LINKAGE AND EXOME SEQUENCE ANALYSES TO IDENTIFY DISEASE GENES

by

Özgecan Ayhan

B.S., Mathematics Teacher Training Education, Marmara University, 2010

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2013

dedicated to my beloved father Ahmet Ayhan

ACKNOWLEDGEMENTS

I would like to sincerely thank my thesis supervisor Prof. Aslıhan Tolun whose guidance, support and encouragement throughout the study helped my transformation of being a mathematician to being a geneticist, and enabled me to gain valuable skills and a broader perspective towards scientific questions. I also extend my appreciation to the members of my thesis committee, Assist. Prof. Necla Birgül İyison and Assoc. Prof. Eda Tahir Turanlı for allocating their time to evaluate this work.

I would like to thank all previous and present members of Kommagene laboratory, especially Çiğdem Köroğlu for her help in lab work and her friendship, Ayşe Güven, Atalay Tok, Sibel Uğur, Esra Yıldız, Nehir Mavioğlu, and Ozan Baytaş for everything. I would like to extend my thanks to all my friends in the department, among whom I would like to mention Gizem Gül, Hilal Kahraman, Tuncay Şeker, and Burçin Duan for their support and great friendship.

We are grateful to TUBITAK Advanced Genomics and Bioinformatics Group (IGBAM) for sharing with us the Turkish exome sequence database.

I would like to convey my great appreciation to all members of my family. I am especially grateful for the endless support and love of my parents.

This work was supported by Bogazici University Research Projects (BAP 6655 and 7695).

ABSTRACT

LINKAGE AND EXOME SEQUENCE ANALYSES TO IDENTIFY DISEASE GENES

Identification of the genetic basis of a hereditary disease provides important information about the molecular mechanism of the disease and the function of the gene and could contribute to the development of therapeutic means. Linkage analysis is one of the most powerful methods used to map a disease gene locus. For a recessive disease, a consanguineous family is important for the identification of the disease locus since the affected individuals share the disease haplotype in the homozygous state, descending from a single ancestor. Whole exome sequencing is a very efficient method for the identification of the causative mutation. After the identification of the disease gene locus by linkage analysis, exome sequence analysis can be applied to find the causative mutation at the candidate locus. In this study, identification of the disease loci and the causative gene defects for six inherited disorders were attempted. The study families were afflicted with *Congenital Disorder of Glycosylation* (CDG), *Myosin Storage Myopathy* (MSM), *Desmin Deficiency Myopathy* (DDM), *Hereditary Brain Tumors* (HBT), *Epilepsy*, or *Muscular Dystrophy* (MD). Stopgain mutation *SRD5A3* p.W19X was found in CDG family, missense mutation p.R1820W in *MHY7* was identified in MSM family, and frameshift mutation p.N116Qfs*2 in *DES* in DDM family. Three mutations were identified in HBT family: frameshift p.E138Vfs*93 in *CARHSP1*, missense p.Y193C in *RGS22* and p.E363G in *METTL22*. In *Epilepsy* family no disease locus could be identified. Synonymous mutation *LMNA* c.346C>T was identified in MD family.

ÖZET

HASTALIK GENLERİ BELİRLEMEK İÇİN BAĞLANTI VE EKSOM DİZİLEME ANALİZLERİ

Kalıtsal bir hastalığın genetik temelinin belirlenmesi hastalığın moleküler mekanizması ile genin işlevi hakkında önemli bilgiler sağlar ve tedavi araçları geliştirilmesine katkıda bulunabilir. Bağlantı analizi hastalık geni bölgesi haritalamasında kullanılan en kuvvetli yöntemlerden biridir. Çekinik bir hastalıkta akraba evliliği yapmış bir aile hastalık geninin bulunduğu bölgenin belirlenmesinde çok yararlıdır, çünkü hastalar hastalık haplotipini homozigot şekilde, aynı ortak atadan kalıtırlar. Tüm eksom dizileme ise hastalığa neden olan mutasyonun belirlenmesi için çok yararlı bir yöntemdir. Bağlantı analizi ile hastalık geninin bölgesi belirlendikten sonra, eksom dizileme analizi aday bölgede hastalığa neden olan mutasyonun bulunması için uygulanabilir. Bu çalışamada, altı kalıtsal hastalıkta hastalık genlerinin bölgelerinin ve hastalığa neden olan gen kusurlarının bulunması hedeflendi. Çalışmaya katılan ailelerdeki hastalıklar şunlardır: *Doğuştan Glikosilasyon Hastalığı* (DGH), *Miyozin Birikim Miyopatisi* (MBM), *Desmin Eksikliği Miyopatisi* (DEM), *Kalıtsal Beyin Tümörleri* (KBT), *Epilepsi* ve *Kas Distrofisi* (KD). DGH için *SRD5A3* geninde anlamsız p.W19X mutasyonu, MBM için *MYH7* geninde yanlış anlamlı yeni p.R1820W mutasyonu ve DEM için *DES* geninde çerçeve kayması yeni p.N116Qfs*2 mutasyonu bulundu. KBT ailesinde ise üç mutasyon belirlendi: *CARHSP1* geninde çerçeve kayması yeni p.E138Vfs*93, *RGS22* geninde yanlış anlamlı yeni p.Y193C ile *METTL22* geninde yanlış anlamlı yeni p.E363G. *Epilepsi* ailesinde hastalık geninin bölgesi bulunamadı. KD ailesinde ise *LMNA* geninde yeni eşanlamlı c.346C>T mutasyonu bulundu.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AD	Autosomal dominant
APS	Ammonium peroxodisulphate
AR	Autosomal recessive
BAM	Binary Alignment/Map
bp	Base pair
BSA	Bovine serum albumine
BWA	Burrows-Wheeler aligner
CDG	Congenital Disorder of Glycosylation
cDNA	Complementary deoxyribonucleic acid
Chr	Chromosome
cM	Centimorgan
DDM	Desmin Deficiency Myopathy
DGV	Database of Genomic Variants
dH_2O	Distilled water
DMSO	Dimethyl sulfate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetate
EVS	Exome Variant Server
HCiE	Homozygosity Comparison in Excel
HHT	Hereditary Head Tumors
HRM	High Resolution Melting
IBD	Identical by descent
kb	Kilobase pair
LOD	Logarithm of odds
MAF	Minor allele frequency
Mb	Mega base
MD	Muscular Dystrophy
min	Minute

mRNA	Messenger ribonucleic acid
MSM	Myosin Storage Myopathy
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
OMIM	Online Mendelian Inheritance in Man
PAR	Pseudoautosomal region
PCR	Polymerase chain ceaction
RNA	Ribonucleic acid
rpm	Revolution per minute
RT-PCR	Reverse-transcription polymerase chain reaction
SAM	Sequence Alignment/Map
SDS	Sodium dodecyl sulphate
sec	second
SEM	Standard error of the mean
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SSCP	Single strand conformation polymorphism
TBE	Tris-boric acid-EDTA
TE	Tris-EDTA
TEMED	N, N, N, N’-Tetramethylethylenediamine
UTR	Untranslated region

1. INTRODUCTION

Identification of a disease gene facilitate the elucidation of the molecular mechanism of the disease and the function of the gene. In general, identification of molecular mechanism of a disease does not facilitate the treatment of the patients in the study family but can contribute to the development of therapeutic means, i.e. gene therapy and medication. However, genetic studies in consanguineous families can help the study family and other future families with mutations in the gene by genetic counseling and prenatal diagnosis.

Genetic research for the purpose of identification of disease genes in five recessive and one dominant disease was performed in this study. Linkage analysis was applied to search for gene loci, and causative mutations were searched by either candidate gene approach or exome sequencing.

1.1. Congenital Disorder of Glycosylation (CDG)

Congenital disorders of glycosylation (CDG) are genetically heterogeneous inborn errors of metabolism due to defects in protein glycosylation. Glycosylation is a process in which a carbohydrate is added, creating a glycoconjugate, and it is the one of the major post-transcriptional protein modification mechanisms. Glycoconjugates are important for cell recognition and adhesion, metabolism, protease resistance, cell migration, antigenicity and host defense (MIM 212065).

CDG is divided into two major groups: CDG type I and CDG type II. In type I, assembly and addition of sugars to the nascent protein are impaired whereas proper trimming and folding of the new glycoprotein are defective in type II (Wolfe and Krasnewich, 2013). Another classification for CDG that also depends on the type of glycosylation defect is N-linked glycosylation defects, O-linked glycosylation defects, lipid glycosylation defects and multiple glycosylation pathways defects (Jaeken, 2011).

The spectrum of clinical presentations of CDG is very broad. About 60 different CDG types and many gene defects causing CDG have been reported since the first clinical description in 1980 (Jaeken, 2011, Matthijs *et al*., 2013). One of the genes responsible for the disease is *SRD5A3,* encoding steroid 5-alpha-reductase 3, that converts the alpha-isoprene unit of polyprenol to dolichol in humans (Cantagrel, 2010). Defects in *SRD5A3* cause congenital disorder of glycosylation type Iq (SRD5A3-CDG; MIM 611715) and Kahrizi syndrome (MIM 611715). Ten different recessive mutations in 15 children with variable phenotypes from 12 families have been reported so far (Al-Gazali *et al.,* 2008; Cantagrel *et al.*, 2010; Morava *et al.*, 2010; Gründahl *et al.*, 2012; Kasapkara *et al.,* 2012). The major clinical symptoms are severe early visual impairment and variable ocular anomalies including coloboma, optic nerve hypoplasia, congenital cataract and glaucoma as well as some degree of intellectual disability. Cerebellar ataxia, cerebellar malformations, ichthyosiform skin disorders, hypertrichosis, congenital heart defects and abnormal coagulation are other clinical features of the disease, and these features vary considerably both among families and within a family (Al-Gazali *et al.*, 2008).

To date, all reported patients afflicted with SRD5A3-CDG were children. In this study, two patients, aged 38 and 40, with an initial diagnosis of cerebellar ataxia and born to consanguineous parents are studied. Four candidate loci were found, and in one of those a homozygous *SRD5A3* nonsense mutation was detected.

1.2. Myosin Storage Myopathy (MSM)

Myosin storage myopathy (MSM; MIM 608358) is a protein aggregate myopathy characterized by subsarcolemmal hyalinized bodies in type I muscle fibers. MSM can be caused by sporadic (Sahgal and Sahgal, 1977; Ceuterick *et al*., 1993; Barohn *et al.*, 1994) or familial mutations in *MYH7,* coding for beta-cardiac myosin heavy chain protein in muscle (MYH7, MIM 160760). *MYH7* defects also cause other disorders, namely, Laing distal myopathy (MIM 160500), familial hypertrophic cardiomyopathy (MIM 192600) and dilated cardiomyopathy (MIM 613426).

Most of the familial *MYH7* mutations are dominant. Dominant *MYH7* mutations causing MSM are as follows: K1784delK, L1793P, R1845W, E1886K, H1901L and X1936WfsX32. There is only one case of MSM that is caused by a recessive, homozygous mutation, p.E1883K; associated features are hypertrophic cardiomyopathy and respiratory failure (Tajsharghi *et al.*, 2007). All mutations reported to cause MSM are located in the last four exons (37-40) of the gene (Tajsharghi and Oldfors, 2013).

Two brothers from a consanguineous family afflicted with autosomal recessive MSM were studied. One of the patients had an unusual combination of myosin storage myopathy, dilated cardiomyopathy and respiratory and scapuloperoneal weakness. Four candidate loci were found, and *MYH7* was located in one of them. The last four exons of *MYH7* were sequenced, and a homozygous missense mutation was identified.

1.3. Desmin Deficiency Myopathy (DDM)

Desmin is a member of muscle-specific intermediate filament protein family. It is coded by *DES* (MIM 125660) and has an important role in the maintenance of sarcomeric integrity and tensile strength in cardiac, skeletal and smooth muscles. Defects in *DES* cause four disorders, namely, limb girdle muscular dystrophy type 2R (LGMD2R, MIM 615325), dilated cardiomyopathy (CMD1I, MIM 604765), myofibrillar myopathy-1 (MFM1, MIM 601419), and Kaeser type neurogenic scapuloperoneal syndrome (SCPNK, MIM 181400). More than 40 mutations most of which were inherited in an autosomal dominant fashion or arose by heterozygous de-novo mutations have been reported (Goldfarb *et al.*, 2004). Additionally, recessive mutations were reported in a few families afflicted with infantile onset cardiomyopathy, severe myopathy, or classical limb-girdle muscular dystrophy (Carlsson and Thornell, 2001; Pinol-Ripoll *et al.*, 2009; Cetin *et al.*, 2013; Henderson *et al.*, 2013).

Studies on mice models showed that desmin null mice were viable and fertile but were afflicted with a multisystem disorder involving cardiac, skeletal and smooth muscle (Milner *et al.*, 1996; Li *et al.*, 1997). In the absence of desmin, the primary muscle formation and regeneration was normal but disorganized, some fibers were

distended and nonaligned, and the fibers were more susceptible to damage during contraction. Additionally, lack of desmin resulted in weaker and more easily fatigued mice (Milner *et al.,* 1996; Li *et al.,* 1997).

A family afflicted with severe fatigable myopathy with infantile onset, generalized muscle weakness in among others facial and extraocular muscles, cardiomyopathy, and respiratory involvement was studied. Since no similar disease was reported at the time, linkage analysis was performed and the maximal logarithm of the odds (LOD) score was found at 2q35-2q37.1, and by exome sequencing a homozygous frame-shift deletion in *DES* was identified.

1.4. Hereditary Head Tumors (HHT)

Cancer is a genetic but mostly not inherited disease that is caused by mutations in genes that play important roles in cell cycle, causing the cells to divide in an uncontrollable fashion. There are two types of genes the defects in which can cause cancer: oncogenes and tumor suppressor genes. Oncogenes promote cell growth and proliferation, and overexpression or increased activity of those genes can cause cancer. Tumor suppressor genes inhibit cell division and survival, and their underexpression or impaired function can lead to cancer. However, cancer is a complicated process and mostly not inherited in a Mendelian fashion. Some genes can function as both an oncogene and a tumor suppressor; an example is *p53*.

Brain tumors are neoplasms that occur within the brain. They are diverse tumors which are either benign or malignant, and may be primary tumors that originated in the nervous system. Meningiomas (MIM 607174) are derived from the arachnoidal (the middle of the 3 meninges) cap cells. Meninges, consisting of 3 connective tissue layers, are the membranes that envelop the brain and spinal cord. Meningiomas are the most common central nervous system neoplasms in humans, usually benign and generally sporadic. However, some multiple and familial cases are reported, and those cases are mostly associated with neurofibromatosis type II (Zang, 2001). Heterozygous mutations in *SMARCE1* (Smith *et al*., 2013), *PTEN* (Staal *et al*., 2002), *SUFU* (Aavikko *et al*.,

2012) and *PDGFB* (Bolger *et al*., 1985), a translocation in *MN1* (Lekanne Deprez *et al*., 1991) and somatic mutations in *NF2* (Wellenreuther *et al*., 1995) were identified in meningioma.

Gliomas (MIM 137800) are also central nervous system neoplasms but are derived from glial cells. World Health Organization (WHO) classified gliomas into astrocytomas, oligodendrogliomas, ependymomas and oligo-astrocytomas (mixed gliomas) (Louis *et al.*, 2007). Astrocytomas are subdivided into pilocytic, grade I; diffuse, grade II; anaplastic, grade III; and glioblastoma multiforme (GBM), grade IV, according to the grade of malignancy (Louis *et al.*, 2001). Glioblastomas are aggressive type of gliomas and account for 60-70% of all gliomas (Wen and Kesari, 2008). Several families with autosomal recessive and dominant inheritance or complex inheritance were reported (Malmer *et al.*, 2005; de Andrade *et al.*, 2001).

Many genes are implicated in gliomas. Further, many loci and a few genes are reported for genetic susceptibility to gliomas. Nonetheless, genetic basis of the majority of gliomas are yet unknown. Any of the numerous genes implicated in cancer can be considered a candidate for glioma as well. One example is *CARHSP1,* encoding calcium regulated heat stable protein 1 that contains a cold-shock domain with two RNA binding motifs (Schafer *et al.*, 2003). Pfeiffer *et al.* (2011) demonstrated that CARHSP1 interacts with an AU rich element (ARE) in the 3’ UTR of tumor necrosis factor alpha (TNF-α) mRNA, increasing its half-life. Knockdown of *CARHSP1* in a mouse macrophage cell line leads to a decrease in *TNF*-α mRNA level, suppressing TNF-α synthesis, suggesting a key feature in the regulation of *TNF-α*. The authors concluded that CARHSP1 is necessary for effective *TNF-α* mRNA stabilization (Pfeiffer *et al.*, 2011). TNF-α is a cytokine that plays essential roles in signaling pathways by binding its receptors TNFR1 or TNFR2. TNF-α has anticancer properties, which are exerted by the activation of caspase cascades and induction of apoptosis. It has been used as an anticancer agent to treat patients with advanced solid tumors. However, TNF-α has a dual role as both an oncogene and a tumor suppressor gene, as is *p53* (Bertazza and Mocellin, 2010).

RGS22 is another gene associated with cancer. The gene encodes a regulator of G-protein signaling 22 that belongs to a family of proteins that regulate heterotrimeric G proteins. G-proteins are responsible for many cellular processes, such as proliferation, differentiation, migration, embryonic development and membrane trafficking, and their regulators are needed for their proper function. The protein product represses cancer metastasis by suppressing cell migration and invasion (Hu *et al.,* 2008 and 2011).

METTL22 is a protein that has a substrate protein linked to cancer. *METTL22* (formerly *C16orf64)* encodes a methyltransferase-like 22 which is a non-histone lysine methyltransferase (Cloutier *et al.,* 2013). Kin17, encoded by *KIN*, is a DNA/RNA binding protein and has a role in DNA repair and replication and mRNA processing (Cloutier *et al.,* 2014). METTL22 protein interacts with Kin17 and is responsible for trimethylation of lysine at residue 135 of Kin17 leading to the detachment of the protein from chromatin, and then the protein is seen in cytoplasm. Cloutier *et al.* (2014) suggested that Kin17 may have a chaperon function in cytoplasm after trimethylation by *METTL22.* Variations in *KIN* expression was observed in several tumor cell lines, and authors argued that Kin17 lysine 135 trimethylation could be perturbed in cancer and, therefore, *METTL22* may represent a novel therapeutic target.

A non-consanguineous family with five members afflicted with either meningioma or glioblastoma was previously studied by Tarık Bozoğlu as a M.S. thesis project (Bozoğlu, 2008). DNA samples of fifteen family members were subjected to genome scan with microsatellite markers. Linkage analysis was performed assuming autosomal dominant inheritance with either full or 70% penetrance. A maximal LOD score of 3.01 was obtained at 16p13.2-p13.12. *GSPT1*, *GRIN2A* and *SOCS1* genes were selected as candidate genes but were not analyzed for mutations.

In the present thesis, locus 16p13.2-p13.12 was confirmed and a nonsense/stopgain mutation in *CARHSP1* was identified by exome sequence analysis. Analysis of the coding exons of *CARHSP1* in a group of individuals with diverse tumors identified one novel nonsynonymous and three novel or very rare synonymous variants. Additionally, a very rare missense mutation in *RGS22* and a novel missense mutation in *METTL22* were detected.

1.5. Epilepsy Syndrome

Epilepsy is a common, heterogeneous neurological disorder affecting approximately 1% of the population and characterized by recurrent and unprovoked seizures (Poduri and Lowenstein, 2011). It is clinically categorized as symptomatic, presumed symptomatic, and idiopathic according to International League Against Epilepsy. Genetics plays an important role in idiopathic epilepsies besides the non-genetic factors, and especially genetic susceptibility contributes to the majority of cases. Many gene defects were identified in epilepsy, and the majority of the genes encode Na+ or K+ channel proteins (Poduri and Lowenstein, 2011). However, in some cases, epilepsy is syndromic, having various other symptoms.

We studied a family afflicted with epilepsy, mental retardation and facial dysmorphism. Linkage analysis was performed in several ways, i.e. excluding some of the patients or assuming an inheritance model of autosomal recessive or X-linked, but no strong candidate locus was found for the syndrome. Exome sequence analysis was performed with the aim of identifying the disease gene, but unfortunately the data for another sample were analyzed by mistake.

1.6. Muscular Dystrophy (MD)

Muscular dystrophies (MD) are heterogeneous disorders characterized by progressive skeletal muscle wasting and weakness of variable distribution and severity (Emery, 2002). Many forms of muscular dystrophies are known; some examples are Duchenne type (MIM 310200), Becker type (MIM 300376), Emery-Dreifuss types (MIM 310300 and 181350), and Limb-girdle muscular dystrophies.

One class of muscle disorders are laminopathies, caused by lamin defects. Two lamin proteins are lamin A and lamin C, coded by *LMNA* (MIM 150330), which are structural protein components of the nuclear lamina, a meshwork of proteins underneath the nuclear envelope. Mutations in the *LMNA* gene can lead to various disorders that are classified in four major types: diseases of striated muscle, lipodystrophy syndromes,

peripheral neuropathy, and premature aging (Worman and Bonne, 2007). Muscular dystrophies caused by *LMNA* gene defects include Emery-Dreifuss muscular dystrophy (MIM 181350), congenital muscular dystrophy (MIM 613205), and limb-girdle muscular dystrophy type 1B (MIM 159001). Inheritance model is either autosomal recessive or dominant.

A family having three children afflicted with MD was studied. Five candidate loci were found by linkage analysis. *LMNA*, residing at the largest candidate locus, was assessed as a candidate gene. Since no variants in *LMNA* was found by exome sequence analysis, all exons except a part of the 3' UTR region were sequenced by Sanger method since some in/dels can be missed by exome sequencing. A synonymous variant was identified in *LMNA*.

1.7. Linkage Analysis

Genetic linkage is the tendency of two loci that are located proximal to each other on a chromosome to be inherited together. DNA recombination takes place when homologous chromosomes exchange parts in meiosis. Two loci that are close to each other on the same chromosome are likely to segregate together. Alternatively, new combinations of alleles are often created by recombination, and two loci are not inherited together; they are said to be not linked. The frequency of a crossover between two loci is the recombination frequency, denoted by θ. The maximum value for θ is 0.5, meaning that the probabilities of recombinant and parental chromosomes are equal and the two loci are far apart from each other. The recombination frequency is directly related to the physical distance between the two loci and increases as the distance between the loci increases. Centimorgan (cM) is the unit of the recombination frequency, and one cM is roughly equal to 1% probability of recombination between two loci in subtelomeric regions.

Various types of genetic markers, commonly variable number tandem repeats (VNTRs), restriction fragment length polymorphisms (RFLPs), short tandem repeats (STRs or microsatellites) and single nucleotide polymorphisms (SNPs) are used to

observe recombination events. SNP genotyping arrays are generally used in linkage analysis for the localization of disease genes because of low cost due to automation, high marker density, and accessibility of information about them via online databases.

There are two types of linkage analysis, nonparametric (model-free) and parametric (model-based). Nonparametric linkage analysis is used in studies on complex traits whereas parametric linkage analysis is applied for monogenic diseases. In parametric linkage analysis, predetermined genetic parameters such as disease allele frequency, penetrance and mutation rates are taken into account. The statistical method to evaluate parametric linkage analysis is the calculation of LOD scores.

1.7.1. LOD Score Analysis

LOD score analysis considers null hypothesis of no linkage ($\theta = 0.5$) and alternative hypothesis of linkage ($\theta < 0.5$) (Morton, 1955). Odds ratio represents the ratio of the probability of observing the distribution of two loci in a pedigree under linkage at θ to the probability of observing it under the hypothesis of no linkage ($\theta = 0.5$) (Risch, 1992). LOD score represents the logarithms of the odds ratio, and a LOD score of 3 is the mostly used threshold for autosomal trait localization and indicates that the observed linkage most likely has not occurred by chance in 1000 to 1 odds (Terwilliger, 1955). However, maximal LOD score depends also on the inbreeding coefficient of the family as the highest maximal LOD score can be <3 in some families. The threshold value for LOD score is related to the power of the pedigree, that is, in a very large family a locus with >3 LOD score may not be the disease gene locus.

Two types of LOD score calculations are used: two-point and multipoint. In two point LOD score calculations cosegregation of each marker with the disease locus is investigated, whereas multipoint LOD score calculations evaluate linkage of a disease to several markers in a row. easyLINKAGE package is a user friendly platform that converts all input information to a standard form for linkage analysis (Hoffman and Lindner, 2005).

1.7.2. Homozygosity Mapping

Homozygosity mapping is applied to find disease loci in consanguineous families afflicted with autosomal recessive diseases, based on the assumption that affected members of a consanguineous family or an inbred population are likely to share a haplotype in the homozygous state that has descended from a recent common ancestor (Lander and Botstein, 1987). Due to the very low frequency of a rare disease, such a disease can only be seen in a consanguineous family or an inbred population, and the affected members of the family share a genotype that is homozygous for the disease haplotype, i.e. the parental disease haplotypes are identical by descent. Homozygosity mapping is a parametric linkage analysis and frequently used to localize the causative gene in a consanguineous family with two or more affected members.

HomozygosityMapper (http://www.homozygositymapper.org/) is an online tool that allows uploading of the genotypes and detecting homozygous regions with specified parameters, such as lower and upper limits for the length of the homozygous regions to be detected and exclusion of the regions that are homozygous also in the controls.

1.8. Candidate Gene Approach

Linkage analysis or homozygosity mapping is the first step in gene hunt since they identify the locus/loci that could possibly harbor the disease gene. The second step is candidate gene approach in which the genes at the identified locus/loci are evaluated to assess which of them could be good candidates to be the disease gene(s). Genes that are associated with similar diseases are considered with priority, and they can be found in the Morbid/Disease lists under NCBI-MapViewer. Also, the pathway that the gene product is involved in, expression in relevant tissues, gene knockout animal models with similar disease phenotypes and the function of the gene product are other criteria for evaluating candidate genes.

1.9. Exome Sequencing

Human exome comprises approximately 1% of the human genome, with a total size of approximately 30 Million bases (Mb). Since approximately 85% of the disease-causing mutations are located in protein coding regions, it is very useful to sequence those regions to find the causative mutation. Exome sequencing is one of the Next Generation Sequencing (NGS) methods that selectively sequences exons of almost all known genes. It has been widely used in gene search for both Mendelian and complex diseases after the advances in the next generation technologies that made exome sequencing more cost-effective and faster as compared to Sanger sequencing of all the candidate genes, whenever the disease locus harbors many genes (Ku *et al*., 2011).

In exome sequencing, DNA is first sheared into fragments by nebulization. Then exons and flanking regions are captured by probes in either solid or liquid phase and amplified by primers. Lastly, the amplified regions are subjected to sequencing on next generation sequencing platforms, creating sequence reads. In the bioinformatics studies, exome sequence reads are aligned to the reference genome, and variants, the sequences differing from the reference sequence, are called. Evaluation of the candidate variants is based on various filtering strategies. Filtering of variants at the locus identified by genetic linkage analysis is generally the first step in filtering the exome variants. Subsequent filtering are genetic filter and functional filter. In the genetic filter, the exonic and splicing variants that are rare with minor allele frequency (MAF) < 0.01 in variant databases such as dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) or exome variant server (EVS, http://evs.gs.washington.edu/EVS/), the alt depth/total depth ratio greater than 0.6 (for recessive diseases), and not present in the other in-laboratory exome sequencing files are selected. In functional filter, variants that are selected are those that can possibly affect protein function (i.e., all exonic/splicing variants except for the synonymous variants), are in good candidate genes associated with phenotypes similar to the patients, an animal knockout model with phenotype similar to the patients is reported, in genes expressed in relevant tissues, substituting an evolutionarily conserved amino acid, and predicted to be damaging by online tools. Those remaining after the final filter are the candidate variants.

Although exome sequencing is a powerful technique, it can present artifacts. First of all, exome sequencing, similarly to other NGS technologies, has a much higher base calling error rate as compared to Sanger sequencing (Ku *et al.*, 2011). Therefore, candidate variants need to be validated by Sanger sequencing. Incomplete coverage of some exons is another drawback. Uncovered regions at candidate loci should be detected, and the ones that are in good candidate genes should be investigated for mutations by Sanger sequencing. Lastly, some or all of the exons of some genes are not included in the capture chip.

1.10. Analysis of Gene Expression

Gene expression refers to the synthesis of the gene product, which is either a protein or a non-protein coding RNA such as a ribosomal RNA (rRNA), transfer RNA (tRNA) and small nuclear RNA (snRNA). Many intrinsic and extrinsic factors regulate gene expression since gene expression is generally tissue and/or developmental stage specific, except for the housekeeping genes that are constitutively expressed. Various approaches can be used to analyze gene expression at the RNA level, such as real-time quantitative PCR (RT-qPCR), northern blotting, and high throughput analyses including microarrays, SAGE (serial analysis of gene expression) and RNA sequencing. There are also online databases such as UniGene (NCBI) and GeneSorter (UCSC Genome Browser) that display information about gene expression profiles with respect to tissue, age or health state. Additionally, access to gene expression profiles has improved with the help of the Encyclopedia of DNA Elements (ENCODE) project. However, the assessment can be wrong or incomplete since they are based on experimental results under special conditions such as a certain tissue or developmental stage.

RT-qPCR is a method to measure the abundance of transcripts of a specific gene in a sample. First total RNA is isolated from the tissue of interest and converted to cDNA by reverse transcriptase using random hexamer or oligo-dT primers. Then cDNA is amplified using intron spanning primers specific to the gene of interest. The amount of the product is measured by fluorescence emitted by DNA intercalating dyes, and the amount of the cDNA template, considered an indicator of the initial amount of RNA, is

deduced. Cycle threshold (C_T) is the cycle that the amplification began its exponential phase, in which the DNA fragment is doubled in each cycle.

RT-qPCR can be either an absolute or a relative quantification. Absolute quantification determines the absolute copy number of transcripts by relating to a standard curve. In relative quantification, the amount of the transcript of the gene of interest is determined relative to a reference, generally transcripts from an unaffected individual. The RNA level of the gene of interest is normalized to the level of an internal control gene which is usually a housekeeping gene, the expression of which is assumed to be constant across tissues. $2^{-\Delta\Delta C}{}_T$ approach is an easy method for comparing gene expression between the control and the patient samples after normalizing each group to the internal control gene (Livak and Schmittgen, 2001).

2. PURPOSE

The purpose of this study was to map the gene loci and identify the causative mutations and thus the genes responsible for six inherited disorders, by linkage analysis using the genome scan data for available family members and subsequent exome sequence analysis for one affected member of the family or by candidate gene approach. The disorders are Congenital Disorder of Glycosylation (CDG), Myosin Storage Myopathy (MSM), Desmin Deficiency Myopathy (DDM), Hereditary Head Tumors (HHT), Epilepsy, and Muscular Dystrophy (MD).

3. MATERIALS

3.1. Subjects

Information on study families are summarized below. Written informed consent was obtained for all participants. Boğaziçi University Institutional Review Board for Research with Human Participants approved the study protocol.

3.1.1. Congenital Disorder of Glycosylation (CDG)

A consanguineous family with 4 unaffected daughters and 2 affected sons with initial diagnosis of ataxia was studied to identify the disease locus and the causative mutation (Figure 3.1). The parents were second cousins once removed. The patients were at the age of 38 and 40 and had progressive ataxic gait and visual impairment since early childhood. Clinical information and peripheral blood samples from the parents and the two affected sibs were provided by Bülent Kara, M.D. at Kocaeli University.

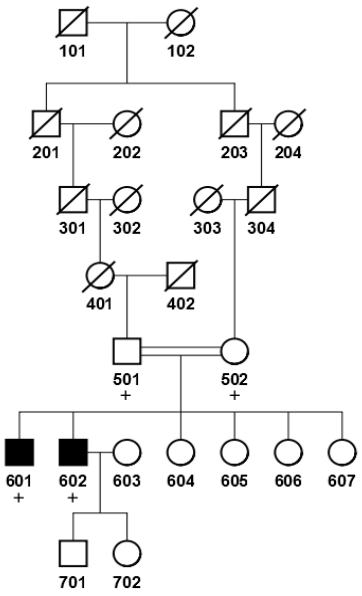

Figure 3.1. Pedigree for CDG family. DNA samples subjected to SNP genotyping are indicated by +.

3.1.2. Myosin Storage Myopathy (MSM)

A family with 2 affected sons afflicted with myosin storage myopathy (MSM) in which the disease locus was previously mapped to 3p22.2-p21.32 using microsatellites (Onengüt *et al.*, 2004) was studied to remap the disease locus using SNP genotype data and subsequently identify the causative mutation (Figure 3.2). The parents denied consanguinity, but because the disease is very rare, parents were assumed third cousins. Peripheral blood samples from the mother, two affected sons and two unaffected daughters were provided by Hatice Karasoy, M.D. at Ege University in İzmir.

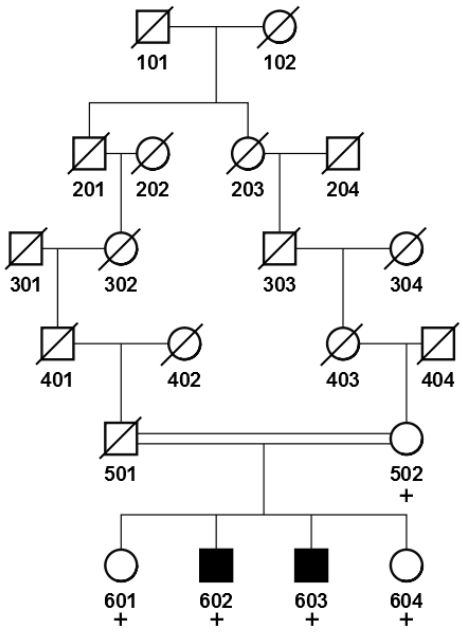

Figure 3.2. Pedigree for MSM family. DNA samples subjected to SNP genotyping are indicated by +.

3.1.3. Desmin Deficiency Myopathy (DDM)

A consanguineous family with two patients having generalized weakness and fatigability was studied (Figure 3.3). Peripheral blood samples from the parents and affected sibs were provided by Hacer Durmuş, M.D. at Istanbul University.

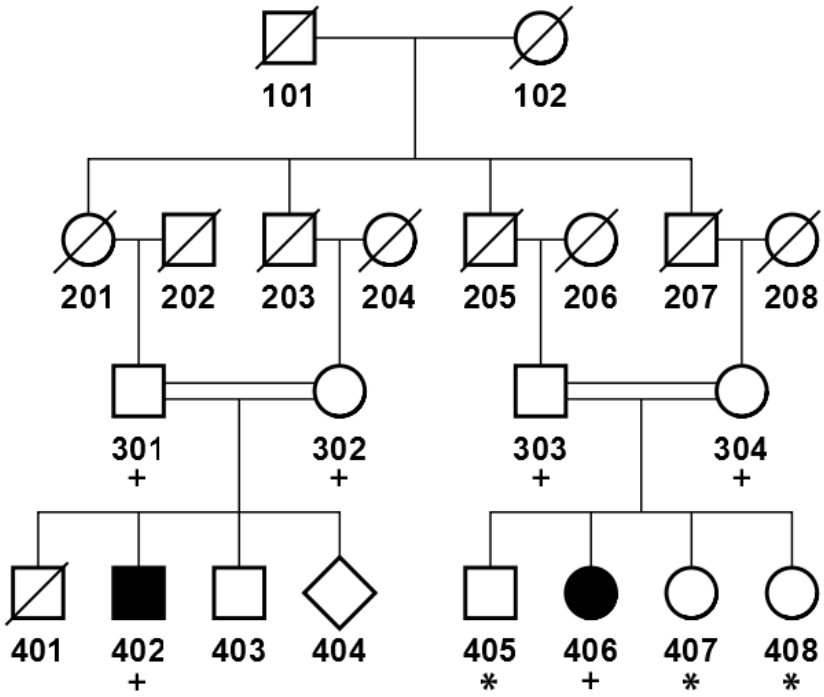

Figure 3.3. Pedigree for DDM family. DNA samples subjected to SNP genotyping are indicated by + and DNA availability is indicated by *.

3.1.4. Hereditary Head Tumors (HHT)

A large consanguineous family afflicted with head tumors in which the disease locus was previously mapped to 16p13.2-p13.12 (Bozoğlu, 2008) was studied with the aim of identifying the causative mutation (Figure 3.4). Tumors were either grade IV astrocytoma (glioblastoma multiforme) or meningioma. Meningioma tumors were in two siblings in the second generation (207 and 209) whereas glioblastoma was in three cousins in the third generation (301, 306 and 310). Unaffected parents 201 and 203 were assumed affected in the statistical analysis, since they need to be carriers of the putative mutation in the assumed dominant inheritance model. Peripheral blood samples from most of the family members had been provided by Dr. Kenan Koç at Kocaeli University. Additionally, DNA samples from 61 patients afflicted with meningioma, 12 patients afflicted with glioblastoma, and 67 DNA samples extracted from tumor tissues had been provided by Hakan Tuna, M.D. at Ankara University.

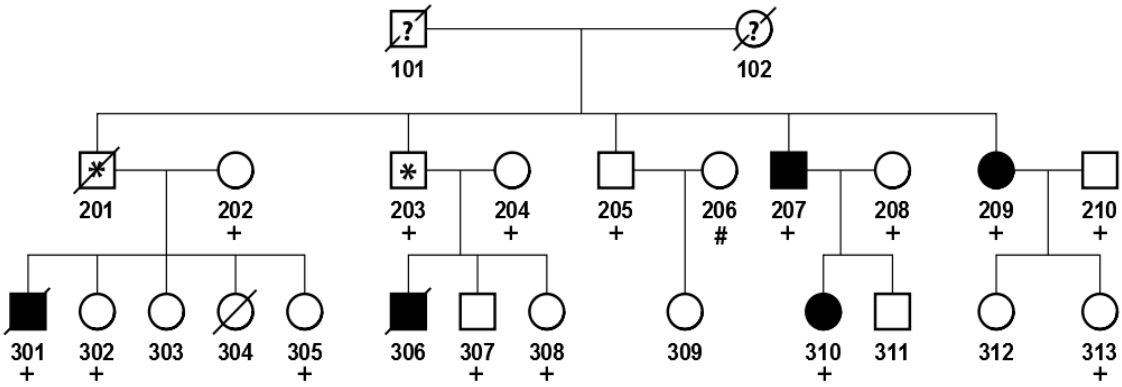

Figure 3.4. Pedigree for HHT family. DNA samples subjected to genome scan with microsatellite markers are indicated by +, and DNA availability is indicated by #. Obligate carriers are shown by *.

3.1.5. Epilepsy

A consanguineous family afflicted with idiopathic generalized epilepsy and mild mental retardation was studied (Figure 3.5). Clinical information and peripheral blood samples from the parents and four affected and five unaffected sibs were provided by Hacer Durmuş, M.D. at Istanbul University.

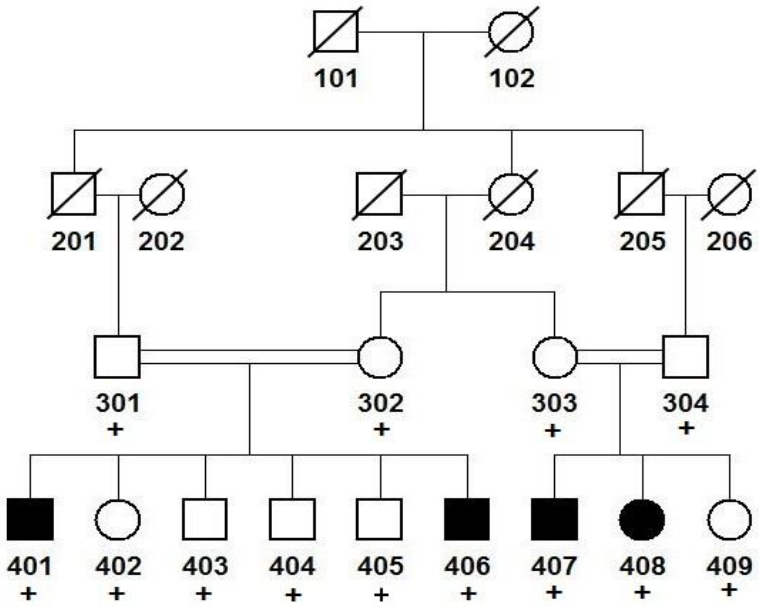

Figure 3.5. Pedigree for epilepsy family. DNA samples subjected to SNP genome scan are indicated by +.

3.1.6. Muscular Dystrophy (MD)

A consanguineous family afflicted with muscular dystrophy was studied (Figure 3.6). Peripheral blood samples from the mother and three affected and three unaffected sibs were provided by Bülent Kara, M.D. at Kocaeli University.

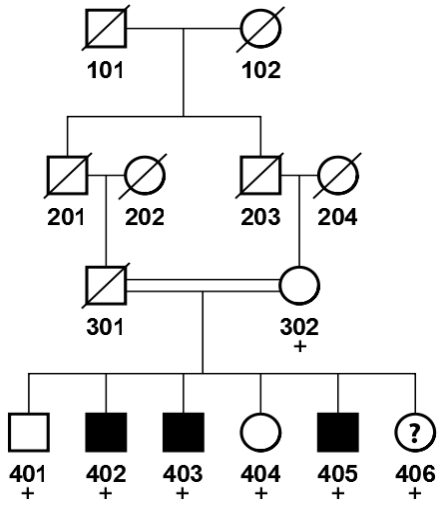

Figure 3.6. Pedigree for MD family. DNA samples subjected to SNP genome scan are indicated by +.

3.2. Chemicals

All solid and liquid chemicals used in this study were purchased from Merck (Germany), Sigma (USA), Riedel de-Häen (Germany), Carlo Erba (Italy) or Biochrom (Germany) unless stated otherwise in the text.

3.3. Buffers and Solutions

3.3.1. DNA Extraction from Peripheral Blood Samples

- Cell Lysis Buffer: 155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM Na_2EDTA (pH 7.4 pH)

- Nucleus Lysis Buffer: 400 mM NaCl, 2 mM Na_2EDTA, 10 mM tris (pH 8.2)
- Proteinase K: 20 mg/ml proteinase K in dH_2O (Sigma, USA)
- Sodium dodecylsulfate (SDS): 10% SDS (w/v) in dH_2O
- Ammonium Acetate: 7.5 M CH_3COONH_4 in dH_2O
- Ethanol: Absolute ethanol
- TE buffer: 1 mM EDTA, 20 mM Tris-HCl (pH 8.0)

3.3.2. Polymerase Chain Reaction (PCR)

- $MgCl_2$: 25 mM $MgCl_2$ (Roche, Germany)
- dNTP: 12.5 mM each of dATP, dCTP, dGTP and dTTP in dH_2O (Roche, Germany and Fermentas, Lithuania)
- DMSO: 6.7% dimethyl sulfoxide
- Betaine: 5 M betaine (Sigma, USA)
- BSA: 50 ng/µl (Sigma, USA)
- 10 X PCR Buffer: 20 mM $MgSO_4$, 100 mM KCl, 200 mM Tris-HCl (pH 8.8), 100 mM $(NH_4)_2SO_4$, 1% triton X-100, 1 mg/ml BSA

3.3.3. Agarose Gel Electrophoresis

- 40% Acrylamide: 40% acrylamide-bisacrylamide (37.5:1) (Stock) in dH_2O
- 8% Acrylamide: 8 or 10% acrylamide-bisacrylamide (37.5:1, non-denaturing) in 0.6 X TBE buffer
- Glycerol: 8% glycerol in gel solution
- APS: 10% ammonium peroxydisulfate
- TEMED: N,N,N,N-tetramethylethylenediamine
- 10X Sample Buffer: 95% formamide, 20 mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue

3.3.4. Single Strand Conformational Polymorphism (SSCP) Analysis

- 40% Acrylamide: 40% acrylamide-bisacrylamide (37.5:1) (Stock) in dH_2O
- 8% Acrylamide: 8 or 10% acrylamide-bisacrylamide (37.5:1, non-denaturing) in 0.6 X TBE buffer
- Glycerol: 8% glycerol in gel solution
- APS: 10% ammonium peroxydisulfate
- TEMED: N,N,N,N-tetramethylethylenediamine
- 10X Sample Buffer: 95% formamide, 20 mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue

3.3.5. Silver Staining

- Staining Buffer: 0.1% $AgNO_3$ in dH_2O
- Developing Buffer: 1.5% NaOH, 0.01% $NaBH_4$ 0.015% formaldehyde in dH_2O

3.4. Enzyme and Kits

Commercial enzyme and kits, manufacturing companies and their usage are listed below.

- Taq DNA polymerase (with supplied PCR buffer): It is used for amplification of DNA templates. It was purchased from Roche (Germany), Kapa Biosystems (USA), Fermantas (Lithuania), or Quanta Biosciences (USA).
- GC-Rich PCR System (Roche, Germany): It is used for amplification of DNA templates that are rich in GC nucleotides.
- Dream Taq PCR Kit (Fermentas, Lithuania): It is used for amplification of DNA templates.
- LightCycler 480 High Resolution Melting Kit (Roche, Germany): It is used for heteroduplex analysis (mutation screening) on LightCycler 480.

- Light Cycler 480 SYBR Green I Master kit (Roche, Germany): It is used for relative quantification of cDNA on LightCycler 480.
- RevertAid First strand cDNA synthesis kit (Fermentas, Lithuania): It is used for reverse transcription-PCR from RNA templates.
- Transcriptor High Fidelity cDNA Synthesis Kit (Roche, Germany): It is used for reverse transcription-PCR from RNA templates.

3.5. Oligonucleotide Primers

Oligonucleotide primers were designed via online program Primer3 and analyzed with Oligo Calc program for secondary structures and self-annealing sites. Primer specificity were analyzed via in-silico PCR on USCS Genome Bioinformatics. The oligonucleotide primers were purchased from Macrogen Inc (South Korea), were received in lyophilized form, and dissolved in one ml dH_2O. Ten μM dilutions were used in all types of reactions.

3.6. DNA Molecular Weight Markers

GeneRuler 1-kb DNA ladder and pUC19 DNA/*Msp*I marker were purchased from Fermentas (Lithuania).

3.7. Equipment

- Computer: Gigabyte X58A-UD5 Motherboard, Intel i7 960 (8X) processor, Kingston 1333 Mhz RAM, CPU @ 3.20GHz, 12328MB Memory, Ubuntu 11.10 Operating System
- Centrifuges: MiniSpin Plus (Eppendorf, Germany), Universal 16R (Hettich, Germany), Allegra X-22R (Beckman Coulter, USA), J2-MC (Beckman Coulter, USA)

- Documentation System: GelDoc Documentation System with Quantity One 4.6.9 Analysis Software (BioRad, USA)
- Electrophoresis Equipment: Horizontal DNA Electrophoresis Gel Box (Bio-Rad, USA), Primo Minicell Horizontal Gel System (Thermo Scientific, USA), DCode Universal Mutation Detection System (Bio-Rad, USA)
- Micropipettes: Pipetman (Gilson, France)
- Power Supplies: Power Pac Model 3000 (Bio-Rad, USA), Fotoforce 250 Electrophoresis Power Supply (Fotodyne, USA), P250A Power Supply (Sigma-Aldrich, USA)
- Spectrophotometers: NanoDrop 1000 (Thermo Scientific, USA)
- Thermal Cyclers: MyCycler (Bio-Rad, USA), PTC-200 (MJ Research, USA), T100 ThermalCycler (Bio-Rad, USA), LightCycler 480 (Roche, Germany)

3.8. Electronic Databases

A list of electronic databases and online tools utilized in this study are given in Table 3.1.

Table 3.1. Electronic databases and online tools used in this study.

Name	**Description**
Databases	
NCBI Genome Resources http://www.ncbi.nlm.nih.gov/projects/genome/guide/human	Provides reference sequences (RefSeqs) and maps of genes for a large collection of genomes.
NCBI Gene http://www.ncbi.nlm.nih.gov/gene	Integrates information on genes in a wide range of species (Nomenclature, RefSeqs, maps, pathways, variations, phenotypes, and links to genome, phenotype, and locus-specific resources).
UCSC Genome Browser http://genome.ucsc.edu	Provides reference sequences and maps of genes for a large collection of genomes.
Ensembl Genome Browser http://www.ensembl.org/index.html	Provides reference sequences and maps of genes for a large collection of genomes.

Table 3.1. Electronic databases and online tools used in this study (cont.).

NCBI UniGene http://www.ncbi.nlm.nih.gov/unigene	Provides expression data in various tissues for a wide collection of genes.
NCBI Homologene http://www.ncbi.nlm.nih.gov/homolo gene	Provides multiple protein alignments, conserved domains and sequence comparison among genes, proteins and species for the same gene.
UniProt (Universal Protein Resource) http://www.uniprot.org	A comprehensive, high-quality and freely accessible resource for protein sequence and functional information.
OMIM (Online Mendelian Inheritance In Man) http://www.omim.org	An online catalog of human genes and genetic disorders and associated phenotypes.
DGV (Database of Genomic Variants) http://dgv.tcag.ca/dgv/app/home	Collects structural variations (genomic alterations that involve segments of DNA that are longer than 50 bp) in the human genome.
HGMD (The Human Gene Mutation Database) http://www.hgmd.org	Collects known gene lesions responsible for human inherited diseases and links to the related articles.
Tools Used in Primer Design	
Primer3 http://bioinfo.ut.ee/primer3-0.4.0	Allows primer design with specified parameters for a specific region.
Oligo Calc (Oligonucleotide Properties Calculator) http://www.basic.northwestern.edu/ biotools/oligocalc.html	An online calculator of oligonucleotide properties
Prediction Tools for Nucleotide Change or Amino Acid Substitutions	
PolyPhen-2 (Polymorphism Phenotyping v2) http://genetics.bwh.harvard.edu/pph2	Predicts possible impact of an amino acid substitution using sequence conservation and structure. Output score ranges from 0 to 1 where 0 means neutral and high numbers close to 1 mean damaging.
MutPred http://mutpred.mutdb.org	Predicts the possible impact of an amino acid substitution as deleterious or neutral. Output score is the probability of deleterious mutation.

Table 3.1. Electronic databases and online tools used in this study (cont.).

SIFT http://sift.jcvi.org	Predicts possible impact of an amino acid substitution based on the degree of conservation in sequence alignments derived from closely related sequences. Output score ranges from 0 to 1, as damaging to neutral.
Human Splicing Finder Version 2.4.1 http://www.umd.be/HSF	Calculates consensus values of potential splice sites and searches for branch points in a given sequence, gene or transcript. It also predicts the effect of a sequence change on pre-mRNA splicing.
Mutation Taster http://www.mutationtaster.org	Predicts possible impact of a nucleotide change for all types of mutations. Output is "benign" or "disease causing" with a probability of prediction. Also amino acid (AA) substitution score, ranging from 0 to 215, is calculated taking into account the physico-chemical characteristics of amino acids. It also predicts possible splice site changes and regulatory feature changes such as histone modification sites, open chromatin or transcription factor binding sites.
Others	
Homozygosity Mapper http://www.homozygositymapper.org	Allows uploading of genotypes and homozygosity mapping of those genotypes with various parameters such as lower and higher limits for homozygous regions and exclusion of homozygous stretches in the control samples.
GeneDistiller 2014 http://www.genedistiller.org	Lists genes for a specific region, phenotype or key word related to disease, based on OMIM reports or PubMed articles to evaluate candidate genes. It also lists genes with expression in defined tissues and with defined cellular localization.
NHLBI Exome Variant Server (ESV) http://evs.gs.washington.edu/EVS	Allows searching for the variants in the 6503 human exome data sequenced for NHLBI Exome Sequencing Project (ESP).

3.9. Bioinformatics Tools

The computer-based programs used for the analysis of genome scan data and exome sequence data in this study are given in Table 3.2.

Table 3.2. Bioinformatics tools used for analyzing genome scan data and exome sequence data.

Name	**Description**
For Genome Scan	
Illumina GenomeStudio Software http://www.illumina.com/informatics/sequencing-microarray-data-analysis/genomestudio.ilmn	A software for visualizing and analyzing data generated on Illumina sequencing and array platforms.
cnvPartition Plug-in v3.2.0 http://support.illumina.com/downloads/cnvpartition_plug-in_v320_for_genomestudio.ilmn	A software library that works with Illumina GenomeStudio data analysis software. It calculates copy numbers with confidence scores and detects CNV regions.
For Analysis of the Next-Generation Sequencing Data	
BWA (Burrows-Wheeler Aligner) http://bio-bwa.sourceforge.net	A software package for mapping low-divergent sequences against a large reference genome, such as the human genome.
SAMtools (Sequence Alignment/Map) http://samtools.sourceforge.net	Provides various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format.
GATK (The Genome Analysis Toolkit) https://www.broadinstitute.org/gatk	A software package for analysis of next-generation sequencing data such as depth of coverage analysis, SNP/indel calling and local realignment.
ANNOVAR http://www.openbioinformatics.org/annovar	A software tool to utilize date-to-date information to functionally annotate genetic variants obtained from next-generation sequencing data. It provides gene based, region-based, and filter-based annotation of variants.

Table 3.2. Bioinformatics tools used for analyzing genome scan data and exome sequence data (cont.).

Picard http://picard.sourceforge.net	Comprises Java-based command-line utilities that manipulate SAM files, and a Java API (SAM-JDK) for creating new programs that read and write SAM files.
BEDTools https://code.google.com/p/bedtools	A suite of utilities for comparing genomic features such as computing coverage and finding feature overlaps in next-generation sequencing data.
BamView http://bamview.sourceforge.net	An interactive display of read alignments in BAM data files.
IGV (Integrative Genomics Viewer) https://www.broadinstitute.org/igv/home	A visualization tool for interactive exploration of large, integrated genomic datasets.

4. METHODS

4.1. DNA Extraction from Peripheral Blood Samples

Genomic DNA was isolated from peripheral blood samples that were saved in sterile tubes containing anticoagulant K_2EDTA. Three ml of cold cell lysis buffer was added per one ml of blood in order to lyse the plasma membranes. After keeping the mixture for 10 minutes (min) at 4^{o}C, the sample was centrifuged at 5000 revolution per minute (rpm) for 10 min at 4^{o}C. The supernatant was discarded, and the leukocyte nuclei in the pellet were suspended in 10 ml of cell lysis buffer using a vortex and centrifuged at 5000 rpm for 10 min at 4^{o}C to wash them. The supernatant was discarded, and 0.3 ml nuclei lysis buffer per one ml of initial blood sample was added to the pellet. The pellet was dissolved by using a vortex, and 5 µl of Proteinase K (20 mg/ml) and 8 µl of 10% SDS were added per one ml of initial blood volume. Then the sample was mixed gently and incubated either for 3 hours (hr) at 56°C or overnight at 37°C to digest nuclear proteins. After incubation, 280 µl of NH_4Ac (9.5 M) was added to the sample per ml of initial blood volume, and the tube was shaken vigorously to salt out proteins. Later the sample was centrifuged at 10,000 rpm for 25 min at room temperature to collect and discard precipitated proteins. The supernatant was transferred into a clean 50 ml tube, and two volumes of ethanol were added. The tube was mixed gently by rotation to precipitate out DNA. DNA was fished out carefully using a micropipette tip and transferred into a clean 1.5 ml tube. After DNA was air-dried for evaporation of residual ethanol, 500 µl of TE buffer for DNA isolated from 10 ml blood sample was added, and DNA was stored at -20°C until use.

4.2. Strategy

This section explains the general strategy we use in gene hunt. The following sections explain the strategies for each family investigated in this study.

4.2.1. Initial Work

Samples are collected, pedigree is constructed, and phenotype is defined. Linkage to known disease genes, if any, are excluded by genotyping the family members with microsatellite markers flanking the known gene. Alternatively, SNP genotyping is performed for one or two patients and the genotypes at the known disease loci are investigated for compatibility with the inheritance pattern in the family. After the exclusion of linkage to the known disease genes, SNP genotyping of affected and informative unaffected family members are performed.

4.2.2. Linkage Analysis

- Programs: Allegro and GeneHunter programs can be used for pedigrees with sizes up to 20 bits. Bit calculation is as follows: B=2n-f where n is a non-founder (a person whose father and mother are specified in the pedigree) and f is a founder (a person whose father and mother are not specified in the pedigree). For large pedigrees, calculations are performed initially using a simplified pedigree (≤ 20 bits), and loci that yield LOD scores above a threshold value are later analyzed by including 1-cM flanking regions and using the actual pedigree.
- Spacing and marker sets: Markers at 0.07, 0.05 or 0.03-cM intervals are selected. If Mb is used instead of cM, a maximum of 0.01 Mb spacing between markers and marker sets of 30, 50 or 100 are used. Smaller marker sets give better results. Optimum parameters are marker spacing of 0.05 or 0.01 and a marker set of 30.
- Penetrance and disease frequency: Assuming incomplete penetrance generally results in finding more candidate regions. In those regions, unaffected relatives may carry the same genotype with the patients. A disease frequency of 0.0001 is assumed for rare diseases and 0.001 or 0.01 for relatively common diseases but keeping in mind that those diseases are genetically heterogeneous.
- Evaluation of Linkage Analysis Results: A threshold LOD score value is set according to the results. All loci yielding scores above that value are listed, and genotypes at each of them are investigated on formatted Excel sheets to assess whether the patients share the same homozygous genotype for a recessive

disease. In a dominant disease, haplotype analysis is applied to assess whether a haplotype segregates with the disease.

4.2.3. Detailed Linkage Analysis (Fine Mapping)

At a locus that yielded maximal LOD scores, linkage analysis is performed using all markers (no space between markers) at the locus plus 1-cM flanking regions. For example, if a locus between 145 and 167 cM on chromosome 2 yielded a maximal LOD score of 2.9 and the haplotypes are consistent with the inheritance pattern, a detailed linkage analysis is performed using all markers from 144 to 168 cM.

4.2.4. Homozygosity Mapping (Optional)

Homozygosity Mapper is an online tool that allows the uploading of genotypes and analyzing those genotypes with various parameters. The uploaded data is accessible at any time via the internet. In contrast to linkage analysis, Homozygosity Mapper can more effectively detect small regions of shared homozygosity that can be missed in the linkage analysis.

4.2.5. Genotype and Haplotype Inspections

The genotypes are formatted on Excel and evaluated at the loci that yielded maximal LOD scores. Also, haplotypes are constructed using easyLINKAGE in “haplotyping on” mode. Additionally, SNPs having 0 results (no call) in patients but read in other family members are evaluated for possible deletions. Maximal homozygosity regions, the delineating SNPs and corresponding genomic positions can be identified.

4.2.6. Deletion and Duplication Analysis

Deletion and duplication analysis is performed in the candidate regions using GenomeStudio CNVpartition program to detect deleted or duplicated regions. If any duplicated or deleted region is found, Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home) is referred to see whether any overlapping deleted or duplicated region is reported.

4.2.7. Searching for Candidate Genes

- Morbid/Disease Maps: Morbid/Disease map is investigated for the candidate homozygous regions for any clinical phenotype that resembles the disease in the study family. Accession is: NCBI → MapViewer → Chr → Region → Cytogenetic maps → Morbid.
- GeneDistiller: Online GeneDistiller 2014 (http://www.genedistiller.org/) program is used to search for candidate genes at candidate loci. A specific phenotype or key word related to the disease of interest can be selected to limit the number of candidate genes. GeneDistiller2 facilitates searching for genes possibly related to those phenotypes or key words in PubMed articles or OMIM reports.

4.2.8. Exome Sequence Analysis

- Annotation: BWA and GATK programs are used.
- Filtering the candidate variants: All variants at the candidate loci are listed, and non-exonic variants (except for splicing variants), variants with alternative depth <60% of total depth, and variants reported in dbSNP or EVS with MAF >0.01 (for rare diseases) are filtered out. For a highly sensitive analysis, heterozygous variants are kept until further analyses, such as visualization on BamView or IGV and Sanger sequencing. dbSNP or EVS are referred for frequency. Allele frequencies of some variants can be found also in 1000 Genomes (Ensembl →

Select Human → type SNP ID → Choose population genetics). The filtered variants are searched in other in-laboratory exome data using SNPchecker program, and those found in at least two other in-laboratory exome files can be eliminated. Variants with MAF >0.01 are eliminated for rare diseases. The filtered variants are inspected using BamView or IGV programs, and those that are in the mid-regions of sequence reads are considered more reliable whereas those that are close to the termini of the sequence reads are possibly false calls. Additionally, candidate regions are searched for variants that have a MAF >0.98 in the list constructed in our laboratory and then searched in the relevant exome data using annotated exome file or IGV to find whether there was a variant for which the rare allele has been designated as the reference allele. Such a variant would not be called in exome sequence analysis.

- Detecting unread exons or exons with low coverage: Coverage for targeted exons is computed using BedTools. Candidate loci are checked for uncovered or partially covered exons. For each candidate gene in the candidate region, the coverage of the gene of interest is investigated. Generally there are always some genes that are not fully covered in exome sequencing, and such regions in candidate genes need to be Sanger sequenced. Also, for genes with multiple isoforms, exome sequencing generally covers the exons of the largest transcript isoform. The exons that are not in that isoform but are specific to some other isoforms need to be Sanger sequenced.

4.2.9. Analysis of Filtered Variants

- Expression: To investigate whether a gene of interest is expressed in relevant tissues, "NCBI → UniGene → Gene of interest → EST profiles" and "genome.ucsc → GeneSorter → Gene of interest → Expression profiles" are used.

- Effect of the variant on the protein: The strategy followed for each type of mutation is explained below.

For nonsynonymous variants**,** online tools SIFT (http://sift.jcvi.org/), Mutation Taster (http://www.mutationtaster.org/), MutPred (http://mutpred.mutdb.org/), and PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) are used to predict whether the amino acid substitution is damaging or benign. Multiple sequence alignment is investigated using PolyPhen2 (Multiple Sequence Alignment section) or NCBI → Homologene → Name of gene of interest → Show Multiple Alignment (in Protein Alignments section).

For splicing variants, online tools Mutation Taster (http://www.mutationtaster.org/) and Human Splicing Variants (http://www.umd.be/HSF/) are used to investigate the possible effect of the variant on the processing of the primary script to messenger RNA.

For synonymous variants, codon usage database (http://www.kazusa.or.jp/codon/) is used to investigate whether the nucleotide substitution converts the codon to a rare codon. Also, whether the variant is in a regulatory region is investigated via genome.ucsc using ENCODE tract.

For deletions and insertions, Mutation Taster is used to assess the effect of a deletion or insertion on the protein. Mutalyzer also can be used for the same purpose.

4.2.10. Validation of the Variants

Primers are designed to amplify the site of the candidate variant, and Sanger sequencing is applied to validate the variants. PCR products up to 250 bp are more useful since they can be used later for HRM analysis.

4.2.11. Population Screening

Population screening for a mutation is performed by either HRM or SSCP. A minimum of 108 samples should be tested in order to achieve 80% power to rule out that the variant is a common polymorphism in the population (Collins and Schwartz, 2002).

- HRM Analysis: PCR products up to 300 bp can be used effectively for HRM analysis. DNA concentration should be 20 ng/ul. DNA sample to be tested is mixed with final 20% control sample DNA (20 ng/ul) in order to distinguish the wildtype homozygous and mutant homozygous genotypes. C_T values should be under 30 for reliable results. Generally, normalization and sensitivity are adjusted automatically by the program. However, manually changing normalization and sensitivity can be useful when homozygous and heterozygous samples cannot be distinguished.
- SSCP Analysis: In the optimization of SSCP analysis, gels with or without glycerol are used and electrophoresis is performed at various temperatures.

4.3. Identification of Disease Loci

4.3.1. Genome Scan

Genotype data generated by whole-genome single nucleotide polymorphism (SNP) scans for participants and the physical and genetic maps of the markers were provided by Yale Center for Genome Analysis (YCGA, USA). Illumina Genome Studio v.1.02 Genotyping Module was used to handle the data for linkage analysis and deletion/duplication analysis. Table 4.1 summarizes the genome scan chips (all Illumina) used in this thesis.

Table 4.1. SNP genome scan chips, their contents, and families genotyped.

Family	**Chip Name**	**Total SNP**	**CNV (approx.)**	**Marker spacing (Kb) Mean/Median**
CDG	370 Duo	370,405	74,000	7.7 / 5.0
MSM, Desmin, MD, Epilepsy*	HumanOmniExpress	730,525	0	4.0 / 2.1
Epilepsy	HumanOmni1-Quad	1,140,419	60,000	2.4 / 1.2

* Only for unaffected individual 405.

Genome scan for Hereditary Head Tumors (HHT) family members had been performed using Marshfield Screening Set 16 that contains 402 microsatellite markers with an average density of 10 cM (Weber and Broman, 2001) at the National Heart, Lung and Blood Institute (NHLBI) Mammalian Genotyping Service as a service grant (Contact Number HV48141).

4.3.2. Linkage Analysis and Haplotype Analysis

easyLINKAGE package version 5.08 was used to calculate parametric multipoint logarithm of the odds (LOD) scores. Homozygosity mapping was performed by online Homozygosity Mapper (http://mitonet.charite.de/HomozygosityMapper/) or Homozygosity Comparison in Excel (HCiE) developed in our laboratory (Cetinkaya, 2010). In HCiE, each of the three genotypes (AA, AB and BB) is colored differently. The markers were ordered in genomic position on an MS Excel macro, allowing the visualization of the genotypes and a comparison of the genotypes in the patients and other family members in the shared homozygosity regions. Homozygosity mapping was used for the five recessive diseases but not for HHT, where dominant inheritance was assumed.

Homozygosity Mapper analyzes the homozygous regions in the genotypes that are uploaded to the server, and the detected homozygous regions are analyzed. Generally the parents or the unaffected siblings are specified as controls whereas the patients are specified as affected. In autosomal recessive disorders, genetic homozygosity is searched and regions homozygous also in controls are not excluded in order not to overlook the non-informative regions. Also, it is best that block length for homozygous stretches is not restricted. Homozygosity Mapper was used in epilepsy family to find shared homozygosity regions in the patients. Analysis was performed in two ways. In the first analysis, all patients were considered as affected. In the second analysis, all patients except 401 were considered as affected since he has an atypical phenotype.

Haplotype construction allows the investigation of haplotype segregation in the family at specified regions to assess whether the shared haplotype is identical by descent

(IBD) or identical in state (IBS). Haplotype construction was performed via Allegro or GeneHunter in the easyLINKAGE package. Each candidate region was analyzed together with approximately 1-cM flanking sequences via one of the mentioned easyLINKAGE package programs to investigate haplotype/genotype sharing among patients. Haplotypes were constructed by using Allegro in the easyLINKAGE package after investigating genotypes by HCiE; an example of it is presented for MD family in the results section.

Multipoint linkage analysis was performed with Allegro v1.2.c for families CDG, MSM, DDM, Epilepsy and MD and with Simwalk v2.91 for HHT. SNPs are selected at 0.07 cM intervals and used in sets of 100 for CDG and DDM families, selected at 0.1 cM intervals and used in sets of 100 for MSM family, and selected at 0.05 cM intervals and used in sets of 30 for MD family. Autosomal recessive inheritance, full penetrance and a disease frequency of 0.001 were assumed for CDG, MSM, DDM and MD families. For MD family additionally recessive X-linked inheritance was assumed. An autosomal dominant inheritance model, either full or 90% penetrance, and a disease allele frequency of 0.001 were assumed for HHT family. Genotypes at candidate loci were investigated by HCiE for all the families except HHT. Haplotypes were constructed by SimWalk for investigating candidate loci.

In Epilepsy family, multipoint parametric linkage analysis was performed with different parameters and different pedigree models due to the phenotypic variation among the patients. Table 4.2 summarizes those different parameters and the different pedigree models assumed.

In the first analysis, linkage analysis was performed using the actual pedigree, assuming recessive inheritance with 70% penetrance. Since the size of the pedigree exceeded the capacity of the program, LOD scores were calculated separately for the two branches of the family, and cumulative LOD scores were obtained.

In the second analysis, only the patients and their parents were included since the full size of the family was too large for the capacity of the linkage program. Autosomal recessive inheritance, full penetrance and a disease frequency of 0.01 were assumed.

The third was X-linked linkage analysis, and only the three male patients were included since the female patient has a milder phenotype that could be due to skewed X inactivation. In calculating multipoint LOD scores, X-linked recessive inheritance with full penetrance was assumed and among the females only the mothers were included.

Considering that the female patient had a milder phenotype, only the three male patients and the parents were included in the linkage analysis in calculating multipoint LOD scores in an autosomal recessive model with 70% penetrance.

The clinician hypothesized that patient 401, with an atypical phenotype as compared to the other patients in the family, could be having another disease; therefore, linkage analysis that did not include him was also performed. Assuming autosomal recessive inheritance with full penetrance, the genotype data of the mothers, fathers, remaining 3 patients (406, 407, and 408) and unaffected brother 403 were used.

Since skewed X-inactivation could explain the milder phenotype of the female patient and atypical patient 401 could be having another disease, X-linked multipoint linkage analysis was performed by not including those patients but including the mothers, fathers, remaining 2 patients (406 and 407) and unaffected brother 403.

4.4. Deletion and Duplication Analysis

cnvPartition is one of the several algorithms developed by Illumina and can be used to detect deleted or duplicated regions in microarray data. cnvPartition can find regions that are aberrant in copy number by using two outputs from Illumina Genome Studio v.1.02; one is logR ratios (LRR) and the other is B allele frequency (BAF). LRR is logged ratio of observed probe intensity to expected intensity, and deviations from zero are indicators of aberrant copy number since a ratio of observed to expected probe intensity is expected to be one for a copy number of one (log1=0). B is the minor allele for a SNP, and BAF is the proportion of major allele to minor allele for a SNP. Thus, 0.0 is the expected BAF value for a SNP homozygous for major allele, 0.5 for a heterozygous SNP, and 1.0 for a SNP homozygous for minor allele. Deviations from 0,

0.5 and 1 for BAF are indicators of aberrant copy number. cnvPartition calculates copy numbers with confidence scores based on LRR and BAF matrices and finds deleted/duplicated regions.

Table 4.2. Linkage analyses for epilepsy family. Different parameters and different pedigree models were used due to the phenotypic variations among the patients.

Analysis Description	**Pedigree/Used genotypes**	**Inheritance Model**	**Penetrance**	**Intervals between Markers (cM)**	**Marker Sets**
Actual pedigree*	All family members	Autosomal recessive	70%	0	100
Only patients and parents	Only patients and parents	Autosomal recessive	100%	0	100
Disregarding female patient	Three male patients and parents	X-linked recessive	100%	0.07	100
Disregarding female patient	Three male patients and parents	Autosomal recessive	70%	0.1	100
Disregarding atypical Patient 401	Parents, patients 406, 407 and 408 and unaffected brother 403	Autosomal recessive	100%	0.07	50
Disregarding atypical Patient 401	Parents, male patients 406 and 407, and unaffected brother 403	X-linked recessive	100%	0.07	50

* Since the size of the pedigree exceeded the capacity of the program, LOD scores were calculated separately for the two branches of the family, and cumulative LOD scores were obtained.

4.5. Exome Sequencing

4.5.1. Exome Capture and Sequencing

Targeted exome capture and sequencing of captured exons were performed at BGI (China) for MD family and at Macrogen Inc (South Korea) for the other four families as a commercial service. Four affected individuals, one each from CDG, Desmin, HHT, and Epilepsy families, were subjected to exome capture with Illumina TruSeq Capture kit (Illumina, USA). DM sample was subjected to exome capture with an Agilent SureSelect Target Enrichment kit (Agilent, USA). Features of TruSeq Exome Capture kit and SureSelect Target Enrichment kit are given in Table 4.3 and Table 4.4, respectively. The captured fragments were later sequenced using Ilumina HiSeq 2000 (Illumina, USA).

Table 4.3. Features of Illumina TruSeq Exome Capture kit.

Exome Capture Statistics	**Value**
Target region size	62 Mb
Number of target genes	20,794
Number of target exons	201,121
Number of probes	340,427
% of RefSeq (refGene) coding exons covered	96.4%
% of CCDS coding exons covered (31.3Mb)	97.2%
Recommended library insert size	300-400 bp

Briefly, exome sequencing method has five main steps: sample preparation, exome enrichment, clustering, sequencing and analysis. In sample preparation, 5-10 µg of genomic DNA is fragmented by nebulization, and 300-400 bp double-stranded DNA fragments are obtained by converting overhangs to blunt ends. Then, an “A” is ligated to the 3’ ends of fragmented DNA that would in turn allow the ligation of an adaptor containing a complementary 5’ overhang “T” nucleotide. After that, DNA is denatured, and strands with ligated adaptors are hybridized with biotinylated capture probes of targeted regions, and unbound strands are washed off.

Table 4.4. Features of Agilent SureSelect Target Enrichment kit.

Exome Capture Statistics	**Value**
Initial bases on target	51,543,125
Initial bases near target	73,252,334
Initial bases on or near target	124,795,459
Total effective reads	49,768,780
Total effective yield (Mb)	4857.71
Average read length (bp)	97.61
Effective sequences on target (Mb)	2759.37
Effective sequences near target (Mb)	1113.96
Effective sequences on or near target (Mb)	3873.33
Number of reads uniquely mapped to target	33,565,876
Number of reads uniquely mapped to genome	47,709,226
Fraction of effective bases on target	56.8%
Fraction of uniquely mapped on target	70.4%
Fraction of effective bases on or near target	79.7%
Average sequencing depth on target	53.54

Streptavidin beads, which can bind to biotinylated probes, are used to capture the biotinylated probes for the targeted regions of interest. In total two hybridization steps and three washing steps are performed in order to enrich the targeted regions and remove non-specifically-bound strands from the beads, respectively. After the hybridization of strands onto the flow cell, which is a solid surface containing forward and reverse primers for the regions of interest, a “bridged” amplification reaction takes place on the flow cell. The amplified fragments are later sequenced on a sequencer.

4.5.2. Analysis of Exome Sequence Results

Bioinformatics analysis results of the exome sequencing data and an annotated list of variants were provided by Macrogen Inc (South Korea) in MS Excel format as part of the exome sequencing service. The company generates this file using BWA and SAMtools with standard parameters.

Bioinformatics analysis of exome sequencing data, which includes alignment, SNP calling, coverage computations and visualization of the reads, was performed also in our laboratory with more relaxed parameters to obtain better results. The raw data for exome sequencing were obtained from Macrogen Inc (South Korea) in paired end ".fastq" files. Human reference genome sequence (assembly GRCh37/hg19) was downloaded from UCSC Genome Bioinformatics Site, and all chromosomes were indexed. Alignment of the paired end to the reference genome was performed using BWA which generates a final alignment in the SAM (Sequence Alignment/Map) format. SAMtools was used in order to convert the SAM format to BAM (Binary Alignment/Map) format, and subsequently to sort and index the resulting file. Finally, nucleotides differing from the reference genome sequence were listed (SNP calling) with SAMtools. The list of variants was annotated with ANNOVAR. Annotation of variants includes determination of chromosomal location, genomic position, dbSNP ID (if any) and frequency (if any), the type of region of the variant (exonic, intronic, UTR, intergenic or non-coding RNA), the type of exonic variants (nonsynonymous, synonymous, stopgain, stoploss, frameshift or non-frameshift indel or splicing), and whether an intronic variant affects splicing. BamView and IGV were used to visualize the alignment of the reads. Table 4.5 shows command lines used in the bioinformatics analysis of exome sequence data.

Table 4.5. Command lines and their functions in bioinformatics analysis of exome sequence data.

Command	**Purpose**
cat *.fa > all_chr.fasta	Concatenates all chromosomes of the reference genome
bwa index –a bwtsw all_chr.fasta	Indexes the reference genome
bwa aln –n 0.01 –t 8 all_chr.fasta file1.fastq > file1.sai	Aligns paired-end reads to reference genome (raw data is presented as two fastq files and aligning is performed for each .fastq file separately)

Table 4.5. Command lines and their functions in bioinformatics analysis of exome sequence data (cont.).

bwa sampe all_chr.fasta file1.sai file2.sai file1.fastq file2.fastq > file.sam	Generates alignment in the SAM (Sequence Alignment/Map) format
samtools view –bS file.sam > file.bam	Converts SAM format to BAM (Binary Alignment/Map) format
samtools sort file.bam	Sorts the BAM file
samtools index file.sorted.bam	Indexes the BAM file
samtools pileup –vcf all_chr.fasta file.sorted.bam > file.pileup.txt	Used for SNP calling
convert2annovarl.pl file.pileup.txt –outfile file_annovar.pileup.txt	Converts the file to a format compatible with ANNOVAR
annotate_variation.pl –buildver hg19 file_annovar.pileup.txt humandb	Performs region based annotation
annotate_variation.pl –filter –dbtype snp131 file_annovar.pileup.txt	Performs filter based annotation
java –mx152m –jar BamView_v1.1.8.jar –a file.sorted.bam –r all_chr.fasta	Used for visualization of the alignment
bamToBed –i file.bam > file.bed coverageBed –a file.bed –b targetedRegions.bed > file_coverage.txt	Computes the coverage for each exon included in the target capture using BEDTools

4.6. Candidate Genes and Mutation Screening

Exome sequence data was filtered and analyzed as explained in Section 1.10 step 8 to detect candidate variants that might possible underlie the pathology in the family. The sites of the candidate variants were amplified by polymerase chain reaction (PCR), and PCR products were sequenced by Sanger method for validation. A validated variant was later screened in the family to investigate segregation with the trait. Additionally, 120 individuals randomly selected from the population were screened in order to achieve a power of 80% to detect a normal sequence variant with a frequency of 0.01 (Collins and Schwartz, 2002).

MYH7 was chosen as the best candidate gene for MSM family, and since mutations residing in the last four exons (exons 37, 38, 39 and 40) of the gene were reported to cause MSM whereas mutations in the other regions cause other muscle myopathies (Tajsharghi and Oldfors, 2013), those four exons together with at least 35 bp sequences in the flanking introns were amplified in patient 602 and subjected to Sanger sequencing.

In the case of MD family, *LMNA* residing at the largest candidate locus was the best candidate gene. However, exome sequence analysis did not reveal any candidate variants in the gene. For this reason, all coding exons of the gene were amplified and sequenced to detect any missed variant.

4.6.1. PCR Amplifications

PCR was performed to amplify specific DNA sequences. Regions of 100-250 bp containing the strong candidate variants listed in the exome sequencing output were amplified in relevant individuals by PCR and subjected to Sanger sequencing for validation. Also, exons of candidate genes were amplified in relevant individuals by PCR and subjected to Sanger sequencing to search for mutations. In PCR, 1X PCR buffer, 0.2 µM of each dNTP, 0.4-0.8 µM of each primer, 50-70 ng of genomic DNA as template, and 0.2 U Taq DNA polymerase were used in a total volume of 25 µl reaction by adding sufficient distilled water.

- Validation of Exome Sequence Variants: In an exome sequence analysis, errors can occur due to inaccuracies in sequencing of the reads, alignment of the reads to the reference genome and variant calling. For these reasons, all candidate variants need to be validated by Sanger sequencing. Table 4.6 presents the primers and optimum conditions for PCR to prepare samples for sequencing. Assembly GRCh37/hg19 was used throughout this thesis.

- Amplification of the exons of candidate genes: The last four exons of *MYH7* and all exons of the *LMNA* longest transcript were amplified by PCR and subjected to Sanger sequencing. Primers and optimal conditions for amplification are given in Table 4.7.

Table 4.6. PCR conditions to amplify candidate exome sequence variants. Primer information, product sizes, amplified regions and annealing temperatures are presented.

Primer Name	**Primer Sequence (5' → 3')**	**P. Size (bp)**	**Amplified Region (bp)**	**Anneal. T (°C)**
Congenital Disorder of Glycosylation				
SRD5A3	F: GGAGGCCGAGCACTCG R: GCAGCCCGGGAGCAG	115	chr4:56212518 -56212632	66→61 /-0.5
Desmin				
DES c344	F: GAGCTGCTGGACTTCTCACTG R: CCTTGAGCCGGTTCACTTC	174	chr2:220283443 -220283617	56
Hereditary Head Tumors				
HHT CARHSP1	F: CAAAGTCTCCCACAAGCACA R: GGCGACGAGGTCACCTATAA	180	chr16:8948976 -8949155	58
HHT LRRCC1	F: GATCAACAAGAGGATCACCTTA R: TTCAACAGGCTATGTGACCG	248	chr8:86041522 -86041769	61→56 /-0.5
HHT SLC7A13_2	F: CCATAATGAACCCGTGAAAAA R: CCTCATTATTTAGCAACCTTCTGA	221	chr8:87229768 -87229988	61→56 /-0.5
HHT RGS22	F: AGATGGAGGCTGCAGTGAG R: ACCCAGTCTACCACCTCCT	229	chr8:101083444 -101083672	66→60 /-0.5
METTL22 c1088	F: AACCCTTCCCATTGATGATG R: GCTCGTAAACCAGGAGCTGT	207	chr16:8738362 -8738568	62→56 /-0.5
Epilepsy				
ESX1	F: GTGACCTGGGTGGCAGAG R: CACCTATGCCACCCATGG	336	chrX:103494992 -103495327	57
ESX1 SET2	F: CTGGCAGGAAAGAACTTTGG R: TCTGGATCCTGCTTTGTGTG	582	chrX:103494865 -103495446	60→54 /-0.5

P. Size: product size; Anneal. T: annealing temperature

LMNA has three transcript isoforms with different first exons, second exons and last exons as compared to the longest transcript. Those regions were amplified by PCR

and subjected to Sanger sequencing. They are given in Table 4.8, and primer information for amplification of those regions is presented in Table 4.9.

Table 4.7. Primers and PCR conditions for exons of candidate genes *MYH7* and *LMNA*. Sequences, PCR product sizes and PCR conditions for primers to amplify the exons are presented.

Region (exon)	**Primer Name**	**Primer Sequence (5' → 3')**	**P. Size (bp)**	**Region (bp)**	**Anneal. T (°C)**
***MYH7* for Myosin Storage Myopathy Family** (Chromosome 14)					
Exon 37	MYH7 Ex35	F: TCCTCAGCTGGTTGTCACTG R: GATGTAAGTCCCCCACTCCA	454	23884139-23884592	57
Exons 38 and 39	MYH7 Ex36+37	F: ATGTGGCTCAAGTGTGTGGA R: CATCTTCACCCCCTGCCTA	491	23882855-23883345	57
Exon40	MYH7 Ex38	F: GGCAGTGAAGAAGAGTCTGGA R: CAGACCCCTCTCACCTTTGT	250	23881919-23882168	57
***LMNA* for Muscular Dystrophy Family** (Chromosome 1)					
Exon 1 (5' UTR included)	LMNA 5UTR+E1 Set2	F: CCCTCTAGCCCAGAAGGTCT R: CCCTCTCACTCCCTTCCTG	932	156084226-156085157	63→56 /-0.5
Exon 2	LMNA E2	F: CTGGCACTGTCTAGGCACAC R: AGGACAGGTGAATGGCTCTG	384	156100347-156100694	62→55 /-0.5
Exons 3 and 4	LMNA E3+E4	F: CTCCTTCCCTGGACCTGTTT R: CTGATCCCCAGAAGGCATAG	784	156104070-156104853	63→53 /-0.5
Exon 5	LMNA E5	F: CTATGCCTTCTGGGGATCAG R: CTTTTCATCCCTCTCCTCCTC	392	156104834-156105225	63→56 /-0.5
Exons 6 and 7	LMNA E6+E7	F: CTCTGGGGAAGCTCTGATTG R: CTCTGAGGGCAAGGATGTTC	764	156105566-156106329	63→53 /-0.5
Exons 8 and 9	LMNA E8+E9	F: GATGGAAGGAGAGGCCTCAAT R: CAGCTGGCTCCGATGTTG	495	156106598-156107092	62→52 /-0.5
Exon 10	LMNA E10	F: CAGAGGACAGAGTAAGCAGCA R: GGGTTCCCTGTTCAAGGTAT	400	156107312-156107711	62→52 /-0.5
Exons 11 and 12	LMNA E11+E12	F: CTTGGGCACAGAACCACAC R: GTGGGCATGAGGTGAGGA	807	156108155-156108961	65→55 /-1

P. Size: product size; Anneal. T: annealing temperature

Table 4.8. Regions that are specific to *LMNA* transcripts other than the largest one.

Transcript no (Ensembl)	**Region**	**Genomic position (bp)**
ENST00000448611	First exon	156,095,951 - 156,096,014
	Last exon	156,109,561 - 156,109,804
ENST00000473598	First exon	156,096,346 - 156,096,442
	Second exon	156,099,618 - 156,099,699
ENST00000368297	First exon	156,095,980 - 156,096,014
	Second exon	156,096,527 - 156,096,706

Table 4.9. PCR conditions for *LMNA* transcript isoforms. The gene is on chromosome 1.

Region of transcript isoform	**Primer Name**	**Primer Sequence (5' → 3')**	**P. Size (bp)**	**Amplified Region (bp)**	**Anneal. T (^{o}C)**
The first exon for ENST00000448611	LMNA_611_5U-E1	F: ATGCCCAGGAGGATACAGG R: AAGACAAAGGGGAGGTGGAG	247	156095845-156096091	62→55 / -0.5
The first exon for ENST00000473598, and the first and second exon for ENST00000368297	LMNA_598_5U1+297_5U-E1	F: AGATGCCCAGGAGGATACAG R: TCAGGTGACTCATGCTCAGG	926	156095843-156096770	64→56 / -1
The second exon for ENST00000473598	LMNA_598_5U2-E1	F: AATAATCAGGGGCCCAGAAA R: GGCACAATATTGACCCTTTCC	299	156099526-156099824	66→56 / -1
The last exon for ENST00000448611	LMNA_611_lastE-3U	F: CTACACCTGGCTGAGGTTCC R: GTACACACAGCGACGTCCAC	497	156109441-156109919	62→56 / -0.5

P. Size: product size; Anneal. T: annealing temperature

4.6.2. Analysis of PCR products

Agarose gels were used to assess the purity of the PCR products and the amplification efficiency for the PCR reactions. One µl 6X loading dye and 5 µl PCR product were mixed and loaded on a 2% agarose gel that contained 10 mg/ml ethidium bromide. Either pUC restriction enzyme fragments or a 1-kb marker ladder was used to assess PCR product size. The gels were subjected to electrophoresis in 0.5X TBE at 100 volts for 25 minutes, and the fragments were visualized under UV light using BIORAD Universal Hood II and Quantity One 4.6.9 software.

4.6.3. DNA Sequence Analysis

PCR amplifications were performed in 50 µl reactions for all regions to be analyzed with Sanger sequencing. A total volume of 50-150 µl per sample, depending on the PCR efficiency, was sent to Macrogen Inc (South Korea) for purification and subsequent Sanger sequencing. Results were obtained in ".ab1" format, and ChromasLite was used to visualize sequences. Sequence of the amplified region was obtained also in ".text" format, and BLAT in USCS Genome Browser was used to align the sequence to the human reference genome.

4.6.4. Single Strand Conformational Polymorphism Analysis

Single Strand Conformational Polymorphism (SSCP) analysis is a method used to detect sequence differences in DNA fragments. It relies on the phenomenon that different single strands of DNA migrate differently on a polyacrylamide gel, depending on the folding conformation. Whether the individual is homozygous wildtype, homozygous mutant or heterozygous for a variant can be assessed by the migration pattern. Therefore, SSCP can be used to test for a variant the family members or control population samples.

In the present study, SSCP analysis was used to screen control samples for the variants identified in *SRD5A3* and *DES*, and to screen the meningioma, glioblastoma and tumor tissue samples for the variant identified in *CARHSP1* exon 1. Moreover, the variant in *RGS22* was screened in HHT family by SSCP analysis to investigate the segregation. The sequences of the primers used for amplification, PCR conditions and SSCP conditions are summarized in Table 4.10.

Population control samples and CDG family members from whom DNA was available were screened for the mutation detected in *SRD5A3* by SSCP analysis. Population control samples and desmin family members from whom DNA was available were screened for the identified *DES* mutation by SSCP analysis.

Table 4.10. PCR and SSCP conditions for population screening.

Primer Name	**Primer Sequence (5' → 3')**	**P. Size (bp)**	**Region (bp)**	**Anneal. T (^{o}C)**	**SSCP Conditions**
Congenital Disorder of Glycosylation					
SRD5A3	F: GGAGGCCGAGCACTCG R: GCAGCCCGGGAGCAG	115	chr4:56212518-56212632	66→61 /-0.5	Gels without Gly, 4^{o}C, 15 W, 3 h
Desmin					
DES_HRM_F and DES_c.344_R	F: GCGGTGAACCAGGAGTTTC R: CCTTGAGCCGGTTCACTTC	148	chr2:220283470-220283617	60→54 /-0.5	Gels without Gly, 25^{o}C, 12W, 3 h
Hereditary Head Tumors					
CARHSP1_Ex1	F: CCTGCTCCAGACTCACGC R: CAGCCTCTTCTCTTTCCAGGTC	200	chr16:8953012-8953211	57	Gels with Gly, 4^{o}C, 11 W, 5 h
RGS22_Set2	F: TGGTGACAGAGTGAGATCCTG R: TCTCCCTGGATCGTGAAAA	206	chr8:101083490-101083695	58→53 /-0.5	Gels with Gly, 4^{o}C, 11 W, 7 h

P. Size: product size; Anneal. T: annealing temperature, Gly: Glycerol, W: watt, h: hour

4.6.5. High Resolution Melting (HRM) Curve Analysis

Control samples was screened for *MYH7* c.5458C>T transition with high resolution melting curve analysis (HRM) using a Roche HRM Master kit in Roche Light

Cycler 480. Additionally, the cancer samples were screened for exons 2 and 3 of *CARHSP1* by HRM analysis. The samples showing aberrant melting patterns were subjected to Sanger sequencing. Primers used in HRM analysis and HRM conditions are given in Table 4.11.

Table 4.11. Conditions for HRM curve analysis.

Primer Name	**Primer Sequence (5' → 3')**	**P. Size (bp)**	**Region (bp)**	**Anneal. T (°C)**	**HRM Conditions**	
					$MgCl_2$ (mM)	**T Range (°C)**
Myosin Storage Myopathy						
MYH7_Ex35_F and MYH7_HRM_R	F: TCCTCAGCTGGTTGTCACTG R: AACAGACCATTAAGGACCT	242	chr14:23884162-23884403	60→54 / -0.5	3.5	88-94
Hereditary Head Tumors						
CARHSP1_EX2	F: CAGACCTGCCGCTGAC R: CCCTGACATTTCCCCC	169	chr16:8952187-8952355	62→52 / -0.5	2.5	82-91
CARHSP1_EX3	F: AAGGGGTGCTTCCAC R: GCTTTATTTATGCCCCC	205	chr16:8949000-8949204	50	3	83-90

P. Size: product size; Anneal. T: annealing temperature

All reactions were carried out in 96-well plates in Light Cycler 480. HRM reactions were carried out using 5 µl Roche High Resolution Melting Master Mix, 200 nM of each primer pair, 2.5, 3 or 3.5 mM $MgCl_2$, 20 ng genomic DNA and sufficient dH_2O to add up to a total volume of 10 µl. The conditions were as follows: 95°C for 10 minutes for initial denaturation, 45 cycles at 95°C for 10 seconds for denaturation, 10 seconds at the appropriate annealing temperature (given in Table 4.11) and 10 seconds for elongation at 72°C. HRM curve data acquisitions were performed in the range of 65-95°C at a rate of 25 acquisitions per °C. Normalization of fluorescence values and generation of melting curves were performed using Light Cycler 480 Gene Scanning software.

4.7. Relative Quantification of *DES* Transcripts

Relative quantification was performed to assess *DES* and *POL2A* transcript levels to investigate whether mutated *DES* mRNA possibly undergoes nonsense mediated decay. A muscle biopsy sample of patient 402 and a positive control (biopsy sample of an individual suffering from another muscle disease) were obtained. Total RNAs were isolated with a Qiagen RNA Isolation Kit and converted to cDNA with random hexamer primers by using a Thermo Reverse Transcription Kit. Real-time quantitative PCR was performed using those cDNA samples as templates and Roche Syber Green I Master kit on Roche Light Cycler 480 in 96-well plates. Samples were assayed in triplicates, and the control sample was set as the calibrator. The reactions were performed using 10 µl Light Cycler 480 SYBR Green Master mix, 200 nM of each primer pair, 60 ng cDNA and sufficient water adding up to a total volume of 20 µl. Primer pairs were intron-spanning and specific to either *DES* or *POL2A*. *POL2A* was the reference house-keeping gene. The sequences of the primer pairs used in relative quantification are given in Table 4.12. The conditions were as follows: 10 minutes for initial denaturation at 95°C, 45 cycles of 10 seconds at 95°C for denaturation, for 15 seconds of annealing at 64°C →58°C touchdown (-0.5°C for 13 cycles) and for 17 seconds of elongation at 72°C. Cycle threshold (C_T) values and $2^{-\Delta\Delta C}{}_T$ method (Livak *et al*., 2001) calculations were performed using Light Cycler 480 Relative Quantification software, and the graphs showing ± standard error of the mean (SEM) were generated by the software.

Table 4.12. Primers used in relative quantification of *DES* transcripts.

Primer Name	**Primer Sequence (5' → 3')**	**Product Size (bp)**	**Primer specificity**
DES_EXP	F: AGCTGCTGGACTTCTCACTGG R: CTGCATCCACGTCCGCT	392 bp	F: Exon 1 R: Exons 2 and 3
POLR2A_RT	F: TGCTCTTCAACATCCACCTG R: ACACCCAGCGTCACATTCTT	247 bp	F: Exon 19 R: Exon 20

5. RESULTS

In this section, searching for disease loci via linkage analysis and for causative mutations via candidate gene approach or exome sequencing is summarized for the five recessive and one dominant disease within the scope of the thesis.

5.1. Congenital Disorder of Glycosylation

SNP genome scan was performed in the CDG family, and candidate disease loci were determined. Later, the causative mutation was detected via exome sequencing. The mutation segregated with the disease in the family.

5.1.1. Linkage Analysis and Haplotype Analysis

The data generated by SNP genome scan in the two affected brothers and their parents were used for multipoint linkage analysis assuming an autosomal recessive model, and multipoint LOD scores were calculated. A maximal LOD score of 2.66 was obtained at seven loci. Genotypes at those loci were investigated by HCiE, and three of the loci were eliminated since the patients were heterozygous. The remaining four loci were considered candidate regions to harbor the disease gene. Multipoint LOD score graphics for loci yielding maximal LOD scores >2 are given in Figure 5.1, and candidate loci with flanking SNPs and their genomic positions are presented in Table 5.1.

Locus 4q12-24 was considered the strongest candidate due to its large size. Nonetheless, all candidate loci were included in the evaluation of the exome sequence results. Patient haplotypes at all four candidate loci were assessed as IBD.

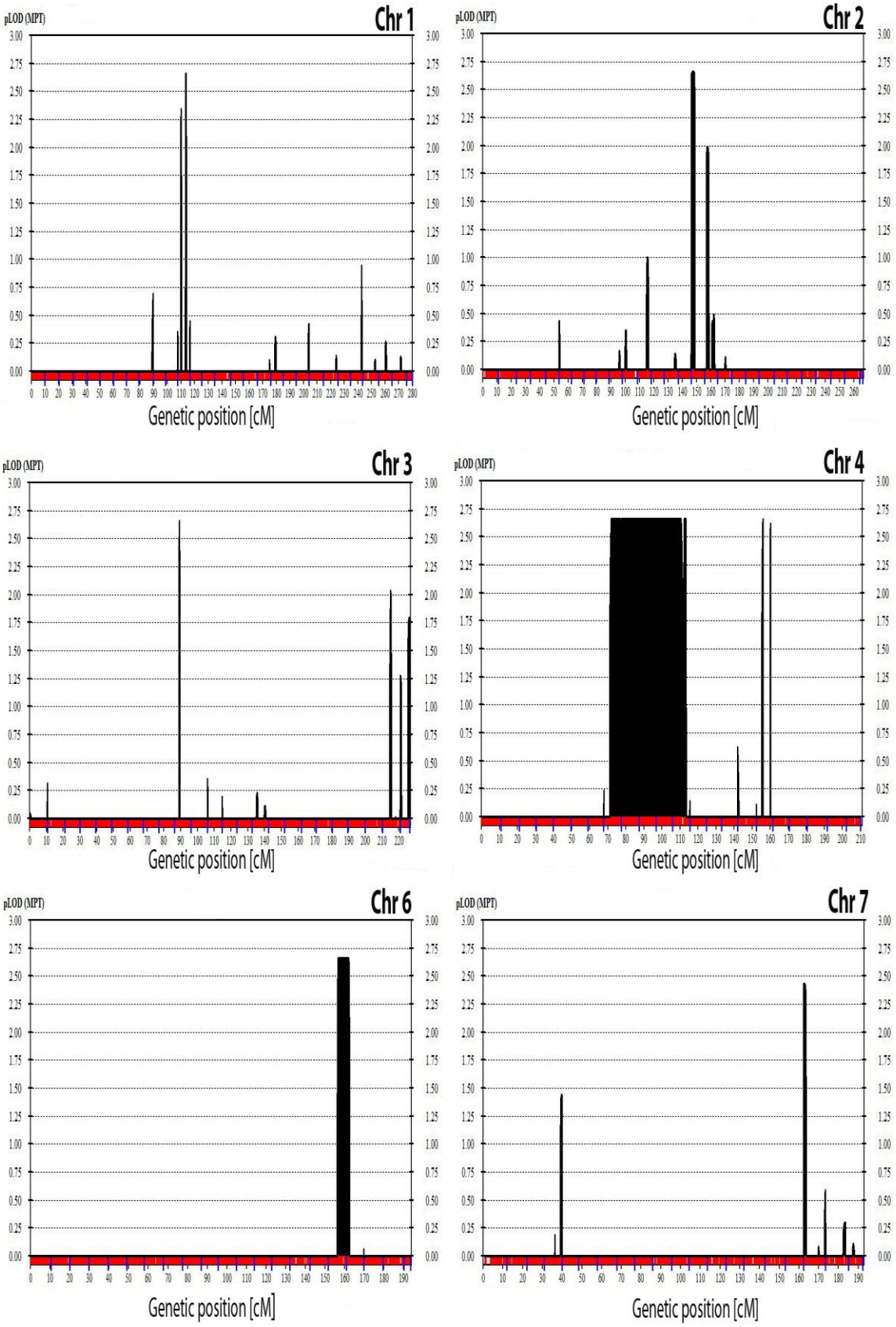

Figure 5.1. Multipoint LOD score graphics for CDG family. Only the chromosomes that yielded maximal LOD scores >2 are presented.

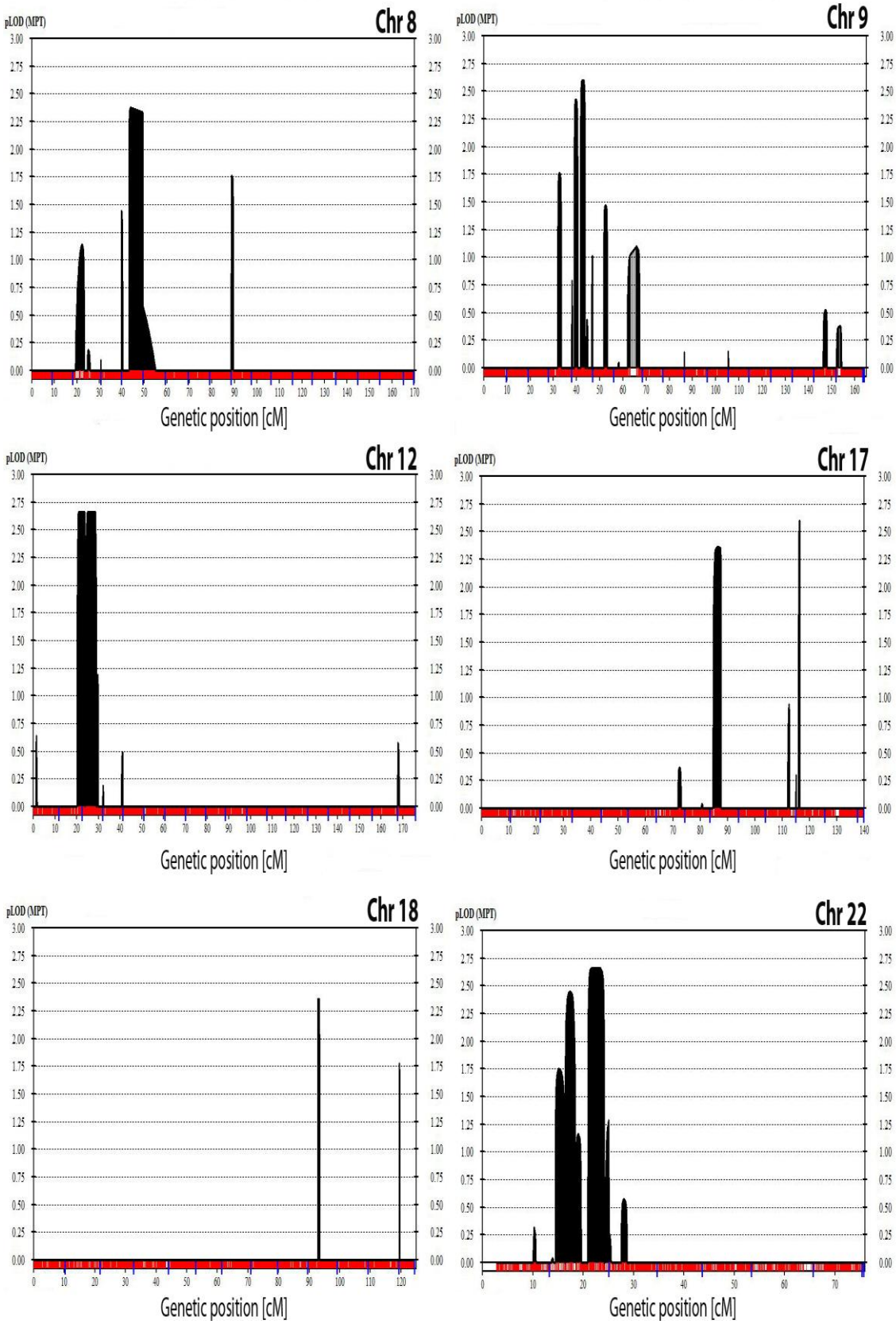

Figure 5.1. Multipoint LOD score graphics for CDG family (cont.). Only the chromosomes that yielded maximal LOD scores >2 are presented.

Table 5.1. Candidate loci for CDG family. They are listed in the order of size (Mb).

Locus	LOD Score	Flanking SNPs (bp)	Size (Mb)
4q12-24	2.66	rs2067951 (55,805,555) - rs11733406 (107,751,089)	51.94
12p13.31-p13.1	2.66	rs1868798 (7,284,081) - rs873218 (12,765,603)	5.48
6q25.1-25.2	2.66	rs912561 (150,492,727) - rs2758767 (152,985,508)	2.49
22q12.1	2.66	rs5760440 (24,865,187) - rs5752043 (25,416,023)	0.55

5.1.2. Exome Sequence Analysis and Evaluation of the Variants

DNA sample of patient 601 was subjected to exome sequence analysis. Alignment of the reads, annotation of variants and variant calling with standard parameters were provided by the company. Exome sequence data at candidate loci were evaluated, and possibly deleterious variants as explained in section 4.2.8 were selected. Table 5.2 shows the filtered variants.

The strongest candidate variant was assessed as nonsense/stopgain *SRD5A3* c.57G>A (p.W19X), considering its harmful effect to the protein product and relevance of the function of the gene to the clinical phenotype. The variant was searched in the other in-laboratory exome sequencing samples and EVS and was not found in any of them. Additionally, BamView program was used to visualize the reads, and the variant was assessed as likely real, as it resided in mid-regions of the sequence reads. The variant was reported as the disease mutation in three unrelated children afflicted with SRD5A3-CDG (Assmann *et al.*, 2001; Prietsch *et al.*, 2010; Gründahl *et al.*, 2012).

The mutation was validated by Sanger sequencing in the DNA sample of patient 601. Chromatograms showing the mutation in the patient and a reference sample are given in Figure 5.2.

Table 5.2. Filtered variants at the candidate loci in CDG patient.

Genomic position in bp	**Ref base /Alt base**	**Hom /Het**	**Qual. score**	**Total depth/ Alt depth**	**dbSNP131 MAF/EVS MAF**	**Region**	**Gene**	**Change**
At 4q12-24, from rs2067951 (55,805,555 bp) to rs11733406 (107,751,089 bp)								
56,212,560	G/A	hom	40	3/3	-/-	exonic	*SRD5A3*	stopgain
89,300,218	G/A	hom	104	20/20	-/-	exonic	*HERC6*	nonsyn
100,349,766	G/A	hom	225	155/154	-/-	exonic	*ADH7*	nonsyn
At 12p13.31-p13.1, from rs1868798 (7,284,081 bp) to rs873218 (12,765,603 bp)								
9,885,707	-/TA AGT	hom	4939	82/56	-/0.75	exonic	*CLECL1*	insertion
9,994,446	TGT /-	hom	1062	146/139	rs33911869 NR/-	exonic	*KLRF1*	deletion
11,244,027	T/C	hom	84	21/21	-/-	exonic	*TAS2R43*	nonsyn
11,244,036	T/G	hom	53	22/22	-/-	exonic	*TAS2R43*	nonsyn
11,244,067	-/TT	hom	218	20/13	-/-	exonic	*TAS2R43*	insertion
11,546,192	G/A	hom	221	165/150	rs10845349 NR/0.16	exonic	*PRB2*	nonsyn
25,294,512	G/T	het	24	5/3	-/-	exonic	*SGSM1*	nonsyn
At 6q25.1-25.2, from rs912561 (150,492,727 bp) to rs2758767 (152,985,508 bp) No variants listed.								
At 22q12.1, from rs5760440 (24,865,187 bp) to rs5752043 (25,416,023 bp) No variants listed.								

Qual: quality; ref: reference; alt: alternative; hom: homozygous; het: heterozygous; nonsyn: nonsynonymous; NR: not reported

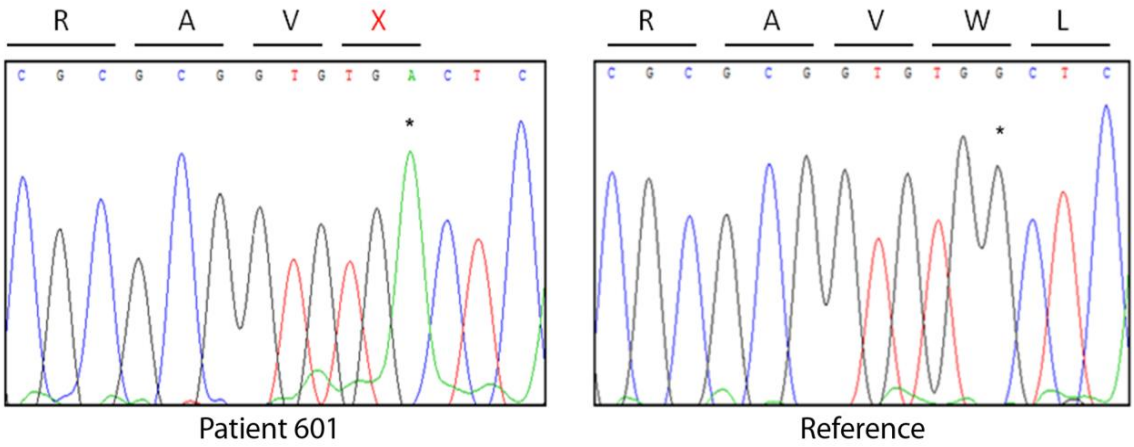

Figure 5.2. Chromatograms showing mutation *SRD5A3* c.57G>A (p.W19X).

5.1.3. Population Screening

The parents, affected sons, and 56 control samples were screened for the mutation by SSCP analysis (Figure 5.3). The parents were found heterozygous whereas the affected brothers were homozygous. The mutation was not found in the population control samples.

The conclusion is that the detected mutation was the causative mutation.

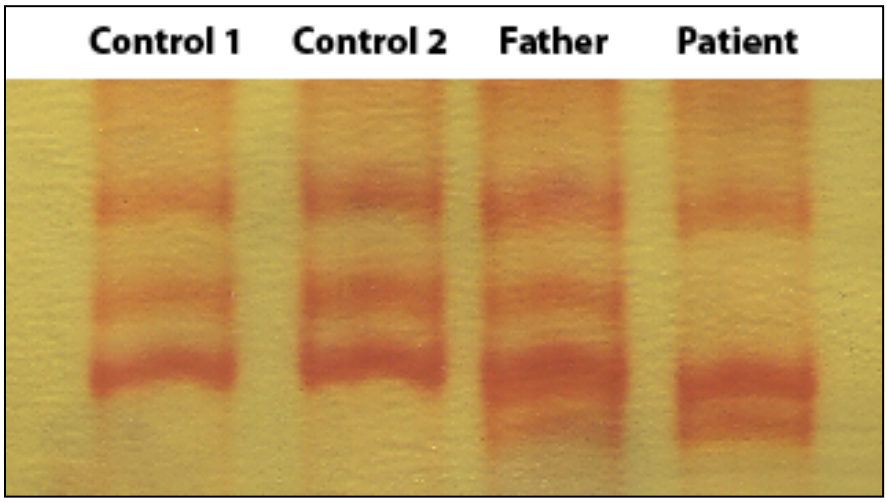

Figure 5.3. Part of a SSCP gel for population screening for *SRD5A3* c.57G>A.

5.2. Myosin Storage Myopathy

Candidate disease loci were found via linkage analysis, and *MYH7* was assessed as the strongest candidate gene. The last four coding exons were subjected to Sanger sequencing in a patient. The family members and 120 population control samples were screened by HRM analysis for the identified mutation.

5.2.1. Linkage Analysis

SNP genome scan was performed for all four children and the mother. Linkage analysis for calculating parametric multipoint LOD scores was performed assuming autosomal recessive inheritance with full penetrance. The loci that yielded maximal LOD scores >2 are presented in Figure 5.4.

Eight loci yielded multipoint LOD scores around 3, and in one of them haplotype segregation did not support IBD. Detailed linkage analysis was performed for the remaining seven loci, and LOD scores increased at four of them and did not change at the remaining three loci (Table 5.3).

In conclusion, seven candidate loci were found, ranging in size from 0.12 to 11.89 cM. Largest locus (by cM) 14q11.2 was assessed as the candidate harboring the disease gene.

5.2.2. Candidate Gene Approach

Morbid/Disease search results for the candidate loci indicated that *MYH7* that resides at the largest (by cM) candidate locus (14q11.2) was the strongest candidate gene because defects in that gene are known to be responsible for MSM and cardiomyopathy (MIM 160760). *MYH6*, encoding alpha heavy chain subunit of cardiac myosin, resides at the same candidate locus and is responsible for familial hypertrophic cardiomyopathy (MIM 613251), dilated cardiomyopathy (MIM 613252), sick sinus syndrome 3 (MIM

614090), and atrial septal defect 3 (MIM 614089). Since the patients had MSM besides cardiomyopathy, we focused on *MYH7*.

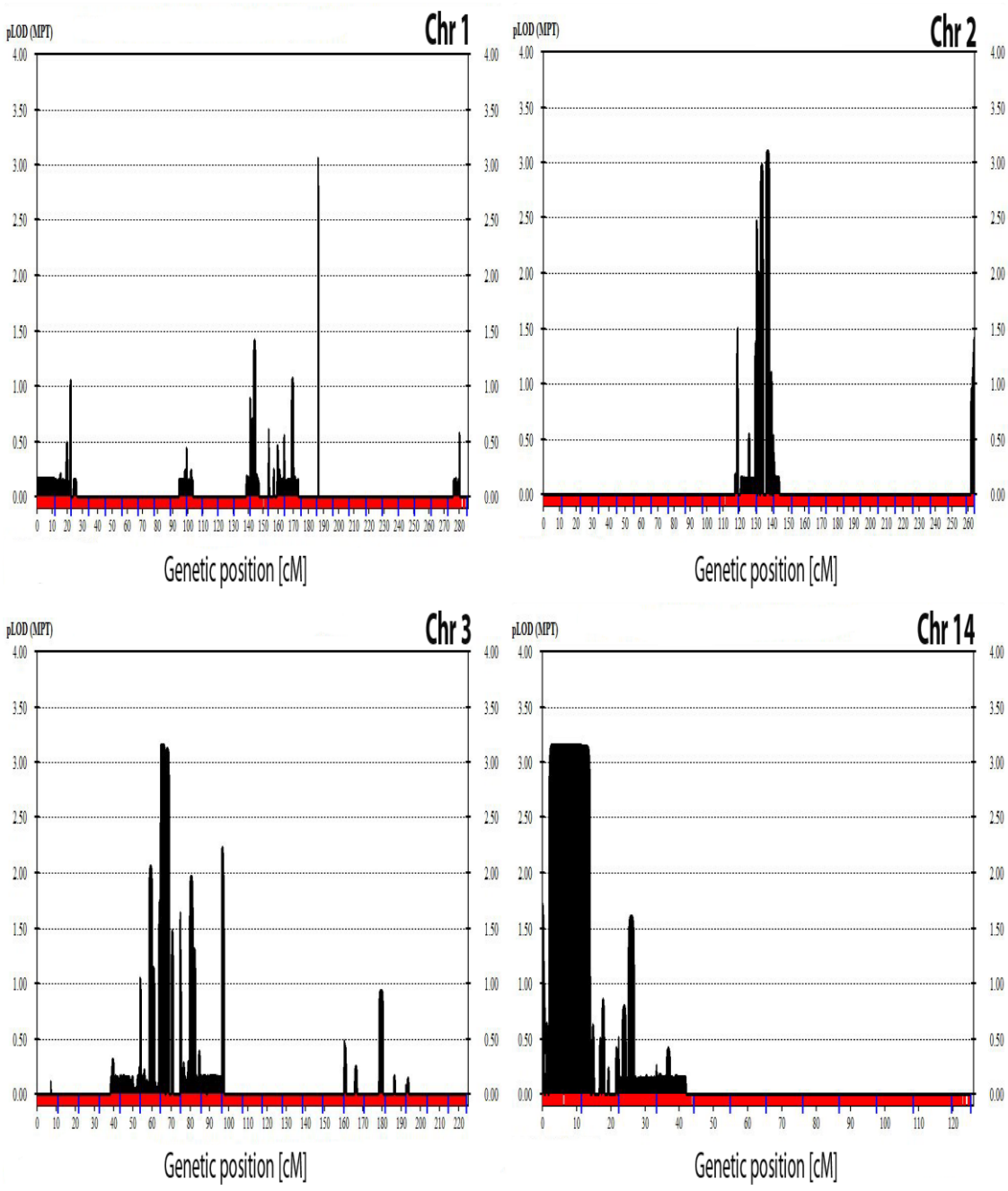

Figure 5.4 Multipoint LOD score graphics for MSM family. Only the chromosomes that yielded maximal LOD scores >2 are presented.

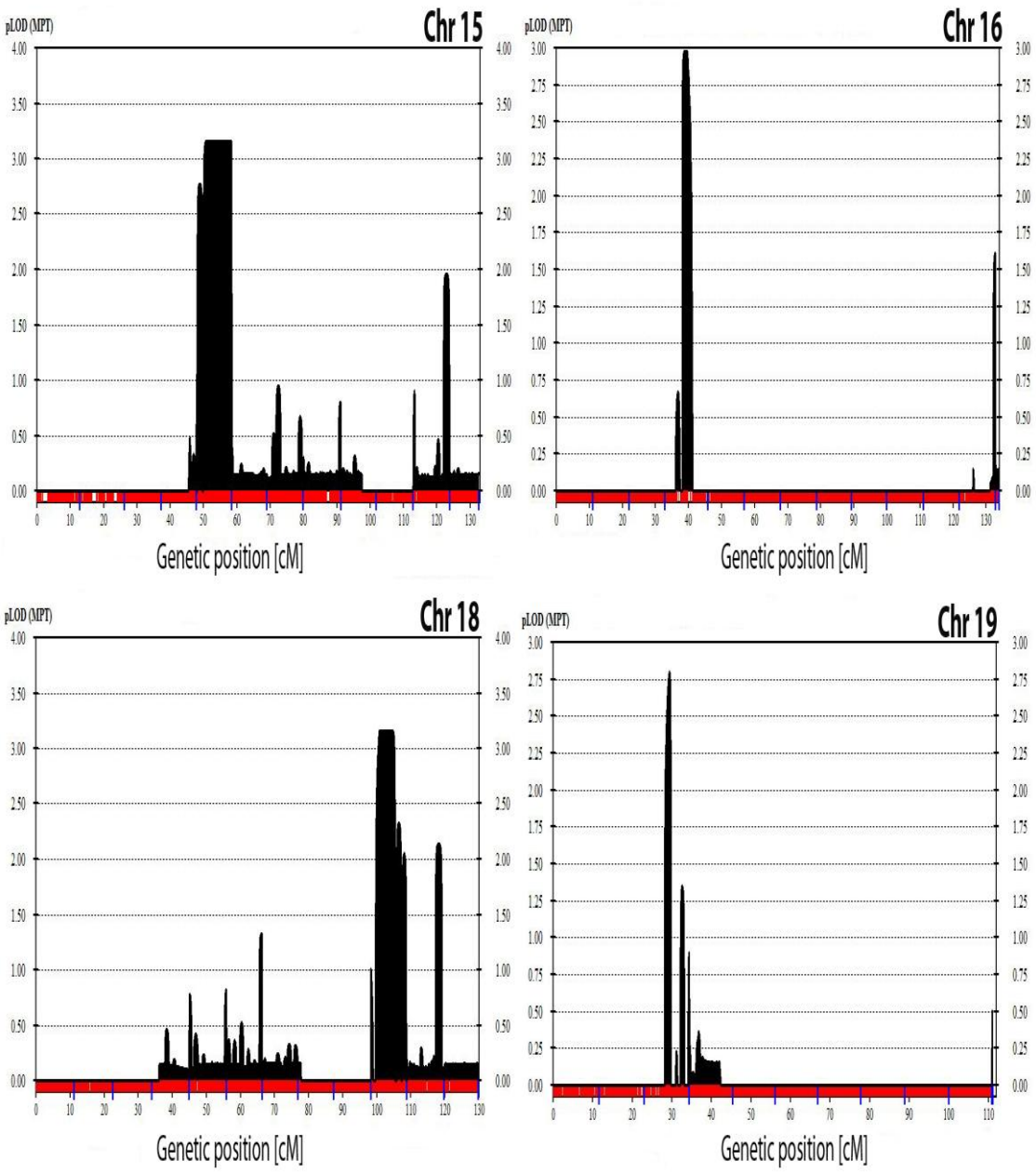

Figure 5.4. Multipoint LOD score graphics for MSM family (cont.). Only the chromosomes that yielded maximal LOD scores >2 are presented.

MYH7 mutations cause various myopathies, namely, cardiomyopathy dilated 1S (MIM 613426); cardiomyopathy, familial hypertrophic 1 (MIM 192600); Laing distal myopathy (MIM 160500); left ventricular noncompaction 5 (MIM 613426); myosin storage myopathy (MIM 608358); and scapuloperoneal syndrome, myopathic type (MIM 181430). However, all *MYH7* mutations associated with MSM to date are in the last four exons (Tajsharghi and Oldfors, 2013). That is why those exons (37, 38, 39 and

40) were analyzed with priority. DNA sample of patient 602 was subjected to Sanger sequencing. The extent of sequencing is given in Table 5.4.

Table 5.3. Candidate regions for MSM family. Regions where only the affected children are homozygous for a haplotype possibly IBD from a recent ancestor are presented.

Locus	**LOD Score***	**Homozygosity in patients**			**Shared parental haplotype**	
		Flanking SNPs (bp)	**Size**		**Flanking SNPs (bp)**	**size (Mb)**
			cM	**Mb**		
14q11.2	3.16 (3.16)	rs1243705 (21,004,037) rs222717 (24,092,124)	11.89	3.09	Same as in patients	
15q15.3-21.2	3.16 (3.21)	rs12593509 (53,376,445) rs4417505 (59,784,648)	8.42	6.40	Same as in patients	
18q22.1-22.2	3.16 (3.16)	rs9636020 (65,286,754) rs2032213 (68,726,792)	5.44	3.44	rs4987788 (60.878.629) rs2032213 (68.726.792)	7.84
3p22.1-3p21.32	3.16 (3.16)	rs2887961 (40,214,049) rs9839994 (44,125,372)	3.36	3.91	Same as in patients	
19p13.2	2.97 (3.16)	rs12150874 (9,414,155) rs2233679 (9,945,354)	1.32	0.53	Same as in patients	
16p13.11	2.97 (3.16)	rs238848 (18,082,349) rs7199186 (18,266,347)	0.42	0.18	Same as in patients	
2q14.3	3.10 (3.16)	rs168604 (127,123,647) rs6747825 (127,372,728)	0.25	0.25	Same as in patients	
		rs7581770 (127,524,084) rs10928992 (127,644,785)	0.12	0.12	Same as in patients	

*LOD scores obtained after detailed calculations are given in parentheses.

Table 5.4. The extent of sequencing in the last four exons of *MYH7*. "-" indicates upstream and "+" indicates downstream of the exon. Completely sequenced introns are indicated.

Exon	**Sequenced part (bp)**
37	-71, +52
38	-35, +111 (intron 38)
39	-111 (intron 38), +66
40	-42, +119

Homozygous nonsynonymous variant NM_000257.2(MYH7):c.5458C>T (p.R1820W) was identified in exon 37. Chromatograms showing the variant in the patient and a reference sequence are given in Figure 5.5. Online tools predicted the variant as damaging; PolyPhen score was 0.98, Mutation Taster predicted it as disease causing with a probability of 0.99, SIFT score was 0 indicating a deleterious substitution, and MutPred predicted it as deleterious with a probability of 0.55.

5.2.3. Population Screen

Family members and 120 population control samples were screened for the mutation identified, by HRM curve analysis to detect a normal sequence variant with a frequency of 0.01 with a power of 80% (Collins and Schwartz, 2002). Both patients were homozygous whereas the mother and unaffected sister 601 were heterozygous for the mutation, and unaffected sister 604 was homozygous wild type. The mutation was not found in the control samples. Melting curves obtained in the HRM analysis are given in Figure 5.6.

In conclusion, nonsynonymous mutation *MYH7* c.5458C>T (p.R1820W) was assessed as the causative mutation in MSM family.

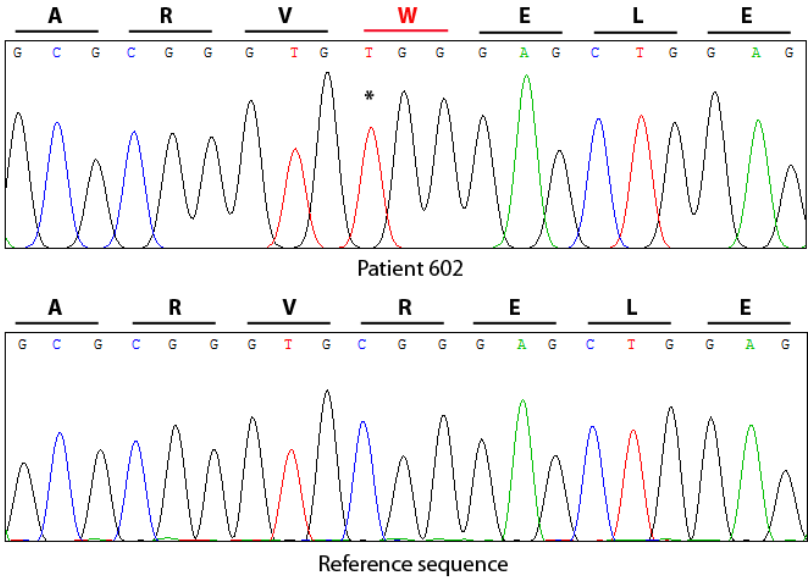

Figure 5.5. Chromatograms showing mutation *MYH7* c.5458C>T (p.R1820W).

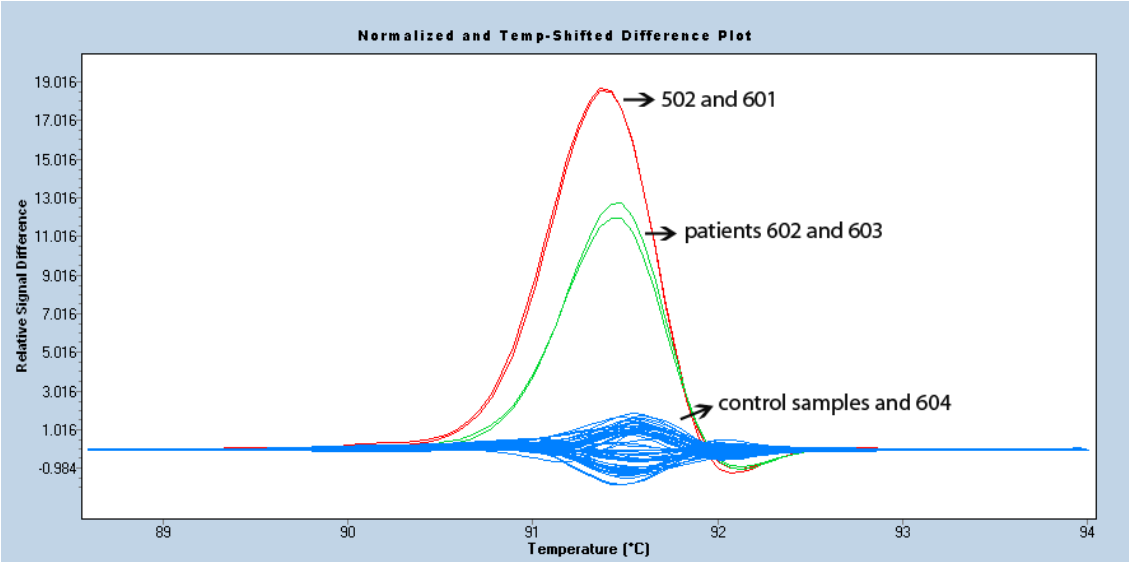

Figure 5.6. HRM analysis results for mutation *MYH7* c.5458C>T in MSM family and the control samples.

5.3. Desmin Deficiency Myopathy

Linkage analysis was performed, and the disease locus was identified. Causative mutation was identified by exome sequencing, and the family members and population control samples were screened. mRNA levels were determined by relative quantification.

5.3.1. Linkage Analysis

Multipoint linkage analysis assuming recessive autosomal inheritance and full penetrance was performed using the data generated by the genome scans in the two affected cousins and their parents. A single locus had a maximal LOD score >3.00, 2q35-37.1. Thus, the disease locus was identified as 2q35-37.1. Multipoint LOD score graphics for chromosomes yielding scores >2 are presented in Figure 5.7. The maximal homozygous region shared by the patients was approximately 16 Mb, flanked by rs2372848 (217,504,112 bp) and rs6747223 (233,211,861 bp).

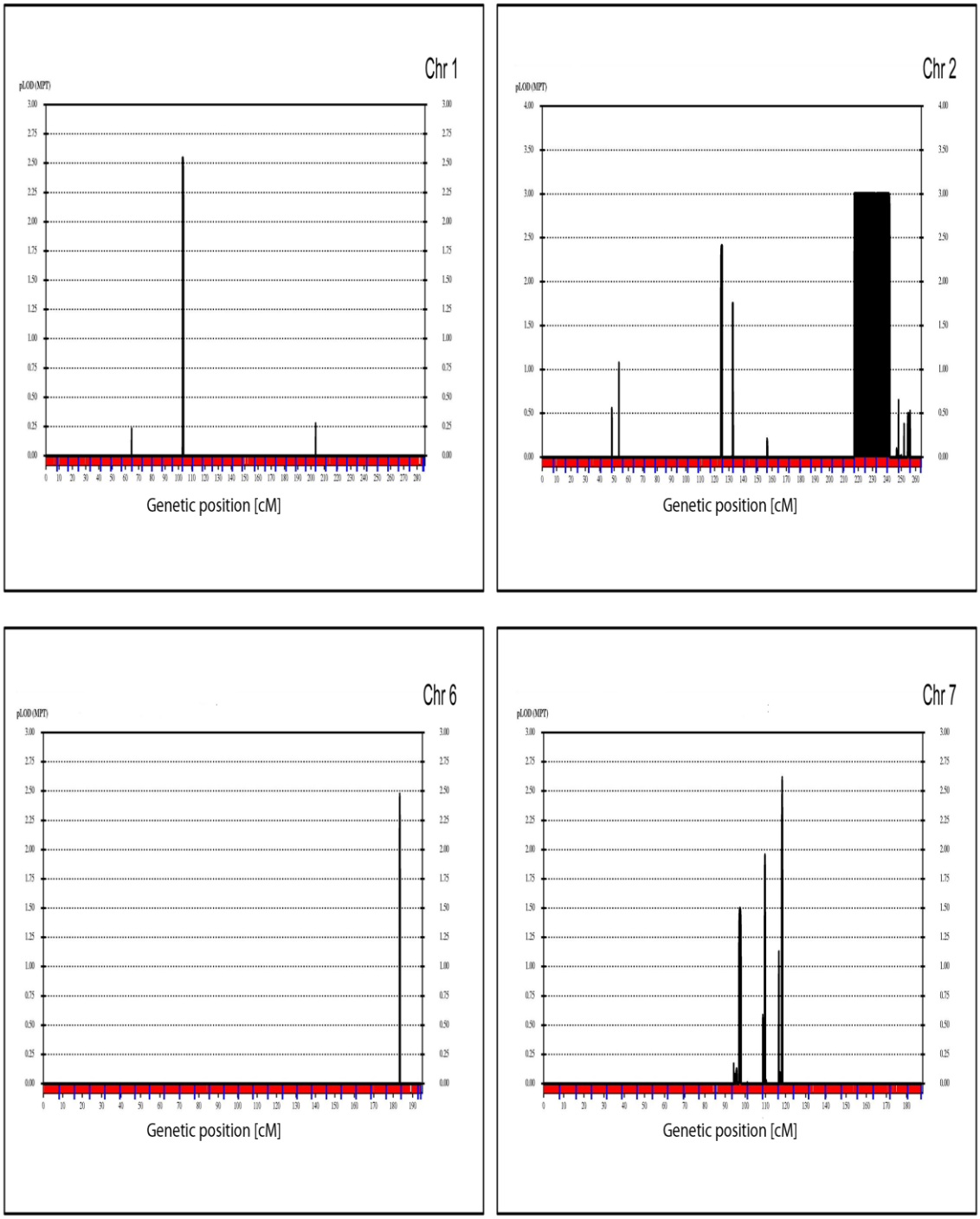

Figure 5.7. Multipoint LOD score graphics for Desmin family. Only chromosomes yielding maximal LOD scores >2 are presented.

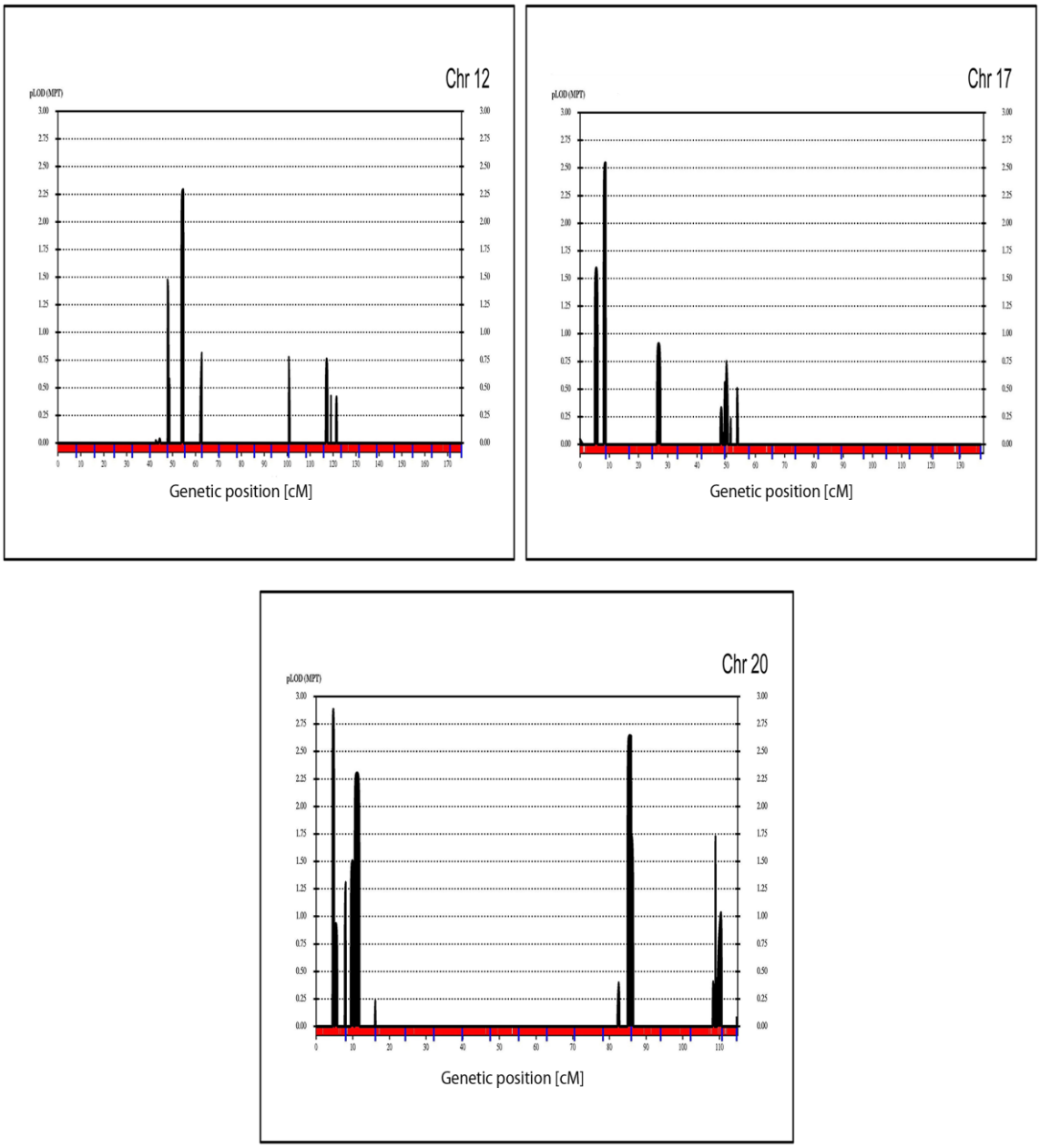

Figure 5.7. Multipoint LOD score graphics for Desmin family (cont.). Only chromosomes yielding maximal LOD scores >2 are presented.

5.3.2. Exome Sequence Analysis

The presence of a large number of genes (196 in total) at the disease locus prompted us to launch exome sequencing. DNA sample of patient 402 was subjected to exome sequencing. The variant data were filtered in the disease gene region, and

novel/rare exonic/splicing variants (except for the synonymous variants) and potentially deleterious variants were selected as described in section 4.2.8. The final variants remaining after the filtration are given in Table 5.5.

Table 5.5. Novel exonic/splicing variants at 2q35-37.1 in Desmin patient.

Genomic position in bp	**Ref base/ Alt base**	**Hom /Het**	**Qual. score**	**Total depth/ Alt depth**	**dbSNP131 MAF/ EVS MAF**	**Region**	**Gene**	**Change**
219,871,207	C/T	hom	90	24/22	rs149635127 0.002/0.0004	exonic	*CCDC108*	nonsyn
220,283,529	-/C	hom	214	27/26	-/-	exonic	*DES*	insertion
231,338,156	G/A	het	94	22/20	rs150147150 0.001/0.002	intronic	*SP100*	splicing

Qual: quality; Ref: reference; alt: alternative; hom: homozygous; het: heterozygous; nonsyn: nonsynonymous

Of the three candidate variants, missense c.4699G>A (p.E1567K) in exon 30 of *CCDC108* (Coiled-coil domain containing protein 108) and splice variant c.1546+1G>A in *SP100* (Nuclear body protein SP100) are in genes that are expressed in most tissues except for muscle (http://www.ncbi.nlm.nih.gov/unigene/EST profiles). Moreover, online tools predicted the missense variant *CCDC108* c.4699G>A as benign; PolyPhen score was 0.058 indicting that it was not harmful, Mutation Taster predicted it as polymorphism with a probability of 0.73, and SIFT predicted that it was tolerated with a score of 0.38. The splice variant *SP100* c.1546+1G>A was predicted as disease causing by Mutation Taster since donor site was predicted to get weaker (wt: 0.98 / mt: 0.31). Remaining variant frameshift c.345dup (p.N116Qfs*2) in *DES* exon 1 was assessed as the plausible disease mutation, as it is deduced to result in a null allele for Desmin, a muscle protein defective in various muscle diseases (MIM 125660).

The identified *DES* variant was validated by Sanger sequencing using DNA sample of 402. Chromatograms showing mutation *DES* c.345dup (p.N116Qfs*2) and a reference sequence are presented in Figure 5.8.

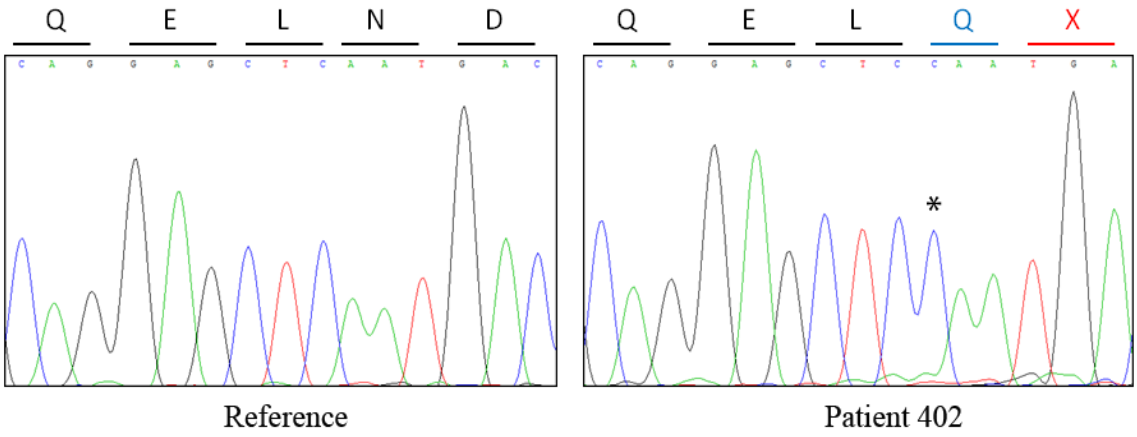

Figure 5.8. Chromatograms showing mutation *DES* c.345dup (p.N116Qfs*2).

5.3.3. Population Screen

SSCP analysis was used to screen DNA samples for the mutation since a HRM analysis could not be optimized for this purpose. Only the patients were homozygous, and the fathers, mothers and unaffected individuals 405 and 408 were heterozygous. The mutation was not found in any of the 70 control samples screened. Figure 5.9 shows a part of the SSCP gel used for mutation screening. Considering all evidence together, the causative mutation was determined as *DES* c.345dup (p.N116Qfs*2).

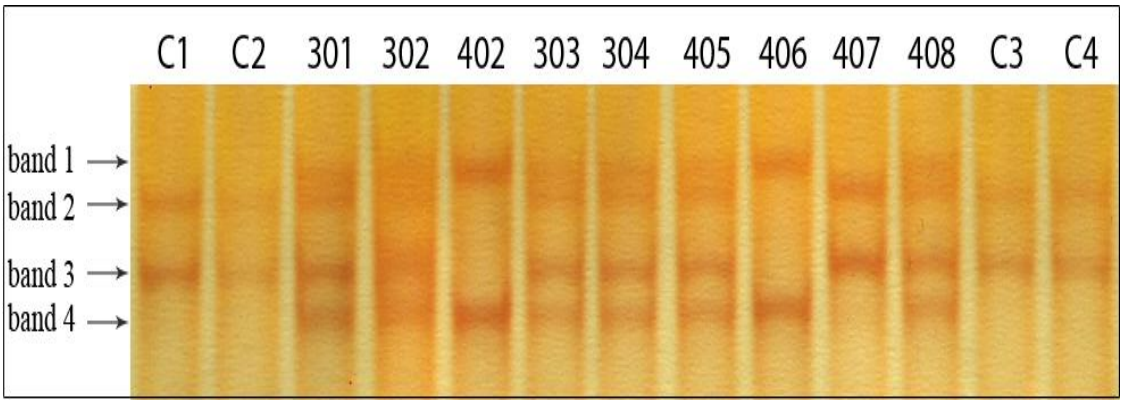

Figure 5.9. A SSCP gel for screening the family members and the population samples for *DES* c.345dup (p.N116Qfs*2).

5.3.4. Relative Quantification of *DES* transcripts

DESMIN p.N116Qfs*2 is deduced to shift the translational reading frame at codon 116 and lead to the creation of a premature termination codon after the synthesis of one nonnative amino acid. Thus, the truncated protein would contain only the first 115 native residues of the 470-amino acid protein. However, the mutant protein is likely not synthesized at all, as the mutant mRNA is expected to undergo nonsense-mediated decay, the premature stop codon being located more than 50 nucleotides upstream of the 3′-most exon–exon junction (Nagy and Maquat, 1998). In order to verify this hypothesis, relative quantification of *DES* mRNA was performed in a muscle biopsy sample of the patient and of a control muscle sample.

The muscle biopsy samples were provided by Dr. Hacer Durmuş at Istanbul University. RNA was extracted, and cDNA was synthesized. RT-PCR was performed on cDNA using an intron-spanning primer pair specific to *DES* mRNA or to *POL2A* mRNA which was the reference house-keeping gene. The *DES* transcript level in the patient normalized to that of *POL2A* was 0.3% as compared to the control, indicating that the mRNA likely undergoes nonsense mediated decay in the patient and that no desmin is synthesized. Figure 5.10 shows the comparison of *DES* transcript levels. In conclusion, the relative quantification assay revealed that *DES* mRNA was extremely decreased in the muscle cells of the patient, verifying the desmin-null phenotype.

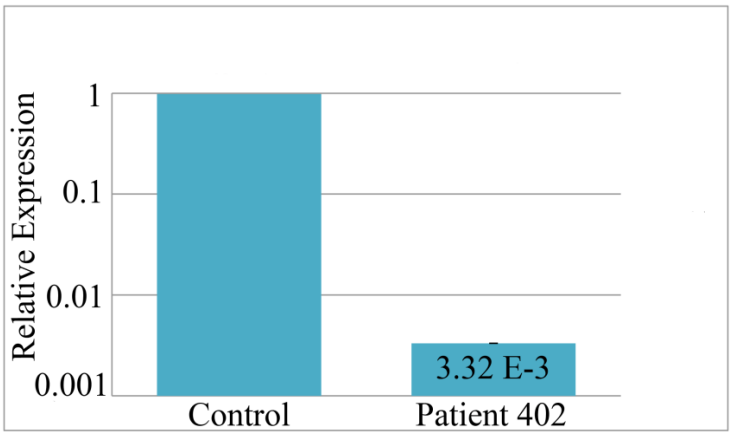

Figure 5.10. Relative *DES* transcript levels in muscle cells. Control sample was set as the calibrator.

5.4. Hereditary Head Tumors

Previously, Hereditary Head Tumors (HHT) family was studied by Tarık Bozoğlu as a M.S. Thesis project (Bozoğlu, 2008), and disease gene was localized to 16p13.2-p13.12. In this study, linkage analysis with the aim of detecting the disease locus was repeated, and similar results were obtained. Then, exome sequence was launched to identify the causative mutation. A frameshift mutation in *CARHSP1* was identified in the meningioma patients of the family, and all coding exons of the gene were screened in the cancer samples by either HRM or SSCP analysis. One nonsynonymous and three synonymous variants in the gene were found by screening of the cancer samples. Additionally, two nonsynonymous variants in different genes were detected by exome sequencing and were validated by Sanger sequencing.

5.4.1. Linkage and Haplotype Analyses

Multipoint linkage analysis was performed with microsatellites using SimWalk v2.91 program under the assumption of autosomal dominant inheritance, either full or 90% penetrance, and 0.001 disease allele frequency. The highest LOD score (3.04 with full penetrance and 2.43 with 90% penetrance) was obtained at 16p13.3-16p13.13 between 15 cM and 28 cM. The shared haplotype contained markers D16S678 (9,292,396 bp) and ATA3A07 (12,139,445 bp) and was flanked by D16S768 (8,371,609 bp) and D16S2619 (13,742,507 bp). Additionally, 8q12.1-23.1 yielded a maximal LOD score of 1.479 between GGAA8G07 (59,708,209 bp, 59.8 cM) and GATA26E03M (107,328,872 bp, 107.3 cM). Multipoint LOD score graphs for all chromosomes in the analysis assuming full penetrance are presented in Figure 5.11 and in the analysis assuming 90% penetrance in Figure 5.12.

Haplotypes were constructed with program SimWalk at 16p13.3-16p13.13 and 8q12.1-23.1 (figures 5.13 and 5.14, respectively), and a haplotype at each locus was found to segregate with head cancers.

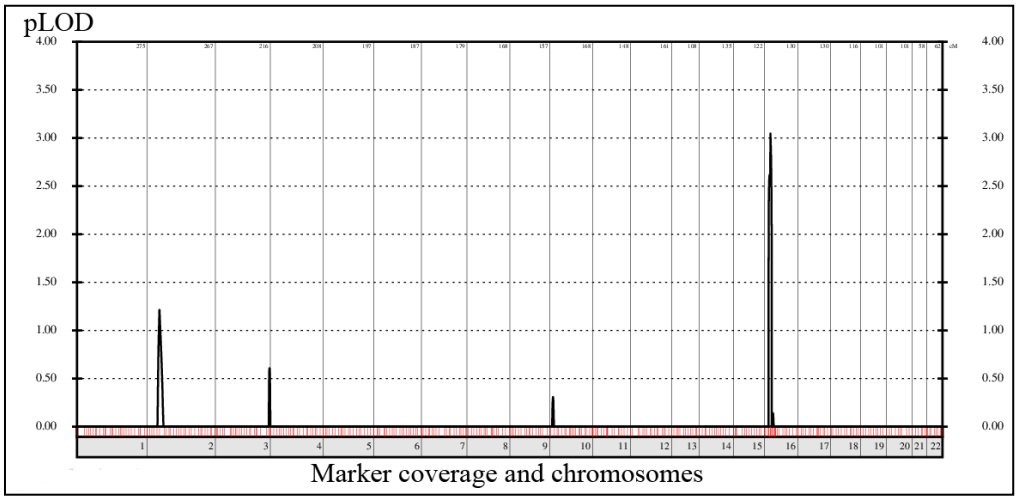

Figure 5.11. Multipoint LOD score graphs for HHT family assuming full penetrance.

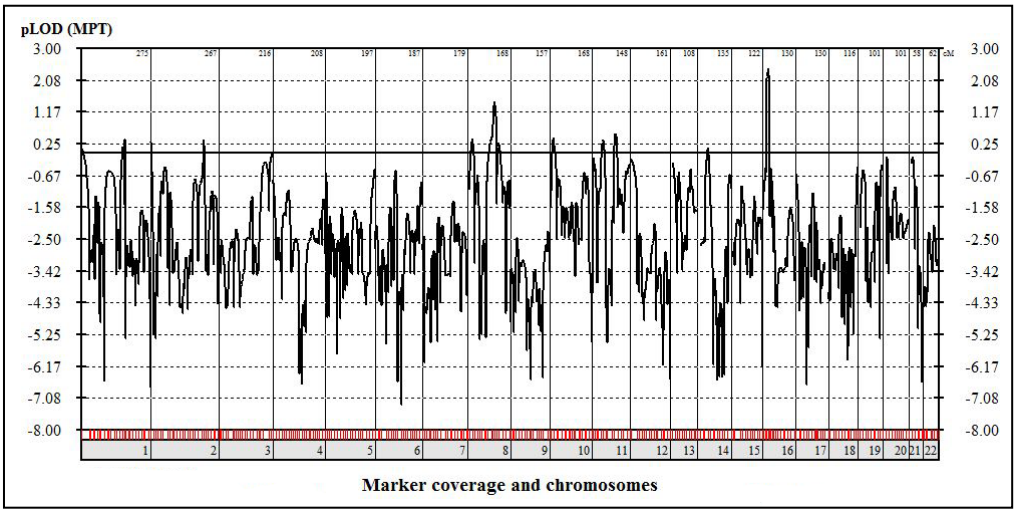

Figure 5.12. Multipoint LOD score graphs for HHT family assuming 90% penetrance.

The haplotype at 16p13.3-16p13.13 (flanked by markers D16S768 and D16S2619) was shared by all patients whose genotypes were known, and also by obligate carriers 201 and 203. Therefore, 16p13.3-16p13.13 was considered as the candidate locus harboring the disease gene with first priority. Since the disease had late onset, analysis were also performed with reduced penetrance and 8q12.1-23.1 was found. Haplotypes at the locus were analyzed, taking into consideration the late onset of the disease and the reduced penetrance. Consistent with the reduced penetrance model, a haplotype at 8q12.1-23.1 was shared by all patients but also by unaffected individuals 205 and 308. The region was considered as a candidate locus harboring the disease gene with second priority.

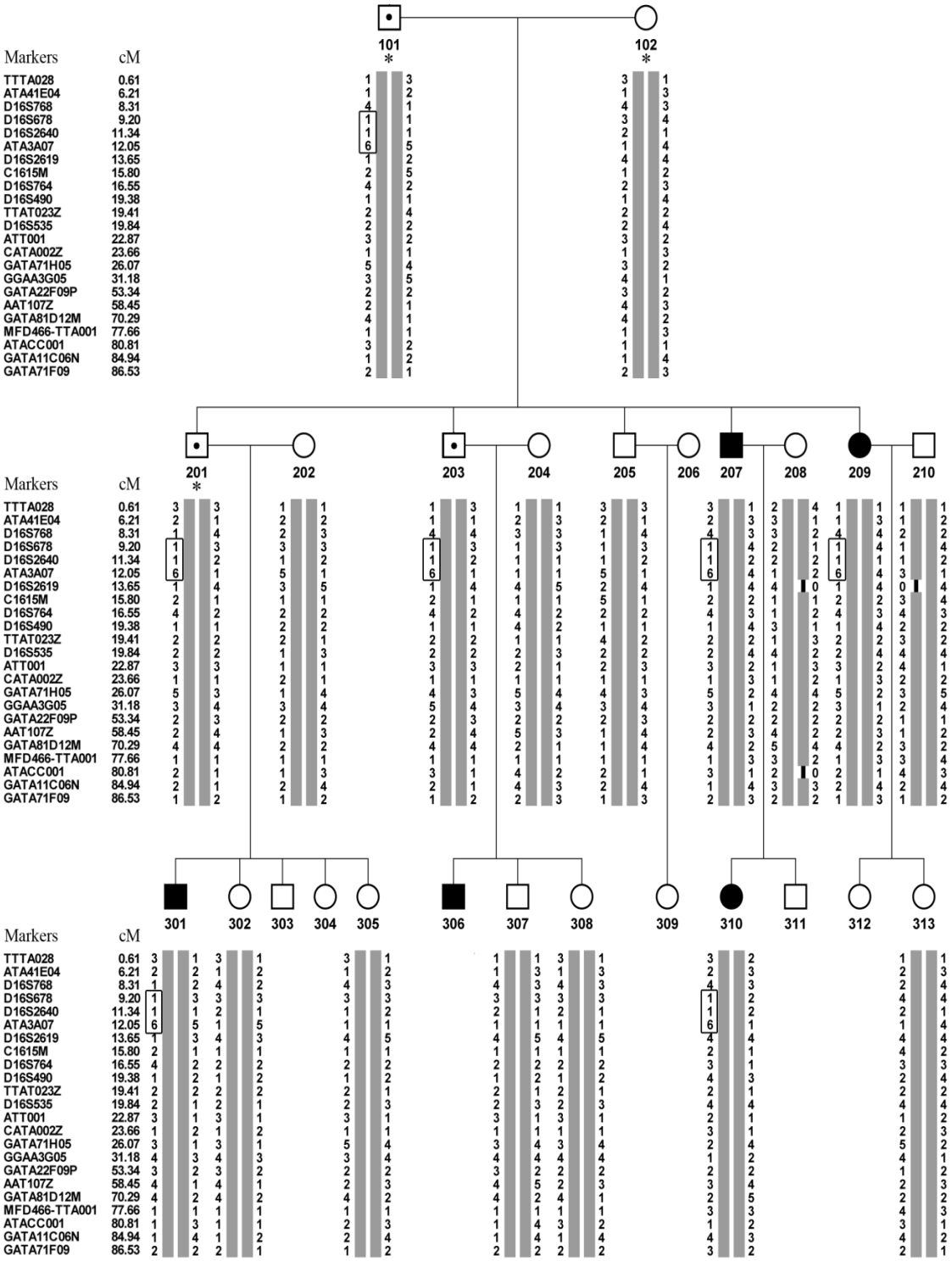

Figure 5.13. Haplotypes for HHT family for chromosome 16. The possible disease haplotype is boxed. Haplotypes of family members 101, 102 and 201 (indicated by asterisks) were deduced. Obligate carriers (indicated by dots) were introduced to the program as affected.

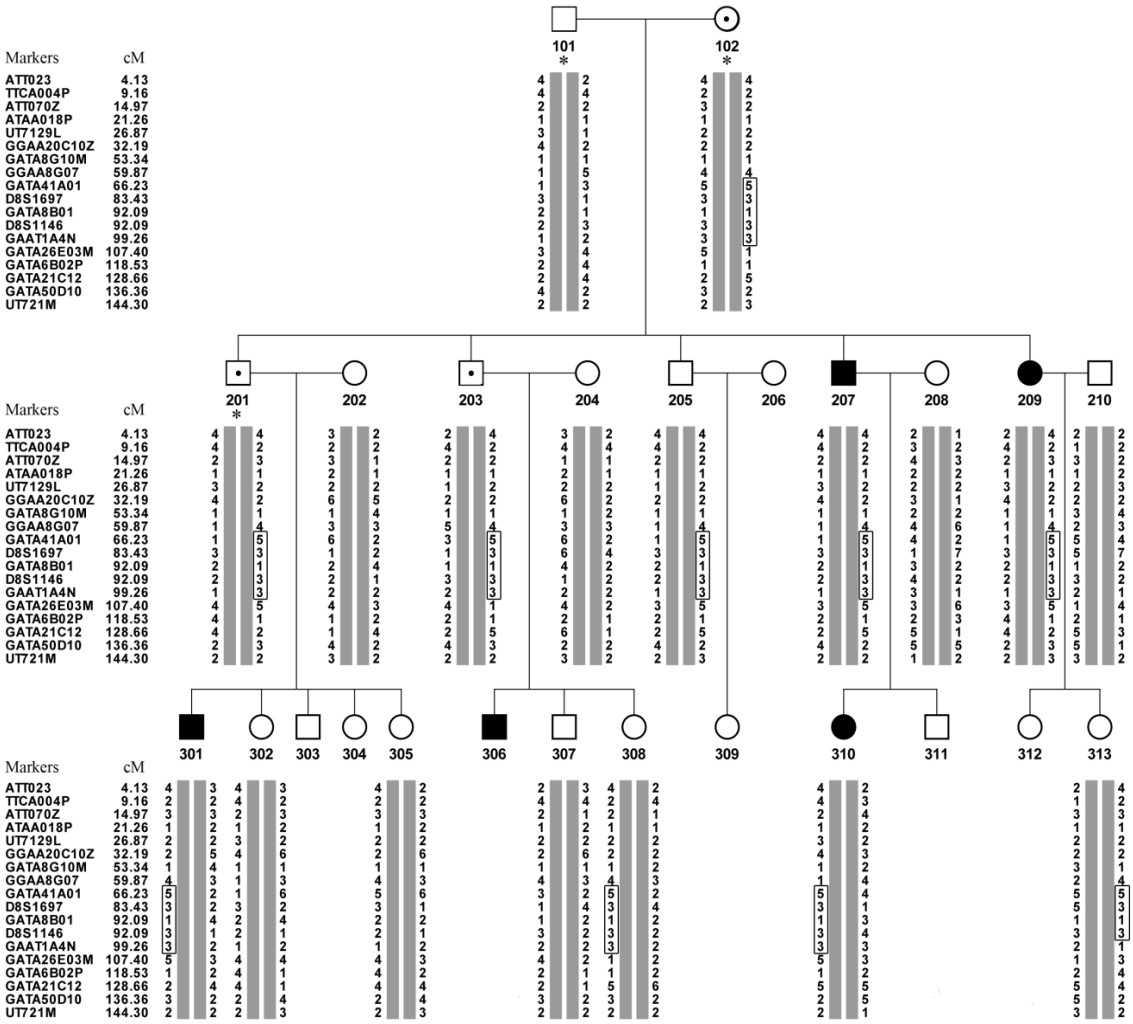

Figure 5.14. Haplotypes for HHT family for chromosome 8. The possible disease haplotype is boxed. Haplotypes of family members 101, 102 and 201 (indicated by asterisks) were deduced. Obligate carriers (indicated by dots) were introduced to the program as affected.

5.4.2. Exome Sequence Analysis

DNA sample of patient 207 was subjected to exome sequencing. Bioinformatics analysis was provided by the company. Exome sequence data were filtered at 16p13.3-p13.13 and 8q12.1-23.1 as described in section 4.2.8. Heterozygous variants were not filtered out since the inheritance model was dominant. Additionally, variants having

MAF >0.98 were investigated in exome sequence data not to miss any rare variants, but no such variants were found. Filtered variants are presented in Table 5.6.

Table 5.6. Novel or rare exonic/splicing variants at candidate loci in HHT patient.

Genomic position in bp	**Ref base/ Alt base**	**Hom/ Het**	**Qual. score**	**Total dept/ Alt depth**	**dbSNP131 MAF/EVS MAF**	**Region**	**Gene**	**Change**
At 16p13.3-16p13.13, between D16S768 (8,371,609 bp) and D16S2619 (13,742,507 bp)								
8,738,491	A/G	het	127	44/22	rs201309316 NR/-	exonic	*METTL22*	nonsyn
8,949,053	-/A	het	32	20/3	-/-	exonic	*CARHSP1*	frameshift insertion
At 8q12.1-23.1, between GGAA8G07 (59,708,209 bp) and GATA26E03M (107,328,872 bp)								
86,041,589	AA/-	het	128	46/4	-/-	exonic	*LRRCC1*	frameshift deletion
101,083,613	T/C	het	169	53/18	-/0.001	exonic	*RGS22*	nonsyn

Qual: quality, ref: reference; alt: alternative; hom: homozygous; het: heterozygous; nonsyn: nonsynonymous; NR: not reported

Sanger sequencing results validated the variants in *METTL22*, *CARHSP1* and *RGS22* but not in *LRRCC1*. Three or four online tools were used to predict the effects of the nonsynonymous variants *RGS22* p.Y193C and *METTL22* p.E363G. The prediction results are presented in Table 5.7.

Frameshift mutation *CARHSP1* c.412_413insT (p.E138Vfs*93) was the strongest candidate variant due to its deleterious effect on the protein. Chromatograms showing the reference sequence and the mutation are presented in Figure 5.15.

The mutation is deduced to lead to the replacement of the last 10 native amino acids of the 147-residue protein with 92 non-native amino acids. Amino acid sequences of the native protein and the mutant are presented in Figure 5.16.

Table 5.7. Prediction of the effects of *RGS22* p.Y193C and *METTL22* p.E363G on the respective proteins via online tools. Output scores of the online tools are explained in Materials section, Table 3.2.

Variant	PolyPhen 2	Mutation Taster	SIFT	MutPred
RGS22 p.Y193C	Probably damaging with a score of 0.998	Polymorphism with a prediction probability of 0.99	Damaging, SIFT score is 0.03	Probability of deleterious mutation: 0.528
METTL22 p.E363G	Probably damaging with a score of 0.651	Disease causing with a prediction probability of 0.99	Damaging, SIFT score is 0.04	Not available*

*Page was not available at the time.

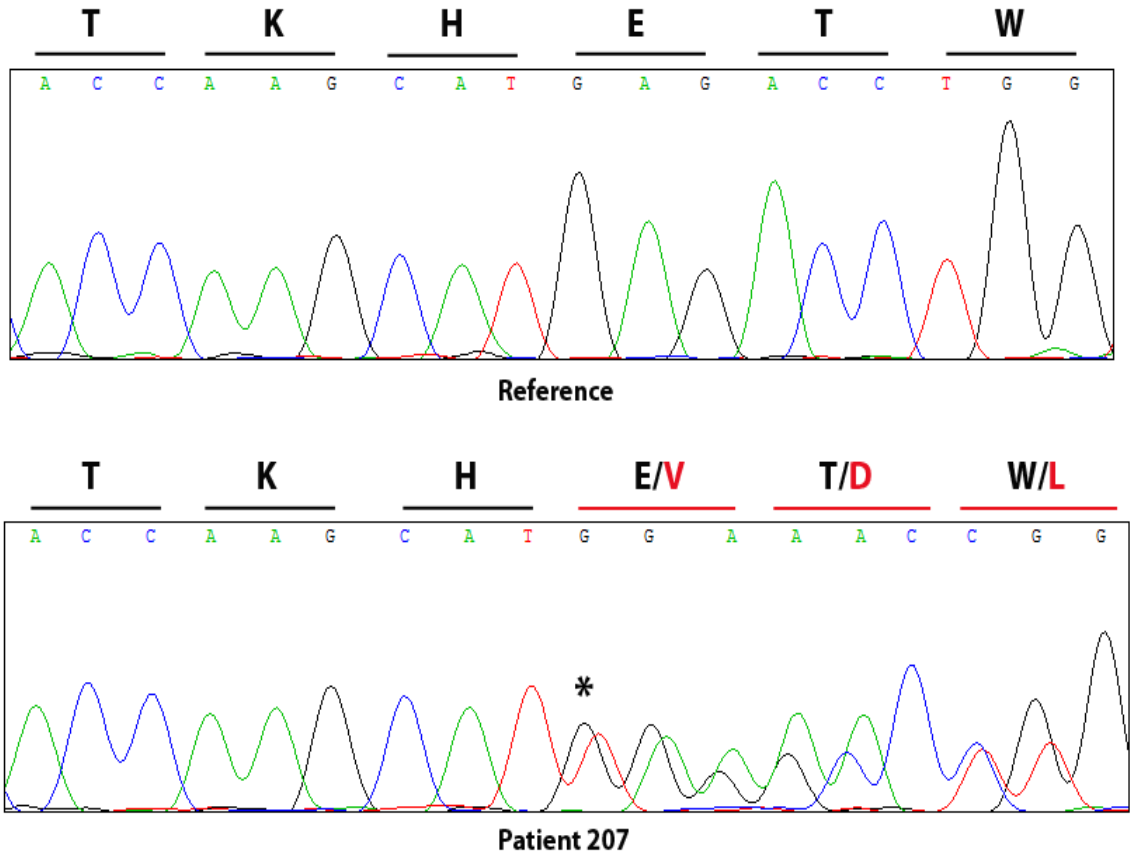

Figure 5.15. Chromatograms for frameshift *CARHSP1* c.412_413insT (p.E138Vfs*93).

Reference protein:

```
  1  MSSEPPPPPQ PPTHQASVGL LDTPRSRERS PSPLRGNVVP SPLPTRRTRT FSATVRASQG
 61  PVYKGVCKCF CRSKGHGFIT PADGGPDIFL HISDVEGEYV PVEGDEVTYK MCSIPPKNEK
121  LQAVEVVITH LAPGTKHETW SGHVISS*
```

Mutant protein:

```
  1  MSSEPPPPPQ PPTHQASVGL LDTPRSRERS PSPLRGNVVP SPLPTRRTRT FSATVRASQG
 61  PVYKGVCKCF CRSKGHGFIT PADGGPDIFL HISDVEGEYV PVEGDEVTYK MCSIPPKNEK
121  LQAVEVVITH LAPGTKHVDL VWTCHQLLGD GGSTPCPVLV GDFAGRRQQT LEMTFFHTRR
181  GFSRAWSLSS ISWRKGYGGQ VWGVGCSRPS AQPMTIATTS HHLKSIKSI*
```

Figure 5.16. The effect of *CARHSP1* c.412_413insT (p.E138Vfs*93) on the protein sequence. Deleted residues are underlined, and non-native residues are highlighted grey.

Another candidate variant found by exome sequencing was *METTL22* c.1088A>G (p.E363G) at 16p13.2. Chromatograms showing the mutation and the reference sequence are presented in Figure 5.17. The family members were screened for the mutation by HRM curve analysis. Family members 207, 209 and 313 were grouped together as having the same genotype whereas all other members were in a different group. Family members 201 and 301 were not included in the analysis since their DNA samples were not available. Melting curves of HRM analysis are presented in Figure 5.18.

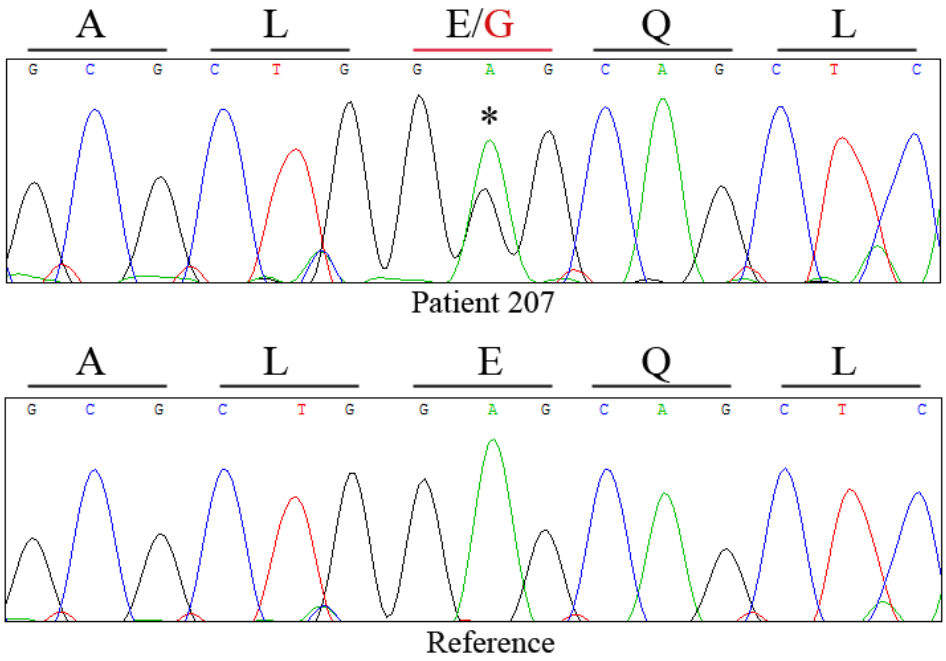

Figure 5.17. Chromatograms for *METTL22* c.1088A>G (p.E363G).

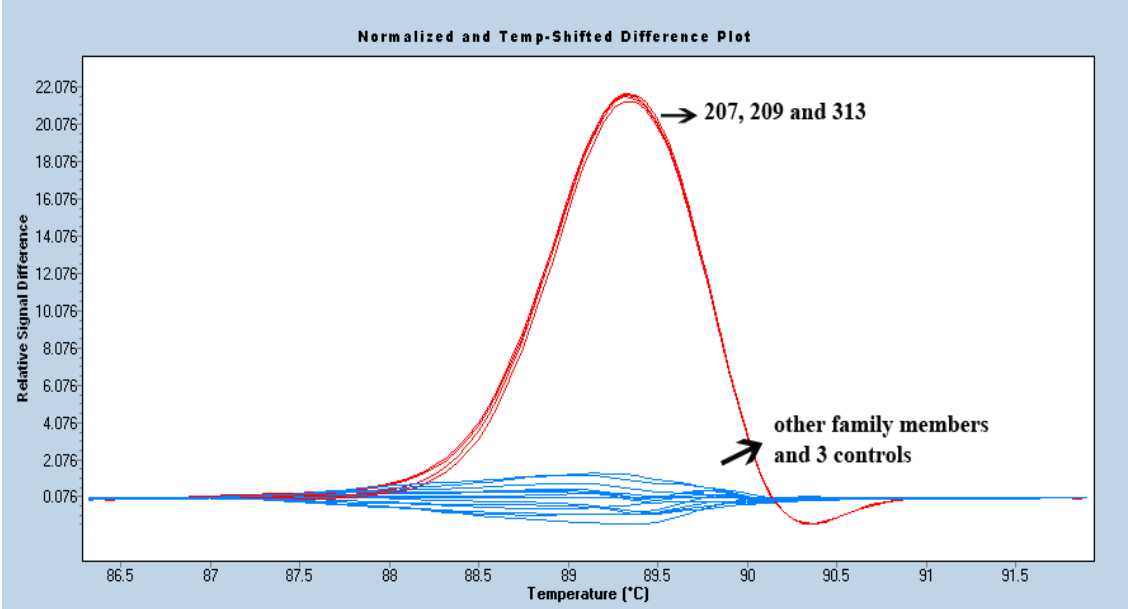

Figure 5.18. HRM melting curves for *METTL22* c.1088A>G (p.E363G).

Considering the relation of the gene to cancer, *RGS22* variant was another possible candidate. The family was screened for the variant by SSCP analysis, and it was found in the two meningioma patients 207 and 209, glioma patient 310, and unaffected individual 205. Chromatogram showing the mutation is given in Figure 5.19, and part of a SSCP gel showing the segregation of the mutation in the family is presented in Figure 5.20.

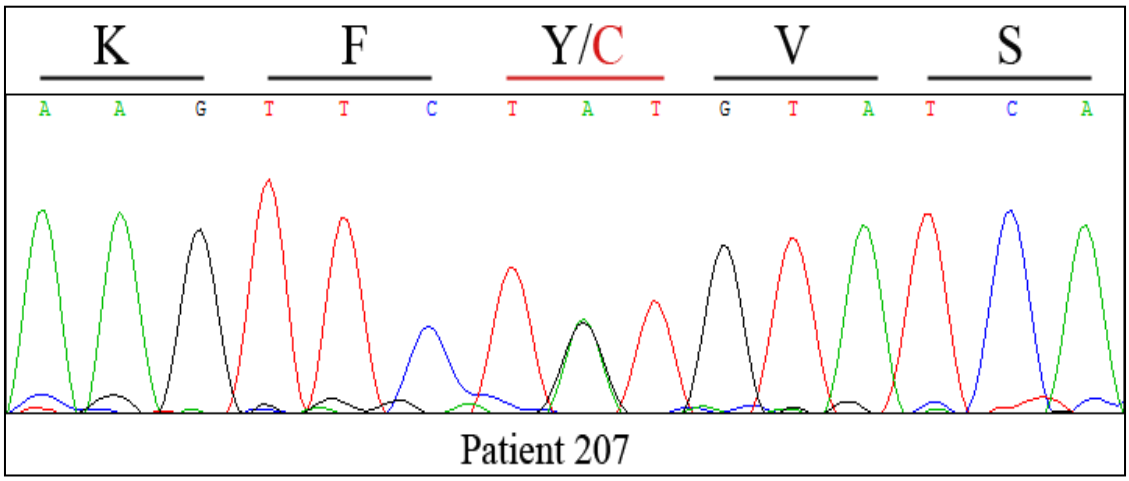

Figure 5.19. Chromatogram showing nonsynonymous mutation *RGS22* c.578A>G (p.Y193C).

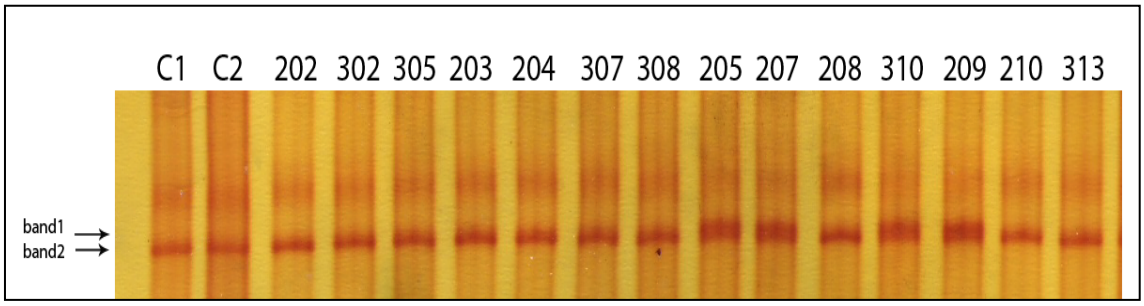

Figure 5.20. A SSCP gel showing segregation of *RGS22* c.578A>G (p.Y193C) in the family members.

5.4.3. Screening the Cancer Samples for the Identified Mutations and any Other Mutations in *CARHSP1*

HRM analysis for mutation *CARHSP1* p.E138Vfs*93 in family members 202, 203, 204, 205, 207, 208, 209, 210, 302, 305, 307, 308, 310 and 313 was performed using primers designed for Sanger sequencing. Family members 207, 209 and 313 were grouped together as having the same genotype whereas all other members were in a different group. Unexpectedly, patient 310 was in the same group as the unaffected family members whereas unaffected individual 313 was in the same group with patients 207 and 209. Melting curves of HRM analysis are presented in Figure 5.21.

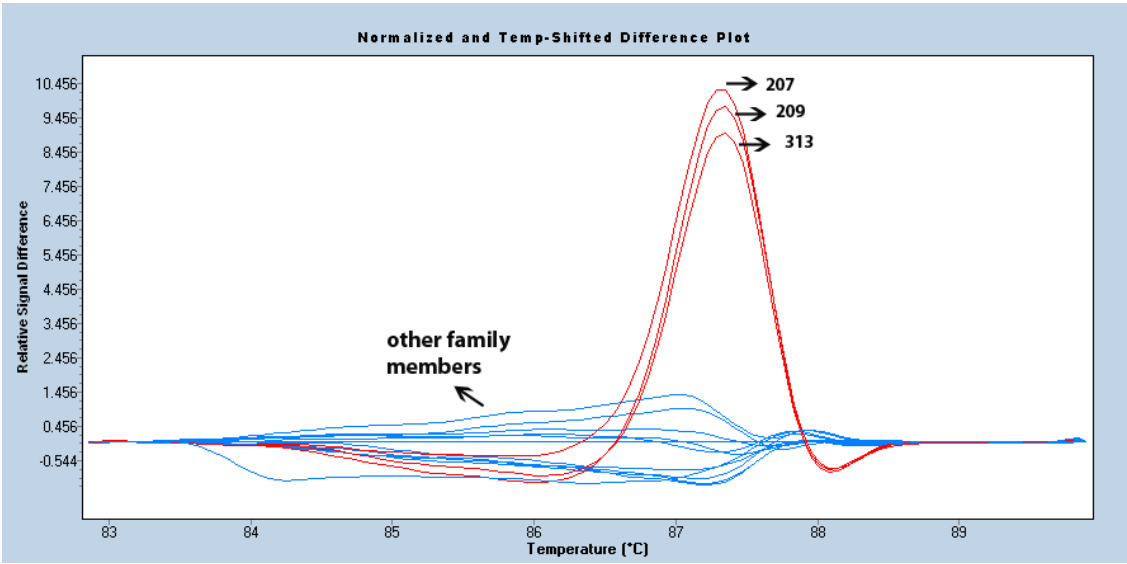

Figure 5.21. HRM melting curves for *CARHSP1* c.412_413insT (p.E138Vfs*93) for the family members.

We then decided to test via HRM analysis the brain tumor samples for the mutation and for other possible mutations in the coding regions of the gene, i.e. exons 2-4. Primer pairs were designed to analyze each coding region and its neighboring sequences.

The identified mutation is in exon 4. Thus, DNA samples were first screened by HRM analysis using primers specific to that exon. Melting curves are presented in Figure 5.22. The samples that were grouped differently were meningioma samples M13, M27, M30, M40, M45 and M48 and glioma tumor tissue sample T13; these samples were subjected to Sanger sequencing to find out what made the difference in the melting curves. One nonsynonymous and three synonymous variants were identified (Table 5.8). All were novel except for 8,949,104 C>T (rs148643770) with MAFs 0.002 in both dbSNP and EVS and 0.004 in TEVS.

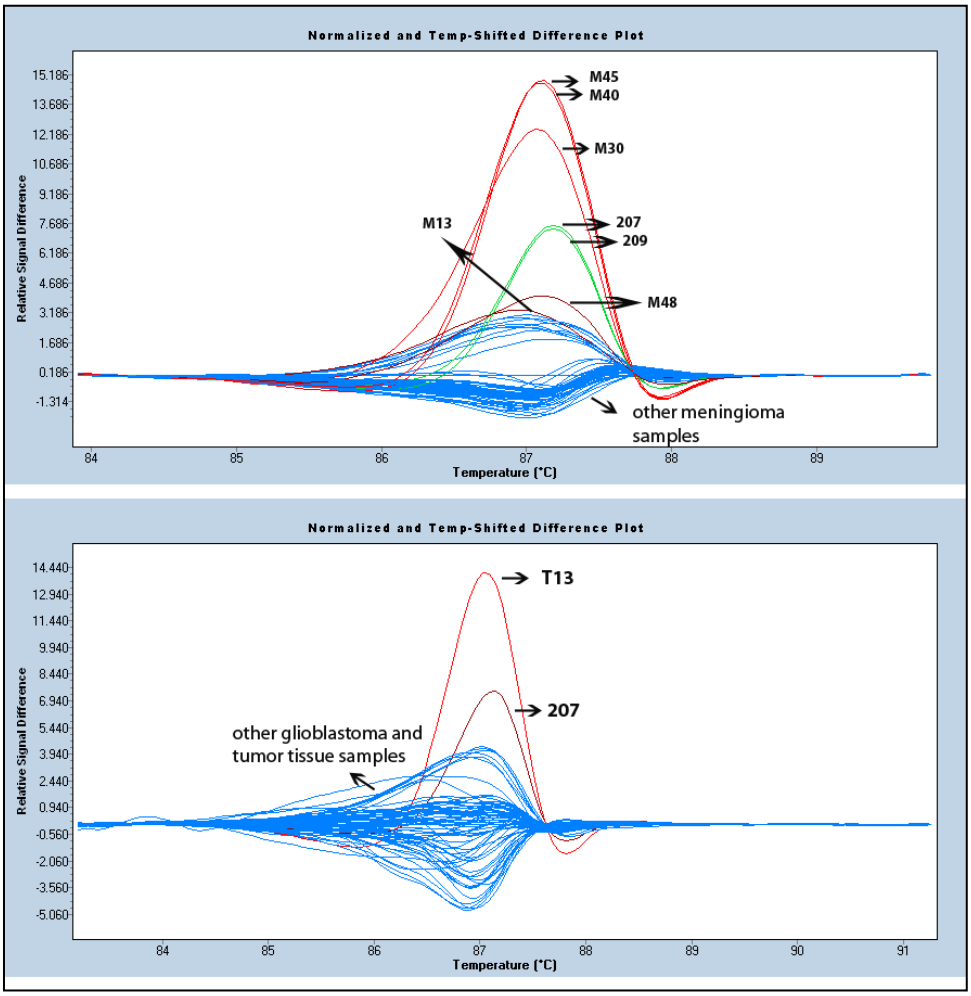

Figure 5.22. HRM analysis results of screening *CARHSP1* exon 4 in the tumor-samples.

Table 5.8. *CARHSP* exon 4 variants in HHT family and cancer samples. The gene is on chromosome 16. Online tools were used to predict the effect of the variants. MAF stands for minor allele frequency, wt for wildtype and mt for mutant.

Genomic position (bp); variation	**Change**	**Online Tools and Results**	**DNA no**
8,949,053 T insertion	frameshift	**Mtlyz:** c.411dup (p.E138*) **MuTa:** Disease causing p=0.99	HHT family 207, 209, 313
8,949,064 het G>A	nonsyn	**Mtlyz:** c.401G>A (p.G134D) **MuTa:** Disease causing p=0.99 **PP:** Possibly damaging with a score of 0.969 **MP:** Probability of deleterious mutation = 0.377 **SIFT:** Tolerated 0.42	M40, M45
8,949,081 het C>T	syn	**Mtlyz:** c.384C>T **MuTa:** Disease causing p=1 **CUD:** wt=20.8 mt=16.0	T13
8,949,104 het C>T	syn	**Mtlyz:** c.361C>T **MuTa:** Disease causing p=1 **CUD:** wt=39.6 mt=12.9	M30
8,949,165 het C>T	syn	**Mtlyz:** c.300C>T **MuTa:** Disease causing p=0.99 **CUD:** wt= 14.5 mt=11.0	M27

Syn:synonymous; nonsyn: nonsynonymous; Mtlyz: Mutalyzer; MuTa: MutationTaster; CUD: codon usage database; PP: PolyPhen; MP: MutPred

Besides the frameshift mutation identified in the family, four other variants possibly affecting the primary structure of the protein were identified/detected. Heterozygous *CARHSP1* c.401G>A (p.G134D) was found in meningioma samples M40 and M45. Figure 5.23 shows the chromatograms for the variant and the reference. Online tools predicted the mutation as damaging (Table 5.8). The remaining three variants were synonymous. They were analyzed via online Codon Usage Database to compare codon usage bias which means differences in the frequency of occurrence of synonymous codons. All mutant codons had lower frequencies as compared to the corresponding wildtype codons. Rare codons can possibly result in a misfolded protein (Spencer *et al.*, 2012). Chromatograms for the synonymous mutations are presented in Figure 5.24.

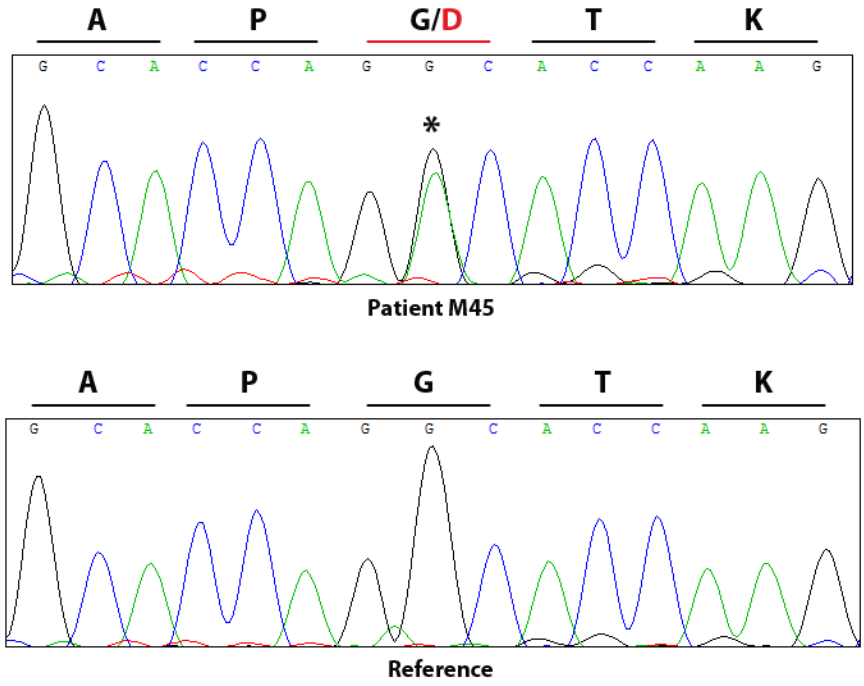

Figure 5.23. Chromatograms showing nonsynonymous *CARHSP1* c.401G>A (p.G134D).

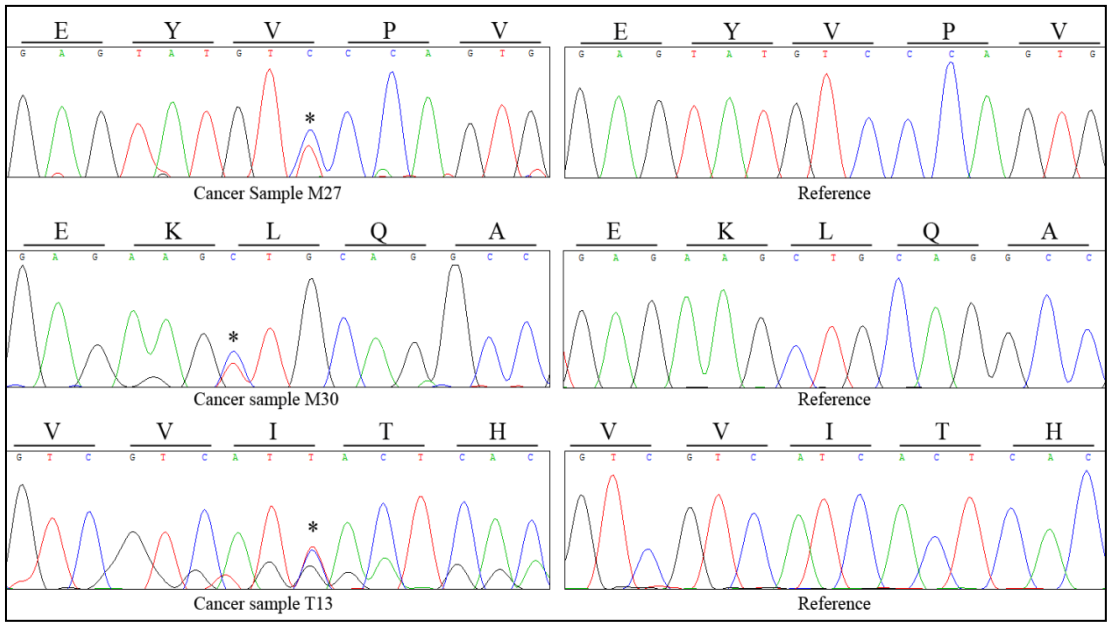

Figure 5.24. Chromatograms showing identified synonymous mutations in cancer samples.

CARHSP1 exons 2 and 3 were also screened by HRM analysis in the cancer samples, and no mutation was found.

In summary, five different novel or rare variants in *CARHSP1* coding regions that were predicted to possibly affect the protein product were identified. Additionally, one novel missense variant each in *RGS22* and *METTL22* were identified. *METTL22* variant segregated with head tumors whereas *RGS22* variant was found in two of the head tumors patients and a brain tumors patient.

5.5. Epilepsy Syndrome

DNA samples of 12 family members were subjected to genome scan, and exome sequencing was performed on patient 406. Multipoint parametric linkage analysis was performed with different parameters and different pedigree models due to phenotypic variations among the patients. (Exome sequence variants were investigated at all candidate loci detected in different analyses.)

5.5.1. Linkage Analysis

- Linkage Analysis Using the Actual Pedigree: Multipoint linkage analysis was performed using the actual pedigree, assuming recessive inheritance with 70% penetrance. Since the size of the pedigree exceeded the capacity of the program, multipoint LOD scores were calculated separately for the two branches of the family, and cumulative LOD scores were obtained. Linkage analysis yielded a maximal LOD score of 3.41 at 8p21.3, 3.33 at 12q24.23, and 3.27 at 6q25.1. Haplotypes were investigated by HCiE for those regions, but at no region the genotypes were homozygous IBD. Patients were heterozygous for the regions at 8p21.3 and 6q25.1, and 12q24.23 was a non-informative region with non-identical genotypes in the patients. Thus, no candidate locus was found.

- Linkage Analysis Including Only the Patients and Parents: Since the size of the family exceeded the capacity of the program, only the patients and their parents were included. Recessive autosomal inheritance, full penetrance and a disease frequency of 0.01 were assumed. Two loci yielded a maximal LOD score of 3.9,

and in total nine loci yielded scores >3. Haplotypes of the family members were investigated by HCiE for those regions, and haplotypes in all the regions were found not to be IBD in the patients; either the genotypes of the patients were heterozygous or the patient genotype was shared also by an unaffected sibling. In conclusion, no candidate locus was found.

- X-linked Linkage Analysis: Only the three male patients were included since the female patient has a milder phenotype that could be due to another disease or skewed X inactivation. In calculating multipoint LOD scores, X-linked recessive inheritance with full penetrance was assumed and no females except the mothers were included. A maximal LOD score of 1.49 was obtained at Xq22.1-24 between 103 and 115 cM, flanked by SNPs rs6621038 (100,559,427 bp) and rs2843588 (114,771,867 bp). Genotypes in this region were consistent with skewed X inactivation in the female patient, as she was heterozygous for the haplotype for which the male patients were all hemizygous and which was different from the unaffected brothers.

- Linkage Analysis that Included only the Male Patients: In the linkage analysis for calculating multipoint LOD scores in an autosomal recessive model with 70% penetrance, only the three male patients and the parents were included, considering that the female patient had a milder phenotype. A maximal LOD score of 3.33 at 3q29 and of 3.28 at 20p13 was obtained. Region 3q29 was eliminated by haplotype inspection since patient 407 was heterozygous, and the region at 20p13 from rs6051417 (2,802,023 bp) to rs3848815 (4,297,569 bp) was a good candidate since the male patients shared the same homozygosity. The female patient shared the homozygosity in the region rs11087565 (3,084,593 bp) - rs170974 (3,599,112 bp) but she had a different maternal allele than the affected siblings. In addition, unaffected siblings 403 and 404 also shared the genotype (both maternal and paternal haplotypes) with the patients.

- Linkage Analysis that did not Include Atypical Patient 401: The clinician hypothesized that patient 401, with an atypical phenotype as compared to the other patients in the family, could be having another disease; therefore, we

decided to perform a new linkage analysis that did not include him. Assuming autosomal recessive inheritance with full penetrance, the genotype data of the remaining 3 patients (406, 407, and 408), unaffected brother 403 and parents were used. The highest maximal LOD score of 3.33 was obtained at 21p11.1, with homozygosity from rs9647078 (17,518,797 bp) to rs377240 (18,258,451 bp). Three patients carried the same homozygosity for about a 130-Kb region from rs2823997 (18,091,248 bp) to rs2178914 (18,221,003). However, this region was excluded because patient 406 carried the same genotype with unaffected sibling 402. The next highest maximal LOD score 2.87 was obtained at 21q22.11, from rs926755 (36,134,690 bp) to rs9610448 (36,608,205 bp); however, this region was also eliminated since the genotypes of patients 405 and 406 were not IBD, and also patient 406 carried the same genotype with unaffected sibling 402. Another high maximal LOD score (2.65) was at 2q33.1 with homozygosity between rs10200021 (196,414,587 bp) and rs2697298 (198,081,584 bp). This region was also excluded because patient 406 carried the same genotype with unaffected sibling 402. In summary, no candidate region was found with linkage analysis including only 3 of the patients (406, 407, and 408) and an unaffected sibling (403).

- X-linked Linkage and Exome Sequence Analyses that did not Include Atypical Patient 401: Since skewed X-inactivation could explain the milder phenotype of the female patient 408 and atypical patient 401 could be having another disease, X-linked multipoint linkage analysis was performed by not including those patients but including the mothers, fathers, remaining 2 patients (406 and 407) and unaffected brother 403.

A maximal LOD score of 1.8 was obtained at Xq26.2-27.3, with shared hemizygosity from rs858618 (131,820,147 bp) to rs12836143 (144,764,370 bp). In this region, male patients 406 and 407 shared the same haplotype which was different than those of the other male relatives. Female patient 408 was heterozygous for the haplotype.

5.5.2. Exome Sequence Analysis

Exome sequence output was filtered for the candidate loci as described in section 4.2.8. Totally ten variants were subjected to Sanger sequencing and none were validated. Then, we suspected that exome sequence data were not for patient 406 but for another individual. We compared the homozygous regions revealed by the SNP genotype data of the patient to those indicated by exome sequence data of two samples which were sent together to the company for exome sequencing. Indeed, there was a mistake in labeling the samples, and the analyzed genome data in fact belonged to another sample. That is why none of the all selected variants could be validated by Sanger sequencing. Analysis of true exome sequence data is one of the future prospects.

In the exome sequence output, one of the selected variants was a nonsynonymous variant in *ESX1*. When we were analyzing the sequencing results of the *ESX1* variants, we realized that there were extra sequences at the very end of the chromatogram. We inspected the reference sequence around the amplified region and saw that the 27-nucleotide sequence "GGGTGGCAGAGGCGCCATGGGCGGCCC" was repeated 12 times with slight sequence differences. We decided to design primers that would facilitate the amplification of the entire repeated region to investigate any duplications and insertions. We found that the 27 nucleotide sequence was repeated 13 times in patient 406. DNA samples of all family members and 10 population controls were amplified at the region, and the PCR products were resolved on a 2% NuSieve agarose gel to estimate the lengths (Figure 5.25). The novel allele with 13 repeats turned out to be just a polymorphism; some family members such as unaffected male 405 and some controls were carrying it. DNA sample of this unaffected male had not been subjected to SNP genome scan; we decided to perform genome scan for this individual also, since his genotype could eliminate or narrow down this candidate region as well as some others. Indeed, his genotype eliminated this region, because he shared the same hemizygous genotype as the male patients.

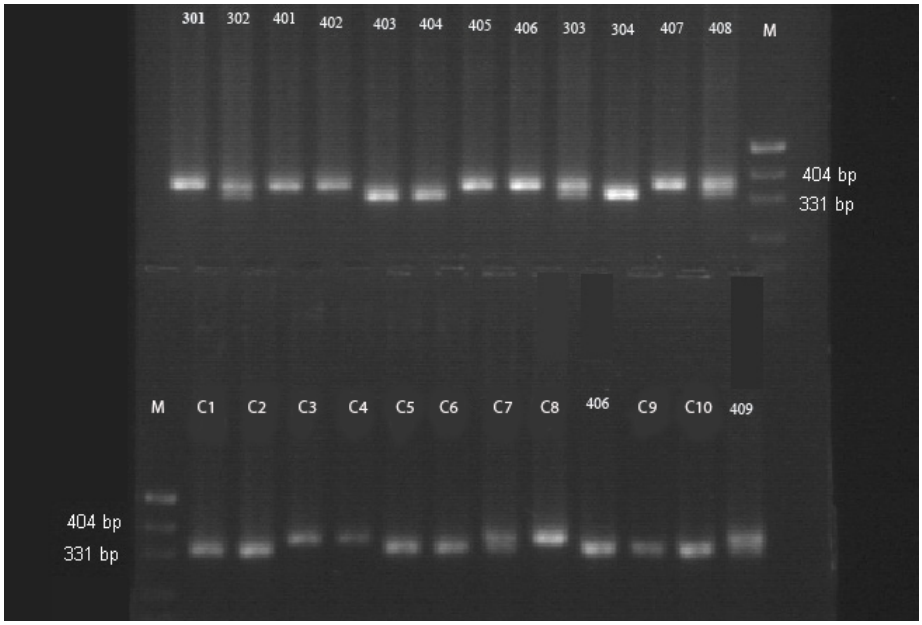

Figure 5.25. PCR products of the *ESX1* region resolved on a 2% NuSieve agarose gel. The lengths of the expected bands are 336 and 363 bp. M is length marker, and samples C1-C10 are control samples.

5.5.3. Excluding Fragile X Syndrome for Patient 406

Fragile X syndrome is an X-linked syndrome characterized by moderate to severe mental retardation, macroorchidism, and distinct facial features that include long face, large ears, and prominent jaw (MIM 300624). The clinician hypothesized that patient 406 could have Fragile X syndrome rather than the disease the other patients have, since his phenotype is very severe.

Correct exome sequence output variants of the patient were searched for novel or rare variants in *FMR1* and *AFF2*, the genes associated with Fragile X syndrome. No candidate variant was found in either gene.

Fragile X syndrome is caused also by expanded CGG trinucleotide repeats in the 5'UTR of *FMR1* (Kremer *et al*., 1991). Higher than 200 repeats causes the vast majority of the cases whereas repeats from 55 to 200 are referred to as 'premutations' (MIM 300624). As exome sequencing did not cover the repeat region site, DNA samples of

patient 406 and his mother 302 were sent to Dr. Hülya Kayserili's laboratory at Istanbul Medical School, and the repeat size was found within the normal range.

5.5.4. Homozygosity Mapping using Online Tool Homozygosity Mapper

Online tool Homozygosity Mapper was used to find shared homozygosity regions in the patients. Analysis was performed in two ways. In the first analysis, all patients were considered as affected. In the second analysis, all patients except 401 were considered as affected since he has an atypical phenotype. In both analyses, parents were considered as controls. No limit was set for maximal and minimal lengths of homozygous blocks, not to miss any candidate regions. Also, homozygous regions in parents were not excluded not to miss any noninformative regions. Haplotypes were investigated in the regions found by those analyses, and some of the regions were excluded since they were not IBD. Table 5.9 shows the regions that were found using Homozygosity Mapper and not excluded by haplotype inspection.

Table 5.9. Homozygous regions in patients as detected by Homozygosity Mapper.

Chr	Flanking SNPs	Size (Mb)	Remarks
A. Analysis 1: Four patients			
2	rs6719550 (188,272,460) - rs2697307 (198,075,465)	9.8	The region has160 homozygous subregions due to interruptions in 406. Nonetheless, the exome data in the total region were investigated.
B. Analysis 2: Three patients (401 is not included)			
2	rs10200021 (196,414,587) - rs2697298 (198,081,584)	1.66	Patient 406 carries the same genotype with unaffected sister 402.
6	rs11758337 (28,260,443) - rs7772289 (28,674,322)	0.41	The region is interrupted by three heterozygous SNPs in patient 406. Patient 401 is heterozygous. Genotypes of patients 407 and 408 are IBD.
3	rs1461131 (116,000,672) - rs7649435 (116,353,532)	0.35	Not IBD.

Also, when we investigated the candidate region at 2q33.1, we realized that the region from rs1157699 (188,258,904 bp) to rs10210380 (193,031,408 bp) would be IBD only if 406 is considered not having the same disease. Since he shared a haplotype with the other patients, we expected that he carried the mutation in the heterozygous state.

Morbid/Disease map did not list in those regions any diseases manifesting with mental retardation or epilepsy. Also, online tool GeneDistiller was used to find candidate genes with neuronal phenotype.

Future prospects are the analysis of true exome sequence output for the candidate loci and the analysis of the regions with incomplete coverage. Since there are many such regions, we can focus only on those that we assess having potential relation to epilepsy, mental retardation or neurological phenotype.

5.5.5. Mitochondrial Variant Analysis

Mitochondrial inheritance and mitochondrial mutations have been related to epilepsy. Such mutations are in several genes, including *POLG, MTTK, MTTL1, MTTH, MTTS1, MTTS2,* and *MTTF,* defects in which cause myoclonic epilepsy associated with ragged-red fibers (MERRF) syndrome (Bindoff and Engelsen, 2012). In mitochondrial inheritance in humans, the mother and all her children could be affected since mitochondria are passed to offspring only from the mother, i.e. maternal inheritance. However, in most cases, not the mother and not all children are affected due to heteroplasmy in the mother. Mitochondrial genome was not included in the annotated exome sequence excel file; thus, mitochondrial variants were detected using IGV program by inspection and annotated manually. Novel or rare (MAF < 0.01) variants that were assessed to potentially affect the protein (except synonymous variants) or the non-coding RNA were selected (not presented).

One candidate variant in patient 406 was subjected to Sanger sequencing, and the variant was found to be a false call. Then, we realize that the exome sequence data

belonged to another sample. Therefore, analysis of mitochondrial variants in the correct exome sequence data is one of the future prospects.

5.5.6. Deletion and Duplication Analysis

cnvPartition program in GenomeStudio platform was used to investigate any deletions or duplications using SNP genotyping data of all family members. All candidate regions mentioned in the previous sections were investigated for any deletions or duplications that segregated with the trait, but no such variants were found.

5.6. Muscular Dystrophy

Genome scan data for the mother, three unaffected sibs and three affected sibs were subjected to linkage analysis, and candidate loci were detected. Exome sequencing was performed in one of the affected children. There was a very strong candidate gene at the largest candidate locus; however, no candidate variant was found by exome sequence analysis, but some of the exons were not fully covered. So, those regions and also all coding exons of the longest transcript were subjected to Sanger sequencing. A novel synonymous mutation in *LMNA* was identified.

5.6.1. Linkage Analysis

When we first started to study the family, affection status of individual 405 was not definite. So, linkage analysis was performed assuming him unaffected and individuals 402 and 403 affected. Multipoint LOD scores were calculated under the assumption of autosomal recessive inheritance with 80% penetrance or assuming an X-linked recessive model (Results not shown).

Later the clinician informed us that disease symptoms began in individual 405, and a new linkage analysis was performed assuming him affected. The youngest

member of the family 406 was not included in the analysis, considering that she could be too young for disease onset. A maximal LOD score of 2.65 was obtained at two loci, and two other loci yielded maximal LOD scores 2.58 and 2.46 (Figure 5.26). Linkage analysis was performed also for the X-chromosome, including only the mother and sons. Table 10 presents the candidate loci for MD family.

Haplotypes were investigated by HCiE and constructed using Allegro program. Candidate regions and haplotype analysis results are presented in Table 5.16. Haplotypes at 1q22-23.2, 10p15.2-0p14, 9p22.2 and 2p16.2 are presented in figures 5.27, 5.28, 5.29 and 5.30, respectively. In haplotype drawing, markers were selected in various cM-intervals. This is the reason for the discrepancy between the results in Table 5.10 and figures 5.27-5.30. The sizes of the parental disease haplotypes were 75.2, 31.9, 0.5, and 0.38 Mb.

Table 5.10. Linkage analysis results for MD family assuming 405 as affected.

Locus	**LOD score**	**Flanking SNPs (bp)**	**Size (Mb)**	**Remarks**
Autosomal recessive inheritance model				
1q22-23.2	2.65	rs2335406 (154,786,686) - rs6690595 (163,815,477)	9.02	Possibly IBD with a crossover in 403 around 155 Mb. Sister 406 shares the genotype up to 160.4 Mb (rs2369611).
10p15.2-p14	2.65	rs12217962 (3,954,953) - rs3740204 (11,505,175)	7.55	Possibly IBD with crossovers in 402 around 3.9 Mb and in 405 around 11.5 Mb. Sister 406 shares the genotype.
9p22.2	2.58	rs1487509 (16,986,415) - rs2383022 (17,485,598)	0.5	IBD. Sister 406 shares the same homozygosity with the patients.
2p16.2	2.46	rs11898505 (54,684,557) - rs2116435 (55,072,474)	0.38	IBD. Sister 406 is heterozygous.
X-linked inheritance model				
Xp22.11-p11.4	0.89	rs5926470 (26,985,616) - rs10127045 (40,745,226)	13.76	Patients share the same hemizygosity which is different from their unaffected brother 401.

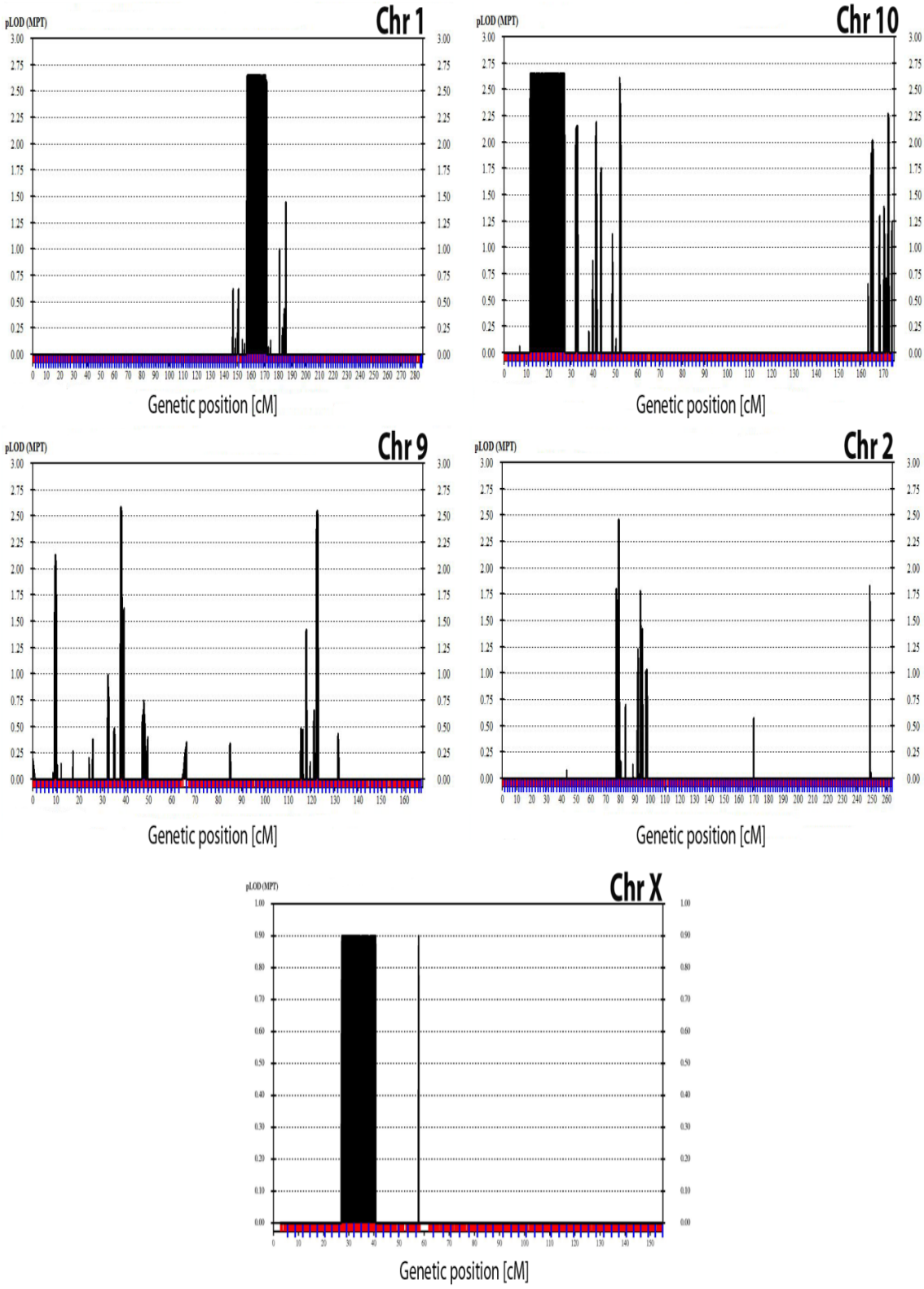

Figure 5.26. Multipoint linkage analysis results for MD family for autosomes and PARs yielding maximal LOD scores >2 and the X-chromosome.

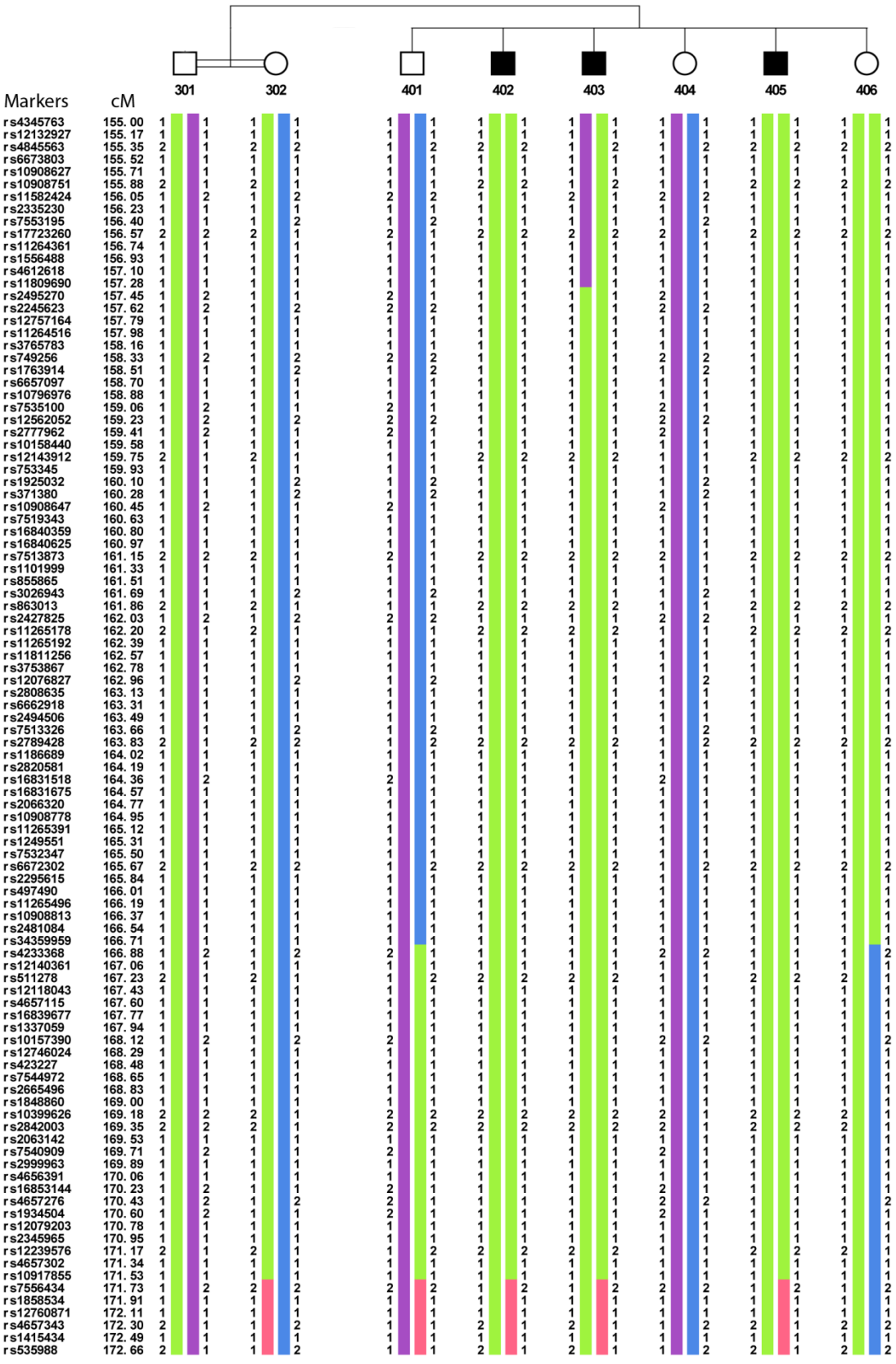

Figure 5.27. Haplotypes for MD family at 1q22-23.2.

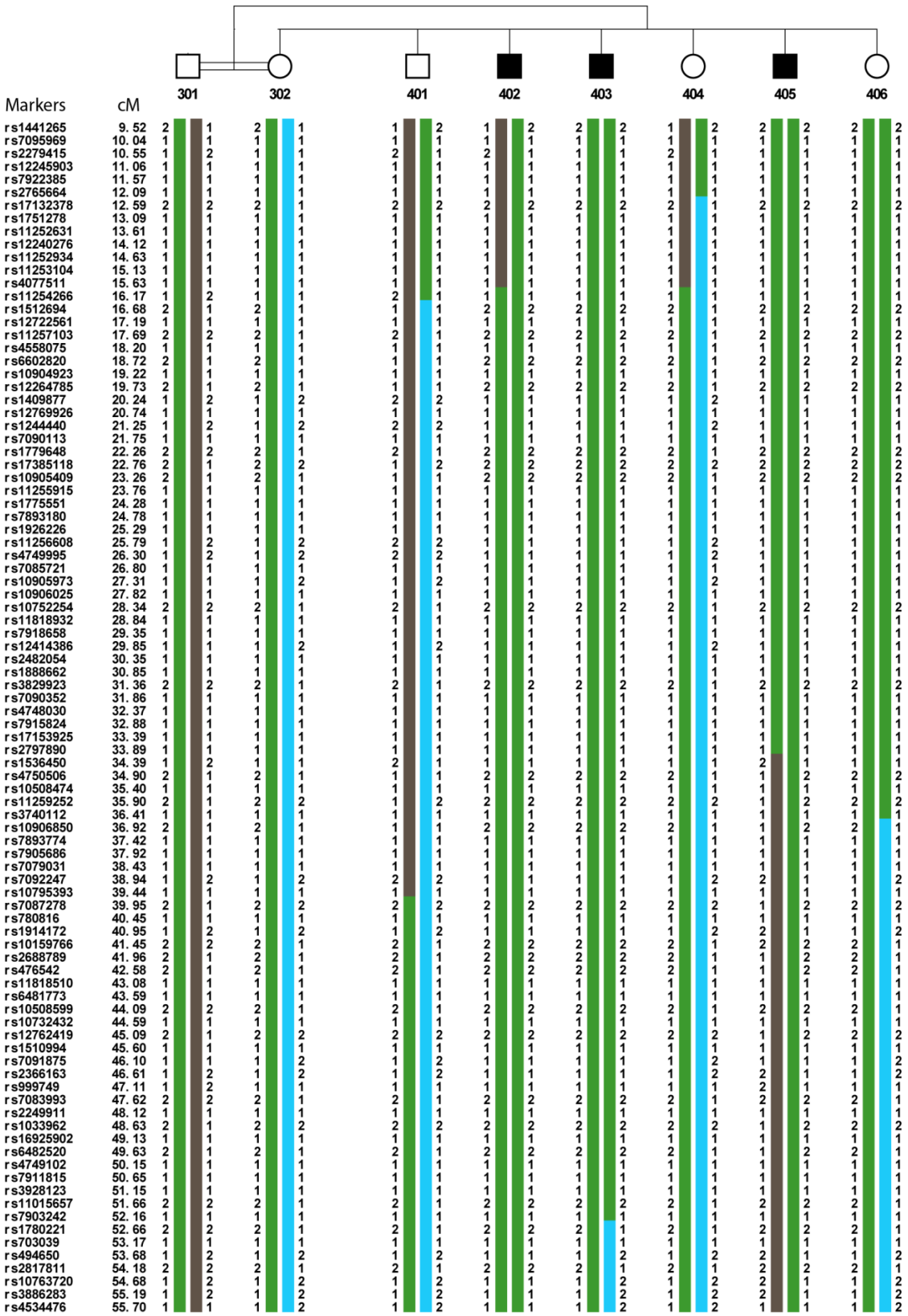

Figure 5.28. Haplotypes for MD family at 10p15.2-p14.

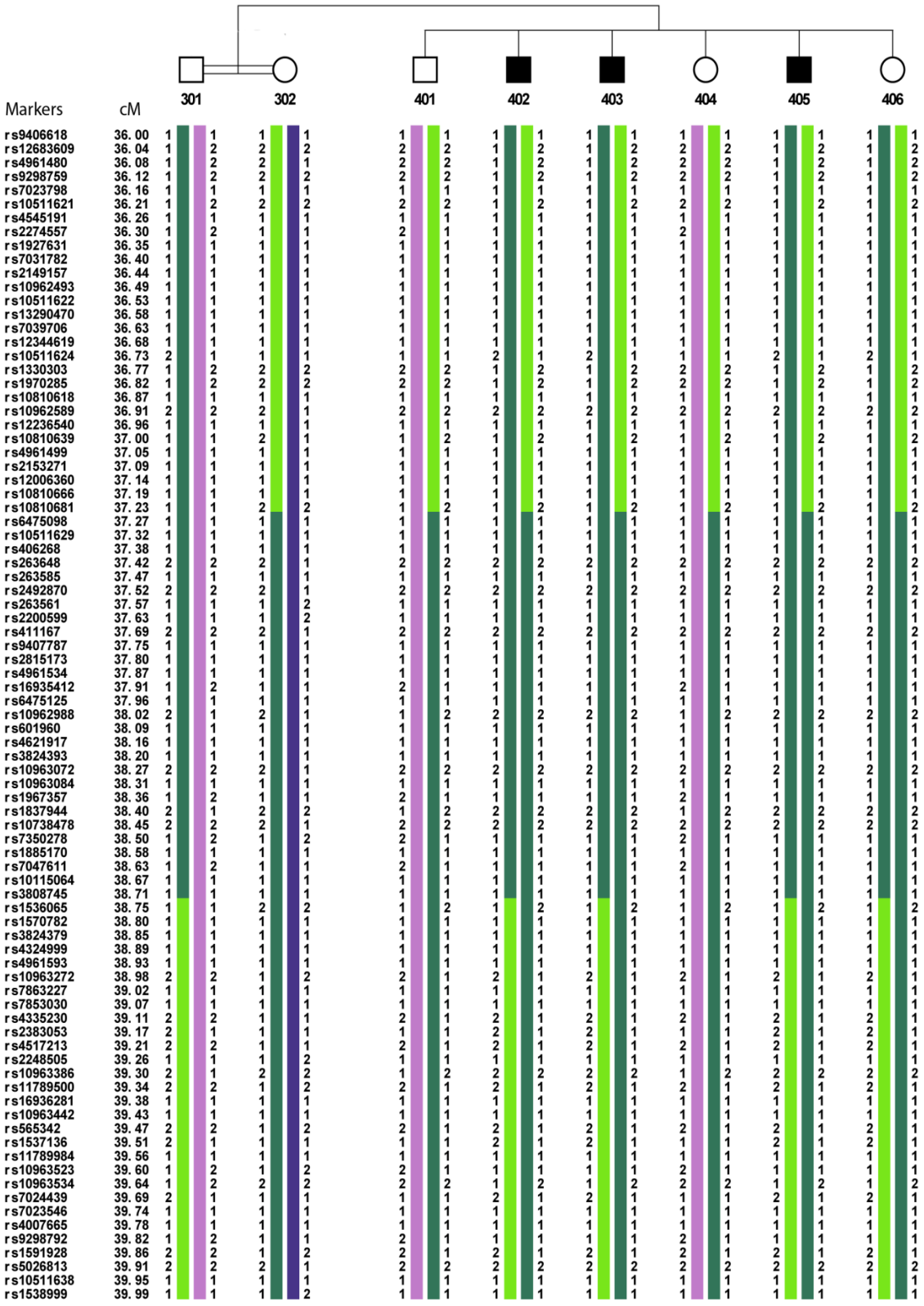

Figure 5.29. Haplotypes for MD family at 9p22.2.

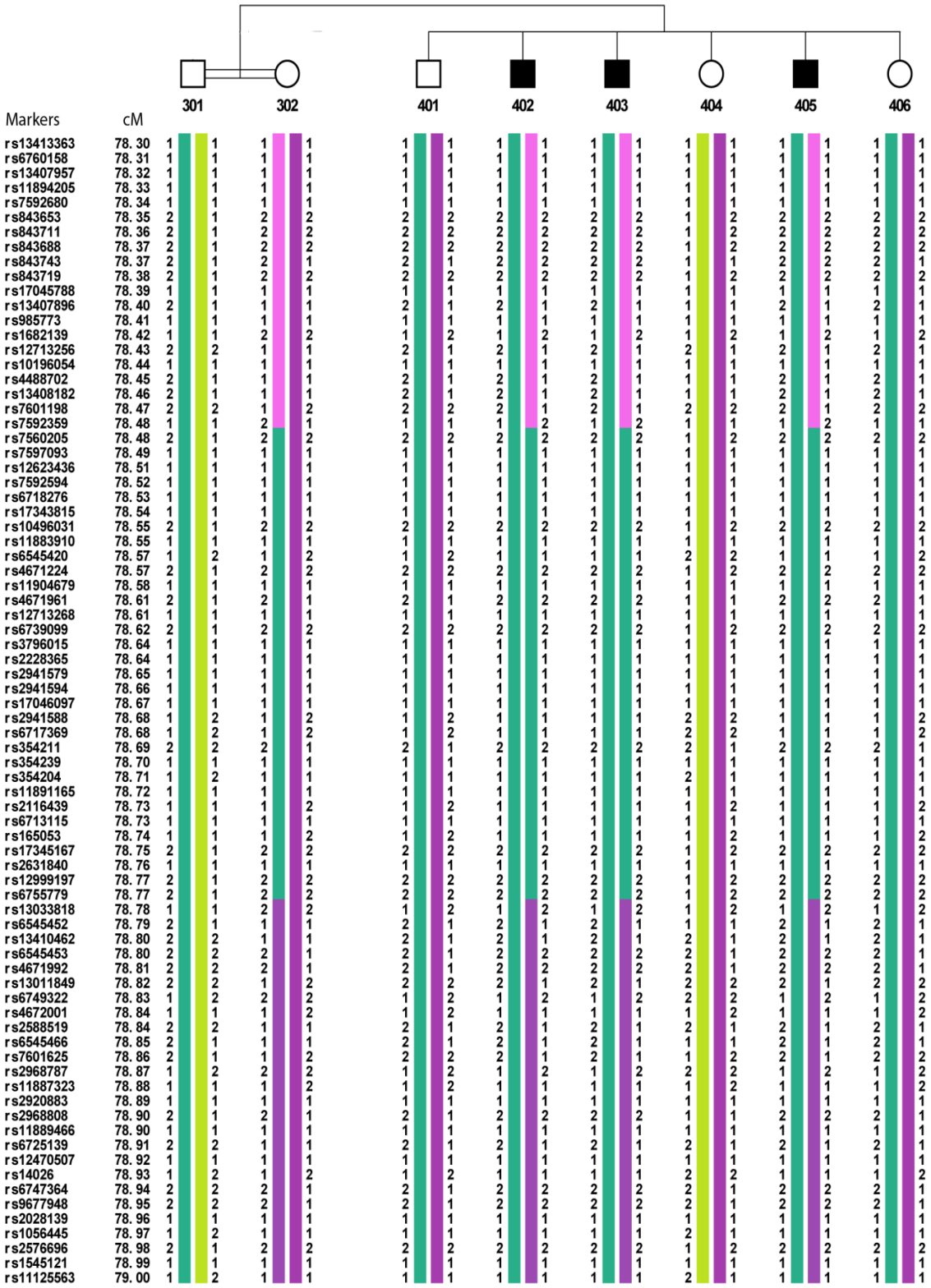

Figure 5.30. Haplotypes for MD family at 2p16.2.

The candidate regions were searched for candidate genes with the help of program GeneDistiller and Morbid/Disease maps by selecting genes with muscular phenotypes. Morbid/Disease lists with muscular phenotypes and candidate genes assuming 405 also affected are presented in Table 5.11.

Table 5.11. Candidate regions, Morbid/Disease maps and candidate genes with muscular phenotypes.

Locus	**Diseases (Morbid)**	**Candidate Genes (GeneDistiller)**
1q22-23.2	Emery-Dreifuss muscular dystrophy (MIM 181350), Muscular dystrophy, congenital (MIM 613205), Muscular dystrophy, limb-girdle, type 1B (MIM 159001)	*SHC1, EFNA1, LMNA, MEF2D, NTRK1, ATP1A2, CASQ1, HSD17B7*
10p15.2-p14	No related diseases.	*PRKCQ*
9p22.2	No related diseases.	No candidate genes.
2p16.2	No related diseases.	*SPTBN1*
Xp22.11-p11.4	Duchenne muscular dystrophy (MIM 310200); Becker muscular dystrophy (MIM 300376); Cardiomyopathy, dilated, 3B (MIM 302045)	*DMD, CYBB, ATP6AP2*

At 1q22-23.2, the largest locus by Mb, the region was narrowed down by a crossover in 403 at rs2335406 (154,786,686 bp). Individual 406, the youngest member of the family, was not included in the linkage analysis, and the finding that she was in part homozygous in the region was disregarded, since her affection status was unknown.

5.6.2. Exome Sequence Analysis

DNA sample of individual 402 was subjected to exome sequencing. Alignment of the reads and variant calling were performed in our laboratory.

Since individual 405 was assumed unaffected at the beginning of the study, exome sequence data were first analyzed at candidate loci 10p14-p12.2 and 17q21.1-

21.32 that yielded the same significant LOD score of 2.18, assuming 405 was unaffected. Exome sequence output was filtered as described in section 4.2.8. Some of the filtered variants were validated by Sanger sequencing (Table 5.12).

Table 5.12. Novel/rare exonic/splicing (except synonymous) validated variants at loci that yielded significant LOD scores, assuming 405 unaffected.

Location	**Ref base/ Alt base**	**Hom/ Het**	**Qual. score**	**Total /Alt depth**	**dbSNP131 MAF/EVS MAF**	**Gene**	**Change**	**Seq. result**
10p14-p12.2, between rs10906734 (14,700,243) and rs2815572 (27,904,646)								
22,498,484	-/AGA	hom	5277	75/64	-/-	*EBLN1*	insertion	hom
25,31,3368	T/-	hom	2037	56/52	-/-	*THNSL1*	deletion	hom
17q21.1-21.32, between rs7222207 (38,955,173) and rs8079740 (48,697,481)								
39,021,137	G/A	het	29	50/49	rs61735165 0.001/0.0009	*KRT12*	nonsyn	hom
42,426,591	G/A	het	139	41/21	-/-	*GRN*	nonsyn	het
43,319,435	CCG/-	hom	2175	45/29	-/-	*FMNL1*	deletion	hom
48,245,011	C/T	het	24	121/120	-/-	*SGCA*	nonsyn	hom

Qual: quality, seq: sequencing; ref: reference; alt: alternative; hom: homozygous; het: heterozygous; nonsyn: nonsynonymous

The nonsynonymous variations in genes *KRT12*, *GRN* and *SGCA* were predicted to be damaging with a probability of 1 by online tool PolyPhen. Heterozygous mutations in *KRT12*, encoding Keratin 12, cause Meesmann corneal dystrophy (MIM 601687). Heterozygous mutations in *GRN*, encoding granulin precursor protein, cause frontotemporal lobar degeneration with ubiquitin-positive inclusions (MIM 607485) and primary progressive aphasia (MIM 607485) while homozygous mutations cause ceroidlipofuscinosis, neuronal, 11 (MIM 138945). Homozygous, heterozygous and compound heterozygous mutations in *SGCA*, encoding alpha sarcoglycan, cause Limb-girdle muscular dystrophy, type 2D (MIM 600119).

When the clinician reported disease symptoms in individual 405, a new linkage analysis was performed assuming him also affected and not including the youngest member of the family (406) because she may be too young for disease onset. So, exome sequence data were analyzed at candidate loci 1q22-23.2, 10p15.2-p14, 9p22.2, 2p16.2 and Xp22.11-p11.4 (Table 5.13).

Table 5.13. Novel or rare exonic/splicing (except synonymous) variants in MD patient in the candidate regions, assuming 405 also affected.

Genomic position in bp	**Ref base/ Alt base**	**Hom/ Het**	**Qual. score**	**Total depth/Alt depth**	**dbSNP131 MAF/EVS MAF**	**Gene**	**Change**
1q22-23.2, between rs2335406 (154,786,686) and rs6690595 (163,815,477)							
154,919,096	C/T	het	60	220/215	-/-	*PBXIP1*	nonsyn c.1054G>A, p.V352M
160,144,519	G/A	hom	225	46/46	rs148265962 NR/0.0000001	*ATP1A4*	nonsyn c.2293G>A, p.V765I
162,343,861	C/T	hom	85	282/282	rs142892903 0.0014/ 0.001	*C1orf111*	nonsyn c.763G>A, p.G255S
At 10p15.2-p14, between rs12217962 (3,954,953) and rs3740204 (11,505,175) No candidate variants listed.							
At 9p22.2, between rs1487509 (16,986,415) and rs2383022 (17,485,598) No candidate variants listed.							
At 2p16.2, between rs11898505 (54,684,557) and rs2116435 (55,072,474) No candidate variants listed.							
At Xp22.11-p11.4, between rs5926470 (26,985,616) and rs10127045 (40,745,226)							
38145525	TT/-	het	29	3/1	-/-	*RPGR*	frameshift deletion c.2726_2727del, p.909_909del

Qual: quality, ref: reference; alt: alternative; hom: homozygous; het: heterozygous; nonsyn: nonsynonymous; NA: not reported

Nonsynonymous variants listed in Table 5.13 were investigated via online tools Mutation Taster, PolyPhen and SIFT to possibly assess whether amino acid substitution is benign or damaging. Prediction results of those online tools are presented in Table 5.14.

Table 5.14. Prediction results of online tools for candidate nonsynonymous variants.

Variant	Mutation Taster	PolyPhen	SIFT
PBXIP1 c.1054G>A, p.V352M	Polymorphism with a prediction probability of 0.99	Benign with a score of 0.110	Tolerated, score = 0.18
ATP1A4 c.2293G>A, p.V765I	Disease causing with a prediction probability of 0.99	Benign with a score of 0.042	Damaging, score = 0 Low confidence prediction with Median conservation 4.28.
C1orf111 c.763G>A, p.G255S	Polymorphism with a prediction probability of 0.99	Benign with a score of 0.002	Damaging, score = 0 Low confidence prediction with Median conservation 4.32.

In summary, three nonsynonymous and one frameshift variants were filtered in exome sequence analysis in MD patient, but none was Sanger sequenced for validation. Future prospects include sequencing and investigating segregation of the validated variants in the family and their frequencies in the population.

5.6.3. Analysis of Candidate Gene *LMNA*

The clinician mentioned that the disease resembled a laminopathy. *LMNA*, coding for Lamin A/C and residing at the largest candidate locus 1q22-23.2, was assessed as a very strong candidate gene, but the exome sequence data did not present

any *LMNA* variants. However, exome sequencing included only the UTRs and coding exons of the longest transcript. *LMNA* encodes 11 different protein coding transcripts, most of them slightly differing at the 5'UTR or 3' UTR of the longest transcript, and three of the transcripts have completely different first exons or last exons as compared to the longest transcript. Thus, primers presented in tables 4.6 and 4.7 were designed for Sanger sequencing to amplify those regions that were not included in the exome sequencing. The extent of sequencing of *LMNA* is presented in Table 5.15.

Table 5.15. The extent of sequencing of *LMNA* in patient 402. "-" indicates upstream and "+" indicates downstream of the region. Completely sequenced introns are indicated.

Region	**Sequenced part (bp)**
Exon1	-185 and +72
Exon 2	-61 and +66
Exon 3	-81 and +276 (intron 3)
Exon 4	-276 (intron 3) and +71
Exon 5*	-100 and +87
Exon 6	-84 and +92 (intron 6)
Exon 7	-92 (intron 6) and +91
Exon 8	-62 and +84 (intron 8)
Exon 9	-84 (intron 8) and +56
Exon 10	-92 and +54
Exon 11	-113 and +324 (intron 11)
Exon 12	+324 (intron 11) and -34

*****Since exon 5 could not be optimized for Sanger sequencing, it was analyzed by exome data via IGV program.

By Sanger sequencing, a novel synonymous *LMNA* c.346C>T variant was identified in patient 402. Then, we used IGV program to inspect the variant in the exome data and verified it. DNA sample of individual 406 was subjected to Sanger sequencing, and she was found to carry the mutation in the homozygous state. Online tool Mutation Taster predicted the variant as disease causing with a prediction probability of 1 since donor site was predicted to get stronger (wild type 0.4401, mutant 0.4585). The variant was not found in EVS and TEVS. Chromatograms showing the reference sequence and

the mutation in patient 402 are presented in Figure 5.31. Online Codon Usage Database was used to compare codon usage bias, and the mutant codon had a lower frequency than the wildtype codon (wildtype: 39.6, mutant: 12.9). Additionally, on online database ENCODE the region of 200 bases covering the mutation site is designated as the active promoter site as well as a regulatory region marked by histone acetylation H3K27Ac which is thought to enhance transcription possibly by blocking the spread of the repressive histone mark H3K27Me3.

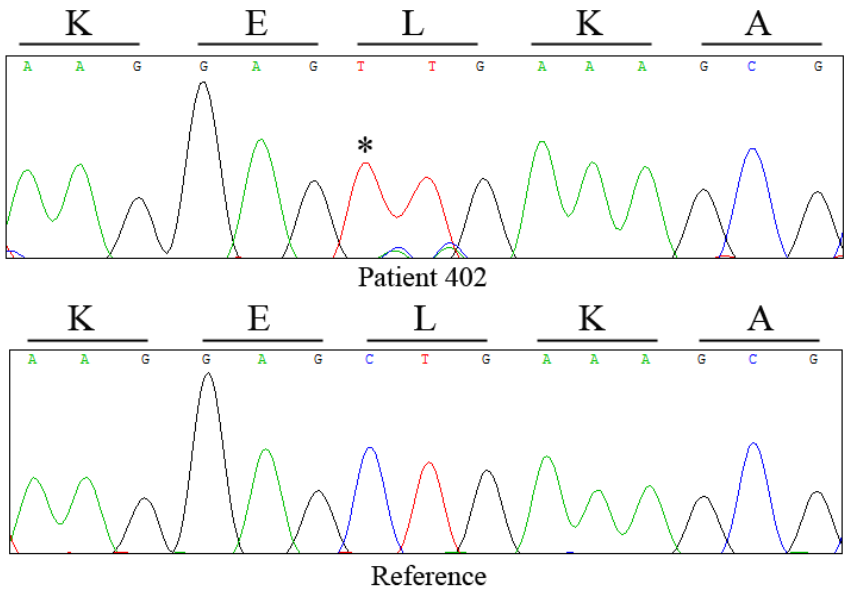

Figure 5.31. Chromatograms showing *LMNA* c.346C>T.

In conclusion, a novel synonymous *LMNA* c.346C>T mutation was identified. Future prospects include investigation of the segregation of the variant in the family and the investigation of whether the variant (synonymous) changes the expression or splicing of the gene.

5.6.4. Deletion and Duplication Analysis

cnvPartition program in GenomeStudio platform was used to investigate any deletions or duplications using the SNP genotyping data for all family members. All candidate regions found by assuming 405 affected were investigated to find deletions or duplications that segregated with the trait, and no such variant was found.

6. DISCUSSION

6.1. Congenital Disorder of Glycosylation (CDG)

We studied a consanguineous family that had two affected sons with initial diagnosis of progressive ataxic gait and visual impairment which began in early childhood, and the older son had a more severe phenotype. The disease was initially assessed as a novel neurologic disorder. Aiming to identify the disease loci/locus, SNP genome scan and subsequent linkage analysis were performed. Four candidate loci were found, and homozygous stopgain *SRD5A3* c.57G>A (p.W19X) variant detected by exome sequencing was assessed as the disease mutation, due to its harmful effect to the protein product and the relevance of the function of the gene to the clinical phenotype. The mutation converts codon 19 (tryptophan) to a premature termination codon that is deduced to lead to the deletion of 300 amino acids from the native 318-amino acid protein, resulting in a truncated protein. The mutation segregated with the disease in the family. Fifty-six unrelated control samples were tested for it, and none of them carried it; however, it was reported in three unrelated children afflicted with SRD5A3-CDG (Assmann *et al.,* 2001; Prietsch 2002; Gründahl *et al.,* 2012).

SRD5A3 encodes steroid 5-alpha-reductase 3 that converts the alpha-isoprene unit of polyprenol to dolichol in humans. Dolichols function as membrane anchors for the formation of the oligosaccharides in the ER membrane and have a very important role in N-glycosylation pathway and protein folding (Cantagrel *et al.*, 2010). Congenital disorders of glycosylation (CDG) are caused by defects in protein glycosylation, and CDG type Iq (SRD5A3-CDG; MIM 611715) is caused by defects in *SRD5A3.*

SRD5A3-CDG has a very broad phenotype, but the most prominent clinical symptoms are severe early visual impairment and variable ocular anomalies as well as some degree of intellectual disability. Cerebellar ataxia, the initial diagnosis of the patients in the study family, is one of the other clinical features of the disease. The diagnosis of SRD5A3-CDG in the study family was verified at Istanbul Medical School

with the abnormal serum transferrin isoelectric focusing pattern in the older brother. All reported patients afflicted with SRD5A3-CDG were children in contrast to our patients, who were 38 and 40 years old. Additionally, our patients had a specific form retinitis pigmentosa, bone spicule pigmentation, which results from the migration of pigment-containing cells to perivascular sites in the inner retina. Moreover, congenital colobomas and glaucoma which are two of the leading clinical pathologies in SRD5A3-CDG were not observed in our patients. Thus, our study widens the spectrum of clinical manifestations of SRD5A3-CDG. Our manuscript reporting an adult phenotype and further phenotypic variability in SRD5A3-CDG was published online in BMC Medical Genetics, with the first two authors contributing equally (Kara and Ayhan *et al.*, 2014).

6.2. Myosin Storage Myopathy (MSM)

Onengut et al. (2004) reported mapping an autosomal recessive MSM in a family to 3p22.2-p.21.32 using microsatellite markers for linkage analysis. In the present study, the same family was studied, and SNPs were used as markers for linkage analysis to validate the locus or search for another locus that was possibly missed using microsatellite genotyping data. One of the seven candidate loci detected included *MYH7*, defects in which are known to be responsible for MSM and cardiomyopathy (MIM 160760). Therefore, *MYH7* was assessed as the strongest candidate gene.

Various myopathies can arise due to *MYH7* mutations: Laing distal myopathy (mutations in exons 32–38; MIM 160500), familial hypertrophic cardiomyopathy (mutations in several exons; MIM 192600), dilated cardiomyopathy (mutations in several exons; MIM 613426) and myosin storage myopathy (mutations in exons 37-40; MIM 608358). Because MSM is caused by mutations in the last four exons (Tajsharghi and Oldfors, 2013), those exons were subjected to Sanger sequencing in one of the patients. Novel homozygous nonsynonymous mutation *MYH7* c.5458C>T (p.R1820W) in exon 37 was identified.

c.5458C>T in *MYH7* substitutes the positively charged, hydrophilic arginine residue with non-polar, hydrophobic tryptophan. This change is predicted as damaging

via all four online tools employed, PolyPhen, Mutation Taster, SIFT and MutPred. Family members and 120 population control samples were screened for the mutation by HRM curve analysis to distinguish a sequence variant with a power of 80% (Collins and Schwartz, 2002). Additionally, the mutation was not found in the Turkish EVS (in 353 unrelated persons) and EVS databases which corresponds to at least 95% power. The results confirmed recessive inheritance, as the patients were homozygous and the sisters were not and the mother was heterozygous.

MYH7 is a 1935-amino acid protein encoded in the 38 of the total 40 exons of the gene. It consists of two globular heads attached to a long alpha-helical coiled-coil (the myosin rod). The head region has roles in ATPase and actin binding properties of myosin whereas the rod region is responsible for the assembly of the filaments (Armel and Leinwand, 2009). Exons 30 to 40 encode the light meromyosin (LMM) domain of the protein. The myosin rod (long alpha-helical coiled-coil region) has a high content of LMM which has a very important role in the assembly of myosin into filaments (Sohn *et al*., 1997; Blair et *al*., 2002). Sixty per cent of the reported mutations lie on the head region, and those mutations were shown to result in cardiomyopathy (Morimoto, 2008).

All reported *MYH7* mutations that cause MSM are heterozygous except for the homozygous p.E1883K (Tajsharghi et al., 2007). The recessive form resulted in MSM with hypertrophic cardiomyopathy and respiratory failure. Our study MSM family is the second one with recessive *MYH7* mutation. The reason for recessive effect is possibly that the protein is not able to assemble into filaments because of the altered rod domain and does not result in gain of toxic function that would lead to a dominant effect. Moreover, one our patients has an unusual combination of dilated cardiomyopathy and MSM with scapuloperoneal distribution. These findings widen the spectrum of clinical phenotype of MSM, having MSM associated with dilated cardiomyopathy but not hypertrophic cardiomyopathy. Tajsharghi *et al.* suggested that cardiomyopathy and type II respiratory failure were due to recessive inheritance, since all previously reported MSM patients were heterozygous for *MYH7* mutations and did not have these clinical features. Our manuscript reporting MSM associated with dilated cardiomyopathy caused by a recessive *MYH7* mutation has been submitted for publication and is under review.

6.3. Desmin Deficiency Myopathy (DDM)

We studied a consanguineous family afflicted with severe fatigable myopathy with infantile onset, generalized weakness that extending to facial and extraocular muscles, cardiomyopathy and respiratory involvement. Since there was no similar disease at the time, we searched for the causative gene in the family. Linkage analysis pointed out to *DES* (*Desmin*) as the strongest candidate gene, as it resided at a candidate locus. We consulted the clinicians whether the disease could be caused by *DES* mutation, but they said it could not. The presence of a large number of genes (196 in total) prompted us to launch exome sequencing. Evaluation of the exome sequencing results showed three homozygous variants as candidate mutations to underlie the disease in the family. Two of those, missense c.4699G>A (p.E1567K) in *CCDC108* (Coiled-coil domain containing protein 108) and splice variant c.1546+1G>A in *SP100* (Nuclear body protein SP100), were in genes that are expressed in most tissues except for muscle (http://www.ncbi.nlm.nih.gov/unigene/, EST profiles). Online tools predicted *CCDC108* c.4699G>A as benign and *SP100* c.1546+1G>A as disease causing. The remaining variant was frameshift c.345dup (p.N116Qfs*2) in exon 1 of *DES*, encoding a muscle protein responsible for various muscle diseases (MIM 125660). This deleterious mutation was assessed as the plausible disease mutation due to its harmful effect on the protein product. The variant segregated with the trait in the family and was not found in the control samples.

The mutation is deduced to lead to a truncated protein, since it shifted the translational reading frame at codon 116 and created a premature termination codon after the synthesis of one nonnative amino acid. Thus, only the first 115 native residues of the 470-amino acid protein would be included in the truncated protein. The mutant mRNA is expected to undergo nonsense-mediated decay, the premature stop codon being located more than 50 nucleotides upstream of the 3′-most exon–exon junction (Nagy and Maquat, 1998); thus, the mutant protein is likely not synthesized at all. Consistent with this hypothesis, relative quantification of *DES* mRNA showed that the *DES* transcript level was about 0.3% in the patient as compared to the normal control. Additionally, immunohistochemical analysis using an antibody against desmin in one of

the patient's muscle biopsy sample (performed in Istanbul Medical School) validated the desmin null phenotype.

Desmin is a major intermediate filament type 3 and important for the tensile strength and maintenance of sarcomere in cardiac, skeletal and smooth muscles. *DES* defects cause various myopathies and are mostly inherited in an autosomal dominant fashion or arise by de-novo mutations (MIM 125660). However, there are a few cases with recessive inheritance: p.R173_E179del in two unrelated patients (Muñoz-Mármol *et al.,* 1998; Piñol-Ripoll *et al.,* 2009), splice site mutation c.1289-2A>G in two siblings (Cetin *et al.,* 2013) and compound heterozygous null mutations p.T76fs*21 and p.E108X in two siblings with fatigability and mitochondrial dysfunction (Henderson *et al.,* 2013). The patients carrying the first two mutations exhibited infantile onset cardiomyopathy and severe myopathy with classical desmin accumulation in the muscle whereas the siblings with the compound heterozygous mutations had a similar phenotype to our patients but with mitochondrial enzyme abnormality. The phenotypic difference between our patients and the other patients carrying recessive *DES* mutations could be due to desmin accumulation in the other patients but not ours. Additionally, desmin-null mice were viable and fertile and their phenotype was very similar to our patients, displaying exercise intolerance, skeletal myopathy and cardiomyopathy (Milner *et al.,* 1996).

In conclusion, we identified a frameshift mutation that resulted in a desmin-null phenotype and lead to a novel autosomal recessive desmin deficiency myopathy with fatigability. Our manuscript is submitted for publication and is under review.

6.4. Hereditary Head Tumors

A non-consanguineous family afflicted with various types of head tumors were studied. Tumors were meningioma in the second generation and glioma in the third generation. 16p13.3-p13.13 yielding the highest maximal LOD score of 3.01 was assessed as a candidate region harboring the disease gene previously in our laboratory (Bozoğlu, 2008). We re-performed the linkage analysis and obtained similar results.

DNA sample of one of the patients was subjected to exome sequence analysis within the scope of this thesis. Three candidate variants were listed in exome sequencing. Novel frameshift *CARHSP1* c.412_413insT (p.E138Vfs*93) was considered as the candidate mutation due to its harmful effect on the protein. The mutation is deduced to lead to the replacement of the last 10 native amino acids of the 147-residue protein with 92 non-native amino acids. Additionally, EVS was investigated for all types of variants in the gene. There was no splicing, nonsense or frameshift variants while there were totally 26 missense variants that were observed in the heterozygous state in 1 to 6 individuals in a total of 6495 samples. Considering the full conservation of the protein sequence between human and chimpanzee and low variation in humans, the frameshift *CARHSP1* mutation is likely pathogenic and thus a strong candidate.

We investigated whether the mutation segregated with head cancer susceptibility in the family. Unfortunately, DNA samples of patients 301 and 306 were not available for the analysis. In the HRM analysis results, patients 207 and 209 and unaffected individual 313 were in the same group while patient 310 was grouped with the other unaffected members of the family. When we investigated the haplotypes in the region, we found that unaffected individual 313 shared a haplotype with patients 207 and 209 but not the haplotype shared by patients 207, 209 and 310 (Figure 5.13). Since the onset of the disease is late, the affection status of individual 313 is uncertain. Moreover, since patients 207 and 209 were afflicted with meningioma and patient 310 was afflicted with glioma, the mutation could be the causative mutation for meningioma only.

We identified one nonsynonymous and three synonymous variants, all in exon 4, in cancer samples. Nonsynonymous c.401 G>A (p.G134D) was predicted as damaging by online tools employed. Additionally, the mutated residue G134 was fully conserved in vertebrates investigated and is within a stretch of 116 amino acids that are fully conserved across mammals. The synonymous variants were investigated via codon usage database for codon bias, and all mutant codons had lower frequencies of occurrence as compared to the wildtype codons. All evidence taken together, *CARHSP1* mutations could be causative for head tumors.

We also analyzed region 8q12.1-23.1 that yielded the second highest maximal LOD score and identified missense mutation *RGS22* c.578A>G (p.Y193C) via exome sequencing. The mutation was predicted as damaging by three of the four online tools employed. Mutated residue Y193 was fully conserved across mammals whereas the entire 1264 amino acid protein is 99.6% conserved between human and chimpanzee. EVS reported 7 splicing, nonsense or exonic frameshift variants in the gene; all were observed only once in the total 4080 samples and all were in the heterozygous state. There were also many missense variants with both low and high frequencies. We investigated the segregation of mutation p.Y193C in the family; obligate carriers 201 and 203 and patients 207, 209, 301 and 310 shared a haplotype near the gene locus (Figure 5.14). This haplotype was also shared by unaffected individuals 205, 308 and 313. The mutation was found in patients 207, 209 and 310 and also unaffected individual 205, consistent with the segregation of the haplotype; the remaining patients 201 and 301 could not be analyzed since DNA samples were not available. Obligate carrier 203 and his unaffected son shared the candidate haplotype, but they did not carry the mutation. This situation can be explained by a possible cross-over in 205 around non-informative microsatellites maximally between D8S1697 and GAAT1A4N centromeric to *RGS22*. Additionally, the observation that unaffected individual 205 carried the mutation can be explained by reduced penetrance or possible late onset. The variant was found in one of total 706 unrelated chromosomes in the Turkish EVS (Personal communication), indicating that it is not a common sequence variant in the Turkish population with a power of 95%.

Another novel variant was identified at 16p13.3-p13.13, missense c.1088A>G (p.E363G) in the gene formerly known as *C16orf68* and recently designated as *METTL22*. Recent studies showed that the gene encodes a member of a family of newly-discovered non-histone lysine methyltransferases. Its substrate is Kin17, which is a DNA/RNA binding protein with a role in DNA repair and replication and mRNA processing (Cloutier *et al.,* 2013). MTTL22 protein is responsible for trimethylation of lysine at residue 135 of Kin17, both proteins localize on the chromatin, and overexpression of *METLL22* leads to displacement of Kin17 from chromatin to cytoplasm (Cloutier *et al.,* 2014). Besides the functions in DNA repair/replication, Kin17 may have a chaperon function in the cytoplasm. Since an increase in the

expression of *KIN*, the gene encoding Kin17, is observed in many tumor cell lines (Despras *et al.,* 2003), the authors argued that if Kin17 lysine 135 trimethylation is perturbed in cancer, *METTL22* may represent a novel therapeutic target. Considering those results and the relations among *KIN*, *METTL22* and cancer, we can speculate that the missense mutation identified, which was predicted as disease causing by all the online tools employed, may lead to a nonfunctional MTTL22 protein perturbing Kin17 lysine 135 trimethylation. The segregation of mutation p.E363G in the family was investigated by HRM curve analysis; patients 207 and 209 and unaffected individual 313 were in the same group while patient 310 was grouped with the other unaffected members of the family, similar to the segregation of *CARHSP1* mutation. Therefore, the same arguments are valid for *METTL22* mutation: Since the onset of the disease is late, the affection status of individual 313 is uncertain. Moreover, since patients 207 and 209 were afflicted with meningioma and patient 310 was afflicted with glioma, the mutation could be the causative mutation for meningioma only.

Possible relation of a *CARHSP1* defect to cancer is more straightforward; the protein product is a regulator of *TNF*, which has dual roles as both oncogene and tumor suppressor (Bertazza and Mocellin, 2010). *CARHSP1* encodes calcium regulated heat stable protein 1 containing a cold-shock domain with two RNA binding motifs (Schafer *et al.,* 2003) and is ubiquitously expressed (NCBI/Unigene). The protein product is necessary for effective *TNF*-α mRNA stabilization (Pfeiffer *et al.,* 2011). We propose that the identified frameshift mutation in *CARHSP1* could lead to cancer by leading to reduced stability in *TNF*-α mRNA decreasing its expression, and thus impairing the tumor suppressor potential. *CARHSP1* mutation is expected to result in a nonfunctional protein that has lost ten most terminal amino acids and gained a long, nonnative terminal region that possibly affects the folding of the protein.

RGS22 encodes a regulator of G-protein signaling 22 and is highly expressed in the human testis during different stages of development (Hu *et al.,* 2008) and is also moderately expressed in the brain (NCBI/Unigene). G-proteins are known to control numerous cellular processes such as proliferation, differentiation, migration, embryonic development and membrane trafficking, and their regulators are as important. Hu *et al.* (2011) recently showed that *RGS22* is expressed in many tumor types and represses

epithelial cancer metastasis by suppressing cell migration and invasion. Although it is highly expressed in the cancers of epithelial origin, it is also expressed in non-epithelial cancers such as sarcoma and lymphoma (Hu *et al.,* 2010). The meningioma is a cancer originate from the meninges and meninges have two different origins: those surrounding the midbrain, hindbrain, and spinal cord originate from the cephalic and somatic mesoderm, and those in the forebrain originate from diencephalic neural crest (Siegenthaler and Pleasure, 2012). In our study family, the *RGS22* mutation may lead to a nonfunctional protein which is a tumor suppressor and thus to carcinogenesis.

This is the first study to associate genes *CARHSP1* and *RGS22* and perhaps *METTL22* with head cancer. Future studies could unravel whether the mutations identified herein indeed underlie the pathogenicity in the study family. In summary, we identified three candidate variants, which all are in the genes related to cancer in some way, possible causing susceptibility. Since the family members are afflicted with either meningioma or glioma, future studies include performing linkage analysis separately for meningioma and glioma and further investigation of the exome sequence data at the new candidate locus/loci.

6.5. Epilepsy Syndrome

A consanguineous family afflicted with mental retardation, facial dysmorphism and epilepsy was studied. Linkage analysis was performed in various ways: assuming different inheritance patterns or penetrance levels and not including some family members. In the first linkage analysis, SNP genome scan data of all family members were used, but no candidate locus was found. In the second linkage analysis, only patients and parents were included to search for shared homozygosity regions among patients, but again no candidate locus was found. Then, we thought that the disease could be X-linked since the female patient had a milder phenotype that could have resulted from skewed X inactivation. A candidate region at Xq22.1-24 and 14.2 Mb in size was found where all male patients shared the same hemizygosity. Exome sequence data was filtered at the region, the listed variants were subjected to Sanger sequencing in one of the patients, but none was validated. However, when we investigated the

chromatogram for a possible *ESX1* variant in a patient, we observed that there were extra sequences. The reference sequence contained a 27-nucleotide sequence "GGGTGGCAGAGGCGCCATGGGCGGCCC" that repeated 12 times with slight sequence differences around the amplified region whereas it was repeated 13 times in the patient. However, NuSieve agarose jel electrophrosis showed that the novel allele with 13 repeats was just a polymorphism; some family members and some controls were carrying it. One of the unaffected males whose DNA sample had not been subjected to genome scan was also carrying it. Therefore, we decided to perform SNP genome scan for him since his genotype could eliminate or narrow down this candidate region as well as others. Indeed, his genotype eliminated this region, because he shared the same hemizygous haplotype as the male patients.

Linkage analysis was also performed including only the male patients, considering the milder phenotype of the female patient. Haplotype inspection at the highest maximal LOD score regions revealed only one candidate region at 20p13. At the region, some of the unaffected individuals shared the same homozygous haplotype with the patients, but we thought that this situation could have resulted from low penetrance. Exome sequence data were investigated at 20p13, filtered variants were subjected to Sanger sequencing, but none was validated.

In another analysis, the patient with atypical phenotype was not included in the linkage analysis hypothesizing that he could be having another disease, and linkage analysis was performed with either autosomal recessive or X-linked recessive inheritance. Only one candidate region was found, Xq26.2-27.3, where all male patients except the one with atypical phenotype shared the same hemizygosity. Exome sequence data was investigated at the region, but all of the two filtered variants were false calls, as revealed by Sanger sequencing.

Since all filtered variants in exome sequence output were false calls according to Sanger sequencing, we suspected that the exome output did not belong to patient 406 but to another sample. Thus, we compared the homozygous regions of the patient revealed by SNP genome scan to the homozygous regions of the exome sequence output of the two samples that were sent together to the company for exome sequence analysis. A

comparison of two exome sequence outputs revealed that the exome sequence output that we analyzed for epilepsy in fact belonged to another sample. That explains why all filtered variants were false calls. Therefore, analysis of the correct exome sequence output is one of the future prospects.

Since the clinician hypothesized that patient 406 with severe phenotype could be afflicted with Fragile X syndrome instead of the disease the other patients have, exome data were investigated for Fragile X syndrome genes *FMR1* and *AFF2* but no novel or rare variant was found in either gene in the correct exome sequence output. Then, DNA samples of patient 406 and his mother were sent to Istanbul University to be investigated for CGG trinucleotide repeat expansion, and the repeat size was found within the normal range. Therefore, Fragile X syndrome was excluded.

We also used online program Homozygosity Mapper to find shared homozygosity regions in the patients. Analyses were performed in two ways: including all patients or not including the patient with the atypical phenotype. In total four regions were found, but at those regions either some patients were partially heterozygous or one or two of the unaffected individuals shared the same genotype with the patients. Future prospects include analysis of the correct exome data for the candidate regions. Analysis of the unread regions of exome data is also one of the future prospects.

To rule out mitochondrial inheritance, we analyzed for mitochondrial variants. Many defects in various mitochondrial genes such as *POLG, MTTK, MTTL1, MTTH, MTTS1, MTTS2,* and *MTTF* cause syndromes associated with epilepsy (Bindoff and Engelsen, 2012). Considering that the patients could carry the mutation in a heteroplasmic state due to hypothesized heteroplasmy in the mother, investigation of the mitochondrial variants in the correct exome data is another future prospect.

No duplication or deletion was found in any candidate region.

To conclude, we found a few regions but none of those regions are good candidates due to sharing of the same genotype between patients and some of the unaffected children or the partial heterozygosity in the patients. Future prospects are the

analysis of the correct exome sequence output for the identified regions and the investigation of mitochondrial variants and the unread regions of the exome data.

6.6. Muscular Dystrophy (MD)

We studied a consanguineous family afflicted with muscular dystrophy resembling laminopathy. At the beginning of the study, the affection status of individual 405 was not definite. Therefore, linkage analysis was performed assuming this individual as unaffected, with either autosomal recessive inheritance or X-linked recessive inheritance. Three candidate regions on autosomes and three others on the X-chromosome were found, and exome data were investigated for those regions. Among nine filtered variants, seven were sequenced, and six of them were validated. Then we were informed that individual 405 developed the disease, and new linkage analysis was performed considering him affected, again assuming either autosomal recessive or X-linked recessive inheritance. Four candidate regions on the autosomes and one candidate region on the X-chromosome were found. In total four variants remained in those regions after filtering the exome sequence data, but none is yet Sanger sequenced. The largest candidate region was at 1q22-23.2 and harbored *LMNA* gene, encoding Lamin A/C protein. Considering that the disease could be a laminopathy, *LMNA* was assessed as the strongest candidate gene. Exome sequence data did not list any *LMNA* variants. When we investigated the sequenced regions of the gene, we found that only the UTRs and coding exons of the longest transcript were included in exome sequencing. *LMNA* encodes 11 different protein coding transcripts, most of them slightly differing at the 5' UTR or 3' UTR from the longest transcript. Three of them have completely different 5' UTR exons, first coding exons, last coding exons or 3' UTRs as compared to the longest transcript, and those exons were not covered in the exome sequencing. Therefore, we sequenced those regions with priority but did not find a variant that could possibly underlie the disease. Then we decided to sequence all exons of the gene. We identified a synonymous *LMNA* c.436C>T variant in exon 1 by Sanger sequencing. Inspection of the exome sequence reads via IGV program verified the variant, but the variant was not listed in the exome sequence output after variant calling using program SAMtools. Methotta yazıyor. That is why we identified the variant so late. Therefore, it is highly

recommended that the exome sequence reads for the strongest candidate gene, if any, should be inspected. The variant was analyzed via online Codon Usage Database to compare the frequencies of occurrence of the wildtype and the mutant codons, and the mutant codon had a lower frequency of occurrence as compared the wildtype. Since rare codons can possibly result in a misfolded protein (Spencer *et al.*, 2012), the variant can lead to misfolded lamin protein that could lead to the laminopathy in the patients. According to database ENCODE, 200 nucleotides around the site of the variant are in the active promoter site. Another possibility is that the variant could affect the recognition of the promoter sequence for some of the transcript isoforms that are essential for muscle, leading to overexpression or underexpresion of the gene. Gene expression studies can be performed if a muscle biopsy sample from a patient is obtained. Additionally, expression of the possibly affected transcript isoforms could be performed in commercial muscle tissue cDNA. The variant was not found in either EVS or TEVS. Future prospects include investigation of segregation of the variant in the family members and analysis of the effect of the variant on gene expression.

Emery–Dreifuss muscular dystrophy (EDMD) is a genetically heterogeneous disease; X-linked recessive, AD and AR forms were described, and it exhibits variable age of onset and disease progression. It is characterized by early contractures, slowly progressive muscle wasting and weakness, and cardiac conduction defects. X-linked recessive EDMD is caused by mutations in emerin gene whereas *LMNA* mutations lead to dominant and recessive EDMD. Both genes encode proteins localized to the nuclear membrane. The majority of the mutations in *LMNA* are dominant and lead to AD-EDMD (Puckelwartz and Mcnally, 2011) whereas in five families recessive mutations are reported either in the homozygous or the compound heterozygous state: H222Y, R225Q, R249W, D461Y, R482Q, R644C, c.IVS7-1G4T (Raffaele Di Barletta *et al*., 2000; Scharner *et al*., 2011; Jimenez-Escrig *et al.*, 2012; Wiltshire *et al*., 2013,). *LMNA* null mice are viable but develop muscular dystrophy and dilated cardiomyopathy soon after birth (Sullivan et al., 1999; Nikolova et al., 2004). The phenotype of the study patients most resemble EDMD among the diseases caused by *LMNA* defects. Therefore, our study family could be the fourth case of AR-EDMD. However, analysis of muscle biopsy sample of one of the patients could unravel whether the mutation identified herein indeed underlie the pathogenicity in the study family.

It should also be emphasized that six mutations, most of them deleterious, were filtered and then validated by Sanger sequencing at the candidate loci found by linkage analysis performed by assuming only two of the children affected before we were informed about the affected status of 405. Thus, it should be kept in mind that each person carries numerous possibly deleterious mutations, but those mutations are not all pathogenic.

7. CONCLUSIONS

In the framework of this study, a mutation in *SRD5A3* that was previously reported in three children afflicted with Congenital Disorder of Glycosylation was found; the mutation resulted in a different phenotype in our study family. Additionally, novel mutations were identified in Myosin Storage Myopathy family and Desmin Deficiency Myopathy family in genes that are already related to the respective diseases. For Hereditary Head Tumors, two or possibly three mutations were identified in genes related to cancer, but whether any of them contributed to the cancer susceptibility in our study family could not be discerned for sure. For DM five candidate loci were found, and a synonymous mutation was identified in a gene related to laminopathy. Investigation of segregation of the variant in the family members is ongoing. In epilepsy family, no candidate region was found and analysis of the correct exome data is a future prospect.

REFERENCES

Aavikko, M., S. P. Li, S. Saarinen, P. Alhopuro, E. Kaasinen, E. Morgunova, Y. Li, K. Vesanen, M. J. Smith, D. G. R. Evans, M. Poyhonen, A. Kiuru, A. Auvinen, L. A. Aaltonen, J. Taipale, and P. Vahteristo, 2012, "Loss of SUFU Function in Familial Multiple Meningioma", *American Journal of Human Genetics*, Vol. 91, pp. 520-526.

Al-Gazali, L., J. Hertecant, K. Algawi, H. El Teraifi, and M. Dattani, 2008, "A New Autosomal Recessive Syndrome of Ocular Colobomas, Ichthyosis, Brain Malformations and Endocrine Abnormalities in an Inbred Emirati Family", *American Journal of Human Genetics*, Vol. 146, pp. 813–819.

Armel, T. Z., and L. A. Leinwand, 2009, "Mutations in the Beta-Myosin Rod Cause Myosin Storage Myopathy via Multiple Mechanisms", *Proceedings of National Academy of Sciences,* Vol. 106, No. 15, pp. 6291-6296.

Assmann, B., R. Hackler, V. Peters, J. R. Schaefer, T. Arndt, E. Mayatepek, J. Jaeken, and G. F. Hoffmann, 2001, "A New Subtype of a Congenital Disorder of Glycosylation (CDG) with Mild Clinical Manifestations", *Neuropediatrics*, Vol. 32, pp. 313-318.

Barohn, R. J., R. A. Brumback, and J. R. Mendell, 1994, "Hyaline Body Myopathy", *Neuromuscular Disorders*, Vol. 4, pp. 257-262.

Bertazza, L., and S. Mocellin, 2010, "The Dual Role of Tumor Necrosis Factor (TNF) in Cancer Biology", *Current Medicinal Chemistry*, Vol. 17, No. 29, pp. 3337-3352.

Bindoff, L. A., and B. A. Engelsen, 2012, "Mitochondrial Diseases and Epilepsy", *Epilepsia*, Vol. 4, pp. 92-97.

Blair, E., C. Redwood, M. de Jesus Oliveira, J. C. Moolman-Smook, P. Brink, V. A. Corfield, I. Ostman-Smith, and H. Watkins, 2002, "Mutations of the Light Meromyosin Domain of the Beta-Myosin Heavy Chain Rod in Hypertrophic Cardiomyopathy", *Circulation Research,* Vol. 90, No. 3, pp. 263-269.

Bolger, G. B., J. Stamberg, I. R. Kirsch, G. F. Hollis, D. F. Schwarz, and G. H. Thomas, 1985, "Chromosome Translocation T(14;22) and Oncogene (C-Sis) Variant in a Pedigree with Familial Meningioma", *New England Journal of Medicine,* Vol. 312, pp. 564-567.

Bozoğlu, T., 2008, *Genetic Analysis In Three Inherited Disorders,* M.Sc. Thesis, Bogazici University.

Cantagrel, V., D. J. Lefeber, B. G. Ng, Z. Guan, J. L. Silhavy, S. L. Bielas, L. Lehle, H. Hombauer, M. Adamowicz, E. Swiezewska, A. P. De Brouwer, and P. Blumel, 2010, "SRD5A3 Is Required For Converting Polyprenol to Dolichol and Is Mutated in a Congenital Glycosylation Disorder", *Cell,* Vol. 142, No. 2, pp. 203-217.

Carlsson L., and L. E. Thornell, 2001, "Desmin-Related Myopathies in Mice and Man", *Acta Physiologica Scandinavica*, Vol. 171, No. 3, pp. 341-348.

Cetin, N., B. Balci-Hayta, H. Gundesli, P. Korkusuz, N. Purali, B. Talim, E. Tan, D. Selcen, S. Erdem-Ozdamar, and P. Dincer, 2013, "A Novel Desmin Mutation Leading to Autosomal Recessive Limb-Girdle Muscular Dystrophy: Distinct Histopathological Outcomes Compared with Desminopathies", *Journal of Medical Genetics*, Vol. 50, No. 7, pp. 437-443.

Cetinkaya, M., 2010, *Gene Hunt in Four Inherited Diseases,* Ph.D. Thesis, Bogazici University.

Ceuterick, C., J. J. Martin, and C. Martens, 1993, "Hyaline Bodies in Skeletal Muscle of a Patient with a Mild Chronic Nonprogressive Congenital Myopathy", *Clinical Neuropathology*, Vol. 12, pp. 79-83.

Cloutier, P., M. Lavallée-Adam, D. Faubert, M. Blanchette, and B. Coulombe, 2013, "A Newly Uncovered Group of Distantly Related Lysine Methyltransferases Preferentially Interact with Molecular Chaperones to Regulate Their Activity", *PLoS Genetics*., Vol. 9, No. 1, e1003210, doi: 10.1371/journal.pgen.1003210.

Cloutier, P., M. Lavallée-Adam, D. Faubert, M. Blanchette, and B. Coulombe, 2014 "Methylation of the DNA/RNA-Binding Protein Kin17 by METTL22 Affects Its Association with Chromatin", *Journal of Proteomics.*, Vol. 100, pp. 115-124.

Collins, J. S., and C. E. Schwartz, 2002, "Detecting Polymorphisms and Mutations in Candidate Genes", *American Journal of Human Genetics*, Vol. 71, pp. 1251-1252.

de Andrade, M., J. S. Barnholtz, C. I. Amos, P. Adatto, C. Spencer, and M. L. Bondy, 2001, "Segregation Analysis of Cancer in Families of Glioma Patients", *Genetical Epidemiology*, Vol. 20, pp. 258–270.

Despras, E., L. Miccoli, C. Créminon, D. Rouillard, J. F. Angulo, and D. S. Biard, 2003, "Depletion of KIN17, a Human DNA Replication Protein, Increases the Radiosensitivity of RKO Cells", *Radiation Research*, Vol. 159, No. 6, pp. 748-758.

Emery, A. E., 2002, "The Muscular Dystrophies", *Lancet*, Vol. 359, No. 9307, pp. 687-695.

Goldfarb, L. G., P. Vicart, H. H. Goebel, and M. C. Dalakas, 2004, "Desmin Myopathy", *Brain*, Vol. 127, pp. 723-734.

Gründahl, J. E. H., Z. Guan, S. Rust, J. Reunert, B. Müller, I. Du Chesne, K. Zerres, S. Rudnik-Schöneborn, N. Ortiz-Brüchle, M. G. Hausler, J. Siedlecka, E. Swiezewska, C. R. H. Raetz, and T. Marquardt, 2012, "Life with Too Much Polyprenol: Polyprenol Reductase Deficiency", *Molecular Genetics and Metabolism*, Vol. 105, pp. 642-651.

Henderson, M., L. De Waele, J. Hudson, M. Eagle, C. Sewry, J. Marsh, R. Charlton, L. He, E. L. Blakely, I. Horrocks, W. Stewart, and R. W. Taylor, 2013, "Recessive Desmin-Null Muscular Dystrophy with Central Nuclei and Mitochondrial Abnormalities", *Acta Neuropathologica*, Vol. 125, No. 6., pp. 917-919.

Hoffmann, K. and T. H. Lindner, 2005, "easyLINKAGE-Plus--Automated Linkage Analyses Using Large-Scale SNP Data", *Bioinformatics*, Vol. 21, No. 17, pp. 3565-3567.

Hu, Y., J. Xing, L. Chen, X. Guo, Y. Du, C. Zhao, Y. Zhu, M. Lin, Z. Zhou, and J. Sha, 2008, "RGS22, a Novel Testis-Specific Regulator of G-Protein Signaling Involved in Human and Mouse Spermiogenesis along with GNA12/13 Subunits", *Biology of Reproduction*, Vol. 79., No. 6, doi: 10.1095/biolreprod.107.067504.

Hu, Y., J. Xing, L. Wang, M. Huang, X. Guo, L. Chen, M. Lin, Y. Zhou, Z. Liu, Z. Zhou, and J. Sha, 2011, "RGS22, a Novel Cancer/Testis Antigen, Inhibits Epithelial Cell Invasion and Metastasis", *Clinical & Experimental Metastasis*., Vol. 28, No. 6, pp. 541-549.

Jaeken J., 2011, "Congenital Disorders of Glycosylation (CDG): It's (Nearly) All in It", *Journal of Inherited Metabolic Disease*, Vol. 34, No. 4, pp. 853-858.

Jimenez-Escrig, A., I. Gobernado, M. Garcia-Villanueva, and A. Sanchez-Herranz, 2012, "Autosomal Recessive Emery-Dreifuss Muscular Dystrophy Caused by a Novel Mutation (R225Q) in the Lamin A/C Gene Identified by Exome Sequencing", *Muscle Nerve*, Vol. 45, No. 4, pp. 605-610.

Kara, B., Ö. Ayhan, G. Gökçay, N. Başboğaoğlu, and A. Tolun, 2014, "Adult Phenotype and Further Phenotypic Variability in SRD5A3-CDG", *BMC Medical Genetics*, Vol. 15, No. 10, doi: 10.1186/1471-2350-15-10.

Kasapkara, C. S., L. Tümer, F. S. Ezgü, A. Hasanoğlu, V. Race, G. Matthijs, and J. Jaeken, 2012, "SRD5A3-CDG: A Patient with a Novel Mutation", *European Journal of Paediatric Neurology*, Vol. 16, pp. 554-556.

Kremer, E. J., M. Pritchard, M. Lynch, S. Yu, K. Holman, E. Baker, S. T. Warren, D. Schlessinger, G. R. Sutherland, and R. I. Richards, 1991, "Mapping of DNA Instability at the Fragile X to a Trinucleotide Repeat Sequence p(CCG)n", *Science*, Vol. 252, No. 5013, pp. 1711-1714.

Ku, C. S., N. Naidoo, and Y. Pawitan, 2011, "Revisiting Mendelian Disorders through Exome Sequencing", *Human Genetics*, Vol. 129, No. 4, pp. 351-70.

Lalonde, E., S. Albrecht, K. C. Ha, K. Jacob, N. Bolduc, C. Polychronakos, P. Dechelotte, J. Majewski, and N. Jabado, 2010, "Unexpected Allelic Heterogeneity and Spectrum of Mutations in Fowler Syndrome Revealed by Next-Generation Exome Sequencing", *Human Mutation*, Vol. 31, No. 8, pp. 918-923.

Lander, E. S. and D. Botstein, 1987, "Homozygosity Mapping: A Way to Map Human Recessive Traits with The DNA of Inbred Children", *Science*, Vol. 236, No. 4808, pp. 1567-1570.

Lekanne Deprez, R. H., N. A. Groen, N. A. van Biezen, A. Hagemeijer, E. van Drunen, J. W. Koper, C. J. J. Avezaat, D. Bootsma, and E. C. Zwarthoff, 1991, "A t(4;22) in a Meningioma Points to the Localization of a Putative Tumor-Suppressor Gene", *American Journal of Human Genet*ics, Vol. 48, pp. 783-790.

Li, Z., M. Mericskay, O. Agbulut, G. Butler-Browne, L. Carlsson, L. E. Thornell, C. Babinet, and D. Paulin, 1997, "Desmin Is Essential for the Tensile Strength and Integrity of Myofibrils but not for Myogenic Commitment, Differentiation, and Fusion of Skeletal Muscle", *Journal of Cell Biology*, Vol. 139, No. 1, pp. 129-144.

Livak, K. J. and T. D. Schmittgen, 2001, "Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta C}{}_T$ Method", *Methods*, Vol. 25, pp. 402-408.

Louis, D. N., E. C. Holland, and J. G. Cairncross, 2001, "Glioma Classification: A Molecular Reappraisal", *American Journal of Pathology*, Vol. 159, No. 3, pp. 779-86.

Louis, D. N., H. Ohgaki, O. D. Wiestler, W. K. Cavenee, P. C. Burger, A. Jouvet, B. W. Scheithauer, and P. Kleihues, 2007, "The 2007 WHO Classification of Tumours of the Central Nervous System", *Acta Neuropathologica*, Vol. 114, No. 2, pp. 97-109.

Malmer, B., S. Haraldsson, E. Einarsdottir, P. Lindgren, and D. Holmberg, 2005, "Homozygosity Mapping of Familial Glioma in Northern Sweden", *Acta Oncologica*, Vol. 44, pp. 114-119.

Matthijs, G., D. Rymen, M. B. Millón, E. Souche, and V. Race, 2013, "Approaches to Homozygosity Mapping and Exome Sequencing for the Identification of Novel Types of CDG", *Glycoconjugate journal*, Vol. 30, No.1, pp. 67-76.

Milner, D. J., G. Weitzer, D. Tran, A. Bradley, and Y. Capetanaki, 1996, "Disruption of Muscle Architecture and Myocardial Degeneration in Mice Lacking Desmin", *Cell Biology*, Vol. 134, No. 5, pp. 1255-1270.

Morava, E., R. A. Wevers, V. Cantagrel, L. H. Hoefsloot, L. Al-Gazali, J. Schoots, A. van Rooij, K. Huijben, C. M. A. Van Ravenswaaij-Arts, M. C. J. Jongmans, J. Sykut-Cegielska, and G. F. Hoffmann, 2010, "A Novel Cerebello-Ocular Syndrome with Abnormal Glycosylation Due To Abnormalities in Dolichol Metabolism", *Brain*, Vol. 133, pp. 3210-3220.

Morimoto, S., 2008, "Sarcomeric Proteins and Inherited Cardiomyopathies", *Cardiovascular Research*, Vol. 77, pp. 659-666.

Morton, N. E., 1955, "Sequential Tests for the Detection of Linkage", *American Journal of Human Genetics*, Vol. 7, No. 3, pp. 277-318.

Muñoz-Mármol, A. M., G. Strasser, M. Isamat, P. A. Coulombe, Y. Yang, X. Roca, E. Vela, J. L. Mate, J. Coll, M. T. Fernández-Figueras, J. J. Navas-Palacios, A. Ariza, and 1 other, 1998, "A Dysfunctional Desmin Mutation in a Patient with Severe Generalized Myopathy", *Proceedings of National Academy of Sciences*, Vol. 95, No. 19, pp. 11312-11317.

Nagy, E., and L. E. Maquat, 1998, "A Rule for Termination-Codon Position within Intron-Containing Genes: When Nonsense Affects RNA Abundance", *Trends in Biochemical Sciences*, Vol. 23, No. 6, pp. 198-199.

Ng, P. C., S. Levy, J. Huang, T. B. Stockwell, B. P. Walenz, K. Li, N. Axelrod, D. A. Busam, R. L. Strausberg, and J. C. Venter, 2008, "Genetic Variation in an Individual Human Exome", *PLoS Genetics*, Vol. 4, No. 8, e1000160.

Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. A. Nickerson, J. Shendure, and M. J. Bamshad, 2010, "Exome Sequencing Identifies the Cause of a Mendelian Disorder", *Nature Genetics*, Vol.42, No. 1, pp. 30-35.

Ng, S. B., E. H. Turner, P. D. Robertson, S. D. Flygare, A. W. Bigham, C. Lee, T. Shaffer, M. Wong, A. Bhattacharjee, E. E. Eichler, M. Bamshad, D.A. Nickerson, and J. Shendure, 2009, "Targeted Capture and Massively Parallel Sequencing of 12 Human Exomes", *Nature*, Vol. 461, No. 7261, pp. 272-276.

Nikolova V., C. Leimena, A. C. McMahon, J. C. Tan, S. Chandar, D. Jogia, S. H. Kesteven, J. Michalicek, R. Otway, F. Verheyen, S. Rainer, and C. L. Stewart, 2004, "Defects in Nuclear Structure and Function Promote Dilated Cardiomyopathy in Lamin A/C-Deficient Mice", *The Journal of Clinical Investigation*, Vol. 113, No. 3, pp. 357-369.

Onengüt, S., S. A. Uğur, H. Karasoy, N. Yüceyar, and A. Tolun, 2004, "Identification of a Locus for an Autosomal Recessive Hyaline Body Myopathy at Chromosome 3p22.2-p21.32", *Neuromuscular Disorders*, Vol. 14, pp. 4-9.

Pfeiffer, J. R., B. L. McAvoy, R. E. Fecteau, K. M. Deleault, and S. A. Brooks, 2011, "CARHSP1 Is Required for Effective Tumor Necrosis Factor Alpha mRNA Stabilization and Localizes to Processing Bodies and Exosomes", *Molecular and Cellular Biol*ogy, Vol. 31, No. 2, pp. 277-86.

Pinol-Ripoll, G., A. Shatunov, A. Cabello, P. Larrode, I. de la Puerta, J. Pelegrin, F. J. Ramos, M. Olive, and L. G. Goldfarb, 2009, "Severe Infantile-Onset Cardiomyopathy Associated with a Homozygous Deletion in Desmin", *Neuromuscular Disorders*, Vol, 19, No. 6, pp. 418-422.

Poduri, A., and D. Lowenstein, 2011, "Epilepsy Genetics: Past, Present, and Future", *Current Opinion in Genetics and Development*, Vol. 21, No. 3, pp. 325-332.

Prietsch, V., V. Peters, R. Hackler, R. Jakobi, B. Assmann, J. Fang, C. Körner, A. Helwig-Rolig, J. R. Schaefer, and G. F. Hoffmann, 2002, "A New Case of CDG-X with Stereotyped Dystonic Hand Movements and Optic Atrophy", *Journal of Inherited Metabolic Di*sease, Vol. 25, pp.126-130.

Puckelwartz M. and E. M. McNally, 2011, "Emery-Dreifuss Muscular Dystrophy", *Handbook of Clinical Neurology*, Vol. 101, pp. 155-166.

Raffaele Di Barletta, M., E. Ricci, G. Galluzzi, P. Tonali, M. Mora, L. Morandi, A. Romorini, T. Voit, K. H. Orstavik, L. Merlini, C. Trevisan, and V. Biancalana, 2000, "Different Mutations in the LMNA Gene Cause Autosomal Dominant and Autosomal Recessive Emery-Dreifuss Muscular Dystrophy", *American Journal of Human Genetics*, Vol. 66, No. 4, pp. 1407-1412.

Risch, N., 1992, "Genetic Linkage: Interpreting Lod Scores", *Science*, Vol. 255, No. 5046, pp. 803-804.

Sahgal, V., and S. Sahgal, 1977, "A New Congenital Myopathy", *Acta Neuropathologica*, Vol. 37, pp. 225-230.

Schafer, C., H. Steffen, K. J. Krzykowski, B. Göke, and G. E. Groblewski, 2003, "CRHSP-24 Phosphorylation Is Regulated by Multiple Signaling Pathways in Pancreatic Acinar Cells", *American Journal of Physiology, Gastrointestinal and Liver Physiology*, Vol. 285, pp. 726-734.

Scharner J., C. A. Brown, M. Bower, S. T. Iannaccone, I. A. Khatri, D. Escolar, E. Gordon, K. Felice, C. A. Crowe, C. Grosmann, M. N. Meriggioli, and A. Asamoah, 2011, "Novel LMNA Mutations in Patients with Emery-Dreifuss Muscular Dystrophy and Functional Characterization of Four LMNA Mutations", *Human Mutation*, Vol. 32, No. 2, pp. 152-167.

Siegenthaler, J. A. and S. J. Pleasure, 2011, "We've Got You "Covered": How the Meninges Control Brain Development", *Current Opinion in Genetics and Development*, Vol. 21, No. 3, pp. 249-255.

Smith, M. J., J. O'Sullivan, S. S. Bhaskar, K. D. Hadfield, G. Poke, J. Caird, S. Sharif, D. Eccles, D. Fitzpatrick, D. Rawluk, D. du Plessis, W. G. Newman, and 1 other, 2013, "Loss-of-Function Mutations in SMARCE1 Cause an Inherited Disorder of Multiple Spinal Meningiomas", *Nature Genetics*, Vol. 45, pp. 295-298.

Sohn, R. L., K. L. Vikstrom, M. Strauss, C. Cohen, A. G. Szent-Gyorgyi, and L. A. Leinwand, 1997, "A 29 Residue Region of the Sarcomeric Myosin Rod Is Necessary for Filament Formation", *Journal of Molecular Biology*, Vol. 266, No. 2, pp. 317-330.

Spencer, P. S., E. Siller, J. F. Anderson, and J. M. Barral, 2012, "Silent Substitutions Predictably Alter Translation Elongation Rates and Protein Folding Efficiencies", *Journal of Molecular Biology*, Vol. 422, No. 3, pp. 328-35.

Staal, F. J. T., R. B. van der Luijt, M. R. M. Baert, J. van Drunen, H. van Bakel, E. Peters, I. de Valk, H. K. P. van Amstel, M. J. B. Taphoorn, G. H. Jansen, C. W. M. van Veelen, B. Burgering, and 1 other, 2002, "A Novel Germline Mutation of PTEN Associated with Brain Tumours of Multiple Lineages", *British Journal of Cancer*, Vol. 86, pp. 1586-1591.

Sullivan T., D. Escalante-Alcalde, H. Bhatt, M. Anver, N. Bhat, K. Nagashima, C. L. Stewart, and B. Burke, 1999, "Loss of A-type Lamin Expression Compromises Nuclear Envelope Integrity Leading to Muscular Dystrophy", *The Journal of Cell Biology*, Vol 147, No. 5, pp. 913-920.

Tajsharghi, H., and A. Oldfors, 2013, "Myosinopathies: Pathology and Mechanisms", *Acta Neuropathologica*, Vol. 125, pp. 3-18.

Tajsharghi, H., A. Oldfors, D. P. Macleod, and M. Swash, 2007, "Homozygous Mutation in MYH7 in Myosin Storage Myopathy and Cardiomyopathy", *Neurology*, Vol. 68, No. 12, pp. 962.

Terwilliger, N. E., 1955, "Sequential Tests for the Detection of Linkage", *American Journal of Human Genetics*, Vol. 7, No. 3, pp. 277-318.

Weber, J. L., and K. W. Broman, 2001, "Genotyping for Human Whole-Genome Scans: Past, Present, and Future", *Advances in Genetics*, Vol. 42, pp. 77-96.

Wellenreuther, R., J. A. Kraus, D. Lenartz, A. G. Menon, J. Schramm, D. N. Louis, V. Ramesh, J. F. Gusella, O. D. Wiestler, and A. von Deimling, 1995, "Analysis of the Neurofibromatosis 2 Gene Reveals Molecular Variants of Meningioma", *American Journal of Pathology*, Vol. 146, pp. 827-832.

Wen, P. Y., and S. Kesari, 2008, "Malignant Gliomas in Adults", *New England Journal of Medicine*, Vol. 359, pp. 492-507.

Wiltshire K. M., R. A. Hegele, A. M. Innes, and A. K. Brownell, 2013, "Homozygous Lamin A/C Familial Lipodystrophy R482Q Mutation in Autosomal Recessive Emery Dreifuss Muscular Dystrophy", *Neuromuscular Disorders*, Vol. 23, No. 3, pp. 265-268.

Wolfe, L. A., and D. Krasnewich, 2013, "Congenital Disorders of Glycosylation and Intellectual Disability", *Developmental Disabilities Research Review*, Vol. 17, No.3, pp. 211-225.

Worman, H. J., and G. Bonne, 2007, "Laminopathies: A Wide Spectrum of Human Diseases", *Experimental Cell Research*, Vol. 313, pp. 2121-2133.

Zang, K. D., 2001, "Meningioma: A Cytogenetic Model of a Complex Benign Human Tumor, Including Data on 394 Karyotyped Cases", *Cytogenetics and Cell Genetics*, Vol. 93, pp. 207-220.