

Division of Genetics, Department of Biosciences
Integrative Life Sciences Doctoral Program
University of Helsinki
Helsinki, Finland

GENETIC BACKGROUND OF LATE-ONSET SPINAL MOTOR NEURONOPATHY

Sini Penttilä

ACADEMIC DISSERTATION

To be presented, with permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki, for public examination in Auditorium XII, Main Building, on 29th August 2018, at 12 noon.

Helsinki 2018

Supervisors

Professor Bjarne Udd, MD, PhD

Neuromuscular Research Center, University of Tampere and Tampere University Hospital; Tampere, Finland

Folkhälsan Institute of Genetics and Department of Medical Genetics, Haartman Institute, University of Helsinki; Helsinki, Finland

Department of Neurology, Vaasa Central Hospital; Vaasa, Finland

Docent Johanna Palmio, MD, PhD

Neuromuscular Research Center, University of Tampere and Tampere University Hospital; Tampere, Finland

Thesis Advisory Committee

Docent Pekka Heino, PhD

Division of Genetics, Department of Biosciences, Faculty of Biological and Environmental Sciences, University of Helsinki; Helsinki, Finland

Docent Ritva Karhu, PhD

Department of Clinical Genetics, Fimlab Laboratories; Tampere, Finland

Reviewers

Professor Jaakko Ignatius, MD, PhD

Department of Clinical Genetics, Turku University Hospital; Turku, Finland

Docent Irma Järvelä, MD, PhD

Department of Medical Genetics, University of Helsinki; Helsinki, Finland

Opponent

Professor Frank Baas, MD, PhD

Department of Clinical Genetics, Leiden University Medical Centre; Leiden, Netherlands

Custos

Professor Ville Mustonen, PhD

Division of Genetics, Department of Biosciences, Faculty of Biological and Environmental Sciences, University of Helsinki; Helsinki, Finland

ISBN 978-951-51-4384-6 (pbk)

ISBN 978-951-51-4385-3 (PDF)

<http://ethesis.helsinki.fi>

Unigrafia

Helsinki 2018

TABLE OF CONTENTS

TABLE OF CONTENTS	4
LIST OF ORIGINAL PUBLICATIONS	6
AUTHOR CONTRIBUTIONS	7
ABBREVIATIONS	8
ABSTRACT	9
TIIVISTELMÄ	10
1 INTRODUCTION	12
2 REVIEW OF THE LITERATURE	13
2.1 Introduction to molecular genetics.....	13
2.1.1 Genes	13
2.1.2 Microsatellites.....	14
2.1.3 Mutations.....	15
2.1.4 Modes of Mendelian inheritance	16
2.2 Identification of the disease-causing mutation	17
2.2.1 Genetic linkage and linkage analysis	17
2.2.1.1 Founder effect	18
2.2.2 Candidate gene approach	19
2.2.3 Sanger sequencing.....	19
2.2.4 Massively parallel sequencing.....	20
2.2.5 Verification of the mutation.....	21
2.3 Spinal muscular atrophies	22
2.3.1 <i>SMN1</i> -related SMA	22
2.3.2 Non-5q forms of SMA	24
2.3.2.1 Distal spinal muscular atrophy / distal hereditary motor neuronopathy	25
2.3.2.2 Proximal spinal muscular atrophy	27
2.3.2.3 Spinal muscular atrophy, Jokela type	28
2.4 <i>CHCHD10</i>	29
2.4.1 Structure and function of <i>CHCHD10</i> protein.....	29
2.4.2 Diseases associated with mutations in <i>CHCHD10</i>	31
3 AIMS OF THE STUDY	34
4 SUBJECTS AND METHODS	35
4.1 Patients and controls.....	35

4.1.1 Linkage analysis (I)	35
4.1.2 Identification of new families and refinement of the linked region (II).....	36
4.1.3 Identification of the disease-causing mutation (III)	36
4.1.4 Prevalence and clinical outcome of <i>CHCHD10</i> mutations in Finland (IV)	36
4.2 Methods	36
4.2.1 DNA isolation (I-IV).....	36
4.2.2 Genotyping (I-III).....	37
4.2.3 Linkage analysis (I) and haplotype construction (I-III)	37
4.2.4 Exome sequencing (II)	37
4.2.5 Sanger sequencing (II-IV).....	37
4.2.6 Whole-genome sequencing (III).....	38
5 RESULTS AND DISCUSSION	39
5.1 Linkage analysis (I).....	39
5.2 Identification of new families and refinement of the linked region (II)	40
5.3 Identification of the disease-causing mutation (III).....	41
5.4 Prevalence and clinical outcome of <i>CHCHD10</i> mutations in Finland (IV)	42
5.5 Limitations and methodological considerations	45
6 CONCLUSIONS AND FUTURE PROSPECTS	46
7 ACKNOWLEDGMENTS	48
8 ELECTRONIC TOOLS AND DATABASES	50
9 REFERENCES	51

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals.

- I **Penttilä S**, Jokela M, Hackman P, Saukkonen AM, Toivanen J, Udd B. Autosomal dominant late-onset spinal motor neuronopathy is linked to a new locus on chromosome 22q11.2-13.2. *Eur J Hum Gen* 2012 Nov;20(11):1193-6.
- II **Penttilä S***, Jokela M*, Huovinen S, Saukkonen AM, Toivanen J, Lindberg C, Baumann P, Udd B. Late-onset spinal motor neuronopathy - a common form of dominant SMA. *Neuromuscul Disord* 2014 Mar;24(3):259-68.
- III **Penttilä S**, Jokela M, Bouquin H, Saukkonen AM, Toivanen J, Udd B. Late onset spinal motor neuronopathy is caused by mutation in *CHCHD10*. *Ann Neurol* 2015 Jan;77(1):163-72.
- IV **Penttilä S**, Jokela M, Saukkonen AM, Toivanen J, Palmio J, Lähdesmäki J, Sandell S, Shcherbii M, Auranen M, Ylikallio E, Tyynismaa H, Udd B. *CHCHD10* mutations and motor neuron disease: the distribution in Finnish patients. *J Neurol Neurosurg Psychiatry* 2017 Mar;88(3):272-77.

*The authors contributed equally to the work.

The articles are reprinted with the permission of their copyright holders. In addition, some unpublished material is presented.

AUTHOR CONTRIBUTIONS

I

Recruitment and evaluation of patients	MJ, AMS, JT, BU
Linkage analysis	SP
Selection of additional microsatellites	SP, PH
Genotyping	SP
Construction of haplotypes	SP
Writing of paper	SP, MJ, BU

II

Recruitment and evaluation of patients	MJ, PB, CL, AMS, JT, BU
Histological examination	SH
Assembling clinical data	MJ
Selection of additional microsatellites	SP
Genotyping	SP
Construction of haplotypes	SP
Sanger sequencing	SP
Inspection of exome sequencing data	SP
Writing of paper	SP, MJ, SH, BU

III

Recruitment and evaluation of patients	MJ, AMS, JT, BU
Genotyping	SP
Inspection of whole genome sequencing data	SP
Sanger sequencing	SP, HB
Analyzing clinical data	MJ
Writing of paper	SP, MJ, BU

IV

Recruitment and evaluation of patients	MJ, AMS, JT, JP, JL, SS, MA, BU
Sanger sequencing of HSP patients	MS, EY, HT
Sanger sequencing of other patients	SP
MtDNA deletion analysis	MS, EY, HT
Assembling clinical data	MJ, JP, MA, EY
Writing of paper	SP, MJ, JP, HT, BU

ABBREVIATIONS

ACMG	American College of Medical Genetics and Genomics
ALS	amyotrophic lateral sclerosis
AMP	Association for Molecular Pathology
bp	base pair
CHCH	coiled-coil-helix-coiled-coil-helix
CK	creatine kinase
cM	centiMorgan
CMT	Charcot-Marie-Tooth disease
CMT2	Charcot-Marie-Tooth disease type 2, axonal form of CMT
ddNTP	dideoxynucleoside triphosphate
dHMN	distal hereditary motor neuropathy
DSMA	distal spinal muscular atrophy
FTD	frontotemporal dementia
FTDALS1	frontotemporal dementia and/or amyotrophic lateral sclerosis 1
GATK	Genome Analysis Toolkit
gnomAD	genome Aggregation Database
HSP	hereditary spastic paraplegia
LOD	logarithm of the odds
LOSMoN	late-onset spinal motor neuropathy
MICOS	mitochondrial contact site and cristae organizing system
MND	motor neuron disease
MPS	massively parallel sequencing
mRNA	messenger RNA
NGS	next-generation sequencing
OMIM	Online Mendelian Inheritance in Man
PCR	polymerase chain reaction
SBMA	spinobulbar muscular atrophy
SISu	Sequencing Initiative Suomi
SMA	spinal muscular atrophy
SMAJ	spinal muscular atrophy, Jokela type
SNP	single nucleotide polymorphism
tRNA	transfer RNA
UTR	untranslated region
WES	whole-exome sequencing
WGS	whole-genome sequencing

In addition, standard abbreviations of amino acids and approved symbols of human genes and proteins are used.

ABSTRACT

The aim of this study was to find and describe the genetic background of a new form of spinal muscular atrophy (SMA). Late-onset spinal motor neuronopathy (LOSMoN), later named spinal muscular atrophy, Jokela type (SMAJ), is a relatively benign autosomal dominant form of a lower motor neuron disorder. Disease onset is after the age of 30-40 years, and SMAJ is characterized by initial painful cramps and fasciculations affecting the proximal and distal muscles of the upper and lower limbs. The disease is slowly progressive, resulting in weakness and mild to moderate muscle atrophy later in life.

SMAJ was originally identified in two families in Eastern Finland. The genome-wide scan study performed for these families showed that the disease is linked to chromosome 22q11.2-q13.2. The disease-associated haplotype was identical in both families, suggesting a founder effect. The founder hypothesis was also confirmed later, as several other unrelated patients carrying the same haplotype were identified.

The disease-causing mutation, c.197G>T p.G66V in *CHCHD10*, was detected by whole-genome sequencing. The mutation was present in all then identified 55 SMAJ patients belonging to 17 families. At the same time, other dominant mutations in *CHCHD10* were described to cause a wide range of neurological disorders. *CHCHD10* was the first SMA-causing gene identified that encodes for a mitochondrial protein.

The prevalence and clinical outcome of *CHCHD10* mutations in Finnish neuromuscular disease patients were clarified in a screening study. The only detected mutation was c.197G>T p.G66V, and all patients carrying this mutation had a phenotype restricted to SMAJ. The prevalence of c.197G>T p.G66V was estimated to be around 4/100 000 in Finland, i.e. approximately 200 symptomatic SMAJ patients.

The results of this study confirm that SMAJ is a genetically distinct entity caused by a dominant mutation c.197G>T p.G66V in *CHCHD10*. This finding enables genetic testing of SMAJ, providing patients with an accurate diagnosis and prognosis. According to our genotyping results, c.197G>T p.G66V is a founder mutation in Finland, all SMAJ patients having common ancestry. SMAJ was shown to be relatively common in Finland. It is clearly the most common *CHCHD10*-related disease reported, and in Finland it may be the most common form of SMA. Because SMAJ seems to be absent from other populations, it can be considered a part of the Finnish disease heritage.

TIIVISTELMÄ

Tämän tutkimuksen tavoitteena oli kuvata LOSMoN-taudin eli myöhään alkavan spinaalisen motoneuronitaudin geneettinen tausta. LOSMoN, joka sittemmin nimettiin Jokela-tyypin spinaaliseksi lihasatrofiaksi (SMAJ), on suhteellisen lievä, autosomissa vallitsevasti periytyvä alemman motoneuronin sairaus. Tauti alkaa 30-40 vuoden iässä, ja sen alkuaireita ovat kivuliaat lihaskrampit ja lihasnykinät, joita esiintyy proksimaalisesti ja distaalisesti sekä ala- että yläraajoissa. Tauti on hitaasti etenevä, ja se johtaa lihasten heikkouteen sekä lievään tai keskivaikeaan lihasten surkastumiseen myöhäisellä iällä.

SMAJ tunnistettiin alun perin kahdessa itäsuomalaisessa perheessä. Koko genomien kartoituksessa näiden perheiden taudin havaittiin kytkeytyvän kromosomialueeseen 22q11.2-q13.2. Tautiin kytkeytyvä haplotyyppi oli molemmissa perheissä samanlainen, mikä viittaa siihen, että perheiden taudilla on yhteinen alkuperä. Tämä perustajavaikutus saatiin vahvistettua, kun tutkimuksessa löydettiin muita, alkuperäisiin perheisiin kuulumattomia potilaita, joilla oli sama tautiin kytkeytyvä haplotyyppi.

Taudin aiheuttava mutaatio, *CHCHD10*-geenin muutos c.197G>T p.G66V, löydettiin koko genomien sekvensoinnilla. Kyseinen mutaatio havaittiin kaikilta siihen mennessä tunnistetuilta 55 SMAJ-potilaalta, jotka kuuluivat 17 eri perheeseen. Samaan aikaan *CHCHD10*-geenin muiden vallitsevasti periytyvien mutaatioiden kuvattiin aiheuttavan erilaisia neurologisia sairauksia. *CHCHD10* oli ensimmäinen mitokondriaalista proteiinia koodaava geeni, jonka on kuvattu aiheuttavan spinaalista lihasatrofiaa.

CHCHD10-geenin mutaatioiden yleisyyttä ja niiden aiheuttamia tauteja selvitettiin suomalaisille neuromuskulaaritautipotilaille toteutetussa seulontatutkimuksessa. Ainoa tutkimuksessa havaittu mutaatio oli c.197G>T p.G66V, ja kaikkien sitä kantavien potilaiden taudinkuva vastasi SMAJ:a. Mutaation c.197G>T p.G66V yleisyydeksi Suomessa arvioitiin noin 4/100 000, mikä tarkoittaa noin 200 oireista SMAJ-potilasta.

Tämä tutkimus osoitti, että SMAJ on geneettisesti erillinen tauti, joka johtuu vallitsevasti periytyvästä *CHCHD10*-geenin c.197G>T p.G66V –mutaatiosta. Tutkimuksen tulosten myötä SMAJ:lle on voitu kehittää geenitesti, jonka avulla potilaat voivat saada taudilleen oikean diagnoosin ja ennusteen. Genotyypitystulokset osoittivat, että c.197G>T p.G66V on perustajamutaatio suomalaisväestössä. Tämän tutkimuksen perusteella SMAJ on Suomessa suhteellisen yleinen tauti. SMAJ on selvästi yleisin *CHCHD10*-geeniin

liitetyistä taudeista, ja Suomessa se on mahdollisesti yleisin SMA:n muoto. Koska SMAJ:a ei ilmeisesti esiinny muissa väestöissä, sitä voidaan pitää osana suomalaista tautiperintöä.

1 INTRODUCTION

Inherited disorders are diseases that have a genetic background. Although individual genetic disorders are rare, together they are present in at least 2 % of all neonates and affect 5 % of the population by the age of 25 (Turnpenny et al., 2005). In monogenic diseases, identification of the causative gene is very important because it enables immediate clinical diagnostics by genetic testing, genetic counseling, and the development of possible therapeutic interventions. Recent advances in next-generation sequencing technologies have unveiled a growing number of new genetic disorders. Identification of the pathogenic mutation may still be complicated if no knowledge of the gene function is available and/or there are not enough patients for segregation studies. Furthermore, it may be challenging to determine the significance of the variants identified, leaving many findings classified as variants of unknown significance (Richards et al., 2015).

Motor neuron diseases (MNDs) are a heterogeneous group of diseases that result from progressive death of motor neurons. MNDs are generally classified according to whether the degeneration affects upper motor neurons, lower motor neurons, or both. The most common MND is amyotrophic lateral sclerosis (ALS), which progressively affects both upper and lower motor neurons, leading to death usually within 3-5 years of onset of symptoms. MNDs affecting only upper motor neurons are classified mainly as hereditary spastic paraplegias (HSPs), whereas MNDs involving specifically lower motor neurons are spinal muscular atrophies (SMAs). However, the clinical and genetic overlap of these diseases is considerable (James and Talbot, 2006; Rossor et al., 2012).

SMAs form a category of inherited disorders characterized by degeneration of motor neurons in the spinal cord, leading to symmetric muscle weakness and atrophy. Despite advances in SMA research, the molecular basis and phenotypic spectrum of SMA are not fully understood. An increasing number of ubiquitously expressed genes have been identified as the genetic cause of SMA. Nevertheless, causative mutations are identified in up to 35.6% of patients (Bansagi et al., 2017), leaving most patients without definitive diagnosis.

Late-onset spinal motor neuronopathy (LOSMoN) is a new form of slowly progressive autosomal dominant SMA that has been described in Finnish patients. Clinical characteristics of LOSMoN do not directly match any of the previously reported autosomal dominant SMA disorders (Jokela et al., 2011). The studies presented in this thesis aim to describe the genetic background of this novel clinical entity.

2 REVIEW OF THE LITERATURE

2.1 Introduction to molecular genetics

The human genome consists of chromosomes, i.e. large DNA molecules that contain the genetic information. The DNA molecule is composed of two chains of nucleotides arranged in a double helix (Watson and Crick, 1953). The nucleotides, namely adenine, thymine, cytosine, and guanine, hold the DNA chains together by pairing with each other and are hence referred to as base pairs (bp). The human genome is approximately 3 000 000 000 bp in size (International Human Genome Sequencing Consortium, 2004). However, only 1.5% of the human genome is estimated to be protein coding, the rest being associated with non-coding RNA molecules, repetitive elements, introns, pseudogenes, regulatory regions, and sequences for which the function is still unknown.

2.1.1 Genes

Genes are genomic units that code for proteins or RNA molecules. The structure of a gene consists of a promoter region, which regulates the expression of the gene, coding parts, called exons, and non-coding parts, called untranslated regions (UTRs) and introns (Figure 1). In addition to the promoter, the gene may have other regulatory elements, such as enhancers and silencers, which are often located far from the actual gene. Regulatory elements control the expression, i.e. transcription of the gene. In transcription, the genetic information is transmitted from DNA to messenger RNA (mRNA). The transcribed pre-mRNA is modified by removing the introns (splicing) and adding a 5' cap and a poly-A tail. In alternative splicing, also some exonic sequences may be removed. A single gene can produce several different mRNA and protein isoforms that may have distinct functions. The mature mRNA is transported from the nucleus to the cytoplasm and translated to amino acids that form the final protein. Only coding parts of exons are translated, whereas the 5' cap, UTRs and poly-A tail regulate the translation. A simple scheme of gene function is presented in Figure 1 (Strachan and Read, 2011).

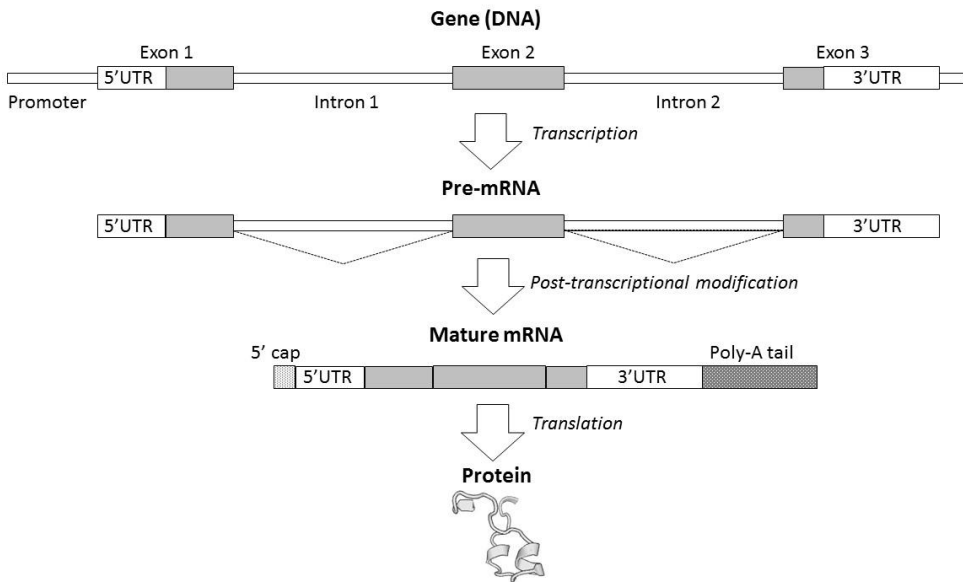


Figure 1. Structure and function of a protein-coding gene. In transcription, both coding and non-coding parts of the gene are transcribed. The resulting pre-mRNA is modified so that introns are spliced away and the 5' cap and poly-A tail are added. The mature mRNA is transported to the cytoplasm and translated to form the final protein product.

2.1.2 Microsatellites

Microsatellites are tandemly repeated DNA sequences. They consist of blocks of sequences with units less than ten nucleotides long. Most of the microsatellites are located in non-coding parts of the genome, but in humans approximately 17% of genes contain repeats within their open reading frames (Gemayel et al., 2010). Microsatellites are extremely unstable because they are prone to errors in both recombination and replication, resulting in an increase or decrease of the repeating unit. Therefore, microsatellites are highly polymorphic and they are often used as markers in linkage analysis. Some repeat expansion mutations of microsatellites are also known to cause human disease, such as the neuromuscular disorders spinal and bulbar muscular atrophy (SBMA) (La Spada et al., 1991), myotonic dystrophy type 1 and 2 (Brook et al., 1992; Liquori et al., 2001) and frontotemporal dementia and/or amyotrophic lateral sclerosis 1 (FTDALS1) (DeJesus-Hernandez et al., 2011; Renton et al., 2011).

2.1.3 Mutations

Mutation is a change in the genetic material. Some mutations are harmless variants, i.e. polymorphisms, whereas others are pathogenic. The vast majority of mutations occur spontaneously through errors in DNA replication or repair, but they can also be caused by a mutagen, such as ionizing radiation, DNA reactive chemicals or viruses. Only mutations affecting germ-line cells can be transmitted to future generations. There are different types of mutations (Table 1), of which substitution being the most common (Turnpenny et al., 2005; Strachan and Read, 2011).

Table 1. Different types of mutations.

Mutation type	Definition	Subtype	Location	Consequence
Substitution	Replacement of a single nucleotide by another	Missense mutation	Coding sequence	Change of an amino acid to another
		Nonsense mutation	Coding sequence	Change of an amino acid coding codon to stop codon
		Synonymous / silent mutation	Coding sequence	No change of an amino acid
		Splicing mutation	Coding / non-coding sequence	Change in splice site recognition
Deletion	Loss of one or more nucleotides, not a multiple of three	Frameshift mutation	Coding sequence	Loss of reading frame
	Loss of one or more multiples of three	In-frame mutation	Coding sequence	Deletion of one or more amino acids
Insertion	Insertion of one or more nucleotides, not a multiple of three	Frameshift mutation	Coding sequence	Loss of reading frame
	Insertion of one or more multiples of three	In-frame mutation	Coding sequence	Insertion of one or more amino acids
Copy number variation	Large deletion / insertion	-	One or more exons / genes	Loss or addition of exons / genes
Repeat number mutation	Deletion / insertion of a repeating unit	-	Coding / non-coding sequence	Decrease or increase of a repeating unit

2.1.4 Modes of Mendelian inheritance

If a trait or disorder is caused by a mutation(s) in a single gene, it exhibits monogenic or Mendelian inheritance. According to the Online Mendelian Inheritance in Man database (OMIM, 16.4.2018), there are 5 219 monogenic human disorders and traits for which the molecular basis is known. If a gene that determines a trait is on an autosome, it is said to show autosomal inheritance. A trait is called dominant if it manifests in heterozygous state, i.e. only one mutated allele causes the phenotype. A parent with an autosomal dominant trait has a 50% risk of transmitting the trait to his or her child. A recessive trait requires both alleles to be mutated, this means either homozygosity or compound heterozygosity. Thus, both parents need to carry at least one mutation in order to have an affected child, the risks being 25% for an affected child, 50% for an unaffected carrier and 25% for an unaffected non-carrier.

Sex-linked inheritance involves a gene mutation on either the X- or Y-chromosome. Y-chromosomal inheritance is very rare, but already 324 X-chromosomal traits are known (OMIM, 16.4.2018). X-linked traits can be dominant or recessive, but the inheritance pattern differs from that of autosomal traits because males only have one X-chromosome and they are hemizygous for all X-chromosomal genes. Therefore, X-linked recessive traits usually manifest in males only. X-linked dominant traits manifest in both sexes alike, but they never exhibit male-to-male inheritance.

In some cases, genetic traits do not seem to follow any of the patterns of inheritance described above. Such traits may originate from so-called *de novo* mutations. A *de novo* mutation is a newly arisen mutation that has not been inherited from either of the parents. Parents who have a child with a *de novo* mutation are not at risk of having another child with the same mutation, but the child can pass on the mutation to his/her own children. In some cases, more than one child in the same family can carry the same apparent *de novo* mutation. These mutations are not actual *de novo* mutations but they are due to gonadal mosaicism in one of the parents. In gonadal mosaicism, the mutation is present in a proportion of gonadal cells, but not in other cells of an individual.

Atypical inheritance patterns may also be caused by reduced penetrance. Penetrance refers to the proportion of individuals harboring a particular pathogenic mutation or genotype who exhibit clinical signs of the associated disorder within a specific and clearly defined time period (Cooper et al., 2013). If this proportion is not 100%, the disorder is said to exhibit reduced penetrance. Reduced penetrance is likely to be a consequence of the combination of a variety of different genetic and environmental factors. It is often seen in late-onset diseases in which it can be age-related.

2.2 Identification of the disease-causing mutation

It is very important to identify the mutation responsible for any inherited disorder. This enables direct genetic testing and thereby fast and accurate diagnosis. When the genetic background of a disease is known, wrong and sometimes harmful treatments can be avoided and the patient can be offered proper management, including genetic counseling, prognosis of the disease, and possibility to enroll in clinical trials. Curative treatments are not currently available for most genetic disorders, but some gene therapies have been developed (Mendell et al., 2017; Mah, 2018), and understanding the genetic background of the disorders is the first step towards the development of future therapies. Identification of a gene for a hereditary disorder begins with investigating the family history of patients and drawing the pedigrees. Based on the pedigrees, an assumption of the mode of inheritance is made and further methods for molecular genetic analyses are selected.

2.2.1 Genetic linkage and linkage analysis

Genetic linkage can be defined as the tendency of DNA sequences located closely in the chromosome to be inherited together. This means that genes or other genetic markers that reside close to each other remain linked during meiosis. In meiosis sister-chromatids exchange homologous segments in a process called crossing over or recombination. The unit of measurement for genetic linkage is the centiMorgan (cM). If two loci are 1 cM apart, then a cross-over occurs between them in 1 of 100 meioses (Turnpenny et al., 2005). Because recombination rarely separates loci that lie very close together, sets of alleles on the same small chromosomal segment tend to be transmitted as a block. Such a block of alleles is known as a haplotype (Strachan and Read, 2011).

Linkage analysis is a powerful tool to detect the chromosomal location of disease genes. In gene mapping, a panel of genetic markers (microsatellites or single nucleotide polymorphisms, SNPs) is used to effectively label the participants' genomes so that the segregation of genetic material can be followed. Linkage analysis consists of studying the pattern of co-inheritance of marker alleles and the presence or absence of a phenotype (Barrett and Teare, 2011). For linkage analysis, a set of recombination fractions and logarithm of the odds (LOD) scores are calculated. Recombination fraction (θ) is defined as the frequency with which a single recombination will take place between two loci in meiosis. If $\theta = 0$, the two loci are in the same genomic position, and if $\theta = 0.5$, the two loci are completely unlinked. LOD score is a measure of the likelihood of linkage, which is calculated as follows (Morton, 1955):

$$\text{LOD} = \log_{10} \frac{\text{probability of birth sequence with a given linkage value}}{\text{probability of birth sequence with no linkage}}$$

$$= \log_{10} \frac{(1-\theta)^N \theta^R}{0,5^{(N+R)}}$$

θ = recombination fraction

N = number of non-recombinant offspring

R = number of recombinant offspring

If $\text{LOD} \geq +3$, the region is considered to be linked. Linkage can be excluded if $\text{LOD} < -2$.

The success and reliability of linkage analysis depends on the markers chosen, the individuals analyzed and the mathematical method used. Microsatellites are usually highly informative markers because they have many possible alleles. SNPs normally have only two alleles, which means that more informative individuals are needed when using SNPs than when using microsatellites. However, SNPs are more densely distributed in the genome, hence providing a means for very high-resolution mapping (Terwilliger et al., 1992; Barrett and Teare, 2011; Ott et al., 2015).

Linkage analysis can be either parametric or non-parametric. Parametric linkage analysis requires specification of the mode of inheritance, whereas non-parametric linkage analysis is model-free (Kruglyak et al., 1996; Barrett and Teare, 2011). Both can be based on two-point calculations where linkage is calculated between the disease locus and the marker, or multipoint calculations where linkage is calculated between the disease locus and more than one marker simultaneously (Kruglyak et al., 1996). There are many freely available software packages for linkage analysis for different kinds of situations. Some examples are Genehunter (Kruglyak et al., 1996), which is widely used for non-parametric linkage analysis, Morgan (Wijsman et al., 2006), which uses Monte Carlo Markov Chain methods and is suited to handle large complex pedigrees, and Merlin (Abecasis et al., 2002), which can cope with very large numbers of marker loci (Barrett and Teare, 2011).

2.2.1.1 Founder effect

In genetically isolated populations derived from a small number of founders, monogenic disorders may be caused by just one mutation originating from a single founder. Consequently, in such a population all patients with a certain disorder share the same mutation as well as the haplotype segregating with the disorder. Founder effect is a typical phenomenon in the Finnish population,

particularly displayed in the Finnish disease heritage, a group of rare hereditary diseases overrepresented in Finland (Norio et al., 1973; Norio, 2003). In genetically isolated populations, founder effect can be assumed in linkage analyses, even though the individuals analyzed are not known to be related.

2.2.2 Candidate gene approach

To identify a disease-causing gene by candidate gene approach, a list of possible genes is made. There are many ways to do this. The list can be based on genes and gene families already identified with similar phenotypes, knowledge of the protein product compatible with the presumed disease mechanism, or other hints for possible genes. More often, however, the chromosomal location of the gene is first defined by linkage analysis, and the genes within the linked region are listed as candidates. Once the list has been composed, a search for those genes is made to identify a gene that carries mutation(s) in patients, but not in controls. For Mendelian conditions, this approach has been very successful (Strachan and Read, 2011).

2.2.3 Sanger sequencing

Sanger sequencing (Sanger et al., 1977) is the conventional method for determining the precise order of nucleotides within a defined DNA region. The method is based on fluorescently labeled dideoxynucleoside triphosphates (ddNTPs) that randomly terminate the elongation of DNA template. The reaction produces a range of different fragments that have a common 5' end but variable 3' ends. The fragments are separated by size using an automated capillary sequencer that detects the fluorescence of the ddNTPs and determines the order of the nucleotides.

Sanger sequencing is widely used for mutation detection in both clinical diagnostics and research. However, the read length in Sanger sequencing is limited (usually 500-800 bp), which means that even the smallest genes usually have to be sequenced in several parts (Strachan and Read, 2011). This is laborious and for large sequencing projects the method is relatively expensive. The method is also prone to errors; if a primer binding site contains a sequence variation, there is a risk of allele drop-out. Therefore, in the absence of heterozygous variants, allele drop-out cannot be excluded (Stevens et al., 2017).

2.2.4 Massively parallel sequencing

The method of massively parallel sequencing (MPS), also known as next-generation sequencing (NGS), can carry out up to billions of sequencing reactions in parallel. There are a number of different MPS platforms (L. Liu et al., 2012). Although they differ in the underlying technology involved, their overall processes are very similar: DNA fragmentation, adaptor ligation, immobilization, amplification, sequencing reaction, and data analysis (Figure 2). Common applications include whole-genome sequencing (WGS), whole-exome sequencing (WES) or targeted gene panels for disease-causing gene discovery, genetic diagnosis, and targeted cancer therapy (Nguyen and Burnett, 2014). At least in theory, MPS provides a way to identify a disease-causing gene without any family data, exact phenotype data, or predefined lists of candidate genes.

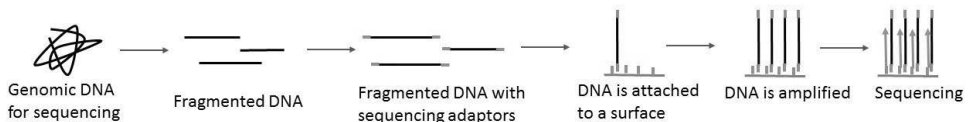


Figure 2. Overview of massively parallel sequencing process.

The most commonly used platform for MPS is Illumina. In Illumina's platform, DNA is fragmented, e.g. by random shearing and adaptors are ligated to the ends of the fragments. These single molecules of DNA are attached to a flat surface (the flow cell) by hybridizing the adaptors to their complementary sequences on the flow cell. The fragments are amplified *in situ* to form DNA clusters. The clusters are used as templates for synthetic sequencing with reversible terminator deoxyribonucleotides labeled with a removable fluorophore. The sequencing is performed in repeated cycles and after each cycle of incorporation of nucleotides the identity of the inserted base is determined by laser-induced excitation of the fluorophores and imaging. The fluorescent dye of the nucleotides is removed and a 3' hydroxyl group is regenerated for the next cycle of nucleotide addition. Images of the surface are analyzed to generate a high-quality sequence (Bentley et al., 2008).

MPS enables sequencing of multiple target regions in one reaction, so that even the whole human genome can be sequenced in one experiment. This significantly reduces the costs and time needed for sequencing. MPS also, at least partly, circumvents the possibility of allele drop-out, as the final sequence is compiled from multiple random fragments so that small variants cannot block out the amplification of the whole allele. There are some limitations in

MPS as well. Sequencing errors are quite frequent, although they are usually concentrated in certain (e.g. GC-rich) genomic regions and they may be identified using certain quality scores. Incomplete coverage, where regions are sequenced poorly or not at all, is also a considerable problem. One of the major challenges of MPS is data analysis, interpretation, and storage. Powerful computational and bioinformatics tools are required to store the data and perform the various steps involved in read alignment, variant calling, and annotation (Nguyen and Burnett, 2014).

2.2.5 Verification of the mutation

When a possible disease-causing mutation has been identified, its pathogenicity must be verified. The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have published standards and guidelines for the interpretation of sequence variants (Richards et al., 2015). In order to determine whether a mutation is pathogenic, its frequency in the general population is ascertained from population databases. There are several publicly available population databases, some of which are presented in Table 2. If the frequency of the mutation is suitable for the assumed mode of inheritance, the segregation of the mutation with the disease is studied. The possible consequence of the mutation may be studied by *in silico* prediction programs, such as MutationTaster, PolyPhen or SIFT, but these are not conclusive. Functional studies can be a powerful tool in support of pathogenicity, although not all functional studies are effective in predicting an impact on a gene or protein function. According to the guidelines of ACMG and AMP, there is strong evidence of the pathogenicity of the mutation if (1) its prevalence in affected individuals is statistically increased over controls in population databases, (2) the same amino acid change is an established pathogenic variant or the variant is predicted to be a null variant in a gene where loss-of-function is a known disease mechanism, (3) well-established functional studies show a deleterious effect, and (4) the variant co-segregates with the disease in multiple affected family members or the variant is confirmed to be *de novo*.

Table 2. Examples of publicly available population databases.

Population database	Internet address	Content
Genome Aggregation Database, gnomAD	http://gnomad.broadinstitute.org/	The provided data set spans 123 136 exome sequences and 15 496 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies. Pediatric disease subjects were excluded.
Exome Variant Server, EVS	http://evs.gs.washington.edu/EVS	Database of variants found during exome sequencing of more than 200 000 individuals of European and African American ancestry.
1000 Genomes	http://www.internationalgenome.org/	Database of variants found during low-coverage and high-coverage genomic and targeted sequencing from 26 populations. Provides more diversity than EVS, but also contains lower quality data and some cohorts contain related individuals.
Sequencing Initiative Suomi, SISu	http://www.sisuproject.fi/	Single-nucleotide variants and insertion/deletions from exomes of over 10 000 Finns sequenced in disease-specific and population genetic studies.

2.3 Spinal muscular atrophies

Spinal muscular atrophies are hereditary disorders that exclusively or predominantly affect lower motor neurons, leading to progressive muscle weakness and atrophy. Worldwide the most common SMA is caused by mutations in the *SMN1* gene on chromosome 5q13. Non-5q SMAs are rare and both clinically and genetically heterogeneous (Zerres and Rudnik-Schöneborn, 2003; Van Den Bosch and Timmerman, 2006; Peeters et al., 2014).

2.3.1 *SMN1*-related SMA

SMN1-related spinal muscular atrophy (SMA1-4, OMIM #253300, #253550, #253400, #271150) is a common autosomal recessive disorder caused by degeneration of anterior horn cells of the spinal cord, leading to symmetric muscle weakness and atrophy. It is the leading inherited cause of infant mortality with a reported incidence of approximately 1 in 10 000 live births and a carrier frequency of 1 in 54 (Sugarman et al., 2012). *SMN1*-related SMA is classically divided into four clinical subtypes based on the maximal functional status achieved (Table 3) (Darras, 2015).

Table 3. Subtypes of SMN1-related SMA.

SMA type	Other name	Age at onset	Life span	Highest motor milestone achieved
Type I	Werdnig-Hoffmann disease	< 6 months	< 2 years	Never sits unsupported
Type II	Intermediate SMA (Dubowitz disease)	6-18 months	> 2 years	Sits independently, never stands or walks
Type III	Kugelberg-Welander disease	> 18 months	Almost normal	Stands and walks
Type IV	Adult-onset SMA	> 21 years	Normal	Normal

The disease is caused by deletions or point mutations in the *SMN1* gene (Lefebvre et al., 1995). *SMN1* lies within an inverted duplication on chromosome 5q13. The centromeric half of the duplication contains *SMN2*, which is an almost identical copy of *SMN1*. However, *SMN2* differentiates from *SMN1* by five nucleotides, one of which creates an exonic splicing suppressor that leads to exclusion of exon 7 in almost all transcripts (Monani et al., 1999). Thus, *SMN2* produces only a small part of functional SMN protein, and this amount is insufficient to compensate for the loss of function of *SMN1*. Patients with mutated *SMN1* can carry a variable number of *SMN2*. The more *SMN2* copies one has, the more SMN protein is expressed and the less severe the phenotype (Feldkötter et al., 2002; Rudnik-Schöneborn et al., 2009).

SMN protein is part of a large multiprotein complex and it interacts with spliceosomal small nuclear ribonucleoprotein particles (Q. Liu et al., 1997). SMN plays a crucial role in the generation of the pre-mRNA splicing machinery and thus in mRNA biogenesis (Pellizzoni et al., 1998). SMN deficiency, similar to that occurring in severe SMA, alters the stoichiometry of small nuclear RNAs and causes widespread pre-mRNA splicing defects in numerous transcripts of diverse genes (Z. Zhang et al., 2008).

SMN1-related SMA is one of the first hereditary disorders for which a seemingly effective gene therapy has been developed. The first on the market was an antisense oligonucleotide drug nusinersen (Spinraza), which is administered intrathecally. It binds to *SMN2* causing inclusion of exon 7 with good effect when started early (Finkel et al., 2017). The US Food and Drug Administration approved nusinersen for the treatment of patients with all subtypes of SMA in December 2016. However, the extreme cost of the medication and the complicated logistical requirements for administering nusinersen have raised difficult ethical and health insurance issues (King and Bishop, 2017).

2.3.2 Non-5q forms of SMA

Non-5q forms of SMA are usually classified on the basis of inheritance pattern and distribution of muscle weakness (proximal, distal, or bulbar) (Darras, 2011). The classification of these disorders is difficult, however, because many genetically defined disease entities show allelic variants and significant phenotypic overlap, and, on the other hand, one specific phenotype may be caused by mutations in more than one gene (Darras, 2011; Rossor et al., 2012).

Despite the growing number of genes implicated in non-5q SMA, no unifying molecular disease mechanism has been identified. Common pathways to non-5q SMA genes include molecular transport, lipid metabolism, and RNA processing and trafficking. It has been hypothesized that large axons of motor neurons have such high metabolic requirements for maintenance, transport over long distances, and precise connectivity that they are particularly vulnerable to defects in ubiquitously expressed proteins (Irobi et al., 2006). There is significant clinical overlap of non-5q SMA genes with other neuromuscular, mainly neurogenic disorders, such as hereditary spastic paraplegia, amyotrophic lateral sclerosis, and Charcot-Marie-Tooth disease (CMT) (Figure 3) (Rossor et al., 2012; Peeters et al., 2014).

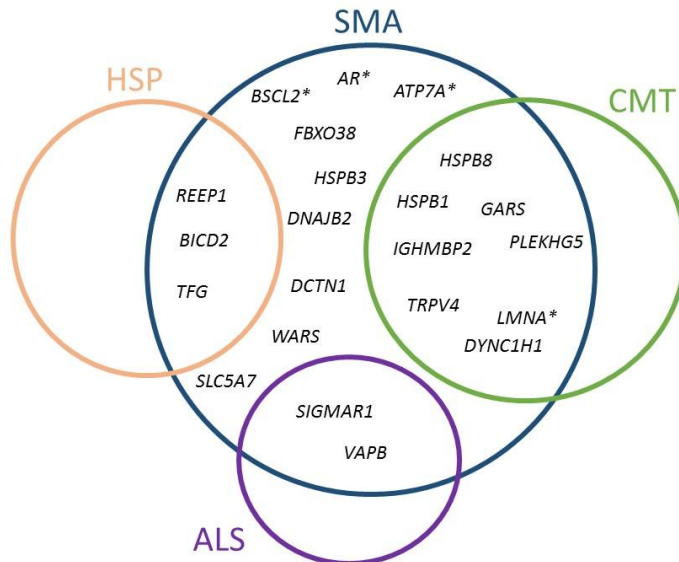


Figure 3. Clinical overlap of causal genes for non-5q SMA with other neuromuscular disorders. Asterisks indicate genes that are also associated with non-neuromuscular diseases. ALS = amyotrophic lateral sclerosis, CMT = Charcot-Marie-Tooth disease, HSP = hereditary spastic paraplegia.

2.3.2.1 Distal spinal muscular atrophy / distal hereditary motor neuropathy

Distal spinal muscular atrophy (DSMA), also called distal hereditary motor neuropathy (dHMN), is a length-dependent, predominantly motor neuropathy. This is in contrast to CMT and hereditary sensory neuropathies where sensory involvement is a significant component. The disease is very slowly progressive and symmetrical, and bulbar involvement is rare (Rossor et al., 2012). Current classification of OMIM (12.1.2018) uses DSMA for recessive and dHMN for dominant forms of the disease. The prevalence of DSMA/dHMN in Northern England has been reported to be 2.14 per 100 000 (Bansagi et al., 2017).

Increasing numbers of ubiquitously expressed genes have been identified as the genetic cause of DSMA/dHMN. So far, 15 causative genes and two loci have been identified (Table 4). The causative genes code for proteins with diverse functions. For example, *HSPB1*, *HSPB8*, and *DNAJB2* are chaperones in the quality control machinery and play a role in regulation of protein folding and turnover (Carra et al., 2005; Ackerley et al., 2006; Blumen et al., 2012), *GARS* and *WARS* are involved in transfer RNA (tRNA) aminoacylation (Antonellis et al., 2003; Tsai et al., 2017), *DCTN1* encodes for an axonal transport protein (Puls et al., 2003), and the protein product of *ATP7A* is a cation channel (Kennerson et al., 2010). Many of the disease genes are shared between DSMA/dHMN and the axonal forms of CMT (CMT2), indicating identical disease mechanisms. Therefore, DSMA/dHMN has been suggested to be a subcategory of CMT (Bansagi et al., 2017).

Table 4. Currently known disease genes and loci for distal SMA. Data adapted from OMIM and GeneTable of Neuromuscular Disorders.

Disease	OMIM	Inheritance	Gene	Locus	Protein	Noteworthy features
DSMA1	#604320	AR	<i>IGHMBP2</i>	11q13.3	DNA-binding protein SMUBP-2	Respiratory distress
DSMA2	#605726	AR	<i>SIGMAR1</i>	9p13.3	Sigma non-opioid intracellular receptor 1	-
DSMA3	%607088	AR	-	11q13.3	-	-
DSMA4	#611067	AR	<i>PLEKHG5</i>	1p36	Pleckstrin homology domain-containing family G member 5	Severe early-onset
DSMA5	#614881	AR	<i>DNAJB2</i>	2q35	DnaJ homolog, subfamily B member 2	
dHMN1	%182960	AD	-	7q34-q36	-	Juvenile onset
dHMN2A	#158590	AD	<i>HSPB8</i>	12q24.23	Heat shock protein beta-8	More pronounced in LL
dHMN2B	#608634	AD	<i>HSPB1</i>	7q11.23	Heat shock protein beta-1	More pronounced in LL
dHMN2C	#613376	AD	<i>HSPB3</i>	5q11.2	Heat shock protein beta-3	More pronounced in LL
dHMN2D	#615575	AD	<i>FBXO38</i>	5q32	F-box only protein 38	Calf-predominant
dHMN5A	#600794	AD	<i>GARS</i>	7p14	Glycine-tRNA ligase	Confined largely to UL
dHMN5A	#600794	AD	<i>BSCL2</i>	11q12	Seipin	Confined largely to UL
dHMN5B	#614751	AD	<i>REEP1</i>	2p11.2	Receptor expression-enhancing protein 1	Primarily affecting intrinsic hand muscles
dHMN7A	#158580	AD	<i>SLC5A7</i>	2q12.31	High affinity choline transporter 1	Vocal cord paralysis
dHMN7B	#607641	AD	<i>DCTN1</i>	2p13	Dynactin subunit 1	Vocal cord paralysis, facial paralysis
dHMN9	#617721	AD	<i>WARS</i>	14q32	Tryptophan-tRNA ligase, cytoplasmic	Juvenile onset
dSMAX	#300489	XR	<i>ATP7A</i>	Xq21	Copper-transporting ATPase 1	-

AR = autosomal recessive, AD = autosomal dominant, XR = X-linked recessive, LL = lower limbs, UL = upper limbs.

2.3.2.2 Proximal spinal muscular atrophy

Proximal SMA is characterized by symmetrical weakness, more pronounced for proximal than distal limb muscles, and generally affecting legs more than arms. The clinical course ranges from static to rapidly progressive, leading to respiratory distress requiring mechanical ventilation. Sensitivity is spared (Peeters et al., 2014). There are also forms of proximal SMA with prominent additional syndromic features, such as arthrogyrosis, myoclonic epilepsy, sensorineural deafness, or pontocerebellar hypoplasia (Peeters et al., 2014), but they are not discussed here.

Currently, seven genes are known to cause proximal SMA (Table 5). The most common form worldwide is SBMA, also known as Kennedy's disease. SBMA is a late-onset X-linked recessive disorder with proximal spinal and bulbar weakness but without pyramidal tract or major sensory impairment (Kennedy et al., 1968). Its prevalence in the Vaasa region of Western Finland was estimated to be 13 per 85 000 male inhabitants (Udd et al., 1998), but has later been shown to be two times more common (Bjarne Udd, personal communication). SBMA is caused by a repeat-expansion mutation in the first exon of the *AR* gene (La Spada et al., 1991). This study concerns the eighth subtype of proximal SMA.

As with distal SMAs, the causative genes of proximal SMAs are mostly ubiquitously expressed. The protein products of *LMNA* are lamin A and lamin C, which are structural components of the nuclear lamina (Stuurman et al., 1998), whereas *TRPV4* encodes for a cation channel (Deng et al., 2010), and *DYNC1H1* and *BICD2* produce molecular motors (Harms et al., 2012; Neveling et al., 2013; Oates et al., 2013; Peeters et al., 2013). Their molecular defects can affect also other tissues, causing e.g. diverse laminopathies (*LMNA*), skeletal dysplasias (*TRPV4*), and malformations of cortical development (*DYNC1H1*) (Peeters et al., 2014).

Table 5. Currently known disease genes for proximal non-5q SMA. Data adapted from OMIM and GeneTable of Neuromuscular Disorders.

Disease	OMIM	Inheritance	Gene	Locus	Protein
Scapuloperoneal SMA	#181405	AD	<i>TRPV4</i>	12q24.11	Transient receptor potential cation channel subfamily V member 4
Lower extremity-predominant SMA type 1	#158600	AD	<i>DYNC1H1</i>	14q32.31	Cytoplasmic dynein 1 heavy chain 1
Lower extremity-predominant SMA type 2	#615290	AD	<i>BICD2</i>	9q22.31	Protein bicaudal D homolog 2
Finkel-type late-onset SMA	#182980	AD	<i>VAPB</i>	20q13.32	Vesicle-associated membrane protein-associated protein B/C
Adult-onset proximal SMA, followed by cardiac involvement	#159001	AD	<i>LMNA</i>	1q22	Lamin-A/C
Hereditary motor and sensory neuropathy, Okinawa type	#604484	AD	<i>TFG</i>	3q12.2	TRK-fused gene protein
Spinal and bulbar muscular atrophy	#313200	XR	<i>AR</i>	Xq12	Androgen receptor

2.3.2.3 Spinal muscular atrophy, Jokela type

Spinal muscular atrophy, Jokela type (SMAJ, OMIM #615048), originally named late-onset spinal motor neuronopathy (Jokela et al., 2011), is a relatively benign form of autosomal dominant SMA identified and described in several Finnish families. The disease onset is after the age of 30-40 years, and it is characterized by painful cramps and fasciculations affecting the proximal and distal muscles of the upper and lower limbs. Other symptoms and signs of the disease are decreased or absent tendon reflexes, elevated creatine kinase (CK), and hand tremor. The disease is slowly progressive, resulting in weakness and mild to moderate muscle atrophy later in life. Patients have remained ambulant for several decades after onset of the disease and their life expectancy is within normal range. Electromyography and muscle biopsy reveal chronic and active neurogenic findings (Jokela et al.,

2011; Jokela, 2015). So far, the disease has not been reported in other populations.

2.4 CHCHD10

CHCHD10 is a gene located on chromosome 22: 23,765,834-23,768,443 (reference sequence GRCh38.p10). *CHCHD10* has four exons and it encodes for five transcripts, three of which are potentially protein coding (Figure 4). The gene has one paralogue, *CHCHD2*, and 79 known orthologues throughout the animal kingdom, including invertebrates. The genomic sequences of *CHCHD2* and *CHCHD10* are approximately 54% identical (Ensembl genome browser 91) (Zerbino et al., 2018).

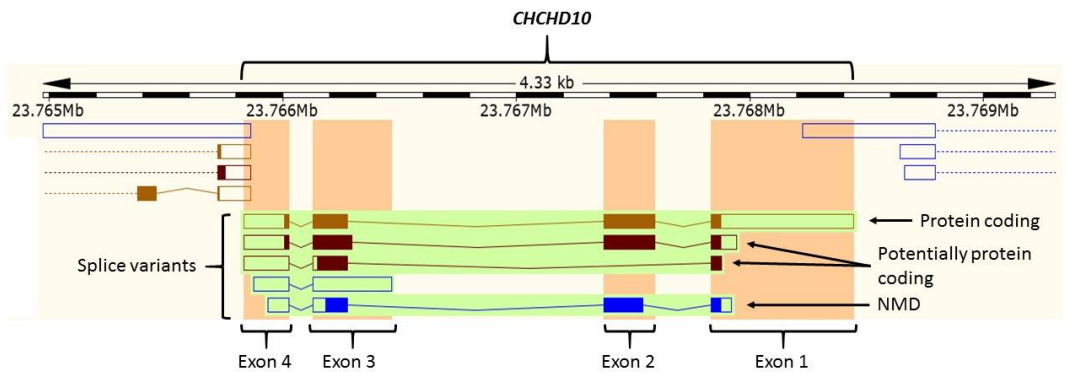


Figure 4. Human CHCHD10 gene. The gene is located on the reverse strand at 23,765,834-23,768,443. Different splice variants are marked with light green background. Open boxes indicate non-coding regions of each transcript. NMD = nonsense mediated decay. Data adapted from Ensembl.

2.4.1 Structure and function of CHCHD10 protein

CHCHD10 encodes a coiled-coil-helix-coiled-coil-helix (CHCH) domain containing protein. *CHCHD10* belongs to a family of mitochondrial proteins characterized by two conserved CX₉C motifs, called twin CX₉C proteins. The CHCH domain forms a helix-turn-helix fold, stabilized by two disulfide bonds (Cavallaro, 2010). The exact structure of *CHCHD10* is not known, but a prediction, modeled with RaptorX structure prediction server (Källberg et al., 2012), is presented in Figure 5.

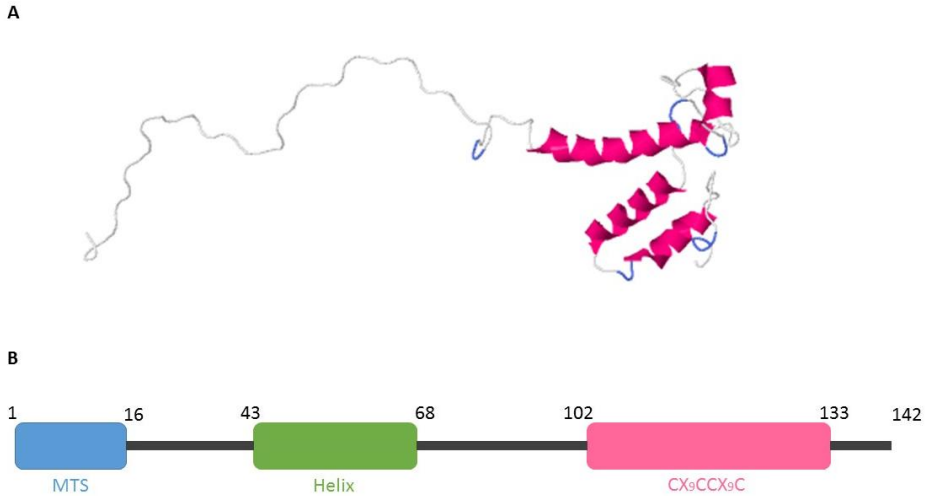


Figure 5. Predicted structure of CHCHD10. A. Predicted tertiary structure of CHCHD10. Secondary structures are colored. *B.* Schematic presentation of CHCHD10 domains. Numbers refer to amino acids. MTS = mitochondrial targeting sequence.

The function of CHCHD10 remains unknown. It has been established to be located in the intermembrane space of mitochondria and to be enriched at cristae junctions (Bannwarth et al., 2014). CHCHD10 is associated with the mitochondrial inner membrane, through either a transmembrane helix or a membrane-bound region (Burstein et al., 2018). CHCHD10 has been reported to be part of the mitochondrial contact site and cristae organizing system (MICOS) complex (Genin et al., 2015), which is crucial for mitochondrial membrane architecture and cristae organization (Rampelt et al., 2017). However, in another study, the association of CHCHD10 with MICOS could not be shown (Straub et al., 2018). Furthermore, no ultrastructural abnormalities of mitochondria were detected in CHCHD10 knockout mice (Burstein et al., 2018) or in patient-derived fibroblasts (Brockmann et al., 2018; Straub et al., 2018), which does not support the hypothesis that CHCHD10 is required for mitochondrial cristae maintenance (Burstein et al., 2018).

Two recent independent studies suggest that CHCHD10 interacts with CHCHD2, both of which interact with p32/C1QBP, a protein with various intra- and extra-mitochondrial functions (Burstein et al., 2018; Straub et al., 2018). There is also evidence that CHCHD10-CHCHD2 complexes are necessary for efficient mitochondrial respiration (Burstein et al., 2018; Straub et al., 2018), and the results of Straub et al. (2018) suggest a key role for CHCHD10 in respiration, particularly under the stress conditions induced by forcing mitochondrial oxidative phosphorylation and increasing oxidative stress. CHCHD10 and CHCHD2 both have rapid turnover, supporting

regulatory rather than structural function (Burstein et al., 2018). A study utilizing *Caenorhabditis elegans* models, mouse cell lines, mouse primary neurons and mouse brains, demonstrated that CHCHD10 exerts a protective role in mitochondrial and synaptic integrity and inhibition of cytoplasmic TDP-43 accumulation (Woo et al., 2017). However, in patient-derived fibroblasts TDP-43 mislocalization could not be detected, and the results demonstrated that mutant alleles CHCHD10^{R15L} and CHCHD10^{G66V} are not expressed in patient cells (Brockmann et al., 2018). CHCHD10 knockdown studies in different cell lines have contradictory results, suggesting that there could be a cell type specificity involving compensatory processes for the loss of CHCHD10 (Burstein et al., 2018). The effect of mutant CHCHD10 has been speculated to cause disease through a gain of toxic function mechanism (Burstein et al., 2018) or haploinsufficiency (Woo et al., 2017; Brockmann et al., 2018; Straub et al., 2018).

2.4.2 Diseases associated with mutations in *CHCHD10*

Dominant mutations in *CHCHD10* have been identified to cause a wide range of neurological disorders. The first reported mutation was c.176C>T p.S59L, described to cause frontotemporal dementia (FTD)-ALS with mitochondrial myopathy (Bannwarth et al., 2014). Two mutations, c.43C>A p.R15S and c.172G>C p.G58R, located *in cis* have been identified as a cause of mitochondrial myopathy (Ajroud-Driss et al., 2015). In addition to these distinct phenotypes, several *CHCHD10* mutations have been associated with different neurological disorders, including ALS, FTD, Alzheimer's disease, and Parkinson's disease (Table 6). Most of the reported patients have been sporadic cases and no segregation studies have been available. Functional studies have shown that mutations p.R15L, p.G58R, p.S59L, and p.G66V alter the protein function (Bannwarth et al., 2014; Ajroud-Driss et al., 2015; Woo et al., 2017; Brockmann et al., 2018; Burstein et al., 2018; Straub et al., 2018), whereas other potentially pathogenic mutations have not been investigated at protein level so far.

Table 6. Possibly pathogenic exonic variants in CHCHD10. Data concerning mutation c.197G>T p.G66V includes the results from this study.

Mutation	Exon	Frequency in gnomAD	Reported phenotypes	Number of reported patients	Patient origin	Pathogenic according to functional studies	Segregates with the disease	References
c.34C>T p.P12S	1	0,00003239	ALS	1	Spain	n.a.	n.a.	(Dols-Icardo et al., 2015)
c.43C>A p.R15S*	2	-	MM	10	Puerto Rico	No	Yes	(Ajroud-Driss et al., 2015)
c.44C>A p.R15L	2	-	ALS	13	Germany, USA, Canada	Yes	No / Yes	(Müller et al., 2014; Johnson et al., 2014; Kurzwelley et al., 2015; M. Zhang et al., 2015)
c.64C>T p.H22Y	2	-	FTD	1	China	n.a.	n.a.	(Jiao et al., 2016)
c.67C>A p.P23T	2	-	FTD	1	Italy	n.a.	n.a.	(M. Zhang et al., 2015)
c.67C>T p.P23S†	2	-	FTD	1	China	n.a.	n.a.	(Jiao et al., 2016)
c.68C>T p.P23L	2	-	ALS, FTD	2	China	n.a.	n.a.	(Jiao et al., 2016; Shen et al., 2017)
c.89C>T p.S30L	2	-	PD	1	China	n.a.	n.a.	(X. Zhou et al., 2018)
c.95C>A p.A32D	2	-	FTD	1	China	n.a.	n.a.	(Jiao et al., 2016)
c.104C>A p.A35D	2	0,00004369	FTLD, AD	2	Italy, China	n.a.	n.a.	(M. Zhang et al., 2015; Xiao et al., 2017)
c.170T>A p.V57E	2	-	FTD	1	China	n.a.	n.a.	(Jiao et al., 2016)
c.172G>C p.G58R*	2	-	MM	10	Puerto Rico	Yes	Yes	(Ajroud-Driss et al., 2015)

c.176C>T p.S59L	2	-	FTD-ALS with MM, FTD	10	France, Spain, Germany	Yes	Yes	(Bannwarth et al., 2014; Chausseot et al., 2014; Blauwendraat et al., 2018)
c.197G>T p.G66V	2	0,00001789	SMAJ, CMT2	93	Finland	Yes	Yes	Present study (III, IV), (Müller et al., 2014; Auranen et al., 2015; Pasanen et al., 2016)
c.239C>T p.P80L	2	0,0002869	ALS	4	Italy, Canada, Belgium	n.a.	Possibly	(Ronchi et al., 2015; M. Zhang et al., 2015; Perrone et al., 2017)
c.244C>T p.Q82X	2	-	FTD	1	Spain	n.a.	n.a.	(Dols-Icardo et al., 2015)
c.285G>C p.Q95H	3	-	ALS	1	China	n.a.	n.a.	(Q. Zhou et al., 2017)
c.285G>A p.Q95=	3	-	ALS	1	China	n.a.	n.a.	(Q. Zhou et al., 2017)
c.322C>T p.Q108X	3	-	FTD	1	Belgium	n.a.	n.a.	(Perrone et al., 2017)

The variants have been annotated according to transcript NM_213720. AD = Alzheimer's disease, ALS = amyotrophic lateral sclerosis, CMT2 = Charcot-Marie-Tooth disease type 2, MM = mitochondrial myopathy, FTD = frontotemporal dementia, FTLD = frontotemporal lobar degeneration, PD = Parkinson's disease, SMAJ = spinal muscular atrophy, Jokela type, n.a. = not assessed, * = mutations *in cis*, † = homozygous mutation.

3 AIMS OF THE STUDY

This work was the genetic part of a larger study aiming at describing a new disease entity both clinically and genetically. Prior to this thesis work, it had long been known that there are Finnish patients with an autosomal dominant, adult-onset lower motor neuron disease that does not really fit any previously defined neuromuscular disease categories. The study started already in 2008, and the first clinical paper was published in 2011 (Jokela et al., 2011).

The aims of this study were as follows:

1. To show that late-onset spinal motor neuronopathy is a new, genetically distinct entity.
2. To provide evidence for a possible founder mutation.
3. To identify the disease-causing mutation.
4. To investigate Finnish neuromuscular disease patients to determine the prevalence and clinical outcome of the mutations in the identified gene

4 SUBJECTS AND METHODS

4.1 Patients and controls

In this study, altogether 437 DNA samples from patients and unaffected family members were included. A summary of all samples is presented in Table 7. The DNA was isolated from peripheral blood leukocytes or saliva. All individuals included in the study were adults and all unaffected family members were at least 35 years old. The age of affected individuals at clinical examination ranged from 29 to 86 years. All individuals were of Finnish origin, although two blood samples were provided from Sweden. In addition, 104 control samples from the normal Finnish population were used.

This study was conducted at Tampere Neuromuscular Research Center. The study protocol was approved by local ethics committees, and samples were obtained in accordance to the Declaration of Helsinki.

Table 7. Summary of the samples used in the study.

	Affected individuals		Unaffected family members		Normal population control samples
	New	From earlier studies	New	From earlier studies	
Study I	11	-	7	-	-
Study II	43	7	11	5	-
Study III	18	37	11	19	104
Study IV	336	-	-	-	-
Total	408		29		104

4.1.1 Linkage analysis (I)

Samples included in Study I were collected from members of two families (F1 and F2) initially identified and evaluated clinically (Jokela et al., 2011). DNA samples of 18 individuals were investigated. Eleven individuals were affected, six were unaffected, and the affection status of one family member was uncertain.

4.1.2 Identification of new families and refinement of the linked region (II)

Patients included in Study II were adult sporadic SMA patients and patients in families with autosomal dominant inheritance and a phenotype compatible with SMAJ (e.g. slowly progressive lower motor neuron disease, absent tendon reflexes, elevated CK). DNA samples were collected from 54 individuals belonging to 26 families. Eleven of these individuals were unaffected family members. In addition, the DNA samples of 12 members (seven affected, five unaffected) of family F1 from Study I were included in the analyses.

4.1.3 Identification of the disease-causing mutation (III)

Study III included 55 patients and 30 unaffected family members from 17 families with adult-onset SMA and autosomal dominant inheritance. The DNA samples of 56 individuals had already been included in Studies I and/or II, whereas 29 DNA samples were new. In this study, also 104 anonymous control DNA samples were used. The control samples had been obtained from Finnish blood donors originally collected for the 'Health 2000 Project' in collaboration with Professor Leena Peltonen-Palotie.

4.1.4 Prevalence and clinical outcome of *CHCHD10* mutations in Finland (IV)

In Study IV, 336 Finnish patients with distinct neurogenic disorders were selected to clarify whether *CHCHD10* mutations are prevalent in selected disease groups. Of these patients, 215 had lower motor neuron syndrome similar to SMAJ, 28 had hereditary spastic paraplegia (HSP), 24 had non-specified neurogenic disorder, 14 had mitochondrial myopathy, and 55 had a diagnosis of ALS.

4.2 Methods

4.2.1 DNA isolation (I-IV)

Genomic DNA was isolated from peripheral blood and muscle biopsies using Puregene DNA Blood Kit (Gentra Systems, Minneapolis, MN, USA) or ArchivePure DNA Purification Kit (5 Prime, Hamburg, Germany). From saliva samples, DNA was isolated using Oragene DNA saliva kits (DNA Genotek, Ottawa, Canada) according to the manufacturer's instructions.

4.2.2 Genotyping (I-III)

For 16 members of families F1 and F2, a genome-wide scan was performed using deCODE 536 microsatellite marker set at an average of 8 cM density (Kong et al., 2002). To confirm and refine the haplotype, seven additional microsatellite markers, D22S264, D22S446, D22S306, D22S686, D22S926, D22S419, and D22S421, were analyzed in Study I. In Study II, six more microsatellite markers (D22S303, SHGC-106816, D22S301, D22S345, D22S343, and D22S925) were added so that up to 13 microsatellite markers were used for genotyping. The fluorescently labeled polymerase chain reaction (PCR) products were analyzed using ABI3130xl automatic DNA Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and Peak Scanner Software v1.0 (Applied Biosystems).

4.2.3 Linkage analysis (I) and haplotype construction (I-III)

Linkage analysis on all autosomes was run with Merlin (Abecasis et al., 2002) using the results of the genome-wide scan and genotyping of markers D22S264, D22S446, D22S306, D22S686, D22S926, D22S419, and D22S421. In linkage analysis, a parametric dominant model was used with a disease allele frequency of 0.0001 and absolute disease risk of 0.9 for carriers. In Study I, haplotypes were constructed using Merlin. In later studies (II, III), haplotypes were constructed manually referring to the founder haplotype.

4.2.4 Exome sequencing (II)

Exome sequencing was performed at the Axeq Technologies laboratory in South Korea. The DNA enrichment was performed using Illumina TruSeq™ Exome Enrichment Kit (Illumina, San Diego, CA, USA). Captured DNA fragments were sequenced on Illumina HiSeq 2000 platform using 100 bp paired-end reads. Sequence reads were aligned to the human reference genome (UCSC hg19) using Burrows-Wheeler Aligner (H. Li and Durbin, 2009), and variant calling was carried out with Genome Analysis Toolkit (GATK; Broad Institute, Cambridge, MA, USA). Bioinformatics was performed at Axeq Technologies. The data were visualized with Integrative Genomics Viewer (Robinson et al., 2011; Thorvaldsdottir et al., 2013).

4.2.5 Sanger sequencing (II-IV)

Sanger sequencing was performed in Study II for the entire coding sequence of the two best candidate genes at the linked region, *SNRPD3* and *SGSM1*. In

Study III, exon 8 in *SLC2A11* and exon 2 in *CHCHD10* were sequenced in three patients (F1:II-16, F6:II-3, and F12:II-2) to verify the potential disease-causing variants found in WGS and to determine whether the variants segregate with the disease. Exon 2 in *CHCHD10* was then sequenced in all subjects in the study. In Study IV, exon 2 in *CHCHD10* was sequenced in all patients. Additionally, in 103 patients, the rest of the coding region was also sequenced in order to detect possible mutations outside exon 2.

Primer sequences for the exons studied were designed using Primer3 software (Koressaar and Remm, 2007; Untergasser et al., 2012) to include the entire exon and exon-intron borders. The regions examined were amplified with PCR (DyNAzyme II, Thermo Scientific Finnzymes, Vantaa, Finland or Thermo Scientific Fermentas PCR Master Mix; MBI Fermentas, Amherst, NY, USA) and directly sequenced using a Big-Dye Terminator v3.1 Kit on an ABI3130xl automatic DNA Genetic Analyzer (Applied Biosystems). Sequences were analyzed with Sequencher 5.1 software (Gene Codes Corporation, Ann Arbor, MI, USA).

4.2.6 Whole-genome sequencing (III)

The DNA sample of patient F1:II-16 was used for whole-genome sequencing. WGS was performed at Beijing Genomics Institute (BGI) using the Illumina HiSeq 2000 platform (500 bp library, 90 bp paired-end reads), resulting in at least 30X haploid coverage. Genome was aligned to hg19 reference genome with Burrows-Wheeler Aligner (H. Li and Durbin, 2009). SNPs were detected by SOAPsnp (R. Li et al., 2009), small insertion/deletions were detected by SAMtools (H. Li et al., 2009), and structure variants were identified by BreakDancer/CREST (Chen et al., 2009). ANNOVAR (Wang et al., 2010) was used to annotate the variant results. Bioinformatics was performed at BGI. The whole linked region (chr22: 23,761,389-24,488,587) was visualized and manually inspected by Integrative Genomics Viewer (Robinson et al., 2011; Thorvaldsdottir et al., 2013).

5 RESULTS AND DISCUSSION

5.1 Linkage analysis (I)

Study I included two families, F1 and F2, which had already been clinically described (Jokela et al., 2011). In the study of Jokela et al. (2011) linkage to all chromosomal loci for hereditary motor neuropathies known at that time had been excluded. However, without further genetic evidence it was still uncertain if the disease in the two families was the same and if it really was a new type of SMA. In the genome-wide scan of Study I, significant linkage was obtained on chromosome 22q11.2-q13.2, the maximum LOD being 3.43 at marker D22S315. This confirmed that the disease in F1 and F2 was a genetically distinct entity.

The linked region was defined by flanking markers D22S686 and D22S276, comprising 18.9 Mb. None of the 402 genes located in the linked region were previously known to be associated with SMAs. Unfortunately, most of the genes inside the linked area were poorly described, and there were no obvious candidate genes. On the other hand, because the genes known to cause motor neuron disorders have highly variable functions, practically no genes could be excluded.

The disease-associated haplotype (Figure 6A) was identical in both families. This provided evidence for a common founder mutation. The families were not known to be related, but they both originated from Northern Karelia, which belongs to the area of late settlement in Finland, populated in the 1500s and thereafter. In this region, the population started to expand only 15-25 generations ago from a small group of founders (Norio, 2003). Therefore, it was reasonable to assume that F1 and F2 should have a common ancestor, and because the disease was dominantly inherited, it was likely that more Finnish patients with the same linked haplotype could be identified.

A				B					
Marker	Location (Mb)	F1:I-16	F2:IV-3	Marker	Location (Mb)	F1:II-16	F6:III-3	F7:III-15	F12:I-2
D22S420	17,86	3 4	3 3	D22S264	20,77	3 8	7 3	4 3	7 8
D22S427	18,59	3 5	5 2	D22S446	22,02	11 11	3 0	9 4	11 11
D22S264	20,77	3 8	7 3	D22S306	22,56	3 3	3 0	3 3	3 3
D22S446	22,02	11 11	11 11	D22S686	23,07	1 5	5 1	5 6	4 5
D22S539	22,26	1 1	1 1	D22S303	23,27	2 3	3 2	3 3	4 3
D22S306	22,56	3 3	3 3	SHGC-106816	23,76	2 1	1 2	1 1	1 1
D22S686	23,07	1 5	5 5	D22S301	23,87	3 4	7 4	4 6	7 4
D22S1174	24,18	4 1	0 1	<i>CHCHD10</i>	24,11				
D22S926	25,82	2 2	3 2	D22S345	24,49	4 1	5 1	4 5	4 2
D22S419	25,94	7 10	6 10	D22S343	25,51	3 6	5 3	6 5	7 4
D22S421	25,95	8 11	8 11	D22S925	25,78	6 6	6 7	6 4	6 7
D22S315	26,02	2 3	6 3	D22S926	25,82	2 2	5 5	2 2	5 2
D22S1154	26,62	4 7	4 7	D22S419	25,94	7 10	8 4	10 4	8 8
D22S531	30,70	2 2	3 2	D22S421	25,95	8 11	8 9	11 8	9 8
D22S276	42,01	3 7	3 7						
D22S928	45,47	8 2	5 5						
D22S1170	48,35	8 3	8 3						

Figure 6. Disease-associated haplotype. **A.** The founder haplotype identified in families F1 and F2 in Study I. The markers shown in boldface were part of the deCODE 536 microsatellite marker set. **B.** Informative recombinations found in Study II (F6:III-3 and F7:III-15) and Study III (F12:I-2). F7:III-15 is unaffected, other individuals presented are affected. The founder haplotype is shown in gray.

5.2 Identification of new families and refinement of the linked region (II)

The disease-associated haplotype identified in Study I enabled a screening study based on a founder effect hypothesis. In Study II, genotyping of the linked area was performed for a cohort of patients with late-onset SMA or undetermined lower motor neuron disease. Altogether 26 patients from nine new families were identified to carry the disease-associated haplotype. The expanded cohort of SMAJ patients enabled more accurate determination of the phenotype and diversification of the characteristics of SMAJ disease. At this point, the number of SMAJ patients was already 37, indicating that the prevalence of SMAJ could be quite high in Finland. Our results suggested that the point prevalence of SMAJ in Northern Karelia (12:100 000 with these patient numbers) already exceeded that of ALS.

The identification of new patients confirmed the founder hypothesis. Especially in diseases belonging to the Finnish disease heritage, strong linkage disequilibrium between disease gene and nearby markers is a rule (Norio, 2003), which seems to be the case in SMAJ as well. The results of Study II showed that the founder haplotype could be used to genetically identify SMAJ patients, at least when enough family members are available to confirm the segregation of the markers. A genotyping test for SMAJ disease based on the founder haplotype was indeed used in clinical diagnostics for some time at

Tampere Neuromuscular Research Center before the disease-causing gene was identified.

There were significant recombinants in the new patient cohort (F6:III-3 and F7:III-15, Figure 6B), which restricted the disease locus by 90% to 1.8 Mb. The refined region was located between markers SHGC-106816 and D22S343 and included 55 genes. To identify the disease-causing gene, we performed WES using the samples of four affected patients and two healthy family members. Surprisingly, however, WES did not reveal any potentially pathogenic variants that would have segregated with the disease. This was due to very low coverage at the linked region; over 60% of exons were covered poorly (< 10X) or not at all. Retrospectively, this kind of methodological limitation of WES was not uncommon at that time.

After unsuccessful WES, we selected two candidate genes with the most potential, *SNRPD3* and *SGSM1*, for Sanger sequencing. *SNRPD3* encodes for small nuclear ribonucleoprotein Sm D3 (snRNP Sm-D3), which interacts with SMN (Friesen and Dreyfuss, 2000). *SGSM1*, in turn, is thought to modulate the small G protein RAP and RAB-mediated neuronal signal transduction and vesicular transport pathways (Yang et al., 2007). Both are thus related to neuronal functions, and mutations in these genes could cause neurological disease. However, no mutations in either *SNRPD3* or *SGSM1* were found, and the disease-causing gene was left unidentified. Diseases caused by mutations in *SNRPD3* or *SGSM1* are still not known.

5.3 Identification of the disease-causing mutation (III)

By the time of Study III, we had identified altogether 55 SMAJ patients from 17 families by genotyping. In family F12, a new informative recombination was observed (Figure 6B) that reduced the linked region to 727 kb between markers SHGC-106816 and D22S345. Because WES in Study II had failed and there were no more good candidate genes left, we performed WGS for one patient (F1:II-16). Inspection of the coding areas of the linked region revealed two heterozygous candidate mutations that were absent from the Exome Variant Server and 1000 Genomes databases. These were a silent mutation c.728C>T in *SLC2A11* and a missense mutation c.197G>T p.G66V in *CHCHD10*.

Sanger sequencing of exon 8 in *SLC2A11* and exon 2 in *CHCHD10* verified that both variants seen in WGS were true findings, but only c.197G>T p.G66V in *CHCHD10* segregated with the disease. The pathogenicity of c.197G>T p.G66V was supported by the finding that it was present in all 55 SMAJ patients and absent from unaffected family members as well as from 104 Finnish normal

population control samples. The mutated glycine residue G66 is completely conserved in *CHCHD10* orthologs, suggesting that it is intolerant to mutations.

During Study III the first disease associated with *CHCHD10* was reported. Dominant mutation c.176C>T p.S59L was described to cause FTD-ALS with mitochondrial myopathy (Bannwarth et al., 2014). Soon afterwards, *CHCHD10* allele c.[43C>A;172G>C] p.[R15S;G58R] was reported to cause mitochondrial myopathy (Ajroud-Driss et al., 2015), and mutation c.44C>A p.R15L was described to be associated with ALS (Johnson et al., 2014; Müller et al., 2014). Also c.197G>T p.G66V was reported by Müller et al. (2014) in one of our Finnish patients described as having slowly ascending progressive motor neuron disease. The phenotype of this patient (F16:III-1) was described in detail in Study III. Expression studies showed that cells expressing *CHCHD10*^{S59L}, *CHCHD10*^{R15S/G58R}, and *CHCHD10*^{G58R} undergo fragmentation of the mitochondria (Bannwarth et al., 2014; Ajroud-Driss et al., 2015). Because amino acid G66 resides at the same evolutionarily conserved region encoding for a highly hydrophobic helix as S59 and G58, these results supported the pathogenic role of the c.197G>T p.G66V mutation.

Functional studies of the pathogenic mechanism of the p.G66V mutation were not performed until very recently (Brockmann et al., 2018). In their study using patient fibroblasts, Brockmann et al. (2018) showed that with regard to the *CHCHD10* p.G66V mutant, altered secondary structure and rapid protein degradation were observed. Their results further indicated that p.G66V, as well as p.R15L, causes a reduction of *CHCHD10* protein levels in mutant patient fibroblasts to approximately 50%, and hence, their pathogenicity was suggestive of haploinsufficiency of *CHCHD10*. However, in muscle biopsies of SMAJ patients, mitochondrial *CHCHD10* was shown to be normally expressed in immunohistochemistry, and mitochondria were also ultrastructurally normal (Jokela et al., 2016).

The results of Study III show that heterozygous mutation c.197G>T p.G66V in *CHCHD10* is the cause of SMAJ. *CHCHD10* was the first SMA-causing gene identified that codes for a mitochondrial protein. The phenotype of SMAJ differs significantly from other reported *CHCHD10*-related phenotypes (FTD-ALS with mitochondrial myopathy, mitochondrial myopathy, ALS), showing that, as with other SMA-causing genes, also *CHCHD10* may cause highly variable phenotypes.

5.4 Prevalence and clinical outcome of *CHCHD10* mutations in Finland (IV)

The results of Study III suggested that the prevalence of SMAJ in Finland is considerable. On the other hand, it seemed possible that mutation c.197G>T

p.G66V could cause other phenotypes besides SMAJ. Already one study had reported c.197G>T p.G66V to cause CMT2 (Auranen et al., 2015), although three patients in that study had already been published in our previous studies (F1:III-27, F14:II-1, F14:II-2) and their phenotype was typical SMAJ without sensory abnormalities compatible with CMT2. In another family, c.197G>T p.G66V had been reported to cause intrafamilial clinical variability ranging from SMAJ to CMT2 (Pasanen et al., 2016). Study IV was performed in order to clarify the presence of different *CHCHD10* mutations as well as their phenotypic manifestations in Finland.

In Study IV, we screened several cohorts of Finnish patients with distinct neurogenic disorders and one cohort of mitochondrial myopathy, a total of 336 patients with undetermined disease. The mutation c.197G>T p.G66V was found in 23 patients, all of whom had a phenotype restricted to SMAJ. Together with the results of Study IV, the number of all reported patients carrying c.197G>T p.G66V was now 93, suggesting that c.197G>T p.G66V is a common mutation in Finland.

In the Sequencing Initiative Suomi database (SISu, 25.1.2018) the variant c.197G>T p.G66V was present in 4/9415 individuals. Thus, allele frequency of the mutation would be 0.000212427 in the Finnish population. Notably, all four mutation carriers in SISu are from Northern Karelia. In SISu, the same person might appear multiple times and persons may be related to each other. Therefore, the allele frequency of c.197G>T p.G66V may be overestimated. On the basis of our determined patients and the SISu frequency, we estimate the prevalence of c.197G>T p.G66V to be around 4/100 000 in Finland. Considering the age of onset, this prevalence suggests altogether some 200 symptomatic SMAJ patients in Finland.

No other possibly pathogenic mutations were found in any of the cohorts, not even in those 103 patients for whom the whole coding region of *CHCHD10* was sequenced. These results support the impression that pathogenic mutations in *CHCHD10* are concentrated in exon 2. Only one mutation in exon 1 and three mutations in exon 3 have been reported (Dols-Icardo et al., 2015; Perrone et al., 2017; Q. Zhou et al., 2017), but their pathogenicity is unclear. According to our results, possible other *CHCHD10* mutations, if present at all, have to be extremely rare in the Finnish population.

There is growing evidence that dominantly inherited mutations in SMA-causing genes may also cause CMT2. At least mutations in *HSPB1*, *HSPB8*, and *DYNC1H1* do not cause distinct entities but a continuum of phenotypes ranging from SMA to CMT2 (Solla et al., 2010; Hoffman and Talbot, 2012; Bombelli et al., 2014). It is notable that *MFN2*, a well-described mitochondrial protein-coding gene, also causes very heterogeneous clinical features. Mutations in *MFN2* primarily cause type 2A CMT, but a continuum between

CMT2 and dHMN has been suggested (Bombelli et al., 2014). These results are supported by a recent study by Bansagi et al. (2017) who suggest that dHMN should be included as a subcategory of CMT. It is debatable whether all patients carrying c.197G>T p.G66V have typical SMAJ or whether the phenotype extends into the category of CMT2. Nevertheless, SMAJ is an important differential diagnostic alternative in patients with a neurogenic disorder suggestive of CMT2 or even ALS because patients with SMAJ have received both of these diagnoses before the final molecular diagnosis.

In general, *CHCHD10*-related disorders seem to be rare worldwide (Table 6). Of the ascertained *CHCHD10*-related disorders, SMAJ is by far the most frequent. This is due to the enrichment of the founder mutation in the Finnish population. Based on the genotype studies, it is evident that all Finnish patients carrying the mutation c.197G>T p.G66V have a common ancestry, probably of Northern Karelian origin (Figure 7). DMLE+ Linkage disequilibrium mapping software (Rannala and Bertorelle, 2001) estimates that our genotyped families had a common ancestor approximately 4-5 generations ago (unpublished data). This estimation is not very accurate, however, because the allele frequencies of the employed markers in the normal population are not known. In any case, it is possible that c.197G>T p.G66V is a relatively recent mutation in the Finnish population.

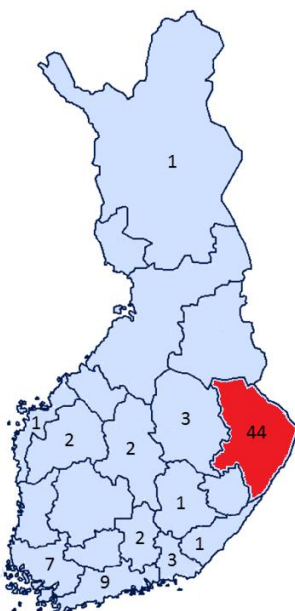


Figure 7. Number of SMAJ patients by Finnish hospital districts found in Studies I-IV (total 78). Two patients were living in Sweden and are not shown on the map. Northern Karelia is indicated in red.

5.5 Limitations and methodological considerations

Although the goals of this work were achieved, there are some limitations of the studies. Considering the number of patients and families available it took a very long time to identify the disease-causing gene. One reason for this was the lack of informative recombinations. The recombination in family F12 that restricted the linked region to 727 kb and 22 genes was only identified after Study II. This hindered the selection of potential candidate genes for sequencing. On the other hand, *CHCHD10* was reported to cause neurological disease only after the SMAJ-causing mutation was identified. Besides, *CHCHD10* was the first SMA-causing gene known to encode for a mitochondrial protein, and based on its function it would have been an unlikely candidate, no matter how short the gene list would have been.

From the WES data, the disease-causing mutation should have been easily found, especially with the number of patients that we used (four affected, two unaffected). However, the quality of our WES data was too poor so that most of the linked area was insufficiently covered. At the time of Study II (in 2012-2013), WES was still a new method and the quality of produced data was highly variable. It was unfortunate that the quality of our WES run was low. However, according to gnomAD *CHCHD10* is still a difficult gene for WES so that even with the present WES methods the coverage of *CHCHD10* is often too low. Possibly we might have identified c.197G>T p.G66V by rerunning WES or adding more sequencing cycles to the original WES, but WGS was probably a more certain method.

Another shortcoming of this study was that functional studies for mutation c.197G>T p.G66V were not performed. Because *CHCHD10* was not a known disease-causing gene when we identified the mutation, we planned for *in vivo* studies in zebrafish. However, before the zebrafish studies were started the first *CHCHD10*-related disease was published. The first publication was followed by numerous letters reporting new *CHCHD10* mutations, which made us hasten the publication of our finding and abandon the planned zebrafish studies. Hence, according to the guidelines of ACMG and AMP the evidence for the pathogenicity of c.197G>T p.G66V is not yet classified as very strong (Richards et al., 2015). The final evidence for the pathogenicity as well as solving the pathomechanism of c.197G>T p.G66V were left for future studies.

6 CONCLUSIONS AND FUTURE PROSPECTS

This study showed that SMAJ is a genetically distinct previously unknown entity, caused by a dominant mutation c.197G>T p.G66V in *CHCHD10*. According to genotyping results c.197G>T p.G66V is a founder mutation in Finland, all SMAJ patients having common ancestry. SMAJ fulfills the criteria of being part of the Finnish disease heritage; it is a rare hereditary disease that is relatively common in Finland and according to present knowledge, totally absent from other populations. The origin of SMAJ seems to be in Northern Karelia, suggesting that the mutation was introduced in the Finnish population quite recently. Genealogical studies might refine the estimation of the age of the mutation.

In Finland, with possibly as many as 200 patients, SMAJ may be the most common form of SMA, its prevalence exceeding even that of *SMN1*-related SMA. This study enabled genetic testing of SMAJ, which is of great importance for the patients in order to provide them with an accurate diagnosis and prognosis. Many SMAJ patients had initially received a diagnosis of ALS, which carries a much grimmer prognosis than SMAJ. At the Tampere Neuromuscular Research Center, 101 SMAJ patients had been genetically diagnosed by January 2018 (unpublished data). This number represents only a proportion of the Finnish SMAJ patients because genetic testing for *CHCHD10* is nowadays offered also in other laboratories and totally new families are still identified. Thus, SMAJ seems to be the most common *CHCHD10*-related disease in the world.

Notably, also one patient homozygous for c.197G>T p.G66V has been identified (unpublished data). The phenotype of the homozygous patient is very similar to SMAJ, albeit much more severe. The onset of the disease was already in childhood and at the age of 36 years the patient became wheel-chair bound. The parents of the homozygous patient both belong to known SMAJ families and have a typical SMAJ phenotype (Manu Jokela, personal communication). This patient is the first individual globally known to carry confirmed pathogenic *CHCHD10* mutations on both alleles.

To date, SMAJ seems to be an exclusively Finnish disease. Two patients of this study live in Sweden and it is very possible that there are more SMAJ patients of Finnish descent outside Finland. Also non-Finnish SMAJ patients may exist, but studies performed thus far have not been able to identify them. According to gnomAD (25.1.2018), mutation c.197G>T p.G66V has no frequency in non-Finnish populations. However, population databases do not contain data from all populations and, for example, the Russian population is underrepresented. This is unfortunate because it is likely that c.197G>T p.G66V is present in Russian Karelia.

Although c.197G>T p.G66V has not been seen in non-Finnish populations, a mutation in an adjacent nucleotide affecting the same amino acid, c.196G>A p.G66S, has been detected in eleven European (non-Finnish) individuals (total allele frequency 0.00004325). The potential clinical phenotype related to c.196G>A p.G66S would be important to investigate.

Mutations in *CHCHD10* seem to have relatively strict genotype-phenotype correlations. However, most studies have concentrated on patients with ALS and/or FTD, and only one non-Finnish *CHCHD10* study has been performed on SMA patients (Morel et al., 2015). To confirm the phenotype-genotype correlation, more studies in patients in different cohorts of undetermined motor neuron disorders, especially CMT2, are needed.

The basic function of CHCHD10 and the mechanism of pathogenesis in *CHCHD10*-related diseases remain unclear. There are also a number of reported mutations with uncertain pathogenicity. Further familial segregation and functional studies are necessary to elucidate which mutations are truly pathogenic and how different mutations result in particular phenotypes. It is also interesting that even though CHCHD10 is a mitochondrial protein and other mutations in *CHCHD10* cause mitochondrial myopathy, no mitochondrial pathology has been observed in SMAJ patients (Jokela et al., 2016). Ongoing studies on muscle biopsy and fibroblasts of c.197G>T p.G66V homozygote are expected to shed light on the pathomechanism of SMAJ.

7 ACKNOWLEDGMENTS

This study was conducted at the Tampere Neuromuscular Research Center (TNMC) during years 2009-2016. TNMC is acknowledged for providing excellent resources and facilities for the research. My work was also mostly funded by TNMC, but additional support from the Finska Läkaresällskapet, the Liv och Hälsa Foundation, and the Jane and Aatos Erkko Foundation is gratefully acknowledged.

My deepest gratitude is owed to my supervisor and our group leader Bjarne Udd, whose enthusiasm towards neuromuscular diseases, research, and new exciting things is limitless. He gave me a fascinating research topic, in addition to inspiring and rewarding work on the genetic diagnostics of neuromuscular diseases. I have learnt a lot during these years and I feel privileged for being able to work with him.

Johanna Palmio, the co-supervisor of my PhD, is thanked for generous support and guidance. She offered me a valuable second opinion during the writing process and her efforts notably improved the manuscript.

The external reviewers of the thesis manuscript, Jaakko Ignatius and Irma Järvelä, are warmly thanked for all of their efforts and sharing their scientific expertise with me. Carol Ann Pelli is acknowledged for her highly careful revision of the language of this thesis. I am also grateful to Pekka Heino and Ritva Karhu, the external members of my thesis advisory committee, for accepting the task and for always being so helpful and supportive. I also thank Ville Mustonen for agreeing to act as a custos for my defence. Ulla Pirvola is thanked for participating the grading committee.

All of the co-authors of the original publications are thanked for their contributions to the projects. Anna Maija Saukkonen and Jari Toivanen are thanked for providing the invaluable patient material that made this work possible. Peter Hackman is thanked for starting me off on the genotyping project. I am especially grateful for Manu Jokela for being responsible for the clinical part of this project, thereby making this work possible. His enthusiasm and genuine interest in neurological diseases provided the driving force of this project.

I thank all past and present co-workers at the Neuromuscular Research Unit. I am especially obliged to Tiina Suominen, Sara Lehtinen, Olayinka Raheem, Raija Çalışkan, Hanna-Liisa Kojo, Henna-Riikka Koskinen, Jaana

Leppikangas, Satu Luhtasela, Emilia Nuortio, Eini Penkkimäki, Eeva Tuulkari, and Veera Viljakainen who all participated in this work one way or another. Tiina and Sara are warmly thanked for invaluable peer support and for sharing the numerous highs and lows of this work. My humblest thanks go to Raija, Hanna-Liisa, Henna, Jaana, Satu, Emilia, Eini, Eeva, and Veera for their technical expertise and ever so helpful attitude towards all problems encountered. Veera is acknowledged for the extensive lab work that she did with the *CHCHD10* sequences. Yinka and also Sanna Huovinen are especially thanked not only for their collaboration, but also for taking me with them on the long and eventful combined conference/pleasure trips around the world. I also thank all of my workmates of the late Finn-Medi 3, 4th floor community, for help with practical issues and for creating the best work atmosphere imaginable. My heartfelt thanks go to each and every one of you; because of you, coming to work has always been a pleasure, despite all major and minor disasters we have confronted together.

I am also grateful to Satu Sandell as well as the Helsinki branch of our group: Meharji Arumilli, Anni Evilä, Mridul Johari, Per-Harald Jonson, Helena Luque, Jaakko Sarparanta, Marco Savarese, Merja Soininen, and Anna Vihola. Thank you for all of your help and expertise and your delightful company during WMS and other conference trips and meetings.

My thanks go also to my new work community at the Department of Clinical Genetics of Fimlab Laboratories. Thank you for having me as part of the team and making me feel so welcome. Especially Mauri Keinänen, Ritva Karhu, Virpi Laitinen, Tiina Lund-Aho and Kirsi Kaukoniemi are warmly thanked for giving me the opportunity to finish my thesis even though there was really no time for it.

Finally, I thank my family and friends for unstinting support and for taking care of life outside of work.

Tampere, May 2018

Sini Penttilä

8 ELECTRONIC TOOLS AND DATABASES

1000 Genomes, <http://www.internationalgenome.org/>

DMLE+ linkage disequilibrium mapping software, <http://dmle.org/>

Ensembl genome browser, release 91, <http://www.ensembl.org/>

Exome Variant Server, <http://evs.gs.washington.edu/EVS/>

GeneTable of Neuromuscular Disorders, <http://www.musclegenetable.fr/>

genome Aggregation Database, gnomAD, <http://gnomad.broadinstitute.org/>

Online Mendelian Inheritance in Man®, OMIM, <https://omim.org/>

RaptorX, protein structure and function prediction server,
<http://raptorx.uchicago.edu/>

Sequencing Initiative Suomi database, SISu, <http://sisuproject.fi/>

9 REFERENCES

- Abecasis, G.R., Cherny, S.S., Cookson, W.O., and Cardon, L.R. (2002). Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* *30*, 97-101.
- Ackerley, S., James, P.A., Kalli, A., French, S., Davies, K.E., and Talbot, K. (2006). A mutation in the small heat-shock protein HSPB1 leading to distal hereditary motor neuropathy disrupts neurofilament assembly and the axonal transport of specific cellular cargoes. *Hum. Mol. Genet.* *15*, 347-354.
- Ajrourd-Driss, S., Fecto, F., Ajroud, K., Lalani, I., Calvo, S.E., Mootha, V.K., Deng, H.X., Siddique, N., Tahmouh, A.J., Heiman-Patterson, T.D., and Siddique, T. (2015). Mutation in the novel nuclear-encoded mitochondrial protein CHCHD10 in a family with autosomal dominant mitochondrial myopathy. *Neurogenetics* *16*, 1-9.
- Antonellis, A., Ellsworth, R.E., Sambuughin, N., Puls, I., Abel, A., Lee-Lin, S.Q., Jordanova, A., Kremensky, I., Christodoulou, K., Middleton, L.T., *et al.* (2003). Glycyl tRNA synthetase mutations in Charcot-Marie-Tooth disease type 2D and distal spinal muscular atrophy type V. *Am. J. Hum. Genet.* *72*, 1293-1299.
- Auranen, M., Ylikallio, E., Shcherbii, M., Paetau, A., Kiuru-Enari, S., Toppila, J.P., and Tyynismaa, H. (2015). CHCHD10 variant p.(Gly66Val) causes axonal Charcot-Marie-Tooth disease. *Neurol. Genet.* *1*, e1.
- Bannwarth, S., Ait-El-Mkadem, S., Chaussonot, A., Genin, E.C., Lacas-Gervais, S., Fragaki, K., Berg-Alonso, L., Kageyama, Y., Serre, V., Moore, D.G., *et al.* (2014). A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain* *137*, 2329-2345.
- Bansagi, B., Griffin, H., Whittaker, R.G., Antoniadi, T., Evangelista, T., Miller, J., Greenslade, M., Forester, N., Duff, J., Bradshaw, A., *et al.* (2017). Genetic heterogeneity of motor neuropathies. *Neurology* *88*, 1226-1234.
- Barrett, J.H., and Teare, M.D. (2011). Linkage analysis. *Methods Mol. Biol.* *760*, 19-33.
- Bentley, D.R., Balasubramanian, S., Swerdlow, H.P., Smith, G.P., Milton, J., Brown, C.G., Hall, K.P., Evers, D.J., Barnes, C.L., Bignell, H.R., *et al.* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* *456*, 53-59.
- Blauwendraat, C., Wilke, C., Simon-Sanchez, J., Jansen, I.E., Reifschneider, A., Capell, A., Haass, C., Castillo-Lizardo, M., Biskup, S., Maetzler, W., *et al.* (2018). The wide genetic landscape of clinical frontotemporal dementia: systematic combined sequencing of 121 consecutive subjects. *Genet. Med.* *20*, 240-249.

Blumen, S.C., Astord, S., Robin, V., Vignaud, L., Toumi, N., Cieslik, A., Achiron, A., Carasso, R.L., Gurevich, M., Braverman, I., *et al.* (2012). A rare recessive distal hereditary motor neuropathy with HSN1 chaperone mutation. *Ann. Neurol.* *71*, 509-519.

Bombelli, F., Stojkovic, T., Dubourg, O., Echaniz-Laguna, A., Tardieu, S., Larcher, K., Amati-Bonneau, P., Latour, P., Vignal, O., Cazeneuve, C., Brice, A., and Leguern, E. (2014). Charcot-Marie-Tooth disease type 2A: from typical to rare phenotypic and genotypic features. *JAMA Neurol.* *71*, 1036-1042.

Brockmann, S.J., Freischmidt, A., Oeckl, P., Muller, K., Ponna, S.K., Helferich, A.M., Paone, C., Reinders, J., Kojer, K., Orth, M., *et al.* (2018). CHCHD10 mutations p.R15L and p.G66V cause motoneuron disease by haploinsufficiency. *Hum. Mol. Genet.* *27*, 706-715.

Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., and Hudson, T. (1992). Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* *68*, 799-808.

Burstein, S.R., Valsecchi, F., Kawamata, H., Bourens, M., Zeng, R., Zuberi, A., Milner, T.A., Cloonan, S.M., Lutz, C., Barrientos, A., and Manfredi, G. (2018). In vitro and in vivo studies of the ALS-FTLD protein CHCHD10 reveal novel mitochondrial topology and protein interactions. *Hum. Mol. Genet.* *27*, 160-177.

Carra, S., Sivilotti, M., Chavez Zobel, A.T., Lambert, H., and Landry, J. (2005). HspB8, a small heat shock protein mutated in human neuromuscular disorders, has in vivo chaperone activity in cultured cells. *Hum. Mol. Genet.* *14*, 1659-1669.

Cavallaro, G. (2010). Genome-wide analysis of eukaryotic twin CX9C proteins. *Mol. Biosyst* *6*, 2459-2470.

Chausse, A., Le Ber, I., Ait-El-Mkadem, S., Camuzat, A., de Septenville, A., Bannwarth, S., Genin, E.C., Serre, V., Auge, G., French research network on FTD and FTD-ALS, *et al.* (2014). Screening of CHCHD10 in a French cohort confirms the involvement of this gene in frontotemporal dementia with amyotrophic lateral sclerosis patients. *Neurobiol. Aging* *35*, 2884.e4.

Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S., McGrath, S.D., Wendl, M.C., Zhang, Q., Locke, D.P., *et al.* (2009). BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* *6*, 677-681.

Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C., and Kehrer-Sawatzki, H. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* *132*, 1077-1130.

- Darras, B.T. (2011). Non-5q spinal muscular atrophies: the alphanumeric soup thickens. *Neurology* 77, 312-314.
- Darras, B.T. (2015). Spinal muscular atrophies. *Pediatr. Clin. North Am.* 62, 743-766.
- DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., *et al.* (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245-256.
- Deng, H.X., Klein, C.J., Yan, J., Shi, Y., Wu, Y., Fecto, F., Yau, H.J., Yang, Y., Zhai, H., Siddique, N., *et al.* (2010). Scapuloperoneal spinal muscular atrophy and CMT2C are allelic disorders caused by alterations in TRPV4. *Nat. Genet.* 42, 165-169.
- Dols-Icardo, O., Nebot, I., Gorostidi, A., Ortega-Cubero, S., Hernandez, I., Rojas-Garcia, R., Garcia-Redondo, A., Povedano, M., Llado, A., Alvarez, V., *et al.* (2015). Analysis of the CHCHD10 gene in patients with frontotemporal dementia and amyotrophic lateral sclerosis from Spain. *Brain* 138, e400.
- Feldkötter, M., Schwarzer, V., Wirth, R., Wienker, T.F., and Wirth, B. (2002). Quantitative analyses of SMN1 and SMN2 based on real-time lightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. *Am. J. Hum. Genet.* 70, 358-368.
- Finkel, R.S., Mercuri, E., Darras, B.T., Connolly, A.M., Kuntz, N.L., Kirschner, J., Chiriboga, C.A., Saito, K., Servais, L., Tizzano, E., *et al.* (2017). Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy. *N. Engl. J. Med.* 377, 1723-1732.
- Friesen, W.J., and Dreyfuss, G. (2000). Specific sequences of the Sm and Sm-like (Lsm) proteins mediate their interaction with the spinal muscular atrophy disease gene product (SMN). *J. Biol. Chem.* 275, 26370-26375.
- Gemayel, R., Vinces, M.D., Legendre, M., and Verstrepen, K.J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu. Rev. Genet.* 44, 445-477.
- Genin, E.C., Plutino, M., Bannwarth, S., Villa, E., Cisneros-Barroso, E., Roy, M., Ortega-Vila, B., Fragaki, K., Lespinasse, F., Pinero-Martos, E., *et al.* (2015). CHCHD10 mutations promote loss of mitochondrial cristae junctions with impaired mitochondrial genome maintenance and inhibition of apoptosis. *EMBO Mol. Med.* 8, 58-72.
- Harms, M.B., Ori-McKenney, K.M., Scoto, M., Tuck, E.P., Bell, S., Ma, D., Masi, S., Allred, P., Al-Lozi, M., Reilly, M.M., *et al.* (2012). Mutations in the tail domain of DYNC1H1 cause dominant spinal muscular atrophy. *Neurology* 78, 1714-1720.

- Hoffman, E.P., and Talbot, K. (2012). A calm before the exome storm: coming together of dSMA and CMT2. *Neurology* 78, 1706-1707.
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature* 431, 931-945.
- Irobi, J., Dierick, I., Jordanova, A., Claeys, K.G., De Jonghe, P., and Timmerman, V. (2006). Unraveling the genetics of distal hereditary motor neuropathies. *Neuromolecular Med.* 8, 131-146.
- James, P.A., and Talbot, K. (2006). The molecular genetics of non-ALS motor neuron diseases. *Biochim. Biophys. Acta* 1762, 986-1000.
- Jiao, B., Xiao, T., Hou, L., Gu, X., Zhou, Y., Zhou, L., Tang, B., Xu, J., and Shen, L. (2016). High prevalence of CHCHD10 mutation in patients with frontotemporal dementia from China. *Brain* 139, e21.
- Johnson, J.O., Glynn, S.M., Gibbs, J.R., Nalls, M.A., Sabatelli, M., Restagno, G., Drory, V.E., Chio, A., Rogaeva, E., and Traynor, B.J. (2014). Mutations in the CHCHD10 gene are a common cause of familial amyotrophic lateral sclerosis. *Brain* 137, e311.
- Jokela, M., Penttilä, S., Huovinen, S., Hackman, P., Saukkonen, A.M., Toivanen, J., and Udd, B. (2011). Late-onset lower motor neuronopathy: A new autosomal dominant disorder. *Neurology* 77, 334-340.
- Jokela, M., Huovinen, S., Raheem, O., Lindfors, M., Palmio, J., Penttilä, S., and Udd, B. (2016). Distinct Muscle Biopsy Findings in Genetically Defined Adult-Onset Motor Neuron Disorders. *PLoS One* 11, e0151376.
- Jokela, M. (2015). Late-onset spinal motor neuronopathy - a new neuromuscular disease. (Turku, Finland: Turun yliopiston julkaisu - Annales universitatis Turkuensis).
- Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., and Xu, J. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* 7, 1511-1522.
- Kennedy, W.R., Alter, M., and Sung, J.H. (1968). Progressive proximal spinal and bulbar muscular atrophy of late onset. A sex-linked recessive trait. *Neurology* 18, 671-680.
- Kennerson, M.L., Nicholson, G.A., Kaler, S.G., Kowalski, B., Mercer, J.F., Tang, J., Llanos, R.M., Chu, S., Takata, R.I., Speck-Martins, C.E., *et al.* (2010). Missense mutations in the copper transporter gene ATP7A cause X-linked distal hereditary motor neuropathy. *Am. J. Hum. Genet.* 86, 343-352.
- King, N.M.P., and Bishop, C.E. (2017). New treatments for serious conditions: ethical implications. *Gene Ther.* 24, 534-538.

Kong, A., Gudbjartsson, D.F., Sainz, J., Jonsdottir, G.M., Gudjonsson, S.A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B., Masson, G., *et al.* (2002). A high-resolution recombination map of the human genome. *Nat. Genet.* *31*, 241-247.

Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* *23*, 1289-1291.

Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., and Lander, E.S. (1996). Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* *58*, 1347-1363.

Kurzweily, D., Kruger, S., Biskup, S., and Heneka, M.T. (2015). A distinct clinical phenotype in a German kindred with motor neuron disease carrying a CHCHD10 mutation. *Brain* *138*, e376.

La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E., and Fischbeck, K.H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* *352*, 77-79.

Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., and Zeviani, M. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* *80*, 155-165.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* *25*, 2078-2079.

Li, R., Li, Y., Fang, X., Yang, H., Wang, J., Kristiansen, K., and Wang, J. (2009). SNP detection for massively parallel whole-genome resequencing. *Genome Res.* *19*, 1124-1132.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* *293*, 864-867.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* *2012*, 251364.

Liu, Q., Fischer, U., Wang, F., and Dreyfuss, G. (1997). The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell* *90*, 1013-1021.

- Mah, J.K. (2018). An Overview of Recent Therapeutics Advances for Duchenne Muscular Dystrophy. *Methods Mol. Biol.* 1687, 3-17.
- Mendell, J.R., Al-Zaidy, S., Shell, R., Arnold, W.D., Rodino-Klapac, L.R., Prior, T.W., Lowes, L., Alfano, L., Berry, K., Church, K., *et al.* (2017). Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med.* 377, 1713-1722.
- Monani, U.R., Lorson, C.L., Parsons, D.W., Prior, T.W., Androphy, E.J., Burghes, A.H., and McPherson, J.D. (1999). A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.* 8, 1177-1183.
- Morel, G., Rouzier, C., Chaussonot, A., Ait-El-Mkadem, S., Bannwarth, S., Genin, E.C., Auge, G., Chabrol, B., Pouget, J., Soriani, M.H., Sacconi, S., and Paquis-Flucklinger, V. (2015). CHCHD10 mutations are not a common cause of SMN1-negative type III/IV spinal motor atrophy. *Ann. Neurol.* 78, 831.
- Morton, N.E. (1955). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7, 277-318.
- Müller, K., Andersen, P.M., Hübers, A., Marroquin, N., Volk, A.E., Danzer, K.M., Meitinger, T., Ludolph, A.C., Strom, T.M., and Weishaupt, J.H. (2014). Two novel mutations in conserved codons indicate that CHCHD10 is a gene associated with motor neuron disease. *Brain* 137, e309.
- Neveling, K., Martinez-Carrera, L.A., Holker, I., Heister, A., Verrips, A., Hosseini-Barkooie, S.M., Gilissen, C., Vermeer, S., Pennings, M., Meijer, R., *et al.* (2013). Mutations in BICD2, which encodes a golgin and important motor adaptor, cause congenital autosomal-dominant spinal muscular atrophy. *Am. J. Hum. Genet.* 92, 946-954.
- Nguyen, L., and Burnett, L. (2014). Automation of molecular-based analyses: a primer on massively parallel sequencing. *Clin. Biochem. Rev.* 35, 169-176.
- Norio, R., Nevanlinna, H.R., and Perheentupa, J. (1973). Hereditary diseases in Finland; rare flora in rare soul. *Ann. Clin. Res.* 5, 109-141.
- Norio, R. (2003). Finnish Disease Heritage I: characteristics, causes, background. *Hum. Genet.* 112, 441-456.
- Oates, E.C., Rossor, A.M., Hafezparast, M., Gonzalez, M., Speziani, F., MacArthur, D.G., Lek, M., Cottenie, E., Scoto, M., Foley, A.R., *et al.* (2013). Mutations in BICD2 cause dominant congenital spinal muscular atrophy and hereditary spastic paraplegia. *Am. J. Hum. Genet.* 92, 965-973.
- Ott, J., Wang, J., and Leal, S.M. (2015). Genetic linkage analysis in the age of whole-genome sequencing. *Nat. Rev. Genet.* 16, 275-284.

- Pasanen, P., Myllykangas, L., Pöyhönen, M., Kiuru-Enari, S., Tienari, P.J., Laaksovirta, H., Toppila, J., Ylikallio, E., Tyynismaa, H., and Auranen, M. (2016). Intrafamilial clinical variability in individuals carrying the CHCHD10 mutation Gly66Val. *Acta Neurol. Scand.* *133*, 361-366.
- Peeters, K., Litvinenko, I., Asselbergh, B., Almeida-Souza, L., Chamova, T., Geuens, T., Ydens, E., Zimon, M., Irobi, J., De Vriendt, E., *et al.* (2013). Molecular defects in the motor adaptor BICD2 cause proximal spinal muscular atrophy with autosomal-dominant inheritance. *Am. J. Hum. Genet.* *92*, 955-964.
- Peeters, K., Chamova, T., and Jordanova, A. (2014). Clinical and genetic diversity of SMN1-negative proximal spinal muscular atrophies. *Brain* *137*, 2879-2896.
- Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* *95*, 615-624.
- Perrone, F., Nguyen, H.P., Van Mossevelde, S., Moisse, M., Sieben, A., Santens, P., De Bleecker, J., Vandenbulcke, M., Engelborghs, S., Baets, J., *et al.* (2017). Investigating the role of ALS genes CHCHD10 and TUBA4A in Belgian FTD-ALS spectrum patients. *Neurobiol. Aging* *51*, 177.e16.
- Puls, I., Jonnakuty, C., LaMonte, B.H., Holzbaur, E.L., Tokito, M., Mann, E., Floeter, M.K., Bidus, K., Drayna, D., Oh, S.J., *et al.* (2003). Mutant dynactin in motor neuron disease. *Nat. Genet.* *33*, 455-456.
- Rampelt, H., Zerbes, R.M., van der Laan, M., and Pfanner, N. (2017). Role of the mitochondrial contact site and cristae organizing system in membrane architecture and dynamics. *Biochim. Biophys. Acta* *1864*, 737-746.
- Rannala, B., and Bertorelle, G. (2001). Using linked markers to infer the age of a mutation. *Hum. Mutat.* *18*, 87-100.
- Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., *et al.* (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* *72*, 257-268.
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., *et al.* (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* *17*, 405-424.
- Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24-26.

Ronchi, D., Riboldi, G., Del Bo, R., Ticozzi, N., Scarlato, M., Galimberti, D., Corti, S., Silani, V., Bresolin, N., and Comi, G.P. (2015). CHCHD10 mutations in Italian patients with sporadic amyotrophic lateral sclerosis. *Brain* 138, e372.

Rossor, A.M., Kalmar, B., Greensmith, L., and Reilly, M.M. (2012). The distal hereditary motor neuropathies. *J. Neurol. Neurosurg. Psychiatry*. 83, 6-14.

Rudnik-Schöneborn, S., Berg, C., Zerres, K., Betzler, C., Grimm, T., Eggermann, T., Eggermann, K., Wirth, R., Wirth, B., and Heller, R. (2009). Genotype-phenotype studies in infantile spinal muscular atrophy (SMA) type I in Germany: implications for clinical trials and genetic counselling. *Clin. Genet*. 76, 168-178.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463-5467.

Shen, S., He, J., Tang, L., Zhang, N., and Fan, D. (2017). CHCHD10 mutations in patients with amyotrophic lateral sclerosis in Mainland China. *Neurobiol. Aging* 54, 214.e10.

Solla, P., Vannelli, A., Bolino, A., Marrosu, G., Coviello, S., Murru, M.R., Tranquilli, S., Corongiu, D., Benedetti, S., and Marrosu, M.G. (2010). Heat shock protein 27 R127W mutation: evidence of a continuum between axonal Charcot-Marie-Tooth and distal hereditary motor neuropathy. *J. Neurol. Neurosurg. Psychiatry*. 81, 958-962.

Stevens, A.J., Taylor, M.G., Pearce, F.G., and Kennedy, M.A. (2017). Allelic Dropout During Polymerase Chain Reaction due to G-Quadruplex Structures and DNA Methylation Is Widespread at Imprinted Human Loci. *G3 (Bethesda)* 7, 1019-1025.

Strachan, T., and Read, A.P. (2011). *Human molecular genetics* (New York, NY: Garland Science).

Straub, I.R., Janer, A., Weraarpachai, W., Zinman, L., Robertson, J., Rogaeva, E., and Shoubridge, E.A. (2018). Loss of CHCHD10-CHCHD2 complexes required for respiration underlies the pathogenicity of a CHCHD10 mutation in ALS. *Hum. Mol. Genet.* 27, 178-189.

Stuurman, N., Heins, S., and Aebi, U. (1998). Nuclear lamins: their structure, assembly, and interactions. *J. Struct. Biol.* 122, 42-66.

Sugarman, E.A., Nagan, N., Zhu, H., Akmaev, V.R., Zhou, Z., Rohlf, E.M., Flynn, K., Hendrickson, B.C., Scholl, T., Sirko-Osadsa, D.A., and Allitto, B.A. (2012). Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: clinical laboratory analysis of >72,400 specimens. *Eur. J. Hum. Genet.* 20, 27-32.

Terwilliger, J.D., Ding, Y., and Ott, J. (1992). On the relative importance of marker heterozygosity and intermarker distance in gene mapping. *Genomics* 13, 951-956.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14, 178-192.

Tsai, P.C., Soong, B.W., Mademan, I., Huang, Y.H., Liu, C.R., Hsiao, C.T., Wu, H.T., Liu, T.T., Liu, Y.T., Tseng, Y.T., *et al.* (2017). A recurrent WARS mutation is a novel cause of autosomal dominant distal hereditary motor neuropathy. *Brain* 140, 1252-1266.

Turnpenny, P.D., Ellard, S., and Emery, A.E.H. (2005). *Emery's elements of medical genetics* (Edinburgh: Elsevier).

Udd, B., Juvonen, V., Hakamies, L., Nieminen, A., Wallgren-Pettersson, C., Cederquist, K., and Savontaus, M.L. (1998). High prevalence of Kennedy's disease in Western Finland -- is the syndrome underdiagnosed? *Acta Neurol. Scand.* 98, 128-133.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 40, e115.

Van Den Bosch, L., and Timmerman, V. (2006). Genetics of motor neuron disease. *Curr. Neurol. Neurosci. Rep.* 6, 423-431.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164.

Watson, J.D., and Crick, F.H. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature* 171, 964-967.

Wijsman, E.M., Rothstein, J.H., and Thompson, E.A. (2006). Multipoint linkage analysis with many multiallelic or dense diallelic markers: Markov chain-Monte Carlo provides practical approaches for genome scans on general pedigrees. *Am. J. Hum. Genet.* 79, 846-858.

Woo, J.A., Liu, T., Trotter, C., Fang, C.C., De Narvaez, E., LePochat, P., Maslar, D., Bukhari, A., Zhao, X., Deonaraine, A., Westerheide, S.D., and Kang, D.E. (2017). Loss of function CHCHD10 mutations in cytoplasmic TDP-43 accumulation and synaptic integrity. *Nat. Commun.* 8, 15558.

Xiao, T., Jiao, B., Zhang, W., Pan, C., Wei, J., Liu, X., Zhou, Y., Zhou, L., Tang, B., and Shen, L. (2017). Identification of CHCHD10 Mutation in Chinese Patients with Alzheimer Disease. *Mol. Neurobiol.* 54, 5243-5247.

Yang, H., Sasaki, T., Minoshima, S., and Shimizu, N. (2007). Identification of three novel proteins (SGSM1, 2, 3) which modulate small G protein (RAP and RAB)-mediated signaling pathway. *Genomics* 90, 249-260.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G., *et al.* (2018). Ensembl 2018. *Nucleic Acids Res.* *46*, D761.

Zerres, K., and Rudnik-Schöneborn, S. (2003). 93rd ENMC international workshop: non-5q-spinal muscular atrophies (SMA) - clinical picture (6-8 April 2001, Naarden, The Netherlands). *Neuromuscul. Disord.* *13*, 179-183.

Zhang, M., Xi, Z., Zinman, L., Bruni, A.C., Maletta, R.G., Curcio, S.A., Rainero, I., Rubino, E., Pinessi, L., Nacmias, B., *et al.* (2015). Mutation analysis of CHCHD10 in different neurodegenerative diseases. *Brain* *138*, e380.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* *133*, 585-600.

Zhou, Q., Chen, Y., Wei, Q., Cao, B., Wu, Y., Zhao, B., Ou, R., Yang, J., Chen, X., Hadano, S., and Shang, H.F. (2017). Mutation Screening of the CHCHD10 Gene in Chinese Patients with Amyotrophic Lateral Sclerosis. *Mol. Neurobiol.* *54*, 3189-3194.

Zhou, X., Liu, Z., Guo, J., Sun, Q., Xu, Q., Yan, X., Tang, B., and Lei, L. (2018). Identification of CHCHD10 variants in Chinese patients with Parkinson's disease. *Parkinsonism Relat. Disord.* *47*, 96-97.