DISEASE GENE SEARCH VIA LINKAGE MAPPING AND EXOME SEQUENCING

by

Çiğdem Köroğlu Altok

B.S., Molecular Biology and Genetics, Boğaziçi University, 2007

Submitted to the Institute of Graduate Studies in

Science and Engineering in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2014

dedicated to my beloved husband Serdar

ACKNOWLEDGEMENTS

I would like to express my sincere respect and gratitude to my thesis supervisor Prof. Aslı Tolun for her guidance, support and valuable criticism in every step of my study and for inspiring me by her work ethic. I would like to extend my appreciation to the members of my thesis committee, Prof. Hande Çağlayan, Assoc. Prof. Eda Tahir Turanlı, Prof. Nejat Dalay and Assoc. Prof. Sibel Uğur İşeri for devoting their time to evaluate this work.

I would like to thank Assoc. Prof. Yusuf Durlu, Hacer Durmuş, M.D., Rezzak Yılmaz, M.D. and Assoc. Prof. Mehmet Seven for the blood samples and clinical evaluations. I thank all the families that participated in the study.

I would like to express my appreciation to Nehir Mavioğlu for her extensive assistance in lab work and to my colleagues in the lab, Özgecan Ayhan and Esra Yıldız for their help. I sincerely thank all previous Kommagene laboratory members I worked with, especially Dr. Sibel Uğur İşeri for sharing her experience and encouraging me during all my graduate years and Ayşe Güven for her help in bioinformatics analyses.

I would also like to thank Prof. Murat Günel and members of his laboratory at Yale University for their help in SNP genotyping analyses and exome sequencing.

I wish to thank all academic staff whose expertise contributed greatly to my graduate education.

I have great pleasure in thanking my husband Serdar for his endless support, patience and assistance. I deeply thank my parents for their support and sympathy through my entire life.

This work was supported by Boğaziçi University Research Fund (5708 and 6655). I was a fellow of the Scientific and Technological Research Council of Turkey from 2008 to 2010.

ABSTRACT

DISEASE GENE SEARCH VIA LINKAGE MAPPING AND EXOME SEQUENCING

Identification of disease genes has a key importance in understanding the molecular mechanisms of disease pathogenesis as well as in expanding our knowledge on gene function. Linkage mapping is a powerful method to identify the regions in the genome that possibly harbor the gene responsible for the disease in a family. Once the disease locus is identified, the genes within that locus are searched for variants that could be associated with the disease. Advances in sequencing technologies in the last decade made mutation search much easier; causative mutations underlying the disease can be identified via whole exome sequencing even in small families. In this study, causative genes for five inherited disorders were identified. In a large consanguineous family with four members afflicted with Cone-Rod Dystrophy (CORD), we identified a homozygous mutation p.R106P in novel disease gene *POC1B*, encoding a centriolar protein. We identified homozygous missense mutation p.Q21P in *CEP290*, encoding another centriolar protein, in four patients afflicted with Optic Atrophy (OPA) in a large inbred family. In three of those patients, we additionally detected a novel and homozygous variant in *ACSS3* possibly underlying the extraocular neurological symptoms shared by at least two of the patients. For the other three diseases we studied, identified mutations were all homozygous and truncating: *ALDH1B1* p.G388EfsX23 for Amyotrophic Lateral Sclerosis (ALS), *P2RX6* p.K331EfsX26 for Rett syndrome and *NALCN* p.Q642X for Infantile Neuroaxonal Dystrophy. Further, one difficulty in studying large families is demonstrated in OPA family, in which two genes are responsible for the variable phenotype. Our findings expand the spectrums of phenotypes of CORD, OPA, ALS, RTT, and INAD, for which we identified novel causative genes, and shed light onto the functions of the causative, novel genes. Screening identified genes for mutations could benefit families afflicted with diseases with similar manifestations as in our respective study families.

ÖZET

BAĞLANTI HARİTALAMASI VE EKSOM DİZİLEME İLE HASTALIK GENİ ARAMA

Bir hastalık geninin tanımlanması hastalık patogenezinin moleküler mekanizmasının anlaşılmasının yanısıra, gen fonksiyonu üzerine bilgimize katkıda bulunması açısından da kilit öneme sahiptir. Bağlantı haritalaması, bir ailede hastalığa yol açan geni içeren genom bölgesini belirlemek için kuvvetli bir yöntemdir. Hastalık lokusu belirlendiğinde, o lokustaki genlerde hastalıkla ilişkili olabilecek varyantlar aranır. Dizileme teknolojilerinde son on yıldaki gelişmeler mutasyon aramayı çok kolaylaştırmıştır; küçük bir ailede bile hastalığa sebep olan mutasyonun belirlenmesi artık mümkündür. Bu çalışmada, beş kalıtımsal hastalıkla ilişkili genler belirlendi. Akraba evliliği yapmış dört Koni ve Çubuk Distrofisi (KÇD) hastalı büyük bir ailede yeni bir hastalık geni olan *POC1B*'de bir homozigot mutasyon (p.R106P) tanımladık. Büyük bir ailenin Optik Atrofi (OPA) hastası dört bireyinde, diğer bir sentriyol protein kodlayan *CEP290* geninde homozigot bir yanlış anlamlı mutasyonu (p.Q21P) belirledik. Ayrıca, bu hastaların üçünde, *ACSS3* geninde, hastaların en az ikisinde görülen göz dışı semptomların temelini oluşturabilecek yeni, homozigot ve muhtemelen zararlı bir varyant bulduk. Çalıştığımız diğer üç nörolojik hastalıkta, hepsi homozigot ve güdük proteine yol açan nedensel mutasyonlar bulduk: *Amyotrofik Lateral Skleroz* (ALS) için *ALDH1B1* p.G388EfsX23, *Rett Sendromu* (RTT) için *P2RX6* p.K331EfsX26 ve *İnfantil Nöroaksonal Distrofi* (INAD) için *NALCN* p.Q642X. Ayrıca, büyük ailelerde çalışmanın bir zorluğu OPA ailesinde gösterilmiştir: değişken fenotipin temelinde iki hastalık geninin birden olması yatmaktadır. Bulgularımız, yeni nedensel genler belirlediğimiz CORD, OPA, ALS, RTT ve INAD hastalıklarının fenotip spektrumlarını genişletmiş ve yeni hastalık genlerinin işlevlerine ışık tutmuştur. Tanımlamış olduğumuz genlerin çalıştığımız ailelerle benzer hastalık taşıyan ailelerde mutasyon için taranması o ailelere yarar sağlayabilir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

A	Alanine
AD	Autosomal dominant
ALS	Amyotrophic lateral sclerosis
APS	Ammonium peroxodisulphate
AR	Autosomal recessive
bp	Base pair
BAM	Binary alignment/map
BSA	Bovine serum albumine
BWA	Burrows-wheeler aligner
C	Cysteine
cDNA	Complementary deoxyribonucleic acid
Chr	Chromosome
cM	Centimorgan
CNV	Copy number variation
CORD	Cone-rod dystrophy
D	Aspartic acid
dH_2O	Distilled water
DGV	Database of Genomic Variants
DMSO	Dimethyl sulfate
DNA	Deoxyribonucleic acid
E	Glutamic acid
EDTA	Ethylenediaminetetraacetate
EVS	Exome Variant Server
F	Phenylalanine
G	Glycine
H	Histidine
HCiE	Homozygosity comparison in Excel
HRM	High resolution melting
IBD	Identical by descent
INAD	Infantile neuroaxonal dystrophy

K	Lysine
kb	Kilobase pair
L	Leucine
LCA	Leber congenital amaurosis
LOD	Logarithm of odds
M	Methionine
MAF	Minor allele frequency
Mb	Mega base
min	Minute
mRNA	Messenger ribonucleic acid
NA	Not available
NCBI	National Heart, Lung and Blood Institute
NGS	Next-generation sequencing
OPA	Optic atrophy
P	Proline
PAGE	Polyacrylamide gel elecrophoresis
PAR	Pseudoautosomal region
PCR	Polymerase chain reaction
PR	Photoreceptor
Q	Glutamine
QPCR	Quantitative polymerase chain reaction
R	Arginine
RNA	Ribonucleic acid
RPE	Retinal Pigment Epithelium
RTT	Rett Syndrome
S	Serine
SAM	Sequence Alignment/Map
SDS	Sodium dodecyl sulphate
sec	Second
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
T	Threonine
TBE	Tris, boric acid, EDTA

TE	Tris, EDTA
TEMED	N, N, N, N'-Tetramethylethylenediamine
UTR	Untranslated region
UV	Ultraviolet
W	Tryptophan
X	Stop codon
XR	X-linked recessive
Y	Tyrosine

1. INTRODUCTION

Within the framework of this study, genes responsible for inherited recessive disorders in six families were searched using linkage analysis/mapping and subsequent exome sequencing or mutation screening. The diseases were Cone-Rod Dystrophy, Leber Congenital Amaurosis, Optic Atrophy, Amyotrophic Lateral Sclerosis, Rett-like Syndrome, and Infantile Neuroaxonal Dystrophy. In the family afflicted with Leber Congenital Amaurosis, a previously reported mutation was identified in *AIPL1*. For the other five diseases, novel causative genes were identified.

1.1. Photoreceptor Degeneration

Retina, the lining of the inner surface of the eye, is a light-sensitive tissue capable of capturing and processing visual signals. It consists of an outer layer of retinal pigment epithelium (RPE) cells and an inner layer of various neuronal cells which is called neural retina (Masland, 2001). The outer segment of neural retina contains photoreceptor (PR) cells that generate light signals and pass them to the inner cell layers. Those light signals are then transmitted to the brain through the optic nerve. RPE separates neural retina from the choroid, a vascular tissue that nourishes RPE and PR cells (Figure 1.1). RPE, PR cells and the choroid are interdependent; dysfunction in any one of them may disrupt the function of the others (Wright *et al*., 2010).

There are two types of photoreceptors, rod and cone. They differ in morphology, pigment content and physiological function. Rods are specialized for vision in low light since they are extremely sensitive and can be triggered by just a few photons. They have inner and outer segments, and outer segments have stacked discs containing opsin protein that absorbs photons. The discs are constantly renewed by RPE through phagocytosis (Michaelides *et al*., 2006). Cone receptors are specialized for color vision and require bright light to function. They do not have discrete disks unlike rod receptors, but their outer segment is highly convoluted to increase the surface area. There are three types of cone photopigments whereas rods have only one type of pigment.

Degeneration of PRs is the major cause of adult blindness (Wright *et al*., 2010). The most common type is retinitis pigmentosa (RP) causing adult onset blindness. In RP, primarily rod cells are degenerated. Secondary cone degeneration is also seen in most cases. Another type of PR degeneration is cone-rod dystrophy (CORD); cone cells are more severely affected than rods (Wright *et al*,. 2010). Leber congenital amaurosis (LCA) also results from PR degeneration and resembles RP; however, it is more severe than RP, and onset is in early childhood. In fact, CORD, RP and LCA are clinically similar diseases, and a few of the causative genes are in common.

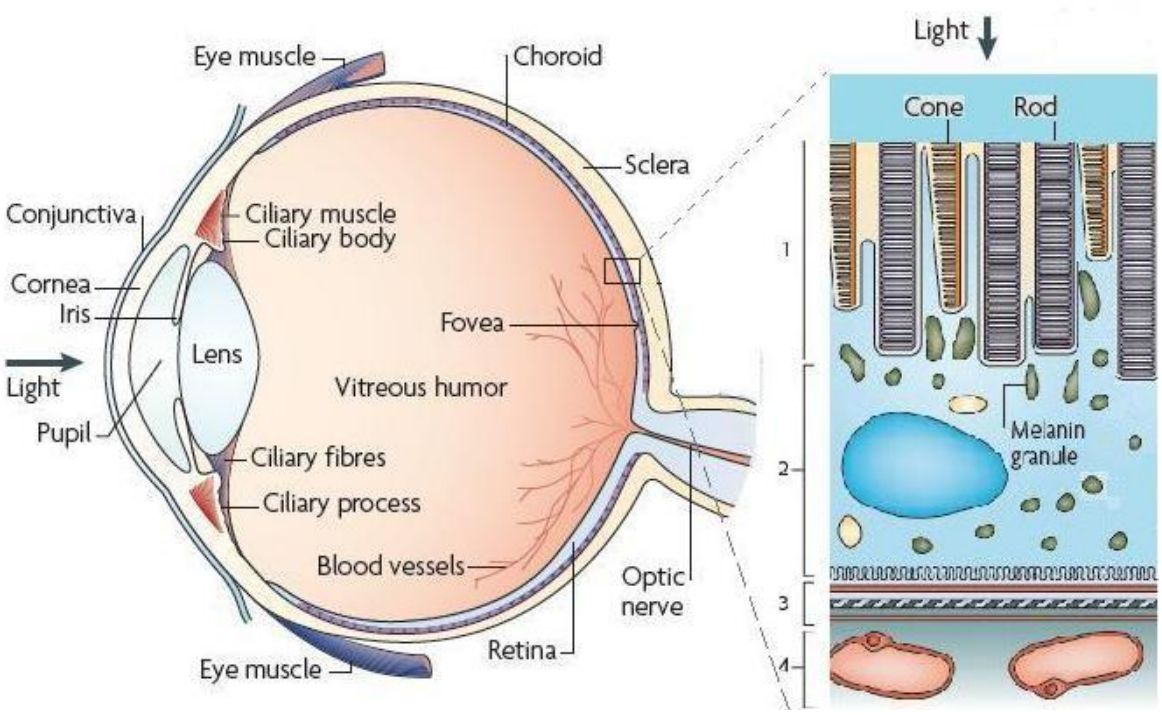

Figure 1.1. Schematic representation of the human eye. Enlarged view of the retina illustrates PR cells (1), the RPE (2), the Bruch's membrane (3), and the choroid (4). Bruch's membrane is a pentalaminar structure consisting of the plasma membrane of the RPE, two collagen layers with an elastin layer in between, and the plasma membrane of the choroidal capillaries (Modified from Wright *et al*., 2010).

1.1.1. Cone-Rod Dystrophy (CORD)

Cone-rod dystrophy (CORD) is a rare, clinically heterogeneous disease that causes blindness in children and young adults. It is characterized by visual field loss, abnormalities in color vision, and variable degrees of nystagmus (a condition of involuntary eye movements) and photophobia caused by progressive degeneration of the rod and cone photoreceptors (Michaelides *et al*., 2006; Traboulsi, 2012). Cones are degenerated, causing an early decrease of visual acuity and color vision. This is followed by rod degeneration leading to nyctalopia (night blindness) and visual field loss (Moore, 1992). Rod involvement is important in differential diagnosis of CORD and RP (also known as progressive cone dystrophy), in which rod degeneration is either not seen or occurs later compared to CORD (Michaelides *et al*., 2006). Electrophysiological tests such as electroretinography (ERG) and electrooculogram (EOG) are helpful in the diagnosis of CORD (Hamel, 2007).

The most common mode of inheritance in familial CORD is autosomal dominant (AD) (Michaelides *et al*., 2006). However, autosomal recessive (AR) and X-linked recessive (XR) inheritance are also observed. The loci and genes identified in CORD are summarized in Table 1.1. Allelic heterogeneity was reported for most of those genes. To date, seven genes have been associated with AR-CORD.

In this study, a large consanguineous family with four members afflicted with CORD was investigated (Figure 1.2). Thirteen members of the family were available for study. Detailed ophthalmologic examinations were performed on the patients plus five of their healthy relatives. The ages and clinical findings in the patients are given in Table 1.2. Clinical examinations included visual acuity tests, visual field examinations, intraocular pressure measurements, fundus examination after pupil dilation, ERG, and EOG. Genotyping was performed on the DNA samples of all available family members using single nucleotide polymorphism (SNP) markers. Via linkage mapping using genotyping data and subsequent mutation screening in the genes at the candidate locus, we identified a homozygous missense mutation in a centriolar gene (*POC1B*, see Section 1.1.3) not previously associated with a disease.

Table 1.1. Chromosomal loci and causative genes for CORD.

Name, OMIM number	Locus	Gene, OMIM number	Inheritance pattern
CORD1, 600624	18q21.1-21.3	Not identified	AR
CORD2, 120970	19q13.1-13.2	*CRX*, 602225	AD
CORD3, 604116	1p22.1	*ABCA4*, 601691	AR
CORD5, 600977	17p13.2	*PITPNM3*, 608921	AD
CORD6, 601777	17p13.1	*GUCY2D*, 600179	AD, AR
CORD7, 603649	6q14	*RIM1* , 606629	AD
CORD8, 605549	1q12-24	Not identified	AR
CORD9, 608194	14q11	*RPGRIP1*, 605446	AR
CORD10, 610283	1q22	*SEMA4A*, 607292	AR
CORD11, 610381	19q13.3	*RAXL1*, 610362	AD
CORD12, 612657	4p15.32	*PROM1*, 604365	AD
CORD14, 602093	6p211	*GUCA1A*, 600364	AD
CORD15, 613660	10q23.1	*CDHR1*, 609502	AR
CORD16, 614500	8q22.1	*C8orf37*, 614477	AR
CORD17, 615163	10q26	Not identified	AD
CORD18, 615374	4p15.33	*RAB28*, 612994	AR
CORD, 604393	17p13.2	*AIPL1*, 604392	AD
CORDX1, 304020	Xp11.4	*RPGR*, 312610	XR
CORD	17q11.2	*UNC119*, 604011	-

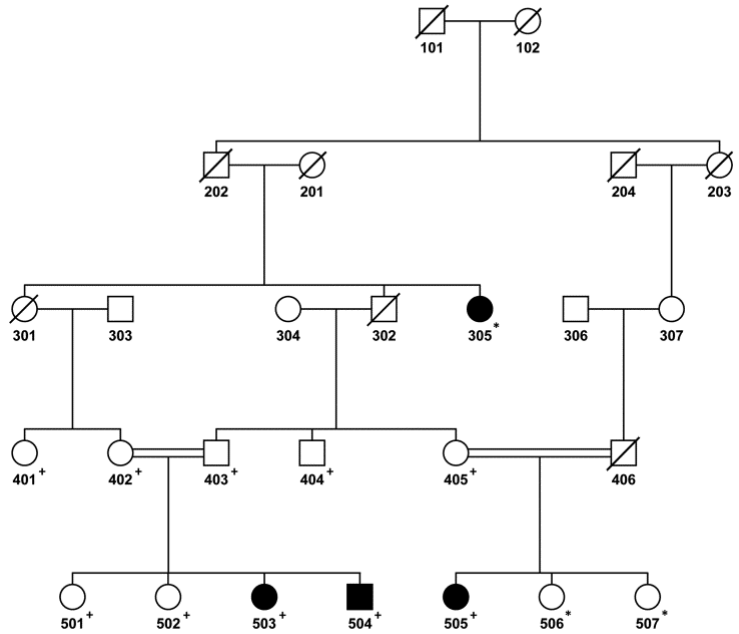

Figure 1.2. Partial pedigree of CORD family. Plus sign indicates SNP genotyping with 370,000 markers and asterisks with 610,000 markers (see Section 4.2.3).

Table 1.2. Ocular Findings in CORD Patients.

Patient	Age (years)	Photophobia	Dyschromatopsia	Nystagmus	Chorioretinal Atrophy
305	61	+	+	-	Moderate in the left eye
503	28	+	+	+	Mild in the left eye
504	27	+	+	+	None
505	17	+	+	-	None

1.1.2. Leber Congenital Amaurosis (LCA)

Leber congenital amaurosis (LCA) is a very severe, nonsyndromic retinal dystrophy with onset within the first year of life. The disease is characterized by early vision loss, nystagmus and amaurotic pupils, i.e., pupils do not contract in response to light. Electrical signal in ERG is usually absent (den Hollander *et al.*, 2008). LCA accounts for over 5% of all inherited retinal dystrophies (Koenekoop, 2004). Inheritance pattern is predominantly autosomal recessive, and a few dominant cases and heterozygous de novo mutations are reported (den Hollander *et al*., 2008). To date 17 genes have been identified (Table 1.3). *CEP290*, *GUCY2D* and *CRB1* are the most frequently mutated LCA genes. Compound heterozygous mutations are reported for most of the LCA genes. Genetic heterogeneity of LCA leads to phenotypic variability in features such as retinal appearance, refractive errors, photoaversion, and nyctalopia among patients with mutations in different genes (den Hollander *et al*., 2008). Disease progress also may vary: LCA patients with *CEP290* or *GUCY2D* mutations remain stable after experiencing significant vision loss whereas patients with *AIPL1* or *RPGRIP1* mutations have progressive vision loss (Dharmaraj *et al*., 2004; Koenekoop *et al*., 2007).

Most of the LCA genes are specifically expressed in the retina, but *CEP290* and *LCA5* are ubiquitously expressed. Products of LCA genes are involved in phototransduction, photoactivation of receptor pigments, photoreceptor development and/or ciliary transport (den Hollander *et al*., 2008).

LCA affects both rod and cone photoreceptors, hence the phenotype is similar to both CORD and RP. Juvenile retinitis pigmentosa is considered to be a less aggressive form of LCA (Gu *et al*., 1997). Indeed, a comparison of Table 1.1 and Table 1.3 reveals

that CORD and LCA have several causative genes in common, namely, *GUCY2D*, *AIPL1*, *RPGRIP1,* and *CRX*. Besides, a few LCA genes are implicated in RP as well.

Table 1.3. Chromosomal loci and causative genes for LCA.

Name, OMIM number	**Locus**	**Gene, OMIM number**
LCA1, 204000	17p13.1	*GUCY2D*, 600179
LCA2, 204100	1p31.3-p31.2	*RPE65*, 180069
LCA3, 604232	14q31.3	*SPATA7*, 609868
LCA4, 604393	17p13.2	*AIPL1*, 604392
LCA5, 604537	6q14.1	*LCA5*, 611408
LCA6, 613826	14q11.2	*RPGRIP1*, 605446
LCA7, 613829	19q13.33	*CRX*, 602225
LCA8, 613835	1q31.3	*CRB1*, 604210
LCA9, 608553	1p36.22	*NMNAT1*, 608700
LCA10, 611755	12q21.32	*CEP290*, 610142
LCA11, 613837	7q32.1	*IMPDH1*, 146690
LCA12, 610612	1q32.3	*RD3*, 180040
LCA13, 612712	14q24.1	*RDH12*, 608830
LCA14, 613341	4q32.1	*LRAT*, 604873
LCA15, 613843	6p21.31	*TULP1*, 602280
LCA16, 614186	2q37.1	*KCNJ13*, 603208
LCA17, 615360	8q22.1	*GDF6*, 601147

We studied a family with three sibs afflicted with LCA. Clinical evaluations were performed on one of the patients, and the clinician stated that the phenotype is compatible with LCA2, caused by mutations in *RPE65*. The patients, their unaffected sister, and the parents were available for study, and SNP genotyping was performed on their DNA samples. Also, a sample of one patient was subjected to whole-exome sequencing. Via linkage analysis and haplotype inspection, we identified several small chromosomal loci where the three patients shared homozygosity. Exome sequencing data were evaluated for homozygous mutations at those loci and later for compound heterozygous mutations in the whole-exome. Finally, Sanger sequencing analysis of a few exons of known LCA genes, which were not sufficiently covered by exome sequencing, revealed a heterozygous nonsense mutation in *AIPL1*. This mutation is reported in the literature.

Parental consanguinity was not definite, thus we assumed the parents as third cousins when performing linkage analysis.

1.1.3. Centriolar Proteins in Photoreceptor Degeneration

Centrioles are cylindrical structures composed of microtubules. Together with pericentriolar material, they form centrosomes that serve as microtubule organizing centers in eukaryotes and function in mitosis. Centrioles also play a key role in cilia formation. They move to plasma membrane during ciliogenesis and serve as template for the extending microtubules that make up cilia (Dawe *et al*., 2007). Those template centrioles are called basal bodies. Cilia are on many cell types in humans. Mutations in centriolar proteins may result in defective ciliogenesis or ciliary function, thereby causing various pathologies such as cystic kidney disease, polydactyly, retinal degeneration, obesity, and chronic respiratory problems (Afzelius, 2004; Badano *et al*., 2006). Several genes coding for centriolar proteins are implicated in retinal degeneration. *RPGR* and *UNC119* are associated with CORD, *C8orf37* with CORD and RP (Table 1.1), *CEP290* with LCA phenotype (den Hollander *et al*., 2006), and *NPHP4* with Senior-Loken syndrome, which has RP as its main feature (Otto *et al*., 2002).

POC1B protein also is part of the centriole proteome and functions in centriole duplication and length control (Keller *et al*., 2009). It is required for primary cilia formation in human RPE cells (Pearson *et al*., 2009).

1.2. Optic Atrophy (OPA)

Optic Atrophy (OPA) can manifest either as an isolated case or as a part of a syndrome. The main characteristics are early onset visual disturbances with moderate to severe loss of visual acuity, color vision deficits and optic disk pallor (Votruba *et al.*, 1998). Autosomal dominant, autosomal recessive and X-linked recessive inheritance patterns are observed. To date, seven loci have been mapped and three causative genes identified for nonsyndromic OPA. *OPA1* gene at chromosome 3q29 is associated with AD-OPA and codes for a mitochondrial protein similar to dynamin-related GTPase, which regulates the stability of mitochondrial network (Yu-Wai-Man *et al.*, 2010). *OPA3* at 19q13.32 is also associated with AD-OPA and codes for a protein of unknown function that localizes to mitochondria (Huizing *et al*., 2010). Mutations in *TMEM126A* at 11q14.1-

q21 cause OPA7, and it is the only known gene for AR-OPA. TMEM126A protein also is a mitochondrial protein of unknown function (Hanein *et al*., 2009).

In many OPA cases, symptoms may be accompanied by extraocular complaints, such as deafness, muscle cramps and ataxia (Yu-Wai-Man *et al.*, 2010). We studied a large consanguineous family with nine individuals afflicted with an autosomal recessive neurological disease (Figure 1.3). Four affected individuals were deceased; two of them were blind. Of the five patients available for study, four were afflicted with bilateral OPA with or without additional findings, such as skeletal deformities and mental retardation. The remaining patient had psychiatric and neurological complaints without visual impairment. Neurological examinations were performed on the five patients and another one before her death. Clinical evaluations are presented in Table 1.4. Blood samples from the five surviving patients plus their fifteen relatives were available. DNA samples for all 18 participants were genotyped with SNP markers (Figure 1.3).

Linkage mapping using SNP genotype data and subsequent exome sequencing performed on one patient sample identified a homozygous missense mutation in *CEP290* (see section 1.1.3) which was previously associated with LCA and some syndromic disorders (MIM#610142). Further linkage analyses were performed in the search for a second disease locus, which could underlie the symptoms other than optic atrophy in the family, and a candidate variant in *ACSS3* gene was detected via exome sequence analysis.

Table 1.4. Clinical Findings in the Patients in OPA Family.

Patient	**Age (years)**	**Optic Atrophy**	**Other Findings**
401	Deceased, NA	NA	NA
403	50	-	Dementia, seizures, psychiatric findings (delusions, delirium), cannot walk, no speech.
405	Died in the fifth decade	-	Dementia, seizures, quadriparesis. Similar prognosis to 403.
408	35	+	-
501	Died at 12	+	NA
502	Died at 11	+	NA
506	20	+	-
508	16	+	Difficulty in walking and speaking, followed by regression (currently bedridden), hearing loss, kidney malfunction, fibrotic changes in liver.
512	7	+	Mental and motor retardation. Similar phenotype to 508.

NA: Data not available

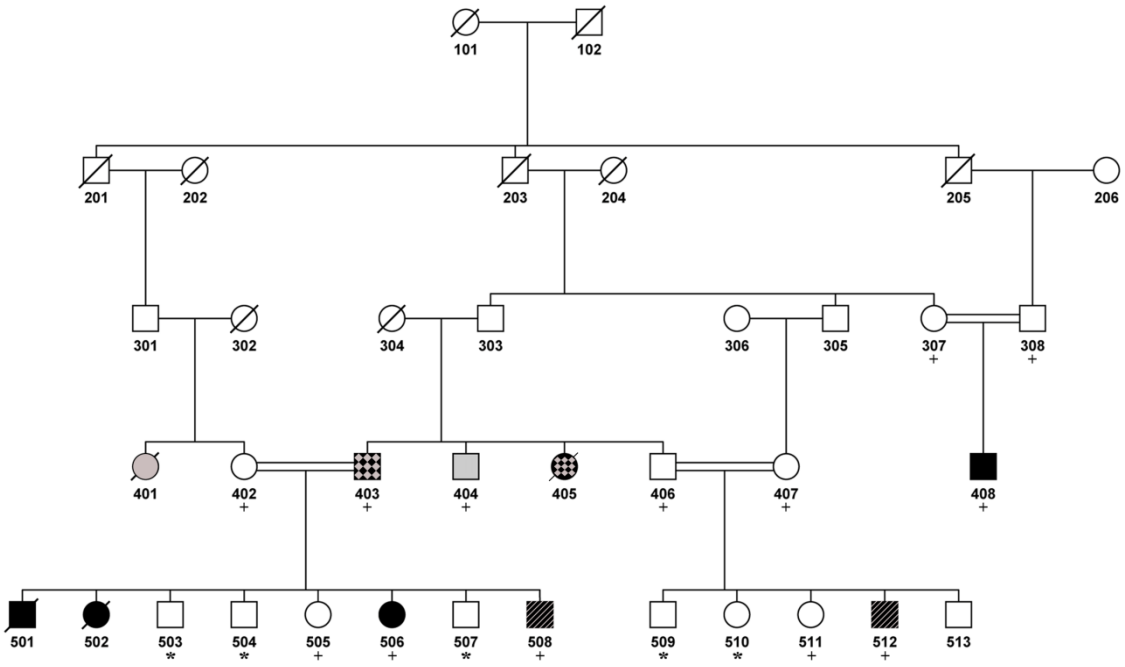

Figure 1.3. Partial pedigree of OPA family. Plus sign indicates SNP genotyping with 1 million markers and asterisks with 700,000 markers. Individuals with unclear disease status is shown in grey; individuals, who do not have OPA but neurological symptoms, in checkered black; and individuals, who have OPA with additional findings, are shown in black with lines.

1.3. Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic Lateral Sclerosis (ALS) is a progressive and lethal neurodegenerative disorder, characterized by degeneration of both upper and lower motor neurons leading to paralysis of voluntary muscles (Pasinelli and Brown, 2006). The worldwide incidence is 0.40 to 1.76 per 100,000 each year (Victor and Ropper, 2005). Approximately 10% of all ALS cases are familial (FALS, MIM#105400). Early manifestation of the disease is weakness in the distal parts of limbs or less frequently in bulbar muscles (Sreedharan and Brown, 2013), although the percentage of the latter increases with increasing age of onset (Haverkamp *et al*., 1995). Those initial symptoms progress to paralysis of those muscles and later other muscles and eventually lead to death, usually due to respiratory failure. The pathological findings in ALS are atrophic motor neurons and accumulation of phosphorylated neurofilaments, Lewy body-like inclusions and/or ubiquitinated material in the axons (Pasinelli and Brown, 2006). The age of onset is usually after 50 years. Early

onset forms of ALS commonly start in the third and fourth decades of life (Sabatelli *et al*., 2008). FALS generally has an earlier onset compared to sporadic ALS (Mulder *et al*., 1986). Juvenile forms of ALS are almost always familial, and onset is before the age of 25 years. Survival is typically three to five years; however, patients with early onset have longer survival than the classic adult onset phenotype (Sabatelli *et al*., 2008).

In FALS, the inheritance pattern is usually autosomal dominant. Autosomal recessive and X-linked forms are also present. About 25 causative genes have been identified for FALS (Sreedharan and Brown, 2013), and only three of them are responsible for AR-FALS, namely, *ALSIN* (MIM#205100), *SPATACSIN* (Orlacchio, 2010), and *SIGMAR1* (MIM#614373). Mutations in *OPTINEURIN* (MIM#613435), *FUS* (MIM#608030) and *SOD1* (MIM#105400) can cause both AD- and AR-FALS. Identification of ALS genes enables a better understanding of disease mechanisms. *SOD1* was the first gene identified for FALS, which codes for the superoxide dismutase 1 enzyme. This protein is ubiquitously expressed and catalyzes the detoxification of cellular superoxide. Gain-of-function mutations in *SOD1* may cause cytotoxicity due to aberrant enzyme and also aggregation of misfolded SOD1 (Cleveland, 1999). This gain-of-function toxicity further results in mitochondrial dysfunction, axonal transport failure, axonal degeneration, and failure in the chaperone and proteosome systems which can explain cytoplasmic inclusions (Cleveland and Rothstein, 2001; Sreedharan and Brown, 2013). Discovery of other ALS genes revealed that defective DNA repair, RNA processing and angiogenesis also play roles in ALS pathogenesis (Dion *et al*., 2009; Sreedharan and Brown, 2013).

We investigated a large consanguineous family with five male members afflicted with ALS (Figure 1.4). The mothers of the patients were sisters. Fifteen members of the family were available for the study, and all of them were given a neurological examination. The ages of patients, disease onset, and their main clinical findings are given on Table 1.5. The mean age of onset was 26 years (range 12-35), and the disease had slow progression. Only one patient had bulbar symptoms. Neurological examinations of the other family members were unremarkable, except that one of the mothers (403 in Figure 1.4) had multiple sclerosis.

By linkage mapping using SNP genotypes of the family members and subsequent exome sequencing analysis, we found in the patients a rare, homozygous truncating mutation in one of the *ALDH* superfamily genes, *ALDH1B1*.

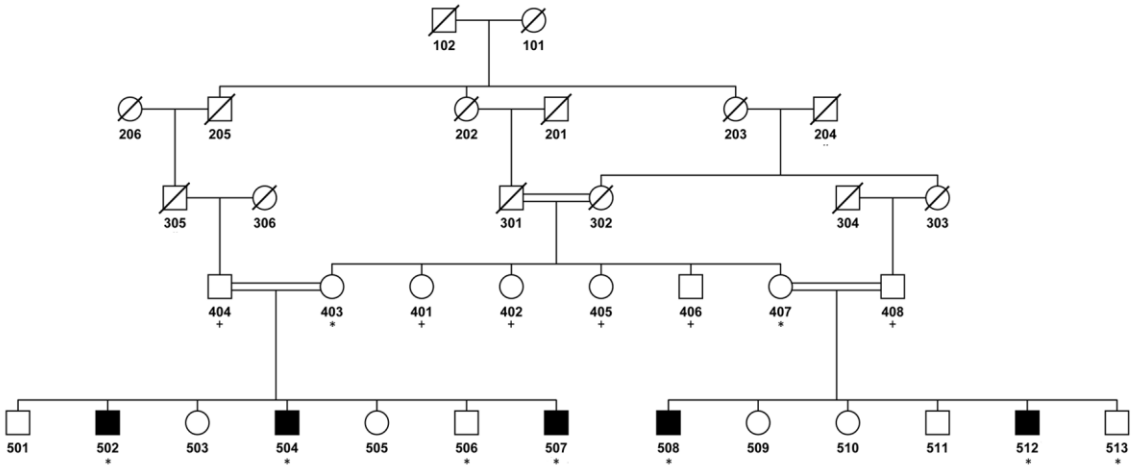

Figure 1.4. Pedigree of ALS family. Asterisks denotes individuals included in SNP genotyping with 1 million markers and plus denotes individuals available for mutation screen only (see Section 4.2.3).

Table 1.5. Clinical Findings in ALS Patients.

Patient	**Age**	**Age of disease onset**	**Main clinical features**	**Loss of ambulation (age in years)**
502	47	31	Atrophy in arms and distal legs, tetraparesis	36
504	44	35	Atrophy in arms and distal legs, distal weakness in all limbs, bulbar symptoms	No
507	37	27	Weakness in legs, atrophy in paravertebral and limb muscles	No
508	43	27	Atrophy in all limbs and paravertebral muscles, weakness in forearms and distal legs	33
512	28	12	Atrophy in legs and paravertebral muscles, generalized weakness,	17

1.4. Rett Syndrome (RTT)

Rett Syndrome (RTT) is a neurodevelopmental disorder with an incidence of 1 in 10,000 female births (Neul *et al*., 2010). It is characterized by microcephaly, loss of acquired skills, stereotypic hand movements, gait abnormalities, seizures, scoliosis, autistic behavior, and mental retardation (Weng *et al*., 2011). Microcephaly is not congenital; it is a result of decelerated head growth that begins after 2-3 months of age (Schultz *et al*., 1993). In classic RTT syndrome, infants generally develop normally in their first year of life but later begin losing speech and purposeful hand movements and develop RTT symptoms. The disease has a broad spectrum of clinical symptoms; however, developmental regression is an essential characteristic for RTT (Neul *et al*., 2010).

Inheritance pattern of RTT is predominantly X-linked dominant. It is almost exclusively seen in females and caused by mutations in *MECP2* at Xq28 (Amir *et al*., 1999). Hemizygous male fetuses usually die in utero. If they survive to term, they develop severe neonatal-onset encephalopathy (Zeev *et al*., 2002). *MECP2* mutations are responsible for classical RTT syndrome; however, mild mutations lead to Zapella variant of RTT, in which speech is preserved (Renieri *et al*., 2009). Two other genes identified for RTT are *CDKL5* at Xp22 and *FOXG1* at 14q12. Heterozygous *CDKL5* mutation can result in both classic RTT and an atypical form with infantile spasms (Scala *et al*., 2005; Russo *et al*., 2009), and heterozygous *FOXG1* mutation causes a congenital variant of RTT (Kortum *et al*., 2011).

We studied a small consanguineous family with two daughters afflicted with RTT (Figure 1.5) and an unaffected daughter. The parents were first cousins. Neurologic examination was performed on the parents and sibs. Symptoms of RTT patients include microcephaly, epileptic seizures, psychomotor retardation, stereotypic hand movements, abnormal gait, thoracic scoliosis, and constipation. In both patients, disease began in infancy with epileptic seizures as the initial symptom. Their sister also had scoliosis, which was the only RTT sign in her. The two severely affected sisters are diagnosed with RTT due to the clinical findings. However, the criteria of regression in head growth and acquired skills were not considered, since the patients never acquired speech or hand skills. Their microcephaly is likely not congenital. Homozygosity mapping using genome scan

data of all available family members including the grandmothers (Figure 1.5) identified candidate disease loci, and a homozygous truncating mutation, which was assessed to cause the disease, was identified by exome sequence analysis.

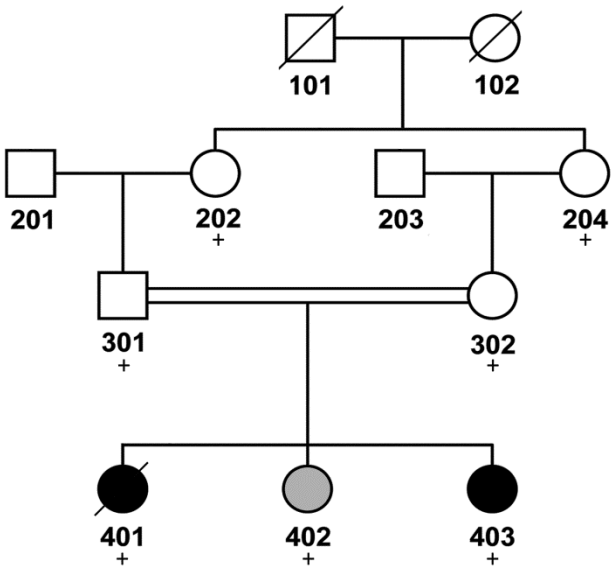

Figure 1.5. Partial pedigree of RTT family. Plus sign indicates individuals whose DNA samples were genotyped with 700,000 SNP markers (see Section 4.2.3). The mildly affected sister is shown in grey.

1.5. Infantile Neuroaxonal Dystrophy (INAD)

Infantile neuroaxonal dystrophy (INAD) is a hereditary disorder with autosomal recessive mode of inheritance. Onset is in infancy, usually between 6 months and 2 years of age. Clinical manifestations include bilateral pyramidal tract signs, truncal hypotonia, cognitive decline, and visual disturbances such as strabismus, nystagmus and optic atrophy (Gordon, 2002). The most typical pathological finding is distal axonal swellings and spheroid bodies, which are detectable in the skin, rectum or conjunctiva biopsies. The disease eventually leads to death, usually before the age of ten years.

To date, the only gene shown to be involved in INAD pathogenesis is *PLA2G6* at 22q13.1 (Khateeb *et al*., 2006; Morgan *et al*., 2006). However, linkage analysis in 12 families indicated locus heterogeneity for INAD, supporting the existence of at least one additional locus (Morgan *et al*., 2006). *PLA2G6* encodes iPLA2-VI, a calcium-independent

phospholipase, which is involved in cell growth via the generation of lipid signals (Hooks and Cummings, 2008) and cell membrane homeostasis, such as regulation of phosphatidylcholine levels (Baburina and Jackowski, 1999). Clinical phenotype of INAD may vary among patients with *PLA2G6* mutation (Kurian *et al*., 2008; Tonelli *et al*., 2010).

In this study, six INAD families were investigated. Patients in all families were born to consanguineous parents, usually first cousins. All families were small with one or two affected sibs, except that Family 1 was large with two affected cousins. Family 4 had two affected sibs with an unusually severe phenotype that led to death of both children before the age of two. In this family, and also in families 1, 2 and 3, mutations in *PLA2G6* were detected. All mutations were novel.

The phenotypes of the affected sibs in Family 5 (Figure 1.6) were atypical for INAD in that they had facial dysmorphism, skeletal anomalies and a much longer life span as compared to classical INAD patients. However, they had severe growth retardation: At the last visit the 18-years old male patient weighed 9.2 kg and had a height of 110.5 cm (Figure 1.7), and the 21-years old female patient weighed 9.7 kg and had a height of 112.5 cm (all measurements are <<-5 SD). Via linkage analysis using genome scan data of family members and subsequent exome sequencing, we identified a truncating mutation in a sodium leak channel gene, *NALCN*, which was previously not associated with a disease.

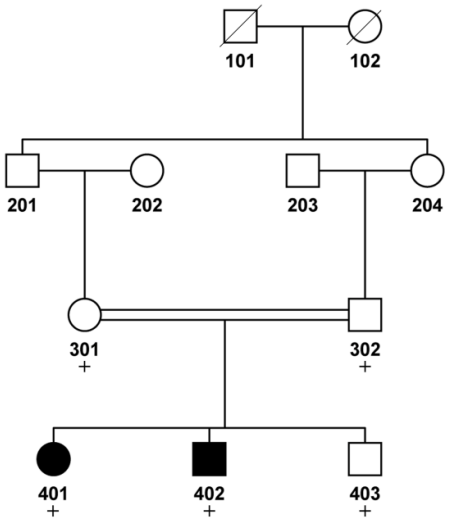

Figure 1.6. Partial pedigree of INAD family 5. Plus sign denotes genotyping with 610,000 SNP markers (see Section 4.2.3).

Figure 1.7. Patient 402 in INAD family 5 at the age of 18 years.

1.6. Linkage Analysis

Genetic linkage analysis is a powerful method to identify the chromosomal region that contains a disease gene in a family. It depends on the principle that two genetic loci are linked if they are inherited together more often than expected under independent segregation. During meiosis, recombination occurs with a probability of 50% between unlinked loci, i.e. loci that are far apart. The closer two loci are on the chromosome, the less likely that a recombination event would occur and prevent cosegregation, hence the recombination probability drops below 50%. This probability is defined as "recombination fraction", which is denoted by θ, and it is the main quantity of interest in the parametric linkage analysis. The maximum value for θ is 0.5, indicating that the two loci are either on different chromosomes or so far apart on the chromosome that they segregate independently of each other. Recombination events along a chromosome can be tracked by using closely spaced genetic markers, enabling a disease gene to be mapped to a relatively small region. The unit centimorgan (cM) is used to define genetic distances. One cM corresponds roughly 1% probability for recombination and 1 Mb at the physical map.

Various types of genetic markers can be used in disease gene mapping. Restriction fragment length polymorphisms (RFLPs), variable number of tandem repeats (VNTRs), and short tandem repeats (STRs, also known as microsatellites) were being utilized until

the last decade. RFLPs are DNA sequence variations, usually single base pairs, which create or abolish restriction enzyme recognition sites. They span the genome with more than 10-cM spacing (Donis-Keller et al., 1987). Since these markers can only have two alleles, they have limited informativeness. VNTRs are short tandem repeats with sizes ranging from 14 to 100 base pairs, and they are multiallelic, providing more information. However, they are not evenly distributed throughout the genome. The drawbacks of both RFLPs and VNTRs were overcome by microsatellites, which are repeating sequences of two to five base pairs. Due to their supreme features, such as being highly polymorphic, evenly distributed and more densely spaced, microsatellites were more useful in gene mapping, compared to RFLPs and VNTRs.

After the discovery of a vast number of single nucleotide polymorphisms (SNPs), mainly as the achievement of the HapMap Project, and the advancements in the microarray technology that allowed cost effective high-throughput SNP genotyping, SNP markers became the preferred markers for gene localization. The main advantage of SNP markers is that they are much denser in the genome as compared to all the other types of genetic markers. At present the commonly used SNP chips contain 300,000 to 2.5 million markers that provide huge data for linkage analysis. However, one drawback of SNP markers is the low informativeness, as they have only two alleles. This drawback can be compensated by choosing the multipoint analysis method for linkage, which is explained below.

1.6.1. Parametric Linkage Analysis and LOD Scores

Linkage analysis can be applied to both Mendelian and complex diseases. Parametric (model-based) linkage analysis is used to study Mendelian diseases. In this approach, the mode of inheritance for the disease is modeled by setting genetic parameters such as penetrance and disease-allele frequency. In complex diseases, however, inheritance pattern cannot be described accurately, since those diseases are caused by interactions of two or more genes and environmental factors. Nonetheless, a susceptibility gene can be localized by searching shared haplotypes that are identical by descent (see Section 1.6.2) in the affected kin. This approach is the basis of nonparametric (model-free) analysis, the method of choice for complex diseases (Dawn Teare and Barrett, 2005).

Parametric linkage analysis utilizes a likelihood ratio test called LOD (logarithm of odds) score analysis introduced by Morton (1955). The null hypothesis that θ, the recombination fraction between the marker and the disease, is 0.5 (no linkage) is tested against the alternative hypothesis θ <0.5 (linkage). The probability of observing the particular distribution of the marker and the disease allele in the family is calculated under the assumption of no linkage versus assuming linkage (Risch, 1992). The logarithm of the ratio of the latter probability to the former, considered as a function of θ, is the LOD score function denoted by Z(θ). The maximum value of Z(θ) is called the LOD score, and the null hypothesis of no linkage is rejected for LOD scores above 3, which means that the data are at least 1000 times more likely to have arisen under linkage than under no linkage. The threshold LOD score 3 is equivalent to a p-value of 0.0001. (Dawn Teare and Barrett, 2005). However, since the LOD score is dependent on the pedigree size as well, i.e. large pedigrees are expected to yield LOD scores >3 whereas small pedigrees with close degree of consanguinity do not have the power to yield a LOD score of 3, the threshold of 3 need to be modified accordingly. The corresponding θ for the LOD score is the maximum likelihood estimate (MLE) of θ. A negative LOD score is evidence for no linkage, and linkage is rejected if the LOD score is below -2, which means that the data are at least 100 times more likely to have arisen under no linkage than under linkage. For LOD scores between -2 and 3, the findings are inconclusive for linkage.

LOD scores are calculated either by two-point analysis or multipoint analysis. Two-point calculations are done by evaluating cosegregation of one marker at a time with the disease locus. This analysis is based on the estimation of the recombination fraction between the tested marker and the disease locus. In multipoint analysis, more than one marker is included at a time in the calculation, and the position of the disease locus is estimated relative to a fixed map of markers with known locations. Thus, in this approach the LOD score is the function of chromosomal position in cM, and not θ. Multipoint analysis is preferred in linkage analysis using SNP markers, since it can overcome the above-mentioned drawback of limited informativeness through considering several markers in a row together as a haplotype.

Various open source software are available for computerized linkage analyses. Each of them is designed to perform either a two-point or a multipoint analysis or both at the

same run. Depending on the algorithm the software uses, the calculation time needed and the maximal usable pedigree size vary. For instance, SimWalk software uses a Markov chain Monte Carlo method for calculations (Sobel and Lange, 1996), which enables the program to run with virtually any pedigree size. However, the time it takes to complete the calculation is much longer as compared to other software such as Allegro, which uses hidden Markov model (Gudbjartsson *et al*., 2000). Even though it is very fast, Allegro can only perform linkage analysis on small pedigrees. The size of the pedigree is measured in bits using the formula 2n-f, where n and f are the numbers of non-founders (individuals whose parents are shown on the pedigree) and founders (parents not shown), respectively. Allegro can handle pedigrees at the most 20 bits in size. Hence, the selection of software depends on the size of the pedigree studied and whether the analysis would be two-point or multipoint.

Each linkage software requires a different input format and is run by unique command lines, usually on a UNIX platform. easyLINKAGE package integrates several software and runs them on Microsoft Windows. It enables one to run different linkage analysis software on a user friendly interface using standardized input data formats. The results are presented in graphic outputs (Hoffman and Lindner, 2005).

1.6.2. Homozygosity Mapping

Homozygosity mapping is an efficient method for mapping autosomal recessive diseases in consanguineous families. The principle of this application is that at the disease locus, affected family members have most likely received the same haplotype in the homozygous state from a single common ancestor (Lander and Botstein, 1987). This method is also an application of parametric linkage analysis and can be used either alone or in parallel with LOD score calculations. Regions with shared homozygosity in affected individuals are detected by using appropriate software or inspecting the genotypes of family members on MS Excel. At those regions, haplotypes are constructed to investigate whether the homozygosity is possibly identical by descent (IBD), which means that both paternal and maternal haplotypes descended from the same common ancestor. It is possible for two individuals to share the same haplotypes that are not IBD. Such haplotypes are said to be identical by state (IBS).

1.7. Whole Exome Sequencing

In the last decade, applications for genome analysis shifted from Sanger sequencing to massively parallel sequencing. Since the Sanger sequencing method is considered as a "first-generation" technology, the new method is widely referred to as next-generation sequencing (NGS). There has been a drastic advance in NGS platforms in the last years, and they have become widely available due to their ability to produce an enormous amount of sequence data inexpensively and time effectively as compared to Sanger sequencing. NGS is able to generate billions of reads in a single run, each 35-250 bp in length (Metzker, 2010); however, these reads are not as accurate as those produced by Sanger sequencing, which is a drawback of this technology. Nevertheless, this drawback is mostly compensated by increasing the target coverage by generating multiple reads for the target genomic regions. The more deeply a region is covered, the more robust the sequencing result becomes. Most NGS platforms typically aim 50X-100X on target coverage, i.e. coverage of sequence reads that can be aligned to the reference sequence.

NGS technology is useful in various applications, such as whole genome sequencing, generation of de novo genomic assemblies of different species, whole transcriptome analysis (RNA-seq), genome-wide profiling of epigenetic marks and chromatin structure (methyl-seq, ChIP-seq), and mutation discovery by sequencing targeted regions of interest (Metzker, 2010). Whole exome sequencing falls into the last category of these applications, in which all the protein-coding sequences in the genome are selectively analyzed. The method consists of two main steps: targeted capture of the exons and their flanking regions with DNA or RNA probes followed by next generation sequencing. Exome capture is performed either as microarray-based (solid phase) or more commonly solution-based (liquid phase), which is the method used throughout this study as well (see Section 4.4.1).

The exome represents a highly enriched subset of the genome useful in a search for disease causing variants, because the majority of disease-causing mutations identified are in protein coding sequences (Stenson *et al*., 2009). Thus, exome sequencing has become a very powerful tool in identifying new Mendelian disease genes and has been successful in

approximately 60% of the projects (Gilissen *et al*., 2012). In our experience, a higher success rate can be achieved through careful bioinformatics analyses.

The main challenge in disease gene identification via exome sequencing is to single out the disease-causing variant from the background of benign polymorphisms, false nucleotide calls due to sequencing and alignment, and bioinformatics analysis errors. Exome sequencing identifies tens of thousands of variants per exome. On average, it detects approximately 20,000 to 24,000 variants in Americans (Bamshad *et al*., 2011). Since only one or two of those variants may explain the pathogenesis of the studied disease, a good variant prioritization strategy is essential. Mutations underlying severe Mendelian disorders are not expected to be common in the population; they should either be unique to the study family or have a very low frequency in the population. Based on this idea, common variants reported in the databases may be excluded. This first filter reduces the number of variants to be investigated by 90-95% (Gilissen *et al*., 2012). Variants outside the coding regions as well as synonymous variants may also be filtered out, since chances of mutations of severe diseases being located within the protein coding regions are higher as compared to other regions, and it is easier to deduce the effect of a variant in the coding region on protein function. Nonsynonymous or stopgain codon changes or small insertion-deletions (indels) are predicted to be deleterious to protein function (Kryukov *et al*., 2007); this makes coding sequences primary investigation sites for mutation identification in Mendelian disorders. This additional filter further downsizes the variant list; however, there would still be too many variants to distinguish the disease causing mutation from. At this point, prior identification of the candidate disease locus via linkage analysis can be very advantageous. This strategy provides the final filter by reducing the candidate variants to a few; one can then more easily identify the disease causing mutation by some further analyses (details are given in the following section). Validation of variants by Sanger sequencing is crucial, since this method is more robust than NGS, as mentioned above.

1.8. Disease Gene Identification, Step by Step

In this section, all the strategies used for disease gene identification throughout this study are summarized. For detailed explanations please see Chapter 4 (Methods).

1.8.1. Beginning a Project

- Written informed consent is obtained from or for all participants of the family to be studied. Study protocol should be approved by the Boğaziçi University Institutional Review Board for Research with Human Participants for the study to be performed.
- Blood samples are collected from the available members of the family. Phenotype of each individual is defined, and a pedigree is drawn.
- The known genes for the disease, if any, are excluded by genotyping the DNA samples of family members with microsatellite markers flanking the known genes. Alternatively, SNP genotyping is performed on one or two patients and known disease loci are inspected on MS Excel to investigate whether the patients carry a shared homozygous genotype at any of those loci.
- After exclusion of the known genes, DNA samples of the selected family members (usually all affected members, available parents, and some unaffected siblings) are genotyped with SNP markers.

1.8.2. Linkage Analysis and Homozygosity Mapping

- Multipoint LOD scores are calculated currently using SNP markers and assuming recessive inheritance with full or reduced penetrance for disease allele. Calculations are performed selecting markers at 0.07-cM spacing and using them in sets of 30 or 100. If less than approximately 300,000 markers are included in the calculation (due to merging of the data from different chips), 0.01 cM spacing are used instead.
- For pedigrees larger than 20 bits in size, LOD score calculations are first performed using a simplified pedigree ($\leq$ 20 bits). A threshold for LOD scores is set according to the results, and then analyses are repeated at the loci yielding LOD scores above the threshold value, using the actual pedigree and including 1-cM flanking regions.
- All loci above the threshold score are listed and genotypes at each of them are inspected on Excel to determine whether the patients share homozygosity, in which case the locus is designated as a candidate disease locus.
- Homozygosity inspection can also be done using program PLINK. The input files are created using a plug-in on GenomeStudio software, and the PLINK output file is

arranged on Excel. Another method is the detection of loss of heterozygosity (LOH) regions via LOH Detector plug-in on GenomeStudio.

- To further investigate the homozygous regions, CNV analysis may be performed using the cnvPartition plug-in on GenomeStudio.
- Multipoint LOD score calculations are repeated using all markers at the candidate loci to assess final LOD scores.
- Haplotypes are constructed at the candidate loci including the 0.5-cM flanking regions in order to investigate possible identity by descent (IBD).

1.8.3. Candidate Variant Detection by Exome Sequencing

- Exome sequencing is performed on the DNA sample of one of the patients. Alignment (by BWA), variant calling (by SAMTools), and annotation (by ANNOVAR) are done in this order.
- All variants at the candidate loci are listed. Non-coding variants except those that affect splicing, synonymous coding variants, heterozygous variants (with alternative depth <60% of total depth), and variants reported in dbSNP with minor allele frequencies (MAF) >0.01 are filtered out. For a highly sensitive analysis, heterozygous variants can be kept until further analyses, such as visualization on software BamView or IGV and validation by Sanger sequencing.
- Remaining variants are checked on other in-house samples, namely, other exome sequencing results in our laboratory. Previously detected variants are filtered out.
- Total exome sequencing results are analyzed for novel or rare, homozygous, and possibly damaging variants. Listed variants are checked on in-house samples.
- Coverages for targeted exons are computed using BEDTools software. Candidate loci are checked for any uncovered exon.
- Variant calling can be repeated via GATK as an optional step.

1.8.4. Evaluation of Candidate Variants

- Candidate variants are prioritized according to their predicted effect on protein function (synonymous SNVs or SNVs on UTRs are not considered with priority),

severity (truncating SNVs are better candidates than missense SNVs), the function of the gene they reside in, and the expression pattern of the gene.

- For novel variants, novelty is further verified by checking them on databases such as NHLBI Exome Variant Server.
- Candidate variants are validated by Sanger sequencing.
- Segregation of the mutation with the disease in the family is verified by Sanger sequencing or high resolution melting (HRM) curve analysis. However, finding the mutation in the homozygous state in a seemingly unaffected member of the family does not necessarily exclude the mutation. Possibility of incomplete penetrance or variable expressivity should be considered.
- Population samples are screened for the mutation via HRM analysis. At least 108 samples should be used in order to achieve 80% power to rule out polymorphism.
- For missense mutations the following analyses are done: The effect of amino acid substitution is predicted by at least three online tools (Mutation Taster, PolyPhen, SIFT, MutPred or MMB PMut). Conservation of the substituted amino acid is evaluated across different species using the HomoloGene database. The nature of the base change is evaluated, i.e. a polar amino acid may be substituted with a non-polar amino acid.
- For truncating mutations the following questions are addressed: How much of the native protein is lost? Which functional domains are lost? Is the premature stop codon located more than 50-nt upstream of the 3'-most exon-exon junction, possibly leading to non-sense mediated decay?
- Splice site mutations are analyzed by online tool Human Splicing Finder.
- Gene expression in different tissues is evaluated in relevant databases. Quantitative PCR experiments for gene transcripts in selected tissues can also be performed.
- The function of the mutated protein is evaluated for the relevancy to the clinical manifestations of the disease.

1.8.5. Deletion Analysis

- Any deletion detected at the candidate loci by genotype inspection on Excel and subsequent validation by PCR assay is checked for novelty in the Database of Genomic Variants. The deletion is also checked whether it contains part of a gene. If

not, the deleted sequences are investigated thoroughly using ENCODE tracks on UCSC Genome Browser for the presence of regulatory elements such as silencers or enhancers.

2. PURPOSE

The purpose of this thesis study was to localize and identify the genes responsible for six rare inherited disorders using genome scan and exome sequencing data for the families. Two of the studied disorders are retinal dystrophies, namely, Cone-Rod Dystrophy (CORD) and Leber Congenital Amaurosis (LCA). The remaining four disorders are neurological: Optic Atrophy (OPA), Amyotrophic Lateral Sclerosis (ALS), Rett Syndrome (RTT), and Infantile Neuroaxonal Dystrophy (INAD). Additionally, the gene responsible for the extraocular symptoms in the OPA family was searched.

Following the identification of causative genes for ALS, RTT and INAD, quantitative PCR assays were performed with the aim of assessing expressions of identified genes in various tissues.

.

3. MATERIALS

3.1. Subjects

Informed consent was obtained for all subjects who participated in this study. The study protocol was approved by the Committee on Research with Human Participants at Boğaziçi University.

For CORD family, peripheral blood samples of 13 individuals were supplied by Assoc. Prof. Yusuf Durlu at Makula Eye Health, Inc. in Istanbul, and clinical evaluations of the four patients and five unaffected relatives were also performed by him. Blood samples from the three siblings afflicted with LCA, their unaffected sister and the parents were provided by the family. Blood samples of 20 members of OPA family were supplied and clinical examinations on the patients were performed by Hacer Durmuş, M.D. at Istanbul University Istanbul Medical School Department of Neurology. Blood samples from fifteen members of ALS family including all patients, and samples from all siblings, the parents, and the grandparents of RTT family were supplied by Rezzak Yılmaz, M.D. at Ankara University School of Medicine Department of Neurology (Ankara). Neurological examinations for the participating members of ALS family and clinical investigations on one patient were performed by Rezzak Yılmaz, M. D. and Mine Hayriye Sorgun, M. D. at the same institution. Medical records of the two RTT patients were available for the study. The oldest sister was deceased before the study was initiated. Rezzak Yılmaz, M. D. evaluated the medical records for both patients and performed additional neurological examinations on the younger RTT sister, the mildly affected sister and the parents. INAD study included five families. Blood samples for families 1-4 were referred to us by clinicians, and samples for Family 5 were provided by Assoc. Prof. Mehmet Seven at Istanbul University Cerrahpaşa Medical School Department of Medical Genetics. He also clinically evaluated the family members.

3.2. Chemicals

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

3.3. Buffers and Solutions

3.3.1. DNA Extraction from Whole Blood

A list of buffers and solutions used for DNA extraction from whole blood is given in Table 3.1.

Table 3.1. The list of buffers and solutions used for DNA extraction from whole blood.

Name	**Ingredients**
Proteinase K	20 mg/ml proteinase K in dH_2O
Sodiumdodecylsulfate (SDS)	10% SDS (w/v) in dH_2O
Ammonium acetate	7.5 M CH_3COONH_4 in dH_2O
Ethanol	Absolute ethanol
Cell lysis buffer	155 mM NH_4Cl, 0.1 mM Na_2EDTA (pH 7.4), 10 mM $KHCO_3$
Nucleus lysis buffer	400 mM NaCl, 2 mM Na_2EDTA, 10 mM Tris (pH 8.2)
TE buffer	1 mM EDTA, 20 mM Tris-HCl (pH 8.0)

3.3.2. Polymerase Chain Reaction (PCR)

A list of buffers and solutions used for PCR assays is given in Table 3.2.

Table 3.2. The list of buffers and solutions used for PCR assays.

Name	**Ingredients**
5X Combinatorial Enhancer Solution (CES)	2.7 M betaine, 6.7% dimethylsulfoxide (DMSO), 6.7 mM dithiothreitol (DTT), 55 µg/ml bovine serum albumin (BSA)
$MgCl_2$	25 mM $MgCl_2$ (Roche, Germany)

Table 3.2. The list of buffers and solutions used for PCR assays (cont.).

Name	**Ingredients**
dNTP	12.5 mM each of dATP, dCTP, dGTP, and dTTP in dH_2O (Roche, Germany and Fermentas, Lithuania)
Betaine	5 M betaine (Promega, USA)
BSA	10 mg/ml BSA in dH_2O
10 X PCR buffer	20 mM $MgSO_4$, 100 mM KCl, 100 mM$(NH_4)_2SO_4$, 1%Triton X-100, 1 mg/ml BSA, 200 mM Tris-HCl (pH 8.8)

3.3.3. Agarose Gel Electrophoresis

A list of buffers and solutions used for agarose gel electrophoresis is given in Table 3.3.

Table 3.3. The list of buffers and solutions used for agarose gel electrophoresis.

Name	**Ingredients**
Agarose	2% agarose (Pronadisa, Spain) in 0.5 X TBE buffer
Ethidium bromide	10 mg/ml in dH_2O
10 X TBE buffer	20 mM EDTA, 0.89 M boric acid, 0.89 M Trizma base (pH8.3)
6 X Loading buffer	50% Glycerol, 60 mM EDTA, 2.5 mg/ml bromophenol blue and/or 2.5 mg/ml xylene cyanol, 10 mM Tris-HCl (pH 7.6)

3.3.4. Polyacrylamide Gel Electrophoresis (PAGE)

A list of buffers and solutions used for polyarcylamide gel electrophoresis (PAGE) is given in Table 3.4.

Table 3.4. The list of buffers and solutions used for PAGE.

Name	**Ingredients**
40% Acrylamide (stock)	40% acrylamide-bisacrylamide (19:1) in dH_2O
8% Instagel (denaturing)	8% acrylamide-bisacrylamide (19:1), 8.3 M urea in 1 X TBE Buffer
APS	10% ammonium peroxidisulfate
TEMED	N,N,N,N-tetramethylethylenediamine

Table 3.4. The list of buffers and solutions used for PAGE (cont.).

Name	**Ingredients**
10 X sample solution	95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol

3.3.5. Silver Staining

A list of buffers and solutions used for silver staining is given in Table 3.5.

Table 3.5. The list of buffers and solutions used for silver staining.

Name	**Ingredients**
Staining solution	0.1% $AgNO_3$ in dH_2O
Developing solution	1.5% NaOH, 0.01% $NaBH_4$, 0.015% formaldehyde in dH_2O

3.4. Enzyme

Taq DNA polymerase was purchased from Roche (Germany), Fermentas (Lithuania), Quanta Biosciences (USA), or Kapa Biosystems (USA) with supplied PCR buffer.

3.5. Kits

A list of commercial kits used in this study is given in Table 3.6.

Table 3.6. The list of kits used in this study, their usage and manufacturing companies.

Name	**Used for**	**Company**
Dream Taq PCR Kit	Amplification of DNA templates	Fermentas, Lithuania
Taq DNA Polymerase	Amplification of DNA templates	Roche, Germany
AccuStart II Taq DNA Polymerase	Amplification of DNA templates	Quanta Biosciences, USA
KapaTaq ReadyMix DNA Polymerase	Amplification of DNA templates	Kapa Biosystems, USA
Expand Long Range, dNTPack	Amplification of long templates	Roche, Germany

Table 3.6. The list of kits used in this study, their usage and manufacturing companies (cont.).

Name	**Used for**	**Company**
QIAquick PCR Purification Kit	PCR clean-up for sequencing	Qiagen, USA
Transcriptor High Fidelity cDNA Synthesis Kit	Reverse transcription PCR	Roche, Germany
RevertAid First Strand cDNA Synthesis Kit	Reverse transcription PCR	Fermentas, Lithuania
LightCycler 480 High Resolution Melting Kit	Heteroduplex analysis (mutation screening) on LightCycler 480 device	Roche, Germany
Accumelt HRM SuperMix	Heteroduplex analysis on LightCycler 480 device	Quanta Biosciences, USA
LightCycler 480 SYBR Green I Master Kit	Relative quantification on LightCycler 480 device	Roche, Germany

3.6. Oligonucleotide Primers

Oligonucleotide primer pairs used in this study were designed via Primer3 software and checked for secondary structures via Oligo Calc tool and for specificity via in-silico PCR analysis on UCSC Genome Bioinformatics site. Primers were purchased from Massachusetts General Hospital (MGH) DNA Synthesis Core (USA) or Macrogen Inc. (Korea). Lyophilized primers were dissolved in 1 ml dH_2O, and 10 μM dilutions were prepared for PCR.

3.7. DNA Molecular Weight Markers

Lambda DNA/*Hind*III, pUC19 DNA/*Msp*I and GeneRuler 1 kb Plus ladders were purchased from Fermentas (Lithuania), and 50 bp DNA ladder was purchased from Roche (Germany).

3.8. Equipment

A list of equipment used in this study is given in Table 3.7.

Table 3.7. The list of equipment used in this study.

Type	**Details**
Autoclave	AMB430T (Astell, UK)
Balance	Electronic balance (Precisa, Switzerland)
Centrifuge	MiniSpin Plus (Eppendorf, Germany), Allegra X-22R (Beckman Coulter, USA), J2-MC (Beckman Coulter, USA), Universal 16R (Hettich, Germany)
Computer	Gigabyte X58A-UD5 Motherboard, Intel i7 960 (8X) processor, Kingston 1333 Mhz RAM, CPU @ 3.20 Ghz, 12328 Mb Memory, Ubuntu 11.10 Operating System
Deep freezer	-20°C (AEG, Turkey), -20°C (Arçelik, Turkey), -20°C (Bosch, Germany), -80°C ultra freezer (Thermo Scientific, USA)
Documentation system	GelDoc Documentation System with Quantity One 4.6.9 Analysis Software (Bio-Rad, 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)
Incubator	Orbital (Gallenkamp, Germany)
Magnetic stirrer	MR3001 (Heidolph, Germany)
Micropipette	Pipetman (Gilson, France)
Minishaker	Rotamax 120 (Heidolph, Germany)
Oven	KD 200 (Nüve, Turkey)
Power supply	Power Pac Model 3000 (Bio-Rad, USA), Fotoforce 250 Electrophoresis Power Supply (Fotodyne, USA), P250A Power Supply (Sigma-Aldrich, USA)
Refrigerator	-4°C (Arçelik, Turkey)
Spectrophotometer	NanoDrop 1000 (Thermo Scientific, USA)
Thermal cycler	MyCycler (Bio-Rad, USA), T100 ThermalCycler (Bio-Rad, USA), PTC-200 (MJ Research, USA), LightCycler 480 (Roche, Germany)
Vortex	Reax vortex mixer (Heidolph, Germany), Lab Dancer Vario (Roth, Germany)
Water Purification System	Ultra Pure Water Purification System (UTES, Turkey and Hach-Lange, Germany)

3.9. Electronic Databases and Bioinformatics Tools

The electronic databases, online bioinformatics tools, and open-source bioinformatics software used for exome sequencing data analysis are given in Table 3.8.

Table 3.8. Electronic databases, online tools, and bioinformatics software utilized in this study.

Name and web site	Description
Basic bioinformatics	
Ensembl http://www.ensembl.org	A database of reference sequences for a large number of genomes
GeneDistiller2 http://www.genedistiller.org/	Lists genes in linkage intervals and supplies genomic and functional information about them
NCBI HomoloGene http://www.ncbi.nlm.nih.gov/homologene/	Lists putative orthologs of a given gene and performs multiple protein alignment
NCBI Human Genome Resources http://www.ncbi.nlm.nih.gov/genome/guide/human/	Supplies reference sequence for human genome
NCBI UniGene http://www.ncbi.nlm.nih.gov/unigene/	Expression data of genes in various tissues
Online Mendelian Inheritance in Man http://www.ncbi.nlm.nih.gov/Omim	A database of human genes and associated phenotypes
UniProt (Universal Protein Resource) http://www.uniprot.org/	A database of protein sequence and structural and functional information
UCSC Genome Browser http://genome.ucsc.edu/	Provides reference sequences for a large number of genomes and portals to the ENCODE project
Annotation of genomic variants	
1000 Genomes http://browser.1000genomes.org/index.html	A catalog of human genetic variation
Database of Genomic Variants (DGV) http://projects.tcag.ca/variation/	A catalog of human genomic structural variation
dbSNP http://www.ncbi.nlm.nih.gov/SNP/	A database of short genetic variations
Human Gene Mutation Database (HGMD) http://www.hgmd.org/	A database of published gene defects responsible for human inherited diseases
Human Genome Variation Society (HGVS) http://www.hgvs.org/	Provides guidelines for mutation nomenclature and access to different mutation databases
Mutalyzer http://www.lovd.nl/mutalyzer	Tool for sequence variant nomenclature and comparison of reference and submitted amino acid sequences
NHLBI Exome Variant Server http://evs.gs.washington.edu/EVS/	Data release from 6500 human exome, sequenced for NHLBI Exome Sequencing Project (ESP)
Amino acid substitution prediction	
Mutation Taster http://www.mutationtaster.org/	Predicts disease-causing potential of all types of variants in coding sequences
MutPred http://mutpred.mutdb.org/	Classifies amino acid substitutions as disease-causing or neutral
Polymorphism Phenotyping (PolyPhen-2) http://genetics.bwh.harvard.edu/pph	Predicts possible effects of amino acid substitutions on the structure and function of a human protein. Output score ranges from 0 to 1, where 0 is benign and 1 is probably damaging
SIFT (Sorting Intolerant From Tolerant) http://blocks.fhcrc.org/sift/SIFT.html	Scores amino acids in a given protein using sequence homology. The scores range from 1 to 0, where 1 is neutral and 0 is damaging

Table 3.8. Electronic databases, online tools, and bioinformatics software utilized in this study (cont.).

Name and web site	Description
Primer design	
Oligo Calc http://www.basic.northwestern.edu/biotools/OligoCalc.html	Calculates melting temperatures and predicts potential self annealing sites or hairpin structures of primers
Primer3 http://frodo.wi.mit.edu/primer3	Pick primers from a given DNA sequence to amplify specified target region
Tandem Repeats Finder http://tandem.bu.edu/trf/trf.html/	Locates and displays tandem repeats in a given DNA sequence
Analysis of exome sequencing data	
ANNOVAR http://www.openbioinformatics.org/annovar/	Performs functional annotations of genetic variants detected via next-generation sequencing
BamView http://bamview.sourceforge.net/	Software for visualizing sequence read alignments in BAM files
Integrative Genomics Viewer https://www.broadinstitute.org/igv/home	Software for visualizing sequence read alignments and read coverage in BAM files
BEDTools http://code.google.com/p/bedtools/	A package of tools for comparing genomic features such as computing read coverage in next-generation sequencing data
BWA (Burrows-Wheeler Aligner) http://bio-bwa.sourceforge.net/	Software for alignment of sequencing reads to the reference genome
GATK (Genome Analysis Tool Kit) http://www.broadinstitute.org/gatk/	A package of NGS data analysis tools, including coverage analysis, SNP/indel calling and local realignment
Picard http://picard.sourceforge.net/	Software for manipulating SAM (Sequence Alignment/Map) files using Java-based command-line
SAMTools http://samtools.sourforge.net	Software for manipulating alignments in SAM format. It performs sorting, merging, and indexing of alignments, generates per-position information in the pileup format, and performs variant calling.
Others	
Human Splicing Finder 2.4.1 http://www.umd.be/HSF/	Calculates potential splices sites in a given sequence and predicts the effect of a sequence variant on pre-mRNA splicing
Laboratory of Statistical Genetics at Rockefeller University http://linkage.rockefeller.edu/	Supplies list of tools for genetic linkage analysis

4. METHODS

4.1. DNA Extraction from Peripheral Blood Samples

All DNA samples used in the present study are isolated from peripheral blood samples of patients and their unaffected relatives. Two to ten ml of blood samples arrived in sterile vacutainer tubes containing K_2EDTA as anticoagulant. Blood sample was transferred into a sterile falcon tube, and 3 ml of cell lysis buffer was added to every 1 ml of blood. After 15 minutes (min) at 4°C to allow for the lysis of plasma membrane, the sample was centrifuged at 2,600 g for 10 min at 4°C. The supernatant was discarded, and 10 ml of cell lysis buffer was added to the pellet of leukocyte nuclei. After vortexing briefly, the sample was centrifuged again at 4°C for 10 min, and the supernatant was discarded. The pellet containing the nuclei was washed with a few drops of cell lysis buffer and resuspended in nucleus lysis buffer by vortexing. Per ml of initial blood volume, 0.3 ml of nuclei lysis buffer was used. For the digestion of nuclear proteins, 5 µl of Proteinase K (20 mg/ml) and 8 µl of 10% SDS were added per ml of initial blood volume. After incubation at either 37°C overnight or 56°C for 3 hours (hr), 280 µl NH4Ac (9.5 M) per ml of initial blood volume was added, and the tube was shaken by hand to salt out the proteins. The sample was centrifuged at 15,300 g at room temperature for 25 min. The supernatant containing the DNA was transferred to a clean 50 ml falcon tube, and two volumes of ethanol was added to precipitate out the DNA. The DNA precipitate was fished out carefully with a sterile micropipette tip and transferred to a 2 ml microtube. DNA was air-dried and then dissolved in TE buffer (500 µl buffer for DNA isolated from 10 ml blood sample). DNA samples were stored at -20°C.

4.2. Linkage Mapping

Linkage analysis is a statistical method used to identify loci harboring disease genes. Evidence for linkage is expressed as LOD scores (Details are given in Section 1.6). In this study, parametric linkage analysis method is used, because the inheritance model for the diseases in all study families could be assessed with confidence. All of the diseases exhibited Mendelian inheritance in a recessive fashion. To perform parametric linkage

analysis, the pedigree of the family, the genotype data of available family members, and the genetic map of markers used for genotyping are needed.

4.2.1. Genome Scan with Microsatellite Markers

A whole genome scan using microsatellite markers for RTT family was performed at the National Health, Lung and Blood Institute (NHLBI) Mammalian Genotyping Service as a service grant (Contract Number HV48141; Weber and Bronnan, 2001). Genotyping included 405 microsatellite markers. The marker set covered all autosomes and sex chromosomes with an average spacing of 10 cM.

All fifteen participating members of ALS family were genotyped for microsatellite markers on the X-chromosome (UniSTS database, NCBI) by amplifying the marker site via PCR and resolving the alleles on polyacrylamide gels. The same method was used in ALS family and INAD families to test linkage to known disease genes as well as for fine mapping at the candidate loci in RTT family. The list of microsatellite markers utilized is given in Table 4.1.

To test linkage to *PLA2G6* locus, the known gene for INAD, in INAD families 1, 2, 3 and 5 a novel microsatellite marker was searched within *PLA2G6* gene. The DNA sequence of the gene was introduced to Tandem Repeats Finder program, which detected a dinucleotide repeat. The primer pair that was designed to amplify this region had sequences 5’-CCAGCACACACACCCAATAG and 5’-CACCAGAGCCTGTGTGATTG.

Table 4.1. List of microsatellite markers used in this study.

ALS
Chromosome 2 (linkage to *ALS2* locus): D2S2392, D2S2214, D2S1384
Chromosome 15 (linkage to *SPG11* locus): GATA50C03, D15S643
Chromosome X (mapping): DXS6807, DXS9902, DXS556, DXS6810, DXS2505, DXS6785, DXS6799, GATA172D05, DXS6805
INAD
Chromosome 22 (linkage to *PLA2G6* locus): D22S1156, D22S272
RTT
Chromosome 7 (fine mapping): D7S2212, D7S2555
Chromosome 14 (fine mapping): D14S139

4.2.2. Polyacrylamide Gel Electrophoresis and Silver Staining

Microsatellite marker alleles were resolved on denaturing polyacrylamide gels using a Sequi-Gen (Bio-Rad, USA) nucleic acid electrophoresis system assembled with 0.4-mm spacers. Thirty-five ml of denaturing instagel was mixed with 250 µl of 10% APS and 25 µl of TEMED and poured between the glass plates of a 21 x 40 cm gel casting apparatus. The gel was prerun for 15 min to heat it and the buffer to 45-50°C. The samples were mixed with 10 X Sample solution in 2:1 (sample:solution) ratio, denatured at 95°C for four minutes, and loaded into slots of a shark tooth comb. The gel was run at a constant W of 35.

In order to stain the gel after electrophoresis, the glass plates were separated gently by allowing the gel to remain on the surface of one of the plates. A piece of filter paper was placed on the gel to pick it up from the plate, and the gel was let soak in staining solution for 10 min. The gel was subsequently incubated in developing solution until DNA bands became visible (Kavaslar *et al*., 2000). Whenever the bands were faint, the staining procedure was repeated with shorter incubation periods and after extensive washing with dH_2O in-between.

4.2.3. Genome Scan with Single Nucleotide Polymorphism (SNP) Markers

Whole-genome single nucleotide polymorphism (SNP) scans were performed at deCODE (Iceland), Yale Center for Genome Analysis (YCGA; USA), and Macrogen Inc (South Korea). Illumina microarrays were used for all genotyping analyses. The types of whole-genome arrays used are presented in Table 4.2. A SNP genotyping requires at least 1 µg DNA sample. We obtained the raw data and reformatted it for Excel using Progeny for CORD and Illumina Genome Studio v.1.02 Genotyping Module software for all other diseases. The actual pedigrees for study familes are given in Chapter 1.

4.2.3.1. CORD. DNA samples from ten individuals including three patients were genotyped using Illumina Human370-Quad BeadChip at DeCode Institute (Iceland). The raw data obtained were reformatted for Excel using Progeny Software. For the patient and two unaffected individuals who were included later in the study, genome scan was

performed using Illumina Human610-Quad BeadChip in YCGA (USA). Illumina Genome Studio v.1.02 Genotyping Module was used to export the genotypes, which were then transferred to Excel sheets. Markers common to the two chips were selected using Excel and used in linkage analysis.

4.2.3.2. LCA. DNA samples of the affected individuals were genotyped using Illumina Human610-Quad BeadChip, and samples of the parents and the unaffected sibling were genotyped using Illumina HumanOmni1-Quad BeadChip. Both genome scans were performed in YCGA (USA). Markers common to the two chips were selected using Excel.

4.2.3.3. OPA. Genome scan was performed for 13 individuals including all participating patients using Illumina HumanOmni1-Quad BeadChip. DNA samples of the five unaffected sibling were later genotyped using Illumina HumanOmniExpress BeadChip. Both genome scans were performed in YCGA (USA). Markers common to the two chips were selected using Excel.

4.2.3.4. ALS. The genome scan was performed for samples from the five patients plus four unaffected individuals using Illumina HumanOmni1-Quad Beadchip at Macrogen Inc (South Korea).

4.2.3.5. RTT. A DNA sample from an affected child was genotyped in YCGA (USA) using Illumina Human610-Quad BeadChip. Later, a genome scan was performed for all seven participating members of the family, namely, the siblings, their parents and grandmothers using Illumina HumanOmniExpress Beadchip in YCGA (USA). Illumina Genome Studio v.1.02 Genotyping Module was used to prepare input files for PLINK 1.07.

4.2.3.6. INAD. Genome scan was performed for all three siblings and their parents in Family 5 in YCGA (USA) using Illumina Human610-Quad Beadchip. Homozygous regions were detected using ChromoZone and Homozygosity Detector (Illumina Inc, USA) plug-ins on Illumina Genome Studio v.1.02 Genotyping Module, and they were visualized via Illumina Genome Viewer.

Table 4.2. The Illumina whole-genome SNP microarray chips used for genotyping.

Chip Name (BeadChip)	Markers per Sample		SNP Spacing in Kb Mean/Median	Family Genotyped
	SNP	CNV		
370-Quad	~370,000	~70,000	7.7/5.0	CORD
610-Quad	~610,000	~100,000	4.4/2.3	CORD, LCA, INAD
Omni1-Quad	~1 million	~130,000	2.4/1.2	LCA, OPA, ALS
OmniExpress	~700,000	-	4.0/2.1	OPA, RTT

4.2.4. Parametric Linkage Analysis and Homozygosity Mapping

The raw data obtained from genotyping analyses were reformatted for the software package easyLINKAGE v5.08, which was used for parametric LOD score calculations in all families studied. easyLINKAGE package is an open source graphical user interface (GUI) that contains programs to detect genotyping errors (PedCheck), to calculate two-point and multipoint LOD scores, and to construct haplotypes. For multipoint LOD score calculations, GeneHunter v2.1r5 and Allegro v1.2c were used for small pedigrees (bit size <20), and SimWalk2 was used for larger pedigrees, for reasons mentioned in the Introduction chapter. For calculations autosomal recessive inheritance and a disease frequency of 0.0001 (0.00001 for INAD) were assumed. Haplotypes were constructed using GeneHunter or Allegro and visualized via HaploPainter 029.5.

Homozygosity Comparison in Excel (HCiE) is a Microsoft Excel macro that was developed in our laboratory (Cetinkaya, 2010). HCiE with a minor improvement was used for homozygosity mapping in all families. For each chromosome, genotypes were transferred to an Excel sheet and markers were listed according to physical positions on the chromosome. Each genotype (AA, BB and AB) were colored differently. This allows the detection of homozygous regions shared by affected individuals and the inspection of genotype sharing among them. All candidate loci identified via linkage analysis were investigated by HCiE.

Besides HCiE, Homozygosity Detector plug-in on Illumina Genome Studio (Illumina Inc, USA) and PLINK 1.07 were also used for homozygosity mapping in INAD and RTT families, respectively. Homozygosity Detector detects homozygous blocks throughout the

genome above the specified threshold (usually 50 SNPs). Those blocks are visualized via Illumina Genome Viewer. PLINK program was used to detect homozygous stretches >1 Mb in a genotyped individual. The input files were created using PLINK Input Report 2.1.3 plug-in on Illumina GenomeStudio 2011.1 software, the program was executed on Windows/MS-DOS, and the output file was arranged on Excel.

Throughout the study, GRCh37/hg19 map was used, except for the analysis of mitochondrial genome in LCA family. For that purpose, GRCh38/hg38 was used.

4.2.4.1. CORD. Initial multipoint LOD score calculations were performed using program GeneHunter v2.1r5. The data of ten individuals genotyped with 370,000 SNP markers were used assuming a simplified pedigree (Figure 4.1) and full penetrance. All markers were included and used in sets of 100. The genetic map of the markers was provided by Macrogen Inc. When three more participants joined the study, they were genotyped with 610,000 markers. Markers common to the two chips were selected, and multipoint LOD scores were calculated on SimWalk v2.91 selecting markers at 0.07-cM spacing and using them in sets of 100. Those calculations were performed only for the four chromosomes that yielded the highest LOD scores in the initial linkage analysis on Allegro.

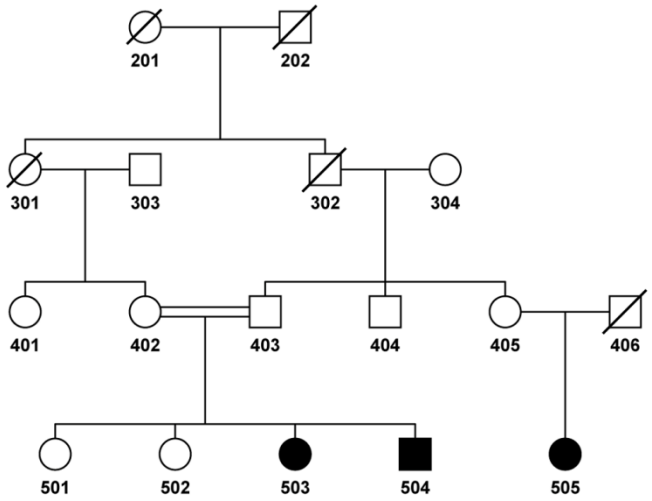

Figure 4.1. The simplified CORD pedigree. It was used for the initial whole-genome multipoint linkage analysis on GeneHunter.

4.2.4.2. LCA. DNA samples of the affected family members were genotyped with 610,000 SNP markers. Later samples from the parents and the healthy sibling were genotyped with 1 million markers. Markers common to the two chips were selected, and multipoint LOD scores were calculated on Allegro assuming full penetrance. Markers were selected at 0.07-cM spacing and used in sets of 30. Later new calculations were performed, selecting markers at 0.01-cM spacing and again using them in sets of 30.

4.2.4.3. OPA. Multipoint LOD score calculations were performed using the genotyping data generated for 1 million SNP markers for eight participants, with program SimWalk and assuming 95% mutation penetrance. Markers were selected at 0.5-cM spacing and used in sets of 100. At 5- to 15-cM long regions around the loci that yielded LOD scores >3, new calculations were performed selecting markers at 0.1-cM spacing. A simplified pedigree was used for the analyses (Figure 4.2).

In the search for a putative second disease locus, a simple pedigree that assumed only individuals 508 and 512 as affected was used (Figure 4.3a). Full penetrance was assumed, and multipoint LOD scores were calculated using Allegro. Markers were selected at 0.07-cM spacing and used in sets of 100. Loci that yielded LOD scores >2.5 were investigated further by HCiE analysis, and new LOD scores were calculated on SimWalk using a larger pedigree (Figure 4.3b).

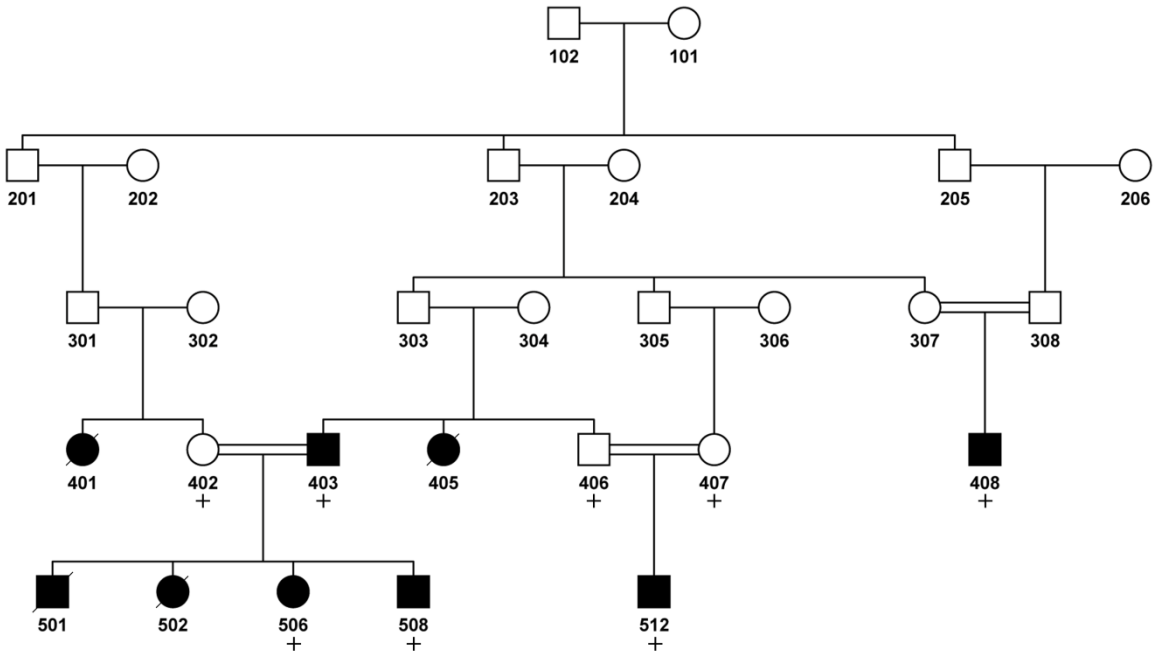

Figure 4.2. Simplified pedigree of OPA family used for linkage analysis. Plus sign denotes genotyped individuals.

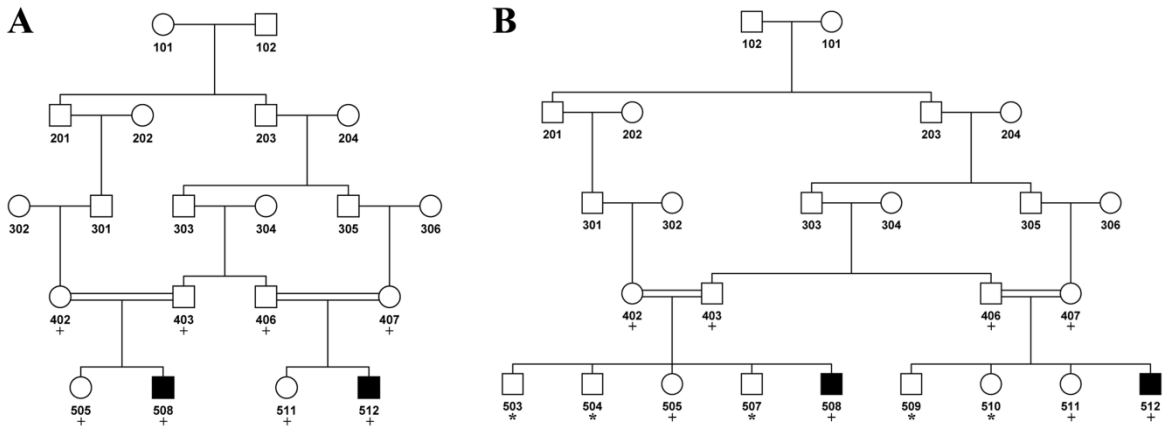

Figure 4.3. Simplified pedigrees for OPA family used in linkage analysis in search for a second disease locus. **a)** Pedigree used by Allegro in whole-genome LOD score calculations, and **b)** The larger pedigree used for detailed calculations via Simwalk. Plus sign denotes individuals genotyped with 1 million SNP markers and asterisk denotes those genotyped with 700,000 markers.

4.2.4.4. ALS. Multipoint LOD score calculations were performed under the assumption of full penetrance. Initial analysis was performed with Allegro v1.2c using sets of 30 markers that were selected at 0.07-cM spacing. Some of the genotyped participants were not

included in this analysis since that would have exceeded the computational limits of Allegro. At loci that yielded LOD scores >3, new LOD score calculations were performed using SimWalk v2.91 including also 0.5-Mb flanking regions. This analysis included all genotyped individuals.

4.2.4.5. RTT. Homozygosity mapping was performed via PLINK to detect homozygous stretches >1 Mb shared by either the three sisters or only the two severely affected sisters. Multipoint LOD score calculations were performed using GeneHunter V.2.1r5 under the assumption of full penetrance. Markers at 0.07-cM spacing were selected and used in sets of 30. Two analyses were performed, one using a pedigree designating two RTT sibs as affected and the other designating all three sibs as affected. After each analysis, new calculations were performed including all markers at loci that yielded LOD scores >1 plus the 0.5-Mb flanking regions.

4.2.4.6. INAD. Multipoint LOD scores were calculated using GeneHunter version 2.1r5 assuming full penetrance. Markers were selected at 0.07-cM spacing and used in sets of 100. Homozygous regions in affected siblings were detected via Homozygosity Detector plug-in and analyzed via HCiE. Another linkage analysis was performed by using all markers at the loci where the affected sibs shared homozygosity.

4.3. Copy Number Variation Analysis

In order to detect copy number variations (CNVs) in genotyped individuals, CNV analysis was performed via cnvPartition v3.1.6 plug-in on Illumina GenomeStudio v.1.02. The generated report files were transferred to Excel for analysis.

SNP markers that were recorded as no call in affected individuals but not in others were considered as indicators of possible genomic deletions. The deletions were investigated for novelty in Database of Genomic Variants (DGV). Whether they contained any genes, any expressed sequences or predicted gene regulatory elements were investigated on UCSC Genome Browser using ENCODE tracks. For validation of the deletions, small regions within the deduced minimal deletion regions were attempted to be amplified by PCR. Additionally, in order to narrow down the maximal deletion region

delineated by the flanking read markers, arbitrarily selected small regions on either side of the minimal deletion sequences were subjected to PCR. Any product obtained was visualized on agarose gels.

Table 4.3. PCR amplifications for deletion analysis in CORD and RTT families.

Primer Name	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temperature (°C)
Deletion at 12q21.33 in CORD Patients				
Validation				
cord_del	F: TTCATAAGGACCGCTTGGTT R: TGGAAATTCACAGCAGCCTA	303	Standard buffer	55.2
Narrowing down the maximal region				
cord_del1	F: TTTCTTCTTTTCCTGGGCACT R: CAGGCTAGAGTGTGCTGGTG	388	Standard buffer	58.5
cord_del2	F: CAGCACCATATCCACACACAC R: CCCCACTGCTATGCTCCTAC	375	Standard buffer	57
cord_del3	F: GAAGGAAACGGAGAACATGC R: CAAAGCCCACATACTCACGA	274	Standard buffer	57
cord_del4	F: GTTTTCATGGTGGCCTGTTT R: TTAAGATGGGATTGTCTAAGCA	198	Standard buffer	54.8
cord_del5	F: TCCTGGAAGCCCTCTTTACA R: GCAGGTGGATCACTTGAGGT	293	Standard buffer + DMSO	53
RTT				
Validation of the deletion at 15q21.1				
RTT_del	F: AATCTGGGAAAAGCAACTGG R: AATCCATGAACAACGCATCTT	350	Standard buffer	65 → 53

4.4. Exome Sequencing Analysis

4.4.1. Exome Enrichment and Sequencing

Exome enrichment and sequencing were performed on a DNA sample of an affected individual in each family. In RTT family, exome sequencing was performed in YCGA (USA). Exome capture was performed using a NimbleGen 2.1M Exome Array V.1.0 (Roche), and sequencing of the exome library was performed on a Genome Analyzer IIx (Illumina) device. In LCA, OPA, ALS and INAD families, exome sequencing was performed at Macrogen Inc (South Korea). Exome capture was performed with TruSeq Exome Capture kit (Illumina), and sequencing was performed on Illumina HiSeq 2000 platforms.

Exome capture protocols for TruSeq Capture kit (Illumina, USA) and NimbleGen 2.1M Exome Array V.1.0 (Roche) are basically the same. Both protocols require 5-10 µg of genomic DNA and begin with random shearing of DNA. Single-strand overhangs are converted to blunt ends to obtain fully double-stranded DNA. Fragments 350-400 bp in size are selected, and adaptors are ligated to the blunt ends. Using primers specific to those adaptors, the fragments are amplified by PCR. Created DNA library is mixed with biotinylated RNA-based capture probes, targeting protein coding sequences and flanking sites. Capture probes are hybridized to the fragments containing target sequences, and streptavidin beads are utilized to pull out the biotinylated capture probes together with the bound fragments. After elusion, the exome enriched library is amplified by PCR.

For sequencing of the enriched library, DNA templates are hybridized onto the flow cell which is a solid support containing forward and reverse primers. By the bridge amplification technique, templates are amplified in clusters, and then they are sequenced via the cyclic reversible termination method. In this method, fluorescently modified nucleotides are incorporated on the primed templates. Each type of nucleotide is labeled with a different fluorescent dye. Due to the modification that blocks the 3' site of the nucleotide, DNA synthesis is terminated after the addition of just one nucleotide. Unincorporated nucleotides are washed off, and imaging is performed. This is followed by the cleavage off of the hindering fluorophore, hence restoring the 3'-OH group necessary for chain elongation. A new cycle begins upon addition of new nucleotides. Overviews of the exome capture and sequencing are shown in Figure 4.4 and in Figure 4.5, respectively.

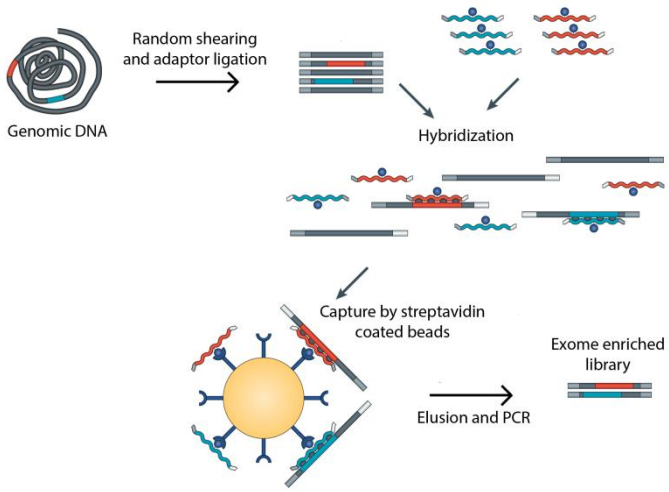

Figure 4.4. Schematic representation of exome capture (Modified from Metzker, 2010).

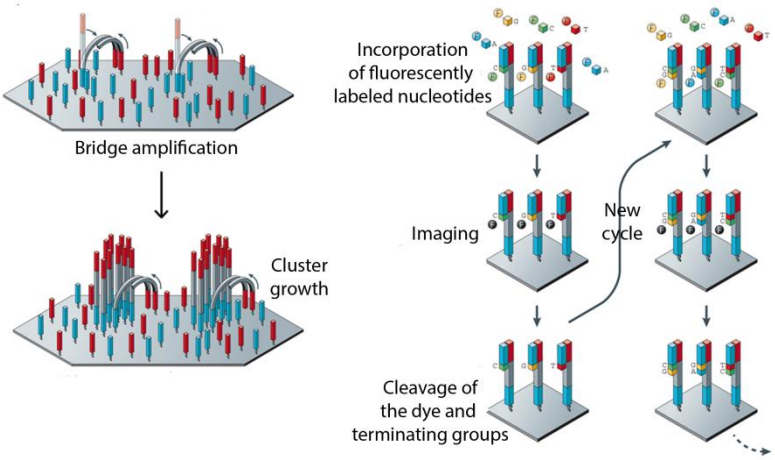

Figure 4.5. Schematic representation of exome sequencing (Modified from Metzker, 2010).

4.4.2. Analysis of Exome Sequencing Results

Bioinformatics analysis of exome sequencing results were performed at Macrogen Inc for the samples analyzed in the company. Bioinformatics analysis of RTT sample was performed in the laboratory of Prof. Murat Günel at Yale University. All analyses were performed using software BWA and SAMTools, and annotated lists of variants were provided to us as Excel files. However, we also performed the bioinformatics analyses ourselves in our laboratory in order to avoid false negative calls of previous analyses and to be able to perform additional analyses such as computation of the coverage of exons or visualizing sequence reads. We requested the raw data for the sequencing results as paired end ".fastq" files and aligned the reads to human genome reference sequence hg19 using BWA software, allowing approximately 2% base-pair mismatch. The reference genome sequence was downloaded from UCSC Genome Bioinformatics Site. Sequence data for individual chromosomes were concatenated to each other and indexed. After alignment, BWA generated a final alignment file in SAM (Sequence Alignment/Map) format. Using SAMTools, the SAM file was converted to BAM (Binary Alignment/Map) format, which is a more compact file format that many software use as input format. The BAM file was sorted and indexed by SAMTools prior to variant calling. Variant calling was also performed by SAMTools. Single nucleotide variations (SNVs) and indels were detected by comparing the sequence reads to the reference sequence. The list of those variants were annotated for novelty after comparing them to databases dbSNP (builds 131 and 132) and 1000 Genomes using ANNOVAR software. After annotation, each variant was described

by its chromosomal location, base change, reading depth, dbSNP ID (if any), region (exonic, intronic, splicing, UTR, non-coding RNA, or intergenic), and for exonic variants the type of the change (synonymous, nonsynonymous, frameshift, stopgain, or stoploss). The list of annotated variants was transferred to Excel for candidate variant search.

For some of the families, alignment and variant calling were also performed using GATK, which uses a different method than BWA. GATK software makes local re-alignments around potential indel sites and evaluates all reads at those regions simultaneously, whereas in BWA, each read is mapped individually to the reference genome. Input files were modified for GATK, and variant calling was performed by Unified Genotyper tool in the GATK package. Since, unlike SAMTools, Unified Genotyper does not call indels, variant calling was repeated also with SAMTools after realignment with GATK. However, the drawback of GATK analysis is that it generates too many variants, and most of them are false positive.

Alignment of the reads to the reference sequence was visualized by BamView software. BEDTools software was used for computing the coverage of each exon in target regions. All bioinformatics analyses were performed on UNIX operating system. The command lines used in the analyses are given in Table 4.4.

Table 4.4. Command lines used in bioinformatics analysis of exome sequencing results and their functions.

Command Line	**Function**
cat chrfile.fa > all_chr.fasta	Concatenates all chromosomes of the reference genome sequence
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 sequence reads to the reference genome (performed for each paired-end read file)
bwa sampe all_chr.fasta file1.sai file2.sai file1.fastq file2fastq > file.sam	Generates file in SAM format using paired-end read files
samtools view -bS file.sam > file.bam	Converts SAM to BAM format
samtools sort file.bam	Sorts the BAM file
samtools index file.sorted.bam	Indexes the sorted BAM file
samtools pileup -vcf all_chr.fasta file.sorted.bam > file.pileup.txt	Calls variants
java -jar ReorderSam.jar I=file.bam O=file_karyo.bam R=ucsc.hg19.fasta	Creates karyotypically ordered BAM file using Picard-Tools software

Table 4.4. Command lines used in bioinformatics analysis of exome sequencing results and their functions (cont.).

Command Line	**Function**
Java –jar GenomeAnalysisTK.jar –T RealignerTargetCreator –R ucsc.hg19.fasta –I file_karyo.sorted.bam –o file_realign.intervals	Determines the intervals to be re-aligned
java -jar -I file_karyo.sorted.bam -R ucsc.hg19.fasta -T IndelRealigner -targetintervals file_realign.intervals -o file_realigned.bam	Makes local re-alignment
java -jar AddOrReplaceReadGroups.jar I=file_realigned.bam O=file_realigned.fixed.bam SORT_ORDER=coordinate RGID=file RGLB=file RGPL=illumine RGPU=file RGSM=file CREATE INDEX=True VALIDATION STRINGENCY=LENIENT	Converts the re-aligned.bam file to an input file compatible to UnifiedGenotyper, a variant calling tool in GATK, using Picard-Tools.
java -jar GenomeAnalysisTK.jar -R ucsc.hg19.fasta -T UnifiedGenotyper -I file_realigned.fixed.bam -o file_GATK.vcf	Calls variants using GATK
convert2annovar.pl file.pileup.txt -outfile file_annovar.pileup.txt	Converts the SAMTools output file to a file compatible to ANNOVAR
convert2annovar.pl file_GATK.vcf -format vcf4 -outfile fileGATK_annovar.pileup.txt	Converts the GATK output file to a file compatible to ANNOVAR
annotate_variation.pl -buildver hg19 file_annovar.pileup.txt humandb	Region based annotation
annotate_variation.pl -filter -dbtype snp131 file_annovar.pileup.txt	Filter based annotation
java -mx152m -jar BamView_v1.1.8.jar -a file.sorted.bam -r all_chr.fasta	Visualizes reads aligned to the reference sequence
bamToBed -i file.bam > file.bed coverageBed -a file.bed -b targetedRegions.bed > file_coverage.txt	Computes the coverage for each exon in the target regions of the exome chip

4.5. Candidate Genes and Mutation Screening

In CORD and ALS families, some of the exons at the candidate loci were analyzed by Sanger sequencing to search for sequence variations. In the remaining families, candidate variants were selected from exome sequencing results by filtering the data in a stepwise manner. First, all novel and rare exonic and splicing variants at the candidate loci were filtered. Then, variants with alternative depth <60% of total depth or found in other in-laboratory exome sequencing samples were filtered out. Among the remaining variants, candidates were selected regarding the function of the genes they reside in and the deduced effect of the variant on the protein. All candidate mutations were validated by Sanger sequencing.

4.5.1. PCR Amplifications

In the search for the mutation underlying CORD, primers were designed to amplify the coding regions of genes *DUSP6, POC1B, GALNT4, ATP2B1, LUM,* and *DCN* and their intronic splice sites so that at least 30 nucleotides flanking the coding sequences could also be read by Sanger sequencing. In ALS family, all coding regions, intronic splice sites, and 5' and 3' untranslated regions (UTRs) of *SIGMAR1* were amplified. Coding regions of known LCA genes, which had not been fully covered by exome sequencing, were amplified for LCA study. For validation of a candidate variant selected from exome sequencing results, primers were designed to amplify a region at least 100-bp long and encompassing the variant.

Primer3 software was used to design primers. Potential self-annealing sites or hairpin structures on the primers were checked by OligoCalc tool. PCR uniqueness was checked via in silico PCR at UCSC database. PCR reactions were carried out using 1X PCR buffer, 400 nM of each primer, 0.2 mM of each dNTP, 20-100 ng of genomic DNA, 0.2 U Taq DNA polymerase, and sufficient water to adjust the final volume to 25 µl. The cycling conditions for amplifications were as follows: An initial denaturation step at 95°C for 3 min, then 35 cycles of denaturation for 30 sec at 95°C, annealing at appropriate temperature for 35 sec (usually a touchdown protocol is used), elongation for 30 sec to 1 min at 72°C, and a final extension step for 6 min at 72°C. The primer sequences and optimum PCR conditions are given in Table 4.5 for candidate genes and in Table 4.6 for candidate variants detected by exome sequencing.

Table 4.5. PCR amplifications for candidate gene screening in CORD and ALS studies.

Exon	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temperature (°C)
CORD				
DUSP6				
1a	F: CGGCTTTACTTCTGTCTCGTC R: CCCGTATTCTCGTTCCAGTC	499	Standard buffer	64 → 53
1b	F: GCCGCAGGAGCTATACGAG R: GCGGAGCAGAGGTATTTTCA	371	Standard buffer	64 → 53
2a	F: CTCCACCATCACCCTGTTTT R: ATTCCTCCAACACGTCCAAG	638	Standard buffer	64 → 53
2b	F: TGTAGGGGGTTCGTAGGATG R: TTAGCAAGCAGCAAGAACGA	649	Standard buffer	64 → 53

Table 4.5. PCR amplifications for candidate gene screening in CORD and ALS studies (cont.).

Exon	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temperature (°C)
CORD				
DUSP6				
3	F: GCTCAACTGCCACACACAAC R: TGCCAAGAGAAACTGCTGAA	488	Standard buffer	64 → 53
POC1B				
1	F: TCTCCTTCCCCATCCTCTC R: GCTACGGACACCTGCCTC	205	Standard buffer, 6% DMSO	64 → 53
2	F: CATTCCCAAGCCGGACTC R: CCGGCCCATGGAGTTTAG	238	Standard buffer	64 → 53
3	F: TACCTGGGCAGCCTTACAAC R: CTTTACCACAAGGGCAGCAT	446	Standard buffer	64 → 53
4	F: CGAAGAGTACAGTACAACATAGCCC R: TTGTTTACTACCGTTCCCGC	499	Standard buffer	64 → 53
5	F: TGGTGTAGAGGTGAGGTGGTG R: CCCGGCCATATTTCAATTTC	285	Standard buffer, 6% DMSO	64 → 53
6	F: TGCGGTGAGCTGACATTG R: AGCCAAACCACACCAACAAG	318	Standard buffer	64 → 53
7	F: GGCACTGGGGTATCTTCAAA R: ACGCCCAGCCACAAATAATA	401	Standard buffer	55
8	F: TGCATAAATTAGGATGAGAAATTG R: TCTTAGCTGAGGTCAAGGATTC	232	Standard buffer, 6% DMSO	64 → 53
9	F: CCCTTCCTACTTTGGGAGACA R: CCTTCAGCAAAAGCCTCAAA	321	Standard buffer	64 → 53
10	F: AGAAAATGCCCTCCAGGTTT R: TTTTATCTCCTCCCCCAACA	266	Dream Taq buffer, Betaine, 1% DMSO	64 → 56
11	F: AAATCTTCAATGTAAGCCTGGG R: GAAATCTCCCCTGCTCTGC	433	Standard buffer	64 → 53
12	F: TTGTCATCCTCACCACCAGA R: TGCCTTTTCTGCCTTTTGTT	324	Dream Taq buffer, Betaine, 6% DMSO	61 → 52
GALNT4				
1a	F: CTGCTGGGAAGTACCTGAGC R: TGTGAATGGTACGGAGCAAA	567	Standard buffer	64 → 53
1b	F: CCCTGCATCGACACATAGAG R: CAGCCATGGTAGGTGATCTG	600	Standard buffer	64 → 53
1c	F: CCATGATTGGTGGGTTTGAC R: GACGAGATCCCTCTACTGCG	547	Standard buffer	64 → 53
1d	F: GAGAGCGGTTGAGATGCAAG R: GAGGCTCTTAAGGCTCTTTGAC	558	Standard buffer	64 → 53
ATP2B1				
1	F: TGGGCCAAAGGTCAAGATAC R: TCATCCCGCCAATCTAAAAC	611	Standard buffer	64 → 53
2	F: TTTTGGTAGGGTCTAGAGTGATGAG R: TGGCCAGCTATTATTATTGTGC	387	Standard buffer	64 → 53
3 & 4	F: CAGGGAACTTGAAAGTGACATAG R: ATGGTGTTTGCTGACAATGG	724	Standard buffer	64 → 53
5	F: TGAAATTTGGTGGTAGCAAGG R: CATACCAGAATCTCTTGAAAGAAGG	297	Standard buffer	64 → 53
6	F: TCTGCATAATTTAGCATGGCTG R: TTGTTGATTTTCCCTGCCAC	218	Dream Taq buffer, Betaine, 1% DMSO	61 → 52
7	F: GCTGAATGCTTATACCTTTACTGG R: CAAAGCAACAATTTCCAGCC	381	Dream Taq buffer, Betaine, 1% DMSO	64 → 56

Table 4.5. PCR amplifications for candidate gene screening in CORD and ALS studies (cont.).

Exon	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temperature (°C)
CORD				
ATP2B1				
8	F: CAATGTCAAGTTATAAGGTGCAAG R: GGGCATCAATGACAACATTTC	526	Standard buffer	64 → 53
9	F: TGAGGTGATTGGTTTTATATGCAG R: CAGTAATGGTGGTCTGAAGACTTA	385	Dream Taq buffer, Betaine, 1% DMSO	61 → 53
10	F: CCCCAGGCTCCTTTCTACC R: GGGGCAAATATCTGGATACC	564	Standard buffer	67 → 56
11	F: TTCTAGGCCTGCCAGTTCAC R: TCTACCCACTTCAAGGCACTG	575	Standard buffer	64 → 53
12	F: GCAGTCGTGTGCAGTGAAAT R: GCTCCATATCAGCCTTTCCA	572	Dream Taq buffer, Betaine, 1% DMSO	62 → 53
13	F: GCCAGATAGCTCATTAGGATTTAAG R: CTGGGAAGCAGAACACATTTC	217	Standard buffer	64 → 53
14	F: CCGACGTCAGGATATTCTTC R: CGAACTAATCTTGAGGGACAGTG	304	Standard buffer	64 → 53
15	F: TGCTAAACCAGAATGTCTTTCC R: CTCAGCAATTCTCAGGAAACC	560	Standard buffer	64 → 53
16	F: TTCCTGAGAATTGCTGAGCC R: TTGGATGATTAGCTTGGAAAAG	340	Standard buffer	64 → 53
17	F: TTTCAGATCAAATAAAAGGTCTAGC R: CAAGTGATTGCTAGAACAAGCTG	396	Dream Taq buffer, Betaine, 1% DMSO	61 → 52
18	F: GGTTGTTCACCTTGTTTTAGAGG R: TCACCACATTAACCATTTAGAATC	323	Dream Taq buffer, Betaine, 1% DMSO	61 → 52
19	F: CGGATACTCTCACCACTTTTAATACC R: AAGGAGCACACTCGAACCTC	363	Standard buffer	64 → 53
20	F: GATGGAATGGTCCTTTCTCTTG R: GAAACGTTAGGTGTGTGAAGTAGG	288	Standard buffer	64 → 53
21	F: TGATTCACTCATACCTAGCTGTTTG R: TGAGGCTCTGAATCTTCTATCC	227	Standard buffer	64 → 53
22	F: TAGCTGTTTGTGGATTCAAACTC R: AGACCCAGTTTCAATTTGTTTC	462	Standard buffer	64 → 53
LUM				
1	F: GGCTGAGCACATTGCACTTA R: AAAAACGCTGTCCCAAACAG	495	Standard buffer	65 → 54
2a	F: GCAAATTGAATGTCTTTTTCTTGA R: CAATCCTTCAAAAGAGCCCA	563	KapaTaq buffer	67 → 56
2b	F: CCACTTCCCAAATCTCTGGA R: TTTAATGGAGCCAGATGCAA	555	Standard buffer	65 → 54
3a	F: TTGTTTTGAGCCAGTGTACTGAA R: AATCCAAGCTTTGAGGAACA	553	Standard buffer	65 → 54
3b	F: GCCTATTTCATCACAAGAACACAC R: AACACCAGGAATGGGAATAATG	701	Standard buffer	65 → 54
DCN				
1	F: CAAGGCTAGAAAGGGTGGAG R: TCACCAAATAATCCCTTTTGC	539	Standard buffer	65 → 54
2	F: TCGTCTTGTCAATAAAGTGATGG R: AGTCCTCACCTGAACCCTGA	350	Standard buffer	65 → 54
3	F: GACACACAGATTAGGCAGTGCT R: AGAGTTGCCATTCCCAAGAA	360	Standard buffer	65 → 54

Table 4.5. PCR amplifications for candidate gene screening in CORD and ALS studies (cont.).

Exon	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temperature (°C)
CORD				
DCN				
4	F: TGATTCTGTAGCCACATGTCC R: TGGCCTTATCTAGGTAGTGTTTTG	248	Standard buffer	65 → 54
5	F: CATGTATGTGGAAGGTTTTGGA R: CCTGCACTTAAAACCCAGTTG	346	Standard buffer	65 → 54
6	F: CTTGCTGAAACAGATACTACAGTG R: CAATTCCAGTTTCCACAAGATTT	280	Standard buffer, CES	62 → 51
7	F: TGTCTTCTTTAAGTTTGTCCACATTC R: GCACTAATATACCTAGCCATTGTTC	227	Standard buffer, CES	62 → 51
8	F: GAAAGGCATCCATGTGTGGT R: GGGGGTCTTGCTTTTTGG	240	Standard buffer, CES	62 → 51
9a	F: ACCTGAAGGGCCTCAACATA R: CTGAAAATGGCAGGCAAAAT	451	Standard buffer	65 → 54
9b	F: TGCATAAAGCCAAATTTCCA R: CAAACCAATGTCTTCTCCCTTC	476	Standard buffer	65 → 54
9c	F: AAGCCTCATTTGAATGTGTGAA R: CATCCCCAGGCTTATTAAATTG	513	Standard buffer	65 → 54
ALS				
SIGMAR1				
1	F: AGGAAATGGTTCAACCGAAG R: CACGATCAGACGAGAGAAGG	500	Dream Taq buffer, Betaine, 2% DMSO	65 → 54
2	F: GGAGCCTAGGGTTCCGAAG R: CCAGCCAAACATCAGAAAGG	361	Standard buffer, BSA (0.2 µg/ml)	62 → 53
3	F: CATGCCCTCCTTTCTGATGT R: CCCAACACACTCCTTTTCCA	267	Standard buffer, BSA (0.2 µg/ml)	62 → 53
4a	F: GCAGAGCTGGCTTTTCACTT R: CGAGCATAGGAGCGAAGAGT	218	Dream Taq buffer, Betaine, 2% DMSO	65 → 54
4b	F: CACATGGATGGTGGAGTACG R: TGCATGGTGTATGTCCCTGT	276	Standard buffer	62 → 53
4c	F: GGCTTGAGCTCACCACCTAC R: TCTCCAGATGGGTGTGAGTG	285	Standard buffer, BSA (0.2 µg/ml)	62 → 53
4d	F: CCCATGGGAACAAATGAGAC R: AGCTCCTCTTCCCCCTCA	294	Standard buffer, 2% DMSO	62 → 53
4e	F: CAGCAATTTGAGGGGATGAG R: GGGCTGTGTGAAAACTGTGA	426	Standard buffer, BSA (0.2 µg/ml)	62 → 53
4f	F: GGGGTATGATGTGGAAGCTG R: GGCTGCTCAGACACACAAAA	254	Dream Taq buffer, Betaine, 2% DMSO	65 → 54
LCA				
GUCY2D				
2a	F: AGGCCGGGGTCTCAGTC R: CTTCTTCGGCGAGCAGC	481	Standard buffer, BSA (0.2 µg/ml)	65 → 54
2b	F: CTTGCGACCCCATCTTCTC R: CATGATCACTGCTGCGGAC	628	Standard buffer, Betaine, 1% DMSO	65 → 54
4	F: GGCTTGACAGGCAGTGAAAG R: GTGGATGGTCCATGGCG	483	Standard buffer	65 → 54
AIPL1				
6	F: GGTCTTGGAGGCTGGTGAG R: TCGAACCAGAAGTGACCAGG	522	Standard buffer	65 → 54

Table 4.5. PCR amplifications for candidate gene screening in CORD and ALS studies (cont.).

Exon	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temperature (°C)
LCA				
CRB1				
3	F: GCTAAATTATGAACACTTTGCTAAAAC R: GGTAAAATAGTTCATGGTCAGGG	319	Standard buffer	65 → 54

Table 4.6. PCR amplifications for sequencing in LCA, OPA, ALS, RTT, and INAD studies.

Variant	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temp (°C)
LCA				
Mitochondrial Variants				
m.2442T>C	F: ACCCCAGAAAACTACGATAGC R: TGCGTAGGGGTTTTAGTTAAATG	167	Standard buffer	63 → 53
m.7968T>C	F: CCATCCCTACGCATCCTTTA R: GGTCGTGTAGCGGTGAAAGT	332	Standard buffer	63 → 53
m.9763T>C	F: TCCAAGCCTACGTTTTCACA R: GACGTGAAGTCCGTGGAAG	679	Standard buffer	63 → 53
m.13651A>G m.13787T>C	F: GCAGCCTAGCATTAGCAGGA R: TTGAGGTCTAGGGCTGTTAGAA	388	Standard buffer	63 → 53
m.14180T>C	F: CTCCACCTCCATCATCACCT R: AGAGGGGTCAGGGTTGATTC	239	Standard buffer	63 → 53
m.15581A>C	F: GCGACCCAGACAATTATACCC R: ATGAGGAGGTCTGCGGCTA	250	Standard buffer	63 → 53
OPA				
ACSS3 **NM_024560**				
Exon 1 c.233G>A	F: CCCCAGGAAGTTGCAAGAG R: CAGGTCCCAGGGAGGTC	448	Standard buffer	63 → 53
PNMT **NM_002686**				
Exon 3 c.586C>T	F: TCCCATAGAGTGGCTGGTTG R: TGCCACTTCAAAGAACAGGG	579	AccuStart II Taq buffer	63 → 53
CEP290 **NM_025114**				
Exon 2 c.62A>C	F: TTGACCAATTAAGCACCTTGG R: CAGAGGTGGAGCACAGTGAA	148	Dream Taq buffer, Betaine, 1% DMSO	61 → 53
ALS				
UNC13B **NM_006377**				
Exon 9 c.790T>C	F: TCTCCTTTCTTTTGTCCTGGTC R: TGTGGCAAGAATGAATCGAG	199	Standard buffer, 5X CES	59.1
GNE **NM_001128227**				
Exon 12 c.2114A>G	F: GGTCTTGGGGTTGTGAACAT R: GTCAACCAAATCCGAAACCA	162	Standard buffer	57.6
EXOSC3 **NM_016042**				
Exon 3 c.556C>G	F: AAGCCAGCAGACATCAGGAC R: TGCTGAGAGGAAATGAACTGG	215	Standard buffer	59.1
ALDH1B1 **NM_000692**				
Exon 2 c.1161delC	F: TGGACAAGGAGCAGTTTGAA R: TGGCAATTCTCATGTCATCC	156	Standard buffer	55.6

Table 4.6. PCR amplifications for sequencing in LCA, OPA, ALS, RTT, and INAD studies.

Variant	Primer Sequence (5' → 3')	Size (bp)	Buffer & Additives	Annealing Temp (°C)
RTT				
***P2RX6* NM_005446**				
Exon 10 c.990dupC	F: CTGACAGTCGTGGGCTGAG R: CCTGGTCAGCAGCACTCAC	163	Standard buffer	52
INAD				
***NALCN* NM_052867**				
Exon 16 c.1924C>T	F: GCAATTAAAGAAGTATTTTGGTACCTG R: ATGCGGACACCAAAGAAAAG	142	Standard buffer, 5X CES	54.1

4.5.2. Analysis of PCR Products

In order to assess the amount of PCR product and whether unspecific products are present, a 5 µl aliquot was mixed with 1 µl of 6X loading dye and loaded on a 2% agarose gel containing 10 mg/ml ethidium bromide. Electrophoresis was performed in 0.5X TBE buffer at 150 volts for 15 minutes. The fragment was visualized under UV light on a Bio-Rad GelDoc Documentation System.

4.5.3. DNA Sequence Analysis

For a sample to be analyzed by Sanger sequencing, a total volume of 50-100 µl PCR product was sent to Macrogen Inc (South Korea), where the sample was purified from primers, nucleotides, salts and polymerase enzyme and subsequently sequenced on an ABI3730xl DNA Sequencer device (Applied Biosystems, USA). Resulting electropherograms were provided to us as ".ab1" files. Chromas Lite 2.01 software was used to visualize the electropherograms.

4.5.4. High Resolution Melting Curve Analysis for Mutation Screening

High resolution melting (HRM) curve analysis was performed to screen family members and the control individuals randomly selected from the population for candidate variants validated by Sanger sequencing. Additionally, all coding regions of *ALDH1B1*, the candidate gene for ALS, were screened for mutations in 90 ALS patients. LightCycler 480 device (Roche) was used for the assays. Real-time PCR reactions were carried out in 96-

well plates. Each reaction mixture contained 5 µl of 2X High Resolution Melting Master Mix, 200 nM of each primer pair, 3 mM of $MgCl_2$, 20 ng of genomic DNA, and sufficient dH_2O to adjust the final volume to 10 µl. Alternatively, 5 µl of 2X AccuMelt HRM SuperMix, 200 nM of each primer pair, 30 ng of genomic DNA, and sufficient dH_2O to adjust the final volume to 10 µl were used for each reaction mixture. In order to discriminate between homozygous wild-type allele and homozygous mutant allele, each sample was spiked by adding some wild-type DNA, except for the control samples from population. Spike DNA constituted 20% of the final DNA amount.

For PCR amplification, a touchdown protocol was used with annealing temperatures ranging from 65°C to 55°C. The cycling conditions were as follows: an initial denaturation step for 10 min at 95°C followed by 40 to 45 cycles of denaturation for 10 sec at 95°C, annealing for 12 to 20 sec at the appropriate temperature and elongation for 10 to 15 sec at 72°C (70°C for AccuMelt HRM SuperMix). After amplification, HRM analysis was performed. Fluorescence signal versus °C graph was constructed using data acquired with 25 measurements per 1°C increase in temperature. LightCycler 480 Gene Scanning software was used for normalization, temperature-shifting and difference plotting of fluorescence data, and to categorize the samples according to melting pattern. Samples with aberrant melting patterns were subjected to Sanger sequencing. Table 4.7 lists the primer sequences used in HRM curve analysis and the assay conditions.

Table 4.7. PCR amplifications and melting temperature ranges for HRM curve analyses.

Region	**Primer Sequence (5' → 3')**	**Size (bp)**	**Annealing Temp (°C)**	**HRM Temp (°C)**
Candidate variant screening				
CORD				
POC1B c.317G>C	F: TCCAGGAGAGGAAAATTCTCAG R: TTGCACTTACTTGGCACAGC	195	65 → 58	70 → 95
OPA				
CEP290 c.62A>C	F: TTGACCAATTAAGCACCTTGG R: CAGAGGTGGAGCACAGTGAA	148	65 → 55	65 → 95
ACSS3 c.233G>A	F: CAGTGGCAGCGAGTACAAGA R: GGCACAGAAGTCACTCACCA	159	65 → 55	65 → 95
ALS				
GNE c.2114A>G	F: GGTCTTGGGGTTGTGAACAT R: GTCAACCAAATCCGAAACCA	162	64 → 57	68 → 91
ALDH1B1 c.1161delC	F: TGGACAAGGAGCAGTTTGAA R: TGGCAATTCTCATGTCATCC	156	64 → 57	68 → 91

Table 4.7. PCR amplifications and melting temperature ranges for HRM curve analyses (cont.).

Region	Primer Sequence (5' → 3')	Size (bp)	Annealing Temp (°C)	HRM Temp (°C)
Candidate variant screening				
RTT				
P2RX6 c.990dupC	F: CTGACAGTCGTGGGCTGAG R: CCTGGTCAGCAGCACTCAC	163	65 → 59	70 → 95
INAD				
NALCN c.1924C>T	F: GCAATTAAAGAAGTATTTTGGTACCTG R: ATGCGGACACCAAAGAAAAG	142	64 → 59	65 → 91
Screening of *ALDH1B1* for mutations				
1a	F: CACCTGTTCACCCTGGTTTC R: TGCATCTTGCCATTCATTGT	189	65 → 58	70 → 95
1b	F: CCAGACATCCCCTACAACCA R: GCTCAGAGGCATCCATCC	208	65 → 58	70 → 95
1c	F: CGACGGTCAACCCTACCA R: TCCAAGTCCAAGGCGTAAGA	256	65 → 58	70 → 95
1d	F: ACCTTGGACAATGGGAAGC R: GATGATCTGGCCACAGACAC	180	64 → 57	68 → 91
1e	F: ATGGATGGCCAGCATTTCT R: GCCAAATACAGGGCAGAGAG	179	65 → 58	70 → 95
1f	F: GTCATGCAGGGTTGGAAACT R: GGCAACTTTGTCAACATCCA	210	65 → 58	70 → 95
1g	F: AACATCATCACGGGGTATGG R: ACTGGCCCATGTTGAAGAAC	247	64 → 56	70 → 95
1h	F: GTAAGAGCCCCAGCATCGT R: GGGTTCCCCACTTTCCTC	190	65 → 58	70 → 95
1i	F: ACCGTGGAGAAAGCAAAGC R: CAGTAGGCTTGATGAAGAAACCA	190	65 → 58	70 → 95
1j	F: TGGACAAGGAGCAGTTTGAA R: AGGCCATACCTGGTGTTGTT	247	65 → 58	70 → 95
1k	F: AGCCCCTGTTCAAGTTCAAG R: GCAGGTGACGATGTTGTAGG	170	65 → 58	70 → 95
1l	F: GTACTTCACCCAGGCACTCC R: GTGGAATTGCTGGACTGTGA	227	65 → 58	70 → 95

4.5. Assessment of Relative Transcript Levels by Quantitative PCR (QPCR)

Abundance of *ALDH1B1* and *NALCN* transcripts and of two transcript isoforms of *P2RX6* were investigated in total RNA samples from 17 different adult human tissues. Total RNA samples from subtantia nigra, pons, putamen, spinal cord, liver, blood, skeletal muscle, bone marrow, adipose tissue, and testis tissues were purchased from Clontech Inc (USA). cDNA synthesis was performed with the RevertAid First Strand cDNA Synthesis kit using oligo(dT) primers (Thermo Scientific, USA). cDNA samples from frontal cortex, parietal cortex, occipital cortex, corpus callosum, cerebellum, and brain stem tissues (Clontech Inc, USA) which were prepared using the same method were kindly provided by Prof. Tayfun Özçelik, Bilkent University, Ankara. A real-time polymerase chain reaction

(PCR) was performed with an intron-spanning primer pair that was specific to either the investigated transcript or housekeeping gene *TUBB* transcript. Primers specific to the transcripts are given in Table 4.8. Real-time PCR reactions were carried out in 96-well plates in a total volume of 20 µl consisting of 10 µl of 2X Light Cycler 480 SYBR Green I Master, 200 nM of each primer pair, 4 ng of cDNA, and sufficient amount of dH_2O to adjust the volume, using a SYBR Green I Master Kit on a LightCycler 480 (Roche). All reactions were performed in triplicates using a touchdown protocol with annealing temperatures ranging from 65°C to 57°C. Reaction conditions are as follows: initial denaturation for 10 minutes at 95°C and 40 to 45 cycles of denaturation for 10 seconds at 95°C, annealing for 12 to 13 seconds at the appropriate temperature, and elongation for 9 to 11 seconds at 72°C. Transcript levels were normalized to *TUBB* via advanced relative quantification analysis on the LightCycler 480 Relative Quantification Software (Roche), using relative standard curves of serial dilutions of cDNAs to calculate PCR efficiencies. PCR products of investigated transcripts were subjected to Sanger sequencing for validation.

Table 4.8. Primers used in qPCR assays.

Transcript	**Primer site**	**Primer Sequence (5'→ 3')**	**cDNA product size (bp)**
ALDH1B1 NM_000692.4	F: Exon 1 R: Exon 2	AGAACCCAAGCGTGATCCT CTGCTGCCGAGGAGTAGC	123
NALCN NM_052867.2	F: Exon 42 R: Exon 43	ACACGAATGCCAACAGTCAG CCAAATTTCCTCTGGGGTTT	139
P2RX6 NM_005446.3 & NM_001159554.1	F: Exons 2 & 3 R: Exon 4	CACCTCAGGGAGAGAACGTG CTCGTCGACCCAGCAGTTAG	119
TUBB NM_178014.2	F: Exon 1 R: Exon 3	ACATCCAGGCTGGTCAGTGT AAAGGACCTGAGCGAACAGA	226

5. RESULTS

5.1. Cone-Rod Dystrophy (CORD)

5.1.1. Multipoint Linkage Analysis and Haplotype Inspection

In order to identify candidate disease loci, whole-genome multipoint linkage analysis was performed using the genome scan data for ten family members. The analysis yielded the same maximal multipoint LOD score of 2.65 at loci 12q21.33, 15q12, 17p12-11.1, and 18p11.23-p11.22 (Figure 5.1). As the original pedigree exceeded the computational limits of GeneHunter, a simplified pedigree was used (Figure 4.1). A second, finer linkage analysis was performed on SimWalk for chromosomes 12, 15, 17 and 18, which included also the genotypes of the three individuals who later joined the study. In this analysis, the original pedigree was used except that individuals 401, 404 and 506 were not included. Locus 12q21.33 yielded a maximal LOD score of 3.92 whereas 15q12 and 18p11.23-p11.22 yielded LOD scores <3 (Figure 5.2) and 17p12-11.1 yielded negative scores.

Haplotype analysis was performed for all four loci that yielded maximal LOD scores of 2.65 in the initial linkage analysis in order to investigate identity by descent among affected individuals. At the 10-Mb locus 17p12-11.1, affected individuals 305 and 505 carried heterozygous SNP genotypes. At 18p11.23-p11.22, all patients were homozygous; however, homozygosities were not IBD according to haplotype investigation. For locus 15q12 (0.7 Mb), CNV genotypes were exported from Software Progeny 7.0, and haplotypes were constructed manually; affected cousin 505 carried a heterozygous CNV genotype. Those three loci were excluded. At the remaining locus 12q21.33, flanked by markers rs12311684 and rs934891 (89,435,177 bp and 90,571,042 bp), all affected children shared homozygosity which was possibly IBD. None of the other family members was homozygous for the haplotype. This locus was 1.14 Mb. Haplotype analysis around it revealed a second locus with shared homozygosity in the patients, which was 0.8 Mb in length and delineated by rs2408366 (91,192,135 bp) and rs2130402 (91,976,653 bp). This region was 0.6 Mb telomeric to the previous locus, and yielded a LOD score of 3.40 in

SimWalk analysis. The haplotypes are presented in Table 5.1. We concluded that the disease gene was likely in one of these two IBD homozygous regions.

Table 5.1. Haplotypes for selected SNP markers at 12q21.33 for CORD family. The genotypes of individual 406 were deduced and are shown in italics. Disease haplotype and ID numbers of affected individuals are highlighted grey (Durlu *et al*., 2014).

Marker ID	*Position (Mb)*	***305***	*401*	*404*	*402*	*403*	*501*	*502*	***503***	***504***	*405*	*406*	***505***	*506*	*507*
rs6538116	88.23	**1 1**	2 2	**1** 1	2 **1**	**1** 1	1 **1**	1 2	**1 1**	**1 1**	**1** 2	*2 2*	2 **1**	2 2	2 2
rs4146751	88.33	**1 1**	1 2	**1** 1	2 **1**	**1** 1	1 **1**	1 2	**1 1**	**1 1**	**1** 1	*1 1*	1 **1**	1 1	1 1
rs12831250	88.57	**2 2**	2 2	**2** 2	2 **2**	**2** 2	2 **2**	2 2	**2 2**	**2 2**	**2** 2	*1 2*	1 **2**	2 2	1 2
rs7976732	88.69	**2 2**	2 2	**2** 2	2 **2**	**2** 2	2 **2**	2 2	**2 2**	**2 2**	**2** 2	*1 2*	1 **2**	2 2	1 2
rs454833	89.20	**1 1**	1 1	**1** 2	1 **1**	**1** 2	2 **1**	2 1	**1 1**	**1 1**	**1** 2	*1 2*	1 **1**	2 2	1 2
rs12311684	89.43	**2 2**	1 1	**2** 2	1 **2**	**2** 2	2 **2**	2 1	**2 2**	**2 2**	**2** 2	*1 2*	1 **2**	2 2	1 2
rs2407806	89.51	**2 2**	1 1	**2** 2	1 **2**	**2** 2	2 **2**	2 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs995727	89.59	**2 2**	1 1	**2** 2	1 **2**	**2** 2	2 **2**	2 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs922350	89.67	**1 1**	2 2	**1** 2	2 **1**	**1** 2	2 **1**	2 2	**1 1**	**1 1**	**1** 1	***1*** *1*	**1 1**	1 1	**1** 1
rs704061	89.77	**2 2**	1 1	**2** 1	1 **2**	**2** 1	1 **2**	1 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs7134172	89.84	**2 2**	1 2	**2** 1	1 **2**	**2** 1	1 **2**	1 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs12426089	89.96	**1 1**	1 1	**1** 2	1 **1**	**1** 2	2 **1**	2 1	**1 1**	**1 1**	**1** 2	***1*** *2*	**1 1**	2 2	**1** 2
rs10858918	90.10	**1 1**	1 1	**1** 2	1 **1**	**1** 2	2 **1**	2 1	**1 1**	**1 1**	**1** 2	***1*** *2*	**1 1**	2 2	**1** 2
rs11105458	90.29	**2 2**	1 2	**2** 2	1 **2**	**2** 2	2 **2**	2 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs10858942	90.41	**2 2**	2 1	**2** 1	2 **2**	**2** 1	1 **2**	1 2	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs11105501	90.48	**2 2**	1 1	**2** 1	1 **2**	**2** 1	1 **2**	1 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs11105520	90.55	**2 2**	2 2	**2** 2	2 **2**	**2** 2	2 **2**	2 2	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs2579103	90.63	**2 2**	1 1	**2** 2	1 **2**	1 2	2 **2**	2 1	1 **2**	1 **2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs2056700	90.70	**2 2**	2 1	**2** 2	1 **2**	2 2	2 **2**	2 1	2 **2**	2 **2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs1222512	90.77	**2 2**	1 1	**2** 2	1 **2**	1 2	2 **2**	2 1	1 **2**	1 **2**	**2** 2	***2*** *2*	**2 2**	2 2	**2** 2
rs1148678	90.85	**1 1**	2 2	**1** 2	2 **1**	2 2	2 **1**	2 2	2 **1**	2 **1**	**1** 2	***1*** *2*	**1 1**	2 2	**1** 2
rs12369618	90.92	**1 1**	2 2	**1** 1	2 **1**	1 2	1 **1**	1 2	2 **1**	2 **1**	**1** 1	***1*** *1*	**1 1**	1 1	**1** 1
rs2520528	90.98	**2 2**	1 1	**2** 1	1 **2**	1 1	1 **2**	1 1	1 **2**	1 **2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs1847443	91.11	**1 1**	1 1	**1** 2	1 **1**	1 2	2 **1**	2 1	1 **1**	1 **1**	**1** 2	***1*** *2*	**1 1**	2 2	**1** 2
rs1354421	91.18	**2 2**	1 2	**2** 1	2 **2**	1 1	1 **2**	1 2	1 **2**	1 **2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs11105838	91.24	**2 2**	2 1	**2** 1	1 **2**	**2** 1	1 **2**	1 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs4523751	91.32	**1 1**	1 1	**1** 2	1 **1**	**1** 2	2 **1**	2 1	**1 1**	**1 1**	**1** 1	***1*** *1*	**1 1**	1 1	**1** 1
rs17018653	91.46	**2 2**	2 2	**2** 1	2 **2**	**2** 1	1 **2**	1 2	**2 2**	**2 2**	**2** 2	***2*** *2*	**2 2**	2 2	**2** 2
rs566806	91.56	**2 2**	1 2	**2** 1	1 **2**	**2** 1	1 **2**	1 1	**2 2**	**2 2**	**2** 2	***2*** *2*	**2 2**	2 2	**2** 2
rs861158	91.65	**2 2**	2 2	**2** 1	2 **2**	**2** 1	1 **2**	1 2	**2 2**	**2 2**	**2** 2	***2*** *2*	**2 2**	2 2	**2** 2
rs1492917	91.73	**2 2**	2 2	**2** 1	2 **2**	**2** 1	1 **2**	1 2	**2 2**	**2 2**	**2** 2	***2*** *2*	**2 2**	2 2	**2** 2
rs10497386	91.82	**2 2**	1 2	**2** 1	1 **2**	**2** 1	1 **2**	1 1	**2 2**	**2 2**	**2** 2	***2*** *2*	**2 2**	2 2	**2** 2
rs2130398	91.89	**2 2**	1 2	**2** 1	1 **2**	**2** 1	1 **2**	1 1	**2 2**	**2 2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs2130402	91.98	**1 1**	2 2	**1** 1	2 **1**	2 1	1 **1**	1 2	2 **1**	2 **1**	**1** 1	***1*** *1*	**1 1**	1 1	**1** 1
rs10506999	92.06	**2 2**	1 1	**2** 1	1 **2**	1 1	1 **2**	1 1	1 **2**	1 **2**	**2** 1	***2*** *1*	**2 2**	1 1	**2** 1
rs7976371	92.14	**2 2**	1 1	**2** 2	1 **2**	1 2	2 **2**	2 1	1 **2**	1 **2**	**2** 2	***2*** *2*	**2 2**	2 2	**2** 2
rs11106293	92.21	**1 1**	1 2	**1** 1	1 **1**	2 1	1 **1**	1 1	2 **1**	2 **1**	**1** 1	***1*** *1*	**1 1**	1 1	**1** 1

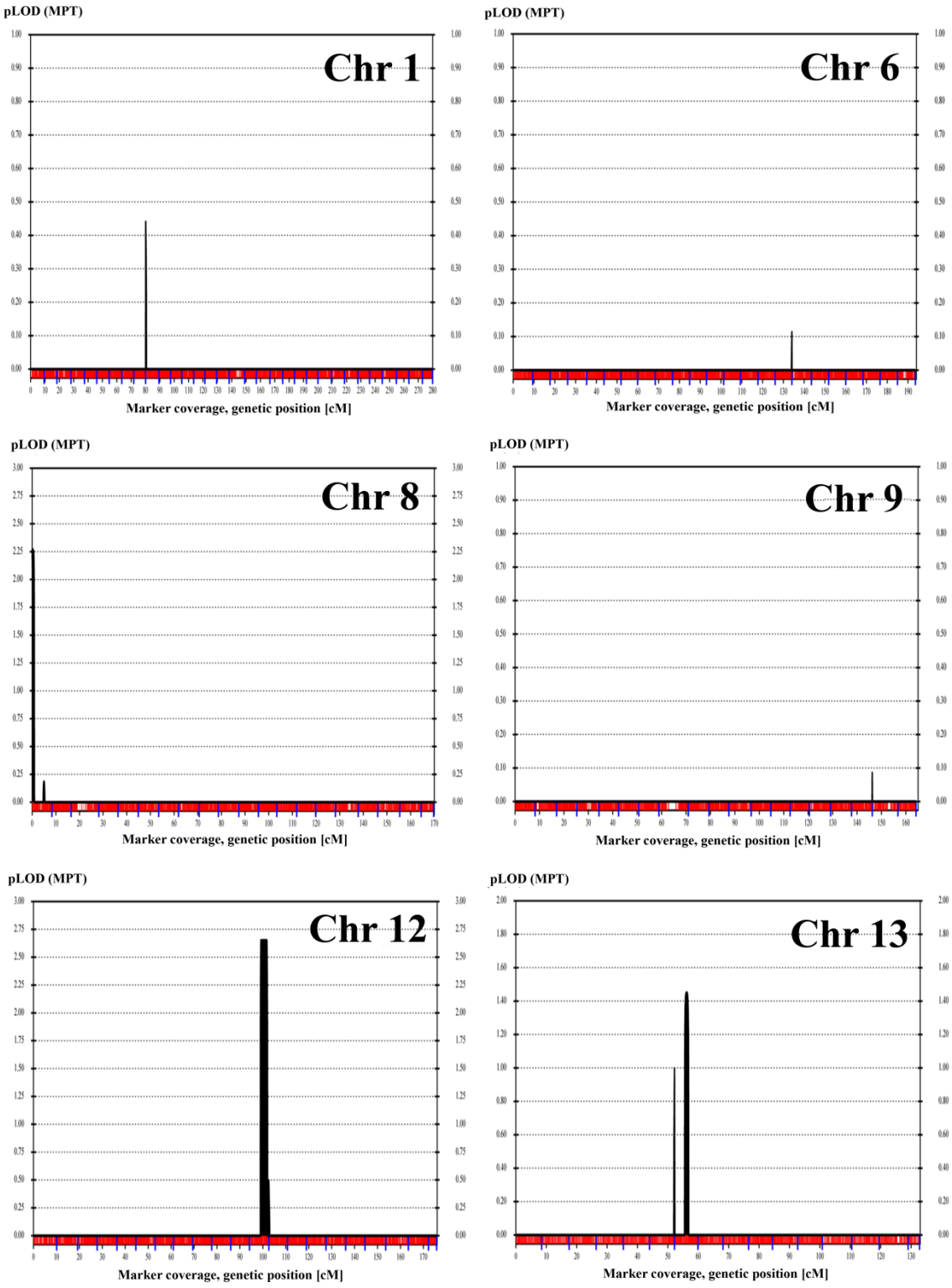

Figure 5.1. Initial multipoint LOD scores for CORD family. The scores were calculated for all autosomes and pseudoautosomal regions (PARs) by GeneHunter using the genome scan data for the initial ten participants. Only chromosomes with scores >0 are presented.

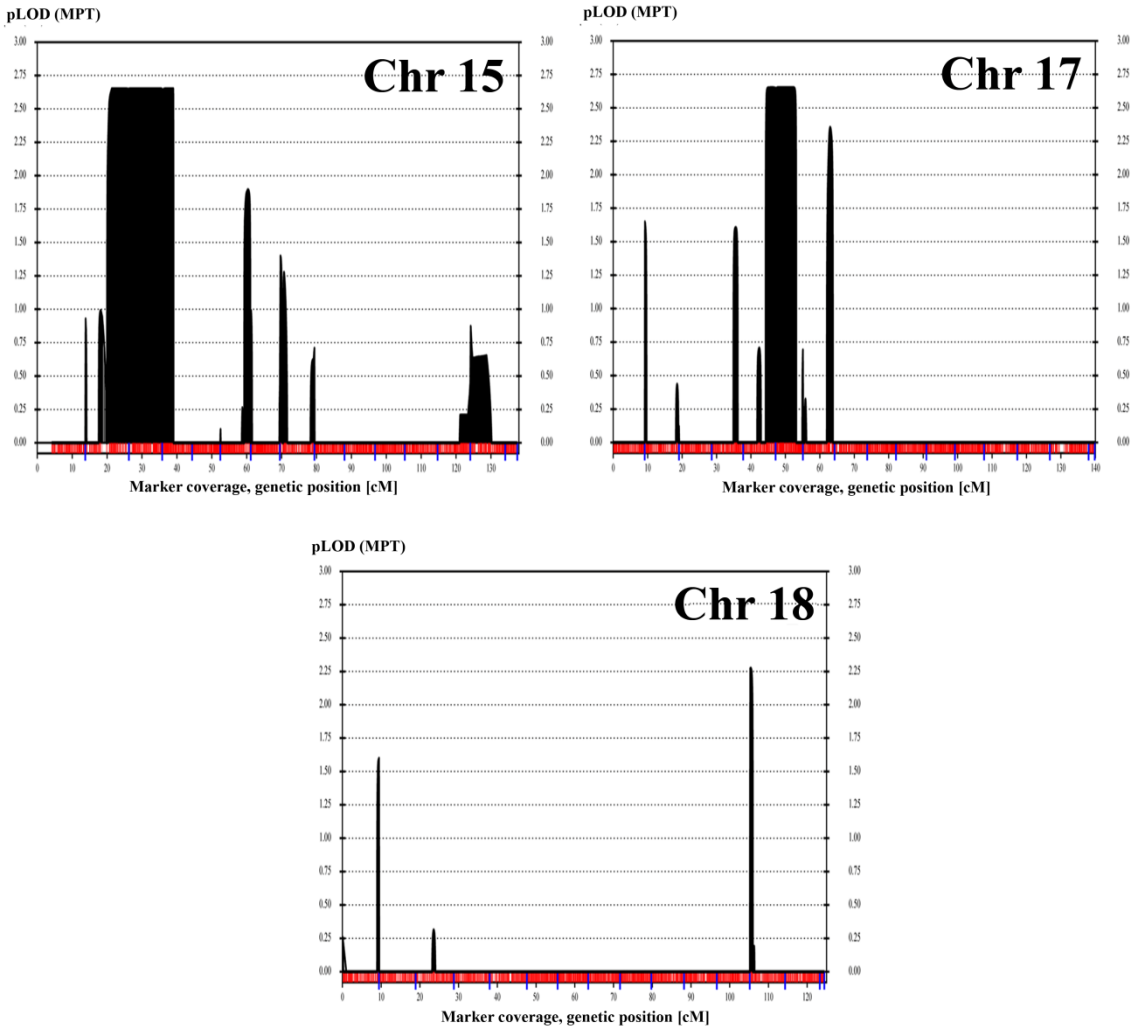

Figure 5.1. Initial multipoint LOD scores for CORD family. The scores were calculated for all autosomes and PARs by GeneHunter using the genome scan of the initial ten participants. Only chromosomes with scores >0 are presented (cont.).

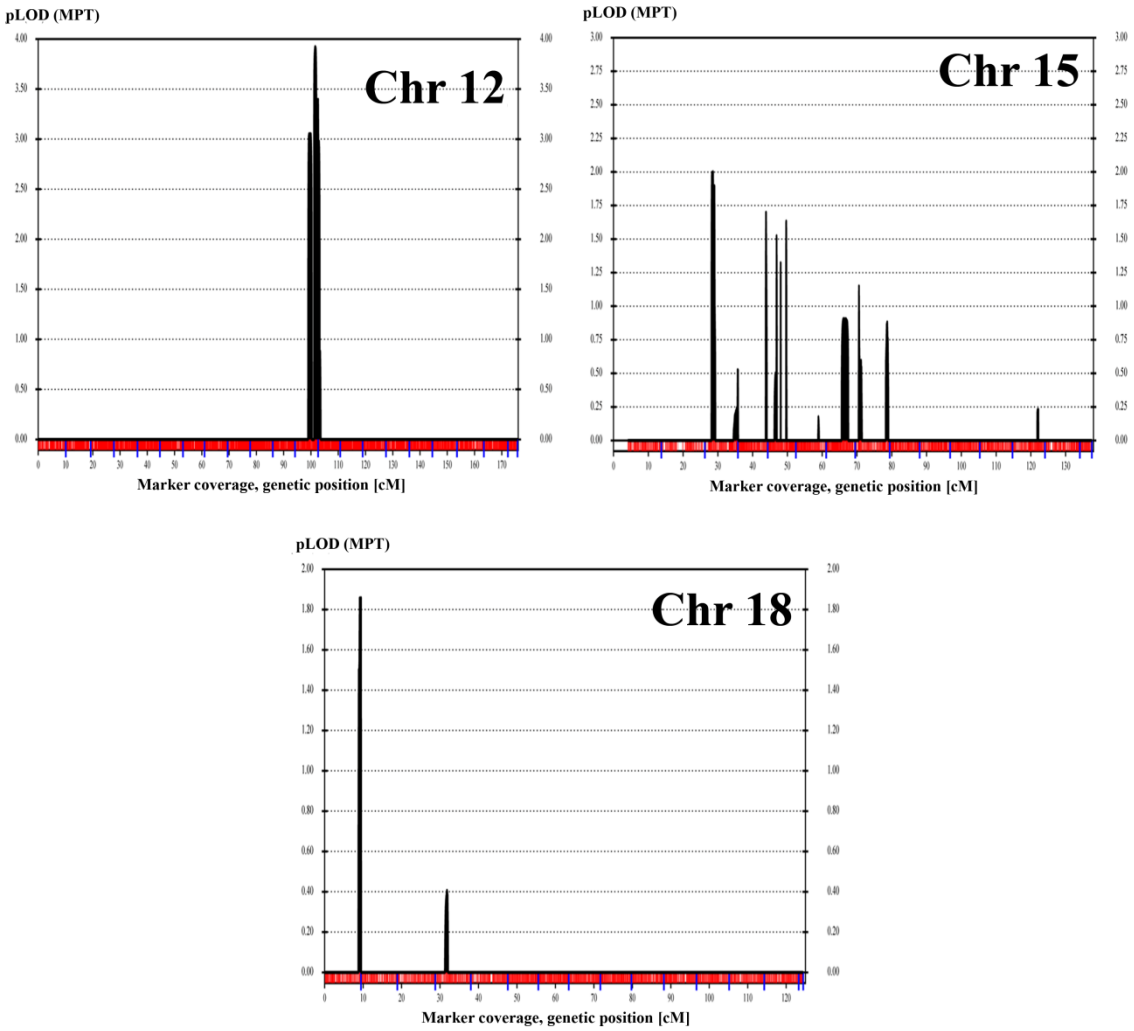

Figure 5.2. Finer multipoint LOD scores for chromosomes 12, 15, and 18 for CORD family. Scores were calculated by SimWalk using SNP genotypes of 13 family members.

5.1.2. Candidate Gene Screening

The two homozygous regions at 12q21.33 contained 11 genes in total. Among them, four were pseudogenes (*MRPS6P4*, *CENPC1P1*, *MRPL2P1*, and *LOC100287505*) and one was a non-coding RNA gene (*LOC338758*). The remaining six genes were protein coding. The first, larger region harbored *DUSP6*, *POC1B*, *GALNT4* and *ATP2B1*, and the smaller region harbored *LUM* and *DCN*. The coding regions of these six genes were analyzed by Sanger sequencing in patient 503. Four sequence variants in *POC1B* and *GALNT4* were detected (Table 5.2). Three of them were reported variants with MAF >0.1. The remaining *POC1B* c.317G>C (p.R106P) (Figure 5.3) was not reported in dbSNP or the NHLBI Exome Sequencing Project (EVS) databases. All 13 participants were analyzed for this

novel variant by Sanger sequencing, and the variant was found to segregate with the disease in the family. A control panel of 113 unrelated control subjects from the population was screened using HRM curve analysis, and none of the individuals was found to carry the variant. This substitution of arginine by proline at codon 106 was predicted to be deleterious by online tool PolyPhen-2 (score: 1) and disease causing by MutationTaster. According to these results, *POC1B* c.317G>C stood out as the best candidate for the mutation causing CORD in the family.

Table 5.2. Variants detected in the six genes analyzed in CORD family. Designations are according to the longest transcript isoforms.

Variant	SNP ID	Frequency
POC1B c.273-31G>T	rs6538189	0.17
POC1B p.R106P	(Novel)	--
POC1B c.1113+7C>T	rs4842658	0.25
GALNT4 p.I270T	rs2230281	0.27

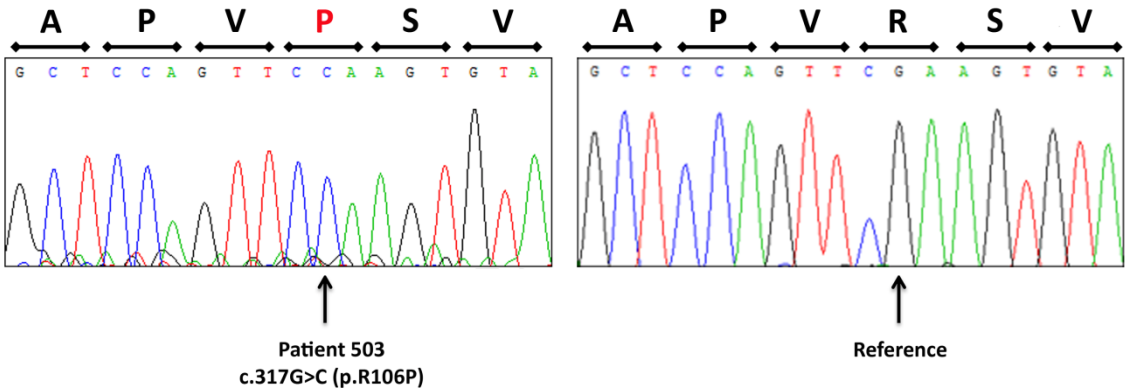

Figure 5.3. Chromatograms showing *POC1B* mutation.

5.1.3. Deletion Analysis

Two adjacent SNP markers rs12817868 (90,488,249 bp) and rs10858953 (90,490,259 bp) at the first candidate locus 12q21.33 were recorded as no call in the patients but not any other family member, suggesting a possible homozygous deletion. Nearest read markers were rs11105501 (90,482,636 bp) and rs10777224 (90,508,309 bp), suggesting that the deletion was maximally 26 Kb. The deletion was validated by a PCR

assay; 303-bp region within those two unread markers could be amplified in a control individual and in the heterozygous father, but patient samples did not yield any products. In order to narrow down the maximal deletion region, arbitrarily selected sequences within that region but outside the minimal deletion region were attempted to be amplified in the patients and in an unrelated control sample. The maximal deletion region was delineated as 90,486,683 - 90,492,571 bp. Thus, the deletion was maximally 5,888 bp and minimally 3,280 bp. Deletion breakpoints could not be identified, since this region was highly AT-rich, rendering further amplification impossible.

The maximal deletion region was investigated in DGV for novelty, and no reported deletion was found. The deletion was hypothesized to possibly harbor regulatory regions necessary for gene expression, since it did not contain any part of a gene. The maximal deletion region was investigated via the UCSC Genome Browser tracks. The only sequences found were two predicted transcription factor binding sites. The prediction was based on computational analysis. There were no genes, expressed sequence tags, H3K27Ac marks, DNaseI hypersensitivity clusters, CpG islands, DNA and CpG methylation sites, or distant-acting enhancer elements. The deletion is unlikely to be the cause of the disease, since there are no gene regulatory sequences in the region (Figure 5.4).

Figure 5.4. UCSC Genome Browser prediction for possible regulatory roles of the sequences in the maximal deletion region at 12q21.33.

5.2. Leber Congenital Amaurosis (LCA)

5.2.1. Multipoint Linkage Analysis and Homozygosity Inspection

Genotypes of the affected sibs generated by genome scan with 610,000 SNP markers were inspected on Excel to investigate whether the sibs shared homozygosity, possibly IBD, at the loci of known LCA genes. There was no homozygosity sharing by the patients at any of the known loci. After merging genome scan data of all genotyped family members, whole-genome multipoint linkage analysis was performed on Allegro using markers at 0.07-cM spacing. Highest LOD scores were obtained at 1q23.2 (3.05) and nearby locus 1q23.3-24.1 (3.06). At those loci, affected individuals carry identical genotypes. However, maximal homozygosity regions they share are very small, 0.16 Mb and 0.29 Mb, respectively. Not finding any exonic variants in those loci prompted us to perform another linkage analysis using markers at 0.01-cM. A total of 19 loci yielded LOD scores >3 (Table 5.3), and they were investigated by HCiE. This analysis showed that the affected sibs shared homozygosity >200 Kb which are possibly IBD at nine loci (Table 5.3). One of those homozygous loci was excluded due to two heterozygous variants within the locus in exome sequencing results. The remaining eight loci were assessed as candidates for disease gene locus.

Table 5.3. Regions with homozygous genotypes assessed as IBD in the LCA sibs. No candidate variant was found in any of them.

Locus	**LOD Score**	**Maximum homozygosity region (location in bp)**	**Size (Kb)**	**Candidate Variants**
1p22.2	3.20	rs12141090 - rs306321 (88,420,455 - 89,005,795)	585	-
1p12-p11.2	3.64	rs7513294 - rs12125273 (115,898,335 - 116,776,537)	878	-
1p11.1	3.59	rs4659128 - rs10802085 (119,464,710 - 119,799,665)	335	-

Table 5.3. Regions with homozygous genotypes assessed as IBD in the LCA sibs. No candidate variant was found in any of them (cont.).

Locus	**LOD Score**	**Maximum homozygosity region (location in bp)**	**Size (Kb)**	**Candidate Variants**
1q21.3	3.55	No homozygosity*	-	-
1q22	3.63	rs4845552 - rs11579790 (153,479,998 - 154,279,183)	799	-
1q23.3	3.62	rs10918338 - rs4626860 (166,047,796 - 166,362,651)	315	-
1q31.3	3.64	rs6427830 - rs4915463 (200,464,803 - 200,879,302)	414	-
1q32.1	3.52	No homozygosity	-	-
4q24	3.64	No homozygosity	-	-
8q12.3	3.42	No homozygosity	-	-
9q21.32	3.63	rs6560613 - rs3824548 (80,340,387 - 80,911,421)	571	-
9q22.1	3.63	No homozygosity	-	-
9q22.31	3.21	No homozygosity	-	-
10p14	3.34	No homozygosity	-	-
13q33.3	3.64	No homozygosity	-	-
15q11.2	3.33	No homozygosity	-	-
17p11.2	3.63	No homozygosity	-	-
22q11.22	3.54	rs2142843 - rs10483117 (25,368,543 - 25,632,266)	264	-
22q11.23	3.44	No homozygosity	-	-

*According to SNP genotyping results, the sibs share homozygosity for a 369-Kb region. However, the region is excluded due to two heterozygous variants within it detected by exome sequencing.

5.2.2. Evaluation of Exome Sequencing Results

DNA sample of a patient was subjected to exome sequencing at Macrogen Inc (South Korea) using Illumina TrueSeq Exome Capture Kit and Illumina Hiseq2000 platform.

Alignment of the sequencing reads to the reference genome, variant calling, and annotation of detected variants were performed at Macrogen Inc and repeated in our laboratory using BWA, SAMTools and ANNOVAR software. Only 14 exonic and splicing variants were detected at the candidate loci. All variants except *VANGL1* c.G641A were reported in dbSNP132 database with frequencies >0.01. Since the variant was predicted to be benign by MutationTaster, and the gene is associated with neural tube defects (MIM#610132), *VANGL1* was not considered as a candidate gene for LCA.

Alignment and variant calling were repeated using GATK software. However, no additional novel or rare exonic variant could be detected at the candidate loci. Coverages of exon reads were computed by BEDTools. There was no unread exon within the candidate loci. We also analyzed the total exome sequence results for novel, homozygous and possibly damaging mutations. Among the 72 such variants, 59 were detected in some of the other 25 in-house exome sequencing samples. The regions of the remaining 13 variants were investigated in SNP genotypes on Excel regardless of their LOD scores to see whether there is genotype sharing among the affected sibs. Only variant *TBC1D2* c.A2636G (p.E879G) on chromosome 9 at 100,961,781 bp, where we obtained a maximal LOD score of 0.69, was located in a very small region of shared homozygosity (136 Kb). The affected sibs but not the unaffected sib shared genotypes. However, TBC1D2 protein is reported as a tumor antigen (MIM#609871), and MutationTaster predicted the variant as benign.

Not finding any candidate variants, we considered the possibility that compound heterozygous mutations in a gene could cause the disease. We first investigated the exome sequencing results to find any compound heterozygous mutations in the 17 known LCA genes (Table 1.3). Patients but not the unaffected sibling shared heterozygous genotypes at three of them; nonetheless, none of the genes at those loci harbored two novel or rare (MAF <0.01) exonic variants. Subsequently, we analyzed the total exome sequence results for all genes in search of at least two rare, heterozygous, and possibly damaging mutations in the same gene. The loci of such genes were then investigated via HCiE to see whether the affected sibs but not the unaffected sib shared genotypes. The only gene that harbored two rare, nonsynonymous, heterozygous mutations and that resides in a region where the affected sibs shared genotypes was *F5*. It encodes coagulation factor V and is associated with the diseases factor V deficiency and thrombophilia (MIM#612309). Finally, the exon

coverages of the known LCA genes were evaluated on the BEDTools output file. Two exons of *GUCY2D* and one exon each of *AIPL1* and *CRB1* were only partially read. Those exons were analyzed by Sanger sequencing,and a heterozygous stopgain variant in the last exon of *AIPL1* was detected (Figure 5.5). At *AIPL1* locus, the three affected sibs but not their unaffected sister shared genotypes. The variant *AIPL1* c.834G>A (p.W278X) is reported in four families afflicted with LCA. In one of those families, the patients were compound heterozygous for p.W278X and another mutation in *AIPL1* (Sohocki *et al*., 2000), suggesting that p.W278X is associated with LCA phenotype in our study family as well but most likely together with another heterozygous mutation in *AIPL1* not identified yet.

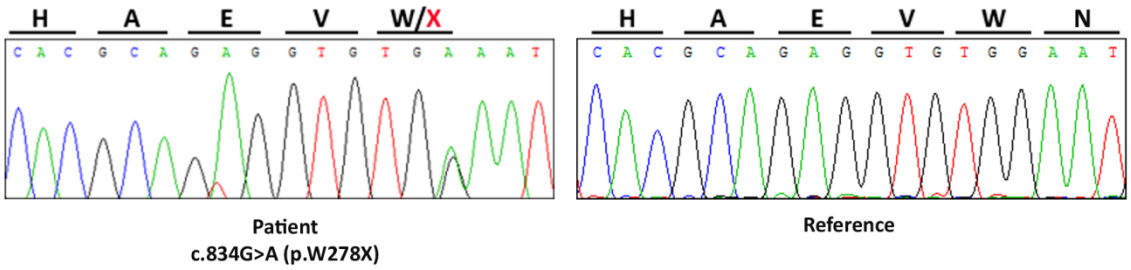

Figure 5.5. Chromatograms showing heterozygous *AIPL1* mutation in a patient, and the reference.

5.2.3. Evaluation of Mitochondrial Variants

Since mitochondrial mutations were reported in optic neuropathies, which causes blindness (Yu-Wai-Man *et al*., 2009), we investigated the exome sequencing results for mitochondrial variants. The reads aligned to the mitochondrial reference genome were investigated using IGV software, and all mismatches were noted. Variants were called after verifying each mismatched base on the map of the mitochondrial genome build hg38 (NCBI, MapViewer). Gene-based annotation was performed using GeneDistiller software, and filter-based annotation was performed by searching the variants in dbSNP138 database. After filtering the remaining variants against other in-house exome sequencing samples, seven novel or rare variants remained (Table 5.4). Three of them, at positions 13651, 14180, and 15581, were validated by Sanger sequencing. However, both the mother

and the unaffected sister also carried all three mutations without any sign of heteroplasmy, as did the patient. Hence, we excluded all mitochondrial variants.

Table 5.4. Novel and rare variants detected in the mitochondrial DNA of the LCA patient.

Position (bp)	Ref. base	Obs. base	Total depth	Depth of obs. base	Gene	Type	Change
2442	T	C	5	5	RNR2	rRNA	n.772T>C
7968	T	C	11	2	COX2	Protein-coding	p.L128P
9763	T	C	5	2	COX3	Protein-coding	p.F186S
13651	A	G	2	2	ND5	Protein-coding	p.T439A
13787	T	C	5	2	ND5	Protein-coding	p.L484P
14180*	T	C	5	5	ND6	Protein-coding	p.Y165C
15581	A	C	8	2	CYTB	Protein-coding	p.T279P

* Reported in dbSNP138 as rs2009333 with rare allele freq of 0.008.
Ref: reference, Obs: observed

5.3. Optic Atrophy (OPA)

5.3.1. Multipoint Linkage Analysis and Homozygosity Inspection

In order to find the disease locus, multipoint LOD scores were calculated using the genotyping data generated by the genome scan with 1 million SNP markers. A simplified pedigree (Figure 4.2) that did not include the unaffected sibs in the last generation and their paternal uncle 404 with unclear disease status was used in the analysis. Even then this new pedigree was too large to allow linkage analysis to be completed in a reasonable period of time. Therefore, linkage analysis was performed in two steps: In the first run, markers were selected at 0.5-cM spacing and multipoint LOD scores were calculated for all autosomes and PAR. Three loci, 6p25.3, 9q21.1 and 12q21.33 yielded LOD scores >3 (Figure 5.6). At those three loci, detailed calculations were performed using markers with 0.1-cM spacing. Negative LOD scores were obtained at 6p25.3 and 9q21.1, and locus 12q21.33 yielded a maximum LOD score of 3.80.

HCiE analysis was performed to investigate any genotype sharing among patients at the three loci that had yielded LOD scores >3 in the first linkage analysis. Three patients at 6p25.3 and all patients at 9p21.1 carry heterozygous genotypes, supporting the results of the detailed LOD score calculations using markers at 0.1-cM. At 12q21.33, four patients

share homozygosity and affected individual 403 is heterozygous along the region. The maximal homozygosity region begins at rs6539851 (85,043,911 bp) and extends to rs7978615 (90,584,821), for approximately 5.5 Mb (Figure 5.7).

Since patient 403 does not have ocular findings but instead has various psychiatric and neurologic symptoms (Table 1.4), we thought that 12q21.33 could be the locus for optic atrophy in this family, even though he is heterozygous. The other four patients who share homozygosity at 12q21.33 have optic atrophy.

Another LOD score calculation was performed for 12q21.33 using markers with 0.1-cM spacing and assuming 403 as unaffected. This analysis yielded a score of 4.85 (Figure 5.8), strongly suggesting linkage to 12q21.33, where the four individuals with OPA share homozygosity.

5.3.2. Evaluation of Exome Sequencing Results, and Analysis for *CEP290* Variant

DNA sample of patient 508 was subjected to exome sequencing. Alignment of the sequencing reads to the reference genome, variant calling, and annotation of detected variants were performed at Macrogen Inc and also in our laboratory. The results were searched for novel or rare (MAF <0.05) exonic SNVs at the candidate locus. The only such SNV was the nonsynonymous c.A62C (p.Q21P) change in gene *CEP290*. Online tool PolyPhen-2 predicted that it is probably damaging to protein function with the highest score of 1, and MutationTaster predicted it as disease causing. However, SIFT predicted that the variant is tolerable (score: 0.12). Sanger sequencing validated the variant. Patients 408, 506, 508, and 512 were homozygous and patient 403 was heterozygous for it. The available 20 members of the family plus a control panel of 118 unrelated control subjects from the population were screened for the variant with HRM curve analysis. The assay showed that the variant segregated with optic atrophy trait in the family (Figure 5.9), and none of the control individuals from the population was found to carry the variant, showing that the variant is not common in our population. Since *CEP290* p.Q21P is found to be a novel variant segregating with the OPA trait in the family and predicted to be disease causing by two tools, it probably underlies OPA in the family.

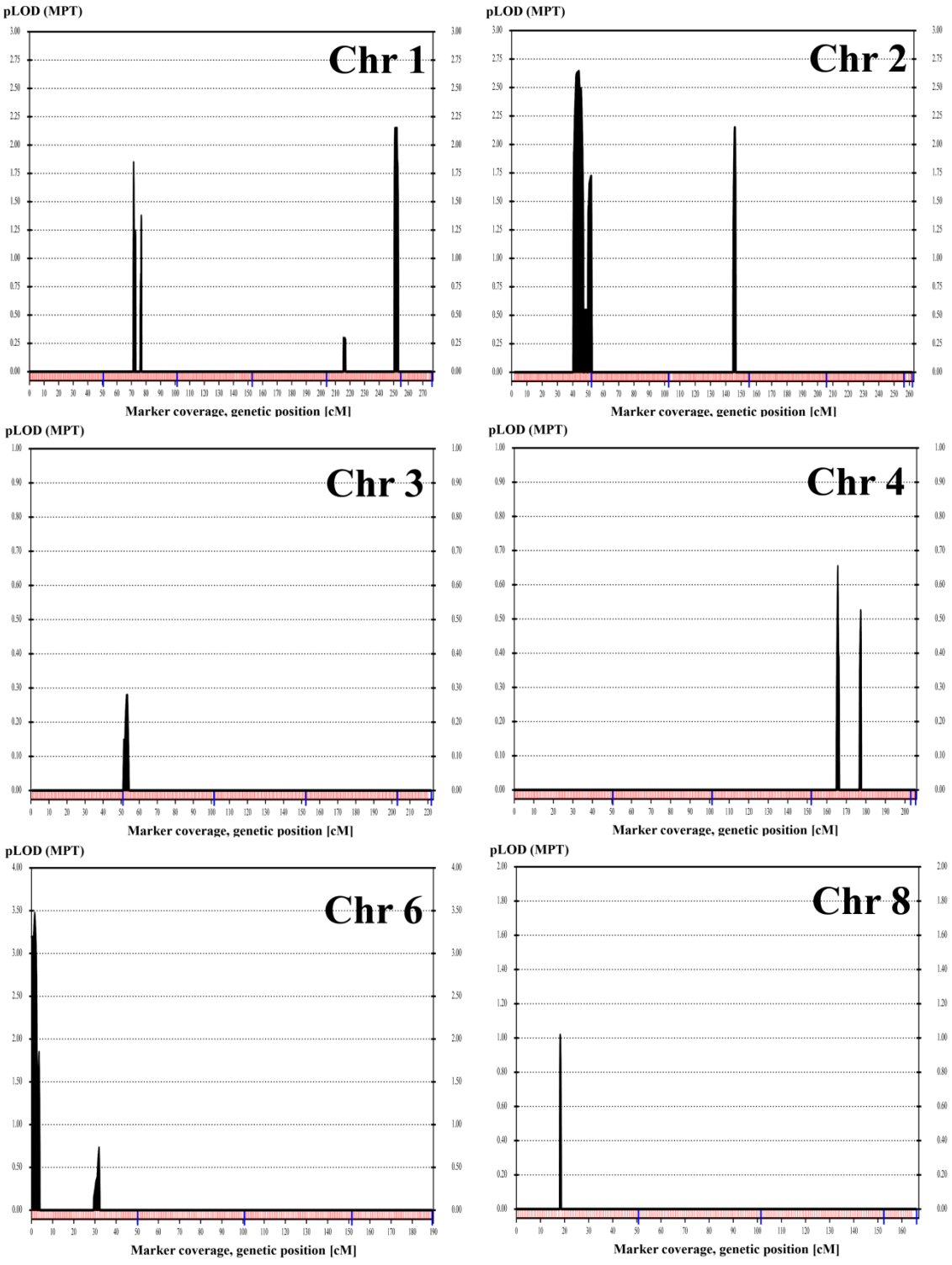

Figure 5.6. Multipoint LOD scores for all autosomes and the PARs for OPA family. Calculations were performed by selecting SNP markers at 0.5-cM spacing. Only the chromosomes with scores >0 are presented.

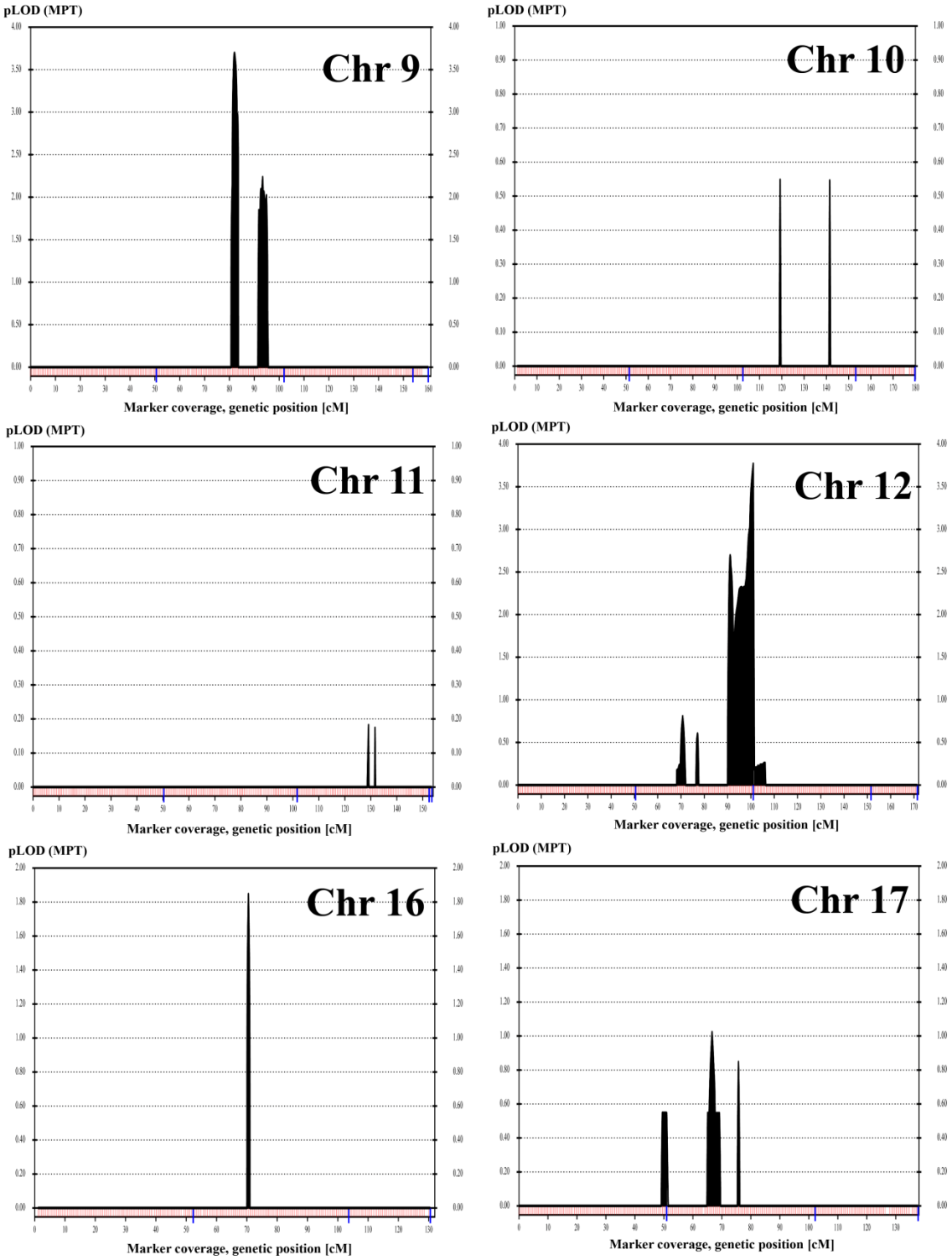

Figure 5.6. Multipoint LOD scores for all autosomes and the PARs for OPA family. Calculations were performed by selecting SNP markers at 0.5-cM spacing. Only the chromosomes with scores >0 are presented (cont.).

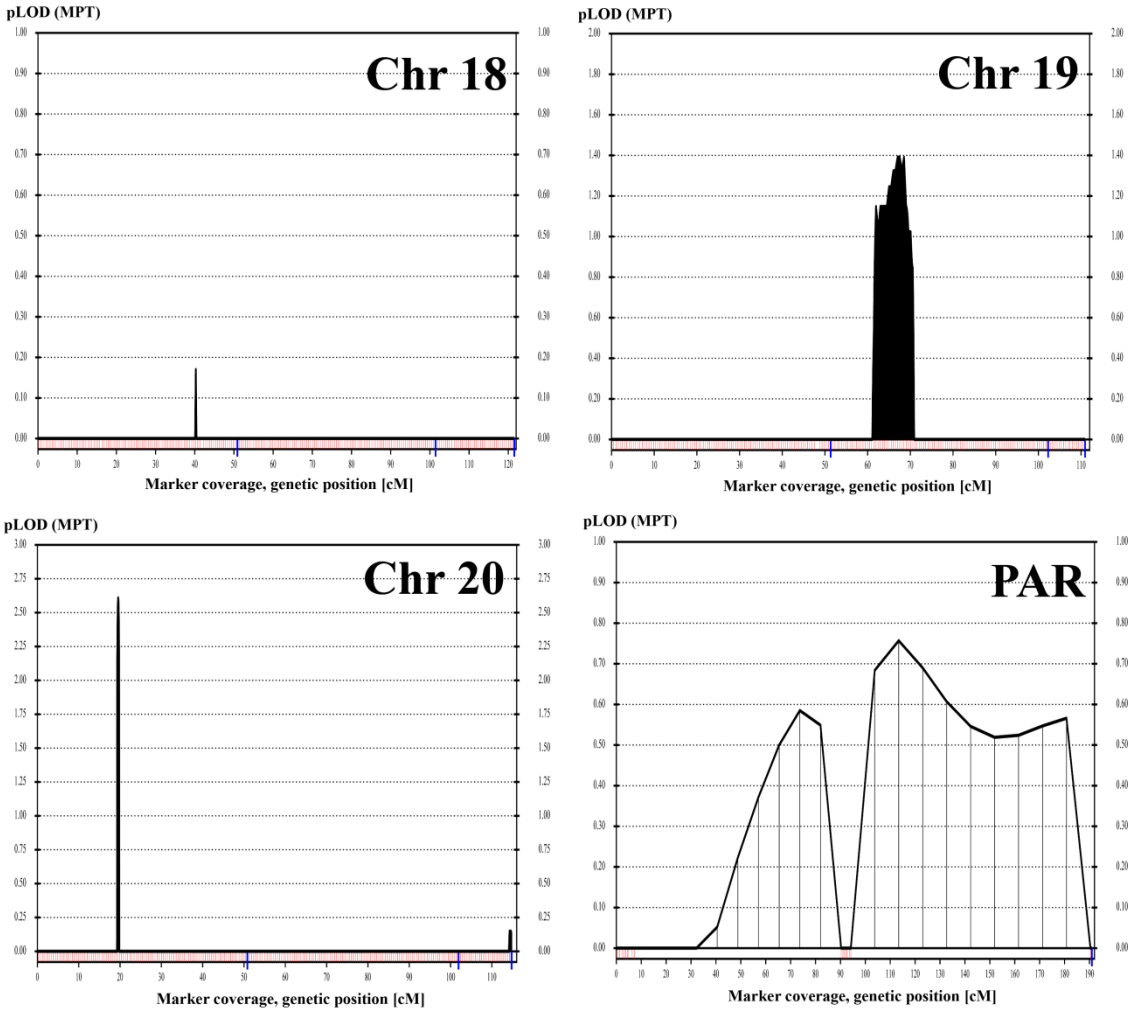

Figure 5.6. Multipoint LOD scores for all autosomes and the PARs for OPA family. Calculations were performed by selecting SNP markers at 0.5-cM spacing. Only the chromosomes with scores >0 are presented (cont.).

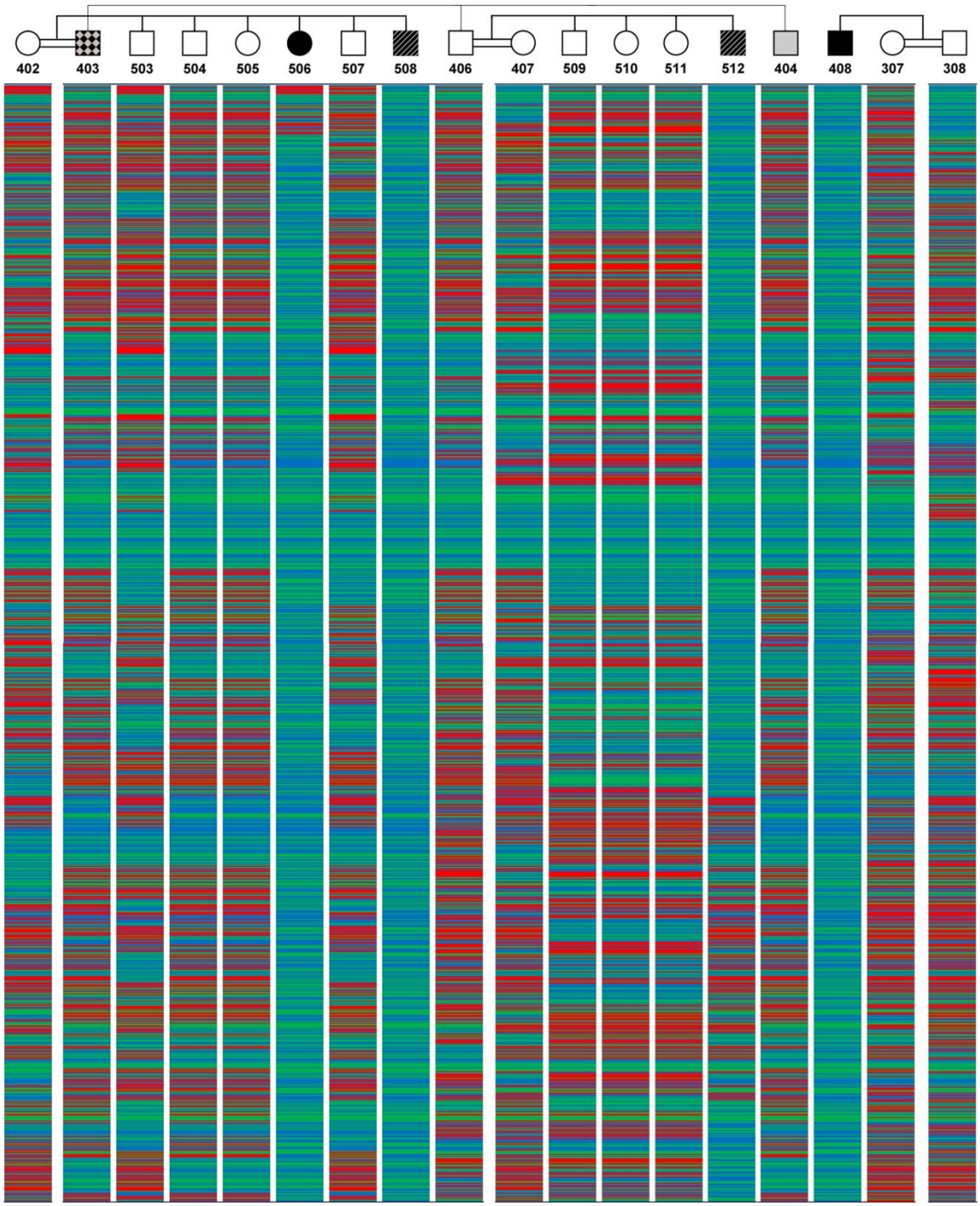

Figure 5.7. Genotypes of OPA family members between markers rs12316323 and rs791223 (84,626,797 bp and 92,860,820 bp). Genotype “AA” is shown in green, “BB” in blue, and “AB” in red. Individual 404 with unclear disease status is shown in grey; 403, who does not have OPA but neurological symptoms, in checkered black; and cousins, who have OPA with additional findings, are shown in black with lines.

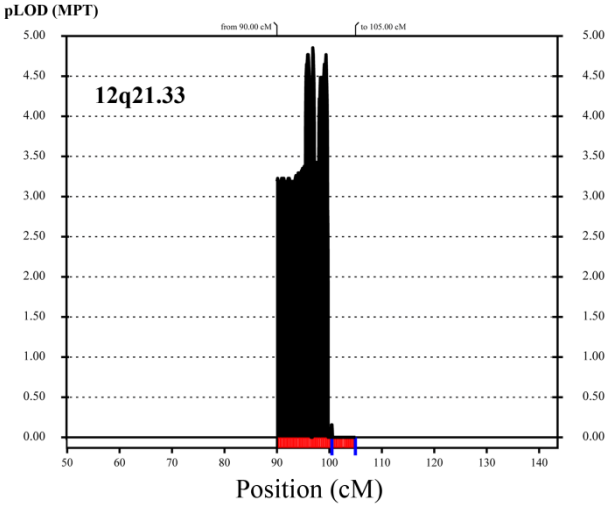

Figure 5.8. Detailed multipoint LOD scores calculated for 12q21.33. Markers were used at 0.1-cM spacing, and the pedigree in Figure 4.2 was used with the modification that father 403 was introduced as unaffected.

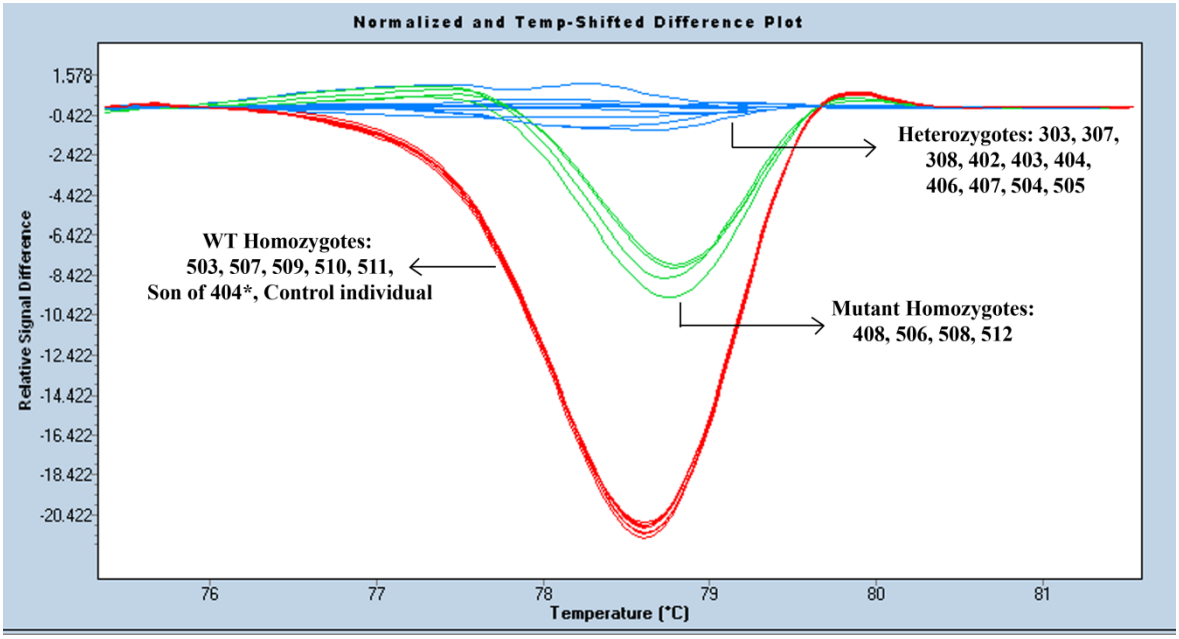

Figure 5.9. High Resolution Melting Curve assay for *CEP290* c.A62C variant. All 20 OPA family members were screened. A spike DNA was added in each sample. Asterisk denotes the individual not shown in the OPA pedigree in Figure 1.3.

5.3.3. Linkage Mapping in Search for a Second Disease Locus

As presented in Table 1.4, the phenotypes of the five patients participated in the study are variable: Patients 408 and 506 have optic atrophy only whereas patients 508 and 512 have additional symptoms such as skeletal deformities, mental retardation and kidney malfunction, and patient 403 has yet another phenotype with seizures and psychiatric findings. We thought that a second gene defect in the family could underlie the additional symptoms that patients 508 and 512 share. In order to find the putative second disease locus, linkage analysis was performed using new pedigrees that assume only 508 and 512 are affected (Figure 4.3). Since individual 403 has no symptoms in common with 508 and 512, he was introduced to the pedigrees as unaffected.

First, multipoint LOD scores were calculated by Allegro using genotyping data of 1 million markers and the small pedigree shown in Figure 4.3A. The analysis revealed 12 loci, mostly on chromosome 12, with LOD scores >2.5, and scores at eight of those loci were >3 (Table 5.5). Genotypes for markers common to the two SNP chips employed were investigated on Excel at the 12 loci that yielded scores >2.5. This investigation revealed that 508 and 512 do not share homozygosity at all the regions, including all the four loci that yielded LOD scores between 2.5 and 3. Those four loci were excluded. Multipoint LOD scores were calculated by SimWalk for the remaining eight loci and 1-cM flanking regions using the larger pedigree in Figure 4.3B and selecting markers at 0.01-cM intervals. The results are presented in Table 5.5. Three loci yielded LOD scores >3.

Exome sequencing data were investigated at the homozygous regions shared at least by individuals 508 and 512 for novel or rare exonic or splicing variants that could possibly damage the structure of the protein product. Two such variants were found: *ACSS3* c.G233A (p.G78D) at 12q15-21.33 and *PNMT* c.C586T (p.R196W) at 17q21.1. *ACSS3* is located approximately 7 Mb centromeric to *CEP290* locus and 408 is also homozygous in the region (Table 5.5). Indeed, Sanger sequencing showed that 508, 512, and 408 are homozygous for *ACSS3* c.G233A, and 506 is heterozygous. HRM assay to test family members for the variant validated this result and showed that no additional family member was homozygous for the variant (Figure 5.10). The variant is not reported in dbSNP138, 1000 Genomes or EVS databases. Also, none of the 114 samples from the population

screened via HRM was found to carry it. Its effect on protein function was investigated with PolyPhen-2 and MutationTaster. MutationTaster predicted it as disease causing and PolyPhen-2 as probably damaging.

Variant *PNMT* c.C586T was reported in dbSNP138 database as rs201626166 with an unknown frequency and in the EVS database in two subjects out of total 6501 samples in the heterozygous state. At *PNMT* locus, only 508 and 512 share homozygosity. However, Sanger sequencing showed that 512 is heterozygous for the variant (Figure 5.11). Among the five sequenced individuals, namely, 403, 408, 506, 508, and 512, only 508 was homozygous. Therefore, this variant was excluded. *ACSS3* c.G233A remained as the only candidate for the mutation underlying the putative second disease.

Alignment and variant calling were repeated using GATK, and the homozygosity regions on Table 5.5 were re-evaluated for any candidate variant undetected by BWA and SAMTools. No such variant was found.

Total exome sequence results were analyzed for novel, homozygous and possibly damaging mutations, and the resulting list was compared against the results of 25 other in-laboratory samples. Among the 40 filtered variants, only three of them, namely, *ACSS3* c.G233A, *CEP290* c.A62C and *PNMT* c.C586T, were located in a region of homozygosity shared by affected cousins 508 and 512. Since *PNMT* c.C586T was excluded due to heterozygosity in 512 and *CEP290* c.A62C was identified as the variant possibly causing OPA, *ACSS3* c.G233A remained as the only candidate after total exome sequence results analysis as well.

Table 5.5. Candidate loci for a second disease gene in OPA family.

Loci	**LOD score (Allegro)**	**LOD score (SimWalk)**	**Maximum region of homozygosity shared by 508 and 512**	**Size (Kb)**	**Candidate variants**	**Other individuals homozygous in the region**
10q21.1	3.28	4.16	rs11006428 - rs11006576 (61,050,653 bp - 61,322,015 bp)	271	None	509
12p12.1	3.50	2.41	No homozygosity	-	-	-
12p11.23	3.44	2.16	rs1862996 - rs10492374 (29,277,563 bp - 29,539,458)	262	None	503

Table 5.5. Candidate loci for a second disease gene in OPA family (cont.).

Loci	LOD score (Allegro)	LOD score (SimWalk)	Maximum region of homozygosity shared by 508 and 512	Size (Kb)	Candidate variants	Other individuals homozygous in the region
12q12	3.47	2.16	rs10506247 - rs11183193 (43,038,952 bp - 44,427,639 bp)	1,389	None	506
12q13.13	3.54	3.84	rs7971175 - rs10783655 (53,266,482 bp - 53,475,139 bp)	209	None	308 and 408
12q15-q21.33	3.54	4.16	rs34699916 - rs7978615 (74,779,321 bp - 89,108,952 bp)	14,330	ACSS3 NM_024560 c.G233A p.G78D	408
17q21.1	3.20	4.16	rs9674546 - rs2134808 (37,476,164 bp - 38,035,609 bp)	559	PNMT NM_002686 c.C586T p.R196W	None
17q21.2-q21.31	3.41	2.29	No homozygosity	-	-	-

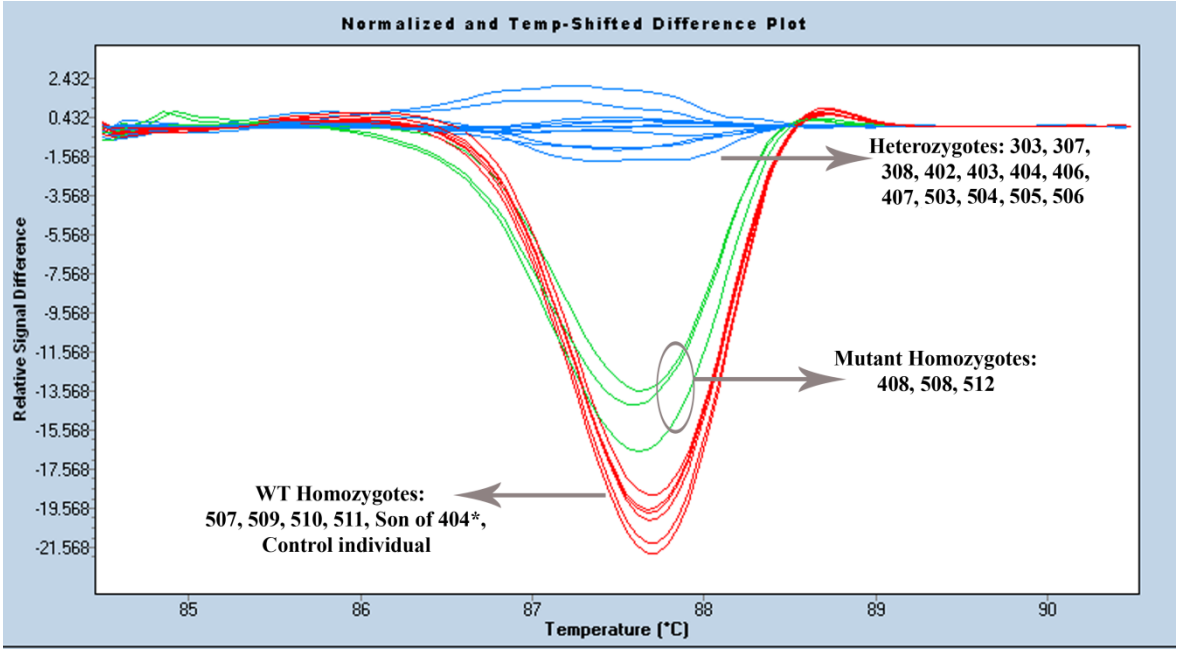

Figure 5.10. High Resolution Melting Curve assay used to screen 20 OPA family members for *ACSS3* c.G233A. A spike DNA was added in each sample. Asterisk denotes the individual not shown in the OPA pedigree in Figure 1.3.

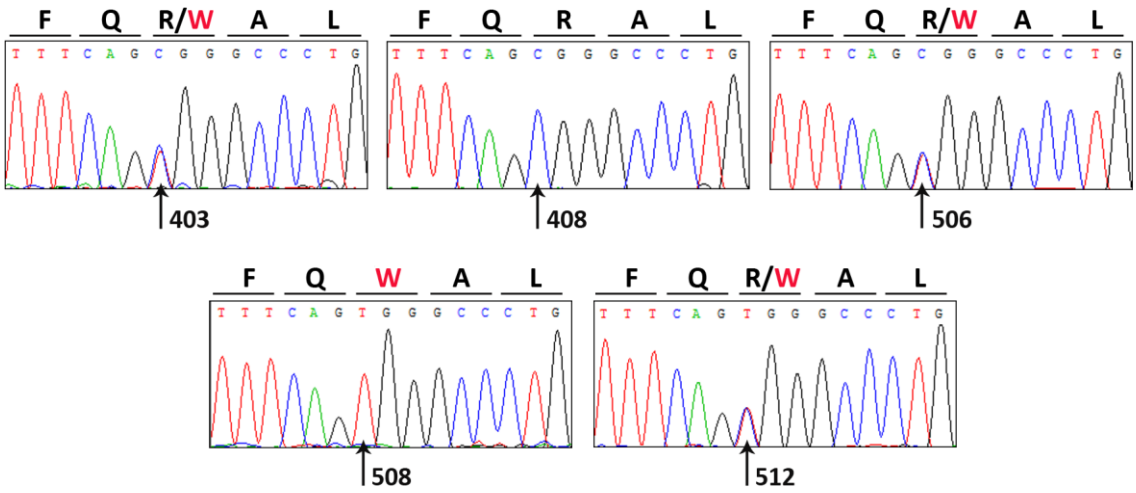

Figure 5.11. Chromatograms showing *PNMT* c.C586T (p.R196W). Individual 508 is homozygous for the variant and individuals 403, 506 and 512 were heterozygous for it. Individual 408 has wild-type alleles.

5.4. Amyotrophic Lateral Sclerosis (ALS)

5.4.1. Investigation of Linkage to Chromosome X and to Known ALS Loci

At the beginning of the study, X-linked recessive inheritance was investigated, since all affected subjects are males and their mothers were sisters. X-chromosomes of all participants were scanned using nine polymorphic microsatellite markers, and haplotypes were constructed (Figure 5.12). There was no haplotype sharing among patients in any region that could possibly be IBD, hence X-linked inheritance was excluded. This finding was later supported by the comparison of X-chromosome genotypes of SNP markers on Excel.

All participants were genotyped with microsatellite markers flanking the two known disease genes for autosomal recessive ALS, *ALSIN* (2q33.1) and *SPATACSIN* (15q21.1). Linkage to any of those gene loci was excluded.

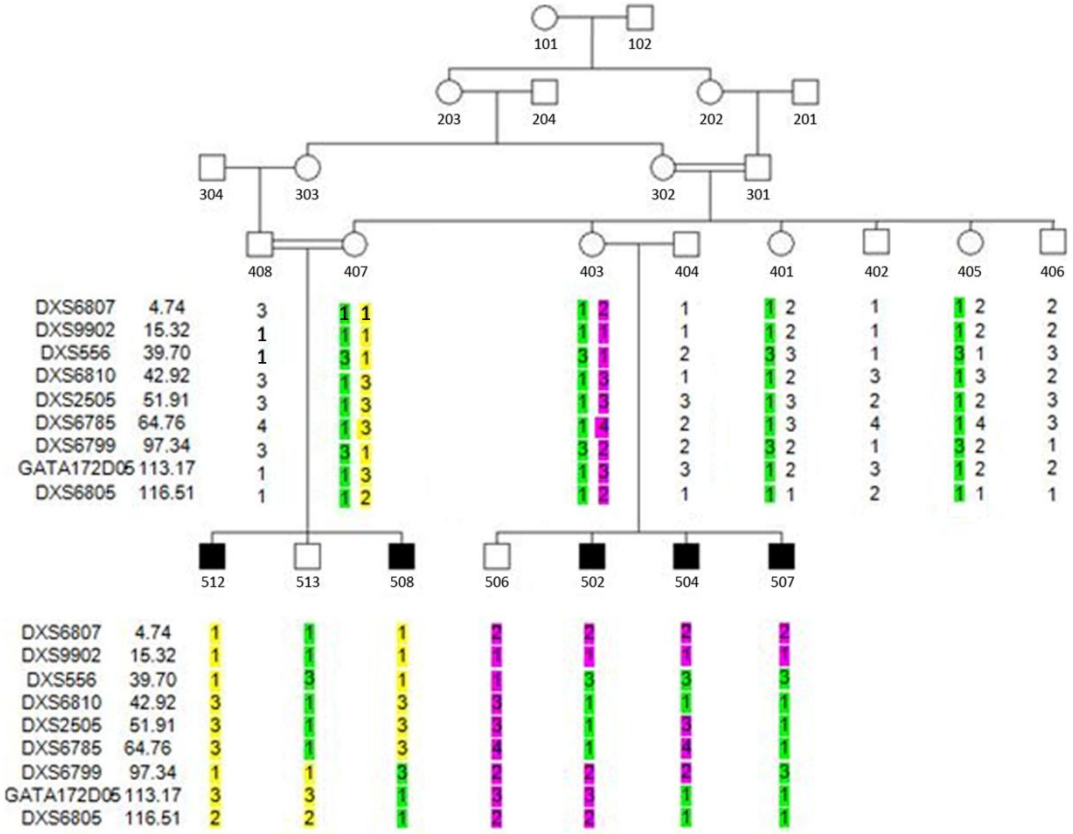

Figure 5.12. Haplotypes for ALS family with nine markers on the X-chromosome.

5.4.2. Multipoint Linkage Analysis

Whole-genome multipoint linkage analysis was performed on Allegro in order to localize the disease gene. Genotyped individuals 506, 512 and 513 were not included in the analysis and the consanguinity of 408 was disregarded due to the computational limits of the program. A maximum LOD score of 3.36 was obtained at 9p21.1-p12. Five other loci (15q14, 19q13.32-13.33, 19q13.33, 19q13.42, and 22q13.33) also yielded scores >3 (Figure 5.13). At all loci that yielded LOD scores >3, detailed calculations were performed using SimWalk. In this analysis we included all genotyped individuals and used the actual pedigree. LOD scores decreased at five loci; 19q13.32-13.33 yielded 2.55, and the remaining four loci yielded negative scores. Locus 9p21.1-p12 yielded a maximal LOD score of 5.25, standing out as the best candidate (Figure 5.14) for disease locus.

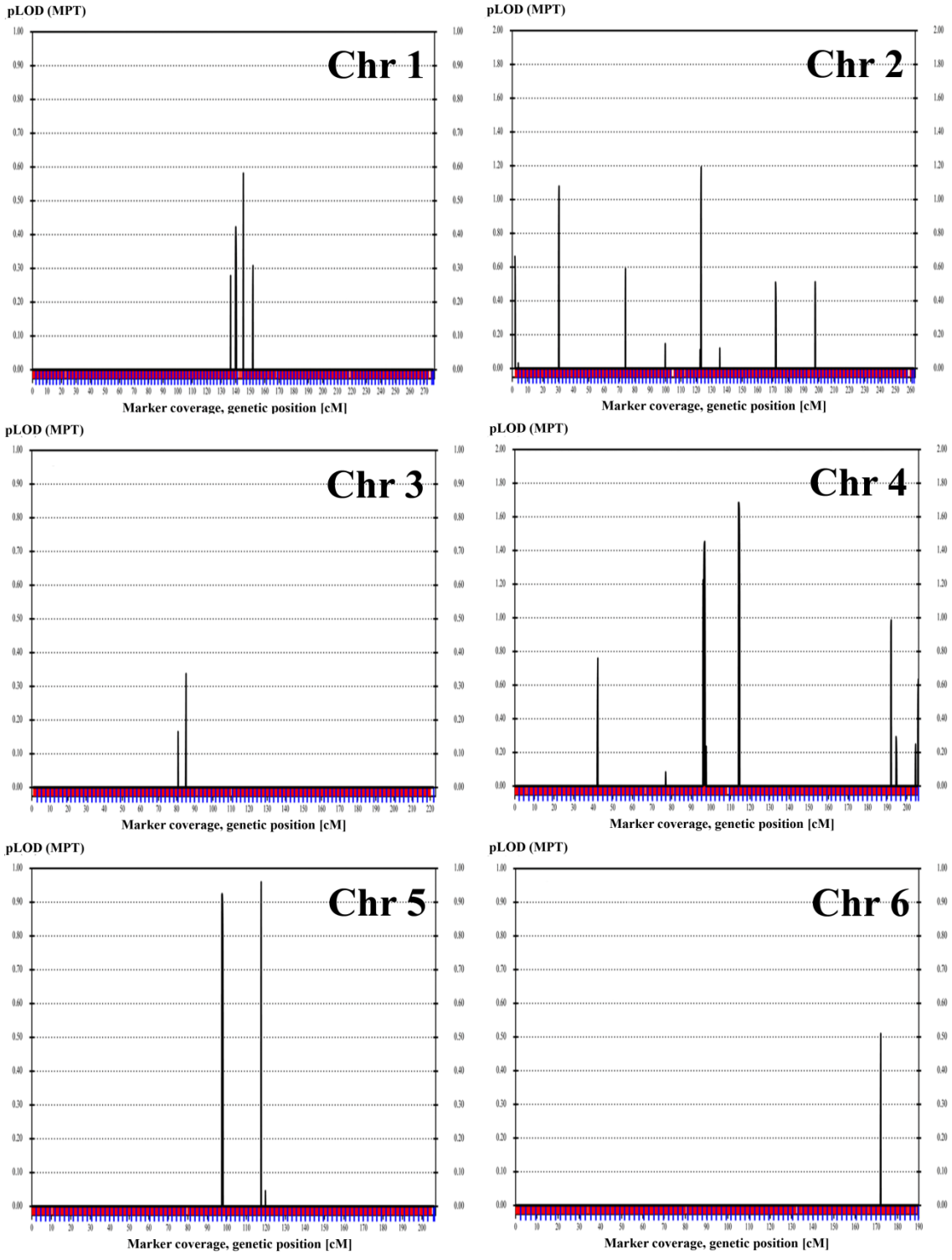

Figure 5.13. Whole-genome multipoint LOD scores for ALS family. Only the chromosomes with scores >0 are presented.

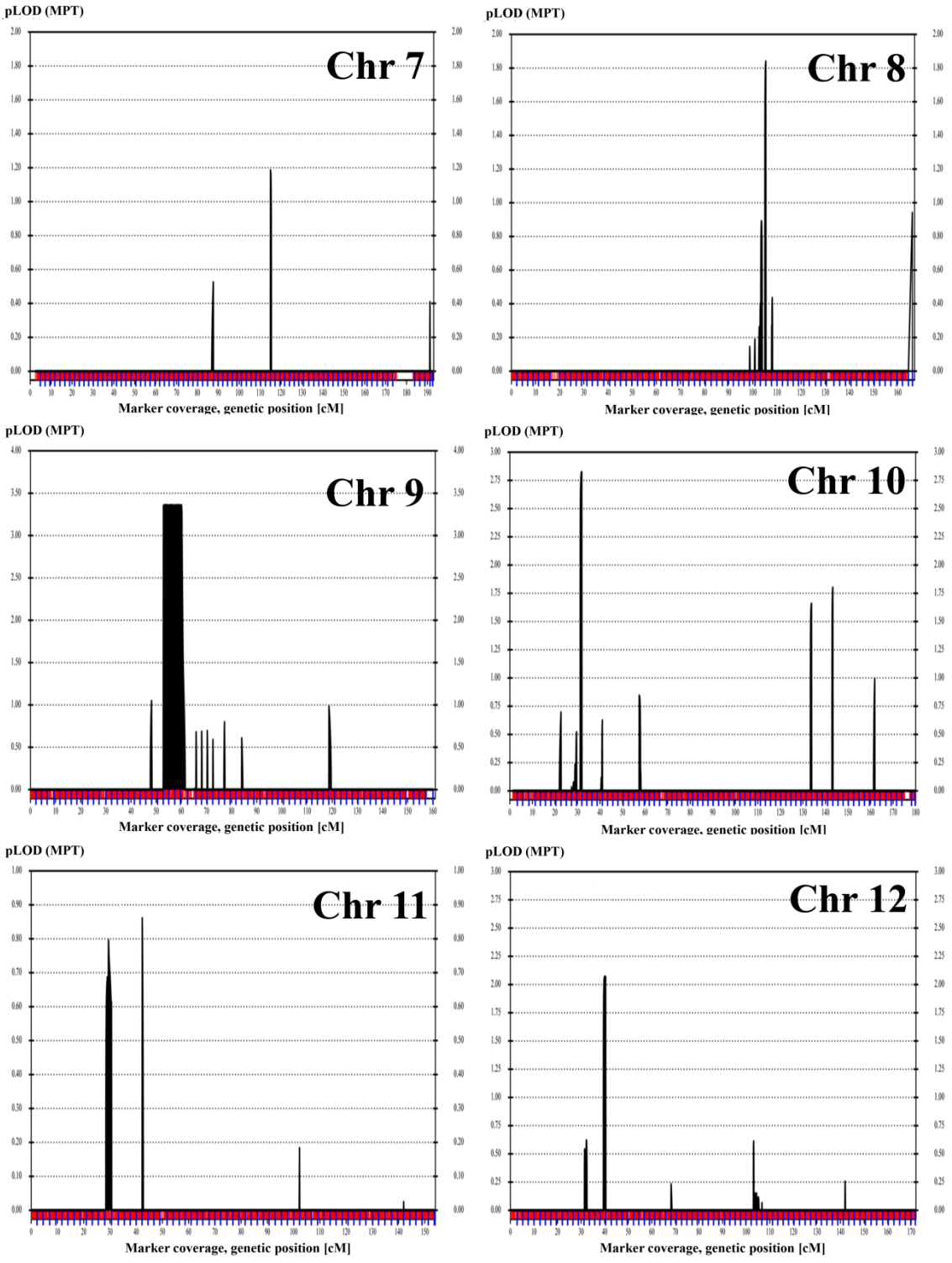

Figure 5.13. Whole-genome multipoint LOD scores for ALS family. Only the chromosomes with scores >0 are presented (cont.).

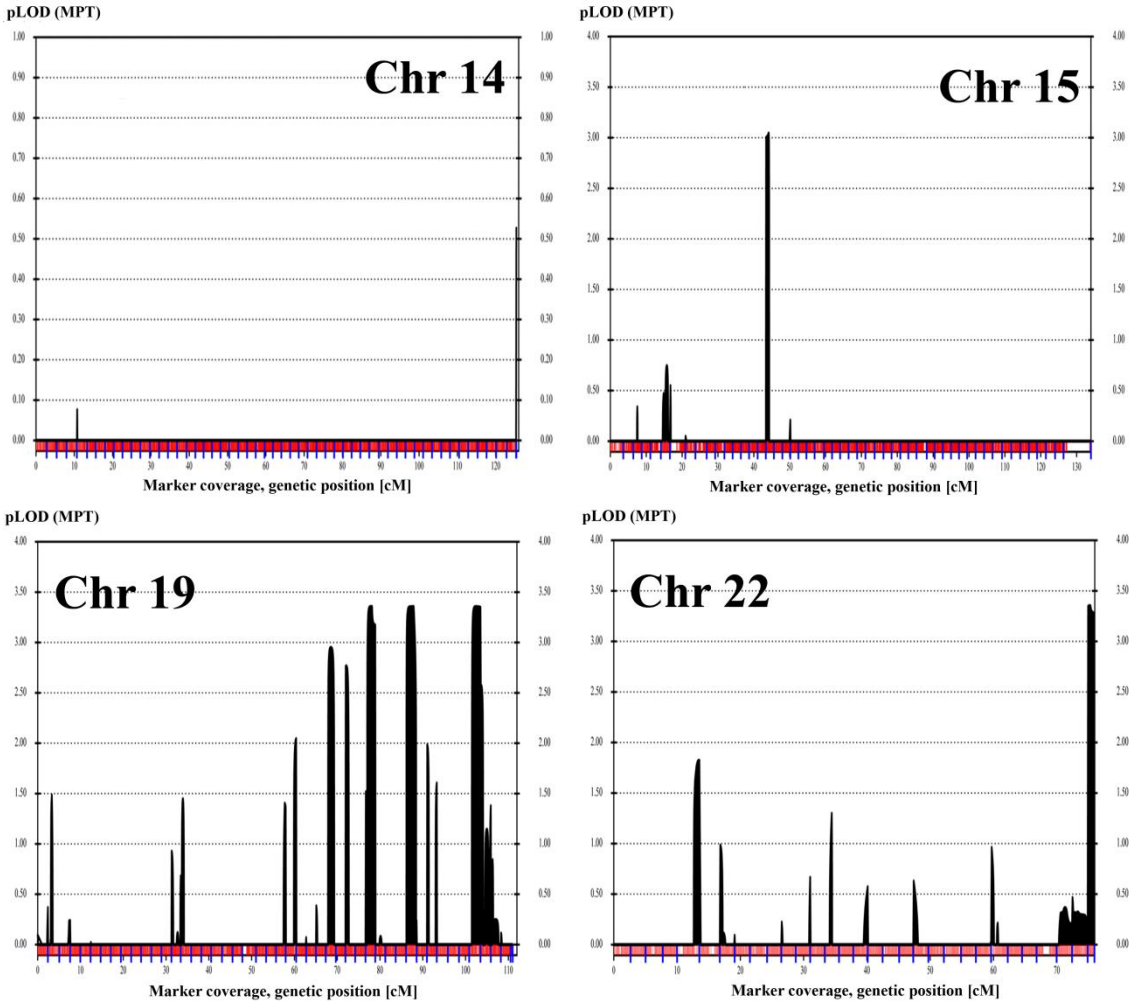

Figure 5.13. Whole-genome multipoint LOD scores for ALS family. Only the chromosomes with scores >0 are presented (cont.).

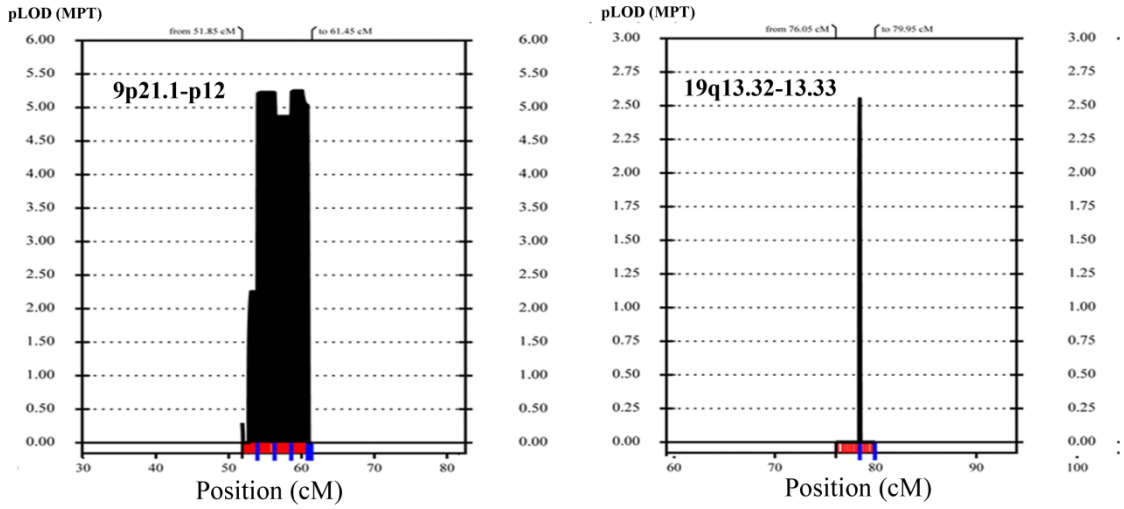

Figure 5.14. Detailed multipoint LOD scores at loci that yielded scores >3 in the initial analysis for ALS family. Consanguinity between 407 and 408 was taken into account. Only the loci with scores >0 are presented.

5.4.3. Haplotype Investigation at 9p21.1-p12

At 9p21.1-p12, genotypes were analyzed using Excel to assess homozygosity sharing among patients. The maximal region of homozygosity shared by all patients was approximately 10 Mb, delineated by markers rs321746 (28,759,983 bp) and rs2993129 (38,563,171 bp). Haplotypes were constructed by SimWalk for this region, and all patients were found to share the same, possibly IBD haplotype (Table 5.6). According to the results of the linkage analysis and the haplotype investigation together, the disease gene we search lies most probably at 9p21.1-p12.

5.4.4. Evaluation of Candidate Variants

Patient 506 was subjected to exome sequencing. Capture probes targeted a total of 62 Mb genomic regions aiming to cover 95% of the CCDS database. The sequencing statistics are given in Table 5.7.

A total of 62 exonic or splicing variants were obtained at the candidate region. Novel and rare (frequency ≤0.05) variants were selected, and the variants found in other 25 in-laboratory exome sequencing samples were filtered out. The remaining five variants are

presented in Table 5.8. *PIGO* c.1430_1431insG was shown to be a false call by Sanger sequencing. The depth of *NPR2* c.1577_1578insCG was only one in 81, and *TLN1* c.C2749T was synonymous. Those variants were not evaluated further. The remaining two variants, namely, *GNE* c.A2114G (p.H705R) and *ALDH1B1* c.1161delC (p.G388EfsX23) were validated by Sanger sequencing (Figures 5.15 and 5.16). Neither of those variants was in the results of 25 other in-laboratory exome sequencing samples.

Alignment and variant calling were performed again using GATK, and the results were very different than those obtained by BWA and SAMTools. *GNE* c.A2114G and *ALDH1B1* c.1161delC were also detected; however, most of the remaining 60 exonic and splicing variants detected at the candidate region by BWA and SAMTools were not detected. Besides the variants in *GNE* and *ALDH1B1*, 66 exonic and splicing variants were detected by GATK. Among them, only three variants had depths >60% of total depth: one variant was synonymous and one variant was found in other in-laboratory exome samples. The remaining nonsynonymous variant *HINT2* c.A61G (p.T21A) had a total depth of only two. Visualization of the relevant reads by BamView did not verify the variant. GATK analysis supported *GNE* c.A2114G and *ALDH1B1* c.1161delC as the best candidate variants.

Segregation of variants *ALDH1B1* c.1161delC and *GNE* c.2114A>G with ALS in the family was investigated via HRM assays. Both variants were shown to segregate with the disease (Figure 5.17). Additionally, all patients, available unaffected siblings and parents were analyzed by Sanger sequencing for the *ALDH1B1* mutation to validate the segregation of the mutation with the disease. Population control samples were screened via HRM for the *ALDH1B1* and *GNE* variants, 175 and 155 subjects, respectively. *GNE* variant was not found in the control samples. *ALDH1B1* variant was detected in three individuals in the heterozygous state (Figure 5.17 and 5.18). When we add 25 exome sequencing samples to the 175 samples analyzed via HRM assay, the frequency of *ALDH1B1* mutation in the Turkish population is calculated as 0.0075 (3 in 400 alleles), which is close to the value given in dbSNP137 and EVS databases (0.007 in both). The variant was not found in the homozygous state among 6253 subjects in EVS. All coding regions of *ALDH1B1* in 90 unrelated ALS patients were also analyzed by HRM curve analysis. No other mutation in this gene was detected. The missense *GNE* variant was

investigated by online tools PolyPhen-2, SIFT, and Mutation Taster to predict possible damage to protein structure and function. Mutation Taster predicted the variant to be damaging, PolyPhen-2 predicted it as benign (score: 0.057), and SIFT predicted it as tolerable (score: 0.38).

The abovementioned analyses did not highlight either variant *GNE* c.2114A>G or *ALDH1B1* c.1161delC as a better candidate for ALS mutation. On the other hand, the *GNE* variant is missense and predicted as benign by PolyPhen whereas the *ALDH1B1* variant is truncating; this which favors the latter to be the cause of the disease. To propose one of them as the causative mutation, functions of the genes and their possible associations with the phenotype in our study family were investigated in the literature (see Section 6.4).

5.4.5. Sequencing *SIGMAR1* in a Patient

While the study was in progress, a homozygous missense mutation in *SIGMAR1* (at 9p13.3) in juvenile ALS in a consanguineous Saudi Arabian family was reported (Al-Saif *et al.*, 2011). Exome sequencing results did not reveal any variant in *SIGMAR1* in our sample. Since this gene resided at our candidate locus, we computed the coverage of its exons using BEDTools to see whether there were uncovered parts in this gene. Two exons were not fully read. Although the uncovered parts were mostly UTRs, we analyzed them as well as the fully covered parts of all four exons by Sanger sequencing in patient 512. We did not detect any mutations.

5.4.6. Assessment of *ALDH1B1* Transcript Levels in Various Tissues

The abundance of *ALDH1B1* transcripts was investigated in 11 brain regions plus six non-neural tissues. *ALDH1B1* transcripts were detected in all neural tissues tested; however, most abundant expression was observed in the liver and the skeletal muscle (Figure 5.19). The results suggested that *ALDH1B1* is ubiquitously expressed.

Table 5.6. Haplotypes for selected SNP markers at 9p21.1-p12 for ALS family. The genotypes of individuals 404 and 408 were deduced and are shown in italics. The disease haplotype and ID numbers of affected individuals are highlighted grey.

Marker ID	*Position (Mb)*	404		403		**502**		**504**		506		**507**		408		407		**508**		**512**		513	
rs7865194	28.23	**2**	*1*	**2**	1	**2**	**2**	**2**	1	1	1	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs6476056	28.33	**2**	*1*	**2**	1	**2**	**2**	**2**	1	1	1	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs1331919	28.44	***1***	*2*	**1**	2	**1**	**1**	**1**	2	2	2	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs11793079	28.68	***1***	*2*	**1**	1	**1**	**1**	**1**	**1**	2	1	**1**	**1**	*2*	***1***	**1**	2	**1**	**1**	**1**	**1**	2	2
rs16914476	29.21	**2**	*2*	**2**	2	**2**	**2**	**2**	**2**	2	2	**2**	**2**	*1*	**2**	**2**	2	**2**	**2**	**2**	**2**	1	2
rs7028834	29.65	**2**	*2*	**2**	2	**2**	**2**	**2**	**2**	2	2	**2**	**2**	*1*	**2**	**2**	2	**2**	**2**	**2**	**2**	1	2
rs4878418	29.82	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*2*	***1***	**1**	2	**1**	**1**	**1**	**1**	2	2
rs1962543	30.01	**2**	*1*	**2**	1	**2**	**2**	**2**	**2**	1	1	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs10118094	30.43	***1***	*2*	**1**	1	**1**	**1**	**1**	**1**	2	1	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs1332233	30.82	**2**	*1*	**2**	2	**2**	**2**	**2**	**2**	1	2	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs10970183	31.18	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*2*	***1***	**1**	2	**1**	**1**	**1**	**1**	2	2
rs10511880	31.54	**2**	*1*	**2**	1	**2**	**2**	**2**	**2**	1	1	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs1887874	31.92	***1***	*2*	**1**	1	**1**	**1**	**1**	**1**	2	1	**1**	**1**	*2*	***1***	**1**	2	**1**	**1**	**1**	**1**	2	2
rs10117454	32.33	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs11788663	32.76	**2**	*1*	**2**	2	**2**	**2**	**2**	**2**	1	2	**2**	**2**	*1*	**2**	**2**	1	**2**	**2**	**2**	**2**	1	1
rs1016674	32.98	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*2*	***1***	**1**	2	**1**	**1**	**1**	**1**	2	2
rs855465	33.67	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*2*	***1***	**1**	2	**1**	**1**	**1**	**1**	2	2
rs216345	33.80	**2**	*1*	**2**	2	**2**	**2**	**2**	**2**	1	2	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs3739690	33.97	***1***	*1*	**1**	2	**1**	**1**	**1**	**1**	1	2	**1**	**1**	*1*	***1***	**1**	1	**1**	**1**	**1**	**1**	1	1
rs12376762	34.16	**2**	*1*	**2**	1	**2**	**2**	**2**	**2**	1	1	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs11547035	34.46	**2**	*2*	**2**	2	**2**	**2**	**2**	**2**	2	2	**2**	**2**	*1*	**2**	**2**	2	**2**	**2**	**2**	**2**	1	2
rs277598	34.78	***1***	*1*	**1**	1	**1**	**1**	**1**	**1**	1	1	**1**	**1**	*2*	***1***	**1**	1	**1**	**1**	**1**	**1**	2	1
rs568300	35.09	***1***	*1*	**1**	2	**1**	**1**	**1**	**1**	1	2	**1**	**1**	*2*	***1***	**1**	1	**1**	**1**	**1**	**1**	2	1
rs10972462	35.43	**2**	*2*	**2**	1	**2**	**2**	**2**	**2**	2	1	**2**	**2**	*2*	**2**	**2**	2	**2**	**2**	**2**	**2**	2	2
rs1570246	35.75	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs1408469	36.01	**2**	*1*	**2**	2	**2**	**2**	**2**	**2**	1	2	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs12376699	36.38	**2**	*2*	**2**	1	**2**	**2**	**2**	**2**	2	1	**2**	**2**	*1*	**2**	**2**	2	**2**	**2**	**2**	**2**	1	2
rs10814430	36.69	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*2*	***1***	**1**	2	**1**	**1**	**1**	**1**	2	2
rs7025013	36.80	***1***	*2*	**1**	1	**1**	**1**	**1**	**1**	2	1	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs3780151	36.93	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs7872067	37.09	***1***	*1*	**1**	1	**1**	**1**	**1**	**1**	1	1	**1**	**1**	*2*	***1***	**1**	1	**1**	**1**	**1**	**1**	2	1
rs10814518	37.30	**2**	*2*	**2**	2	**2**	**2**	**2**	**2**	2	2	**2**	**2**	*1*	**2**	**2**	2	**2**	**2**	**2**	**2**	1	2
rs1571234	37.49	**2**	*2*	**2**	1	**2**	**2**	**2**	**2**	2	1	**2**	**2**	*1*	**2**	**2**	2	**2**	**2**	**2**	**2**	1	2
rs4878169	37.67	***1***	*2*	**1**	1	**1**	**1**	**1**	**1**	2	1	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs1082865	37.93	***1***	*1*	**1**	2	**1**	**1**	**1**	**1**	1	2	**1**	**1**	*2*	***1***	**1**	1	**1**	**1**	**1**	**1**	2	1
rs10738992	38.12	**2**	*1*	**2**	2	**2**	**2**	**2**	**2**	1	2	**2**	**2**	*2*	**2**	**2**	1	**2**	**2**	**2**	**2**	2	1
rs1885490	38.31	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*1*	***1***	**1**	2	**1**	**1**	**1**	**1**	1	2
rs10973948	38.63	***1***	*2*	**1**	2	**1**	**1**	**1**	**1**	2	2	**1**	**1**	*1*	*2*	**1**	2	**1**	**1**	2	**1**	1	2
rs6476761	39.00	***1***	*1*	**1**	2	**1**	**1**	**1**	**1**	1	2	**1**	**1**	*1*	*2*	**1**	2	**1**	**1**	2	**1**	1	2

Table 5.7. Exome sequencing statistics for the ALS patient.

Target regions (bp)	62,085,286
% Coverage of target regions (more than 1X)	94.3%
% Coverage of target regions (more than 10X)	82.7%
Median read depth of target regions	43X
Mean read depth of target regions	48.4X
Number of total SNPs found	64,770
Number of coding SNPs found	18,240
Number of indels found	13,890
Number of coding indels found	738

Table 5.8. Novel and rare variants at the disease locus 9p21.1-p12 in ALS family. The variants were detected by analyzing exome sequencing results using BWA, SAMTools and ANNOVAR.

Location	Reference base	Observed base	Call	Total depth	Depth of observed base	Variant and change	SNP ID and frequency
35090636	-	C	Het	18	2	*PIGO* NM_001201484 c.1430_1431insG, frameshift insertion	Not in databases
35715061	G	A	Hom	47	47	*TLN1* NM_006289 c.C2749T, syn.	Not in databases
35801942	-	CG	Het	81	1	*NPR2* NM_003995 c.1577_1578insCG, frameshift insertion	Not in databases
36217510	T	C	Hom	70	70	*GNE* NM_001128227 c.A2114G, nonsyn.	Not in databases
38396906	C	-	Hom	68	59	*ALDH1B1* NM_000692 c.1161delC, frameshift deletion	rs201408956, 0.007

Syn: synonymous SNV
Nonsyn: Nonsynonymous SNV
NA: not available

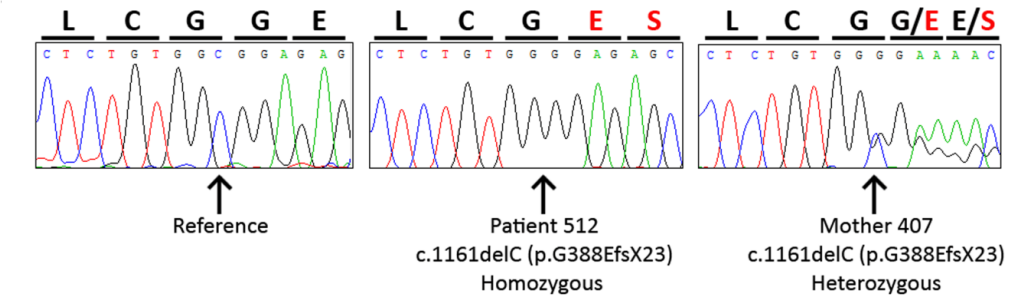

Figure 5.15. Chromatograms showing *ALDH1B1* c.1161delC (p.G388EfsX23). The reference sequence and the mutation in a homozygous individual and the heterozygous mother are shown.

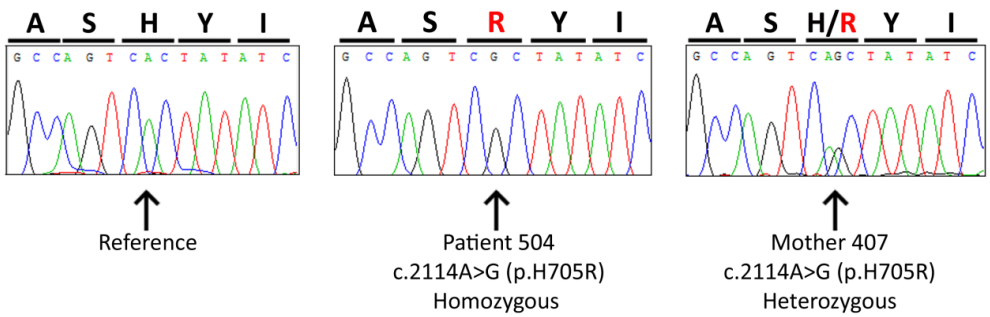

Figure 5.16. Chromatograms showing *GNE* c.2114A>G (p.H705R). The reference sequence and the mutation in a homozygous individual and the heterozygous mother are shown.

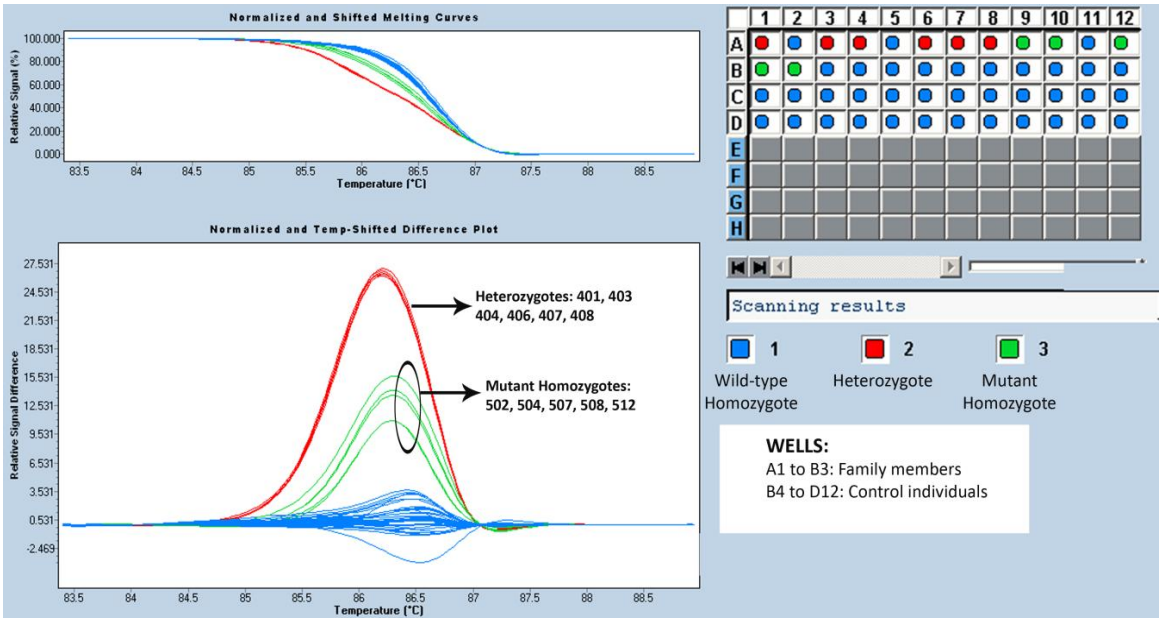

Figure 5.17. High Resolution Melting Curve assay for *ALDH1B1* c.1161delC testing. Control individuals selected from population and 15 family members were screened. A spike DNA was added in each sample.

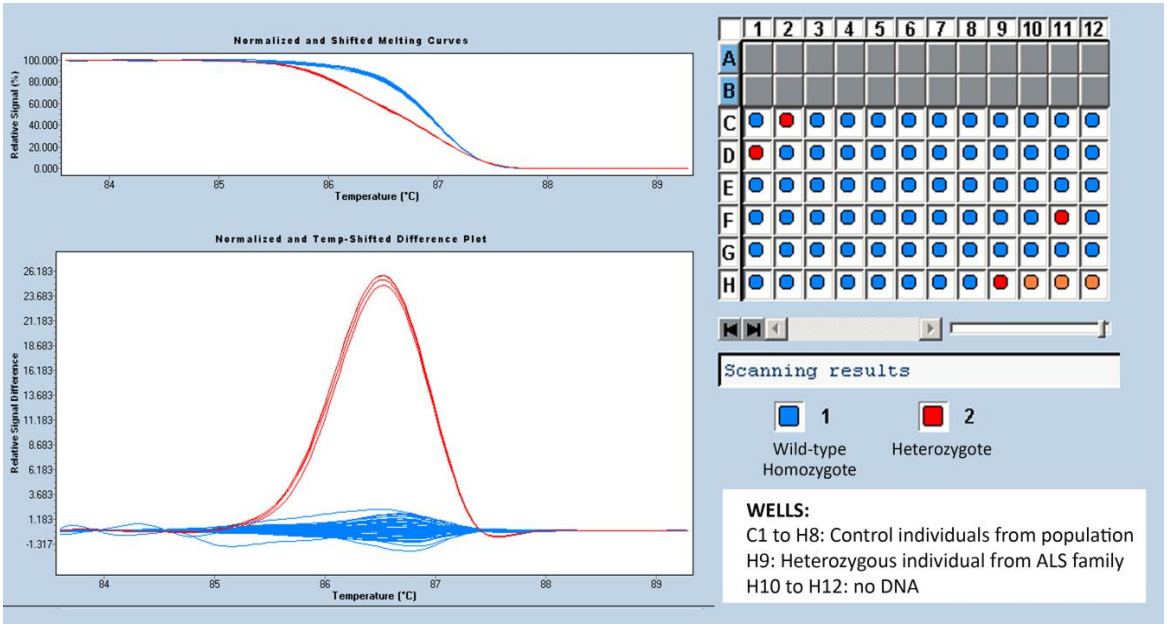

Figure 5.18. High Resolution Melting Curve assay for *ALDH1B1* c.1161delC in the population samples. The results of 68 samples are presented.

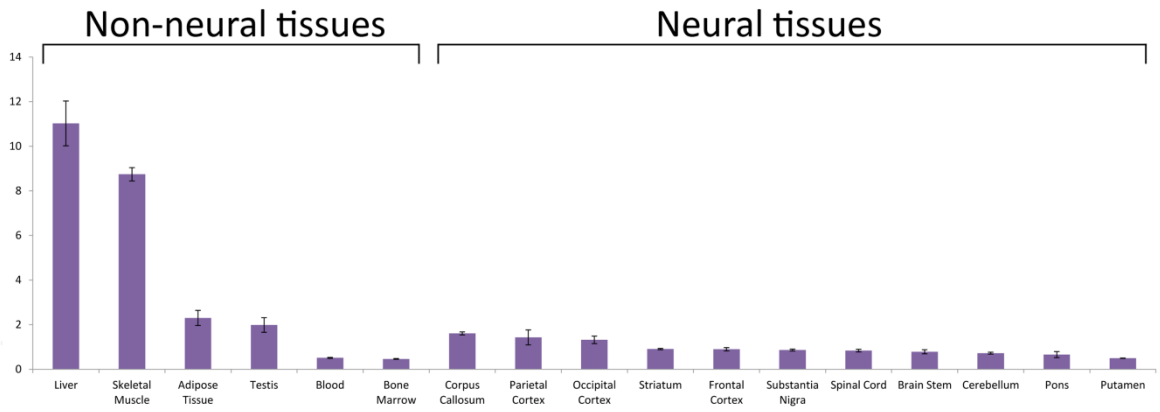

Figure 5.19. Quantification of transcript levels for *ALDH1B1* relative to *TUBB* in various tissues.

5.5. Rett Syndrome (RTT)

5.5.1. Multipoint Linkage Analysis and Haplotype Inspection

DNA samples of the three children and their parents had been previously genotyped using microsatellite markers. Using the generated genotype data, whole-genome multipoint LOD score analysis was performed on SimWalk v2.91 under the assumption of autosomal recessive inheritance, full penetrance, and a disease gene frequency of 1 in 10,000. Locus 22q11.1 yielded the highest LOD score of 1.83 (Figure 5.20). All loci that yielded LOD scores >1 were analyzed further. There were four such loci, 7q11.22-21.11, 7q35, 14q22.1, and 22q11.1. Haplotypes were constructed via Allegro v1.2c for chromosomes 7, 14 and 22. At 7q35 and 22q11.1, the two patients shared the same homozygous genotype whereas the unaffected sib carried a heterozygous genotype. At 7q11.22-21.11 and 14q22.1, all three sibs shared homozygous genotypes. Family members were genotyped with additional microsatellite markers: D7S2212 and D7S2555 at 7q21.11-21.12, and D14S139 at 14q21.3. At least one patient was heterozygous for each of the three markers utilized. Genotypes were added to the previous data set, and linkage analysis was repeated with the same parameters. Maximal LOD scores dropped below 1 at 7q11.22-21.11 and 14q22.1, eliminating those loci. The genome scan data of patient 401 for 610,000 SNP markers at loci 7q35 and 22q11.1 were analyzed to investigate homozygosity. The patient carried a continuous homozygous SNP genotype at 22q11.1 between markers rs12157537 and rs9624480 (16114244 bp and 24871942 bp); in contrast, at 7q35 she had a heterozygous

genotype throughout the region, eliminating the region. Thus, 22q11.1 remained as the only candidate locus. Haplotypes for chromosome 22 are presented in Figure 5.21.

Later a genome scan with 700,000 SNP markers was performed for all seven participating members of the family. Homozygosity mapping was performed assuming a recessive model of inheritance. First, homozygous loci >1 Mb in the two sisters severely afflicted with RTT (401 and 403) were searched by using program PLINK 1.07. Six loci were detected (Table 5.9). Also, assuming reduced expressivity, homozygous loci >1 Mb shared by all sisters were searched as well. Five such loci were detected after this second analysis (Table 5.9). To find any additional loci that could possibly be candidates to harbor the disease gene, we launched multipoint linkage analysis using GeneHunter, first assuming only RTT sisters affected. Three loci yielded a maximum LOD score of 1.92, and eight other loci yielded LOD scores >1 (Figure 5.22). Haplotypes were constructed using GeneHunter, and haplotype inspection revealed that at those three loci plus six of the eight loci with LOD scores >1, RTT sisters have IBD homozygous genotypes. However, only three of those loci, which were previously detected by PLINK, were >1 Mb (Table 5.9). New calculations were performed assuming all sisters affected. This time, seven loci yielded LOD scores >1 (Figure 5.23). At six of them, all three sisters were found to share homozygous and IBD genotypes by haplotype inspection. One locus was <1 Mb, and thus eliminated. The remaining five loci had been detected previously by PLINK. Overall, linkage analysis did not indicate additional loci >1 Mb and yielding LOD scores >1, assuming either all sisters affected or only the two RTT sisters affected.

At the total 11 candidate loci detected, detailed calculations were performed with the same parameters but including all markers in the homozygosity region plus the 0.5 Mb flanking regions to obtain final LOD scores. The maximal LOD score reached 2.40 at some of the loci (Table 5.9). At the three loci yielding negative scores, the haplotype inspection revealed that the common haplotype of the RTT sisters descended from the grandfathers, who were not kin according to family history. Although those haplotypes could possibly be common haplotypes in the inbred village and thus are less likely to be candidates for disease locus, we included them in the further analysis.

Exome sequencing results indicated that only 22q11.21 contained any rare, deleterious exonic variant (Table 5.9). Haplotypes at 22p11.1-22q11.21 are presented in Figure 5.24. RTT sisters are homozygous between markers rs1041770 and rs140281 (14,603,201 bp and 22,663,865 bp). A crossover in the unaffected sister narrowed down the large disease haplotype, as she is homozygous between rs474117 (21,019,273 bp) and rs140281 (22,663,865 bp). 22q11.21 is one of the loci yielding the highest LOD score of 2.40 and the only locus with a rare, possibly deleterious variant. Moreover, all the sisters share homozygosity. Considering all evidence together, we thought that most likely it is the disease locus.

5.5.2. Exome Sequencing and Evaluation of Variants

DNA sample of patient 401 was subjected to exome sequencing in YCGA (USA) using Roche NimbleGen Exome Array. Alignment, SNV calling and variant annotation were performed in our laboratory using BWA, SAMTools and ANNOVAR software. The variants at the candidate loci were selected, and those that were missense, truncating, or splicing were filtered. Variants with alternative depth <40% of total depth were filtered out, and those that were novel or rare (frequency <0.01) according to dbSNP139 database were selected. Variants that were annotated as novel were further checked for novelty in the EVS database and in the results of 25 other in-laboratory exome samples.

At the candidate loci, a total of 66 possibly deleterious variants were detected. All but one variant was filtered out: 55 variants had reported frequencies >0.01, nine variants had depths <40% of total depth (three of them had depths of one), and one variant was a false call as was shown by Sanger sequencing. The remaining single candidate variant was *P2RX6* c.990dupG (p.K331EfsX26) residing at the candidate locus 22q11.21, where all sisters shared homozygosity (Table 5.9).

Coverage of exons was computed using BEDTools. Known RTT genes *MECP2* and *FOXG1* were found not to be sufficiently covered. Unread coding sequences of those genes were analyzed by Sanger sequencing. Although they were not located in a homozygosity region shared at least by RTT sisters, we wanted to rule out parental germinal mosaicism for a defect in those genes. No mutation was detected in *MECP2* or *FOXG1*.

All sisters and the parents were tested for the identified variant *P2RX6* c.990dupG (p.K331EfsX26). The mother and father were heterozygous whereas all sisters were homozygous for it. Via HRM assay, 145 samples from the population were screened to investigate whether *P2RX6* c.990dupG is a common variant in the Turkish population. None of the tested samples carried the mutation. *P2RX6* c.990dupG was reported only in NHLBI EVS, in one subject out of total 6258 and in the heterozygous state.

According to these results, the rare variant *P2RX6* c.990dupG remained as the sole candidate variant detected at the candidate loci; all sisters are homozygous for it, suggesting that it could underlie RTT in the family.

5.5.3. Assessment of *P2RX6* Transcript Levels in Various Tissues

The abundance of *P2RX6* transcripts was investigated in 11 neural and five non-neural tissues. Variable amounts of transcripts were found in all neural tissues tested whereas the highest expression was in blood and skeletal muscle (Figure 5.25).

5.5.4. Deletion Analysis

SNP genotype data were analyzed on Excel to search for deletions shared by at least the two RTT sisters. A deletion was inferred at 15q21.1, located minimally 127,307 bp downstream to gene *UNC13C*. It is reported in DGV with a frequency of 0.0005 (Shaikh *et al*., 2009). Maximal boundaries of the deletion were rs2062640 and rs992348 (55,029,974 bp and 55,099,758 bp) and minimal boundaries were rs7176868 and rs1813846 (55,048,113 bp and 55,095,241 bp). This 47,128 to 69,784 bp deletion was validated by PCR assays. All sisters were homozygous for it whereas the other members of the family were not. The sequences in the maximal region were investigated using ENCODE tracks on UCSC Genome Browser. No expressed sequences or predicted gene regulatory elements were found within the region, suggesting that this rare deletion does not underlie RTT pathology.

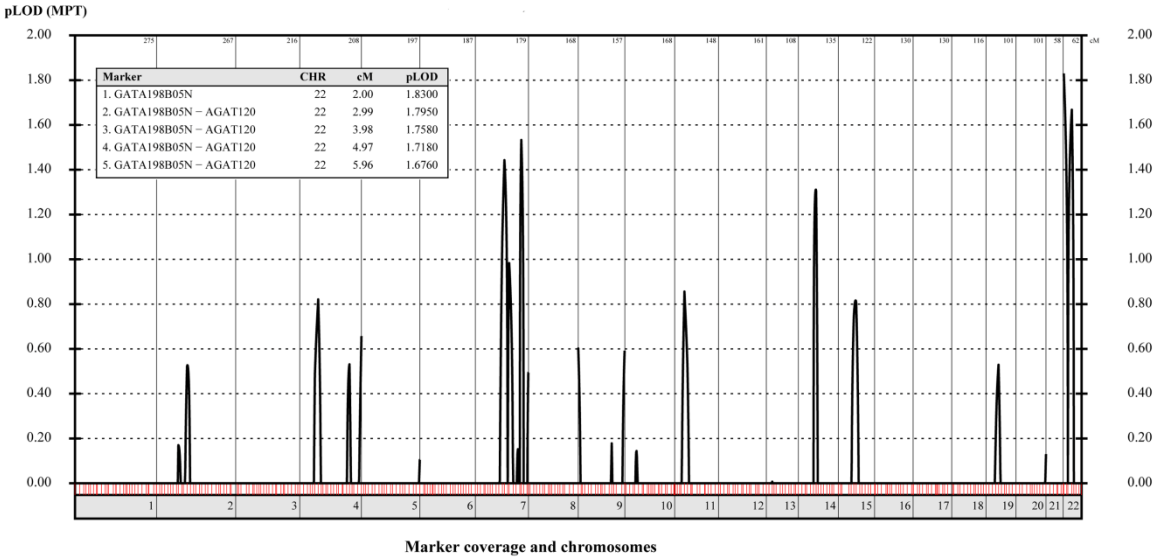

Figure 5.20. Whole-genome multipoint LOD scores for RTT family calculated with microsatellite genotype data. LOD scores are plotted in the order of chromosomal position.

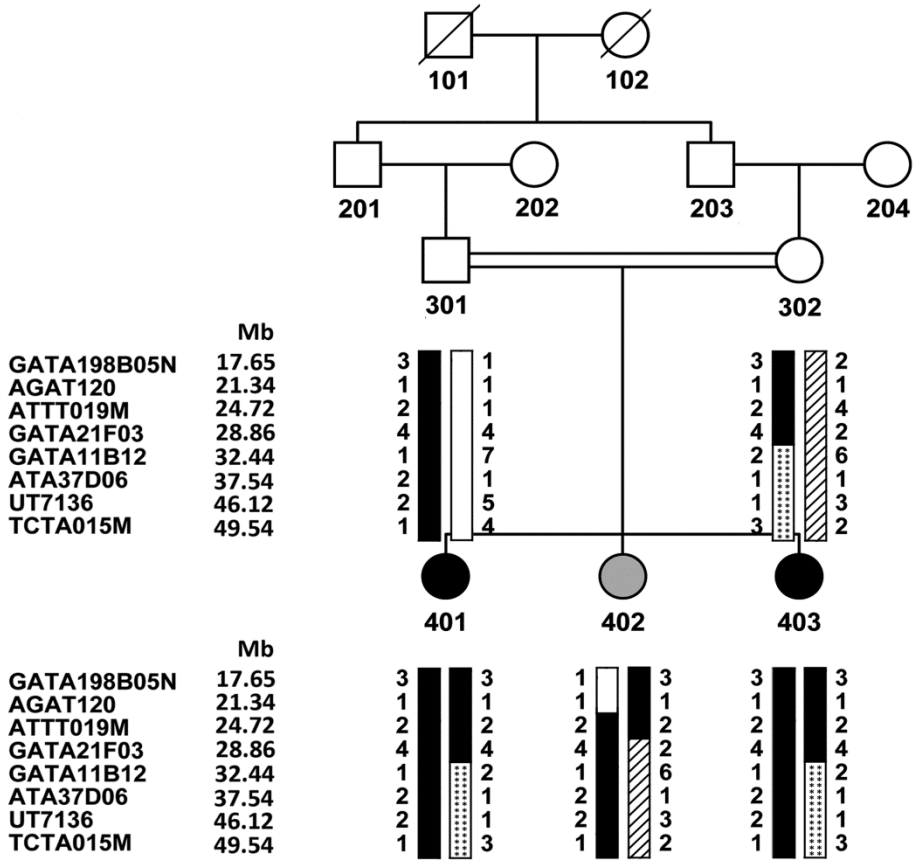

Figure 5.21. Haplotypes of RTT family for chromosome 22 constructed using microsatellite genotypes.

Table 5.9. Regions with homozygous genotypes assessed as IBD in the indicated sisters and the candidate variants harbored.

Locus	**LOD Score**	**Minimal region (location in bp)**	**Size (Mb)**	**Candidate Variants**
IBD in the Two RTT Sisters				
22p11.1-22q11.21	1.92	ter* - rs520698 (0 - 21,024,880)	21.024	-
3q25.2-26.1	1.92	rs12765 – rs2367209 (154,900,690 - 160,398,885)	5.498	-
11p15.1-p14.3	1.28	rs1546329 – rs1915061 (19,137,947 - 24,130,487)	4.994	-
12q24.13-24.21	-1.08	rs3177979 – rs17824050 (111,841,592 - 113,036,817)	1.195	-
4p12	-1.09	rs12508460 – rs3747690 (47,816,387 - 48,988,450)	1.172	-
12q21.31	-2.81	rs17740709 - rs7307305 (81,947,471 - 82,964,479)	1.017	-
IBD in All Three Sisters				
15q21.1-22.31	2.40	rs1897028 - rs304985 (54,356,060 - 70,138,158)	15.790	-
11p14.3-p13	1.50	rs1915061 – rs1883959 (24,130,487 - 34,234,145)	10.104	-
22q11.21	2.40	rs520698 - rs6001563 (21,024,880 - 22,599,764)	1.575	*P2RX6* c.990dupG
3q25.2	2.40	rs6767565 – rs12765 (153,259,376 - 154,900,690)	1.641	-
13q33.3	2.40	rs9555491 - rs6492218 (109,320,091 - 110,325,749)	1.006	-

*The first SNP on the chip for chromosome 22 is rs1041770 at 14.60 Mb.

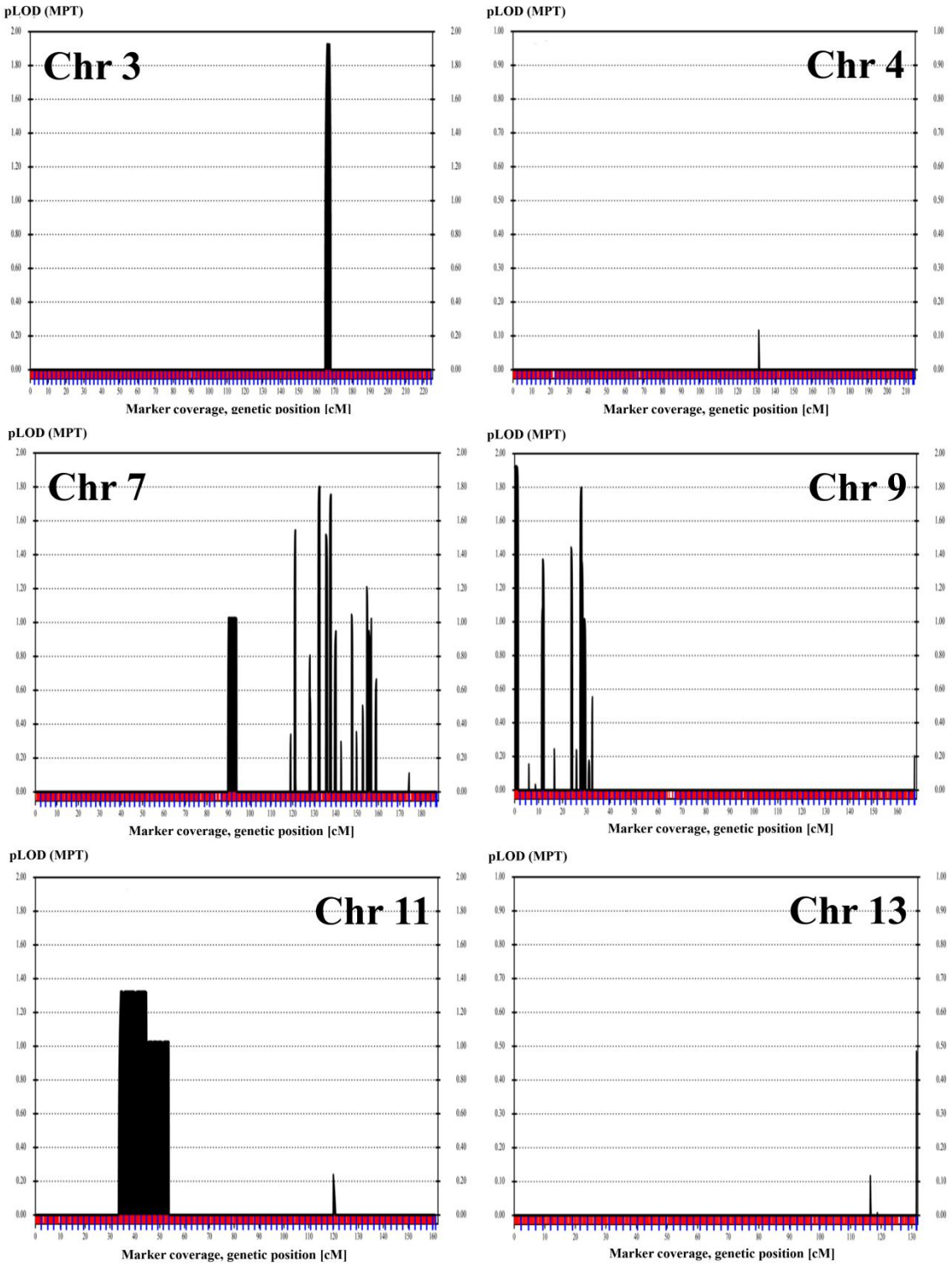

Figure 5.22. Whole-genome multipoint LOD scores for RTT family, calculated by assuming only the two RTT sisters as affected. Only the chromosomes with scores >0 are presented.

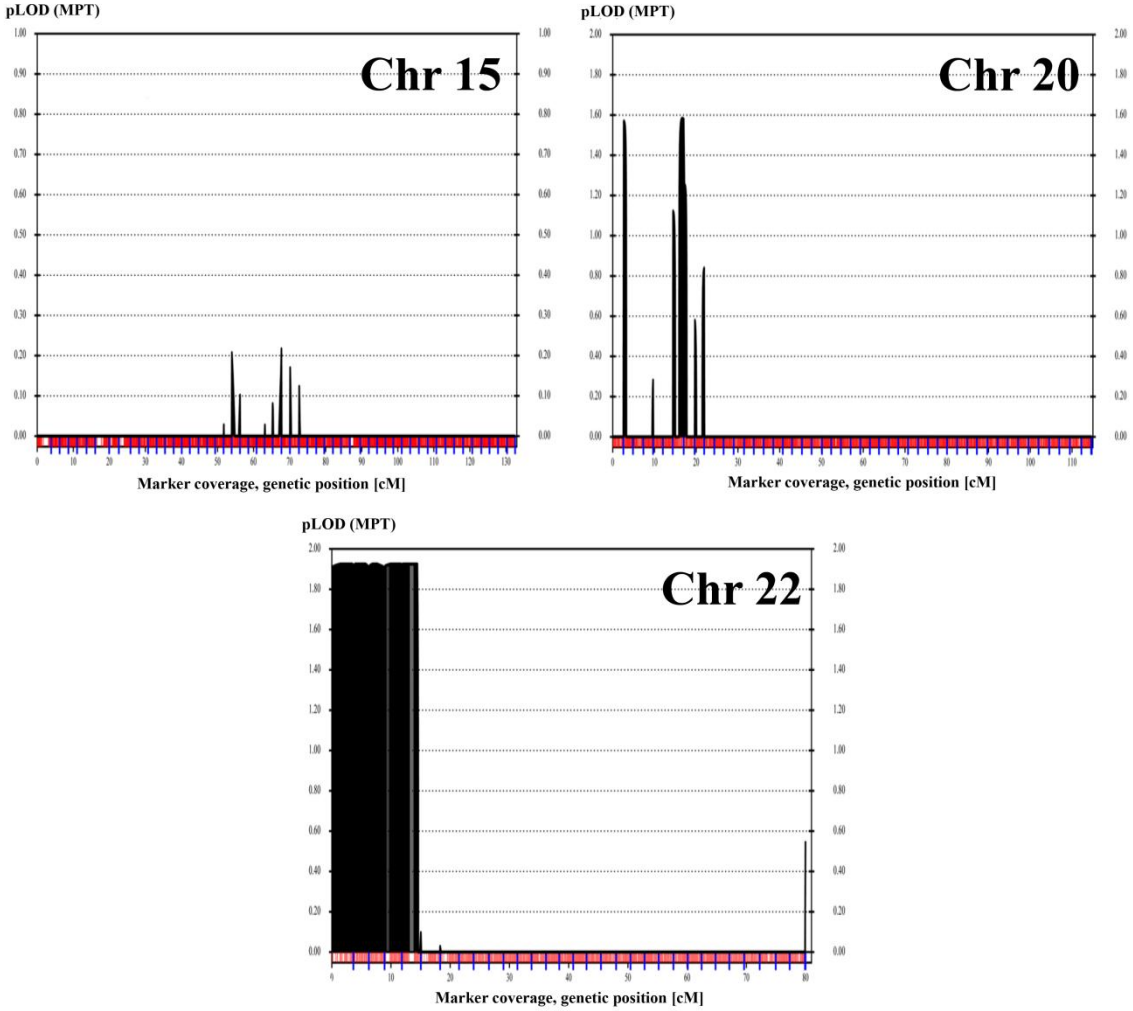

Figure 5.22. Whole-genome multipoint LOD scores for RTT family, calculated by assuming only the two RTT sisters as affected. Only the chromosomes with scores >0 are presented (cont.).

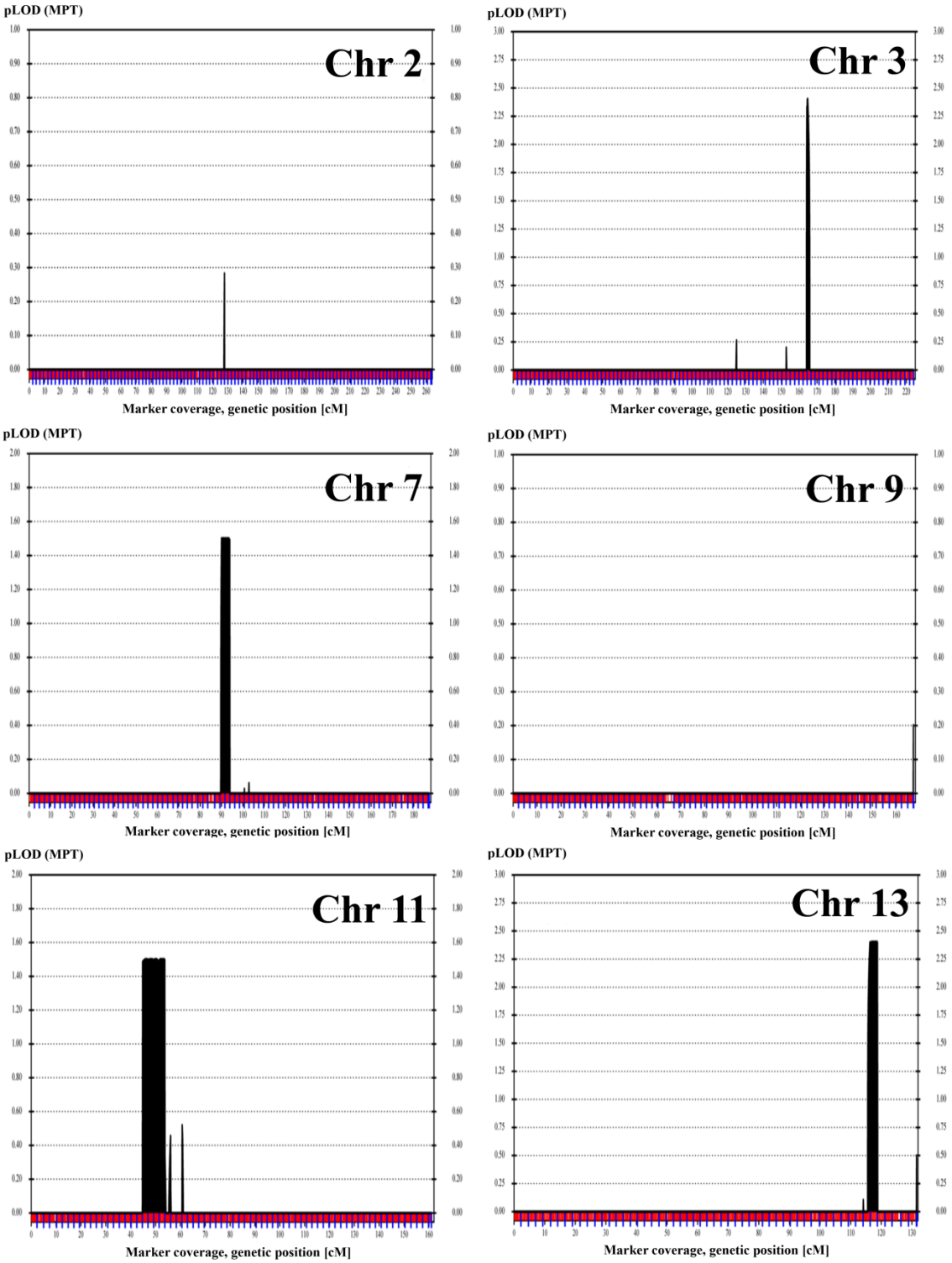

Figure 5.23. Whole-genome multipoint LOD scores for RTT family, calculated by assuming all three sisters as affected. Only the chromosomes with scores >0 are presented.

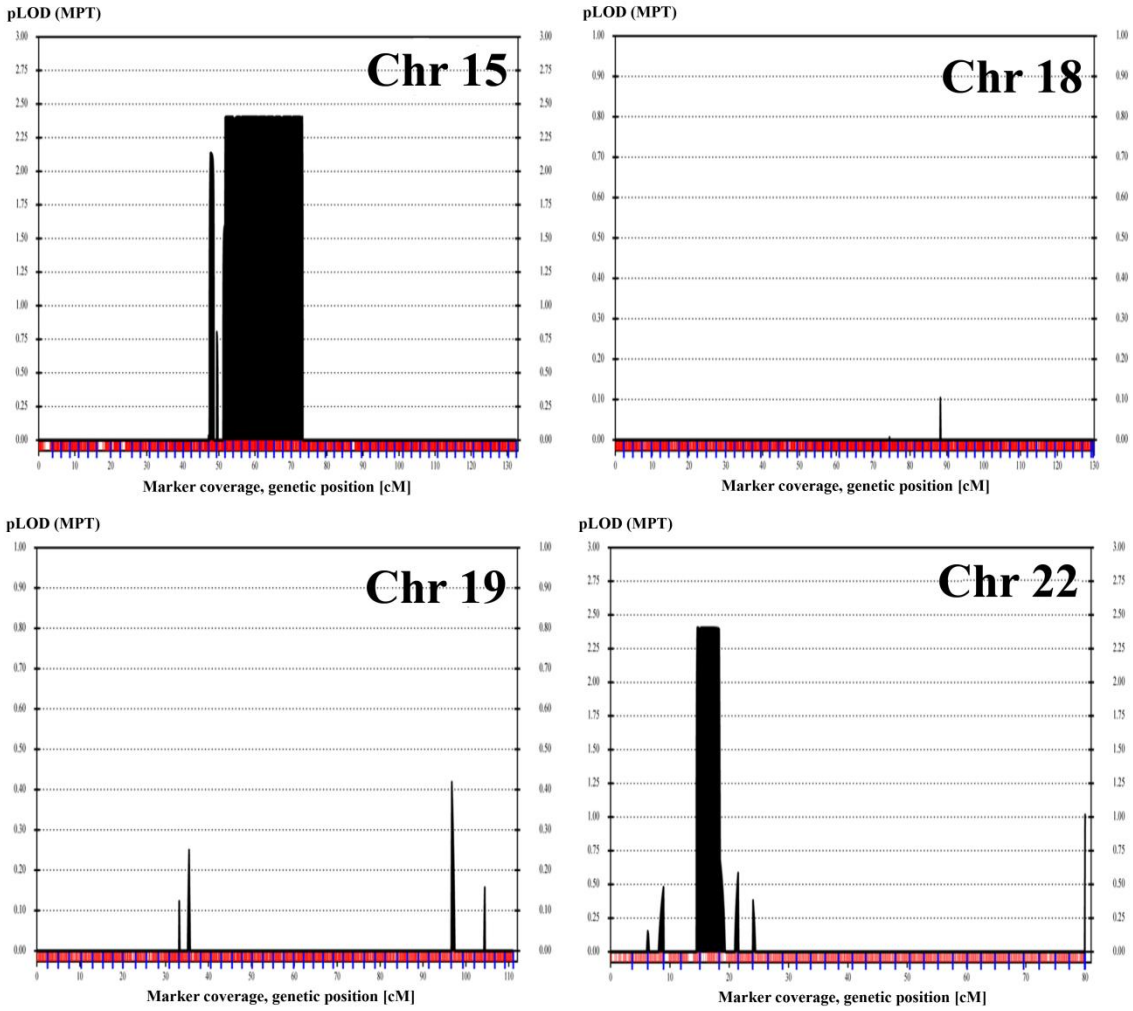

Figure 5.23. Whole-genome multipoint LOD scores for RTT family, calculated by assuming all three sisters as affected. Only the chromosomes with scores >0 are presented (cont.).

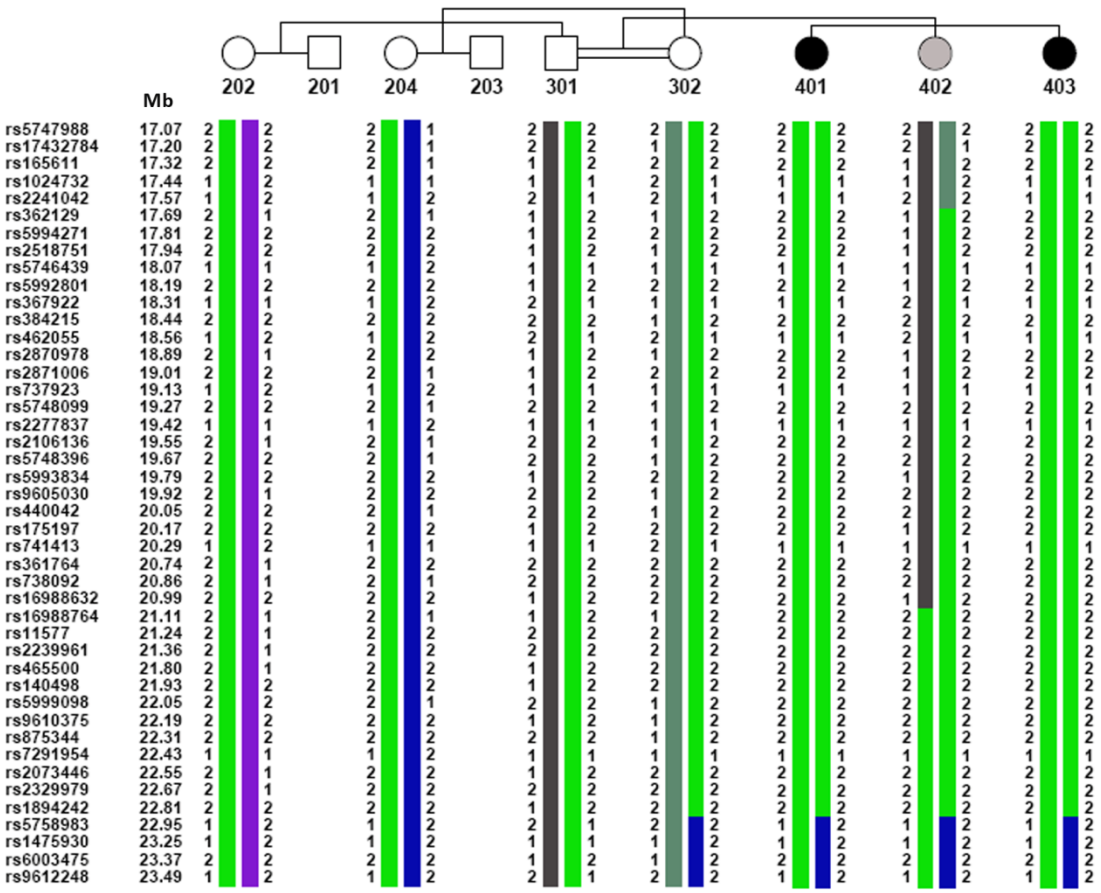

Figure 5.24. Haplotypes for selected SNP markers at 22p11.1-q11.21 for RTT family. IBD disease haplotype is shown in green.

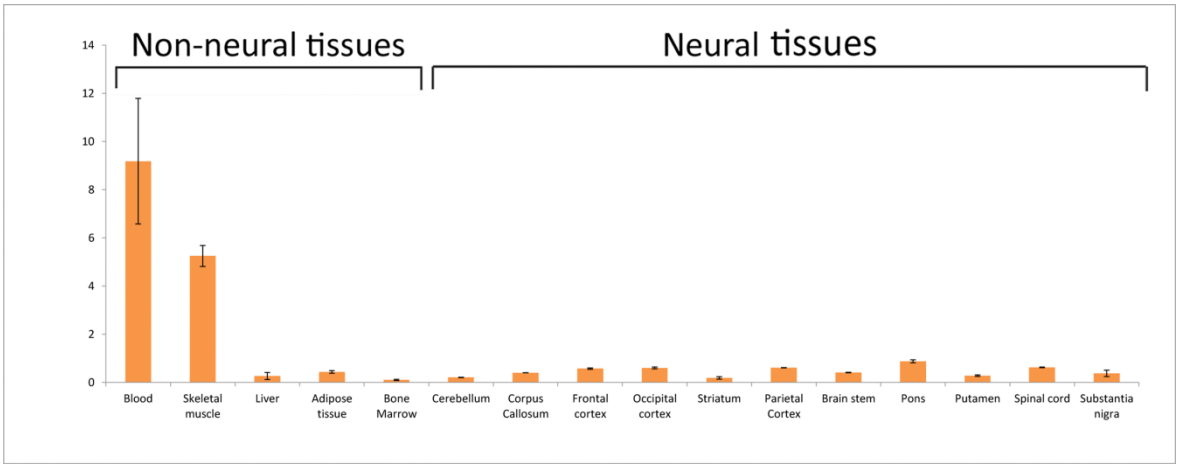

Figure 5.25. Quantification of *P2RX6* transcripts relative to *TUBB* transcripts in 16 different tissues.

5.6. Infantile Neuroaxonal Dystrophy (INAD)

5.6.1. Investigation of Linkage to *PLA2G6* Locus and Analysis of *PLA2G6* Gene

Linkage to *PLA2G6* locus (22q13.1), the only known locus for INAD, was investigated in families 1, 2, 3 and 5 using three microsatellite markers. Co-segregation of marker alleles with the disease was observed in all families tested, except in Family 5. The affected sibs in this family were heterozygous for each marker genotyped, excluding linkage. In families 2 and 3, there is just one affected child, homozygous for the markers employed, and there was no unaffected sib. Therefore, linkage to 22q13.1 was inconclusive in these families. We analyzed all coding regions of *PLA2G6* in all six families, including Family 5 via Sanger sequencing, and mutations in four families were identified (Figure 5.26). All mutations were novel and in the homozygous state in the patients. Two missense mutations were found; T661M in exon 11 in Family 1 and D484G in exon 14 in Family 2. T661M was predicted as damaging by both PolyPhen-2 and SIFT, the latter with high confidence (PolyPhen score: 1; SIFT score: 0), and disease causing by MutationTaster. PolyPhen predicted D484G as benign (score 0.4), SIFT as damaging with low confidence (score 0.04), and MutationTaster as disease causing. The other two mutations were truncating. The identified mutations are given in Table 5.10. No mutation in *PLA2G6* was detected in Family 5 and 6.

Table 5.10. *PLA2G6* mutations in INAD families 1-4.

Family	Gene	Exon	Nature	Effect on cDNA	Effect on protein
1	*PLA2G6*	Exon 11	Missense	c.1451A>G	p.D484G
2	*PLA2G6*	Exon 3	Frameshift	c.402delC	p.F135SfsX79
3	*PLA2G6*	Exon 14	Missense	c.1982C>T	p.T661M
4	*PLA2G6*	Exon 11	Frameshift	c.1545_1546insGC	p.G517RfsX29

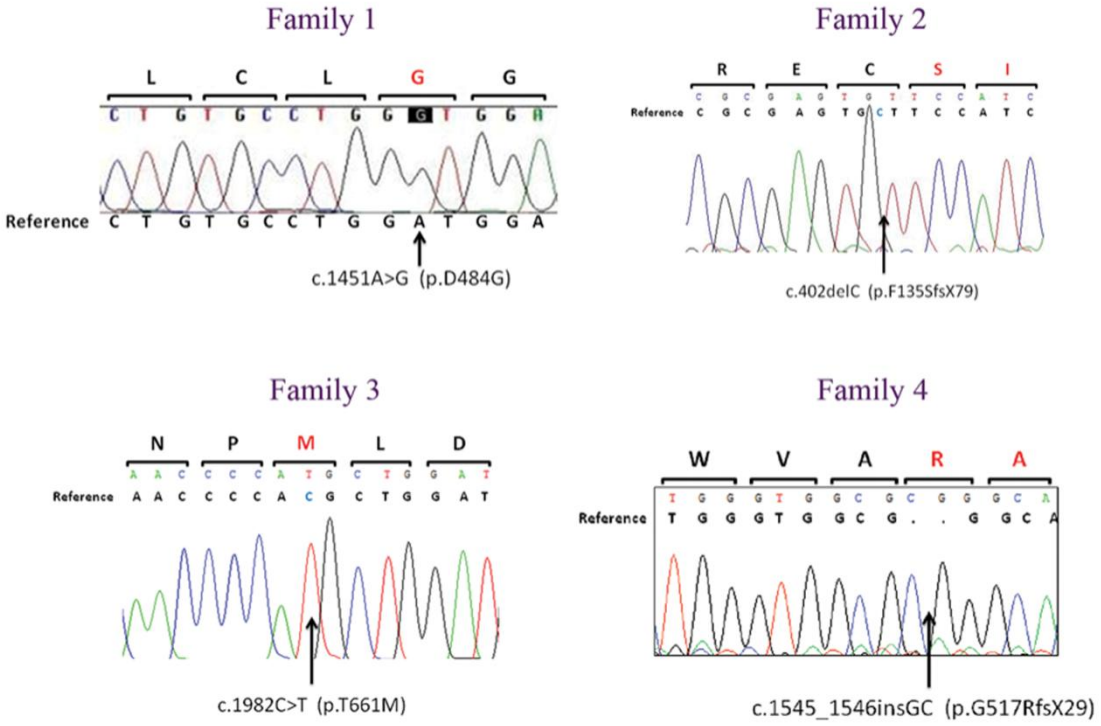

Figure 5.26. Chromatograms showing *PLA2G6* mutations in INAD families 1-4 (Koroglu *et al*, 2013). Reference sequences are given below the read sequences.

5.6.2. Linkage Analysis, CNV Analysis, and Haplotype Inspection for Family 5

Family 5 was subjected to multipoint linkage analysis using genotyping data generated by 610,000 SNP markers. Markers were used at 0.07-cM spacing. Twelve loci yielded LOD scores >1. These are 1p34.1, 4p15.33, 8q24.23, 11p15.2, 14q21.1, 15q25.3, 16q24.1, 20p12.3, 9p24.1, 13q32.3-33.2, 18q12.2, and 20p12.1 (Figure 5.27). The latter four loci yielded the highest score of 1.92. Through comparison of genotypes on Excel, four of the loci were selected for further investigation, 4p15.33, 9p24.1, 13q32.3-33.2 and 18q12.2. At those regions, the affected sibs shared homozygosity whereas the unaffected sib was heterozygous, indicating that the loci could possibly harbor a disease gene. The affected sibs were not homozygous at the other eight loci, hence those loci were excluded.

Homozygosity investigation was performed using LOH Detector plug-in on Illumina GenomeStudio 2011.1 software. The plug-in detects loss of heterozygosity (LOH) regions by assessing B-allele frequencies, which is a measure of the allelic intensity ratios of

marker alleles. A value of 1 or 0 indicates AA or BB genotype that is the complete absence of A or B allele, and 0.5 indicates AB genotype that is the equal presence of both alleles. Regions homozygous for >50 SNP markers in both affected sibs were selected. In addition to the abovementioned four loci (4p15.33, 9p24.1, 13q32.3-33.2, and 18q12.2), five other loci with LOD scores >1 were detected (Figure 5.28 and Table 5.11). To assess the final scores at those total nine loci, genotypes were analyzed using Excel, and new LOD score calculations were performed with the same parameters but using all markers. Haplotype segregation was investigated via HaploPainter. All except one locus (1q32.1) yielded a maximal LOD score of 1.92. Locus 1q32.1 was excluded, because haplotypes of affected sibs did not support identity by descent. The remaining eight loci were considered candidates to harbor the disease gene (Table 5.11).

Finally, we performed CNV analysis at the eight candidate loci via cnvPartition plug-in on GenomeStudio and detected at 6p21.31 non-identical CNV genotypes in patients. After exclusion of that locus, seven loci remained, one of which was possibly the gene locus. Those loci are presented in Table 5.11.

The largest of the candidate loci was 13q32.3-33.2. Affected sibs shared homozygosity for a 10.7-Mb region between rs1412934 and rs7334038 (94,893,706 bp and 105,630,414 bp). The unaffected sib also shared the homozygous genotype in part, between rs1412934 and rs998733, narrowing down the candidate locus to approximately 6.5 Mb, between markers rs998733 and rs7334038 (99,089,156 bp and 105,630,414 bp). The next largest locus was 9p24.1, encompassing a 4.1-Mb region. Locus 1p34.1-1p33 was <1.5 Mb, and the remaining four loci were even smaller, 0.7 to 0.85 Mb (Table 5.11).

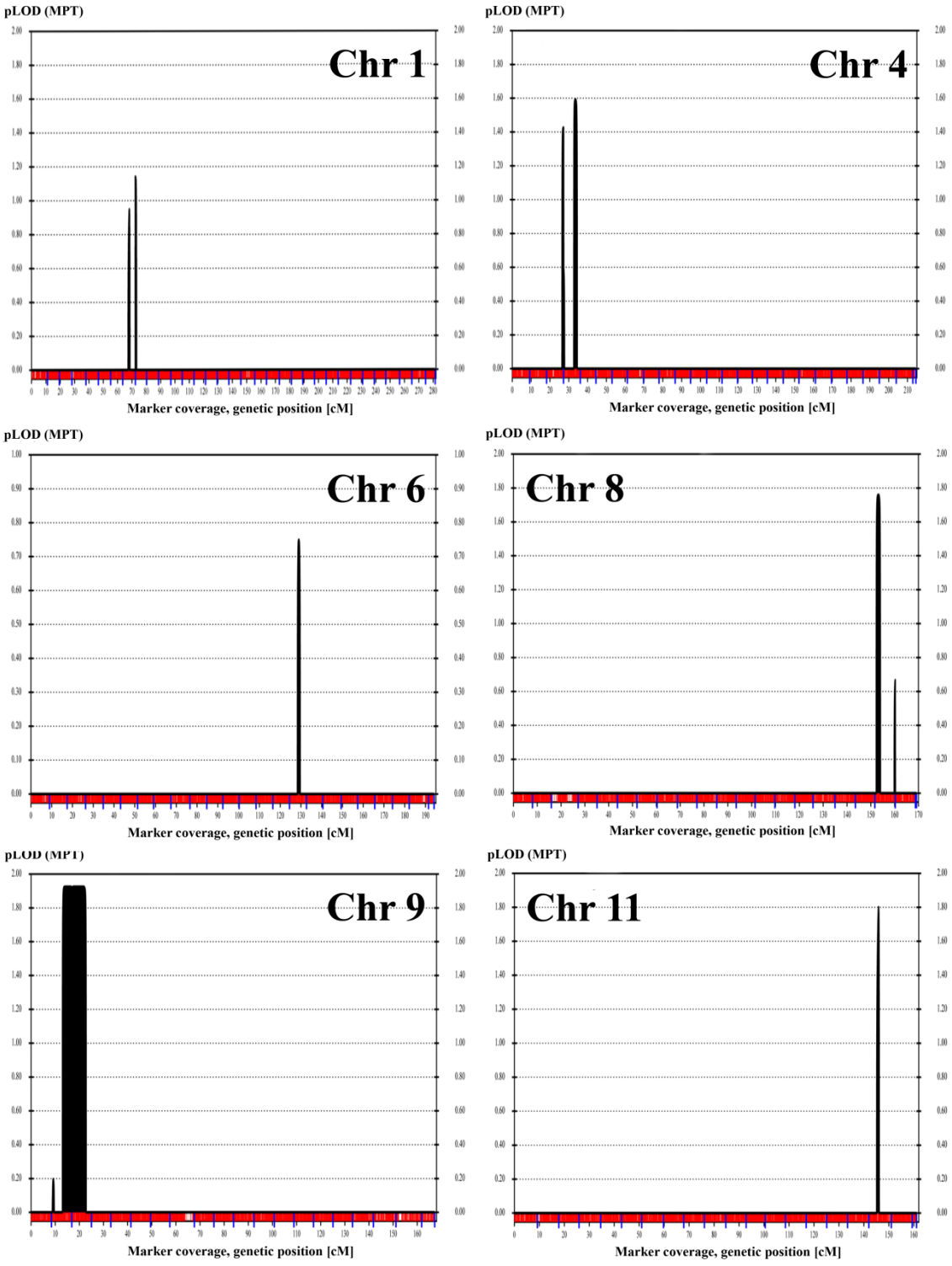

Figure 5.27. Whole-genome multipoint LOD scores for INAD Family 5. Only the chromosomes with scores >0 are presented.

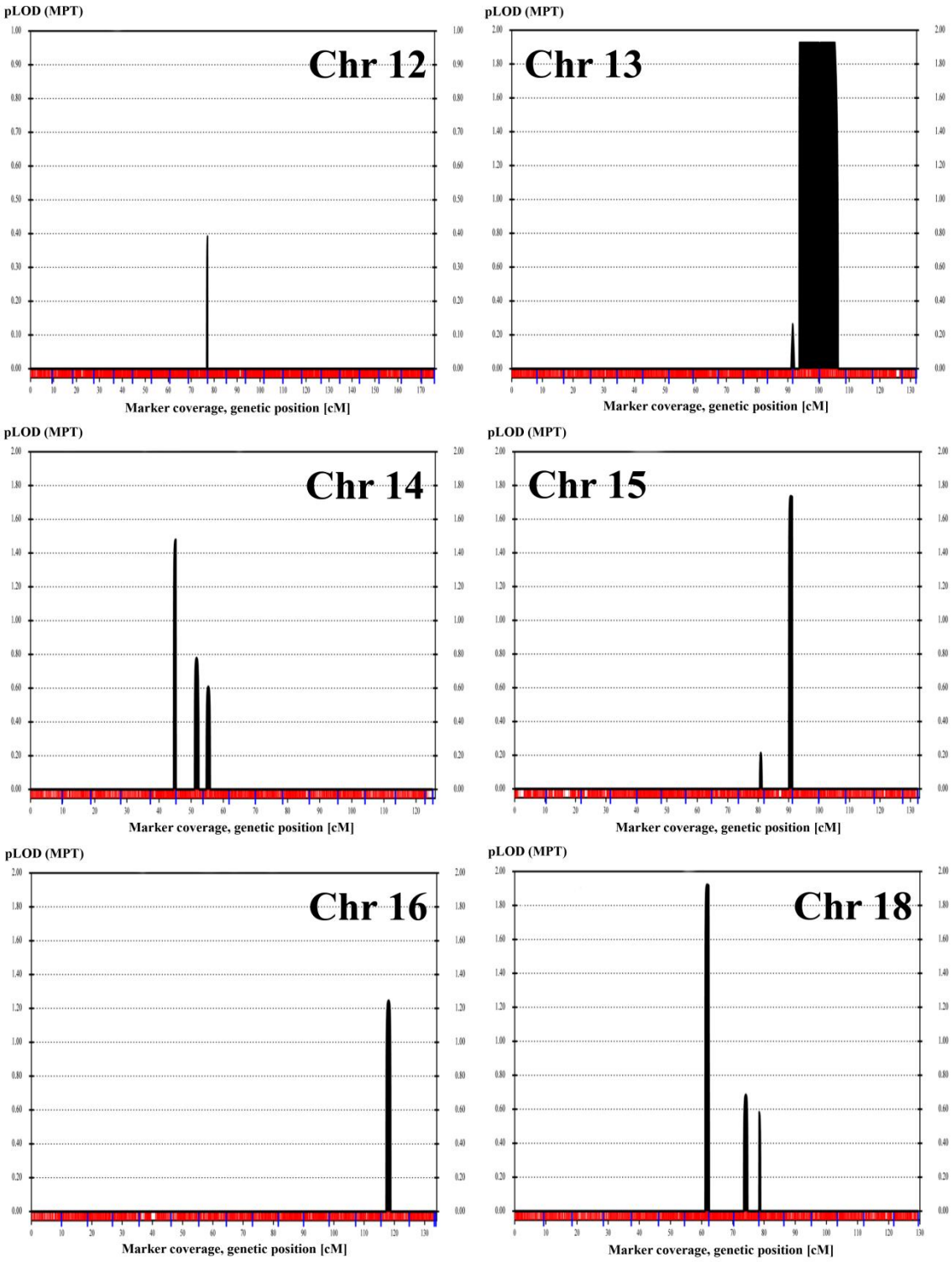

Figure 5.27. Whole-genome multipoint LOD scores for INAD Family 5. Only the chromosomes with scores >0 are presented (cont.).

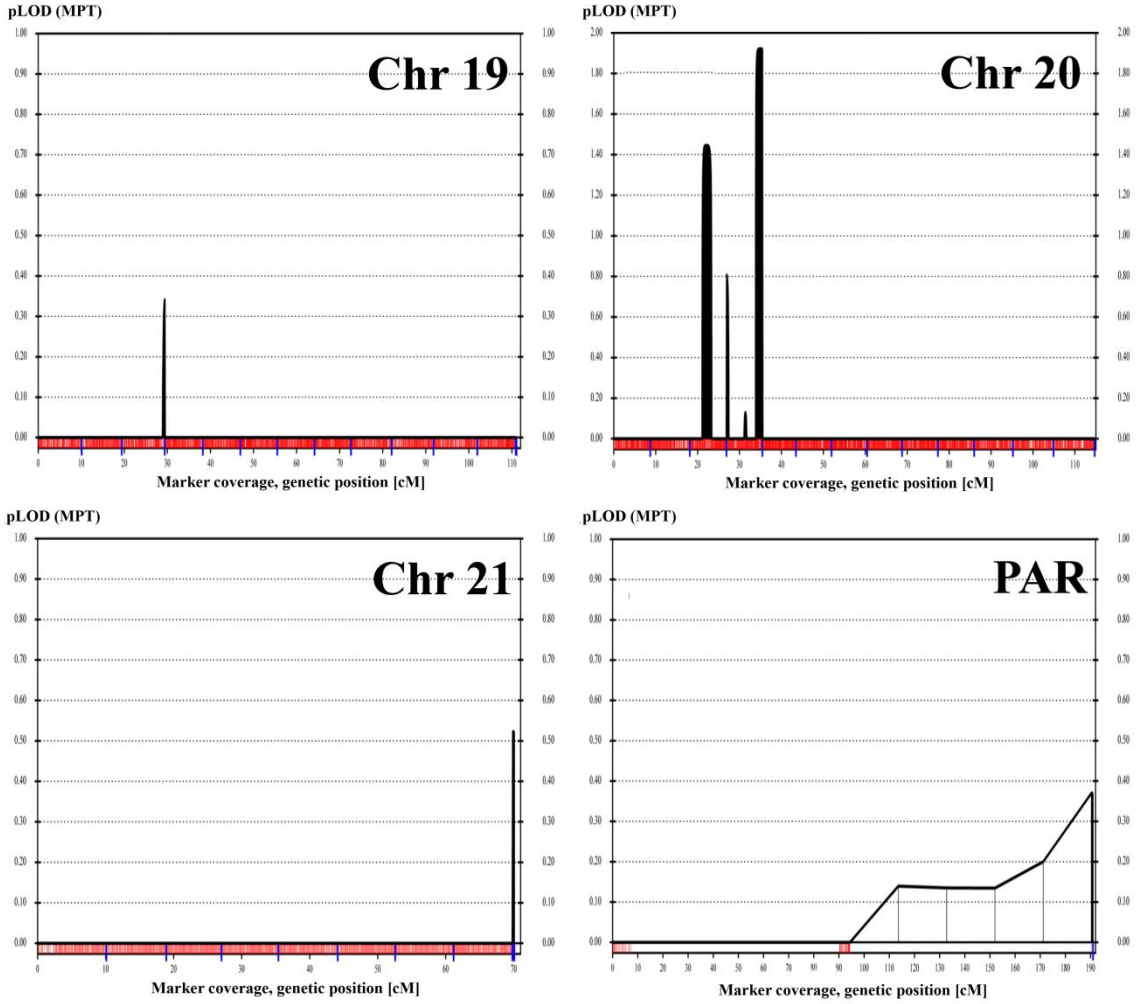

Figure 5.27. Whole-genome multipoint LOD scores for INAD Family 5. Only the chromosomes with scores >0 are presented. PAR denotes pseudoautomal regions (cont.).

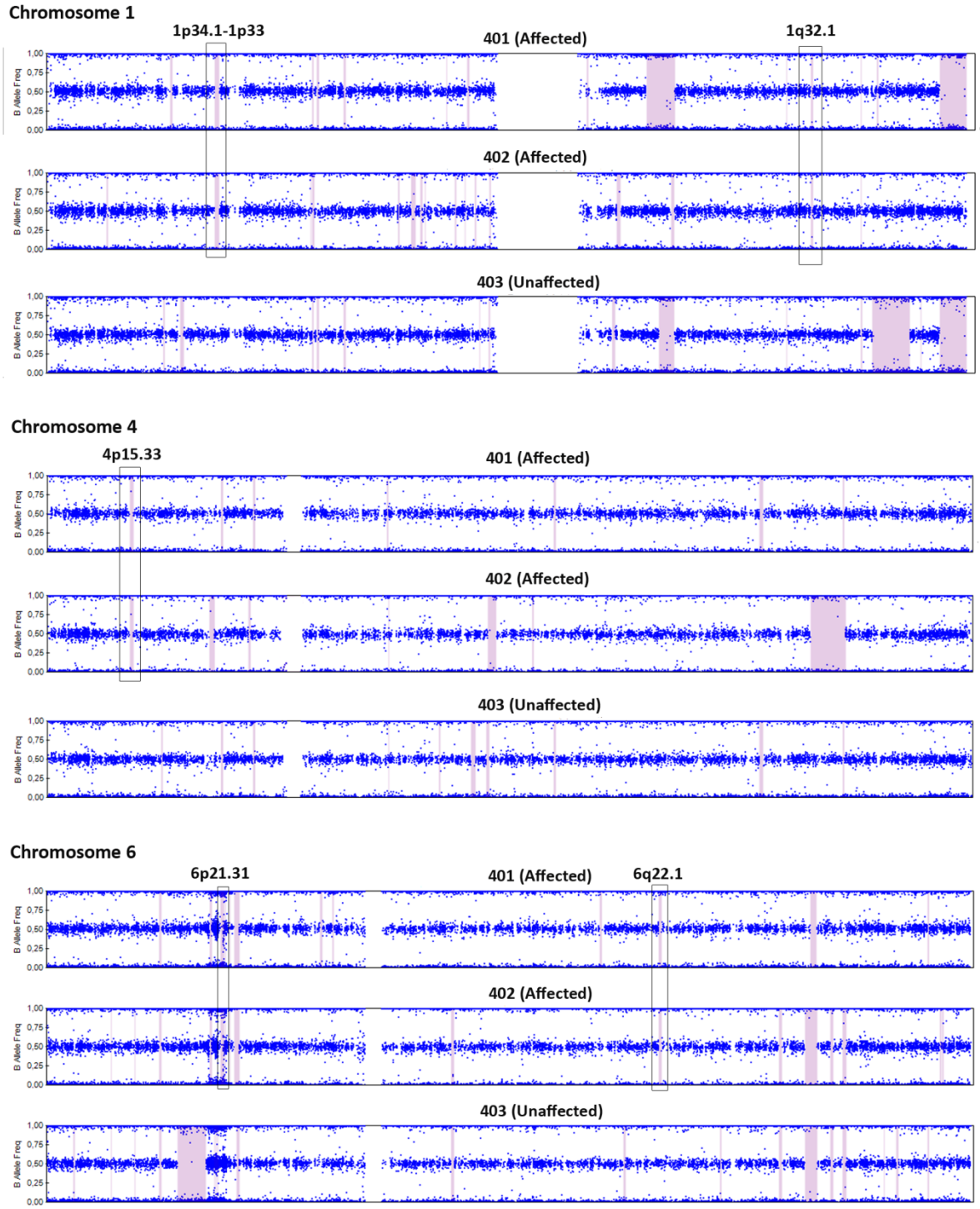

Figure 5.28. Homozygous regions shared by affected sibs only in INAD family 5. For each sib, regions of LOH >50 SNP markers were detected by LOH Detector and highlighted pink.

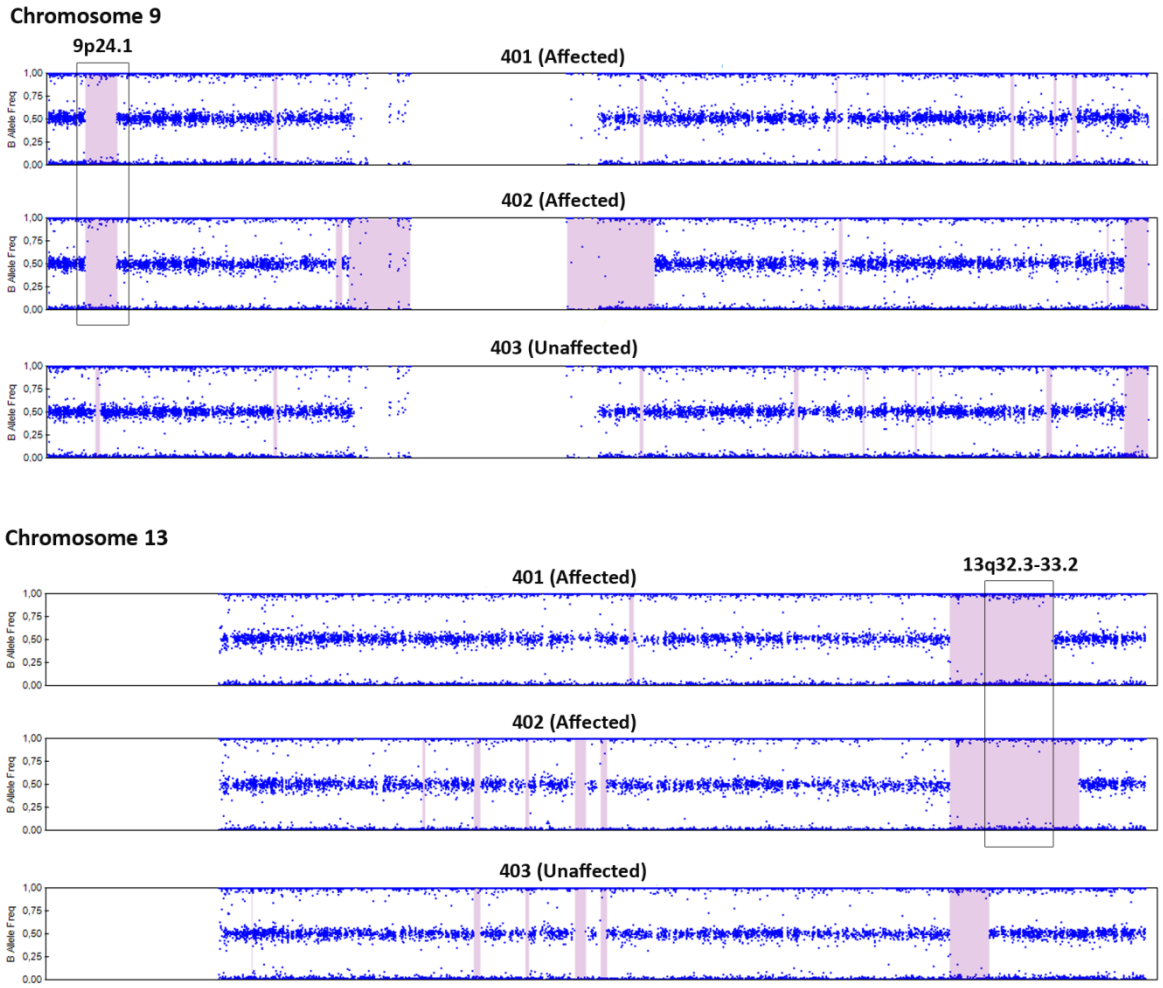

Figure 5.28. Homozygous regions shared by affected sibs only in INAD family 5. For each sib, regions of LOH >50 SNP markers were detected by LOH Detector and highlighted pink (cont.).

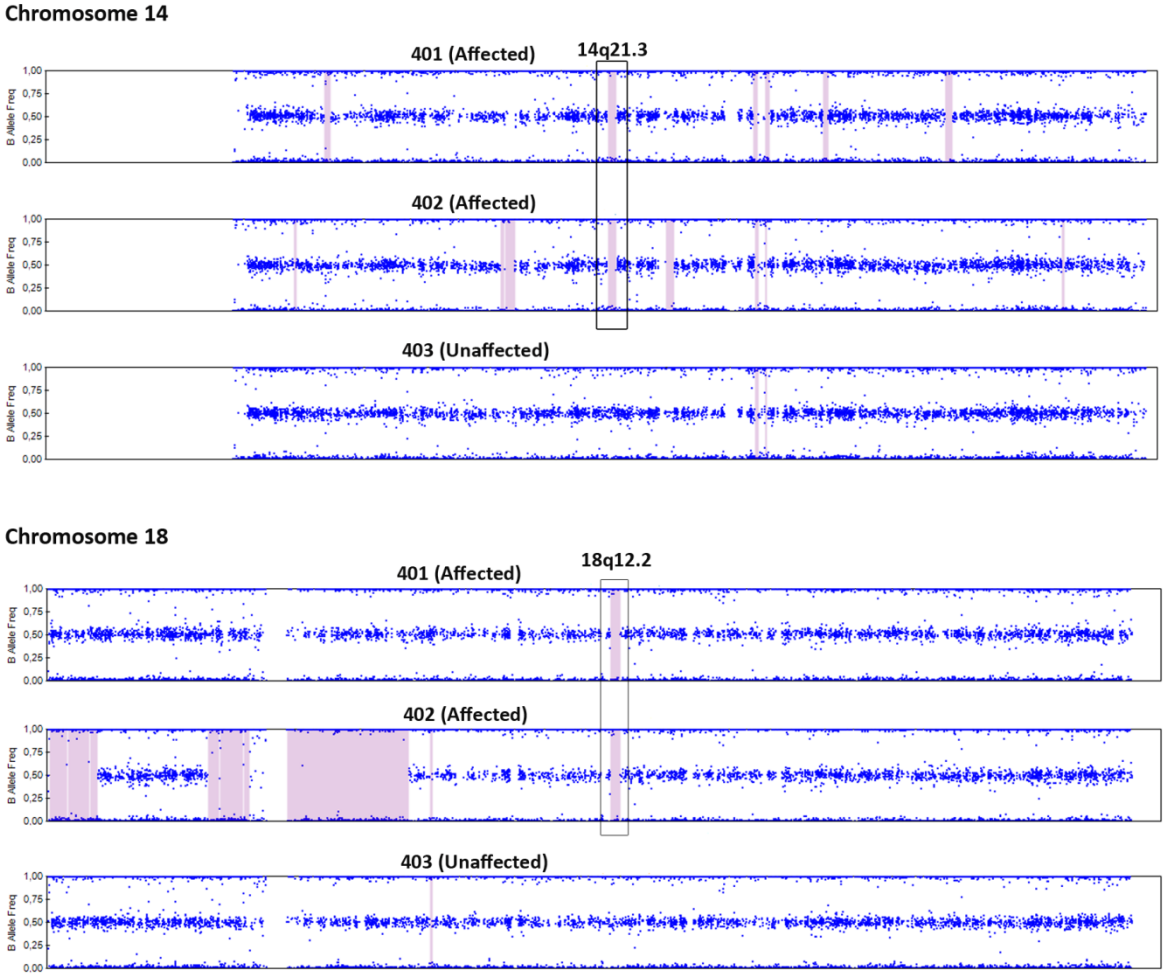

Figure 5.28. Homozygous regions shared by affected sibs only in INAD family 5. For each sib, regions of LOH >50 SNP markers were detected by LOH Detector and highlighted pink (cont.).

Table 5.11. Loci where INAD sibs only are homozygous for >50 SNP markers. Flanking markers and their physical positions, locus size, LOD scores and the conclusion are given for each locus.

Locus	Maximum region of shared homozygosity (markers/physical position as base-pair)	Size (Kb)	LOD Score (markers with 0.07cM spacing)	LOD Score (markers with no spacing)	Conclusion
1p34.1-1p33	rs12022162 - rs785475 (44,934,432 - 46,382,323)	1,448	-0.98	1.92	Candidate locus
1q32.1	rs1126618 - rs2761424 (205,364,303 - 205,960,405)	596	-1.07	-1.07	Excluded by haplotype analysis
4p15.33	rs4698560 - rs1503872 (16,900,494 - 17,740,343)	840	1.59	1.92	Candidate locus
6p21.31	rs28986321 - rs17500468 (32,278,958 - 32,819,156)	540	-0.17	1.92	Excluded by CNV analysis

Table 5.11. Loci where INAD sibs only are homozygous for >50 SNP markers. Flanking markers and their physical positions, locus size, LOD scores and the conclusion are given for each locus (cont.).

Locus	Maximum region of shared homozygosity (markers/physical position as base-pair)	Size (Kb)	LOD Score (markers with 0.07cM spacing)	LOD Score (markers with no spacing)	Conclusion
6q22.1	rs4945909 - rs9320442 (113,105,769 - 113,793,354)	688	-1.07	1.92	Candidate locus
9p24.1	rs4740801 - rs2196066 (4,790,165 - 8,894,910)	4,105	1.92	1.92	Candidate locus
13q32.3-33.2	rs1412934 - rs7334038 (94,893,706 - 105,630,414)	10,737	1.92	1.92	Candidate locus
14q21.3	rs11628843 - rs6573040 (54,250,093 - 55,097,568)	847	0.78	1.92	Candidate locus
18q12.2	rs10502817 - rs1456605 (39,503,140 - 40,221,499)	718	1.92	1.92	Candidate locus

5.6.3. Exome Sequencing and Evaluation of Variants for Family 5

Exome sequencing was performed at Macrogen Inc on a DNA sample from patient 402. Capture probes targeted a total of 62 Mb genomic regions aiming to cover 95% of the CCDS database. The sequencing statistics are given in Table 5.12.

Table 5.12. Exome sequencing statistics for INAD patient 402.

Target regions (bp)	62,085,286
% Coverage of target regions (more than 1X)	93%
% Coverage of target regions (more than 10X)	79.7%
Median read depth of target regions	39X
Mean read depth of target regions	46.9X

The total seven candidate loci harbored a total of 80 exonic and splicing variants, and 74 of them either had MAF >0.01 or were found in other 25 in-laboratory exome sequencing samples. Those variants were excluded. The remaining six variants are presented in Table 5.13. All except *NALCN* c.1924C>T had depths <40% of total depth. *NALCN* c.1924C>T (p.Q642X) was the only homozygous variant, and it was considered as the only candidate mutation. It was validated by Sanger sequencing (Figure 5.29), which also showed that the mutation is segregated with the disease in the family. Population

screening with 110 subjects was performed via HRM assay. None of the tested individuals carried the mutation, showing that the variant is not common in Turkish population.

Besides investigating the variants at candidate loci, we also analyzed the total exome sequence results for novel, homozygous and possibly damaging variants. Of the total 20 such variants (Table 5.14), only *NALCN* p.Q642X was located in a region of homozygosity shared by the two affected sibs. This analysis verified that the *NALCN* variant is the best candidate for the mutation underlying INAD in Family 5.

Table 5.13. Novel and rare variants detected at the candidate loci for INAD.

Chr	Position (bp)	Ref. base	Obs. base	Call	Total depth	Depth of obs. base	Variant and change	SNP ID and frequency
1	45469395	G	A	Het	8	3	*HECTD3* NM_024602 c.C2447T, nonsyn.	rs200632297, 0.0005
13	95860159-95860160	GT	-	Het	23	1	*ABCC4* NM_001105515 c.805_806del, frameshift deletion	Not in databases
13	96589189	C	G	Het	18	3	*UGGT2* NM_020121 c.G1966C, nonsyn.	Not in databases
13	100190097	-	G	Het	32	1	*TM9SF2* NM_004800 c.696_697insG, frameshift deletion	Not in databases
13	101721053	G	T	Het	27	4	*NALCN* NM_052867 c.C4324A, nonsyn.	Not in databases
13	101797163	G	A	Hom	40	40	*NALCN* NM_052867 c.C1924T, stopgain	Not in databases

Ref: reference
Obs: observed
Nonsyn: Nonsynonymous SNV

Table 5.14. Novel and possibly damaging variants in total exome sequence results of INAD sample. Only those called as homozygous are included.

Chr	Position (bp)	Ref. base	Obs. base	Total depth	Depth of obs. base	Variant	Change
2	32463365	C	A	21	21	*NLRC4* NM_02120 c.G2357T	Nonsynonymous
6	42231239-42231244	GGCGGA	-	14	8	*TRERF1* NM_033502 c.1698_1703del	Nonframeshift deletion
7	12409801	C	-	36	33	*VWDE* NM_001135924 c.2131delG	Frameshift deletion
7	74225480	T	C	6	6	*GTF2IRD2* NM_173537 c.A755G	Nonsynonymous
9	5185581	A	T	8	8	*INSL6* NM_007179 c.T22A	Nonsynonymous
9	126133060	T	G	6	6	*CRB2* NM_17368 c.T1728G	Nonsynonymous
10	90530701	G	A	53	53	*LIPN* NM_001102469 c.G772A	Nonsynonymous
10	101997844	-	A	10	9	*CWF19L1* NM_018294 c.1189_1190insT	Stopgain
10	114203278	A	T	32	32	*ZDHHC6* NM_022494 c.T354A	Stopgain

Ref: reference
Obs: observed

Table 5.14. Novel and possibly damaging variants in total exome sequence results of INAD sample. Only those called as homozygous are included.

Chr	Position (bp)	Ref. base	Obs. base	Total depth	Depth of obs. base	Variant	Change
10	118969015	-	T	8	8	*KCNK18* NM_181840 c.360_361insT	Frameshift insertion
11	4703314	C	T	13	13	*OR51E2* NM_030774 c.G628A	Nonsynonymous
11	62457948	A	G	5	5	*BSCL2* NM_00113070 c.T1280C	Nonsynonymous
12	101759271	G	A	53	53	*UTP20* NM_014503 c.G5993A	Nonsynonymous
13	101797163	G	A	40	40	*NALCN* NM_052867 c.C1924T	Stopgain
18	214849	T	C	37	37	*THOC1* NM_005131 c.A1751G	Nonsynonymous
19	35434202	A	G	52	52	*ZNF30* NM_001099437 c.A335G	Nonsynonymous
19	48305603	G	A	5	5	*TPRX1* NM_198479 c.C665T	Nonsynonymous
19	50730169	A	G	6	6	*MYH14* NM_001077186 c.A820G	Nonsynonymous
19	50957348	T	A	15	15	*MYBPC2* NM_004533 c.T1821A	Nonsynonymous
19	50958484	C	T	37	37	*MYBPC2* NM_004533 c.C2134T	Nonsynonymous

Ref: reference
Obs: observed

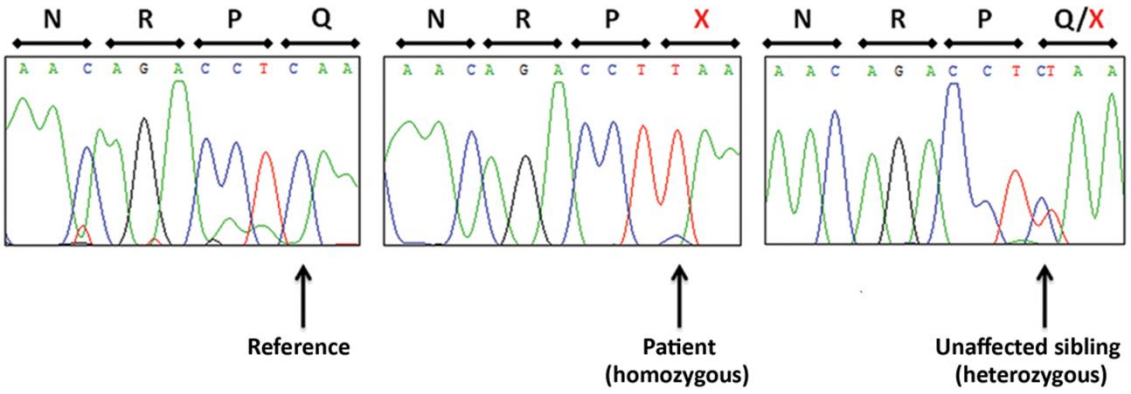

Figure 5.29. Chromatograms showing *NALCN* c.1924C>T (p.Q642X) in patient 402 and unaffected sib 403 (Koroglu et al, 2013).

5.6.4. Assessment of *NALCN* Transcript Levels in Various Tissues

The abundance of *NALCN* transcripts was investigated in 11 neural and five non-neural tissues. Variable amounts were observed, and the highest expression was in occipital lobe, followed by cerebellum and putamen. There was no expression in any of the non-neural tissues, indicating neural specificity for *NALCN* expression (Figure 5.30.).

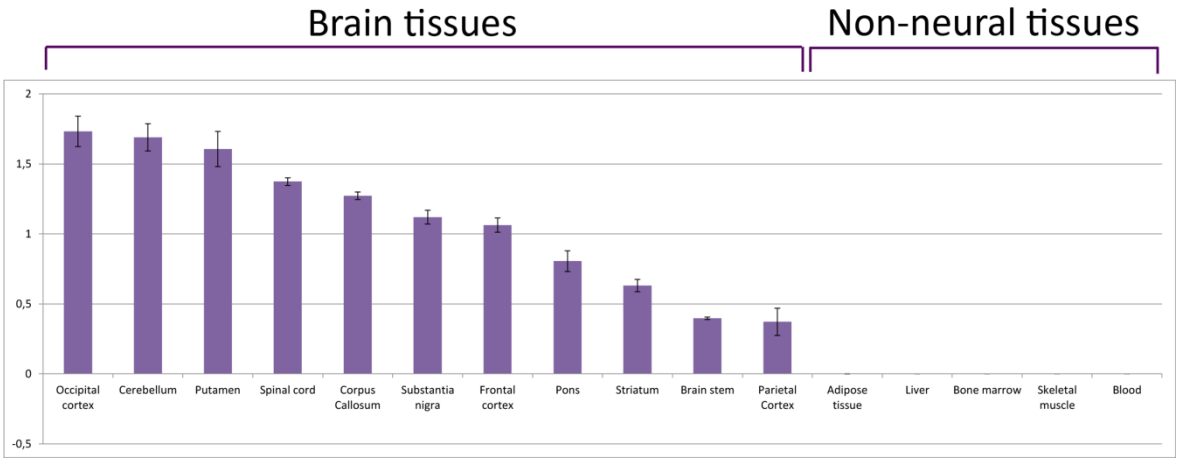

Figure 5.30. Transcript levels for *NALCN* relative to *TUBB* in various tissues (Köroğlu *et al*., 2013).

6. DISCUSSION

6.1. Cone-Rod Dystrophy (CORD)

One of the ocular diseases we studied was an autosomal recessive cone-rod dystrophy in a large consanguineous family. Four members were afflicted with decreased visual acuity, severe photophobia and color vision disturbances which are typical symptoms of CORD. However, another common symptom, nyctalopia was not prominent despite the rod degeneration all patients have, and the disease progression was assessed by the clinician as slow compared to other types of CORD. Via linkage mapping and subsequent analysis by Sanger sequencing of the coding sequences of the six protein-coding genes at the candidate locus, a novel homozygous missense mutation was identified in *POC1B* (MIM#614784). Our manuscript reporting a new type of CORD and a novel CORD gene, *POC1B,* was published online in journal JAMA Ophthalmology.

We localized the disease gene to a >2-Mb region at 12q21.33, where the patients shared interrupted homozygosity (Table 5.1). There was no known CORD locus on chromosome 12. Of the two proximal homozygous regions at 12q21.33, the centromeric one was the larger, and we first evaluated the deletion we had identified at this larger region for novelty and for any functional elements. The maximal deletion region is ~6 Kb, and no deletion is reported in this region in DGV. The region contains neither a part of a gene nor any predicted gene regulatory elements. Hence, we concluded that this novel deletion most likely did not underlie the disease.

We then analyzed all the genes residing in the shared homozygosity regions, except the pseudogenes and the hypothetical genes by Sanger sequencing all the coding regions. The only novel/rare variant we detected was *POC1B* c.317G>C (p.R106P). The mutation segregated with CORD in the family, and the HRM assay showed that *POC1B* c.317G>C was not a common sequence variant in the Turkish population with 80% power (Collins and Schwartz, 2002). *POC1B* c.317G>C mutation resulted in the substitution of the basic amino acid arginine by nonpolar cyclic proline in the protein product; the substitution is expected to have an adverse effect on the protein structure. This prediction was supported

by two online tools. We also investigated the larger isoform of POC1B protein for conserved sequences in ten different species on HomoloGene database (http://www.ncbi.nlm.nih.gov/homologene). The entire 478 amino acid sequence except for the single residue at position 402 was conserved in full between human and chimpanzee. The mutated residue R106 is within a stretch of 278 amino acids that are fully conserved among human, chimpanzee and rhesus monkey, in a stretch of 20 amino acids that are fully conserved across mammals, and in a stretch of four amino acids that are conserved in all organisms investigated (Figure 6.1).

POC1B is a centriolar protein that plays a key role in centriole duplication and length control (Keller *et al*., 2009). It is essential for both motile and immotile cilia formation, including primary cilia in human RPE cells (Pearson *et al*., 2009). Through its seven WD repeats (amino acids 1–298), POC1B localizes to basal bodies that make up cilia (Keller *et al*., 2009). A defective WD repeat is expected to prevent proper localization to basal bodies and consequently to impair ciliogenesis. Therefore, we propose that alteration of the conserved residue R106 would impair cilia formation in RPE, which in turn would interfere with phagocytosis of the outer segments of photoreceptor (PR) cells, one of the important functions of RPE cilia (Finnemann *et al*., 1997). Outer segments of PR cells are renewed in a daily basis via phagocytosis in order to prevent the accumulation of toxic photo-oxidative products. Excessive accumulation of these products is believed to cause degenerative retinal diseases (Kevany and Palczewski, 2010). Thus, proper function of RPE cilia is essential for PR survival, and mutations in several genes associated with ciliary structure and function are known to underlie PR degeneration (Wright *et al*., 2010). Three genes encoding centriolar proteins, namely, *RPGR, UNC119* and *C8orf37* are associated with CORD (Table 1.1). We therefore concluded that the mutation we identified in *POC1B,* encoding another centriolar gene, underlies the CORD phenotype in the family.

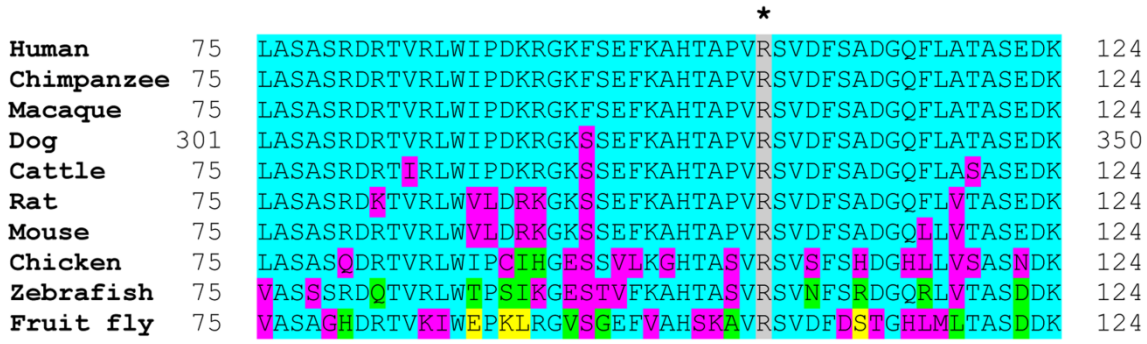

Figure 6.1. Multiple alignment of partial POC1B amino acid sequences across ten Species. Residue R106 is highlighted in grey and marked above with an asterisk.

6.2. Leber Congenital Amaurosis (LCA)

In the family with three blind sibs with LCA diagnosis, we considered a recessive inheritance model for the disease and investigated the loci of all known LCA genes for homozygosity sharing among affected sibs as the first step of the study. Since we could not find any locus with a LCA gene where all affected sibs share a homozygous genotype, we launched linkage analysis in order to identify a novel gene locus. However, the candidate loci that were determined via LOD score calculations and HCiE analysis were all very small and according to exome sequencing results, none of them harbored a gene defect that could underlie the disease. This result led us to consider possibilities other than a homozygous mutation in an autosome as responsible for LCA in our study family. We hypothesized that a mitochondrial mutation or compound heterozygous mutations could underlie the disease. Our motivation for the former was based upon the similarity of symptoms of LCA with Leber's hereditary optic neuropathy (LHON). LHON is an adult onset ocular disease that usually causes complete vision loss as in LCA and is caused by missense mutations in various mitochondrial genes (Yu-Wai-Man *et al*., 2009). Therefore, we investigated exome sequencing results for mitochondrial variants and identified three missense variants after validating them in a patient by Sanger sequencing. We expected heteroplasmy in the mother and possibly the unaffected sister, which would confirm an mitochondrial inheritance model for the disease; however, via Sanger sequencing both participants were found to carry all three variants apparently in the homozygous state, as the patient. We excluded mitochondrial genes upon these findings.

Another possibility for the genetic defect underlying the disease was the presence of two heterozygous mutations in a known or novel gene for LCA. Compound heterozygous mutations are reported in several LCA genes; therefore, we thought that this could be the case for our family as well. In order to search for such mutations, we analyzed the total exome sequence results for a gene that harbored at least two rare, heterozygous, and possibly damaging variants and subsequently investigated via HCiE to see whether the affected sibs but not the unaffected sib shared genotypes at the loci of such genes. We could not find any gene that harbored two deleterious variants, had a function relevant to the disease pathogenesis, and resided in a region of genotype sharing among affected individuals. Finally, we re-focused on the known LCA genes and investigated whether all their exons were covered by exome sequencing. We identified a heterozygous stopgain mutation (p.W278X) in *AIPL1*, the gene for LCA4 in a region not read by exome sequencing. This mutation was previously reported in four families, two Pakistani and two of European descent. The affected members of three of these families were homozygous for the mutation and of the remaining family were compound heterozygous for p.W278X and a 2-bp deletion (Sohocki *et al*., 2000). Based on this report, we hypothesized that our patients could harbor another heterozygous, damaging *AIPL1* variant which had not been detected by exome sequencing. Especially the 5' and 3' untranslated regions (UTRs) of targeted genes are usually not sufficiently covered by exome sequencing, and sometimes the sequencing results are not fully accurate. Therefore, all coding regions and UTRs of *AIPL1* should be analyzed via Sanger sequencing to investigate another possible heterozygous variant. We do not expect heterozygous *AIPL1* c.834G>A mutation exerts its effect dominantly. This is not only because any dominant mutation was reported to date for LCA, but also there is not any affected member in the upper generations of our family as we know. However, the possible absence of a second mutation would not necessarily rule out *AIPL1* c.834G>A (p.W278X), the mutation we identified, from contributing to the disease phenotype in our family. Wiszniewski *et al*., 2011 identified in seven unrelated LCA patients two heterozygous mutations in one LCA gene and an additional heterozygous mutation in another LCA gene; for example, one patient had two mutations in *CEP290* and another in *AIPL1*. A similar triallelic inheritance could underlie the LCA phenotype in the family we studied. If this is our case, Sanger sequencing of the coding regions and the UTRs of *GUCY2D* and *CRB1* that lie in a region where the patients share genotypes could reveal possible other heterozygous mutations in either of these genes.

AIPL1 encodes aryl-hydrocarbon interacting protein-like 1 and is expressed exclusively in retina (van der Spuy *et al*., 2002). Recessive mutations in this gene are associated with a relatively severe form of LCA (Dharmaraj *et al*., 2004). The protein product plays essential roles in survival and functioning of rod and cone photoreceptor cells (van der Spuy *et al*., 2002; Kirschman *et al*., 2010) through interacting with farnesylated proteins within photoreceptors (Ramamurthy *et al*., 2003). Mutations in *AIPL1* may account for approximately 20% of recessive LCA (Sohocki *et al*., 2000), and compound heterozygous mutations in this gene were reported, including p.W278X that we identified, as mentioned above.

AIPL1 protein contains three tetratricopeptide repeats (TPR) constituting a 34-amino-acid motif related to nuclear transportation and chaperone activity (Ma and Whitlock, 1997). The G to A conversion that occurred in the last exon of *AIPL1* in our patients creates a premature translational termination signal at codon 278, resulting in the deletion of the terminal 107 amino acids (28%) of the longest native protein isoform with 384 amino acids. The resulting truncated protein would lack 14 amino acids in the third TPR motif, a region conserved between human, rat, and mouse (Sohocki *et al*., 2000). There is also the possibility that translation would not occur at all. Because, the premature stop codon is located more than 50 nucleotides upstream of the 3′-most exon–exon junction (Nagy and Maquat, 1998) and this would induce the nonsense-mediated decay mechanism that degrades the mutant mRNA.

We concluded that *AIPL1* is associated with the disease phenotype in our family, since it is a known LCA gene and the severe mutation we identified was reported previously. However, the heterozygous mutation *AIPL1* c.834G>A is unlikely to cause LCA by itself. Analysis of all exons of *AIPL1* could fully unravel another possible genetic defect in the patients. The outcomes of this study also demonstrate the major drawback of exome sequencing: all regions targeted to be sequenced may not be fully covered, as in our case. Via investigating whether all exons of the genes in the candidate loci and of the known genes of the studied disease were fully covered, one might avoid missing the disease mutations. Exome sequencing is also not a convenient method to investigate large deletions, such as the deletion of a whole exon, or duplications. CNV analysis could be performed to investigate such variations. However, the genotyping chip we used for this

family did not contain any CNV marker and contained only three SNP markers at *AIPL1* locus. Via quantitative PCR assay, exons of *AIPL1* could be analyzed in the family to investigate a possible variation in exon copy number.

6.3. Optic Atrophy (OPA)

Five affected members of the large consanguineous family participated in the study. Four of them were diagnosed with OPA, which was accompanied by extraocular findings in two of them. Via linkage analysis and evaluation of exome sequencing results, we identified homozygous missense *CEP290* p.Q21P variant in the four patients having OPA and homozygous missense *ACSS3* p.G78D variant in two of those also having extraocular symptoms.

Initial linkage analysis was performed introducing all five patients as affected. The pedigree used was the largest one we analyzed throughout this thesis study, and the completion of the analysis would require an extremely long time. Therefore, markers were selected at 0.5-cM spacing. This analysis yielded three loci with LOD scores >3, and after fine mapping at those loci with markers selected at 0.1-cM spacing, two loci yielded negative scores and were eliminated. The significant drop in the scores could be due to the low intensity of the markers selected in the first analysis, as this could have caused some informative markers to be left out. Since SNP markers are not as polymorphic as microsatellites, they need to be used in much denser spacing in order to obtain a good resolution in linkage analysis. However, linkage programs are not capable of analyzing large pedigrees using dense maps of SNP markers. The two-step analysis we performed seemed to be the best way to overcome this difficulty.

The remaining locus 12q21.33 yielded LOD scores >3 in fine-mapping and was considered the disease locus. There, four patients with OPA shared homozygous *CEP290* c.A62C (p.Q21P). The mutation was not reported in dbSNP138, NHLBI EVS, or CEP290base (http://medgen.ugent.be/cep290base), which is a CEP290 gene-specific mutation database that links mutations with phenotypes. Online tools PolyPhen-2 and MutationTaster predicted the mutation as probably disease causing whereas SIFT predicted it as tolerable. We expected that the substitution of polar glutamine with hydrophobic

proline would affect the protein function adversely. We also investigated the residue for conservation across ten species on HomoloGene database. The mutated residue Q21 was conserved except in cattle and zebrafish, and it was located within a stretch of 99 amino acids that are fully conserved in human, chimpanzee and rhesus monkey (Figure 6.2).

The centrosomal protein encoded by *CEP290* has a role in ciliary assembly and ciliary trafficking, and it interacts with several other centrosomal proteins. *CEP290* is expressed in both ciliated and non-ciliated cells of various tissues including those of the retina (Coppieters *et al*., 2010). It is one of the few ciliary genes associated with retinal degeneration (see Section 1.1.3) and is responsible for LCA and isolated blindness. Its deficiency causes also various other diseases such as Joubert, Meckel-Grüber and Bardet-Biedl syndromes (MIM#610142). Other ciliary genes are thought to have modifier effects in each of those diseases (Coppieters *et al*., 2010). We concluded that it is the causative gene for the OPA phenotype in our study family, since its association with visual impairment has been well established.

As the phenotypes of the affected members of the family are not uniform, we thought that yet another gene defect could be underlying the skeletal deformities, hearing loss and kidney malfunction in the two patients, namely, 508 and 512. In order to identify that putative gene, we performed linkage analysis using a pedigree that assumed only those two individuals as affected and found six candidate loci. Subsequent analysis of the exome sequencing results revealed one novel and one rare, homozygous, missense variants at two of the loci. Variant *PNMT* p.R196W (rs201626166; frequency: 0.00015 according to EVS) was excluded because 512 is heterozygous for it. The other variant *ACSS3* c.G233A (p.G78D) is novel, and patients 508, 512 as well as 408 are homozygous for it. It was predicted to be deleterious to protein function by two online tools. The encoded protein was investigated for conserved sequences across nine species on HomoloGene database. The first 78 amino acids of ACSS3 are fully conserved between human and chimpanzee, and the whole protein is highly conserved between them; only three residues differ in a total of 686 (Figure 6.3). ACSS3 protein is one of the three acetyl-CoA short chain synthetase enzymes (ACSSs) and localizes to mitochondria. ACSSs are involved in the activation of fatty acids via formation of thioesters with CoA (Castro *et al*., 2012). *ACSS3* is expressed in neural and adipose tissues at high levels. Moderate expression was detected

in muscle, kidney, and liver (NCBI UniGene). Its function has not been unraveled yet, and as the other two ACSSs, it has not been associated with a disease. Therefore, it is hard to conclude that the variant we identified in *ACSS3* could cause the extraocular symptoms described in the family we studied. Besides, the only diagnosis made for the family was OPA. Implication of a second disease is possible but not clearly manifested. Whether 408, also homozygous for the *ACSS3* variant, is affected by that disease or non-penetrant for the mutation is another issue to be solved. He should be clinically re-evaluated. *ACSS3* p.G78D is the only candidate variant found via linkage mapping and exome sequencing and might exert its effect as skeletal deformities, kidney malfunction and a few other symptoms in the family.

Figure 6.2. Multiple alignment of amino acid sequences of the amino terminal of CEP290 across ten species. Residue Q21 is highlighted in grey and marked above with an asterisk.

Figure 6.3. Multiple alignments of partial ACSS3 amino acid sequence across nine species. Residues G78 is highlighted in grey and marked above with an asterisk.

6.4. Amyotrophic Lateral Sclerosis (ALS)

In the consanguineous family afflicted with recessive, and early onset-juvenile ALS that we investigated with linkage mapping and subsequent exome sequencing analysis, we identified in the patients a rare, deleterious, homozygous mutation in *ALDH1B1* and a novel missense variant in *GNE*.

Both mapping the disease gene locus and identifying the causative variant was challenging. We first investigated X-linked recessive inheritance pattern, since all affected subjects are male. However, there was no haplotype sharing among patients in any region on the X-chromosome that could possibly be IBD. We then considered autosomal recessive inheritance. It may seem that the probability of all five patients being male is very low; however, considering that there are a total of nine male and only four female offspring in the last generation, the probability of five affected individuals being male is calculated as approximately 10%. The family is one of the rare examples of autosomal recessive FALS. In comparison to numerous mutations in over 20 genes associated with dominant FALS, only few mutations in the six known genes are reported in recessive FALS: 12 in *SPATACSIN* (Orlacchio, 2010), five in *ALSIN* (Hadano, *et al*., 2001; Yang *et al*., 2001; Kress *et al*., 2005), two in *OPTINEURIN* (Maruyama *et al*., 2010), and one in each *SOD1* (Andersen *et al*., 1995), *SIGMAR1* (Al-Saif *et al*., 2011) and *FUS* (Kwiatkowski *et al*., 2009).

Before SNP genotyping for the family, we excluded linkage to *ALSIN* and *SPATACSIN* loci using microsatellite markers flanking these loci. We then performed linkage analysis using SNP genome scan data in order to map the disease and faced the main challenge of the study: Chromosome 9p21.1-p12 yielded the highest LOD scores, and all affected males shared homozygosity for approximately a 10-Mb region; however, heterozygous genotype of mother 407 (Figure 1.4), who was reported as affected at the time, fragmented that long homozygosity. Being unaware of the fact that she was afflicted with multiple sclerosis instead of ALS, we searched for a region where the affected males and the mother share homozygosity, and excluded the best variants detected by exome sequencing in *ALDH1B1* and *GNE*, since the mother was heterozygous for them. After the correction in the diagnosis, we were able to map with confidence the disease gene to 9p21.1-p12, and we re-evaluated the variants at this locus. No doubt, it is constructing the correct pedigree with correct clinical phenotypes which is the first and the most important step in disease gene identification.

The disease locus we identified overlaps with *ALS16* (MIM#614373) locus, which was identified while our study was in progress. *ALS16* is a juvenile form of FALS and recessively inherited. A missense mutation in *SIGMAR1* at 9p13.3 was shown to underlie this disease in a consanguineous family (Al-Saif *et al*., 2011). *ALS16* and our disease were both slow progressive but they have several different features. The onset of *ALS16* was much earlier (1-2 years of age), the patients did not show weakness in bulbar muscles and their upper extremities were affected much later after onset. Nonetheless, we wanted to make sure that our patients did not have a mutation in *SIGMAR1*. Exome sequencing had not revealed any variant in this gene; however, UTRs were not fully covered as the BamView visualization revealed. We analyzed all four exons of the gene by Sanger sequencing in order to detect any variant possibly missed by exome sequencing, as especially large insertions can often escape detection by exome sequencing. We did not detect any mutation in *SIGMAR1*.

The second big challenge of the study was to identify which of the two good candidates, namely, *GNE* c.A2114G and *ALDH1B1* c.1161delC, was the pathogenic mutation. Both were assessed as deleterious and segregated with the disease in the family. *GNE* c.A2114G (p.H705R) is not reported in dbSNP or EVS databases and predicted to be

disease causing by MutationTaster but benign by the other tools PolyPhen-2 and SIFT. The closest reported variant p.Y706H (c.2116T>C) yielded similar prediction results; MutationTaster predicted it to be disease causing and PolyPhen-2 and SIFT to be benign. *GNE* encodes udp-n-acetylglucosamine 2-epimerase/n-acetylmannosamine kinase and is associated with dominant sialuria (MIM#269921) and two recessive muscle diseases, Nonaka distal myopathy (MIM#605820) and inclusion body myopathy 2 (IBM2; MIM#600737). Since ALS is a neuron disease rather than a muscle disease, we did not consider *GNE* as the disease gene. However, our manuscript was rejected, mainly because exclusion of *GNE* based on clinical description was not found convincing. One of the reviewers stated that the phenotype in the family is compatible with IBM2 in several aspects. IBM2 onsets early (before the fourth decade of life), is slowly progressive, and affects mainly leg muscles (Argov and Yarom, 1984; Massa *et al*., 1991; Eisenberg *et al*., 2001), similar to the disease in our patients. On the other hand, ALS diagnosis is probably correct for our patients, since they all presented upper motor signs. The clinicians claimed that upper motor signs are absent in IBM2, and the clinical evaluation is compatible with the diagnostic criteria of ALS (El Escorial World Federation of Neurology Criteria). The electrophysiological test results also supported ALS diagnosis. We therefore considered that *GNE* is unlikely to be responsible for the disease, since the pathology does not originate from muscles. However, all clinical findings need to be re-evaluated before excluding *GNE*, and the manuscript should be revised to explain the clinical evaluations in more detail.

The other candidate variant is *ALDH1B1* c.1161delC (p.G388EfsX23), which is reported in dbSNP137 as rs201408956, with a frequency of 0.007. HRM assay revealed a similar frequency (0.0075; 3 in 400 alleles) in our population. The estimated homozygous frequency of the mutation is approximately 1 in 20,000. This frequency is low enough to associate the variant with FALS, which is a rare disease. *ALDH1B1* c.1161delC is deduced to shift the translational reading frame and result in the truncation of the protein product after the synthesis of 22 non-native amino acids. This would lead to the loss of 25% (terminal 130 residues) of the native protein consisting of 517 residues and damage protein function.

ALDH1B1 (MIM#100670) encodes a mitochondrial enzyme that belongs to the aldehyde dehydrogenase (ALDHs) family of proteins and converts aldehydes to carboxylic acids. Aldehydes are generated in cells, mainly through the breakdown of alcohol, amino acids and lipids, and they are highly reactive compounds and toxic to cells. They may impair cellular homeostasis and lead to DNA damage and cell death (Singh *et al*., 2013). Aldehyde toxicity has been implicated in neurodegenerative diseases such as Parkinsonism and Alzheimer diseases (Voulgaridou *et al*., 2011). Regarding the fact that cytotoxicity is one of the major causes of ALS pathology, *ALDH1B1* can be considered as a good candidate gene for ALS, since a nonfunctional enzyme may lead to the accumulation of aldehydes in the cell.

ALDH1B1 enzyme is involved in various other reactions as well, such as retinoic acid synthesis and oxidation of some products of lipid peroxidation (Stagos *et al*., 2010). *ALDH1B1* has not been linked to any disease but several other members of the *ALDH* superfamily have been linked to various diseases. Gene variants in *ALDH2*, encoding another mitochondrial ALDH enzyme with 73% homology to ALDH1B1 (Hsu and Chang, 1991), are linked to acute alcohol sensitivity in Asian populations (Harada *et al*., 1980). ALDH1B1 deficit could lead to cytotoxicity due to accumulation of aldehydes in the cell. Oxidative stress is known to be a major cause of ALS pathology (Barber *et al*., 2006; Rothstein, 2009). Mutations in ALS genes *SOD1* (Cu,Zn superoxide dismutase), a ubiquitous enzyme that catalyzes the detoxification of superoxide, cause ALS via oxidative damage due to aberrant enzymatic function (Cleveland and Rothstein, 2001), impaired mitochondrial metabolism, axonal degeneration and axonal transport failure (Sreedharan and Brown, 2013). Mutations in *TDP-43 (TARDBP;* TAR DNA-binding protein, 43-KD), another gene responsible for FALS, cause mitochondrial dysfunction as well (Wang *et al*., 2013). We hypothesize a similar pathogenesis for ALDH1B1 deficit: it could increase oxidative stress in mitochondria and thus lead to mitochondrial dysfunction and subsequent cellular stress, which would reduce cell viability and axonal maintenance. We concluded that the deleterious *ALDH1B1* mutation underlies ALS in our study family. The transcript abundance analysis for *ALDH1B1* via qPCR assay showed highest transcript levels in the liver tissue followed by skeletal muscle (Figure 5.19). This result was in agreement with the previous reports on *ALDH1B1* expression (Stagos *et al*., 2010). We detected *ALDH1B1* transcript in all tissues tested.

We repeated alignment and variant calling using GATK with the aim of verifying the candidate variants that were found via SAMTools analysis as well as possibly finding other variants that escaped detection in the analysis by SAMTools. Since the GATK tool for variant calling, Unified Genotyper, uses a different method for alignment, most of the variants called by it were different than those called by SAMTools. However, as we experienced in this and other studies, GATK produces too many false positives. Nevertheless, the analysis with GATK provided verification for *GNE* c.A2114G and *ALDH1B1* c.1161delC, which were the only deleterious variants detected by both SAMTools and GATK.

6.5. Rett Syndrome (RTT)

Via homozygosity mapping and exome sequencing analysis in the small consanguineous family with two sibs diagnosed with RTT, we identified a rare, homozygous, truncating mutation in *P2RX6* in the RTT patients plus their mildly affected sister.

RTT diagnosis for the two sisters of the family was based on the findings of microcephaly, epilepsy, psychomotor retardation, stereotypic hand movements, scoliosis, speech inability, and constipation. However, it was not clear whether microcephaly was congenital or acquired. The first symptom of the disease was seizures that had been noticed in the older sister by the age of six months and in the younger sister by ten months. This onset is earlier than the expected onset for classical type of RTT. Besides, the patients lack a history of regression of acquired skills, which is essential for RTT diagnosis according to the criteria proposed by the RettSearch Consortium (Neul *et al*., 2010). The two sisters never acquired speech or hand skills. On the other hand, the lack of development of those skills may be due to the very early onset of the disease. Nonetheless, the phenotype of the sisters did not fit to the classical form of RTT.

The family had another daughter (402), whom we considered as mildly affected as RTT. She had scoliosis as the only RTT sign, which made it difficult to categorize her as unaffected or mildly affected based solely on this finding. Adolescent idiopathic scoliosis

is a rather common spinal deformity that affects 2% of the population (Ogilvie, 2010). However, scoliosis in 402 began before adolescence; it was noticed at the age 8 years. The prevalence of scoliosis in the age-group of six to eight years was 0.1% and in the age-group of nine to 11 years was 0.3% (Stirling *et al*., 1996). We took into account the possibility of scoliosis in 402 being related to RTT and accordingly considered two models in our analyses: Two RTT sisters are assumed affected in one model and all three sisters are assumed affected in the other.

We first wanted to rule out the very low possibility of a germ line mosaicism shared by the two RTT sisters for a mutation in any known RTT gene, *MECP2*, *CDKL5* and *FOXG1*. Exome sequencing results were investigated to see whether all coding regions of those genes were covered. The coding sequences of *MECP2* and *FOXG1* that had not been covered were analyzed by Sanger sequencing, and consequently, parental germinal mosaicism in any those three genes was excluded. We then searched for the disease locus via homozygosity mapping and linkage analysis.

Before SNP genotyping, the family members had been subjected to whole genome scan using 405 microsatellite markers. Initial linkage analysis using microsatellite genotyping data and assuming only RTT sisters as affected identified 22q11.1 as the only candidate disease locus. However, the site of the crossover in 402 that narrowed down the disease haplotype could not be discerned (Figure 5.21) since microsatellites are not as dense as SNPs, and the parents were non-informative for one marker (AGAT120) within the disease haplotype. Therefore, we thought that *P2RX2* is located in a region where only the RTT sisters shared homozygosity and non-RTT sister 402 had a another genotype until we obtained and analyzed SNP genotyping data and Sanger sequencing results for the *P2RX6* mutation. After we found out that the other sister carried the mutation also in the homozygous state, we re-evaluated the disease status of 402 and diversified our assumptions for the subsequent analyses.

Via analysis of exome sequencing results, we identified only one homozygous, rare, deleterious mutation, *P2RX6* p.K331EfsX26 (c.990dupG) at 22q11.21. *P2RX6* was not previously associated with a disease. The mutation was not reported in dbSNP138 but reported in NHLBI EVS database, in one subject out of total 6258 and in the heterozygous state, suggesting that the mutation is very rare. We did not detect the mutation in the 145

population control subjects tested, showing that it is not a common variant in the Turkish population with 80% power (Collins and Schwartz, 2002). *P2RX6* p.K331EfsX26 is a frameshift mutation that arose by the duplication of a G at codon 331, and it is predicted to result in premature termination of translation following the synthesis of 25 non-native amino acids after residue 330. As a result, the mutant peptides would lack 86 of the native 441 amino acids in isoform 1 and of the 415 residues in isoform 2, if the protein is translated. However, nonsense-mediated decay mechanism would probably degrade the mutant mRNA, since the premature stop codon is located more than 50 nucleotides upstream of the 3′-most exon–exon junction (Nagy and Maquat, 1998). If degradation does not occur, *P2RX6* p.K331EfsX26 mutation would lead to a truncated protein lacking one of the transmembrane domains (residues 334 to 354) as well as all downstream residue. This deleterious mutation was a good candidate to underlie the pathology in the family.

P2RX6 encodes P2X6, a purinoreceptor protein subunit belonging to the P2X purinergic receptor family, members of which are ATP-gated ion channels that are permeable to small cations. All seven members of the P2X receptor gene family encode P2X receptor subunits ($P2X_{1-7}$), each having two transmembrane domains separated by an extracellular domain (North, 2002). ATP acts as an extracellular messenger via binding to P2X receptors and inducing a change in the intracellular cation concentrations, such as calcium. P2X receptors are expressed in the neurons of mouse hypothalamus (Fu *et al*., 2009), suggesting that those receptors are important for the nervous system. Again, in mouse, alternative splicing of *P2rx6* was shown both in developing brain and during in vitro neuronal differentiation (da Silva *et al*., 2007). Another study reported that cultured rat neural stem cells had high concentrations of P2X2 and P2X6 proteins, suggesting that P2X2/6 receptors have a role in the induction of neuronal differentiation (Schwindt et al., 2011). Based on these findings, we hypothesize that P2RX6 deficit is compatible with RTT, which is a neurodevelopmental disease. P2RX6 was suggested to have other physiological functions such as transmission of nociceptive information (Bardoni *et al*., 1997) and regulation of enteric nervous system functioning (Yu *et al*., 2010). Our patients' pain reception seemed normal; however, their complaint of persistent constipation could be a sign of abnormality in gastrointestinal tract neurons resulting from the P2RX6 deficit.

P2RX6 (MIM#608077) is known to be expressed in mouse and rat brains (Collo *et al*., 1996; da Silva *et al*., 2007) and in mouse skeletal muscles (Nawa et al., 1998). Consistent with the reported data, our qPCR assay showed wide expression in adult human neuronal tissues and high expression in skeletal muscle. Variable amount of *P2RX6* transcripts was found in all 11 neuronal tissues (Figure 5.25), suggesting that P2RX6 is most likely essential for the nervous system. Predominant expression of *P2RX6* in skeletal muscle indicates an important role of this gene for skeletal muscle. Indeed, P2RX6 was shown to be involved in the proliferation and/or differentiation of skeletal muscle cells (Urano *et al*., 1997). However, the condition of the RTT patients we studied did not support that hypothesis; there was no muscle involvement.

Our finding that the mildly affected sister, who has only scoliosis as a RTT sign, is homozygous for the same deleterious mutation as her severely affected sisters was unexpected. However, such mutations with variable expressivity are not very rare (Cooper et al., 2013). One good example for this phenomenon is the case of SHFM6, in which one of the sibs of the family afflicted with split-hand/foot malformation carried the homozygous *WNT10B* mutation identified in the patients, despite the lack of any bone malformation in her. She had decreased bone density, observable only via radiographic investigation (Ugur and Tolun, 2008). It is difficult to detect variable expressivity in a family, since seemingly unaffected members do not usually undergo detailed physical examinations and diagnostic tests, and those individuals can even be excluded from the study. Therefore, mutations with variable expressivity are most likely underdetected.

Based on the results of the homozygosity mapping, exome sequencing and expression analyses, and the reported data emphasizing the importance of P2RX6 for neuronal differentiation, we concluded that the deleterious *P2RX6* p.K331EfsX26 mutation underlies RTT phenotype in our study family. Screening *P2RX6* in patients affected by RTT spectrum disorders without any mutation in *MECP2, CDKL5* or *FOXG1* genes could identify other mutations in *P2RX6.*

6.6. Infantile Neuroaxonal Dystrophy (INAD)

We studied six families afflicted with INAD, and in four of them identified novel mutations in *PLA2G6*, the known INAD gene encoding calcium-independent phospholipase iPLA2-VI. Two of the mutations were missense, and two others were truncating (Table 5.10). PolyPhen, SIFT and MutationTaster predicted the effect of T661M mutation to be damaging to protein function. Moreover, residue T661 is within a stretch of 47 amino acids that is fully conserved in mammals. Predictions for the other missense mutation, D484G, were variable. PolyPhen predicted it as benign and SIFT as damaging but with low confidence whereas MutationTaster predicted it as disease causing. This latter prediction seems more likely to be the correct one because the substitution of acidic aspartic acid with small, neutral glycine is expected to have an adverse effect on protein function. Residue 484 is within a stretch of 103 amino acids that is fully conserved among human, chimpanzee and macaque. Of the truncating mutations, p.F135SfsX79 (c.402delC), is very severe. It is deduced to result in the incorporation of 78 non-native residues after residue 134 and the loss of 74% of the native protein. However, the mutant mRNA is not expected to be translated due to nonsense-mediated decay mechanism. The two sibs homozygous for c.402delC mutation had the most severe phenotype among all patients we studied; they died at about 2 years of age. The phenotypes of the other patients with *PLA2G6* mutations were typical INAD. We could not detect any mutation in Families 5 and 6. For Family 6, we did not have enough DNA samples to continue study by SNP genotyping, and the affected sibs were deceased.

Neurological examination on the two affected sibs in Family 5 revealed severe psychomotor retardation, muscular atrophy, truncal hypotonia, and total loss of audiovisual functions. A nerve biopsy sample of one patient showed spheroid formations. Based on those findings, the sibs had been diagnosed with INAD. However, they had some features atypical to INAD such as facial dysmorphism and longer life span. At the last visit, patients were 18 and 21 years of age; however, they had very severe developmental retardation (see Section 1.5). We did not detect any mutation in *PLA2G6* in this family and hypothesized that a defect in a new gene underlies the disease. Presence of an additional INAD locus in humans was demonstrated by linkage analysis (Morgan *et al*., 2006), and a

mutation in *MFN2* gene was shown to underlie canine fetal-onset neuroaxonal dystrophy in another study (Fyfe *et al*., 2011).

Via linkage mapping and subsequent analysis of exome sequencing we identified homozygous *NALCN* c.1924C>T (p.Q642X) at the largest candidate locus 13q32.3-33.2. The mutation was novel, and we did not find it among the 110 population samples either, showing with 80% power that *NALCN* c.1924C>T is not a common variant in the Turkish population (Collins and Schwartz, 2002). The C to T conversion occurred in exon 16 and created a premature translational termination signal at codon 642, resulting in the deletion of 1097 amino acids (63%) of the native protein with 1738 amino acids (Figure 6.4). The mutant mRNA is expected to undergo nonsense-mediated decay. NALCN protein is fully conserved between human and chimpanzee and highly conserved among other mammals. Only 24 (1.4%) of the total 1738 residues differs between human and rat (HomoloGene).

NALCN is expressed predominantly in the central nervous system (Lu *et al*., 2007). Moderate expression in heart and weak expression in pancreas are reported as well (Lee *et al*., 1999). The results of our qPCR assay are consistent with the previous studies. Variable *NALCN* expression was observed in all brain regions tested, but there was no expression in any of the non-neural tissues, suggesting that *NALCN* is largely neuron specific.

NALCN encodes a voltage-independent cation channel permeable to sodium, potassium and calcium (MIM#611549). It is a member of the sodium/calcium channel gene family, and its voltage-independent and nonselective nature is unique among the total 21 members of the family of voltage-gated sodium-selective or calcium-selective channel proteins (Lu *et al*., 2007). NALCN protein is composed of four homologous domains, each with six transmembrane segments (Figure 6.4). The channel regulates neuronal excitability (Lu *et al*., 2007), and no other channel with this function is known in mammals. Regarding the structural and functional uniqueness of the protein together with its high sequence conservation among mammals, we hypothesized that the deficit of this protein would be pathogenic for the organism.

NALCN regulates neuronal excitability through forming a complex with UNC79 and UNC80 proteins upon changes in the extracellular calcium level (Lu *et al*., 2010). This

activity requires the two intracellular domains of the protein, namely, the carboxy-terminus tail and the loop connecting repeats II and III (Figure 6.4). NALCN protein lacking the last 202 residues was shown to be nonfunctional. Even the deletion of the last amino acid (I1738) had an adverse effect on protein function; it rendered the channel insensitive to changes in calcium level (Lu *et al*., 2010). Based on those findings, Q642X mutation is assessed as deleterious, if the mutant mRNA gets translated.

Animal models with mutations in the orthologous genes showed that the protein is involved in the regulation of rhythmic behaviors in the studied animals. *Nalcn* mutant mice had disrupted respiratory rhythm that led to death soon after birth (Lu *et al*., 2007), mutant fruit flies had inverted day and night locomotor activities (Nash *et al*., 2002), and round worms lost the ability to switch between crawling and swimming behavior when they moved from solid to aqueous environment (Pierce-Shimomura *et al*., 2008). Our patients did not have disruption in respiratory rhythm. We investigated for any abnormality in heart rhythm as well, regarding that *NALCN* has moderate expression in the heart. Electrocardiogram of the patients did not show altered heart rates or arrhythmia. Our findings suggest that the phenotypes of animal models do not match human phenotype for NALCN deficiency.

In the light of the evidences mentioned above, we concluded that *NALCN* mutation underlies the disease in Family 5. We published our manuscript reporting *NALCN* as the gene for INAD with facial dysmorphism (Koroglu *et al*., 2013). The disease phenotype in our family together with the phenotype seen in other two families with homozygous *NALCN* mutations (Al-Sayed *et al*., 2013) that was reported a few months after our manuscript were given the name infantile hypotonia with psychomotor retardation and characteristic facies (IHPRF; MIM#615419) in OMIM database. The families in the study of Al-Sayed *et al*. (2013) were afflicted with a syndrome characterized by facial dysmorphism, hypotonia, speech impairment, chronic constipation, and intellectual disability. Phenotypic similarity of those patients to ours verifies the implication of *NALCN* in the severe neurological disease designated as IHPRFs.

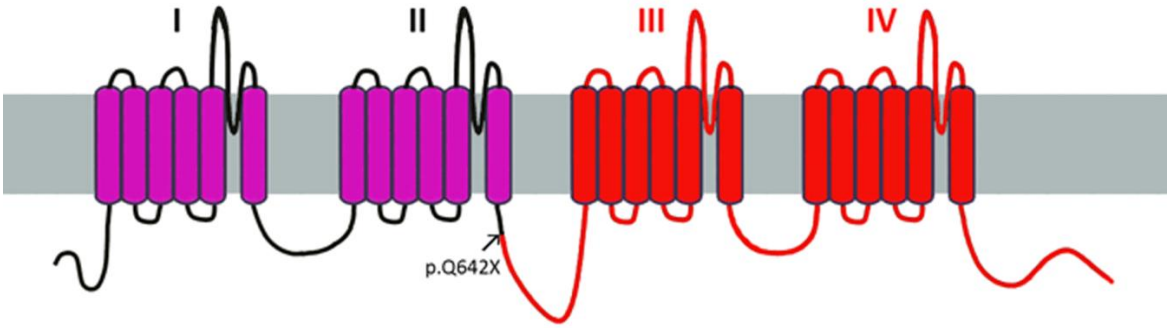

Figure 6.4. A structural model of NALCN. The protein has four homologous domains. The mutation is deduced to delete the latter two domains which are shown in red. The cell membrane is in grey (Koroglu *et al.*, 2013).

7. CONCLUSION

In this study, novel genes are identified as causative for five autosomal recessive disorders: *POC1B* for Cone-Rod Dystrophy, *CEP290* for Optic Atrophy, *ALDH1B1* for Amyotrophic Lateral Sclerosis, *P2RX6* for Rett Syndrome, and *NALCN* for Infantile Neuroaxonal Dystrophy. Additionally, a previously reported mutation was identified in *AIPL1* in a family afflicted with Leber Congenital Amaurosis, and a novel, homozygous variant in *ACSS3* was detected in three patients in Optic Atrophy family, possibly underlying the extraocular neurological symptoms shared by at least two of the patients.

REFERENCES

Afzelius, B. A., 2004, "Cilia-Related Diseases", *The Journal of Pathology*, Vol. 240, No. 4, pp. 470-477.

Al-Saif, A., F. Al-Mohanna, and S. A. Bohlega, 2011, "Mutation in Sigma-1 Receptor Causes Juvenile Amyotrophic Lateral Sclerosis", *Annals of Neurology*, Vol. 70, pp. 913-919.

Amir, R. E., I. B. Van den Veyver, M. Wan, C. Q. Tran, U. Francke, and H. Y. Zoghbi, 1999, "Rett Syndrome is Caused by Mutations in X-Linked MECP2, Encoding Methyl-CpG-Binding Protein 2", *Nature Genetics*, Vol. 23, No. 2, pp. 185-188.

Andersen, P. M., P. Nilsson, V. Ala-Hurula, M. L. Keränen, I. Tarvainen, T. Haltia, L. Nilsson, M. Binzer, L. Forsgren, and S. L. Marklund, 1995, "Amyotrophic Lateral Scleros is Associated with Homozygosity for an Asp90Ala Mutation in CuZn-Superoxide Dismutase", *Nature Genetics*, Vol. 10, No. 1, pp. 61-66.

Argov, Z., R. Yarom, 1984, "Rimmed Vacuole Myopathy Sparing the Quadriceps. A Unique Disorder in Iranian Jews", *Journal of the Neurological Sciences*, Vol. 64, No. 1, pp. 33-43.

Baburina, I., and S. Jackowski, 1999, "Cellular Responses to Excess Phospholipid", *Journal of Biological Chemistry*, Vol. 274, No. 14, pp. 9400-9408.

Badano, J. L., N. Mitsuma, P. L. Beales, and N. Katsanis, 2006, "The Ciliopathies: an Emerging Class of Human Genetic Disorders", *Annual Review of Genomics and Human Genetics*, Vol. 7, pp. 125-148.

Bamshad, M. J., S. B. Ng, A. W. Bigham, H. K. Tabor, M. J. Emond, D. A. Nickerson, and J. Shendure, 2011, "Exome Sequencing as a Tool for Mendelian Disease Gene Discovery", *Nature Reviews Genetics*, Vol. 12, No. 11, pp. 745-755.

Barber, S. C., R. J. Mead, and P. J. Shaw, 2006, "Oxidative Stress in ALS: a Mechanism of Neurodegeneration and a Therapeutic Target", *Biochimica et Biophysica Acta*, Vol. 1762, pp. 1051-1067.

Bardoni, R., P. A. Goldstein, C. J. Lee, J. G. Gu, and A. B. MacDermott, 1997, "ATP P2X Receptors Mediate Fast Synaptic Transmission in the Dorsal Horn of the Rat Spinal Cord", *The Journal of Neuroscience*, Vol. 17, No. 14, pp. 5297-5304.

Castro, L. F., M. Lopes-Marques, J. M. Wilson, E. Rocha, M. A. Reis-Henriquez, M.M. Santos, and I. Cunha, 2012, "A Novel Acetyl-CoA Synthetase Short-Chain Subfamily Member 1 (Acss1) Gene Indicates a Dynamic History of Paralogue Retention and Loss in Vertebrates", *Gene*, Vol. 497, No. 2, pp. 249-255.

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

Cleveland, D. W., 1999, "From Charcot to SOD1: Mechanisms of Selective Motor Neuron Death in ALS", *Neuron*, Vol. 24, No. 3, pp. 515-520.

Cleveland, D. W., and J. D., Rothstein, 2001, "From Charcot to Lou Gehrig: Deciphering Selective Motor Neuron Death in ALS", *Nature Reviews Neuroscience*, Vol. 2, No. 11, pp. 806-819.

Collins, J. S., and C. E. Schwartz, 2002, "Detecting Polymorphisms and Mutations in Candidate Genes", *American Journal of Human Genetics*, Vol. 71, No. 5, pp. 1251-1252.

Collo, G., R. A. North, E. Kawashima, E. Merlo-Pich, S. Neidhart, A. Surprenant, and G. Buell, 1996, "Cloning of P2X5 and P2X6 Receptors and the Distribution and

Properties of an Extended Family of ATP-gated Ion Channels", *The Journal of Neuroscience*, Vol. 16, No. 8, pp. 2495-2507.

Cooper, D. N., M. Krawczak, C. Polychronakos, C. Tyler-Smith, and H. Kehrer-Sawatzki, 2013, "Where Genotype is not Predictive of Phenotype: Towards an Understanding of the Molecular Basis of Reduced Penetrance in Human Inherited Disease", *Human Genetics*, Vol. 132, No. 10, pp. 1077-1130.

Coppieters, F., S. Lefever, B. P. Leroy, and E. De Baere, 2010, "CEP290, a Gene with Many Faces: Mutation Overview and Presentation of CEP290base", *Human Mutation*, Vol. 31, No. 10, pp. 1097-1108.

da Silva, R. L., R. R. Resende, and H. Ulrich, 2007, "Alternative Splicing of P2X6 Receptors in Developing Mouse Brain and During in Vitro Neuronal Differentiation", *Experimental Physiology*, Vol. 92, No. 1, pp. 139-145.

Dawe, H. R., H. Farr, and K. Gull, 2007, "Centriole/basal Body Morphogenesis and Migration During Ciliogenesis in Animal Cells", *Journal of Cell Science*, Vol. 120, pp. 7-15.

Dawn Teare, M., and J. H. Barrett, 2005, "Genetic Linkage Studies", *Lancet*, Vol. 366, No. 9490, pp. 1036-1044.

Dion P. A., H. Daoud, and G. A. Rouleau, 2009, "Genetics of Motor Neuron Disorders: New Insights into Pathogenic Mechanisms", *Nature Reviews Genetics*, Vol. 10, No. 11, pp. 769-782.

den Hollander, A. I., R. K. Koenekoop, S. Yzer, I. Lopez, M. L. Arends, K. E. J. Voesenek, M. N. Zonneveld, T. M. Strom, T. Meitinger, H. G. Brunner, C. B. Hoyng, L. I. van den Born, K. Rohrschneider, and F. P. M. Cremers, 2006, "Mutations in the CEP290 (NPHP6) Gene are a Frequent Cause of Leber Congenital Amaurosis", *American Journal of Human Genetics*, Vol. 79, pp. 556-561.

den Hollander, A. I., R. Roepman, R. K. Koenekoop, and F. P. M. Cremers, 2008, "Leber Congenital Amaurosis: Genes, Proteins and Disease Mechanisms", *Progress in Retinal and Eye Research*, Vol. 27, pp. 391-419.

Dharmaraj, S., B. P. Leroy, M. M. Sohocki, R. K. Koenekoop, I. Perrault, K. Anwar, S. Khaliq, R. S. Devi, D. G. Birch, E. De Pool, N. Izquierdo, L. Van Maldergem, M. Ismail, A. M. Payne, G. E. Holder, S. S. Bhattacharya, A. C. Bird, J. Kaplan, and I. H. Maumenee, 2004, "The Phenotype of Leber Congenital Amaurosis in Patients with AIPL1 Mutations", *Archives of Ophthalmology*, Vol. 22, No. 7, pp. 1029-1037.

Donis-Keller, H., P. Green, C. Helms, S. Cartinhour, B. Weiffenbach, K. Stephens, T. P. Keith, D. W. Bowden, D. R. Smith, E. S. Lander, D. Botstein, G. Akots, K. S. Rediker, T. Gravius, V. A. Brown, M. B. Rising, C. Parker, J. A. Powers, D. E. Watt, E. R. Kaufman, A. Bricker, P. Phipps, H. Muller-Kahle, T. R. Fulton, S. Ng, J. W. Schumm, J. C. Braman, R. G. Knowlton, D. F. Barker, S. M. Crooks, S. E. Lincoln, and M. J. Daly, 1987, "A Genetic Linkage Map of the Human Genome", *Cell*, Vol. 51, pp. 319-337.

Durlu, Y. K., Ç. Köroğlu, and A. Tolun, 2014, "Novel Recessive Cone-Rod Dystrophy Caused by POC1B Mutation", *JAMA Ophthalmology*, http://archopht.jamanetwork.com/article.aspx?articleid=1882218, accessed at June 2014.

Eisenberg, I., N. Avidan, T. Potikha, H. Hochner, M. Chen, T. Olender, M. Barash, M. Shemesh, M. Sadeh, G. Grabov-Nardini, I. Shmilevich, A. Friedmann, G. Karpati, W. G. Bradley, L. Baumbach, D. Lancet, E. B. Asher, J. S. Beckmann, Z. Argov, and S. Mitrani-Rosenbaum, 2001, "The UDP-N-Acetylglucosamine 2-Epimerase/N-Acetylmannosamine Kinase Gene is Mutated in Recessive Hereditary Inclusion Body Myopathy", *Nature Genetics*, Vol. 29, No. 1, pp. 83-87.

Finnemann, S. C., V. L. Bonilha, A. D. Marmorstein, and E. Rodriguez-Boulan, 1997, "Phagocytosis of Rod Outer Segments by Retinal Pigment Epithelial Cells Requires Alpha(v)beta5 Integrin for Binding but not for Internalization", *Proceedings of the*

National Academy of Sciences of the United States of America, Vol. 94, No. 24, pp. 12932-12937.

Fu, J., Q. Yu, W. Guo, C. He, G. Burnstock, and Z. Xiang, 2009, "P2X receptors are Expressed on Neurons Containing Luteinizing Hormone-Releasing Hormone in the Mouse Hypothalamus", *Neuroscience Letters*, Vol. 458, No. 1, pp. 32-36.

Fyfe, J. C., R. A. Al-Tamimi, J. Liu, A. A. Schaffer, R. Agarwala, and P. S. Henthorn, 2011, "A Novel Mitofusin 2 Mutation Causes Canine Fetal-Onset Neuroaxonal Dystrophy", *Neurogenetics*, Vol. 12, pp. 223-232.

Gilissen, C., A. Hoischen, H. G. Brunner, and J. A. Veltman, 2012, "Disease Gene Identification Strategies for Exome Sequencing", *European Journal of Human Genetics*, Vol. 20, No. 5, pp. 490-497.

Gordon, N., 2002, "Infantile Neuroaxonal Dystrophy (Seitelberger's Disease)", *Developmental Medicine and Child Neurology*, Vol. 44, No. 12, pp. 849-851.

Gu, S., D. A. Thompson, C. R. S. Srikumari, B. Lorenz, U. Finckh, A. Nicoletti, K. R. Murthy, M. Rathmann, G. Kumaramanickavel, M. J. Denton, and A. Gal, 1997, "Mutations in RPE65 Cause Autosomal Recessive Childhood-Onset Severe Retinal Dystrophy", *Nature Genetics*, Vol. 17, No.2, pp. 194-197.

Gudbjartsson, D. F., K. Jonasson, M. L. Frigge, and A. Kong, 2000, "Allegro, a New Computer Program for Multipoint Linkage Analysis", *Nature Genetics*, Vol. 25, No. 1, pp. 12-13.

Hadano, S., C. K. Hand, H. Osuga, Y. Yanagisawa, A. Otomo, R. S. Devon, N. Miyamoto, J. Showguchi-Miyata, Y. Okada, R. Singaraja, D. A. Figlewicz, T. Kwiatkowski, B. A. Hosler, T. Sagie, J. Skaug, J. Nasir, R. H. Brown Jr, S. W. Scherer, G. A. Rouleau, M. R. Hayden, and J. E. Ikeda, 2001, "A Gene Encoding a Putative GTPase Regulator is Mutated in Familial Amyotrophic Lateral Sclerosis 2", *Nature Genetics*, Vol. 29, No. 2, pp. 166-173.

Hamel, C. P., 2007, "Cone Rod Dystrophies", *Orphanet Journal of Rare Diseases,* Vol. 2, No. 2, pp. 1-7.

Hanein S., I. Perrault, O. Roche, S. Gerber, N. Khadom, M. Rio, N. Boddaert, M. Jean-Pierre, N. Brahimi, V. Serre, D. Chretien, N. Delphin, L. Fares-Taie, S. Lachheb, A. Rotig, F. Meire, A. Munnich, J. L. Dufier, J. Kaplan, and J. M. Rozet, 2009, "TMEM126A, Encoding a Mitochondrial Protein, is Mutated in Autosomal-Recessive Nonsyndromic Optic Atrophy", *American Journal of Human Genetics,* Vol. 84, No. 4, pp. 493-498.

Harada, S., S. Misawa, D. P. Agarwal, and H. W. Goedde, 1980, "Liver Alcohol Dehydrogenase and Aldehyde Dehydrogenase in the Japanese: Isozyme Variation and its Possible Role in Alcohol Intoxication", *American Journal of Human Genetics,* Vol. 32, pp. 8-15.

Haverkamp, C. J., V. Appel, and S. H. Appel, 1995, "Natural History of Amyotrophic Lateral Sclerosis in a Database Population: Validation of Scoring System and a Model for Survival Prediction", *Brain,* Vol. 118, pp. 707-719.

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.

Hooks, S. B., and B. S. Cummings, 2008, "Role of Ca2+-Independent Phospholipase A2 in Cell Growth and Signaling", *Biochemical Pharmacology,* Vol. 76, No. 9, pp. 1059-1067.

Hsu, L. C., and W. C. Chang, 1991, "Cloning and Characterization of a New Functional Human Aldehyde Dehydrogenase Gene", *The Journal of Biological Chemistry,* Vol. 266, No. 19, pp. 12257-12265.

Huizing M., H. Dorward, L. Ly, E. Klootwijk, R. Kleta, F. Skovby, W. Pei, B. Feldman, W. A. Gahl, and Y. Anikster, 2010, "OPA3, Mutated in 3-Methylglutaconic Aciduria Type III, Encodes Two Transcripts Targeted Primarily to Mitochondria", *Molecular Genetics and Metabolism*, Vol. 100, pp. 149-154.

Kavaslar, G. N., S. Önengüt, O. Derman, A. Kaya and A. Tolun, 2000, "The Novel Genetic Disorder Microhydranencephaly Maps to Chromosome 16p13.3-12.1", *American Journal of Human Genetics*, Vol. 66, No. 5, pp. 1705-1709.

Keller, L. C., S. Geimer, E. Romijn, J. Yates 3rd, I. Zamora, and W. F. Marshall, 2009, "Molecular Architecture of the Centriole Proteome: The Conserved WD40 Domain Protein POC1 is Required for Centriole Duplication and Length Control", *Molecular Biology of the Cell*, Vol. 20, No. 4, pp. 1150-1166.

Kevany, B. M., K. Palczewski, 2010, "Phagocytosis of Retinal Rod and Cone Receptors", *Physiology (Bethesda)*, Vol. 25, No. 1, pp. 8-15.

Khateeb, S., H. Flusser, R. Ofir, I. Shelef, G. Narkis, G. Vardi, Z. Shorer, R. Levy, A. Galil, K. Elbedour, and O. S. Birk, 2006, "PLA2G6 Mutation Underlies Infantile Neuroaxonal Dystrophy", *American Journal of Human Genetics*, Vol. 79, pp. 942-948.

Kirschman, L. T., S. Kolandaivelu, J. M. Frederick, L. Dang, A. F. Goldberg, W. Baehr, and V. Ramamurthy, 2010, "The Leber Congenital Amaurosis Protein, AIPL1, is Needed for the Viability and Functioning of Cone Receptor Cells", *Human Molecular Genetics*, Vol. 19, No. 6, pp. 1076-1087.

Koenekoop, R. K., 2004, "An Overview of Leber Congenital Amaurosis: a Model to Understand Human Retinal Development", *Survey of Ophthalmology*, Vol. 49, pp. 379-398.

Koenekoop, R. K., I. Lopez, A. I. Den Hollander, R. Allikmets, and F. P. M. Cremers, 2007, "Genetic Testing for Retinal Dystrophies and Dysfunctions: Benefits,

Dilemmas and Solutions", *Clinical & Experimental Ophthalmology*, Vol. 35, No. 5. pp. 473-485.

Köroğlu, Ç., M. Seven, and A. Tolun, 2013, "Recessive Truncating NALCN Mutation in Infantile Neuroaxonal Dystrophy with Facial Dysmorphism", *Journal of Medical Genetics*, Vol. 50, No. 8, pp. 515-520.

Kortüm, F., S. Das, M. Flindt, D. J. Morris-Rosendahl, I. Stefanova, A. Goldstein, D. Horn, E. Klopocki, G. Kluger, P. Martin, A. Rauch, A. Roumer, S. Saitta, L. E. Walsh, D. Wieczorek, G. Uyanik, K. Kutsche, and W. Dobyns, 2011, "The Core FOXG1 Syndrome Phenotype Consists of Postnatal Microcephaly, Severe Mental Retardation, Absent Language, Dyskinesia, and Corpus Callosum Hypogenesis", *Journal of Medical Genetics*, Vol. 48, No.6, pp. 396-406.

Kress, J. A., P. Kühnlein, P. Winter, A. C. Ludolph, J. Kassubek, U. Müller, and A. D. Sperfeld, 2005, "Novel Mutation in the ALS2 Gene in Juvenile Amyotrophic Lateral Sclerosis", *Annals of Neurology*, Vol. 58, No. 5, pp. 800-803.

Kryukov, G. V., L. A. Pennacchio, and S. R. Sunyaev, 2007, "Most Rare Missense Alleles are Deleterious in Humans: Implications for Complex Disease and Association Studies", *American Journal of Human Genetics*, Vol. 80, No. 4, pp. 727-739.

Kurian, M. A., N. V. Morgan, L. MacPherson, K. Foster, D. Peake, R. Gupta, S. G. Philip, C. Hendriksz, J. E. Morton, H. M. Kingston, E. M. Rosser, E. Wassmer, P. Gissen, and E. R. Maher, 2008, "Phenotypic Spectrum of Neurodegeneration Associated with Mutations in the PLA2G6 Gene (PLAN)", *Neurology*, Vol. 70, pp. 1623-1629.

Kwiatkowski, T. J. D. A. Bosco Jr, A. L. Leclerc, E. Tamrazian, C. R. Vanderburg, C. Russ, A. Davis, J. Gilchrist, E. J. Kasarskis, T. Munsat, P. Valdmanis, G. A. Rouleau, B. A. Hosler, P. Cortelli, P. J. de Jong, Y. Yoshinaga, J. L. Haines, M. A. Pericak-Vance, J. Yan, N. Ticozzi, T. Siddique, D. McKenna-Yasek, P. C. Sapp, H. R. Horvitz, J. E. Landers, and R. H. Brown Jr, 2009, "Mutations in the FUS/TLS

Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis", *Science*, Vol. 323, No. 5918, pp. 1205-1208.

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.

Lee, J. H., L. L. Cribbs, and E. Perez-Reyes, 1999, "Cloning of a Novel Four Repeat Protein Related to Voltage-Gated Sodium and Calcium Channels", *FEBS Letters*, Vol. 445, pp. 231-236.

Lu, B., Y. Su, S. Das, J. Liu, J. Xia, and D. Ren, 2007, "The Neuronal Channel NALCN Contributes Resting Sodium Permeability and is Required for Normal Respiratory Rhythm", *Cell*, Vol. 129, No. 2, pp. 371-383.

Ma, Q., and J. P. Whitlock Jr, 1997, "A Novel Cytoplasmic Protein that Interacts with the Ah Receptor, Contains Tetratricopeptide Repeat Motifs, and Augments the Transcriptional Response to 2,3,7,8-Tetrachlorodibenzo-p-Dioxin", *The Journal of Biological Chemistry*, Vol. 272, No. 14, pp. 8878-8884.

Maruyama, H., H. Morino, H. Ito, Y. Izumi, H. Kato, Y. Watanabe, Y. Kinoshita, M. Kamada, H. Nodera, H. Suzuki, O. Komure, S. Matsuura, K. Kobatake, N. Morimoto, K. Abe, N. Suzuki, M. Aoki, A. Kawata, T. Hirai, T. Kato, K. Ogasawara, A. Hirano, T. Takumi, H. Kusaka, K. Hagiwara, R. Kaji, and H. Kawakami, 2010, "Mutations of Optineurin in Amyotrophic Lateral Sclerosis", *Nature*, Vol. 465, No. 7295, pp. 223-226.

Masland, R. H., 2001, "The Fundamental Plan of the Retina", *Nature Neuroscience*, Vol. 4, pp. 877-886.

Massa, R., B. Weller, G. Karpati, E. Shoubridge, and S. Carpenter, 1991, "Familial Inclusion Body Myositis Among Kurdish-Iranian Jews", *Archives of Neurology*, Vol.48, No. 5, pp. 519-522.

Metzker, M. L., 2010, "Sequencing Technologies - the Next Generation", *Nature Reviews Genetics*, Vol. 11, No. 1, pp. 31-46.

Michaelides, M., A. J. Hardcastle, D. M. Hunt, and A.T. Moore, 2006, "Progressive Cone and Cone-Rod Dystrophies: Phenotypes and Underlying Molecular Genetic Basis", *Survey of Ophthalmology*, Vol. 51, No. 3, pp. 232-258.

Moore, A. T., 1992, "Cone and Cone-Rod Dystrophies", *Journal of Medical Genetics*, Vol. 29, No. 5, pp. 289-290.

Morgan, N. V., S. K. Westaway, J. E. Morton, A. Gregory, P. Gissen, S: Sonek, H. Cangul, J. Coryell, N. Canham, N. Nardocci, G. Zorzi, S. Pasha, D. Rodriguez, I. Desguerre, A. Mubaidin, E. Bertini, R. C. Trembath, A. Simonati, C. Schanen, C. A. Johnson, B. Levinson, C. G. Woods, B. Wilmot, P. Kramer, J. Gitschier, E. R. Maher, and S. J. Hayflick, 2006, "PLA2G6, Encoding a Phospholipase A(2), is Mutated in Neurodegenerative Disorders with High Brain Iron", *Nature Genetics*, Vol. 38, No. 7, pp. 752-754.

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

Mulder, D. W., L. T. Kurland, K. P. Offord, and C. M. Beard, 1986, "Familial Adult Motor Neuron Disease: Amyotrophic Lateral Sclerosis", *Neurology*, Vol. 36, No. 4, pp. 511-517.

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.

Nash, H. A., R. L. Scott, B. C. Lear, and R. Allada, 2002, "An Unusual Cation Channel Mediates Photic Control of Locomotion in Drosophila", *Current Biology*, Vol. 12, No. 24, pp. 2152-2158.

Nawa, G., T. Urano, T. Tokino, T. Ochi, and Y. Miyoshi, 1998, "Cloning and Characterization of the Murine P2XM Receptor Gene", *Journal of Human Genetics*, Vol. 43, No. 4, pp. 262-267.

Neul, J. L., W. E. Kaufmann, D. G. Glaze, J. Christodoulou, A. J. Clarke, N. Bahi-Buisson, H. Leonard, M. E. Bailey, N. C. Schanen, M. Zappella, A. Renieri A, P. Huppke, and A. K. Percy; RettSearch Consortium, 2010, "Rett Syndrome: Revised Diagnostic Criteria and Nomenclature", *Annals of Neurology*, Vol. 68, No. 6, pp. 944-950.

North, R. A., 2002, "Molecular Physiology of P2X Receptors", *Physiological Reviews*, Vol. 82, No. 4, pp. 1013-1067.

Ogilvie, J., 2010, "Adolescent Idiopathic Scoliosis and Genetic Testing", *Current Opinion in Pediatrics*, Vol. 22, pp. 67-70.

Orlacchio, A., C. Babalini, A. Borreca, C. Patrono, R. Massa, S. Basaran, R. P. Munhoz, E. A. Rogaeva, P. H. St George-Hyslop, G. Bernardi, and T. Kawarai, 2010, "SPATACSIN Mutations Cause Autosomal Recessive Juvenile Amyotrophic Lateral Sclerosis", *Brain*, Vol. 133, pp. 591-598.

Otto, E, J. Hoefele, R. Ruf, A. M. Mueller, K. S. Hiller, M. T. F. Wolf, M. J. Schuermann, A. Becker, R. Birkenhager, R. Sudbrak, H. C. Hennies, P. Nurnberg, and F. Hildebrandt, 2002, "A Gene Mutated in Nephronophthisis and Retinitis Pigmentosa Encodes a Novel Protein, Nephroretinin, Conserved in Evolution", *American Journal of Human Genetics*, Vol. 71, pp. 1161-1167. Erratum: 2002, *American Journal of Human Genetics*, Vol. 71, pp. 1489.

Pasinelli, P., and R. H. Brown, 2006, "Molecular Biology of Amyotrophic Lateral Sclerosis: Insights from Genetics", *Nature Reviews Neuroscience*, Vol. 7, No. 9, pp. 710-723.

Pearson, C. G., D. P. Osborn, T. H. Giddings Jr, P. L. Beales, and M. Winey, 2009, "Basal Body Stability and Ciliogenesis Requires the Conserved Component Poc1", *The Journal of Cell Biology*, Vol. 187, No. 6, pp. 905-920.

Pierce-Shimomura, J. T., B. L. Chen, J. J. Mun, R. Ho, R. Sarkis, and S. L. McIntire, 2008, "Genetic Analysis of Crawling and Swimming Locomotory Patterns in C. elegans", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, No. 52, pp. 20982-20987.

Ramamurthy, V., M. Roberts, F. van den Akker, G. Niemi, T. A. Reh, and J. B. Hurley, 2003, "AIPL1, a Protein Implicated in Leber's Congenital Amaurosis, Interacts with and Aids in Processing of Farnesylated Proteins", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 100, No. 22, pp. 12630-12635.

Renieri, A., F. Mari, M. A. Mencarelli, E. Scala, F. Ariani, I. Longo, I. Meloni, G. Cevenini, G. Pini, G. Hayek, and M. Zappella, 2009, "Diagnostic Criteria for the Zappella Variant of Rett Syndrome (the Preserved Speech Variant)", *Brain & Development*, Vol. 31, No. 3, pp. 208-216.

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

Rothstein, J. D, 2009, "Current Hypotheses for the Underlying Biology of Amyotrophic Lateral Sclerosis", *Annals of Neurology*, Vol. 65, pp. S3-S9.

Russo, S., M. Marchi, F. Cogliati, M. T. Bonati, M. Pintaudi, E. Veneselli, V. Saletti, M. Balestrini, B. Ben-Zeev, and L. Larizza, 2009, "Novel Mutations in the CDKL5 Gene, Predicted Effects and Associated Phenotypes", *Neurogenetics*, Vol. 10, No. 3, pp. 241-250.

Sabatelli, M., F. Madia, A. Conte, M. Luigetti, M. Zollino, I. Mancuso, M. Lo Monaco, G. Lippi, and P. Tonali, 2008, "Natural History of Young-Adult Amyotrophic Lateral Sclerosis", *Neurology*, Vol. 71, No. 12, pp. 876-881.

Shaikh, T. H., X. Gai, J. C. Perin, J. T. Glessner, H. Xie, K. Murphy, R. O'Hara, T. Casalunovo, L. K. Conlin, M. D'Arcy, E. C. Frackelton, E. A. Geiger, C. Haldeman-Englert, M. Imielinski, C.E. Kim, L. Medne, K. Annaiah, J. P. Bradfield, E. Dabaghyan, A. Eckert, C. C. Onyiah, S. Ostapenko, F. G. Otieno, E. Santa, J. L. Shaner, R. Skraban, R. M. Smith, J. Elia, E. Goldmuntz, N. B. Spinner, E. H. Zackai, R. M. Chiavacci, R. Grundmeier, E. F. Rappaport, S. F. Grant, P. S. White, and H. Hakonarson, 2009, "High-Resolution Mapping and Analysis of Copy Number Variations in the Human Genome: a Data Resource for Clinical and Research Applications", *Genome Research*, Vol. 19, pp. 1682-1690.

Schultz, R. J., D. G. Glaze, K. J. Motil, D. D. Armstrong, D. J. del Junco, C. R. Hubbard, and A. K. Percy, 1993, "The Pattern of Growth Failure in Rett Syndrome", *American Journal of Diseases of Children*, Vol. 147, No. 6, pp. 633-637.

Scala, E., F. Ariani, F. Mari, R. Caselli, C. Pescucci, I. Longo, I. Meloni, D. Giachino, M. Bruttini, G. Hayek, M. Zappella, and A. Renieri, 2005, "CDKL5/STK9 is Mutated in Rett Syndrome Variant with Infantile Spasms", *Journal of Medical Genetics*, Vol. 42, No. 2, pp. 103-107.

Schwindt, T. T., C. A. Trujillo, P. D. Negraes, C. Lameu, and H. Ulrich, 2011, "Directed Differentiation of Neural Progenitors into Neurons is Accompanied by Altered Expression of P2X Purinergic Receptors", *Journal of Molecular Neuroscience*, Vol. 44, No. 3, pp. 141-146.

Singh S., C. Brocker, V. Koppaka, Y. Chen, B. C. Jackson, A. Matsumoto, D. C. Thompson, and V. Vasiliou, 2013, "Aldehyde Dehydrogenases in Cellular Responses to Oxidative/Electrophilic Stress", *Free Radical Biology & Medicine*, Vol. 56, pp. 89-101.

Sobel, E., and K. Lange, 1996, "Descent Graphs in Pedigree Analysis: Applications to Haplotyping, Location Scores, and Marker-Sharing Statistics", *American Journal of Human Genetics*, Vol. 58, No. 6, pp. 1323-1337.

Sohocki, M. M., S. J. Bowne, L. S. Sullivan, S. Blackshaw, C. L. Cepko, A. M. Payne, S. S. Bhattachary, S. Khaliq, S. Qasim Mehdi, D. G. Birch, W. R. Harrison, F. F. Elder, J. R. Heckenlively, and S. P. Daiger, 2000, "Mutations in a New Photoreceptor-Pineal Gene on 17p Cause Leber Congenital Amaurosis", *Nature Genetics*, Vol. 24, No. 1, pp. 79-83.

Sreedharan, J., and R. H. Brown Jr, 2013, "Amyotrophic Lateral Sclerosis: Problems and Prospects", *Annals of Neurology*, Vol. 74, No. 3, pp. 309-316.

Stagos, D., Y. Chen, C. Brocker, E. Donald, B. C. Jackson, D. J. Orlicky, D. C. Thompson, and V. Vasiliou, 2010, "Aldehyde Dehydrogenase 1B1: Molecular Cloning and Characterization of a Novel Mitochondrial Acetaldehyde-Metabolizing Enzyme", *Drug Metabolism and Disposition*, Vol. 38 pp. 1679-1687.

Stenson, P. D., E. V. Ball, K. Howells, A. D. Phillips, M. Mort, and D. N. Cooper, 2009, "The Human Gene Mutation Database: Providing a Comprehensive Central Mutation Database for Molecular Diagnostics and Personalized Ggenomics", *Human Genomics*, Vol. 4, No. 2, pp. 69-72.

Stirling, A. J., D. Howel, P. A. Millner, S. Sadiq, D. Sharples, and R. A. Dickson, 1996, "Late-Onset Idiopathic Scoliosis in Children Six to Fourteen Years Old. A Cross-Sectional Prevalence Study", *The Journal of Bone and Joint Surgery*, Vol. 78, No. 9, pp. 1330-1336.

Tonelli, A., R. Romaniello, R. Grasso, A. Cavallini, A. Righini, N. Bresolin, R. Borgatti, and M. T. Bassi, 2010, "Novel Splice-Site Mutations and a Large Intragenic Deletion in PLA2G6 Associated with a Severe and Rapidly Progressive Form of Infantile Neuroaxonal Dystrophy", *Clinical Genetics*, Vol. 78, pp. 432-440.

Traboulsi, E. I., 2012, “Cone Dysfunction Syndrome, Cone Dystrophies, and Cone-Rod Degenerations”, *Genetic Diseases of the Eye*, Oxford University Press, New York, NY, USA, pp. 410-420.

Ugur, S.A., A. Tolun, 2008, “Homozygous WNT10b Mutation and Complex Inheritance in Split-Hand/Foot Malformation”, *Human Molecular Genetics*, Vol. 17, No. 17, pp. 2644-2653.

Urano, T., H., Nishimori, H. Han, T. Furuhata, Y. Kimura, Y. Nakamura, and T. Tokino, 1997, “Cloning of P2XM, a Novel Human P2X Receptor Gene Regulated by p53”, *Cancer Research*, Vol. 57, No. 15, pp. 3281-3287.

Vasiliou, V., D. C. Thompson, C. Smith, M. Fujita, and Y. Chen, 2013, “Aldehyde Dehydrogenases: from Eye Crystallins to Metabolic Disease and Cancer Stem Cells”, *Chemico-biological Interactions*, Vol. 202, pp. 2-10.

Victor, M., and A. Ropper, 2005, “Diseases of the Peripheral Nerves”, *Principles of Neurology*, McGraw-Hill, New York, NY, USA, pp. 939-945.

van der Spuy, J., J. P. Chapple, B. J. Clark, P. J. Luthert, C. S. Sethi, and M. E. Cheetham, 2002, “The Leber Congenital Amaurosis Gene Product AIPL1 is Localized Exclusively in Rod Receptors of the Adult Human Retina”, *Human Molecular Genetics*, Vol. 11, No. 7, pp. 823-831.

Votruba, M., F. W. Fitzke, G. E. Holder, A. Carter, S. S. Bhattacharya, and A. T. Moore, 1998, “Clinical Features in Affected Individuals from 21 Pedigrees with Dominant Optic Atrophy”, *Archives of Ophthalmology*, Vol. 116, No. 3, pp. 351-358.

Voulgaridou, G. P., I. Anestopoulos, R. Franco, M. I. Panayiotidis, and A. Pappa, 2011, “DNA Damage Induced by Endogenous Aldehydes: Current State of Knowledge”, *Mutation Research*, Vol. 711, pp. 13-27.

Wang, W., L. Li, W. L. Lin, D. W. Dickson, L. Petrucelli, T. Zhang, and X. Wang, 2013, "The ALS Disease-Associated Mutant TDP-43 Impairs Mitochondrial Dynamics and Function in Motor Neurons", *Human Molecular Genetics*, Vol. 22, pp. 4706-4719.

Weber, J. L., and K. W. Broman, 2001, "Genotyping for Human Whole-Genome Scans: Past, Present, and Future", *Advances in Genetics*, Vol. 42, No. 1, pp. 77-96.

Weng, S. M., M. E. Bailey, and S. R. Cobb, 2011, "Rett Syndrome: from Bed to Bench", *Pediatrics and Neonatology*, Vol. 52, No. 6, pp. 309-316.

Wiszniewski, W., R. A. Lewis, D. W. Stockton, J. Peng, G. Mardon, R. Chen, and J. R. Lupski, 2011, "Potential Involvement of More than One Locus in Trait Manifestation for Individuals with Leber Congenital Amaurosis", *Human Genetics*, Vol. 129, No. 3, pp. 319-327.

Wright, A. F., C. F. Chakarova, M. M. Abd El-Aziz, and S. S. Bhattacharya, 2010, "Photoreceptor Degeneration: Genetic and Mechanistic Dissection of a Complex Trait", *Nature Reviews Genetics*, Vol. 11, pp. 273-284.

Yang, Y., A. Hentati, H. X. Deng, O. Dabbagh, T. Sasaki, M. Hirano, W. Y. Hung, K. Ouahchi, J. Yan, A. C. Azim, N. Cole, G. Gascon, A. Yagmour, M. Ben-Hamida, M. Pericak-Vance, F. Hentati, and T. Siddique, 2001, "The Gene Encoding Alsin, a Protein with Three Guanine-Nucleotide Exchange Factor Domains, is Mutated in a Form of Recessive Amyotrophic Lateral Sclerosis", *Nature Genetics*, Vol. 29, No. 2, pp. 160-165.

Yu-Wai-Man, P., G. Griffiths, G. Hudson, and P. F. Chinnery, 2009, "Inherited Mitochondrial Optic Neuropathies", *Journal of Medical Genetics*, Vol. 46, No. 3, pp. 145-158.

Yu-Wai-Man, P., G. Griffiths, G. S. Gorman, C. M. Lourenco, A. F. Wright, M. Auer-Grumbach, A. Toscano, O. Musumeci, M. L. Valentino, L. Caporali, C. Lamperti, C.

M. Tallaksen, P. Duffey, J. Miller, R. G. Whittaker, M. R. Baker, M. J. Jackson, M. P. Clarke, B. Dhillon, B. Czermin, J. D. Stewart, G. Hudson, P. Reynier, D. Bonneau, W. Marques Jr, G. Lenaers, R. McFarland, R. W. Taylor, D. M. Turnbull, M. Votruba, M. Zeviani, V. Carelli, L. A. Bindoff, R. Horvath, P. Amati-Bonneau, and P. F. Chinnery, 2010, "Multi-System Neurological Disease is Common in Patients with OPA1 Mutations", *Brain*, Vol. 133, pp. 771-786.

Zeev, B. B., Y. Yaron, N. C. Schanen, H. Wolf, N. Brandt, N. Ginot, R. Shomrat, and A. Orr-Urtreger, 2002, "Rett Syndrome: Clinical Manifestations in Males with MECP2 Mutations", *Journal of Child Neurology*, Vol. 17, No. 1, pp. 20-24.