Assessment of copy number variations in the nebulin gene and other nemaline myopathy-causing genes

Kirsi Kiiski

The Folkhälsan Institute of Genetics and
the Department of Medical and Clinical Genetics, Medicum

Division of Genetics, Department of Biosciences
Faculty of Biological and Environmental Sciences

Integrative Life Sciences Doctoral Programme

University of Helsinki
Helsinki, Finland

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

folkhälsan

ACADEMIC DISSERTATION

To be presented for public examination, with the permission of the Faculty of
Biological and Environmental Sciences of the University of Helsinki,
in Lecture Hall 3, Biomedicum Helsinki, on 6th November 2015, at 12 o’clock.

Helsinki, 2015

Supervisors:

Docent Carina Wallgren-Pettersson, MD
The Folkhälsan Institute of Genetics and the Department of Medical and Clinical Genetics, Medicum, University of Helsinki
Docent Katarina Pelin, PhD
Division of Genetics, Department of Biosciences, University of Helsinki

Thesis Advisory Committee:

Docent Nina Horelli-Kuitunen, PhD
Yhtyneet Medix Laboratoriot Oy, Helsinki, Finland
Docent Maija Wessman, PhD
The Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland
Institute for Molecular Medicine Finland (FIMM), University of Helsinki

Reviewers:

Docent Janna Saarela, MD, PhD
Institute for Molecular Medicine Finland (FIMM), University of Helsinki
Docent Maija Wessman, PhD
The Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland
Institute for Molecular Medicine Finland (FIMM), University of Helsinki

Official Opponent:

Assistant Professor Nicole de Leeuw, PhD
Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Medical Centre, Nijmegen, the Netherlands

Custos:

Professor Minna Nyström, PhD
Division of Genetics, Department of Biosciences, University of Helsinki

ISBN 978-951-51-1621-5 (paperback)
ISBN 978-951-51-1622-2 (PDF)
http://ethesis.helsinki.fi

Unigrafia Oy
Helsinki, 2015

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications. In addition, some unpublished results (U) are presented.

I. **Kiiski K**, Laari L, Lehtokari V-L, Lunkka-Hytönen M, Angelini C, Petty R, Hackman P, Wallgren-Pettersson C, Pelin K. Targeted array comparative genomic hybridization – a new diagnostic tool for the detection of large copy number variations in nemaline myopathy-causing genes. Neuromuscul Disord. 2013 Jan; 23(1):56-65.

II. Lehtokari V-L, **Kiiski K**, Sandaradura S, Laporte J, Repo P, Frey JA, Donner K, Marttila M, Saunders C, Barth P, den Dunnen J, Beggs A, Clarke N, North KN, Laing N, Romero NB, Winder T, Pelin K, Wallgren-Pettersson C. Mutation update: the spectra of nebulin variants and associated myopathies. Hum Mutat. 2014 Dec; 35(12):1418-26.

III. **Kiiski K**, Lehtokari V-L, Löytynoja A, Ahlstén L, Laitila J, Wallgren-Pettersson C, Pelin K. A recurrent copy number variation of the *NEB* triplicate region: only revealed by the targeted nemaline myopathy CGH array. Eur J Hum Genet. 2015 Jul 22. doi: 10.1038/ejhg.2015.166. Epub ahead of print.

IV. **Kiiski K**, Lehtokari V-L, Manzur AY, Sewry C, Zaharieva I, Muntoni F, Pelin K, Wallgren-Pettersson C. A large deletion affecting *TPM3*, causing severe nemaline myopathy. J Neuromuscul Dis. 2015 doi: 10.3233/JND-150107. Epub ahead of print.

The publications are referred to in the text by their Roman numerals.

The articles are reprinted with the permission of the copyright owners.

ABBREVIATIONS

A	adenine
aCGH	array comparative genomic hybridization
ACTA1	the gene encoding skeletal muscle-specific α-actin
AD	autosomal dominant inheritance
AR	autosomal recessive inheritance
ATP	adenosine triphosphate
bp	base pair
C	cytosine
cDNA	complementary DNA
CBS	circular binary segmentation algorithm
CFL2	the gene encoding cofilin 2
CGH	comparative genomic hybridization
CNP	copy number polymorphism
CNV	copy number variation
Condel	Consensus deleteriousness software
DECIPHER	DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources
del	deletion
DGV	Database of Genomic Variants
dHPLC	denaturing high-performance liquid chromatography
DSB	double-stranded break
DMD	Duchenne muscular dystrophy
DNA	deoxyribonucleic acid
dup	duplication
ECARUCA	European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations
FoSTeS	Fork Stalling and Template Switching
G	guanine
GRCh37	Genome Reference Consortium Human Build 37
hg19	Human Genome Build 19
HGVS	Human Genome Variation Society
IGV	Integrative Genomics Viewer
indel	(small) insertion or deletion
kb	kilobase

KBTBD13	the gene encoding kelch repeat and BTB domain-containing protein 13
kDa	kiloDalton
KLHL40	the gene encoding kelch-like protein 40 (*KLHL40 =KBTBD5*)
KLHL41	the gene encoding kelch-like protein 41 (*KLHL41=KBTBD10*)
LCR	low-copy repeat
LINE	long interspersed nuclear element
LMOD3	the gene encoding leimodin-3
MLPA	multiplex ligation-dependent probe amplification
MMBIR	microhomology-mediated break-induced replication
MMEJ	microhomology-mediated end joining
NAHR	non-allelic homologous recombination
NEB	the gene encoding nebulin
NGS	next generation sequencing
NHEJ	non-homologous end joining
NM	nemaline myopathy
nt	nucleotide
OMIM	Online Mendelian Inheritance in Man database
PCR	polymerase chain reaction
Polyphen	Polymorphism Phenotyping software
RNA	ribonucleic acid
SD	segmental duplication
SIFT	Sorting Intolerant from Tolerant software
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
SSCP	single-stranded conformation polymorphism
T	thymine
TNNT	the gene(s) encoding troponin T
TPM2	the gene encoding β-tropomyosin
TPM3	the gene encoding α-tropomyosin$_{\text{slow}}$
TRI	triplicate region of nebulin covering exons 82-105
TTN	the gene encoding titin
VUS	variant of unknown significance
WES	whole-exome sequencing
WGS	whole-genome sequencing
YBX3	the gene encoding Y box binding protein 3 (=*CSDA*).

ABSTRACT

Nemaline myopathy (NM) and related disorders constitute a heterogeneous group of congenital myopathies. Mutations in the nebulin gene (*NEB*) are the main cause of the recessively inherited form. *NEB* is one of the largest genes in the human genome consisting of 249 kb of genomic sequence. *NEB* contains 183 exons and a 32 kb homologous triplicate region (TRI) where eight exons are repeated three times.

The aims of this Doctoral Thesis study were to develop and implement into diagnostics new efficient variant analysis methods for *NEB* and other NM-causing genes. The first aim was to design and validate a custom copy number microarray targeting the NM-causing genes for the detection of copy number variations. MLPA (multiplex ligation-dependent probe amplification) and Sanger sequencing were also used. The second aim was to utilise whole-exome sequencing to search for novel disease-causing variants in the known NM genes and try to identify novel NM genes. Lastly, the aim was to collect more data in order to try to find genotype-phenotype correlations of *NEB*-caused NM.

The design and validation of the NM-CGH microarray was successful. Of the total sample cohort of 356 NM families, 196 NM families were studied using the custom-made NM-CGH array. Nine different novel large causative variants were identified in ten NM families. The size of these variants varies greatly, covering only a part of one *NEB* exon on up to dozens of *NEB* exons (72bp - 133 kb). In addition, a novel recurrent variation of the *NEB* TRI region was identified in 13% of the NM families and in 10% of the studied 60 control samples. Deviations of one copy are suggested to be benign but gains of two or more copies might be pathogenic. One novel homozygous deletion was also identified in another NM gene, *TPM3*, in a patient with severe NM. Furthermore, ten samples were studied using exome sequencing, and for six of those samples, novel disease-causing variant(s) were identified. Two variants were identified in one family in a novel, putative NM gene that is currently under further investigation.

165 NM families from the total cohort of 356 NM families have been identified thus far with two pathogenic *NEB* variants. Altogether 220 different pathogenic variants were identified in these 165 families, accentuating that the patients in the majority (84%) of the families are compound heterozygous for two different *NEB* variants. Most of the variants are small, containing splice-site mutations (33%), small indels (33%), nonsense (22%) and missense mutations (7%). Large variants are the smallest category (5%), however, copy number variations are much more frequent than previously thought. Genotype-phenotype correlations between the type of *NEB* mutation and the NM subtypes remained, however, unobtainable.

The NM-CGH microarray has been implemented into molecular diagnostics of NM. Using the NM-CGH microarray followed by exome-sequencing has accelerated mutation detection. This combination has increased the coverage of the NM genes and thus improved the diagnostics of NM and NM-related disorders.

TIIVISTELMÄ

Nemaliinimyopatia (NM) ja samankaltaiset taudit on heterogeeninen tautiryhmä synnynnäisten myopatioiden joukossa. Nebuliini-geenin (*NEB*) mutaatiot ovat yleisin resessiivisen NM:n aiheuttaja. *NEB* on kooltaan 249 kb, eli yksi ihmisen suurimmista geeneistä. *NEB* sisältää 183 eksonia ja se kattaa myös toistojaksoja, kuten homologisen triplikaatioalueen (TRI), jossa kahdeksan eksonia toistuu kolme kertaa.

Väitöskirjatutkimuksen tarkoituksena oli kehittää ja ottaa diagnostiseen käyttöön uusia mutaatioanalyysimenetelmiä *NEB*-geeniä sekä muita NM-geenejä varten. Tavoitteena oli suunnitella ja validoida NM-geeneihin kohdennettu mikrosiru, jolla voidaan tutkia kopiolukuvariaatioita näistä geeneistä. Lisäksi käytettiin MLPA-menetelmää (multiplex ligation-dependent probe amplification) ja Sanger-sekvensointia. Eksomisekvensointia hyödynnettiin tautia aiheuttavien varianttien löytämiseksi tunnetuista NM-geeneistä sekä uusista geeneistä. Tavoitteena oli lisäksi kerätä lisätietoa nebuliinimutaatioiden aiheuttaman nemaliinimyopatian genotyyppi-fenotyyppi -korrelaatiosta.

NM-CGH-mikrosirun kehittäminen sekä validointi onnistui hyvin. Koko 356 NM-perheen näytekohortista 196 perhettä tutkittiin NM-CGH-sirulla, joista tunnistettiin yhteensä yhdeksän uutta suurta patogeenistä varianttia kymmenestä eri NM-perheestä. Näiden mutaatioiden koko vaihtelee suuresti, kattaen vain osan yhdestä *NEB*-geenin eksonista aina yli puoleen koko geenistä (72 bp – 133 kb). Lisäksi osoitettiin että 13 % tutkituista NM-perheistä sekä 10 % tutkituista 60 kontrollinäytteestä sisältää *NEB*:n triplikaatioalueen kopiolukuvariaation. Tutkimustulosten perusteella yhden kopioluvun lisäys tai vähenemä olisi harmitonta mutta mikäli ylimääräisiä kopioita on kaksi tai enemmän, se voisi olla tautia aiheuttavaa. Lisäksi tunnistettiin homotsygoottinen suuri deleetio toisesta tunnetusta NM-geenistä, *TPM3*. Eksomisekvensoinnilla löydettiin puolestaan toinen tai molemmat tautia aiheuttavat variantit kuudelle kymmenestä tutkitusta NM-potilaasta. Yhdessä NM-perheessä tunnistettiin kaksi uutta varianttia potentiaalisessa uudessa NM-geenissä, ja tätä löydöstä tutkitaan parhaillaan tarkemmin.

Tutkimuskohorttimme 356 NM-perheestä 165 perheelle on nyt tunnistettu kaksi tautia aiheuttavaa *NEB*-varianttia. Näissä perheissä esiintyi yhteensä 220 eri patogeenistä *NEB*-varianttia eli suurin osa potilaista (84 %) on yhdistelmäheterotsygootteja. Pääosa mutaatioista on splice-site -mutaatioita (33 %), pieniä insertioita tai deleetioita (33 %), nonsense- (22 %) ja missense-mutaatioita (7 %). Harvinaisimpia ovat suuret kopiolukumuutokset (5 %) mutta näiden osuus on kuitenkin huomattavasti suurempi kuin on aiemmin oletettu. *NEB*-mutaatioiden ja NM-fenotyypin välille ei kuitenkaan onnistuttu saamaan genotyyppi-fenotyyppi –korrelaatiota.

NM-CGH-mikrosirumenetelmä on otettu osaksi nemaliinimyopatian molekyyligeneettistä diagnostiikkaa. NM-CGH-mikrosiruanalyysin ja eksomisekvensoinnin yhdistelmä on tehostanut NM-geenien kattavuutta, edistänyt mutaatioiden löytymistä, ja näin ollen parantanut nemaliinimyopatian ja muiden samankaltaisten tautien diagnostiikkaa.

REVIEW OF THE LITERATURE

1. Human genome variation

The human genome consists of a large nuclear genome of 3.1 Gb and a small separate mitochondrial genome of 16.6 kb. The ~26 000 genes of the nuclear genome are packed into 46 chromosomes that contain 22 autosome pairs and the sex chromosomes, XX for females and XY for males. The gene-rich regions of DNA have a high level of methylated CpG islands which can be shown as light bands on Giemsa staining (G banding) of chromosomes. Protein-coding genes vary greatly in size, differing from <1 kb to >2 Mb. The average number of exons in the protein-coding genes is estimated to be 10 but the largest genes include more than 300 exons. Most genes also include introns that differ greatly in size. However, all genes do not encode proteins but for example pseudogenes and retrogenes as well as non-coding RNAs that are involved in protein synthesis, RNA maturation, DNA synthesis, gene regulation and transposon control (Lander et al., 2001; McPherson et al., 2001; Strachan and Read, 2011; Venter et al., 2001). Altogether, the human genome is highly diverse and the various sections and parts have different features that are all important for a properly functioning genome.

It has been estimated that the human genome is approximately 99.5-99.9% identical between different individuals. Consequently, the remaining 0.1-0.5% of DNA accounts for all individual differences including normal variation and susceptibility to disease (Kruglyak and Nickerson, 2001).

1.1. Normal variation

1.1.1. Polymorphisms

Polymorphisms are normal variants in the human genome. Single nucleotide polymorphisms (SNPs) are changes of one nucleotide that are found in the general population with > 1% frequency. Furthermore, every individual is estimated to carry ~3 million SNPs in their genome. (Kim et al., 2009; Tong et al., 2010) The database for SNPs (dbSNP) hosted by the National Center for Biotechnology Information (NCBI) included altogether circa 150 million different SNPs in the human genome (db SNP build 144 in June 2015).

Copy number polymorphisms (CNP), also called benign copy number variations, are large structural variations. They are benign duplications, deletions or inversions that are not known to be associated with a disease or a disorder. Copy number changes have originally been defined as changes of the DNA copy number in a segment of DNA more than 1 kb in size (Redon et al., 2006). The Database of Genomic Variants carries 490 000 CNPs collected from 67 studies of healthy individuals (July 2015).

1.1.2. Repetitive DNA

Human DNA also includes many different types of repetitive sequences; both internal and external to gene sequences. The various repeat elements of the human genome can be divided into two groups: low-copy repeats (LCR) and high-copy repeats.

Interspersed repetitive elements are the most common high-copy repeats and they are scattered throughout the genome. Moreover, they are estimated to cover ~45% of the human genome (Chen et al., 2014; Lander et al., 2001). The most common repetitive element in human is the LINE-1 repeat (long interspersed element) that constructs ~17% of the human genome. This family is capable of autonomous transposition and still has actively transposing members. LINE-1 elements are ~6 kb long elements that encode proteins essential for the transposition, such as nucleic acid binding protein and protein with endonuclease and reverse transcriptase activities (Beck et al., 2011). *Alu* repeat elements belong to the family of SINE elements (short interspersed elements) and they have been named after the *Alu*I restriction site found in their sequence. *Alu* repeats are ~280 kb long elements that construct ~10% of the genome. *Alu* repeats are non-autonomous transposons that have been shown to use the LINE element machinery for transposing (Beck et al., 2011).

Low-copy repeats (LCR), also known as segmental duplications (SD) compose approximately 5% of the human genome. SDs are repeats that occur twice or a few times in the genome. SDs are typically 10-300 kb in size, they share 95-97% sequence similarity and are usually separated by 50 kb – 10 Mb of intervening sequence (Gu et al., 2008; Sharp et al., 2006).

1.2. Pathogenic variation

Mutations of DNA occur, for example, in every DNA replication event, but most are corrected by cellular DNA repair mechanisms. Mutations that happen in somatic cells affect only that individual, but mutations that occur in the gametocytes can be inherited by the offspring. Mutations are the major driving force of evolution. The mutation rate in the human genome is estimated to be approximately ~1.5 x 10^{-8} per site per generation (Conrad et al., 2011; Lynch, 2010; Samocha et al., 2014). Mutations can create modifications of the DNA that enable better adaptation of the individual to the environment. Mutations are more often silent, which means that they do not cause an effect on the protein level. However, sometimes mutations can cause adverse effects such as a disease. Mutations can be described as heritable changes at the DNA level that can cause errors in the gene product, such as proteins that they encode. There are different types of mutations that can be categorized, for example based on their size or origin.

1.2.1. Point mutations

Point mutations alter only one nucleotide of DNA sequence. Depending on the change, this can cause an amino acid substitution, a premature stop-codon, abnormal splicing, or a silent mutation. A missense mutation causes the encoded amino acid to be substituted with another. The change of one amino acid can be harmful if it resides in a conserved DNA sequence or of it changes an important functional domain of the protein or the protein conformation. A nonsense mutation causes a premature stop codon which can cause a truncated protein product to be produced, or more often, nonsense-mediated mRNA decay. A silent mutation changes the nucleotide but does not change the amino acid. All of these different types of point mutations can also cause splicing errors when they occur in splicing donor or acceptor sites. Splicing errors may result in splicing of exons (exon skipping) or splicing at cryptic splice sites within introns or exons. Point mutations can also create novel donor or acceptor splice sites within exons or introns.

1.2.2. Copy number variations (CNVs)

Structural changes such as translocations and inversions are large changes that modify the structure of one or more chromosomes. These may be balanced, i.e. they do not change the copy number of the DNA segment. They may also be unbalanced, creating a copy number variation (CNV). It has been estimated that the mutation rate for *de novo* locus-specific CNVs is higher compared with nucleotide substitutions (Redon et al., 2006). Most commonly there are two copies of a certain gene, one in each allele. Copy number gains of one additional copy are called duplications and copy number losses are called deletions. If they occur inside a gene, they may change the reading frame and a premature stop codon may arise. However, if a deletion or duplication causes an in-frame mutation, it can produce a shorter or a longer gene product. Some CNVs can cover several megabases and can thus easily contain an entire gene or numerous genes (Zhang et al., 2009).

1.2.2.1. Mechanisms creating copy number variations

Many different mechanisms can cause CNV formation. A summary of different mechanisms is presented in Table 1. Non-allelic homologous recombination (NAHR) is thought to be one of the most common mechanisms. NAHR is caused by misalignment and cross-over of non-allelic homologous DNA segments. These homologous DNA segments can be repetitive sequences such as segmental duplications (SD). NAHR requires so called minimal efficient processing segments such as SDs to take place. The homology of the SDs and the distance between the two segments affect the NAHR efficiency (Gu et al., 2008; Hastings et al., 2009; Sharp et al., 2006).

Malfunction of the DNA repair mechanisms may also cause loss or gain of DNA segments. Non-homologous end joining (NHEJ) is a common mechanism to correct pathological double-stranded DNA breaks (DSBs). It is effective throughout the cell cycle and does not require a homologous chromosome or particular sequences to take place. NHEJ is flexible as the result of different nuclease, polymerase and ligase activities. NHEJ includes four steps: detection of the double-stranded break, molecular bridging of the broken DNA ends, modification of the ends, and ligation (Gu et al., 2008; Lieber, 2008).

Microhomology-mediated end joining (MMEJ) is a rather recently suggested DNA-repair and CNV-formation mechanism. It is also called alternative NHEJ (alt-NHEJ). Microhomology, a small segment of DNA that is homologous between the joined DNA sites, is required for MMEJ. This microhomology is used to align the DNA sequences before joining the segments. NHEJ may also use microhomology (~1-4 nucleotides), but for MMEJ microhomology it is obligatory and the homologous stretch is usually larger (~5-25 nucleotides) (Lieber, 2008; McVey and Lee, 2008).

Fork Stalling and Template Switching (FoSTeS) is a replication-based mechanism that causes CNVs. If the replication fork stalls during DNA replication, the lagging strand can disengage from the original template and anneal to another replication fork and then continue the DNA synthesis. If the strand switches to a fork located downstream (forward invasion), this causes a deletion. If it switches to an upstream-located fork (backward invasion), this causes a duplication (Lee et al., 2007).

Microhomology-mediated break-induced replication (MMBIR) is another replication-based repair mechanism. In MMBIR the 3’ overhang of the broken DNA strand invades another chromosome, such as the sister chromatid or the homologue, using microhomology, and continues the replication from there, up until the end of the chromosome (Bauters et al., 2008; Hastings et al., 2009; Vissers et al., 2009).

Table 1. Mechanisms creating copy number variations. This table summarises different CNV-creating mechanisms and their characteristics.

Name	Abbreviation	Mechanism	Functions in	Special features	Result
Non-allelic homologous recombination	NAHR	Non-allelic homologous DNA segments misalign and crossing over occurs	Meiosis and mitosis	Homologous non-allelic sequence	deletion, duplication, inversion, and mosaic rearrangements
Non-homologous end joining	NHEJ	Detects a double-stranded DNA break, builds a molecular bridge, modifies the ends, and ligates	Meiosis and mitosis	Any sequence, microhomology can be used (1-4 bp), may leave an information scar	deletion, duplication, translocation
Microhomology-mediated end joining (=alternative NHEJ)	MMEJ (=alt-NHEJ)	The broken DNA is joined using microhomology	Mitosis	Microhomology required (5-25 bp), may leave an information scar	deletion, translocation
Fork Stalling and Template Switching	FoSTeS	The replication fork stalls, the lagging strand disengages from the template and anneals to another replication fork	DNA-replication	Microhomology	deletion, duplication, triplication, inversion, complex rearrangement
Microhomology-mediated break-induced replication	MMBIR	3' overhang of a broken DNA strand invades sister chromatid or the homologue and continues replication	DNA-replication	Microhomology required, may leave an information scar	duplication, deletion, inversion, translocation, triplication, and loss of heterozygosity
Chromothripsis	-	One catastrophic event of simultanous DNA breaks is repaired	DNA repair	Repair of up to hundreds of breakpoints	complex rearrangements

References: Gu et al., 2008; Kloosterman et al., 2011; McVey and Lee, 2008; Vissers et al., 2009, Liu et al., 2012.

Chromosome shattering, also called chromothripsis, is a recently discovered phenomenon that is thought to be caused by one catastrophic event that results in complex rearrangements. Typically there is strong clustering of breakpoints. Chromothripsis is thought to be caused by many simultaneous double-stranded DNA breaks that are then repaired through non-homologous mechanisms (Kloosterman et al., 2011; Stephens et al., 2011).

1.2.2.2. Repeat elements producing copy number variations

Several repeat elements are known to be involved in creating CNVs. *Alu* repeats can be involved in homologous recombination, via two suggested ways. First, they may serve as binding sites for proteins necessary for homologous recombination. Second, they can promote DNA strand exchange directly themselves (Kolomietz et al., 2002). Different studies have suggested that *Alu* repeats could also mediate chromosomal rearrangements via non-homologous mechanisms such as NHEJ, FosTeS, or MMBIR (Shaw and Lupski, 2005; Vissers et al., 2009). Repetitive sequences such as *Alu* repeats may predispose the rearrangement to additional deletions at the breakpoints (Kolomietz et al., 2002). Furthermore, it has been suggested that certain *Alu* elements play an important role in the constitutional as well as evolutionary chromosomal rearrangements (Shaw and Lupski, 2005). The *DMD* gene encoding dystrophin has been shown to carry different CNV deletions with scattered breakpoints that include *Alu* and tandem repeats (Nobile et al., 2002). In another study of congenital aberrations, it was shown that a repetitive element was identified in 70% of the studied breakpoints (42/60). These included different SINEs (such as different *Alu* repeats), LINEs, DNA repeats and long terminal repeats (Vissers et al., 2009). Altogether, there is a broad spectrum of different variations and mechanisms caused by repetitive elements of the genome.

1.2.2.3. Copy number variations and disease

Copy number variations may be benign or cause harmful effects, especially when including genes. CNVs are common in congenital as well as in acquired disorders. For example, when a CNV contains genes that are dosage-sensitive, or a deletion occurs in a region including haploinsufficient or imprinted genes, they are more likely to affect the phenotype. Even very small CNVs can cause problems if they disrupt a gene. Furthermore, CNVs may be pathogenic even if they do not specifically contain annotated disease-causing genes, but instead they may carry, for example, their transcription factors.

Various microarray techniques have revealed a great number of novel CNVs during the last decade, however, many disease-related CNVs are likely yet to be discovered. Understanding the pathogenetic mechanism of a disorder is always important in every disease, whether it concerns a congenital disorder or an acquired disease such as cancer. The different CNV-inducing mechanisms (Table 1) work in different settings and can thus give an indication of the stage where the pathogenic rearrangement occurred.

As mentioned above, many genomic disorders and syndromes are known to be caused by NAHR-induced CNVs. NAHR can occur in meiotic recombination and create either inherited or sporadic disorders. A well-known inherited CNV example is the chromosomal region 17p12 in which a deletion causes hereditary neuropathy with liability to pressure palsies (HNPP) and duplication of the same region causes Charcot-Marie-Tooth disease type 1 (CMT1A). NAHR can also cause sporadic disorders due to recurrent *de novo* rearrangements. A deletion in the chromosomal region 17p11.2 causes Smith-Magenis syndrome and a duplication Potocki-Lupski syndrome. This region of chromosome 17p is remarkably rich in LCR segments that predispose these particular regions for NAHR. NAHR is also known to occur in mitotic cells, causing mosaic rearrangements which are especially common in cancer. The same pair of LCRs may utilize both meiotic and mitotic events, but not necessarily with the same frequency (Gu et al., 2008).

NHEJ is effective throughout the cell cycle and tolerates some nucleotide loss or addition in the breakpoint. This explains the breakpoint heterogeneity as well as the so-called information scars that are often left in the repaired sites. Furthermore, these CNVs occur more randomly throughout the genome and are thus usually non-recurrent. NHEJ is also used to repair the physiologic DSBs that occur during the somatic recombination of the antigen receptors of the lymphocytes. The flexibility and imprecision of NHEJ further enhances antigen receptor diversity and the adaptive immune system. Moreover, inherited defects in this mechanism can cause severe combined immune deficiency syndrome (SCID) (Gu et al., 2008; Lieber, 2008). MMEJ (or alt-NHEJ) requires microhomology for alignment of broken ends and can thus only cause deletions in the breakpoint region. Both NHEJ and MMEJ are known to create translocations and rearrangements that are common in cancer cells (Bennardo et al., 2008; Gu et al., 2008; McVey and Lee, 2008).

From the replication-based mechanisms, FoSTeS is also thought to use microhomology, however, it can cause duplications as well as deletions. Because the replication fork can switch the template several times, FoSTeS is also thought to be capable of causing large complex rearrangements (Gu et al., 2008; Liu et al., 2012). Furthermore, MMBIR can cause many types of rearrangements, including duplications, deletions, inversions, translocations, triplications, and loss of heterozygosity, and thus also imprinting disorders. MMBIR is thought to form non-recurrent CNVs, and it has been suggested to contribute to chromosomal instability, such as somatic changes in cancer cells and tumour formation (Hastings et al., 2009; Vissers et al., 2009).

Chromothripsis was first described in cancer (Stephens et al., 2011), and shortly afterwards in constitutional diseases (Kloosterman et al., 2011). In constitutional cases, more than one chromosome is usually involved and the number of breakpoints is less than 25. In cancer, chromothripsis involves one or multiple chromosomes and there can be dozens or even hundreds of breakpoints (Kloosterman and Cuppen, 2013).

1.2.2.4. Breakpoint analysis of copy number variations

The different CNV mechanisms are effective in different environments (Table 1). For example, homology-dependent NAHR and homology-independent NHEJ are effective during meiosis and mitosis, MMEJ acts in mitosis and FoSTeS and MMBIR during replication. Consideration must be taken, when studying the CNV breakpoint, regarding the fact that NAHR and NHEJ usually correct double-stranded breaks, whereas replication-based mechanisms correct single-stranded breaks. Furthermore, several mechanisms can use microhomology, such as MMEJ, MMBIR, FoSTeS, and NHEJ. The molecular fingerprint can indicate the replication mechanism. A molecular scar of inserted nucleotides at the breakpoint may indicate NHEJ, but also for example MMBIR. Many mechanisms may cause deletions and duplications, but MMEJ can only cause deletions, and MMBIR is versatile also causing inversions, translocations, triplications, and loss of heterozygosity. On the other hand, FoSTeS and chromothripsis may also cause complex rearrangements (Conrad et al., 2010; Vissers et al., 2009).

The only way to elucidate the origin of the rearrangement is to reveal the exact breakpoints of the CNV. This may often be more difficult than anticipated. PCR-based sequencing has been used in previous studies but it can be extremely laborious. It also requires previous knowledge or estimation of the structure of the rearrangement, such as the orientation of the duplicated segment. This is why genome-wide shotgun sequencing has become popular in resolving the exact breakpoints of the CNVs. However, this is a rather expensive method to be used to further delineate already identified CNVs, especially in a large sample cohort. All in all, no method is perfect alone. For example, high copy number repeats and heterochromatin regions are extremely difficult to catch and verify with sequencing (Conrad et al., 2010). Unique parts of sequences are required to align the different pieces of sequence properly and this may not be achieved when dealing with long repetitive sequences.

Even when the exact breakpoint has been identified, defining the causative method behind the CNV or rearrangement may be difficult. This is due to the fact that many mutational mechanisms can create similar breakpoint signatures. For example, microhomology can be found at breakpoints created by MMEJ, MMBIR, NHEJ, and FoSTeS (Conrad et al., 2010; Vissers et al., 2009). This demonstrates the current challenge when interpreting the CNV breakpoints and their origin. Nevertheless, this field of research has expanded in recent years and novel data will undoubtedly shed new light on how to best unravel these mechanisms in the future.

1.3. Variant detection

The identification of disease-causing variant(s) in each affected family is often important in the case of monogenic disorders. For many diseases, the identification of the pathogenic variant(s) is needed to confirm the diagnosis. It is also essential for genetic counselling as it helps to determine the mode of inheritance and thus the recurrence risk in each family. Identifying novel disease-causing variants may further help to establish possible genotype-phenotype correlations of the disease. Characterizing new mutations may also help to elucidate the gene functions and to understand the pathogenetic mechanisms of the disease. Understanding the pathogenesis is a prerequisite for the development of specific therapies.

Different variant detection methods are usually optimal for finding only certain types of variants. Heterogeneous diseases like nemaline myopathy (NM) that have several causative genes, some of them also lacking proper mutational hotspots, can make variant analysis very cumbersome. Even though the DNA samples from families with this muscle disorder have been extensively studied, many families remain where one or both pathogenic variants are yet to be identified. In some cases the suspected diagnosis might be incorrect. This may prohibit the identification of the disease-causing variants, if the appropriate genes are not tested. Even if the appropriate genes are tested, the variant detection methods may be limited in finding all types of mutations. Furthermore, it is also likely that there are novel genes yet to be identified. This accentuates the importance of developing novel variant detection methods.

1.3.1. Variant screening methods and Sanger sequencing

Variants can be sought by direct gene sequencing from PCR products, but for large genes, such as *NEB* including 183 exons, a screening method preceding sequencing can be useful. The SSCP (single-stranded conformation polymorphism) or dHPLC (denaturing high-performance liquid chromatography) methods have previously been used to pre-screen the genes (Jones et al., 1999; Orita et al., 1989; Sheffield et al., 1993; Underhill et al., 1996). These screening methods can help to point out the region in a large gene where a DNA change might be located and that needs to be sequenced further. This has been efficient in identifying small, heterozygous variants, but additional methods are needed since both or the second disease-causing variant of many patients remain unidentified after analysis by dHPLC followed by sequencing. Nowadays sequencing techniques have become much more powerful, and such screening methods are seldom used anymore. In addition to next generation sequencing, other methods have been developed that allow the examination of the entire human genome even in a single experiment.

1.3.2. Microarray

The microarray method became available in the late nineties when comparative genomic hybridization (CGH), which was used for fixed metaphase chromosomes (Kallioniemi et al., 1993), was developed into probe-based arrays (Pinkel et al., 1998; Solinas-Toldo et al., 1997). This method allows determination of the copy number variation between the sample and a reference genome.

Nowadays microarray-based methods are very commonly used and there are many different variations and applications. Gene expression microarrays from cDNA allow comparisons of different gene expression patterns between individuals or different tissues. Microarrays can also be used for micro-RNA profiling and studying protein interactions or epigenetic modifications. However, one of the most commonly used microarray applications is still the DNA-based aCGH (array comparative genomic hybridization), which detects copy number variations of different sizes in the genome. It can be based for example on SNP or CNV probes. A two-colour aCGH method is described in Figure 1. It is based on attaching thousands of probes to a surface, such as

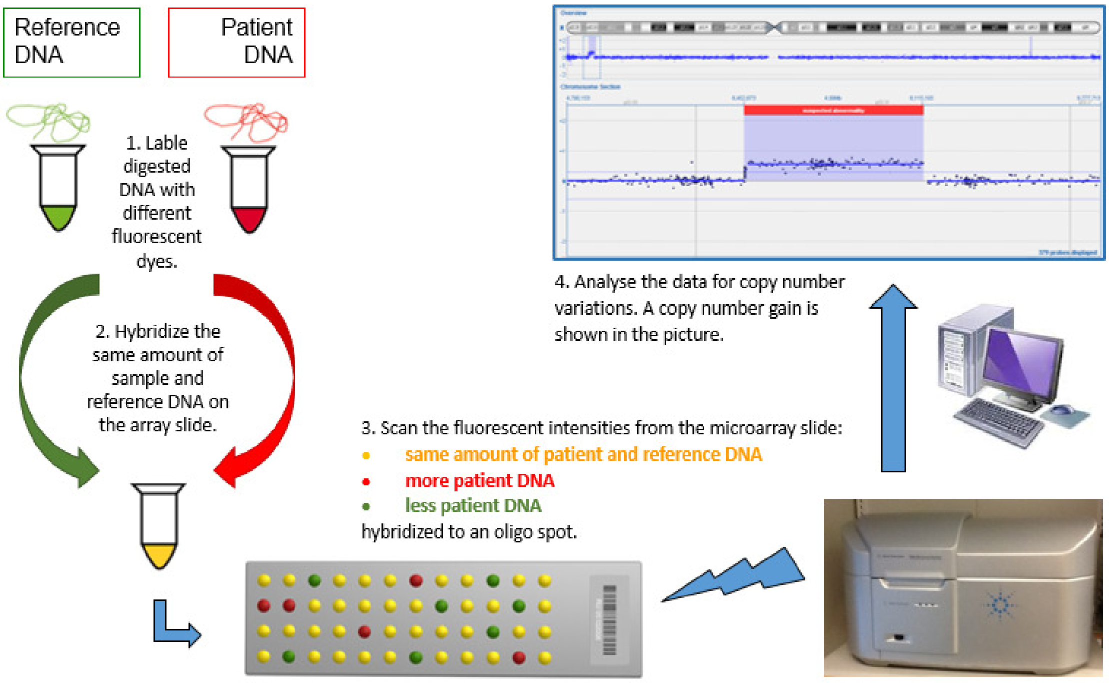
Reference DNA
Patient DNA
1. Lable digested DNA with different fluorescent dyes.
2. Hybridize the same amount of sample and reference DNA on the array slide.
3. Scan the fluorescent intensities from the microarray slide:
● same amount of patient and reference DNA
● more patient DNA
● less patient DNA
hybridized to an oligo spot.
4. Analyse the data for copy number variations. A copy number gain is shown in the picture.

Figure 1. A schematic overview of the array-CGH method used in this study. 1) The same quantity (1000 ng) of patient DNA and reference DNA are digested and labelled with different fluorescent dyes. 2) The differently labelled DNAs are hybridized together on a glass slide. 3) The microarray slide is washed and the fluorescent intensities are scanned with a laser scanner. 4) The intensity values are transformed into a text format using the Feature Extraction Software (Agilent Technologies) and transformed and analysed in a graphic format using the CytoSure Interpret Software (Oxford Gene Technology).

a glass slide. The sample and reference DNA are labelled with different fluorescent dyes and they are hybridized together on the slide. The different fluorescent intensities are then measured and the copy numbers of each probe can be analysed. The resolution of the microarray depends on the number of probes and the targeting of the microarray.

Chromosomal microarray (aCGH of the entire genome) is currently recommended as a first-tier diagnostic test for patients with unexplained intellectual disability, developmental delay, autism spectrum disorders, and multiple congenital anomalies. This is due to the much higher diagnostic yield (15-20%) compared with the conventional chromosomal karyotyping using G-banding (3%) (Miller et al., 2010).

Microarrays are also commonly used in cancer research and diagnostics; they can help to identify cancer-specific variants or altered gene expression. This aids diagnosis, classifying, and estimating the prognosis of different malignancies (Shinawi and Cheung, 2008).

Microarrays can cover the whole genome or be more targeted, for example towards known syndromes. Furthermore, it has become possible to design targeted custom arrays to densely cover only the genes of interest. One example of a high-density custom array is the DMD-CGH array targeted for variant detection in the gigantic dystrophin gene where mutations cause dystrophinopathies (Bovolenta et al., 2008). In recent years various microarray methods have shown great success in identifying previously characterized as well as novel copy number variations.

1.3.3. Multiplex Ligation-dependent Probe Amplification

The multiplex ligation-dependent probe amplification (MLPA) method has successfully been used in variant detection and diagnostics of several genes. The MLPA method is based on multiplex PCR amplification of selected DNA regions of different lengths. The products are then separated by fragment analysis and the copy number of each fragment can be analysed compared to the reference DNA (Schouten et al., 2002; Schwartz and Duno, 2004; Sulek et al., 2011). Even though the MLPA technique allows, in an optimal situation, for detecting of the copy numbers of tens of different genomic regions, there are commercially available kits only for a selection of genes. However, self-designed synthetic MLPA sets can be designed for almost any region of interest (Stern et al., 2004).

1.3.4. Next Generation Sequencing

Next generation sequencing (NGS) techniques have revolutionized sequencing possibilities. While the costs of sequencing have decreased, throughput has increased. NGS includes exome sequencing, whole-genome sequencing, and targeted sequencing approaches. There are different NGS sequencing techniques, however, they follow the same principle. The first step is template preparation. A DNA library is created by fragmenting genomic DNA and amplifying it. Using synthetic oligonucleotides, the fragments can be attached to a sequencing media (such as a flow cell, slides, beads etc.) and amplified creating a fragment or a mate-paired DNA library. Sequencing of the fragments can be accomplished using many different approaches. Sequencing can be based on differentially labelled nucleotides, change in voltage or release of photons after adding nucleotides in predefined order, or other methods that allow identifying the added nucleotides to create the sequence. After sequencing, the reads are aligned to a known reference sequence or assembled to create a consensus sequence for variant analysis (Desai and Jere, 2012; Metzker, 2010; Ng et al., 2009).

For example, in the Agilent Sure Select approach the DNA library is created by fragmenting the DNA and ligating adapters to both ends of each fragment. At this stage, the library can be hybridized with a selected capture kit to select the regions, genes or exons of interest. For example when using the Illumina sequencing platform (MiSeq or HiSewq), the fragments are then attached to a flow cell coated with primers. Solid-phase bridge PCR amplification is done creating millions of clusters on the flow cell. The sequencing is done using four differently labelled nucleotides (A, T, G, C) that emit different fluorescence after laser excitation that can be identified to create the sequence. The error rate of MiSeq nucleotide substitutions has been reported to be ~1%, which is similar to Sanger sequencing with the Genome Analyzer (Desai and Jere, 2012; May et al., 2015).

The newest forms of sequencing technology are the single-molecule sequencing techniques that do not require PCR amplification. This provides the next step of sequencing technology that is unrestricted by the limitations of PCR. It allows the generation of long stretches of DNA and sequencing also of such types of DNA regions that are difficult to amplify. This could possibly reduce sample handling time as well as required quantities of DNA and avoid errors produced by the amplification (Xuan et al., 2013). Sequencing through nanopores is a currently emerging technique that has also been called fourth-generation sequencing. The MinION platform by Oxford Nanopore Technologies is the first nanopore sequencer to have been commercialized, although, still being further developed by the users. The MinION technique is based on directly measuring the changes in electrical current as one individual strand of DNA sequence passes through one of the 500 pores on the nanopore platform. In theory, this allows the analysis of DNA stretches of unlimited size. The current limitations of these methods include requirement of high-quality DNA, low sequencing capacity, and high error rate up to ~10-30% (personal experience). Currently, read lengths of ~40 kb have been reported which are a great deal longer than traditional sequencing can provide. Using this approach, the copy number for a cancer-testis gene family (CT47) in human chromosomal region Xq24 was sequenced even though it had been inaccessible due to the high repetitiveness thus far. This, in particular, shows the potential of the fourth-generation sequencing techniques that do not require PCR (Jain et al., 2015; Loman and Watson, 2015).

All in all, NGS techniques have made sequencing exponentially more efficient. They are also optimal methods for hunting novel disease-causing genes and variants, because they do not require previous knowledge of the genes. However, when trying to identify novel pathogenic genes, it is highly beneficial if family trios, including parents and affected child, or even larger family sample sets are available for testing. Including several samples makes it possible to exclude the majority of the discovered benign variants when comparing the sample from the Index Patient to samples from the non-affected members of the family.

Whole-exome sequencing (WES) is currently widely used in research and is steadily making its way into diagnostics as well. This method allows sequencing at the level of an entire exome at once (Ng et al., 2009). However, the choice of exome capture or enrichment kit defines which genes and exons are included in the study. Five commonly used patforms, Illumina Nextera Rapid Capture Exome, Illumina TruSeq Exome, Agilent SureSelect XT Human All Exon, Agilent SureSelect QXT, and NimbleGen SeqCap EZ Human Exome Library, all use DNA or RNA baits for targeting. However, they differ in genomic fragmentation method, target region selection, bait length and density, as well as the molecules used for capture. This causes differences in their gene and exon coverage. Different platforms currently target 40-60 Mb of the human genome (Chilamakuri et al., 2014; Shigemizu et al., 2015). In a recent study, Agilent SureSelect XT Human All Exon platform achieved the highest sequence coverage in the coding region, the Illumina platform showed the highest performance in detecting medically interesting variants, whereas Nimblegen was superior in detecting variations in the untranslated regions (Shigemizu et al., 2015).

Whole-genome sequencing includes, at least in theory, the entire genome, yet it is a lot more expensive and it produces a lot more data. However, there are genomic regions such as highly repetitive segments of DNA that cannot be covered because they lack unique sequence, which would be required to align and locate the sequences. In addition, NGS can also be done on targeted genomic region(s) of interest. The NGS approach can thus be chosen based on the study expectation, number of genes of interest, the total size of the regions, whether exonic data is adequate or whether intronic regions are also needed etc. The targeted sequencing approach provides the most cost-effective way of covering the genes of interest in high read depth. This approach is used for example for genetic diagnostics of the Usher syndrome. In this disorder, 12 causative genes have been identified that cover 80% of the disease-causing variants (Jamuar and Tan, 2015; Krawitz et al., 2014). There are even commercially available panels targeting, for example, cancer genes and assisting in oncology decision making (Weiss et al., 2015). However, a study comparing WGS, WES and targeted sequencing suggested that WES currently remains the preferred choice when searching for the genetic cause of intellectual disabilities, where many of the disease-causing variants remain yet to be identified (Sun et al., 2015). WGS currently offers only limited advantages over WES and when cost-efficiency and turnaround time are taken into consideration, WES and targeted panels outperform WGS in a clinical setting (Sun et al., 2015).

In any next generation sequencing approach, the analysis of the large amounts of generated data is currently the most challenging part of the process. This is also a major reason why the NGS technique has been implemented to the clinical setting slowly and with caution. In clinical work it is important that the samples can be analysed promptly and interpreted easily and with great certainty and reliability (Desai and Jere, 2012). So-called unsolicited or incidental findings create a great challenge, and the first recommendations on how to handle these findings have been published. For example, the American College of Medical Genetics and Genomics (ACMG) has listed genes in which pathogenic variants are known to cause medically actionable disorders, and thus should be reported, unless the individual chooses to opt out of such analysis (ACMG Board of Directors, 2015; Green et al., 2013). On the other hand, the European Society of Human Genetics (ESHG) recommends a much more cautious approach in reporting unsolicited findings and favours targeted approaches over WES and WGS only when necessary (van El et al., 2013). Currently, the practice seems to differ between different laboratories. Nevertheless, it is highly important that each institution reaches a consensus regarding the ethical issues when implementing NGS into clinics.

1.4. Assessing variant pathogenicity

It is not always easy to interpret the effect that a variant may cause at the functional level. The variant itself combined with the location and its surroundings are important in this estimation. Variants including genes or residing inside genes, in their exons and conserved DNA sequences are more likely to cause a pathogenic effect. Variants outside genes or inside the gene introns, excluding splice sites, are less likely to cause an effect. This is why every identified DNA change needs to be evaluated carefully to determine whether they are benign polymorphisms or disease-causing variants.

A number of databases can be used to interpret the potential pathogenicity of the variants, such as Leiden Open Variation Database (LOVD), Exome Variant Server (EVS), or Exome Aggregation Consortium (ExAC) Browser. The consequences that the change creates at the DNA level can be estimated using different softwares, such as Polyphen (Polymorphism Phenotyping), SIFT (Sorting Intolerant from Tolerant), MutationTaster, Mutation Assessor (MASS), FatHMM, Combined Annotation Dependent Depletion (CADD), and Condel (Consensus Deleteriousness score) (Adzhubei et al., 2010; Gonzalez-Perez and Lopez-Bigas, 2011; Kircher et al., 2014; Ng and Henikoff, 2003; Reva et al., 2011; Schwarz et al., 2014; Shihab et al., 2013). The softwares estimate the variant pathogenicity by measuring sequence conservation, assessing the impact on protein structure and function, or quantifying the pathogenic potential using diverse genomic information (Grimm et al., 2015).

The potential pathogenicity of copy number changes can also be estimated using various databases, such as Database of Genomic Variants (DGV), DECIPHER (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources), and ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations), presenting information about benign and pathogenic variations described previously. The changes can be further studied at the RNA, amino acid, and protein levels using functional studies.

Despite using several databases for variant interpretation, variants of unknown significance (VUS) may remain and they can for example be categorized as likely benign or likely pathogenic (Jamuar and Tan, 2015). When more data has been collected, the significance of these variants may be resolved in the future. Guidelines have been published on how to write the reports and which variants should be reported (MacArthur et al., 2014).

2. Skeletal muscle

There are three types of muscles in the human body; skeletal muscle, cardiac muscle and smooth muscle. Skeletal muscles are striated muscles that enable voluntary muscle movements. Cardiac muscle is also striated, but cardiac movements are non-voluntary and this type of muscle is found only in the heart. Smooth muscle is non-striated and its contractions are non-voluntary. Smooth muscles are found, for example, in the gastrointestinal tract and the walls of blood vessels.

There are several hundred skeletal muscles in human and the number varies between different estimations. Proximal muscles are the ones close to the trunk and distal muscles are further away from the trunk. The muscles vary in shape and size, depending on their function in the body. One muscle fibre consists of a single cell that has emerged from fused adjacent cells and thus one fibre usually contains multiple nuclei. Several muscle fibres form fascicles which bundle to form the muscle (Stone and Stone, 2011). Figure 2A shows the structure of the striated muscle.

2.1. Skeletal muscle fibre types

Skeletal muscles consist of different muscle fibre types. Type 1, slow fibres and type 2A fast fibres use oxidative phosphorylation for generating ATP (adenosine triphosphate) which makes them optimal for endurance. Type 2B, ultrafast fibres use mainly glycolysis which makes them optimal for short-duration maximal performance. Most human muscles include a typical combination of Type 1 and 2 fibres, depending on the muscle. In myopathies the fibre size can vary abnormally (Spangenburg and Booth, 2003). The fibres can be hypotrophic (smaller than normal because of failure to grow normally), atrophic (smaller than normal because of degeneration) or hypertrophic (larger than normal). In human, usually both fibre types are present roughly in equal proportions, although there is some variability between different types of muscle. In a state called fibre type disproportion the distribution is not equal. This means that the diameter of the type 1 fibres are at least 25% smaller than that of the type 2 fibres. This is a common phenomenon in the congenital myopathies (Brooke and Engel, 1969; Clarke and North, 2003; Jungbluth and Wallgren-Pettersson, 2013).

2.2. The sarcomere

One muscle fibre cell consists of a bundle of myofibrils enveloped by a cell membrane, the sarcolemma. Each myofibril is made up of adjacent sarcomeres. One sarcomere is approximately 2-3 µm long and 1-2 µm in diameter and it is the basic functional unit of the muscle. The striated appearance of skeletal muscle is formed by the organised alignment of its different bands (Figure 2B). Z-discs separate one sarcomere from the adjacent one. The I-bands surrounding the Z-discs are formed by the thin actin filaments, and proteins such as tropomyosins and troponin complexes bound to it. Figure 2C shows how two nebulin molecules span each thin filament. The A-bands are formed by thick myosin filaments. The M-band connects the myosins to the titin filaments in the middle of the sarcomere. Actin and myosin are responsible for the transduction of chemical energy to mechanical force during muscle contraction (Craig and Padron, 2004; Dubowitz et al., 2013).

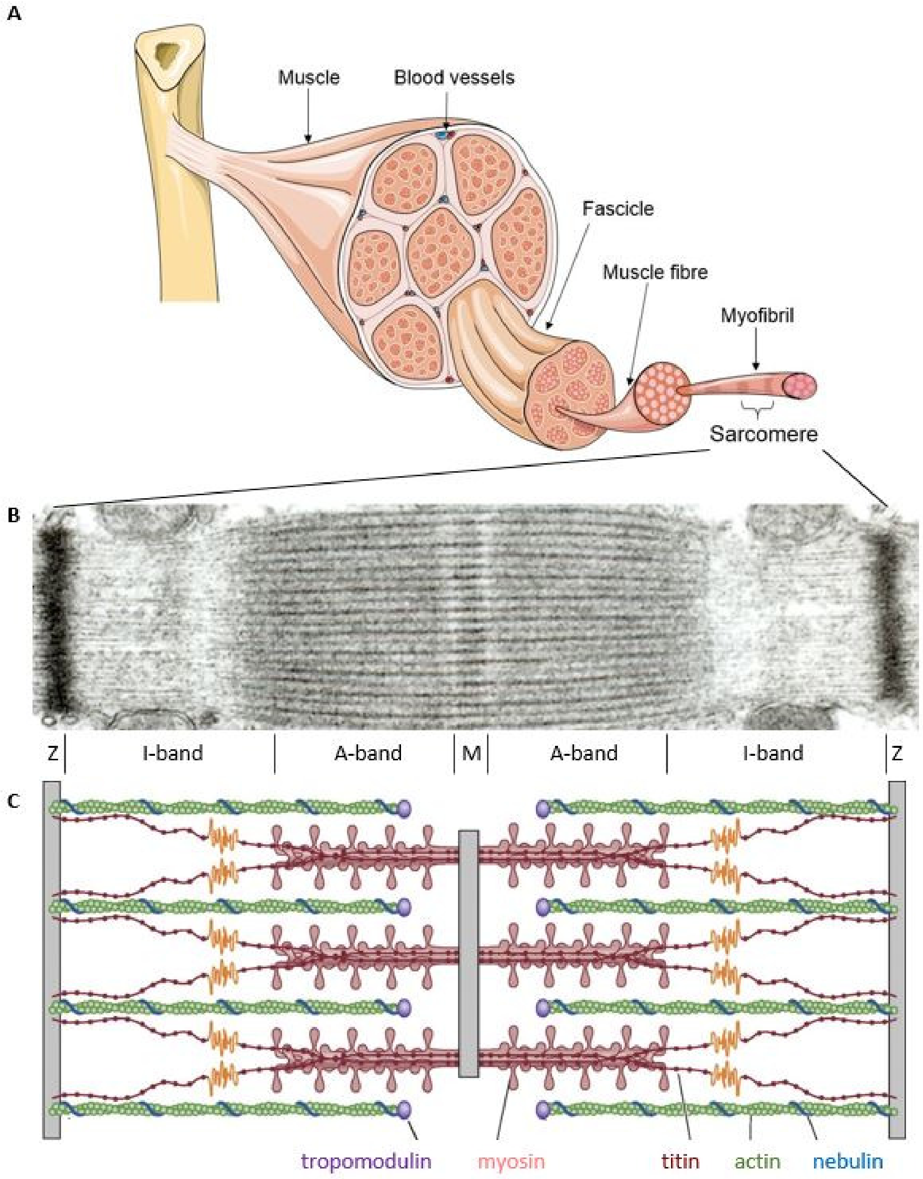

Figure 2. A schematic picture of the muscle and the sarcomere. A) The muscle organization. B) An electron-microscopic photograph of the sarcomere. C) A schematic picture of the sarcomere.

The pictures are reprinted and modified with the permission of their copyright owners: A) Servier Medical Art (www.servier.com) B) Ottenheijm et al, Respiratory Research 2008, 9:12; licensee BioMed Central Ltd. C) Ottenheijm et al, Physiology Published 2010, 25:304-310; licensee the American Physiological Society.

2.3. Muscle contraction

Muscle contraction starts when a motor neuron action potential reaches the neuromuscular junction and releases Ca^{2+} from the sarcoplasmic reticulum into the muscle fibre. Troponin binds the released Ca^{2+}, causing a change in the troponin-tropomyosin complex. This exposes the myosin binding sites, allowing actin-myosin interaction. The myosin ATPase hydrolyses ATP resulting in a conformational change in the globular head of myosin. This allows myosin to move along the actin filament. Myosin heads are released from actin as the next ATP molecule binds to myosin. When this occurs simultaneously in several myofibrils, it shortens the muscle fibre I-bands and the muscle contracts. The muscle contraction is released when Ca^{2+} is withdrawn from the sarcoplasm (Dubowitz et al., 2013; Stone and Stone, 2011).

2.4. Sarcomeric proteins

Sarcomeres are very complex units containing numerous proteins and their subunits. The functions of some of the proteins that are also known to be involved in nemaline myopathy are presented here. The genes and their locations are marked according to the Genome Reference Consortium Human Build 37 (GRCh37/hg19).

2.4.1. Actin

Actin is one of the most abundant proteins in human cells. There are six types of actin proteins that are expressed differently in different cells. β- and γ-actins are expressed virtually in all cells, as they are part of the cytoskeleton of the cell. The α-skeletal, α-cardiac, α-smooth muscle, and γ-enteric actin are tissue-specific. Skeletal α-actin has a central role in muscle contraction. Skeletal α-actin monomers polymerise forming a filamentous helical structure creating the backbone of the muscle thin (actin) filament of the sarcomere. Skeletal α-actin has several binding sites for other proteins, such as α-actinin, nebulin, tropomyosin, and the troponin complex. This explains why the different isoforms of actin are very homologous, especially in terms of their binding sites. The skeletal muscle α-actin-encoding gene, *ACTA1* is located in chromosomal region 1q42 (GRCh37/hg19). *ACTA1* is particularly conserved and rarely tolerates any mutations (Hanauer et al., 1983; Kabsch and Vandekerckhove, 1992; Laing et al., 2009; Nowak et al., 1999). Mutations can interfere with folding and polymerization, create aggregates, affect expression and cause changes in myosin force generation (Feng and Marston, 2009).

2.4.2. Nebulin

Nebulin is a gigantic structural protein (600-900 kDa) located mainly in the thin (actin) filaments of the sarcomeres in striated muscle. Two α-helical nebulin molecules span the entire thin filament (Pfuhl et al., 1996). Furthermore, nebulin determines the minimum length of the thin filament and regulates its contractility by actomyosin interactions and force generation (Bang et al., 2006; Chandra et al., 2009; Witt et al., 2006).

Nebulin is a highly repetitive protein; it periodically binds actin, calmodulin, tropomyosin and the troponin complex. Most of nebulin consists of 30-35 amino acid long simple repeats, each containing one actin-binding site. Nebulin binds 179-239 actin monomers of the thin filament, depending on the isoform. Moreover, simple repeats are arranged into 22-30 super repeats (Chandra et al., 2009; Labeit et al., 1991; Pfuhl et al., 1996; Wang, 1996). Every super repeat binds to the tropomyosin-troponin complex that forms a calcium-linked regulatory complex. The vast majority of nebulin is expressed in striated muscle, but it has also been detected at a low level in the heart and brain (Joo et al., 2004; Kazmierski et al., 2003; Laitila et al., 2012). Nebulin is encoded by *NEB* located in the chromosomal region 2q23.3. Mutations in the nebulin gene have been shown to alter the affinity in nebulin-actin, as well as nebulin-tropomyosin interactions (Marttila et al., 2014).

2.4.3. Tropomyosins

The tropomyosins are a family of proteins that bind to actin filaments. Four different genes have been characterized encoding different tropomyosin proteins and isoforms. α-Tropomyosin$_{\text{fast}}$ is mainly expressed in the heart and in type 2 fast skeletal muscle fibres. It is encoded by the *TPM1* gene that is located in chromosomal region 15q22.2. β-Tropomyosin is mainly expressed in skeletal muscle, mainly in slow, type 1 muscle fibres, but to a lesser extent also in type 2 fast fibres. It is encoded by *TPM2* located in chromosomal region 9p13.3. α-Tropomyosin$_{\text{slow}}$ is expressed in slow, type 1 muscle fibres and it is encoded by *TPM3* located in the chromosomal region 1q21.3. Muscle-specific α- and β-tropomyosin are alpha-helical coiled-coil proteins that form dimers which bind head-to-tail along the thin actin filament. The *TPM4* gene located in the chromosomal region 19p13.12 encodes tropomyosin expressed in tissues other than muscle. The separate isoforms have differences in their functional domains that bind for example actin, tropomodulin, and troponin. Tropomyosins act together with the troponin complex regulating the Ca^{2+}-dependent actin-myosin interaction during muscle contraction. *TPM1* and *TPM3* have their own internal promoters for producing low molecular weight tropomyosin isoforms also in other tissues (Gunning et al., 2005; Laing et al., 1995; Perry, 2001). Mutations in *TPM2* and *TPM3* have been identified all along the genes. The majority of the mutations have been shown to affect tropomyosin-actin interaction. Mutations causing increased Ca^{2+} sensitivity resulting in hypercontractile molecular phenotypes have also been described. In addition, a minority of the mutations have been shown to interfere with tropomyosin head-to-tail binding, affecting the polymerisation of tropomyosin (Marttila et al., 2014).

2.4.4. The troponin complex

The troponin complex is formed by three different troponin protein subunits. The various troponins have been named after their first function discovered. Troponin I inhibits the actin-myosin interaction, troponin T binds tropomyosin, and troponin C binds calcium, in addition to their other interactions and functions. There are several genes encoding troponins. Their different isoforms are usually specific for a certain skeletal muscle fibre type or for cardiac muscle (Tiso et al., 1997). The only troponin currently associated with nemaline myopathy is troponin T encoded by the *TNNT1* gene in the chromosomal region 19q13.42 (Johnston et al., 2000). A recessive truncating mutation found in the Old Order Amish population in *TNNT1* causes lower affinity of troponin T for tropomyosin and thus inefficient incorporation of the troponin complex into the thin filament (Wang et al., 2005).

2.4.5. Cofilin

Cofilins belong to a family of proteins that are known to regulate actin filament dynamics. Cofilin 2 is expressed in skeletal muscle and it controls actin polymerization and depolymerization in a reversible manner. Cofilin 2 is encoded by *CFL2* located in the chromosomal region 14q13.1 (Thirion et al., 2001). It has been suggested that mutations in *CFL2* may exert their effects by decreasing regenerative repair and increasing apoptosis and mitochondrial dysfunction in skeletal muscle (Morton et al., 2015).

2.4.6. Kelch repeat-containing proteins

Kelch repeat-containing proteins belong to a family of proteins that contain a BTB domain as well as several Kelch repeats. The Kelch repeat forms a so called beta-propeller found in several different proteins allowing protein-protein interaction. The BTB domain helps protein interactions and is found, for example, in some Kelch proteins and/or proteins that bind actin or nuclear DNA (Adams et al., 2000; Albagli et al., 1995). These proteins have been discovered to have many different functions such as being transcription factors and participating in cytoskeleton regulation. Recently three of these proteins have been found to carry pathogenic variants that cause nemaline myopathy. Kelch repeat and BTB domain-containing protein 13 was the first to be associated with NM. The encoding *KBTBD13* gene is located in the chromosomal region 15q22.31 (Sambuughin et al., 2010). Recently, two other Kelch-like proteins have been associated with NM: Kelch-like protein 40 and 41, encoded by *KLHL40 (=KBTBD5)* in the chromosomal region 3p22.1 (Ravenscroft et al., 2013) and *KLHL41 (=KBTBD10)* in the chromosomal region 2q31.1 (Gupta et al., 2013), respectively. Recently, it was shown that KLHL40 promotes NEB and LMOD3 stability, and the abundance of these proteins was reduced in the skeletal muscles of *Klhl40*$^{-/-}$ mice (Garg et al., 2014). KLHL40 has also recently been shown to be essential for muscle myogenesis and all three proteins have been shown to be connected to the ubiquitin proteasome pathway. This may be a link to NM pathogenesis (Gong et al., 2015; Sambuughin et al., 2012).

2.4.7. Leiomodin-3

Leiomodins form a subfamily within the group of tropomodulins. LMOD1 is found in smooth muscle and LMOD2 in cardiac muscle (Conley et al., 2001). Leiomodin-3 (LMOD3) is expressed in striated muscle and it localizes to the pointed end of the sarcomeric thin filaments. It includes predicted tropomyosin- and actin-binding sites and is predicted to contribute into stabilizing the thin filament. Recently, mutations in *LMOD3* have been identified in in patients with severe NM (Yuen et al., 2014). LMOD was recently shown to localize to the A band of the sarcomere and to be essential in sarcomere assembly, embryonic myofibrillogenesis, and sarcomere integrity (Garg et al., 2014; Nworu et al., 2015; Tian et al., 2015).

3. Nemaline myopathy and related disorders

3.1. Nemaline myopathy

Nemaline (rod) myopathy (NM) (MIM IDs from the database of Online Mendelian Inheritance in Man: NEM1 #609284, NEM2 #256030, NEM3 #161800, NEM4 #609285, NEM5 #605355, NEM6 #609273, NEM7 #610687, NEM8 #615348, NEM9 #615731, NEM10 #616165) is one of the most common congenital myopathies. Two groups described NM as a novel disease entity in 1963 (Conen et al., 1963; Shy et al., 1963). The typical histological findings include the presence of nemaline bodies (rods) in the muscle fibres. Nemaline bodies are aggregates of Z-disc and thin-filament proteins of the muscle sarcomere (Conen et al., 1963; Schroder et al., 2003; Shy et al., 1963; Wallgren-Pettersson et al., 1995).

The severity of the disease varies widely and thus NM can be described as a heterogeneous group of muscle disorders. NM is typically characterized initially by proximal muscle weakness (Figure 3), however, the phenotypes vary from neonatally lethal forms to mild muscle weakness. Four different NM patients presenting typical or severe NM are shown in Figure 2. They have mutation either in the *NEB* or *ACTA1* gene. Because of the clinical heterogeneity, NM has been divided into six clinical subgroups; typical, intermediate and severe congenital NMs, the mild childhood/juvenile onset, the adult-onset form and the so-called other forms of NM (Wallgren-Pettersson et al., 1999; Wallgren-Pettersson and Laing, 2000). Table 2 presents these different NM subgroups and their clinical criteria in more detail. Pathogenic variants in ten different genes have been published as causing NM. The disease can be inherited as an autosomal recessive (AR) or autosomal dominant (AD) trait, or it can be due to a *de novo* dominant mutation (Jungbluth and Wallgren-Pettersson, 2013).

Table 2. NM phenotype categories. This table summarises the clinical criteria for different forms of NM (modified from Wallgren-Pettersson et al., 2004).

Form of NM	**Clinical inclusion criteria**
Severe	**Onset at or before birth**; no spontaneous movements; no spontaneous respiration, or with severe contractures or fractures at birth
Intermediate	**Infantile onset**; patient breathing and moving at birth, but unable to maintain respiratory independence, or to sit and walk independently; use of wheelchair before the age of 11 years; contractures developing in early childhood
Typical	**Infantile onset**; typical distribution of muscle weakness (weakness most pronounced in facial, bulbar, and respiratory muscles, neck flexors and limb-girdle muscles; initially proximal, later also distal involvement); motor milestones delayed but reached; course slowly progressive or non-progressive
Mild childhood or juvenile onset	**Childhood or juvenile onset**
Adult onset	**Adult onset**
Other forms of NM	Unusual associated features such as cardiomyopathy, ophthalmoplegia, or unusual distribution of muscle weakness
Distal myopathy	Mainly distal involvement

3.2. The nebulin gene (NEB)

The main cause of recessively inherited NM are mutations of the nebulin gene (*NEB*, MIM *161650). *NEB* consists of 183 exons spanning 249 kb of genomic sequence, making it one of the largest genes in the human genome. There is only one initiation and termination codon in *NEB*, but several alternatively spliced exons theoretically give rise to hundreds of different nebulin isoforms (Donner et al., 2004; Pelin and Wallgren-Pettersson, 2008). It has been hypothesized that the large variety of different nebulin isoforms is needed because of the diverse requirements on muscle tissue at different stages of development, in different muscles and fibre types (Kazmierski et al., 2003; Labeit et al., 2011; Laitila et al., 2012). Alternatively spliced exons include

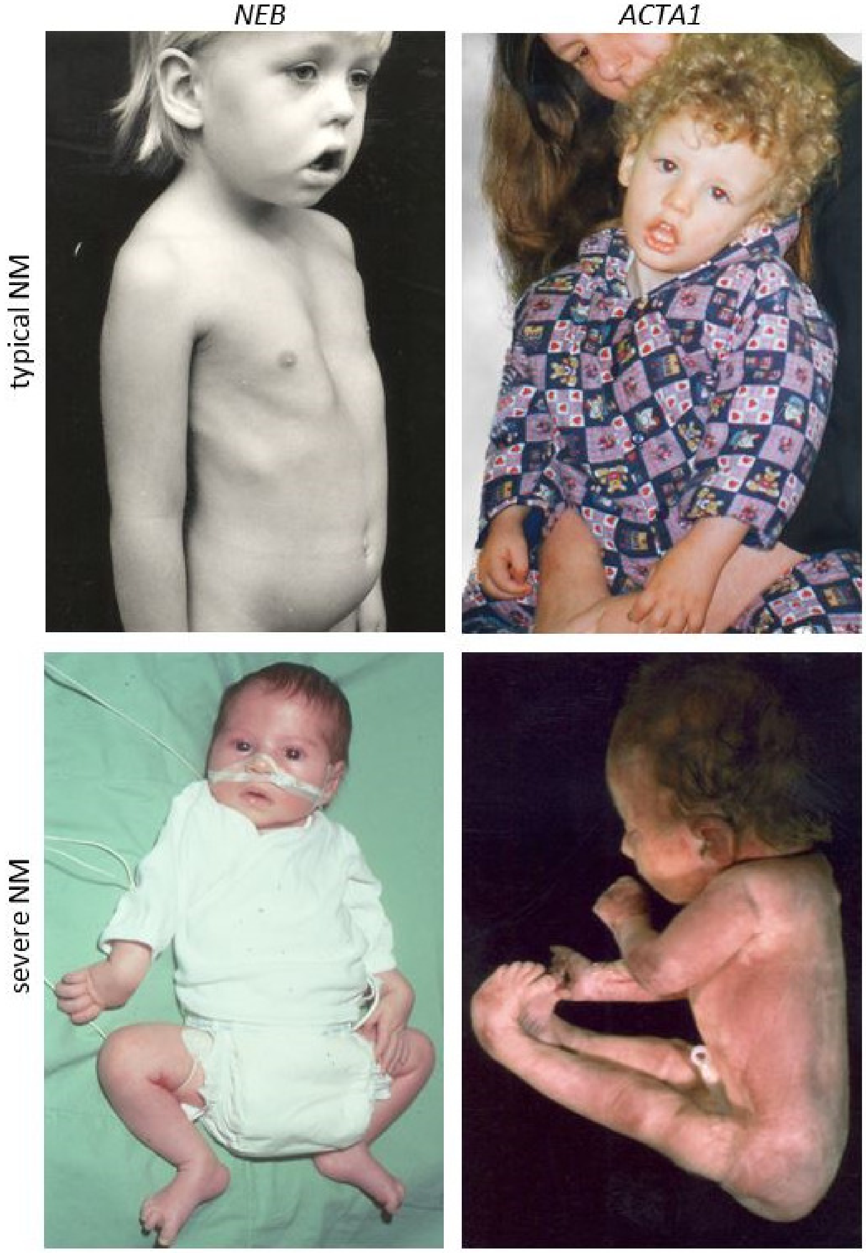

Figure 3. Nemaline myopathy patients with typical and severe NM caused by two recessive *NEB* mutations or one dominant *ACTA1* mutation (Wallgren-Pettersson, 1989; Wallgren-Pettersson, 1990; Wallgren-Pettersson et al., 2004).

The pictures are reprinted and modified with the permission of their copyright owners: Wallgren-Pettersson et al, 1989, Journal of the Neurological Sciences, 89,1-14, Elsevier Limited and Wallgren-Pettersson et al, 2004, Neuromuscular Disorders, 14, 461-470, Elsevier Limited.

exons 63-66, 143-144, and 167-177. Exons 63-66 form a cluster of exons which is either present in or absent from the transcript. The expression of the exon pair 143 and 144 is mutually exclusive, while exons 167-177 are independently spliced. In addition, there is a triplicate region (TRI), a block of eight exons that is repeated three times (82-89, 90-97, 98-105) in the middle of the gene. This region is also thought to undergo alternative splicing (Donner et al., 2004; Pelin and Wallgren-Pettersson, 2008). The region is difficult to study due to its repetitiveness and high homology. Overall, variant analysis of *NEB* is demanding due to the large size and complexity of the gene. Furthermore, the mutations are spread all along the large gene, which contributes to making variant analysis very laborious.

3.3. Other nemaline myopathy-causing genes

Many of the sporadic NM cases are caused by *de novo* mutations of the *ACTA1* (MIM *102610) gene. These most commonly cause the severe form of NM. Also familial AD, AR, and somatic mosaic mutations have been identified. Mutations are spread all along the six coding exons (Laing et al., 2009; Nowak et al., 1999). In *TPM2* (MIM *190990) and *TPM3* (MIM *191030), AD mutations are most common, usually causing mild or typical NM (Donner et al., 2002; Laing et al., 1995; Marttila et al., 2014; Tajsharghi et al., 2007). AR mutations in *TPM3* have also been identified causing severe forms of NM (Donner et al., 2002; Lehtokari et al., 2008; Marttila et al., 2014). In *TNNT1* (MIM *191041), a recessive nonsense mutation has been identified in the Old Order Amish population causing severe NM when present in a homozygous form (Johnston et al., 2000). Furthermore, two NM families outside the Amish population have been characterized with pathogenic variants in *TNNT1*. A Hispanic patient was identified with another homozygous nonsense mutation, as well as a Dutch pedigree in which the affected members carried either compound heterozygous or homozygous variants in *TNNT1* (Marra et al., 2015; van der Pol et al., 2014). In *CFL2* (MIM *601443), two different homozygous missense mutations have been identified, one in a sibling pair affected by NM, and another in a sibling pair with a congenital myopathy with features of NM and myofibrillar myopathy (Agrawal et al., 2007; Ockeloen et al., 2012). In addition, a homozygous frameshift mutation has been identified in a patient with severe NM (Ong et al., 2014). The most recent NM genes identified encode Kelch-family and Kelch-like-family proteins. In *KBTBD13* (MIM *613727), three different dominant mutations were characterized in four NM families (Sambuughin et al., 2010). In *KLHL40* (=*KBTBD5*, MIM *615340), 19 different AR mutations were identified in 28 probands with a very severe form of NM characterized by fetal akinesia (Ravenscroft et al., 2013).

Table 3. Nemaline myopathy genes and modes of inheritance (modified from Jungbluth and Wallgren-Pettersson, 2013).

Gene	Chromosomal region	Protein	Mode of Inheritance	Form of NM (phenotype)	OMIM number
ACTA1	1q42.13	Actin, α-skeletal	de novo/ AD/AR	severe most common, also intermediate, typical, mild and unusual ("other") forms	MIM *102610
NEB	2q23.3	Nebulin	AR	typical most common, all other forms also	MIM *161650
TPM3	1q21.3	Tropomyosin, α-slow	AD/AR	AD variable, AR usually severe	MIM *191030
TNNT1	19q13.42	Troponin T, slow	AR	severe ("Amish type", with tremors)	MIM *191041
TPM2	9p13.3	Tropomyosin, β	AD	variable	MIM *190990
CFL2	14q13.1	Cofilin 2	AR	unusual ("other") form (severity as in typical form, unusual distribution of muscle weakness)	MIM *601443
KBTBD13	15q22.31	Kelch repeat and BTB domain-containing protein 13	AD	childhood onset	MIM *613727
KLHL40	3p22.1	Kelch-like protein 40	AR	severe	MIM *615340
KLHL41	2q31.1	Kelch-like protein 41	AR	severe and milder	MIM *607701
LMOD3	3p14.1	Leiomodin-3	AR	severe	MIM *616112

Chromosomal regions are marked according to GRCh37. Abbreviations: AD=autosomal dominant, AR=autosomal recessive, NM=nemaline myopathy.

Five patients were identified with homozygous or compound heterozygous mutations in *KLHL41* (=*KBTBD10*, MIM *607701), with truncating mutations causing severe and missense mutations a milder form of NM (Gupta et al., 2013). Recently, the tenth causative NM gene, *LMOD3* (MIM *616112) was identified; 21 patients in 14 families were shown to have homozygous or compound heterozygous mutations in this gene, causing the severe form of NM (Yuen et al., 2014). Table 3 summarises all the known genes and their modes of inheritance in causing NM.

3.4. Nemaline myopathy-related disorders

Pathogenic variants in *NEB* can also cause other disorders besides nemaline myopathy. These include core-rod myopathy and various distal myopathies with or without nemaline bodies.

Distal nebulin myopathy has been diagnosed in patients in four Finnish families who presented only distal weakness, and no nemaline bodies could be detected in the routine histological examination. The identified pathogenic *NEB* variants in these families are also known to cause NM when they appear in compound heterozygous form with more disruptive *NEB* mutations (Wallgren-Pettersson et al., 2007).

Distal nemaline myopathy has been diagnosed in two families with *NEB* mutations, presenting distal weakness and showing nemaline bodies (Lehtokari et al., 2011).

Core-rod myopathy is another disorder in which pathogenic *NEB* variants have been identified in one family. This disorder presents with severe generalized muscle weakness and the muscle biopsy showing cores in addition to nemaline bodies. Cores are a typical finding in another congenital myopathy, central core disease (Romero et al., 2009).

Childhood-onset distal myopathy with rods and cores has been characterized in one patient. It is a novel disease entity caused by compound heterozygous mutations in *NEB*, the patients presenting with a distal distribution of limb muscle weakness, and showing nemaline rods and cores in the muscle biopsy (Scoto et al., 2013).

Mutations in *ACTA1, TPM2, TPM3,* and *CFL2* also underlie other myopathies. These include actin myopathy, congenital fibre type disproportion, central core disease, multi-minicore myopathy, intranuclear rod myopathy, cap myopathy, NM-myofibrillar myopathy, distal myopathies, distal arthrogryposis, and Escobar syndrome (Jungbluth and Wallgren-Pettersson, 2013).

This variety of mutations in different genes causing the same disorder, and mutations in the same gene causing different disorders, shows the remarkable clinical and genetic overlap of the congenital myopathies. In addition, there is currently no clear genotype-phenotype correlations for example for the pathogenic *NEB* variants, and the severity of the disease can vary even within the same family. Therefore, histological studies are important in the diagnosis of these patients. There is still much to learn about these disorders, thus mutation detection and the discovery of new genes is a valuable pursuit. Functional studies are of course the following step after mutation identification, to reveal the pathogenetic mechanisms of these disorders. Only after that is the development of potential specific therapies possible.

AIMS OF THE STUDY

This Doctoral Thesis is a part of a larger study of nemaline myopathy and related muscular disorders. It was aimed to design and implement novel variant detection methods for use in NM diagnostics and reveal novel disease-causing variants in the known NM genes, as well as to identify novel NM genes.

The aims of this study were to:

1. Design and validate a custom copy number microarray targeting the NM genes
2. Identify novel pathogenic variants in *NEB* and the other known NM-causing genes
3. Identify novel disease-causing NM gene(s) using whole-exome sequencing
4. Contribute by variant identification to the search for genotype-phenotype correlations in NM patients with *NEB* mutations

MATERIALS AND METHODS

4. Patient and control samples

4.1. *Patient samples*

This study has been approved by the Ethics Committee of the Children's Hospital, Helsinki University Central Hospital. The patients or their guardians have given their consent for their samples and data to be included in the study.

The sample cohort of this PhD study consists altogether of 356 families worldwide with diagnosed or suspected NM or a related muscle disorder (Figure 4). Study I includes the first 43 patient samples from 36 NM families that were studied using the NM-CGH array during the validation phase. This cohort is also included in the larger Study III that includes all the 266 samples from 196 NM families that had been studied using the NM-CGH array. Study IV presents one family with a novel pathogenic variant in *TPM3* that was identified among the studied 196 NM families using the NM-CGH array. Study II is a mutational update that presents all the 159 families that had been characterized with two pathogenic *NEB* variants using different methods from a cohort of altogether 309 families.

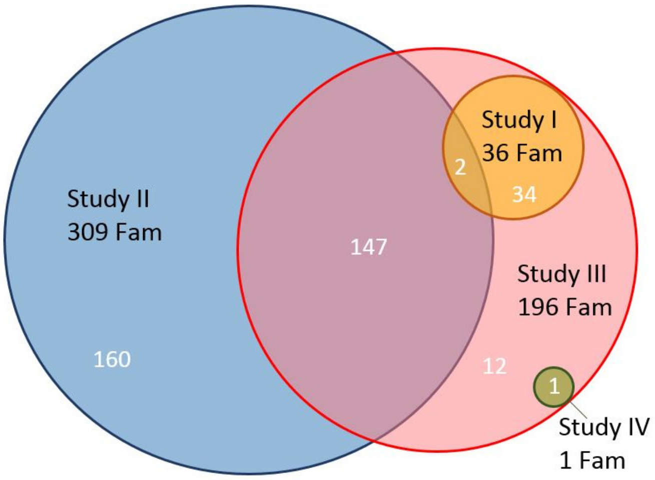

Figure 4. The number of studied families. This figure shows the number of studied families (n=356) with diagnosed or suspected NM or related muscle disorder in each publication (black numbers). All studies have samples shared with other studies (white numbers). Study III contains all the samples run by the NM-CGH array, including also the samples of Studies I and IV. Study II presents all the families with both identified *NEB* mutations, overlapping with samples included in Studies I and III. Abbreviations: Fam=Families.

The studied NM families are of different ethnic origin, including samples from Finland and other European countries, Asia, Australia, Africa, and America. The samples were received either as isolated DNA or as blood, cell lines, muscle, or skin biopsies from which DNA extraction was done using appropriate methods.

4.2. Control samples

The NM-CGH validation (Study I) included four positive control samples from two NM families with previously-identified disease-causing *NEB* variants and five healthy control samples. In addition, one sample with a known variant in the *TTN* gene was used as a control. To date, 60 healthy control samples have been studied with the NM-CGH array. These anonymous control samples were received from the Finnish Red Cross and the Centre d'Etude du Polymorphisme Humain (CEPH).

5. Methods

The main methods that have been used in the different studies are summarized in Table 4 and presented in more detail in the following chapters.

Table 4. Methods used in the different studies

Method	**Studies**
NM-CGH array	I-IV
Whole-exome sequencing	II, III, IV
MLPA	I, II
SSCP / dHPLC + Sanger sequencing	I-IV
PCR / RT-PCR + Sanger sequencing	I-IV

Abbreviations: NM-CGH array=nemaline myopathy comparative genomic hybridization array, MLPA=multiplex ligation-dependent probe amplification, SSCP=single-stranded conformation polymorphism, dHPLC=denaturing high-performance liquid chromatography, RT-PCR=reverse transcription PCR.

5.1. NM-CGH microarray

The NM-CGH 8x60k microarray (Oxford Gene Technology IP Limited, Oxford, UK) was designed (Human reference sequence, GRCh37/Hg19) to target the causative genes for NM known at the time. The NM-CGH array version 1 was validated including the genes *NEB, ACTA1, TPM3, TPM2, TNNT1, CFL2, KBTBD13* as well as the control gene *TTN*, encoding titin, densely covered with altogether ~53 000 probes. The 60mer oligonucleotide probes were designed to cover each of the genes including the exons, introns and exon-intron boundaries and ~25kb upstream and downstream of each gene. A tiling approach was used to achieve extremely high resolution for all the genes, albeit avoiding the most repetitive regions of these genes. In the NM-CGH array version 1 there is one probe starting at every 10 bp interval. In version 2, the *TTN* control gene was removed and three new nemaline myopathy-associated genes were included; *KLHL40, KLHL41*, and *LMOD3*. Moreover, the probe interval was reduced from 10 bp to

20 bp in the intronic regions for every gene except for *NEB*. No other significant modifications were made in the array update. In the NM-CGH array version 3, one novel putative NM gene, *YBX3* (=*CSDA*), has been included. Detailed information of each gene is presented in Table 5. The remaining ~3 600 probes were spread across the entire genome to yield a low-resolution backbone to the array. The targeted oligonucleotides were designed as replicate pairs for every 60mer sequence. The replicates were not identical, as the first replicate was complementary to the forward and the other to the reverse strand of the target sequence. The replicate approach was used for increasing the reliability of aberration calling. An 8x60k array platform was chosen, where eight different samples can be analysed simultaneously with an identical design of 60 000 probes on one array slide.

The labelling, hybridization, scanning and analysis was done according to the manufacturer's protocol (Oxford Gene Technology Ltd, Cytosure Genomic DNA labelling kit protocol, 8x60k format, version 1, 990097). The microarray was scanned using the Agilent DNA Microarray Scanner G2505C with 2 µm resolution, Agilent Scan Control version A.8.5.1 (Agilent Technologies Inc, Santa Clara, CA, USA), and normalised, filtered and further analysed in Feature Extraction software v10.7.3.1-12.0 (Agilent Technologies). The Cytosure Software v.3.4.9-v4.6.85 (Hg19) (Oxford Gene Technology Ltd) was used for graphic analysis of the data. The CBS (circular binary segmentation)

Table 5. The coverage of the NM-CGH microarray. The genes (presented according to GRCh37/hg19) and the number of probes targeting each gene in the NM-CGH microarray design versions 1, 2 and 3.

Gene	Chromosomal region	Start (bp)	Stop (bp)	Length (bp)	Probe count Design 1	Probe count Design 2	Probe count Design 3
ACTA1	1q42	229 566 992	229 569 845	2 853	1 090	918	918
NEB	2q23.3	152 341 850	152 591 001	249 151	33 866	33 836	33 836
TPM3	1q21.3	154 127 784	154 167 124	39 340	5 604	3 808	3 808
TNNT1	19q13.42	55 644 162	55 660 606	16 444	1 700	1 254	1 254
TPM2	9p13.3	35 681 989	35 691 017	9 028	2 218	1 790	1 790
CFL2	14q13.1	35 179 593	35 183 896	4 303	1 282	1 162	1 162
KBTBD13	15q22.31	65 369 154	65 372 276	3 122	1 310	1 278	1 278
KLHL40	3p22.1	42 727 011	42 734 036	7 025	0	1 233	1 233
KLHL41	2q31.1	170 366 212	170 382 772	16 560	0	1 668	1 668
LMOD3	3p14.1	69 156 023	69 172 183	16 160	0	1 771	1 771
YBX3	12p13.2	10 851 812	10 875 911	24 099	0	0	2 452
TTN (Ctrl)	2q31.2	179 390 716	179 695 529	304 813	5 616	0	0
Total				**692 898**	**52 686**	**48 718**	**51 170**

bp= base pair, Ctrl=control gene

algorithm was used for CNV calling and specific thresholds determined to distinguish the calling of aberrations. The threshold of +0.03 or +0.04 was applied to duplications and threshold of -0.06 or -0.07 for deletions. A minimum of five to ten probes was used as a threshold for making a positive call for an aberration depending on the data quality. The targeted genes and the aberration calls were then manually checked.

In addition to the custom NM-CGH microarray, the Cytosure ISCA+SNP 4x180k microarray (Oxford Gene Technology Ltd) was used for a couple of samples to study whole genomes including copy number variations and loss of heterozygosity (LOH). LOH regions larger than 10 Mb can be detected due to the SNP probes that are used on this array along with the CNV probes and visualized using the B-allele frequency plot. This method was used according to the manufacturer's protocol (Cytosure Genomic DNA labelling kit protocol, 4x44k/4x180k format, version 1, 9900107) and analysed in the similar way as the NM-CGH array data.

5.2. Whole-exome sequencing

Exome capture and sequencing was done by OGT (Oxford Gene Technology Ltd, Oxford, UK) using the Agilent SureSelectXT All Exon 50Mb target enrichment kit (protocol v1.2; Agilent Technologies) on an Illumina HiSeq2000 platform using TruSeq v3 chemistry (Illumina Inc, San Diego, CA, USA). Exome analysis was completed using the OGT exome sequencing pipeline including the Burrows-Wheeler Aligner (BWA) package v.0.6.2 to map the reads to the human genome build 19 (hg19). Genome Analysis Tool Kit (GATK) v.1.6 was used to ensure a minimum number of the mismatches across the reads. Picard software versions 1.89-1.107 were used to mark duplicate reads likely to be resulting from PCR bias. BAM files were additionally processed with Samtools v.0.1.18. GATK was used for base quality scoring and indel variant calling. Variants were annotated with Ensembl data and dbSNP release 135 was used to determine novel SNPs. The data was then manually checked and Integrative Genomics Viewer (IGV) version 2.3.5 was used to visualize the data. Polyphen (Polymorphism Phenotyping) version 2 was used to predict possible impacts of an amino acid substitution on the structure and function of a protein, SIFT (Sorting Intolerant from Tolerant) was used to predict whether an amino acid substitution affects protein function and Condel (Consensus Deleteriousness score) to estimate the outcome of non-synonymous single nucleotide variants (Adzhubei et al., 2010; Gonzalez-Perez and Lopez-Bigas, 2011; Ng and Henikoff, 2003).

5.3. Multiplex ligation-dependent probe amplification

MLPA analyses were done targeting the *NEB* gene exons of interest with self-designed oligonucleotide probes using the MLPA SALSA kit according to the manufacturer's protocol (MRC-Holland, Amsterdam, the Netherlands) as described in study I. The fragment separation and analysis was done using ABI-3730-XL DNA analyzer and the GeneMapper 4.0 software (Applied Biosystems, CA, USA). The final copy number analysis was done using Coffalyser v.7 software (MRC-Holland, Amsterdam, the Netherlands) or an in-house Excel-based software.

5.4. Denaturing high-performance liquid chromatography

The majority of the variant analyses with dHPLC had been done in our group prior to this thesis project. In this method, all *NEB* exons of a DNA sample, excluding the TRI region (exons 82-105), were amplified in 1-2 exon sets using PCR. The PCR samples were denatured and then slowly reannealed to form heteroduplexes and homoduplexes. The melting profiles of these duplexes were analysed using the Transgenomic WAVE Nucleic Acid Fragment Analysis System and the Navigator software (Transgenomic, San Jose, CA, USA) as previously described (Lehtokari et al., 2006).

5.5. Sanger sequencing

Until the year 2012 in our research group, Sanger sequencing was mainly used following aberrant SSCP or dHPLC screening results and later as an individual method where *NEB* exons were amplified in 1-2 exon sets. Sanger sequencing was performed on PCR or RT-PCR products using an ABI-3730-XL DNA analyser (Applied Biosystems, Fosters City, CA, USA) and analysed using the Sequencher software versions 4.0-5.0 (Gene Codes corporation, Ann Arbor, USA) as previously described (Lehtokari et al., 2006).

5.6. Bioinformatics methods

Bioinformatic methods were used to analyse the breakpoints of the highly repetitive *NEB* TRI region. The data was analyzed using the GLAD package (Hupe et al., 2004) for the R environment (R Core Team, 2013) and the breakpoints and copy numbers were inferred using the daglad function. The resulting breakpoints, along with the original signal intensities and the genome annotations for the target region, were visualized using the GenomeGraphs (Durinck et al., 2009) and rtracklayer (Lawrence et al., 2009) packages.

RESULTS AND DISCUSSION

6. New variant detection methods

In this study, new variant detection methods have revealed several novel disease-causing variants as well as mutation types causative of NM. The main focus was on the huge *NEB* gene in which we identified copy number variations, such as deletions and duplications, using the self-designed and validated NM-CGH microarray. Using whole-exome sequencing (WES), we identified small variants including nonsense, missense and frameshift mutations. All pathogenic *NEB* variants are reported according to the reference sequence NM_001271208.1. The *ACTA1* variants are marked according to the coding sequence of CR536516.1 and *TPM3* variants according to the reference sequence NG_008621.1. The numbers of different novel pathogenic variants that have been identified using these novel methods in our group are summarized in Table 6.

Table 6. Novel pathogenic variants identified using new methods. This table presents the number of different novel pathogenic variants and the number of nemaline myopathy families that they were identified in. Each different variant has been counted only once and the variants have been designated to the study in which they were been published for the first time.

	NM-CGH array		**Whole-exome sequencing**	
Published in	**Families**	**Variants**	**Families**	**Variants**
Study I	2	2	0	0
Study II	5	4	2	3
Study III	8	4	3	3
Study IV	1	1	0	0
Unpublished	2	2	2	3
Total	**18**	**13**	**7**	**9**

6.1. NM-CGH microarray (I-IV)

When this study began, only one copy number variation, i.e. a mutation larger than 1 kb, had been characterized in *NEB.* The deletion of the entire *NEB* exon 55 was mainly found in the Ashkenazi Jewish population as a founder mutation (Anderson et al., 2004; Lehtokari et al., 2009). The number of *NEB* copy number variations is now steadily increasing as new methods, such as the NM-CGH microarray, enable much better detection of large variants. The Database of Genomic Variants (DGV) indicated a few copy number variations (CNV) in *NEB*, but these studies have mainly focused on finding CNVs on a genome-wide scale with different approaches and detection resolution (Alkan et al., 2009; Conrad et al., 2010; Perry et al., 2008). However, the incidence of these *NEB* variations and their possible phenotypic relevance remains unclear based on these control study groups. Therefore, a more focused study was initiated to elucidate this.

A new custom tiling NM-CGH microarray platform was designed to target the known NM genes with a high resolution. The NM-CGH microarray was validated using the only previously known copy number variation, the deletion of *NEB* exon 55 (Anderson et al., 2004). Using the NM-CGH array, an index patient affected with the intermediate form of NM (Family 105) was unambiguously shown to be homozygous and both parents heterozygous for this 2.5 kb *NEB* exon 55 deletion. The array profiles between the heterozygous and homozygous deletions were distinct. These results were verified using MLPA. (Study I)

6.1.1. Novel pathogenic *NEB* copy number variations (I-III, U)

To date, 266 samples from 196 NM families have been studied with the NM-CGH array. We have identified nine different novel copy number variations in *NEB* in ten different families. One of these novel variations was detected in two different families. All identified pathogenic variations are listed in Table 7 according to the size of the variant by NM-CGH. From the majority of families, samples only from the index patient with or without the parental samples were available. Each family is described in more detail below.

Family 255 (sample 2553) had no previously identified mutations. To the best of our knowledge, this patient is the first one with NM with onset definitely in adulthood found to have a pathogenic variant in *NEB*. The NM-CGH array suggested a rather small deletion covering only a small part of the 3’ end of *NEB* exon 15 (Figure 5A). Sanger sequencing verified this mutation showing a 72 bp in-frame deletion in exon 15 (c.1291_1362del) deleting 24 amino acids but the last amino acid of this exon is present in the sequence. This variation is interpreted as likely pathogenic because it deletes several amino acids and disrupts an actin-binding site. Parental samples were not available. In this patient, a second pathogenic *NEB* variant remains to be identified. This deletion is the smallest mutation identified by the NM-CGH array to date, and it is at the threshold of the method. It shows that in an optimal situation, this custom array is capable of detecting rather small copy number variations. (Study U)

Table 7. The novel pathogenic *NEB* variants identified or further studied using the NM-CGH array

Family	Patient	Gene	Aberration	Exon Count	Size (bp)	Verification	Breakpoint	Study
F255	2553	*NEB*	ex 15 partly del	part of ex 15	72 bp	PCR+ Sanger seq	ex15 -> T-microhomology ->ex15	U
F318	3183	*NEB*	ex 73 partly del	3' end of ex 73	914 bp	PCR+ Sanger seq	ex73 -> T-microhomology ->int73	II
F45	5451	*NEB*	ex 53-54 del	2	1084 bp	MLPA, PCR+ Sanger seq	int52 -> AGCT-microhomology ->int54	I, II
F211	2113	*NEB*	ex 43-45 del	3	2.2 kb	PCR+ Sanger seq	int42-> AT-microhomology ->int45	II
F333	3333	*NEB*	ex 43-45 del	3	2.2 kb	PCR+ Sanger seq	int42-> AT-microhomology ->int45	II, III
F321	3213	*NEB*	ex 69 partly-71 del (+int65 partly)	2+	4.2 kb del + 110 bp ins	PCR+ Sanger seq	ex69-> AATT-linker ->int65…int65-> GG-microhomology-> int71	III
F181	1813	*NEB*	ex 1-24 del	24	53-64 kb	MLPA	exact breakpoints unknown	I, II
F309	3093#	*NEB*	ex 1-51 dup (starts 5' of *NEB*)	51	80 kb from *NEB* (minimum of 103 kb in total)	PCR+ Sanger seq int 51 breakpoint, 5' breakpoint under study	exact breakpoints unknown	II
F396	3962, 3963, 3964	*NEB*	ex 14-81, 82-105 del	68+TRI	88 kb	MLPA, WES	exact breakpoints unknown	U
F410	4103	*NEB*	ex 1-81, 82-105 dup	81+TRI	133 kb	WES by group Laporte	exact breakpoints unknown	II

Abbreviations: bp=base pair, del=deletion, dup=duplication, ex=exon, int=intron, MLPA=multiplex ligation-dependent probe amplification, NM-CGH array=nemaline myopathy comparative genomic hybridization array, Sanger seq=Sanger sequencing, TRI= triplicate region of the nebulin gene, WES= whole-exome sequencing.

Family 318 (sample 3183) had one previously identified disease-causing variant, a del_TCAA (c.24480_24483del; p.Gln8161fs) in *NEB* exon 173, causing a frameshift, inherited from the mother. The index case was affected by the typical form of NM. The NM-CGH array revealed a 0.9 kb deletion covering only a small part of the 3' end of *NEB* exon 73 and most of intron 73 (c.10798_10872+839del) (Figure 5B), creating a truncating mutation. This was verified by Sanger sequencing, defining the mutation breakpoints and the deletion size as 914 bp. (Study II)

Family 45 (sample 5451) was analysed to search for the second pathogenic *NEB* variant. The index case had the typical form of NM. There was a previously identified splice-site mutation in *NEB* intron 36 caused by an inversion (c.3987+1_+2inv) in the donor splice site (Lehtokari et al., 2006). NM-CGH analysis revealed a 1.1 kb deletion covering *NEB* exons 53-54 (c.6916-163_7431+211del) in the NM-CGH array profiles of the index case and the father (Figure 5C). The deletion of these two exons is most likely pathogenic.

Even though Family 45 as well as Family 105, carrying the previously identified deletion of *NEB* exon 55 (Ashkenazi founder mutation), have deletion breakpoints in intron 54, they do not share an identical breakpoint. There is an approximately 1.7 kb distance between the two breakpoints in this intron. (Studies I, II)

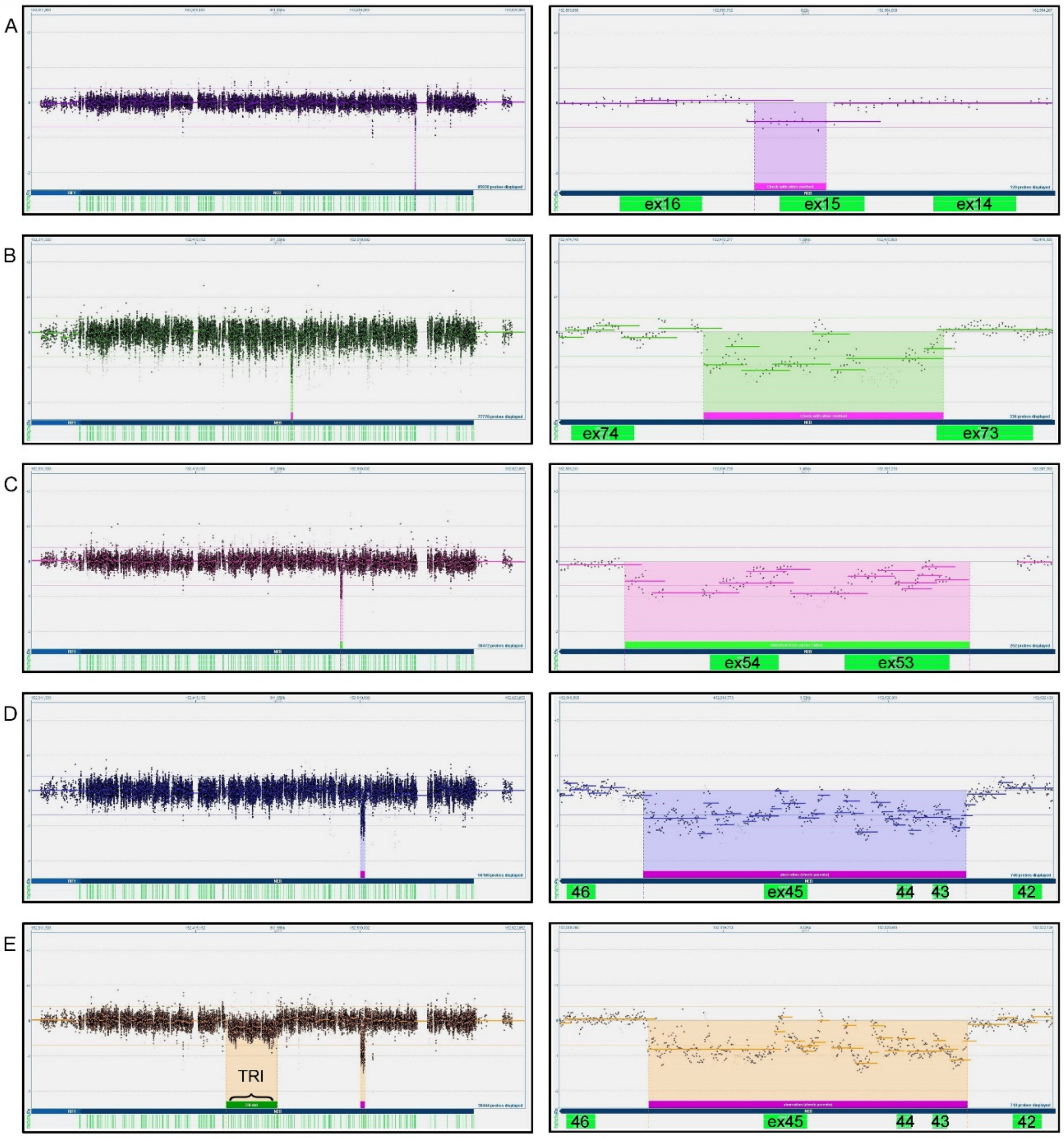

Figure 5. NM-CGH profiles of the *NEB* gene (on the left) and the zoomed aberrations (on the right) of five different nemaline myopathy patient samples. *NEB* is located in the reverse strand of the genomic DNA and therefore the exon numbering is ascending from right to left. The aberrations are marked with highlighting. A) Sample 2553 shows a 0.1 kb deletion in the 3' end of *NEB* exon 15. B) Sample 3183 shows a 0.9 kb deletion including the 3' end of *NEB* exon 73. C) Sample 5451 shows a 1 kb deletion covering the *NEB* exons 53-54. D) Sample 2113 shows a 2.2 kb deletion covering the *NEB* exons 43-45. E) Sample 3333 shows a 2.2 kb deletion covering the *NEB* exons 43-45 and a loss of one copy in the 32 kb *NEB* TRI region.

Family 211 (sample 2113) was also analysed to search for the second disease-causing variant. The index case had the mild form of NM and a previously identified nonsense mutation in *NEB* exon 155 (c.22489C>T; p.Arg7497*), causing a premature stop codon. A 2.2 kb deletion covering *NEB* exons 43-45 was discovered using the NM-CGH array (c.5238+335_5764-407del) (Figure 5D). This was verified by Sanger sequencing that revealed the exact breakpoints of the deletion in introns 42 and 45 and gave the exact deletion size of 2249 bp. The deletion of these three exons is most likely pathogenic. Parental samples were not available for NM-CGH analysis. (Study II)

Family 333 (sample 3333) had one previously identified pathogenic variant. The index patient had a mild form of NM with unusual features. He had started to have muscle pain and weakness at the age of two years, and weakness of the facial muscles was absent. The previously identified pathogenic variant was a frameshift mutation in *NEB* exon 130 (c.19992_19999dup; p.Asp6667fs). The NM-CGH array revealed a 2.2 kb mutation which was very similar to the one detected in family 211, a deletion covering exons 43-45 (c.5238+335_5764-407del) (Figure 5E). Sanger sequencing verified the mutations in these two families to be exactly the same. In addition, a deletion in the *NEB* TRI region was identified. This variation will be discussed in more detail in the next section covering the *NEB* TRI results (6.1.3). Parental samples were not available. (Studies II and III)

Family 321 (sample 3213) had no previously identified pathogenic variants. The index case likely had the typical form of NM. The NM-CGH array revealed a ~4.1 kb deletion including exons 70-71 starting at the end of exon 69 or the beginning of intron 69 (Figure 6A). However, Sanger sequencing revealed that this was not a simple deletion. The deletion turned out to cover 4.2 kb starting at the 3′ end of exon 69 and continuing with an AATT linker to intron 65. There is a 114 bp sequence from intron 65 after which the sequence continues normally from intron 71. The 114 bp duplication of intron 65 did not show on the NM-CGH array because of a gap between probes in this region due to an *Alu* repeat sequence. Parental samples were not available. In addition, a duplication of the *NEB* TRI was identified. This variation will be discussed in more detail in the next section covering the *NEB* TRI results (6.1.3). (Study III)

Family 181 (sample 1813) was analysed to verify and further characterize the variation suspected based on the previous MLPA results. MLPA had indicated a large deletion in the 5′ end of *NEB*. The NM-CGH array revealed a large deletion of *NEB* exons 1-24 in the studied patient sample and the mother (Figure 6B). The size of the deletion lies somewhere between 53 and 64 kb. The exact breakpoints of this variation are difficult to identify due to lack of probes, especially at the deletion breakpoint residing upstream of the 5’UTR of *NEB*. The mutated allele is lacking the promoter and is, thus, not transcribed. This family included two children affected with an unclassified form of NM

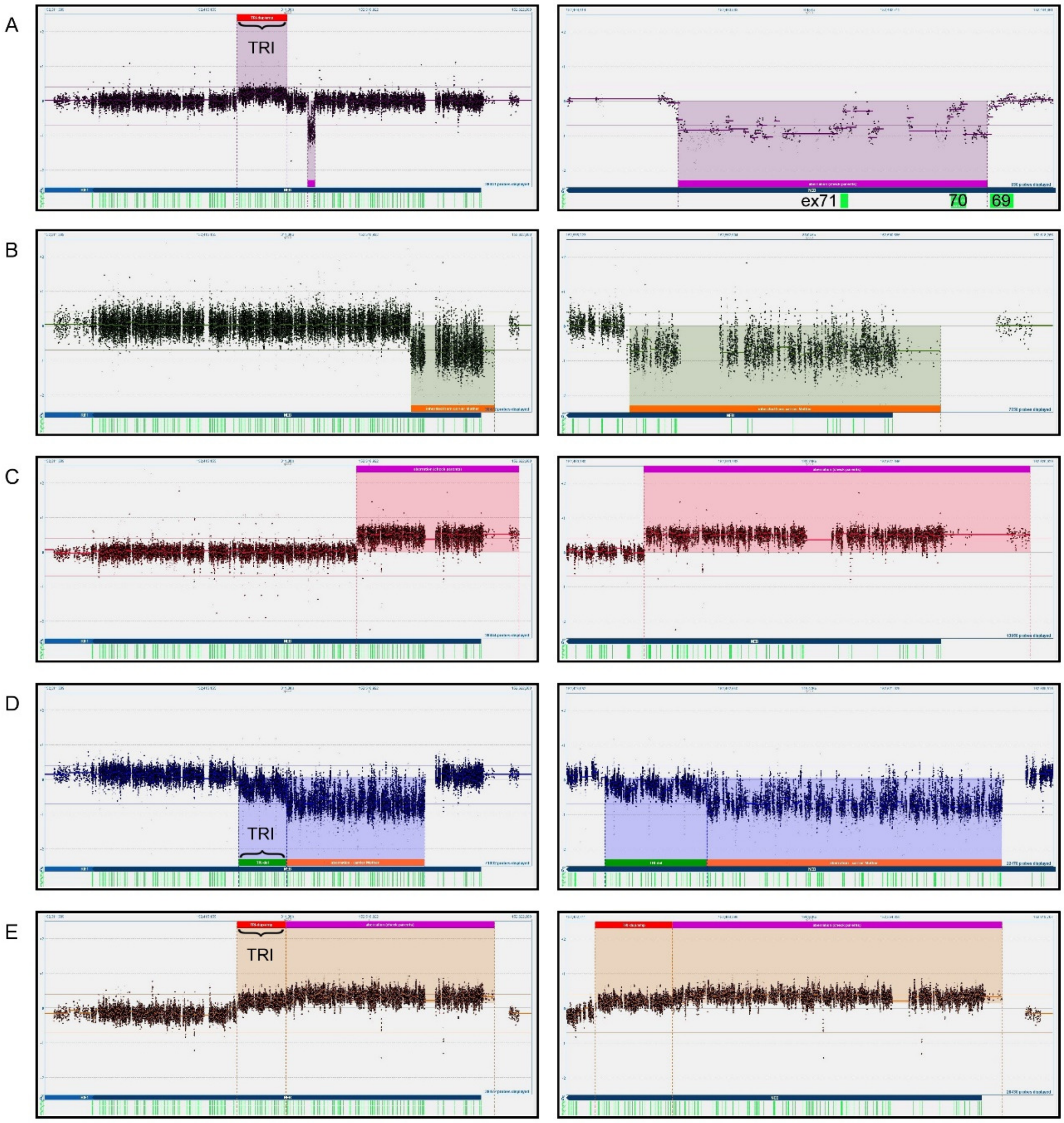

Figure 6. NM-CGH profiles of the *NEB* gene (on the left) and the zoomed aberrations (on the right) of five different nemaline myopathy patient samples. *NEB* is located in the reverse strand of the genomic DNA and therefore the exon numbering is ascending from right to left. The aberrations are marked with highlighting. A) Sample 3213 shows a 4.1 kb deletion covering at least the *NEB* exons 70-71 and a gain of one copy in the 32 kb *NEB* TRI. B) Sample 1813 shows a 53-64 kb deletion covering the *NEB* exons 1-24. C) Sample 3093 shows a duplication covering the *NEB* exons 1-51 (80 kb), altogether the duplication covers >103 kb. D) Sample 3963 shows an 88 kb deletion covering the *NEB* exons 14-81 and a loss of one copy in the 32 kb *NEB* TRI. E) Sample 4103 shows a 133 kb duplication covering the *NEB* exons 1-81 and a gain of one copy in the 32 kb *NEB* TRI.

(clinical data lacking). Previously, a point mutation in *NEB* exon 129 had been identified (c.19913G>C), causing an amino acid shift (p.Arg6638Pro) (Lehtokari et al., 2006). This had not been inherited from the mother while a sample from the father was unavailable. (Studies I, II)

Family 309 (sample 3093) had one previously identified disease-causing variant, a nonsense mutation in *NEB* exon 78 (c.11610C>A; p.Tyr3870*). This patient had a mild form of NM with unusual distribution of muscle weakness. The NM-CGH array identified a large duplication including *NEB* exons 1-51 (Figure 6C). The breakpoint in intron 51 was further characterized with Sanger sequencing to be in the proximity of nucleotide c.6807+573. However, the second breakpoint is located more than 25 kb upstream of the *NEB* gene and thus could not be identified with the NM-CGH array due to low probe density. This breakpoint and the orientation of the duplication are under further investigation. The size of this duplication may be much larger than can be estimated by the NM-CGH array. Parental samples were unavailable. (Study II)

Family 396 (samples 3962, 3963 and 3964) had no previously identified disease-causing variants. The phenotype of the index patient (3964) was mild, with distal weakness as the presenting feature, which is atypical of NM. He was initially diagnosed as having tibial muscular dystrophy. However, on follow-up, the child was found to have a more generalized muscle weakness, with facial, bulbar and neck flexor weakness as well as weakness of the axial and distal limb muscles. There were also two other affected family members, the mother (3963) and maternal grandmother (3962) of the index patient, indicating autosomal dominant inheritance. The mother and grandmother presented later with mild distal weakness. The NM-CGH array showed in all of these three affected family members a large deletion in *NEB* including the exons 14-81 and continuing across the *NEB* TRI exons 81-105 (Figure 6D). This deletion was verified by MLPA. Other family members were studied and their NM-CGH profiles were normal. The cDNA study from the 3964 patient showed that there are heterozygous SNPs outside the deletion region of *NEB* suggesting that this mutated allele is expressed. Despite several attempts with different PCR primers and settings, PCR of the deleted allele has not been successful and thus, Sanger sequencing has not been possible. WES was done using the samples of the index patient (3964) and the mother (3963) but no additional pathogenic variants were identified in *NEB*. This is the first family that might indicate dominant inheritance of a disorder caused by a variant of the *NEB* gene. However, this variant and family warrant further investigation. (Study U)

Family 410 (sample 4103) had one previously identified splice-site mutation in *NEB* exon 129 (c.19944G>A). The phenotype of the index case was the distal form of core-rod myopathy. The second variant required further investigations. The whole-exome sequencing previously done by our collaborators (Jocelyn Laporte /France) had indicated lower read depths in the 3' end of *NEB*. In fact, the NM-CGH array showed that there was a duplication of *NEB* exons 1-81 continuing as a one-copy gain of *NEB* TRI exons 81-105 (Figure 6E). The estimated size of this duplication is 133 kb, not including the *NEB* TRI gain of one copy. The exact duplication breakpoints have remained unidentified. The size of this duplication makes it difficult to design appropriate PCR experiments. This duplication is the largest mutation identified thus far using the NM-CGH array and it requires further investigation. (Study II)

6.1.1.1. Breakpoints of the novel pathogenic NEB variations (I-III, U)

The exact aberration breakpoints could be identified for 60% of the families (6/10) (F255, F318, F45, F211, F333 and F321) with novel aberrations in *NEB*. All the identified aberrations were deletions that had microhomology in the deletion breakpoints. The microhomology differed between 1-4 nucleotides (Table 7). This type of microhomology might indicate non-homologous end joining (NHEJ) or microhomology-mediated break-induced replication (MMBIR) as the causative mechanism. Both of these mechanisms use microhomology in repairing DNA, and they may create aberrations in this repair process (Hastings et al., 2009; Liebert, 2008). Microhomology-mediated end joining (MMEJ) might also be a possible mechanism, but it usually involves longer stretches (5-25 bp) of microhomology (McVey and Lee, 2008). Family 321 had a slightly more complex aberration, including an AATT-linker and insertion of additional sequence from a nearby intron 65 in the other breakpoint and a GG-microhomology at the other breakpoint. These multiple consecutive switches of template reading may indicate Fork Stalling and Template Switching (FoSTeS) or MMBIR as the causative mechanism. FoSTeS is thought to be mediated by microhomology of the original and invaded site and it is known to be capable of inducing complex rearrangements (Lee et al., 2007). On the other hand, MMBIR can also switch the template multiple times and it is also capable of inserting DNA sequence stretches from elsewhere to the repaired site, creating complex rearrangements (Hastings et al., 2009). On the other hand, the information scar of the additional AATT-linker could also suggest NHEJ (Shaw and Laski, 2004). However, these possible mechanisms behind the aberrations are currently hypotheses only, and require further investigations. Data regarding the different mechanisms is accumulating rapidly and more samples are being analyzed in detail, which will help to elucidate the exact mechanisms behind these aberrations as well. (Study U)

Four of the families (F181, F309, F396 and F410) with novel *NEB* aberrations have remained without characterization of the exact breakpoints thus far. Therefore, estimating the mechanisms behind these aberrations is difficult. The alterations include large deletions and duplications containing breakpoints that are difficult to study. Some breakpoints are embedded in sites of probe gaps in the NM-CGH array design. Inside the genes, the probe gaps mainly consist of sites of repetitive sequences. In addition, some aberrations extend outside the targeted gene regions. In these types of situations, the elucidation of the exact breakpoints is much more laborious with downstream analyses.

In general, only a small part of the identified CNV breakpoints have been clarified to the single-nucleotide level in different studies thus far. Traditional PCR verification can be laborious, especially if no estimation of the CNV formation, such as the orientation of a duplicated segment is available. Even so, complex rearrangements can be extremely difficult to catch by traditional PCR methods. Genome-wide sequencing is possible for example using whole-genome NGS methods, but it is an expensive method to use only for verifying identified CNVs. Repetitive DNA poses its own difficulties that make PCR verification followed by sequencing very challenging. The PCR product itself can be difficult to obtain, and because sequencing is usually done in rather small segments (such as 300 bp) this makes studying of large duplications rather difficult. The breakpoints embedded in large repetitive segments will often remain unidentified (Conrad et al., 2010). Nevertheless, the study for deciphering the breakpoints of the identified novel *NEB* aberrations will continue. The characterization of these breakpoints may help to recognize some pattern in the aberration formation and possible sites exposed for mutations in the large nebulin gene.

6.1.2. *NEB* triplicate region copy number variations (II, III)

In addition to novel *NEB* variants, a frequent copy number variation of the *NEB* triplicate region (TRI) was identified using the NM-CGH array. The homologous *NEB* TRI region includes eight exons that are repeated three times (exons 82-89, 90-97, and 98-105). The normal copy number of the *NEB* TRI is thought to be six, i.e. three copies in each allele. The TRI variations were analyzed manually based on the logarithmic scale of the NM-CGH microarray results. Based on the results of patients and controls, it seems that deviations of one copy are tolerated but 2-4 copy gains might be pathogenic.

Of the 196 NM families studied with the NM-CGH array, 5% (9/196) showed a copy number loss (deletion) and 8% (16/196) a copy number gain (duplication) (Figures 5E and 6A) of *NEB* TRI. In addition, one family (F268) was identified including members with a gain, a loss, or a normal *NEB* TRI copy number (Figure 7). All in all, *NEB* TRI variations were identified in 13% of the NM families and in 10% of the control samples in our study cohort (Table 8). This makes *NEB* TRI the most common *NEB* variation characterized so far.

All the identified losses in nine different NM families are deletions of one TRI copy (5/6 copies present) and this aberration does not seem to segregate with the disorder. Moreover, the loss of one copy appears to be even more common in the studied control population. One-copy losses were identified in 8% of the control population samples (5/60) and in only 5% of the studied NM families (9/196). Five of the NM families have both disease-causing variants characterized, two families have one identified pathogenic *NEB* variant and two families remained without any being identified. The loss of one copy would cause an estimated 1458 bp shortening of *NEB* transcripts and a 486 amino acid shortening of the translated protein, which corresponds to two nebulin super repeats (Donner et al., 2004). The two remaining TRI copies are estimated be enough for the allele to produce a functional protein. Many shorter transcripts of *NEB* are known to be produced in the normal muscle, and therefore, this alteration might not cause a drastic effect on the sarcomere thin actin filament structure.

Among control samples, only one copy number gain (2%, 1/60) was identified. This sample showed a one-copy gain (7/6 copies present). Among NM families, copy number gains were much more common and contained gains of 1-4 additional copies.

Table 8. Summary of *NEB* triplicate region (TRI) variations identified with the NM-CGH microarray in the NM family and control cohorts

	NM tamiliesi		**Controls**	
All	**100 %**	**(196)**	**100 %**	**(60)**
***NEB* TRI variations**	**13.3 %**	**(26)**	**10.0 %**	**(6)**
Losses	**4.6 %**	**(9)**	**8.3 %**	**(5)**
1 copy loss	4.6 %	(9)	8.3 %	(5)
Gains	**8.2 %**	**(16)**	**1.7 %**	**(1)**
1 copy gain	4.6 %	(9)	1.7 %	(1)
2-4 copy gain	3.6 %	(7)	0.0 %	(0)
losses and gains	**0.5 %**	**(1)**	**0.0 %**	**(0)**

Abbreviations: *NEB* TRI=triplicate region of the nebulin gene, NM=nemaline myopathy, NM-CGH array=nemaline myopathy comparative genomic hybridization array.

Approximately 5% (9/196) of the NM families harboured a gain of one additional TRI copy (7/6 copies present). This variant did not seem to segregate with the disorder. In one of these families (F335), the disease-causing variants had previously been identified in *LMOD3* (Yuen et al., 2014). A pathogenic frameshift variant c.24475_24479dupCACAA was also identified in exon 173 of *NEB*, in another family (F407) segregating on the same allele as the TRI gain. The consanguineous healthy parents are both carriers of one TRI copy gain, as well as a *NEB* frameshift variant, and the patient is homozygous for both variants. One or both causative pathogenic variants remained unknown in the remaining seven families with a one-copy gain of the *NEB* TRI. Based on this data, it seems that deviations of one copy in a *NEB* allele, loss or gain, are tolerated.

Furthermore, 4% of NM families (7/196) had patients with a gain of more than one TRI copy (8-10/6 copies present). All of these families had unidentified pathogenic variant(s), and where parental samples were available for analysis, this aberration seemed to segregate with the disorder. In addition, this type of aberration was not detected in the control population, however, the difference is not statistically significant (Fisher's exact test). Moreover, three samples from patients with this type of *NEB* TRI gain were further tested using whole-exome sequencing. No further variants in *NEB* or any other NM gene were identified. One family (F268) is described here in more detail.

Family 268: The index case 2683 (typical NM) has inherited a *NEB* TRI four-copy gain from the mother and a *NEB* exon 28 frameshift variant (c.2784delT) from the father. The unaffected father has a frameshift variant (c.2784delT) on one allele and a TRI deletion (one-copy loss) on the other allele, which supports the conclusion that one-copy deletions of TRI would not be pathogenic. We suggest that the four-copy gain carried by the healthy mother, and inherited by the index patient, might be pathogenic (Figure 7). The pedigree shows that the *NEB* TRI gain segregates with the disorder, and only the combination of the *NEB* TRI gain and the frameshift variant result in the NM phenotype. Gains of more than one copy could disrupt the stability or secondary structure of the mRNA, and the transcription process, because of excess copies of the homologous region. In this case, no nebulin protein would be produced from this allele. If the transcription would work, the duplication would cause lengthening of the transcript and the nebulin protein. For example, a gain of four copies would add 5.8 kb to the mRNA and 1944 amino acids or eight super repeats to the protein. This could interrupt the normal repeat structure and disturb the interactions with the binding partners and also thin filament formation, especially if the duplications are inverted. We suggest that loss or gain of one *NEB* TRI copy would be benign, but deviations of more than one copy from the normal three copies in one allele may be pathogenic.

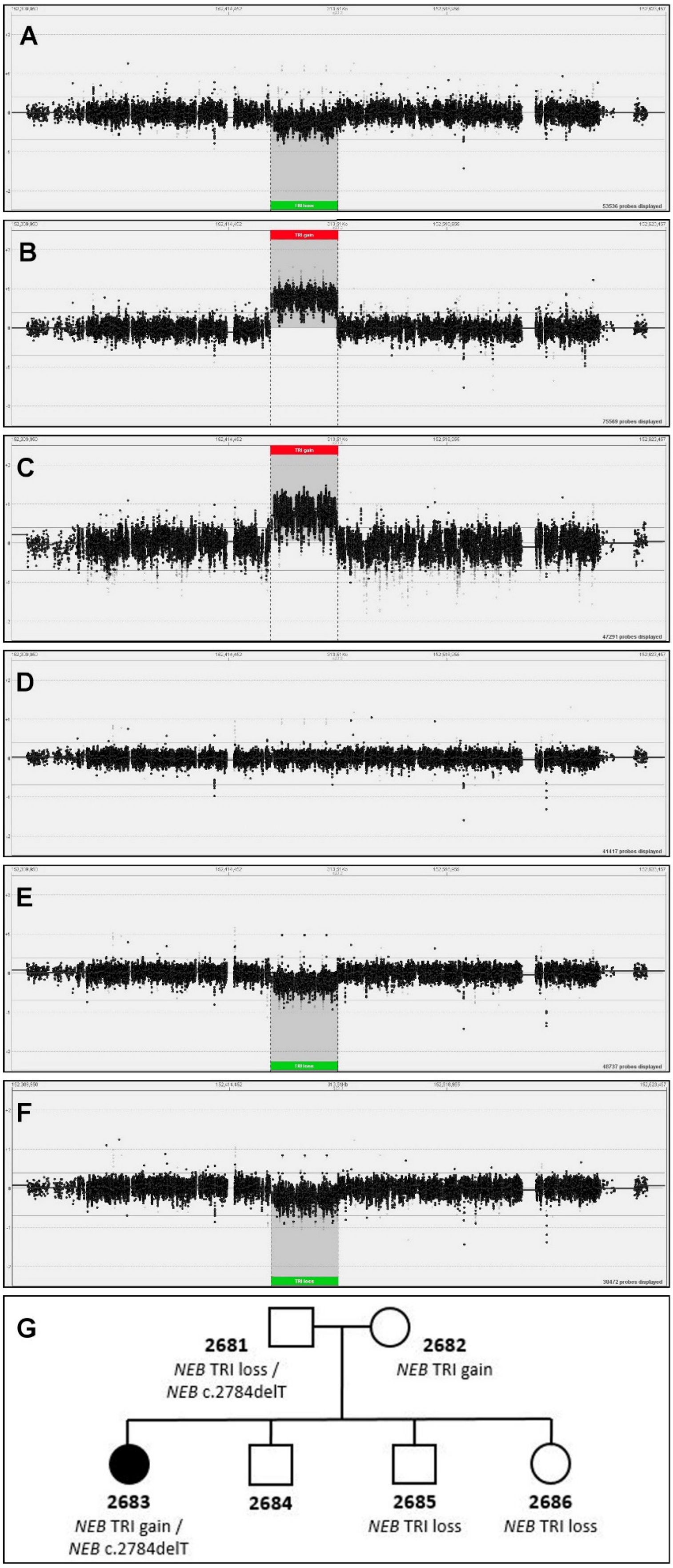

Figure 7. NM-CGH profiles of *NEB* in family 268. The index patient (C) (typical form of nemaline myopathy) has inherited a gain of the triplicate region of the nebulin gene (*NEB* TRI) from the mother (B) and a *NEB* (c.2784delT) frameshift mutation from the father (A) (not shown in the NM-CGH array). The father (A) and two siblings (E, F) have a one-copy deletion and one brother (D) the normal copy number of *NEB* TRI. The index patient has the typical form of NM and all other family members are unaffected. The family pedigree is shown in Figure 7G.

Figure 7 is reprinted and modified with the permission of the copyright owner: Kiiski et al, 2015, Eur J Hum Genet. doi:10.1038/ejhg.2015.166. Licensee the Nature Publishing Group.

6.1.2.1. Breakpoints of the NEB triplicate region copy number variations (III)

The breakpoints of the TRI variations are difficult to characterize. The breakpoint analyses of two samples 526 (F2) with a one-copy loss and sample 2721 (F272) with a four-copy gain are shown in Figure 8A-B. In the normal situation, the whole TRI region size is ~32 kb and gains would only make this region larger, as each gain would add ~10 kb. The one-copy losses would settle the region around 20 kb. The repetitiveness and homology of the TRI region have proven to be very challenging especially for PCR studies. The last introns of each TRI repeat (introns 89, 97, and 105) contain repetitive elements, such as *Alu* and LINE repeats. These transposable elements are known to be capable of being involved in NAHR (non-allelic homologous recombination) that is

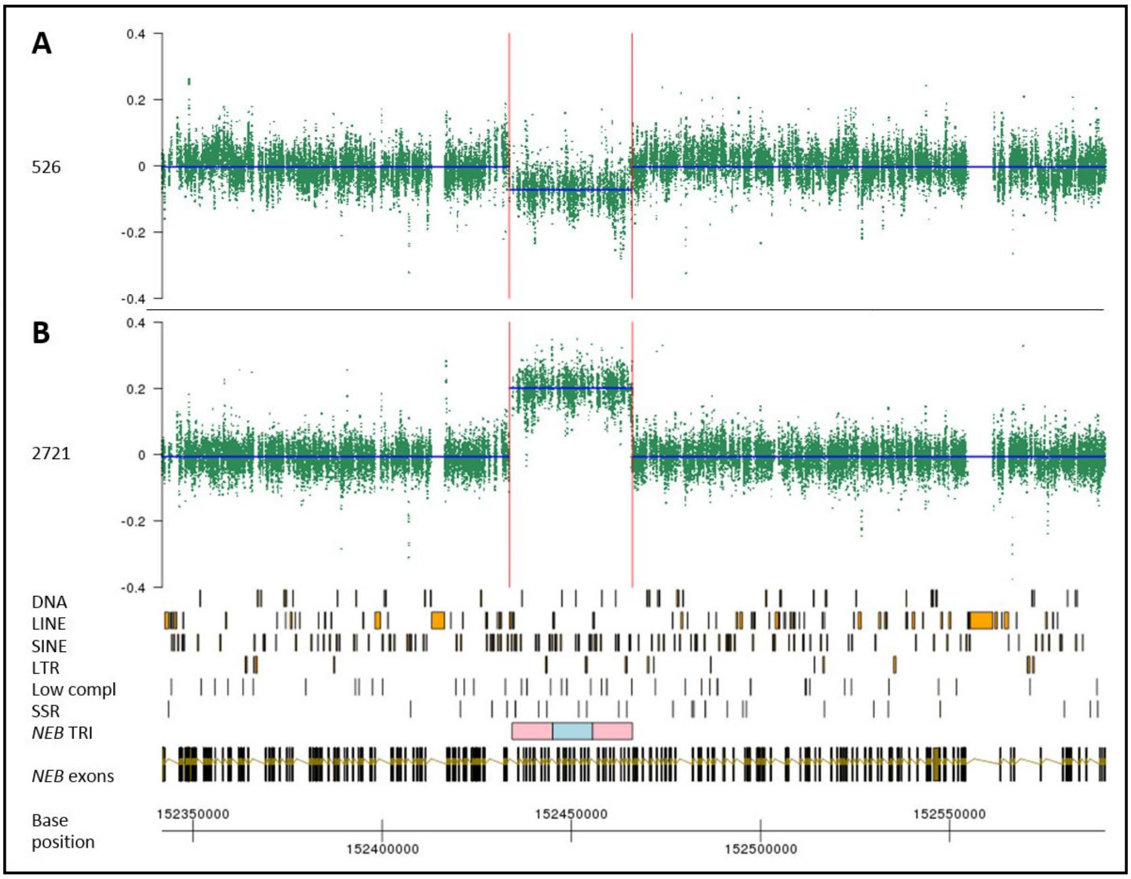

Figure 8A-B. Breakpoint analysis of *NEB* TRI copy number loss and gain. Sample 526 from family 2 shows a one-copy loss (A) and sample 2721 from family 272 a four-copy gain (B) of the *NEB* TRI. Breakpoint analysis shows repeat elements in the breakpoint regions. In the upper part of the figure, the inferred breakpoints are shown with red vertical lines and the inferred copy number of the corresponding fragment is indicated with blue horizontal lines. The tracks in the lower part of the figure show the different repeat elements, the *NEB* triplicate region and the gene structure. Abbreviations: DNA=DNA repeat elements, LTR=long terminal repeats, Low Compl=Low complexity DNA sequences, SSR=Simple sequence repeats=microsatellite DNA. Figure 8A-B is reprinted and modified with the permission of the copyright owner: Kiiski et al, 2015, Eur J Hum Genet. doi:10.1038/ejhg.2015.166. Licensee the Nature Publishing Group.

thought to be the most common underlying mechanism for recurrent CNVs (Gu et al., 2008; Kolomietz et al., 2002). Several studies have suggested that *Alu* repeats could also mediate chromosomal rearrangements via non-homologous mechanisms such NHEJ, FosTeS, or MMBIR (Shaw and Lupski, 2005; Vissers et al., 2009). The lack of further information on these TRI variations makes it difficult to estimate the exact causative mechanism. Viewing it from an evolutionary aspect, the human *NEB* TRI region is thought to have emerged from two duplication events (Bjorklund et al., 2010). For example, the mouse *Neb* includes only one copy of this eight-exon set and is lacking the LINE-L2 elements. Precisely these different repeat elements might explain the susceptibility of the human *NEB* TRI region to recurrent copy number changes, making this the most common variant characterized in *NEB* to date.

6.1.3. A novel pathogenic *TPM3* variant (IV)

In addition to *NEB* variants, one variation was identified in another NM gene, using the NM-CGH array. This is a pathogenic homozygous deletion affecting *TPM3*, identified in one NM family (F366).

Family 366 (sample 3663) had one previously identified pathogenic variant in *NEB* exon 42 (c.5060G>A; p.W1687X). The index case had the severe form of NM and deceased at the age of 17.5 months. Muscle biopsy at five months showed myopathic features with abnormal variation in fibre size. The Gömöri trichrome stain identified red-staining inclusions in several fibres, which electron microscopy confirmed as nemaline rods. The NM-CGH array showed a 17-20 kb deletion region in *TPM3*. The deletion region was first interpreted as two deletions separated by a normal region in between, but using the specific settings created for GC rich probes in the Cytosure analysis software, the results indicated that this was indeed one single deletion, according to HGVS nomenclature (Human Genome Variation Society) a homozygous deletion of hg19 chr1:g.(154,156,325_154,156,028)_(154,173,059_154,177,712). This homozygous deletion removes the promoter and the exons 1a and 2b of *TPM3* from both alleles (Figure 9). However, based on the NM-CGH array and Sanger sequencing, *TPM3* exon 1b and its promoter are present, indicating that the non-muscle isoforms are expressed. This is most likely the case as non-muscle isoforms of *TPM3* are essential for embryonic development (Hook et al., 2004).

The deletion also covers a micro-RNA encoding gene, *MIR190B* and the last two exons of the *C1orf189* gene. The role of miR-190b and the uncharacterized protein C1orf189 in skeletal muscle, if any, is unknown. The deletion starts upstream of the *TPM3* gene and the breakpoint resides in a 4.7 kb region that is lacking probes, because this region contains various *Alu* sequence repeats hindering the designing of unique probes. The identification of the exact breakpoints by PCR and sequencing has not been successful so far which means that the mechanism behind this mutation is difficult to determine. Further experiments are ongoing to resolve this.

This sample has also been studied with whole-exome sequencing, which failed to detect any new causative variants in the known NM genes, but verified the NM-CGH results and the previously identified nonsense variant in *NEB* exon 42. This reinforces the interpretation that the homozygous *TPM3* variation is most likely the causative mutation in this patient. Because both the promoter and exon 1a are deleted from both alleles, we presume that the muscle-specific *TPM3* isoform could not be produced.

One patient with severe NM has previously been characterized with a homozygous nonsense mutation in the muscle-specific exon 1a of *TPM3*. The authors hypothesized that either no *TPM3* peptide would be produced from the mutant allele due to instability of the mutant mRNA, or, if a truncated peptide was produced, it could act as a dominant negative protein, preventing proper formation of the tropomyosin polymer

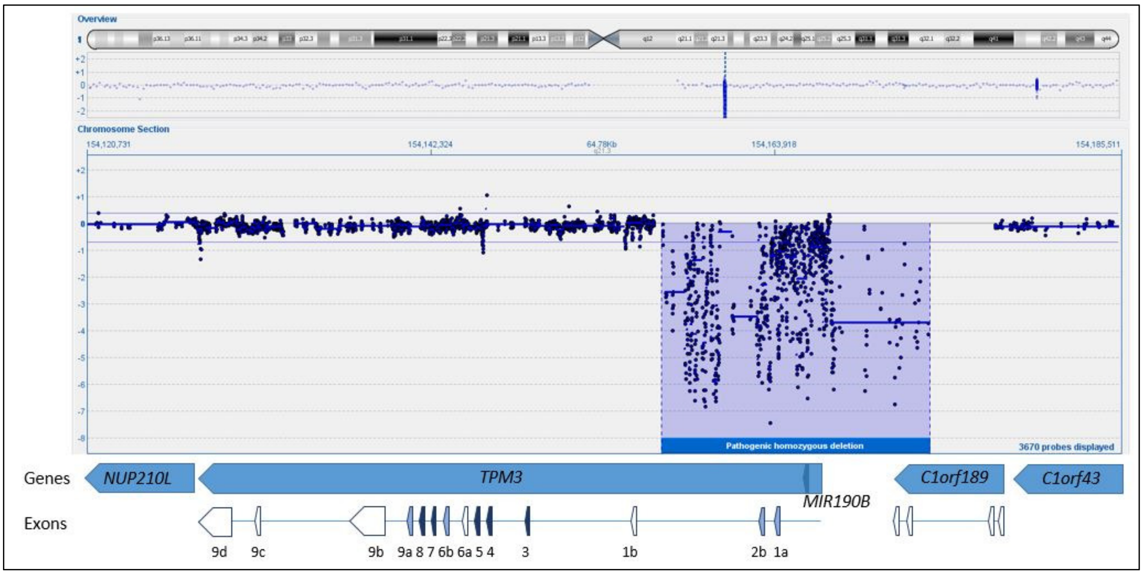

Figure 9. NM-CGH array profile of the *TPM3* gene. NM patient 3663 shows a 17-20 kb homozygous deletion in chromosome region 1q21.3 covering the promoter and the exons 1a and 2b of *TPM3*, *MIR190B* as well as the last two exons of *C1orf189*. Note the different Y axis scale compared with the previous figures due to the homozygous deletion. The genes, exons and their orientation are indicated in the lower part of the figure. The exons expressed in all *TPM3* isoforms are marked with dark blue, exons not expressed in striated muscle are marked with white and alternatively spliced exons present in the striated muscle isoform are marked with pale blue colour. Figure 9 is reprinted and modified with the permission of the copyright owner: Kiiski et al, 2015, J Neuromuscul Dis. 2015 doi: 10.3233/JND-150107. Licensee the IOS Press Copyright.

chain (Tan et al., 1999). Even though the mutational mechanism in this case is different, we hypothesize that the outcome is the same, i.e. a null state for the muscle-specific isoform. Parental samples were not available for analysis but the existing information on the parents suggested consanguinity. This was further shown using whole-genome SNP array, in which the DNA sample of the index patient indicated loss of heterozygosity (LOH) for many regions of the genome including *TPM3* in chromosome region 1q21.3, but not for *NEB* in chromosome region 2q23.3. This explains why the deletion in *TPM3* is homozygous but the variant in *NEB* exon 42 is heterozygous.

In conclusion, as this *TPM3* deletion is the only large CNV identified in a gene other than *NEB*, large copy number variations in other known NM genes are likely a very rare cause of NM. In our data cohort this kind of aberration was encountered in only one family out of 196 (0.5%). However, the possibility of such large aberrations still needs to be taken into consideration.

6.1.4. Features of the NM-CGH array (I-IV)

One of the advantages of a custom CGH microarray is flexibility. This method allows choosing any genes of interest to be analysed in a single experiment. Regarding the NM-CGH array this allows the simultaneous analysis of all NM-causing genes. In addition, the gene content of the microarray is available for modifications. This is useful, for example, when new genes are identified. In the first modification, three novel genes (*KLHL40, KLHL41* and *LMOD3*), and in the second modification, a novel putative NM gene (*YBX3*) were added to the NM-CGH microarray platform. However, positive controls are often unavailable for the testing of these novel genes due to lack of previously identified CNVs in them. The verification is then based on analysing the performance of the probes in general in these newly added genes. The field of NM research is undergoing fast development and the NM-CGH array will be updated when new NM genes are identified. Consequently, samples that have been analysed with a previous array version, need to be subsequently analysed for the novel genes or analysed again with an updated array version. In addition to the flexibility of the array content, the NM-CGH method is fast and cost-efficient, and it requires a rather low quantity of DNA (500-1000 ng) per sample.

As the probes used on this array are 60mer long oligonucleotides, this method is not able to detect very small aberrations. The theoretical functional resolution of the NM-CGH array allows the detection of aberrations as small as 100-150 bp in the regions with optimal tiling resolution, one probe starting every 10 bp and 5-10 consecutive probes used as an aberration threshold. However, the smallest mutation detected by the NM-CGH array so far is even smaller, a deletion of 72 bp inside *NEB* exon 15, which was verified with another method. This mutation is definitely at the detection threshold of the method, and it was identified comparing this aberration with the in-house aberration database, the data for which were gathered during this project. However, we classify this as a CNV, even though the size is smaller than typically defined as the minimum size of a CNV, ~1 kb (Redon et al., 2006). This shows that in an optimal situation, the NM-CGH custom tiling design allows for the identification of very small copy number variations, much smaller compared with conventional whole-genome arrays.

Despite the tiling approach and the very high resolution of the NM-CGH microarray, the method does not always reveal the breakpoints of the identified aberrations very precisely, as they may reside in regions of repetitive sequence or outside the targeted genes. Properly functioning unique probes cannot be designed targeting repetitive sequences and therefore some gaps without any probes remain in the design. Different sequencing methods can be used to further map the exact breakpoints of these particular aberrations.

The breakpoints fairly often reside in these regions that are lacking probes, i.e. repetitive sequences. It is known that repetitive sequences can predispose for rearrangements and copy number aberrations. We hypothesize that this might be the case for some of the detected pathogenic variants and *NEB* TRI variations. Further studies are ongoing to elucidate this.

6.2. Whole-exome sequencing (II-IV, U)

Whole-exome sequencing (WES) is the most recent addition to the variant analysis methods used in this study. This part of the project is currently in its early stages. However, samples from ten NM patients who had remained without identification of one or both pathogenic variants, have been studied. Thus far, we have identified novel pathogenic variants in known NM genes for six of the ten studied families using WES. The identified mutations are summarized in Table 9. Furthermore, we have identified two novel variants in a putative NM gene, *YBX3*, in one NM family. These variants were shown to segregate with the disorder and the potential pathogenicity of the variants is currently being investigated further. In addition, four samples with two identified variants were tested. This included three samples with putative pathogenic *NEB* TRI copy Novel pathogenic variants in the NM genes (II-III) number gains (section 6.1.3) and sample 3663 with the novel *TPM3* deletion (6.1.5). Using WES, no additional pathogenic variants were identified in these samples.

Family 2 (sample 523) includes two siblings with typical NM. One disease-causing variant had been identified previously, a missense mutation in *NEB* intron 122 (c.19097G>T). This was verified with Sanger sequencing and it had been inherited from the mother. We studied the sample of one affected family member using WES and identified a splice-site variant (c.508-7T>A) in *NEB* intron 7. This creates a novel acceptor splice site inserting five nucleotides, leading to a frameshift p.Val170fs. The variant was verified using Sanger sequencing which showed it to be inherited from the father. The effect of the variant on pre-mRNA splicing was verified using a *NEB* minigene construct. (Study III)

Family 4 (sample 545) includes a sibling pair with typical NM. Pathogenic variants had not been identified previously, even though the whole *NEB* gene had been screened using dHPLC. We studied one affected family member with WES and identified a splice-site mutation in *NEB* intron 65 (c.9414+1G>T) and a missense mutation in *NEB* exon 61 (c.8381A>T; p.Tyr2794Phe). These were verified with Sanger sequencing and the splice-site mutation had been inherited from the father and the missense mutation from the mother. The missense mutation disrupts an actin-binding site. (Study II)

Table 9. Variants in the NM families studied by whole-exome sequencing. The previously identified variants were detected using dHPLC, Sanger sequencing or the NM-CGH array. After these analyses, whole-exome sequencing (WES) revealed yet novel variants for 6/10 of the studied NM families. The allele frequencies in the ExAC database are reported. Reference sequences *NEB:* NM_001271208.1, *ACTA1:* CR536516.1.

Family	Patient	Variants identified using NM-CGH array, SSCP/dHPLC+Sanger sequencing	Detected using WES	Allele frequency in ExAC	Variants identified using whole-exome sequencing	Allele frequency in ExAC	Study
F2	523	*NEB* exon 122 c.19097G>T (M) [A]	yes	0.00008935			III
		NEB TRI loss (5/6 copies) (F)	no	NA	*NEB* intron7, c.508-7T>A (F)	NA	
F4	545	-			*NEB* intron 65, c.9414+1G>T (M)	NA	II
					NEB exon 61, c.8381A>T (F)	0.00001665	
F6	564				*NEB* intron 35, c.3879+1G>A (M)	0.00002583	II
		NEB exon 156, c.22746delG (F) [A]	yes	NA			
F16	5163	-			*ACTA1* intron 6, c.990+1G>T (M)	0.000008244	U
					ACTA1 intron 6, c.990+1G>T (F)	0.000008244	
F407	4073	*NEB* TRI gain (7/6 copies) (M)	no	NA	*NEB* exon 173, c.24475_24479dup (M)	NA	III
		NEB TRI gain (7/6 copies) (F)	no	NA	*NEB* exon 173, c.24475_24479dup (F)	NA	
F411	4113	*NEB* TRI loss (5/6 copies) (M), *NEB* exon 86; c.13134delA (M)	no no	NA			III
					NEB exon 172, c.24372_24375del (F)	0.00002915	

Abbreviations: A=published in Lehtokari et al. 2006, ExAC database=Exome Aggregation Consortium Browser Database, (F)=inherited from the father, (M)=inherited from the mother, NA=not available, NEB TRI=triplicate region of the nebulin gene.

Family 6 (Sample 564) includes a pair of siblings affected with typical NM. Previously, a delG causing a frameshift mutation in *NEB* ex 156 (c.22746del; p.Met7582Ilefs*5) had been identified by SSCP and sequencing. This had been inherited from the healthy father. We identified the second pathogenic variant with WES. This was an essential splice-site mutation in *NEB* intron 35 (c.3879+1G>A) inherited from the mother. This variant was verified with Sanger sequencing. (Study II)

Family 16 (sample 5163) had a sibling pair affected with severe NM and no previously identified disease-causing variants. We studied the sample of one affected child and identified a homozygous essential splice-site mutation in *ACTA1*, at the donor splice site of intron 6, c.990+1G>T (GenBank CR536516.1). This is a pathogenic variant interpreted to have caused the nemaline myopathy in this patient. (Study U)

Family 407 (sample 4073) had two siblings affected with severe NM. There were no previous pathogenic variants except a gain of *NEB* TRI (8/6 copies) that was inherited from both parents, who had one gained TRI copy each (7/6 copies). We studied the sample of one affected child and identified a homozygous CACAA insertion in *NEB* exon 173 (c.24475_24479dup). This was verified by Sanger sequencing. (Study III)

Family 411 (sample 4113) includes a child with typical NM. No previous disease-causing variants had been identified, except a loss of *NEB* TRI (5/6 copies), detected by the NM-CGH array, that had been inherited from the mother (5/6 copies). Using WES, we identified a deletion of four nucleotides in *NEB* exon 172, delAAGA, (c.24372_24375del) that was shown with Sanger sequencing to have been inherited from the father. After WES, the *NEB* TRI exons were further studied using Sanger sequencing, because this region is not covered in WES. A frameshift mutation in *NEB* exon 86 was identified (c.13134delA). (Study III)

Families 2, 4, 6, and 16 had been previously studied for *NEB* and *ACTA1* variants, but these splice-site and missense mutations had remained unidentified. This is mostly due to the previously used analysis methods. For example, families 4, 6, and 16 had been screened for *NEB* variants with SSCP and dHPLC but these did not indicate any change in the sequences (this data was checked again after the WES finding). Therefore these regions had not been Sanger sequenced. The formation of the novel splice site in intron 7 of *NEB* had been missed in the analysis of family 2 when Sanger sequenced. This emphasizes that even though a family would have been studied for NM gene variants previously, this does not exclude the possibility of finding mutations in these genes with more efficient methods.

6.2.1. Novel putative disease-causing variants outside the known NM genes (U)

In addition to these characterized pathogenic variants in the known NM genes, two variants were identified in one family (F361) in a novel putative NM gene, *YBX3*, encoding Cold-Shock Domain Protein A. The patient whose NM has been assigned to the category 'other forms of NM' was shown to carry compound heterozygous missense variants in conserved sites of this gene; p.Ser34Arg, of which no allele frequency was available in ExAC (poor coverage in exome data) and p.Arg129Trp with an allele frequency of 0.00005766 in the ExAC database (protein reference sequence AAH15913.1). *YBX3* is highly expressed in skeletal muscle (Kudo et al., 1995). *YBX3* variants have not been previously associated with NM but these variants may turn out to be the disease-causing mutations in this family. Further investigations are ongoing to resolve this.

This, in particular, shows the potential of whole-exome sequencing. All the known NM genes can be studied at once, and with better detection rate than before. Furthermore, the method allows all other genes of the genome to be studied as well. This is important because so many NM families have been extensively studied with a number of different methods and still one or both disease-causing variants may have remained unidentified. This implies that there still are additional NM genes to be discovered. In this perspective, WES is the logical and most potential method of choice, as it does not require previous knowledge of any potential novel genes. However, the interpretation of WES results requires experience to be able to infer whether a finding is a benign polymorphism or pathogenic variation.

With next-generation sequencing techniques becoming more common in the diagnostic field, it is worth noting that no technique alone reveals all variants, especially of the gigantic nebulin gene. There are regions in *NEB* that are not covered in adequate reading depths, and regions of great homology and repetitiveness that cause difficulties in sequencing. For example, most of the intronic variants and some copy number variations would be missed using WES alone. Specifically, the *NEB* triplicate region appears to be missing from the exome-sequencing kits, because of the great homology of the region. However, small variants in the TRI region have been identified using Sanger sequencing of single TRI exons, and copy number variants of the entire *NEB* triplicate region have been identified with the NM-CGH microarray. This is why the NM-CGH microarray and exome sequencing methods complement each other well in NM diagnostics and research.

7. *NEB* variants causing nemaline myopathy

7.1. NEB mutation update (I-III)

To date, both pathogenic *NEB* variants have been identified for 165 different NM families, including 220 different mutations. The number of NM families included in each study and their status regarding the number of identified pathogenic *NEB* variants at the time of publication of each study is presented in Table 10. In addition to these *NEB* variants, disease-causing variants were identified in other NM genes in 15 families. In one family, pathogenic variants were identified in *TPM3*, in five in *ACTA1*, in three in *LMOD3*, while six families showed pathogenic variants in *KLHL40*.

Counting each different 220 *NEB* variant only once, the most common types of pathogenic variants are splice-site mutations (33%) and small deletions or insertions (<20 bp) (33%), mostly causing a frameshift, but also one in-frame mutation. The third most common are nonsense mutations (22%). Missense mutations account for 7% of the disease-causing variants. Only those missense mutations affecting nebulin-actin or nebulin-tropomyosin interactions were interpreted as pathogenic. Various types of copy number variations, i.e. large deletions or duplications, constitute the rarest group (5%) (Figure 10A); however, the occurrence of copy number variations was found to be much more common than previously estimated. Viewing the data on the 165 NM families in which two pathogenic *NEB* variants have been identified, no less than 16% of the families (26/165) had a pathogenic CNV. This shows that even though the different CNVs are rather rare per se, some of them are recurrent, such as the exon 55 deletion and *NEB* triplicate region variations, making pathogenic CNVs a rather frequent finding among NM families.

The great majority (84%) of families (138/165) are compound heterozygous for two different mutations. Furthermore, 41% of families (67/165) have their own two unique mutations that are not shared with any other family. However, there are four different

Table 10. The number of nemaline myopathy families with both, one and no identified pathogenic *NEB* variants in each study

Study	NM Families	2x*NEB*	1x*NEB*	0x*NEB*
study I	36	2	20	14
study II	309	159	58	92
study III	196	36	50	110
study IV	1	1 homozygous deletion in *TPM3*		
Tot	**356**			

variants in our data set which may be regarded as founder mutations. The most common one is the Ashkenazi Jewish founder mutation deleting the 2.5 kb region including exon 55 (Anderson et al., 2004; Lehtokari et al., 2006; Lehtokari et al., 2009) that was identified in 13 families in this study. Moreover, there are three different variants identified in Finnish families that may be regarded as founder mutations. These are a missense mutation in exon 122 (p.Ser6366Ile), another missense mutation in exon 151 (p.Thr7417Pro), and a frameshift mutation in exon 122 (p.Thr6350Profs*4) identified in eleven, five, and two families, respectively. In addition to the founder mutations, there were 11 different pathogenic variants shared between three or more families, and 26 different mutations shared between two families. These families share one mutation, but carry a separate second mutation. The haplotyping results show that the Finnish

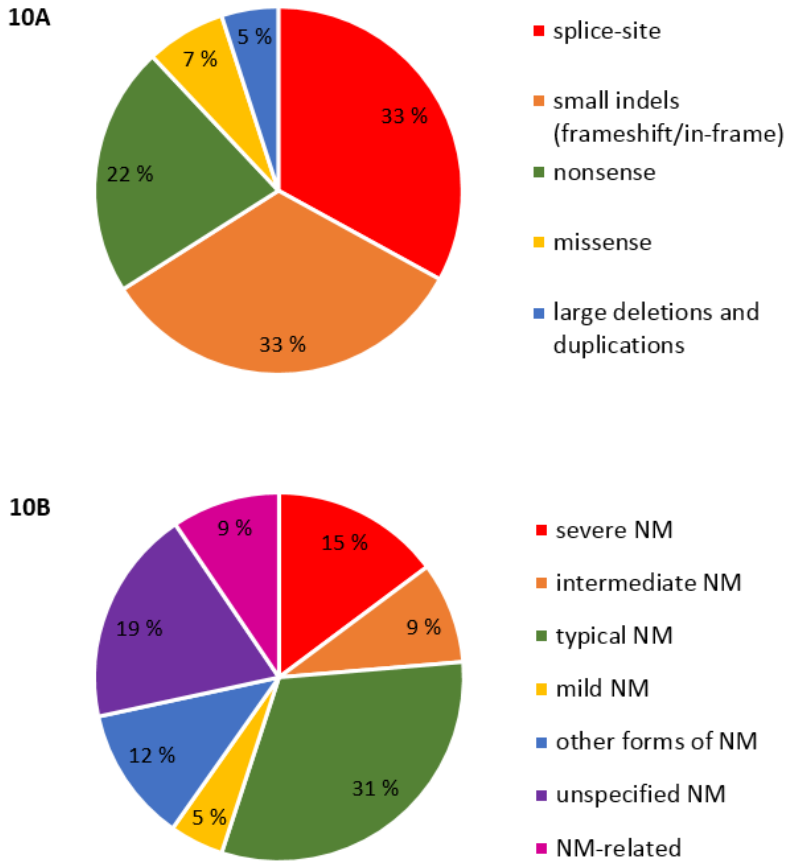

Figure 10. A) The Distribution of Different Types of Pathogenic *NEB* Variants. Counting each different pathogenic *NEB* variant once, the most common types of variants are splice-site and small indel mutations followed by nonsense and missense mutations. Large copy number variations are the rarest. **B) The Clinical Entities Caused by *NEB* Mutations.** *NEB* variants most commonly cause typical NM, then severe, intermediate and most rarely mild NM. In addition it causes other forms of NM, and NM-related disorders. For a fifth of the patients clinical data was not available to determine the form of NM. Abbreviations: indel=insertion/deletion.

families sharing a certain pathogenic *NEB* variant also have the same haplotype. This is also the case regarding the *NEB* exon 55 deletion. Some of the remaining shared pathogenic variants in *NEB* seem to be recurrent, whereas others have a common ancestral haplotype. Only one location in *NEB* is considered to be a mutational hotspot, i.e. the donor splice site of intron 32, including nucleotides c.3255+1 and c.3255+2, that were mutated in altogether 10% of families in this study cohort (17/165).

7.2. Genotype-phenotype correlations (I-III)

Among the families with two identified *NEB* mutations, sufficient clinical data has been provided for 149 families to be able to determine the NM form of the patients (including 143 families from studies I and II, and 6 families from study III): 25 families included patients with severe NM, 15 with intermediate, 53 with typical, 8 with mild, and 20 with "other forms" of NM (in four families there were siblings that had either the typical or the intermediate form of NM) (Figure 10B).

Finding genotype-phenotype correlations in NM patients with *NEB* mutations proved to be more difficult than originally thought. Although the number of *NEB* mutations is now substantial, a statistically significant genotype-phenotype correlation between the type of mutation and the NM subtypes was unobtainable. First of all, statistical data analysis requires that the NM families are grouped into different categories. As most families have two private mutations, there is no simple and correct way to do this. The clinical heterogeneity of NM is underlined by the fact that even affected family members sharing the same mutations may present different forms of the disease.

Some interesting phenotypic findings were made: for example, some of the families that had mutations in the alternatively spliced exons of *NEB* (exons 63-66, 143-144, and 167-177) had exceptional clinical/histological features and they were thus all assigned to the group "other forms of NM". Three out of six families (Families F270, F284 and F386) with pathogenic variants in exons 63-66 included patients with pronounced weakness of axial muscles, with relative sparing of limb muscles. Usually the axial, and the proximal limb muscles, and later the distal limb muscles are involved in patients with NM. Regarding the alternatively spliced exons 143 and 144, one adult sib pair (Family 187) had a pathogenic variant in exon 143, and they both had ophthalmoplegia, which is very unusual in NM. Family 390 had a truncating mutation in exon 144 and showed atypical dystrophic biopsy features and normal strength of the neck flexors, which are almost always weak in NM. Some of the patients with mutations of the alternatively spliced exons 167-177 had fasciculations, selective axial weakness, rigid neck and/or spine, or other atypical features of NM. We hypothesize that these atypical findings may reflect some tissue- or muscle-specific functions of the alternatively spliced exons that are currently unknown. However, more data is needed to draw conclusions on this matter.

7.3. Mutation frequency (II)

Exome Variant Server (http://evs.gs.washington.edu/EVS/) data was searched for information about *NEB* variants in the general population. This databank included exome-sequencing data of 6503 DNA samples from individuals of African American and European American origin with no known neuromuscular disorders. The data contained 1295 variants in the *NEB* coding region and conserved splice sites (February 2014). These included benign missense variants, intronic variants, synonymous changes etc. However, 69 individuals of those approximately 6000 individuals whose samples were successfully exome sequenced, were heterozygous carriers of a pathogenic *NEB* variant. The carrier frequency was thus 1/87, yielding a disease incidence of approximately 1 in 30,000 for autosomal recessive myopathies caused by pathogenic *NEB* variants. The large ExAC database was not publicly available at the time of this study. Currently, individual variations of the nebulin gene can be searched for in the ExAC database. However, due to the large size of the gene, the database does not enable studying the entire *NEB* gene all at once.

Nemaline myopathy caused by *NEB* variants seems unexpectedly uncommon compared with the estimated mutation rate of the genome, and the size of the *NEB* gene (249 kb of genomic sequence). Moreover, it has been shown that there are other important factors affecting the mutability of a certain gene, such as the local sequence context (Samocha et al., 2014). This may partly explain the observation that *NEB* tolerates mutations well. Moreover, patients with severe NM who die at birth or soon after, may be left undiagnosed. Additionally, pathogenic *NEB* variants can also cause other NM-related disorders, such as distal myopathies. These disorders may be milder and thus underdiagnosed. Overall, most of the missense mutations are not, in fact, pathogenic variants. We propose that only if the missense mutation is located in a conserved binding site, such as a nebulin-actin or nebulin-tropomyosin binding site, it should be regarded as a pathogenic variant (Lehtokari et al., 2014; Marttila et al., 2014). This could explain why mutations in such a large gene result in such rare disorders.

CONCLUSIONS AND FUTURE PROSPECTS

In this Doctoral Thesis study, a novel variant detection method was developed for genes causative of nemaline myopathy. Currently, the molecular genetic verification of a clinical and histological diagnosis of NM is cumbersome due to the lack of one specific method for this. Using a combination of different methods has turned out to be a necessity for identifying the large variety of pathogenic variants in the NM genes. The importance of studying *NEB* variants is underlined by the fact that pathogenic *NEB* variants are by far the most common cause of recessively inherited NM, probably also the most common cause of NM in general. Furthermore NM is one of the most common congenital myopathies.

In this study, a new custom targeted tiling NM-CGH microarray was designed which revealed copy number variations in the NM-causing genes in the study cohort of 196 NM families. In ten different families, nine novel disease-causing variations were identified in *NEB* in ten different families. Furthermore, a recurrent copy number change of the *NEB* triplicate region was found in 13% of the NM families (26/196) and the *NEB* triplicate region CNV was interpreted to be possibly pathogenic in 4% of the NM families (8/196). In addition, a novel pathogenic homozygous deletion was found in another NM gene, *TPM3*, in one family. All in all, the NM-CGH array is quick and cost-effective, and it has revealed large novel disease-causing or putative pathogenic copy number variations in altogether 10% of the studied NM families (19/196). New samples that are sent to our group for mutation analysis are currently first tested with the NM-CGH array, and the downstream methods are chosen depending on the array findings.

Further studies are ongoing to decipher the exact breakpoints of each copy number variation, especially for the *NEB* TRI region. For this, testing has commenced on a fourth-generation sequencing technique: Nanopore sequencing of genomic DNA, using the MinION platform that does not require PCR. In addition, functional analyses are needed in order to determine the consequences of the changes at the protein level. This will help to resolve the molecular mechanisms as well as the potential pathogenicity of some of the recurrent *NEB* TRI variations.

Whole-exome sequencing (WES) has made its way to diagnostics and it has proven to be efficient in identifying novel disease-causing variants. It allows for studying the known NM genes as well as most of the other genes of the exome in a single experiment. Novel pathogenic variants in the known NM genes were identified in approximately half of the studied families (6/10). In addition, one family was identified with two variants in a novel, putative NM gene, *YBX3*. These variants were shown to segregate with the disease in the family and functional studies are ongoing.

Despite the indisputable advantages of WES, even this method is unable to detect all types of variants. For example, *NEB* TRI and other regions of great homology are not covered by exome sequencing kits because of lack of unique DNA sequence in those regions. Thus, the NM-CGH microarray and WES methods complement each other well in NM diagnostics.

A total of 165 patient families have now been identified with two pathogenic *NEB* variants. The majority (84%) of the families are compound heterozygous for two different *NEB* pathogenic variants. When counting each different mutation only once, the most common type of mutations in this cohort are splice-site mutations (33%) and small indels (33%), followed by nonsense (22%) and missense mutations (7%). Various large mutations are rare (5%), however, more common than previously thought. Moreover, among these 165 NM families with two pathogenic *NEB* variants, in a total of 16% (26/165), one of the mutations identified was a pathogenic copy number variation.

It has become evident that it is extremely difficult to discern any detailed genotype-phenotype correlations in *NEB*-caused NM, even considering the now large number of patients with clarified genotypes. Patients sharing the same pathogenic variants may have different disease severities, even within the same family. Thus, the identified variants may not be the only factors determining the severity of the disease. Modifying genetic factors might play a role in the severity of the phenotype. These are currently being revealed by the use of new techniques such as exome and whole-genome sequencing. Furthermore, epigenetics and environmental factors play a role in modifying our genome, making genotype-phenotype correlations more complex.

Finding the causative mutations constitutes the foundation for the study of pathogenetic mechanisms, which, in turn, is a prerequisite for developing specific therapies. Thus the hunt for the unidentified disease-causing variants in NM families will continue as new samples from NM families around the world are being sent to us for study. It is expected that novel pathogenic variants in the currently known NM genes will continue to be found with these new mutation detection methods. At the same time we are on the lookout for the next novel NM-causing gene.

ACKNOWLEDGEMENTS

This Doctoral thesis work was carried out at the Folkhälsan Institute of Genetics, the Department of Medical and Clinical Genetics, University of Helsinki during the years 2011-2015. I thank Professor Anna-Elina Lehesjoki as well as the former Heads of the Institute for providing excellent resources and facilities for scientific research.

I thank the Integrative Life Sciences Doctoral Programme and the following foundations and organisations for funding this Doctoral thesis project: the Folkhälsan Research Foundation, the Finska Läkaresällskapet, the Sigrid Jusélius Foundation, the Liv och Hälsa Foundation, the Academy of Finland and the Association Française Contre les Myopathies. In addition the Helsinki University Chancellor's Travel Grant, the Registered Association for Geneticist in Healthcare (TG), the World Muscle Society, European Society of Human Genetics, the Oskar Öflund Foundation, Magnus Ehrnrooth Foundation and the Finnish Concordia Fund, are thanked for financially supporting my participation in international congresses.

I am grateful to my thesis supervisors and our group leaders Carina Wallgren-Pettersson (PI) and Katarina Pelin (co-PI) for their enthusiasm and knowledge in the research field of nemaline myopathy and other neuromuscular disorders. I thank Katarina for all the practical hints with the methods and the endless enthusiasm whenever setting up something new in the lab. I thank Carina for guidance into the clinical aspects of the research and reminding why this research is done in the end – the patients and families. I am very grateful to both for giving me the opportunity to participate in this research and most of all for believing in me throughout this project.

All the co-authors of the original publications are warmly thanked for their contribution to this project; performing laboratory experiments, providing patient samples and clinical data as well as their participation in writing of the manuscripts. I owe my gratitude to all our collaborators throughout the world.

I thank my Thesis Advisory Committee members Nina Horelli-Kuitunen and Maija Wessman for your valuable instructions and supporting outlook.

I thank warmly all the current and former members of the NEM Group (Research Group for Nemaline Myopathy and Other Neuromuscular disorders): Vilma-Lotta Lehtokari for sharing all her knowledge, help in mutation analyses, reviewing this thesis and all the support; Jenni Laitila and Minttu Marttila, my fellow PhD students, for sharing their knowledge, ideas and this entire experience; Marilotta Turunen for excellent technical assistance with the laboratory work and for all the practical things; Liina Laari for her valuable contribution especially during the set-up and validation of the NM-CGH array; Sampo Koivunen for his enthusiasm in our new Nanopore sequencing project; Pauliina Repo for sharing her molecular genetic skills; Kati Donner, Mikaela Grönholm, Mubashir Hanif, and all the other former members of the NEM group for leading us the way and making it easy for us to join the group. Thank you all for the friendship and memorable times and trips that we have had together – everything from Korkeasaari to Perth Australia. We've had so much fun together (let's continue the same way)!

Many thanks to all my fellow PhD students at FIG! Thanks to my current roommates Anni L., Inken, Salla and Otto for a supporting atmosphere in the office. Special thanks to Anni L. and Anni E. for sharing experiences regarding WES, Maria S. for MLPA troubleshooting and Sara for helping me to get started with the arrays at FIG. Thanks to Peter for inspiring attitude in muscle research and Jaakko, Anna, PH, Saara and Kaisa for sharing their expertise, Merja for help with practical things and to Ann-Liz, Teija and Markku for lab management and keeping us in order.

I thank Stephan Keskinen, Madeleine Avellan, Marjatta Valkama, Jaana Welin-Haapamäki and Nina Forss for running the practical and economical things. Thanks to all the FIG members for creating the Folkhälsan atmosphere and a great place to work!

I thank warmly my Clinical Laboratory Geneticist Training Supervisors Harriet von Koskull and Nina Horelli-Kuitunen. Your mentoring, inspiration and encouragement has been indispensable and I am truly grateful of it! A kind thank you to the laboratories that provided me the Clinical Laboratory Geneticist Training; United Medix Laboratories, and HUSLAB Laboratory of Genetics, my superiors Anna-Kaisa Anttonen and Kristiina Aittomäki, and the wonderful colleagues and personnel in both of these laboratories!

Thanks to all the wonderful friends who have supported me during this time. Special thanks to Noora and Eeva for your continuing support since the Master's thesis project. Many thanks to all my friends for reminding me from time to time that there is life also outside work and studying… and renovating!

I am deeply grateful to my family for all the love and support during this project. I thank my parents Eila and Leo for always believing in me whatever it is that I decide to put my mind into and my brother Mikko for always being there for me. I thank my extended family Tuire, Annukka and Tiia for encouragement and Veeti, Anni and Kaisa for your inspiring enthusiasm towards everything! I'm truly grateful to my 'sis' Kristiina who has always been there for me, during this project and all the other times.

I owe my deepest gratitude to my husband Antti, for his love and support, and sharing his life as my teammate for the past thirteen years. I could not have done this without you (and your endless patience with the figures and all my IT problems)!

This work is dedicated to my family and Malla who inspired me to join the research of neuromuscular disorders as well as all the families that have participated in the different research projects making it possible for us to study the genetics of nemaline myopathy.

This thesis project has been a fascinating journey that is now coming to an end. However, my journey exploring the field of medical and clinical genetics has just begun and I'm eager to learn how the story continues.

Helsinki, October 2015

Kirsi

REFERENCES

Electronic References

The Database of Genomic Variants (DGV):
http://dgv.tcag.ca/dgv/app/home

DECIPHER (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources):
https://decipher.sanger.ac.uk/

ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations):
http://ecaruca.net

Exome Aggregation Consortium (ExAC) Browser Database:
http://exac.broadinstitute.org/

The Exome Variant Server, NHLBI Exome Sequencing Project (ESP):
http://evs.gs.washington.edu/EVS/

Leiden Open Variation Database (LOVD), Leiden Muscular Dystrophy Pages, *NEB:*
http://www.dmd.nl/nmdb/home.php?select_db=NEB

National Center for Biotechnology Information (NCBI) SNP database:
http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi

Onlinde Mendelian Inheritance in Man database:
http://www.omim.org/

References

ACMG Board of Directors. (2015). ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. Genet. Med. *17,* 68-69.

Adams, J., Kelso, R., and Cooley, L. (2000). The kelch repeat superfamily of proteins: propellers of cell function. Trends Cell Biol. *10,* 17-24.

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7,* 248-249.

Agrawal, P.B., Greenleaf, R.S., Tomczak, K.K., Lehtokari, V.L., Wallgren-Pettersson, C., Wallefeld, W., Laing, N.G., Darras, B.T., Maciver, S.K., Dormitzer, P.R., and Beggs, A.H. (2007). Nemaline myopathy with minicores caused by mutation of the CFL2 gene encoding the skeletal muscle actin-binding protein, cofilin-2. Am. J. Hum. Genet. *80,* 162-167.

Albagli, O., Dhordain, P., Deweindt, C., Lecocq, G., and Leprince, D. (1995). The BTB/POZ domain: a new protein-protein interaction motif common to DNA- and actin-binding proteins. Cell Growth Differ. *6,* 1193-1198.

Alkan, C., Kidd, J.M., Marques-Bonet, T., Aksay, G., Antonacci, F., Hormozdiari, F., Kitzman, J.O., Baker, C., Malig, M., Mutlu, O.*, et al.* (2009). Personalized copy number and segmental duplication maps using next-generation sequencing. Nat. Genet. *41,* 1061-1067.

Anderson, S.L., Ekstein, J., Donnelly, M.C., Keefe, E.M., Toto, N.R., LeVoci, L.A., and Rubin, B.Y. (2004). Nemaline myopathy in the Ashkenazi Jewish population is caused by a deletion in the nebulin gene. Hum. Genet. *115,* 185-190.

Bang, M.L., Li, X., Littlefield, R., Bremner, S., Thor, A., Knowlton, K.U., Lieber, R.L., and Chen, J. (2006). Nebulin-deficient mice exhibit shorter thin filament lengths and reduced contractile function in skeletal muscle. J. Cell Biol. *173,* 905-916.

Bauters, M., Van Esch, H., Friez, M.J., Boespflug-Tanguy, O., Zenker, M., Vianna-Morgante, A.M., Rosenberg, C., Ignatius, J., Raynaud, M., Hollanders, K.*, et al.* (2008). Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. Genome Res. *18,* 847-858.

Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. Annu. Rev. Genomics Hum. Genet. *12,* 187-215.

Bennardo, N., Cheng, A., Huang, N., and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet. *4,* e1000110.

Bjorklund, A.K., Light, S., Sagit, R., and Elofsson, A. (2010). Nebulin: a study of protein repeat evolution. J. Mol. Biol. *402,* 38-51.

Bovolenta, M., Neri, M., Fini, S., Fabris, M., Trabanelli, C., Venturoli, A., Martoni, E., Bassi, E., Spitali, P., Brioschi, S.*, et al.* (2008). A novel custom high density-comparative genomic hybridization array detects common rearrangements as well as deep intronic mutations in dystrophinopathies. BMC Genomics *9,* 572.

Brooke, M.H., and Engel, W.K. (1969). The histographic analysis of human muscle biopsies with regard to fiber types. 4. Children's biopsies. Neurology *19,* 591-605.

Chandra, M., Mamidi, R., Ford, S., Hidalgo, C., Witt, C., Ottenheijm, C., Labeit, S., and Granzier, H. (2009). Nebulin alters cross-bridge cycling kinetics and increases thin filament activation: a novel mechanism for increasing tension and reducing tension cost. J. Biol. Chem. *284,* 30889-30896.

Chen, L., Zhou, W., Zhang, L., and Zhang, F. (2014). Genome architecture and its roles in human copy number variation. Genomics Inform. *12,* 136-144.

Chilamakuri, C.S., Lorenz, S., Madoui, M.A., Vodak, D., Sun, J., Hovig, E., Myklebost, O., and Meza-Zepeda, L.A. (2014). Performance comparison of four exome capture systems for deep sequencing. BMC Genomics *15,* 449.

Clarke, N.F., and North, K.N. (2003). Congenital fiber type disproportion--30 years on. J. Neuropathol. Exp. Neurol. *62,* 977-989.

Conen, P.E., Murphy, E.G., and Donohue, W.L. (1963). Light and Electron Microscopic Studies of "Myogranules" in a Child with Hypotonia and Muscle Weakness. Can. Med. Assoc. J. *89,* 983-986.

Conley, C.A., Fritz-Six, K.L., Almenar-Queralt, A., and Fowler, V.M. (2001). Leiomodins: larger members of the tropomodulin (Tmod) gene family. Genomics *73,* 127-139.

Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C., Turner, D.J., and Hurles, M.E. (2010). Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nat. Genet. *42,* 385-391.

Conrad, D.F., Keebler, J.E., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V.*, et al.* (2011). Variation in genome-wide mutation rates within and between human families. Nat. Genet. *43,* 712-714.

Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P.*, et al.* (2010). Origins and functional impact of copy number variation in the human genome. Nature *464,* 704-712.

Craig, R.W., and Padron, R. (2004). Molecular Structure of the Sarcomere. In Myology, 3rd edition, Engel, AC, and Franzini-Armstrong, C., McGraw-Hill, pp. 129-66.

Desai, A.N., and Jere, A. (2012). Next-generation sequencing: ready for the clinics? Clin. Genet. *81,* 503-510.

Donner, K., Ollikainen, M., Ridanpaa, M., Christen, H.J., Goebel, H.H., de Visser, M., Pelin, K., and Wallgren-Pettersson, C. (2002). Mutations in the beta-tropomyosin (TPM2) gene--a rare cause of nemaline myopathy. Neuromuscul. Disord. *12,* 151-158.

Donner, K., Sandbacka, M., Lehtokari, V.L., Wallgren-Pettersson, C., and Pelin, K. (2004). Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts. Eur. J. Hum. Genet. *12,* 744-751.

Dubowitz, V., Sewry, C., and Oldfors, A. (2013). Muscle Biopsy - A practical approach, 4th edition, Elsevier Health Sciences.

Durinck, S., Bullard, J., Spellman, P.T., and Dudoit, S. (2009). GenomeGraphs: integrated genomic data visualization with R. BMC Bioinformatics *10,* 2-2105-10-2.

Feng, J.J., and Marston, S. (2009). Genotype-phenotype correlations in ACTA1 mutations that cause congenital myopathies. Neuromuscul. Disord. *19,* 6-16.

Garg, A., O'Rourke, J., Long, C., Doering, J., Ravenscroft, G., Bezprozvannaya, S., Nelson, B.R., Beetz, N., Li, L., Chen, S., *et al.* (2014). KLHL40 deficiency destabilizes thin filament proteins and promotes nemaline myopathy. J. Clin. Invest. *124,* 3529-3539.

Gong, W., Gohla, R.M., Bowlin, K.M., Koyano-Nakagawa, N., Garry, D.J., and Shi, X. (2015). Kelch Repeat and BTB Domain Containing Protein 5 (Kbtbd5) Regulates Skeletal Muscle Myogenesis through the E2F1-DP1 Complex. J. Biol. Chem. *290,* 15350-15361.

Gonzalez-Perez, A., and Lopez-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am. J. Hum. Genet. *88,* 440-449.

Green, R.C., Berg, J.S., Grody, W.W., Kalia, S.S., Korf, B.R., Martin, C.L., McGuire, A.L., Nussbaum, R.L., O'Daniel, J.M., Ormond, K.E., *et al.* (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet. Med. *15,* 565-574.

Grimm, D.G., Azencott, C.A., Aicheler, F., Gieraths, U., MacArthur, D.G., Samocha, K.E., Cooper, D.N., Stenson, P.D., Daly, M.J., Smoller, J.W., Duncan, L.E., and Borgwardt, K.M. (2015). The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. Hum. Mutat. *36,* 513-523.

Gu, W., Zhang, F., and Lupski, J.R. (2008). Mechanisms for human genomic rearrangements. Pathogenetics *1,* 4.

Gunning, P.W., Schevzov, G., Kee, A.J., and Hardeman, E.C. (2005). Tropomyosin isoforms: divining rods for actin cytoskeleton function. Trends Cell Biol. *15,* 333-341.

Gupta, V.A., Ravenscroft, G., Shaheen, R., Todd, E.J., Swanson, L.C., Shiina, M., Ogata, K., Hsu, C., Clarke, N.F., Darras, B.T., *et al.* (2013). 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.

Hanauer, A., Levin, M., Heilig, R., Daegelen, D., Kahn, A., and Mandel, J.L. (1983). Isolation and characterization of cDNA clones for human skeletal muscle alpha actin. Nucleic Acids Res. *11,* 3503-3516.

Hastings, P.J., Ira, G., and Lupski, J.R. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet. *5,* e1000327.

Hook, J., Lemckert, F., Qin, H., Schevzov, G., and Gunning, P. (2004). Gamma tropomyosin gene products are required for embryonic development. Mol. Cell. Biol. *24,* 2318-2323.

Hupe, P., Stransky, N., Thiery, J.P., Radvanyi, F., and Barillot, E. (2004). Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics *20,* 3413-3422.

Jain, M., Fiddes, I.T., Miga, K.H., Olsen, H.E., Paten, B., and Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. Nat. Methods *12,* 351-356.

Jamuar, S.S., and Tan, E.C. (2015). Clinical application of next-generation sequencing for Mendelian diseases. Hum. Genomics *9,* 10.

Johnston, J.J., Kelley, R.I., Crawford, T.O., Morton, D.H., Agarwala, R., Koch, T., Schaffer, A.A., Francomano, C.A., and Biesecker, L.G. (2000). A novel nemaline myopathy in the Amish caused by a mutation in troponin T1. Am. J. Hum. Genet. *67,* 814-821.

Jones, A.C., Austin, J., Hansen, N., Hoogendoorn, B., Oefner, P.J., Cheadle, J.P., and O'Donovan, M.C. (1999). Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. Clin. Chem. *45,* 1133-1140.

Joo, Y.M., Lee, M.A., Lee, Y.M., Kim, M.S., Kim, S.Y., Jeon, E.H., Choi, J.K., Kim, W.H., Lee, H.C., Min, B.I., Kang, H.S., and Kim, C.R. (2004). Identification of chicken nebulin isoforms of the 31-residue motifs and non-muscle nebulin. Biochem. Biophys. Res. Commun. *325,* 1286-1291.

Jungbluth, H., and Wallgren-Pettersson, C. (2013). The Congenital (Structural) Myopathies. In Emery and Rimoin's Principles and practice of medical genetics, 6th edition, Emery, A. E. H., Korf, B. R., Pyeritz, R. E. and Rimoin, D. L., San Diego: Elsevier Science.

Kabsch, W., and Vandekerckhove, J. (1992). Structure and function of actin. Annu. Rev. Biophys. Biomol. Struct. *21,* 49-76.

Kallioniemi, O.P., Kallioniemi, A., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F., and Pinkel, D. (1993). Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors. Semin. Cancer Biol. *4,* 41-46.

Kazmierski, S.T., Antin, P.B., Witt, C.C., Huebner, N., McElhinny, A.S., Labeit, S., and Gregorio, C.C. (2003). The complete mouse nebulin gene sequence and the identification of cardiac nebulin. J. Mol. Biol. *328,* 835-846.

Kim, J.I., Ju, Y.S., Park, H., Kim, S., Lee, S., Yi, J.H., Mudge, J., Miller, N.A., Hong, D., Bell, C.J.*, et al.* (2009). A highly annotated whole-genome sequence of a Korean individual. Nature *460,* 1011-1015.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46,* 310-315.

Kloosterman, W.P., and Cuppen, E. (2013). Chromothripsis in congenital disorders and cancer: similarities and differences. Curr. Opin. Cell Biol. *25,* 341-348.

Kloosterman, W.P., Guryev, V., van Roosmalen, M., Duran, K.J., de Bruijn, E., Bakker, S.C., Letteboer, T., van Nesselrooij, B., Hochstenbach, R., Poot, M., and Cuppen, E. (2011). Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum. Mol. Genet. *20,* 1916-1924.

Kolomietz, E., Meyn, M.S., Pandita, A., and Squire, J.A. (2002). The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. Genes Chromosomes Cancer *35,* 97-112.

Krawitz, P.M., Schiska, D., Kruger, U., Appelt, S., Heinrich, V., Parkhomchuk, D., Timmermann, B., Millan, J.M., Robinson, P.N., Mundlos, S., Hecht, J., and Gross, M. (2014). Screening for single nucleotide variants, small indels and exon deletions with a next-generation sequencing based gene panel approach for Usher syndrome. Mol. Genet. Genomic Med. *2,* 393-401.

Kruglyak, L., and Nickerson, D.A. (2001). Variation is the spice of life. Nat. Genet. *27,* 234-236.

Kudo, S., Mattei, M.G., and Fukuda, M. (1995). Characterization of the gene for dbpA, a family member of the nucleic-acid-binding proteins containing a cold-shock domain. Eur. J. Biochem. *231,* 72-82.

Labeit, S., Gibson, T., Lakey, A., Leonard, K., Zeviani, M., Knight, P., Wardale, J., and Trinick, J. (1991). Evidence that nebulin is a protein-ruler in muscle thin filaments. FEBS Lett. *282,* 313-316.

Labeit, S., Ottenheijm, C.A., and Granzier, H. (2011). Nebulin, a major player in muscle health and disease. FASEB J. *25,* 822-829.

Laing, N.G., Dye, D.E., Wallgren-Pettersson, C., Richard, G., Monnier, N., Lillis, S., Winder, T.L., Lochmuller, H., Graziano, C., Mitrani-Rosenbaum, S*., et al.* (2009). Mutations and polymorphisms of the skeletal muscle alpha-actin gene (ACTA1). Hum. Mutat. *30,* 1267-1277.

Laing, N.G., Wilton, S.D., Akkari, P.A., Dorosz, S., Boundy, K., Kneebone, C., Blumbergs, P., White, S., Watkins, H., and Love, D.R. (1995). A mutation in the alpha tropomyosin gene TPM3 associated with autosomal dominant nemaline myopathy. Nat. Genet. *9,* 75-79.

Laitila, J., Hanif, M., Paetau, A., Hujanen, S., Keto, J., Somervuo, P., Huovinen, S., Udd, B., Wallgren-Pettersson, C., Auvinen, P., Hackman, P., and Pelin, K. (2012). Expression of multiple nebulin isoforms in human skeletal muscle and brain. Muscle Nerve *46,* 730-737.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W*., et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409,* 860-921.

Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25,* 1841-1842.

Lee, J.A., Carvalho, C.M., and Lupski, J.R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell *131,* 1235-1247.

Lehtokari, V.L., Greenleaf, R.S., DeChene, E.T., Kellinsalmi, M., Pelin, K., Laing, N.G., Beggs, A.H., and Wallgren-Pettersson, C. (2009). The exon 55 deletion in the nebulin gene--one single founder mutation with world-wide occurrence. Neuromuscul. Disord. *19,* 179-181.

Lehtokari, V.L., Kiiski, K., Sandaradura, S.A., Laporte, J., Repo, P., Frey, J.A., Donner, K., Marttila, M., Saunders, C., Barth, P.G*., et al.* (2014). Mutation update: the spectra of nebulin variants and associated myopathies. Hum. Mutat. *35,* 1418-1426.

Lehtokari, V.L., Pelin, K., Donner, K., Voit, T., Rudnik-Schoneborn, S., Stoetter, M., Talim, B., Topaloglu, H., Laing, N.G., and Wallgren-Pettersson, C. (2008). Identification of a founder mutation in TPM3 in nemaline myopathy patients of Turkish origin. Eur. J. Hum. Genet. *16,* 1055-1061.

Lehtokari, V.L., Pelin, K., Herczegfalvi, A., Karcagi, V., Pouget, J., Franques, J., Pellissier, J.F., Figarella-Branger, D., von der Hagen, M., Huebner, A*., et al.* (2011). Nemaline myopathy caused by mutations in the nebulin gene may present as a distal myopathy. Neuromuscul. Disord. *21,* 556-562.

Lehtokari, V.L., Pelin, K., Sandbacka, M., Ranta, S., Donner, K., Muntoni, F., Sewry, C., Angelini, C., Bushby, K., Van den Bergh, P*., et al.* (2006). Identification of 45 novel mutations in the nebulin gene associated with autosomal recessive nemaline myopathy. Hum. Mutat. *27,* 946-956.

Lieber, M.R. (2008). The mechanism of human nonhomologous DNA end joining. J. Biol. Chem. *283,* 1-5.

Liu, P., Carvalho, C.M., Hastings, P.J., and Lupski, J.R. (2012). Mechanisms for recurrent and complex human genomic rearrangements. Curr. Opin. Genet. Dev. *22,* 211-220.

Loman, N.J., and Watson, M. (2015). Successful test launch for nanopore sequencing. Nat. Methods *12,* 303-304.

Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. Proc. Natl. Acad. Sci. U. S. A. *107,* 961-968.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A*., et al.* (2014). Guidelines for investigating causality of sequence variants in human disease. Nature *508,* 469-476.

Marra, J.D., Engelstad, K.E., Ankala, A., Tanji, K., Dastgir, J., De Vivo, D.C., Coffee, B., and Chiriboga, C.A. (2015). Identification of a novel nemaline myopathy-Causing mutation in the troponin T1 (TNNT1) gene: A case outside of the old order amish. Muscle Nerve *51,* 767-772.

Marttila, M., Hanif, M., Lemola, E., Nowak, K.J., Laitila, J., Gronholm, M., Wallgren-Pettersson, C., and Pelin, K. (2014). Nebulin interactions with actin and tropomyosin are altered by disease-causing mutations. Skelet Muscle *4,* 15. eCollection 2014.

Marttila, M., Lehtokari, V.L., Marston, S., Nyman, T.A., Barnerias, C., Beggs, A.H., Bertini, E., Ceyhan-Birsoy, O., Cintas, P., Gerard, M*., et al.* (2014). Mutation update and genotype-phenotype correlations of novel and previously described mutations in TPM2 and TPM3 causing congenital myopathies. Hum. Mutat. *35,* 779-790.

May, A., Abeln, S., Buijs, M.J., Heringa, J., Crielaard, W., and Brandt, B.W. (2015). NGS-eval: NGS Error analysis and novel sequence VAriant detection tooL. Nucleic Acids Res. *43,* W301-5.

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K*., et al.* (2001). A physical map of the human genome. Nature *409,* 934-941.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. *24,* 529-538.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat. Rev. Genet. *11,* 31-46.

Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J*., et al.* (2010). Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. *86,* 749-764.

Morton, S.U., Joshi, M., Savic, T., Beggs, A.H., and Agrawal, P.B. (2015). Skeletal muscle microRNA and messenger RNA profiling in cofilin-2 deficient mice reveals cell cycle dysregulation hindering muscle regeneration. PLoS One *10,* e0123829.

Ng, P.C., and Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. *31,* 3812-3814.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E*., et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. Nature *461,* 272-276.

Nobile, C., Toffolatti, L., Rizzi, F., Simionati, B., Nigro, V., Cardazzo, B., Patarnello, T., Valle, G., and Danieli, G.A. (2002). Analysis of 22 deletion breakpoints in dystrophin intron 49. Hum. Genet. *110,* 418-421.

Nowak, K.J., Wattanasirichaigoon, D., Goebel, H.H., Wilce, M., Pelin, K., Donner, K., Jacob, R.L., Hubner, C., Oexle, K., Anderson, J.R*., et al.* (1999). Mutations in the skeletal muscle alpha-actin gene in patients with actin myopathy and nemaline myopathy. Nat. Genet. *23,* 208-212.

Nworu, C.U., Kraft, R., Schnurr, D.C., Gregorio, C.C., and Krieg, P.A. (2015). Leiomodin 3 and tropomodulin 4 have overlapping functions during skeletal myofibrillogenesis. J. Cell. Sci. *128,* 239-250.

Ockeloen, C.W., Gilhuis, H.J., Pfundt, R., Kamsteeg, E.J., Agrawal, P.B., Beggs, A.H., Dara Hama-Amin, A., Diekstra, A., Knoers, N.V., Lammens, M., and van Alfen, N. (2012). Congenital myopathy caused by a novel missense mutation in the CFL2 gene. Neuromuscul. Disord. *22,* 632-639.

Ong, R.W., AlSaman, A., Selcen, D., Arabshahi, A., Yau, K.S., Ravenscroft, G., Duff, R.M., Atkinson, V., Allcock, R.J., and Laing, N.G. (2014). Novel cofilin-2 (CFL2) four base pair deletion causing nemaline myopathy. J. Neurol. Neurosurg. Psychiatry. *85,* 1058-1060.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. U. S. A. *86,* 2766-2770.

Pelin, K., and Wallgren-Pettersson, C. (2008). Nebulin--a giant chameleon. Adv. Exp. Med. Biol. *642,* 28-39.

Perry, G.H., Ben-Dor, A., Tsalenko, A., Sampas, N., Rodriguez-Revenga, L., Tran, C.W., Scheffer, A., Steinfeld, I., Tsang, P., Yamada, N.A.*, et al.* (2008). The fine-scale and complex architecture of human copy-number variation. Am. J. Hum. Genet. *82,* 685-695.

Perry, S.V. (2001). Vertebrate tropomyosin: distribution, properties and function. J. Muscle Res. Cell. Motil. *22,* 5-49.

Pfuhl, M., Winder, S.J., Castiglione Morelli, M.A., Labeit, S., and Pastore, A. (1996). Correlation between conformational and binding properties of nebulin repeats. J. Mol. Biol. *257,* 367-384.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y.*, et al.* (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. *20,* 207-211.

Ravenscroft, G., Miyatake, S., Lehtokari, V.L., Todd, E.J., Vornanen, P., Yau, K.S., Hayashi, Y.K., Miyake, N., Tsurusaki, Y., Doi, H.*, et al.* (2013). Mutations in KLHL40 Are a Frequent Cause of Severe Autosomal-Recessive Nemaline Myopathy. Am. J. Hum. Genet. *93,* 6-18.

R Core Team, (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W.*, et al.* (2006). Global variation in copy number in the human genome. Nature *444,* 444-454.

Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. *39,* e118.

Romero, N.B., Lehtokari, V.L., Quijano-Roy, S., Monnier, N., Claeys, K.G., Carlier, R.Y., Pellegrini, N., Orlikowski, D., Barois, A., Laing, N.G.*, et al.* (2009). Core-rod myopathy caused by mutations in the nebulin gene. Neurology *73,* 1159-1161.

Sambuughin, N., Swietnicki, W., Techtmann, S., Matrosova, V., Wallace, T., Goldfarb, L., and Maynard, E. (2012). KBTBD13 interacts with Cullin 3 to form a functional ubiquitin ligase. Biochem. Biophys. Res. Commun. *421,* 743-749.

Sambuughin, N., Yau, K.S., Olive, M., Duff, R.M., Bayarsaikhan, M., Lu, S., Gonzalez-Mera, L., Sivadorai, P., Nowak, K.J., Ravenscroft, G.*, et al.* (2010). Dominant mutations in KBTBD13, a member of the BTB/Kelch family, cause nemaline myopathy with cores. Am. J. Hum. Genet. *87,* 842-847.

Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnstrom, K., Mallick, S., Kirby, A.*, et al.* (2014). A framework for the interpretation of de novo mutation in human disease. Nat. Genet. *46,* 944-950.

Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. (2002). Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. *30,* e57.

Schroder, R., Reimann, J., Salmikangas, P., Clemen, C.S., Hayashi, Y.K., Nonaka, I., Arahata, K., and Carpen, O. (2003). Beyond LGMD1A: myotilin is a component of central core lesions and nemaline rods. Neuromuscul. Disord. *13,* 451-455.

Schwartz, M., and Duno, M. (2004). Improved molecular diagnosis of dystrophin gene mutations using the multiplex ligation-dependent probe amplification method. Genet. Test. *8,* 361-367.

Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nat. Methods *11,* 361-362.

Scoto, M., Cullup, T., Cirak, S., Yau, S., Manzur, A.Y., Feng, L., Jacques, T.S., Anderson, G., Abbs, S., Sewry, C., Jungbluth, H., and Muntoni, F. (2013). Nebulin (NEB) mutations in a childhood onset distal myopathy with rods and cores uncovered by next generation sequencing. Eur. J. Hum. Genet. *21,* 1249-1252.

Sharp, A.J., Hansen, S., Selzer, R.R., Cheng, Z., Regan, R., Hurst, J.A., Stewart, H., Price, S.M., Blair, E., Hennekam, R.C.*, et al.* (2006). Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat. Genet. *38,* 1038-1042.

Shaw, C.J., and Lupski, J.R. (2005). Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. Hum. Genet. *116,* 1-7.

Shaw, C.J., and Lupski, J.R. (2004). Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum. Mol. Genet. *13 Spec No 1,* R57-64.

Sheffield, V.C., Beck, J.S., Kwitek, A.E., Sandstrom, D.W., and Stone, E.M. (1993). The sensitivity of single-strand conformation polymorphism analysis for the detection of single base substitutions. Genomics *16,* 325-332.

Shigemizu, D., Momozawa, Y., Abe, T., Morizono, T., Boroevich, K.A., Takata, S., Ashikawa, K., Kubo, M., and Tsunoda, T. (2015). Performance comparison of four commercial human whole-exome capture platforms. Sci. Rep. *5,* 12742.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L., Edwards, K.J., Day, I.N., and Gaunt, T.R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum. Mutat. *34,* 57-65.

Shinawi, M., and Cheung, S.W. (2008). The array CGH and its clinical applications. Drug Discov. Today *13,* 760-770.

Shy, G.M., Engel, W.K., Somers, J.E., and Wanko, T. (1963). Nemaline Myopathy. A New Congenital Myopathy. Brain *86,* 793-810.

Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. (1997). Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer *20,* 399-407.

Spangenburg, E.E., and Booth, F.W. (2003). Molecular regulation of individual skeletal muscle fibre types. Acta Physiol. Scand. *178,* 413-424.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A*., et al.* (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144,* 27-40.

Stern, R.F., Roberts, R.G., Mann, K., Yau, S.C., Berg, J., and Ogilvie, C.M. (2004). Multiplex ligation-dependent probe amplification using a completely synthetic probe set. BioTechniques *37,* 399-405.

Stone, J., and Stone, R. (2011). Atlas of Skeletal Muscles. 7th edition McGraw-Hill Higher Education.

Strachan, T., and Read, A. (2011). Human Molecular Genetics. 4th edition Garland Science.

Sulek, A., Elert, E., Rajkiewicz, M., Zdzienicka, E., Stepniak, I., Krysa, W., and Zaremba, J. (2013). Screening for the hereditary spastic paraplaegias SPG4 and SPG3A with the multiplex ligation-dependent probe amplification technique in a large population of affected individuals. Neurol. Sci. *34,* 239-242.

Sun, Y., Ruivenkamp, C.A., Hoffer, M.J., Vrijenhoek, T., Kriek, M., van Asperen, C.J., den Dunnen, J.T., and Santen, G.W. (2015). Next-generation diagnostics: gene panel, exome, or whole genome? Hum. Mutat. *36,* 648-655.

Tajsharghi, H., Ohlsson, M., Lindberg, C., and Oldfors, A. (2007). Congenital myopathy with nemaline rods and cap structures caused by a mutation in the beta-tropomyosin gene (TPM2). Arch. Neurol. *64,* 1334-1338.

Tan, P., Briner, J., Boltshauser, E., Davis, M.R., Wilton, S.D., North, K., Wallgren-Pettersson, C., and Laing, N.G. (1999). Homozygosity for a nonsense mutation in the alpha-tropomyosin slow gene TPM3 in a patient with severe infantile nemaline myopathy. Neuromuscul. Disord. *9,* 573-579.

Thirion, C., Stucka, R., Mendel, B., Gruhler, A., Jaksch, M., Nowak, K.J., Binz, N., Laing, N.G., and Lochmuller, H. (2001). Characterization of human muscle type cofilin (CFL2) in normal and regenerating muscle. Eur. J. Biochem. *268,* 3473-3482.

Tian, L., Ding, S., You, Y., Li, T.R., Liu, Y., Wu, X., Sun, L., and Xu, T. (2015). Leiomodin-3-deficient mice display nemaline myopathy with fast-myofiber atrophy. Dis. Model. Mech. *8,* 635-641.

Tiso, N., Rampoldi, L., Pallavicini, A., Zimbello, R., Pandolfo, D., Valle, G., Lanfranchi, G., and Danieli, G.A. (1997). Fine mapping of five human skeletal muscle genes: alpha-tropomyosin, beta-tropomyosin, troponin-I slow-twitch, troponin-I fast-twitch, and troponin-C fast. Biochem. Biophys. Res. Commun. *230,* 347-350.

Tong, P., Prendergast, J.G., Lohan, A.J., Farrington, S.M., Cronin, S., Friel, N., Bradley, D.G., Hardiman, O., Evans, A., Wilson, J.F., and Loftus, B. (2010). Sequencing and analysis of an Irish human genome. Genome Biol. *11,* R91.

Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J., and Cavalli-Sforza, L.L. (1996). A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc. Natl. Acad. Sci. USA. *93,* 196-200.

van der Pol, W.L., Leijenaar, J.F., Spliet, W.G., Lavrijsen, S.W., Jansen, N.J., Braun, K.P., Mulder, M., Timmers-Raaijmakers, B., Ratsma, K., Dooijes, D., and van Haelst, M.M. (2014). Nemaline myopathy caused byTNNT1 mutations in a Dutch pedigree. Mol. Genet. Genomic Med. *2,* 134-137.

van El, C.G., Cornel, M.C., Borry, P., Hastings, R.J., Fellmann, F., Hodgson, S.V., Howard, H.C., Cambon-Thomsen, A., Knoppers, B.M., Meijers-Heijboer, H*., et al.* (2013). Whole-genome sequencing in health care. Recommendations of the European Society of Human Genetics. Eur. J. Hum. Genet. *21 Suppl 1,* S1-5.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A*., et al.* (2001). The sequence of the human genome. Science *291,* 1304-1351.

Vissers, L.E., Bhatt, S.S., Janssen, I.M., Xia, Z., Lalani, S.R., Pfundt, R., Derwinska, K., de Vries, B.B., Gilissen, C., Hoischen, A*., et al.* (2009). Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. Hum. Mol. Genet. *18,* 3579-3593.

Wallgren-Pettersson, C. (1990). Congenital nemaline myopathy: a longitudinal study. Commentationes Physico-Mathematicae III 1990, Dissertationes no. 30, the Finnish Society of Sciences and Letters, Helsinki, p102.

Wallgren-Pettersson, C. (1989). Congenital nemaline myopathy. A clinical follow-up of twelve patients. J. Neurol. Sci. *89,* 1-14.

Wallgren-Pettersson, C., Jasani, B., Newman, G.R., Morris, G.E., Jones, S., Singhrao, S., Clarke, A., Virtanen, I., Holmberg, C., and Rapola, J. (1995). Alpha-actinin in nemaline bodies in congenital nemaline myopathy: immunological confirmation by light and electron microscopy. Neuromuscul. Disord. *5,* 93-104.

Wallgren-Pettersson, C., and Laing, N.G. (2000). Report of the 70th ENMC International Workshop: nemaline myopathy, 11-13 June 1999, Naarden, The Netherlands. Neuromuscul. Disord. *10,* 299-306.

Wallgren-Pettersson, C., Lehtokari, V.L., Kalimo, H., Paetau, A., Nuutinen, E., Hackman, P., Sewry, C., Pelin, K., and Udd, B. (2007). Distal myopathy caused by homozygous missense mutations in the nebulin gene. Brain *130,* 1465-1476.

Wallgren-Pettersson, C., Pelin, K., Hilpela, P., Donner, K., Porfirio, B., Graziano, C., Swoboda, K.J., Fardeau, M., Urtizberea, J.A., Muntoni, F.*, et al.* (1999). Clinical and genetic heterogeneity in autosomal recessive nemaline myopathy. Neuromuscul. Disord. *9,* 564-572.

Wallgren-Pettersson, C., Pelin, K., Nowak, K.J., Muntoni, F., Romero, N.B., Goebel, H.H., North, K.N., Beggs, A.H., Laing, N.G., and ENMC International Consortium On Nemaline Myopathy. (2004). Genotype-phenotype correlations in nemaline myopathy caused by mutations in the genes for nebulin and skeletal muscle alpha-actin. Neuromuscul. Disord. *14,* 461-470.

Wang, K. (1996). Titin/connectin and nebulin: giant protein rulers of muscle structure and function. Adv. Biophys. *33,* 123-134.

Wang, X., Huang, Q.Q., Breckenridge, M.T., Chen, A., Crawford, T.O., Morton, D.H., and Jin, J.P. (2005). Cellular fate of truncated slow skeletal muscle troponin T produced by Glu180 nonsense mutation in amish nemaline myopathy. J. Biol. Chem. *280,* 13241-13249.

Weiss, G.J., Hoff, B.R., Whitehead, R.P., Sangal, A., Gingrich, S.A., Penny, R.J., Mallery, D.W., Morris, S.M., Thompson, E.J., Loesch, D.M., and Khemka, V. (2015). Evaluation and comparison of two commercially available targeted next-generation sequencing platforms to assist oncology decision making. Onco Targets Ther. *8,* 959-967.

Witt, C.C., Burkart, C., Labeit, D., McNabb, M., Wu, Y., Granzier, H., and Labeit, S. (2006). Nebulin regulates thin filament length, contractility, and Z-disk structure in vivo. EMBO J. *25,* 3843-3855.

Xuan, J., Yu, Y., Qing, T., Guo, L., and Shi, L. (2013). Next-generation sequencing in the clinic: promises and challenges. Cancer Lett. *340,* 284-295.

Yuen, M., Sandaradura, S.A., Dowling, J.J., Kostyukova, A.S., Moroz, N., Quinlan, K.G., Lehtokari, V.L., Ravenscroft, G., Todd, E.J., Ceyhan-Birsoy, O.*, et al.* (2014). Leiomodin-3 dysfunction results in thin filament disorganization and nemaline myopathy. J. Clin. Invest. *124,* 4693-4708.

Zhang, F., Gu, W., Hurles, M.E., and Lupski, J.R. (2009). Copy number variation in human health, disease, and evolution. Annu. Rev. Genomics Hum. Genet. *10,* 451-481.