Aus dem Physiologischen Institut
Lehrstuhl: Physiologie - zelluläre Physiologie (im BioMedizinischen Centrum München)
der Ludwig-Maximilians-Universität München
Vorstand: Frau Prof. Dr. Claudia Veigel

Mapping the Chromosomal Locus of Oculopharyngodistal Myopathy with Microsatellite Markers and Next Generation Sequencing

Dissertation
zum Erwerb des Doktorgrades der Medizin
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von

Matias Wagner

aus München

2018

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Herr Prof. Dr. med. Michael Meyer
Mitberichterstatter:	Herr Prof. Dr. med. Benedikt Schoser
	Herr Prof. Dr. med. Günter Rudolph
	Herr Prof. Dr. med. Florian Heinen
	Frau Prof. Dr. med. Ortrud Steinlein
Mitbetreuung durch den promovierten Mitarbeiter	Herr Prof. Dr. med. Hanns Lochmüller
Dekan	Herr Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung	28.06.2018

meiner Familie

Acknowledgements

I wish to thank Prof Michael Meyer for the supervision of my thesis. I very much appreciate the time and energy he spent on my work and the thoroughness with which he corrected the manuscript.

Also, I'm deeply obliged to Prof Hanns Lochmüller who supervised my work as a PI in Newcastle and who very much took care of my clinical and scientific education and who gave invaluable advice for my career plannings.

For patient and highly competent advice and guidance as well as the amount of time I have been pestering him with questions I want to thank Dr Steven Laval. I am also obliged to Dr Juliane Müller who taught me everything I needed to know in the lab and who is also responsible that I had such a great time in Newcastle.

Additionally, I thank Jon Ingledew, Dr Debbie Hicks, Catherine Jepson, Liz Greally, Dr Helen Griffin, Dr Alison Blain, Dr Alasdair Wood, Dr Morten Ritso and Dr Kevin Gillinder from the Institute of Genetic Medicine for scientific advice, help in trouble shooting and simply for having me in their lab.

Special thanks belong to Christine Weidhaas, not only for a great time in England, but also for her extraordinary help with the linguistic editing of this thesis.

Finally I want to thank my family, my mother Irmi, my father Ernst, my brother Sebastian and my sister Stephanie for enabling my university studies and who are always there for me.

Contents

List of Figures

List of Tables

List of Abbreviations

ACMG	American College of Medical Genetics and Genomics, page 11.
AD	Autosomal dominant, page 58.
AR	Autosomal recessive, page 58.
ATP	Adenosine triphosphate, page 56.
BWA	Burrows-Wheeler aligner, page 20.
CCD	Central core disease, page 61.
CCD	Charge coupled device, page 6.
CGH	Comparative genomic hybridization, page 49.
CK	Creatine kinase, page 1.
cM	Centimorgan, page 19.
CNV	Copy number variation, page 52.
DNA	Deoxyribonucleic acid, page 4.
dNTP	Deoxynucleoside triphosphate acid, page 13.
DOS	Disk operating system, page 17.
EDTA	Ethylenediaminetetraacetic acid, page 13.
EST	Expression sequence tag, page 48.
ExAC	Exome aggregation consortium, page 56.
GATK	Genome analysis toolkit, page 8.
GERP	Genomic evolutionary rate profiling, page 9.
GTEx	Genotype-tissue expression, page 44.
HPA	Human protein atlas, page 44.
ID	Identifier, page 27.
indel	insertion or deletion, page 8.
LGMD	Limb-girdle muscular dystrophy, page 52.
LMU	Ludwig-Maximilians-Universität, page 13.
LOD	Logarithm of odds, page 17.
LoF	Loss of function, page 61.
MAF	Minor allele frequency, page 38.
MFM	Myofibrillar myopathy, page 64.
MmD	Multi-minicore disease, page 61.
MUP	Muscle unit potentials, page 2.
NCBI	National Center for Biotechnology Information, page 29.

OMIM	Online Mendelian Inheritance in Man, page 3.
OPDM	Oculopharyngodistal myopathy, page 1.
OPMD	Oculopharyngeal muscular dystrophy, page 1.
PCR	Polymerase chain reaction, page 4.
pLI	Probability of LoF intolerance, page 61.
RNA	Ribonucleic acid, page 27.
S	Svedberg, page 53.
SNP	Single nucleotide polymorphism, page 7.
SNV	Single nucleotide variant, page 20.
STS	Sequence tagged site, page 18.
TAE	Tris-acetate-EDTA, page 15.
TEMED	Tetramethylethylenediamine, page 13.
UTR	Untranslated region, page 9.
VCPDM	Vocal cord and pharyngeal dysfunction with distal myopathy, page 42.
WES	Whole exome sequencing, page 3.
WGS	Whole genome sequencing, page 52.

1. Introduction

1.1. Oculopharyngodistal Myopathy

Oculopharyngodistal Myopathy (OPDM, MIM #164310) was first described by Satayoshi and Kinoshita in 1977 as an autosomal dominant muscle condition with onset in late adulthood and symptoms of ptosis, slowly progressive dysphagia and predominantly distal limb-girdle muscle weakness [1]. Although most reported families show an autosomal dominant inheritance, patients from Japan [2], the Netherlands [3], China [4] and Turkey [5] are reported with an autosomal recessive inheritance. To this day, 82 patients from 34 unrelated families of different ethnic backgrounds have been reported, with an apparent accumulation of cases in Japan and Turkey. There had been a discussion, whether OPDM is a clinicopathological distinct entity or a variant of oculopharyngeal muscular dystrophy (OPMD; MIM #164300) [6] with the majority of authors claiming, that OPDM is clinically distinguishable from similar muscle conditions. Until the causative genetic changes are found, the diagnosis has to be made based on clinical and histopathological changes as well as on the exclusion of similar genetic disorders. Furthermore, until the genetic cause or causes are found, it is not certain if variations in more than one gene are related to a single distinct phenotype we call OPDM today.

1.1.1. Clinical Symptoms

There is great variability for the age of onset of OPDM, ranging from 7 to 66 years. In general, Japanese patients tend to develop first symptoms much later than individuals from other countries. Usually, a bilateral ptosis is the initial symptom, followed by swallowing difficulties, dysarthria, ophthalmoparesis and predominantly distal limb-girdle muscle weakness. Chinese Patients, however, usually present first with distal muscle impairment [4]. About half of the published cases show respiratory muscle involvement with some of them needing nocturnal non-invasive positive pressure ventilation. Only two reported patients exhibited cardiac involvement - common in other muscle conditions such as muscular dystrophies, myofibrillar myopathies, congenital myopathies and metabolic myopathies - namely hypertrabeculation and myocardial thinning [7], [8]. CK-levels in those cases are normally slightly increased ranging from normal to eight-fold of the upper limit [5]. Almost all examined individuals showed nonspecific myopathic changes in electromyography such as reduced amplitude and duration of muscle unit potentials (MUPs) due to the reduced number of functioning muscle fibres. Some publications have reported myotonic discharges [5], [9], [4], [10].

1.1.2. Histopathological findings

Light microscopy of all patients reported so far showed myopathic changes such as fibre size variation, angulated fibres, internal nuclei, interstitial fibrosis and fatty connective tissues as shown in Figure 1.1. Rimmed vacuoles, seen in both fibre types and appearing with a red margin in Gomori-Trichrome staining, seem to have a high sensitivity [6], [5] but low specificity in diagnosing OPDM. They can also be present in similar muscle conditions like OPMD or inclusion body myositis [11]. Ragged-red fibres or signs of inflammation were not seen in any of the patients.

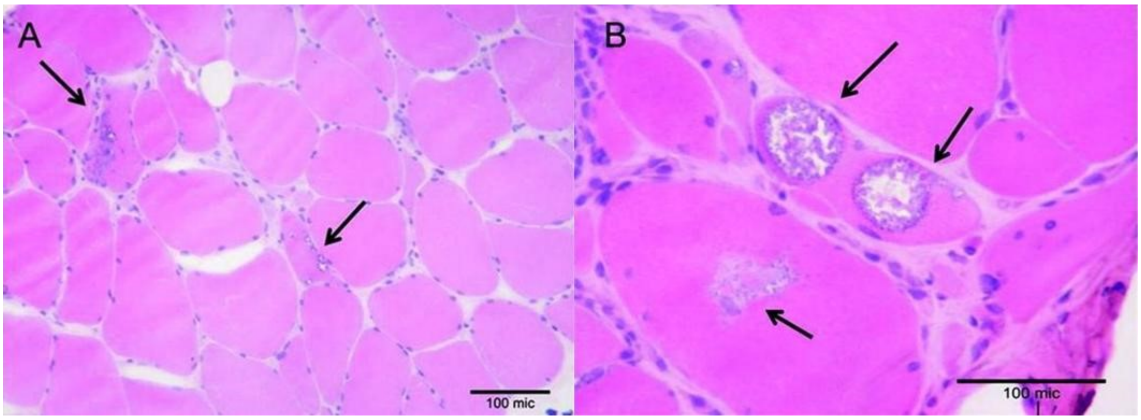


Figure 1.1.: Haematoxylin and eosin staining, biopsy taken from tibialis anterior muscle. **(A)**Non-specific myopathological changes with variation in fibre size, increase in connective tissue and increased number of internal nuclei. Also visible are rimmed vacuoles (arrows) in angular muscle fibres. **(B)** Foamy rimmed vacuoles in small and larger muscle fibres. (Figure taken from Durmus et al. 2011 [5])

Electron microscopy studies have shown a greater variation of findings. Some authors report cytoplasmic filaments [2], [3], which is an unspecific finding in neuromuscular disorders. Myelin figures [5], [9] in rimmed vacuoles seem to be a common finding in OPDM, subsarcolemmal masses of lipofuscin were only observed in one patient [9]. Intranuclear aggregations of tubular filaments, thought to be specific for OPMD, were found in two families [7], [10]. These changes are shown in Figure 1.2. In summary, it is hard to distinguish between OPDM and OPMD based on histopathological analysis. Therefore, some experts argue that OPDM is a subcategory of OPMD with normal GCG-repeats in the nuclear mRNA binding protein *PABPN1*, which are causative for OPMD.

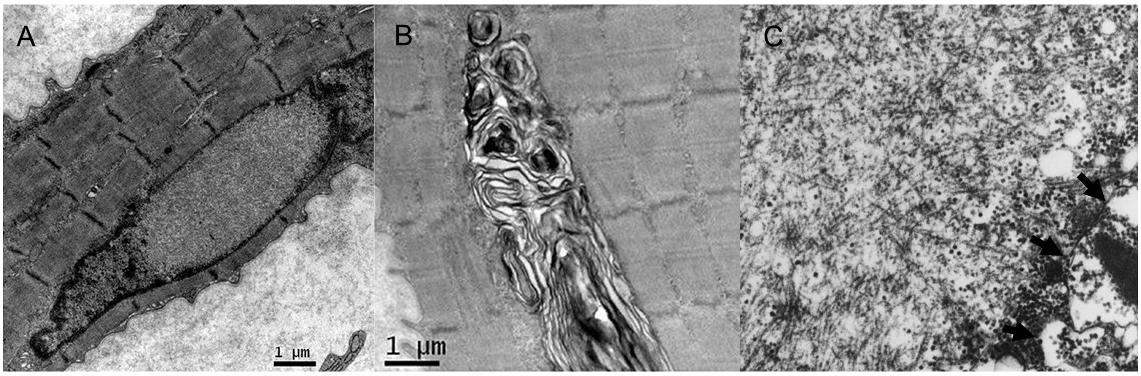


Figure 1.2.: Electron microscopy lead citrate and uranyl acetate staining. (A) Tubulofilamentous inclusions in the nucleus. (B) Numerous myelin figures were aggregated. (C) (magnification: x96.500) Cytoplasmic filaments which are located close to rimmed vacuoles (the latter indicated by arrows). The filaments are 16-18 nm in diameter. (Figure taken from Lu et al. 2008 and Uyama et al. 1998 [10], [2])

1.2. Exome Sequencing

1.2.1. The Concept

The human genome consists of around 3 Gb (giga-basepairs) [12] but only 1% of this vast number constitutes the protein coding part we call the exome - i.e. the protein coding regions. Nevertheless, it is the region where about 85% of all disease causing mutations occur [13]. Before the introduction of whole exome sequencing (WES) , the common approach in order to identify the underlying genetic variation of inherited diseases included performance of linkage analysis in families with known shared genetic heritage, followed by Sanger-sequencing of the genomic region of interest or, alternatively, a candidate gene approach. This is costly and time-consuming and success in identifying disease underlying mutations has been varying [14]. Hence, focussing on the exome is a reasonable approach when trying to identify mutations in Mendelian disorders as this massively reduces time, computational capacity and the problems with identifying a vast number of intronic and intergenic variants of unknown significance. In recent years, the hard- and software for WES have improved immensely while costs and hands-on time have decreased, thus making it an ideal method to target rare inherited conditions of unknown genetic cause. OMIM (Online Mendelian Inheritance in Man) lists more than 6000 presumably monogenic disorders but for more than two thirds of these the molecular basis has not yet been detected [13]. Since the establishment of exome enrichment strategies in 2007 [15], the first genetic diagnosis based on whole-exome sequencing in 2009 [16] and the first identification of a genetic cause of a Mendelian disorder in Miller syndrome [17], many cases have been solved in this short period of time.

1.2.2. Method

To perform whole-exome sequencing, the protein coding regions of the genome which only comprise around 1% of the human genome have to be enriched and amplified. This is followed by the sequencing step, usually done by modern massive parallel high-throughput sequencing-by-synthesis machines. Figure 1.3 briefly summarizes the work flow of whole-exome sequencing. There are three possible main principles, by which exome enrichment can be carried out: By in-solution capture, by hybridisation to an array or by polymerase chain reaction (PCR) . The hybridisation method uses single stranded oligonucleotides attached to the surface of a chip. These are complementary to the exonic regions of the human genome. The probe DNA is sheared to create double stranded fragments and a universal priming sequence is added. There fragments are then hybridized to the oligonucleotides on the array and unhybridised DNA is washed away. The remaining fragments are amplified by PCR and sequenced. Roche NimbleGen is the first and most popular system applying this method. [18]

The most common method for target enrichment, such as the exome, is in-solution capture. Similarly, to the method described above, it uses a pool of custom oligonucleotides attached to magnetic beads that can hybridize with the targeted region. Next, they are ferromagnetically pulled down and washed followed by PCR amplification and sequencing. [20]

Quite recently, Life Technologies have improved the method of exome enrichment by PCR and solved many of its problems, such as low read depth and coverage and combines this with the advantages of being time-saving and needing as little as 50 ng of template DNA. It is an in-solution capture method and uses primer pairs for around 300.000 amplicons which are multiplexed in 12 pools of 24.000. Each of the pools works as one individual multiplex PCR reaction and is then combined for the sequencing reaction. The sequencing-by-synthesis in this kit works by measuring H^+-ions released during base incorporation as opposed to fluorescence or chemiluminescence as used in the most common sequencing systems.

The two most common other sequencing techniques were introduced by Illumina and Roche (454 technique). In both cases, the sequencing library itself has to be created by PCR-amplification of the DNA-template. Roche then uses an emulsion PCR to create millions of clonal amplifications that are then attached to sequencing beads. This is followed by pyrosequencing, carried out in cycles. During each cycle, only one of the deoxyribonucleotides is offered. If it matches the base on the complementary strand, it gets incorporated and pyrophosphate is split off. Together with adenosine phosphosulfate (APS) it forms ATP which then, when luciferin is added

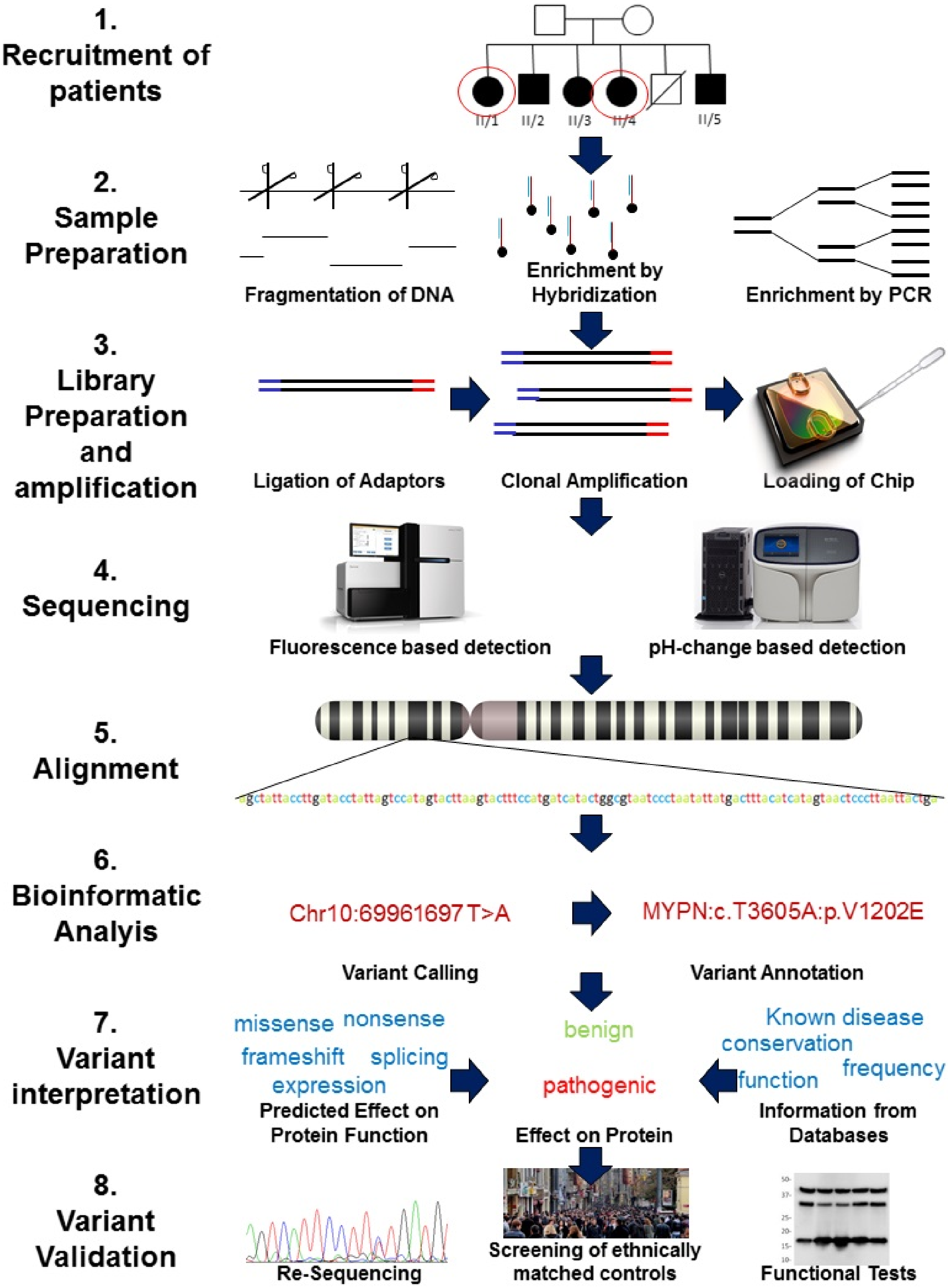


Figure 1.3.: Simplified work flow of a whole-exome sequencing analysis. Adapted from Lohmann et al. 2014, [19]

forms oxy-luciferin and light in the presence of luciferase which can be detected with a CCD (charge coupled device) -camera, as displayed in Figure 1.4

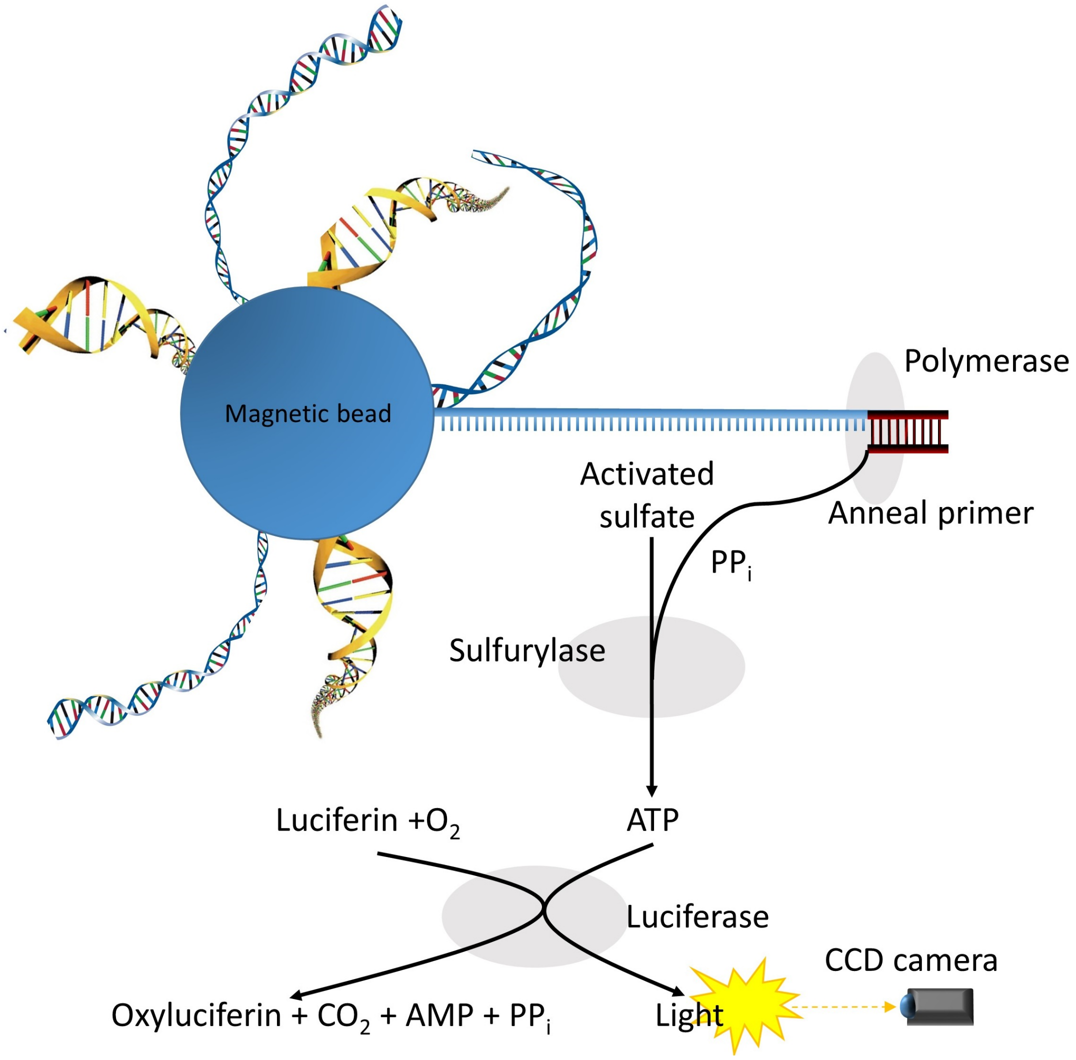


Figure 1.4.: Schematic illustration of pyrosequencing as used by the Roche 454 sequencers.

Illumina, however uses bridge PCRs to create locally differentiated clusters of identical PCR amplicons. DNA and primers get attached to flow cells and the two oligonucleotides hybridize to form a bridge as shown in Figure 1.5. After a certain number of PCR cycles, a cluster of amplicons forms. A step of denaturation leaves single stranded templates anchored to the surface. The sequencing primer anneals to the adaptor sequence of each DNA fragment and the sequencing can be done by a technique called cyclic reversible termination. It uses deoxyribonucleotides which contain a fluorophore and a reversible blocking group. The four nucleotides have four different fluorophores attached, emitting at different wave lengths. In each cycle the polymerase extends the strand by one base and the blocking group terminates

DNA syntheses. Unbound nucleotides are washed away and the array is imaged to determine the incorporated nucleotide. In a final step, the blocking group gets removed before a new cycle begins. (See figure 1.6)

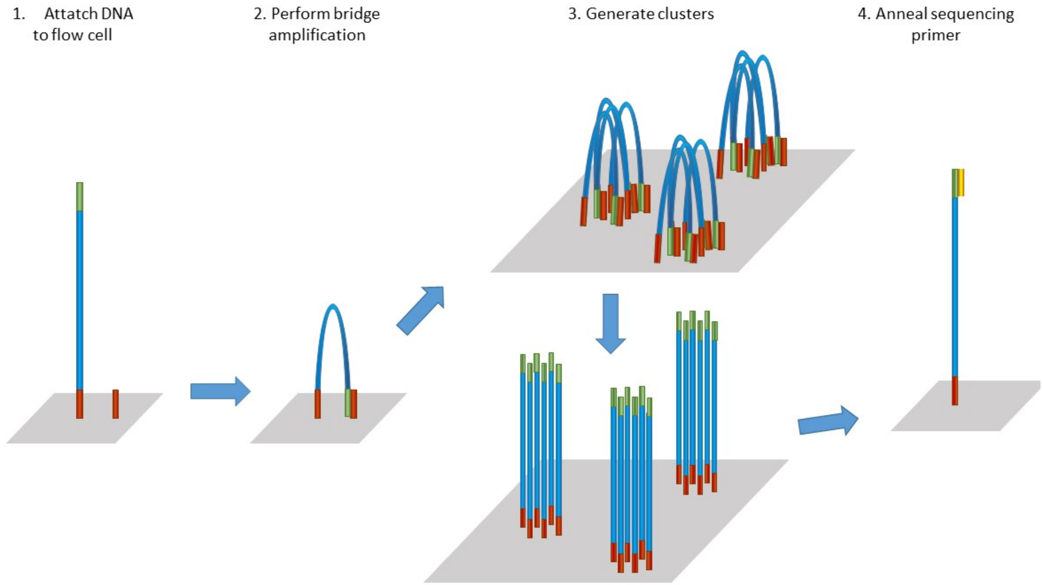

1. Attatch DNA to flow cell
2. Perform bridge amplification
3. Generate clusters
4. Anneal sequencing primer

Figure 1.5.: Schematic illustration of Illumina exome enrichment and amplification. Single DNA fragments and primers (red), complementary to the adaptor-sequence (green), are attached to flow cells. In each amplification step, bridges form and leave clusters of homogenous DNA fragments after the denaturation step. A universal primer binds to the adaptors for the following sequencing reaction (adapted from `https://www.eurofinsgenomics.eu/de/eurofins-genomics/produkt-faqs/next-generation-sequencing.aspx`).

The common output is FASTQ files, which is a format for sharing both the sequencing read combined with a quality score for each base [21]. The read sequences now need to be aligned, which means that each short read is mapped to a reference genome. There is a large number of software tools available that all come with advantages and disadvantages. The most important aspect in discussing different alignment algorithms is accuracy. When it comes to single nucleotide polymorphisms (SNPs) , SOAP is the most accurate aligner, followed by Bowtie [22], BWA and Novoalign. Yet, SOAP fails to align any reads with indels greater than 6bp which limits its use dramatically, considering, that the average pathogenic deletion has a size of 10bp. Here Novoalign performs best, especially when it comes to greater indels of 10-16bp. BWA only produces accurate alignment when a threshold value to remove unfavourable reads is introduced. Another important aspect is runtime. Here, Novoalign performs very good, even when processing large genomes. However, Bowtie performs better for high sequencing depth and therefore greater read

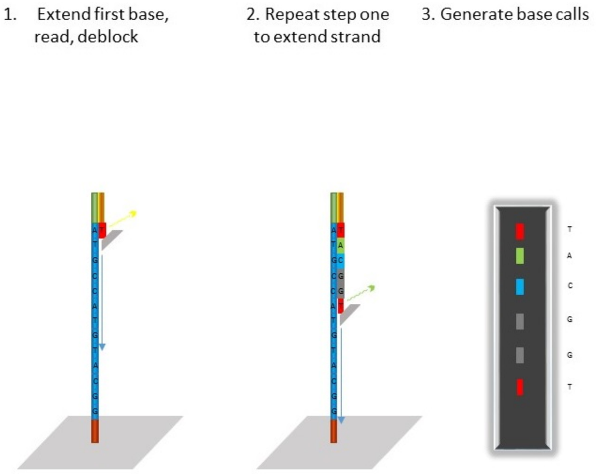

1. Extend first base, read, deblock
2. Repeat step one to extend strand
3. Generate base calls
T
A
C
G
G
T

Figure 1.6.: Schematic illustration of sequencing by cyclic reversible termination. Sequencing of generated clusters is performed by DNA-replication with reversible dye terminators. These are deoxyribonucleotides carrying different fluorophores for the 4 bases respectively and a blocking group. Since the blocking group terminates DNA synthesis the strand is extended only by one labelled nucleotide. The surface is then washed to remove non-incorporated nucleotides. In order to identify incorporated nucleotide, the fluorescent signal is analysed. Subsequently, the fluorophore and the blocking group is cleaved from the nucleotide and the next cycle begins. (adapted from `https://www.eurofinsgenomics.eu/de/eurofins-genomics/produkt-faqs/next-generation-sequencing.aspx`)

counts. All in all, researchers have to choose the alignment program according to their computational infrastructure as well as their scientific aims and questions.

Post-alignment processing of the data comprises removal of duplicated reads, indel (insertion or deletion) realignment for a better detection of insertions and deletions, a base quality score recalibration and finally the variant calling [23]. Studies in the past have demonstrated substantial disagreement between variant calls made by different pipelines. This illustrates the problematic nature of interpreting the data [24][25][26]. Especially detection of indels is still challenging even though algorithms have improved over the years. Common platforms are genome analysis toolkit (GATK) [27], Dindel [28], Platypus [29], SAMtools [30] as well as VarScan [31] for indels. The easiest way of detecting variants is by mapping reads to a reference genome and then scanning for systematic differences [32]. A slightly more complex way is to reconstruct haplotypes that are well supported by the data to identify true variants [33]. The advantage is, that this approach ensures semantic consistency, which means that there can be no logic contradiction in variant calling such as different bases on the same allele in a detected variant. However, mapping algorithms also have a number of disadvantages. Firstly, this method focuses on SNPs and very short indels which leads to errors around larger indels or other variants. Secondly, it often fails in regions of high similarity where misalignment creates a systematic error [29].

Another approach when aiming to avoid these limitations is a reference-free assembly. It does not rely on a reference genome and is therefore variant agnostic. It also copes well with difficult to align regions and reconstructs the haplotypes to call the variants making this algorithm highly specific. Especially indels are detected more sensitively. However, it requires high computational power and has a lower sensitivity for SNPs compared to mapping algorithms [34]. In addition, some variant callers also borrow information across a number of samples to support a change in one sample, if it matches the information contained in many others. Conclusively, a weakly confidential variant can be called if it is confidently identified in another sample or samples. [29].

Depending on alignment and variant calling programs, one is left with around 50,000-100,000 variants from whole exome sequencing including intronic, intergenic and changes found in the untranslated regions (UTRs) of which around 20,000 are on-target meaning in exonic regions [23]. Annotation programs are used to add information such as genomic feature, gene symbol, exonic function and amino acid change of each variant. The program Annovar, that was used in this project, also adds information from dbSNP, 1000genomes and ESP6500 for assessing the minor allele frequencies, integrates data retrieved from Phylop and Genomic Evolutionary Rate Profiling (GERP) and employes different tools to predict the pathogenicity of variants.

These called and annotated variants then have to go through a filtering pipeline to reduce this high number to just a few candidate variants. The filtering steps are based on certain assumptions: First, the disease causing variant is rare, meaning only present in affected individuals. Second, only homozygous or heterozygous mutations in one single gene are required to cause the observed phenotype. Third, these mutations are 100% penetrant and have a large effect size usually affecting protein sequence (insertions/deletions as well as missense, nonsense, frameshift or splice-site aberrations). The assumption of a high penetrance is needed when interpreting allele frequencies. Variants with a small effect size usually have a higher frequency and are associated with polygenic conditions such as diabetes. Therefore, filtering for indels and nonsynonymous, nonsense or splice-site changes and for those with a minor allele frequency of less than 1% is usually the first step [35]. Depending on the project, variants of affected individuals from one family can then be intersected to see, which are present in all and excluded if they are also found in healthy family members. Finally, the number can be further reduced based on the pattern of inheritance. To give an example, heterozygous – but not compound heterozygous – as well as X-chromosomal variants can be excluded for an autosomal recessive model.

1.2.3. Limitations

Even though WES provides a fast and relatively affordable method, there are a couple of limitations when trying to detect disease causing variants. Around 15% of all known causative mutations for Mendelian conditions cannot be found within the protein coding regions. These variants comprise for example mutations in the promoter regions, in the introns which form cryptic splice acceptor sites or in the untranslated regions. It is also possible, that variants in non-coding genes might affect cellular pathways and are therefore pathogenic. These cases are difficult as researchers usually spend much time with non-causative, rare variants which segregate with the disease and try to prove their pathogenicity. Additionally, the coverage is not yet satisfactory. Especially GC-rich regions are hard to enrich resulting in poor read depth as well as the problem that some regions are not covered at all [36], [35]. Furthermore, duplicated regions such as pseudogenes usually cause a large number of false positive calls as the alignment programs map these reads to all similar regions and consequently, the differences are called as genomic changes. This can be problematic for scientists when the list of candidate genes after the filtering steps is unsatisfactory because one cannot dismiss the possibility that the exonic pathogenic variant is simply not covered. Even if the coverage would be 100%, one would still face difficulties with copy number variations and larger insertions or deletions. Copy number variations are usually undetectable by WES, as read depth varies a lot in different chromosomal regions and indels are a common cause, why reads cannot be aligned as they differ too much from the reference sequence. If a variant in a gene, which is not yet annotated was disease causing, it would either not be enriched before sequencing or filtered out in the bioinformatic pipeline [13]. On top of this, one often faces a large number of possible disease causing variants. Especially with smaller families and dominant models it is then hard to decrease this number by linkage analysis or homozygosity mapping. These variants are often difficult to interpret [19]. Also, when detecting changes that are likely disease-causing, WES cannot prove that a specific variant underlies the condition. Often it is then difficult to prove pathogenicity [35]. Usually, researchers try to find mutations in the same gene of other affected individuals from different families. If that is not possible, e.g. in very rare diseases, it is necessary to look at cell and animal models to find the pathomechanism. Also, this problem harbours a second challenge: It can be easy to fall for a compelling but false causative variant when being confronted with a large number of rare, protein altering genetic changes. This effect is called the "narrative-potential" of human genomes [37], [38]. Conclusively, researchers have to be careful when claiming an association between genetic variants and disease and consider certain guidelines such as these published by MacArthur et al. in 2014 [39]. Taken together, the limitations of whole exome sequencing only result in a success rate of around 25% [40], [41].

1.2.4. Ethical Implications

In addition, researchers have to consider ethical implications, when using WES. Some of these are rather theoretical but still have to be discussed, others occur in everyday laboratory practice. Most importantly, there are no generally accepted guidelines determining how to proceed, when detecting pathogenic genetic variants like mutations in *BRCA1*, commonly causing breast- and other gynaecological cancers. Green et al. from the American College of Medical Genetics and Genomics (ACMG) have published a list of 56 genes and suggested to inform the patients when detecting mutations in one of these [42]. These genes are usually associated with treatable/curable conditions, nevertheless this topic is still subject of debate. Yet, informing the index patient about an important pathogenic variation is difficult as relatives who have a right not to know and never gave their consent to the project might also be affected. Most problems, however, deal with data protection. DNA contains a lot of information which may be of interest for insurers, employers or the police. Once it is sequenced and decoded it is hard to tell what is going to happen in the future. Some sequences from WES and WGS can even be downloaded from open access libraries and their future use is unspecified [14]. DNA harbours not only information of the individual who has given informed consent to publishing the data but also of all his relatives as these share some of his variants[13]. It might even be possible to identify this individual based on the information from his DNA [14]. In conclusion, it is difficult to obtain informed consent since it cannot be predicted what information is gathered from the genomic sequence and how it may be used in the future [43], [44]. However, when dealing with rare diseases, research is often the only hope for patients to get a genetic diagnosis. This is important for them due to implications for relatives as well as family planning. In some rare cases, a correct genetic diagnosis also poses treatment options such as in Brown-Vialetto-Van Laere syndrome where high dose riboflavin substitution results in a tremendous improvement of symptoms [45]

2. Aim of the Study

This study aims to identify the causative genetic variant responsible for OPDM. The first descriptions of this disease entity have been published many years ago [1] [46] and all efforts to find the underlying mutations up to the present have been unsuccessful. Patients and their families suffer tremendously from OPDM as this condition affects their ability to perform everyday tasks, eating, swallowing, walking and eventually breathing. It also causes high costs for public health care or insurers, so it is an important task to take the next step to provide support for the patients and identify the condition's genetic cause. Most interestingly, different inheritance patterns and phenotypes have been reported, indicating that a complicated pathomechanism might be underlying [5], [7], [3]. A similar condition, OPMD, is caused by a repeat expansion resulting in a dominant or recessive pattern of inheritance depending on the number of additional triplets [11]. Finding the causative mutation in OPDM would answer a number of questions such as whether this condition is a genetically homogeneous disease and what the underlying pathomechanism is. Depending on the disease mechanism researchers could establish diagnostic algorithms and eventually find a treatment as has been the case in Pompe's Disease where the defective enzyme alpha glucosidase can be substituted intravenously [47]. Also, the knowledge that a certain mutation results in a specific phenotype can provide information on cellular and tissue physiology. Therefore this study tried to map the disease locus using linkage analysis, homozygosity mapping and haplotyping in previously reported families from Turkey [5]. In order to identify the mutation causing OPDM whole exome sequencing was carried out for 2 patients from a large dominant family, 3 patients from a recessive family (both from Turkey [5]) and unpublished patients from Canada, Finland and the United Kingdom. Candidate mutations were evaluated by segregation analyses in the corresponding families as well as tests for their occurrence in ethnically matched control individuals.

3. Materials and Methods

3.1. Patients and Controls

Patients were assessed by their referring clinicians and genomic DNA was sent to the Institute of Genetic Medicine, Newcastle with the patient's informed consent. DNA from 70 ethnically matched healthy individuals was provided by the "Friedrich-Baur-Institut der Medizinischen Fakultät an der Neurologischen Klinik und Poliklinik der LMU München". All studies have been approved by the local ethics committee in Newcastle-upon-Tyne, UK.

3.2. Chemicals

Chemicals used can be found in table 3.1.

3.3. Molecular Biological Methods

3.3.1. Quantification of Nucleic Acid Concentrations

Two different systems were used to determine DNA concentrations. For larger sample numbers the Nanodrop® spectrophotometer (Thermo Scientific), using an ultraviolet-absorbance method, was used according to the manufacturer's instruction [48]. A fast alternative for smaller sample quantities is the Qubit® system (Life Technologies) that applies fluorescent dyes to quantify DNA concentrations.

3.3.2. Polymerase Chain Reaction

3.3.2.1. The Main Principle

Introduced by Karl Mullis in 1983, the polymerase chain reaction (PCR) is a fast and cheap biochemical process to amplify a target genomic DNA sequence [49], [50]. The method relies on thermal cycling, consisting of cycles to separate the two DNA strands, for the annealing of the sequence specific primers and the elongation by a heat-stable DNA polymerase (Usually isolated from Thermus aquaticus).

Chemical	Supplier
Agarose	Sigma-Aldrich
Tris base	Sigma-Aldrich
Glacial acetic acid	Sigma-Aldrich
EDTA	Sigma-Aldrich
Safeview	abm
Taq polymerase	Molzym
MgCl “PCR enhancer”	Molzym
dNTPs	ThermoFisher
Pfu polymerase	ThermoFisher
Primers	MWG Eurofins
Restriction endonucleases	New England BioLabs (if not stated otherwise)
Nuclease-free water	Qiagen
EB-Buffer	Qiagen
Acrylamide/Bis-acrylamide	Sigma-Aldrich
TEMED	Sigma-Aldrich
Ammonium persulfate	Sigma-Aldrich
Ethidium bromide	Sigma-Aldrich
Boric acid	Sigma-Aldrich

Table 3.1.: Chemicals used in this study and the supplying companies.

Double distilled water	35 μl
10 mM dNTP mix	2 μl
10X Moltaq PCR buffer	5 μl
Moltaq PCR enhancer	4 μl
Forward primer at 50 pmol/μl	1 μl
Reverse primer at 50 pmol/μl	1 μl
Genomic DNA template at 100 ng/μl	1 μl
Moltaq Polymerase	1 μl
	50 μl

Table 3.2.: Amounts of chemicals used for a standard PCR with a Taq-polymerase

3.3.2.2. Standard Protocol

Standard PCR was done using the Moltaq® Taq Polymerase (Molzyme) with the amount of reagents which are listed in table 3.2 on page 14.

Standard PCR was performed in a Sensoquest® thermal cycler according to the protocol found in table 3.3.

The melting temperature of primers increases with higher GC content and length. It was calculated with the UCSC In-Silico-PCR software (`https://genome.ucsc.edu/cgi-bin/hgPcr`) and the annealing temperature was adjusted depending on the primer melting temperature.

1. 94°C: 5min	Denaturation
2. 40 cycles	
(1) 94°C: 15s	Separation of DNA strands
(2) 52-62°C: 30s	Primer annealing
(3) 72°C: 2min	DNA elongation
3. 72°C: 7min	Final elongation

Table 3.3.: Standard PCR-cycler settings with 40 cycles.

Acrylamide/Bis-acrylamide, 30% solution	4,8 ml
Water	4,8 ml
5x TBE buffer	2,4 ml
10% ammonium persulfate	200 μl
TEMED (Tetramethylethylendiamine)	10 μl

Table 3.4.: Protocol for casting polyacrylamide gels.

3.3.2.3. Modification of standard PCR protocol

The Phusion ® High-Fidelity DNA Polymerase (New England Biolabs) was used for target sequences with a high GC content or repeat-rich parts of the DNA according to the manufacturer's instructions.

3.3.2.4. PCR Primers

PCR Primers were designed using primer3 (`http://primer3.ut.ee/`) based on the hg19/GRCh37 assembly (`http://genome.ucsc.edu/cgi-bin/hgNear`), introduced in 2009. They were synthesised by Eurofins MWG Operon, Ebersberg, Germany using a HPSF (High Purity Salt Free) purification protocol. A list of all primers used can be found in the addendum (A.1 on page 93).

3.3.2.5. Gel Electrophoresis

1%-3% agarose gels were used for electrophoreses depending on the PCR product's length. Standard agarose concentration for PCR reactions was 2%. The electrophoresis was done in 1X TAE (Tris-acetate-EDTA) buffer. DNA was made visible with Safeview (NBS biologicals). TAE was made in 10x stock solutions using 48.4 g of Tris base [tris(hydroxymethyl)aminomethane], 11.4 ml of glacial acetic acid (17.4 M) and 3.7 g of EDTA, disodium salt in 1l of deionised water.

To determine the length of microsatellites 12% polyacrylamide gels were cast according to the following protocol:

The gels were run in a vertical electrophoresis system and 1X TBE buffer. 10X TBE stock solution was made using 108 g Tris, 55 g Boric acid, 9.3g EDTA and volume was adjusted to 1 liter using deionised water. DNA was stained with ethidium bromide according to the manufacturer's instructions.

3.3.3. PCR DNA Clean-up

DNA was either purified directly after the PCR using the QIAquick PCR Purification Kit or extracted from agarose gels using the QIAquick Gel Extraction Kit (both Qiagen). 96-well plate PCR products were purified using the GenElute^{TM} 96 Well PCR Clean-Up Kit (Sigma-Aldrich) according to standard protocol.

3.3.4. Cleavage of DNA with Restriction Endonucleases

Whenever possible and reasonable, restriction endonucleases were used for segregation analysis of variants. For *DdeI* (Promega), 3μl of the supplied 10X buffer D, 25μl water and 20 units in 2μl of the restriction enzyme were added to purified PCR products and incubated at room temperature overnight. *BsrDI*-digest was performed with 15μl purified PCR-products, 2μl 10X Buffer 2, 2μl 10X BSA (Bovine Serum Albumin) as well as 5 units in 1μl restriction enzyme and incubated for 2 hours at 65°C. Digested DNA fragments were analysed on a 3% agarose gel.

3.3.5. DNA Sequencing and Sequence Analysis

Sequencing was done by Eurofins MWG Operon (Ebersberg, Germany) by cycle sequencing, a modification of Sanger sequencing. Sanger sequencing, first introduced in 1977 [51] uses the chain-termination method with fluorescently or radioactively labelled dideoxynucleotides (ddNTP). Cycle Sequencing however uses a heat-stable polymerase and fluorescently labelled ddNTPs emitting at different wavelengths and can therefore be performed in one tube with much less template DNA. The electrophoresis is done in modern 96-capillaries sequencers.
The results were analysed with Chromas (`http://chromas.software.informer.com/`) and BLAST (`http://blast.ncbi.nlm.nih.gov/Blast.cgi`).

3.4. Bioinformatic methods

3.4.1. SNP-Array

SNP-arrays for 36 affected and unaffected individuals from 8 Turkish families were done by Source BioScience (Nottingham, UK). Each sample was normalised to a concentration of 50ng/μl. 200ng (4μl) of each normalised sample was amplified and subsequently prepared for hybridisation (fragmentation, precipitation, re-suspension).

The samples were hybridised to Infinium II Human Linkage-12 arrays from Illumina for 24 hours. Following hybridization, the Infinium arrays were stained, washed and finally scanned.

3.4.2. Linkage Analysis

The program Merlin was used to perform parametric multipoint linkage analyses [52]. Linkage analyses were carried out for each individual family as well as for all families together using a dominant, a co-dominant and a recessive model. The affection status of the youngest generation in Family 1 (see figure 4.1 on page 22 for pedigree) was defined as "unknown" to prevent the result of false positive or negative high linkage loci. Merlin requires 4 input-files, a data file (*parametric.dat*), a pedigree file (*parametric.ped*), a map file (*parametric.map*) and a model file (parametric.model). The data file contains all the single nucleotide polymorphisms that the linkage analysis is based on. The map file links these markers to positions on each chromosome in cM (CentiMorgan). The pedigree file combines information from all patients namely their affection status, their gender, their generation and their parents. To calculate the LOD (logarithm of odds)- score for different patterns of inheritance, a model file is used to provide data about the estimated penetrance and the minor allele frequency of the mutation (For example input files see figure A.2 on page 104 in the addendum.). To verify that input files are being interpreted correctly, the program Pedstats is used [53] by prompting the following command in DOS :

```
prompt> pedstats -d parametric.dat -p parametric.ped
```

If the pedstats output produced a correct summary of the families, merlin was used to perform a parametric linkage analysis by prompting the following command:

```
prompt> merlin -d parametric.dat -p parametric.ped -m parametric.map
--model parametric.model --step 3 --pdf
```

The –step 3 option of the Merlin program was used, which adds an computed calculation of the LOD-score at three steps between two consecutive markers. This improves the analysis as the LOD-score tends to decrease around marker locations. By adding –pdf the data is shown in pdf files. All input files were created individually for each chromosome and for different combinations of families. Genetic linkage is indicated by a so called LOD score, which is the logarithm (base 10) of odds of the likelihood of obtaining the test data if the two loci are linked compared to the likelihood of observing the same data purely by chance [54]. By convention, a LOD score greater than 3 is considered evidence for linkage whereas a LOD-score of -2 is considered an exclusion of linkage.

3.4.3. Homozygosity Mapping

Homozygosity mapping was performed based on the data from SNP-arrays and Exome-Sequencing of individuals II/2, II/3 and II/5 using the program Homozygositymapper (`http://www.homozygositymapper.org/`) [55]. This web-based program stores marker data in a database into which SNP genotype files can be directly uploaded. The files to be uploaded must be tabular with the samples as columns and the SNPs as rows. Genptypes were written as follows: "AA" for wildtype; "AB" for heterozygous variant; "BB" for homozygous variant and "–" if this SNP has not been detected. Doing homozygosity mapping with exome sequencing data is not reasonable as homozygous wildtype polymorphisms are not being detected by variant calling programs. In consequence, one can not differentiate between homozygosity for the wildtype allele and lack of alignment for this region. In these cases, the genotype was declared unknown if the polymorphism was called in one or two siblings. In order to receive better results the program was provided with 14 control samples. Genotype data was downloaded from the website of the International Hapmap Project (`http://hapmap.ncbi.nlm.nih.gov/downloads/index.html.en`). Unfortunately, there is data from only a few ethnic groups and none perfectly matches the background of the cohort the study at hand is based on. Due to the fact, that Turkish people share a majority of their haplogroups with their caucasian neighbours, Italian individuals were chosen as a control group.[56] Default settings were used in Homozygositymapper and the built-in candidate gene search engine GeneDistiller (`http://www.genedistiller.org.`) [57] was used to analyse genes in homozygous regions. An excel spreadsheet with information on the genes, their expression in skeletal muscle and information from the OMIM database [58] found inside homozygous regions was downloaded from the website and further analysed. Lists of genes from the SNP genotyping array and the exome sequencing data were both concatenated and intersected to retrieve the maximum as well as most likely number of candidate genes.

3.4.4. Microsatellite analysis

Microsatellite analysis was done using Simple Sequence Length Polymorphisms (SSLPs) on Chromosome 10. These markers were selected based on the information found on UniSTS (`http://www.ncbi.nlm.nih.gov/unists`) a database listing sequence tagged sites (STSs) . STSs are defined by PCR primer pairs and are associated with additional information such as genomic position, genes, and sequences. The markers listed in table 3.5 on page 19 were amplified using fluorescently labelled primers and their length was determined in capillaries in the laboratory of Professor Angela Hübner, Klinik- und Poliklinik für Kinder- und Jugendmedizin, TU-Dresden, Germany.

Marker	Position on Genethon Map: (cM)	Position in Mb
D10S1787	70,90	49,8
D10S1793	70,90	50,1
D10S1766	72,00	50,7
D10S196	72,50	52,1
D10S568	74,20	53,7
D10S1643	77,00	55,3
D10S1756	78,40	59,1
D10S589	81,70	61,5
D10S1652	83,30	64,4
D10S561	83,30	65,1
D10S1743	84,90	67,4
D10S1665	92,20	71,3
D10S1688	94,00	72,6
D10S1650	95,60	73,3
D10S1730	103,20	78,9
D10S201	105,90	81,0
D10S1777	105,90	81,1
D10S1686	109,20	85,6

Table 3.5.: List of microsatellites and their genomic position used for haplotype analyses.

3.4.5. Whole Exome Sequencing

3.4.5.1. Target enrichment and sequencing

Exome sequencing for Patients I/5 (OPDM1) and I/10 (OPDM2) was done by Eurofins MWG Operon (Ebersberg, Germany) based on the in-solution hybridization Agilent SureSelect Exome kit with Illumina HiSeq 2000 sequencing. DNA from patients OPDM 3 - 8 were sent to Otogenetics (Atlanta, USA) for exome sequencing. The Agilent V4 51Mb kit was used for target enrichment and sequencing was done on an Illumina HiSeq 2000 sequencer.

3.4.5.2. Bioinformatic workflow

The raw data was downloaded from the company's servers as FASTQ files and aligned to the hg19 assembly using programs working with various algorithms. This was done by the Burrows-Wheeler Aligner (BWA) and Bowtie, both using a FM-index as well as Novoalign and MOSAIK, both hashing the reference [59]. SAMtools was used to create a sorted BAM file. Picard, a program that comprises Java-based command-line utilities that manipulate SAM files, was used to remove duplicate reads to decrease the number of false positive heterozygous calls. SAMtools was used to create a BAI file and SNVs were called by VarScan whereas indels were called by Dindel. The calls were filtered for variants which are 'on-target' (Truseq 62Mb target coordinates +/- 500bp), seen on both DNA strands and for a minimum coverage of 5. SNVs found in more than 25% of the reads were declared heterozygous, when found in more than 85% they were considered homozygous. Finally, Annovar was used for gene based annotation of the changes. Given a list of variants from whole-exome or whole-genome sequencing, it generates an Excel-compatible file with gene annotation, the nucleotide- as well as the amino acid change, SIFT scores [60], PolyPhen2 scores [61], LRT scores [62], MutationTaster scores [63], PhyloP conservation scores [64], GERP++ conservation scores, dbSNP identifiers, 1000 Genomes Project allele frequencies, ESP 6500 exome project allele frequencies and other information. All variants were filtered for those with a minor allele frequency of less than 1% according to dbSNP (`http://www.ncbi.nlm.nih.gov/SNP/`), 1000 Genomes Project (`http://www.1000genomes.org/`) and the Exome Variant Server (`http://evs.gs.washington.edu/EVS/`) as well as the Newcastle University Inhouse MAF list. Any changes that were not found in exonic regions, the UTRs or putative splice site mutations were filtered out together with those in duplicated regions (>92% similarity). Variants found in samples “OPDM1” and “OPDM2” as well as “OPDM 3” - “OPDM 5” were intersected and analysed.

4. Results

4.1. Clinical Findings and Pedigrees

Most of the patients that were subject of the study at hand, come from a region at the Black Sea in Turkey and were assessed in Istanbul at the Department of Neurology (as described by Durmus et al. [5]). These 47 patients derive from 9 unrelated families whose pedigrees can be found in figure 4.1. However, affection status of individuals I/14 and I/15 could not finally be determined by the clinicians as they were very young when the study was performed [5]. Therefore, they were not considered in the linkage- and the haplotype analysis. Apparently, Families 1, 3, 4, 6 and 8 show a dominant pattern of inheritance, whereas Families 2 and 7 are clearly recessive. The pedigree of Family 5 however implies incomplete penetrance with an underlying dominant inheritance, as X-chromosomal dominant heredity or a mitochondrial disease can be excluded due to the unaffected male conductor in the third generation. OPDM in the single patient from Family 9 could occur sporadically or be caused by either a recessive mutation or a dominant one with incomplete inheritance. OPDM-patient DNA was also provided by Dr Tanya Stojkovic in Paris, France, Professor Dotti in Siena Italy ([9]), Professor Bjarne Udd in Tampere, Finland, Professor Bernard Brais in Montreal, Canada, from Professor Patrick Chinnery in Newcastle-upon-Tyne, UK and from Dr Paul Maddison in Nottingham, UK. Unfortunately, clinical data was not provided for any of the non-Turkish and non-Italian patients but a clinical diagnosis and a genetic exclusion of other common neuromuscular disorders with a similar phenotype like Oculopharyngeal Muscle Dystrophy was confirmed.

4.2. Linkage Analysis

Parametric linkage analysis was performed based on SNP genotyping array data (Infinium II Human Linkage-12 array by Illumina) as described in chapter 3.4.2 on page 17. It was done for each of the Families 1, 2, 3, 5, 6 and 8 individually as well as all families 1-9 together. Also single families have been excluded to see how this affects the LOD-score. If it was to drop in a high linkage region this would be a lead that OPDM might be caused by mutations in different genes in different

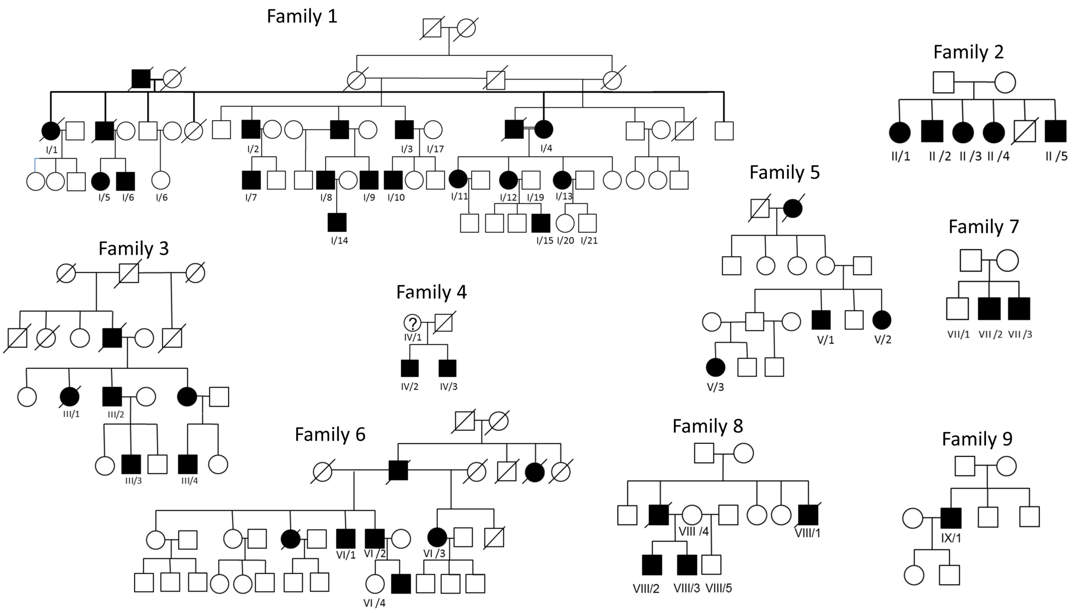


Figure 4.1.: Pedigrees of 9 Turkish families affected by oculopharyngodistal myopathy. Adapted from Durmus et al. 2011 [5]

families. Figures 4.2 and 4.3 summarise the most important findings. For no other chromosome apart from chromosome 2 and 10 a LOD-score greater than 1.0 has been observed for any combination of families. For a recessive model, Family 2 shows a region around markers at 150cM where the LOD-score reaches 1.5 which is significant, considering that only 5 individuals have been genotyped. However, this region can almost certainly be excluded as one shared disease locus for all families because the LOD-score reaches -2 for markers at this position in Family 1. This is reflected in the linkage analysis of all families where the LOD-score is 0. Nevertheless, if all but Family 1 are considered, this increases the LOD score for this region on chromosome 2 beyond 1.7 indicating that OPDM in other families might also be caused by a mutation in this region. The linkage peak in Family 1 on chromosome 2 for a recessive model does not have to be of concern as the pattern of inheritance is clearly dominant. For a dominant model only chromosome 10 exhibited a positive linkage. If a parametric linkage analysis for the largest clearly dominant Family 1 is performed, the resulting LOD-score on chromosome 10 is surprisingly low with values of around 1. Similarly, the LOD-score does not exceed 1 in Family 2 for this region. However, if all families are analysed together, the LOD-score almost reaches 3 but it stays at values between 1,5 and 2, when Family 1 is excluded. This implies that the disease locus for families showing a dominant inheritance could be successfully linked since the addition of other families increased the linkage score considerably. The relatively high linkage on chromosome 10 for a recessive model was interpreted to be a by-product of the correctly mapped area of interest in some families. Conclusively, the area at around 100cM - which is approximately at the genomic position Chr10:80,000,000 - was considered to be the most likely disease locus for dominant families.

4.3. Homozygosity Mapping

Given a high linkage to chromosome 2 for Family 2 which shows a recessive pattern of inheritance, homozygosity mapping was used to further map the disease locus. The web-based program was used as described previously in chapter 3.4.3 on page 18. It does not only detect homozygous regions but also combines them with information on allele frequency to provide an estimation score how likely the disease causing gene is to be found in a region. For example, if an SNP has a very low frequency and both alleles are wildtype in all affected individuals, this would result in a lower score. However, a very rare homozygous polymorphism would result in high scores. Figure 4.4 on page 26 shows results from homozygosity mapping from both SNP genotyping array (**A**)(Individuals II/1, II/2, II/3, II/4 and II/5) and exome sequencing data (**B**) (Indiciduals II/2, II/3 and II/5) and finally when data was combined prior to analysis (**C**). **D** provides further information on homozygous stretches from combined data results. Apparently, no larger homozygous regions could be detected which would be in line with findings in a consanguineous family. Additionally, results from the SNP array data differ from those retrieved through exome sequencing as there is no convincing homozygous region shared by both analyses.

When both information is combined and analysed (figure 4.4, **C**, **D**) the output reaches higher scores than when analysed separately. Nevertheless, no larger homozygous blocks could be identified. Most interestingly, one large region seems to be identical in all affected individuals. Both the data from the SNP genotyping array and the data derived from whole exome sequencing show that all 5 patients are homozygous and heterozygous for the same polymorphisms in the region between rs12711538 and rs344689. This might be due to compound heterozygosity for alleles shared by all patients and could be the disease causing locus. Genotypes for all 5 patients for this region can be found in figure 4.5 on page 27. A list of all genes in this region which are expressed in skeletal muscle can be found in table 4.1 on page 28. Expression data was derived from the Expression Atlas providing data on tissue expression of all protein coding genes (`https://www.ebi.ac.uk/gxa/home`). Among these, Myosin VII B would be a viable candidate gene. Additionally, the

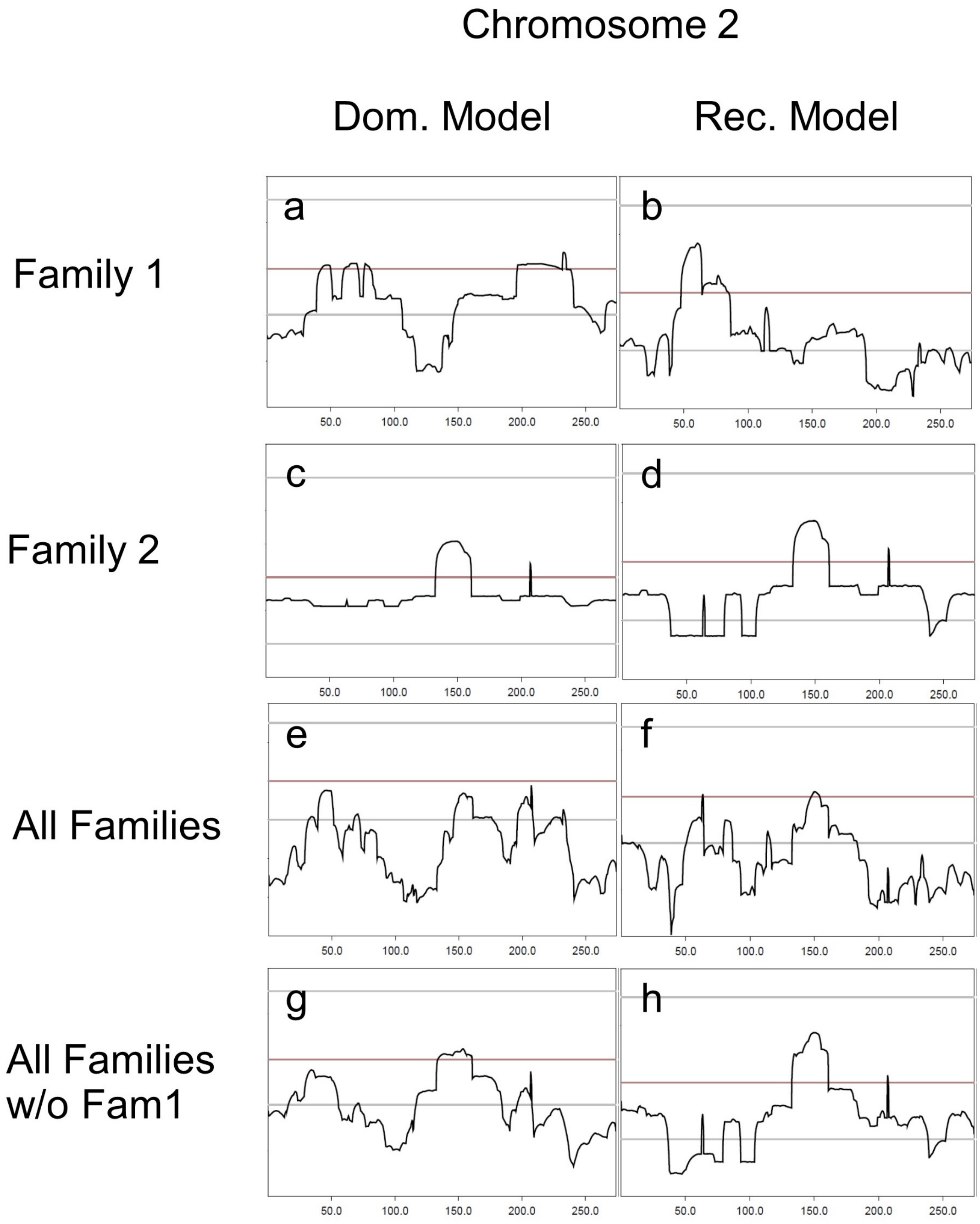


Figure 4.2.: Parametric linkage analysis of Family 1 (a and b), Family 2 (c and d), Families 1 to 9 (e and f) and Families 2 to 9 (g and h). a, c, e and g show linkage analysis for a dominant model and b, d, f and h for a recessive model of chromosome 2. The location on each chromosome in Centimorgan (cM) is displayed on the x-axis and the LOD score on the y-axis. The top grey line in each field marks a LOD-score of 3, the purple line marks a LOD score of 0 and the lower grey bar of -2.

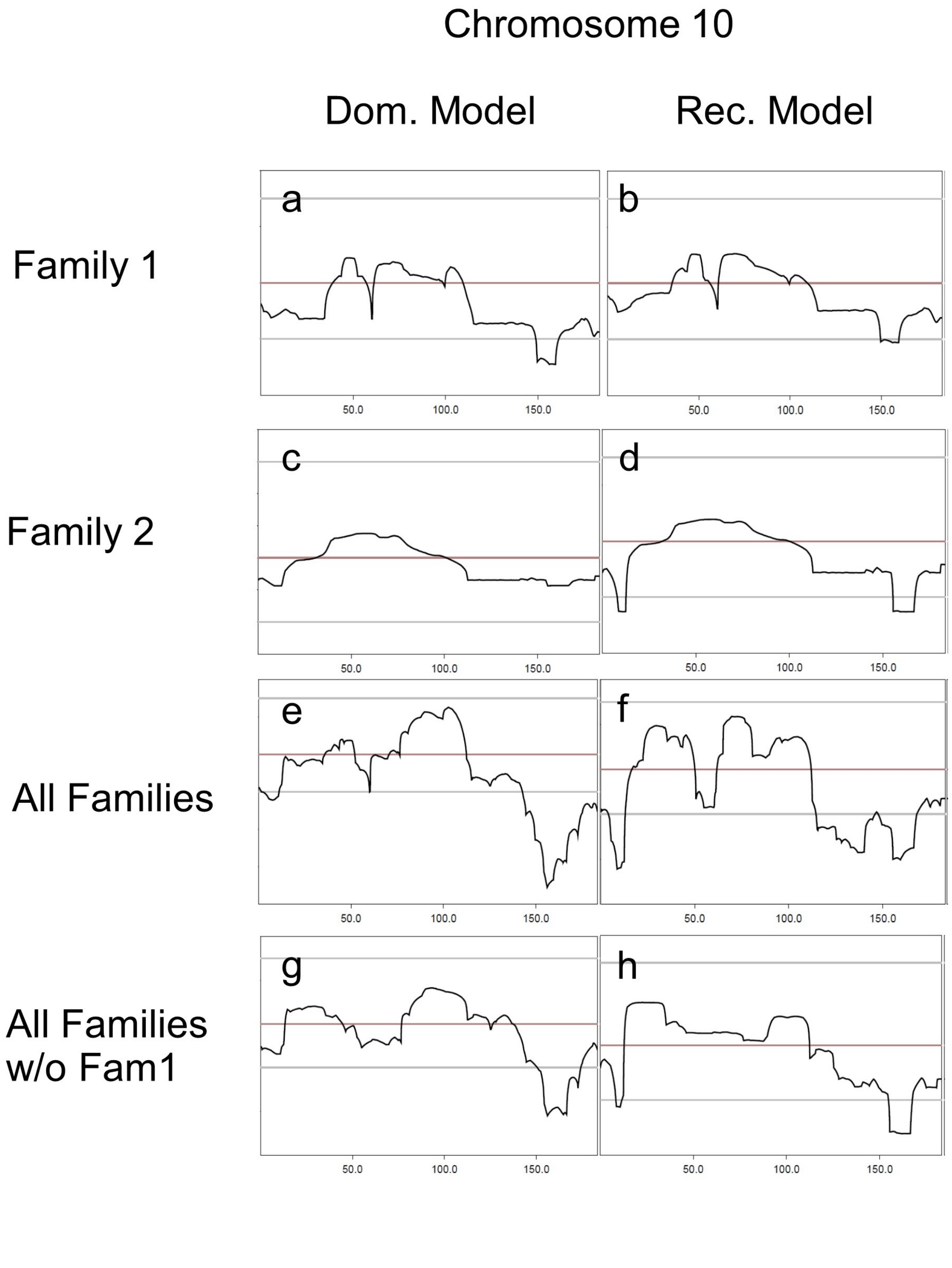
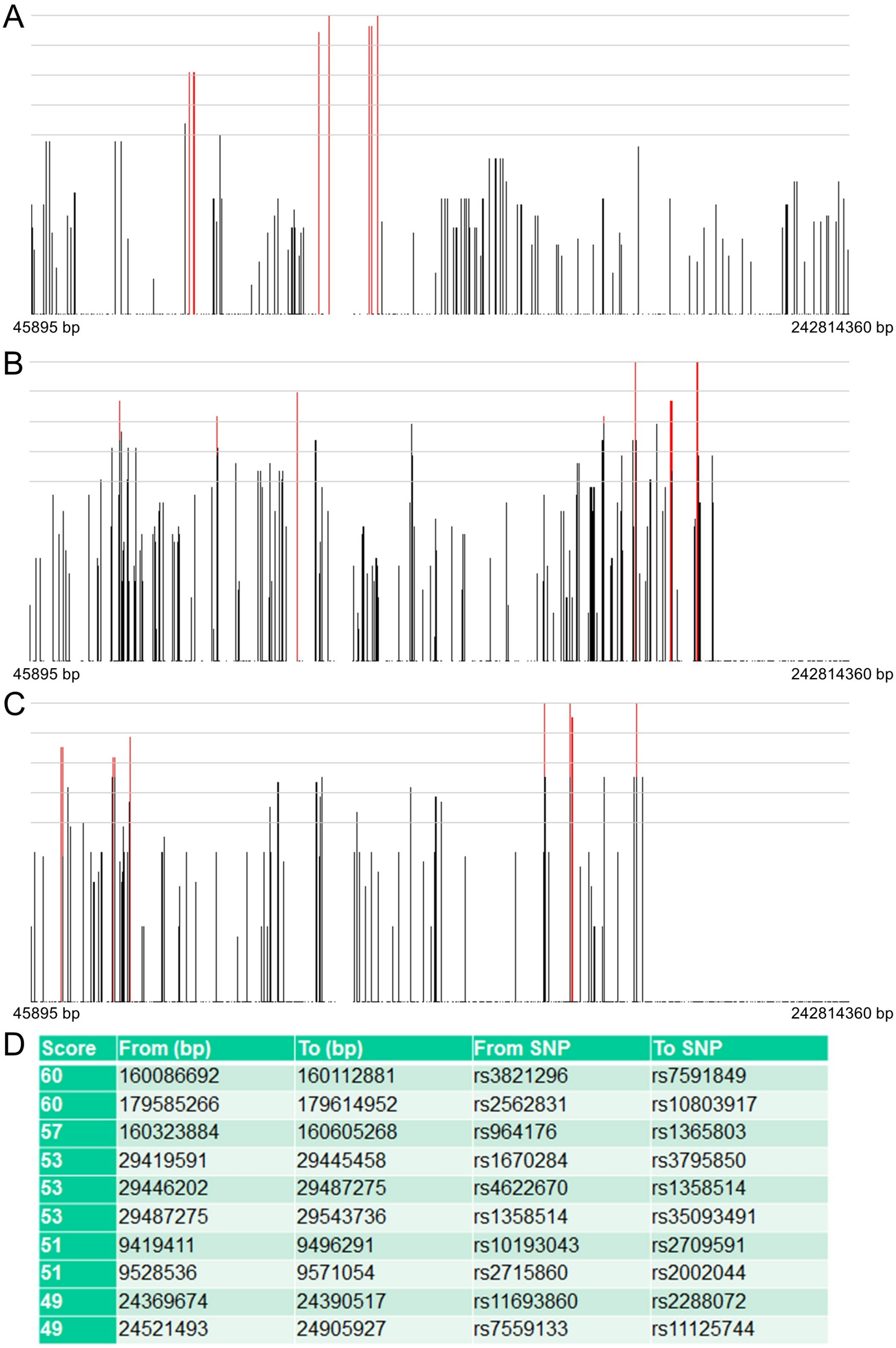


Figure 4.3.: Parametric linkage analysis of Family 1 (a and b), Family 2 (c and d), Families 1 to 9 (e and f) and Families 2 to 9 (g and h). a, c, e and g show linkage analysis for a dominant model and b, d, f and h for a recessive model of chromosome 10. The location on each chromosome in Centimorgan (cM) is displayed on the x-axis and the LOD score on the y-axis. The top grey line in each field marks a LOD-score of 3, the purple line marks a LOD score of 0 and the lower grey bar of -2.



Score	From (bp)	To (bp)	From SNP	To SNP
60	160086692	160112881	rs3821296	rs7591849
60	179585266	179614952	rs2562831	rs10803917
57	160323884	160605268	rs964176	rs1365803
53	29419591	29445458	rs1670284	rs3795850
53	29446202	29487275	rs4622670	rs1358514
53	29487275	29543736	rs1358514	rs35093491
51	9419411	9496291	rs10193043	rs2709591
51	9528536	9571054	rs2715860	rs2002044
49	24369674	24390517	rs11693860	rs2288072
49	24521493	24905927	rs7559133	rs11125744

Figure 4.4.: Homosygosity mapping of individuals from family 2 performed with HomozygosityMapper. **A**, **B** and **C** show the genome-wide homozygosity scores. **A** is based on the SNP genotyping array of individuals II/1, II/2, II/3, II/4 and II/5, whereas **B** displays the results based on the exome sequencing data from patients II/2, II/3 and II/5. **C** shows the results, when both data is combined. Scores are shown as bars. Red coloured bars indicate the most promising genomic regions. Note, that grey lines do not show absolute score values but display the relation to the highest score, which is plotted to the top grey line. These regions are further specified in **D** for the combined analysis data from the SNP genotyping array and exome sequencing.

RNA polymerase *POLR2D* can be found among these which is striking, because the gene *POLR3A* is located inside the disease locus for the dominant Family 1.

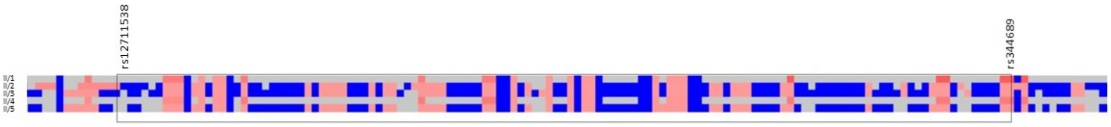


Figure 4.5.: Genotypes for Individuals II/1, II/2, II/3, II/4 and II/5 from rs12711539 and rs344689 (chromosome 2: 121.837.519 - 140.100.106). This figure displays the single genotypes of all samples. Each marker position is depicted as a coloured box. Blue codes for heterozygosity, grey for unknown and red for homozygosity of a certain genotype where longer homozygous stretches are drawn in a darker shade of red than single homozygous markers. The black rectangular surrounds the region from rs12711539 to rs344689 which is heterozygous and shared by all family members. This implies that OPDM might be caused by compound heterozygous mutations in this region.

4.4. Exome Sequencing

4.4.1. OPDM I and II

4.4.1.1. Workflow

After mapping the disease locus in Family 1 to a region around Chromosome 10: 80,000,000bp, whole exome sequencing of affected individuals I/5 (OPDM1) and I/10 (OPDM2) was carried out by MWG Eurofins. The data was downloaded, aligned and variants were called and filtered as previously described in chapter 3.4.5 on page 19. Table 4.2 on page 29 describes the applied filtering steps and the number of remaining variants afterwards. Among the 66,130 variants found in I/5 and the 60,478 in I/10, 24 changes were detected that were rare (allele frequency <1% in the EVS and 1000genomes project), on target (exonic or splice site), heterozygous, protein altering, not in duplicated regions and shared by both patients. Duplicated regions were excluded because variants found here are most likely false positive calls due to misalignment. Among these, 18 variants were found in genes that are expressed in skeletal muscle (information taken from `https://www.ebi.ac.uk/gxa/home`).

4.4.1.2. Variants detected

Table 4.3 on page 30 summarises 18 variants left after intersection and filtering of variants from WES of individuals I/5 and I/10 found in genes expressed in skeletal muscle. Interestingly, most of them do have a dbSNP ID, meaning that they have already been reported as common polymorphisms. Nevertheless, they are very rare with most changes not being listed in EVS or the 1000genomes. Additionally, the

genesymbol	description	startpos
GLI2	GLI family zinc finger 2	121493441
TFCP2L1	transcription factor CP2-like 1	121974163
CLASP1	cytoplasmic linker associated protein 1	122095352
TSN	translin	122513121
GYPC	glycophorin C (Gerbich blood group)	127413426
BIN1	bridging integrator 1	127805599
ERCC3	excision repair cross-complementation group 3	128014866
MAP3K2	mitogen-activated protein kinase kinase kinase 2	128056245
PROC	protein C (inactivator of coagulation factors Va and VIIIa)	128175996
MYO7B	myosin VIIB	128293378
LIMS2	LIM and senescent cell antigen-like domains 2	128395996
GPR17	G protein-coupled receptor 17	128403439
WDR33	WD repeat domain 33	128461808
POLR2D	polymerase (RNA) II (DNA directed) polypeptide D	128603840
SAP130	Sin3A-associated protein, 130kDa	128698791
UGGT1	UDP-glucose glycoprotein glucosyltransferase 1	128848754
HS6ST1	heparan sulfate 6-O-sulfotransferase 1	129023054
RAB6C	RAB6C, member RAS oncogene family	130737235
SMPD4	sphingomyelin phosphodiesterase 4, neutral membrane (neutral sphingomyelinase-3)	130908965
MZT2B	mitotic spindle organizing protein 2B	130939248
IMP4	IMP4, U3 small nucleolar ribonucleoprotein	131100470
PTPN18	protein tyrosine phosphatase, non-receptor type 18 (brain-derived)	131113580
ARHGEF4	Rho guanine nucleotide exchange factor (GEF) 4	131594489
FAM168B	family with sequence similarity 168, member B	131805449
PLEKHB2	pleckstrin homology domain containing, family B (evectins) member 2	131862420
WTH3DI	RAB6C-like	132118065
MZT2A	mitotic spindle organizing protein 2A	132227298
TUBA3D	tubulin, alpha 3d	132233580
C2orf27A	chromosome 2 open reading frame 27A	132479973
C2orf27B	chromosome 2 open reading frame 27B	132552534
GPR39	G protein-coupled receptor 39	133174147
LYPD1	LY6/PLAUR domain containing 1	133402337
MGAT5	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase	134877502
CCNT2	cyclin T2	135676363
MAP3K19	mitogen-activated protein kinase kinase kinase 19	135722076
RAB3GAP1	RAB3 GTPase activating protein subunit 1 (catalytic)	135809835
R3HDM1	R3H domain containing 1	136289036
UBXN4	UBX domain protein 4	136499189
LCT	lactase	136545415
MCM6	minichromosome maintenance complex component 6	136597196
DARS	aspartyl-tRNA synthetase	136664252
CXCR4	chemokine (C-X-C motif) receptor 4	136871919
HNMT	histamine N-methyltransferase	138721808

Table 4.1.: List of genes in possible disease locus on chromosome 2 which are expressed in skeletal muscle and their genomic position on chromosome 2.

changes in *CEP152* and *ELANE* are predicted to be benign by the program MutationTaster making it less likely that one of them is causative for OPDM in patients I/5 and I/10.

Filtering step	**Number of variants**	
	OPDM1	OPDM2
1. All detected variants	66,130	60,478
2. ...filtered for exonic and splice site variants	10,266	10,813
3. ...filtered by the pattern of inheritance (autosomal, heterozygous)	6,384	6,622
4. ...filtered for a frequency of <1% in 1000genome and EVS	1,269	1,173
5. ...filtered excluding synonymous variants	915	848
6. ...filtered excluding variants in duplicated regions	761	719
7. Variants shared by both patients	24	
8. Variants expressed in skeletal muscle (EMBL-EBi-GXA)	18	

Table 4.2.: Filtering pipeline applied to whole-exome sequencing data from individuals OPDM1 and OPDM2. Coding exons and splice sites were defined based on the NCBI annotation.

4.4.1.3. Salient Variants

4.4.1.3.1. *MYPN* c.3605T>A, p.(V1202E) Among the detected 18 variants, especially the change c.3605T>A p.(V1202E) in the gene *MYPN* (NM_032578.3) seems to be striking. This position is highly conserved in other species and the change replaces valine by glutamate, meaning the substitution of a hydrophobic for a positively charged amino acid. This variant is not listed in the EVS, 1000genomes and dbSNP and is predicted to be disease causing by the program MutationTaster. The protein myopalladin is a 147kDa muscle scleroprotein which is located in the I- and Z-discs as well as in the nucleus of cardiac and skeletal muscle cells. It interacts with alpha-actinin as well as nebulin in skeletal muscle and nebulette in cardiac muscle with central and C-terminal domains [65]. These complexes tether actin and titin to the Z-disc. The cardiac ankyrin repeat protein (CARP) is bound by the N-terminus and is considered to be responsible for the control of muscle gene expression [66]. Mutation in *MYPN* are also described to be causative for hypertrophic, dilatative and/or restrictive cardiomyopathy [67].

Chr	Start	Gene	Change	ExonicFunc	ESP5400	dbSNP135	MT Score	MT Pred
chr3	184,429,154	MAGEF1	c.456_457insGGA: p.(L152delinsLE)	nonframeshift insertion		rs34995413		
chr4	119,947,965	SYNPO2	c.441_443del: p.(147_148del)	nonframeshift deletion		rs70944826		
chr5	149,374,881	TIGD6	c.1031delA: p.(Q344fs)	frameshift deletion		rs3832324		
chr6	57,512,565	PRIM2	c.1393T>C: p.(S465P)	nonsynonymous SNV		rs4294008		
chr6	152,737,544	SYNE1	c.6049A>C: p.(T2017P)	nonsynonymous SNV			0,9026	D
chr7	137,612,918	CREB3L2	c.295_297del: p.(99_99del)	nonframeshift deletion		rs3217268		
chr10	69,961,697	MYPN	c.3605T>A: p. (V1202E)	nonsynonymous SNV			0,997522	D
chr10	70,196,983	DNA2	c.1431T>G: p. (S477R)	nonsynonymous SNV		rs80315637		
chr10	79,759,804	POLR3A	c.2551A>G: p.(T851A)	nonsynonymous SNV			0,999613	D
chr11	209,898	RIC8A	c.624_626del: p.(208_209del)	nonframeshift deletion		rs4029256		
chr14	64,468,726	SYNE2	c.3713A>C: p. (D1238A)	nonsynonymous SNV	0,007827	rs75997034	0,958797	D
chr15	49,054,827	CEP152	c.2323T>G: p.(S775A)	nonsynonymous SNV			0,391996	B
chr16	3,119,303	IL32	c.487_488GGA:p.(F163Gfs*21)	frameshift substitution				
chr17	21,319,208	KCNJ12	c.554C>T: p.(A185V)	nonsynonymous SNV		rs73979896	0,999725	D
chr17	21,319,230	KCNJ12	c.576G>C: (p.Q192H)	nonsynonymous SNV		rs1657742	0,993485	D
chr19	856,130	ELANE	c.770C>T: p.(P257L)	nonsynonymous SNV	0,00781	rs17216663	0,005241	B
chr19	49,573,365	KCNA7	c.1324_1326del: p.(442_442del)	nonframeshift deletion		rs3840913		
chr19	50,493,004	VRK3	c.838G>C: p.(A280P)					

Table 4.3.: List of variants shared by I/5 and I/10 after filtering steps. Column 1 shows chromosome number, column 2 the exact position in base-pairs, column 3 the gene symbol, column 4 the exact variant description, column 5 the genetic mechanism. If listed on the EVS server, the allele frequency is provided in column 6 followed by dbSNP ID in column 7 if available. Columns 8 and 9 display pathogenicity prediction by the program MutationTaster.

This change is also in close proximity to the high linkage region (see chapter 4.2 on page 21). It is for these reasons that it was considered the most likely candidate to be causative for OPDM in Family 1 and further analysis was initiated.

First, exon 18 of *MYPN* was amplified by PCR (primer sequences can be found in the addendum A.1 on page 93) and the c.3605T>A variant was confirmed by Sanger sequencing in both patients I/5 and I/10 (see figure 4.7 **A** on page 32 for the sequence of patient I/5)

Subsequently, segregation of the *MYPN* c.3605T>A in all individuals where DNA was available was tested using the restriction endonuclease *DdeI*. Results can be found in figure 4.6 on page 32. Patient I/5 was used as a positive control and a healthy individual as a negative control as presence and absence of the change were previously confirmed by Sanger sequencing. The PCR product has a length of 259bp and *DdeI* cleaves off 50bp in all individuals as the recognition motive 5'-CTNAG-3' is present once in wildtype PCR products and twice in those with the *MYPN* c.3605T>A change. The variant is present, when the remaining 209bp fragment is further digested into a 116bp and a 93bp fragment. All these can be visualised by gel electrophoresis. The presence of the *MYPN* c.3605T>A could be excluded in all unaffected individuals and confirmed in all affected individuals except patient I/15. The referring clinicians were consulted again and asked to state how certain the affection status could be determined as the patient had been assessed at only 15 years of age and had only presented with minimal ptosis [5]. Eventually it was decided to exclude this patient from further studies as the affection status could not be ascertained beyond doubt.

At a later point in the study, DNA from patient I/13 was sent back from Dresden where microsatellite length analysis had been carried out. PCR and Sanger sequencing of *MYPN* exon 18 was done and the absence of the *MYPN* c.3605T>A could be confirmed in this affected individual as shown in figure 4.7 on page 32. As this patient is most certainly affected, presenting with ptosis, ophthalmoparesis, swallowing difficulties, facial atrophy and limb-girdle weakness [5], the *MYPN* c.3605T>A does not segregate with the disease in Family 1.

Before DNA from patient I/13 was available for segregation analysis, the *MYPN* c.3605T>A variant was further evaluated by estimating the frequency in ethnically matched control samples. Therefore, *MYPN* exon 18 was amplified by PCR in 74 individuals from Turkey and frequency was estimated by restriction fragment length polymorphism analysis of PCR products with the restriction endonuclease *DdeI*. Apparently, none of the control individuals harboured the c.3605T>A change, showing that most likely it doesn't constitute a common polymorphism in Turkey.

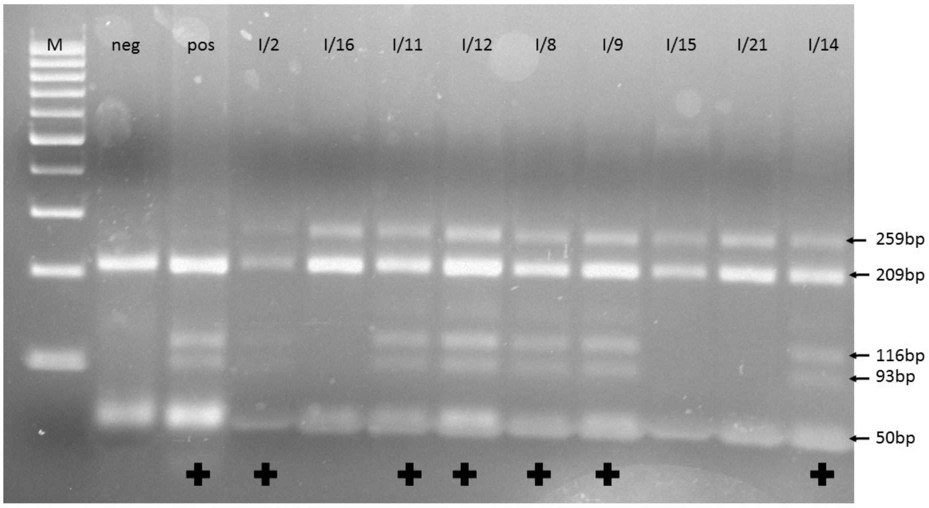


Figure 4.6.: Gel electrophoresis for segregation of the MYPN c.3605T>A change. The *MYPN* ex18 PCR products for members of Family 1 (I/2, I/16, I/11, I/12, I/8, I/9, I/15, I/21, I/14) were completely digested with *DdeI* and the products were analyzed by 2% agarose gel electrophoresis. Lane M indicates DNA marker, 100bp DNA ladder. “pos” and “neg” lanes were loaded with DNA with confirmed presence or absence of the MYPN change. + indicates cleavage of DNA by *DdeI*.

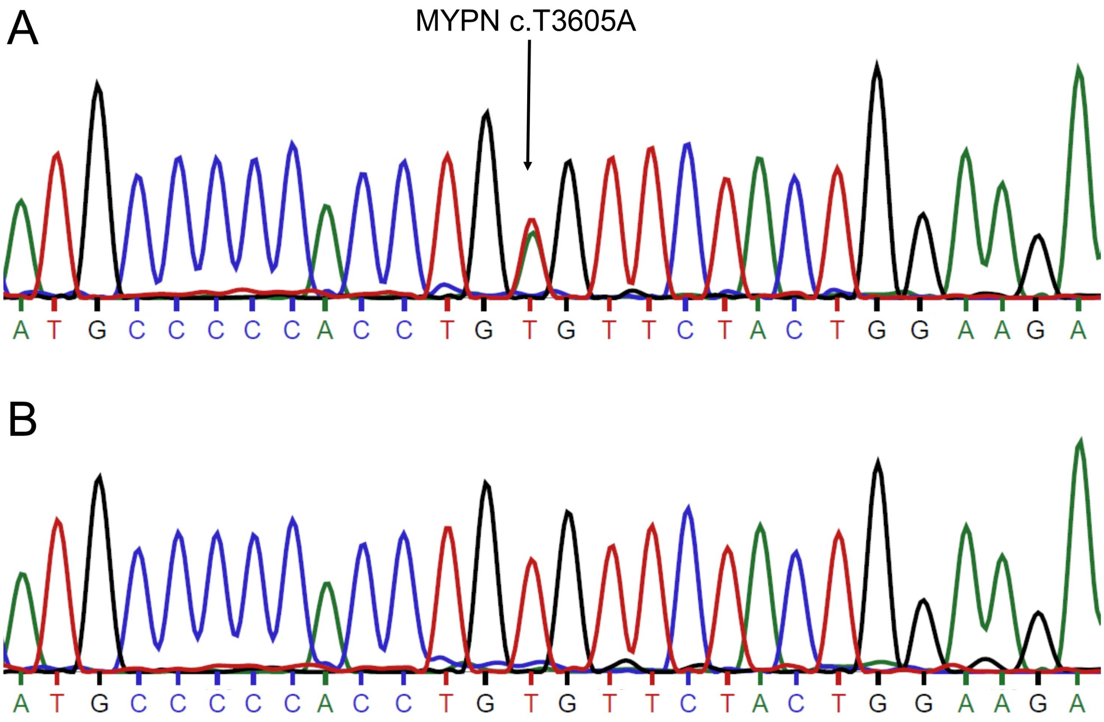


Figure 4.7.: Sanger sequencing of *MYPN* exon 18 of individuals I/5 (**A**)and I/13 (**B**). Presence of the *MYPN* c.3605T>A could be confirmed in I/5 and excluded in I/13.

Next, all exons of *MYPN* were sequenced in one patient each from all other Turkish Families 2,3,4,5,6,7 and 8 (published by Durmus et al. in 2001 [5]) as well as in patients from Finland (FIN/1), France (FRA/1) and Nottingham, UK (NOT). Primer sequences can be found in the addendum A.1 on page 93. All changes detected had a minor allele frequency of at least 1% and could therefore be excluded as possible causes for OPDM in these patients.

4.4.1.3.2. *POLR3A* c.2551A>G, p.(T851A) A second variant, detected by whole exome sequencing of patients I/5 and I/10, is the missense mutation *POLR3A* (NM_007055.3) c.2551A>G; p.(T851A). The encoded protein of *POLR3A* is a subunit of the RNA polymerase III, synthesizing small RNAs [68]. It can also recognise foreign DNA and initiating a consecutive immune response. Mutations in this gene are reported to cause recessive hypomyelinating leukodystrophy [69]. This variant substitutes a highly conserved amino acid - even conserved in C.elegans and Drosophila - and replaces the polar threonine with the hydrophobic alanine, most likely disrupting the protein structure. It is also predicted to create a new splice donor site at the genomic position g.29,499. Accordingly, MutationTaster predicts a disease causing effect. It is also not listed in dbSNP, EVS and 1000genomes.

Presence of the *POLR3A* c.2551A>G variant was confirmed in patients I/5 and I/10 by Sanger sequencing (data not shown). Single patients from all other Turkish families were analysed for the presence of this variant by Sanger sequencing. Apparently, affected individuals from Families 1, 4, 6 and 8 were also carriers of the c.2551A>G variant.

To further analyse the variant, segregation analysis was done by sequencing of *POLR3A* exon 19 in affected and unaffected Family 1 members. PCR products from individuals I/2, I/6, I/8, I/9, I/11, I/12, I/14, I/15, I/16, I/17, I/18 and I/21 were sequenced. Apparently, the variant segregates well with the disease in all individuals but I/15 and I/6. These two belong to the younger generation and according to Durmus et al. 2011 [5], do not have any weaknesses, only minimal ptosis. The age of onset as well as the initial symptom could not be determined in I/6. Therefore, the clinical data was not strong enough to exclude the *POLR3A* c.2551A>G variant from further analyses. Hence, one affected family member from all Turkish families has been screened for the *POLR3A* c.2551A>G variant using Sanger sequencing with the result that it was present in individuals IV/3, VI/2 and VIII/2. Segregation of the variant with the disease in families 4, 6 and 8 as well as the presence in one sporadic patient from Nottingham, UK, France and Canada each was tested using the restriction endonuclease *BsrDI* as described in chapter 3.3.4 on page 16. *BsrDI* cleaves the 236bp PCR fragment when the POLR3A c.2551A>G is present producing a 152bp and a 84bp fragment. There is no recognition site for the restriction endonuclease in the wildtype PCR product. Results can be found in figure 4.8 on page 34. Apparently, the change segregates in all families with the disease, given that individual IV/1 is indeed unaffected. The *POLR3A* c.2551A>G variant could not be detected in any of the non-Turkish patients from Nottingham, UK, Finland and France.

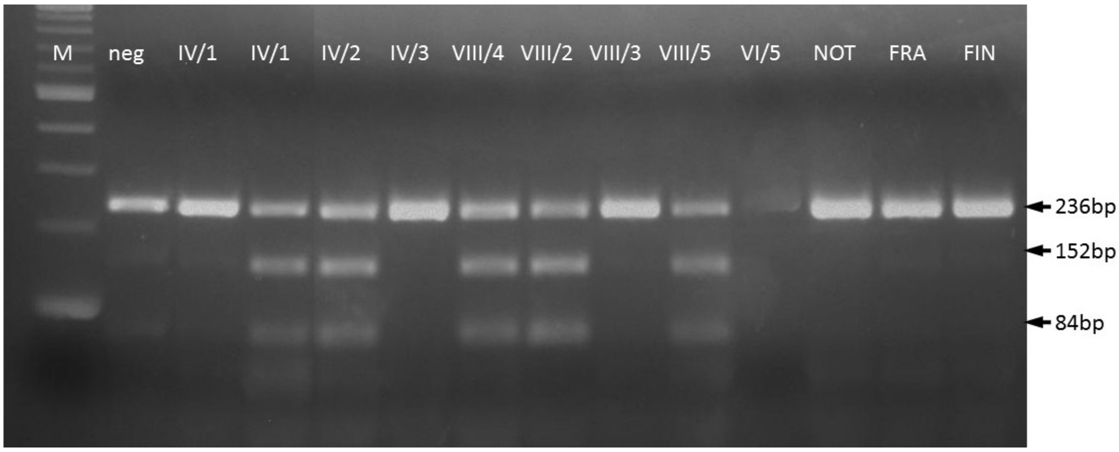


Figure 4.8.: Gel electrophoresis for segregation of the *POLR3A* c.2551A>G change. The *POLR3A* exon 19 PCR products for family 4, 6 and 8 Family members as well as sporadic patients from Nottingham (NOT), France (FRA) and Finland (FIN), named in the top row, were completely digested with *BsrDI* and the products were analyzed by 2% agarose gel electrophoresis. Lane M indicates the 100bp DNA Ladder. "neg" lanes were loaded with DNA with confirmed absence of the c.2551A>G change.

Even though the *POLR3A* c.2551A>G variant did not perfectly segregate in Family 1, the frequency of this change was determined in 58 Turkish control samples. PCR Products of *POLR3A* exon 19 were digested with the restriction endonuclease *BsrDI* and evaluated by gel electrophoresis as described before. Apparently, the change was present in 22 out of 58 individuals resulting in a minor allele frequency of more than 37%. Therefore, the *POLR3A* c.2551A>G could be excluded as being causative for OPDM. However, the presence of this polymorphism in families with a dominant pattern of inheritance implies, that these families share a disease allele including the c.2551A>G variant.

4.4.2. Exome Sequencing of Individuals OPDM III-VIII

4.4.2.1. Exome Sequencing Results OPDM III-V

As exome sequencing of individuals I/5 and I/10 did not reveal any likely candidate genes, the study was extended and whole exome sequencing was carried out on 5 more patients. Three siblings from Family 2 (II/2, II/3 and II/5 called OPDM3, OPDM4 and OPDM5 respectively) as well as sporadic patients from Nottingham, UK (OPDM6), from Finland (OPDM7) and from Canada (OPDM8) were chosen and DNA was sent to Otogenetics, Atlanta, US, where the exome enrichment was done with the Agilent V4 kit and sequencing done on a Illumina HiSeq 2000. Bioinformatic analysis was done as previously described in chapter 3.4.5.2 on page 19. Additionally, an “in-house frequency” database was used to filter out false positive calls. This database consists of all exome sequencing data sets from the Institute of Genetic Medicine in Newcastle-upon-Tyne, UK, and provides a minor allele frequency for every detected variant.

Table 4.4 summarises filtering steps of individuals OPDM3, OPDM4 and OPDM5 and the number of variants left after each of them.

Filtering step	**Number of variants**		
	OPDM3	OPDM4	OPDM5
1. All detected variants	79,915	77,898	77,877
2. ...filtered for exonic and splice site variants	18,250	16,924	17,565
3. ...filtered by the pattern of inheritance (homozygous/comp. het.)	7,028	6,201	6,561
4. ...filtered for a frequency of <1% in 1000genome and EVS	197	185	178
5. ...filtered excluding synonymous variants	187	181	167
6. ...filtered excluding variants in duplicated regions	180	177	163
7. Variants shared by both patients	104		
8. Variants expressed in skeletal muscle (EMBL-EBi-GXA)	84		
9. Variants with an in-house frequency of <2%	10		

Table 4.4.: Filtering pipeline applied to whole-exome sequencing data from individuals OPDM3, OPDM4 and OPDM5. Coding exons and splice sites were defined based on the NCBI annotation.

After all variants of individuals OPDM3, OPDM4 and OPDM5 are filtered and intersected 84 genetic changes remain. Among these 84, 78 are listed in dbSNP. However, no single change is listed in the EVS and the 1000genome database. When the in-house minor allele frequency of these genetic changes is determined, it turns out, that almost all missense variants are very commonly detected in whole exome sequencing data, thus most likely being false positive calls specific to the bioinformatical pipeline. When all variants with an in-house frequency greater than 2% are filtered out, no single homozygous variant remains in the list and only 10 changes could be identified, where two or more can be found in one gene, making compound heterozygosity possible. These variants can be found in table 4.5 on page 37.

As it is very likely that a compound heterozygous mutation on Chromosome 2 is causing OPDM in Family 2, genes with two or more heterozygous changes within the heterozygous region described in chapter 4.3 on page 23 have been identified. This region expands from rs12711538 to rs344689 (121,747,406 - 140,100,106bp according to the hg19 assembly) on Chromosome 2. No single variant could be detected within this region that is shared or not covered by all three siblings.

Accordingly, special focus was put on the high linkage region on Chromosome 10. No rare, protein altering changes could be detected between chr10: 49,968,432 and 93,220,242 in any of the three siblings that are shared among them.

4.4.2.2. Exome Sequencing Results OPDM VI

Table 4.6 summarises filtering steps of individual OPDM6 from Nottingham, UK (patient was assessed and DNA provided by Dr Paul Maddison) and the number of variants left after each of them.

As the pattern of inheritance cannot be determined in sporadic patients, variants were filtered for both a dominant and a recessive model. Special focus was put on variants in the high linkage regions on Chromosome 2 and 10. The only detected variants near the high linkage locus are a heterozygous *AGAP5*(NM_001144000.1) c.673A>G, p.(M225V) change at the chromosomic position 75,435,676 and a heterozygous *PIK3AP1* (NM_152309.2) c.775G>A; p.(V259I) at the chromosomic position 98,411,346. AGAP5 is an ankyrin repeat and GTPase domain and *PIK3AP1* a phosphoinositide-3-kinase adaptor protein. Mutations in both genes are very likely not causing damage to skeletal muscle tissue. On Chromosome 2 two variants could be detected in *NEB* (NM_001164507.1), coding for nebulin, a giant protein component of the cytoskeletal matrix that coexists with the thick and thin filaments within the sarcomeres of skeletal muscle [70]. Mutations in this gene are associated with recessive nemaline myopathy [71]. These heterozygous changes are c.21044C>G, p.(S7015C) and c.22122C>G, p.(D7374E). However, none of them are listed in the Leiden Open Variation Database (LOVD), a database that lists all published disease causing variants in genes known to cause neuromuscular disorders (`http://www.dmd.nl`). Additionally, the c.15941C>G variant is listed in ClinVar, a database that summarises reports of the relationships among human variations and phenotypes, and is labelled ‘likely benign’ (`http://www.ncbi.nlm.nih.gov/clinvar`).

Chr.	Gene	AAChange	OPDM3	OPDM4	OPDM5	OPDM6	OPDM7	OPDM8	1000g
chr11	MUC2	c.5014T>G, p.(S1672A)	R/V	R/V	R/V	R/V	R/V	R/V	
chr11	MUC2	c.4876A>T, p.(I1626F)	R/V	R/V	R/V	R/V	R/V	R/V	
chr19	MBD3L3	c.622A>G, p.(R208G)	R/V	R/V	R/V	V/V	R/V	R/V	
chr19	MBD3L3	c.619T>G, p.(C207G)	R/V	R/V	R/V	R/V	R/V	R/V	
chr19	FBN3	c.6397G>A, p.(G2133S)	R/V	R/V	R/V	R/R	R/R	R/R	0,0009
chr19	FBN3	c.5399G>A, p.(G1800D)	R/V	R/V	R/V	R/R	R/R	R/R	
chr2	TTN	c.92522G>A, p.(C30841Y)	R/V	R/V	R/V	R/R	R/R	R/R	
chr2	TTN	c.87147T>A, p.(D29049E)	R/V	R/V	R/V	R/R	R/R	R/R	0,0014
chr2	TTN	c.11138C>G, p.(T3713S)	R/V	R/V	R/V	R/R	R/R	R/R	0,0014
chr2	TTN	c.1267A>C, p.(S423R)	R/V	R/V	R/V	R/R	R/R	R/R	

Table 4.5.: List of possible compound heterozygous variants shared by OPDM3, OPDM4 and OPDM5. Additionally, genotypes for individuals OPDM6, OPD7 and OPDM8 for these changes are provided.'R' stands for reference allele and 'V' for variant, meaning that 'R/R' implies the presence of two reference alleles, 'R/V' heterozygosity and 'V/V' homozygosity for the alternative allele. Coding positions refer to the following transcript variants: NM_002457.3 for *MUC2*, NM_001164425.1 for *MBD3L3*, NM_032447.3 for *FBN3*, NM_133378.4 for the *TTN* variants

Filtering Step	**Number of variants**
1. All detected variants	76,103
2. . . . filtered for exonic and splice site variants	16,769
3. . . . filtered for a frequency of less than 2% in 1000genomes and ExomeVariantServer	1,569
4. . . . filtered excluding synonymous variants	1,203
5. . . . filtered excluding variants in duplicated regions	1,099
6. . . . filtered excluding variants with an in-house frequency of less than 2%	676
7. variants for a dominant model	509
8. variants for a recessive model	141

Table 4.6.: Filtering pipeline applied to whole-exome sequencing data from individual OPDM6. Coding exons and splice sites were defined based on the NCBI annotation. The dominant model includes all heterozygous variants with a frequency of less than 1% in EVS, 1000genomes and in-house minor allele frequency (MAF) . The recessive model comprises all variants with a frequency of less than 2% in the databases mentioned above that are either homozygous or where two or more changes were detected in one gene.

Also, a total number of 4 heterozygous genetic changes in the gene *TTN* could be detected. This gene encodes a large structural protein of striated muscle. It expands from the Z-disc of the sarcomere with its N-terminus to the M-line with its C-terminus. It also possesses binding sites for other muscle associated genes and acts as a template for the contractile machinery of muscle fibres [70]. As it is one of the largest genes in the human genome great variability exists, especially in the Z-disk-, the M-line- and the I-band regions. Mutations in titin are associated with a number of pathologies such as dilated cardiomyopathy, autosomal dominant tibial muscular dystrophy as well as autosomal recessive limb girdle muscular dystrophy 2J [72], [73]. These variants in the *TTN*-gene (NM_003319.4) are: c.41935C>T, p.(P13979S), c.18052C>T, p.(R6018W), c.11491A>T, p.(I3831F) and c.1492G>A, p.(V498I). However, none of them is listed in the LOVD database but the first two and the last one are listed as ‘benign’ or ‘likely benign’ in ClinVar.

One interesting variant could be detected in the *RYR1* gene (NM_000540.2), a heterozygous c.8382C>G, p.(Y2794X) nonsense mutation. This variant is not listed in the EVS and the 1000genome database and is most certainly disrupting the protein structure as a premature stop-codon leads to the loss of almost half the protein chain. This large gene, counting 106 exons, located at 19q13.2, encodes the protein Ryanodin Receptor 1, functioning as a calcium release channel in the sarcoplasmic reticulum in skeletal muscle. Its function is also to connect the sarcoplasmic reticulum to the transverse tubule. Mutations in this gene reportedly cause autosomal dominant or recessive central core myopathy, autosomal recessive minicore myopathy with external ophthalmoplegia and malignant hyperthermia susceptibility [74],

[75],[76],[77]. Hence, it needs to be discussed, if central core myopathy could mimic an OPDM phenotype and if this variant could be responsible for it.

4.4.2.3. Exome Sequencing Results OPDM VII

Table 4.7 summarises filtering steps of individual OPDM7 from Finland and the number of variants left after each of them (patient was assessed and DNA provided by Professor Bjarne Udd).

Filtering Step	**Number of variants**
1. All detected variants	89,346
2. . . . filtered for exonic and splice site variants	16,702
3. . . . filtered for a frequency of less than 2% in 1000genomes and ExomeVariantServer	1642
4. . . . filtered excluding synonymous variants	1,275
5. . . . filtered excluding variants in duplicated regions	1,181
6. . . . filtered excluding variants with an in-house frequency of less than 2%	715
7. variants for a dominant model	530
8. variants for a recessive model	132

Table 4.7.: Filtering pipeline applied to whole-exome sequencing data from individual OPDM7. Coding exons and splice sites were defined based on the NCBI annotation. The dominant model includes all heterozygous variants with a frequency of less than 1% in EVS, 1000genomes and in-house MAF. The recessive model comprises all variants with a frequency of less than 2% in the databases mentioned above that are either homozygous or where two or more changes were detected in one gene.

Variants were analysed on the basis of a recessive and a dominant model as previously described in chapter 4.4.2.2 on page 36. Special focus was put on the high linkage regions on Chromosome 2 and 10 as well as on changes in genes which are associated with neuromuscular disorders. Among the 3 homozygous variants left after all filtering steps, none were found in the high linkage regions on Chromosome 2 or 10 or otherwise interesting. Two variants in the gene *GLI2*, encoding a zinc finger and transcription factor of Sonic hedgehog signaling could be detected on Chromosome 2 [78]. Variants in this gene are associated with various phenotypes of malformation [79]. The identified heterozygous variants in *GLI2* (NM_005270.4) are: c.4332G>A, p.(M1444I) and c.4333C>T, p.(L1445F). However for a dominant model, two variants could be detected on Chromosome 10 that are located in the high linkage region. These are *POLR3A* (NM_007055.3) c.275G>C, p.(C92S) and *NRG3* (NM_001010848.3):c.1951G>A, p.(E651K) at the chromosomic positions 79,785,423 and 84,745,221 respectively. Both variants cause the substitution of conserved amino acids and are predicted to be deleterious by various prediction tools such as SIFT and MutationTaster. Both changes are listed in the EVS and the 1000genome database but have a very low frequency. The *POLR3A* change is annotated with frequencies of 0,000093 in the EVS and 0,0005 in the 1000genome database and the *NRG3* change with 0,002417 and 0,0023 accordingly. The protein encoded by *POLR3A* has been described previously in 4.4.1.3.2 on page 33. *NRG3* encodes a ligand for the transmembrane tyrosine kinase ERBB4 which is a member of the epidermal growth factor receptor family [80]. Neuregulin 3 is thought to influence neuroblast proliferation, migration and differentiation through ERBB4 [81]. It is susceptible to be associated with schizoaffective disorders and schizophrenia [82]. Other interesting variants include a heterozygous change in the *RYR1* gene which was described above. This c.785C>T, p.(A262V) change is not listed in the EVS and has a frequency of 0,0005 according to the 1000genomes database. It is found in a highly conserved position but mutation prediction tools are inconsistent with some predicting a deleterious effect (PolyPhen2, PhyloP) whereas other claim that is most likely benign to the protein structure (MutationTaster, LRT). This is most likely, because the amino acid substitution replaces a non-polar alanine with the likewise non-polar valine. The presence of this variant was confirmed by Sanger sequencing in this patient. Nevertheless, this is the second individual with a variant in the gene *RYR1* and it needs to be discussed, if these patients are affected with central core myopathy instead of OPDM.

4.4.2.4. Exome Sequencing Results OPDM VIII

Table 4.8 summarises filtering steps of individual OPDM8 from Canada and the number of variants left after each of them (patient was assessed and DNA provided by Professor Bernard Brais).

As in OPDM VI and OPDM VII, especially variants within the high linkage regions on Chromosome 2 and Chromosome 10 as well as those within genes associated with neuromuscular disorders were taken account of. For a recessive model, none of the three homozygous variants (*NDUFS7* c.T617A, p.(L206H) on Chromosome 19, *LILRB1* c.893C>A, p.(S298Y) on Chromosome 19 and *IFNA10* c.496G>A, p.(V166I) on Chromosome 9) seem to be likely disease-causing. However, three variants were detected in the *TTN* gene on Chromosome 2, namely c.4332G>A, p.(M1444I), c.12571G>A, p.(V4191M) and c.10366G>A, p.(V3456I). Out of these, only the c.12571G>A change is listed with a frequency of 0,007965 in the ESP5400 and 0,0046 in the 1000genomes database. None of these variants is listed in the LOVD database. The first variant is a substitution of the non-polar amino acid methionine with the likewise non-polar isoleucine. Similarly, the non-polar valine is

Filtering Step	Number of variants
1. All detected variants	75,682
2. ...filtered for exonic and splice site variants	18,011
3. ...filtered for a frequency of less than 2% in 1000genomes and ExomeVariantServer	1,800
4. ...filtered excluding synonymous variants	1,345
5. ...filtered excluding variants in duplicated regions	1,253
6. ...filtered excluding variants with an in-house frequency of less than 2%	667
7. variants for a dominant model	516
8. variants for a recessive model	119

Table 4.8.: Filtering pipeline applied to whole-exome sequencing data from individual OPDM8. Coding exons and splice sites were defined based on the NCBI annotation. The dominant model includes all heterozygous variants with a frequency of less than 1% in EVS, 1000genomes and in-house MAF. The recessive model comprises all variants with a frequency of less than 2% in the databases mentioned above that are either homozygous or where two or more changes were detected in one gene.

changed to methionine in the second change and valine replaced with isoleucine in the third. Thus, it is very hard to predict a pathogenic potential for any of these variants. Close to the locus on Chromosome 10, two heterozygous variants were detected in the genes *PLAU* and *WAPAL* at chromosomic locations 75,675,086 and 88,230,804 respectively. *PLAU* encodes a urinary plasminogen activator, involved in thrombolysis and mutations in this gene are associated with Quebec Platelet Disorder [83]. *WAPAL* (wings-apart-like homolog from Drosophila) is involved in the removal of cohesins from the mitotic human chromosomes and therefore act to protect from segregation errors and aneuploidy [84]. Considering the gene function as well as the genomic locations, both changes are an improbable cause of OPDM in this individual.

However, three heterozygous changes in genes, known to be associated with neuromuscular disorders, could be detected. The first one is a c.1564G>A, p.(G522R) change in the gene *MEGF10*. The protein encoded by this gene is involved in cell motility, proliferation as well as adhesion. It also plays a role in cell phagocytosis during apoptosis and amyloid-beta uptake in the brain [85]. Mutations in this gene cause either autosomal recessive early-onset myopathy, areflexia, respiratory distress, and dysphagia (EMARDD) or congenital myopathy with minicores [86], [87]. The c.1564G>A, p.(G522R) variant replaces a non-polar glycine with the positively charged basic polar amino acid arginine, most likely altering the protein structure. It has a frequency of 0.002231 in ESP5400 and 0.0009 in the 1000genomes database and is predicted to be deleterious by all prediction tools.

The second change is a heterozygous c.655C>T, p.(R219X) nonsense variant in *MYOT*, a gene encoding for myotilin, which has been confirmed by Sanger sequencing. This protein binds several skeletal muscle structural proteins such as F-actin and alpha-actinin and plays a crucial role in stabilising and anchoring of thin filaments of the sarcomere [88]. Mutations in this gene are associated with a number of autosomal dominant neuromuscular conditions; namely limb-girdle muscular dystrophy, Type 1A [89], myofibrillar myopathy 3 [90] and spheroid body myopathy [91]. This variant has not been published yet and is not listed in the LOVD. As the nonsense mutation most certainly affects protein integrity it needs to be discussed, if this mutation could be the underlying genetic defect for this patient's phenotype.

In addition to the changes mentioned above, a heterozygous c.313C>T, p.(R105C) variant in the gene *MATR3* at the genomic position chr5:138,660,985 in the hg19 assembly was detected. Matrin 3 is a nuclear matrix protein that binds DNA and RNA [92]. It is known to cause autosomal dominant Amyotrophic Lateral Sclerosis 21, formerly called vocal cord and pharyngeal dysfunction with distal myopathy (VCPDM) [93], [94]. This variant is a substitution of the polar and positively charged amino acid arginine with the non-polar cysteine and is conclusively most likely altering the protein structure. However, this variant is located at the exon-intron boundary of intron 13 and exon 14 in only some of the common transcript variants (e.g. NM_018834.5). It also changes the strength of two splice acceptor sites at positions g.51556 (wt:0.42/mu:0.56) and g.51547 (wt:0.24/mu:0.36) according to the calculations of the program MutationTaster as shown in figure 4.10 on page 43. Other splice site prediction tools have been used to further validate this *in silico* analysis, ASSP (alternative splice site prediction; `http://wangcomputing.com/assp/`) and Fruitfly (`http://www.fruitfly.org/seq_tools/splice.html`) [95], [96]. The score for the alternative splice acceptor site (Exon 14a alt. in figure 4.10 **c**) increases from 7.012 to 7.609 according to ASSP. The score for the splice acceptor site, resulting in exon 14a (figure 4.10 **b**), decreases from 8.424 to 8.020. Fruitfly predicts no change in the strength for this site (0.89 -> 0.90) but an increase form 0.50 to 0.70 for the cryptic acceptor site resulting in exon14a alt.

Therefore it needs to be discussed if the frequency of this particular transcript variant is increased by the altered splice acceptor site strength and, if that is the case the missense mutation disrupts the protein structure.

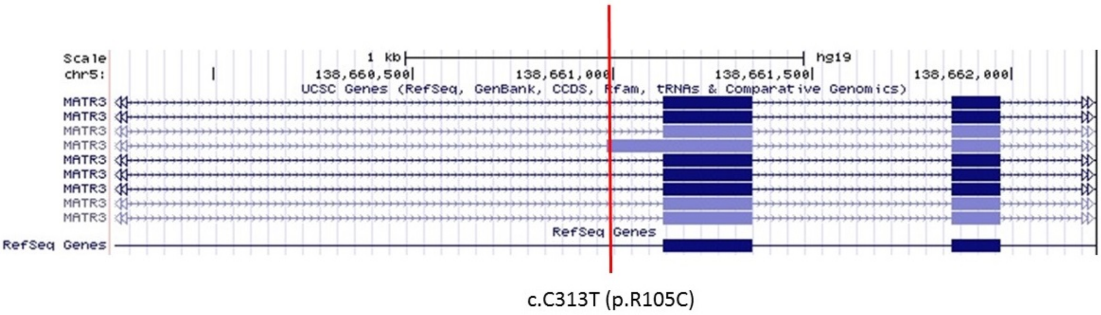


Figure 4.9.: Position of the variant in *MATR3*: c.313C>T, p.(R105C). The nine lines comprise nine different UCSC transcript variants. Transcripts in darker blue accord the consensus coding sequence (`https://www.ncbi.nlm.nih.gov/projects/CCDS/CcdsBrowse.cgi`). Note that the variant only affects the protein sequence of transcript variant NM_018834.5. However, it alters the strengths of splice acceptor sites and might therefore increase the frequency of NM_018834.5.

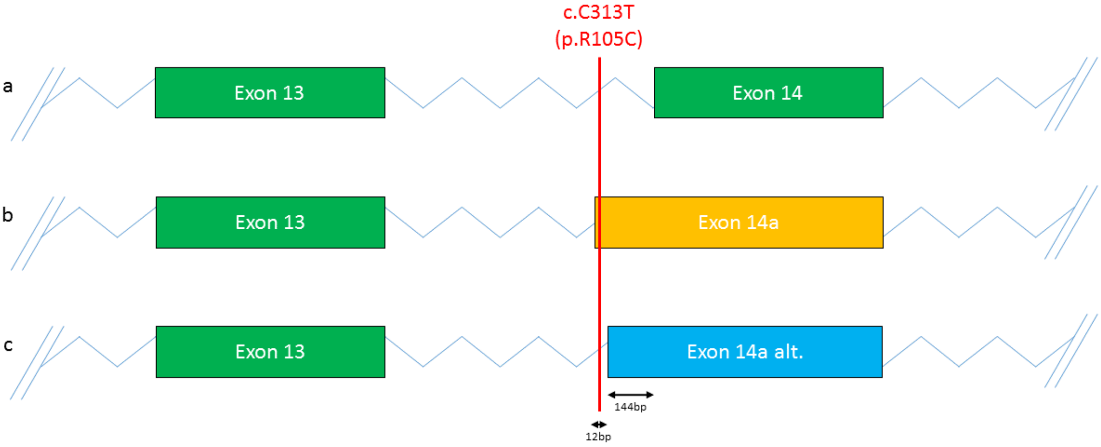


Figure 4.10.: Predicted isoforms of *MATR3* and the *in silico* predicted effect of the c.313C>T variant. **a** shows the most common and **b** the 144bp longer alternative transcript variant where the c.313C>T, p.(R105C) is located in exon 14a. This variant, however, also creates a cryptic splice acceptor site resulting in a 14bp smaller alternative exon 14a, displayed in **c**.

4.5. Fine mapping the disease locus for Family 1

4.5.1. Fine mapping of the locus on Chromosome 10

Fine mapping of the locus on Chromosome 10 was undertaken by haplotype construction and identification of recombinant haplotypes by use of genotyping data for the 53 markers that span the region of interest. Among these, 18 are microsatellites and 35 are SNP markers with a frequency greater than 5%. A list of the microsatellite markers can be found in table 3.5 on page 19. As the affection status of the youngest generation could not be determined with confidence, a total of 9 affected individuals from generations 3 and 4 were studied and several recombinant haplotypes identified. Reconstruction was carried out in a way that ensured that the largest possible chromosomic region was covered. The most telomeric microsatellite marker, D10S1686, turned out to be uninformative. Therefore, the telomeric recombination point located at chromosomic position 85,566,388bp had to be determined based on SNP data to rs1188786, most likely overestimating the size of the locus (all positions referring to the hg19 assembly). The centromeric recombination point could be identified by sequencing a synonymous *POLR3A* (NM_007055.3) c.2829C>T change that did not segregate with the disease in all patients and is therefore determined at chromosomic position 79,750,884bp. The identified disease locus consequently spans from position 79,750,884 to 85,566,388 within Chromosome 10. The reconstruction of recombinant haplotypes and the recombination points can be found in figure 4.11 on page 46.

4.5.2. Characterization of the locus

Given, that a conserved haplotype could be identified, the region on Chromosome 10: 79,750,884 - 85,566,388 was further analysed. Figure 4.12 shows a schematic Chromosome 10 q22.3q23.1 area linked to autosomal dominant OPDM as well as all genes located in this region. A total number of 44 genes could be found of which 28 are protein-coding. 19 of them are reported to be expressed in skeletal muscle according to either the in Common Fund's Genotype-Tissue Expression (GTEx) [93]database (`http://www.gtexportal.org/home/`) or the Human Protein Atlas (HPA) (`http://www.proteinatlas.org/`) [97], [98]. A list of these can be found in table 4.9 on page 45 and 45.

Gene Symbol	Expression in GTEx	Expression in HPA	Gene Function
POLR3A	X		Catalytic component of RNA polymerase III, which synthesizes small RNAs. Also acts as a sensor to detect foreign DNA and

			trigger an innate immune response.
RPS24	X		Encodes a ribosomal protein that is a component of the 40S subunit. Mutations result in Diamond-Blackfan anemia.
PLAC9		X	Homo sapiens placenta-specific 9
ZMIZ1	X		Member of the PIAS (protein inhibitor of activated STAT) family of proteins. It regulates the activity of various transcription factors, including the androgen receptor, Smad3/4, and p53. It may also play a role in sumoylation. A translocation between this locus on Chromosome 10 and the protein tyrosine kinase ABL1 locus on chromosome 9 has been associated with acute lymphoblastic leukemia
DYDC1	X	X	Member of a family of proteins that contains a DPY30 domain. It is involved in acrosome formation during spermatid development.
PPIF	X		Member of the peptidyl-prolyl cis-trans isomerase (PPIase) family. PPIases catalyse the cis-trans isomerization of proline imidic peptide bonds and accelerate the folding of proteins. Part of the mitochondrial permeability transition pore in the inner mitochondrial membrane. Activation of this pore may be involved in the induction of apoptotic and necrotic cell death.
ZCCHC24	X	X	Zinc finger, contains a CCHC domain
SFTPD	X		Part of the innate immune response, protecting the lungs against inhaled microorganisms and chemicals. May also be involved in surfactant metabolism
SH2D4B		X	SH2 domain containing 4B
EIF5AL1	X		Eukaryotic translation initiation factor 5A-like 1
TMEM254	X		Transmembrane protein 254.
ANXA11	X	X	Member of the annexin family, a group of calcium-dependent phospholipid-binding proteins. It is recognized by sera from patients with various autoimmune diseases.
FAM213A	X	X	Involved in redox regulation of the cell. Acts as an antioxidant. Inhibits TNFSF11-induced NFKB1 and JUN activation and osteoclast differentiation. May affect bone resorption and help to maintain bone mass.
TSPAN14	X	X	Tetraspanin 14.

Table 4.9.: List of genes and their function in the refined locus on Chromosome 10: 77,991,127 - 84,563,458. The gene information was extracted from: `http://www.ncbi.nlm.nih.gov/mapview/`. Source for gene-expression information is: `https://www.ebi.ac.uk/gxa/experiments/`.

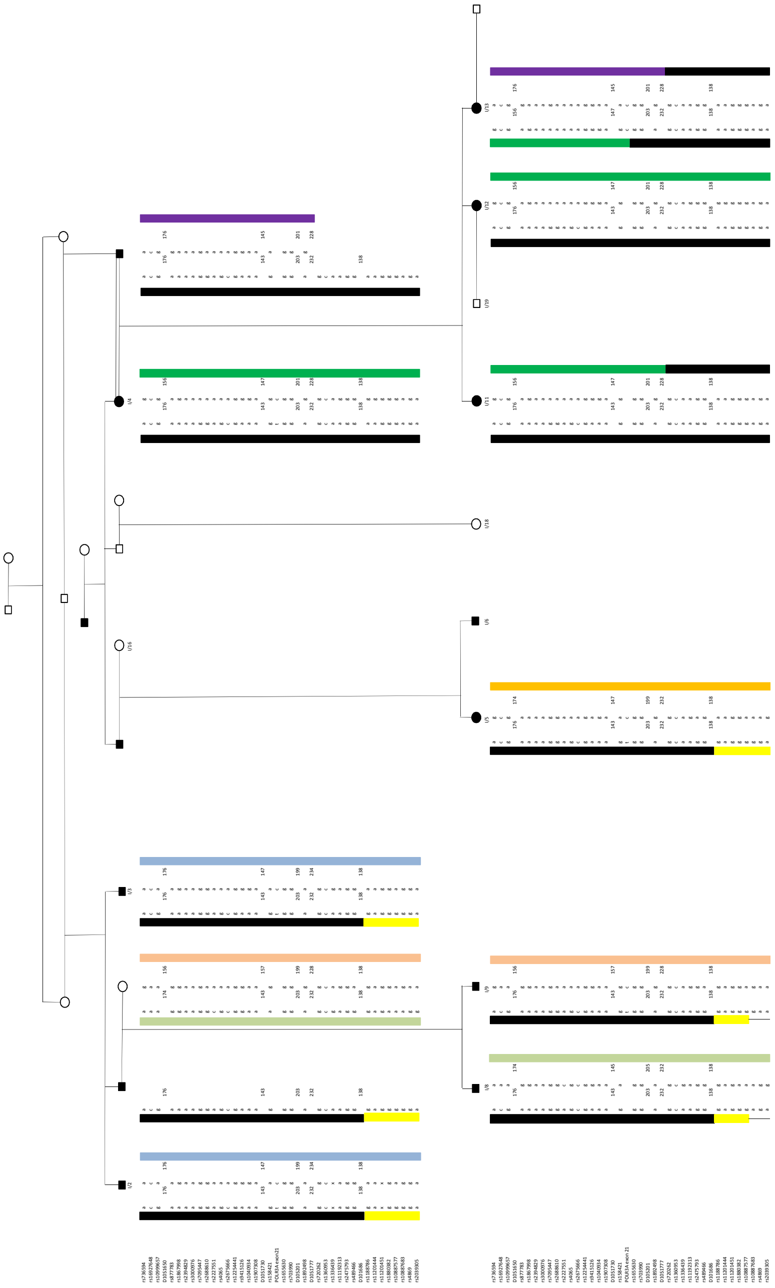


Figure 4.11.: Reconstruction of Family 1 recombinant haplotypes for Chromosome 10: 77,991,127-84,563,458 based on 50 markers. The disease haplotype is shown in black colour. The youngest generation was not considered as their affection status could not be determined with confidence. Haplotypes were reconstructed manually to result in the largest possible shared haplotype.

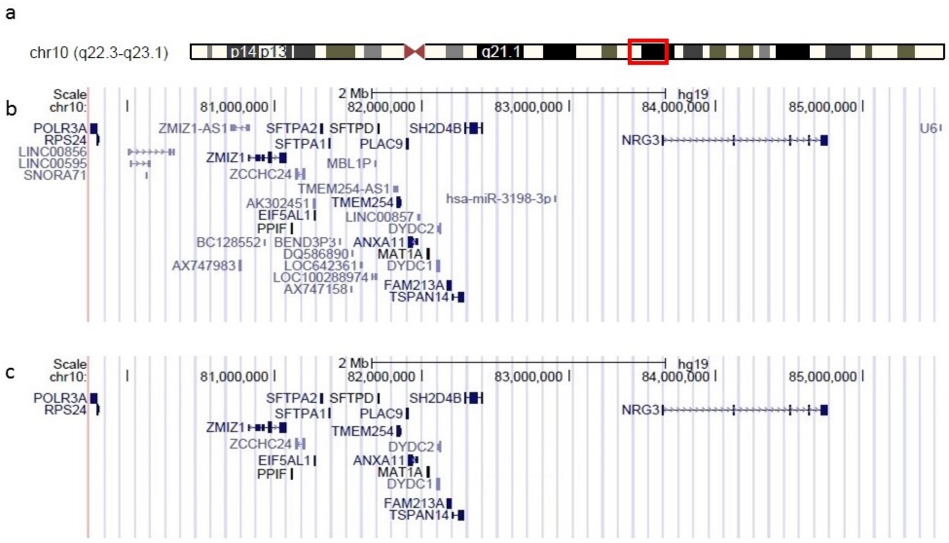


Figure 4.12.: Characterisation of the shared haplotype region Chromosome 10: 77,991,127-84,563,458. **(A)** Chromosomal location of the disease locus indicated by the red box: Chr10 (q22.3q23.1). **(B)** Map of all genes found in the locus. **(C)** Map containing all protein coding genes inside the locus.

4.6. Chromosome 10 Locus Sanger Sequencing

Since whole exome sequencing did not cover all coding exons of the 44 genes in the disease locus for Family 1 with a read depth of at least 10, Sanger sequencing was used to screen all of them for undetected variants. A list of primers used can be found in the addendum A.1 on page 93. Uncovered exons were amplified by PCR and sent to MWG Eurofins for Sanger-sequencing. These were found in genes *NRG3, RPS24, ZMIZ1, PPIF, EIF5AL1, C10orf57, PLAC9, ANXA11, SH2D4B, AK302451, AX747983, ZCCHC24, FAM213A, MAT1A, GHITM, CDHR1, LRIT2, LRIT1*. No single mutation could be detected that is not listed in the dbSNP database with a frequency greater than 1% apart from a *EIF5AL1* c.254A>G, p.(K85R) variant.

4.6.1. MicroRNAs Within the Disease Locus for Family 1

As the results so far do not show any convincing variants in the coding sequences of genes within the disease locus for Family 1, the two microRNAs found in this region, *MIR_554* (chr10: 83,467,245-83,467,350) and *hsa-miR-3198-3p* (chr10: 82,904,458-82,904,477), were analysed by Sanger sequencing. Primer sequences can be found in the addendum in A.1 on page 93. No variants that are not listed with a frequency of less than 1% in dbSNP could be detected in these two microRNAs.

4.6.2. *EIF5AL1* c.254A>G, p.(K85R)

When uncovered exons within the locus were sequenced, a rare missense variant in the gene *EIF5AL1* (NM_001099692.1) was discovered: c.254A>G, p.(K85R). This change is listed in dbSNP with the ID rs201647668 but was neither found in the ESP or the 1000genomes databases. It is reported to have a minor allele frequency of 0.3% according to dbSNP build 146. The amino acid change replaces the basic polar positively charged lysine with the likewise basic polar positively charged arginine. Nevertheless, the variant is predicted to be "disease causing" by the program MutationTaster. The gene *EIF5AL1* encodes the eukaryotic translation initiation factor 5A-like 1, a homologue of the eukaryotic translation initiation factor 5A. The function of EIF5AL1 is still unknown, EIF5A, however, plays a role in the elongation phase and, more specifically, stimulates the production of proteins containing runs of consecutive proline residues. It is the only known protein where a lysine residue is post-translationally modified to hypusine [99]. It is predicted to be expressed in skeletal muscle by the Illumina body map (`https://www.ebi.ac.uk/gxa/home`) but not by the UniGene database, which offers an expression sequence tag (EST)-based expression profile (`http://www.ncbi.nlm.nih.gov/est/`). This variant was detected in patient I/9 and confirmed in patient I/5. Therefore it is most likely present in all affected individuals, as it is located on the conserved haplotype. To see if the change segregates in other OPDM families as well, affected individuals from all other Turkish pedigrees (III/2, IV/3, V/1, VI/1, VII/2, VIII/2) as well as three patients from France, and two from the United Kingdom were genotyped for this change. Apparently, the variant could be detected in all of them. 36 Turkish control samples were screened for the c.254A>G, p.(K85R) variant as well and it was present in all. Most likely, the primers bind to other homologues of *EIF5A* with homology in all bases except for this, where the variant was assumed to be.

4.6.3. Triple Repeat Analysis

As OPMD, which is phenotypically very similar to OPDM, is caused by a triple-repeat expansion in the gene *PABPN1*, intronic triple- and hexarepeats within the disease locus on Chromosome 10 were analysed. Five repeats were found in the genes *NRG3*, and one in *LOC219347*, *ZMIZ1*, *LOC100132987* and *TSPAN14* each. They were amplified by PCR in individual I/9 followed by Sanger sequencing. PCR reactions did not work for the *TSPAN14*- as well as for the *NRG3* repeats 2 and 5. The first triple-repeat in intron 1 of *NRG* located at chr10: 83,665,964-83,665,990 in the hg19 assembly appears to be homozygous for a total number of eight ATC repeats $(ATC)_8$. The average length in control individuals is 8.7. However, larger repeat expansions and complete deletions cannot be detected by Sanger sequencing. The third, a CAC-repeat within *NRG3*, was amplified, sequenced and analysed with the program Chromas and the website BLAST. Interestingly, the alignment maps to a region on Chromosome 11: 70,895,189-70,895,773. The primer pairs were reevaluated but are specific to the repeat in intron 2 of *NRG3*, mispriming can almost be excluded. A TTA-repeat within the *NRG3* gene, located in intron 4 (Chr10: 84,429,004 - 84,429,044) shows an average length of $(TTA)_{13.3}$. In patient I/9 it could be determined as 11 and 13. Additionally, triple-repeats in non-coding genes within the disease locus were analysed. The length of the intronic TTG-repeat in *LOC100132987* was found to be $(TTG)_9$, while the average length in controls is 11.7. The sequence appeared to be homozygous. The intronic AAT-repeat in the gene *LOC219347* appeared to be heterozygous with lengths of 12 and 15. The average length is reported being $(AAT)_{14}$. The intronic TTA-repeat in the gene *ZMIZ1* located at chr10:81,038,152 - 81,038,182 appeared to be homozygous for a length of 11 triplets whereas the average in controls is 9.7.

4.7. Array-CGH

DNA from patient I/5 was sent to the “Medizinisch Genetisches Zentrum”, Munich for comparative genomic hybridization on a microarray (array CGH). Array CGH is a molecular cytogenic method to study copy number variations [100]. It is employed to uncover deletions, amplifications, breakpoints and ploidy abnormalities which might be causing OPDM as no single mutation could be detected in the coding regions within the disease locus for Family 1. Unfortunately, the analysis failed due to bad quality of the DNA sample. As patients live in a very remote area of Turkey, no new blood samples could be collected for a reanalysis.

5. Discussion

5.1. Is OPDM a genetically heterogeneous disease?

To identify the genetic cause or causes of OPDM mapping by linkage analysis followed by reconstruction of recombinant haplotypes was performed and exome sequencing was done to detect possible causative variants. However, it is not certain yet, if OPDM is a homogeneous disease and this needs to be discussed first, as different families with different genetic background were included in this study.

Studying the literature of all OPDM cases reported so far revealed that there is some variation in the clinical as well as the histological presentation. For example, two patients with autosomal dominant OPDM were also diagnosed with dilated cardiomyopathy – a finding not common among other OPDM patients [7]. Some patients show tubulofilamentous inclusions in the nuclei but some do not in electron microscopy ultrastructural studies [6]. Most importantly, both autosomal dominant and autosomal recessive inheritance reportedly imply genetic heterogeneity [5]. Nevertheless, OPMD, witch is phenotypically a very similar neuromuscular disorder, presents with both dominant and recessive inheritance, depending on the length of the GCG repeat expansion in the gene *PABPN1* and a similar mechanism could be responsible in OPDM.

Therefore, linkage analysis was performed separately for the dominant Family 1 and the recessive Family 2 and results show that there is a clear dominant disease locus on chromosome 10. The LOD-score for this region, however, is around 0 for family 2 for both a dominant and a recessive model. The genotyping array and microsatellite analysis revealed that OPDM could be mapped to a locus on chromosome 2 for Family 2, for which all patients should be compound heterozygous. Yet, linkage analysis of Family 1 revealed a negative LOD-score for this region. In summary, this study shows that OPDM is a genetically heterogeneous disease and all families should be analysed with special focus on the loci identified on chromosome 2 and 10.

5.2. Possible Genetic Causes for OPDM in Family 1

5.2.1. Intronic and UTR Repeat expansions

As exome sequencing of two individuals from Family 1 did not result in any obvious candidate genes, it needs to be discussed, which genetic variants could be causative for OPDM but missed by this technology. The most plausible would be repeat expansions either in coding or non-coding regions due to the similarity to other neuromuscular disorders, especially OPMD. Repeat expansions can either be pathogenic on RNA or protein level. In OPMD, oligomerisation of polyalanine expanded PABPN1 results in nuclear protein aggregation and causes cell death [101]. Depending on the number of expanded GCG-repeats, OPMD is inherited with a autosomal dominant or recessive trait which also conforms to the fact that most families affected by OPDM stem from a small region in Turkey and show both dominant and recessive inheritance. The idea of whole exome sequencing is to detect all protein altering mutations and therefore only the coding regions are enriched. Repeat expansions in the untranslated regions (UTRs) of genes and the introns are commonly missed if they are not close to the exons. Additionally, larger expansions cannot be identified by the methods used in this study (Illumina HiSeq 2000) because the additionally inserted bases cause too many mismatches in comparison to the reference genome. For a more detailed description see chapter 5.7.1 on page 67.

5.2.2. Transcription-Reducing Variants

A second conceivable mechanism to explain which genetic defect could lead to OPDM is mutations that reduce transcription of a certain gene. These are commonly found in the promoter region e.g. in the RNA-polymerase binding site or regulatory elements. There are not many reported cases with this underlying genetic mechanism because firstly, they are rare and secondly difficult to identify as they are not enriched by whole exome sequencing. One example would be a study of patients with limb-girdle muscular dystrophy type 2O (OMIM #613157) where a 9bp deletion was detected in the promoter region of *POMGNT1* resulting in reduced expression [102]. As the promoter regions are commonly not covered by WES this could be the genetic aberration leading to OPDM. A second mechanism how transcription could be reduced are mutations in the UTRs. Even though they do not change the amino acid structure of the protein, they are transcribed and might affect post-transcriptional regulation. There is a number of processes involved that control mRNA half-life and conclusively protein translation. First, there is capping of the 5-prime end of the mRNA to protect it from 5' exonuclease as well as polyadenylation of the 3-prime end which adds adenine bases to the 3' end to buffer the effects of the 3' exonuclease. Long poly(A) tails therefore correlate with a long half-life [103]. Additionally, a long poly(A) tail can increase translation by binding of poly(A) binding proteins (PABP) that initiate the translation through interaction with the eukaryotic initiation factors EIF4E and EIF4G [104]. Second, a process called RNA editing can alter the sequence of mRNA molecules by deamination of adenosine to inosine bases. This reaction is catalysed by 'Adenosine Deaminase Acting on RNA' (ADAR) enzymes and can alter the splicing and translation machineries, the double-stranded RNA structures and the binding affinity between RNA and RNA-binding proteins with unpredictable effects [105]. And third, microRNA mediated regulation controls the expression of about 60% of all protein coding genes [106]. These genes' mRNAs have conserved binding sites, mostly found in the 3' UTR, for microRNAs that reduce expression. Conclusively, mutations in this region can either create a microRNA binding site or lose one and therefore alter the expression of a gene. Recently, a study has demonstrated that mutations in the 3' UTR of *GFPT1* creates a new binding site for *miR-206** resulting in repression of translation and causing congenital myasthenic syndrome in the affected individuals [107]. These genetic mechanisms should be considered possible causes for OPDM.

5.2.3. Copy Number Variation

A third group of genetic aberration that can cause disease in mammals is copy number variation (CNVs). This is a phenomenon where sections of the genome are repeated – usually duplicated, seldom triplicated or quadruplicated – or deleted. There are a number of disease phenotypes associated with CNVs most of which are congenital malformations and mental retardations. Some are caused by loss of gene function due to under- or overexpression such as deletions in *TBX1* causing Velocardiofacial Syndrome (OMIM #192430) [108], or *PMP22* resulting in Charcot-Marie-Tooth disease type 1A (OMIM #118220) [109]. Others are caused by overexpression and protein aggregation such as duplications of the *APP* gene in Alzheimers disease or *SNCA* in Parkinsons disease [110], [111], [112]. Quite recently, Ankala and colleagues have studied 41 genes by next-generation sequencing and array CGH and discovered a rate of 5 CNV out of 70 patients presenting with congenital muscular dystrophies (CMD) and 8 CNV out of 193 patients presenting with limb-girdle muscular dystrophies (LGMD) . Conclusively, it seems to be a common genetic mechanism in neuromuscular disorders [113]. However, detection of copy number variation is challenging from whole-exome sequencing data, as the coverage is much more variable compared to whole-genome sequencing [114] and computational algorithms are still being improved to increase specificity and sensitivity of CNV prediction from NGS-data [115].

5.2.4. Candidate Genes in the Locus on chromosome 10

As fine-mapping defined a clear disease locus for Family 1 by reconstruction of recombinant haplotypes (4.5 on page 44) and no missense or nonsense variants were identified in this region, it was discussed which other genetic mechanisms could be underlying OPDM in these patients. Thus, genes within the locus have to be evaluated for a candidate gene approach.

One of the genes, located within the disease locus on chromosome 10 is *NRG3*. Neuregulin 3 is a ligand for the Erbb4 transmembrane tyrosine kinase receptor and can signal in an autocrine, paracrine and juxtacrine fashion [116]. Erbb4 activation mediates cell migration, control of cell proliferation, cell stratification, and cell adhesion in developmental as well as pathogenic processes in the nervous system, heart, kidney, and mammary gland [116], [117]. It has been shown to be involved in embryonic mammary gland development [118] but little is known about its function in the developing brain, where it is highly expressed, and in skeletal muscle tissue. Common genetic variation in *NRG3* is thought to increase the risk of schizophrenia [82].This gene, even though not expressed in skeletal muscle is particularly interesting because its introns contain a large number of repeats, which might be causing OPDM when expanded. A recent study has shown that a microsatellite repeat expansion within intron 7 in the *NRG3* gene correlates with reduced levels of *NRG3* expression. However, the phenotypical correlation was impaired mammary gland development in mice, reducing the possibility of an association with a muscle phenotype [119]. Nevertheless, repeats within this gene should be screened for expansions in further studies to identify the mutation causing OPDM in this family.

POLR3A is a subunit of the RNA polymerase III which transcribes genes encoding ribosomal 5S RNA, tRNAs, U6 small nuclear RNA, mitochondrial RNA-processing RNA, H1 RNA, Y RNAs, and 7SK RNA [68]. 5S RNA is imported into mitochondria but, more importantly, is also an essential component of the large ribosomal 60S subunit and reduced or increased expression might affect protein biosynthesis resulting in a muscle phenotype with rimmed vacuoles in histological studies representing protein depositions [120]. The U6 small nuclear RNA is involved in splice site detection and conclusively, impaired or excessive transcription could result in a huge amount of alternatively spliced mRNA and therefore cause disease [121]. Mutations in the gene *POLR3A* have been associated with recessive hypomyelinating leukodystrophy (MIM #607694) [69] but it cannot be excluded that an OPDM phenotype is a variant specific phenotype resulting from mutations in the *POLR3A* gene. Conclusively, further studies such as whole-genome sequencing should be done to detect or exclude genetic variation within the *POLR3A* gene.

A third interesting gene, found within the detected disease locus is *PPIF*, encoding Cyclophilin D, a mitochondrial peptidyl-prolyl cis/trans isomerase. Cyclophilins catalyse the cis to trans isomerisation of certain proline imidic peptide bonds but Cyclophilin D is also known to be an activator of the mitochondrial permeability transition pore (MPTP) [122], [123]. Under certain conditions like oxidative stress or calcium overload, the pore opens and allows free passage of smaller molecules over the mitochondrial membranes resulting in ATP depletion by uncoupling of oxidative phosphorylation and conclusively necrotic cell death. *PPIF* deficient mice (*PPIF-/-*) showed protection against reperfusion injury after ischemia of heart and brain tissue due to the reduced activity of the MPTP, whereas overexpression of *PPIF* in cardiac tissue lead to mitochondrial swelling and spontaneous cell death [124]. Knockout of *PPIF* in a mouse model for sarcoglycanopathies (*SGCD-/-*), a limb-girdle muscular dystrophy showed markedly less dystrophic disease in both skeletal muscle and cardiac muscle compared to a single knockout of *SGCD* [125]. Conclusively, mutations leading to a gain of function of *PPIF* could result in mitochondrial swelling and necrotic cell death of muscle fibres and might be causative for OPDM.

5.3. Possible Genetic Causes for OPDM in Family 2

5.3.1. Recessive or Dominant Trait?

When discussing the possible underlying genetic defect responsible for OPDM in Family 2 it first needs to be deliberated whether the disease in this family follows a dominant or recessive inheritance pattern. At first glance, this is easy to answer - as the parents are reported to be consanguineous and none of them are affected, it is recessive. However, the chances of being homozygous for the mutation is 25% and one would not expect $^5/_6$ of the descendants to be affected by the disease. Additionally, one would assume that the patients in this family would be homozygous in the regions with a high LOD-score in the linkage analysis. Nonetheless, haplotyping for chromosome 2, as well as homozygosity mapping revealed, that the patients all inherited the same allele from their mother but a different one from their father that is likewise shared by all affected siblings (see figure 4.5 on page 27 and figure 5.1 on page 55). If the disease is not caused by a homozygous mutation, it is also imaginable that the underlying pattern of inheritance is dominant and the

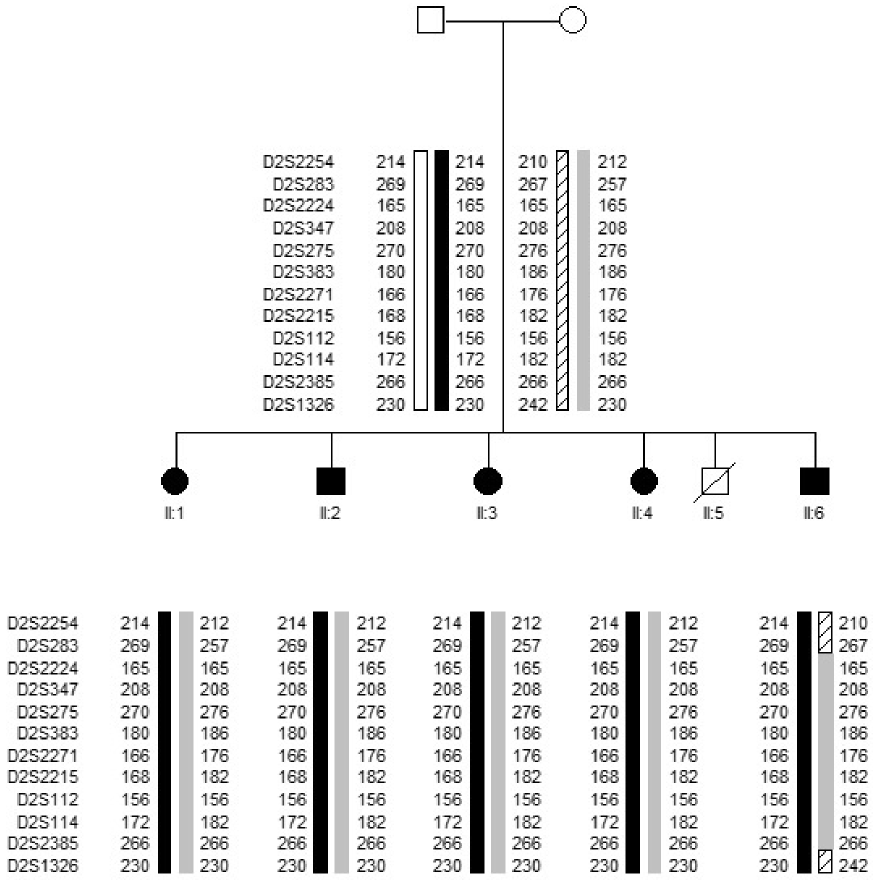

D2S2254 214 214 210 212
D2S283 269 269 267 257
D2S2224 165 165 165 165
D2S347 208 208 208 208
D2S275 270 270 276 276
D2S383 180 180 186 186
D2S2271 166 166 176 176
D2S2215 168 168 182 182
D2S112 156 156 156 156
D2S114 172 172 182 182
D2S2385 266 266 266 266
D2S1326 230 230 242 230
II:1
II:2
II:3
II:4
II:5
II:6
D2S2254 214 212 214 212 214 212 214 212 214 210
D2S283 269 257 269 257 269 257 269 257 269 267
D2S2224 165 165 165 165 165 165 165 165 165 165
D2S347 208 208 208 208 208 208 208 208 208 208
D2S275 270 276 270 276 270 276 270 276 270 276
D2S383 180 186 180 186 180 186 180 186 180 186
D2S2271 166 176 166 176 166 176 166 176 166 176
D2S2215 168 182 168 182 168 182 168 182 168 182
D2S112 156 156 156 156 156 156 156 156 156 156
D2S114 172 182 172 182 172 182 172 182 172 182
D2S2385 266 266 266 266 266 266 266 266 266 266
D2S1326 230 230 230 230 230 230 230 230 230 242

Figure 5.1.: Reconstruction of haplotype alleles in Family 2 for chromosome 2 from microsatellite marker from D2S2254 to D2S1326 (Chr2: 119,988,825 - 139,923,024). Apparently, all patients inherited the same set of alleles (from D2S283 to D2S1326: Chr2: 121,643,494 - 139,923,024) further narrowing down the possible disease locus on chromosome 2. These results imply, that OPDM could be caused by compound heterozygosity in this family.

causative genetic variant is heterozygous. Since none of the parents are affected, either incomplete penetrance, a germ-line mutation or genetic mosaicism would have to underlie the condition in this family. Still, chromosome 2 at around 150cM would be a good candidate region, presenting with a LOD-score greater than 1 (top score for a dominant model).

5.3.2. Candidate Genes on chromosome 2

In summary, special attention was put on the list of genes inside the conceivable disease locus on chromosome 2 mapped to the region from rs12711539 and rs344689 (chromosome 2: 121,837,519 - 140,100,106) by genotype analysis based on SNP genotyping array and whole exome sequencing data (see Figure 4.5 on page 27) which was further mapped by recombinant haplotype analysis done in Dresden by Prof.

Angela Hübner to D2S283 to D2S1326 (chromosome 2: 121,643,494 - 139,923,024) as displayed in figure 5.1 on page 55. None of the genes found in table 4.1 on page 28 are striking candidate genes according to their gene function, yet the most likely would be *MYO7B*, encoding the protein myosin VIIB. Myosins are molecular motors that use energy from adenosine triphosphate (ATP) hydrolysis to generate mechanical force upon their interaction with actin filaments. There are seven vertebrate myosin classes, the conventional myosin II and the unconventional myosins I, V, VI, VII, IX and X. Unconventional myosins have a structurally conserved head that moves along actin filaments. Their highly divergent tails are presumed to enable them to transport cargo [126]. *MYO7B* is expressed in small amounts in skeletal muscle, its main function, still not fully understood, however seems to be in the small and large intestines as well as the kidneys, where it is highly expressed [127]. Exome sequencing of three siblings from Family 2 did not identify any rare, protein altering variants within the locus from rs12711539 and D2S1326 (chromosome 2: 121,837,519 - 139,923,024) that are shared by all three patients.

5.3.3. Possible Compound Heterozygous Variants

Exome sequencing revealed a number of variants with two or more of them being found in one single gene. Therefore, these could be compound heterozygous and causative for a recessive inheritance. Namely, these are variants in the genes *MUC2*, *MBD3L3*, *FBN3* and *TTN* and are listed in Table 4.5 on page 37. *MUC2* encodes a member of the mucin protein family, which are large glycoproteins produced by many epithelial tissues. Mucin 2 is secreted by the gut mucosa and forms an insoluble barrier to protect the intestines [128]. The two detected variants are c.5014T>G, p.(S1672A) and c.4876A>T, p.(I1626F) (transcript variant NM_002457.3). In 4362 control alleles from the ExAC browser, the c.50T>G variant can be found twice in a heterozygous state whereas the c.4876A>T is not listed. However, there is a huge number of homozygous missense variants listed in the ExAC browser. In a nutshell, *MUC2* is not a good candidate gene by it's function and by the identified variants.

MBD3L3 encodes the protein methyl-CpG-binding domain protein 3 like 3 and there are no studies up to date about its cellular function. Methyl-CpG binding domain protein 3, however, is a subunit of the NuRD (Nucleosome Remodeling Deacetylase), a multisubunit complex with ATP-dependent chromatin remodeling and histone deacetylase activities [129]. This complex is crucial for the regulation of chromatin structure and promotion of transcriptional repression [130]. Variants in this gene have not been associated with any diseases yet. The two identified variants in *MBD3L3* (NM_001164425.1), located on chromosome 19, are c.622A>G, p.(R208G) and c.619T>G, p.(C207G) both of which are not listed in the dbSNP-, EVS5400- and the 1000genomes database. The first variant is a substitution of the positively charged amino acid arginine with the hydrophobic glycine. This position is the last amino acid before the stop codon in *MBD3L3* and conserved in rhesus and dog. The second change alters a hydrophobic cysteine to a likewise hydrophobic glycine. Eventually, a disulfide bond could get lost and therefore alter the protein's tertiary structure. These variants are both not listed in the ExAC-browser database but ExAC-coverage of this gene is very bad in general (average: 5.947) resulting in unreliable data. However, the linkage analysis for this region on chromosome 19 results in a negative LOD score of less than -2. Most interestingly, individuals from Finland (OPDM7) and Canada (OPDM8) carry the same pair of variants, whereas an individual from the UK (OPDM6) is homozygous for the c.622A>G, p.(R208G) change. Even though the linkage analysis implies that the other two affected individuals from Family 2 do not carry the same two variants Sanger sequencing should be done to verify the changes followed by segregation analysis in Family 2.

Two other variants were detected in the gene *FBN3* (NM_032447.3) which is located on chromosome 19 and encodes the protein fibrillin 3. This extracellular matrix macromolecule assembles into microfibrils in a vast number of connective tissues, especially during fetal development [131]. Polycystic ovary syndrome susceptibility was linked to a dinucleotide repeat expansion in Intron 55 by a genomewide association study (OMIM: %184700) [132]. The two variants are c.6397G>A, p.(Gly2133Ser) and c.5399G>A, p.(Gly1800Asp). The first one substitutes the non-polar amino acid glycine with the uncharged polar serine and the second one substitutes glycine with the likewise uncharged polar asparagine. Amino acid position 2133 is highly conserved but some other species, including chicken, zebrafish and xenopus tropicalis, express serine instead of glycine at position 1800. The c.6397G>A variant can be found 21 times in the ExAC browser, and c.5399G>A 20 times. Linkage analysis revealed a LOD-score of around -1 for this region on chromosome 19, implying that the other two affected siblings might not carry the same pair of variants within *FBN3*. Additionally, it is more likely that mutations in a cytoplasmic or nuclear protein are causative for OPDM as microscopic studies on muscle tissue showed a clear intracellular pathology. Deleterious missense mutations are expected to result in a loss of function and result in a gene-function related phenotype. Therefore, these two variants were not investigated any further.

Finally, a number of 4 heterozygous rare and protein altering variants in the gene *TTN* (NM_133378.4) were detected, shared by OPDM3, OPDM4 and OPDM5: c.92522G>A, p.(C30841Y) (listed 12 times in a heterozygous state in the ExAC-browser), c.87147T>A, p.(D29049E) (listed 345 times in a heterozygous state in the ExAC-browser), c.11138C>G, p.(T3713S) (listed 359 times in a heterozygous state in the ExAC-browser) and c.1267A>C, p.(S423R) (listed once in a heterozygous state in the ExAC-browser). Additionally, c.87147T>A and c.11138C>G are listed in dbSNP and have a frequency of about 0,24% and c.73031G>A of 0,02% according to the ESP5400 database, the other one is not listed in any of these databases as shown in table 4.5 on page 37.

Mutations in the titin gene are associated with a number of neuromuscular disorders such as autosomal recessive (AR) limb-girdle muscular dystrophy 2J (MIM #608807) [133] and it's milder form, the autosomal dominant (AD) tibial muscular dystrophy (MIM #600334) [73], where patients are heterozygous for the mutations causing AR LGMD2J. They can also cause early-onset myopathy with fatal cardiomyopathy (MIM #611705) [134]. Cases with inherited hypertrophic (MIM #613765) or dilative (MIM #604145) cardiomyopathy have also been described [72]. Most of these skeletal muscle titinopathies are caused by truncation and other loss of function alleles in the most distal M-band (C-Terminus) region of titin, commonly with autosomal recessive inheritance [135].

Evaluation of assigning pathogenicity to single variants in the titin gene is challenging as genetic polymorphisms are common and associated with a number of conditions - a study by Herman et al. showed, that heterozygous truncating mutations (nonsense-, frameshift- and splicing mutations) in the titin gene occur in about 3% of apparently healthy individuals [72]. For truncating variants, a length dependent algorithm has been established to estimate the chance of being causative for nonischemic dilated cardiomyopathy [136]. In neuromuscular disorders, however, this is much more difficult, as the number of patients with conditions caused by aberrations in the titin gene are rare and most are missense and not nonsense mutations. A second problem, especilly with next-generation sequencing of titin, is the challenge of mapping short reads against such a repetitive sequence leading to false positive variants being called [135]. Filtering out variants that occur often – the so called "in-house frequency" – tries to tackle this issue but cannot completely solve it.

The c.11138C>G, p.(T3713S) and c.1267A>C, p.(S423R) variants are located towards the N-terminal end of the titin protein and c.92522G>A, p.(C30841Y) as well as c.87147T>A, p.(D29049E) within the elastic I-band region. Therefore they are most likely not causing any of the neuromuscular conditions described above. Additionally, some of the variants are listed with a higher frequency in the ExAC-browser, decreasing the likelihood of pathogenicity. Conclusively, these variants were also excluded from further studies.

5.3.4. Summary

In Family 2, a very likely disease locus has been identified by reconstruction of recombinant haplotype alleles which was mapped to chromosome 2: 121,837,519 - 139,923,024 and patients would be expected to be compound heterozygous as they inherited different alleles from their parents. Homozygosity mapping did not uncover any larger shared homozygous regions on chromosome 2 providing evidence, that OPDM can be mapped to the locus described above. However, no candidate genes could be identified, indicating, that similarly to Family 1, a more complex genetic reason is underlying OPDM in this family. Conceivable mechanisms would be larger deletions as well as copy number variations or repeat expansions as found in myotonic dystrophies or OPMD [137], [138], [139]. Larger insertions cannot be detected by next generation sequencing techniques as they provide too many mismatches to the reference assembly for alignment tools. Conclusively, further studies such as sequencing of repeats will be necessary to find the causative mutation for OPDM in this family.

5.4. Possible Genetic Causes for OPDM in Individual OPDM VI

5.4.1. *NEB* c.21044C>G, p.(S7015C) and c.22122C>G, p.(D7374E)

Whole exome sequencing identified 2 variants of unknown significance in the gene *NEB* (NM_001164508.1) encoding nebulin, c.21044C>G, p.(S7015C) and c.22122C>G, p.(D7374E) which are located within the high linkage area on Chromosome 2. As previously mentioned, mutation in the *NEB* gene are associated with autosomal recessive Nemaline myopathy (OMIM #256030), a muscle condition where patients usually exhibit generalised hypotonia at birth also affecting respiratory muscles. Proximal limb muscles are usually weaker initially, but distal limb muscle weakness eventually occurs. The facies is commonly myopathic with a high-arched palate and extraocular muscles spared. Chest deformities, hyperlordosis and scoliosis develop in some cases at puberty. Deep tendon reflexes are usually decreased or absent. None of the patients initially reported showed cardiac involvement [140]. Histopathological studies usually show nemaline bodies which are thread- or rod-like structures. Also, in 2007 a novel entity caused by homozygous missense variants in *NEB* was described in four Finnish families where affected individuals only presented with mild distal myopathy and no nemaline bodies were seen in histological examinations[141]. The *NEB* c.21044C>G, p.(S7015C) and c.22122C>G, p.(D7374E) variants were rare with frequencies of 0.004122 and 0.009407 in the ESP respectively but are also listed in dbSNP. However, the ExAC-browser lists 411 heterozygous and 4 homozygous carriers for the c.21044C>G-variant as well as 1085 heterozygous and 13 homozygous control individuals for the c.22122C>G-variant. To confirm the compound-heterozygous state of the variants the changes would have to be segregated in the families or tested by cloning and sequencing *NEB*. However, DNA from the parents was not available. Looking at both the Nemaline-Myopathy phenotype and the frequencies, it is unlikely that the variants mentioned above are the underlying genetic defects in this patient.

5.4.2. Variants in the *TTN* gene

A total number of four variants within the *TTN*-gene, encoding titin, a gigantic structural protein of muscle fibres and the largest human protein , could be identified: c.42310C>T, p.(P14104S), c.18427C>T, p.(R6143W), c.11866A>T p.(I3956F) and c.1492G>A, p.(V498I) (NM_133432.3). Difficulties with assigning pathogenicity to mutations in the TTN gene are discussed in chapter 5.3.3 on page 56.

Most of these skeletal muscle titinopathies are being caused by truncation and other loss of function alleles in the most distal M-band (C-Terminus) region of titin, commonly with autosomal recessive inheritance [135]. The c.18427C>T, p.(R6143W), c.11866A>T p.(I3956F) and c.1492G>A, p.(V498I) variants, however, are located towards the N-terminal end of the titin protein and the c.41935C>T, p.(P13979S) change within the elastic I-band region most likely not causing one of the conditions described in chapter 5.3.3 on page 56. Additionally, the phenotype of tibial muscular dystrophy or limb-girdle muscular dystrophy 2J differ decisively from that of OPDM. To further evaluate these variants, their allele frequency was determined based on the data from the ExAC-browser which lists exome data from around 60,000 controls. c.42310C>T is found in 407 healthy individuals in a heterozygous- and once in a homozygous state. The c.18427C>T-variant is rare and was only found 6 times in a heterozygous state. There were 659 heterozygous and 3 homozygous controls listed for the c.11866A>T-change and 1801 heterozygous and 21 homozygous individuals for the c.1492G>A-variant. These frequencies almost certainly exclude a causative association between the variants and the patient's phenotype. Therefore, carrier status of the parents was not tested to see which changes are compound heterozygous.

5.4.3. *RYR1*: c.8382C>G, p.(Y2794X)

Exome sequencing detected a truncating variant in *RYR1*, encoding the protein ryanodin receptor 1. Mutations in this gene consisting of 106 exons are associated with autosomal dominant central core myopathy (MIM #117000), autosomal recessive minicore myopathy with external ophthalmoplegia (MIM #255320) and malignant hyperthermia susceptibility (MIM #145600) [74], [75],[76],[77].

Central core disease (CCD) is an autosomal dominant myopathy presenting in infancy and involving predominantly proximal muscles [142]. Muscle weakness of the lower limb is the most important feature and can be slow- or non-progressive. Diagnosis is made by muscle biopsy showing amorphous central areas (cores). Mutations in *RYR1* identified so far are missense variants and small in-frame deletions which are mostly located in the C-terminal domain. Multi-minicore disease (MmD) is quite similar with the only distinct differences being the pattern of inheritance and the different size of the histological lesions in skeletal muscle fibres [74]. Most likely, patient OPDM VI does not have a congenital myopathy with cores in the muscle biopsy, otherwise DNA would not have been provided as an OPDM patient. Additionally, the nonsense-mutation p.(Y2794X) would lead to nonsense-mediated decay of the mRNA and therefore exclude a dominant negative effect. The ExAC-Browser lists all variants in about 120,000 control alleles (`http://exac.broadinstitute.org`). It lists 50 loss of function variants (LoF) in the *RYR1*-gene in healthy individuals, resulting in a pLI (probability of LoF intolerance) of 0.00. Therefore, it is unlikely, that the variant described above results in haplotype insufficiency and is causing the muscle phenotype in this patient. As the parents of the patients were reported to be healthy, it would also be crucial to test them for the carrier status and confirm that the variant is *de novo* to argue for a deleterious effect. It might also be the case that a second mutation within the *RYR1* gene was missed by whole exome sequencing leading to autosomal recessive MmD - but even then the phenotype would not fit to OPDM.

5.4.4. Summary

All in all, none of the variants described above is very likely to be responsible for the muscle condition in this patient if she presents with a OPDM phenotype including those on chromosome 2. There were no mutations detected within the novel disease locus on chromosome 10. As the most complex and largest genes in the human genome including *TTN*, *RYR1*, *NEB* and *DMD* are associated with neuromuscular disorders it is foreseeable, that variants are being detected here. A recent study has identified variants of unknown significance in 32% of 177 samples in these and other large NMD-genes [143]. If mutations in the same gene are responsible for the disease in Family 1 and OPDM VI it it understandable that they are missed by whole exome sequencing.

5.5. Possible Genetic Causes for OPDM in Individual OPDM VII

5.5.1. Variants Within High Linkage Areas

A number of variants were identified within the high linkage area on chromosome 2 and the defined disease locus on chromosome 10. Two changes within the gene *GLI2* (NM_005270.4) were identified, c.4332G>A, p.(M1444I) and c.4333C>T, p.(L1445F) at genomic locations 121,747,822 and 121,747,823 on chromosome 2. Mutations in *GLI2* – encoding a zinc finger and transcription factor of Sonic hedgehog signaling – are associated with various kinds of malformation (OMIM #610829, #615849) [78], [79]. Both the c.4332G>A and the c.4333C>T change can be found 20 times in a homozygous state in healthy individuals according to the ExAC-browser, therefore excluding pathogenicity.

Two variants were identified within the shared recombinant haplotype region on chromosome 10. The first is a change in the third exon of *POLR3A* (NM_007055.3): c.275G>C, p.(C92S) and thus located very close to the centromeric recombination point of the disease haplotype. Even though mutations in *POLR3A* are associated with recessive hypomyelinating leukodystrophy (OMIM #607694) [69], it is conceivable that they are also responsible for a completely different phenotype like OPDM. However, this variant has a frequency of 0.0002 according to the 1000genomes database as it was identified on 1 out of 5008 alleles. The ExAC browser lists 34 individuals carrying the variant in a heterozygous state. Considering an autosomal dominant trait or *de novo* status, the minor allele frequency of 0.0002801 is too high to consider this variant to be causative for the patient's phenotype.

The second change found within the disease locus is *NRG3*: c.901G>A, p.(E301K). A detailed description of the gene function as well as phenotypes associated with mutations can be found in chapter 5.2.4 on page 53. This variant is reported to have a minor allele frequency of 0.002417 in the ESP and 0.0023 in the 1000genomes database, meaning, that around 1 in 218 individuals is heterozygous for this variant, which is far too many for a dominant trait. The ExAC browser also lists 9 homozygotes, thus confirming that c.901G>A is a benign polymorphism.

5.5.2. *RYR1*: c.10025C>T, p.(A3342V)

A variant within the *RYR1* (NM_000540.2) was detected in patient OPDM VII, c.10025C>T, p.(A3342V). Most interestingly, this is the second variant found in a patient with suspected diagnosis of OPDM. This variant substitutes the non-polar amino acid alanine with the likewise non-polar valine which most likely does not disrupt the protein structure. This variant is not listed in the ESP- and has a minor allele frequency of 0.0002 according to the 1000genomes database. The ExAC browser lists 52 control alleles carrying this variant resulting in a minor allele frequency of 0.0004323. This number is far too high for an autosomal dominant trait or *de novo* status. Therefore, this variant was considered not to be associated with the patient's muscle condition and consequently not segregated in the patient's family.

5.6. Possible Genetic Causes for OPDM in Individual OPDM VIII

5.6.1. Variants in the *TTN*-Gene

Whole exome sequencing uncovered three variants in gene *TTN* (NM_001267550.1) namely c.107098G>A, p.(D35700N), c.16303G>A, p.(V5435M) and c.10879G>A, p.(V3627I). The ExAC browser lists 673 heterozygotes and 7 homozygotes for the c.16303G>A, p.(V5435M) variant resulting in a minor allele frequency of 0.005636. Therefore, this change has to be benign to the gene. The other two variants are not listed in in the 1000genomes-, the EVS- and the ExAC- database. However, without the patient's parents' DNA it was not possible to proof biallelic location of the c.107098G>A and the c.10879G>A change. Similarly to the discussion of the changes within the titin gene found in patients OPDM III, OPDM IV, OPDM V and OPDM VI (5.3.3 on page 56), the latter were not thought to be the underlying genetic cause for this patient's muscular condition.

5.6.2. *MYOT*: c.655C>T, p.(R219X)

Exome sequencing uncovered a variant in the gene *MYOT*, associated with Limb-Girdle Muscular dystrophy type 1A (LGMD1A, MIM #159000), Myofibrillar Myopathy (MIM #609200) and Spheroid Body Myopathy (MIM #182920), all of which are inherited in an autosomal dominant way. Patients with LGMD1A exhibit a proximal pattern of muscle weakness progressing to include distal limb-girdle muscles. CK levels are elevated up to 9-fold of the normal upper limits. Biopsies of affected individuals show myopathic changes such as variations in fiber size, fiber splitting, and other hallmarks of degeneration as well as a large number of rimmed vacuoles. Z-line streaming, similar to that seen in nemaline myopathy, was also observed [144]. Although some individuals with LGMD1A exhibit a distinctive nasal, dysarthric pattern of speech [144], the predominantly proximal pattern of muscle involvement is quite distinct from the phenotypical presentation of OPDM.

Myofibrillar myopathies (MFMs) are a genetically heterogeneous group of muscular disorders. They are characterised by a pathologic pattern of myofibrillar degradation and accumulation of Z disc proteins [145]. MFM due to mutations in *MYOT* (myotilin) includes progressive distal muscle weakness and peripheral neuropathy with hyporeflexia. The age of onset is usually in the fifties or sixties. Cardiac involvement, as seen in some OPDM patients occurs in a number of individuals. Muscle biopsies show abnormal muscle fibers deposits consisting of amorphous granular and/or hyaline material. Some hyaline structures are thought to comprise beta-pleated amyloid sheets. Electron microscopy studies show smear like aggregates of dense material emerging from Z discs [90]. Although patients exhibit a distal limb-girdle weakness, the phenotype of myofibrillar myopathy caused by mutations in *MYOT* differs from that of OPDM as these patients typically do not have any pharyngeal or ocular involvement.

A subgroup of MFM caused by mutations in the *MYOT* gene is called spheroid body myopathy due to accumulation of myofilamentous material within individual muscle fibers [146], [91]. Patients present first in adolescence and proceed to some motor incapacitation, but life span is not shortened. Muscle weakness is predominantly proximal, almost excluding a misdiagnosis of OPDM.

The variant found in patient OPDM VIII inside the myotilin gene is not easy to evaluate as it is the heterozygous nonsense mutation c.655C>T, p.(R219X) in exon 5 (of 10) of the *MYOT* gene (transcript variant NM_006790.2). To discuss, if a heterozygous missense mutation can cause a dominant inherited condition the molecular basis of dominance has to be understood. First, protein levels can be reduced by a phenomenon called haploinsufficiency. This is, when the monoallelic expression of a gene is not enough to result in a "normal" phenotype [147]. This phenomenon, common in cancer progression, however, is rare in other fields of medicine when focussing on single nucleotide alterations and smaller deletions. Second, protein function can be altered, producing a gain or loss of function. A loss of function could be exemplary explained, when both the wildtype and the mutated protein get incorporated into a structural protein causing a lack of stability such as in Ehlers-Danlos syndrome [148]. Gain of function, such as in Chronic mucocutaneous candidiasis disease caused by mutations in the gene *STAT1*, results in this condition through increased activation of some cytokines thus inhibiting the development of a subgroup of T-cells [149]. The missense variant p.(R219X) is very likely not altering the protein function as in most cases with nonsense mutations the mRNA is eliminated by nonsense-mediated decay resulting in no protein production at all [150].

The only remaining explanation how this variant can cause a neuromuscular phenotype is haploinsufficiency. This is a phenomenon that appears to be extremely rare in autosomal dominant neuromuscular disorders with only a small number of publications where authors claim that haploinsufficiency is the underlying genetic mechanism as by Benedetti et al. 2007 [151]. A rough prediction, if loss of function variants result in haploinsufficiency can be derived from the pLI-score (probability of LoF intolerance) provided by the ExAC browser. It is assumed that there are three classes of genes with respect to tolerance to LoF variation: null (complete tolerance to LoF), recessive (heterozygous LoFs tolerance), and haploinsufficient (where heterozygous LoFs are not tolerated). Observed and expected LoF variants counts are used to determine the probability of LoF intolerance (pLI). The closer pLI is to one, the more LoF intolerant the gene appears to be. A pLI ≥ 0.9 is considered an extremely LoF intolerant set of genes [152]. *MYOT* has a pLI of 0.00, therefore it is very likely to tolerate the loss of one allele which makes it unlikely that the heterozygous *MYOT* p.(R219X) variant could cause OPDM in this patient.

5.6.3. *MEGF10*: c.1564G>A, p.(G522R)

A second heterozygous variant, uncovered by whole exome sequencing in this patient is c.1564G>A, p.(G522R) in the gene *MEGF10*. Mutations in *MEGF10* are associated with autosomal recessive early-onset myopathy, areflexia, respiratory distress, and dysphagia (EMARDD, MIM #614399) as well as recessive congenital myopathy with minicores [86], [87]. Both conditions are not very likely to cause a phenotype similar to that of OPDM. Additionally, the c.1564G>A, p.(G522R) change was found twice in a homozygous status in control individuals according to the ExAC browser. Therefore, this variant is most likely not associated with the muscle condition in this patient.

5.6.4. *MATR3*: c.313C>T, p.(R105C)

5.6.4.1. Can Distal Myopathy Mimic OPDM?

As a change in the gene *MATR3*, associated with vocal cord pharyngeal distal myopathy (VCPDM, OMIM #606070), was detected in patient OPDM VIII, it first needs to be discussed, if VCPDM can mimic the phenotype of OPDM. In VCPDM patients the mean age of onset is 42.2 years (range: 30-55 years) in [153]. Patients exhibit a specific pattern of muscle weakness: Legs seem to be more severely affected than the arms. Weakness in the distal limbs is commonly more pronounced than in the proximal compartments. Still, most patients remain ambulant for a long time. Dysphagia and voice pathology is common but ocular muscle involvement has not been described so far. Interestingly, the Achilles reflex was absent in all patients assessed by Müller et al.. Histopathology shows myopathic changes such as fiber size variation, minor fatty replacement and internal nuclei in all patients. Additionally, subsarcolemmal rimmed vacuoles, also observed in hereditary inclusion body myositis [154], OPMD [155], and other myopathies as well as atrophic fibers consistent with denervation, can be seen in most patients. Ultrastructural studies showed sparse and small tubular aggregates but no filamentous inclusions [156]. Patients with VCPDM usually present with myopathic changes in electromyographical assessments [157]. Pathological spontaneous activity was found in some patients [153]. Creatine kinase (CK) serum levels were generally within twice the upper limits of normal levels. [153].

All these pathologic findings are consistent with the clinical diagnosis of OPDM except that ocular muscles are usually spared in patients with VCPDM [157]. Therefore, it is not farfetched that patients with VCPDM can clinically be diagnosed with OPDM and consequently, the detected mutation in the gene *MATR3* should be considered a good candidate gene in this patient.

5.6.4.2. Evaluation of the Mutation

The c.313C>T, p.(R105C) in the Matrin 3 gene, transcript variant (NM_018834.5) found in this patient was predicted to alter the strength of two splice acceptor sites and cause a substitution of the polar and positively charged amino acid arginine with the nonpolar cysteine. However, this variant is located in a non-canonical transcript and the proportion of expression has not been published, yet. If this region would be coding in only a small percentage it would be difficult to argue that this mutation is deleterious. However, if the altered strength of two splice acceptor sites caused by this variant is shifting the ratio towards the transcripts including this region to the protein coding part it would most likely disrupt the protein structure. Only one mutation in the *MATR3* gene, found in patients from Tennessee., USA, Bulgaria, Germany and Japan, namely p.(S85C), is reported to cause autosomal dominant vocal cord pharyngeal distal myopathy (VCPDM) [94]. Most interestingly, the variant is not listed in the ExAC browser and the pLI is 1.00 as there is not a single LoF variant in the *MATR3* gene found in control individuals. This implies, that the observed variant is most likely deleterious if this transcript is expressed in skeletal muscle. No specific diagnostic criteria for VCPDM has been determined yet as immunohistochemical staining shows no differences between patients and controls concerning subcellular location inside the nuclei of muscle cells as well as expression levels determined by real-time PCR and Western-Blot analyses [153]. To further investigate the case, cDNA analyses to determine the ratio of transcript variants as well as quantification of Matrin 3 protein as well cDNA have to be done. In summary, the c.313C>T, p.(R105C) variant detected in the gene *MATR3* is considered a good candidate gene in this individual.

5.6.5. Variants Within the High-Linkage Areas

In addition to variants in genes known to cause neuromuscular diseases, special interest was put on the mapped disease locus on chromosome 10 as well as on the high-linkage area on chromosome 2. No single rare and protein-altering change was seen at chromosome 10: 79,750,884 - 85,566,388 and only the *TTN*-changes discussed above were interesting among those detected on chromosome 2. Conclusively, the best candidate variant in this patient is *MATR3* c.313C>T, p.(R105C) and needs to be further evaluated by expression analyses, immunohistochemical stainings and screening of patients with a similar phenotype for this particular mutation.

5.7. How Disease Causing Variants can be Missed by Whole-Exome Sequencing

5.7.1. Technical Issues

As exome sequencing of patients from two larger families as well as 2 individual patients did not detect convincing candidate genes it needs to be discussed, how disease causing variants can be missed by this technology. First, there are a number of technical issues which should be considered when analysing WES data. Obviously, only the protein-coding parts of the genome are covered by this method and variants in the gene's noncoding as well as intergenic regions cannot be identified. Additionally, some coding regions of rarely expressed transcript variants might not be targeted by exome sequencing libraries. Also, only around 92% of the exome is currently enriched by modern kits as it is difficult to design hybridisation probes for some coding regions, especially repetitive elements and GC rich sequences. Furthermore, even coverage for targeted coding exons is not 100% accurate with some regions being poorly and some being not covered at all. Modern established sequencing pipelines reach a coverage of around 97-98% with a minimum read depth of 20x which is required for confident variant calling [36], [158]. As the bioinformatical pipeline used in this study has not been validated and tested against different approaches for coverage as well as sensitivity and specificity of variants being called the coverage is expected to be less than 97%. The importance of these technical issues was highlighted by a study displaying that especially some exons of genes associated with neuromuscular disorders are difficult to enrich for next-generation sequencing [159]. They also reported that approximately 10-20% of the 92% of targeted exons had low or zero coverage in their whole exome database [159]. The main reason for the incomplete coverage is GC-rich regions which are a challenge for exome enrichment kits that use clonal amplification of templates. These commonly comprise the first exons of protein coding genes. As PCR is required for these techniques, AT-rich and GC-rich target sequences may be underrepresented in genome alignments and assemblies which may result in very low coverage [160], [161], [162]. All in all, this may well be an explanation, why this study was not able to identify the genetic reason for OPDM. However these issues are mitigated by the fact, that the disease could be mapped and uncovered exons within the linkage area were analysed by Sanger sequencing.

5.7.2. Repeat Expansions

A second issue, why exome sequencing studies might fail to identify causative genetic aberration, could be that a repeat expansion is underlying the inherited condition. This is likely in OPDM as discussed in chapter 5.2 on page 51. Identification of large expansions can fail on different stages during the process of next-generation sequencing [163]. First, hybridisation of patient DNA to the probe for target enrichment could be impaired and result in no coverage. Second, as target enrichment is based on PCR, the polymerase might struggle which results in poor coverage. Third, sequencing itself is challenging for repeat regions because most next-generation sequencing platforms, as the Illumina sequencing system used in our studies, rely on reading signal from bulk DNA populations, like Sanger sequencing does. Therefore they are limited by the loss of sequence phase coherence - a particular problem of the often GC-rich repeat regions. Conclusively, even the best amplification based NGS technology cannot sequence alleles with expansions of around 100 repeats [164]. And finally, reads covering an expanded repeat might not be aligned by software tools due to the number of mismatches to the reference genome as the maximum read length is 100bp using an Illumina platform. In summary, standard approaches for NGS enrichment will not always work in high-repeat genomic regions and preferential use of alternative technologies such as the PacBio system should be considered when a repeat expansion is likely to be the underlying genetic cause of a disease.

5.7.3. Indels

Additionally insertions and deletions (so-called indels) could be causative for OPDM and be missed by whole-exome sequencing. There are two main difficulties in identifying indels: First, next-generation sequencing technology is susceptible to produce indel artifacts, especially 1 bp heterozygous indels inserted to or deleted from long poly-A or poly-T runs (homopolymers) [165], [166]. And second, alignment of reads spanning indels is challenging with many of them being misaligned, resulting in bad coverage which might not be enough for a confident variant call [167], [165]. Unsurprisingly, a recent study has demonstrated that the concordance between different software tools to call indels is as little as 30% [25]. Thus, if the causative genetic variation for OPDM was an indel, it is likely to be missed by whole-exome sequencing.

5.7.4. Copy Number Variations

A special subset of indels are structural variants called copy number variation meaning very large insertions or deletions ranging up to several megabases in size. In recent years as detection tools improved, their role in inherited diseases as well as in cancer has been highlighted [168], [169], [170]. There are two main options to identify copy number variations. One is detecting linkage disequilibrium by SNP genotyping array data meaning that certain SNPs flanking CNV regions are inherited combined (linkages) more or less often than would be expected judging by their distance apart in the genome [171]. The other is a read-depth approach, where the the mapping ratio of next-generation sequencing read counts relative to a reference genome is determined for detection of copy number variations [172]. At the time of bioinformatical processing of the exome-sequencing data for this study, many of the software tools for detection of copy number variation were not available. Therefore these structural variations could not be identified, if present at all, and programs like

CNV-seq, Pindel or ExomeDepth should be included in a reanalysis of whole-exome sequencing data [173], [174], [175].

5.7.5. Bioinformatic Difficulties

Also, bioinformatical analysis of exome-sequencing raw data poses problems and needs to be discussed as the choice of annotation tool and variant callers can result in various discrepancies. In one study, two different transcript sets – RefSeq and Ensemble – were used as a basis for annotation and the authors only found 44% agreement in annotations for loss of function variants [176]. The same study also showed, that only 65% of loss of function changes and 87% of all variants in the coding regions were matching when comparing results from two annotation software tools annovar and VEP (Variant Effect Predictor, [177]), implying that there is a huge number of false negative variants. A similar study demonstrates that combination of different read aligners with variant calling software tools vary in performance [178]. They also show huge discrepancies for indels and SNPs, as all alignment tools use different algorithms coming with diverging strengths and weaknesses. For example the usage of gapped alignment algorithms - i.e. the ability of allowing a gap in a read compared to the reference sequence - is important for indel detection on the cost of efficiency and time [59]. Faster alignment tools will therefore identify less indels and might miss the causative genetic change.

Additionally, the effect of grouped or single sample variant calling needs to be discussed as the exome sequencing data has been analysed one at a time. Pooled sample variant calling allows the use of reads across all samples of a batch at a position to determine the presence of a polymorphism, markedly improving the sensitivity. One study showed that grouped sample variant calling resulted in 4.30% more detected SNPs [179]. To sum it up, the perfect bioinformatical pipeline does not exist and all combinations of software tools, like Mosaik Aligner and Varscan/Dindel variant caller as used in this study, have their weaknesses and will not be able to detect all present variants.

5.7.6. Familial Locus Heterogeneity

A recent study highlighted the problem of familial locus heterogeneity, meaning that two or more disease-causing mutations are responsible for the same phenotype in a larger family. Rehman et al. presented 10 consanguineous Pakistani families with autosomal recessive hearing loss due to mutations in two or more genes [180]. They concluded that familial locus heterogeneity occurs in around 15% of families in their collection, making it a common cause of failure in next-generation sequencing projects. In such cases, linkage analysis would result in decreased LOD-scores. It might therefore be a good strategy to obtain the MLOD score for each pedigree from simulations using information on the pedigree structure, mode of inheritance, affection status, penetrance and availability of genotype data. The MLOD score is the highest possible LOD score obtained for all replicates. If the actual LOD score does not reach the level of the simulated MLOD score it is reasonable to assume familial locus heterogeneity and analyses should be repeated for the different branches of a family. This could be the underlying cause of the low LOD score in Family 1 which was expected to be higher due to the size of the pedigree. Further analyses as suggested above should be performed to exclude this phenomenon in Family 1.

5.8. Options to Find the Genetic Cause of OPDM

5.8.1. Upcoming Advances in Next-Generation Sequencing

As this study was unable to identify the genetic cause or causes of OPDM even though two disease loci could be identified, further options need to be discussed how to solve this case.

First, some of the technical issues, discussed in chapter 5.7.1 on page 67 are being tackled everyday and there will be advances in next-generation sequencing technology that will solve many of these issues. One of the most interesting advances will be the improvement of real-time DNA sequencing from single polymerase molecules such as the PacBio Sequencing platforms [181]. This sequencing platform produces reads with an average length of 10kb with half the amount of reads being longer than 20kb [182]. Therefore it would be ideal for the detection of larger aberrations such as indels or repeat expansions and should be considered the number one choice for further studies on OPDM. An alternative sequencing platform, which is currently being established and improved, is the the Oxford Nanopore which shows great potential regarding accuracy as well as cost- and time efficiency. From first studies, a proof of concept and improvements by using genetically improved biological pores in 2009 to the introduction of the minION sequencing device of the size of a palm being commercially available since May 2015 this technology is expected to revolutionise the field of high-throughput sequencing [183], [184]. If first studies prove a higher sensitivity in comparison with established platforms from Illumina and Roche it might be a good option for further sequencing projects on OPDM.

5.8.2. Analysis for Copy Number Variations

As discussed in chapter 5.2.3 on page 52, copy number variations are an underestimated genetic basis of neuromuscular disorders and the bioinformatic pipeline used in this study is not able to detect them. There are two main strategies how to call CNVs, one is from paired data where the ratio of read counts from a test- and a control sample is used to detect regions that deviate from one; the other is from pooled data where CNVs are detected as difference of read counts from the average depth of coverage (DOC) profile in a region [185]. The raw data should therefore be reanalysed including the use of programs like exome2CNV or PropSeg in order to identify CNVs from the exome sequencing data, which is cheaper and faster than additional usage of the array CGH technology [186], [187].

5.8.3. Whole Genome Sequencing or Target Sequencing of the Disease Locus

When whole-exome sequencing fails to identify disease-causing variants one possible explanation is that the mutation is located in an area of the genome which is not targeted. Intronic as well as promoter regions are conceivable locations where genetic changes could result in a disease phenotype as discussed in chapter 5.2 on page 51. Therefore, it is a reasonable step to perform whole genome sequencing next in order to identify the mutation causing OPDM. Alternatively, target resequencing for the defined disease locus on chromosome 10 could be done, which might be cheaper and should provide a higher sensitivity since a recent study showed, that target sequencing of genes associated with neuromuscular disorders resulted in 20-30% more detected variants compared to whole exome sequencing [143]. Apart from coverage of intronic as well as intergenic regions, WGS comprehends the advantages of reliable detection of copy number variations as well as better identification of repeat expansions compared to WES [188]. However, variant detection from whole genome sequencing is still limited by the ability to annotate and interpret non-coding sequence variants. Additionally, there are only few data sets of high coverage reference genomes to compare results to [37]. In the worst case one would identify a large number of small genetic variants within the disease locus and would not be able to find the one true disease causing mutation. Still, whole-genome sequencing is the most obvious next step in order to identify the genetic cause of OPDM.

5.8.4. RNA-Sequencing

RNA-seq (RNA-sequencing), also called whole transcriptome shotgun sequencing (WTSS) is a method to sequence the set of all messenger RNA in a population of cells. Apart from being capable of detecting mutations, it comprises a number of advantages over DNA sequencing: First, it is able to identify alternative spliced transcripts, especially, when a cryptic splice acceptor site is missed by WES. Second, it can detect gene fusions, which is more important in cancer research, but might be underlying an inherited disease like OPDM. And third, it is widely used to analyse gene expression which might be affected by mutations in regulatory elements such as the promoter [189]. If WES and WGS fail to detect good candidate variants, this technology could be used to study the differences in gene expression with special focus on those located within the defined disease locus on chromosome 10 and on chromosome 2 between patient derived material and control samples. However, this would require either muscle tissue from a biopsy or fibroblasts from a skin biopsy which could be transformed into myoblasts by MyoD virus transduction [190]. This would be challenging, as patients live in Turkey and also because invasively taking tissue samples for research reasons rises ethical questions.

5.8.5. Immunohistochemical Staining for Candidate Gene Products

A possibility to get around the issue of taking further tissue samples would be immunohistochemical studies of paraffin embedded tissue at hand. As there are only 14 protein coding genes in the mapped disease locus on chromosome 10, immunohistochemical staining for their encoded proteins could identify aberrant staining patterns as well as reduced or missing transcription. The human protein atlas (`http://www.proteinatlas.org`) summarises information on tissue expression, sub-cellular localisation and provides antibodies for most known protein coding gene products [98]. For all 14 proteins mentioned above, there are antibodies available and could be used for further studies.

6. Summary

Oculopharyngodistal myopathy (OPDM) is an inherited adult onset muscle disease with both dominant and recessive patterns of inheritance. Ptosis is usually the initial symptom, followed by distal weakness and swallowing difficulties. Histopathological findings include chronic myopathic changes and rimmed vacuoles. Despite all efforts, the underlying genetic cause for OPDM could not yet be identified.

The aim of the study described in this thesis was to identify the mutation responsible for OPDM and is based on 49 individuals from unrelated Turkish families as well as three sporadic patients from England, Finland and Canada.

Linkage analysis was done using SNP-genotyping array data and showed high LOD-scores for regions on chromosome 2 for a recessive family and on chromosome 10 for a large dominant family. To further map the disease loci finemapping was carried out by microsatellite analysis. Two separate disease loci could be identified and reconstruction of recombinant haplotypes mapped them to chr2 q14.2-q22.1 for the recessive family and chr10 q22.3-q23.1 for the dominant family.

Subsequently, whole exome sequencing and Sanger sequencing of five individuals from two unrelated families as well as three sporadic patients was used in order to identify the underlying genetic cause of OPDM. Intriguingly, no likely disease causing variants or candidate genes could be identified in the Turkish families. However, in one sporadic patient, a heterozygous variant in the *MATR3* gene was identified, which is predicted to cause the substitution of a conserved amino acid as well as result in aberrant splicing. Another variant in *MATR3*, detected in 5 unrelated families, has previously been associated with vocal cord pharyngeal distal myopathy. This finding could replicate the association of *MATR3* with a myopathy and expand the phenotypical presentation of patients with *MATR3*-related disorders.

It has been discussed that there is a number of possible explanations why the underlying genetic cause of OPDM has remained elusive despite WES analysis including the possibilities of larger deletions or duplications, intronic or intergenetic mutations and also repeat expansions. There are also a number of problems arising from the sequencing process and the bioinformatical analysis. Any of these issues may be the reason, why the underlying genetic cause for OPDM has not yet been identified and further studies including technologies like RNA-Seq or whole genome sequencing will be necessary to finally understand the genetic basis of OPDM.

7. Zusammenfassung

Die Oculopharyngodistale Myopathie (OPDM) ist eine erbliche Muskelerkrankung des Jungend- und Erwachsenenalters, die sowohl autosomal dominant als auch rezessiv vererbt werden kann. Typischerweise entwickeln Patienten initial eine Ptose, die von einer distalen Gliedergürtelschwäche und einer Dysphagie gefolgt wird. Histopathologisch lassen sich chronische myopathische Veränderungen und vakuoläre Einschlüsse nachweisen. Trotz intensiver Bemühungen konnte die genetische Ursache der OPDM noch nicht identifiziert werden.

Das Ziel dieser Arbeit war es, die genetische Veränderung, die für die OPDM verantwortlich ist zu finden und zu erforschen. Die Studie basierte auf 49 Individuen aus 9 nicht-verwandten türkischen Familien und sporadischen Patienten aus England, Finnland und Kanada.

Mittels SNP-Array wurde eine Kopplungsanalyse durchgeführt, die hohe LOD-Scores für Marker auf Chromosom 2 für eine Familie mit rezessivem und auf Chromosom 10 für eine Familie mit dominantem Erbang ergab. Um den Genlokus für die OPDM weiter einzugrenzen, wurde eine Untersuchung von Mikrosatelliten durchgeführt, um die Haplotypen in den gekoppelten Regionen zu rekonstruieren. Hier konnten zwei unterschiedliche Krankeitslozi identifiziert werden, Chromosom 2q14.2-q22.1 für die rezessive und Chromosom 10q22.3-q23.1 für die dominante Familie.

Anschließend wurde von fünf betroffenen Individuen aus zwei nicht-verwandten und von drei sporadischen Patienten eine Exom-Sequenzierung durchgeführt, um die der Erkrankung zugrunde liegenden genetische Ursache zu identifizieren. Zu unserem Erstaunen konnten wir in den türksichen Familien keinerlei genetische Veränderungen identifizieren, die möglicherweise mit der Erkrankung assoziiert werden können. Dennoch fand sich bei einem sporadischem Patienten eine heterozygote Variante im *MATR3*-Gen. Auf Grund der Variante wird der Austausch einer konservierten Aminosäure und aberrantes Spleißen vorhergesagt. In 5 nicht-verwandten Familien mit einer Stimmband-Pharyngealen distalen Myopathie konnte eine Mutation im *MATR3*-Gen als genetische Ursache identifiziert werden. Dieses Ergebnis könnte eine Replikation der Assoziation von MATR3-Varianten mit Myopathien darstellen und das phänotypische Spektrum dieser Erkrankungsgruppe erweitern.

Es wurde diskutiert, dass es eine Reihe möglicher Erklärungen gibt, weshalb der, der OPDM zugrundeliegende, genetische Defekt bisher noch nicht identifiziert werden konnte, trotz der Sequenzierung des gesamten Exoms. Beispielsweise könnten Deletionen oder Duplikationen, intronische oder intergene Mutationen und auch Repeat-Expansionen eine Erkrankung hervorrufen. Es gibt außerdem einige Schwierigkeiten beim Sequenzieren und bei der bioinformatischen Auswertung, die bedacht werden müssen. All diese Punkte könnten ein Grund sein, weshalb die genetische Ursache der OPDM bisher noch nicht identifiziert worden ist und Folgestudien mit weitergehenden Methoden wie der Transkriptom- und Genomsequenzierung sind nötig, um den komplexen genetischen Mechanismus zu verstehen, der für die OPDM verantwortlich ist.

Bibliography

[1] Satoyoshi, E. and Kinoshita, M. "Oculopharyngodistal myopathy". *Archives of Neurology* 34.2 (1977), pp. 89–92.

[2] Uyama, E., Uchino, M., Chateau, D., and Tomé, F. M. "Autosomal recessive oculopharyngodistal myopathy in light of distal myopathy with rimmed vacuoles and oculopharyngeal muscular dystrophy". *Neuromuscular disorders* 8.2 (1998), pp. 119–125.

[3] van der Sluijs, B M, ter Laak, H. J., Scheffer, H., van der Maarel, S M, and van Engelen, B G M. "Autosomal recessive oculopharyngodistal myopathy: a distinct phenotypical, histological, and genetic entity". *Journal of Neurology, Neurosurgery, and Psychiatry* 75.10 (2004), pp. 1499–1501.

[4] Zhao, J. et al. "Clinical and muscle imaging findings in 14 mainland chinese patients with oculopharyngodistal myopathy". *PloS one* 10.6 (2015), e0128629.

[5] Durmus, H. et al. "Oculopharyngodistal myopathy is a distinct entity: clinical and genetic features of 47 patients". *Neurology* 76.3 (2011), pp. 227–235.

[6] Minami, N. et al. "Oculopharyngodistal myopathy is genetically heterogeneous and most cases are distinct from oculopharyngeal muscular dystrophy". *Neuromuscular Disorders* 11.8 (2001), pp. 699–702.

[7] Thevathasan, W. et al. "Oculopharyngodistal myopathy–a possible association with cardiomyopathy". *Neuromuscular Disorders* 21.2 (2011), pp. 121–125.

[8] Finsterer, J. and Stöllberger, C. "Oculopharyngodistal myopathy and acquired noncompaction". *Neuromuscular Disorders* 21.7 (2011), 523–4; author reply 524–5.

[9] Mignarri, A. et al. "The first Italian patient with oculopharyngodistal myopathy: case report and considerations on differential diagnosis". *Neuromuscular Disorders* 22.8 (2012), pp. 759–762.

[10] Lu, H. et al. “The clinical and myopathological features of oculopharyngodistal myopathy in a Chinese family”. *Neuropathology* 28.6 (2008), pp. 599–603.

[11] Brais, B. “Oculopharyngeal muscular dystrophy: a late-onset polyalanine disease”. *Cytogenetic and genome research* 100.1-4 (2003), pp. 252–260.

[12] Lander, E. S. et al. “Initial sequencing and analysis of the human genome”. *Nature* 409.6822 (2001), pp. 860–921.

[13] Rabbani, B., Tekin, M., and Mahdieh, N. “The promise of whole-exome sequencing in medical genetics”. *Journal of Human Genetics* 59.1 (2014), pp. 5–15.

[14] Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A., and Jabado, N. “What can exome sequencing do for you?” *Journal of Medical Genetics* 48.9 (2011), pp. 580–589.

[15] Hodges, E. et al. “Genome-wide in situ exon capture for selective resequencing”. *Nature Genetics* 39.12 (2007), pp. 1522–1527.

[16] Choi, M. et al. “Genetic diagnosis by whole exome capture and massively parallel DNA sequencing”. *Proceedings of the National Academy of Sciences of the United States of America* 106.45 (2009), pp. 19096–19101.

[17] Ng, S. B. et al. “Exome sequencing identifies the cause of a mendelian disorder”. *Nature Genetics* 42.1 (2010), pp. 30–35.

[18] Albert, T. J. et al. “Direct selection of human genomic loci by microarray hybridization”. *Nature Methods* 4.11 (2007), pp. 903–905.

[19] Lohmann, K. and Klein, C. “Next generation sequencing and the future of genetic diagnosis”. *Neurotherapeutics* 11.4 (2014), pp. 699–707.

[20] Bashiardes, S. et al. “Direct genomic selection”. *Nature Methods* 2.1 (2005), pp. 63–69.

[21] Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., and Rice, P. M. “The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants”. *Nucleic Acids Research* 38.6 (2010), pp. 1767–1771.

[22] Langmead, B. and Salzberg, S. L. “Fast gapped-read alignment with Bowtie 2”. *Nature Methods* 9.4 (2012), pp. 357–359.

[23] Bao, R. et al. “Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing”. *Cancer Informatics* 13.Suppl 2 (2014), pp. 67–82.

[24] Yu, X. and Sun, S. “Comparing a few SNP calling algorithms using low-coverage sequencing data”. *BMC Bioinformatics* 14 (2013), p. 274.

[25] O’Rawe, J. et al. “Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing”. *Genome Medicine* 5.3 (2013), p. 28.

[26] Pirooznia, M. et al. “Validation and assessment of variant calling pipelines for next-generation sequencing”. *Human Genomics* 8 (2014), p. 14.

[27] McKenna, A. et al. “The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data”. *Genome Research* 20.9 (2010), pp. 1297–1303.

[28] Albers, C. A. et al. “Dindel: accurate indel calls from short-read data”. *Genome Research* 21.6 (2011), pp. 961–973.

[29] Rimmer, A. et al. “Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications”. *Nature Genetics* 46.8 (2014), pp. 912–918.

[30] Li, H. et al. “The Sequence Alignment/Map format and SAMtools”. *Bioinformatics* 25.16 (2009), pp. 2078–2079.

[31] Koboldt, D. C. et al. “VarScan: variant detection in massively parallel sequencing of individual and pooled samples”. *Bioinformatics* 25.17 (2009), pp. 2283–2285.

[32] Li, H., Ruan, J., and Durbin, R. “Mapping short DNA sequencing reads and calling variants using mapping quality scores”. *Genome Research* 18.11 (2008), pp. 1851–1858.

[33] Garrison, E. and Marth, G. “Haplotype-based variant detection from short-read sequencing”. *arXiv* (2012).

[34] Iqbal, Z., Caccamo, M., Turner, I., Flicek, P., and McVean, G. “De novo assembly and genotyping of variants using colored de Bruijn graphs”. *Nature Genetics* 44.2 (2012), pp. 226–232.

[35] Lyon, G. J. et al. “Exome sequencing and unrelated findings in the context of complex disease research: ethical and clinical implications”. *Discovery Medicine* 12.62 (2011), pp. 41–55.

[36] Li, M. H. et al. “Utility and limitations of exome sequencing as a genetic diagnostic tool for conditions associated with pediatric sudden cardiac arrest/sudden cardiac death”. *Human Genomics* 9 (2015), p. 15.

[37] Lek, M. and MacArthur, D. “The Challenge of Next Generation Sequencing in the Context of Neuromuscular Diseases”. *Journal of Neuromuscular Diseases* 1.2 (2014), pp. 135–149.

[38] Goldstein, D. B. et al. “Sequencing studies in human genetics: design and interpretation”. *Nature Reviews Genetics* 14.7 (2013), pp. 460–470.

[39] MacArthur, D. G. et al. “Guidelines for investigating causality of sequence variants in human disease”. *Nature* 508.7497 (2014), pp. 469–476.

[40] Yang, Y. et al. “Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders”. *The New England Journal of Medicine* 369.16 (2013), pp. 1502–1511.

[41] Biancalana, V. and Laporte, J. “Diagnostic use of Massively Parallel Sequencing in Neuromuscular Diseases: Towards an Integrated Diagnosis”. *Journal of Neuromuscular Diseases* 2.3 (2015), pp. 193–203.

[42] Green, R. C. et al. “ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing”. *Genetics in Medicine* 15.7 (2013), pp. 565–574.

[43] McCormack, P. et al. “‘You should at least ask’. The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research”. *European Journal of Human Genetics* 24.10 (2016), pp. 1403–1408.

[44] Gainotti, S. et al. “Improving the informed consent process in international collaborative rare disease research: effective consent for effective research”. *European Journal of Human Genetics* 24.9 (2016), pp. 1248–1254.

[45] Foley, A. R. et al. “Treatable childhood neuronopathy caused by mutations in riboflavin transporter RFVT2”. *Brain* 137.Pt 1 (2014), pp. 44–56.

[46] Jaspar, H. H. et al. “Oculopharyngodistal myopathy with early onset and neurogenic features”. *Clinical Neurology and Neurosurgery* 80.4 (1977), pp. 272–282.

[47] Hout, H. van et al. “Recombinant human α-glucosidase from rabbit milk in Pompe patients”. *The Lancet* 356.9227 (2000), pp. 397–398.

[48] Desjardins, P. and Conklin, D. “NanoDrop microvolume quantitation of nucleic acids”. *Journal of Visualized Experiments* 45 (2010).

[49] Saiki, R. K. et al. “Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia”. *Science* 230.4732 (1985), pp. 1350–1354.

[50] Mullis, K. et al. “Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction”. *Cold Spring Harbor Symposia on Quantitative Biology* 51 Pt 1 (1986), pp. 263–273.

[51] Sanger, F., Nicklen, S., and Coulson, A. R. “DNA sequencing with chain-terminating inhibitors”. *Proceedings of the National Academy of Sciences of the United States of America* 74.12 (1977), pp. 5463–5467.

[52] Abecasis, G. R., Cherny, S. S., Cookson, W. O., and Cardon, L. R. “Merlin–rapid analysis of dense genetic maps using sparse gene flow trees”. *Nature Genetics* 30.1 (2002), pp. 97–101.

[53] Wigginton, J. E. and Abecasis, G. R. “PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data”. *Bioinformatics* 21.16 (2005), pp. 3445–3447.

[54] Morton, N. E. “Sequential tests for the detection of linkage”. *American Journal of Human Genetics* 7.3 (1955), pp. 277–318.

[55] Seelow, D., Schuelke, M., Hildebrandt, F., and Nurnberg, P. “Homozygosity Mapper–an interactive approach to homozygosity mapping”. *Nucleic Acids Research* 37.Web Server issue (2009), W593–9.

[56] Cinnioğlu, C. et al. “Excavating Y-chromosome haplotype strata in Anatolia”. *Human Genetics* 114.2 (2004), pp. 127–148.

[57] Seelow, D., Schwarz, J. M., and Schuelke, M. “GeneDistiller–distilling candidate genes from linkage intervals”. *PloS one* 3.12 (2008), e3874.

[58] Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. “OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders”. *Nucleic Acids Research* 43.Database issue (2015), pp. D789–98.

[59] Li, H. and Homer, N. “A survey of sequence alignment algorithms for next-generation sequencing”. *Briefings in Bioinformatics* 11.5 (2010), pp. 473–483.

[60] Sim, N. et al. “SIFT web server: predicting effects of amino acid substitutions on proteins”. *Nucleic Acids Research* 40.W1 (2012), W452–W457.

[61] Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. “Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2”. In: *Current Protocols in Human Genetics*. 1994, pp. 7.20.1–7.20.41.

[62] Chun, S. and Fay, J. C. “Identification of deleterious mutations within three human genomes”. *GenomeResearch* 19.9 (2009), pp. 1553–1561.

[63] Schwarz, J., Rödelsperger, C., Schuelke, M., and Seelow, D. “MutationTaster evaluates disease-causing potential of sequence alterations”. *Nature Methods* 7.8 (2010), pp. 575–576.

[64] Cooper, G. M. “Distribution and intensity of constraint in mammalian genomic sequence”. *Genome Research* 15.7 (2005), pp. 901–913.

[65] Bang, M. L. et al. “Myopalladin, a novel 145-kilodalton sarcomeric protein with multiple roles in Z-disc and I-band protein assemblies”. *The Journal of Cell Biology* 153.2 (2001), pp. 413–427.

[66] Miller, M. K. et al. “The muscle ankyrin repeat proteins: CARP, ankrd2/Arpp and DARP as a family of titin filament-based stress response molecules”. *Journal of Molecular Biology* 333.5 (2003), pp. 951–964.

[67] Purevjav, E. et al. “Molecular basis for clinical heterogeneity in inherited cardiomyopathies due to myopalladin mutations”. *Human Molecular Genetics* 21.9 (2012), pp. 2039–2053.

[68] Sepehri, S. and Hernandez, N. “The largest subunit of human RNA polymerase III is closely related to the largest subunit of yeast and trypanosome RNA polymerase III”. *Genome Research* 7.10 (1997), pp. 1006–1019.

[69] Bernard, G. et al. “Mutations of POLR3A encoding a catalytic subunit of RNA polymerase Pol III cause a recessive hypomyelinating leukodystrophy”. *American Journal of Human Genetics* 89.3 (2011), pp. 415–423.

[70] Labeit, S. and Kolmerer, B. “The complete primary structure of human nebulin and its correlation to muscle structure”. *Journal of Molecular Biology* 248.2 (1995), pp. 308–315.

[71] Pelin, K. et al. “Mutations in the nebulin gene associated with autosomal recessive nemaline myopathy”. *Proceedings of the National Academy of Sciences of the United States of America* 96.5 (1999), pp. 2305–2310.

[72] Herman, D. S. et al. “Truncations of Titin Causing Dilated Cardiomyopathy”. *New England Journal of Medicine* 366.7 (2012), pp. 619–628.

[73] Hackman, P. et al. “Tibial Muscular Dystrophy Is a Titinopathy Caused by Mutations in TTN, the Gene Encoding the Giant Skeletal-Muscle Protein Titin”. *The American Journal of Human Genetics* 71.3 (2002), pp. 492–500.

[74] Monnier, N. et al. “A homozygous splicing mutation causing a depletion of skeletal muscle RYR1 is associated with multi-minicore disease congenital myopathy with ophthalmoplegia”. *Human Molecular Genetics* 12.10 (2003), pp. 1171–1178.

[75] Zhang, Y. et al. “A mutation in the human ryanodine receptor gene associated with central core disease”. *Nature Genetics* 5.1 (1993), pp. 46–50.

[76] Quane, K. A. et al. “Mutations in the ryanodine receptor gene in central core disease and malignant hyperthermia”. *Nature Genetics* 5.1 (1993), pp. 51–55.

[77] Robinson, R., Carpenter, D., Shaw, M., Halsall, J., and Hopkins, P. “Mutations in RYR1 in malignant hyperthermia and central core disease”. *Human Mutation* 27.10 (2006), pp. 977–989.

[78] Cohen, M. et al. “Ptch1 and Gli regulate Shh signalling dynamics via multiple mechanisms”. *Nature Communications* 6 (2015), p. 6709.

[79] Bertolacini, C. D., La Ribeiro-Bicudo, Petrin, A., Richieri-Costa, A., and Murray, J. C. “Clinical findings in patients with GLI2 mutations–phenotypic variability”. *Clinical Genetics* 81.1 (2012), pp. 70–75.

[80] Zhang, D. et al. “Neuregulin-3 (NRG3): A novel neural tissue-enriched protein that binds and activates ErbB4”. *Proceedings of the National Academy of Sciences* 94.18 (1997), pp. 9562–9567.

[81] Anton, E. S. et al. “Receptor tyrosine kinase ErbB4 modulates neuroblast migration and placement in the adult forebrain”. *Nature Neuroscience* 7.12 (2004), pp. 1319–1328.

[82] Kao, W.-T. et al. “Common genetic variation in Neuregulin 3 (NRG3) influences risk for schizophrenia and impacts NRG3 expression in human brain”. *Proceedings of the National Academy of Sciences* 107.35 (2010), pp. 15619–15624.

[83] Paterson, A. D. et al. “Persons with Quebec platelet disorder have a tandem duplication of PLAU, the urokinase plasminogen activator gene”. *Blood* 115.6 (2010), pp. 1264–1266.

[84] Haarhuis, J. H. et al. "WAPL-mediated removal of cohesin protects against segregation errors and aneuploidy". *Current Biology* 23.20 (2013), pp. 2071–2077.

[85] Singh, T. et al. "MEGF10 functions as a receptor for the uptake of amyloid-β". *FEBS Letters* 584.18 (2010), pp. 3936–3942.

[86] Logan, C. et al. "Mutations in MEGF10, a regulator of satellite cell myogenesis, cause early onset myopathy, areflexia, respiratory distress and dysphagia (EMARDD)". *Nature Genetics* 43.12 (2011), pp. 1189–1192.

[87] Boyden, S. E. et al. "Mutations in the satellite cell gene MEGF10 cause a recessive congenital myopathy with minicores". *Neurogenetics* 13.2 (2012), pp. 115–124.

[88] Salmikangas, P. et al. "Myotilin, the limb-girdle muscular dystrophy 1A (LGMD1A) protein, cross-links actin filaments and controls sarcomere assembly". *Human Molecular Genetics* 12.2 (2003), pp. 189–203.

[89] Hauser, M. A. et al. "Myotilin is mutated in limb girdle muscular dystrophy 1A". *Human Molecular Genetics* 9.14 (2000), pp. 2141–2147.

[90] Selcen, D. and Engel, A. G. "Mutations in myotilin cause myofibrillar myopathy". *Neurology* 62.8 (2004), pp. 1363–1371.

[91] Foroud, T. et al. "A mutation in myotilin causes spheroid body myopathy". *Neurology* 65.12 (2005), pp. 1936–1940.

[92] Salton, M. et al. "Matrin 3 binds and stabilizes mRNA". *PloS one* 6.8 (2011), e23882.

[93] Johnson, J. O. et al. "Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis". *Nature Neuroscience* 17.5 (2014), pp. 664–666.

[94] Senderek, J. et al. "Autosomal-dominant distal myopathy associated with a recurrent missense mutation in the gene encoding the nuclear matrix protein, matrin 3". *American Journal of Human Genetics* 84.4 (2009), pp. 511–518.

[95] Wang, M. and Marín, A. "Characterization and prediction of alternative splice sites". *Gene* 366.2 (2006), pp. 219–227.

[96] Reese, M. G., Eeckman, F. H., Kulp, D., and Haussler, D. "Improved splice site detection in Genie". *Journal of Computational Biology* 4.3 (1997), pp. 311–323.

[97] Carithers, L. J. et al. “A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project”. *Biopreservation and Biobanking* 13.5 (2015), pp. 311–319.

[98] Uhlen, M. et al. “Tissue-based map of the human proteome”. *Science* 347.6220 (2015), p. 1260419.

[99] Dever, T. E., Gutierrez, E., and Shin, B. S. “The hypusine-containing translation factor eIF5A”. *Critical Reviews in Biochemistry and Molecular Biology* 49.5 (2014), pp. 413–425.

[100] Pinkel, D. et al. “High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays”. *Nature Genetics* 20.2 (1998), pp. 207–211.

[101] Fan, X., Dion, P., Laganiere, J., Brais, B., and Rouleau, G. A. “Oligomerization of polyalanine expanded PABPN1 facilitates nuclear protein aggregation that is associated with cell death”. *Human Molecular Genetics* 10.21 (2001), pp. 2341–2351.

[102] Raducu, M., Baets, J., Fano, O., van Coster, R., and Cruces, J. “Promoter alteration causes transcriptional repression of the POMGNT1 gene in limb-girdle muscular dystrophy type 2O”. *European Journal of Human Genetics* 20.9 (2012), pp. 945–952.

[103] Weill, L., Belloc, E., Bava, F., and Méndez, R. “Translational control by changes in poly(A) tail length: recycling mRNAs”. *Nature Structural & Molecular Biology* 19.6 (2012), pp. 577–585.

[104] Park, E. H. et al. “Multiple elements in the eIF4G1 N-terminus promote assembly of eIF4G1*PABP mRNPs in vivo”. *The EMBO Journal* 30.2 (2011), pp. 302–316.

[105] Tomaselli, S., Locatelli, F., and Gallo, A. “The RNA editing enzymes ADARs: mechanism of action and human disease”. *Cell and Tissue Research* 356.3 (2014), pp. 527–532.

[106] Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. “Most mammalian mRNAs are conserved targets of microRNAs”. *Genome Research* 19.1 (2009), pp. 92–105.

[107] Dusl, M. et al. “A 3’-UTR mutation creates a microRNA target site in the GFPT1 gene of patients with congenital myasthenic syndrome”. *Human Molecular Genetics* 24.12 (2015), pp. 3418–3426.

[108] Driscoll, D. A. et al. "Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome". *American Journal of Medical Genetics* 44.2 (1992), pp. 261–268.

[109] Raeymaekers, P. et al. "Duplication in chromosome 17p11.2 in Charcot-Marie-Tooth neuropathy type 1a (CMT 1a). The HMSN Collaborative Research Group". *Neuromuscular Disorders* 1.2 (1991), pp. 93–97.

[110] Rovelet-Lecrux, A. et al. "APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy". *Nature Genetics* 38.1 (2006), pp. 24–26.

[111] Chartier-Harlin, M. et al. "α-synuclein locus duplication as a cause of familial Parkinson's disease". *The Lancet* 364.9440 (2004), pp. 1167–1169.

[112] Ibáñez, P. et al. "Causal relation between α-synuclein locus duplication as a cause of familial Parkinson's disease". *The Lancet* 364.9440 (2004), pp. 1169–1171.

[113] Ankala, A. et al. "A comprehensive genomic approach for neuromuscular diseases gives a high diagnostic yield". *Annals of Neurology* 77.2 (2015), pp. 206–214.

[114] Hehir-Kwa, J. Y., Pfundt, R., and Veltman, J. A. "Exome sequencing and whole genome sequencing for the detection of copy number variation". *Expert Review of Molecular Diagnostics* 15.8 (2015), pp. 1023–1032.

[115] Chang, L. C. et al. "RefCNV: Identification of Gene-Based Copy Number Variants Using Whole Exome Sequencing". *Cancer Informatics* 15 (2016), pp. 65–71.

[116] Falls, D. "Neuregulins: Functions, forms, and signaling strategies". *Experimental Cell Research* 284.1 (2003), pp. 14–30.

[117] Veikkolainen, V. et al. "ErbB4 modulates tubular cell polarity and lumen diameter during kidney development". *Journal of the American Society of Nephrology* 23.1 (2012), pp. 112–122.

[118] Kogata, N., Zvelebil, M., and Howard, B. A. "Neuregulin 3 and Erbb Signalling Networks in Embryonic Mammary Gland Development". *Journal of Mammary Gland Biology and Neoplasia* 18.2 (2013), pp. 149–154.

[119] Howard, B., Panchal, H., McCarthy, A., and Ashworth, A. "Identification of the scaramanga gene implicates Neuregulin3 in mammary gland specification". *Genes & Development* 19.17 (2005), pp. 2078–2090.

[120] Smirnov, A. et al. “Mitochondrial enzyme rhodanese is essential for 5 S ribosomal RNA import into human mitochondria”. *The Journal of Biological Chemistry* 285.40 (2010), pp. 30792–30803.

[121] Kandels-Lewis, S. and Seraphin, B. “Role of U6 snRNA in 5’ splice site selection”. *Science* 262.5142 (1993), pp. 2035–2039.

[122] Nguyen, T. T. et al. “Cyclophilin D modulates mitochondrial acetylome”. *Circulation Research* 113.12 (2013), pp. 1308–1319.

[123] Halestrap, A. P. “What is the mitochondrial permeability transition pore?” *Journal of Molecular and Cellular Cardiology* 46.6 (2009), pp. 821–831.

[124] Baines, C. P. et al. “Loss of cyclophilin D reveals a critical role for mitochondrial permeability transition in cell death”. *Nature* 434.7033 (2005), pp. 658–662.

[125] Millay, D. P. et al. “Genetic and pharmacologic inhibition of mitochondrial-dependent necrosis attenuates muscular dystrophy”. *Nature Medicine* 14.4 (2008), pp. 442–447.

[126] Li, J., Lu, Q., and Zhang, M. “Structural Basis of Cargo Recognition by Unconventional Myosins in Cellular Trafficking”. *Traffic* 17.8 (2016), pp. 822–838.

[127] Chen, Z. et al. “Myosin-VIIb, a Novel Unconventional Myosin, Is a Constituent of Microvilli in Transporting Epithelia”. *Genomics* 72.3 (2001), pp. 285–296.

[128] Round, A. N. et al. “Lamellar structures of MUC2-rich mucin: a potential role in governing the barrier and lubricating functions of intestinal mucus”. *Biomacromolecules* 13.10 (2012), pp. 3253–3261.

[129] Xue, Y. et al. “NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities”. *Molecular Cell* 2.6 (1998), pp. 851–861.

[130] Shimbo, T. et al. “MBD3 Localizes at Promoters, Gene Bodies and Enhancers of Active Genes”. *PLoS Genetics* 9.12 (2013).

[131] Corson, G. M., Charbonneau, N. L., Keene, and Sakai, L. Y. “Differential expression of fibrillin-3 adds to microfibril variety in human and avian, but not rodent, connective tissues”. *Genomics* 83.3 (2004), pp. 461–472.

[132] Urbanek, M., Sam, S., Legro, R. S., and Dunaif, A. “Identification of a polycystic ovary syndrome susceptibility variant in fibrillin-3 and association with a metabolic phenotype”. *The Journal of Clinical Endocrinology and Metabolism* 92.11 (2007), pp. 4191–4198.

[133] Bushby, K. M. and Beckmann, J. S. “The 105th ENMC sponsored workshop: pathogenesis in the non-sarcoglycan limb-girdle muscular dystrophies, Naarden, April 12-14, 2002”. *Neuromuscular Disorders* 13.1 (2003), pp. 80–90.

[134] Carmignac, V. et al. “C-terminal titin deletions cause a novel early-onset myopathy with fatal cardiomyopathy”. *Annals of Neurology* 61.4 (2007), pp. 340–351.

[135] Watkins, H. “Tackling the Achilles’ Heel of Genetic Testing”. *Science Translational Medicine* 7.270 (2015), 270fs1.

[136] Roberts, A. M. et al. “Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease”. *Science Translational Medicine* 7.270 (2015), 270ra6.

[137] Mahadevan, M. et al. “Myotonic dystrophy mutation: an unstable CTG repeat in the 3’ untranslated region of the gene”. *Science* 255.5049 (1992), pp. 1253–1255.

[138] Liquori, C. et al. “Myotonic Dystrophy Type 2 Caused by a CCTG Expansion in Intron 1 of ZNF9”. *Science* 293.5531 (2001), pp. 864–867.

[139] Brais, B. et al. “Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy”. *Nature Genetics* 18.2 (1998), pp. 164–167.

[140] Wallgren-Pettersson, C. et al. “Clinical and genetic heterogeneity in autosomal recessive nemaline myopathy”. *Neuromuscular Disorders* 9.8 (1999), pp. 564–572.

[141] Wallgren-Pettersson, C. et al. “Distal myopathy caused by homozygous missense mutations in the nebulin gene”. *Brain* 130.Pt 6 (2007), pp. 1465–1476.

[142] Patterson, V. H., Hill, T. R. G., Fletcher, P. J. H., and Heron, J. R. “Central core disease: clinical and pathological evidence of progression within a family”. *Brain* 102.3 (1979), pp. 581–594.

[143] Savarese, M. et al. “MotorPlex provides accurate variant detection across large muscle genes both in single myopathic patients and in pools of DNA samples”. *Acta Neuropathologica Communications* 2 (2014), p. 100.

[144] Hauser, M. A. et al. “myotilin Mutation Found in Second Pedigree with LGMD1A”. *The American Journal of Human Genetics* 71.6 (2002), pp. 1428–1432.

[145] Selcen, D. “Myofibrillar myopathies”. *Neuromuscular Disorders* 21.3 (2011), pp. 161–171.

[146] Goebel, H. H., Muller, J., Gillen, H. W., and Merritt, A. D. “Autosomal dominant "spheroid body myopathy"”. *Muscle & Nerve* 1.1 (1978), pp. 14–26.

[147] Seidman, J. G. and Seidman, C. “Transcription factor haploinsufficiency: when half a loaf is not enough”. *The Journal of Clinical Investigation* 109.4 (2002), p. 451.

[148] Richards, A. J. et al. “A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type II”. *Journal of Medical Genetics* 35.10 (1998), pp. 846–848.

[149] Liu, L. et al. “Gain-of-function human STAT1 mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis”. *The Journal of Experimental Medicine* 208.8 (2011), pp. 1635–1648.

[150] Baker, K. E. and Parker, R. “Nonsense-mediated mRNA decay: Terminating erroneous gene expression”. *Current Opinion in Cell Biology* 16.3 (2004), pp. 293–299.

[151] Benedetti, S. et al. “Phenotypic clustering of lamin A/C mutations in neuromuscular patients”. *Neurology* 69.12 (2007), pp. 1285–1292.

[152] Lek, M. et al. “Analysis of protein-coding genetic variation in 60,706 humans”. *Nature* 536.7616 (2016), pp. 285–291.

[153] Müller, T. et al. “Phenotype of matrin-3-related distal myopathy in 16 German patients”. *Annals of Neurology* 76.5 (2014), pp. 669–680.

[154] Askanas, V. and Engel, W. K. “New advances in inclusion-body myositis”. *Current Opinion in Rheumatology* 5.6 (1993), pp. 732–741.

[155] Tome, F. M., Chateau, D., Helbling-Leclerc, A., and Fardeau, M. “Morphological changes in muscle fibers in oculopharyngeal muscular dystrophy”. *Neuromuscular Disorders* 7 Suppl 1 (1997), S63–9.

[156] Yamashita, S. et al. “Clinicopathological features of the first Asian family having vocal cord and pharyngeal weakness with distal myopathy due to a

MATR3 mutation". *Neuropathology and Applied Neurobiology* 41.3 (2015), pp. 391–398.

[157] Feit, H. et al. "Vocal Cord and Pharyngeal Weakness with Autosomal Dominant Distal Myopathy: Clinical Description and Gene Localization to 5q31". *The American Journal of Human Genetics* 63.6 (1998), pp. 1732–1742.

[158] Rehm, H. L. et al. "ACMG clinical laboratory standards for next-generation sequencing". *Genetics in medicine* 15.9 (2013), pp. 733–747.

[159] Valencia, C. A. et al. "Comprehensive mutation analysis for congenital muscular dystrophy: a clinical PCR-based enrichment and next-generation sequencing panel". *PloS one* 8.1 (2013), e53083.

[160] Metzker, M. "Sequencing technologies – the next generation". *Nature Reviews Genetics* 11.1 (2010), pp. 31–46.

[161] Dohm, J. C., Lottaz, C., Borodina, T., and Himmelbauer, H. "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing". *Nucleic Acids Research* 36.16 (2008), e105–e105.

[162] Harismendy, O. et al. "Evaluation of next generation sequencing platforms for population targeted sequencing studies". *Genome Biology* 10.3 (2009), p. 1.

[163] Mueller, P., Lyons, J., Kerr, G., Haase, C., and Isett, R. B. "Standard enrichment methods for targeted next-generation sequencing in high-repeat genomic regions". *Genetic Medicine* 15.11 (2013), pp. 910–911.

[164] Loomis, E. W. et al. "Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene". *Genome Research* 23.1 (2013), pp. 121–128.

[165] Li, H. "Toward better understanding of artifacts in variant calling from high-coverage samples". *Bioinformatics* 30.20 (2014), pp. 2843–2851.

[166] Challis, D. et al. "The distribution and mutagenesis of short coding INDELs from 1,128 whole exomes". *BMC Genomics* 16 (2015), p. 143.

[167] DePristo, M. A. et al. "A framework for variation discovery and genotyping using next-generation DNA sequencing data". *Nature Genetics* 43.5 (2011), pp. 491–498.

[168] Beroukhim, R. et al. "The landscape of somatic copy-number alteration across human cancers". *Nature* 463.7283 (2010), pp. 899–905.

[169] Gilissen, C. et al. "Genome sequencing identifies major causes of severe intellectual disability". *Nature* 511.7509 (2014), pp. 344–347.

[170] Lupski, J.R. “Structural variation in the human genome”. *The New England Journal of Medicine* 356.11 (2007), pp. 1169–1171.

[171] Eckel-Passow, J. E., Atkinson, E. J., Maharjan, S., Kardia, S. L., and Andrade, M. de. “Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform”. *BMC Bioinformatics* 12 (2011), p. 220.

[172] Jo, H.-Y. et al. “Application of whole–exome sequencing for detecting copy number variants in CMT1A/HNPP”. *Clinical Genetics* 90.2 (2016), pp. 177–181.

[173] Xie, C. and Tammi, M. T. “CNV-seq, a new method to detect copy number variation using high-throughput sequencing”. *BMC Bioinformatics* 10 (2009), p. 80.

[174] Ye, K., Schulz, M. H., Long, Q., Apweiler, R., and Ning, Z. “Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads”. *Bioinformatics* 25.21 (2009), pp. 2865–2871.

[175] Plagnol, V. et al. “A robust model for read count data in exome sequencing experiments and implications for copy number variant calling”. *Bioinformatics* 28.21 (2012), pp. 2747–2754.

[176] McCarthy, D. J. et al. “Choice of transcripts and software has a large effect on variant annotation”. *Genome Medicine* 6.3 (2014), p. 26.

[177] McLaren, W. et al. “Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor”. *Bioinformatics* 26.16 (2010), pp. 2069–2070.

[178] Hwang, S., Kim, E., Lee, I., and Marcotte, E. M. “Systematic comparison of variant calling pipelines using gold standard personal exome variants”. *Scientific Reports* 5 (2015), p. 17875.

[179] Meynert, A., Ansari, M., FitzPatrick, D., and Taylor, M. “Variant detection sensitivity and biases in whole genome and exome sequencing”. *BMC Bioinformatics* 15.1 (2014), p. 247.

[180] Rehman, A. et al. “Challenges and solutions for gene identification in the presence of familial locus heterogeneity”. *European Journal of Human Genetics* 23.9 (2015), pp. 1207–1215.

[181] Eid, J. et al. “Real-Time DNA Sequencing from Single Polymerase Molecules”. *Science* 323.5910 (2009), pp. 133–138.

[182] Rhoads, A. and Au, K. F. “PacBio Sequencing and Its Applications”. *Genomics, proteomics & bioinformatics* 13.5 (2015), pp. 278–289.

[183] Howorka, S., Cheley, S., and Bayley, H. “Sequence-specific detection of individual DNA strands using engineered nanopores”. *Nature Biotechnology* 19.7 (2001), pp. 636–639.

[184] Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G., and Bayley, H. “Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore”. *Proceedings of the National Academy of Sciences* 106.19 (2009), pp. 7702–7707.

[185] Kadalayil, L. et al. “Exome sequence read depth methods for identifying copy number changes”. *Briefings in Bioinformatics* 16.3 (2015), pp. 380–392.

[186] Valdes-Mas, R., Bea, S., Puente, D. A., Lopez-Otin, C., and Puente, X. S. “Estimation of copy number alterations from exome sequencing data”. *PloS one* 7.12 (2012), e51422.

[187] Rigaill, G. J. et al. “A regression model for estimating DNA copy number applied to capture sequencing data”. *Bioinformatics* 28.18 (2012), pp. 2357–2365.

[188] Koboldt, D. C. et al. “VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing”. *Genome Research* 22.3 (2012), pp. 568–576.

[189] Han, Y., Gao, S., Muegge, K., Zhang, W., and Zhou, B. “Advanced Applications of RNA Sequencing and Challenges”. *Bioinformatics and Biology Insights* 9.Suppl 1 (2015), pp. 29–46.

[190] Am Roest, P. et al. “New possibilities for prenatal diagnosis of muscular dystrophies: Forced myogenesis with an adenoviral MyoD-vector”. *The Lancet* 353.9154 (1999), pp. 727–728.

A. Anhang

A.1. List of Primers

List of primers used:

MYPN:

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
MYPN ex 18 fwd	agcctgggtgacagagcaagac	64	6818
MYPN ex 18 rev	ccagactcaaatagcagcagac	60,3	6706
MYPN ex18 fwd2	ctcagccaaagaggtgaagaa	57,9	6497
MYPN ex1 fwd	cataggtgctgtcctgctatgg	62,1	6957
MYPN ex1 rev	tgccctgttatcaaaacacact	56,5	6638
MYPN ex2 fwd	tgggtgacaaagtgagaccttc	60,3	6799
MYPN ex2 rev	ataaactggagctgttttcctg	58,4	6765
MYPN ex3 fwd	atggtttgaatactgccaactc	56,5	6709
MYPN ex3 rev	agaagagcgcatggtagaggag	62,1	6922
MYPN ex4 fwd	aaatcttatgtcgtgtttaggaacc	58,1	7670
MYPN ex4 rev	ggagccacccttcttaagttc	59,8	6357
MYPN ex5 fwd	cacctgtaagcagtgatgcc	59,4	6101
MYPN ex5 rev	tgcaagatggtcatggtcac	57,3	6157
MYPN ex6 fwd	ttgggatgcatttcatat	54	6411
MYPN ex6 rev	ccgtacatacagaagaccaaatc	58,9	6994
MYPN ex7 fwd	atgcacatccacagaactgaag	58,4	6721
MYPN ex7 rev	tggaaaggtgttcattaaatgttg	55,9	7461
MYPN ex8 fwd	aatatccatcctgtccctgttg	58,4	6636
MYPN ex8 rev	tgctggaattacagacatgagc	58,4	6783
MYPN ex9 fwd	aattgttttgaccaattgttttc	51,7	7009
MYPN ex9 rev	ttttagaagagccaagccagc	57,9	6439
MYPN ex10 fwd	tgtgaacactttcccatttgtg	56,5	6691
MYPN ex10 rev	gtgtgagcaactgtgcctagc	61,8	6462
MYPN ex11 fwd	tctgaacattgtttgaaaggtg	54,7	6804
MYPN ex11 rev	gagatttggtttgcacagagg	57,9	6541
MYPN ex12 fwd	tgtcatttcaaccactctgatttc	57,6	7228
MYPN ex12 rev	gtatccgaggactgaatcaagt	60,3	6768
MYPN ex13 fwd	gcttcctcaattgtactgatgg	58,4	6716
MYPN ex13 rev	agaccttcttgaaggcactg	57,3	6116

MYPN ex14 fwd	tcacttaaaagatggcagttgg	56,5	6798
MYPN ex14 rev	tttcctcagcaatccttagtaactc	59,7	7526
MYPN ex15 fwd	atttcacggtgttctggtcc	57,3	6089
MYPN ex15 rev	atccagtactttggtgctcacg	60,3	6701
MYPN ex16 fwd	tgttttacatcagctccacacc	58,4	6605
MYPN ex16 rev	tgacattaaatactccaaacaagcc	58,1	7586
MYPN ex17 fwd	agcaaggataaagaattcagcc	56,5	6785
MYPN ex17 rev	atgaaggaattctggcagagg	57,9	6559
MYPN ex19 fwd	ttcctggaaccctaaatttgac	56,5	6669
MYPN ex19 rev	accttgcctgacccatttatc	57,9	6292
MYPN ex20 fwd	gtgaaggacagaatgcacctc	59,4	6151
MYPN ex20 rev	gcttggaaaccaccaagtctg	59,8	6415
MYPN ex11b fwd	cgaagtatttcttcccctccac	60,3	6581
MYPN ex11b rev	gagagccctgtttcagatcaag	60,3	6759
MYPN ex2b fwd	ataaccctcgaagtcccacc	59,4	5990
MYPN ex2b rev	aaaccaggtgcttaaatgataatac	56,4	7682
MYPN ex3 fwd (2)	ttaagagaatatctggagctgtct	57,6	7406
MYPN ex3 rev (2)	tgctgtatcctcattgcctaga	58,4	6676
MYPN ex2a fwd (2)	ttgagctttaatttctaacgagtc	55,9	7332
MYPN ex9 fwd (2)	gccagctttttatattgactttg	55,3	7010
MYPN ex9 rev (2)	ggaatggaaacacaaaatctgc	56,5	6785
MYPN ex8 fwd (2)	taatatccatcctgtccctgtt	56,5	6611
MYPN ex8 rev (2)	tttattcatctcagtgtaacttcatt	55,3	7876
MYPN ex18 fwd (3)	agaatgacccttctcttgctca	60,4	
6645 MYPN ex16 fwd (2)	gttctctaggtctgtagccatgc	62,4	7021

Table A.1.: List of Primers *MYPN*

POLR3A:

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
POLR3A ex19 fwd	ttagctcccagctgccaaag	59,4	6061
POLR3A ex19 rev	attacagctcatgtgcaaaacg	56,5	6727
POLR3A ex1 fwd	gcgagtagcggaagaggaag	61,4	6305
POLR3A ex1 rev	atctctgaccctgcaagacc	59,4	6021
POLR3A ex2 fwd	aggttggttatggtgggcta	57,3	6259
POLR3A ex2 rev	gctcctttcaatctggtaagtca	58,9	6989
POLR3A ex3 fwd	gcaaaagaaatgattctgtgtca	55,3	7095
POLR3A ex3 rev	ctagatgtatccccccaccactc	62,1	6575
POLR3A ex4 fwd	tcacgtggttaagggtacaaaa	56,5	6807
POLR3A ex4 rev	aaaagctgactcccgaacatta	56,5	6696

POLR3A ex5 fwd	ggacctctcatctttcattgct	58,4	6627
POLR3A ex5 rev	ttttggaagaaagtgggtgtct	56,5	6860
POLR3A ex6 fwd	aaacatagtgaaggaaaaccttgc	57,6	7402
POLR3A ex6 rev	tttttctcacattttcttgacca	53,5	6905
POLR3A ex7 fwd	gcctcccatttccttgtaagtt	58,4	6627
POLR3A ex7 rev	gagaagctggacagacactcct	32,1	6753
POLR3A ex8 fwd	agtctctccgttcttattgttcc	58,9	6922
POLR3A ex8 rev	tttctactgcctgttgtttgc	55,9	6360
POLR3A ex9 fwd	caggatgcctctctttctccta	60,3	6612
POLR3A ex9 rev	tgtggctgagtatgaccacagt	60,3	6790
POLR3A ex10 fwd	tcctgatctgaagagggagaaa	58,4	6832
POLR3A ex10 rev	agaagtccactgtttagcactga	58,9	7047
POLR3A ex11 fwd	ttttaatgtttcaaaacagagaagc	54,8	7688
POLR3A ex11 rev	tggtgttttcatgtaagtttcctt	55,9	7345
POLR3A ex12 fwd	aaaccttgtgattcaggctttg	56,5	6740
POLR3A ex12 rev	tgaatcactatgaacgaggaaca	57,1	7089
POLR3A ex13 fwd	tgttggtcatggttcaaatttat	53,5	7074
POLR3A ex13 rev	cactcatttcaccagtctaccc	60,3	6550
POLR3A ex14 fwd	ggggtagactggtgaaatgagt	60,3	6919
POLR3A ex14 rev	cagaaacaatgaatttgcttgc	54,7	6742
POLR3A ex15 fwd	gctttgaggagaatttctgtttg	57,1	7115
POLR3A ex15 rev	gggatgaaatggcagtaaaaga	56,5	6905
POLR3A ex16 fwd	gcaggcataaactgtatttagtagg	59,7	7745
POLR3A ex16 rev	tacctctattcatggctcagca	58,4	6645
POLR3A ex17 fwd	gcatcttgcctcagtattttca	56,5	6651
POLR3A ex17 rev	gctgtgactatcacattttctgg	58,9	7020
POLR3A ex18 fwd	ctgttttacccttccaatctgc	58,4	6587
POLR3A ex18 rev	ccaacggtctttgatctgaata	56,5	6709
POLR3A ex19 fwd	tttctgatttgcgtggatttca	52,8	6761
POLR3A ex19 rev	tgtgcaaaacgtgtactcaatac	57,1	7071
POLR3A ex20 fwd	tgcttgtaaccttgagactcttg	58,9	7020
POLR3A ex20 rev	aactgcaattgatagtccaaaca	55,3	7024
POLR3A ex21 fwd	gctaaaagctcaccttgggtaa	58,4	6743
POLR3A ex21 rev	cccttgcaaacagagttcaa	57,9	6381
POLR3A ex22 fwd	cagtatccagattgggtccttt	58,4	6716
POLR3A ex22 rev	tgtacacatggggaaacagaag	58,4	6841
POLR3A ex23 fwd	gctgggactcacatcctaattt	58,4	6685
POLR3A ex23 rev	ccagggagcacaaaactcttta	58,4	6712
POLR3A ex24 fwd	ggtgataaccagaagcctctcc	62,1	6704
POLR3A ex24 rev	atgccacccagagtttaagaca	58,4	6712

POLR3A ex25 fwd	cacttgggtttaacaaagcagta	57,1	7071
POLR3A ex25 rev	tggcagctgatttttacacttc	56,5	6691
POLR3A ex26 fwd	aagcagtcgtgtgctcttagg	59,8	6477
POLR3A ex26 rev	tgcttagctcttgccctagttt	58,4	6658
POLR3A ex27 fwd	gggtgcttagaacaaacctgac	60,3	6768
POLR3A ex27 rev	gggataaggccaagaagaaatta	57,1	7178
POLR3A ex28 fwd	agctggggtgatcaaggtga	59,4	6262
POLR3A ex28 rev	tgcagatggcacaaggaaga	57,3	6224
POLR3A ex29 fwd	acagggtttgctttgaaactg	55,9	6476
POLR3A ex29 rev	tctatgatggtcctcacagcag	60,3	6710
POLR3A ex30 fwd	ggaggatttttgttgattgtattg	55,9	7474
POLR3A ex30 rev	gcctagccatggtctatttgta	58,4	6716

Table A.2.: List of Primers *POLR3A*

Microsatellites:

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
D10S606 fwd	tttgaacctgggagacg	52,8	5250
D10S606 rev	catggacattctgctgc	52,8	5161
D10S1136 fwd	gtgggctgaaactctgctt	56,7	5834
D10S1136 rev	gtggggaaacagacaaacc	56,7	5879
D10S1730 fwd	gtgcagccactgttgagag	58,8	5868
D10S1730 rev	aagtttgagaaccactggtctatc	59,3	7351
D10S1164 fwd	ggtgctgaggtgggaagat	58,8	5988
D10S1164 rev	gaggtgtaaggaaagcacga	57,3	6264
D10S201 fwd	agctcatgggatggaagcat	57,3	6206
D10S201 rev	agctaaaaggctgctggaga	57,3	6215
D10S1774 fwd	ctcttgtccacttggcctca	59,4	5994
D10S1774 rev	cctgccttcacactgctctg	61,4	5979
D10S523 fwd	tggaggttgtggtgagctg	58,8	5970
D10S523 rev	ccattctagactgcggctg	58,8	5779
D10S583 fwd	tctgaccaaaataccaaaagaac	55,3	7002
D10S583 rev	agagactccagatgtttgatga	56,5	6798
D10S577 fwd	ttgcacaccagcctaag	52,8	5139
D10S577 rev	gcccaagagttggagac	55,2	5244

Table A.3.: List of Primers Microsatellites

UNC5B

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
UNC5B ex9 fwd	gctcagactggaactcagcac	61,8	6400
UNC5B ex9 rev	ctctggtctgggtaccacca	61,4	6068

Table A.4.: List of Primers *UNC5B*

NRG3

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
NRG3 ex1-1 fwd	cagggagcggatttgcat	56,0	5579
NRG3 ex1-1 rev	atgaagccgatgaacagaggta	58,4	6841
NRG3 ex1-2 fwd	ggctcctaggatgagtgaagg	61,8	6551
NRG3 ex1-2 rev	aagtggtagtggtggtggtctc	62,1	6877
NRG3 ex1-3 fwd	ctcagcctcatgcttctcaaat	58,4	6605
NRG3 ex1-3 rev	agcgtgctgctactgaagaac	59,8	6455
NRG3 ex1-4 fwd	caccaccactaccacttccac	60,8	6200
NRG3 ex1-4 rev	ggaggaggaagaagaggaagaa	60,3	7004
NRG3 ex1-5 fwd	ggcatacgctacctcctccta	61,8	6302
NRG3 ex1-5 rev	agggggcttgctagaaaacag	59,8	6544
NRG3 ex1-1 fwd (2)	cggctcctaggatgagtgaag	61,8	6511
NRG3 ex1-1 rev (2)	gaacagaggtaccacgcacag	61,8	6458
NRG3 ex1-2 fwd (2)	tgaagccgatgaacagaggta	57,9	6528
NRG3 ex1-2 rev (2)	atgaagccgatgaacagaggt	57,9	6528
NRG3 ex1-3 fwd (2)	aagaccggctcctaggatgagt	62,1	6784
NRG3 ex1-3 rev (2)	aaggtgaagaccggctccta	59,4	6151
NRG3 ex2 fwd	cattttcccaggaggtgtttag	58,4	6756
NRG3 ex2 rev	ctgagggccctgtcaataatg	59,8	6406
NRG3 ex2 fwd (2)	gagggttggagctgtctgtcta	62,1	6837
NRG3 ex2 rev (2)	aaacggtgggactgtgtgtatc	60,3	6830
NRG3 ex1-1 fwd (3)	tcttccgagctccttaccg	58,8	5690
NRG3 ex1-1 rev (3)	agaggaagaaggggtcctg	58,8	5966
NRG3 ex1-1 fwd (4)	ccctcttccgagctccttac	61,4	5939
NRG3 ex1-1 rev (4)	atgtaagccgatgaacagaggta	58,4	6841
NRG3 ex2 fwd (2)	agcagtcatttttgagagcaca	56,5	6758
NRG3 ex2 rev (2)	cagatttttcccctcttttcct	56,5	6553
NRG3 ex9 fwd	ggccacaacaagtctactgga	59,8	6424
NRG3 ex9 rev	tcactgaattctcacagcaacc	58,4	6623

Table A.5.: List of Primers *NRG3*

Micro RNAs

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
MIR_554 fwd	tcaaaaatgaaaatatgctggatg	54,2	7432
MIR_554 rev	tttaacagttcccatgcacttg	56,5	6660
hsa-miR-3198-3p fwd	tggccctagaattgtaatccat	56,5	6709
hsa-miR-3198-3p rev	actcccccataaacctgaaagt	58,4	6632

Table A.6.: List of Primers miRNAs

RPS24

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
RPS24 fwd	gggctgtggcaagtatttacag	60,3	6830
RPS24 rev	ggagaagaaggtggagagatga	60,3	6986

Table A.7.: List of Primers *RPS24*

ZMIZ1

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
ZMIZ1 3'-UTR fwd	gttccttttcactgtctgtgg	57,9	6385
ZMIZ1 3'-UTR fwd (2)	gggcgagttgattcacttactc	60,3	6741
ZMIZ1 3'-UTR rev	gggacactttaagggaaaaacc	58,4	6801
ZMIZ1 ex9 fwd	ctatggccaatgccaacaac	57,3	6054
ZMIZ1 ex9 rev	actgctgcagcgccttatct	59,4	6043
ZMIZ1 ex10 fwd	caagtggcacaaatgaatgg	55,3	6199
ZMIZ1 ex10 rev	caccctaatgcagtcagctctc	62,1	6615
ZMIZ1 ex11 fwd	tccctccctgcactttcaat	57,3	5938
ZMIZ1 ex11 rev	acacctcctcaagtccctcaag	62,1	6584
ZMIZ1 ex12 fwd	gtgacctggctatgtgacgtt	59,8	6468
ZMIZ1 ex12 rev	aacacacgcagggtcagagt	59,4	6160
ZMIZ1 ex21 fwd	aggtcacctgggtgtctgtc	61,4	6139
ZMIZ1 ex21 rev	gccaccatcagcacagaaat	57,3	6063
ZMIZ1 ex21 fwd (2)	tgtggtgagagtgggagcag	61,4	6318
ZMIZ1 ex21 rev (2)	gggggatgtgttacttctctct	60,3	6763
ZMIZ1 ex23 fwd	ggttgtgttgggtttcattttc	56,5	6784
ZMIZ1 ex23 rev	ctcacacccaccccttctct	61,4	5868

Table A.8.: List of Primers *ZMIZ1*

PPIF

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
PPIF ex1 fwd	caggggtagtccacggacag	63,5	6192
PPIF ex1 rev	cattctcagaaatggggaaact	56,5	6767
PPIF ex2 fwd	ttggatgtttattgaccccttt	54,7	6697
PPIF ex2 rev	aatgctgagacagcctacagtg	60,36	6768

Table A.9.: List of Primers *PPIF*

EIF5AL1

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
EIF5AL1 ex1 fwd	tcatatgaaagacgtgtaaaatgc	55,9	7408
EIF5AL1 ex1 rev	acgaaggtcctctggtacctc	61,8	6382
EIF5AL1 ex1 fwd (2)	gtgtaaaatgcctgggtagagg	60,3	6879
EIF5AL1 ex1 rev (2)	cttgccaaggtctccctcag	61,4	6028
EIF5AL1 ex1 fwd (3)	aagatcgtggagatgtctgctt	58,4	6805
EIF5AL1 ex1 rev (3)	ggtggggaaaaccaaaataaaa	54,7	6858

Table A.10.: List of Primers *EIF5AL1*

C10orf57

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
C10orf57 fwd	gagactcgctctcagggactt	61,8	6098
C10orf57 rev	gattgtgcttgcacgacttc	57,3	6446

Table A.11.: List of Primers *C10orf57*

PLAC9

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
PLAC9 fwd	ggttctctcgagccagaaagt	59,8	6446
PLAC9 rev	cagctctctctccgtctctctc	64,0	6524
PLAC9 fwd (2)	gctcgtaacaaacccctgac	59,4	6030
PLAC9 rev (2)	cattccttcctcgccatct	56,7	5625

Table A.12.: List of Primers *PLAC9*

ANXA11

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
ANXA11 ex11 fwd	cccatctactgagccatgtgt	59,8	6357
ANXA11 ex11 rev	caggctctgctttgtgtcct	59,4	6065

Table A.13.: List of Primers *ANXA11*

SH2D4B

Oligo Name	Sequence (5’->3’)	Tm [°C]	MW [g/mol]
SH2D4B ex1 fwd	aactgacaatgctgcacagaga	58,4	6761
SH2D4B ex1 rev	cctggccctgctaatttttct	57,9	6314
SH2D4B ex1 fwd (2)	tgggtagaggagatgagttcgt	60,3	6910
SH2D4B ex1 rev (2)	atcttgaagaagggcacagc	57,3	6175
SH2D4B ex4 fwd	ggttcctggactattaggttgg	60,3	6812
SH2D4B ex4 rev	ccaaactacacagcaaatctgg	58,4	6681

Table A.14.: List of Primers *SH2D4B*

AK302451

Oligo Name	Sequence (5’->3’)	Tm [°C]	MW [g/mol]
AK302451 fwd	gaactcaaacactccctccatc	60,3	6568
AK302451 rev	tgaggagattcatgtgaaggtg	58,4	6894

Table A.15.: List of Primers *AK302451*

AX747983

Oligo Name	Sequence (5’->3’)	Tm [°C]	MW [g/mol]
AX747983 ex1 fwd	gacctctgttattccagcaacc	60,3	6630
AX747983 ex1 rev	tgggtaatcaatcccctttatg	56,5	6700
AX747983 ex2 fwd	aagatggctgtggaaactgatt	56,5	6838
AX747983 ex2 rev	gaattcttggctgaactgtgtg	58,4	6796
AX747983 ex2 fwd (2)	gaggatcagttgagtccaggag	62,1	6864
AX747983 ex2 rev (2)	ggctgaactgtgtgcaatagaa	58,4	6823

Table A.16.: List of Primers *AX747983*

ZCCHC24

Oligo Name	Sequence (5’->3’)	Tm [°C]	MW [g/mol]
ZCCHC24 ex1 fwd	aggaagagcgggtcagacag	61,4	6629
ZCCHC24 ex1 rev	aaaagtttcctgcccaactttc	56,6	5941
ZCCHC24 ex2 fwd	agcagggacaaaagggtagag	59,8	6602
ZCCHC24 ex2 rev	cccaaggcagaggctgtagtat	62,1	6784
ZCCHC24 ex2 fwd (2)	atttgaactcaggcttctggag	58,4	6765
ZCCHC24 ex2 rev (2)	aatcccagcagggacaaaag	57,3	6153

ZCCHC24 ex4 fwd	gcttttgcttgcctttgtcc	57,3	6031
ZCCHC24 ex4 rev	ctctccctcactgtgtctgtca	62,1	6588
ZCCHC24 ex4 fwd (2)	ctcatcgggtgtgtgtctctc	61,8	6395
ZCCHC24 ex4 rev (2)	ctggacatgggcttttgctt	57,3	6129

Table A.17.: List of Primers *ZCCHC24*

FAM213A

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
FAM213A ex1 fwd	caaccagcaccatcttctcc	59,4	5941
FAM213A ex1 rev	accagtatgcttgctctcattg	58,4	6676
FAM213A alternate ex fwd	catctacttgggaggctgagg	61,8	6502
FAM213A alternate ex rev	acccactgaaagagaagcagag	60,3	6795
FAM213A alternate ex fwd (2)	cacgtgtagtcccatctacttg	60,3	6661
FAM213A alternate ex rev (2)	tgaaagagaagcagagacacaga	58,9	7172

Table A.18.: List of Primers *FAM123A*

MAT1A

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
MAT1A ex9 fwd	gctgtgttacagttcgttgctc	60,3	6723
MAT1A ex9 rev	tgacaggacaggctaaatgaga	58,4	6841

Table A.19.: List of Primers *MAT1A*

GHITM

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
GHITM ex2 fwd	tttggttggttttgccttttt	52,0	6421
GHITM ex2 rev	aggagggaccagaatgatacaa	58,4	6850

Table A.20.: List of Primers *GHITM*

CDHR1

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
CDHR1 ex1 fwd	gagccgtgtcatcctcttagc	61,8	6373
CDHR1 ex1 rev	aggaagatggaaggacttctcc	60,3	6808
CDHR1 alternate exon fwd	gataaatggatggagctgctg	60,3	6821
CDHR1 alternate exon rev	tggtgggtagggaagtattcag	60,3	6910

Table A.21.: List of Primers *CDHR1*

LRIT2

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
LRIT2 ex2 fwd	gtttgagatccaagaccctcag	60,3	6719
LRIT2 ex2 rev	tggccagatgttagagggttat	58,4	6845
LRIT2 ex3 fwd	gataaatggatggagctgctg	57,9	6550
LRIT2 ex3 rev	ccccggaatcaatacttatgct	58,4	6654
LRIT2 alternate exon fwd	ctgacagagcagtgtcttctcc	62,1	6686
LRIT2 alternate exon rev	ggcacttcctgaagctcataat	58,4	6694

Table A.22.: List of Primers *LRIT2*

LRIT1

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
LRIT1 ex2 fwd	gtgataacaggcagaactggag 60,3	6857	
LRIT1 ex2 rev	aagaccccaggtgaaggttg 59,4	6191	
LRIT1 ex3 fwd	cttcagccagcttgaactgag 59,8	6406	
LRIT1 ex3 rev	aagagccactgtcattgttgaa 56,5	6758	
LRIT1 ex4 fwd	ctgtgaacttggccctgaaag 59,8	6446	
LRIT1 ex4 rev	gtcagctcctcctttgtgct 59,4	6025	

Table A.23.: List of Primers *LRIT1*

MYOT

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
MYOT ex5 fwd	gaacttaccagggctgttcaaa	58,4	6743
MYOT ex5 rev	ttcccctgtgatagttttgatg	56,5	6722

Table A.24.: List of Primers *MYOT*

FKRP

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
FKRP fwd	ctctacgaggagcgctggac	63,5	6143
FKRP rev	gtactgcacgcggaaaaagt	57,3	6175

Table A.25.: List of Primers *FKRP*

RYR1

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
RYR1 fwd	tccctaagacccttagcttgttc	60,6	6925
RYR1 rev	atgtgaaattgcctcactcctc	58,4	6645
RYR1 ex93 fwd	gaatggttttgaatgaatgaactc	55,9	7430
RYR1 ex93 rev	caaggtgagcaggagaggtg	61,4	6296

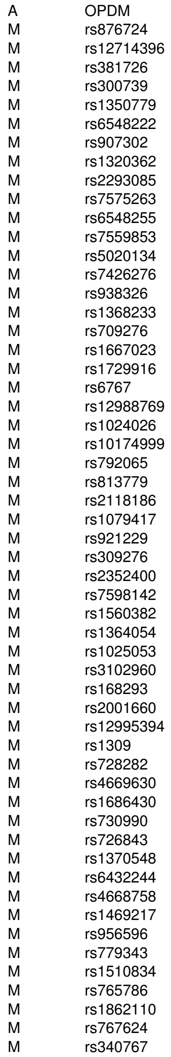
Table A.26.: List of Primers *RYR1*

Intronic repeats

Oligo Name	Sequence (5'->3')	Tm [°C]	MW [g/mol]
NRG3 repeat1 fwd	gggcaaggagactcttctaggt	62,1	6815
NRG3 repeat1 rev	tagcaattgaatgaaggaggag	56,5	6896
NRG3 repeat2 fwd	agcttctttcttgttgtgagga	56,5	6762
NRG3 repeat2 rev	gtggtggtgcatgtctgtagtc	62,1	6828
NRG3 repeat3 fwd	tttatgtgctcttggattgctg	56,5	6753
NRG3 repeat3 rev	caaataggagggatgtgcaagt	58,4	6872
NRG3 repeat4 fwd	caaattgaaagtctgccatcct	56,5	6678
NRG3 repeat4 rev	gtgtccaacccaagaaaatgat	56,5	6736
NRG3 repeat5 fwd	aaggaaaatgacaggctgagaa	56,5	6874
NRG3 repeat5 rev	tgtagtcccagctactcggaag	62,1	6735
LOC100132987 rep fwd	gtcagggttctgcagctctaaa	60,3	6750
LOC100132987 rep rev	ggcaacacagcaagatgtagtc	60,3	6777
ZMIZ1 rep fwd	ccttggtcataagccctttgta	58,4	6676
ZMIZ1 rep rev	ctctgcctaggaaaaccagaga	60,3	6737
LOC219347 rep fwd	acctggatccaatgtacacaag	58,4	6632
LOC219347 rep rev	tagccaaggtgagtcagtgaaa	58,4	6832
TSPAN14 rep fwd	tttgagacagggtcttgctgt	57,9	6483
TSPAN14 rep rev	ggtgaaaccccatctctacaaa	58,4	6672

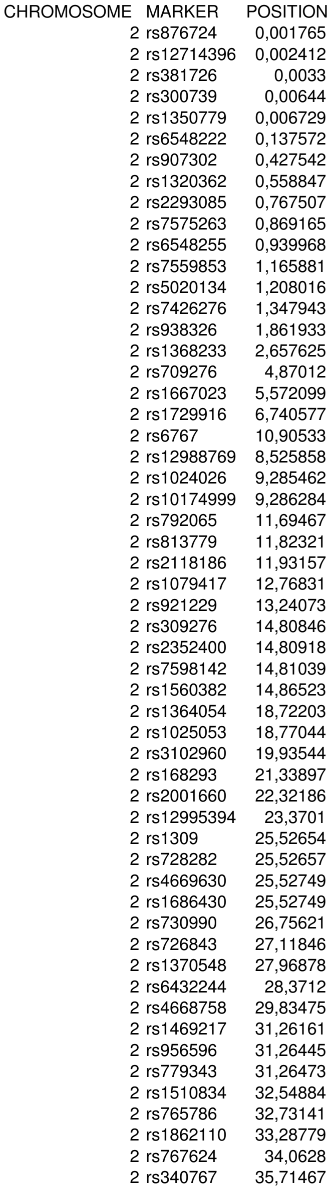
Table A.27.: List of Primers Intronic repeats

A.2. Merlin Input Files



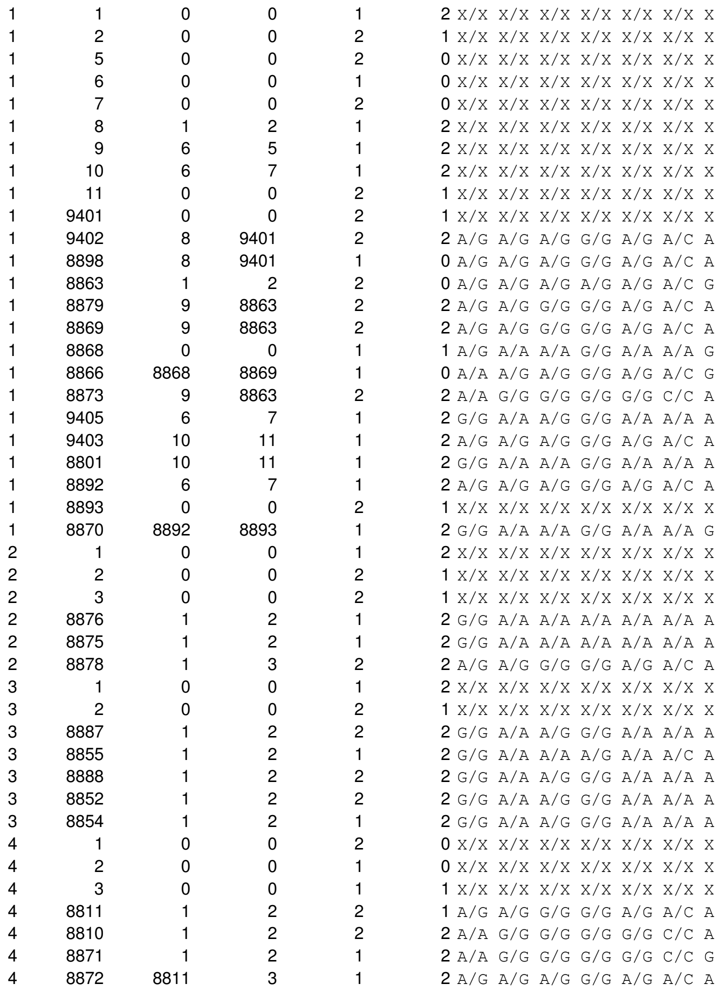
```
A        OPDM
M        rs876724
M        rs12714396
M        rs381726
M        rs300739
M        rs1350779
M        rs6548222
M        rs907302
M        rs1320362
M        rs2293085
M        rs7575263
M        rs6548255
M        rs7559853
M        rs5020134
M        rs7426276
M        rs938326
M        rs1368233
M        rs709276
M        rs1667023
M        rs1729916
M        rs6767
M        rs12988769
M        rs1024026
M        rs10174999
M        rs792065
M        rs813779
M        rs2118186
M        rs1079417
M        rs921229
M        rs309276
M        rs2352400
M        rs7598142
M        rs1560382
M        rs1364054
M        rs1025053
M        rs3102960
M        rs168293
M        rs2001660
M        rs12995394
M        rs1309
M        rs728282
M        rs4669630
M        rs1686430
M        rs730990
M        rs726843
M        rs1370548
M        rs6432244
M        rs4668758
M        rs1469217
M        rs956596
M        rs779343
M        rs1510834
M        rs765786
M        rs1862110
M        rs767624
M        rs340767
```

Figure A.1.: Exemplary .dat Merlin input file. Left column indicates that all SNPs are markers to be considered by the program. Right column lists the SNPs which have been genotyped. Note, that only a few markers are shown to fit this figure to one page.



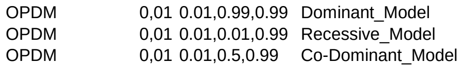
CHROMOSOME	MARKER	POSITION
2	rs876724	0,001765
2	rs12714396	0,002412
2	rs381726	0,0033
2	rs300739	0,00644
2	rs1350779	0,006729
2	rs6548222	0,137572
2	rs907302	0,427542
2	rs1320362	0,558847
2	rs2293085	0,767507
2	rs7575263	0,869165
2	rs6548255	0,939968
2	rs7559853	1,165881
2	rs5020134	1,208016
2	rs7426276	1,347943
2	rs938326	1,861933
2	rs1368233	2,657625
2	rs709276	4,87012
2	rs1667023	5,572099
2	rs1729916	6,740577
2	rs6767	10,90533
2	rs12988769	8,525858
2	rs1024026	9,285462
2	rs10174999	9,286284
2	rs792065	11,69467
2	rs813779	11,82321
2	rs2118186	11,93157
2	rs1079417	12,76831
2	rs921229	13,24073
2	rs309276	14,80846
2	rs2352400	14,80918
2	rs7598142	14,81039
2	rs1560382	14,86523
2	rs1364054	18,72203
2	rs1025053	18,77044
2	rs3102960	19,93544
2	rs168293	21,33897
2	rs2001660	22,32186
2	rs12995394	23,3701
2	rs1309	25,52654
2	rs728282	25,52657
2	rs4669630	25,52749
2	rs1686430	25,52749
2	rs730990	26,75621
2	rs726843	27,11846
2	rs1370548	27,96878
2	rs6432244	28,3712
2	rs4668758	29,83475
2	rs1469217	31,26161
2	rs956596	31,26445
2	rs779343	31,26473
2	rs1510834	32,54884
2	rs765786	32,73141
2	rs1862110	33,28779
2	rs767624	34,0628
2	rs340767	35,71467

Figure A.2.: Exemplary .map Merlin input file. This file connects all markers from the .dat file to a position on a Chromosome in Centimorgan. Note, that only a few markers are shown to fit this figure to one page.



```
1     1     0     0   1   2 X/X X/X X/X X/X X/X X/X X
1     2     0     0   2   1 X/X X/X X/X X/X X/X X/X X
1     5     0     0   2   0 X/X X/X X/X X/X X/X X/X X
1     6     0     0   1   0 X/X X/X X/X X/X X/X X/X X
1     7     0     0   2   0 X/X X/X X/X X/X X/X X/X X
1     8     1     2   1   2 X/X X/X X/X X/X X/X X/X X
1     9     6     5   1   2 X/X X/X X/X X/X X/X X/X X
1    10     6     7   1   2 X/X X/X X/X X/X X/X X/X X
1    11     0     0   2   1 X/X X/X X/X X/X X/X X/X X
1  9401     0     0   2   1 X/X X/X X/X X/X X/X X/X X
1  9402     8  9401   2   2 A/G A/G A/G G/G A/G A/C A
1  8898     8  9401   1   0 A/G A/G A/G G/G A/G A/C A
1  8863     1     2   2   0 A/G A/G A/G A/G A/G A/C G
1  8879     9  8863   2   2 A/G A/G G/G G/G A/G A/C A
1  8869     9  8863   2   2 A/G A/G G/G G/G A/G A/C A
1  8868     0     0   1   1 A/G A/A A/A G/G A/A A/A G
1  8866  8868  8869   1   0 A/A A/G A/G G/G A/G A/C G
1  8873     9  8863   2   2 A/A G/G G/G G/G G/G C/C A
1  9405     6     7   1   2 G/G A/A A/G G/G A/A A/A A
1  9403    10    11   1   2 A/G A/G A/G G/G A/G A/C A
1  8801    10    11   1   2 G/G A/A A/A G/G A/A A/A A
1  8892     6     7   1   2 A/G A/G A/G G/G A/G A/C A
1  8893     0     0   2   1 X/X X/X X/X X/X X/X X/X X
1  8870  8892  8893   1   2 G/G A/A A/A G/G A/A A/A G
2     1     0     0   1   2 X/X X/X X/X X/X X/X X/X X
2     2     0     0   2   1 X/X X/X X/X X/X X/X X/X X
2     3     0     0   2   1 X/X X/X X/X X/X X/X X/X X
2  8876     1     2   1   2 G/G A/A A/A A/A A/A A/A A
2  8875     1     2   1   2 G/G A/A A/A A/A A/A A/A A
2  8878     1     3   2   2 A/G A/G G/G G/G A/G A/C A
3     1     0     0   1   2 X/X X/X X/X X/X X/X X/X X
3     2     0     0   2   1 X/X X/X X/X X/X X/X X/X X
3  8887     1     2   2   2 G/G A/A A/G G/G A/A A/A A
3  8855     1     2   1   2 G/G A/A A/A A/G A/A A/C A
3  8888     1     2   2   2 G/G A/A A/G G/G A/A A/A A
3  8852     1     2   2   2 G/G A/A A/G G/G A/A A/A A
3  8854     1     2   1   2 G/G A/A A/G G/G A/A A/A A
4     1     0     0   2   0 X/X X/X X/X X/X X/X X/X X
4     2     0     0   1   0 X/X X/X X/X X/X X/X X/X X
4     3     0     0   1   1 X/X X/X X/X X/X X/X X/X X
4  8811     1     2   2   1 A/G A/G G/G G/G A/G A/C A
4  8810     1     2   2   2 A/A G/G G/G G/G G/G C/C A
4  8871     1     2   1   2 A/A G/G G/G G/G G/G C/C G
4  8872  8811     3   1   2 A/G A/G A/G G/G A/G A/C A
```

Figure A.3.: Exemplary .ped Merlin input file. First column names the family, second column the individual's name. Since the program requires a consistent pedigree, parents which have not been genotyped have to be added and numbered. Third column names the father and fourth column the mother of each individual giving the program all the information to reconstruct the pedigree. Fifth column indicates disease status, "1" stands for unaffected, "2" for affected and "0" for unknown. Then, SNP data is added after each individual according to the order of SNPs in the .map file. If a family member has not been genotyped "X/X" is used.



```
OPDM          0,01  0.01,0.99,0.99   Dominant_Model
OPDM          0,01  0.01,0.01,0.99   Recessive_Model
OPDM          0,01  0.01,0.5,0.99    Co-Dominant_Model
```

Figure A.4.: Exemplary .model Merlin input file. This file is required for parametric linkage analysis and consists of 4 fields per line: an affection status label (matching the data file), a disease allele frequency, a probability of being affected for individuals with 0, 1 and 2 copies of the disease allele (penetrances), and finally a label for the analysis model.

A.3. List of homozygous variants shared by OPDM3, OPDM4 and OPDM5

Chr	Position	Ref	Obs	Variant	Func
chr1	109792750	-	CGC	CELSR2:c.49_50insCGC:p.L17delinsPL	exonic
chr1	31905904	-	CAG	SERINC2:c.1116_1117insCAG: p.Q372delinsQQ	exonic
chr10	7605079	C	-	ITIH5:c.2154delG:p.M718fs	exonic
chr10	97920100	-	C	ZNF518A:c.4021_4022insC:p.L1341fs	exonic
chr11	111853108	-	C	DIXDC1:c.181_182insC:p.L61fs	exonic
chr11	118898437	C	-	SLC37A4:c.527delG:p.W176fs	exonic
chr11	118939941	-	C	VPS11:c.222_223insC:p.S74fs	exonic
chr11	125452303	-	C	EI24:c.735_736insC:p.P245fs	exonic
chr11	14101494	-	C	SPON1:c.602_603insC:p.A201fs	exonic
chr11	3661588	-	TGG	ART5:c.71_72insCCA:p.P24delinsPT	exonic
chr11	67786065	-	C	ALDH3B1:c.231_232insC:p.N77fs	exonic
chr11	67795380	-	C	ALDH3B1:c.1268_1269insC:p.P423fs	exonic
chr11	76751543	T	-	B3GNT6:c.948delT:p.L316fs	exonic
chr11	76751605	T	-	B3GNT6:c.1010delT:p.L337fs	exonic
chr12	124824739	-	GCCGCTGCT	NCOR2:c.5470_5471insAGCAGCGGC: p.S1824delinsSSGS	exonic
chr12	6777111	CTG	-	ZNF384:c.1153_1155del:p.385_385del	exonic
chr12	6938024	-	G	P3H3:c.419_420insG:p.R140fs	exonic
chr12	7080212	-	C	EMG1:c.126_127insC:p.S42fs	exonic
chr12	76424952	CTG	-	PHLDA1:c.568_570del:p.190_190del	exonic
chr12	9994450	GTT	-	KLRF1:c.377_379del:p.126_127del	exonic
chr14	24646413	-	AAG	REC8:c.688_689insAAG:p.A230delinsEA	exonic
chr14	53619494	-	CGCCGC	DDHD1:c.323_324insGCGGCG: p.S108delinsSGG	exonic
chr14	73957982	-	C	C14orf169:c.260_261insC:p.A87fs	exonic
chr15	35230936	-	GTTA	AQR:exon10:c.718+2-TAAC	splicing
chr15	93198687	GAGCTG	-	FAM174B:c.198_203del:p.66_68del	exonic
chr16	138773	-	G	NPRL3:c.930_931insC:p.T310fs	exonic
chr16	2059625	C	-	ZNF598:c.124delG:p.G42fs	exonic
chr16	3602230	G	-	NLRC3:c.2318delC:p.A773fs	exonic
chr17	26699368	-	C	SARM1:c.315_316insC:p.C105fs	exonic
chr17	26727723	A	-	SLC46A1:c.1142delT:p.I381fs	exonic
chr17	43192550	-	C	PLCD3:c.1622_1623insG:p.R541fs	exonic
chr17	61660895	G	-	DCAF7:c.561delG:p.G187fs	exonic
chr17	6555548	-	G	C17orf100:c.315_316insG:p.R105fs	exonic
chr17	7470288	A	-	SENP3:c.1308delA:p.K436fs	exonic
chr17	7750216	-	ACCACC	KDM6B:c.791_792insACCACC: p.P264delinsPPP	exonic
chr17	79614938	AACT	-	TSPAN10:c.682_685del:p.228_229del	exonic
chr17	8725216	-	G	PIK3R6:c.1826_1827insC:p.S609fs	exonic
chr18	19100762	-	CTT	GREB1L:c.5586_5587insCTT:p.L1862delinsLL	exonic
chr18	43833704	-	CTG	C18orf25:c.757_758insCTG:p.G253delinsAG	exonic
chr18	74090964	G	-	ZNF516:c.3106delC:p.P1036fs	exonic
chr19	16268213	A	-	HSH2D:c.668delA:p.K223fs	exonic
chr19	21299776	-	AAT	ZNF714:c.306_307insAAT:p.Y102delinsYN	exonic
chr19	2340156	-	C	SPPL2B:c.824_825insC:p.P275fs	exonic
chr19	30500143	TGA	-	URI1:c.798_800del:p.266_267del	exonic
chr19	36258940	G	-	PROSER3:c.1193delG:p.G398fs	exonic
chr19	41123095	-	G	LTBP4:c.3034_3035insG:p.V1012fs	exonic
chr19	41173904	TTGCTG	-	NUMBL:c.1294_1299del:p.432_433del	exonic
chr19	4954680	-	C	UHRF1:c.2015_2016insC:p.A672fs	exonic
chr19	51835893	-	G	VSIG10L:c.2576_2577insC:p.A859fs	exonic;splicing
chr19	56599452	GTC	-	ZNF787:c.1087_1089del:p.363_363del	exonic
chr19	58718361	-	G	ZNF274:c.216_217insG:p.E72fs	exonic
chr2	31805882	-	G	SRD5A2:c.88_89insC:p.P30fs	exonic
chr2	95847047	GCG	-	ZNF2:c.348_350del:p.116_117del	exonic
chr20	21186163	-	G	KIZ:c.987_988insG:p.R329fs	exonic
chr20	278701	GGC	-	ZCCHC3:c.474_476del:p.158_159del	exonic
chr20	32664865	-	AGC	RALY:c.642_643insAGC:p.A214delinsAS	exonic
chr21	34166190	A	T	C21orf62:c.T543A:p.F181L	exonic
chr3	12942852	C	-	IQSEC1:c.2976delG:p.L992fs	exonic
chr3	14561629	-	G	GRIP2:c.1309_1310insC:p.P437fs	exonic
chr3	16926642	A	G	PLCL2:c.A94G:p.T32A	exonic
chr3	16926648	T	G	PLCL2:c.T100G:p.S34A	exonic
chr3	50251835	-	G	SLC38A3:c.103_104insG:p.V35fs	exonic;splicing
chr3	50306757	-	C	SEMA3B:c.85_86insC:p.H29fs	exonic
chr4	140651610	TGC	-	MAML3:c.2277_2279del:p.759_760del	exonic
chr4	177605086	CAT	-	VEGFC:c.1252_1254del:p.418_418del	exonic
chr4	184367560	TGC	-	CDKN2AIP:c.723_725del:p.241_242del	exonic

chr22	37602586	-	C	SSTR3:c.1257_1258insG:p.X419delinsX	exonic
chr5	140568035	-	A	PCDHB9:c.1143_1144insA:p.T381fs	exonic
chr5	176930176	AGG	-	DOK3:c.555_557del:p.185_186del	exonic
chr5	77745854	-	A	SCAMP1:c.730_731insA:p.I244fs	exonic
chr6	160211649	GTT	-	MRPL18:c.30_32del:p.10_11del	exonic
chr6	161519381	CTG	-	MAP3K4:c.3596_3598del:p.1199_1200del	exonic
chr6	170871039	-	GCA	TBP:c.155_156insGCA:p.Q52delinsQQ	exonic
chr6	28239933	-	G	ZSCAN26:c.236_237insG:p.C79fs	exonic
chr6	30558478	-	A	ABCF1:c.2424_2425insA:p.X808delinsX	exonic
chr7	128533516	-	C	KCP:c.940_941insG:p.G314fs	exonic
chr7	128550685	C	-	KCP:c.46delG:p.G16fs	exonic
chr7	149426307	-	C	KRBA1:c.1656_1657insC:p.A552fs	exonic
chr7	150713903	-	C	ATG9B:c.2295_2296insG:p.E765fs	exonic
chr7	28997597	-	C	TRIL:c.66_67insG:p.L22fs	exonic
chr8	145106943	CC	-	OPLAH:c.3497_3498del:p.1166_1166del	exonic
chr8	145738769	G	-	RECQL4:c.2296delC:p.P766fs	exonic
chr8	30620844	-	T	UBXN8:c.625_626insT:p.X209delinsL	exonic
chr8	38827187	C	-	PLEKHA2:c.1164delC:p.A388fs	exonic
chr8	86126830	-	ATTAAC	C8orf59:c.262_263insGTTAAT: p.V88delinsVNV	exonic
chr9	123476563	GCGGCG	-	MEGF9:c.69_74del:p.23_25del	exonic

Table A.28.: List of homozygous variants in genes expressed in skeletal muscle shared by OPDM3, OPDM4 and OPDM5.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt,

dass ich die vorliegende Dissertation mit dem Thema

Mapping the Chromosomal Locus of Oculopharyngodistal Myopathy with Microsatellite Markers and Next Generation Sequencing

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 02.07.2018	Matias Wagner
Ort, Datum	Unterschrift Doktorand