c.3976C > T / unknown	Pathogenic
10	c.1326T > G	VUS	g.129498870T > G	r.(?)	p.(Cys442Trp)	–	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102360	M	United States	MDC1A	Deficiency	Het. / c.3976C > T / unknown	Pathogenic
								102549	M	United States	MDC1A	–	Het. / c.7658delC / unknown	Pathogenic

(Continues)

TABLE 2 (Continued)

Exon/ Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Bioinformatic assessment [b]	Patient-ID in LSDB	Gender	Geographic origin	Phenotype	IHC for LAMA2 in muscle/skin	Zygosity/second variant/orientation (cis, trans, or unknown)	Interpretation of the second variant
11i	c.1609-41_1609-7inv	VUS	g.129513784_129513818inv	r.(spl?)	p.(?)	–	-Effect on splicing predicted in 4/4 of tools tested	102463	F	United States	MDC1A	Deficiency	Het. / c.283+2delT / unknown	Likely Pathogenic
17	c.2370T > A	VUS	g.129591816T > A	r.(?)^ r.(spl?)	p.(790 =) ^p.(?)	–	Possibly affects function: -Effect on splicing predicted in 3/4 of tools tested (all except NNSplice)	102395	M	United States	MDC1A	Deficiency	Het. / c.8669dupT / trans	Pathogenic
23	c.3235T > G	Pathogenic	g.129634066T > G	r.(?)	p.(Cys1079 Gly)	–	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102726	F	Portugal	Late-onset LAMA2-related MD	Normal	Het. / c.7750-1713_7899-2154del / unknown	Pathogenic
								102718	M	Portugal	Late-onset LAMA2-related MD	Partial deficiency	Het. / c.3085C > T / trans	Pathogenic
31	c.4523G > A	Pathogenic	g.129670529G > A	r.(?)^r.(spl?)	p.(Arg1508 Lys)^p.(?)	ClinVar (RCV000483171.1); dbSNP (rs770084568); gnomAD (0.00041%)	Possibly affects function: -Missense variant not tolerated in 2/4 predictors (UMD and PolyPhen-2) -Effect on splicing predicted in 4/4 of tools tested	102457 [1]	F	United States	MDC1A	–	Het. / c.2049_2050del / trans	Pathogenic
								102484 [1]	F	United States	MDC1A	–	Het. / c.2049_2050del / trans	Pathogenic
								102656	M	United States	MDC1A	–	Hom. / n.a. / n.a.	n.a.
32	c.4654G > A	VUS	g.129674439G > A	r.(?)	p.(Ala1552 Thr)	dbSNP (rs771891309) gnomAD (0.0012%)	-Missense variant not tolerated in 2/4 predictors (UMD and PolyPhen-2)	102383	M	Canada	MDC1A	–	Het. / c.3829C > T / trans	Likely Pathogenic

(Continues)

TABLE 2 (Continued)

Exon/Intron	DNA variant (NM_000426.3)	Interpretation [a]	DNA variant (NC_000006.11) hg19	RNA variant	Predicted effect on protein	External variant databases	Bioinformatic assessment [b]	Patient-ID in LSDB	Gender	Geographic origin	Phenotype	IHC for LAMA2 in muscle/skin	Zygosity/second variant/orientation (cis, trans, or unknown)	Interpretation of the second variant
46	c.6548T>G	VUS	g.129774251T>G	r.(?)	p.(Leu2183Arg)	–	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used	102399	M	Iran	CMD	–	Hom./n.a./n.a.	n.a.
47	c.6707G>A	VUS	g.129775433G>A	r.(?)^ r.(spl?)	p.(Arg2236Lys)^p.(?)	–	Possibly affects function: -Missense variant not tolerated in 4/4 predictors used -Effect on splicing predicted in 3/4 of tools tested (all except GeneSplicer)	103971	U[1]	Portugal	LGMD/EDMD[2]	–	Het./c.2461A>C/unknown	Pathogenic
54	c.7571A>T	VUS	g.129799957A>T	r.(?)^ r.(spl?)	p.(Glu2524Val)^p.(?)	–	Possibly affects function: -Missense variant not tolerated in 3/4 predictors used (all except MutPred2) -Effect on splicing predicted in 3/4 of tools tested (all except GeneSplicer)	102379	F	United States	MDC1A	Partial deficiency	Het./c.6588dupT/unknown	Likely pathogenic

Notes. CMD: congenital muscular dystrophy; F: female; Het.: heterozygous; Hom.: homozygous; ID: identification; IHC: immunohistochemistry; M: male; MD: muscular dystrophy; MDC1A: congenital muscular dystrophy type 1A; n.a.: not applicable; U: unknown; VUS: variant of unknown (clinical) significance; [1]: Siblings; [2]: Variants identified through an anonymized screening performed in genetically uncharacterized LGMD/ EDMD patients; [a]: According to the ACMG guidelines. [b]: More detailed information available in Supporting Information Table S2. References sequences used to describe variants: NM_000426.3 and NC_000006.11.
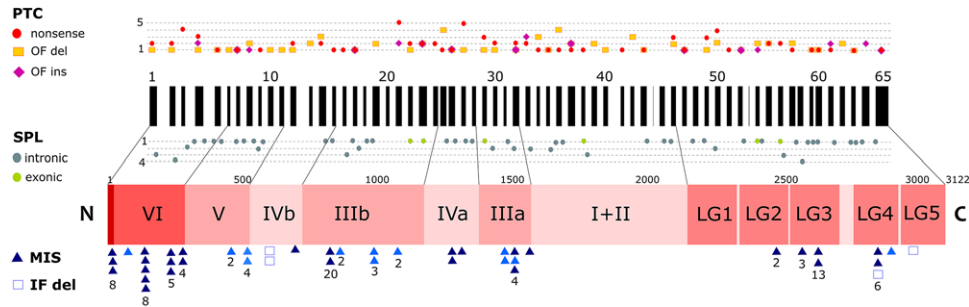
FIGURE 1 Point variants recorded in the laminin-α2 (*LAMA2*)-LOVD. Top layer unveils the number of unique variants that originate premature termination codons (PTC): nonsense, out-of-frame (OF) deletions (DEL) or insertions (INS), per *LAMA2* exon (black rectangles). Middle layer shows splice-site variants (SPL), also indicating the number of unique variants per region: intronic (grey) or exonic (light green). Laminin-211 protein domains: I to VI, and Laminin G-like (LG) are shown in light pink to red boxes, from the C-terminal (C) to N-terminal region (N). Bottom layer displays missense (MIS) changes or single codon in-frame (IF) deletions. Light blue triangles indicate substitution of a cysteine. Variants are clustered per exonic region, and numbers below each symbol indicate the total number of changes

FIGURE 2 Large deletions and duplications listed in the laminin-α2 (*LAMA2*)-LOVD. Large duplications (DUP) are shown in the top layer as yellow rectangles encompassing affected gene regions, and deletions (DEL) are shown in the bottom part of the picture as red rectangles. Black rectangles represent the *LAMA2* gene exons. Grey boxes indicate undetermined breakpoints and numbers the total of entries in the database (otherwise only one entry is present)

the binding of laminin-α2 to the O-linked carbohydrate chains of α-DG, whereas LG2 and LG3 bind to integrin α7/β1. Rare missense variants and IF deletions pose a problem for genetic data interpretation. Four IF codon deletions have been reported so far and, since there are no functional analysis strategies currently available for laminin-α2 variants, their impact remains unclear. A second aspect to be considered, as highlighted before, is that some changes predicted to be missense may instead have an effect on mRNA splicing. In addition to the application of bioinformatics tools used to assess the pathogenicity of missense changes (such as those mentioned in 3.1), it is advisable to consider if their location coincides with the potential hotspots outlined here.

Since our previous assessment (Oliveira et al., 2014) only two novel large deletions have been reported (Bhowmik, Dalal, Matta, Sundaram, & Aggarwal, 2016; Ding et al., 2016), totaling 17 deletions and two duplications (Figure 2). There are two apparent mutational "hotspots" for large deletions, the first region includes exons 3 and 4, and the second is in the 3' end of *LAMA2* gene (exons 56 to 65).

Considering the distribution of disease-associated variants, exons 14, 21, 22, 26, 27, 36, 38, and 56 contain over 25 variant entries. In contrast, seven exons (namely 20, 28, 44, 45, 48, 53, and 58) have no disease-causing variants reported so far. Fourteen disease-

associated variants are among the most prevalent in the *LAMA2*-LOVD database, with at least 10 independent entries each (Table 3). The most frequent across different ethnical backgrounds are: c.2049_2050del (p.Arg683Serfs*21), c.3085C > T (p.Arg1029*), and c.3976C > T (p.Arg1326*). Interestingly, these variants are also represented in population variant databases such as gnomAD (ExAC), found in heterozygosity with frequencies ranging from 0.012 to 0.001%. Other variants such as c.1854_1861dup (p.Leu621Hisfs*7), seem to be population- or ethnic group-specific, exhibiting a relatively high frequency (0.23%) within control alleles from the "Latino" population (gnomAD).

5 | CLINICAL RELEVANCE: THE EXPANDING DISEASE SPECTRUM OF *LAMA2*-RELATED MD

5.1 | Genotype–phenotype correlations

The severest end of the spectrum of *LAMA2*-related MD—MDC1A—corresponds to a neonatal onset disease that gives rise to hypotonia and compromised normal motor development. In *LAMA2*-related MD

TABLE 3 List of the most frequent pathogenic variants in the *LAMA2* LSDB (variants with 10 or more entries in LOVD)

Exon/ Intron	DNA variant (NM_000426.3)	DNA variant (hg19)	RNA variant	Predicted effect on protein	Number of independent entries in LOVD	Geographic origin of patients (LOVD)	gnomAD/ExAC data: population, nr alleles/ total alleles (frequency)
13	c.1854_1861dup	g.129571328_ 129571335dup	r.1854_1861dup	p.Leu621Hisfs*7	12	France, Portugal, Brazil, Spain	Latino: 2/838 (0.23%)
14	c.2049_2050del	g.129573393_ 129573394del	r.2049_2050del	p.Arg683Serfs*21	42	Several countries	All populations except Ashkenazi Jewish: 34/277,008 (0.012%)
18	c.2461A > C	g.129601216A > C	r.2461a > c	p.Thr821Pro	18	Portugal	–
22	c.3085C > T	g.129621928C > T	r.(?)	p.(Arg1029*)	24	Portugal, Spain, United States	Latino: 1/34,418 (0.003%); European (Non-Finnish): 1/126,676 (0.001%)
26i	c.3924+2T > C	g.129637097T > C	r.3736_3924del	p.Leu1246_Glu1308del	33	Saudi Arabia, Sudan, United States	–
27	c.3976C > T	g.129637234C > T	r.(?)	p.(Arg1326*)	24	Portugal, Sweden, United States, Spain, Denmark	European (Non-Finnish): 14/126,524 (0.001%); European (Finnish): 1/25,788 (0.004%)
32	c.4645C > T	g.129674430C > T	r.[4645c > u, 4580_4717del]	p.[Arg1549*, Cys1527_Val1572del]	10	Australia, Italy, United States	South Asian: 2/30,780 (0.006%)
36i	c.5234+1G > A	g.129712799G > A	r.5072_5234del	p.Val1765Serfs*21	10	Portugal, Canada, United States	Latino: 1/34,376 (0.003%); European (Non-Finnish): 2/126,266 (0.002%)
38	c.5476C > T	g.129722399C > T	r.(?)	p.(Arg1826*)	11	China, Saudi Arabia, United Kingdom	East Asian: 2/17,240 (0.012%); European (Non-Finnish): 5/111,398 (0.0045%)
38i	c.5562+5G > C	g.129722490G > C	r.[5446_ 5562del, 5562_5563ins5562	p.[Lys1816_Asp1854del, Tyr1855Valfs*24]	14	United Kingdom, United States	European (Non-Finnish): 7/125,864 (0.0056%); European (Finnish): 1/25,408 (0.0039%)
46	c.6488del	g.129774191delA	r.(?)	p.(Lys2163Argfs*12)	15	Qatar, Saudi Arabia, United States	–
55	c.7732C > T	g.129802567C > T	r.(?)	p.(Arg2578*)	12	China, Denmark, Mexico, Russian Federation, United States	Latino: 3/34,380 (0.0087%); European (Non-Finnish): 10/126,598 (0.0079%); East Asian: 1/18,834 (0.0053%); South Asian: 1/30,782 (0.0032%)
55i_ 56i	c.7750-1713_ 7899-2154del	g.129805906_ 129810892del	r.7750_7898del	p.Ala2584Hisfs*8	17	Portugal	–
58i	c.8244+1G > A	g.129813629G > A	r.8076_8244del	p.Pro2693Valfs*12	12	Germany, Portugal, Tunisia, United States	European (Non-Finnish): 1/111,114 (0.0009%)

Note. nr: number.

the locomotion attainment has been considered an important clinical measure of disease severity. In a series of 26 MDC1A patients only two had acquired independent locomotion (Oliveira et al., 2008). Interestingly, all patients that harbored variants inducing PTC in both disease alleles were unable to achieve independent walking. In contrast, the two patients that were able to walk had a missense or a single codon deletion in one of the disease genes. LSDB content and other studies reported in the literature (Geranmayeh et al., 2010) further corroborated our findings. However, there are exceptions to this rule; for example, a patient with a homozygous nonsense variant

(p.Arg1549*) was able to reach ambulation and even climb stairs (Geranmayeh et al., 2010). This particular variant was reported in association with partial deficiency of laminin-α2 in several unrelated patients (Di Blasi et al., 2000; Geranmayeh et al., 2010; Pegoraro et al., 1998). Here, an explanation for this discrepancy is the fact that the variant is located within exon 32 that undergoes alternative splicing (Pegoraro et al., 2000). The exon removal leads to an IF deletion at the mRNA level, thereby restoring the reading frame from the PTC created by the nonsense variant.

Geranmayeh et al. (2010) provided further genotype–phenotype correlations with prognostic clinical implications. Statistically significant differences were identified between patients with complete deficiency and those with partial deficiency of laminin-α2. Patients with absence of laminin-α2 had earlier onset ($P = 0.0073$), lack of independent ambulation ($P = 0.0215$), and were more prone to requiring artificial feeding ($P = 0.0099$) or respiratory support ($P = 0.0354$; Geranmayeh et al., 2010). Within MDC1A, there is a subset of patients with early onset phenotype but a "milder" disease progression and with partial laminin-α2 deficiency. This partial deficiency is often associated with missense variants, IF deletions, and splicing variants (leaky or inducing IF exon-skipping; Allamand & Guicheney, 2002; Quijano-Roy et al., 2012). One of the earliest such cases reported had a homozygous variant (p.Cys996Arg) that affects domain IIb of laminin-α2 (Nissinen et al., 1996).

Despite the general consistency between phenotype, the type of variant and the IHC status, some exceptions have been documented in the literature. These include patients with complete laminin-α2 deficiency and missense variants that achieved independent locomotion (Geranmayeh et al., 2010), although this could be attributed to IHC sensitivity issues. Intrafamilial clinical variability has also been reported, such as that found among patients from one large Kenyan kindred of Asian ancestry. Here, patients shared the same genotype (homozygous missense variant located in the G-domain of laminin-α2) but locomotion was not achieved in all cases (Geranmayeh et al., 2010).

As previously mentioned, a very small fraction of *LAMA2*-MD patients have brain structural defects, which are frequently associated with intellectual disability (ID) and/or refractory seizures (Geranmayeh et al., 2010, Vigliano et al., 2009). However, there are also reports of patients, with these structural defects who, apparently have no seizures or ID. The opposite also holds true in the case of seizures (and to a lesser extent ID) since they have been reported in patients without cerebral structural changes. Based on the reassessment of data available in the LOVD and reported in the literature, no association was found between epilepsy, cognitive function or brain anomalies, and a particular set of *LAMA2* genotypes/variants. The variants found in these cases are diverse in terms of their impact, ranging from those causing PTC to missense changes, and are apparently dispersed with no obvious hotspot along the gene. Furthermore, phenotypical discrepancies have been found in patients sharing with same genotype. For example, two siblings reported by Di Blasi et al. (2001) and case #2 from Nelson et al. (2015) share the same genotype (a homozygous nonsense variant p.Arg744*), but cortical polymicrogyria and lissencephaly were only reported in the latter patient. It is conceivable that other genetic factors besides *LAMA2* variants are contributing to these phenotypes.

Over the last few years there has been a significant increase in reports of late-onset *LAMA2*-related MD patients (Ding et al., 2016; Gavassini et al., 2011; Harris et al., 2017; Kevelam, van Engelen, van Berkel, Küsters, & van der Knaap, 2014; Kim et al., 2017; Løkken, Born, Duno, & Vissing, 2015; Marques et al., 2014; Nelson et al., 2015; Rajakulendran, Parton, Holton, & Hanna, 2011). Most of these patients have heterozygous or homozygous missense or splice variants. Their clinical presentation is also variable but often overlapping with a childhood-onset LGMD, consisting of proximal muscle weakness and delayed motor milestones, but in all cases achieving independent ambulation. Rigid spine syndrome with joint contractures has been also reported in some patients (Nelson et al., 2015).

5.2 | Additional cases of late-onset *LAMA2*-related MD sharing the p.Thr821Pro variant

Phenotypic variability in *LAMA2*-related MD has been clearly underestimated so far, with only a limited number of patients with this later-onset phenotype reported in the literature. As for establishing further genotype–phenotype correlations, the cases are still relatively scarce and there is a vast diversity of genetic defects and/or genotypes, which makes it difficult to stratify patients into homogeneous groups.

To address some of these limitations, and resorting to our large *LAMA*-related MD patient cohort, the clinical and genetic characterization of six additional patients with a late-onset phenotype from four unrelated families is reported (Table 4). They all share the same missense variant: p.Thr821Pro. In five cases the genotype was similar in that, besides this missense substitution, the second allele was a truncating variant: c.7750-1713_7899-2154del (p.Ala2584Hisfs*8) in patients P1 and P2, c.3976C > T (p.Arg1326*) in P3 and P4, and c.1854_1861dup (p.Leu621Hisfs*7) in P5. The sixth patient (P6) represents the first documented case with a homozygous p.Thr821Pro missense variant. Most of these patients were only diagnosed during adulthood, which reflects the diagnostic difficulties concerning non-MDC1A cases. All have a very mild muscle weakness (as compared with typical MDC1A) with lower limb weakness resulting in gait disturbances. In the oldest patient (P6) this weakness culminated in loss of ambulation during the sixth decade of life. In four patients brain MRI was performed (P1, P2, P3, and P6), revealing WMC like those usually found in *LAMA2*-related MD (Figure 3a–c). These findings were pivotal for conducting *LAMA2* gene analysis in three of the cases. Patient P6, who developed dementia over the last 2 years, also had hypothalamus and pons alterations (data not shown). Five patients were subjected to a muscle biopsy. These showed myopathic or dystrophic features (Figure 3d–f), and IHC analysis for laminin-α2 revealed apparently normal labeling ($n = 3$, Figure 3g–i) or partial deficiency ($n = 1$, data not shown).

5.3 | Prevalence of p.Trp821Pro variant in a genetically uncharacterized MD patient cohort

The missense variant p.Trp821Pro is one of the most frequent genetic causes of late-onset *LAMA2*-MD in a population-specific (Portuguese)

TABLE 4 Additional *LAMA2*-related muscular dystrophy patients sharing the p.Thr821Pro variant

Family patient #	Genotype	Predicted effect on protein	Age (gender)	Age of first symptoms onset	Phenotype at onset	Pattern of muscle weakness; other clinical features	Cardiac involvement?	Contractures?	Independent locomotion? (age)	Loss of ambulation? (age)	CK levels (U/l)	Muscle biopsy [1]	Laminin-α2 IHC [1]	Brain changes (MRI)
F.I-P1	c.2461A > C + c.7750-1713_7899-2154del	p.Thr821Pro + p.Ala2584Hisfs*8	42 yrs (F)	Third decade	Migraine-like headaches for 8 mo. Complaints of limb weakness and walking difficulties	Mild generalized muscular atrophy and tetraparesis (4+/5 grade), feet dorsiflexion (4/5). Paresis of trunk and neck flexion, cannot do sit-ups or lift head while in supine position.	N	N	Y	N	n.p.	Moderate myopathic changes, with discrete endomysial fibrosis	Normal	WMC
F.I-P2	c.2461A > C + c.7750-1713_7899-215del	p.Thr821Pro + p.Ala2584Hisfs*8	50 yrs (M)	Forth decade	Running difficulties and progressive lower limb weakness	Proximal paresis (4/5 grade) in upper limbs, distal and proximal paresis in lower limbs. Paresis of trunk and neck flexion (grade 2 and 4+, respectively). Lordosis and myopathic gait with slight steppage.	N	N	Y	N	n.p.	Myopathic changes with fiber necrosis, also "ragged red fibers" and large number of COX negative fibers.	Normal	WMC
F.II-P3	c.2461A > C + c.3976C > T	p.Thr821Pro + p.Arg1326*	15 yrs (F)	18 mo	Running difficulties and stairs	Proximal paresis (2/5 grade) in upper limbs and in lower limbs (3/5 grade), neck flexion (grade 2/5).	N	Y	Y (15 mo)	N	839	Dystrophic changes	Normal	WMC
F.II-P4	c.2461A > C + c.3976C > T	p.Thr821Pro + p.Arg1326*	11 yrs (F)	5 yrs	Facial fatigue	Proximal paresis (2/5 grade) in upper limbs and in lower limbs (4/5 grade), neck flexion (grade 1/5).	N	N	Y (14 mo)	N	2466	n.p.	n.p.	n.p.
F.III-P5	c.2461A > C + c.1854_1861dup	p.Thr821Pro + p.Leu621Hisfs*7	33 yrs (F)	Childhood	Running difficulties. Difficulty in getting out of a bed	Proximal paresis (4-3/5 grade) in lower limbs. Paresis of trunk and neck flexion (grade 3/5). Lordosis and myopathic gait.	N	N	Y	N	~2,000	Dystrophic changes	Partial deficiency	n.p.
F.IV-P6	c.2461A > C (hom.)	p.Thr821Pro	71 yrs (M)	Childhood	Gait impairment	LGMD initially suspected. Proximal tetraparesis (grade 4/5, 4-/5 in lower limbs). Moderate intellectual disability (last 2 yrs).	Y [2]	N	Y	Y (last 2 years)	579	Moderate dystrophic changes associated to angulated atrophic fibers and nuclear clumps	Normal	WMC, thalamus and pons involvement

Notes. F: female; hom.: homozygous; LGMD: limb-girdle muscular dystrophy; M: male; MD: muscular dystrophy; mo: months; N: no; n.p.: not performed; U: unknown; WMC: white matter changes; Y: yes; yrs: years. [1]: Clone: Mer3/22B2 (Leica Biosystems, Newcastle upon Tyne, United Kingdom); [2]: Left ventricle hypokinesia of unknown cause, normal ejection fraction.
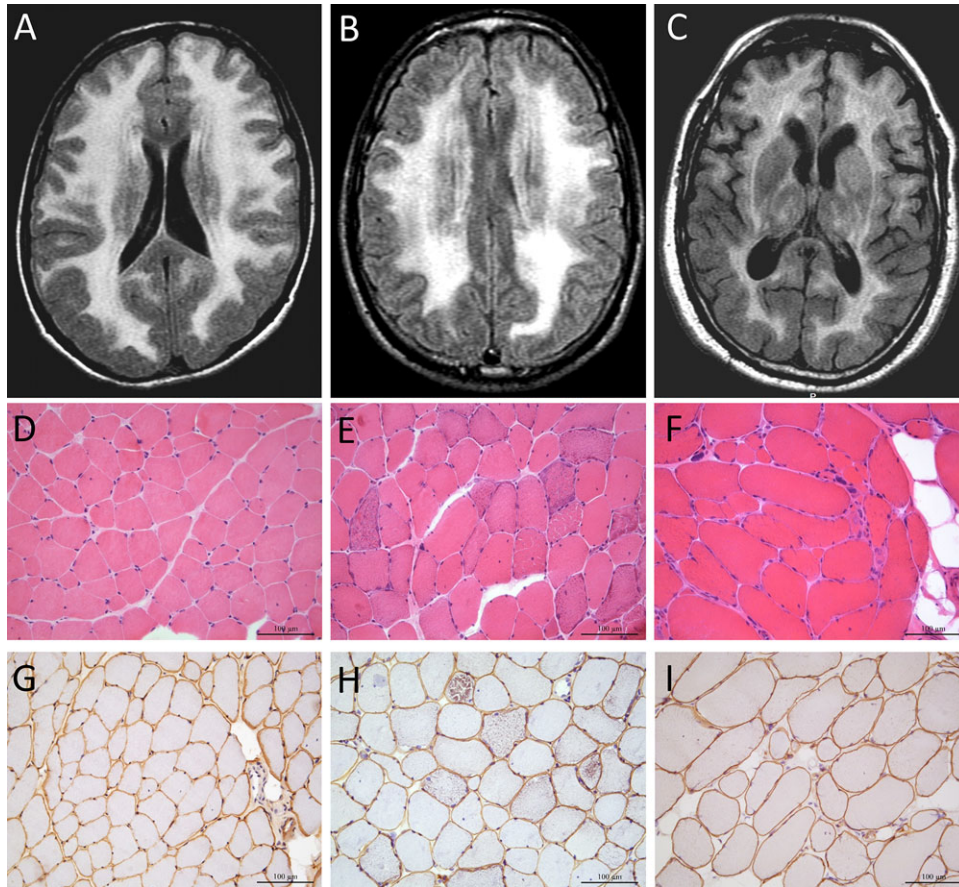
FIGURE 3 Brain magnetic resonance imaging (MRI) and muscle neuropathology results. Patient P1: (a) brain MRI (FLAIR) shows typical white matter changes (WMC) with normal structural cerebral cortex changes; (d) moderate myopathic changes with discrete endomysial fibrosis in hematoxylin eosin (h & e) staining, and (g) normal immunohistochemistry (IHC) for laminin-α2. Patient P2: (b) brain MRI (FLAIR) with WMC and normal cerebral cortex; (e) myopathic changes with necrotic fibers and several "ragged red fibers" (h & e), and (h) normal IHC for laminin-α2. Patient P6: (c) WMC in brain MRI (FLAIR); (f) dystrophic changes (necrotic fibers under myophagocytosis, fiber splitting and hypersegmentation, and fat substitution) and mild neurogenic features (atrophic angulated fibers and nuclear clumps), and (i) normal IHC for laminin-α2

patient cohort. This missense substitution was initially identified in two patients with atypical presentations (Marques et al., 2014), prior to the six patients described above. A further three patients, from two unrelated families with Portuguese ancestry, have also been reported by other groups: one patient from Canada (with #102482 in LAMA2-LOVD), and two brothers studied in France (Nelson et al., 2015). Thus far, all patients are reported to have milder muscle weakness and the majority were initially classified as possible LGMD or EDMD. To evaluate if this missense variant could account for additional uncharacterized cases, we screened an irreversibly anonymized group of 239 myopathic Portuguese patients with clinical presentation is compatible with LGMD or EDMD. Variant screening was performed by restriction fragment length analysis (RFLA, Supporting Information IV), since the c.2461A > C change creates a new restriction site for HpyCH4III (Supporting Information Figure S4). Positive samples were confirmed by Sanger sequencing. A total of seven patients carried this missense substitution (2.9% of the cohort), three of which were homozygotes and four were heterozygotes (2% of all disease alleles). To further ascertain the genotype of the four patients carrying the c.2461A > C variant in heterozygosity, the entire coding sequence of the LAMA2 gene was sequenced. In all patients an additional heterozygous variant was detected. Three were previously identified in other (MDC1A) patients: c.4739dup (p.Leu1581Profs*5; Oliveira et al., 2008), c.3372dup (p.Cys1125Metfs*4; patient #103970 in Table 1) and a missense variant c.32T > C (p.Leu11Pro) listed in ClinVar (RCV000157587.1) as being disease associated. The fourth variant, c.6707G > A, is also new and was interpreted as a VUS; it predictably gives rise to a missense change (p.Arg2236Lys) and/or may have an effect on splicing (r.spl?; Table 2, patient #103971).

Since the c.2461A > C variant was not listed in variant population databases, its prevalence was estimated in control individuals using the aforementioned RFLA-screening strategy. For this study, we randomly selected and irreversibly anonymized 1,100 out of a total of 11,000 samples previously analyzed in the laboratory. These were residual samples from genetic studies for diseases unrelated with neuromuscular disorders that are performed on a nationwide basis. The c.2461A > C variant was identified in one of these samples, in heterozygosity. Its allelic frequency in the general population was estimated as 0.0452% (1/2,200), which explains the relatively high prevalence of this variant among Portuguese patients with LAMA2-MD.

Overall, the presented data reinforces that it is diagnostically important to consider LAMA2 gene involvement not only in CMD
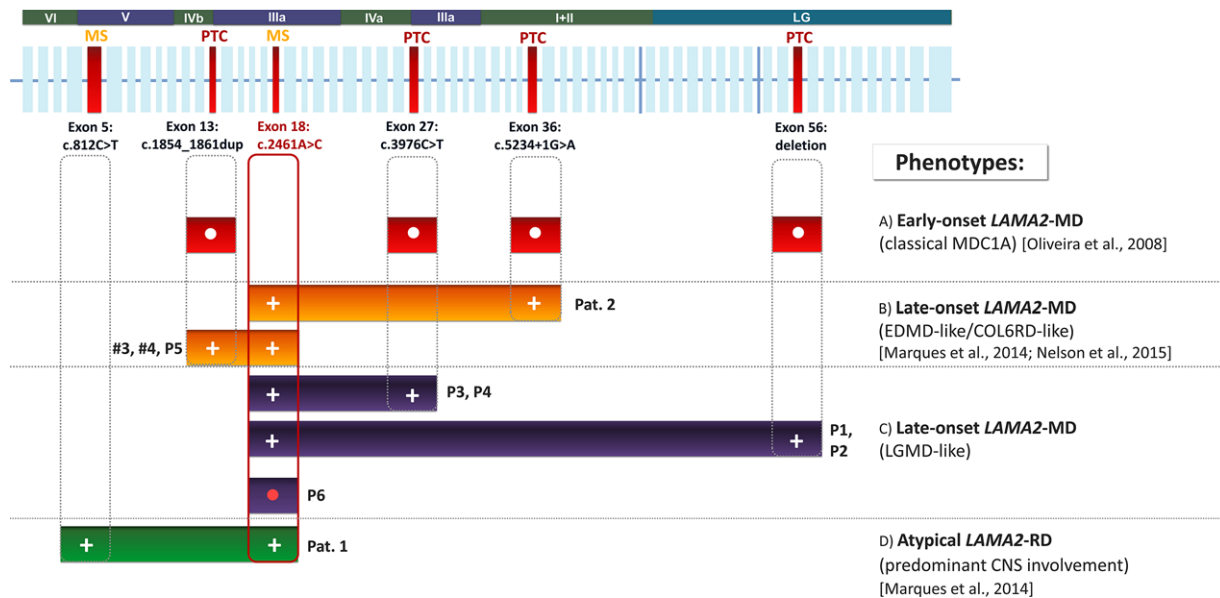
FIGURE 4 Different phenotypes in laminin-α2-related muscular dystrophy (*LAMA2*-MD) found in association with the c.2461A > C (p.Trp821Pro) variant. (a) Early-onset (classical muscular dystrophy type 1A [MDC1A]): no independent ambulation; muscle biopsy shows dystrophic features and no labeling for laminin-α2 in immunohistochemistry (IHC). Several patients reported in Oliveira et al. (2008). (b) Late-onset (Emery–Dreifuss muscular dystrophy [EDMD]/COL6-RD-like): rigid spine syndrome; cardiac involvement in some patients; walking difficulties; dystrophic features in MD, normal and irregular laminin-α2 staining in IHC. Pat.2—patient 2 (Marques et al., 2014); cases #3, #4 (Nelson et al., 2015); P5 (this work). (c) Late-onset (limb-girdle muscular dystrophy (LGMD)-like): slow progression; dystrophic features, normal and irregular laminin-α2 staining in IHC, walking difficulties later in life; P1–4, P6 (this work). (d) Atypical *LAMA2*-RD: predominant central nervous system (CNS) involvement, (occipital agyria, white matter changes (WMC), epilepsy); increased variability of muscle fiber diameter and irregular laminin-α2 staining in IHC. Pat.1—patient 1 in Marques et al. (2014). Genotype–phenotype correlations suggest that the classical MDC1A presentation is explainable by variants causing premature termination codons (PTC) in both disease alleles. While late-onset *LAMA2*-MD are more likely to be associated with missense (MS) substitutions

Note. ●: homozygous; +: Heterozygous.

patients, but also as a possible cause of MD with onset beyond childhood and even in adulthood. The association between *LAMA2* and this later onset phenotype was not fully established, considering the limited number of cases reported so far. Nonetheless, it is advisable to include *LAMA2* in the list of candidate genes for MDs (LGMD or EDMD). The p.Trp821Pro missense variant constitutes an interesting genotype–phenotype linker, as it may give rise to different phenotypes depending on the variant found in the second allele (Figure 4).

6 | DIAGNOSTIC RELEVANCE

Molecular defects in the *LAMA2* gene are the main genetic causes (∼30%) of CMDs in most countries, except for Japan where Fukuyama-type CMD has the highest prevalence, due to a frequent founder mutation in the *FKTN* gene (Kobayashi et al., 1998). Besides the clinical examination, the clinical diagnostic workup of CMDs conventionally relies upon performing a muscle biopsy (Bönnemann et al., 2014). In addition to standard staining methods, muscle pathology analysis includes a panel of antibodies for IHC against proteins involved in MD (laminin-α2, sarcoglycans, dystrophin, and dysferlin). Three different commercial antibodies are currently available for laminin-α2 IHC studies: clone 5H2 detects the 80 kDa protein (C-terminal region),

clone Mer3/22B2 detects the 300 kDa product (N-terminal region), and clone 4H8-2 clone, which also recognizes the N-terminal domain. The diagnostic sensitivity of IHC is extremely high for typical MDC1A cases, where complete deficiency would be detectable regardless of the antibody used for analysis. The milder *LAMA2*-MD cases are more challenging as often only a partial deficiency is often documented. Moreover, depending on the underlying molecular defects, this IHC deficiency may not be consistent for the different antibodies (Cohn, Herrmann, Sorokin, Wewer, & Voit, 1998). N-terminal antibodies usually have higher sensitivity for cases with partial laminin-α2 deficiency, as there was a relatively intact labeling with the antibody for the 80 kDa fragment, when compared with that using the other antibodies (Cohn et al., 1998). It is therefore advisable to include at least two different antibodies against laminin-α2 in order to increase IHC sensitivity. In a small fraction of CMD patients there is also irregular labeling or partial laminin-α2 deficiency. There is some degree of genetic heterogeneity among these patients, depending on whether it is a primary or a secondary deficiency. To distinguish between these two possibilities, antibodies against glycosylated residues of α-DG and laminin-α4/5 may be effective. If changes are detected in α-DG, this would indicate a defective glycosylation pathway and involvement of other loci. On the other hand, normal α-DG labeling and overexpression of laminin-α4/5 (a compensatory gene expression mechanism) is suggestive of a primary laminin-α2 deficiency.

Brain MRI performed beyond the first 6 to 12 months is also an important diagnostic resource for CMDs. As previously mentioned all LAMA2-related MD patients have brain WMC, consisting in bilateral hyperintensity signal on T2-weighted and FLAIR MRI, in periventricular areas and subcortical cerebral hemisphere (Quijano-Roy et al., 2012). These findings alone should be an indication to perform LAMA2 genetic testing. As demonstrated by this work and previously suggested by Gavassini et al. (2011), it is diagnostically relevant to perform brain MRI in uncharacterized LGMD patients. This could be performed even during adulthood, as these typical brain changes will persist throughout life. Brain MRI is especially relevant for "atypical" or mild MD cases where IHC for laminin-α2 has a lower diagnostic yield.

Considering the size and number of exons in the LAMA2 gene, its genetic analysis has been simplified, more than a decade ago, with the introduction of automatized sequencers (fragment analyzers) for Sanger sequencing and the use of universal-tailed primers. Gene sequencing is undoubtedly the approach with the highest sensitivity for LAMA2 analysis, detecting approximately 80% of disease-associated variants. Based on the variant data collected, there is a significant frequency ($\sim$18%) of large deletions and duplications. The genetic study should therefore be complemented with other molecular techniques such as MLPA or array-CGH.

Data available in the LSDB and population-specific cohorts can help to optimize the LAMA2 genetic analysis. This was exemplified in a Portuguese CMD patient cohort where a 3-tier genetic test was proposed (Oliveira et al., 2014): (a) sequencing a small set of selected exons where the majority of point mutations are located (based on a specific population or ethnical group variant data); (b) sequencing the remaining LAMA2 exons; and (c) MLPA analysis or array-CGH.

The introduction of next-generation sequencing technology (NGS, or massive parallel sequencing) has remodeled genetic analysis strategies, especially in genetically heterogeneous conditions such as the MDs. Distinct NGS applications such as gene panels or whole-exome sequencing (WES) can be extremely useful to address diagnostically difficult cases. In fact, novel cases with milder LAMA2-related phenotypes recently reported in the literature have been solved resorting to NGS (Dean, Rashid, Kupsky, Moore, & Jiang, 2017; Ding et al., 2016; Harris et al, 2017; Kim et al., 2017).

The impact of NGS technology is also reflected in five patients described in this work (ID#s: 102662, 132012, 132013, 132015, 132025 in table I), whose disease-associated variants listed were identified by NGS gene panels. One further patient (ID# 103207 in Table 1) demonstrates the utility of NGS to address genetic and clinical heterogeneity. This is a patient with an LGMD phenotype and ID, who has remained without genetic characterization for several years. The patient, currently 14 years of age, had delayed motor development (started walking at 31 months of age), lumbar lordosis, and elevated CK levels ($\sim$1300 U/l). Muscle biopsy performed at 6 years of age (in another clinical center where she was initially followed) revealed dystrophic features and normal IHC results for dystrophin and sarcoglycans. Genetic analysis of FKRP, CAPN3, LMNA, and DMPK genes were negative. The patient was studied by WES as previously reported in a similar research (Oliveira, Martins, Pinto Leite, Sousa, & Santos, 2017). As a first approach, WES data analysis was restricted to a set of genes

known to be associated with muscle diseases (Supporting Information IV). Within the list of filtered-in variants, two heterozygous variants were identified in LAMA2 (Supporting Information Figure S5). The first was the c.1854_1861dup variant, previously reported as disease associated in several MDC1A patients, and the second was a novel splicing variant c.819+2T > C located in the donor splice-site of intron 5 (Table 1). To further characterize the effect of this splice-site variant, a muscle fragment available from patient ID# 102735 (shares the same variant) was used (Supporting Information II). LAMA2 transcript analysis by RT-PCR showed the presence of multiple aberrant products that, upon sequencing, were attributed to multiple skipping events involving exons 5 to 7 (Table 1, Supporting Information Figures S1 and S2). Study of the patient's parents confirmed compound heterozygosity, as each progenitor carried a different LAMA2 variant. Brain MRI performed after WES analysis, revealed WMC but not configuring the typical pattern found in MDC1A cases. In this patient, axial T2 and FLAIR revealed small focal white matter hyperintensities in the subcortical part of brain, more specifically in the frontal, temporal-anterior, parahyppocampus, and insula regions with sparing of the internal capsule and corpus callosum (data not shown).

Finally, as demonstrated by five cases listed in Table 1 (ID#s 102324, 102369, 102378, 132014, and 132015), a small percentage of patients were found to have only one heterozygous LAMA2 disease-causing variant. This could be attributed to deeply placed intronic variants affecting splicing or variants located in the gene's promotor region, both of which are not covered by conventional sequencing, gene panels, or even WES. Here, a more comprehensive NGS approach such as WES and/or RNA sequencing may ultimately provide a final answer to such cases with incomplete genotyping.

7 | FUTURE PROSPECTS

Although there is an increasing recognition of the involvement of LAMA2 disease-associated variants in the genetic etiology of muscular dystrophies, the incidence is probably still underestimated. To improve the diagnosis of these cases, it is necessary to include both brain MRI and to evaluate the expression of laminin-α2 in muscle by IHC. These two approaches are often not considered in the clinical workup of patients with (non-congenital) myopathies. NGS can also contribute toward the identification of further cases. However, the interpretation of variants from such studies often leads to their classification as VUS, which considerably limits the clinical utility of these genetic data.

As for further research, it is necessary not only to continue to document clinical data and LAMA2 variants to obtain further genotype–phenotype correlations, but also to develop strategies for functional analysis and validation of new variants, especially those predictably of the missense type. This task may be complex, as variants might affect several key aspects of the laminin-211 life-cycle: (a) posttranslational modification, (b) protein translocation and secretion process, (c) interaction with membrane-specific receptors, and (d) variety of molecular partners in the BM, possibly some yet to be identified. One strategy for functional analysis would imply obtaining a biological sample from the patient by an invasive procedure (e.g., muscle or

skin biopsy), expanding cells through *in vitro* culture, and performing protein–protein interaction studies, such as pull-down assays using a battery of different bait-proteins known to interact with laminin-α2. Failure to detect a particular interaction would indicate a deleterious effect. To enable such studies, further research should primarily focus in a comprehensive search for domain-specific interactions, which could be accomplished by high-throughput proteomics analyses. An assay for those variants specifically affecting domains involved in laminin polymerization has been reported (Cheng, Champliaud, Burgeson, Marinkovich, & Yurchenco, 1997; Hussain, Carafoli, & Hohenester, 2011). Basically, a mixture of wild-type with the mutated form of laminin would show a failure in establishing normal polymerization levels. Here, the limiting step would be generating and purifying sufficient amounts of proteins to conduct these *in vitro* studies.

As laminin-α2 is not confined to muscle or brain cells, in a transgenic mouse model with deficient laminin-α2 it was shown that the loss of this protein caused disruption of the apical ectoplasmic specialization-blood–testis barrier, and leading to male infertility (Häger, Gawlik, Nyström, Sasaki, & Durbeej, 2005). The laminin-α2 in testis was further implicated in the regulation of an axis that functionally links the BM to the blood–testis barrier of Sertoli cells (Gao et al., 2017). Considering that human infertility has not been linked to laminin-α2, it would be relevant to evaluate male reproductive issues in late-onset *LAMA2*-MD patients.

One of the most important aspects concerning *LAMA2*-MD is the development of a suitable treatment for this condition. Several approaches have been proposed, developed, and tested in laminin-α2-deficient mice and zebrafish models (reviewed by Durbeej, 2015; Wood & Currie, 2014). One particularly effective approach targets extracellular matrix modulation as a way to ameliorate MDC1A. Here, strategies aim to improve muscle viability, through the augmentation of residual functionality within the cellular system, such as upregulation of other laminins (α4 or α1) and integrin-α7 (Wood & Currie, 2014). However, with laminin-411 there are some limitations for BM repair, since this laminin only forms a trimeric structure, lacking capacity to further self-polymerize into superstructures such as those derived from laminins-211 or -111. Overall there are some hurdles toward its applicability, namely the large size of laminins, which make its delivery to target locations extremely challenging. An effective way to address this problem is to use shorter engineered proteins, such as the chimeric laminin/nidogen protein or mini-agrin, shown to be effective in a *LAMA2*-MD mouse model (dy^W/dy^W; McKee et al., 2017; Reinhard et al., 2017). Probably in a near future, we will witness a new generation of laminin-binding proteins that, depending on the underlying genetic defects, are able to replace defective domains of laminin and promote the assembly of a stable and fully functional BM.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

ORCID

Jorge Oliveira http://orcid.org/0000-0003-3924-6385

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, 7(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Allamand, V., & Guicheney, P. (2002). Merosin-deficient congenital muscular dystrophy, autosomal recessive (MDC1A, MIM#156225, *LAMA2* gene coding for alpha2 chain of laminin). *European Journal of Human Genetics*, 10(2), 91–94. https://doi.org/10.1038/sj.ejhg.5200743

Aumailley, M., Bruckner-Tuderman, L., Carter, W. G., Deutzmann, R., Edgar, D., Ekblom, P., … Yurchenco, P. D. (2005). A simplified laminin nomenclature. *Matrix Biology*, 24(5), 326–332. https://doi.org/10.1016/j.matbio.2005.05.006

Bhowmik, A. D., Dalal, A. B., Matta, D., Sundaram, C., & Aggarwal, S. (2016). Targeted next generation sequencing identifies a novel deletion in *LAMA2* Gene in a merosin deficient congenital muscular dystrophy patient. *Indian Journal of Pediatrics*, 83(4), 354–355. https://doi.org/10.1007/s12098-015-1822-3

Bönnemann, C. G., Wang, C. H., Quijano-Roy, S., Deconinck, N., Bertini, E., Ferreiro, A., … North, K. N. (2014). Diagnostic approach to the congenital muscular dystrophies. *Neuromuscular Disorders*, 24(4), 289–311. https://doi.org/10.1016/j.nmd.2013.12.011

Brett, F. M., Costigan, D., Farrell, M. A., Heaphy, P., Thornton, J., & King, M. D. (1998). Merosin-deficient congenital muscular dystrophy and cortical dysplasia. *European Journal of Paediatric Neurology*, 2(2), 77–82.

Celli, J., Dalgleish, R., Vihinen, M., Taschner, P. E., & den Dunnen, J. T. (2012). Curating gene variant databases (LSDBs): Toward a universal standard. *Human Mutation*, 33(2), 291–297. https://doi.org/10.1002/humu.21626

Chan, S. H., Foley, A. R., Phadke, R., Mathew, A. A., Pitt, M., Sewry, C., & Muntoni, F. (2014). Limb girdle muscular dystrophy due to *LAMA2* mutations: Diagnostic difficulties due to associated peripheral neuropathy. *Neuromuscular Disorders*, 24(8), 677–683. https://doi.org/10.1016/j.nmd.2014.05.008

Cheng, Y. S., Champliaud, M. F., Burgeson, R. E., Marinkovich, M. P., & Yurchenco, P. D. (1997). Self-assembly of laminin isoforms. *The Journal of Biological Chemistry*, 272(50), 31525–31532.

Cohn, R. D., Herrmann, R., Sorokin, L., Wewer, U. M., & Voit, T. (1998). Laminin alpha2 chain-deficient congenital muscular dystrophy: Variable epitope expression in severe and mild cases. *Neurology*, 51(1), 94–100.

Dean, M., Rashid, S., Kupsky, W., Moore, S. A., & Jiang, H. (2017). Child neurology: *LAMA2* muscular dystrophy without contractures. *Neurology*, 88(21), e199–e203. https://doi.org/10.1212/WNL.0000000000003958

Deodato, F., Sabatelli, M., Ricci, E., Mercuri, E., Muntoni, F., Sewry, C., … Guzzetta, F. (2002). Hypermyelinating neuropathy, mental retardation and epilepsy in a case of merosin deficiency. *Neuromuscular Disorders*, 12(4), 392–398.

Di Blasi, C., He, Y., Morandi, L., Cornelio, F., Guicheney, P., & Mora, M. (2001). Mild muscular dystrophy due to a nonsense mutation in the *LAMA2* gene resulting in exon skipping. *Brain*, 124(Pt 4), 698–704.

Di Blasi, C., Mora, M., Pareyson, D., Farina, L., Sghirlanzoni, A., Vignier, N., … Morandi, L. (2000). Partial laminin alpha2 chain deficiency in a patient with myopathy resembling inclusion body myositis. *Annals of Neurology*, 47(6), 811–816.

Di Muzio, A., De Angelis, M. V., Di Fulvio, P., Ratti, A., Pizzuti, A., Stuppia, L., … Uncini, A. (2003). Dysmyelinating sensory-motor neuropathy in merosin-deficient congenital muscular dystrophy. *Muscle & Nerve*, 27(4), 500–506. https://doi.org/10.1002/mus.10326

Ding, J., Zhao, D., Du, R., Zhang, Y., Yang, H., Liu, J., … Xiong, H. (2016). Clinical and molecular genetic analysis of a family with late-onset *LAMA2*-related muscular dystrophy. *Brain & Development*, 38(2), 242–249. https://doi.org/10.1016/j.braindev.2015.08.005

Durbeej, M. (2015). Laminin-α2 chain-deficient congenital muscular dystrophy: Pathophysiology and development of treatment. *Current Topic in Membranes*, 76, 31–60. https://doi.org/10.1016/bs.ctm.2015.05.002

Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F., & den Dunnen, J. T. (2011). LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*, 32(5), 557–563. https://doi.org/10.1002/humu.21438

Gao, Y., Mruk, D., Chen, H., Lui, W. Y., Lee, W. M., & Cheng, C. Y. (2017). Regulation of the blood-testis barrier by a local axis in the testis: Role of laminin α2 in the basement membrane. *FASEB Journal*, 31(2), 584–597. https://doi.org/10.1096/fj.201600870R

Gavassini, B. F., Carboni, N., Nielsen, J. E., Danielsen, E. R., Thomsen, C., Svenstrup, K., … Pegoraro, E. (2011). Clinical and molecular characterization of limb-girdle muscular dystrophy due to *LAMA2* mutations. *Muscle & Nerve*, 44(5), 703–709. https://doi.org/10.1002/mus.22132

Geranmayeh, F., Clement, E., Feng, L. H., Sewry, C., Pagan, J., Mein, R., … Muntoni, F. (2010). Genotype-phenotype correlation in a large population of muscular dystrophy patients with *LAMA2* mutations. *Neuromuscular Disorders*, 20(4), 241–250. https://doi.org/10.1016/j.nmd.2010.02.001

Häger, M., Gawlik, K., Nyström, A., Sasaki, T., & Durbeej, M. (2005). Laminin {alpha}1 chain corrects male infertility caused by absence of laminin {alpha}2 chain. *The American Journal of Pathology*, 167(3), 823–833.

Harris, E., McEntagart, M., Topf, A., Lochmüller, H., Bushby, K., Sewry, C., … Straub, V. (2017). Clinical and neuroimaging findings in two brothers with limb girdle muscular dystrophy due to *LAMA2* mutations. *Neuromuscular Disorders*, 27(2), 170–174. https://doi.org/10.1016/j.nmd.2016.10.009

Helbling-Leclerc, A., Zhang, X., Topaloglu, H., Cruaud, C., Tesson, F., Weissenbach, J., … Tryggvason, K. (1995). Mutations in the laminin alpha 2-chain gene (*LAMA2*) cause merosin-deficient congenital muscular dystrophy. *Nature Genetics*, 11(2), 216–218.

Hussain, S. A., Carafoli, F., & Hohenester, E. (2011). Determinants of laminin polymerization revealed by the structure of the α5 chain amino-terminal region. *EMBO Reports*, 12(3), 276–282. https://doi.org/10.1038/embor.2011

Jones, J. C., Dehart, G. W., Gonzales, M., & Goldfinger, L. E. (2000). Laminins: An overview. *Microscopy Research and Technique*, 51(3), 211–213. https://doi.org/10.1002/1097-0029(20001101)51:3<211::AID-JEMT1>3.0.CO;2-P

Jones, K. J., Morgan, G., Johnston, H., Tobias, V., Ouvrier, R. A., Wilkinson, I., … North, K. N. (2001). The expanding phenotype of laminin alpha2 chain (merosin) abnormalities: Case series and review. *Journal of Medics Genetics*, 38(10), 649–657.

Kevelam, S. H., van Engelen, B. G., van Berkel, C. G., Küsters, B., & van der Knaap, M. S. (2014). *LAMA2* mutations in adult-onset muscular dystrophy with leukoencephalopathy. *Muscle & Nerve*, 49(4), 616–617. https://doi.org/10.1002/mus.24147

Kim, M. W., Jang, D. H., Kang, J., Lee, S., Joo, S. Y., Jang, J. H., … Lee, J. H. (2017). Novel Mutation (c.8725T>C) in two siblings with late-onset *LAMA2*-related muscular dystrophy. *Annals of Laboratory Medicine*, 37(4), 359–361. https://doi.org/10.3343/alm.2017.37.4.359

Kobayashi, K., Nakahori, Y., Mizuno, K., Miyake, M., Kumagai, T., Honma, A., … Toda, T. (1998). Founder-haplotype analysis in Fukuyama-type congenital muscular dystrophy (FCMD). *Human Genetics*, 103(3), 323–327.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7), 1073–1081. https://doi.org/10.1038/nprot.2009.86

Lamer, S., Carlier, R. Y., Pinard, J. M., Mompoint, D., Bagard, C., Burdairon, E., … Vallée, C. (1998). Congenital muscular dystrophy: Use of brain MR imaging findings to predict merosin deficiency. *Radiology*, 206(3), 811–816.

Løkken, N., Born, A. P., Duno, M., & Vissing, J. (2015). *LAMA2*-related myopathy: Frequency among congenital and limb-girdle muscular dystrophies. *Muscle & Nerve*, 52(4), 547–553. https://doi.org/10.1002/mus.24588

Marques, J., Duarte, S. T., Costa, S., Jacinto, S., Oliveira, J., Oliveira, M. E., … Evangelista, T. (2014). Atypical phenotype in two patients with *LAMA2* mutations. *Neuromuscular Disorders*, 24(5), 419–424. https://doi.org/10.1016/j.nmd.2014.01.004

Martinello, F., Angelini, C., & Trevisan, C. P. (1998). Congenital muscular dystrophy with partial merosin deficiency and late onset epilepsy. *European Neurology*, 40(1), 37–45. https://doi.org/10.1159/000007954

McKee, K. K., Crosson, S. C., Meinen, S., Reinhard, J. R., Rüegg, M. A., & Yurchenco, P. D. (2017). Chimeric protein repair of laminin polymerization ameliorates muscular dystrophy phenotype. *The Journal of Clinical Investigation*, 127(3), 1075–1089. https://doi.org/10.1172/JCI90854.

Menezes, M. J., McClenahan, F. K., Leiton, C. V., Aranmolate, A., Shan, X., & Colognato, H. (2014). The extracellular matrix protein laminin α2 regulates the maturation and function of the blood-brain barrier. *The Journal of Neuroscience*, 34(46), 15260–15280. https://doi.org/10.1523/JNEUROSCI.3678-13.2014

Mercuri, E., Gruter-Andrew, J., Philpot, J., Sewry, C., Counsell, S., Henderson, S., … Muntoni, F. (1999). Cognitive abilities in children with congenital muscular dystrophy: Correlation with brain MRI and merosin status. *Neuromuscular Disorders*, 9(6-7), 383–387.

Mora, M., Moroni, I., Uziel, G., di Blasi, C., Barresi, R., Farina, L., … Morandi, L. (1996). Mild clinical phenotype in a 12-year-old boy with partial merosin deficiency and central and peripheral nervous system abnormalities. *Neuromuscular Disorders*, 6(5), 377–381.

Nelson, I., Stojkovic, T., Allamand, V., Leturcq, F., Bécane, H. M., Babuty, D., … Bonne, G. (2015). Laminin α2 deficiency-related muscular dystrophy mimicking Emery-Dreifuss and collagen VI related diseases. *Journal of Neuromuscular Diseases*, 2(3), 229–240. https://doi.org/10.3233/JND-150093

Niroula, A., & Vihinen, M. (2016). Variation interpretation predictors: Principles, types, performance, and choice. *Human Mutation*, 37(6), 579–597. https://doi.org/10.1002/humu.22987

Nissinen, M., Helbling-Leclerc, A., Zhang, X., Evangelista, T., Topaloglu, H., Cruaud, C., … Guicheney, P. (1996). Substitution of a conserved cysteine-996 in a cysteine-rich motif of the laminin alpha2-chain in congenital muscular dystrophy with partial deficiency of the protein. *American Journal of Human Genetics*, 58(6), 1177–1784.

Oliveira, J., Gonçalves, A., Oliveira, M. E., Fineza, I., Pavanello, R. C., Vainzof, M., … Sousa, M. (2014). Reviewing large *LAMA2* deletions and duplications in congenital muscular dystrophy patients. *Journal of Neuromuscular Diseases*, 1(2), 169–179. https://doi.org/10.3233/JND-140031

Oliveira, J., Martins, M., Pinto Leite, R., Sousa, M., & Santos, R. (2017). The new neuromuscular disease related with defects in the ASC-1 complex:

Report of a second case confirms ASCC1 involvement. *Clinical Genetics*, *92*(4), 434–439. https://doi.org/10.1111/cge.12997

Oliveira, J., Santos, R., Soares-Silva, I., Jorge, P., Vieira, E., Oliveira, M. E., … Bronze-da-Rocha, E. (2008). *LAMA2* gene analysis in a cohort of 26 congenital muscular dystrophy patients. *Clinical Genetics*, *74*(6), 502–512. https://doi.org/10.1111/j.1399-0004.2008.01068.x

Pegoraro, E., Marks, H., Garcia, C. A., Crawford, T., Mancias, P., Connolly, A. M., … Hoffman, E. P. (1998). Laminin alpha2 muscular dystrophy: Genotype/phenotype studies of 22 patients. *Neurology*, *51*(1), 101–110.

Pegoraro, E., Fanin, M., Trevisan, C. P., Angelini, C., Hoffman, E. P. (2000). A novel laminin alpha2 isoform in severe laminin alpha2 deficient congenital muscular dystrophy. *Neurology*, *55*(8), 1128–1134.

Pejaver, V., Mooney, S. D., & Radivojac, P. (2017). Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. *Human Mutation*, *38*(9), 1092–1108. https://doi.org/10.1002/humu.23258

Philpot, J., Cowan, F., Pennock, J., Sewry, C., Dubowitz, V., Bydder, G., … Muntoni F. (1999). Merosin-deficient congenital muscular dystrophy: The spectrum of brain involvement on magnetic resonance imaging. *Neuromuscular Disorders*, *9*(2), 81–85.

Pini, A., Merlini, L., Tomé, F. M., Chevallay, M., & Gobbi, G. (1996). Merosin-negative congenital muscular dystrophy, occipital epilepsy with periodic spasms and focal cortical dysplasia. Report of three Italian cases in two families. *Brain & Development*, *18*(4), 316–322.

Quijano-Roy, S., Renault, F., Romero, N., Guicheney, P., Fardeau, M., & Estournet, B. (2004). EMG and nerve conduction studies in children with congenital muscular dystrophy. *Muscle & Nerve*, *29*(2), 292–299. https://doi.org/10.1002/mus.10544

Quijano-Roy, S., Sparks, S. E., & Rutkowski, A. (2012). *LAMA2*-related muscular dystrophy. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews®* [Internet]. Seattle, WA: University of Washington. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK97333/

Rajakulendran, S., Parton, M., Holton, J. L., & Hanna, M. G. (2011). Clinical and pathological heterogeneity in late-onset partial merosin deficiency. *Muscle & Nerve*, *44*(4), 590–593. https://doi.org/10.1002/mus.22196

Reinhard, J. R., Lin, S., McKee, K. K., Meinen, S., Crosson, S. C., Sury, M., … Rüegg, M. A. (2017). Linker proteins restore basement membrane and correct *LAMA2*-related muscular dystrophy in mice. *Science Translational Medicine*, *9*(396). https://doi.org/10.1126/scitranslmed.aal4649

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … Rehm, H. L. (2015). ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetic Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Salgado, D., Desvignes, J. P., Rai, G., Blanchard, A., Miltgen, M., Pinard, A., … Béroud, C. (2016). UMD-predictor: A high-throughput sequencing compliant system for pathogenicity prediction of any Human cDNA substitution. *Human Mutation*, *37*(5), 439–446. https://doi.org/10.1002/humu.22965

Sewry, C. A., Philpot, J., Sorokin, L. M., Wilson, L. A., Naom, I., Goodwin, F., … Muntoni, F. (1996). Diagnosis of merosin (laminin-2) deficient congenital muscular dystrophy by skin biopsy. *Lancet*, *347*(9001), 582–584.

Shorer, Z., Philpot, J., Muntoni, F., Sewry, C., & Dubowitz, V. (1995). Demyelinating peripheral neuropathy in merosin-deficient congenital muscular dystrophy. *Journal of Child Neurology*, *10*(6), 472–475. https://doi.org/10.1177/088307389501000610

Sunada, Y., Edgar, T. S., Lotz, B. P., Rust, R. S., & Campbell, K. P. (1995). Merosin-negative congenital muscular dystrophy associated with extensive brain abnormalities. *Neurology*, *45*(11), 2084–2089.

Tomé, F. M., Evangelista, T., Leclerc, A., Sunada, Y., Manole, E., Estournet, B., … Fardeau, M. (1994). Congenital muscular dystrophy with merosin deficiency. *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie*, *317*(4), 351–357.

Tsao, C. Y., Mendell, J. R., Rusin, J., & Luquette, M. (1998). Congenital muscular dystrophy with complete laminin-alpha2-deficiency, cortical dysplasia, and cerebral white-matter changes in children. *Journal of Child Neurology*, *13*(6), 253–256. https://doi.org/10.1177/088307389801300602

Vihinen, M., den Dunnen, J. T., Dalgleish, R., & Cotton, R. G. (2012). Guidelines for establishing locus specific databases. *Human Mutation*, *33*(2), 298–305. https://doi.org/10.1002/humu.21646

Vigliano, P., Dassi, P., Di Blasi, C., Mora, M., & Jarre, L. (2009). *LAMA2* stop-codon mutation: Merosin-deficient congenital muscular dystrophy with occipital polymicrogyria, epilepsy and psychomotor regression. *European Journal of Paediatric Neurology*, *13*(1), 72–76. https://doi.org/10.1016/j.ejpn.2008.01.010

Wood, A. J., & Currie, P. D. (2014). Analysing regenerative potential in zebrafish models of congenital muscular dystrophy. *The International Journal of Biochemistry & Cell Biology*, *56*, 30–37. https://doi.org/10.1016/j.biocel.2014.10.021

Xiong, H., Tan, D., Wang, S., Song, S., Yang, H., Gao, K., … Wu, X. (2015). Genotype/phenotype analysis in Chinese laminin-α2 deficient congenital muscular dystrophy patients. *Clinical Genetics*, *87*(3), 233–243. https://doi.org/10.1111/cge.12366

Yurchenco, P. D. (2015). Integrating activities of laminins that drive basement membrane assembly and function. *Current Topic in Membranes*, *76*, 1–30. https://doi.org/10.1016/bs.ctm.2015.05.001

Zhang, X., Vuolteenaho, R., & Tryggvason, K. (1996). Structure of the human laminin alpha2-chain gene (*LAMA2*), which is affected in congenital muscular dystrophy. *The Journal of Biological Chemistry*, *271*(44), 27664–27669.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

npg

ORIGINAL ARTICLE

New massive parallel sequencing approach improves the genetic characterization of congenital myopathies

Jorge Oliveira[1,2], Ana Gonçalves[1,2], Ricardo Taipa[3], Manuel Melo-Pires[3], Márcia E Oliveira[1,2],
José Luís Costa[4,5], José Carlos Machado[4,5], Elmira Medeiros[6], Teresa Coelho[7], Manuela Santos[8],
Rosário Santos[1,2,9,12] and Mário Sousa[2,10,11,12]

Congenital myopathies (CMs) are a heterogeneous group of muscle diseases characterized by hypotonia, delayed motor skills and muscle weakness with onset during the first years of life. The diagnostic workup of CM is highly dependent on the interpretation of the muscle histology, where typical pathognomonic findings are suggestive of a CM but are not necessarily gene specific. Over 20 loci have been linked to these myopathies, including three exceptionally large genes (*TTN*, *NEB* and *RYR1*), which are a challenge for molecular diagnosis. We developed a new approach using massive parallel sequencing (MPS) technology to simultaneously analyze 20 genes linked to CMs. Assay design was based on the Ion AmpliSeq strategy and sequencing runs were performed on an Ion PGM system. A total of 12 patients were analyzed in this study. Among the 2534 variants detected, 14 pathogenic mutations were successfully identified in the *DNM2*, *NEB*, *RYR1*, *SEPN1* and *TTN* genes. Most of these had not been documented and/or fully characterized, hereby contributing to expand the CM mutational spectrum. The utility of this approach was demonstrated by the identification of mutations in 70% of the patients included in this study, which is relevant for CMs especially considering its wide phenotypic and genetic heterogeneity.
Journal of Human Genetics advance online publication, 4 February 2016; doi:10.1038/jhg.2016.2

INTRODUCTION

Congenital myopathies (CMs) are a highly heterogeneous and continuously expanding, group of muscle diseases with an estimated incidence of around 6 per 100 000 live births.[1] On the severest end of the disease spectrum, CMs are characterized by muscle weakness and hypotonia with neonatal onset or during the first years of life, which often give rise to respiratory and feeding difficulties.[1,2] Typical features may include weakness of facial and bulbar muscles, and also involvement of extraocular muscles (ophthalmoparesis or ophthalmoplegia).[3] Typical forms of CM may also present during childhood as motor development delay and/or waddling gait.[1] At the other end of the scale, it is now clear that some patients with subtle congenital muscle weakness might remain undiagnosed until adulthood, when muscle strength deteriorates or respiratory insufficiency settles in.[1,4]

The classification and the diagnostic workup of CMs are highly reliant upon the identification of distinct structural abnormalities detected in muscle biopsies.[3,5] Based on pathognomonic findings CMs

are subdivided into four major groups: (i) CMs with rods in nemaline myopathy (NM), (ii) CMs with cores, which includes central core disease, multiminicore disease and multicore myopathy, (iii) CMs with central nuclei (centronuclear myopathy) and (iv) congenital fiber-type size disproportion.[5] Other less common subtypes such as cap or core-rod myopathies, which might represent overlapping myopathological entities, have been also reported.[6,7] Although these findings are of extreme diagnostic value, in some patients the muscle biopsy may display only unspecific myopathic features such as type 1 fiber predominance/uniformity.[8]

More than 20 different genes are currently known to be associated with CMs (http://musclegenetable.fr/, August 2015). The majority of these loci encode basic structural components of the sarcomere or proteins involved in calcium homeostasis, both crucial for normal muscle structure and function.[1] Other genetic defects give rise to abnormal triad structure due to aberrant tubulogenesis and/or abnormal membrane recycling.[9] Among these loci there are three particularly large genes: Titin (*TTN*) with 363 exons, Nebulin (*NEB*)

[1]Unidade de Genética Molecular, Centro de Genética Médica Dr Jacinto Magalhães, Centro Hospitalar do Porto, Porto, Portugal; [2]Unidade Multidisciplinar de Investigação Biomédica (UMIB), Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal; [3]Unidade de Neuropatologia, Centro Hospitalar do Porto, Porto, Portugal; [4]Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal; [5]Faculdade de Medicina da Universidade do Porto, Porto, Portugal; [6]Departamento de Neurologia, Hospital Egas Moniz, Centro Hospitalar de Lisboa Ocidental, Lisboa, Portugal; [7]Unidade Clínica de Paramiloidose, Centro Hospitalar do Porto, Porto, Portugal; [8]Consulta de Doenças Neuromusculares, Serviço de Neuropediatria, Centro Hospitalar do Porto, Porto, Portugal; [9]UCIBIO/REQUIMTE, Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal; [10]Departamento de Microscopia, Laboratório de Biologia Celular, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal and [11]Centro de Genética da Reprodução Prof. Alberto Barros, Porto, Portugal
[12]These authors contributed equally to this work.
Correspondence: Dr R Santos, Unidade de Genética Molecular, Centro de Genética Médica Dr Jacinto de Magalhães, Centro Hospitalar do Porto (CHP), Praça Pedro Nunes, 88, Porto 4099-028, Portugal.
E-mail: rosario.santos@chporto.min-saude.pt
Received 9 September 2015; revised 28 November 2015; accepted 5 January 2016

with 183 and the Ryanodine receptor 1 (*RYR1*) with 106. Consequently, conventional Sanger sequencing of these genes is extremely laborious and costly. These aspects might explain the lack of thorough studies for large genes such as *TTN* and the limited number of patients reported in the literature.[10] The genetic study workup is complex considering that the same clinical entity can be caused by mutations in different genes, as may the same defective gene give rise to distinct myopathies. *RYR1*-related myopathies are among the best examples. Nearly all subtypes of CMs (central core disease, centronuclear myopathy, congenital fiber-type size disproportion and core-rod myopathy) have been reported as linked to *RYR1* mutations.[3,4] In addition, although muscle histology is important for the diagnostic workup of CMs, pathognomonic findings (such as rods in NM) are not necessarily gene specific.[11] The considerable genetic and clinical heterogeneity is reflected by the large number of genetic studies reported to be inconclusive. Essentially owing to these challenging diagnostic and technical difficulties, a significant proportion of CM patients are still genetically unsolved.[2]

With the advent of the novel massive parallel sequencing (MPS) technologies, we have developed a new targeted resequencing approach, which allows the simultaneous analysis of 20 genes implicated in CMs. Besides contributing to expand the mutational spectrum of CMs, our study enabled the identification of mutations in 7 CM patients (out of 10 who were prospectively studied), thereby demonstrating its clinical utility.

MATERIALS AND METHODS

Patients
A total of 12 patients were included in this study. Two patients with mutations previously identified by conventional sequencing were used for assay validation. The first patient used as a positive control (C1) was diagnosed with a centronuclear myopathy due to a heterozygous mutation (c.1393C>T, p.Arg465Trp) in the *DNM2* gene. The second patient (C2) has a mild core myopathy with adulthood onset, caused by two heterozygous *RYR1* missense mutations (patient 1 in Duarte *et al.*[12]).

For the prospective part of this study, 10 genetically uncharacterized patients (P1–P10) were selected based on features compatible with CMs, namely (i) early disease onset (neonatal or up to early childhood) and (ii) muscle histology suggestive of CMs or with structural defects. Eight of these patients had structural changes on muscle biopsy: three with central nuclei (P1B, P2 and P8), three with rods (P3, P5 and P7) and two with minicores (P4 and P6). The remaining two patients (P9 and P10) had unspecific myopathic changes seen on muscle biopsy, mainly type 1 fiber predominance. Parents and other family members, when available, were studied to demonstrate compound heterozygosity in autosomal recessive cases and/or to ascertain the parental origin of the mutations. This research work was approved by the ethics committee of Centro Hospitalar do Porto.

Muscle histology analysis
The routine diagnostic workup consisted of an open biopsy (deltoid, quadriceps or gastrocnemius muscle) and histological (hematoxylin and eosin stain, periodic acid-Schif and gomori trichrome) and histochemical (reduced NADH, succinic dehydrogenase, cytochrome oxidase and adenosine triphosphatase) studies performed on frozen material. Semi-thin sections at 1 μm from resin embedded muscle were also routinely used.

AmpliSeq assay design
The first task consisted in the development of a targeted resequencing approach based on semiconductor technology to simultaneous analyze 20 genes linked to CMs, namely *ACTA1, BIN1, CFL2, CNTN1, DNM2, KBTBD13, KLHL40, MEGF10, MTM1, MYBPC3, MYH2, MYH7, NEB, RYR1, SEPN1, TNNT1, TPM2, TPM3, TRIM32* and *TTN* (Supplementary Data I). Assay design was performed using the Ion AmpliSeq Designer software (Life Technologies, Foster

City, CA, USA) pipeline 2.2.1. Based on a list of representative transcripts for these loci (Supplementary Data I), target regions were selected from the University of California Santa Cruz, California, USA (UCSC) Genome Browser (http://genome.ucsc.edu/index.html). Targets included exonic regions (with 50 bp into flanking introns) and untranslated regions. The obtained custom Ampliseq assay covered 92.6% of these regions comprising a total of ~ 320 thousand base pairs (bp) and consisted of 2077 different amplicons (ranging from 64 to 239 bp, 174 bp on average), divided into two independent primer pools for multiplex PCR.

Library and template preparation
Sample quality of patient genomic DNA was evaluated by gel electrophoresis and quantified using Qubit dsDNA HS Assay Kit (Life Technologies). A total of 10 ng of genomic DNA were used in each multiplex PCR reaction. Library amplifications, digestion, bar-coding and purification was performed according to the Ion AmpliSeq Library Kit version 2.0 (Life Technologies) instructions. Library quantification was performed using the Qubit dsDNA HS Assay. All libraries were diluted to the same concentration and pooled to ensure an equal representation of the different samples.

The diluted and combined libraries were subjected to amplification by emulsion PCR using Ion Personal Genome Machine (PGM) Template OT 2 200 Kit (Life Technologies) and prepared on an Ion OneTouch 2 Instrument (Life Technologies) according to the manufacturer's protocol. Enrichment of template Ion Sphere particles was performed using the Ion OneTouch 2 enrichment system (Life Technologies).

Semiconductor sequencing
Sequencing was carried out on a PGM system platform based on semiconductor technology.[13] Two independent experiments were performed, using the Ion 316 and 318 chips, allowing four or eight bar-coded samples, respectively. The Ion Sequencing Kit v2.0 (Life Technologies) was used to perform sequencing runs, following the manufacturer's recommended protocols. Torrent Server version 4.0.2 was used to obtain the basic run metrics and to generate Binary Alignment Map and Variant Caller Format files. Torrent Variant Caller was set for germline – low stringency – calls.

Data analysis and interpretation
Variant annotation and filtering steps were performed using Ion Reporter v4.2 software (http://ionreporter.lifetechnologies.com/) (Life Technologies). Variant Caller Format files from each patient were added to Ion Reporter and were annotated using the 'annotate variants single sample' workflow incorporated in this software. A filter chain was created to restrict the number of variants, which consisted in the removal of single-nucleotide polymorphisms (SNPs) present on the UCSC common SNPs list (those with at least 1% minor allele frequency). Alamut Visual v2.4 software (Interactive Biosoftware, Rouen, France) was used to assist variant analysis and interpretation. The software also incorporates algorithms to evaluate the impact of missense mutations and the effect on splicing, and allows the simultaneous visualization of Variant Caller Format and Binary Alignment Map files. GenomeBrowse v2.0.0 (Golden Helix, Bozeman, MT, USA) was used for visual inspection of Binary Alignment Map files.

Variant validation and expression analysis
Clinically relevant variants (classified as pathogenic) were confirmed by a new PCR using in-house designed primers (sequences in Supplementary Data II) and the resulting amplicons sequenced by the Sanger method.

Variants possibly affecting splicing were evaluated by expression studies at the mRNA level. Total RNA was obtained from cryopreserved muscle samples (P4 and P7) using the PerfectPure RNA Fibrous Tissue kit (5 PRIME, Hilden, Germany) or from whole blood (patient P3) extracted with the PerfectPure RNA Blood Kit (5 PRIME). After conversion to cDNA with the High Capacity cDNA Reverse Transcription kit (Life Technologies), these samples were amplified by PCR using specific primers annealing to different regions of *TTN* and *NEB* transcripts (primers sequences in Supplementary Data II). PCR products were purified with Illustra ExoStar (GE Healthcare, Little Chalfont, UK). A new PCR was prepared with BigDye Terminator kit V3.1 chemistry

(Life Technologies). Sequencing reactions were run on an ABI 3130xl genetic analyser (Life Technologies).

RESULTS

Development of a new sequencing approach for CMs

The initial task was the development of a new gene panel based on MPS, designed using the Ion AmpliSeq software. An overall coverage of 95% was obtained for the genomic coding sequence (CDS) that was considered suitable for this research. The CDS coverage of each locus ranged between 77 and 100% with the majority of the genes ($n = 15$) having a coverage above 90% (Supplementary Data I). For the remaining loci, all except the *NEB* gene could be explained by the presence of high GC content regions, having higher impact in smaller genes. These regions are amenable to filling in by Sanger sequencing. In the case of *NEB* the lower coverage (85%) is explained by the existence of a triplicated region (three almost identical sequence blocks) encompassing exons 82–89, 90–97 and 98–105. Since the majority of MPS approaches are based on the alignment of short sequences, the obtained reads and variants will not be mapped with precision.

The sequencing runs performed on the total of 12 samples generated 6.5 million sequence reads, with an average coverage depth of $257x$ (Supplementary Data III) and with 97.0% of the target regions successfully covered. A total of 2535 sequence variants were called: 2348 single-nucleotide variants and 187 insertion/deletions (INDEL), corresponding to an average of 211.4 variants per individual (Supplementary Data III).

The high number of variants obtained in this work required the development of a computational filtering strategy to facilitate the analytic process and to concentrate on the variants that most likely correlated with the patients' phenotypes. Single-nucleotide variants listed as common SNPs in the UCSC database were excluded from further analysis. As a result, 86% of the detected variants were filtered out, meaning that an average of ∼ 29 variants per patient remained to be analyzed in detail. Data interpretation implied evaluating: (i) the variant predictable impact, (ii) the frequency in publically available genetic variant databases (dbSNP and exome variant server (EVS)) and (iii) the data available in locus specific databases (Leiden Muscular Dystrophy Pages).

A total of 14 pathogenic variants listed in Table 1 were detected in nine patients whose clinical data are shown in Table 2. The three missense mutations previously identified by conventional sequencing in the two positive controls (C1 and C2) were successfully detected. The remaining 11 mutations were identified in seven patients of the prospective study; these were located in *RYR1* ($n = 4$), *NEB* ($n = 4$) and *TTN* ($n = 3$). One additional heterozygous mutation in the selenoprotein N 1 gene (*SEPN1*) was also identified. In this case, and so as to exclude the presence of a second pathogenic variant, this gene was completely resequenced by the Sanger method. In one of the patients (P2) with a novel heterozygous nonsense mutation in *RYR1*, gene analysis was complemented by conventional sequencing, leading to the identification of a second heterozygous mutation located in a region not covered by the assay (part of exon 101).

In this work, a total of 67 variants (the majority intronic or located in untranslated regions) were found to be unclassified variants; these included both ultra-rare sequence changes (reported in genetic databases with a frequency bellow 0.1%) and novel variants. Finally, 43 known polymorphisms were also identified, reported in the EVS database with frequencies higher than 0.1% and thus not listed in the USCS common SNP table.

Difficulties and potential pitfalls

During the validation of the MPS gene panel for CMs, the zygosity of one of the mutations (chr19:g.38986918C>G, c.6612C>G) in positive control C2 was not correctly established. After reviewing Sanger sequencing data and the MPS alignment, a rare SNP (c.6549-51C>T) was identified in the intronic region where the respective MPS forward primer anneals. This variant prevented the amplification of the normal allele and caused a PCR allele dropout, thereby explaining the apparent genotype discrepancy (Supplementary Data IV).

It is noteworthy that during the optimization of the analytical process, an intronic mutation (chr2:g.152417626_152417632del) was initially filtered out. The reference SNP (rs) cluster identification of a known variant was attributed to this 7 bp deletion during the annotation performed by Ion Reporter (Supplementary Data IV). By default the Ion Reporter algorithm associates variants to the rs number based on the genomic coordinates involved rather than the specific variant itself. As a consequence, if a deletion or duplication coincides with the genomic position of a known SNP, the software attributes the same rs number. Considering this potential pitfall, all filtered-out INDEL variants were manually rechecked.

About 9% of the total numbers of variants called were classified as false positives. These false-positive variants resulted from: (i) sequencing artifacts in homopolymeric regions (consistently found in several samples tested), (ii) variants in shorter reads due to mis-priming events (similar to those reported by McCall et al.[14] and usually showing a biased proportion of mutated/normal reads) and to a lesser extent (iii) variants located in regions with low coverage depth.

RYR1-related myopathies

Three patients (P1, P1B and P2) from two families were identified has having an autosomal recessive congenital myopathy due to defects in the ryanodine receptor 1 gene (Figure 1). All three patients had marked hypotonia and respiratory distress during the neonatal period but there was some degree of clinical variability. The two brothers from the first family are both wheelchair dependent (P1 never walked) and have a severe tetraparesis, whereas patient P2 achieved independent locomotion at 3.5 years of age and currently (at 28 years of age) walks without support. The patients' muscle biopsies showed considerable fiber size variability and frequent fibers with abnormal centrally placed nuclei – pathognomonic features for a central nuclear myopathy (Figure 1a). Four novel mutations were identified in *RYR1* (Figure 1b); these comprised two nonsense mutations (p.Gln4004* and p.Arg3053*) and two missense mutations (p.Met4881Ile and p.Pro1267Arg) affecting highly conserved residues. Bioinformatic analysis resorting to two distinct algorithms corroborated the pathogenic impact of these missense variants. The genotype was concordant between all three cases, since there is compound heterozygosity of a missense and a nonsense mutation (Figure 1c).

NEB-related myopathies

Patients P3, P5 and P7 were seen to have rods in their muscle biopsy, evocative of an NM (Figure 2a). Interestingly, in the first biopsy of P5 there were features compatible with congenital fiber-type size disproportion, with a focal area showing structures resembling rods. A second biopsy performed 2 years later showed larger areas and number of the typical rods. In terms of their clinical presentation, phenotypes were rather dissimilar, varying from a very mild distal myopathy diagnosed during adulthood to a congenital form of NM with bilateral club foot and requiring non-invasive ventilation.

Table 1 Pathogenic sequence variants detected by the massive parallel sequencing gene panel

Patient Id/ gender	Gene/accession number	Gene location (exon/ intron)	Genomic cDNA DNA change (zygosity)	Protein	RNA change	Mutation type	Mutation origin	Family history	Frequency in ESP	Mutation reference
C1 M	DNM2 NM_001005360.2	11	Chr19:g.10909219C>T c.1393C>T (heterozygous)	p.Arg465Trp	r.(?)	Missense	De novo? (unaffected parents were not tested)	Sporadic	N.p.	Bitoun et al.[32]
C2 F	RYR1 NM_000540.2	40	Chr19:g.38986918C>G c.6612C>G (heterozygous) (*)	p.His2204Gln	r.(?)	Missense	Inherited (parent #1)	Affected brother (same genotype)	N.p.	Duarte et al.[12]
		98	Chr19:g.39068613G>A c.14228G>A (heterozygous)	p.Gly4743Asp	r.(?)	Missense	Inherited (parent #2)		N.p.	Duarte et al.[12]
P1 F	RYR1 NM_000540.2	87	Chr19:g.39034513C>T NM_000540.2:c.12010C>T (heterozygous)	p.Gln4004 (*)	r.(?)	Nonsense	Inherited (maternal)	Affected brother (P1B, with the same genotype)	N.p.	New
		101	c.14643G>A (**) (heterozygous)	p.Met4881Ile	r.(?)	Missense	Inherited (paternal)		N.p.	New
P2 F	RYR1 NM_000540.2	61	Chr19:g.39002235C>T c.9157C>T (heterozygous)	p.Arg3053 (*)	r.(?)	Nonsense	Inherited? (paternal inferred)	Sporadic	N.p.	New
		28	Chr19:g.38964051C>G c.3800C>G (heterozygous)	p.Pro1267Arg	r.(?)	Missense	Inherited (maternal)		1/13006	New
P3 M	NEB NM_001271208.1	129	Chr2:g.152408252C>T c.19944G>A (homozygous)	p.Asn6649_Ile8560delins27	r.[19944g>a; 19944_19445ins19944 +1_19944+120, =]	Affects splicing	Inherited? (untested parents)	Affected brother (not studied)	N.p.	Lehtokari et al.[15] /new (splicing effect)
P4 M	TTN NM_001267550.2	Intron 273	Chr2:g.179473930C>A c.52102+5G>A (heterozygous)	p.Gly17345_Leu17367del	r.52034_52102del	Affects splicing	Inherited (maternal)	Affected sib (same genotype, deceased in the neonatal period)	N.p.	New
		326	Chr2:g.179434749_179434750del c.76109_76110del (heterozygous)	p.Ile25370Argfs*6	r.(?)	Frameshift deletion	Inherited (paternal)		N.p.	New
P5 M	NEB NM_001271208.1	171	Chr2:g.152354789_152354792dup c.24294_24297dup (homozygous)	p.Glu8100Serfs*5	r.(?)	Frameshift duplication	Inherited? (consanguineous parents)	Affected sister (not studied, deceased)	N.p.	Pelin et al.[16], Lehtokari et al.[17], Lehtokari et al.[15]
P6 M	TTN NM_001267550.2	304	Chr2: g.179454772_179454773insGG c.61679_61680insCC (heterozygous)	p.Ser20561Leufs*17	r.(?)	Frameshift insertion	De novo	Sporadic	N.p.	New
P7 M	NEB NM_001271208.1	Intron 122	Chr2:g.152417626_152417632del c.19102-4_19102-10del (heterozygous)	p.[Val6368_Ser6402del, Val6368_Ile8560delins25]	r.[19102_19206del, 19101_19102ins19101 +1_19102-1, =]	Affects splicing	Inherited? (maternal inferred)	Affected relative (distant cousin) (shares p.Arg7026* mutation)	N.p.	Lehtokari et al.[15] /new (splicing effect)
		140	Chr2:g.152394412G>A c.21076C>T (heterozygous)	p.Arg7026*	r.(?)	Nonsense	Inherited (paternal inferred)		N.p.	Lehtokari et al.[15]
	SEPN1 NM_020451.2	10	Chr1:g.26139280T>G c.1384T>G (heterozygous)	p.Scy462Gly	r.(?)	Selenocystein codon (TGA) loss	?		N.p.	Ferreiro et al.[33]

Abbreviations: C, control; ESP, NHLBI GO Exome Sequencing Project, data accessed through the exome variant server (http://evs.gs.washington.edu/EVS/); Id, identification; N.p., variant not present in the database; P, patient; (*), zygosity incorrectly determined due to allele dropout (see Results section); (**), mutation not detected by gene panel, region not included in the assay.

Table 2 Clinical data of the patients presented in this work

Patient/ gender	Age (years)	Onset	Hypotony	Muscle weakness	CK (IU/l)	Ambulant?	Facial involvement?	Cardiac involvement?	Respiratory insufficiency?	Scoliosis?
C1 M	20	Early childhood	N	Ophthalmoparesis; tetraparesis; weakness predominantly in lower limbs	47	Y	Y	N	N	N
C2 F (*)	43 (**)	Adulthood	N	Only myalgia	456	Y	N	NA	N	NA
P1 F	28	Neonatal	Y	Ophthalmoparesis; bulbar weakness; proximal tetraparesis; axial muscle weakness	NP	N (never walked)	Y	N	Y (refused NIV)	Y (corrected by surgery 13 yr)
P1B M	34	Neonatal	Y	Ophthalmoparesis; bulbar weakness; proximal tetraparesis; axial muscle weakness	18	N (never walked)	Y	N	Y (refused NIV)	Y (corrected by surgery 15 yr)
P2 F	23	Neonatal	Y	Ophthalmoparesis; bulbar weakness; proximal tetraparesis; axial muscle weakness	5	Y (3.5 yr)	Y	N	Y (NIV since 16 yr of age)	Y (corrected by surgery 12 yr)
P3 M	75	Early childhood	N	Severe and early bilateral foot drop; later with milder facial, cervical and proximal upper limb weakness	Norm.	Y	Y	N	N	N
P4 M	Deceased (at 10 yr of age)	Congenital (arthrogryposis)	Y	Facial and bulbar weakness; severe tetraparesis; axial muscle weakness	Norm.	N (never walked)	Y	Y (dilated cardi- omyopathy)	Y (IV during neonatal period)	Y (not corrected)
P5 M	16	Infancy	N	Early respiratory muscle involvement; facial and bulbar weakness; proximal tetraparesis and later distal involvement; axial muscle weakness	137	Y	Y	N	Y (NIV since 4 yr of age)	Y (corrected by surgery 15 yr)
P6 M	17	Infancy	N	Facial weakness; proximal tetraparesis	54	Y	Y	N	N	N
P7 M	12	Congenital (bilateral club foot)	Y	Facial diparesis and bulbar weakness; tetraparesis predominantly distal	NP	Y (2 yr)	Y	N	Y (NIV since 8 yr of age)	N

Abbreviations: F, female; IV, invasive ventilation; M, male; N, no; NA, not assessed; NIV, non-invasive ventilation; norm., within normal range; NP, not performed; Y, yes; yr, years; (*), reported in Duarte et al.[12]; (**), age at the time of publication.

The first patient (P3) is a 75-year-old man with a slowly progressive distal myopathy. His motor and cognitive development was described as normal. After his 60s there were complaints of proximal upper limb and cervical weakness. There is no respiratory or cardiac involvement. All the variants identified in the NEB gene of this patient were detected in homozygosity including one mutation (c.19944G>A) located in the last base of exon 129 (Figure 2b). Since no known parental consanguinity was reported, the possibility of a large deletion spanning the entire NEB gene was excluded by the analysis of the amplicon's normalized coverage depth (Supplementary Data V). The c.19944G>A mutation was previously reported as pathogenic and possibly affecting splicing,[15] which was corroborated by our bioinformatic analysis. However, until now, its effect had not been experimentally proven. This mutation causes the disruption of the canonical splice site (naturally occurring), and a cryptic splice site located within intron 129 is used instead (Figure 2c). The outcome of this splice-site change is a partial intronic retention (120 bp) culminating in a premature termination codon and a shortened nebulin protein (p.Asn6649_Ile8560delins27). The patient has a brother with similar muscle complaints but he was unavailable for study; no other family members were known to be affected.

The second NEB mutation identified in this work (c.24294_24297dup) was detected in patient P5 (Figure 2b), probably in homozygosity since his parents are first degree cousins. This frameshift duplication was previously found in heterozygosity in patients with NM.[15–17] The neonatal period of patient P5 was uneventful with only initial feeding difficulties reported. There was significant respiratory involvement starting at 2 years of age. Two years later the respiratory difficulties progressed, requiring non-invasive ventilation. Corrective surgery for scoliosis was performed at 15 years of age. He currently has a proximal tetraparesis and walks without

assistance. The patient had an older brother (deceased and without genetic studies) with a similar phenotype.

Patient P7 had the severest NM presentation of the three cases included in this work. He had a congenital disease onset, presenting at birth with hypotonia and predominantly distal muscle weakness with bilateral club foot. The patient required non-invasive ventilation since the age of 8. He has two heterozygous mutations in the NEB gene (Figure 2b). Although his parents were not studied, these mutations are likely to have distinct origins, since only one (p.Arg7026*) is in common with a distant cousin also affected with NM. This nonsense mutation has been previously identified in a third Portuguese patient with a (NEB-related) NM.[15] Surprisingly, this patient is reported to carry the same nonsense mutation as found in our patient P7 and the splicing mutation (c.19944G>A) identified in our patient P3. She was described as having a typical NM, which is milder than patient P7 but more severe than P3.

The other NEB mutation found in patient P7, a 7 bp deletion located at the end of intron 122 (c.19102-4_19102-10del), was also previously described[15] but not evaluated at the mRNA level. We demonstrated that it affects the NEB splicing process (Supplementary Data VI). Two abnormal transcripts were identified in muscle-derived total RNA: one is consequential of the exon 123 skipping (r.19102_19206del) and the other is the total retention of intron 122 (r.19101_19102ins19101+1_19102-1) in mature mRNA, predictably originating a shorter polypeptide. Patient P7 is also a carrier of a pathogenic variant (c.1384T>G, p.Scy462Gly) in the SEPN1 gene. Although it may not influence the patient's phenotype, this variant should be considered for genetic counseling purposes in this family.

TTN-related myopathies

Mutations in the TTN gene were identified in two unrelated cases (P4 and P6) (Figure 3). Patient P4 had a severe neonatal hypotonia
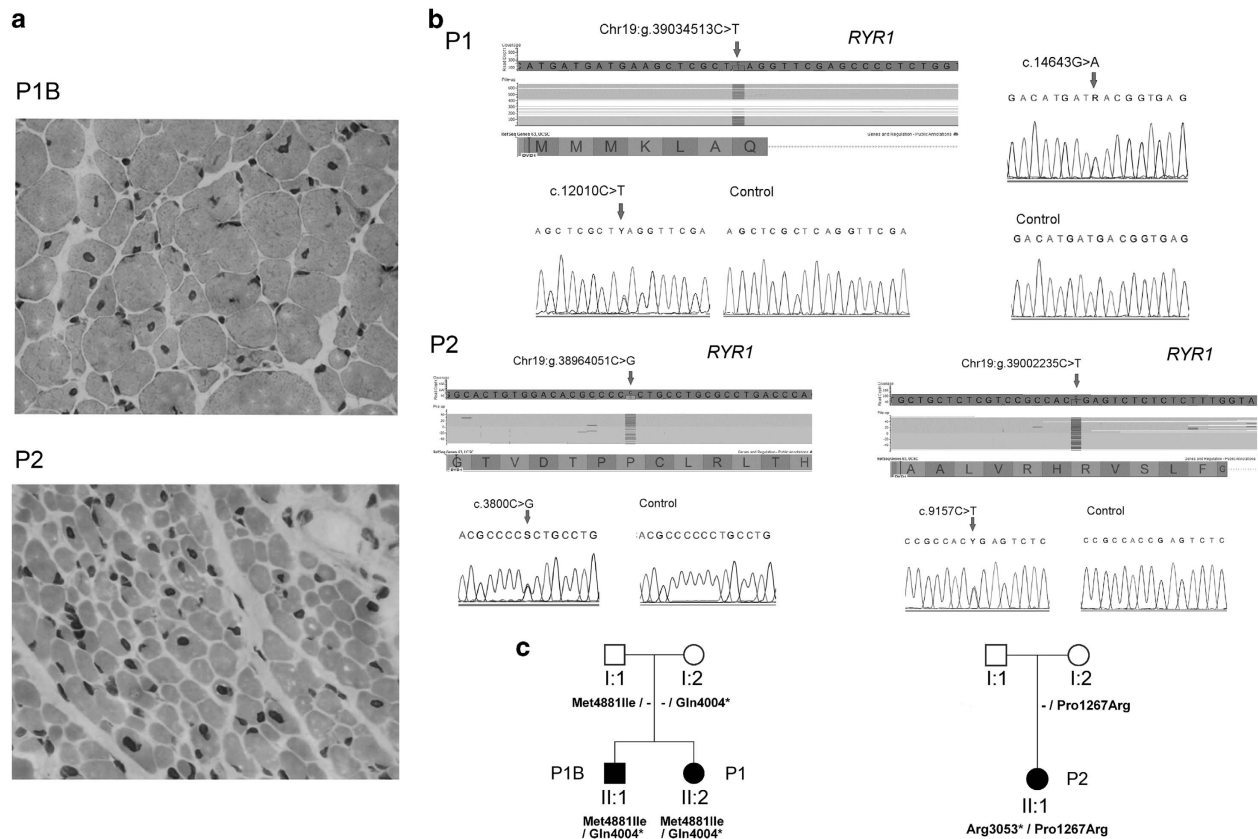
Figure 1 *RYR1*-related centronuclear myopathy. (a) Patients P1B and P2 histology of deltoid muscle biopsies (hematoxylin and eosin with x400 magnification). In both there is abnormal fiber size variability, with atrophic fibers and central nuclei. In patient P1B the majority of atrophic and hypotrophic fibers were type 1 fibers resembling congenital fiber-type size disproportion (not shown). (b) Heterozygous *RYR1* pathogenic variants identified in patient P1: (i) nonsense mutation c.12010C>T (p.Gln4004*) was identified by MPS (top part) and confirmed by Sanger sequencing (bottom); (ii) missense mutation c.14643G>A p.(Met4881Ile). Mutations identified in patient P2: (i) nonsense mutation c.9157C>T (p.Arg3053*) and (ii) missense mutation c.3800C>G (p.Pro1267Arg). (c) Patients' family pedigrees suggesting an autosomal inheritance pattern. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

with ophthalmoparesis, facial and bulbar weakness, dilated cardiomyopathy (cause of death at 10 years of age), requiring invasive ventilation and a feeding tube. The youngest of his two brothers had a similar phenotype and died during the neonatal period. P4's muscle biopsy showed myopathic features and multiple minicores (Figure 3a). Two novel heterozygous mutations in *TTN* were identified in patient P4. One of the mutations is a 2 bp deletion (c.76109_76110del; Figure 3b) of paternal origin, which presumably gives rise to a premature termination codon and a shorter polypeptide (p.Ile25370Argfs*6). The second mutation, maternally inherited, is an intronic single-nucleotide variant located near the splice site (c.52102+5G>A). As predicted by bioinformatic analysis, muscle RNA studies revealed that this variant is disease causing: the canonical splice site is abolished and a cryptic splice site located within exon 273 is alternatively used, originating an in-frame deletion of 69 bp (Figure 3c). The analysis of the other family members demonstrated the autosomal recessive inheritance pattern (Figure 3d).

Patient P6 had a milder phenotype, presenting with delayed walking during infancy. He had facial and limb-girdle weakness, mostly scapular weakness. Muscle biopsy showed myopathic changes and multicores (Figure 3a). The patient is currently 17 years old and shows no respiratory or cardiac involvement. The genetic study revealed a *de novo* heterozygous mutation: a 2 bp insertion in exon 304 (c.61679_61680insCC) (Figure 3b) that predictably gives rise to a

premature stop codon in the A-band of Titin (p.Ser20561Leufs*17). Additionally, a novel silent variant was identified in exon 358 (c.105648G>A), but its effect at the mRNA level has yet to be determined.

DISCUSSION

We report the development of a novel and efficient method for mutation detection based on MPS and its application in the genetic characterization of patients with CMs.

Several centers reported low positivity rates in the molecular characterization of CM patients. These numbers may be seen in light of the fact that these studies were restricted to a relatively small number of candidate genes that were known, until recently.[3] Additionally, although some clinical and histological features allow some diagnostic guidance to the genetic study, in a subset of cases neither can orientate towards a (single) specific gene. Since the pathognomonic morphological abnormalities may show progressive or age-related changes,[5] the diagnostic value of the muscle biopsy depends when it is performed. As a final point, the severest end of the CM disease spectrum includes cases with prenatal onset and neonatal death, where the scarcity of clinical symptoms and the limited number of additional tests often cut short the diagnostic workup.

The introduction of new sequencing technologies (next-generation sequencing or MPS) circumvented some of these limitations, enabling
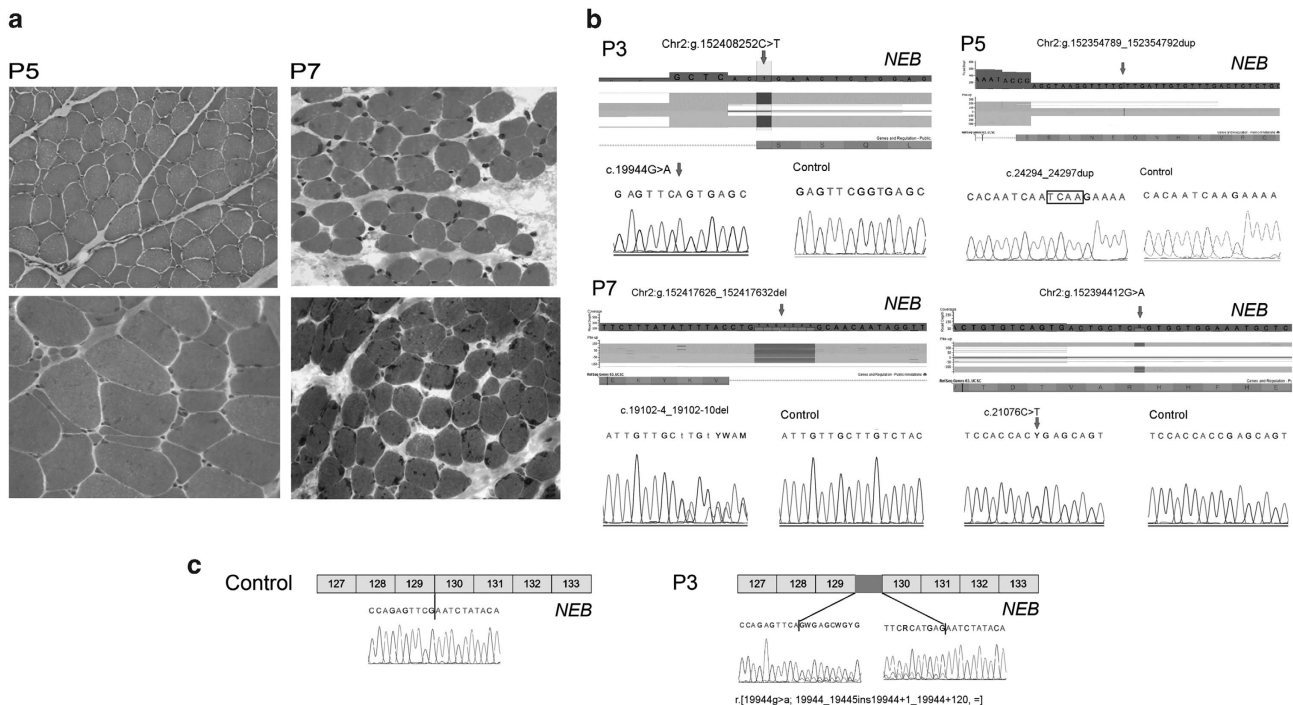
Figure 2 *NEB*-related nemaline myopathy. (a) Histological features of deltoid muscle biopsies of patients P5 and P7 (hematoxylin and eosin with x200 magnification in P5 and x400 in P7; GT with x400 magnification). In patient P5 there was fiber size variability, with severely atrophic fibers, corresponding to the majority to type 1 fibers. On the GT stain, there was a small area with some structures resembling rods. In P7 case, the typical rods were clearly seen on the GT stain. (b) Homozygous *NEB* mutations identified in patient P3 (c.19944G>A) and P5 (c.24294_24297dup), identified by MPS (top part) and confirmed by Sanger sequencing. Two heterozygous *NEB* mutations identified in patient P7: (i) an intronic deletion of 7 bp (c.19102-4_19102-10del) affecting the acceptor splice site and (ii) a nonsense mutation c.21076C>T (p.Arg7026*). (c) Characterization of the effect of the c.19944G>A mutation at the mRNA level. This mutation abolishes the normal donor splice site, shifting towards the use of an alternative (cryptic) splice site in intron 129, which predictably originates a premature stop codon and a shorter polypeptide. Abbreviation: GT, gomori trichrome. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

genetic diagnosis of diseases which had been considered extremely difficult to study by traditional methods. Different strategies based on MPS have been proposed for the genetic characterization of hereditary myopathies. These range from gene panels with a limited number of loci aiming to obtain high coverage depth,[18,19] to larger panels targeting several loci known to be implicated in myopathies[20] or even those with a broader scope, such as that covering all known neuromuscular disorders.[21] To address the genetic heterogeneity of these diseases, several successful applications of whole-exome or -genome sequencing were reported in the literature.[22–25] Irrespective of the approach taken, the histopathological findings should be correlated with the genetic data that are generated, as these can help narrow down the genetic studies considerably.[26]

Our strategy consisted in the development of a well delimited gene panel for CMs which in the present cohort led to an overall mutation detection rate of 70%. This is relevant considering that the assay does not include all the genomic regions, nor does it detect the presence of large genomic rearrangements. In patient P6 only one heterozygous frameshift mutation was identified. This was a *de novo* mutation (not detected in the patient's parents and paternity was confirmed). *TTN*-related myopathies are currently an expanding group of diseases, seen to have a wide clinical presentation that ranges from congenital to late adulthood onset, and with autosomal recessive or dominant inheritance.[10] The histological data of patient P6 showed the presence of multicores in NADH staining and by electron microscopy areas of sarcomeric disruption with irregularities and smearing of the Z line and disorganization of the myofibrils (data not shown). Additional

experimental evidence is being gathered to clarify the impact of the apparently silent variant (c.105648G>A) and to rule out or confirm the presence of an additional *TTN* mutation, not detectable by the methods used.

About 25% of the variants detected in this group of patients required further studies (in particular, expression analysis at the mRNA level). Although bioinformatic analysis was highly suggestive of their pathogenic nature, their impact was ultimately demonstrated by these additional studies. The best example was the study of patient P3, where *NEB* gene analysis revealed the presence of an apparently silent homozygous variant (c.19944G>A), described in previous reports as likely to be deleterious.[16] Given that the patient's biopsy was performed several years ago and there were no muscle fragments to obtain RNA, the possibility of using blood-derived transcripts was evaluated. In fact, we detected suitable expression of nebulin in whole-blood nucleated cells, which enabled the study of the c.19944G>A variant and its effect on splicing. This mutation abolishes the normal splice-site consensus sequence leading to the use of an alternative (cryptic) splice in the adjacent intron. It is predictable that this partial intronic retention creates a premature termination codon and thus a shorter nebulin protein. The presence of a residual amount of normal transcripts, possibly due to a leaky mutational effect, explains the milder phenotype in this patient.

Even though the application of this gene panel was very successful, some technological limitations should be mentioned. First, one variant in a positive control was incorrectly genotyped due to an allele dropout during PCR, owing to the presence of a rare SNP in the

Figure 3 *TTN*-related myopathies. (**a**) Histological features of patients P4 and P6 deltoid muscle biopsies (hematoxylin and eosin with x400 magnification in P4 and x200 in P6, NADH with x400 magnification and toluidine blue semi-thin sections with x630 magnification in P4 and x400 in P6). In both cases there was fiber size variability and multiple small minicores seen on oxidative stains and semi-thin sections. (**b**) Heterozygous *TTN* mutations identified in patient P4: (i) splice-site mutation in intron 279 (c.52102+5G>A) and (ii) an out-of-frame deletion (c.76109_76110del) in exon 326. In patient P6 a heterozygous insertion of two nucleotides was detected (c.61679_61680insCC), which leads to a shift on the *TTN* reading frame (p.Ser20561Leufs*17). (**c**) The c.52102+5G>A mutation was further characterized at the mRNA level, where it was shown that it abolishes the donor splice site and an cryptic splice site in exon 273 is used in alternative, leading to an in-frame deletion (r.52034_52102del). (**d**) Pedigree of patient P4 family is compatible with an autosomal recessive inheritance pattern, where the two affected sibs share the genotype. Fw, forward; Rv, reverse. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

patient. Such events are not unusual in conventional sequencing. In an attempt to avoid this pitfall, the bioinformatic process for primer design used in the Ampliseq strategy takes into account positions of known SNPs. However, *de novo*, ultra-rare or population-specific variants will not be considered in the pipeline. This false-positive result had no practical implication in our study, because all the variants identified by MPS were confirmed by Sanger sequencing. Our main concern is the scenario where the allele dropout prevents the amplification of the mutated allele, generating a false-negative result. Data analysis is one of the biggest hurdles in this technology, given the different possibilities of error that may lead to false-positive or -negative results.[27] As shown here, besides the development of the MPS gene panel it was necessary to deploy a strategy to deal with the large number of variants by using specific variant filters. During this process, one of the mutations detected in this work was initially filtered out due to incorrect annotation. It is highly recommended, especially in large variant datasets, that different sources of variant annotation be used so as to ensure to allow accuracy and consistency. Variant annotation is one additional layer of variability in MPS[28] which very often remains disregarded.

As a future perspective, the intention is to expand this gene panel to include loci that have recently been linked to CMs, namely *KLHL41*, *LMOD3* and *SPEG*.[29–31] It would also be relevant to perform whole-exome sequencing in patients who remain genetically unsolved after screening with this gene panel.

In conclusion, the wide phenotypic heterogeneity and huge size of the candidate genes makes diagnosis of CM complex, costly and labor intensive. We have developed and demonstrated the clinical utility of a new MPS gene panel for the screening of these patients. In 7 of the 10 undiagnosed cases, pathogenic variants were identified, most of which contributed towards widening both the mutational and the phenotypic spectra of CMs.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

1 Jungbluth, H. & Wallgren-Pettersson, C. in *Emery and Rimoin's Principles and Practice of Medical Genetics* 6th edn (eds Rimoin, D.L., Pyeritz, R.E. & Korf, B.) Ch. 127, 1–51 (Academic Press, 2014).

2 Oliveira, J., Oliveira, M. E., Kress, W., Taipa, R., Pires, M. M., Hilbert, P. *et al.* Expanding the MTM1 mutational spectrum: novel variants including the first multi-exonic duplication and development of a locus-specific database. *Eur. J. Hum. Genet.* **21**, 540–549 (2013).

3 North, K. N., Wang, C. H., Clarke, N., Deconinck, N., Bertini, E., Ferreiro, A. *et al.* International Standard of Care Committee for Congenital Myopathies. Approach to the diagnosis of congenital myopathies. *Neuromuscul. Disord* **24**, 97–116 (2014)

4 Snoeck, M., van Engelen, B. G., Küsters, B., Lammens, M., Meijer, R., Molenaar, J. P. *et al.* RYR1-related myopathies: a wide spectrum of phenotypes throughout life. *Eur. J. Neurol.* **22**, 1094–1112 (2015).

5 Romero, N. B & Clarke, N. F. Congenital myopathies. *Handb. Clin. Neurol.* **113**, 1321–1336 (2013).

6 Romero, N. B., Lehtokari, V. L., Quijano-Roy, S., Monnier, N., Claeys, K. G., Carlier, R. Y. *et al.* Core-rod myopathy caused by mutations in the nebulin gene. *Neurology* **73**, 1159–1161 (2009).

7 Piteau, S. J., Rossiter, J. P., Smith, R. G & MacKenzie, J. J. Congenital myopathy with cap-like structures and nemaline rods: case report and literature review. *Pediatr. Neurol.* **51**, 192–197 (2014).

8 Rocha, J., Taipa, R., Melo Pires, M., Oliveira, J., Santos, R. & Santos, M. Ryanodine myopathies without central cores—clinical, histopathologic, and genetic description of three cases. *Pediatr. Neurol.* **51**, 275–278 (2014).

9 Dowling, J. J., Lawlor, M. W. & Dirksen, R. T. Triadopathies: an emerging class of skeletal muscle diseases. *Neurotherapeutics* **11**, 773–785 (2014).

10 Chauveau, C., Rowell, J. & Ferreiro, A. A rising titan: TTN review and mutation update. *Hum. Mutat.* **35**, 1046–1059 (2014).

11 Wallgren-Pettersson, C., Sewry, C. A., Nowak, K. J. & Laing, N. G. Nemaline myopathies. *Semin. Pediatr. Neurol.* **18**, 230–238 (2011).

12 Duarte, S. T., Oliveira, J., Santos, R., Pereira, P., Barroso, C., Conceição, I. *et al.* Dominant and recessive RYR1 mutations in adults with core lesions and mild muscle symptoms. *Muscle Nerve* **44**, 102–108 (2011).

13 Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352 (2011).

14 McCall, C. M., Mosier, S., Thiess, M., Debeljak, M., Pallavajjala, A., Beierl, K. *et al.* False positives in multiplex PCR-based next-generation sequencing have unique signatures. *J. Mol. Diagn.* **16**, 541–549 (2014).

15 Lehtokari, V. L., Kiiski, K., Sandaradura, S. A., Laporte, J., Repo, P., Frey, J. A. *et al.* Mutation update: the spectra of nebulin variants and associated myopathies. *Hum. Mutat.* **35**, 1418–1426 (2014).

16 Pelin, K., Donner, K., Holmberg, M., Jungbluth, H., Muntoni, F. & Wallgren-Pettersson, C. Nebulin mutations in autosomal recessive nemaline myopathy: an update. *Neuromuscul. Disord.* **12**, 680–686 (2002).

17 Lehtokari, V. L., Pelin, K., Sandbacka, M., Ranta, S., Donner, K., Muntoni, F. *et al.* Identification of 45 novel mutations in the nebulin gene associated with autosomal recessive nemaline myopathy. *Hum. Mutat.* **27**, 946–956 (2006).

18 Kondo, E., Nishimura, T., Kosho, T., Inaba, Y., Mitsuhashi, S., Ishida, T. *et al.* Recessive RYR1 mutations in a patient with severe congenital nemaline myopathy with ophthalomoplegia identified through massively parallel sequencing. *Am. J. Med. Genet. A* **158A**, 772–778 (2012).

19 Ankala, A., da Silva, C., Gualandi, F., Ferlini, A., Bean, L. J., Collins, C. *et al.* A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield. *Ann. Neurol.* **77**, 206–214 (2015).

20 Savarese, M., Di Fruscio, G., Mutarelli, M., Torella, A., Magri, F., Santorelli, F. M. *et al.* MotorPlex provides accurate variant detection across large muscle genes both in single myopathic patients and in pools of DNA samples. *Acta Neuropathol. Commun.* **2**, 100 (2014).

21 Vasli, N., Böhm, J., Le Gras, S., Muller, J., Pizot, C., Jost, B. *et al.* Next generation sequencing for molecular diagnosis of neuromuscular diseases. *Acta Neuropathol.* **124**, 273–283 (2012).

22 Bögershausen, N., Shahrzad, N., Chong, J. X., von Kleist-Retzow, J. C., Stanga, D., Li, Y. *et al.* Recessive TRAPPC11 mutations cause a disease spectrum of limb girdle muscular dystrophy and myopathy with movement disorder and intellectual disability. *Am. J. Hum. Genet.* **93**, 181–190 (2014).

23 Böhm, J., Vasli, N., Malfatti, E., Le Gras, S., Feger, C., Jost, B. *et al.* An integrated diagnosis strategy for congenital myopathies. *PLoS ONE* **8**, e67527 (2013).

24 Couthouis, J., Raphael, A. R., Siskind, C., Findlay, A. R., Buenrostro, J. D., Greenleaf, W. J. *et al.* Exome sequencing identifies a DNAJB6 mutation in a family with dominantly-inherited limb-girdle muscular dystrophy. *Neuromuscul. Disord.* **24**, 431–435 (2014).

25 Oliveira, J., Negrão, L., Fineza, I., Taipa, R., Melo-Pires, M., Fortuna, A. M. *et al.* New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing. *J. Hum. Genet.* **60**, 305–312 (2015).

26 Uruha, A., Hayashi, Y. K., Oya, Y., Mori-Yoshimura, M., Kanai, M., Murata, M. *et al.* Necklace cytoplasmic bodies in hereditary myopathy with early respiratory failure. *J. Neurol. Neurosurg. Psychiatry.* **86**, 483–489 (2015).

27 Robasky, K., Lewis, N. E. & Church, G. M. The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Genet.* **15**, 56–62 (2014).

28 McCarthy, D. J., Humburg, P., Kanapin, A., Rivas, M. A., Gaulton, K., Cazier, J. B. *et al.* Choice of transcripts and software has a large effect on variant annotation. *Genome Med.* **6**, 26 (2014).

29 Gupta, V. A., Ravenscroft, G., Shaheen, R., Todd, E. J., Swanson, L. C., Shiina, M. *et al.* Identification of KLHL41 mutations implicates BTB-Kelch-mediated ubiquitination as an alternate pathway to myofibrillar disruption in nemaline myopathy. *Am. J. Hum. Genet.* **93**, 1108–1117 (2013).

30 Yuen, M., Sandaradura, S. A., Dowling, J. J., Kostyukova, A. S., Moroz, N., Quinlan, K. G. *et al.* Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. *J. Clin. Invest.* **124**, 4693–4708 (2014).

31 Agrawal, P. B, Pierson, C. R., Joshi, M., Liu, X., Ravenscroft, G., Moghadaszadeh, B. *et al.* SPEG interacts with myotubularin, and its deficiency causes centronuclear myopathy with dilated cardiomyopathy. *Am. J. Hum. Genet.* **95**, 218–226 (2014).

32 Bitoun, M., Bevilacqua, J. A., Prudhon, B., Maugenre, S., Taratuto, A. L., Monges, S. *et al.* Dynamin 2 mutations cause sporadic centronuclear myopathy with neonatal onset. *Ann. Neurol.* **62**, 666–670 (2007).

33 Ferreiro, A., Quijano-Roy, S., Pichereau, C., Moghadaszadeh, B., Goemans, N., Bönnemann, C. *et al.* Mutations of the selenoprotein N gene, which is implicated in rigid spine muscular dystrophy, cause the classical phenotype of multiminicore disease: reassessing the nosology of early-onset myopathies. *Am. J. Hum. Genet.* **71**, 739–749 (2002).

npg

ORIGINAL ARTICLE

New splicing mutation in the choline kinase beta (*CHKB*) gene causing a muscular dystrophy detected by whole-exome sequencing

Jorge Oliveira[1,2,3], Luís Negrão[4], Isabel Fineza[5], Ricardo Taipa[6], Manuel Melo-Pires[6], Ana Maria Fortuna[3,7], Ana Rita Gonçalves[1,3], Hugo Froufe[8], Conceição Egas[8], Rosário Santos[1,3,10,11] and Mário Sousa[2,3,9,11]

Muscular dystrophies (MDs) are a group of hereditary muscle disorders that include two particularly heterogeneous subgroups: limb-girdle MD and congenital MD, linked to 52 different genes (seven common to both subgroups). Massive parallel sequencing technology may avoid the usual stepwise gene-by-gene analysis. We report the whole-exome sequencing (WES) analysis of a patient with childhood-onset progressive MD, also presenting mental retardation and dilated cardiomyopathy. Conventional sequencing had excluded eight candidate genes. WES of the trio (patient and parents) was performed using the ion proton sequencing system. Data analysis resorted to filtering steps using the GEMINI software revealed a novel silent variant in the choline kinase beta (*CHKB*) gene. Inspection of sequence alignments ultimately identified the causal variant (*CHKB*:c.1031 +3G>C). This splice site mutation was confirmed using Sanger sequencing and its effect was further evaluated with gene expression analysis. On reassessment of the muscle biopsy, typical abnormal mitochondrial oxidative changes were observed. Mutations in *CHKB* have been shown to cause phosphatidylcholine deficiency in myofibers, causing a rare form of CMD (only 21 patients reported). Notwithstanding interpretative difficulties that need to be overcome before the integration of WES in the diagnostic workflow, this work corroborates its utility in solving cases from highly heterogeneous groups of diseases, in which conventional diagnostic approaches fail to provide a definitive diagnosis.
Journal of Human Genetics advance online publication, 5 March 2015; doi:10.1038/jhg.2015.20

INTRODUCTION

Muscular dystrophies (MD) are a set of highly heterogeneous muscle diseases sharing distinctive dystrophic features on muscle histology, namely degenerative fibers and fibrosis, variable myofiber size and abnormally internalized nuclei.[1] MD are classically subdivided into five major groups: congenital MD (CMD), Duchenne/Becker MD, Emery–Dreifuss MD, facioscapulohumeral MD and limb-girdle MD (LGMD). Two of these groups, CMD and LGMD, are particularly heterogeneous from both a clinical and a genetic perspective. Until now 31 different loci have been reported for the LGMDs and 28 genes in CMD.[2,3] Notably, at least seven genes are shared between these two groups, and this list is increasing over time. Concerning CMD, mutations in the laminin-α2 (*LAMA2*) gene are a major cause of the disease, corresponding to ~58% of patients referred for genetic studies in this disease group.[4] In patients with inconclusive

imunohistochemistry studies, the diagnostic workup may be more complex and is usually steered by other muscle biopsy features, clinical evaluation, electromyogram, and brain and muscle imaging. Specific clinical signs are important to consider, namely the distribution and progression of muscle weakness, the presence of brain structural changes (with or without mental retardation) and other features such as the presence of contractures, rigid spine and hyperlaxity.[3] All these features have been used to classify CMD into further subtypes and to establish robust diagnostic algorithms. Besides laminin-α2 deficiency or changes for α-dystroglycan, both detected by imuno-histochemistry, other clinically relevant histological findings (such as structural changes) are rare.

Nishino *et al.*[5] reported a subtype of CMD with mitochondrial structural abnormalities, presumably of autosomal recessive inheritance. The four patients described shared some clinical features with

[1]Unidade de Genética Molecular, Centro de Genética Médica Dr Jacinto de Magalhães, Centro Hospitalar do Porto E.P.E., Porto, Portugal; [2]Departamento de Microscopia, Laboratório de Biologia Celular, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal; [3]Unidade Multidisciplinar de Investigação Biomédica (UMIB), Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal; [4]Consulta de Doenças Neuromusculares, Hospitais da Universidade de Coimbra, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal; [5]Unidade de Neuropediatria, Centro de Desenvolvimento da Criança Luís Borges, Hospital Pediátrico de Coimbra, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal; [6]Unidade de Neuropatologia, Centro Hospitalar do Porto, Porto, Portugal; [7]Serviço de Genética Médica, Centro de Genética Médica Dr Jacinto Magalhães, Centro Hospitalar do Porto E.P.E., Porto, Portugal; [8]Unidade de Sequenciação Avançada, Biocant, Parque Tecnológico de Cantanhede, Núcleo 04, Cantanhede, Portugal; [9]Centro de Genética da Reprodução Professor Alberto Barros, Porto, Portugal and [10]UCIBIO/REQUIMTE, Departamento de Ciências Biológicas, Laboratório de Bioquímica, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal
[11]These authors contributed equally to this work.
Correspondence: Dr R Santos, Unidade de Genética Molecular, Centro de Genética Médica Dr Jacinto de Magalhães, Centro Hospitalar do Porto E.P.E., Praça Pedro Nunes, 88, 4099-028 Porto, Portugal.
E-mail: rosario.santos@chporto.min-saude.pt
Received 21 October 2014; revised 14 January 2015; accepted 2 February 2015

2

other CMD forms, but had marked mental retardation without structural brain changes. Disease progression was slow and some patients developed dilated cardiomyopathy. Creatine kinase (CK) levels were mildly elevated. Muscle phenotype included dystrophic features and a striking finding was observed with succinate dehydrogenase and cytochrome *c* oxidase staining, revealing reduction of mitochondrial content in the center of the myofiber and enlarged mitochondria at the periphery, assuming a 'megaconial' appearance.[5] This entity was later demonstrated to be caused by mutations in the choline kinase beta (*CHKB*) gene, leading to loss of activity for this enzyme in rostrocaudal MD mice[6] and in humans in a very limited number of patients reported to date.[7–11] CHKB together with two isoforms of choline kinase alfa (CHKA) phosphorylates choline into phosphocholine, the initial step of the Kennedy's pathway. This step is essential in the biosynthesis of phosphatidylcholine—an important component of the phospholipidic membrane layer of eukaryotes—and also a key precursor for many signaling molecules with relevant cellular functions such as cell growth signaling and proliferation.[12] The disease mechanism has been attributed to the decreased levels of phosphatidylcholine in the mitochondria, leading to its dysfunction and subsequently mitophagy and clearance in the skeletal muscle.[13,14] Immunoprecipitation studies performed in the mouse liver suggested that chk activity was mainly due to heterodimeric CHKA/CHKB (60%), although some of the activities were due to homodimeric forms.[6] However, in human skeletal muscle CHKA is not expressed; therefore, the CHKB isoform is essential to sustain normal phosphatidylcholine levels.[7]

The development and application of new sequencing technologies, frequently coined next-generation sequencing (NGS), is currently accelerating research in the field of clinical genetics, essentially given its capacity to generate and analyze more genomic sequences at a lower cost and at a faster pace.[15] Using this technology it is possible to analyze several specific genes simultaneously, comprising a disease panel, or at the limit it may include all the exonic regions of the human genome—whole-exome sequencing (WES) or even the entire genome (individual genome sequencing). Whole-exome or -genome sequencing may also contribute to the identification of new genes associated with genetic diseases. Nevertheless, its introduction into routine diagnostics still faces some challenges, particularly considering the amount of data to be analyzed, and the possibility of obtaining false-positive and -negative results should not be neglected.

In this work we report the use of WES in a MD patient, which contributed to the identification of a novel splicing mutation in the *CHKB* gene. Our study corroborates the potential of WES to identify genetic defects involved in rare neuromuscular diseases and illustrates the difficulties that may be encountered using this technology, especially regarding data analysis.

MATERIALS AND METHODS

Patient
The patient is the first child of a nonconsanguineous healthy couple; pregnancy and neonatal period were uneventful. She was first observed by neuropediatrics at the age of 4 years because of motor problems: independent walking at 2.5 years, but with progressive lower limb weakness and difficulties in climbing stairs and running. She is presently unable to rise from a chair, the arms' reach do not rise to the shoulder level and she walks with braces. The CK values, ascertained over a period of 30 years, were slightly elevated, ranging from 299 to 1857 U l^{-1} (average 849 U l^{-1}; normal <200 U l^{-1}). Electromyography performed at the of age 7 showed a myopathic pattern without evidence of fibrillation potentials or myotonic discharges. She also presented behavioral abnormalities and cognitive deficit, which prevented her from attaining the basic academic goals.

At the end of the first decade dilated cardiomyopathy was diagnosed, adequately controlled with medication (digoxin and furosemide). Standard genetic screening tests due to delay in psychomotor development (including speech delay), performed at the age of 22, were all normal. The patient had no signs of ichthyosis. Informed consent was obtained and the research was approved by the ethics committee from the Centro Hospitalar do Porto.

Sanger sequencing
Eight genes known to be defective in MDs were previously analyzed by conventional (Sanger) sequencing, namely: *ANO5, CAPN3, LARGE, FKRP, FKTN, POMGNT1, POMT1* and *POMT2*. Genomic sequencing included all coding regions and adjacent intronic sequences. Results were analyzed using the SeqScape software V2.5 (Life Technologies, Foster City, CA, USA). Variants were described according to the Human Genome Variation Society guidelines for mutation nomenclature (version 2.0).[16] Large heterozygous deletions and duplications were also screened in several genes using the multiplex ligation-dependent probe amplification technique. Different commercial kits from MRC-Holland (Amsterdam, the Netherlands) were used: *SGCA, SGCB, SGCD, SGCG* and *FKRP* (P116); *FKTN, LARGE* and *POMT2* (P326); *POMT1* and *POMGNT1* (P061). Genomic DNA (150 ng) from the patient and at least three normal controls were used in the procedure. Multiplex ligation-dependent probe amplification products were resolved by capillary electrophoresis on an ABI 3130xl genetic analyzer (Life Technologies). Data analysis was conducted using the GeneMarker Software V1.5 (SoftGenetics LLC, State College, PA, USA). After using the population normalization method, data were plotted using the probe ratio option.

Exome-sequencing
Seventy-five nanograms of high-quality DNA from each individual (patient and both parents) were amplified in 12 primer pools of 200-bp amplicons with the Ion AmpliSeq Exome Library Preparation kit (Life Technologies). Samples were barcoded with the Ion Xpress Barcode Adapter (Life Technologies) in order to pool two exomes per chip. The libraries were evaluated for quality with High Sensitivity DNA Kit in Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantified using the Ion Library Quantitation Kit (Life Technologies). Library fragments were clonally amplified with emulsion PCR using the Ion PI Template OT2 200 kit v2 and the Ion OneTouch 2 System (Life Technologies), and the positive-ion sphere particles enriched in the Ion OneTouch ES machine (Life Technologies). Finally, the positive spheres were loaded in Ion PI chip v2 and sequenced in the Ion Proton System (Life Technologies). All procedures were carried out according to the manufacturer's instructions.

Exome mapping and variant calling
The data generated from the three exomes (patient and parents) were processed with the Torrent Suite Software 4.1 (Life Technologies). Reads were mapped against the human reference genome hg19 using Torrent Mapping Alignment Program version 4.0.6 (Life Technologies). Variant calling was performed by running Torrent Variant Caller plugin version 4.0, using the optimized parameters for exome-sequencing recommended for AmpliSeq sequencing (Life Technologies). The variant call format (VCF) files from the trio were combined using vcftools version 0.1.9.0.[17] and all variants were annotated and prioritized with GEMINI.[18] Variant filtering was performed using a list of 108 candidate genes known to be implicated in hereditary myopathies (Supplementary Data I). Each gene was screened for variants matching the recessive disease model, being: (i) homozygous in the patient and parents carrying the same variant in heterozygosity ('autosomal recessive') or (ii) two heterozygous variants in the patient (compound heterozygosity, 'comp_hets' function) and each having a distinct parental origin. Candidate variants were manually checked on the Binary Alignment Map files through Integrative Genomics Viewer (IGV) version 2.3[19] and GenomeBrowse V2.0.0 (Golden Helix Inc., Bozeman, MT, USA).

Confirmation of variants with sanger sequencing
To verify the candidate variant detected in the *CHKB* gene of the three individuals, specific primers were designed (9 F-5′-GTGGAGTCTGGAAGG

AATGGC-3′; 9 R-5′-TTTAACTTCTCCCCCACTGTCC-3′) to amplify the region of interest using PCR. Resulting PCR products were purified using Illustra ExoStar (GE Healthcare, Little Chalfont, UK) and subjected to a new cycling reaction prepared with BigDye Terminator kit V3.1 chemistry (Life Technologies). Sequencing reactions were analyzed in an ABI 3130xl genetic analyzer (Life Technologies).

Gene expression analysis

Total RNA was obtained from the peripheral blood (patient and a control) using the PerfectPure RNA Blood Kit (5 PRIME, Hamburg, Germany) and from muscle (only available from a control) with the PerfectPure RNA Fibrous Tissue kit (5 PRIME). RNA was converted into cDNA using the High Capacity cDNA Reverse Transcription kit (Life Technologies). For the analysis of *CHKB* transcripts, PCR amplification was performed using primers annealing to the regions corresponding to exons 7–11 (6/7 F-5′-TCCAGGAAGGGAACA TCTTG-3′; 11 R-5′-GGTGGAGTCAGGATGAGGAG-3′, on the basis of the reference sequence with accession number NM_005198.4). As an experimental control, the 3′ region of the transcripts for protein-O-mannosyltransferase 1 (*POMT1*) was simultaneously amplified, using the following primers: 18 F-5′-CGGCGAAGAAATGTCCATGAC-3′ and 3′-UTR-R-5′-GAGCTTTTCAATGA GACCCCC-3′ (on the basis of reference sequence with accession number NM_007171.3). Resulting PCR products were resolved in a 2% agarose gel; visible bands were excised and purified using the PureLink Quick Gel Extraction kit (Invitrogen by Life Technologies). Products were sequenced as described above.

Muscle histology

Two muscle biopsies were performed, at 9 and 20 years of age. In the first biopsy, histochemical studies were not performed and paraffin sections showed nonspecific changes. The second biopsy revealed striking variation in fiber size, with type 1 and type 2 fiber atrophy, necrotic fibers, several basophilic fibers and endomysial fibrosis. Imunohistochemistry showed normal labeling for sarcoglycans, dystrophin and laminin-α2. succinate dehydrogenase and nicotinamide adenine dinucleotide staining was uneven in central areas of the myofiber. The conclusion at that time (1995) was that these findings pointed toward a possible MD.

RESULTS

Preceding molecular genetic analysis

The patient was followed by neuropediatrics and later neurology, with the clinical suspicion of LGMD or possibly a dystroglycanopathy. Eight candidate genes were initially analyzed using Sanger sequencing as part of the routine genetic diagnostic workup. Several variants were detected in these studies, including a novel silent heterozygous variant (NM_024301.4:c.474C>G) in the fukutin-related protein (*FKRP*) gene, which prompted further characterization at the complementary DNA (cDNA) level. As no splicing defects were identified, this unclassified variant was considered nonpathogenic. These studies

were complemented by screening for large genomic deletions or duplications at several loci by multiplex ligation-dependent probe amplification; however, again no causal mutations were detected. These continuous molecular studies were inconclusive and unable to identify the genetic defect responsible for the patient's phenotype. The high number of candidate genes that were screened and the detailed clinical characterization consistent with a genetically determined MD led us to consider WES as an appropriate alternative approach to identify the underlying defect.

Exome-sequencing metrics

Overall quality parameters of the exome trio are summarized in Table 1. On average, each exome generated a total of 37 million sequence reads, 99% of which successfully aligned against the human reference genome. Accordingly, the exome's target regions were on average covered 105.6 times and 93% of these had a minimum coverage depth of 20-fold. These values are relatively high and above the threshold considered appropriate for this study. Approximately 83% of all sequenced bases had a Phred quality score above Q20 (that is, a 99% accuracy of the base call). The size of the total captured regions was ∼58 Mbp. For each sequenced exome one Binary Alignment Map and VCF files were generated. An overview of the variants called for each exome is described in Table 2. WES identified 48 847 single-nucleotide variants in the index case and 48 870 and 47 878 single-nucleotide variants in each parent. More than 2300 insertions/deletions were also detected in each individual exome. The patient's exome contains 1820 variants not previously reported in single nucleotide polymorphism database (dbSNP) (http://www.ncbi. nlm.nih.gov/SNP/).

Exome analysis and variant filtering

To deal with the large amount of variants identified using WES (>50 000 variants in each individual exome), the analysis approach was based on the patient's clinical phenotype and the most probable inheritance pattern for the disease. The first filtering strategy focused on the known genetic causes for hereditary myopathies (including MDs), where a total of 108 candidate genes were selected (Supplementary Data I), on the information available in the Muscle Gene Table website (http://musclegenetable.fr/, accessed April 2014). The GEMINI database framework was used to annotate and filter the large number of sequence variants listed in the three generated VCF files. This software has different built-in analysis tools that are useful for identifying *de novo* mutations, but also to filter variants meeting autosomal recessive or autosomal-dominant inheritance patterns in family-based studies.[18] As no other relatives of our patient were reported as being affected with a muscle disease, the autosomal recessive model was used, which in GEMINI is subdivided into two

Table 1 Exome-sequencing quality metrics

Individual	Patient	Mother	Father
Total reads (number)	41 776 885	38 365 995	31 441 741
Aligned reads passed filtering (number)	41 271 480	37 922 459	31 102 979
Aligned reads passed filtering (%)	99	99	99
Mean coverage of target region (x)	117.3	109.5	89.96
Bases with $Q>20$ (%)	83.02	83.37	83.60
Total captured region size (Mb)	57.74	57.74	57.74
Captured regions with coverage >20 (%)	93.19	93.42	91.17

Sequencing and mapping metrics are presented per individual.

Table 2 Exome variant metrics

Individual	Patient	Mother	Father
Total number of SNVs	48 847	48 870	47 878
Total number of indels	2473	2375	2332
Homozygous variants	19 510	19 231	18 958
Heterozygous variants	31 810	32 014	31 252
Novel variants[a]	1820	1749	1610

Abbreviation: SNV, single-nucleotide variant.
Variants (with quality ⩾50) are presented per individual. Variant calling files from each sequenced individual were combined and analyzed with GEMINI.
[a]Not listed in dbSNP (no 'rs' number attibuted).

a NGS: Chr22:g.51018153C>G, [c.1031+3G>C]   **b** SANGER: c.1031+3G>C
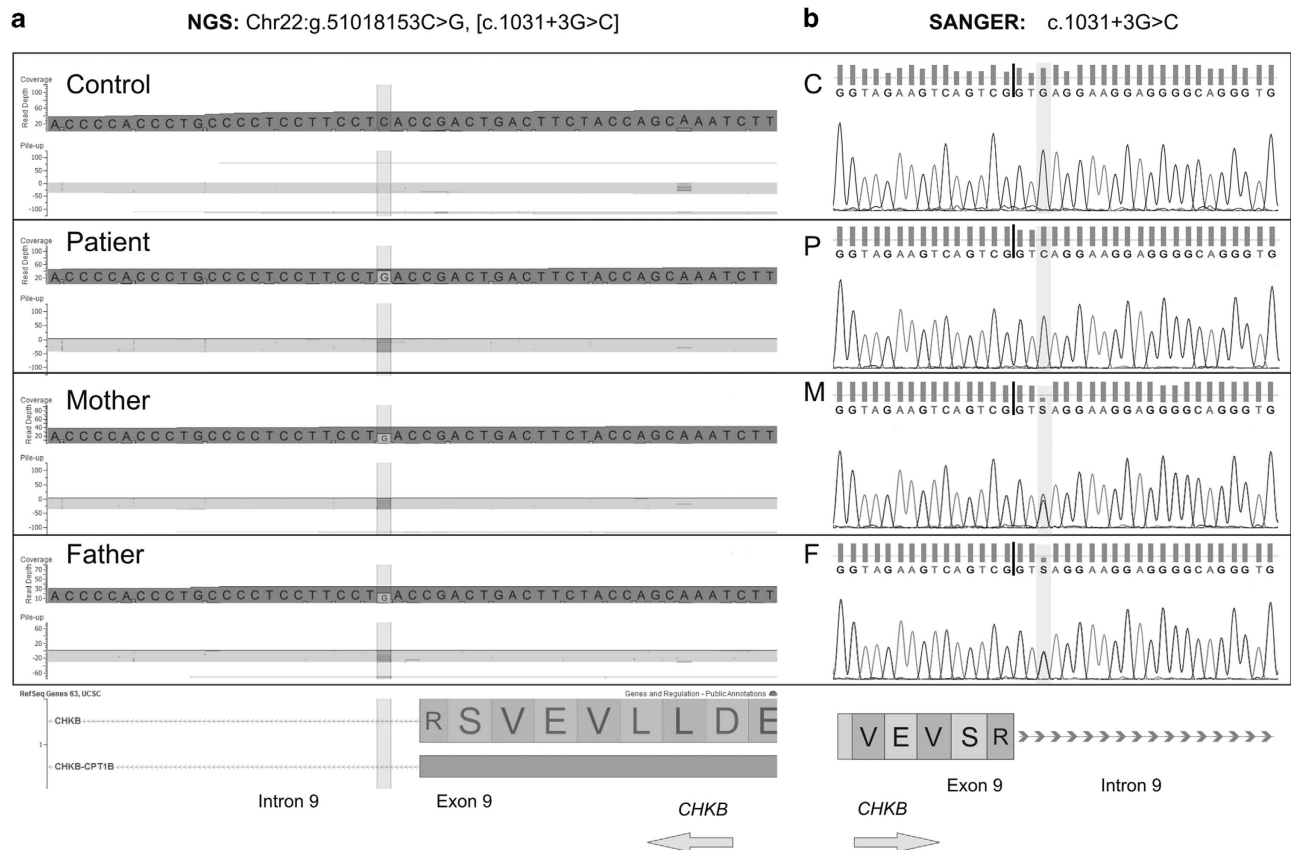
Figure 1 Novel splice mutation (c.1031+3G>C) identified in the *CHKB* gene. (**a**) Sequencing results from next-generation sequencing (NGS) showing this variant in the patient and in both parents. For this region of the gene, only reversely orientated reads were generated. (**b**) Although annotated in the variant call format (VCF) file as being present in heterozygosity, Sanger sequencing confirmed this variant to be homozygous in the patient (P); the patient's mother (M) and father (F) are both heterozygous carriers. cDNA reference sequence: NM_005198.4. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

different functions: 'comp_hets' and 'autosomal_recessive'. Considering that parental consanguinity was not reported, the first tool used was 'comp_hets'. This function is suitable to identify potential compound heterozygote genotypes. A list of compound heterozygous variants in the selected 108 genes was generated for the patient (Supplementary Data II). A total of 29 heterozygous variants in 11 different genes were identified, 24 are considered polymorphisms (frequency higher than 1.0%) and five had not been documented in dbSNP. After manually checking the read alignments on those genomic coordinates (for the trio and by comparing with a control exome), three variants were considered sequencing artifacts. One heterozygous missense variant (NM_004369.3:c.1927C>A, p.Leu643Ile) was detected in the α3 chain of the collagen VI gene (*COL6A3*); however, it was seen to have been inherited from the father, who was homozygous for this change. A second variant (NM_001267550.1: c.33513_33515dup) located in the titin gene (*TTN*), predictably resulting in the insertion of a new codon, was considered nonpathogenic as it was previously reported in the Exome Variant Server (http://evs.gs.washington.edu/EVS/) with an overall frequency of 1.3%.

As no suitable candidates were found, a second filtering approach ('autosomal_recessive') was used. This identifies canonical recessive mutations by generating a list of homozygous variants present in the patient and inherited from each parent, heterozygotes for those changes. In this analysis fewer variants were identified (*n* = 14) present in a total of eight different genes (one variant

Table 3 Data details for the main candidate variant

Individual	Patient	Mother	Father
Chromosome	22	22	22
Position	51 018 153	51 018 153	51 018 153
Gene name	*CHKB*	*CHKB*	*CHKB*
Depth	47	39	34
Reference allele	C	C	C
Number of reads with reference allele (frequency)	2 (0.04)	16 (0.41)	14 (0.41)
Alternate allele	G	G	G
Number of reads with alternate allele (frequency)	45 (0.96)	23 (0.59)	20 (0.49)
Mutation type	Splice site	Splice site	Splice site
Refseq accession number	NM_005198.4	NM_005198.4	NM_005198.4
Mutation DNA	c.1031+3G>C	c.1031+3G>C	c.1031+3G>C
Confirmed by Sanger sequencing	Yes	Yes	Yes

Abbreviation: BAM, Binary Alignment Map.
Data derived from manual inspection of the BAM file and correction of artifacts for the candidate position in each of the parents and the patient.

each in *CHKB*, *CNTN1*, *DNM2*, *ETFDH*, *ITGA7*, *NEB* and *TIA1* and seven variants in *SYNE1*; Supplementary Data II). Thirteen variants were already listed in publically available sequence variant databases (frequency higher than 1.0%) and one was a novel
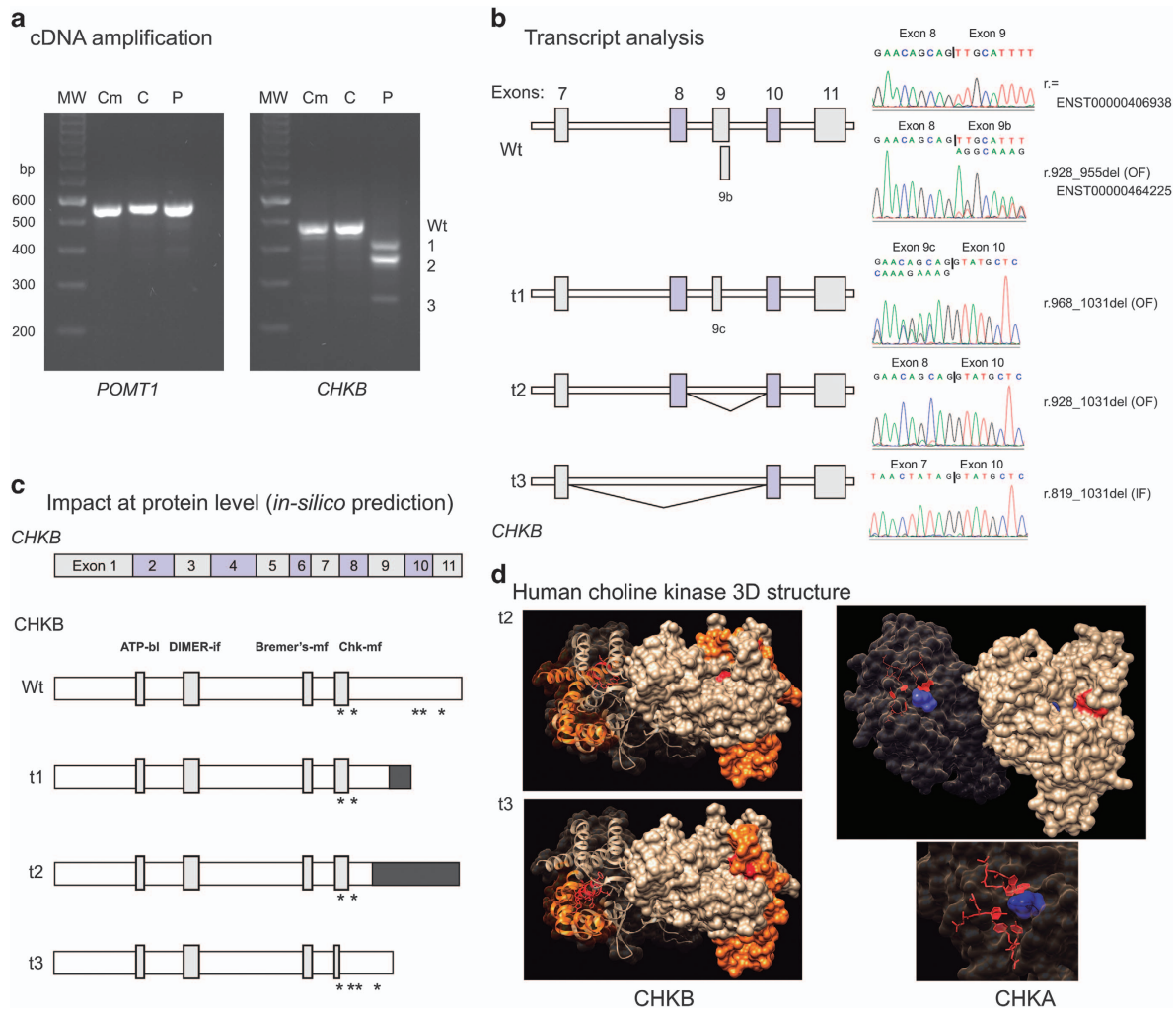
Figure 2 *CHKB* gene expression analysis. (**a**) Reverse transcriptase-polymerase chain reaction (RT-PCR) experiments using total RNA obtained from the peripheral blood (C, control; P, patient) and from the muscle (available only for the control sample, Cm). Amplification of *POMT1* transcripts was performed to show successful amplification in the different samples. Normal *CHKB* transcripts (Wt) were seen in the control sample, whereas in the patient only three aberrant transcripts (1–3) were present. (**b**) Resulting PCR products were sequenced for further transcript characterization. Besides the normal full-length transcript, there is also an alternatively spliced isoform (9 b). Three aberrant transcripts were identified in the patient: t1, resulting from the use of a cryptic splice site in exon 9, causing an out-of-frame (OF) deletion; t2, full exon 9 skipping causing a frameshift; and t3, skipping of exons 8 and 9 leading to corresponds to an in-frame (IF) deletion. (**c**) Bioinformatic analysis of aberrant transcripts. Gray rectangles draw attention to relevant CHKB domains: (i) ATP-binding loop (bl); (ii) dimer interface (if); (iii) Bremer's motif (mf); and (iv) Chk-mf.[26] Asterisks highlight important aromatic residues involved in stabilization of the positively charged amine of choline. Black regions represent residues predictably altered as a result of the frameshift. (**d**) Crystallographic three dimensional (3D) structures of the human choline kinase beta (CHKB) with Protein Data Bank (PDB): 2IG7 and also the α isoform (choline kinase alfa (CHKA)) with PDB: 2CKQ, showing its interaction with the subtract (choline). Both structures were visualized using the Chimera software V1.9.[27] In one of the chains of the CHK homodimers, transparency was increased to show relevant secondary structures and/or residues. The regions affected by the deletions identified in transcripts t2 and t3 are depicted in orange. Red residues indicate aromatic amino acids known to interact with choline, which is shown in blue.

sequence change located in the *CHKB* gene (NM_005198.4: c.843T>C).

Variant data interpretation

The novel homozygous variant c.843T>C located in exon 8 of the *CHKB* gene (Supplementary Data III) was evaluated by bioinformatic tools and prompted a detailed analysis of the alignments encompassing this gene. An additional novel variant was detected (NM_005198.4: c.1031+3G>C), located in the donor splice site of intron 9 (Figure 1a and Table 3). In this region, the read depth was ~40-fold and only reverse sequences were successfully aligned (Supplementary Data IV). It should be noted that the c.1031+3G>C variant was listed as heterozygous in the VCF files of all three exomes, thereby explaining

why it was filtered-out with GEMINI's 'autosomal_recessive' function. Conventional (Sanger) sequencing confirmed it to be homozygous in the patient and heterozygous in both parents (Figure 1b). In addition, this variant was not found in homozygosity in the patient's clinically unaffected sister (data not shown). Taking into account the possibility of other false-negatives in the remaining candidate loci, all heterozygous variants were reassessed. The genotype quality score of 90 was determined as a suitable cutoff value for this reassessment (Supplementary Data V). Eleven sequence variants (including the c.1031+3G>C change in *CHKB*) were retrieved that had been called as heterozygous in all three samples and having a genotype quality score <90 in the patient. With the exception of the *CHKB* variant, all were polymorphisms and thus not considered as possible candidates.
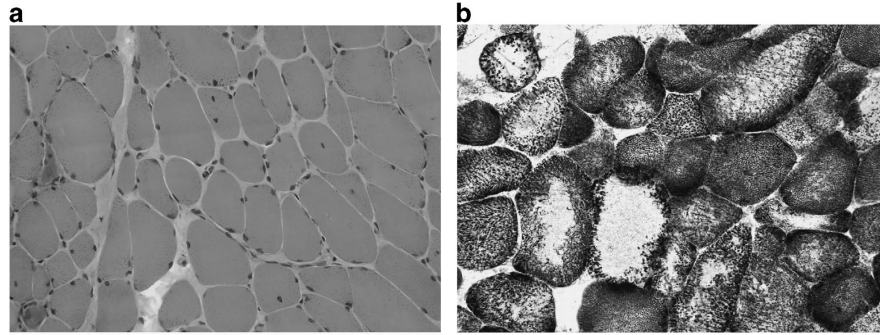
Figure 3 Muscle biopsy showing. (**a**) variation in fiber size, basophilic fibers and endomysial fibrosis (HE, magnification x200) and (**b**) large pale central areas with large mitochondria toward the periphery, seen with oxidative enzyme stain (succinate dehydrogenase (SDH), magnification x200). A full color version of this figure is available at the *Journal of Human Genetics* journal online.

The location of the c.1031+3G>C variant, at the exon–intron junction, leads us to suspect a possible effect on the splicing process; ~95% of mammalian transcripts have an A or a G nucleotide in the +3 position of the donor splice site.[20] This, together with the clinical presentation, led us to consider c.1031+3G>C as a candidate variant accountable for the disease.

CHKB gene expression analysis
Further evidence supporting the pathogenicity of the c.1031+3G>C variant was obtained through bioinformatic analysis. Five different algorithms were used to evaluate splicing: SpliceSiteFinder-Like, MaxEntScan,[21] NNSPLICE,[22] GeneSplicer[23] and Human Splicing Finder,[24] incorporated in the Alamut Visual V2.4 software (Interactive Biosoftware, Rouen, France). In all algorithms there was a reduction of probability scores, and particularly in two (NNSPLICE and GeneSplicer), the results were below the value set as a minimum threshold for splicing (Supplementary Data VI). This effect was experimentally determined by the analysis of *CHKB* transcripts. cDNA derived from the peripheral blood was used, given that no muscle specimens of the patient were available for study. CHKB is ubiquitously expressed and enzymatic activity for this kinase was previously reported in different tissues.[25] Moreover, in control individuals there were no differences in *CHKB* expression between cDNA samples derived from the blood and those derived from muscle (Figure 2a). In the patient there was a drastic effect on splicing; no normal expression for *CHKB* was detected as the expected normal transcript (with 468 bp, as seen to be present in control samples) was not observed. Instead, she presented at least three aberrant transcripts not detected in the controls. These were further characterized by sequencing (Figure 2b). All aberrant transcripts were due to splicing defects related with the mutation-derived weakening of the donor splice site. The first transcript (r.968_1031del) corresponds to a 64-bp out-of-frame deletion, where a cryptic splice site located in the middle of exon 9 is used as an alternative. The other two transcripts originate by exon 9 skipping (r.928_1031del) leading to a 104-bp out-of-frame deletion and by exons 8 and 9 skipping (r.819_1031del) leading to a 213-bp in-frame deletion. Should these mutated transcripts be translated, they predictably result in the loss of critical portions of the enzyme (Figure 2c and d).

The crystallographic structures are available in the Protein Data Bank (PDB) for the human CHKB (PDB: 2IG7) and the α isoform CHKA (PDB: 2CKQ).[26] On the basis of these structures it is predictable that in the mutated polypeptides derived from the transcripts 1 and 2, there is loss of some aromatic residues known to stabilize the positively charged amine of choline; the interaction with this substrate would thus be affected (Figure 2c and d). In addition, the formation of the choline-binding pocket site itself is probably also compromised. The mutated polypeptide with an in-frame deletion derived from transcript 3 (Figure 2c and d) is predicted to abolish part of the choline kinase motif located in the C-terminal region of CHKB, also involved in the binding of the choline moiety.[26]

Muscle histology
The results of WES analysis prompted the revision of the patient's muscle biopsy. Oxidative enzyme staining (succinate dehydrogenase) revealed not only large pale central areas but also abnormal oxidative activity, particularly toward the periphery of the myofiber (Figure 3). Electron microscopy shows the presence of slightly enlarged mitochondria (Supplementary data VII).

DISCUSSION
We describe the identification of a novel intronic splicing mutation in the *CHKB* gene, resorting to the innovative analysis of an exome trio (patient and respective parents). Mutations in this locus were previously reported as associated with an extremely rare form of CMD, with only 21 patients accounted for in the literature[7–11] and at least three additional cases presented in conference proceedings.[28,29] Defects in CHKB have been shown to cause phosphatidylcholine deficiency in muscle cells.[13] The c.1031+3G>C variant located in the donor splice site near exon 9 was considered pathogenic upon *CHKB* expression analysis. It was experimentally demonstrated that in the patient reported herein, there are no normal full-length *CHKB* transcripts, and only aberrant transcripts were identified. If these transcripts escape nonsense-mediated decay, they will probably fail to code for a functional CHKB enzyme. Our experimental findings are similar to those observed in a Turkish CMD patient with a different mutation (c.1031+1G>A) affecting the same donor splice site.[7]

The major clinical features found in the patient (progressive muscle weakness, mental retardation and dilated cardiomyopathy) are in agreement with the phenotype previously described for this entity (OMIM no. 602541). In the first clinical report of 15 patients,[7] findings included mildly to moderately elevated CK levels, floppiness during the neonatal period (in 9/14 patients), mental retardation (15/15), cardiomyopathy (6/14) and independent walking (11/15). Muscle histology was systematically dystrophic and mitochondrial enlargement was observed in all patients. An isolated case with a homozygous nonsense mutation in *CHKB* was identified in a child with a similar clinical and muscle phenotype.[8] The phenotypic spectrum linked to CHKB deficiency has been recently expanded with additional reports, including a female patient with a milder phenotype

resembling LGMD[9] and the oldest known patient (50 years old) having a progressive 'limb-girdle' myopathy.[29] Both patients presented missense mutations (two compound heterozygotes and one homozygote in the second report). Muscle histology showed large mitochondria located at the periphery and central areas in the myofiber, devoid of activity. In fact, it should be emphasized that the majority of patients described in the literature harboring *CHKB* mutations were initially identified by the existence of this specific histological hallmark: the presence of enlarged mitochondria in the periphery of the myofibers and the absence of this organelle in the center of the cells.[5] These enlarged mitochondria were correlated with the reduced levels of phosphatidylcholine in the mitochondrial membrane. This seems to be the consequence of a compensatory mechanism resulting from the depletion of functionally compromised mitochondria by autophagy.[14,30] A common aspect, especially in the atypical cases, is that several muscle biopsies needed to be performed. In the oldest patient reported to date, these mitochondrial changes were more evident in the last biopsy (from a total of three), which was performed at the age of 49 years.[29] This pathognomonic finding (megaconial) is highly specific for the defective gene; however, their detection requires awareness and high-quality histological preparations. Therefore, when MD (congenital or progressive) presents in association with mental retardation and cardiac involvement, the muscle biopsy should be carefully reassessed and the *CHKB* gene considered a possible candidate.

As previously referred, NGS technology is improving the capabilities to perform molecular genetic diagnostics. The most rapidly growing application of this technology is the development of NGS gene panels (targeted resequencing), which can help to deal with high genetic and clinical heterogeneity found in several diseases, such as neuromuscular disorders.[31] By performing targeted resequencing it is possible to obtain higher redundancy and thus more reliable variant data. This is achievable because the extension of these assays (covered regions) is usually quite smaller when compared with other (broader) NGS approaches such as WES or whole-genome sequencing. Accordingly, NGS-targeted resequencing will rapidly replace conventional sequencing for routine genetic diagnosis. However, there is still lack of consensus as to which genes should be included (or not) in such disease-specific gene panels and each laboratory implementing these techniques has been defining their own sets of genes. Another important aspect is that the application of NGS gene panels in some diseases such as hereditary myopathies will perpetuate the long-established clinical classification, which is currently being challenged by the increasing reports of the same gene being involved in several different subtypes. The ryanodine receptor 1 (*RYR1*) gene is an excellent example of this paradigm, considering its association with different diseases, ranging from congenital myopathies and MDs to malignant hyperthermia susceptibility, as well as by presenting dissimilar muscle phenotypes: myopathic with or without structural changes (such as central cores, minicores or centrally placed nuclei), dystrophic signs or apparently normal histology but with compromised function (defective intracellular calcium homeostasis causing hypermetabolic response when exposed to anesthetic agents).[32]

In contrast to NGS-targeted resequencing, WES has been proposed as an important research tool for gene discovery or further expanding the genetic causes of a particular disease.[33] Probably in the near future WES will be used as a first-tier diagnostic tool for several medical specialties, allowing the identification of genetic defects linked to monogenic or even complex multigenic disorders. Our initial analysis strategy was orientated toward the genes known to be implicated in hereditary myopathies and, among the attempted approaches, an autosomal recessive model was assumed. It is interesting to note that, although the mutation in the *CHKB* gene was successfully identified by WES, its zygosity was not called correctly in the VCF file used for automated analysis. A different unreported variant in the *CHKB* gene drew our attention to this locus and enabled us to identify the causative mutation. At this stage, considering the novelty aspect of the work, there is no indication whether this is a common pitfall of WES data analysis, and the overall impact on filtering strategies is as yet unknown. To ensure that no additional false-negative results were obtained upon filtering, a strategy based on the genotype quality score (threshold < 90) was used. Presently, WES should be seen as a screening technique, and it does not provide the answer to all aspects of clinical molecular genetic testing. As shown here, several other techniques and biological samples are very often required to clarify the impact on the phenotype of novel sequence variants. Moreover, different molecular techniques are required for the detection of more complex mutations such as repeat expansions or structural (copy number) variants that are not readily identified by NGS.[34] Another WES analysis constraint is the occurrence of false-positives. Some of these sequence artifacts are not easily flagged; therefore, they may be interpreted by the software as true variants. Even though they are later excluded by conventional sequencing, they can certainly create extra burden on the analysis.

In conclusion, we present a case resolved by WES that enabled improvement of patient follow-up and management and genetic counseling of family members. A novel *CHKB* gene mutation linked to a rare form of MD was identified by WES, in which previously the conventional diagnostic approaches had failed to provide a definitive diagnosis.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

1 Shieh, P. B. Muscular dystrophies and other genetic myopathies. *Neurol. Clin.* **31**, 1009–1029 (2013).
2 Nigro, V. & Savarese, M. Genetic basis of limb-girdle muscular dystrophies: the 2014 update. *Acta Myol.* **33**, 1–12 (2014).
3 Bönnemann, C. G., Wang, C. H., Quijano-Roy, S., Deconinck, N., Bertini, E. & Ferreiro, A. *et al.* Diagnostic approach to the congenital muscular dystrophies. *Neuromuscul. Disord.* **24**, 289–311 (2014).
4 Oliveira, J., Gonçalves, A., Oliveira, M. E., Fineza, I., Pavanello, R. C. M. & Vainzof, M. *et al.* Reviewing large LAMA2 deletions and duplications in congenital muscular dystrophy patients. *J. Neuromuscul. Dis.* **1**, 169–179 (2014).
5 Nishino, I., Kobayashi, O., Goto, Y., Kurihara, M., Kumagai, K. & Fujita, T. *et al.* A new congenital muscular dystrophy with mitochondrial structural abnormalities. *Muscle Nerve* **21**, 40–47 (1998).
6 Sher, R. B., Aoyama, C., Huebsch, K. A., Ji, S., Kerner, J. & Yang, Y. *et al.* A rostrocaudal muscular dystrophy caused by a defect in choline kinase beta, the first enzyme in phosphatidylcholine biosynthesis. *J. Biol. Chem.* **281**, 4938–4948 (2006).
7 Mitsuhashi, S., Ohkuma, A., Talim, B., Karahashi, M., Koumura, T. & Aoyama, C. *et al.* A congenital muscular dystrophy with mitochondrial structural abnormalities caused by defective de novo phosphatidylcholine biosynthesis. *Am. J. Hum. Genet.* **88**, 845–851 (2011).
8 Gutiérrez Ríos, P., Kalra, A. A., Wilson, J. D., Tanji, K., Akman, H. O. & Area Gómez, E. *et al.* Congenital megaconial myopathy due to a novel defect in the choline kinase Beta gene. *Arch. Neurol.* **69**, 657–661 (2012).
9 Quinlivan, R., Mitsuahashi, S., Sewry, C., Cirak, S., Aoyama, C. & Mooore, D. *et al.* Muscular dystrophy with large mitochondria associated with mutations in the CHKB gene in three British patients: extending the clinical and pathological phenotype. *Neuromuscul. Disord.* **23**, 549–556 (2013).
10 Castro-Gago, M., Dacruz-Alvarez, D., Pintos-Martínez, E., Beiras-Iglesias, A., Delmiro, A. & Arenas, J. *et al.* Exome sequencing identifies a CHKB mutation in Spanish patient with megaconial congenital muscular dystrophy and mtDNA depletion. *Eur. J. Paediatr. Neurol.* **18**, 796–800 (2014).

11 Cabrera-Serrano, M., Junckerstorff, R. C., Atkinson, V., Sivadorai, P., Allcock, R. J. & Lamont, P. *et al.* Novel CHKB mutation expands the megaconial muscular dystrophy phenotype. *Muscle Nerve* **51**, 140–143 (2014).

12 Aoyama, C., Liao, H. & Ishidate, K. Structure and function of choline kinase isoforms in mammalian cells. *Prog. Lipid Res.* **43**, 266–281 (2004).

13 Mitsuhashi, S., Hatakeyama, H., Karahashi, M., Koumura, T., Nonaka, I. & Hayashi, Y. K. *et al.* Muscle choline kinase beta defect causes mitochondrial dysfunction and increased mitophagy. *Hum. Mol. Genet.* **20**, 3841–3851 (2011).

14 Mitsuhashi, S. & Nishino, I. Phospholipid synthetic defect and mitophagy in muscle disease. *Autophagy* **7**, 1559–1561 (2011).

15 Buermans, H. P. & den Dunnen, J. T. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta* **1842**, 1932–1941 (2014).

16 den Dunnen, J. T. & Antonarakis, S. E. Nomenclature for the description of human sequence variations. *Hum. Genet.* **109**, 121–124 (2001).

17 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E. & DePristo, M. A. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

18 Paila, U., Chapman, B. A., Kirchner, R. & Quinlan, A. R. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput. Biol.* **9**, e1003153 (2013).

19 Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).

20 Shapiro, M. B. & Senapathy, P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174 (1987).

21 Yeo, G. & Burge, C. B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **11**, 377–394 (2004).

22 Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol.* **4**, 311–323 (1997).

23 Pertea, M., Lin, X. & Salzberg, S. L. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res.* **29**, 1185–1190 (2001).

24 Desmet, F. O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M. & Béroud, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* **37**, e67 (2009).

25 Aoyama, C., Ohtani, A. & Ishidate, K. Expression and characterization of the active molecular forms of choline/ethanolamine kinase-alpha and -beta in mouse tissues, including carbon tetrachloride-induced liver. *Biochem. J.* **363**, 777–784 (2002).

26 Malito, E., Sekulic, N., Too, W. C., Konrad, M. & Lavie, A. Elucidation of human choline kinase crystal structures in complex with the products ADP or phosphocholine. *J. Mol. Biol.* **364**, 136–151 (2006).

27 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S, Greenblatt, D. M & Meng, E. C. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

28 Nascimento, A., Jou, C., Ortez, C., Hayashi, Y.K., Nishino, I. & Olivé, M. *et al.* P.15.11 Megaconial congenital muscular dystrophy in two children with mutations in the CHKB Gene. *Neuromuscul. Disord.* **23**, 822 (2013).

29 Behin, A., Laforêt, P., Malfatti, E., Pellegrini, N., Hayashi, Y. & Carlier, R. Y. *et al.* P.15.10 Megaconial myopathy presenting as a progressive limb-girdle myopathy. *Neuromuscul. Disord.* **23**, 821(2013).

30 Mitsuhashi, S. & Nishino, I. Megaconial congenital muscular dystrophy due to loss-of-function mutations in choline kinase β. *Curr. Opin. Neurol.* **26**, 536–543 (2013).

31 Valencia, C.A., Ankala, A., Rhodenizer, D., Bhide, S., Littlejohn, M. R. & Keong, L. M. *et al.* Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel. *PLoS ONE* **8**, e53083 (2013).

32 Brislin, R. P. & Theroux, M. C. Core myopathies and malignant hyperthermia susceptibility: a review. *Paediatr. Anaesth.* **23**, 834–841 (2013).

33 Ku, C. S., Cooper, D. N., Polychronakos, C., Naidoo, N., Wu, M. & Soong, R. Exome sequencing: dual role as a discovery and diagnostic tool. *Ann. Neurol.* **71**, 5–14 (2012).

34 Vasli, N. & Laporte, J. Impacts of massively parallel sequencing for genetic diagnosis of neuromuscular disorders. *Acta Neuropathol.* **125**, 173–185 (2013).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

WILEY CLINICAL GENETICS

SHORT REPORT

The new neuromuscular disease related with defects in the ASC-1 complex: report of a second case confirms *ASCC1* involvement

J. Oliveira[1,2] | M. Martins[3] | R. Pinto Leite[3] | M. Sousa[2,4,5†] | R. Santos[1,2,6†]

[1]Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto Magalhães, Centro Hospitalar do Porto, Porto, Portugal

[2]Unidade Multidisciplinar de Investigação Biomédica (UMIB), Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal

[3]Centro Hospitalar de Trás-os-Montes e Alto Douro, Unidade de Genética, Vila Real, Portugal

[4]Laboratório de Biologia Celular, Departamento de Microscopia, Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Porto, Portugal

[5]Centro de Genética da Reprodução Prof. Alberto Barros, Porto, Portugal

[6]UCIBIO\REQUIMTE, Departamento de Ciências Biológicas, Faculdade de Farmácia, Universidade do Porto, Porto, Portugal

Correspondence
Rosário Santos, PhD, Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto de Magalhães, Centro Hospitalar do Porto, Praça Pedro Nunes, 88, Porto 4099-028, Portugal.
Email: rosario.santos@chporto.min-saude.pt
Jorge Oliveira, MSc, Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto de Magalhães, Centro Hospitalar do Porto, Praça Pedro Nunes, 88, Porto 4099-028, Portugal.
Email: jorge.oliveira@chporto.min-saude.pt

Next-generation sequencing technology aided the identification of the underlying genetic cause in a female newborn with a severe neuromuscular disorder. The patient presented generalized hypotonia, congenital bone fractures, lack of spontaneous movements and poor respiratory effort. She died within the first days of life. Karyotyping and screening for several genes related with neuromuscular diseases all tested negative. A male sibling was subsequently born with the same clinical presentation. Whole-exome sequencing was performed with variant filtering assuming a recessive disease model. Analysis focused on genes known to be related firstly with congenital myopathies, extended to muscle diseases and finally to other neuromuscular disorders. No disease-causing variants were identified. A similar disorder was described in patients with recessive variants in two genes: *TRIP4* (three families) and *ASCC1* (one family), both encoding subunits of the nuclear activating signal cointegrator 1 (ASC-1) complex. Our patient was also found to have a homozygous frameshift variant (c.157dupG, p.Glu53Glyfs*19) in *ASCC1*, thereby representing the second known case. This confirms *ASCC1* involvement in a severe neuromuscular disease lying within the spinal muscular atrophy or primary muscle disease spectra.

KEYWORDS
ASC-1 complex, *ASCC1*, bone fractures, congenital, neuromuscular, whole-exome sequencing

1 | INTRODUCTION

Perinatal manifesting neuromuscular diseases include a wide variety of clinical entities such as fetal akinesia/hypokinesia, arthrogryposis multiplex congenita, spinal muscular atrophy (SMA) type I, congenital

myopathies and muscular dystrophies, and congenital myasthenic syndromes.[1,2] There is some degree of clinical overlap between these diseases and also wide genetic heterogeneity, as reflected by the vast number of genes known to be involved and the diversity of mutation types. Genes such as *TTN*, *NEB* and *RYR1* are exceptionally large, adding complexity to the genetic studies. In addition to the clinical evaluation, muscle biopsy, brain magnetic resonance imaging and electromyogram are necessary steps in the diagnostic algorithm, in order to guide the genetic study. However, these exams may not be possible in patients at the severest end of the disease spectrum with a lethal outcome. Such limitations contribute towards the high number of cases that remain elusive in terms of diagnosis. Massive parallel sequencing technology, commonly referred to as next-generation sequencing, now offers novel diagnostic possibilities for such cases. Current genetic research is applying this sequencing technology, especially with applications such as gene panels and whole-exome sequencing (WES), in an effort to solve complex muscle diseases, identifying new genetic causes and expanding mutational profiles.[3,4]

Here we report one such research study that identified the second known case originating from genetic defects in *ASCC1*, thereby confirming its involvement in a severe congenital neuromuscular disease with bone fractures.

2 | MATERIALS AND METHODS

2.1 | Human subjects

This work was approved by the ethics committee of Centro Hospitalar do Porto. Written informed consent was obtained from the participants and legal tutors. Besides the parental samples, biological samples and detailed phenotype information was only available from one of the patients (II.2, Figure 1A).

2.2 | Routine genetic tests

The screening of homozygous *SMN1* deletions was performed using polymerase chain reaction (PCR) of exons 7 and 8 followed by restriction enzyme digestion to differentiate *SMN1* from *SMN2*. Pathogenic expansions in *DMPK* were excluded by PCR and southern-blot.

Analysis of *ACTA1*, *MTM1*, *TNNT1*, *TPM2* and *TPM3* was performed using Sanger sequencing.

2.3 | Whole-exome sequencing

Whole-exome sequencing (WES) was performed using the Ion AmpliSeq Exome kit (Thermo Fisher Scientific, Foster City, CA) as previously described.[3] This strategy uses 12 primer pools (293 903 primer pairs in total) to amplify target regions covering >97% of the consensus coding sequence annotation. Sequencing was performed in the Ion Proton System (Thermo Fisher Scientific). Reads were aligned in the Ion Torrent server against the human genome assembly hg19. Sequence variants were called using the Torrent Variant Caller plugin version 4.4. Variant filtering resorted to Ion Reporter software and in-house adapted LOVD databases (version v.3.0) that store rare sequence variants and sequencing artifacts detected in our studies. Candidate variants were manually inspected on the Binary Alignment Map (BAM) file through GenomeBrowse V2.1.1 (Golden Helix Inc., Bozeman, MT).

2.4 | *ASCC1* targeted variant analysis

Conventional (Sanger) sequencing was carried out in the patient and in both parents to confirm the variant identified by WES in the *ASCC1* locus. Exon 3 was amplified using PCR oligonucleotides: 5′-CTTCCCAGGGTTCCAAGACTG-3′ (forward) and 5′-CCACTTCCCGC TGAGTTTCC-3′ (reverse). Amplicons were subjected to a further PCR using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and the resulting products were sequenced on a 3130xl Genetic Analyzer (Applied Biosystems). The reference sequence with accession number NM_001198799.2 was used to describe variants according to the Human Genome Variation Society (HGVS) nomenclature.

3 | RESULTS

A healthy apparently non-consanguineous Portuguese couple referred to genetic counseling had 1 late intrauterine fetal death, of unknown cause, at 32 weeks of gestation. This first pregnancy was followed by 2 live births—1 female and 1 male (Figure 1A). Both newborns had severe neonatal hypotonia, lack of spontaneous movements (including respiratory apparatus) and died within a few
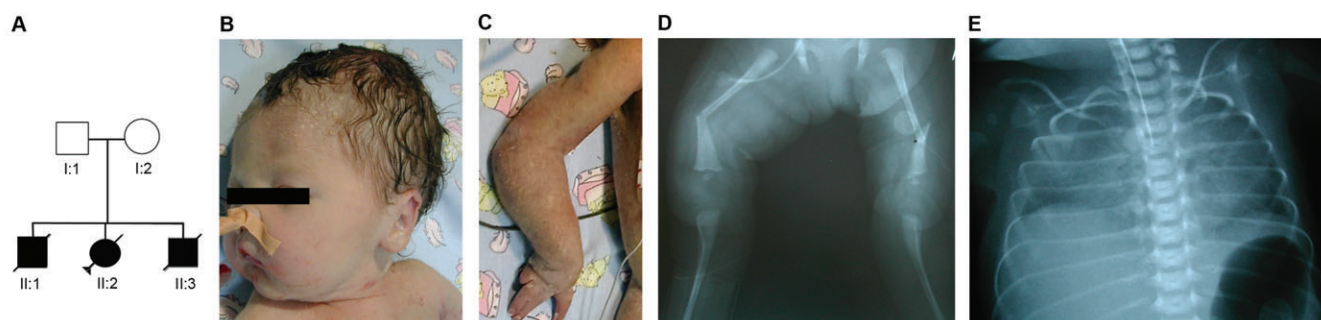
FIGURE 1 Family pedigree and patient's main clinical findings. A, Family pedigree compatible with an autosomal recessive inheritance pattern. B, Myopathic appearance, including high arched palate, tent-shaped mouth and microretrognathia. C, Hand showing signs of arthrogryposis. D, Bilateral congenital femoral fractures. E, Chest radiograph showing thin, gracile ribs
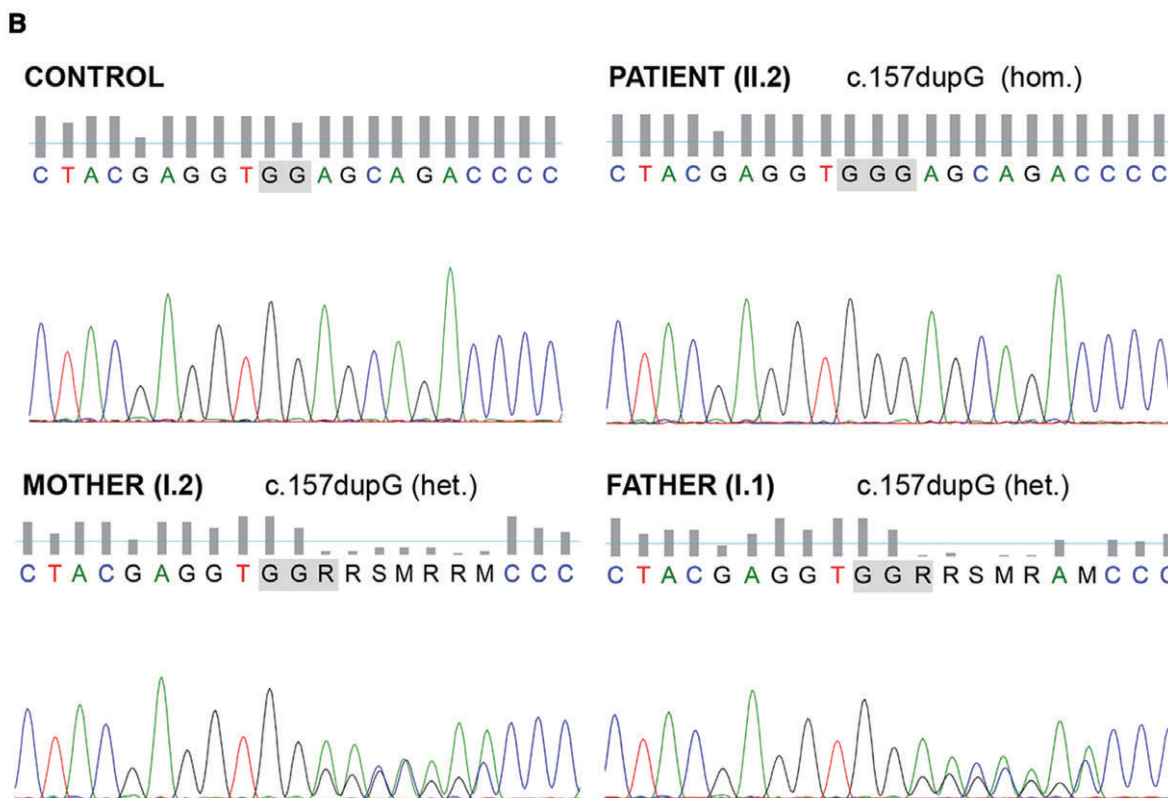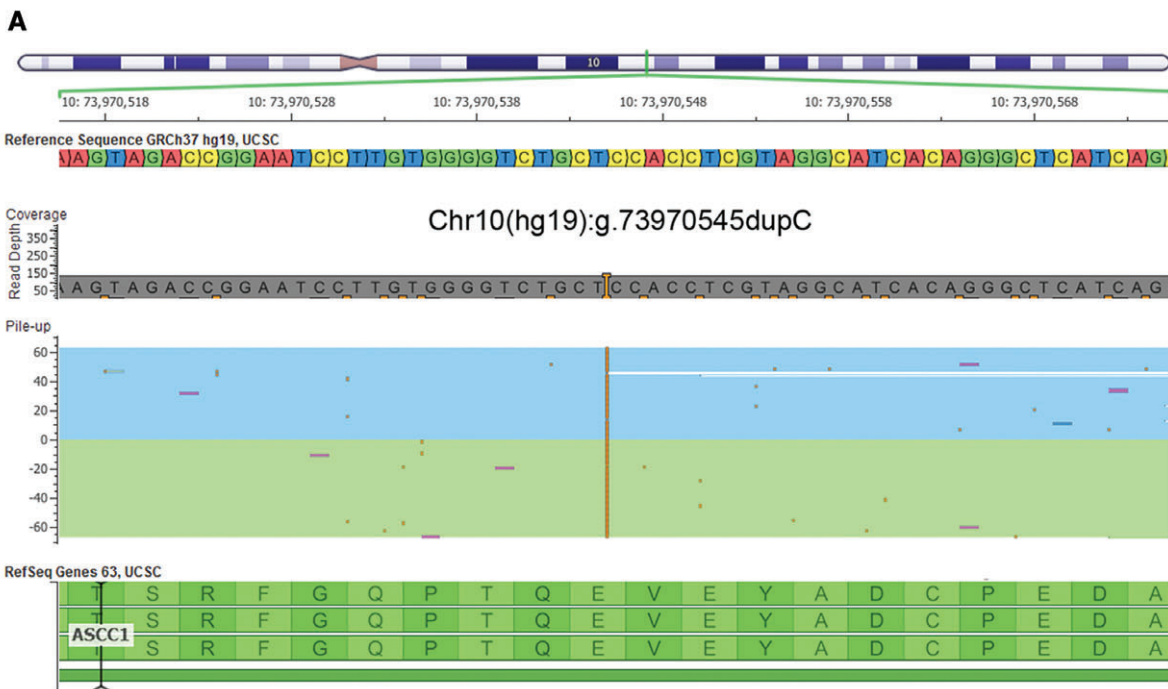
FIGURE 2 Identification of NM_001198799.2:c.157dupG variant in *ASCC1* gene. A, Visualization of the frame-shift variant located in *ASCC1* in the BAM file obtained from whole-exome sequencing data. B, Confirmation of variant by Sanger sequencing in the patient (homozygosity) and in both parents (heterozygosity)

days of life. Detailed clinical and anatomopathological data was available only for the female patient (II.2). She had a 'myopathic' appearance, with tent-shaped mouth, microretrognathia and arthrogryposis (Figure 1B,C), bilateral femoral fractures (Figure 1D) and thin, gracile ribs (Figure 1E). Reports on the histological analysis of skeletal muscle, performed in the context of autopsy, describe atrophic muscle fibers (data not shown).

Diagnostic workup tests included karyotyping and mutation screening for several genes implicated in different monogenic neuromuscular diseases, namely SMA (MIM: 25330) (*SMN1*, MIM: 600354), myotonic dystrophy (MIM: 160900) (*DMPK*, MIM: 605377) and congenital myopathy genes (*ACTA1*, MIM: 102610; *MTM1*, MIM: 310400; *TNNT1*, MIM: 191041; *TPM2*, MIM: 190990; and *TPM3*, MIM: 191031). These all tested negative.

TABLE 1 Clinical features of the patient reported in this work and comparison with published cases with ASC-1-related neuromuscular diseases

	Family 1 Davignon et al[12]	Family A Knierim et al[5]	Family B Knierim et al[5]	Family C Knierim et al[5]	Family D Knierim et al[5]	This report
Number of patients (n)	4	2	2	1	2	1 + 1[a]
Country	France	Kosovo	Kosovo	Albania	Turkey	Portugal
Locus	TRIP4	TRIP4	TRIP4	TRIP4	ASCC1	ASCC1
Pathogenic variant[b]	c.890G>A (p.Trp297*) homo.	c.760C>T (p.Arg254*) homo.	c.832C>T (p.Arg278*) homo.	c.760C>T (p.Arg254*) + c.832C>T (p.Arg278*)	c.157dupG (p.Glu53Glyfs*19) homo.	
Reduced/absent fetal movements	U	Y	Y	U	Y	Y
Poly/hydramnios	Y (n = 1)	N	N	U	Y	N
Oligohydramnios	N	Y	Y	U	N	N
Premature delivery (<37 wk)	U	Y	Y	U	Y	N
Neonatal hypotonia	Y (mainly axial)	Y	Y	Y	Y	Y
Neonatal respiratory distress	Y (n = 3)	Y	Y	Y	Y	Y
Congenital bone fractures	N	Y	Y (n = 1)	U	Y	Y
Joint contractures	Y (n = 2) (mild/mod.) after 14 years of age	Y (arthrogryposis)	Y (arthrogryposis)	Y (arthrogryposis)	Y (arthrogryposis)	Y (mild arthrogryposis)
High-arched palate	Y (n = 1)	Y (n = 1)	Y (n = 1)	U	N	Y
Micro/retrognathia	U	Y	Y	U	N	Y
Muscle weakness and atrophy	Y	Y	Y	U	Y	Y
Delayed motor milestones	Y	U	U	U	U	NA
Best motor achievement	Walk (n = 2); sit (n = 2)	U	U	U	U	NA
Cardiomyopathy	N	Y (n = 1)	Y (n = 1)	U	N	N
Skin changes	Y (hyperelasticity, dryness, follicular hyperkerastosis)	U	U	U	U	N
Brain imaging	Normal (n = 1) (CT scan)	U	U	U	Abnormal cortical gyration (MRI)	N (transfontanel ultrasonography)
Skeletal muscle histology	Fiber size variation. Frequent minicores and cap lesions, few fibers with central nuclei. (n = 3)	Fiber size variation and atrophy (n = 1). Type 1 fiber grouping	Fiber size variation and atrophy (n = 2). Type 1 fiber grouping	U	Fiber size variation and atrophy (n = 1). Type 1 fiber grouping	Fiber atrophy (limited analysis, in the context of autopsy)

[a] Patient's sibling not studied (similar phenotype with a neonatal death and fractures).

[b] cDNA reference sequence: TRIP4 (NM_016213.4); ASCC1 (NM_001198799.2).

CT, computed tomography; homo., homozygous; MRI, magnetic resonance imaging; mod, moderate; N, no; NP, not performed; NA, not applicable; U, unknown; Y, yes.

WES performed on the genomic DNA of patient II.2 targeted a total of 57.7 Mb of target regions. The mean read depth (target base coverage) of these captured regions was 145.6 (fold), with 94.51% of these regions covered by at least 20 reads (metrics determined with Torrent Suite variant caller plugin). A total of 52 830 variants were called, including 49 681 single and multiple nucleotide variants (SNVs and MNPs) and 3149 insertion/deletions (indels). For data analysis, variant filtering assumed an autosomal recessive disease model: homozygous variants or two heterozygous variants in the same locus. Variants were excluded if the minor allele frequency was above 0.1% in the population variant databases (dbSNP, 1000 genome project, exome sequencing project [ESP] and exome aggregation consortium [ExAC]). This high stringency was adopted on the premise that the disease incidence/prevalence would be extremely low. The primary strategy focused on genes known to be related with congenital myopathies (n = 24) extended to those linked to muscle diseases (n = 146, Table S1, Supporting information) and finally to all genes known to be involved in neuro-muscular diseases (n = 407). No plausible disease-causing variants were identified in any of these sets, and exome analysis proceeded with the interpretation of variants in the remaining loci. The extensive list of candidate variants included a homozygous single base pair duplication located in the coding region of ASCC1 (Figure 2A). Meanwhile, Knierim et al reported four families with a severe

prenatal SMA with respiratory distress and congenital bone fractures.[5] All cases showed an autosomal recessive inheritance pattern and were said to be caused by loss-of-function variants in two different genes (*ASCC1* and *TRIP4*), both encoding subunits of the nuclear activating signal cointegrator 1 (ASC-1) transcriptional complex.[5] These cases shared several similar features with our patient II.2 (Table 1). The variant reported in *ASCC1* (NM_001198799.2:c.157dupG, p.Glu53Glyfs*19), identified in the single family of Turkish origin, was exactly the same as that found in our patient. Our WES results were confirmed using Sanger sequencing of *ASCC1* exon 3 (Figure 2B). Similarly, this region was interrogated in the patient's parents, revealing that both are heterozygotes for the NM_001198799.2:c.157dupG variant (Figure 2B). No specimens were available for analysis of the patient's siblings (II.1 and II.3).

4 | DISCUSSION

ASC-1 is a complex composed of 4 proteins encoded by different genes: thyroid hormone receptor interactor 4 (*TRIP4*, MIM: 604501), subunit 1 (*ASCC1*, MIM: 614215), subunit 2 (*ASCC2*, MIM: 614216) and subunit 3 (*ASCC3*, MIM: 614217) of the ASC-1 complex.[5,6] This poorly understood complex is known to interact with nuclear receptors and other transcription factors.[7] More recently, based on the domains or motifs found in ASC-1 subunits, it was suggested to be a ribonucleoprotein complex involved in transcriptional coactivation of a wide range of genes and in RNA processing.[5]

Although an *ASCC1* variant had been picked up as a possible candidate in our study, information on related diseases or phenotypes was scarce. The first disease association reported was a germline *ASCC1* missense variant predisposing to a form of cancer (Barrett esophagus and esophageal adenocarcinoma).[8,9] The ASC-1 complex was also proposed to be an inhibitor of the nuclear factor kappa B (NF-κB). This transcription factor plays an important role in the development and the function of the immune system[10] and, when inadequately active, is known to be a cause of chronic inflammatory disorders and cancer. In particular, a heterozygous *ASCC1* nonsense variant detected in patients with rheumatoid arthritis was shown to abrogate the NF-κB-inhibiting capacity of ASCC1, thereby possibly acting as a risk factor for rheumatoid arthritis and modulating the disease outcome.[11]

Recently, two independent groups reported a combined total of four families with *TRIP4*-related neuromuscular disease[5,12] and a further family with an *ASCC1*-related form.[5] The pathophysiological mechanisms of these newly identified clinical entities arising from pathogenic variants in *ASCC1* and *TRIP4* remain elusive. Ascc1 and trip4 zebrafish morphants have been used to study the ASC-1-related neuromuscular phenotypes.[5] In these disease models, where gene expression knockdown is mediated by antisense morpholino oligonucleotides, there was a severe disturbance in neuromotor unit development. More specifically, the axonal outgrowth, neuromuscular junction density and the myotome were found to be compromised.[5] However, the second report of additional patients with *TRIP4* mutations, sharing clinical features with the previous patients but without bone fractures, added further speculation on the underlying disease

mechanism(s).[12] *Ex vivo* and in *in vitro* experiments performed on murine myoblastic C2C12 cell lines and on muscle primary cultures obtained from patients with *TRIP4* variants, led Davignon et al to propose that ASC-1 might be a key player in muscle development, having a role in late myogenesis and/or myotube growth.[12]

Our patient shows clinical features strikingly similar to those of the siblings described by Knierim et al,[5] namely: a neonatal (most likely pre-natal) onset, severe hypotonia, lack of spontaneous movements, contractures and congenital bone fractures.

Congenital fractures of long bones have been reported in some perinatal-manifesting neuromuscular diseases, such as SMA and primary myopathies.[13,14] These fractures are rare and are attributed to fetal akinesia or hypokinesia, where the decrease of mechanical use affects bone development.[15] As an isolated clinical finding, bone fractures are a nonspecific clinical feature. However, in conjunction with other signs, such as muscle histopathological markers, it may provide some diagnostic guidance. One such example is KLHL40-related myopathy, where muscle histology enables classification as nemaline myopathy and bone fractures (reported in approximately half of the patients) provides a high likelihood of *KLHL40* involvement.[16]

It is noteworthy that, besides sharing a similar phenotype, the variant identified in our patient is exactly the same as that found in the Turkish family reported by Knierim et al. There is no known ethnic connection between the two cases. Our patient's parents are both Portuguese and originating from the same village. Homozygosity mapping using WES data of our patient, revealed only two runs of homozygosity, with 1.8 and 0.2 Mb (in chromosomes 11 and 19, respectively) and neither region contained the disease-causing gene. This is not in agreement with high consanguinity, and it is unlikely that the couple shares a recent a common ancestor.

In conclusion, we report the second known case with a pathogenic variant in *ASCC1*. This independent description provides the confirmation that homozygous truncating variants in *ASCC1* give rise to a severe neuromuscular disease, possibly within the SMA or primary muscle diseases spectra. As for future research, it will be important to identify additional ASC-1-related patients, especially those with milder disease presentations, so as this should enable further characterization of this new clinical entity.

Conflict of interest

The authors declare no conflict of interest.

ORCID

J. Oliveira http://orcid.org/0000-0003-3924-6385

REFERENCES

1. Todd EJ, Yau KS, Ong R, et al. Next generation sequencing in a large cohort of patients presenting with neuromuscular disease before or at birth. *Orphanet J Rare Dis*. 2015;10:148.
2. Ravenscroft G, Davis MR, Lamont P, Forrest A, Laing NG. New era in genetics of early-onset muscle disease: Breakthroughs and challenges. *Semin Cell Dev Biol*. 2016;16:30241-30245.
3. Oliveira J, Negrão L, Fineza I, et al. New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing. *J Hum Genet*. 2015;60(6):305-312.
4. Oliveira J, Gonçalves A, Taipa R, et al. New massive parallel sequencing approach improves the genetic characterization of congenital myopathies. *J Hum Genet*. 2016;61(6):497-505.
5. Knierim E, Hirata H, Wolf NI, et al. Mutations in subunits of the activating signal cointegrator 1 complex are associated with prenatal spinal muscular atrophy and congenital bone fractures. *Am J Hum Genet*. 2016;98(3):473-489.
6. Jung DJ, Sung HS, Goo YW, et al. Novel transcription coactivator complex containing activating signal cointegrator 1. *Mol Cell Biol*. 2002;22(14):5203-5211.
7. Kim HJ, Yi JY, Sung HS, et al. Activating signal cointegrator 1, a novel transcription coactivator of nuclear receptors, and its cytosolic localization under conditions of serum deprivation. *Mol Cell Biol*. 1999;19(9):6323-6332.
8. Orloff M, Peterson C, He X, et al. Germline mutations in MSR1, ASCC1, and CTHRC1 in patients with Barrett esophagus and esophageal adenocarcinoma. *JAMA*. 2011;306(4):410-419.
9. van Nistelrooij AM, Dinjens WN, Wagner A, Spaander MC, van Lanschot JJ, Wijnhoven BP. Hereditary factors in esophageal adenocarcinoma. *Gastrointest Tumor*. 2014;1(2):93-98.
10. Hayden MS, Ghosh S. NF-κB in immunobiology. *Cell Res*. 2011;21(2):223-244.
11. Torices S, Alvarez-Rodríguez L, Grande L, et al. A truncated variant of ASCC1, a novel inhibitor of NF-κB, is associated with disease severity in patients with rheumatoid arthritis. *J Immunol*. 2015;195(11):5415-5420.
12. Davignon L, Chauveau C, Julien C, et al. The transcription coactivator ASC-1 is a regulator of skeletal myogenesis, and its deficiency causes a novel form of congenital muscle disease. *Hum Mol Genet*. 2016;5(8):1559-1573.
13. Ryan MM, Schnell C, Strickland CD, et al. Nemaline myopathy: a clinical study of 143 cases. *Ann Neurol*. 2001;50(3):312-320.
14. Grotto S, Cuisset JM, Marret S, et al. Type 0 spinal muscular atrophy: further delineation of prenatal and postnatal features in 16 patients. *J Neuromuscul Dis*. 2016;3(4):487-495.
15. Rodríguez JI, Garcia-Alix A, Palacios J, Paniagua R. Changes in the long bones due to fetal immobility caused by neuromuscular disease. A radiographic and histological study. *J Bone Joint Surg Am*. 1988;70(7):1052-1060.
16. Ravenscroft G, Miyatake S, Lehtokari VL, et al. Mutations in KLHL40 are a frequent cause of severe autosomal-recessive nemaline myopathy. *Am J Hum Genet*. 2013;93(1):6-18.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

Exonization of an Intronic LINE-1 Element Causing Becker Muscular Dystrophy as a Novel Mutational Mechanism in Dystrophin Gene

Ana Gonçalves [1,2,†], Jorge Oliveira [1,2,*,†] (iD), Teresa Coelho [3], Ricardo Taipa [4] (iD), Manuel Melo-Pires [4], Mário Sousa [2,5,6] and Rosário Santos [1,2,7,*] (iD)

[1] Unidade de Genética Molecular, Centro de Genética Médica Dr. Jacinto Magalhães,
    Centro Hospitalar do Porto, 4050-106 Porto, Portugal; ana.goncalves@chporto.min-saude.pt
[2] Unidade Multidisciplinar de Investigação Biomédica (UMIB),
    Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto,
    4050-313 Porto, Portugal; msousa@icbas.up.pt
[3] Serviço de Neurofisiologia, Departamento de Neurociências, Centro Hospitalar do Porto,
    4099-001 Porto, Portugal; tcoelho@netcabo.pt
[4] Unidade de Neuropatologia, Centro Hospitalar do Porto, 4099-001 Porto, Portugal;
    ricardotaipa@gmail.com (R.T.); melopires@hotmail.com (M.M.-P.)
[5] Departamento de Microscopia, Laboratório de Biologia Celular,
    Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, 4050-313 Porto, Portugal
[6] Centro de Genética da Reprodução Prof. Alberto Barros, 4050-313 Porto, Portugal
[7] UCIBIO/REQUIMTE, Departamento de Ciências Biológicas, Laboratório de Bioquímica,
    Faculdade de Farmácia, Universidade do Porto, 4050-313 Porto, Portugal
*   Correspondence: jorge.oliveira@chporto.min-saude.pt (J.O.); rosario.santos@chporto.min-saude.pt (R.S.);
    Tel.: +351-226070330 (J.O. & R.S.)
†   These authors contributed equally to this work.

Abstract: A broad mutational spectrum in the dystrophin (*DMD*) gene, from large deletions/duplications to point mutations, causes Duchenne/Becker muscular dystrophy (D/BMD). Comprehensive genotyping is particularly relevant considering the mutation-centered therapies for dystrophinopathies. We report the genetic characterization of a patient with disease onset at age 13 years, elevated creatine kinase levels and reduced dystrophin labeling, where multiplex-ligation probe amplification (MLPA) and genomic sequencing failed to detect pathogenic variants. Bioinformatic, transcriptomic (real time PCR, RT-PCR), and genomic approaches (Southern blot, long-range PCR, and single molecule real-time sequencing) were used to characterize the mutation. An aberrant transcript was identified, containing a 103-nucleotide insertion between exons 51 and 52, with no similarity with the *DMD* gene. This corresponded to the partial exonization of a long interspersed nuclear element (LINE-1), disrupting the open reading frame. Further characterization identified a complete LINE-1 (~6 kb with typical hallmarks) deeply inserted in intron 51. Haplotyping and segregation analysis demonstrated that the mutation had a de novo origin. Besides underscoring the importance of mRNA studies in genetically unsolved cases, this is the first report of a disease-causing fully intronic LINE-1 element in *DMD*, adding to the diversity of mutational events that give rise to D/BMD.

Keywords: Becker muscular dystrophy; cDNA; *DMD*; Dystrophin; LINE-1

1. Introduction

Duchenne or Becker muscular dystrophies (D/BMD), caused by pathogenic variants in the Dystrophin (*DMD*) gene, are among the most common inherited diseases of muscle, with an estimated prevalence of ~1/3800 live male births [1]. A broad mutational spectrum for D/BMD has been thoroughly described in the literature, ranging from large multi-exonic deletions/duplications to smaller single nucleotide variants [2]. More complex and rarer *DMD* mutations, such as large rearrangements and gene disruption mediated by retrotransposition activity, have also been reported [3,4]. The genetic heterogeneity, size, and complexity of the *DMD* gene demands expertise in a vast number of molecular techniques, besides the routinely used multiplex-ligation probe amplification (MLPA) and genomic sequencing. Since dystrophinopathies are now amenable to therapy, the genetic characterization of these patients has gained relevance beyond clinical follow-up and genetic counselling purposes.

We previously reported the characterization of 308 dystrophinopathy patients, from 284 unrelated families, leading to the identification of 175 distinct mutations [5]. This 91% positivity rate (284 of 312 families) was achieved in a cohort with strict inclusion criteria. Since then, and considering all referrals with clinical suspicion of D/BMD, over 100 cases remain unsolved at the genetic level.

This report describes a wide combination of genetic studies performed on a patient presenting a mild BMD phenotype, where a unique mutational event involving the insertion of a long interspersed nuclear element 1 (LINE-1) was identified.

2. Materials and Methods

2.1. Patient Samples

Formal written informed consent for publication of this case report was obtained from the patient and other family members whose data is presented. The study was conducted in accordance with the Declaration of Helsinki, and with approval of the institutional (CHP) ethics committee (Code: 336-13(196-DEFI/285-CES); date of approval: 11 December 2013).

2.2. RNA Studies

Total RNA extracted from patient and control muscle samples with PerfectPure RNA Fibrous Tissue kit (5 PRIME) was converted to cDNA using the High Capacity RNA-to-cDNA Kit (Thermo Fisher Scientific, Waltham, MA, USA). *DMD* transcripts were amplified by PCR covering the entire coding region. Amplicons were purified with Illustra ExoStar (GE Healthcare, Little Chalfont, UK) and sequenced using BigDye™ Terminator Cycle Sequencing Kit V3.1 (Thermo Fisher Scientific). Reference sequence for variant description: NM_004006.2.

2.3. Bioinformatics

Genome similarity sequence search was conducted using the Basic Local Alignment Search Tool (BLASTN v2.2.32) [6]. Analysis of repetitive elements was performed using CENSOR [7], RepeatMasker [8] and L1Xplorer [9]. Potential acceptor splice-sites and branchpoints were assessed using different algorithms available in Alamut Visual software (v2.8, Interactive Biosoftware, Rouen, France) (Supplementary Materials Figure S1).

2.4. LINE-1 Characterization

In order to identify the 5′ insertion site, primers were designed against five candidate target regions within intron 51 (Supplementary Materials Data S1). For the 3′ insertion site, a forward primer was designed against a conserved region of the LINE-1 3′UTR (L1-F: AAATTAGGTATTGATGGGACGTATT) and a reverse primer within intron 51 (51int-R: GAGAAGATGACAGTTAAATCAAAGC) (Supplementary Materials Data S1). Resultant amplicons

were sequenced as described above. LINE-1 was genotyped by single molecule real-time sequencing (PacBio RS II system, Pacific Biosciences, San Francisco, CA, USA) using custom DNA libraries (Supplementary Materials Data S1). FASTA/Q files were mapped against a LINE-1 reference and consensus sequence was obtained from BAM files using Samtools mpileup command (Supplementary Materials Data S1). Sequence artifacts and ambiguous sites were clarified via long-range PCR followed by Sanger sequencing.

3. Results

We report the genetic characterization a 50-year-old male patient with clinical features compatible with BMD, namely, onset at 13 years of age with progressive proximal weakness of lower limbs, electromyography showing myopathic signs, and high creatine kinase levels. The patient was initially referred for *DMD* molecular testing in 2001, but multiplex PCR and Southern blot failed to detect pathogenic variants. The case was re-evaluated in the context of genetic counselling of the patient's daughter. A muscle biopsy performed in the patient revealed dystrophic features and irregular staining for Dystrophin (Figure 1).
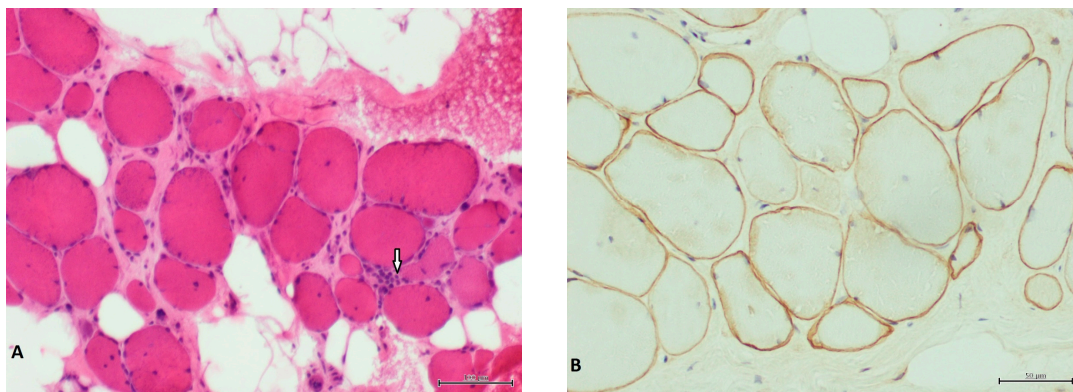
Figure 1. Patient's muscle biopsy. (**A**) Hematoxylin and eosin stain showing severe fibrosis and fat substitution. Arrow indicates a necrotic fiber. Scale bar corresponds to 100 μm. (**B**) Dys2 antibody showing irregular and faint staining of dystrophin. Scale bar corresponds to 50 μm.

Given that large deletions/duplications and point mutations were not detected by current routine genetic studies (MLPA and *DMD* genomic sequencing), complete *DMD* cDNA analysis of the muscle specimen was performed. An insertion of 103 nucleotides (not traceable in *DMD*) was identified between exons 51 and 52 (r.7542_7543ins(103), Figure 2A), predictably shifting the *DMD* open reading frame (ORF). Besides the predominant mutated transcript, a residual amount of wild-type transcript was detectable (Figure 2A). To identify the origin of the mutated sequence, a BLASTN query was performed against the human nucleotide collection. Identity of 95.1% (98/103 base pairs (bp)) was retrieved against two human LINE-1 sequences: L1.21 and L1.14 (GenBank accession numbers U93570 and U93566, respectively) (Supplementary Materials Figure S2). Comparative analysis also showed high similarity with other human genomic sequences containing LINE-1 elements. Further confirmation was obtained using CENSOR software where 98/103 bp had 100% similarity with the consensus sequence of the human LINE-1 element (Supplementary Materials Figure S2).
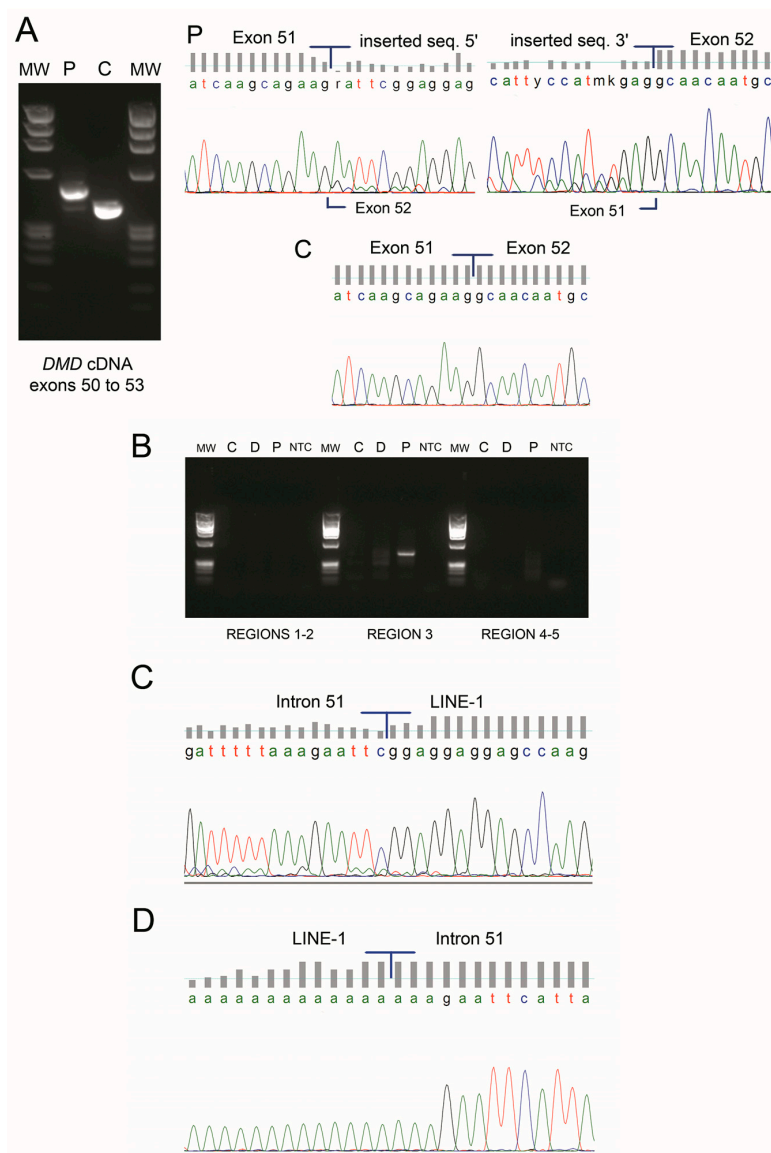
Figure 2. (**A**) cDNA analysis of *DMD* transcript revealed an abnormal PCR product with higher molecular weight in the patient (P) which was not detectable in a control sample (C). Sequencing electropherogram shows an insertion of 103 nucleotides between exons 51 and 52. A residual wild-type transcript is present in the patient sample (faint PCR band). (**B**) PCR amplification of five candidate regions to identify the long interspersed nuclear element (LINE-1) insertion site in intron 51 of *DMD* (C—control, D—control with a deletion of intron 51, P—patient, MW—molecular weight marker, NTC—no template control). A single amplicon was obtained in the patient sample for the candidate region 3. (**C**) Upon sequencing, the exact location of the LINE-1 in intron 51 was mapped to lie between positions c.7542+8951 and c.7542+8952 (NM_004006.2). (**D**) The 3′ insertion site was confirmed through a specific PCR followed by sequencing; the LINE-1 poly-A tail is also seen.

A strategy was delineated to identify the genomic insertion site in the *DMD* gene. The first five nucleotides (AATTC), having no correspondence to the LINE-1 consensus sequence, were presumed to belong to intron 51 of *DMD*. This ~44 Kb intron was scanned for the sequence AG/AATTC (AG being the canonical dinucleotide for acceptor splice-sites); a total of eight such sequences were found. Composite splice-site analysis narrowed this down to five potential sites—those presenting high splice-site scores and suitable (cryptic) branch-points located nearby (Supplementary Materials Figure S1). To identify the LINE-1 5′ insertion site by PCR, three forward

oligonucleotides were designed to encompass these five regions of interest and a single reverse oligonucleotide annealing to the known inserted LINE-1 sequence (detected by cDNA analysis). PCR experiments and subsequent sequencing showed that the LINE-1 was inserted at position NM_004006.2:c.7542+8951_c.7542+8952 of intron 51 (Figure 2B,C). The 3' end was then identified, containing a poly-A tail and a stretch of 9 bp (AAAGAATTC) consistent with a flanking target site duplication (TSD) (Figure 2D). Southern blot and hybridization was performed to estimate the size of insertion. Results revealed a ~6 Kb size increase, thus corresponding to a complete or almost complete LINE-1 element (Supplementary Materials Figure S3).

The patient's daughter was seen to be a carrier of the LINE-1 insertion mutation. Additional family members (the patient's healthy brother and two sisters presenting at-risk haplotypes) were also screened for the mutation using a LINE-1-specific PCR (Figure 3). Only the patient´s daughter tested positive, suggesting a de novo event.
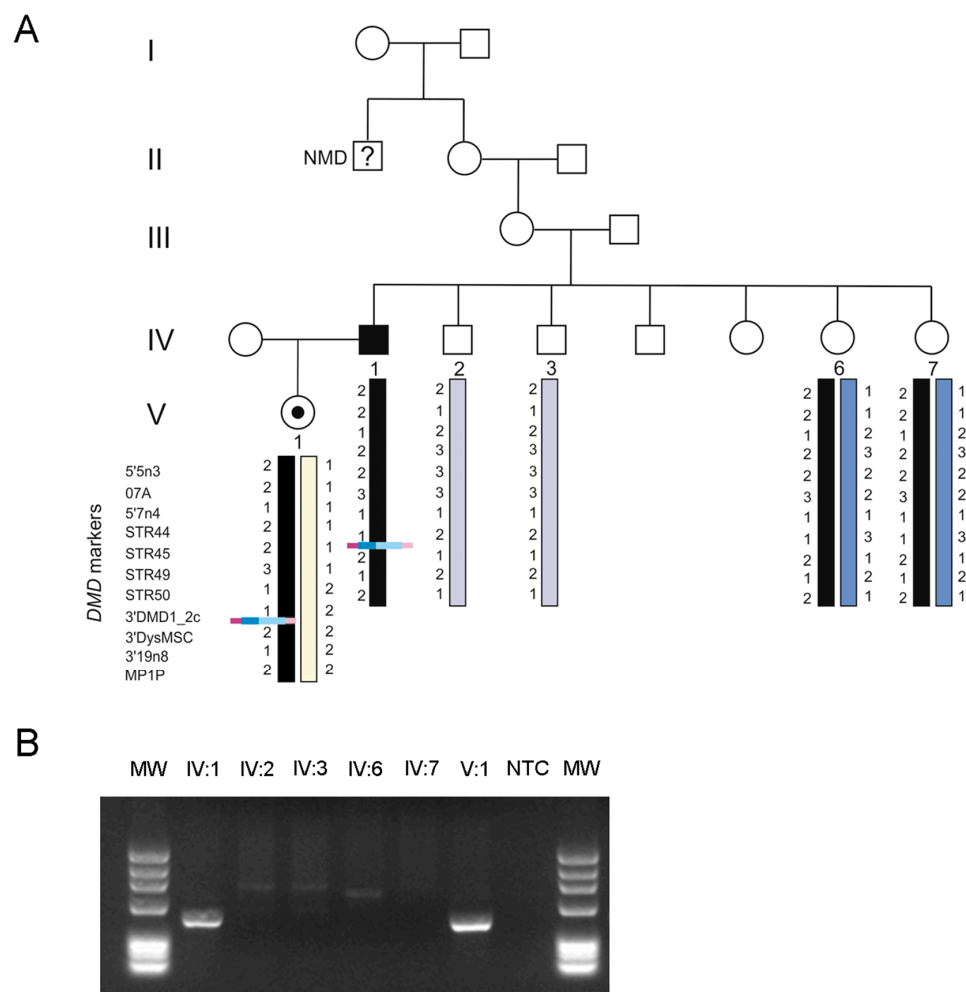
Figure 3. (**A**) Patient's family tree highlighting the segregation of the LINE-1 insertion and *DMD* haplotyping. This insertion is present in the patient (IV:1) and his daughter (V:1), and not detectable in the patient's sisters (IV:6 and IV:7 both carriers of the same at-risk haplotype). Interestingly, this family was initially thought to have an X-linked transmission, since one of the patient's maternal great-uncles (deceased) was suspected to have a neuromuscular disease (NMD). (**B**) LINE-1 specific PCR used to screen for additional carriers.

Further genotyping confirmed that a full-length LINE-1 was present (sequence available in Supplementary Materials Data S2 and submitted to GenBank, accession number MF421743).

L1Xplorer and RepeatMasker tools classified the element as a member of the L1HS subfamily, as it had all the typical hallmarks of these retrotransposons: a 5′-untranslated region (UTR), two non-overlapping ORFs (ORF1 and ORF2), a short 3′UTR and a poly-A tail (Figure 4).
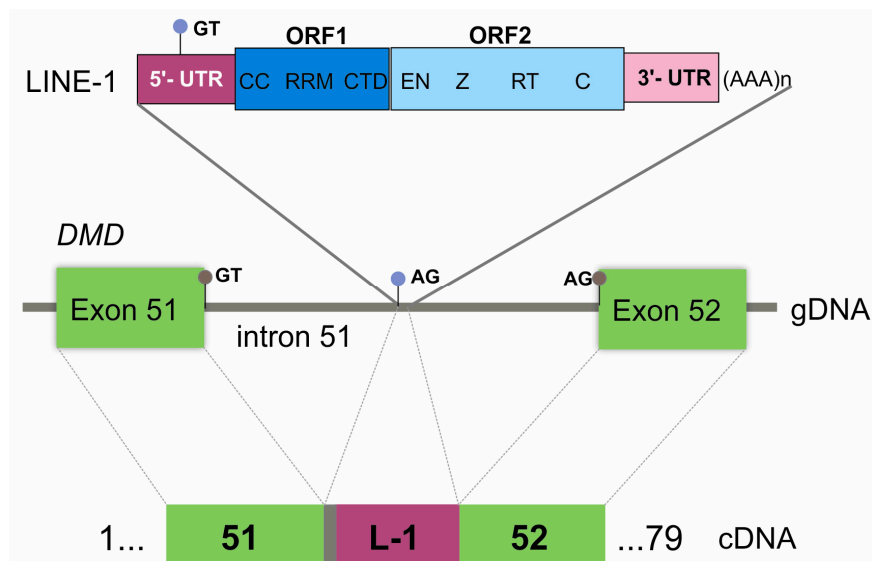
Figure 4. Graphical representation of the pathogenic LINE-1 insertion. At the genomic level (gDNA), the integration site is located in intron 51 of *DMD*. LINE-1 is displayed showing its characteristic features: 5′UTR, open reading frame (ORF) 1 (CC—coiled coiled domain, RRM—RNA recognition motif, CTD—C-terminal domain), ORF2 (EN—endonuclease, Z domain, RT—reverse transcriptase, C—cysteine-rich), and 3′ UTR with a poly-A tail. At the cDNA level, the retention of five base pairs from intron 51 (grey box), and the partial exonization of LINE-1 (L-1) (dark pink box) is explainable by the recognition of cryptic splice-sites (light blue circles) located in intron 51 (acceptor splice-site) and in the LINE-1 sequence itself (donor splice-site). Gray circles correspond to the canonical splice-sites.

4. Discussion

LINE-1s are the most abundant type of retrotransposable elements, accounting for nearly 17% of the human genome [10]. Typically, they are ~6 kb in length and exhibit characteristic components (Figure 4). ORF1 encodes an RNA-binding protein while ORF2 encodes a protein with endonuclease and reverse transcriptase activity. Although the transcriptional mechanism of LINE-1 is not fully understood, it has been proposed to involve target-site primed reverse transcription. The cDNA originated by this process recombines with genomic DNA, giving rise to characteristic signatures: a 7–20 bp direct repeat of the endonuclease target flanking the inserted LINE-1 (TSD) [11]. Only 80–100 LINE-1s in the human genome (0.1% of total) are believed to be capable of active retrotransposition [11].

Despite their importance in evolution and genome diversity, the insertion of a LINE-1 within a gene could have a deleterious effect, giving rise to disease. To date, only 30 such insertions have been reported, the majority located in exonic regions and causing frame-shifts or exon skipping [11]. In contrast, intronic LINE-1 insertions are rarely reported.

Regarding the *DMD* gene, only five pathogenic insertions have been described. Exonic disruptions, giving rise to a DMD phenotype, have been reported twice in exon 44 [12,13], and also in exons 48 and 67 [14,15]. A further pathogenic insertion was detected in two unrelated Japanese families with X-linked dilated cardiomyopathy, where a 5′-truncated form of a LINE-1 was integrated in the *DMD* 5′UTR, thought to affect the transcription or the stability of muscle transcripts [16]. A different repetitive element (Alu-like) was also reported to cause dilated cardiomyopathy, activating a cryptic acceptor splicing site in intron 11 of *DMD* [17]. The mutational event in our patient is

completely distinct in two aspects: it is a deep-intronic insertion and a full LINE-1 sequence is present. This LINE-1 was classified as a member of the L1HS subfamily, responsible for the majority of the documented LINE-1 retrotransposition events. Our results showed its partial exonization at the cDNA level, due to the recognition of a cryptic 3′ splice-site located in intron 51 and a 5′ splice-site within this element (Figure 4). This presumably gives rise to a truncated polypeptide (p.Ala2515Asnfs*21). The presence of a residual wild-type transcript explains the patient's milder dystrophinopathy (BMD phenotype). This LINE-1 sequence has 100% identity with a LINE-1 element located in chromosome 2 (GenBank accession number AC216112), which could constitute its original source. It has a near-complete identity (only one bp difference) with another pathogenic LINE-1 (GenBank accession number AF149422) inserted in the hemoglobin-beta *locus*; a seemingly active retrotransposable element of the human genome [18].

Intronic LINE-1 insertions causing exonization have only been described in three other cases: chronic granulomatous disease (*CYBB* gene) [19], Chanarin-Dorfman syndrome (*ABHD5* gene) [20], and familial retinoblastoma (*RB1* gene) [21]. The rarity of LINE-1-mediated pathogenic insertions described in the literature and in variant databases is attributable mostly to their low activity throughout the genome and to the technical difficulties in their detection (especially the full-length insertions). With massive parallel sequencing technology, there are also considerable limitations, particularly through short reads sequencing, and here, tailored bioinformatics analysis tools and strategies are required [22]. In the case of intronic LINE-1 insertions, detection may be hampered by the intron's length and the fact that it mainly affects transcriptional events (e.g., intronic retentions or exonization events). One possibility is therefore to conduct mRNA studies in yet uncharacterized B/DMD patients, as previously suggested [5,23].

It is known that repetitive transposable elements such as short interspersed nuclear elements (e.g., Alu sequences) or LINE sequences are frequent in *DMD* intronic regions. However, besides being the underlying cause of some gross rearrangements, they may also influence gene expression by mediating alternative splicing. One can thus speculate that this would ultimately interfere with the efficacy of RNA-based therapies, especially those designed to restore the reading frame. To our knowledge, this is the first report of a deep-intronic insertion of a LINE-1 element in the *DMD* gene shown to cause disease. Besides its scientific relevance, while expanding the mutational mechanisms underlying B/DMD, this finding also reinforces the need to develop comprehensive approaches to identify LINE-1 insertion profiles in the human genome.

References

1. Mendell, J.R.; Shilling, C.; Leslie, N.D.; Flanigan, K.M.; al-Dahhak, R.; Gastier-Foster, J.; Kneile, K.; Dunn, D.M.; Duval, B.; Aoyagi, A.; et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Ann. Neurol.* **2012**, *71*, 304–313. [CrossRef] [PubMed]

2. Flanigan, K.M. Duchenne and Becker muscular dystrophies. *Neurol. Clin.* **2014**, *32*, 671–688. [CrossRef] [PubMed]

3. Solyom, S.; Ewing, A.D.; Hancks, D.C.; Takeshima, Y.; Awano, H.; Matsuo, M.; Kazazian, H.H., Jr. Pathogenic orphan transduction created by a nonreference LINE-1 retrotransposon. *Hum. Mutat.* **2012**, *33*, 369–371. [CrossRef] [PubMed]

4. Ishmukhametova, A.; Chen, J.M.; Bernard, R.; de Massy, B.; Baudat, F.; Boyer, A.; Méchin, D.; Thorel, D.; Chabrol, B.; Vincent, M.C.; et al. Dissecting the structure and mechanism of a complex duplication-triplication rearrangement in the *DMD* gene. *Hum. Mutat.* **2013**, *34*, 1080–1084. [CrossRef] [PubMed]

5. Santos, R.; Gonçalves, A.; Oliveira, J.; Vieira, E.; Vieira, J.P.; Evangelista, T.; Moreno, T.; Santos, M.; Fineza, I.; Bronze-da-Rocha, E. New variants, challenges and pitfalls in *DMD* genotyping: Implications in diagnosis, prognosis and therapy. *J. Hum. Genet.* **2014**, *59*, 454–464. [CrossRef] [PubMed]

6. Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezhuk, Y.; McGinnis, S.; Madden, T.L. NCBI BLAST: A better web interface. *Nucleic Acids Res.* **2008**, *36*, W5–W9. [CrossRef] [PubMed]

7. Kohany, O.; Gentles, A.J.; Hankus, L.; Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinform.* **2006**, *7*, 474. [CrossRef] [PubMed]

8. RepeatMasker. Available online: http://www.repeatmasker.org (accessed on 3 March 2017).

9. Penzkofer, T.; Dandekar, T.; Zemojtel, T. L1Base: From functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res.* **2005**, *33*, 498–500. [CrossRef] [PubMed]

10. Kim, Y.J.; Lee, J.; Han, K. Transposable Elements: No More 'Junk DNA'. *Genomics Inform.* **2012**, *10*, 226–233. [CrossRef] [PubMed]

11. Hancks, D.C.; Kazazian, H.H., Jr. Roles for retrotransposon insertions in human disease. *Mob. DNA* **2016**, *7*, 9. [CrossRef] [PubMed]

12. Narita, N.; Nishio, H.; Kitoh, Y.; Ishikawa, Y.; Ishikawa, Y.; Minami, R.; Nakamura, H.; Matsuo, M. Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Investig.* **1993**, *91*, 1862–1867. [CrossRef] [PubMed]

13. Musova, Z.; Hedvicakova, P.; Mohrmann, M.; Tesarova, M.; Krepelova, A.; Zeman, J.; Sedlacek, Z. A novel insertion of a rearranged L1 element in exon 44 of the dystrophin gene: Further evidence for possible bias in retroposon integration. *Biochem. Biophys. Res. Commun.* **2006**, *347*, 145–149. [CrossRef] [PubMed]

14. Awano, H.; Malueka, R.G.; Yagi, M.; Okizuka, Y.; Takeshima, Y.; Matsuo, M. Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA. *J. Hum. Genet.* **2010**, *55*, 785–790. [CrossRef] [PubMed]

15. Holmes, S.E.; Dombroski, B.A.; Krebs, C.M.; Boehm, C.D.; Kazazian, H.H., Jr. A new retrotransposable human L1 element from the *LRE2* locus on chromosome 1q produces a chimaeric insertion. *Nat. Genet.* **1994**, *7*, 143–148. [CrossRef] [PubMed]

16. Yoshida, K.; Nakamura, A.; Yazaki, M.; Ikeda, S.; Takeda, S. Insertional mutation by transposable element, L1, in the *DMD* gene results in X-linked dilated cardiomyopathy. *Hum. Mol. Genet.* **1998**, *7*, 1129–1132. [CrossRef] [PubMed]

17. Ferlini, A.; Muntoni, F. The 5′ Region of intron 11 of the dystrophin gene contains target sequences for mobile elements and three overlapping ORFs. *Biochem, Biophys. Res. Commun.* **1998**, *242*, 401–406. [CrossRef] [PubMed]

18. Kimberland, M.L.; Divoky, V.; Prchal, J.; Schwahn, U.; Berger, W.; Kazazian, H.H., Jr. Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum. Mol. Genet.* **1999**, *8*, 1557–1560. [CrossRef] [PubMed]

19. Meischl, C.; Boer, M.; Ahlin, A.; Roos, D. A new exon created by intronic insertion of a rearranged LINE-1 element as the cause of chronic granulomatous disease. *Eur. J. Hum. Genet.* **2000**, *8*, 697–703. [CrossRef] [PubMed]

20. Samuelov, L.; Fuchs-Telem, D.; Sarig, O.; Sprecher, E. An exceptional mutational event leading to Chanarin-Dorfman syndrome in a large consanguineous family. *Br. J. Dermatol.* **2011**, *164*, 1390–1392. [CrossRef] [PubMed]

21. Rodríguez-Martín, C.; Cidre, F.; Fernández-Teijeiro, A.; Gómez-Mariano, G.; de la Vega, L.; Ramos, P.; Zaballos, Á.; Monzón, S.; Alonso, J. Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the *RB1* gene. *J. Hum. Genet.* **2016**, *61*, 463–466. [CrossRef] [PubMed]

22. Tica, J.; Lee, E.; Untergasser, A.; Meiers, S.; Garfield, D.A.; Gokcumen, O.; Furlong, E.E.; Park, P.J.; Stütz, A.M.; Korbel, J.O. Next-generation sequencing-based detection of germline L1-mediated transductions. *BMC Genomics* **2016**, *17*, 342. [CrossRef] [PubMed]

23. Tuffery-Giraud, S.; Miro, J.; Koenig, M.; Claustres, M. Normal and altered pre-mRNA processing in the *DMD* gene. *Hum. Genet.* **2017**. [CrossRef] [PubMed]

Homozygosity Mapping using Whole-Exome Sequencing: A Valuable Approach for Pathogenic Variant Identification in Genetic Diseases

Jorge Oliveira[1,2,*], Rute Pereira[1,*], Rosário Santos[2] and Mário Sousa[1]

[1]Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto,
R. Jorge de Viterbo Ferreira nº 228, Porto, Portugal

[2]Centro de Genética Médica Dr. Jacinto Magalhães, Centro Hospitalar do Porto,
Praça Pedro Nunes nº88, Porto, Portugal

{jorge.oliveira, rosario.santos}@chporto.min-saude.pt, ruterpereira@gmail.com, msousa@icbas.up.pt

Keywords: Whole-Exome Sequencing, Homozygosity Mapping, Next-Generation Sequencing, Clinical Genetics.

Abstract: In the human genome, there are homozygous regions presenting as sizeable stretches, or 'runs' of homozygosity (ROH). The length of these ROH is dependent on the degree of shared parental ancestry, being longer in individuals descending from consanguineous marriages or those from isolated populations. Homozygosity mapping is a powerful tool in clinical genetics. It relies on the assumption that, due to identity-by-descent, individuals affected by a recessive disease are likely to have homozygous markers surrounding the disease locus. Consequently, the analysis of ROH shared by affected individuals in the same kindred often helps to identify the disease-causing gene. However, scanning the entire genome for blocks of homozygosity, especially in sporadic cases, is not a straight-forward task. Whole-exome sequencing (WES) has been shown to be an effective approach for finding pathogenic variants, particularly in highly heterogeneous genetic diseases. Nevertheless, the huge amount of data, especially variants of unknown clinical significance, and the presence of false-positives due to sequencing artifacts, makes WES analysis complex. This paper briefly reviews the different algorithms and bioinformatics tools available for ROH identification. We emphasize the importance of performing ROH analysis using WES data as an effective way to improve diagnostic yield.

1 INTRODUCTION

Mendelian diseases are caused by pathogenic variants in genes that follow the biological inheritance laws originally proposed by Gregor Mendel. The disease-causing gene may be in an autosome or in a sex-chromosome, and may be dominant or recessive. Individually, these diseases are considered to be rare, but collectively they occur at a high rate, with an estimated 7.9 million children being born annually with a serious birth defect of genetic origin (Christianson et al., 2006).

Sanger sequencing has been the gold standard in molecular diagnostics of Mendelian diseases and is still the first choice to confirm a suspected diagnosis, enabling accurate genetic counselling. However, for diseases with genetic heterogeneity, such as hereditary myopathies or primary ciliary dyskinesia, gene-by-gene Sanger sequencing is not the most cost-effective or efficient approach (Oliveira et al., 2015; Pereira et al., 2015). Technological advances over the past decade has led to the development of high-throughput sequencing platforms, revolutionizing the sequencing capabilities and boosting the use of this so called "next-generation sequencing" (NGS) both in research and in clinical diagnostic settings. With this technology, the human genome can be completely sequenced, allowing the simultaneous analysis of multiple genes. The application of NGS in Mendelian diseases focuses firstly on exonic regions of DNA, as the majority of disease-causing mutations are found in exons or in the flanking intronic regions (Ng et al., 2010).

Marriage between close biological relatives increases the probability of the offspring inheriting two deleterious copies of a recessive gene. Thus, children from consanguineous couples have a higher incidence of autosomal recessive disorders (Bittles, 2001). In addition, as alleles are parts of haplotypes, not only will the affected descendant have two identical copies of the ancestral allele, but the

---

* Equally contributing authors.

surrounding DNA segment (haplotype) will also be homozygous. Thus, the child will be homozygous for that segment, the so-called runs of homozygosity (ROH) (McQuillan et al., 2008). These homozygous segments that are identical by descent (IBD) are generally longer in consanguinity cases. Nevertheless, even in the absence of known recent inbreeding, ROH can be detected in geographically isolated populations and historical bottlenecks (Pemberton et al., 2012). The ROH length is dependent on the degree of shared parental ancestry and its age. Recent inbreeding events/parental consanguinity tend to have longer ROH (measuring tens of Mb) since there are fewer recombination events interrupting the segments that are IBD. Conversely, older ROH are generally much shorter because the homozygous stretches have been split down by repeated meioses over the generations, with the exception of genomic regions where the recombinetion rates are lower (McQuillan et al., 2008).

2   HOMOZYGOSITY MAPPING

Homozygosity mapping (also known as autozygosity mapping), consists in the identification of homozygous regions in the genome. This is a powerful strategy to associate new genes to diseases (Alkuraya 2010; Goodship et al., 2000). As mentioned, affected individuals are likely to have two IBD alleles at markers located in the vicinity of the disease locus and thus will be homozygous for these markers. This method relies on the search for ROH that are shared by affected individuals in the same family. However, scanning the genome for blocks of homozygosity, although simple and efficient, requires sophisticated techniques such as the use of numerous microsatellite markers or high-density single nucleotide polymorphism (SNP) genotyping. For most autozygosity mapping projects, multipoint linkage analyses under a recessive disease model have been performed, with software such as GENEHUNTER (Kruglyak et al., 1996), SIMWALK2 (Sobel et al., 2002), MERLIN (Abecasis et al., 2002) or ALLEGRO (Gudbjartsson et al., 2000).

In general, haplotypes are inspected manually for homozygous regions that are shared by all affected individuals, and can be inferred to be IBD if genotypes from the parents or other close relatives are available. But, in practical terms, conventional parametric methods of multipoint linkage analysis for large datasets in complex consanguineous families are often difficult, because of the time and computational power required, since homozygous blocks among affected individuals tend to be large (mean 4.4 Mb) and contain dozens or hundreds of genes. Moreover, genotyping errors of SNP array platforms and poorly represented chromosomal regions also limit the potential of this technology. If ROH are incorrectly mapped by the introduction of erroneous heterozygous genotypes, the analysis of the causative gene will be compromised.

3   NEXT-GENERATION SEQUENCING

NGS, and in particular whole-exome sequencing (WES), prompted great advances in the study of genetic diseases and gene discovery (Boycott et al. 2013; Ng et al., 2010). However, this technology has still some limitations (Sirmaci et al., 2012). Firstly, some deleterious variations may be in non-coding regions, which cannot be detected by WES. Genetic and phenotypic heterogeneity in affected individuals makes exome sequencing difficult to interpret. Sequencing errors related to poor capture efficiency, mechanical and analytical errors, as well as misalignment of repetitive regions, lead to erroneous results. These hamper the analysis and impose the need to validate candidate causative variants by Sanger sequencing. Moreover, WES analysis imposes the need of applying filtering strategies. This step is critical as it will limit data analysis and consequently influence the results. For instance, almost half of the variants may be excluded for being synonymous. Despite the fact that these are usually not considered deleterious, numerous synonymous mutations have been implicated in human diseases (Sauna & Kimchi-Sarfaty, 2011). In addition, during WES analysis a frequency filter is usually set to exclude variants with minor allele frequency above a certain threshold (above 1% in most cases). This threshold should be set in accordance with the expected prevalence of the disease. Even so, considering that in recessive disorders carriers do not show any signs of the disease, the frequency of damaging alleles in populational variant databases can still be higher than the established threshold. This would lead to the erroneous exclusion of this variant during filtering.

Table 1: List of bioinformatic tools developed for ROH detection and their main features.

Software	OS	UI	Algorithm	Main features / experimental design	Input data files	ROH size range (Mb)	Other features	Ref.
Homozygosity-Mapper	Unix/Linux (web server [a])	GUI	Detection of homozygous blocks of selectable length	Perform autozygosity mapping from SNP arrays and NGS data	VCF files and SNP geno-types	> 1.5	• Independent of parameters like family structure/allele frequencies. • Robust against genotyping errors. • Integrated with GeneDistiller (candidate gene search engine).	Seelow et al., 2009
PLINK	Unix/Linux, Mac OS, Win.	CLI & GUI	Sliding-window	WGAS analysis tool set. Estimation and use of IBD in the context of population-based studies	BED, PED, and FAM files	[0.5 - 1.5] > 1.5	• Makes a variety of standard association tests. • Maps disease loci that contain multiple rare variants in a population-based linkage analysis. • Integrated with Haploview.	Purcell et al., 2007
GERMLINE	Unix/Linux	CLI	Sliding-window	Designed for genome-wide discovery of IBD segments shared within large populations (SNP arrays)	PED, MAP and Hap-map files	> 1.5	• Overcomes the computational barrier of pairwise analysis and can scale the analysis linearly with the sample size.	Gusev et al., 2008
HomSI	Unix/Linux, Mac OS, Win.	GUI	Sliding-window	Identify ROH in consangui-neous families from NGS data	VCF files	> 1.5	• Takes into account the distribution of the variants within genomic coordinates. • Reported to be consistent with data derived from SNP microarrays.	Gormez et al., 2014
H³M²	Unix/Linux	CLI	Heterogeneous hidden Markov model	Analyse medium and short ROH obtained from WES data	BAM files	< 0.5 [0.5-1.5] > 1.5	• Reported to be more accurate than GERMLINE and PLINK, especially in the detection of short and medium ROHs.	Magi et al., 2014
Agile-Genotyper and Agile-VariantMapper	Win.	GUI	User-controllable visualization of homozygous regions	ROH analysis WES data and SNP genotyping file data	SAM; tab-delimited text files	> 1.5	• AgileVariantMapper uses the genotypes of all positions found to be polymorphic. • AgileGenotyper deduces genotypes from positions previously found to be polymorphic in the 1000 Genomes Project data set.	Carr et al., 2013

Footnote: [a]-http://www.homozygositymapper.org; CLI- command-line interface; GUI- graphical user interface; NGS-Next-generation sequencing; OS- operative system; Ref.- references; ROH- runs of homozygosity; SNP- single nucleotide polymorphism; UI- user interface; VCF- variant call format; WES- whole-exome sequencing; WGAS- whole-genome association studies; Win.- Windows.

3.1 Homozygosity Mapping using WES Data

Recent studies have clearly demonstrated the power and the effectiveness of applying homozygosity mapping to WES data, in an attempt to identify causative genes for Mendelian disorders (Gillespie et al., 2014; Shamseldin et al., 2015). This approach has the advantage of unraveling the causal variant irrespective of the gene involved. The homozygosity map allows narrowing down of the target data sets; examination at the base-pair level then enables identification of candidate causative variants.

This approach would start by mapping WES reads against a reference genome (human hg19 in our examples). Data derived from whole-genome and RNA sequencing can also be used. This step is usually performed through a bioinformatic pipeline that generates SAM/ BAM (Sequence /Binary Alignment Map) and, at the end of the process, a variant call format (VCF) file. These files are then used as input for homozygosity mapping analysis. The position and zygosity of the obtained sequence variants can be used to retrieve/infer ROH regions. As all bioinformatic approaches, there is an error rate associated that is difficult to estimate, since it is highly dependent on the WES metrics and the sequencing platform used.

Table 1 lists some of the tools available to perform ROH. For instance, the web-based tool HomozygosityMapper (Seelow & Schuelke 2012), allows users to interactively analyze NGS data for homozygosity mapping. Furthermore, PLINK (Purcell et al., 2007) and GERMLINE (Gusev et al., 2008), originally developed for the analysis of SNP array data, are tools based on sliding-window algorithms. In a sliding window analysis, the statistics are calculated for a small frame of the data. The window incrementally advances across the region of interest and, at each new position, the reported statistics are calculated. In this way, chromosomes are scanned by moving a window of a fixed size along their entire length and variation in genetic markers across the region of interest can be measured. This type of analysis reveals how variation patterns change across a surveyed genomic segment (Srinuandee & Satirapod 2015). EXome-HOMozygosity is an example in which a sliding-window algorithm (PLINK) is applied for WES-based ROH detection (Pippucci et al., 2011).

However, the sliding-windows approaches cannot be used easily with short/medium ROH sizes. In order to solve this issue, Magi et al proposed a new algorithm, H^3M^2, that is capable of detecting smaller ROH (Magi et al., 2014). AgileGenotyper (Carr et al., 2013) and HomSI (Gormez et al., 2014) are other tools that can be used for the graphical visualization of ROHs.

3.2 Case Studies

The following examples are shown to elucidate the relevance and limitations of WES-based ROH. Two patients listed in Table 1 have recessive conditions caused by homozygous pathogenic variants identified by WES. There was no indication of parental consanguinity, and patient P2 had a sibling affected by the same condition.

In both cases, retrospective analysis was based on genome-wide homozygosity mapping performed using HomozygosityMapper algorithm.

Data obtained from patient P1 revealed the presence of two long stretches of homozygous SNPs in chromosomes 1 and 17 (with 19 and 17 Mb, respectively) (Figure 1, top panel). The affected gene (*CCDC103*) is located in one of these regions, specifically in chromosome 17. In this example the proposed approach would be suitable to generate a considerably shorter list of candidate *loci*, and consequently reduce the number of variants to be evaluated in terms of their pathogenicity. In the second example the same strategy was applied in a patient with a rare neuromuscular disease.

Table 2: Cases selected to illustrate the use of homozygosity mapping using WES data in patients with rare genetic diseases.

| Patient	Phenotype	Genotype	Reference
P1, male adult	Infertility due to total sperm immotility. Clinical features compatible with Kartagener syndrome. No known parental consanguinity.	Homozygous pathogenic variant in *CCDC103* gene: Chr17(h19):g.42978470G>C	Pereira et al., 2015.
P2, female neonate	Severe neuromuscular disease, congenital hypotonia, respiratory distress and bone fractures. No known parental consanguinity.	Homozygous pathogenic variant in *ASCC1* gene: Chr10(hg19):g.73970545dupC	Oliveira et al., (submitted).
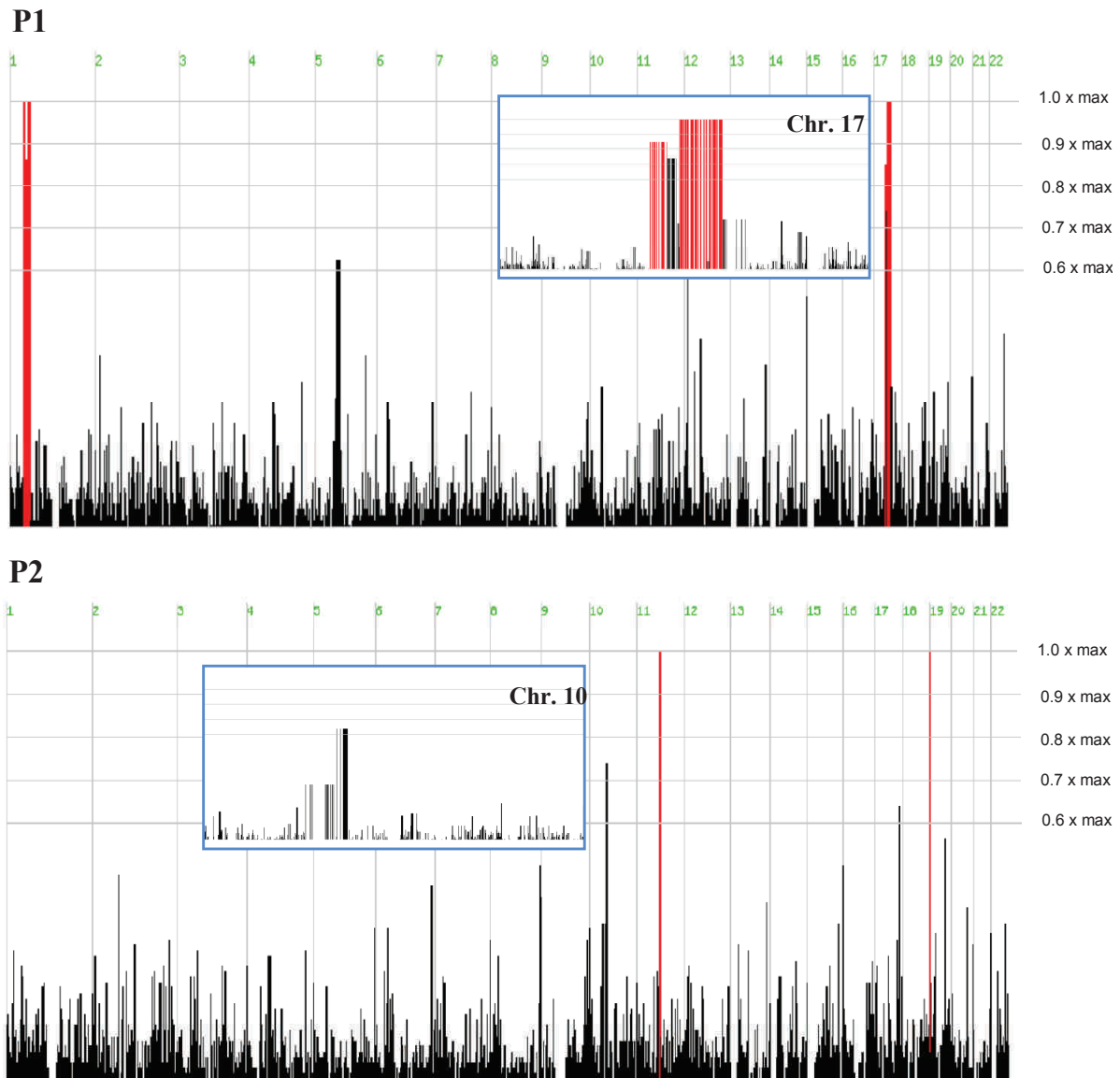
Figure 1: Genome-wide homozygosity mapping using WES and the HomozygosityMapper software. Results are shown for patients P1 and P2. Longer ROH are shown in red color. Inserted blue boxes show the genomic regions in more detail where the disease-related gene is located.

Results for patient P2 revealed two smaller ROH (1.8 and 0.2 Mb) in chromosomes 11 and 19 respectively (Figure 1, bottom panel). However, the *ASCC1* gene, carrying the disease-causing variant, is located in chromosome 10. In this last example, and considering the algorithm used, the analysis based on the assumption of a homozygous variant located in a ROH would clearly fail. The causal gene would not be included in the list of candidate genes, most likely due to the size of the actual ROH where the genetic defect is located. In the first example, the long ROH can be attributed to consanguinity

(although it was not formally confirmed), while in the second case, a distant common ancestor could explain the presence of a rare pathogenic variant in a smaller ROH tract.

4 CONCLUSIONS

The aim of this position paper is to revisit homozygosity mapping as an important tool for clinical genetics in cases where a recessive disease is suspected. We have reviewed the proposed

algorithms and available bioinformatic tools designed for ROH detection based on WES data. Selection of appropriate algorithms should mainly consider specific features of the case under study, such as the genetic context, the ROH size and the number of relatives affected by the same condition.

Monogenic disorders have been studied by classical approaches aiming to unravel several disease-causing genes. The presence of numerous genes in a candidate genomic region was a limiting factor, considering the costs and the time required, to screen mutations by Sanger sequencing. Another limitation with the "traditional" approaches is more evident when they are used to study trios (only the parents and their affected child) or larger families with only one or two affected members. The genotype data extracted in these familial contexts generally remain statistically insufficient for classical analytical approaches (Jorde, 2000).

As compared with conventional homozygosity mapping that uses known SNPs, WES has the added advantage of allowing the identification of the actual disease-causing variant. Therefore, instead of using two different procedures, one for identifying candidate loci and the other to identify the genetic defect itself at the nucleotide level, both can be performed in a single step by WES. Nonetheless, there are still limitations and further bioinformatic developments are required. Considering the examples presented, there are sensitivity issues especially if the genetic defect is in a small ROH. Finally, we consider that it would be useful to develop a bioinformatic tool that combines variant filtering and homozygosity mapping (Appendix), which currently need to be performed separately.

ACKNOWLEDGEMENTS

REFERENCES

Abecasis, G. R. et al., 2002. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, 30(1), pp.97–101.

Alkuraya, F. S., 2010. Homozygosity mapping: One more tool in the clinical geneticist's toolbox. *Genetics in Medicine*, 12(4), pp.236–239.

Bittles, A. H., 2001. Consanguinity and its relevance to clinical genetics. *Clinical genetics*, 60(2), pp.89–98.

Boycott, K. M. et al., 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, 14(10), pp.681–691.

Carr, I. M. et al., 2013. Autozygosity Mapping with Exome Sequence Data. *Human Mutation*, 34(1), pp.50–56.

Christianson, A., Howson, C.P. & Modell, B., 2006. *Global report on birth defects: the hidden toll of dying and disabled children*, New York. Available at: http://www.marchofdimes.org/materials/global-report-on-birth-defects-the-hidden-toll-of-dying-and-disabled-children-full-report.pdf [Accessed November 15, 2016].

Gillespie, R. L., Lloyd, I. C. & Black, G. C. M., 2014. The Use of Autozygosity Mapping and Next-Generation Sequencing in Understanding Anterior Segment Defects Caused by an Abnormal Development of the Lens. *Human Heredity*, 77(1–4), pp.118–137.

Goodship, J. et al., 2000. Report Autozygosity Mapping of a Seckel Syndrome Locus to Chromosome 3q22.1-q24. *Am. J. Hum. Genet*, 67, pp.498–503.

Gormez, Z., Bakir-Gungor, B. & Sagiroglu, M. S., 2014. HomSI: a homozygous stretch identifier from next-generation sequencing data. *Bioinformatics*, 30(3), pp.445–447.

Gudbjartsson, D. F. et al., 2000. Allegro, a new computer program for multipoint linkage analysis. *Nature Genetics*, 25(1), pp.12–13.

Gusev, A. et al., 2008. Whole population, genome-wide mapping of hidden relatedness. *Genome Research*, 19(2), pp.318–326.

Jorde, L. B., 2000. Linkage disequilibrium and the search for complex disease genes. *Genome research*, 10(10), pp.1435–1444.

Kruglyak, L. et al., 1996. Parametric and Nonparametric Linkage Analysis: A Unified Multipoint Approach. *Am. J. Hum. Genet*, 58, pp.1347–1363.

Magi, A. et al., 2014. H3M2: detection of runs of homozygosity from whole-exome sequencing data. *Bioinformatics (Oxford, England)*, 30(20), pp.2852–2859.

McQuillan, R. et al., 2008. Runs of homozygosity in European populations. *American journal of human genetics*, 83(3), pp.359–372.

Ng, S. B. et al., 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*, 42(1), pp.30–35.

Oliveira, J. et al., 2015. New splicing mutation in the choline kinase beta (CHKB) gene causing a muscular dystrophy detected by whole-exome sequencing. *Journal of Human Genetics*. pp. 305-312.

Oliveira, J. et al. 2017. Confirmation of a neuromuscular phenotype related with defects in the ASC-1 complex: report of the second case due to *ASCC1* pathogenic variants. *Submitted*.

Pemberton, T.J. et al., 2012. Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics,* 91(2), pp.275–292.

Pereira, R. et al., 2015. Mutation analysis in patients with total sperm immotility. *Journal of assisted reproduction and genetics*, pp.1–10.

Pippucci, T. et al., 2011. EX-HOM (EXome HOMozygosity): A Proof of Principle. *Human Heredity*, 72(1), pp.45–53.

Purcell, S. et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81(3), pp.559–575.

Sauna, Z. E. & Kimchi-Sarfaty, C., 2011. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics*, 12(10), pp.683–691.

Seelow, D. et al., 2009. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic acids research*, 37(Web Server issue), pp.W593-9.

Seelow, D. & Schuelke, M., 2012. Homozygosity Mapper2012—bridging the gap between homozygosity mapping and deep sequencing. *Nucleic acids research*, 40(W1), pp.W516–W520.

Shamseldin, H.E. et al., 2015. Identification of embryonic lethal genes in humans by autozygosity mapping and exome sequencing in consanguineous families. *Genome Biology*, 16(1), p.116.

Sirmaci, A. et al., 2012. Challenges in Whole Exome Sequencing: An Example from Hereditary Deafness I. Schrijver, ed. *PLoS ONE*, 7(2), p.e32000.

Sobel, E., Papp, J.C. & Lange, K., 2002. Detection and Integration of Genotyping Errors in Statistical Genetics. *Am. J. Hum. Genet*, 70, pp.496–508.

Srinuandee, P. & Satirapod, C., 2015. Use of genetic algorithm and sliding windows for optimising ambiguity fixing rate in GPS kinematic positioning mode. *Survey Review*, 47(340), pp.1–6.

APPENDIX
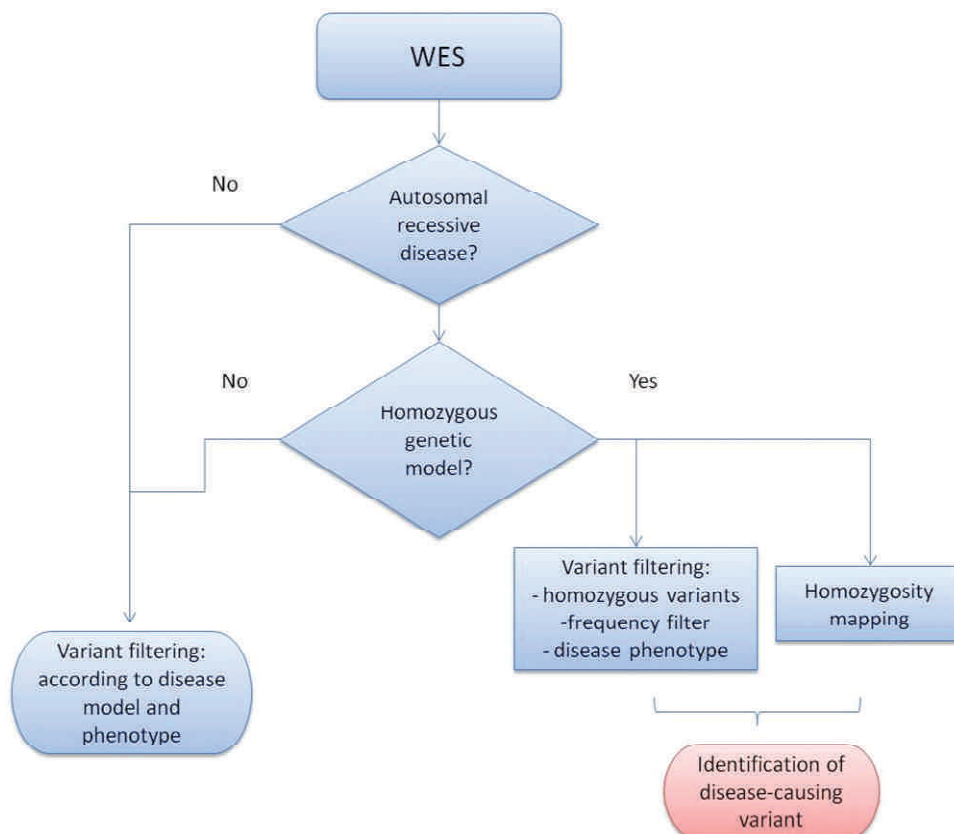
Appendix: Integration of homozygosity mapping in the analysis workflow of whole-exome sequencing (WES).