DISEASE GENE IDENTIFICATION VIA LINKAGE ANALYSIS AND EXOME SEQUENCING

by

Ayşe Güven

B.S., Molecular Biology and Genetics, Bilkent University, 2009

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

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2012

DISEASE GENE IDENTIFICATION VIA

LINKAGE ANALYSIS AND EXOME SEQUENCING

APPROVED BY:

Prof. Aslıhan Tolun .....

(Thesis Supervisor)

Prof. Nazlı Başak .....

Prof. Uğur Özbek .....

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis supervisor, Prof. Aslihan Tolun, for her guidance and encouragement throughout the study, which enabled me to gain valuable skills and a broader perspective towards scientific questions. I also extend my appreciation to the members of my thesis committee, Prof. Nazlı Başak and Prof. Uğur Özbek for allocating their time to evaluate this work.

I would like to thank all previous and present members of Kommagene laboratory, especially Çiğdem Köroğlu for her help in lab work and her friendship, Yeşerin Yıldırım, Murat Çetinkaya, Sibel Uğur, Özgecan Ayhan, Atalay Tok and Sara Suna Yücel for everything. I would like to extend my thanks to all my friends in the department, among whom I would like to mention Murat Atasoy, Nehir Banaz, Sercan Sayın, Burçin Duan, Emine Dindar, Damla Kaptan, İbrahim Taştekin and Sena Ağım for their support and great friendship.

I would also like to thank Prof. Murat Günel and members of his team at Yale University for their help in CNV analyses and exome sequencing. I also thank Erşan Taşan for his contribution in computational analyses.

I would like to convey my great appreciation to all members of my family. I am especially grateful for the endless support and love of my parents Huriye Şen Güven and Rahmi Güven. I also sincerely thank Tuğrul Cem Bıçak, for his love and friendship that kept me motivated throughout this period.

This work was supported by the Scientific and Technological Research Council of Turkey (110T252) and Bogazici University Research Fund (BAP 5708).

ABSTRACT

DISEASE GENE IDENTIFICATION VIA LINKAGE ANALYSIS AND EXOME SEQUENCING

Understanding the genetic basis of hereditary diseases elucidates the molecular mechanisms underlying disease manifestations and uncovers valuable information about gene function. Consanguineous families are very useful for mapping hereditary disease loci. In addition, recent advances in the technologies of genotyping arrays and next generation sequencing as well as their decreasing costs made disease gene search easier. In this thesis, three families afflicted with different recessive hereditary diseases, autosomal recessive ataxia, azoospermia and progressive myoclonic epilepsy with dystonia (PMED), were studied. Ataxia is a neurological disorder, mainly characterized by lack of coordination in muscle movements. A consanguineous family with two brothers afflicted by ataxia, associated with spasticity, leukodystrophy, oculomotor apraxia and hearing loss, was studied. Four candidate loci were found, novel variants at the loci were investigated via exome sequencing, and the disease gene was identified at 3q21.1-22.1. Azoospermia is the condition in males where there is no detectable spermatozoon in the semen. Two candidate loci at 4p16.2-p16.1 and 18q12.1 were found, the loci were examined via exome sequencing, and the disease gene was identified at 18q12.1. PMED is a novel form of progressive myoclonic epilepsy, similar to FIME (MIM 605021) but with a much more severe clinical course. The disease locus is at 16pter-p13.3. A 2-bp deletion in exon 3 of *TBC1D24*, c.969_970delGT (p.S324TfsX3), was identified. The mutation resulted in frameshift and was predicted to lead to a premature termination codon. The mutation was predicted to affect only one isoform and relative expression of different isoforms was assessed in seven different human brain regions. Meanwhile, two novel transcript isoforms, both harboring exon 3, were identified.

ÖZET

BAĞLANTI ANALİZİ VE EKSOM DİZİLEME YÖNTEMLERİYLE HASTALIK GENLERİNİN BULUNMASI

Kalıtsal hastalıkların genetik temellerinin ortaya çıkarılması hastalık belirtilerinin altında yatan moleküler mekanizmaları aydınlatır ve genlerin işlevleriyle ilgili bilgi verir. Akraba evliliği yapmış aileler hastalık genlerinin haritalanmasında çok yararlıdır. Ayrıca, genotipleme ve yeni nesil dizileme teknolojilerinde son yıllarda meydana gelen gelişmeler ve bu teknolojilerin maliyetlerinin düşmesi hastalık genleri aramayı oldukça kolaylaştırmıştır. Bu tezde, yeni hastalık genleri belirlenmesi amacıyla, farklı çekinik kalıtımsal hastalık taşıyan üç aile çalışıldı: ataksi, azospermia ve distonya ile seyreden ilerleyici (progresif) miyoklonik epilepsi (progressive myoclonic epilepsy with dystonia, PMED). Ataksi kas hareketlerinde koordinasyon bozukluğu ile tanımlanan bir nörolojik hastalıktır. Bu tezde çalışılan ailede anne-baba akrabadır ve iki hasta erkek kardeş ataksi, spasitisite, lökodistrofi, okülomotor apraksi ve işitme kaybı sergilemektedir. Analizler sonucunda dört aday bölge belirlendi, bölgelerdeki yeni varyantlar eksom dizileme ile araştırıldı ve hastalık geni 3q21.1-22.1'de bulundu. Azospermi erkeklerde semende gözlenebilir spermatozoa bulunmaması durumudur. Bu çalışmada 4p16.2-p16.1 ve 18q12.1'de iki aday bölge bulundu. Bölgelerin eksom dizileme ile incelenmesi sonucunda hastalık geninin 18q12.1'de olduğu saptandı. PMED ailesel miyoklonik çocukluk epilepsisine (FIME, MIM 605021) benzeyen ancak çok daha hızlı ilerleyen yeni bir progresif miyoklonik epilepsi hastalığıdır. Bu çalışmadaki ailede hastalık geninin 16pter-p13.3'te olduğu biliniyordu. *TBC1D24* geni ekson 3'te 2 nükleotidlik bir silinme (c.969_970delGT, p.S324TfsX3) bulundu. Bu silinme çerçeve kaymasına sebep olduğundan, biraz aşağısında bir stop kodonu oluştuğu yordandı. Bu mutasyon yalnızca bir transkript izoformunu etkilediğinden izoformların insan beyninin yedi ayrı bölgesindeki göreli anlatımları belirlendi. Bu çalışma sırasında ekson 3'ü içeren iki yeni transkript izoform daha bulundu.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AOA	Ataxia-Oculomotor Apraxia
APS	Ammonium peroxodisulphate
AZF	Azoospermia factor
bp	Base pair
BAM	Binary Alignment/Map
BWA	Burrows-Wheeler aligner
cDNA	Complementary deoxyribonucleic acid
Chr	Chromosome
cM	Centimorgan
CNV	Copy number variation
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetate
FIME	Familial Infantile Myoclonic Epilepsy
FRDA	Friedreich's Ataxia
HRM	High Resolution Melting
IBD	Identical by descent
kb	Kilobase pair
LOD	Logarithm of odds
MAF	Minor Allele Frequency
MERRF	Myoclonic Epilepsy with Ragged-Red Fibers
min	Minute
NCL	Neuronal Ceroid Lipofuscinoses
NGS	Next-Generation Sequencing
NOA	Non-Obstructive Azoospermia
OA	Obstructive Azoospermia
PAR	Pseudoautosomal region
PCR	Polymerase Chain Reaction
PMED	Progressive Myoclonic Epilepsy with Dystonia
RNA	Ribonucleic acid

rpm	Revolution per minute
RT-PCR	Reverse-Transcription Polymerase Chain Reaction
SAM	Sequence Alignment/Map
SDS	Sodium dodecyl sulphate
sec	second
SEM	Standard Error of the Mean
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SSCP	Single Strand Conformation Polymorphism
TEMED	N, N, N, N'-Tetramethylethylenediamine
UTR	Untranslated region

1. INTRODUCTION

Three families afflicted with different recessive heritable diseases were studied within the framework of this thesis: ataxia, azoospermia and progressive myoclonic epilepsy with dystonia (PMED).

1.1 Ataxia

Ataxia (from Greek, meaning lack of order) is a clinical condition where affected individuals lack coordination of muscle movements. Ataxia is considered as a sign of impairment in nervous system, mostly indicating dysfunction of the cerebellum. Hereditary ataxias are a group of genetic disorders characterized by slowly progressive lack of coordination of gait accompanied by impairment in speech and hand and eye movements. Autosomal recessive ataxias constitute a subgroup of hereditary ataxias and exhibit a high degree of phenotypic variability. Typical symptoms include balance abnormalities, kinetic and postural tremor and dysarthria (speech impairments); however, peripheral neuropathy, ophthalmological abnormalities and non-neurological signs may also be present (Fogel and Perlman, 2007).

Friedreich's ataxia (FRDA, MIM 229300) is the most common hereditary ataxia with an estimated frequency of 1:30,000 to 1:50,000 (Pandolfo, 2008). Disease onset is usually first or second decade of life. Characteristic clinical symptoms of FRDA are gait instability, lack of deep tendon reflexes, Babinski response and loss of position and vibration senses. Recessive mutations in *Frataxin* (*FXN*) at 9q13-21.1 are the cause of Friedreich's ataxia, and the most common mutation is a GAA expansion which occurs in intron 1 of the gene (Al-Mahdawi *et al.*, 2006). *FXN* mutations impair the mitochondrial iron metabolism, leading to increased oxidative damage (Pandolfo, 2009).

Ataxia Telangiectasia (AT, MIM 208900) is a progressive autosomal recessive disorder caused by mutations in *ATM* at 11q22-23 (Savitsky *et al.,* 1995). Clinical characteristics include cerebellar ataxia, telangiectasia (formation of small, dilated blood vessels) in various body parts, immunodeficiency and increased sensitivity to ionizing radiation which causes a very high risk of developing cancer, principally lymphoid tumors (Taylor and Byrd, 2005). The disease onset is before 3 years of age, and life expectancy is poor. Atm protein is an important controller of cell cycle checkpoint pathway and regulates numerous proteins involved in cell proliferation and DNA damage repair (Shiloh, 2003).

Ataxia-oculomotor apraxia (AOA) is another autosomal recessive ataxia. It has two subtypes: ataxia, early-onset, with oculomotor apraxia and hypoalbuminemia (EAOH, AOA1, MIM 208920) and spinocerebellar ataxia, autosomal recessive 1 (SCAR1, AOA2, MIM 606002). The clinical manifestations of AOAs show high resemblance to ataxia telangiectasia, but have additional signs such as oculomotor apraxia (limitation of ocular movements on command) and peripheral neuropathy. AOA1 is an early onset slowly progressive cerebellar ataxia, coupled with peripheral neuropathy, oculomotor apraxia and hypoalbuminemia (Aicardi *et al.,* 1988). Mutations in *Aprataxin* (*APTX)* at 9p21.1, encoding a member of the histidine triad (HIT) superfamily, are found to be responsible for AOA1 (Date *et al.,* 2001, Moreira *et al.,* 2001). Aprataxin protein has been reported to interact with poly(ADP-ribose) polymerase-1 (PARP1), which has a key role in detecting single-strand DNA breaks (Harris *et al.,* 2009). AOA2 is a clinically heterogeneous form of autosomal recessive cerebellar ataxia with peripheral neuropathy and occasionally oculomotor apraxia. *SETX* gene at 9q34.13 is responsible for AOA2 and encodes a nuclear protein involved in the defense against oxidative DNA damage (Moreira *et al.,* 2004, Suraweera *et al.,* 2007).

Rare recessive ataxias include ataxia with vitamin E deficiency (AVED, MIM 277460) with clinical symptoms that are very similar to Friedreich's ataxia, including progressive trunk and limb ataxia, loss of deep tendon reflexes, vibratory-sense disturbances, dysarthria, muscle weakness and Babinski sign (Fogel, 2007). Mutations in *TPPA*, encoding alpha-tocopherol transfer protein, are reported to cause AVED (Ouahchi *et al.,* 1995). *TPPA* has a role in transport of alpha tocopherol (the most active vitamin E

isomer) to very-low-density lipoproteins (VLDLs) in the liver, which are released to the blood circulation. Low levels of this protein causes tissue degeneration, presumably due to the oxidative stress caused by low VLDLs (Di Donato *et al.*, 2010). Two very rare autosomal recessive hereditary ataxias Refsum disease (MIM 266500) and Marinesco-Sjögren syndrome (MSS, MIM 248800) have additional clinical features of retinitis pigmentosa and peripheral neuropathy in the former and intellectual disability, cataract, short stature and hypotonia in the latter. Mutations in *PHYH* at 10p13 or *PEX7* at 6q23.3 were reported to cause Refsum disease (Mihalik *et al.*, 1997, Braverman *et al.*, 2002), whereas *SIL1* at 5q31.2 is responsible for Marinesco-Sjögren syndrome (Anttonen *et al.*, 2005). Another rare hereditary ataxia is autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS, MIM 270550). ARSACS is an early onset neurodegenerative disease characterized by progressive spastic ataxia, dysarthria, distal muscle wasting, nystagmus and defects in ocular movements (Bouchard *et al.*, 1978). Engert et al. (2000) showed that *Sacsin* (*SACS*) at 13q12 was mutated in patients with ARSACS. The mechanism of the pathogenesis of *SACS* mutations is not known, but Sacsin protein has been shown to interact with Ataxin-1 and to act as a chaperone in protein folding (Parfitt *et al.*, 2009).

Coenzyme Q10 deficiency (MIM 607426) is a rare autosomal recessive disorder with very heterogeneous clinical signs, comprising of encephalomyopathic, cardiomyopathic and cerebellar impairments. *CABC1* at 1q42.13, *COQ2* at 4q21.23, *COQ9* at 16q21, *PDSS1* at 10p12.1 and *PDSS2* at 6q21 were shown to be mutated in the patients (Quinzii and Hirano, 2011).

Niemann-Pick disease type C (NPC1, MIM 257220; NPC2, MIM 607625) is an autosomal recessive lipid storage disorder. Major clinical symptoms are progressive neurodegeneration, ataxia, grand mal seizures, loss of previously learned speech and spasticity and myoclonic jerks. However, the clinical spectrum is very broad, including a neonatal onset disorder that rapidly leads to a fatal outcome and an adult-onset chronic neurodegenerative disease (reviewed in Vanier, 2010). Most of the cases are caused by mutations in *NPC1* at 18q11.2 (Carstea *et al.*, 1997), but mutations in *NPC2* at 14q24.3 are also known to cause the disease. (Naureckiene *et al.*, 2000). NPC1 and NPC2 proteins have roles in cholesterol trafficking, and loss of function of either of them causes an

accumulation of cholesterol in late endosomes/lysosomes, leading to neurodegeneration (Peake and Vanje, 2010).

The patients in the ataxia family investigated in this study had cerebellar ataxia associated with spasticity, leukodystrophy, oculomotor apraxia and hearing loss. In the light of these symptoms, ARSACS, Coenzyme Q10 deficiency and Niemann-Pick disease type C were considered as possible diagnoses. In the framework of this study, all suggested disorders were excluded by linkage analysis, and subsequently the locus and the novel disease gene were identified. The gene is in the same metabolic pathway as the gene responsible for Niemann-Pick disease type C.

1.2. Azoospermia

Azoospermia is a condition where a male does not have a measurable level of spermatozoa in his semen. Azoospermia is categorized as obstructive azoospermia (OA) where ejaculatory pathway is blocked and non-obstructive azoospermia (NOA) in which spermatogenic failure occurs. Depending on the etiology, azoospermia can also be classified as pretesticular, testicular and posttesticular. Pretesticular azoospermia is typically characterized by abnormal hormone levels, especially low FSH level leading to insufficient stimulation for sperm production. Testicular azoospermia includes testicular abnormalities and impairment in spermatogenesis mainly arising from genetic defects. In posttesticular azoospermia, sperm production is normal; however, ejaculation is defective and generally the cause is an obstruction in the ejaculatory duct.

OA may occur due to genetic or external factors. Vasectomy, trauma and infections can lead to blockage in vas deferens, epididymis and ejaculatory ducts. The main genetic cause of OA is congenital bilateral aplasia of the vas deferens (CBAVD, MIM 277180), which may occur as a manifestation of cystic fibrosis or in isolation. *CFTR* gene at 7q31.2 is mutated in 60-90 per cent of patients with CBAVD (O'Flynn O'Brien *et al.*, 2010).

NOA can be associated with genetic defects or hormonal abnormalities. Genetic defects underlie 29 per cent of azoospermia cases and include chromosomal abnormalities, mutations in nuclear or mitochondrial genes and epigenetic alterations (Dohle *et al.,* 2002). Robertsonian translocations and Klinefelter syndrome represent the most common chromosomal defects in azoospermia (Chandley, 1979). Another chromosomal abnormality, 46,XX male syndrome, also causes azoospermia. Individuals affected by 46,XX male syndrome have sex-determining region Y (SRY) translocated to X chromosome or to an autosome, but they lack the AZF region, and this leads to azoospermia (Oates, 2008).

AZF region resides at Yq11 and contains three subregions, namely, AZFa, AZFb and AZFc, commonly deleted in azoospermic males (Vogt *et al.,* 1996). Microdeletions in these regions give rise to a wide range of infertile phenotypes, and AZFc is the most frequently deleted region, due to its distal location and larger size (Yen, 1998). AZFc harbors *USP9Y,* and both point mutations or single gene deletions of *USP9Y* were shown to cause either azoospermia or severe oligospermia, a condition where sperm concentration in the semen is low (Sun *et al.*, 1999).

Genetic alterations in genes related to hormonal regulation also were shown to cause NOA. Mutations in *AR* at Xq12 cause androgen insensitivity syndrome (Ferlin *et al.*, 2007). Polymorphisms in *SHBG* (sex-hormone-binding globulin) at 17p13.1, *ESR1* at 6q25.1 and *FSHR* at 2p16.3 are associated with defects in spermatogenesis (Lazaros *et al.*, 2008, Guarducci *et al.,* 2006, Tapanainen *et al.,* 1997).

Spermatogenesis occurs in several steps that are tightly regulated. Renewal and differentiation of spermatogonial stem cells give rise to rapidly proliferating spermatogonia, spermatocytes (meiotic sperm cells) and spermatids (haploid, round, elongating cells) that mature into spermatozoon released into seminiferous tubules. A large number of genes are specifically expressed in the male germline, and their functions may account for the complexity of spermatogenic process (Kimmins *et al.,* 2004, Matzuk and Lamb, 2008). Heterozygous mutations in various male germline specific genes were

reported to underlie many idiopathic infertility cases. Such mutations in *SYCP3* at 12q23.2, encoding a component of the synaptonemal complex, are associated with spermatogenic failure (SPGF4, MIM 270960) (Miyamoto *et al.*, 2003). Mutations in *NR5A1* at 9q33.3, encoding a transcriptional activator involved in sex determination, are reported in patients with idiopathic, severe spermatogenic failure (SPGF8, MIM 613957) (Bashamboo *et al.*, 2010). Mutations in *AURKC* at 19q13.43 (SPGF5, MIM 243060), *SPATA16* at 3q26.31 (SPGF6, MIM 102530), *CATSPER* at 11q13.1 (SPGF7, MIM 612997) and *DPY19L2* at 12q14.2 (SPGF9, MIM 613958) are shown to cause impairments in sperm motility, acrosome malformations or sperm polyploidy.

Studies on mice models contributed substantially to the identification of genes involved in spermatogenesis. Testis specific *Tex11* in mice was shown to promote the maintenance of synapsis and the formation of crossovers (Yang *et al.*, 2008). Polymorphisms in human ortholog *TEX11* at Xq13.1 were also associated with oligozoospermia and azoospermia (Aston *et al.*, 2010). Another X-linked gene *Taf7l* also showed testis specific expression in mice, and sequence alterations in human ortholog *TAF7L* at Xq22.1 were suggested as risk factors for spermatogenic failure (Akinloye *et al.*, 2007). Likewise, knocking out the testis specific *Msy2* in mice was reported to cause infertility (Yang *et al.*, 2005), and sequence variations in human ortholog *YBX2* at 17p13.1 were suggested to contribute to idiopathic spermatogenic failure (Deng *et al.*, 2008, Hammoud *et al.*, 2009).

To date, no autosomal recessive locus or gene has been associated with spermatogenic failure. In this thesis, a family afflicted with autosomal recessive azoospermia, with normal karyotype and no Y-microdeletion was studied. The disease locus is mapped with linkage analysis and the gene responsible is identified.

1.3. Progressive Myoclonic Epilepsy with Dystonia (PMED)

Epilepsies comprise one of the most common and broad groups of neurological disorders, mainly characterized by recurrent and unprovoked seizures. It affects one per

cent of the world population and 0.5 % of the children (Cowan, 2002, Poduri and Lowenstein, 2011). According to clinical classifications, epilepsy can be divided into symptomatic, presumed symptomatic and idiopathic epilepsies (Poduri and Lowenstein, 2011). Idiopathic epilepsies affect approximately 0.4 % of the general population, and genetic factors are thought to contribute greatly to its etiology (Poduri and Lowenstein, 2011).

Progressive myoclonic epilepsies (PMEs) are a group of symptomatic (with evidence of brain lesions) generalized epilepsies. Their main features are myoclonic seizures, tonic-clonic seizures and progressive neurological deterioration (Shahwan *et al.*, 2005). In PMEs, clinical manifestations, age of onset and frequency of the seizures are highly variable. Environmental factors and predisposing genetic factors contribute to the etiology of the disease (Elmslie *et al.*, 1996, Liu *et al.*, 1995), but several genes exhibiting Mendelian inheritance are also reported as responsible for the disease. Unverricht-Lundborg disease (epilepsy, progressive myoclonic 1 or ULD, MIM 254800) is an autosomal recessive disorder and results from mutations in *CSTB*, also known as *EPM1*, at 21q22.3. The most common *EPM1* mutation found in ULD patients is the expansion of a dodecamer repeat in the 5' UTR region (Lafrenière *et al.*, 1997). Lafora's disease (MIM 254780) is also one of the common progressive myoclonic epilepsies, characterized by epilepsy, myoclonus and the presence of inclusions called "Lafora bodies" in neurons, heart, skeletal muscles and sweat-gland duct cells. Lafora's disease is an autosomal recessive disorder and around 80 per cent of the patients carry mutations in *EPM2A (Laforin)* gene at 6q24.3 (Minassian *et al.*, 1998). Mutations in *NHLRC1 (Malin)* at 6p22.3 are also reported in the disease (Chan *et al.*, 2003). In addition, *PRICKLE1* (12q12) mutations are associated with progressive myoclonic epilepsy-1B (EPM1B, MIM 612437) and *KCTD7* at 7q11.21 is responsible for EPM3 (epilepsy, progressive myoclonic 3; MIM 611726) (Van Bogaert *et al.*, 2007). Other genetic progressive myoclonic epilepsies include myoclonic epilepsy with ragged red fibers (MERRF, MIM 545000), neuronal ceroid lipofuscinoses (NCLs) and sialidoses. MERRF is caused by sporadic or familial mutations in tRNA gene encoded by maternally inherited mitochondria, namely *MTTK* (Rosing *et al.*, 1985). The clinical characteristics of MERRF are myoclonus, generalized epilepsy and ragged red fibers detected in muscle biopsy (Berkovic *et al.*, 1989). Neuronal

ceroid lipofuscinoses (NCLs) constitute a clinically heterogeneous group of disorders characterized by the accumulation of high amounts of autofluorescent lipopigments in lysosomes. There are five types of NCLs causing PME: classic late infantile NCL (CLN2, MIM 204500) due to mutations in *TPP1* at 11p15, juvenile NCL (CLN3**,** MIM 204200) due to mutations in *CLN3* at 16p11.2, adult NCL (CLN4A, MIM 204300 and CLN4B, 162350) due to mutations in *CLN6* at 15q23 or *DNAJC5* at 20q13.33, late infantile variant NCL (MIM 601780) due to mutations in *CLN6* and late infantile Finnish variant NCL (CLN5, MIM 256731) due to mutations in *CLN5* at 13q22.3. Other genetic defects causing PMEs are two types of sialidoses, which are rare hereditary lysosomal storage diseases leading to abnormal accumulation of complex carbohydrates. Sialidosis type I or cherry-red spot myoclonus syndrome (MIM 256550) is characterized by deficiency of α-neuraminidase, and sialidosis type II (MIM 256540) is caused by deficiency of both α-neuraminidase and β-galactosialidase; mutations in *NEU1* at 6p21.33 and *PPCA* at 20q13.12 are associated with these disorders, respectively (Pshezhetsky *et al.*, 1997, Galjart *et al.*, 1988).

Mutations in genes encoding ion channels and neurotransmitters underlie many of the known genetic epilepsies. Sodium channel subunit genes *SCN1B*, *SCN1A* and *SCN2A*, calcium channel subunit genes *CACNA1A* and *CACNB4*, chloride channel gene *CLCN2*, potassium channel genes *KCNQ2*, *KCNQ3* and *KCNMA1*, and sodium/potassium transporter ATPase *ATP1A2* are examples of genes mutated in generalized idiopathic epilepsies and epileptic encephalopathies. Genes involved in neurotransmission, such as GABA receptor subunit genes *GABRG2*, *GABRA1* and *GABRD*, cholinergic receptors *CHRNA4*, *CHRNA7* and *CHRNB2*, and a syntaxin binding protein *STXBP1* are associated with other various forms of epilepsies.

TBC1D24 is a recently identified recessive epilepsy gene. Three mutations have been reported in two families with myoclonic epilepsy, referred as familial infantile myoclonic epilepsy (FIME, MIM 605021). One of the families was of Italian origin and presented symptoms including myoclonic and generalized tonic-clonic seizures, photosensitivity, and normal neurological and mental development (Zara *et al.*, 2000, de Falco *et al.*, 2001, Falace *et al.*, 2010). The other one was an Arab family, and the clinical

characteristics were focal seizures with prominent eye blinking and facial and limb jerking. The symptoms arose at infancy and persisted throughout life. Convulsive seizures and moderate intellectual disability were also observed (Corbett *et al.,* 2010). In both families, seizures were responsive to medication. The Italian patients were reported as compound heterozygous for two missense mutations whereas the Arab patients were carrying another missense mutation in the homozygous state.

The family investigated in this thesis was described with a novel form of autosomal recessive progressive myoclonic epilepsy (Duru *et al.,* 2010). The disease has neonatal onset and manifests with severe progressive myoclonic epilepsy. The clinical features also included first paroxysmal, later persistent dystonia. Episodic phenomena such as postictal enduring hemipareses, autonomic involvements, and periods of obtundation and lethargy also accompanied the myoclonic seizures. Severe developmental and neurological retardation was observed; and with systemic infections the disease progressed to a full deterioration. All patients died before the age of 8. The disease characteristics were distinct from previously described progressive myoclonic epilepsies; therefore, it was assessed as a novel disease and designated as PMED (Duru *et al.,* 2010). In the initial clinical examinations, MERRF, NCL, sialidosis and type II gangliosidosis seemed plausible diagnoses. Sialidoses and type II gangliosidosis were excluded due to the absence of characteristic symptoms such as cherry-red spots, facial dysmorphism and hepatosplenomegaly. MERRF was considered as unlikely based on the absence of ragged-red fibers in muscle biopsies of the two patients investigated. No accumulation of autofluorescent lipopigments were observed in leukocytes, rendering NCL less feasible as the diagnosis. Finally, the disease was mapped to 16pter-p13.3, which excluded all known genetic PMEs (Duru *et al.,* 2010).

In this thesis, the mutation responsible for PMED was identified in *TBC1D24*. Novel transcript isoforms of the gene also were identified. In addition, expression of different TBC1D24 transcript isoforms in various brain parts was investigated.

1.4 Linkage Analysis

Genetic linkage is the tendency of two loci on the same chromosome to be inherited together. Two loci on the same chromosome that are close enough to each other will segregate together; otherwise, crossing over in meiosis takes place between homologous chromosomes and new combinations of alleles are created. Recombination frequency (θ) represents the frequency of a single crossover between two loci in a meiosis event. The probability of a recombination to take place increases as the distance between the two loci increases. The maximum value for θ is 0.5, which denotes that the two loci are so far apart that they behave as if they are on different chromosomes. The unit of recombination frequency is centimorgan (cM), and one cM indicates one per cent probability that two loci will undergo recombination.

Recombination events can be observed using various types of genetic markers, including restriction fragment length polymorphisms (RFLP), variable number tandem repeats (VNTRs), microsatellites or short tandem repeats (STR) and single nucleotide polymorphisms (SNP). Cost effectiveness of SNP genotyping arrays with high marker density and accessibility of information about markers through online databases facilitate mapping of recombination events and localization of disease genes with high efficiency.

Linkage analysis can be either nonparametric (model-free) or parametric (model-based). Nonparametric linkage analysis is applied to studies on complex traits which are due to interactions of two or more genes and may also be affected by environmental factors. A nonparametric method searches a shared haplotype presumably harboring the disease gene and is identical by descent in affected individuals, regardless of the mode of inheritance. A parametric method uses the genetic information about disease-allele frequency, penetrance and mutation rates to assess the mode of inheritance. The statistical test that is used in parametric linkage analysis is LOD (logarithm of odds) score analysis.

1.4.1 LOD Score Analysis

LOD score analysis discriminates between null hypothesis of no linkage ($\theta = 0.5$) and alternative hypothesis of linkage ($\theta<0.5$) (Morton, 1955). In other words, it compares the likelihood of two loci being actually linked with the likelihood of observing the same event purely by chance. Odds ratio denotes the ratio of the probability of observing the distribution of two traits in a pedigree under linkage at θ, to the probability of observing it under hypothesis of no linkage ($\theta = 0.5$) (Risch, 1992). The logarithm of odds ratio $Z(\theta)$ is calculated for a range of θ values, and the value that gives the maximum lod score is reported as the maximum likelihood estimate. A lod score of 3 is the threshold used in human linkage analysis to accept linkage and indicates 1000 to 1 odds that the observed linkage has not occurred by chance (Terwilliger and Ott, 1994).

LOD score analysis is either two-point or multipoint. Two-point LOD score analysis calculates the cosegregation of each marker with the disease locus, whereas multipoint LOD score analysis calculates cosegregation of each marker with more than two loci at a time and outputs LOD scores for all possible positions on a given map of genetic markers. Several open source softwares for two-point and multipoint LOD score calculations are available for computerized linkage analysis. Different softwares require different commands and input formats, and easyLINKAGE package converts all input information into a standard form to create a more user friendly platform (Hoffman and Lindner, 2005).

1.4.2 Homozygosity Mapping

Homozygosity mapping is based on the fact that affected individuals in a consanguineous family or inbred population are likely to receive a shared haplotype that has descended from a common ancestor, in the homozygous state (Lander and Botstein, 1987). This method is very efficient for mapping autosomal recessive diseases in consanguineous families. Usually the frequency of such a disease is so low that it can only arise from inbreeding, and affected individuals in the family are homozygous by descent.

This method is also part of the parametric linkage analysis and is widely used in gene localization studies in consanguineous families with two or more affected sibs.

1.5 Next Generation Sequencing

In the past few years, next generation sequencing (NGS) technologies (also known as massively parallel sequencing), have improved tremendously, due to the increasing demand to obtain faster, cheaper and more accurate genome information. Traditional Sanger sequencing relies on DNA amplification, and final sequence obtained represents a population of DNA templates. NGS, on the other hand, has the power to sequence single DNA molecules. It generates very short reads (40-400 bp) but in billions from a human genome template in a very short time (Pettersson *et al.,* 2009). This increase in throughput pays off with the accuracy of each short read generated, which is very low compared to Sanger sequencing. Therefore, multiple reads are generated for the same region, a concept called sequencing depth, with the aim of obtaining more reliable sequence information. The more deeply covered a region is, the more accurate the sequence will be. Although deep coverage uncovers most of the information on a region, validation via Sanger sequencing of all variants reported in the NGS results is necessary.

NGS technologies have a very broad field of application. In addition to sequencing of a whole genome, NGS platforms can analyze whole transcriptomes (RNA-seq) and genome-wide profiling of epigenetic marks (methyl-seq) and can perform high throughput chromatin immunoprecipitation assays (ChIP-seq) (Wold and Myers, 2008). For sequencing of genomic DNA, the region to be analyzed can be restricted as in exome sequencing by custom designed targeted capture sequencing. Exome sequencing is a strategy in which coding sequences of a genome are selectively sequenced. Exons and flanking regions as well as UTR sequences are captured with probes in either liquid or solid phases and subjected to next generation sequencing platforms. Since approximately 85 per cent of the disease-causing mutations reside in protein coding regions, exome sequencing provides a very powerful technique to identify disease genes responsible for Mendelian disorders (Choi *et al.,* 2009).

Alignment of exome sequencing reads to the reference genome and subsequent variant calling generates a high amount of variants differing from reference genome. In an average Caucasian exome 5000-10000 of missense and nonsense variants leading to protein truncation are estimated to be present, and majority of those are benign (Ng *et al.*, 2008, Ng *et al.*, 2009). In addition, nearly 400-700 of these variants are not included in databases such as dbSNP, HapMap and 1000 genomes (Ng *et al.*, 2008, Ng *et al.*, 2009, Lalonde *et al.*, 2010, Ng *et al.*, 2010). Therefore, several filters need to be applied in the evaluation of the candidate variants. At this point, linkage analysis provides great help by assigning certain regions as candidate loci. In addition, the candidate gene should be involved in a pathway related to the pathogenesis of the disease and should be expressed in the relevant tissues. The overview of approaches used to confer candidate variants generated by exome sequencing data are given at Figure 1.1.

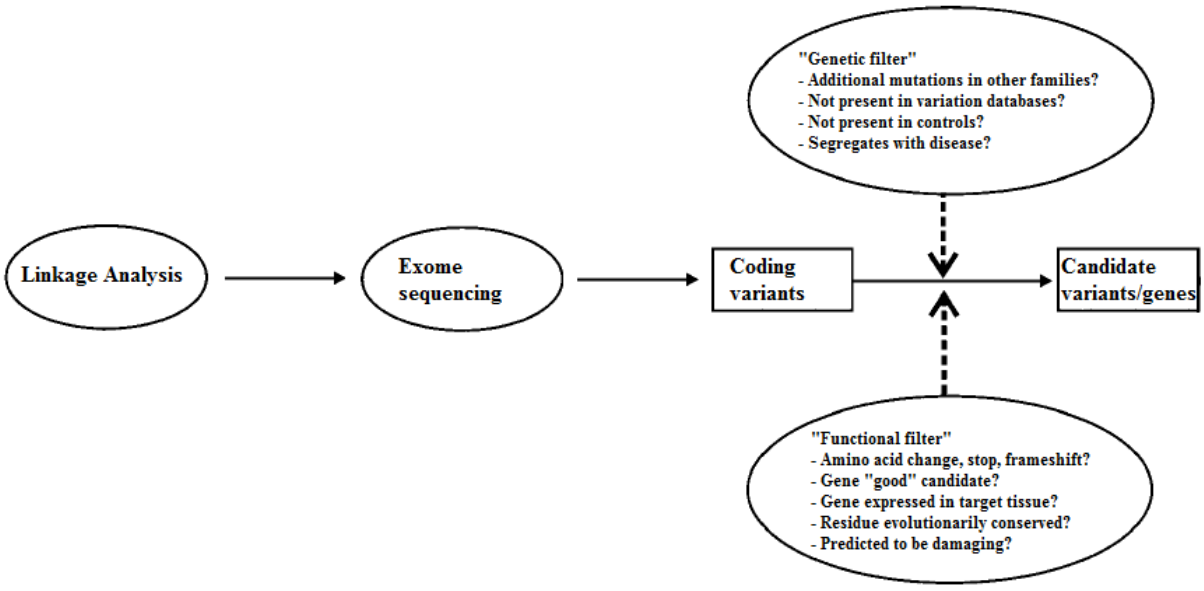


Figure 1.1. The overview of strategy used in identifying candidate variants.

(Modified from Kuhlenbäumer *et al.*, 2011).

1.6 Gene Expression Assays

Gene expression is the synthesis of a functional gene product from a genetic code. The expression of a gene is regulated by several intrinsic and extrinsic factors in order to strictly control its tissue specificity, timing and amount. There are various approaches to analyze the expression of a gene at the RNA level, including northern blotting, real-time quantitative PCR and high throughput analyses such as microarrays, SAGE (serial analysis of gene expression) and RNA-seq. Online databases such as UniGene (NCBI) and GeneSorter (UCSC Genome Browser) store information on expression of genes by tissues, age and health group,. These repositories provide a large collection of expression data for numerous genes; however, they rely on experimental results under specific conditions. Therefore, when designing an experiment on the expression of a gene, it is beneficial to validate the reported expression of the gene under the conditions designed, since the expression may depend on specific temporal and spatial factors.

Real-time quantitative PCR measures the abundance of specific transcripts in a sample. Prior to the assay, all RNA transcripts (mRNA or total RNA) are converted to cDNA by reverse transcription, and then cDNAs of interest are specifically amplified with real-time PCR. Fluorescence emitted by DNA intercalating dyes allows the visualization of cDNA amplification and to the increase in the amount of cDNA, representing the starting amount of RNA studied. The cycle in which the cDNA amplification begins its exponential phase is called cycle threshold (C_T). In relative quantification of gene expression, the RNA level of a target gene is calculated relative to the RNA level of a reference gene that is assumed to be constant across tissues, such as a housekeeping gene. Furthermore, target/reference ratios can be normalized by using a calibrator sample, allowing the comparison of several different PCR reactions. $2^{-\Delta\Delta C_T}$ approach provides an easy method for normalizing the reactions to a calibrator sample (Livak *et al.*, 2001).

Investigating the pattern of expression of a disease gene provides valuable information about its function. In addition, a mutation in a disease gene may also affect its expression and the expression in diseased and healthy individuals can be compared whenever relevant samples are available.

2. PURPOSE

The aim of the studies within the framework of his thesis was to identify genes responsible for three recessive inherited diseases. Ataxia and progressive myoclonic epilepsy with dystonia (PMED) are very rare neurological diseases, whereas azoospermia is a form of male infertility with a high frequency in all populations. In ataxia and azoospermia, the purpose was first to map the disease loci via linkage analysis using genotyping data generated by SNP genome scan. After the identification of the loci, the disease genes were to be searched via exome sequencing, within the candidate loci. In PMED the disease locus was already known, so the goal was to find the disease gene residing at the locus. If the function of the disease gene was not known, as in case of PMED, gene expression assays were to be performed to understand tissue specificity.

3. MATERIALS

3.1. Subjects

Informed consent has been obtained from/for all subjects participated in the study. The research was approved by the Committee on Research with Human Participants at Boğaziçi University.

3.1.1. Ataxia

A consanguineous family with two siblings afflicted with a novel form of autosomal recessive ataxia was studied to search for the disease locus and the underlying mutation (Figure 3.1). Peripheral blood samples from two affected brothers and their healthy sibling were provided by Bülent Kara at Kocaeli University in Kocaeli.

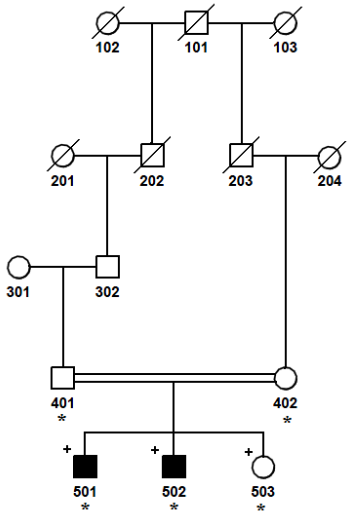


Figure 3.1. Pedigree diagram for Ataxia family. DNA samples subjected to SNP genotyping are indicated with +, and DNA availability is indicated with *.

3.1.2. Azoospermia

A consanguineous family afflicted with azoospermia was studied for locus and subsequent mutation identification (Figure 3.2). Peripheral blood samples from four affected brothers and one healthy sibling were provided by Mahmut Balkan at Dicle University in Diyarbakır. In addition, DNA samples from 45 azoospermic men and 15 oligozoospermic men that had been investigated for karyotype and Y-chromosome deletions also by Mahmut Balkan were studied.

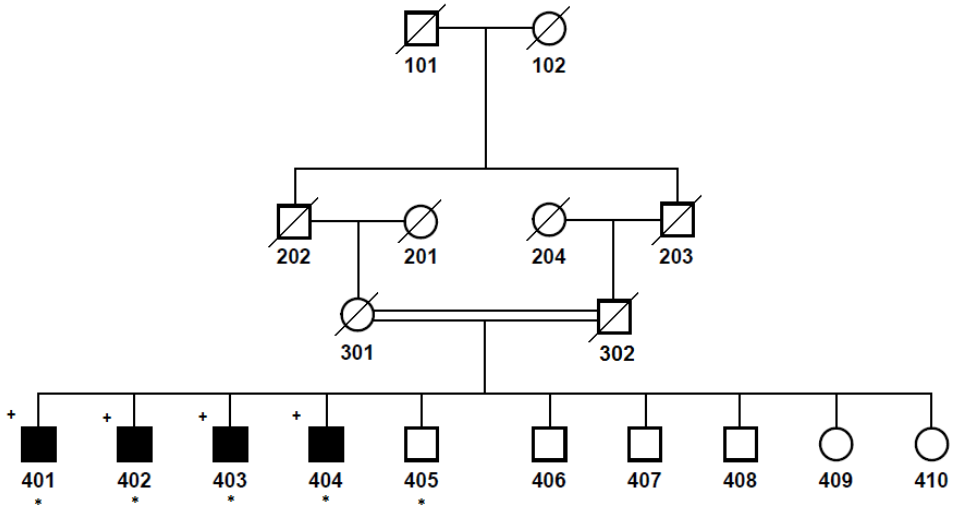


Figure 3.2. Pedigree diagram for Azoospermia family. DNA samples subjected to SNP genotyping are indicated with +, and DNA availability is indicated with *.

3.1.3. PMED

A large consanguineous family (Figure 3.3) with progressive myoclonic epilepsy with dystonia (PMED) in which the disease locus was previously mapped to 16pter-p13.3 (Duru, 2006; Duru *et al.*, 2010) was studied with the aim of identifying the causal mutation. Peripheral blood samples had been provided by Nilgün Selçuk Duru at Pediatric Clinic, Haseki Training and Research Hospital in Istanbul.

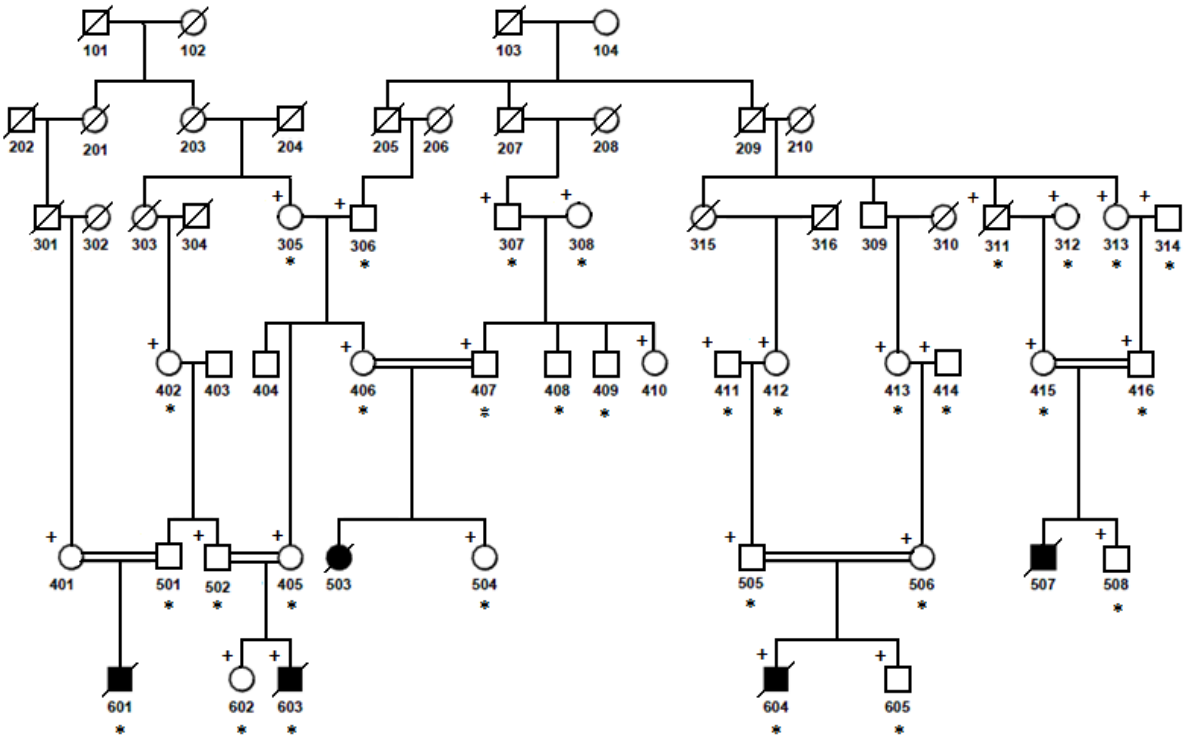


Figure 3.3. Pedigree diagram for PMED family. DNA samples subjected to genome scan are indicated with +, and DNA availability is indicated with *.

3.2. Chemicals

Solid and liquid chemicals used in this study were purchased from Merck (Germany), Sigma (USA), Riedel de-Häen (Germany), Carlo Erba (Italy) or Biochrom (Germany).

3.3. Buffers and Solutions

3.3.1. DNA Extraction from Peripheral Blood Samples

Table 3.1. Buffers and solutions used in DNA extraction from peripheral blood samples

Cell Lysis Buffer	:	155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM Na_2EDTA (pH 7.4 pH)
Nucleus Lysis Buffer	:	400 mM NaCl, 2 mM Na_2EDTA, 10 mM Tris (pH 8.2)
Proteinase K	:	20 mg/ml Proteinase K in dH_2O (Sigma, USA)
Sodium dodecylsulfate (SDS)	:	10 per cent SDS (w/v) in dH_2O
Ammonium Acetate	:	7.5 M CH_3COONH_4 in dH_2O
Ethanol	:	Absolute Ethanol
TE buffer	:	1 mM EDTA, 20 mM Tris-HCl (pH 8.0)

.2.2. Polymerase Chain Reaction (PCR)

Table 3.2. Buffers and solutions used in PCR assays.

10 X PCR Buffer A	:	20 mM $MgCl_2$, 500 mM KCl, 100 mM Tris-HCl (pH 8.3)
10 X PCR Buffer B	:	20 mM $MgSO_4$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 1 per cent Triton X-100, 1 mg/ml BSA, 200 mM Tris-HCl (pH 8.8)
5X Combinatorial Enhancer Solution (CES)	:	2.7 M Betaine, 6.7 per cent dimethyl sulfoxide (DMSO), 6.7 mM dithiothreitol (DTT), 55 µg/ml bovine serum albumin (BSA)
$MgCl_2$	:	25 mM $MgCl_2$ (Roche, Germany)
dNTP	:	12.5 mM each of dATP, dCTP, dGTP and dTTP (Roche, Germany and Fermentas, Lithuania)
Betaine	:	5 M Betaine (Promega, USA)

3.3.3 Agarose Gel Electrophoresis

Table 3.3. Buffers and solutions used in agarose gel electrophoresis

Agarose	:	2 per cent Agarose (Sigma, USA) in 0.5 X TBE buffer
10X TBE Buffer	:	20 mM EDTA, 0.89 M Boric Acid, 0.89 M Trizma base (pH 8.3)
6 X Loading Buffer	:	10 mM Tris-HCl (pH 7.6), 50 per cent Glycerol, 60 mM EDTA, 2.5 mg/ml Bromophenol Blue and/or 2.5 mg/ml Xylene Cyanol
Ethidium Bromide	:	10 mg/ml in dH_2O

3.3.4. Single Strand Conformational Polymorphism (SSCP) Analysis

Table 3.4. Buffers and solutions used in SSCP analysis.

40 per cent Acrylamide	:	40 per cent Acrylamide-Bisacrylamide (37.5:1) (Stock) in dH_2O
8 per cent Acrylamide	:	8 or 10 per cent Acrylamide-Bisacrylamide (37.5:1, non-denaturing) in 0.6 X TBE buffer
Glycerol	:	8 per cent Glycerol in gel solution
APS	:	10 per cent Ammonium Peroxydisulfate
TEMED	:	N,N,N,N-tetramethylethylenediamine
10X Sample Buffer	:	95 per cent formamide, 20 mM EDTA, 0.05 per cent Xylene Cyanol, 0.05 per cent Bromophenol Blue

3.3.5. Silver Staining

Table 3.5. Buffers and solutions used in silver staining.

Staining Buffer	:	0.1 per cent $AgNO_3$ in dH_2O
Developing Buffer	:	0.5 per cent NaOH, 0.01 per cent $NaBH_4$, 0.015 per cent Formaldehyde in dH_2O

3.4. Enzyme

Taq DNA polymerase with supplied PCR buffer was purchased from Roche (Germany) or Fermantas, Dream Taq (Lithuania).

3.5. Kits

The commercial kits used in this study are given in Table 3.6.

Table 3.6. The list of commercial kits, their usage and manufacturing companies.

Name	**Used For**	**Company**
GC-Rich PCR System	Amplification of templates rich in GC nucleotides	Roche, Germany
Dream Taq PCR Kit	Amplification of DNA templates	Fermentas, Lithuania
QIAQuick Gel Extraction kit	Elution of DNA fragments from agarose gels	Qiagen, USA
QIAQuick PCR Purification kit	Purification of PCR products	Qiagen, USA
LightCycler 480 High Resolution Melting Kit	Heteroduplex analysis (mutation screening) on LightCycler 480	Roche, Germany
Light Cycler 480 SYBR Green I Master kit	Relative quantification on LightCycler 480	Roche, Germany
RevertAid First strand cDNA synthesis kit	Reverse Transcription-PCR	Fermentas, Lithuania
Transcriptor High Fidelity cDNA Synthesis Kit	Reverse Transcription-PCR	Roche, Germany

3.6. Oligonucleotide Primers

Oligonucleotide primer pairs were designed via Primer3 software and checked for secondary structures and self annealing sites with OligoCalc (Table 3.2). The oligonucleotide primers used in this study were custom made in Massachusetts General Hospital (MGH) DNA Synthesis Core (USA). The primers were obtained in lyophilized form and later dissolved in 1 ml dH_2O. For PCR reactions, 10 μM dilutions were used.

3.7. DNA Molecular Weight Markers

Lambda DNA/*Hind*III and pUC19 DNA/*Msp*I markers were purchased from Fermentas (Lithuania).

3.8. Equipment

Table 3.7. Equipment used in the study.

Autoclave	:	AMB430T (Astell, UK)
Balance	:	Electronic Balance (Precisa, Switzerland)
Centrifuges	:	MiniSpin Plus (Eppendorf, Germany), Universal 16R (Hettich, Germany), Allegra X-22R (Beckman Coulter, USA) J2-MC (Beckman Coulter, USA)
Computer	:	Intel i7 960 (8X) processor, Gigabyte X58A-UD5 Motherboard, Kingston 1333 Mhz RAM, CPU @ 3.20GHz, 12328MB Memory, Ubuntu 11.10 Operating System
Deep Freezers	:	-20°C (Arçelik, Turkey), -80°C Ultra Freezer (Thermo Scientific, USA)
Documentation System	:	GelDoc Documentation System with Quantity One 4.6.9 Analysis Software (BioRad, USA)

Table 3.7. Equipment used in the study (cont.).

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)
Micropipettes	:	Pipetman (Gilson, France)
Minishaker	:	Rotamax 120 (Heidolph, Germany)
Ovens	:	KD 200 (Nüve, Turkey)
Power Supplies	:	Power Pac Model 3000 (Bio-Rad, USA), Fotoforce 250 Electrophoresis Power Supply (Fotodyne, USA), P250A Power Supply (Sigma-Aldrich, USA)
Spectrophotometers	:	NanoDrop 1000 (Thermo Scientific, USA)
Thermal Cyclers	:	MyCycler (Bio-Rad, USA), 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)
Waterbath	:	Grant LTD 6G Thermostatic Water Bath (Grant, Germany)
Water Purification System	:	Ultra Pure Water Purification system (UTES, Turkey and Hach-Lange, Germany)

3.9. Electronic Databases

Electronic databases and online tools used in this study are given in Table 3.8.

Table 3.8. Electronic databases and tools utilized in this study.

Name	**Description**
Online Mendelian Inheritance In Man (OMIM) http://www.ncbi.nlm.nih.gov/Omim/	A database of human genes, human diseases and associated phenotypes.
NCBI Genome Resources http://www.ncbi.nlm.nih.gov/genome/	A database of sequences and maps of a large number of reference genomes.
UCSC Genome Browser http://genome.ucsc.edu/	A database of sequences and maps of a large number of reference genomes.
Ensembl http://www.ensembl.org/	A database of sequences and maps of a large number of reference genomes.
NCBI UniGene http://www.ncbi.nlm.nih.gov/unigene/	Expression data of a wide collection of genes in various tissues.
UniProt (Universal Protein Resource) http://www.uniprot.org/	A comprehensive resource of protein sequence and functional information.
Gene Cards http://www.genecards.org	A database unifying the information about human genes on genomic, proteomic and functional level.
Polymorphism Phenotyping (PolyPhen-2) http://genetics.bwh.harvard.edu/pph2/	Predicts the structural and functional effects of amino acid substitutions on human proteins.
SIFT (Sorting Intolerant From Tolerant) http://sift.jcvi.org/	Calculates position specific scores for amino acids by utilizing sequence homology. The scores range from 0 to 1, which represents the effect from damaging to neutral.
Molecular Modelling & Bioinformatics (MMB) http://mmb2.pcb.ub.es:8080/PMut	Predicts the pathological effects of nonsynonymous mutations. Mutations with scores greater that 0.5 are pathological.
SNPs3D http://www.snps3d.org/	Uses structure and sequence analysis to assess functional effects of non-synonymous SNPs.
Biology Workbench http://workbench.sdsc.edu	An online tool package for storing and analyzing DNA sequences. Contains Primer3 software.
OligoCalc (Oligonucleotide Properties Calculator) http://www.basic.northwestern.edu/biotools/oligocalc.html	Calculates the melting properties and predicts potential hairpin structures or self annealing sites of primers.
Human Splicing Finder 2.4.1 http://www.umd.be/HSF/	An online tool that calculates consensus values of potential splice sites and searches for branch points in given sequences. It also predicts the effect of a sequence change on pre-mRNA splicing.

3.10. Bioinformatics Tools

The programs used in this study for analysis of exome sequencing data are given in Table 3.9. All bioinformatics programs used for exome sequence analysis are open-source and available for free download.

Table 3.9. Bioinformatics tools utilized for analyzing exome sequencing data.

Name	**Description**
BWA (Burrows-Wheeler Aligner) http://bio-bwa.sourceforge.net/	A program for aligning short nucleotide sequences against a reference sequence, such as the human genome.
SAMTools http://samtools.sourceforge.net/	A program for manipulating alignments in the SAM (Sequence Alignment/Map) format, including sorting, indexing and variant calling.
GATK (Genome Analysis Tool Kit) http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit	A software library of next-generation sequencing data analysis tools, including depth of coverage analysis, SNP/indel calling and local realignment.
ANNOVAR http://www.openbioinformatics.org/annovar/	A program performing functional annotations of genetic variants obtained from next-generation sequencing data. The annotations can be gene based, region based or filter based.
Picard http://picard.sourceforge.net/	A program comprising of Java-based command-line utilities for manipulating SAM files.
BEDTools http://code.google.com/p/bedtools/	A suite of utilities for comparing genomic features such as finding feature overlaps and computing coverage in next-generation sequencing data.
BamView http://bamview.sourceforge.net/	An interactive program for visualizing read alignments in BAM files.
CNVision (former Combined_CNV) http://futo.cs.yale.edu/mw/index.php/CNVision	A program for merging, annotating and visualizing CNVs, using SNP genotyping data obtained from Illumina GenomeStudio.

4. METHODS

4.1 DNA Extraction from Peripheral Blood Samples

A genomic DNA sample was isolated from a peripheral blood sample that was collected in a sterile tube containing K_2EDTA as anticoagulant. For plasma membrane lysis, thirty ml of cold cell lysis buffer was added per 10 ml of blood, and the mixture was kept at 4°C for 10 minutes (min). The sample was centrifuged at 5000 revolution per minute (rpm) at 4°C for 10 min, the supernatant was discarded, and the pellet containing leukocyte nuclei was washed by suspending in 10 ml of cell lysis buffer. Again, the sample was centrifuged at 4°C for 10 min, and the supernatant was discarded. The pellet containing the nuclei was resuspended in 0.3 ml of nuclei lysis buffer per ml of initial blood volume (3 ml for 10 ml blood) and vortexed extensively until the pellet dissolved entirely. For the digestion of nuclear proteins, 50 µl of Proteinase K (20 mg/ml) and 80 µl of 10 per cent SDS were added and the sample was mixed gently. The sample was incubated either at 37°C overnight or at 56°C for 3 hours (hr). After incubation, 2.8 ml of NH_4Ac (9.5 M) was added, and the tube was shaken vigorously to salt out proteins. To collect the precipitated proteins, the sample was centrifuged at 10,000 rpm at room temperature for 25 min. The supernatant was transferred into 50 ml tubes, and two volumes of ethanol were added to precipitate out DNA. After mixing the tube very gently, DNA floating in the supernatant was fished out carefully with a micropipette tip and transferred into a clean 1.5 ml tube. The precipitate was air-dried for the residual ethanol to evaporate. DNA was dissolved in 500 µl of TE buffer and stored at -20°C.

4.2. Linkage Analysis

Linkage analysis is a statistical test that associates a disease gene with a locus by calculating LOD (logarithm of odds) scores. In order to perform parametric linkage

analysis, the pedigree of the family, the genome scan data of family members included in the study and the genetic map of the markers used are required.

4.2.1. Genome Scan

4.2.1.1. Ataxia. Genome scan was performed for two affected brothers and their healthy sister at MacroGen Inc (South Korea) with Illumina Human Omni 1M Quad BeadChip including one million SNP markers at a median marker spacing of 1.25 kb. Illumina Genome Studio v.1.02 Genotyping Module was used to process genotype data. Homozygous regions were extracted using ChromoZone and Homozygosity Detector (Illumina Inc, USA) plug-ins, and the regions were visualized via Illumina Genome Viewer.

4.2.1.2. Azoospermia. Genome scan was performed for four affected brothers at MacroGen Inc (South Korea) with Illumina Human 610 Quad BeadChip including 610,000 SNP markers at a median marker spacing of 2.7 kb. Illumina Genome Studio v.1.02 Genotyping Module was used to handle genotyping data. Homozygous regions were extracted using ChromoZone and Homozygosity Detector plug-ins (Illumina Inc, USA), and the regions were visualized via Illumina Genome Viewer.

4.2.2. Parametric Linkage Analysis and Homozygosity Mapping

easyLINKAGE package version 5.08 was used for parametric LOD score calculations in ataxia and azoospermia families. easyLINKAGE package is an open source GUI (graphical user interface) that contains the programs given in Table 4.1.

Table 4.1. The programs included in easyLINKAGE software.

Program	**Description**
PedCheck	Detects genotyping errors and checks the pedigree file for Mendelian incompatibilities.
SimWalk2	Calculates parametric LOD scores and constructs haplotypes for any size of pedigree, including complex families (bit size > 19).
SuperLink	Calculates two-point parametric LOD scores.
Allegro	Calculates both two-point and multi-point parametric LOD scores in small pedigrees (bit size <19) and constructs haplotypes.
GeneHunter	Calculates multi-point parametric LOD scores and constructs haplotypes in small pedigrees (bit size <19).
SPLink	Performs linkage analysis using affected sib pairs, with the method of maximum likelihood.

Homozygosity mapping was performed in two diseases using Homozygosity Detector plug-in in Illumina Genome Studio (Illumina Inc, USA) and Homozygosity Comparison in Excel (HCiE) developed in our laboratory (Cetinkaya, 2010). HCiE is an MS Excel macro, in which three different genotypes (AA, BB or AB) obtained from genotyping of subjects, are differently colored and the markers are sorted in the order of their locations. This allows the user to detect homozygous genotypes shared by all affecteds, as well as to observe whether the unaffected individuals have different genotypes than those of the affecteds.

4.2.2.1 Ataxia. Multipoint LOD score calculations were performed via program Allegro, using all markers in sets of 100 and with 0.01 cM spacing. Autosomal recessive inheritance with full penetrance was assumed and disease frequency was set to 0.0001. The physical and genetic maps of markers were provided by MacroGen Inc. Candidate loci were investigated by Homozygosity Detector plug-in and visualized in Illumina Genome Viewer. The loci were also confirmed by HCiE (Cetinkaya, 2010). Haplotypes were constructed using program GeneHunter.

4.2.2.2 Azoospermia. Multipoint LOD scores were calculated using program Allegro, assuming autosomal recessive inheritance with full penetrance and a disease frequency of 0.0001. Sets of 100 markers with 0.01 cM spacing were selected. The physical and genetic

maps of markers for Omni 1M Quad BeadChip, provided by MacroGen Inc, were used in calculations. Haplotypes were constructed using GeneHunter. Homozygosity mapping was performed via HCiE (Cetinkaya, 2010). Fine calculations were performed in loci with maximum LOD scores.

4.3 Copy Number Variation (CNV) Analysis

It is possible to detect copy number variations (CNVs) using high density SNP arrays (McCarroll *et al*. 2008). To detect CNVs, a perl based program called Combined_CNV was developed by Stephan Sanders and Chris Mason in the State Lab at Yale University and provided to our laboratory by Prof. Murat Günel at Yale University. Combined_CNV uses three CNV detection algorithms, namely, QuantiSNP (Colella *et al.*, 2007), PennCNV (Wang *et al.*, 2007) and GNOSIS for analyzing high density SNP genotyping data. CNV detection relies on the LogR ratio values, which is logarithm (base 2) ratio of the normalized intensity (R value) for the SNP tested, divided by the expected normalized intensity. Combined_CNV program takes Final Reports generated by Illumina GenomeStudio as input files and outputs annotated files containing the genomic locations and flanking SNPs of CNVs. The program also annotates CNVs previously reported in Database of Genomic Variants (DGV).

4.4 Exome Sequencing

4.4.1 Exome Capture

Targeted capture of the exome and subsequent sequencing was performed at MacroGen Inc (South Korea) as a commercial service. Two affected individuals, one each from ataxia family and azoospermia family, were subjected to exome capture with Illumina TruSeq Exome Capture kit (Illumina, USA), followed by massively parallel paired end sequencing with Illumina HiSeq 2000 (Illumina, USA). The features of TruSeq Exome Capture kit are given in Table 4.2.

Table 4.2. The features of Illumina TruSeq Exome Capture kit.

Capture target	62 Mb
Number of target genes	20,794
Number of target exons	201,121
Number of probes	340,427
Size of probes	95 bp

Obtaining sequence information from exome sequencing occurs in various steps that generate a list of variants within or near the coding regions of the genome. First, a sample library is prepared using Illumina TruSeq DNA Sample Preparation kit (Illumina, USA), using 5-10 µg of genomic DNA as starting material. The aim of this step is to add adapter sequences to the ends of DNA fragments to generate multiplexed paired end sequencing libraries. In this step, firstly the genomic DNA is sheared and overhangs are converted to blunt ends in order to obtain double stranded DNA fragments of an average of 300-400 bp in size. To prevent the ligation of blunt fragments to one another, a single "A" nucleotide is added to the 3' ends. Multiple indexing adapters, containing a complementary "T" nucleotide at the 3' end, are ligated to the blunt fragments. After ligation of the adapters, biotinylated capture probes of targeted regions are hybridized with blunt fragments and unbound fragments are washed. Streptavidin beads pull out the biotinylated capture target, and the fragments are hybridized onto the flow cell, which is a solid surface containing forward and reverse primers in spatially separated clusters. With these primers, the captured target is amplified by bridge PCR in clusters, and subsequent sequencing reaction takes place. An overview of the target capture enrichment and cluster generation is shown in Figure 4.1.

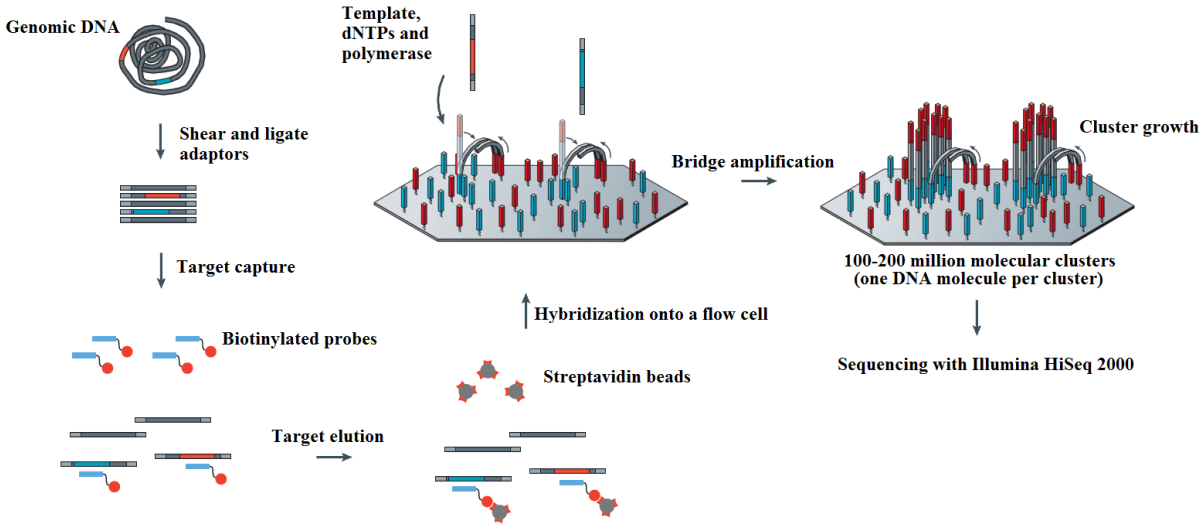


Figure 4.1. Schematic representation of exome capture enrichment workflow (modified from Metzker, 2010).

4.4.2 Analysis of the exome sequencing results

Bioinformatics analysis of exome sequencing data was included in the service given by MacroGen Inc (South Korea), and annotated list of variants were provided in an MS Excel file. This analysis was performed using programs BWA and SAMTools with standard parameters.

In order to avoid false negative calls from MacroGen's standard analysis, we performed alignment, SNP calling and coverage computations with more relaxed parameters. Raw reads generated from sequencing were obtained from MacroGen Inc (South Korea) in paired end ".fastq" files. Human reference genome sequence (assembly GRCh37/hg19) was downloaded from UCSC Genome Bioinformatics Site (described in Table 3.2). Individually assembled chromosomes were concatenated to each other and indexed. Program BWA was used to align paired end to the reference genome and to generate a final alignment in the SAM (Sequence Alignment/Map) format. The SAM file was converted to BAM (Binary Alignment/Map) format with SAMTools. The resulting file was sorted, indexed and subjected to SNP calling with SAMTools, which outputs a list of

nucleotides that were different from the reference genome. The list of variants were annotated with ANNOVAR, which denotes chromosomal location, dbSNP ID, the type of the region of the nucleotide change (exonic, intronic, UTR, non-coding RNA or intergenic) and the type of the exonic nucleotide change (synonymous, non-synonymous, frameshift, stopgain or stoploss) and whether an intronic nucleotide change is a splicing mutation. For visualizing the alignment, BamView program was used. The command lines in all steps are given in Table 4.3.

Table 4.3. Command lines used in bioinformatics analysis of exome sequencing data and their purposes.

Purpose	**Command**
Concatenating all chromosomes of the reference genome	cat *.fa > all_chr.fasta
Indexing the reference genome	bwa index –a bwtsw all_chr.fasta
Aligning paired-end reads to reference genome (performed for each .fastq file separately)	bwa aln –n 0.01 –t 8 all_chr.fasta file1.fastq > file1.sai
Generating alignment in the SAM (Sequence Alignment/Map) format	bwa sampe all_chr.fasta file1.sai file2.sai file1.fastq file2.fastq > file.sam
Converting SAM to BAM (Binary Alignment/Map) format	samtools view –bS file.sam > file.bam
Sorting the BAM file	samtools sort file.bam
Indexing the BAM file	samtools index file.sorted.bam
SNP calling	samtools pileup –vcf all_chr.fasta file.sorted.bam > file.pileup.txt
Converting the file to a format compatible with ANNOVAR	convert2annovarl.pl file.pileup.txt –outfile file_annovar.pileup.txt

Table 4.3. Command lines used in bioinformatics analysis of exome sequencing data and their purposes (cont.).

Purpose	**Command**
Region based annotation	annotate_variation.pl –buildver hg19 file_annovar.pileup.txt humandb
Filter based annotation	annotate_variation.pl –filter –dbtype snp131 file_annovar.pileup.txt
Visualization of the alignment	java –mx152m –jar BamView_v1.1.8.jar –a file.sorted.bam –r all_chr.fasta
Computing the coverage for each exon included in the target capture using BEDTools	bamToBed –i file.bam > file.bed coverageBed –a file.bed –b targetedRegions.bed > file_coverage.txt

4.5 Candidate Genes and Mutation Screening

Among the variants generated by exome sequencing results, the mutations within loci that stood out as the best candidates were selected. In selecting the candidate mutations, the variants that had the highest coverage and highest quality scores, as well as the effect of the mutation on the protein, were considered with priority. The threshold for sufficient coverage was assumed as 4 reads and the minimum quality score accepted was 40. Variants not fitting into these conditions were regarded more likely as false positives. Nevertheless, all variants were taken into consideration. The variants that were previously reported in dbSNP and 1000 genomes were eliminated, and the remaining novel variants were prioritized with respect to gene function, severity of the mutation, expression in relevant tissues and evidence in animal models.

After identifying a candidate mutation in exome sequencing results, it was verified by Sanger sequencing of the PCR amplified products. In the case of PMED, *TBC1D24* was selected as the best candidate gene upon evidence from two previously reported families having familial epilepsy (Corbett *et al.*, 2010; Falace *et al.*, 2010), and all exons were sequenced directly.

4.5.1 PCR Amplifications

All strong candidate variants resulting from exome sequencing and all exons of *TBC1D24* were subjected to PCR amplification in relevant individuals for subsequent DNA sequencing. PCR for each primer pair was carried out using 1X PCR buffer, 0.4 µM of each primer, 0.2 µM of each dNTP, 20-100 ng of genomic DNA, 0.2 U Taq DNA polymerase and sufficient water adding up to a total volume of 25 µl. Primer sequences and optimum PCR conditions are given in Tables 4.4 and 4.5.

Table 4.4 Sequences, PCR product sizes and PCR conditions for primer pairs amplifying *TBC1D24* exons.

Region	**Primer Name**	**Primer Sequence (5'→ 3')**	**Product Size (bp)**	**Buffer**	**Annealing Temperature (°C)**
Exon 1	TBC1D24 _5utr	F: ACTGCCGTTGCCTCCTC R: CCCTCCCGTAGGCCAAG	439	GC-Rich PCR kit	57
Exon 2.1	TBC124D _Ex1-1	F: GAGGCAGGAAGATGGTTCAA R: AGAAGGAGATGTCGGGGAA	700	Dream Taq Buffer	62→55
Exon 2.2	TBC124D _Ex1-2	F: GAAAGGTGTACCAGCGCC R: CAGGAAGACGTCAAAGACCC	545	Dream Taq Buffer	64→56
Exon 2.3	TBC124D _Ex1-3	F: CCTGGTGAACAAGTACTGCC R: TGAGCTCACGCCAGACAC	485	Dream Taq Buffer	62→55
Exon 3	TBC1D24 _Ex2B	F: GGGGGATCGGTACTCACAC R: AGCATCAGTCCGGCTCAG	237	Buffer B	55
Exon 4	TBC124D _Ex2	F: GCTCTGGGGCATACCTCG R: TCTGTGGGCAGGACACG	299	Buffer B	55
Exon 5	TBC124D _Ex3	F: AGTTCCCAGGTCTTCGCC R: TAGGCCCTGAAGCCCATC	263	Buffer B	55
Exon 6	TBC124D _Ex4	F: TAGTCTGGAGCACAGGGACG R: GGTGCTCCTGGAGGGATG	228	Buffer B	55
Exon 7	TBC124D _Ex5	F: AGGGGGCTTCATCTGCTC R: GTCCACCCTGACACCCAG	506	Dream Taq Buffer	62→55
Exon 8.1	TBC124D _Ex6-1	F: CCTGGGTCAGTGCTGATAGG R: CTAGAGAGGCTCAGGGGCAG	552	Dream Taq Buffer	62→55
Exon 8.2	TBC124D _Ex6-2	F: CACCCAGAGCTGGCATAG R: GGGCTACCTCTTCAGATGGG	564	Dream Taq Buffer	64→56
Exon 8.3	TBC124D _Ex6-3	F:GAGGAGGGTGGGATTTGTG R: CACCTACCTGCGTCTGGC	589	Dream Taq Buffer	64→56
Exon 8.4	TBC124D _Ex6-4	F: CAACTGCTTCTCTGCACCAG R: TGAGAGGGATTCACCCCAT	732	Dream Taq Buffer	64→56
Exon 8.5	TBC124D _Ex6-5	F: GTAGCCTCTGACTGGGCG R: ACCCAGTTCAGGAAGTTTGC	546	Dream Taq Buffer	64→56
Exon 8.6	TBC124D _Ex6-6	F: CCGCTGTAGTGCTCACCC R: AAAATTTAGCTGGGCGTGGT	776	GC-Rich PCR kit	58.5

Table 4.5. Sequences, PCR product sizes and PCR conditions for primer pairs amplifying the candidate variants obtained from exome sequencing results.

Region	Primer Name	Primer Sequence (5'→ 3')	Product Size (bp)	Buffer and Additives	Annealing Temperature (°C)
ATAXIA					
OSBPL11					
Exon 5 (c.511C→T)	OSBPL11 _E5	F: TGGTAGTTTGATGGTTAGAAATCC R: GACAAAACCAAGCAGGTAAAATC	308	Buffer A	55
COL6A6					
Exon 34 (g.101976G →A)	COL6A6_ E34	F: GAGCTGAGTTCAATCTTACCACC R: GGTTTCCAAGTATCCATGTTAGC	507	Buffer A	55
Exon 34 (g.101976G →A)	COL6A6_ pop	F: GTGTTTTCCCTTGGCCCTAT R: CTCCCTTCTACCATCCGTCA	316	Buffer B	55
AZOOSPERMIA					
TAF4B					
Exon 9 (c.1831C→ T)	TAF4B	F: TTCTGTCACTTCAAGCATCTCC R: CGGCATGATGATTACACAGG	287	Buffer A	55

4.5.2 Analysis of PCR Products

For all PCR reactions the efficiency of amplification and the purity of the products were checked on agarose gels. Five µl of a PCR product was mixed with 1 µl of 6X loading dye and loaded on a two per cent agarose gel containing 10 mg/ml ethidium bromide. The gel was subjected to electrophoresis in 0.5X TBE at 120 volts for 15 minutes, and the fragments were visualized under UV light, using BIORAD Universal Hood II and Quantity One 4.6.9 software.

4.5.3 DNA Sequence Analysis

A region to be analyzed with Sanger sequencing was amplified by PCR in 50 µl reactions. Total volume per sample was 50-200 µl, depending on the efficiency of the reaction. PCR products were purified from nucleotides, primers, salts and polymerases using QIAQuick PCR Purification kit (QIAGEN) in the early stages of the work. In the later stages, purification was not necessary as it was done at MacroGen Inc (South Korea), where the sequence analysis was performed for all our projects.

4.5.4 SSCP Analysis for Mutation Screening

Polyacrylamide gels with crosslinking ratios of 2 per cent, with or without eight per cent glycerol, were prepared. Glass plates of 20 cm x 20 cm in size were assembled using 0.75 mm spacers. Six ml of 40 per cent acrylamide:bisacrylamide stock (37.5:1) was mixed with 1.8 ml of 10 X TBE, and the volume was adjusted to 30 ml by dH_2O. For gels with glycerol, 2.4 ml of glycerol was added before adjusting the volume to 30 ml. Then two hundred and fifty µl of ten per cent APS and 25 µl of TEMED were added to the solution and was gently mixed, and poured between the glass plates. A 20-well comb was inserted, and the gel was left to polymerize for at least one hour. After polymerization, the gel was cooled at 4°C for at least half an hour before use. The sequences of the primers used for PCR amplifications are given in Tables 4.4 and 4.5. For *COL6A6* g.101976G→A the second pair (designated as COL6A6_pop) was preferred over the first one, as the PCR product was much smaller. PCR products were mixed with equal volumes of sample buffer, denatured at 95°C for 5 min and immediately placed on ice. Seven to 12 µl of each sample was loaded onto the gel and subjected to electrophoresis at a constant power of 10 to 30 Watts in 0.6 X TBE buffer at 4°C in BioRad DCode Universal Mutation detection system or BioRad Protean IIxi Cell. The time for electrophoresis was adjusted according to the length of the PCR product. After electrophoresis, the gels were stained with silver nitrate to visualize DNA. Samples having aberrant migration patterns were subjected to DNA sequencing.

A panel of 120 unrelated subjects was screened for mutation *TBC1D24* c.969_970delGT, variant *COL6A6* g.101976G→A and *TAF4B* c.1831C→T transition using SSCP analysis. Family members were also screened with the same method. In screening for *TBC1D24* c.969_970delGT, alleles were resolved on gels without glycerol at 10 watts and 4^oC for 14 hours. In screening for *COL6A6* g.101976G→A, gels with glycerol were run at 10 watts and 4^oC for 12 hours. In screening for *TAF4B* c.1831C→T, gels with glycerol were run at 10 watts for 18 hours.

Fourty-five azoospermia and fifteen oligospermia patients were screened for mutations in the first 281 nucleotides of exon 1 (designated as TAF4B_Ex1A) and exon 10 of *TAF4B* (designated as TAF4B_Ex1A) by SSCP analysis. Region TAF4B_Ex1A has a high GC-content and exon 10 contains a repetitive sequence; therefore, they could not be amplified with HRM method. The samples were run on 10 per cent acrylamide gels with glycerol at 20 W for 6 hours.

Table 4.6. Sequences, PCR product sizes and PCR conditions for primer pairs amplifying two exons of *TAF4B* for SSCP analysis.

Region	Primer Name	Primer Sequence (5'→ 3')	Product Size (bp)	Buffer	Annealing Temperature (°C)
Exon 1.1	TAF4B_Ex1A	F: CCTCCCCTCACCTCTGCT R: GCCACTTTGGTCACCAGG	281	GC-Rich PCR kit	56
Exon 10	TAF4B_Ex10	F: CACACACAACCTAAAATGTATAACTTT R: CAAACAGACATCTCTATCAGGTCA	259	Buffer B	53

4.5.5 Silver Staining

Glass gel plates were gently separated, letting the gel remain on one of them. The gel was detached from the glass by lifting with a piece of filter paper and transferred to a tray containing 300 ml of staining buffer. After incubation in staining buffer for 10 minutes, the gel was soaked in development buffer until bands appeared (Kavaslar *et al.*, 2000). In the case of inefficient staining, the procedure was repeated after an extensive washing step. After, the image was taken with a scanner and the gel was kept in sealed plastic sheets.

4.5.6 High Resolution Melting (HRM) Curve Analysis

A panel of 120 control individuals was screened for *OSBPL11* c.511C→T transition with high resolution melting curve analysis in Roche Light Cycler 480. In addition, all coding exons of *TAF4B*, except the first 281 nucleotides of exon 1, were screened in 45 azoospermia and 15 oligozoospermia patients. Each sample was "spiked" by the addition of 0.2 µl control DNA, to make a final concentration of 12 per cent control

DNA in the total template sample, to distinguish a homozygous sequence variation by creating an artificial heteroduplex. The samples showing aberrant melting patterns were subjected to DNA sequence analysis. The primers used for screening the population for *OSBPL11* c.511C→T transition and those for screening *TAF4B* exons are given in Table 4.7.

Table 4.7. Primer pairs utilized in HRM curve analysis.

Region	**Primer Name**	**Primer Sequence (5'→ 3')**	**Product Size (bp)**	**Annealing Temperature (°C)**
ATAXIA				
OSBPL11				
Exon 5	OSBPL11_pop,	F: GGCATTTTGACTTGGTCTCC R: TGGTAGTTTGATGGTTAGAAATCC	157	64→56
AZOOSPERMIA				
TAF4B				
Exon 1.2	TAF4B_Ex1B	F: GCCGTGACTAAGGCTCCTG R: AGACAAGGCTGGAAAGGGTT	244	55
Exon 2	TAF4B_Ex2	F: GTGGCTCAGAGGTCACTTTTT R:GCTGAGAAACAAAGAAAGATGAAA	228	56
Exon 3	TAF4B_Ex3	F: TGTTATCTTGCTCCCCTCCA R: AACAACAAGTGCACGAATGG	170	55
Exon 4	TAF4B_Ex4	F: TGAAATGAGGCACCCACTAA R: AAAAGAACTTGCAAGGAACAAGA	280	64→57
Exon 5	TAF4B_Ex5	F: CCTTGCAAGTTCTTTTGAATTGT R: GCATTTCCCAATTACAACATGT	262	59→55
Exon 6	TAF4B_Ex6	F: TTGTTTCTATTTCTCCCATCCC R: GCACTCAGAGGTTTCTCCCA	247	64→57
Exon 7.1	TAF4B_Ex7A	F: TTCCATTCCTCTGCATTTCC R: CCAGCAAGTGGGTTCAAAGT	252	63→58
Exon 7.2	TAF4B_Ex7B	F: TGTCCTCAAGCCAGTCTGAA R: CTGGCTTTTCAGGTTGCAGT	255	63→58
Exon 7.3	TAF4B_Ex7C	F: ACAGCTGGAACTGGTTTGC R: GCTTGTCAGATGTGGTCCC	220	62→57
Exon 7.4	TAF4B_Ex7D	F: TAACACTGTCCCTTCCAGCA R: AACTGGGGAAGAAACAAAATGT	235	64→57
Exon 8	TAF4B_Ex8	F: TTATGGCTGGGGGAACAC R: GGACTGAGATGGACCAGCAT	260	64→57
Exon 9	TAF4B_Ex9	F: TCCTAATGTGGTGTGATTGCTT R: GGTATTATTCTGTCATCTCTCCTTTTT	240	59→55
Exon 11	TAF4B_Ex11	F: TGGATTTTCTCTGCTTCTGC R: AATCCACGCCCAAAAGACTA	218	60→55
Exon 12	TAF4B_Ex12	F: TAGAATGTGTGGCCCAGGAT R: TCATAAATCACACTGGCCCTT	204	64→57
Exon 13	TAF4B_Ex13	F: AATGGCAACTATGGTGTTCAGA R: AGGAGAGAGCTGCCAACCTT	181	55
Exon 14	TAF4B_Ex14	F: TCCCTCTGAGCATGAATGTG R: TTGGCTTGCTTTCTTGACAT	212	55
Exon 15	TAF4B_Ex15	F: TTTCAAATTGCTTTTCCTTTTATT R: GATCTGGATGGAAGAGTGGAG	224	55

For population screening for *OSBPL11* c.511C→T, the reactions were carried out in a total volume of 20 µl, comprising of 10 µl of Roche High Resolution Melting Master Mix, 200 nM of each primer pair (Table 4.6), 2.5 mM of $MgCl_2$, 20 ng of total genomic DNA and dH_2O. Real-time PCR reactions were carried out in 96-well plates. The conditions were as follows: 95°C for 10 minutes for initial denaturation, 45 cycles of 95°C for 10 seconds, 64°C→56°C touchdown (-0.5°C for 16 cycles) for 10 seconds of annealing and 72°C for 10 seconds of elongation. High resolution melting curve data acquisitions were done at the range of 63°C-95°C at a rate of 25 acquisitions per °C.

$MgCl_2$ concentrations and the range of temperature for HRM data acquisitions used for each scanned region of *TAF4B* are given in Table 4.8. The reactions were carried out in a total volume of 20 µl, including 10 µl of Roche High Resolution Melting Master Mix, 200 nM of each primer pair (Table 4.7), 20 ng of total genomic DNA and dH_2O. The conditions were the same as in population screen for *OSBPL11* c.511C→T, except for the annealing temperatures (Table 4.7). Light Cycler 480 Gene Scanning software was used to normalize the fluorescence values and to generate melting curves.

Table 4.8. Reaction conditions for scanning of *TAF4B* exons.

Primer Name	**$MgCl_2$ concentration (mM)**	**HRM temperature range (°C)**
TAF4B_Ex1B	2	80-97
TAF4B_Ex2	3.5	73-90
TAF4B_Ex3	3	73-90
TAF4B_Ex4	3	73-90
TAF4B_Ex5	3	71-86
TAF4B_Ex6	2.5	68-88
TAF4B_Ex7A	3.5	74-90
TAF4B_Ex7B	3	76-92
TAF4B_Ex7C	3	74-92
TAF4B_Ex7D	3	74-92
TAF4B_Ex8	2.5	69-85
TAF4B_Ex9	3	66-80
TAF4B_Ex11	3	70-86
TAF4B_Ex12	2.5	67-87
TAF4B_Ex13	3	71-86
TAF4B_Ex14	3	71-86
TAF4B_Ex15	3	74-91

4.6 Relative Quantification of *TBC1D24* transcript isoforms

4.6.1 Real-Time Quantitative PCR

The expression of three transcript isoforms of *TBC1D24* was compared among themselves in seven brain tissues. Human cDNA samples from adult cerebellum, corpus callosum, frontal cortex, parietal cortex, striatum, occipital cortex and brain stem and additionally fetal brain (Clontech Inc, USA) were kindly provided by Prof. Tayfun Özçelik, Bilkent University, Ankara, Turkey. Fermentas RevertAid First strand cDNA synthesis kit had been used to synthesize first strand cDNA from total RNA of all samples, using oligo(dT) primers. The starting total RNA volume was 1 µg. Real-time quantitative PCR was performed on Roche Light Cycler 480 in 96-well plates, using Roche SYBR Green I Master kit. Primers specific to each transcript isoform are given in Table 4.9. The reactions were performed in a total volume of 20 µl, consisting of 10 µl Light Cycler 480 SYBR Green I Master, 200 nM of each primer pair, 0.8 µl of 1:10 diluted cDNA and dH_2O. All reactions were carried out in triplicate, under the following conditions: 95°C for 10 minutes for initial denaturation, 45 cycles of 95°C for 10 seconds, 63°C→57°C touchdown (-0.5°C for 12 cycles) for 10 seconds of annealing and 72°C for 10 seconds of elongation. Cycle threshold (C_T) values were retrieved from Light Cycler 480 Relative Quantification software and calculations were performed with MS Excel, using $2^{-\Delta\Delta C_T}$ method (Livak *et al.*, 2001). Transcript isoform 2 (NM_020705.2) was set as the calibrator for each tissue, and housekeeping gene *HPRT1* was used as the reference. Finally, graphs showing ± standard error of the mean (SEM) were generated.

Table 4.9. Primers used in relative quantification of *TBC1D24* transcript isoforms.

Transcript Isoform	Primer name	Primer Sequence (5'→ 3')	cDNA product size (bp)	Spanning Exons
TBC1D24				
NM_020705.2	TBC_var1_F TBC_var1_R	TAAAAGGCAGTTTGTACACTTGG CACTCCAGTCTGTGGACAGG	263	F: Exon 3 and 4 R: Exon 6
NM_001199107.1	TBC_var2_F TBC_var1_R	GCAGAAGAGGCAGTTTGTACA CACTCCAGTCTGTGGACAGG	266	F: Exon 2 and 4 R: Exon 6
Novel isoform 3	TBC_var3_F TBC_var1_R	CTAAAAGGTTCTACTTCCAGTGTGA CACTCCAGTCTGTGGACAGG	105	F: Exon 3 and 5 R: Exon 6
HPRT1				
NM_000194.2	HPRT1-F-RNA HPRT1-R-RNA	TGACCTTGATTTATTTTGCATACC CGAGCAAGACGTTCAGTCCT	102	F: Exon2 R: Exon 3

4.6.2 Primary structure analysis of novel *TBC1D24* transcript isoforms

Two novel *TBC1D24* transcript isoforms identified in this study were subjected to DNA sequence analysis from exon 3 up till 3'UTR sequences, using primers whose sequences are given in Table 4.10. For novel isoform 3, human cerebellum cDNA was diluted in a 1:10 ratio, whereas for novel isoform 4, human striatum cDNA was diluted in a 1:10 ratio.

Table 4.10. Primers used in nucleotide analysis of novel *TBC1D24* transcript isoforms

Transcript Isoform	**Primer name**	**Primer Sequence (5'→ 3')**	**cDNA product size (bp)**	**Buffer and Additives**	**Annealing Temperature (°C)**
Isoform 3	TBC_var3_F TBC_3UTR _R	CTAAAAGGTTCTACTTCCAGTGTGA GGTTTCATCTGCTCTCTGCT	631	Dream Taq Buffer	56
Isoform 4	TBC_var4_F TBC_3UTR _R	CTGGCCAGGTGTGTGGT GGTTTCATCTGCTCTCTGCT	568	Dream Taq Buffer, DMSO	61

5. RESULTS

In this section, the results of linkage analysis and identification of disease genes via exome sequencing in ataxia and azoospermia are presented. In addition, the results of mutation screening and relative quantification of transcript isoforms in PMED are given.

5.1 Ataxia

Genome scan data generated by 1 million SNP markers in three siblings were subjected to linkage analysis, and candidate disease loci were identified. Later, the mutation at one of those loci was identified via exome sequencing.

5.1.1 Multipoint Linkage Analysis

The genotyping data generated by the genome scan using 1 million SNPs in the two affected brothers and the healthy sister were subjected to linkage analysis, and multipoint LOD scores were calculated with program Allegro. An autosomal recessive disease model with full penetrance was assumed. A maximum LOD score of 2.5076 was obtained in various regions in most of the chromosomes. The lod score did not reach the critical value of 3 due to the small size of the family. Loci >2 cM were selected as the strongest candidates. Multipoint LOD score graphics for all chromosomes are given in Figure 5.1, for candidate loci 3q21.1-22.1 and 8q24.13-24.22 are given in Figure 5.2 and for candidate loci 2p12-q11.2 and 15q12 in Figure 5.3. The locations and sizes of the candidate loci are given in Table 5.1.

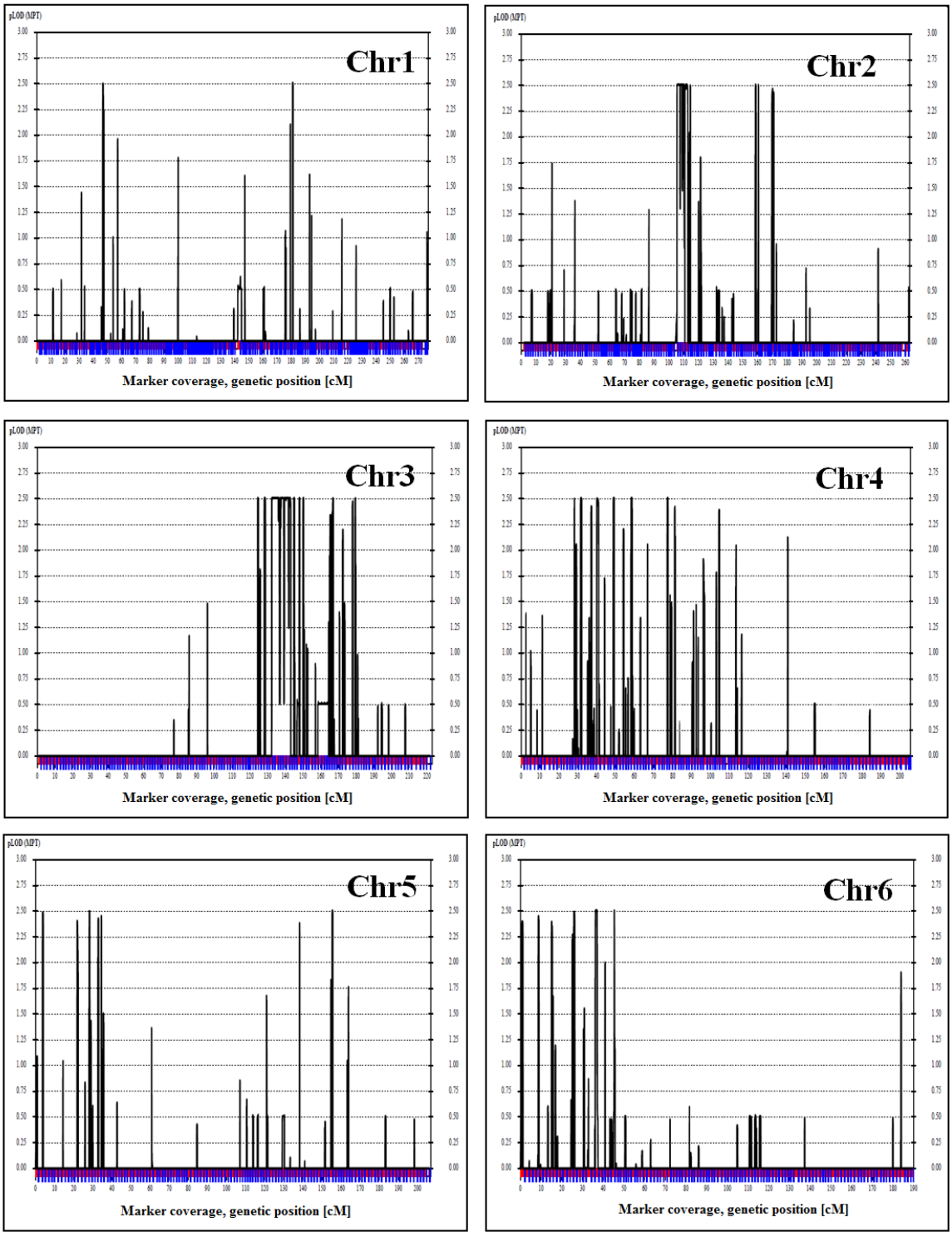


Figure 5.1 Multipoint LOD score results for ataxia for all autosomal chromosomes and PAR regions.

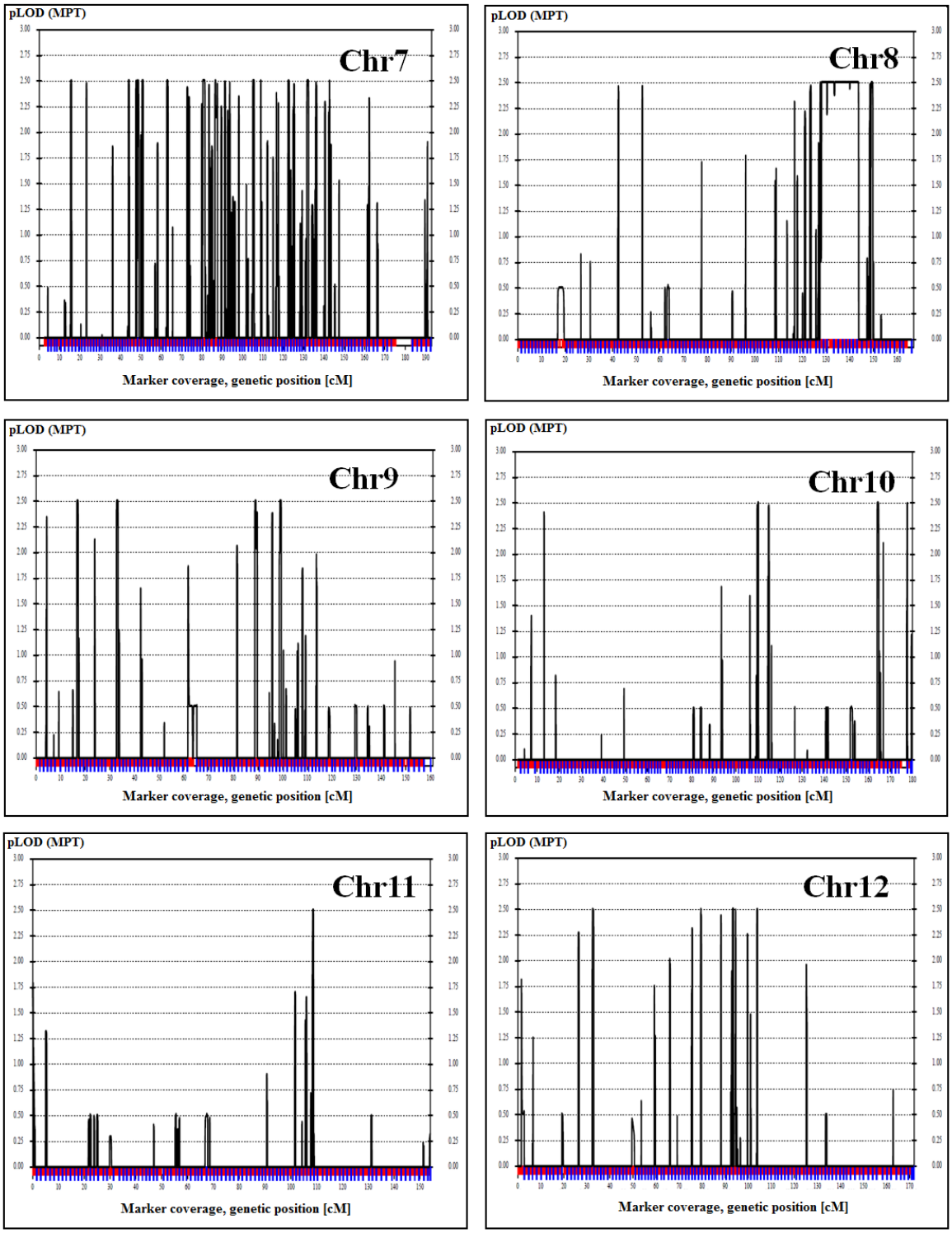


Figure 5.1 Multipoint LOD score results for ataxia for all autosomal chromosomes and PAR regions (cont.).

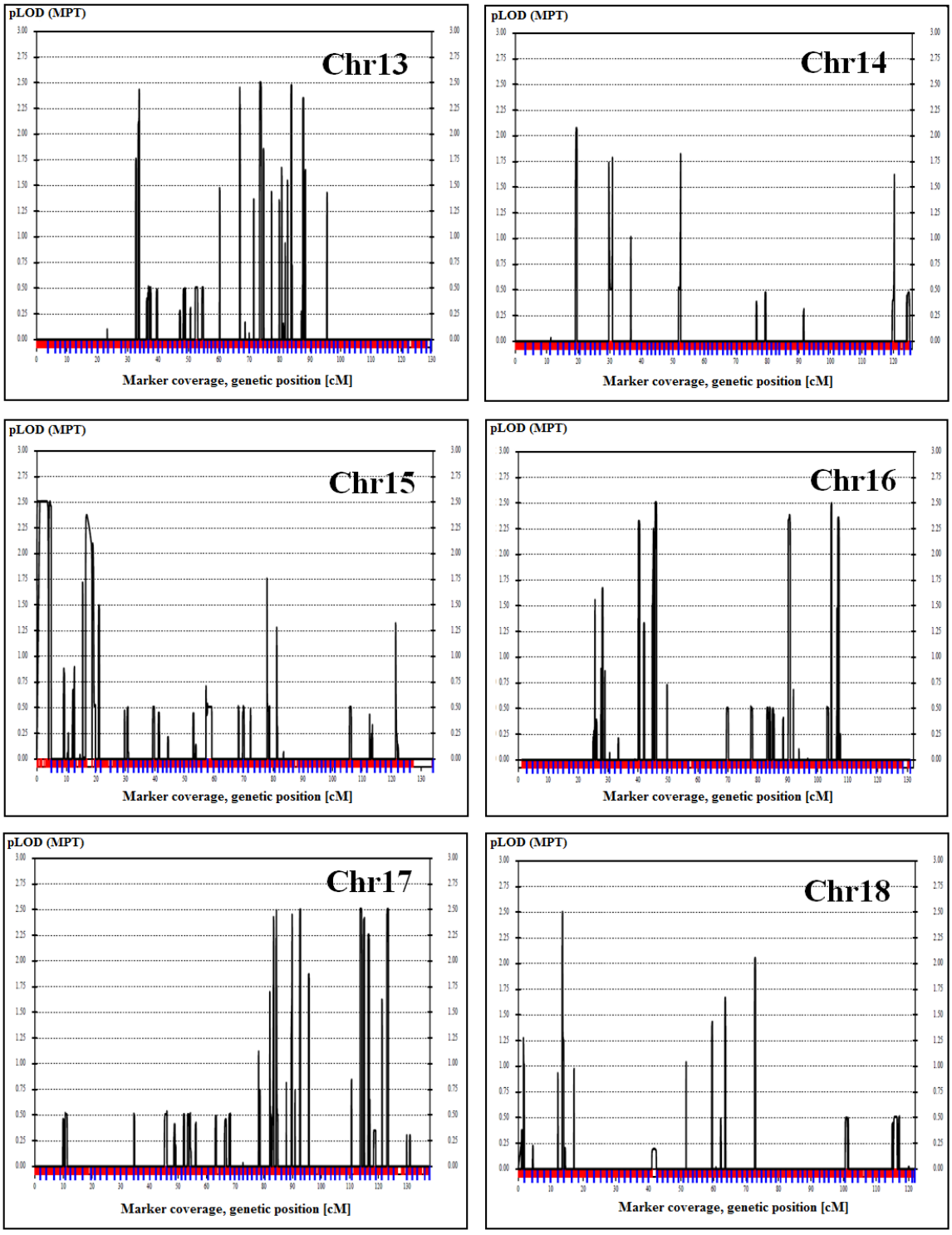


Figure 5.1 Multipoint LOD score results for ataxia for all autosomal chromosomes and PAR regions (cont.).

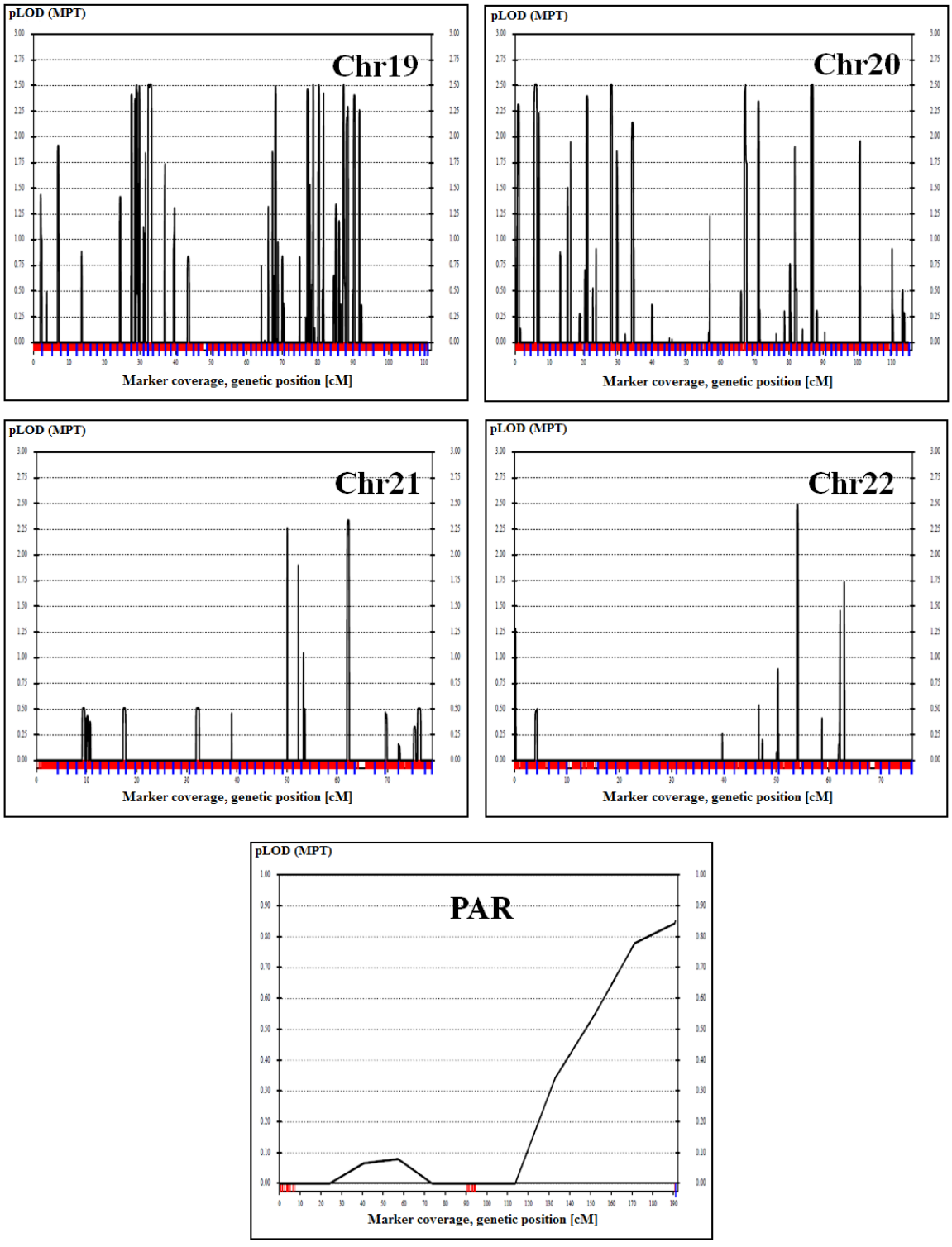


Figure 5.1. Multipoint LOD score results for ataxia for all autosomal chromosomes and PAR regions (cont.).

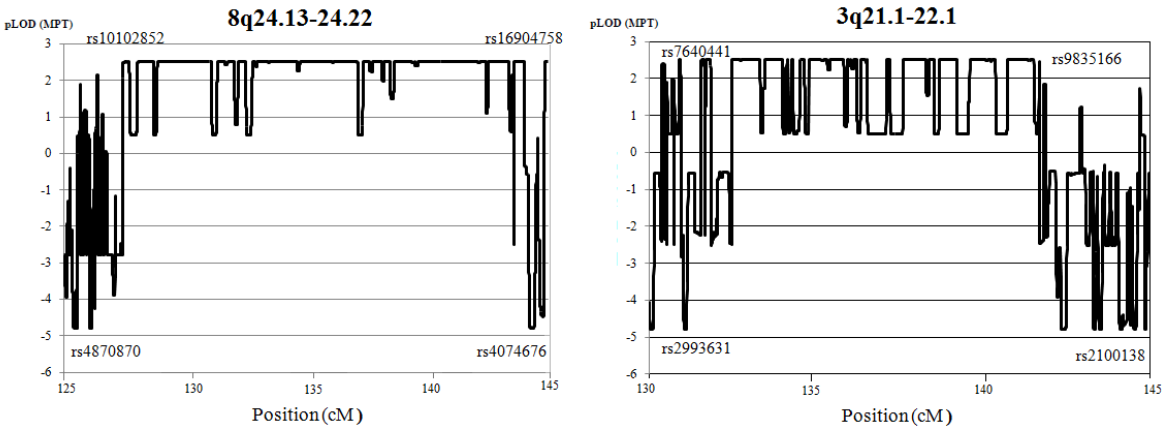


Figure 5.2. Multipoint LOD score curves at 8q24.13-24.22 and 3q21.1-22.1 for ataxia, using all markers in the region.

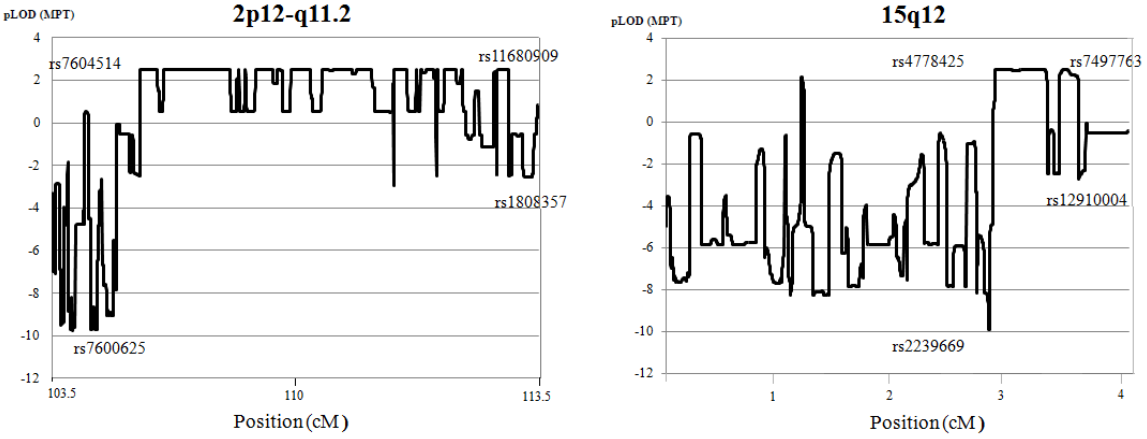


Figure 5.3. Multipoint LOD score curves at 2p12-q11.2 and 15q12 for ataxia, using all markers in the region.

Table 5.1. The strongest candidate loci for ataxia, listed in the order of size in cM.

Chromosome	LOD score	Flanking SNPs	Size	
			bp	cM
8	2.5076	rs4870870 (127.49 cM) - rs4074676 (143.79 cM)	9130701	16.3
3	2.5076	rs2993631 (132.31 cM) - rs2100138 (143.12 cM)	9493182	10.81
2	2.5076	rs7600625 (104.05 cM) - rs1808357 (112.72 cM)	16085698	8.67
15	2.5076	rs2239669 (0 cM) - rs12910004 (3.91 cM)	1094992	3.91

Fine calculations were performed in all four regions and the LOD score of 2.5076 was again obtained. The candidate region at 2p12-p11.2, flanked by rs7600625 (104.05 cM) and rs1808357 (112.72 cM), was spanning the centromere, and so the recombination frequency in this region is expected to be low. However, its genetic size was 8.67 cM, and therefore, it was included in further analyses. Loci 3q21.1-22.1 and 8q24.13-24.22 were assessed as stronger candidates due to their larger sizes. Nevertheless, all four regions were evaluated for exome sequencing.

5.1.2. Evaluation of homozygous regions

First homozygous regions were detected via Illumina Genome Viewer LOH bookmark for further validation of the shared homozygosity between affecteds. A filter of 500 SNPs was applied for extracting homozygous loci, and those loci were painted pink. In the second step, the candidate loci shared only by affected brothers were investigated via HCiE (Cetinkaya, 2010), and the homozygous genotypes at all four candidate loci mentioned above were found to be shared by the affected siblings and not the sister (Figures 5.4, 5.5, 5.6 and 5.7). The locus at 2p12-p11.2 had two small homozygous loci, suggesting the presence of two double crossovers (Figure 5.4). However, haplotypes were constructed and the small loci were found to be non-informative regions. In addition, chromosome 15 harbors two other loci with shared homozygosity, but the sizes of those regions are very small (Figure 5.7).

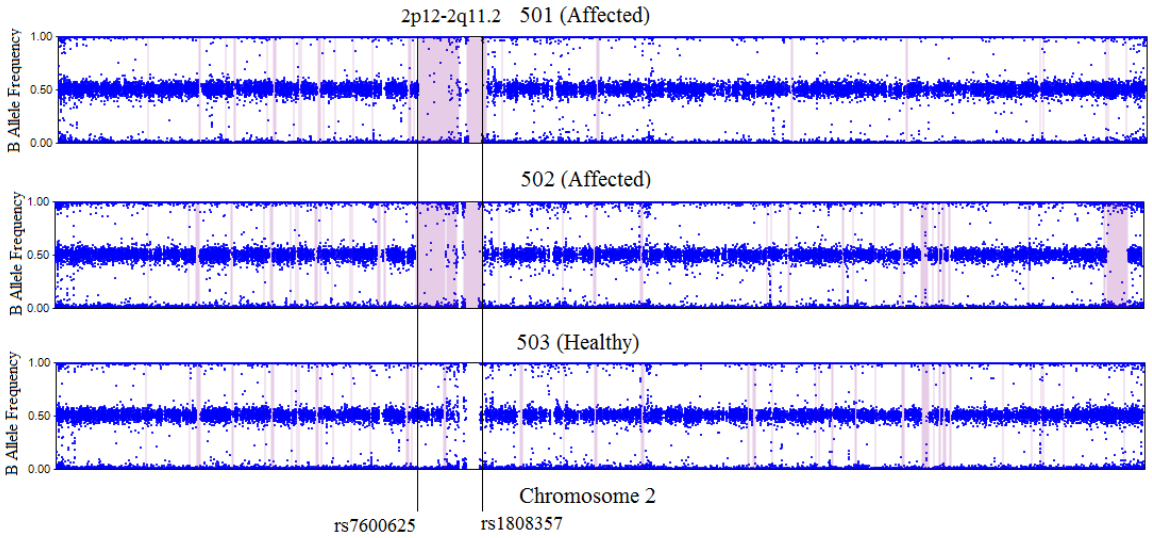


Figure 5.4. Homozygous regions at 2p12-q11.2 in ataxia.

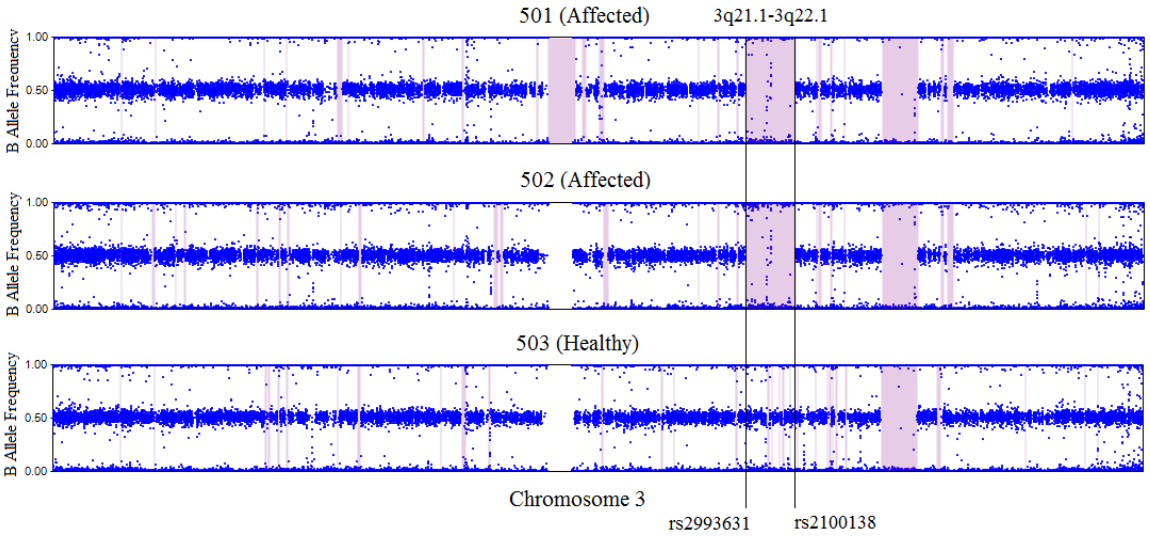


Figure 5.5. Homozygous regions at 3q21.1-22.1 in ataxia.

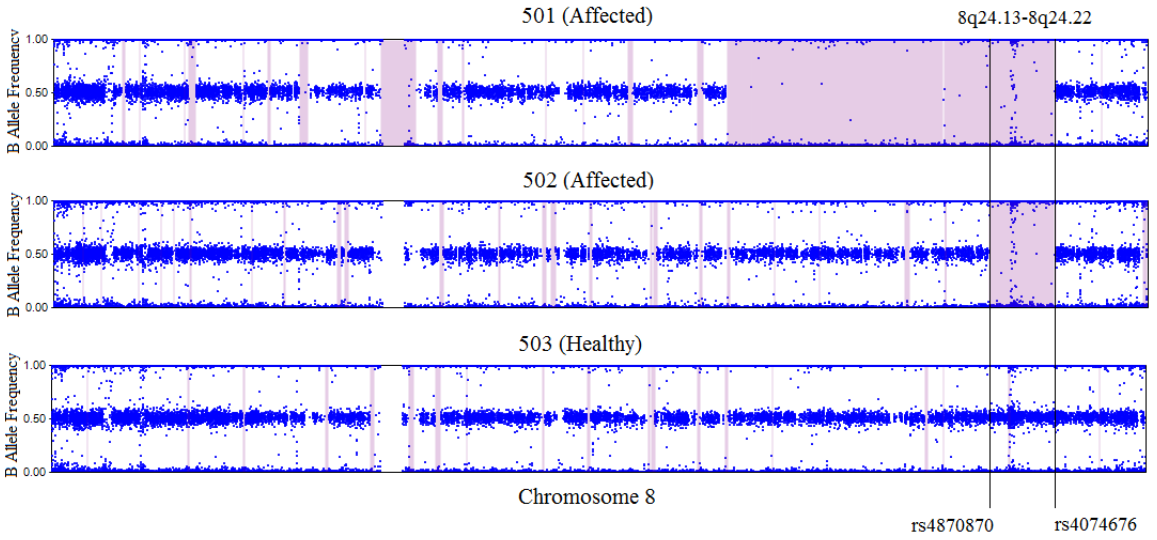


Figure 5.6. Homozygous regions at 8q24.13-24.22 in ataxia.

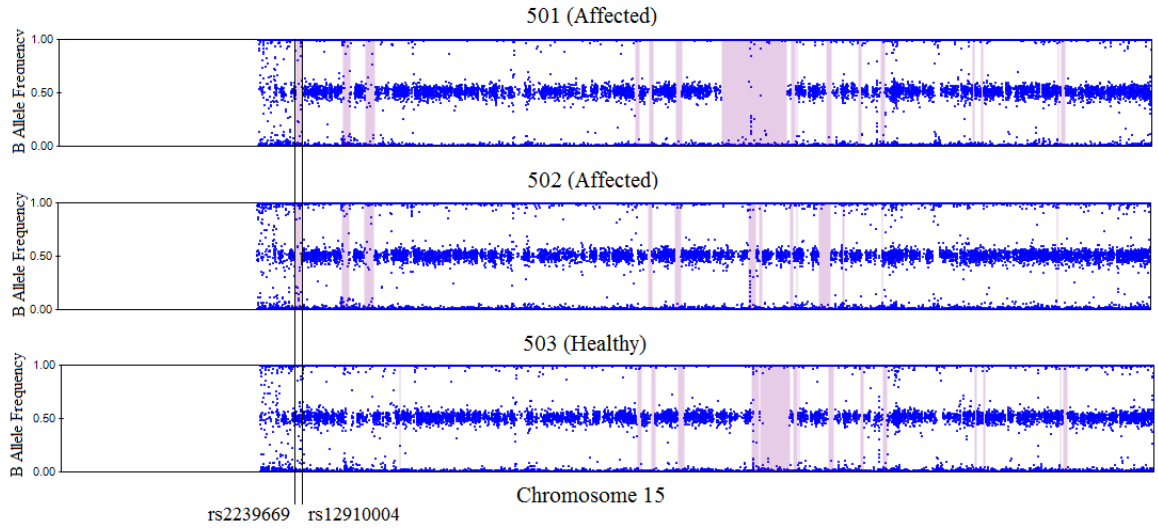


Figure 5.7. Homozygous regions at 15q12 in ataxia.

5.1.3 Exome sequencing and evaluation of the variants

Affected individual 501 was subjected to exome sequencing with Illumina TruSeq Exome Capture. Alignment of the reads, variant calling and annotation of the variants were performed at MacroGen Inc. (South Korea) with standard parameters. The sequencing statistics are given in Table 5.3 and Appendix A.

Table 5.2. Exome sequencing statistics in ataxic brother 501.

Target regions (bp)	62085286
% Coverage of target regions (more than 1X)	94.3%
% Coverage of target regions (more than 10X)	83.0%
Median read depth of target regions	44X
Mean read depth of target regions	51X
Number of total SNPs found	66,321
Number of coding SNPs found	18,689
Number of indels found	13,938
Number of coding indels found	750

The four strongest candidate regions obtained via linkage analysis were investigated for novel mutations harbored, and two candidate variants, one c.511C→T in *OSBPL11* and the other g.101976G→A in *COL6A6*, were observed in the region at 3q21.1-22.1 (Table 5.3). Another novel variant, *KIAA1257* c.223T→C was also found in the region, but the gene had an unknown function. This variant was predicted as damaging by online tool PolyPhen and was not observed in other exome sequencing results in our laboratory. Also a 23-bp deletion in *RHO* was found in 3q21.1-22.1, which was annotated as "splicing". But the deleted sequence was at the end of the 3' UTR and was observed in the results of most other exome sequencing results in our laboratory. Therefore, this variant is considered an artifact, since GC-rich regions such as UTRs are difficult to capture and sequence. Two novel nonsynonymous variants, *LRRTM1* c.1531A→C and *POLR1A* c.2801C→T, were detected at 2p12-11.2. Online tool PolyPhen predicted the effects of both variants on the protein function as neutral while the other two tools SIFT and MMB predicted *LRRTM1* c.1531A→C as damaging with low reliability and *POLR1A* c.2801C→T as neutral. Loci 8q24.13-24.22 and 15q12 did not harbor any novel exonic variant.

Table 5.3. Novel variants in the ataxia candidate regions obtained by exome sequencing. Only the variants with coverage >3 are given.

Location	Reference base	Observed base	Hom/ Het	Quality Score	Total Depth	Observed Base's Depth	Region	Gene
Chromosome 3, rs2993631 (124896668 bp) - rs2100138 (134389850 bp)								
124948050	TT	-	het	592	26	6	UTR3	ZNF148
124951172	TT	-	het	322	59	14	UTR3	ZNF148
124951173	T	-	het	322	59	8	UTR3	ZNF148
124998098	-	AAA	het	581	29	14	Intronic	ZNF148
125295188	G	A	hom	225	91	91	Exonic	OSBPL11, nonsynonymous SNV, c.511C→T, p.R171W
126174021	A	-	hom	241	8	7	UTR3	ZXDC
128292063	T	-	het	121	55	5	UTR3	C3orf27
128338859	A	-	het	282	82	6	UTR3	RPN1
128356613	G	T	hom	76	11	11	Intronic	RPN1
128711925	A	G	hom	57	8	8	Exonic	KIAA1257, nonsynonymous SNV, c. 223T→C p.C75R
128781258	-	T	het	153	9	4	Intergenic	-
128996445	T	-	het	70	44	6	UTR3	COPG
129254170	GGAATG GGAAAA ACCCCA ACTTT	-	hom	1403	30	19	Splicing	RHO
129367533	-	T	hom	2097	85	46	ncRNA	TMCC1
129369123	AA	-	het	905	101	10	ncRNA	TMCC1
129814807	G	-	hom	155	5	4	Intronic	ALG1L2
129947080	GAAA	-	hom	3455	52	45	ncRNA	COL6A4P2
129982990	G	-	hom	187	6	5	ncRNA	COL6A4P2
130381153	G	A	hom	225	64	64	Splicing	COL6A6
130447531	A	-	het	74	16	4	Intronic	PIK3R4
131186894	A	-	hom	729	31	21	Intronic	MRPL3
132179073	T	-	het	111	57	7	Intronic	DNAJC13
132363753	A	-	het	47	35	4	ncRNA	NPHP3-ACAD11
134204834	CCC	-	hom	1982	32	25	ncRNA	ANAPC13
Chromosome 8, rs4870870 (124862057 bp) - rs4074676 (133992758 bp)								
125325218	A	-	het	133	63	6	UTR3	TMEM65
125565176	C	G	hom	225	68	68	UTR3	MTSS1
130874615	A	-	het	20	44	4	Intronic	FAM49B
131064985	-	TG	hom	283	7	4	UTR3	ASAP1
131792607	-	AAAA	het	223	25	7	UTR3	ADCY8
133584735	C	G	het	73	11	6	Intronic	LRRC6

Table 5.3. Novel variants in the ataxia candidate regions obtained by exome sequencing (cont.). Only the variants with coverage >3 are given.

Location	Reference base	Observed base	Hom/ Het	Quality Score	Total Depth	Observed Base's Depth	Region	Gene
Chromosome 8, rs4870870 (124862057 bp) - rs4074676 (133992758 bp)								
133769555	GA	-	hom	2391	38	31	Intronic	TMEM71
133920669	A	-	hom	239	7	7	Intronic	TG
Chromosome 15, rs2239669 (23811592 bp) - rs12910004 (24906584 bp)								
20771609	T	C	het	39	71	31	ncRNA	GOLGA8C
20778594	TT	-	het	938	232	22	ncRNA	GOLGA8C
21935217	AAAAC	-	het	6603	147	91	ncRNA	LOC646214
21937747	-	AGAACC TTA	het	18412	372	169	ncRNA	LOC646214
21938209	CGGAGG CGC	-	het	481	39	8	ncRNA	LOC646214
22014407	T	-	het	211	36	6	Intergenic	-
22709885	G	-	hom	647	18	16	ncRNA	GOLGA8DP
23048393	AC	-	het	1876	56	27	UTR3	NIPA1
23261167	T	G	het	51	39	15	ncRNA	GOLGA8IP
23261170	T	C	het	46	34	12	ncRNA	GOLGA8IP
23283056	CCCGCTC TT	-	het	822	25	12	ncRNA	HERC2P2
Chromosome 2, rs7600625 (80335727 bp) - rs1808357 (96421425 bp)								
80529414	T	G	het	29	21	17	Exonic	LRRTM1, nonsynonymous SNV, c.1531A→C p.T511P
85537305	C	G	het	22	23	11	UTR3	TCF7L1
85772062	T	-	het	21	47	4	UTR3	MAT2A
86272825	G	A	hom	186	21	20	Exonic	POLR1A, nonsynonymous SNV, c.2801C→T p.S934L
86367351	A	-	het	348	62	4	UTR3	PTCD3
86442148	TT	-	het	329	51	4	UTR3	REEP1
86998660	T	-	hom	79	5	4	Intronic	RMND5A
89065483	A	G	het	49	24	14	ncRNA	ANKRD36BP2
89075987	-	ATTAAT CA	het	1104	50	16	ncRNA	ANKRD36BP2
89082296	-	TC	het	2248	190	36	ncRNA	ANKRD36BP2
89082340	ATTG	-	het	3287	209	52	ncRNA	ANKRD36BP2
95513688	-	TA	het	260	78	7	ncRNA	ANKRD20A8P
95942430	T	G	het	23	16	8	Intronic	PROM2

Candidate variants were further analyzed by realignment of the raw reads to the reference genome (hg19/GRCh37 assembly), allowing approximately two per cent base-pair mismatch in read alignments. In addition, the alignment was repeated with program GATK, but no novel variants were detected. Alignments were visualized and validated via program BamView. Novel variants *OSBPL11* c.511C→T and *COL6A6* g.101976G→A were further validated by Sanger sequencing in all siblings. Affected brothers were homozygous for both variants whereas the sister was a heterozygous carrier (Figure 5.8 and 5.9).

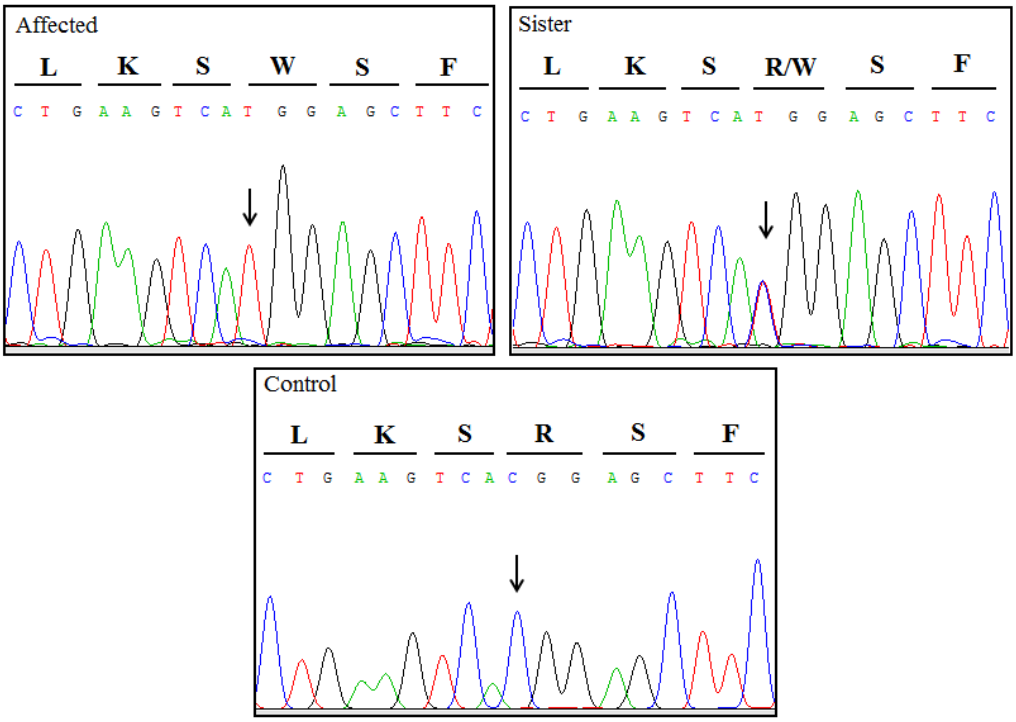


Figure 5.8. Chromatograms showing mutation *OSBPL* c.511C→T in an ataxic brother, the sister and the reference sequence.

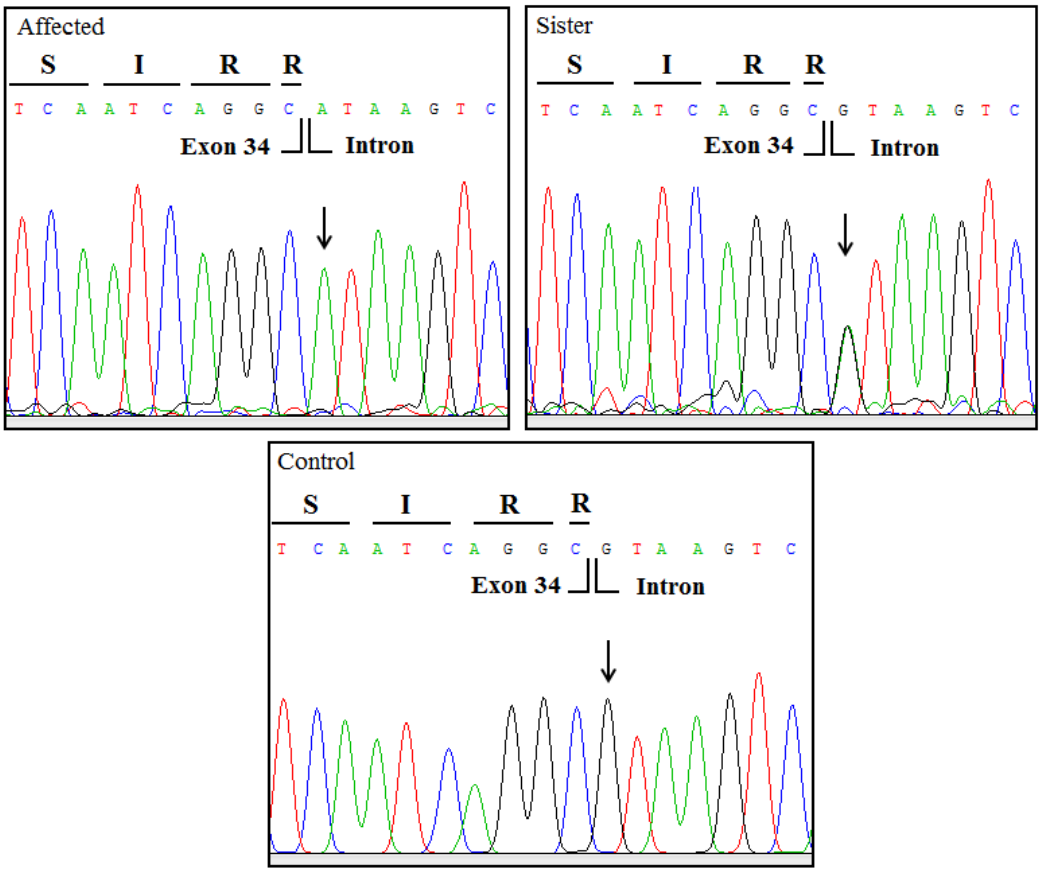


Figure 5.9. Chromatograms showing mutation *COL6A6* g.101976G→A in an ataxic brother, the sister and the reference sequence.

OSBPL11 c.511C→T (p.R171W) leads to the substitution of the positively charged, hydrophilic arginine at position 171 with non-polar, hydrophobic tryptophan. The effect of this change at the protein level was investigated with three online tools, and all predicted the mutation as strongly affecting protein function. The outputs of online tools are given at Table 5.4. In addition, arginine at position 171 was found to lie in a stretch of 48 amino acids ultra conserved in mammals (Figure 5.10).

Table 5.4. The effect of *OSBPL11* c.511C→T (p.R171W) mutation, as predicted by three online tools.

Online Tool	Score	Prediction
Polyphen2	1.00/1.00	Probably damaging
MMB	0.8498/1.00	Pathological
SIFT	0/0	Damaging

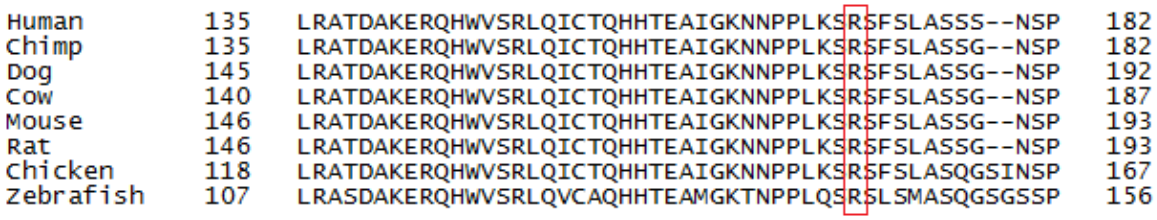
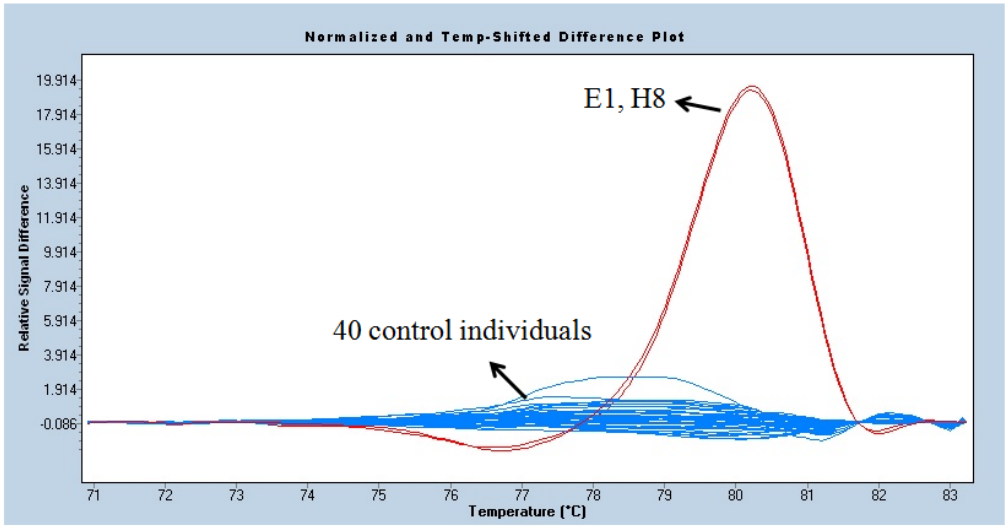
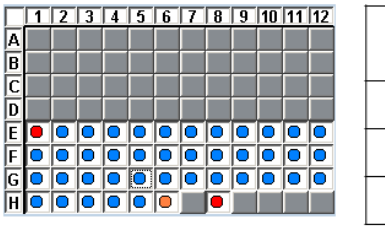


Figure 5.10. The conservation *OSBPL11* R171 across species

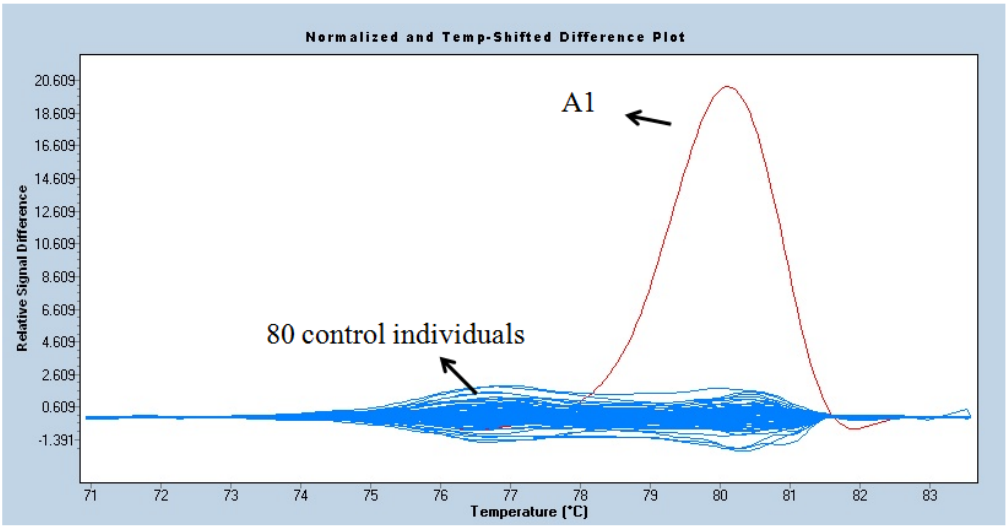
A population of 120 individuals was screened for *OSBPL11* c.511C→T using HRM analysis, to reach at least 80 per cent power to distinguish a sequence variant (Collins and Schwartz, 2002), and the mutation was not found in any. Melting curves of 120 individuals are presented in Figure 5.11.

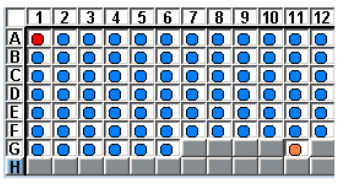




Well number	Genotyped individuals
E1 and H8	Heterozygote
E2 to H5	Population DNAs
H6	No DNA

Group1 Group2 Negative





Well number	Genotyped individuals
A1	Heterozygote
A2 to G6	Population DNAs
G11	No DNA

Group1 Group2 Negative

Figure 5.11. HRM analysis for *OSBPL11* c.511C→T mutation in 120 healthy individuals.

COL6A6 has 36 exons, and g.101976G→A variant affects the first nucleotide of intron 34. The guanine at this position is highly conserved across vertebrates (Figure 5.12), and Human Splicing Finder program predicts the variant to affect mRNA splicing.

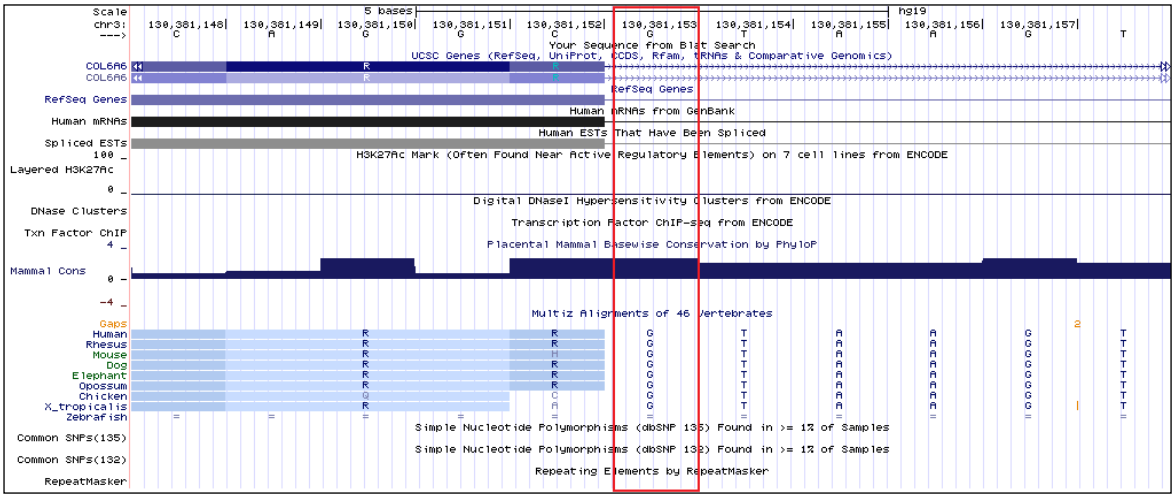


Figure 5.12. Chromosomal position and evolutionary conservation of *COL6A6* g.101976G→A (retrieved from UCSC Genome Browser).

A population of 120 healthy individuals was screened for *COL6A6* g.101976G→A with SSCP analysis, and the mutation was found in two individuals in the heterozygous state (P10 and P13, Figure 5.13). Due to this frequency of 2 in 240 chromosomes, the mutation was considered a variant in our population, and thus the gene was not considered as a candidate disease gene any longer.

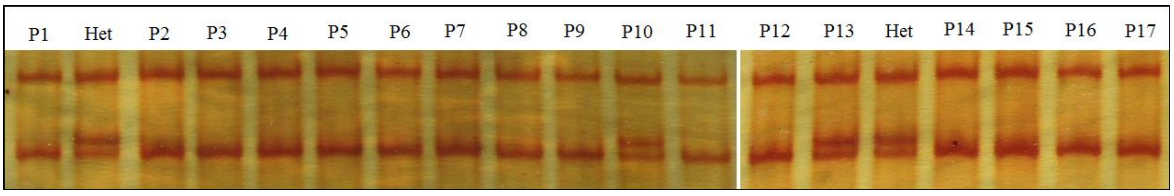


Figure 5.13. SSCP results of population screen of *COL6A6* 15 g.101976G→A in 15 healthy individuals. P1-17 denotes subjects from the population.

To investigate whether any good candidate gene was not included in the capture kit, genes covered with Illumina TruSeq capture kit were extracted. No gene that could

possibly be associated with the disease phenotype was found to be not included in the kit. Additionally, the coverage of each exon included in the kit was calculated with BEDTools, and coverage in exons within candidate regions was found sufficient. The exons with no coverage or incomplete coverage are given in Table 5.5.

Table 5.5. Exons with no/incomplete coverage at the ataxia candidate loci

Chromosome	Start	Stop	Gene and exon	Coverage	# of covered bases	Length (bp)	Fraction of covered bases
3	129612336	129612403	TMCC1, 5'UTR	0	0	67	0.0000000
8	133787603	133787891	PHF20L1, 5'UTR	1	17	288	0.0590278
2	88355144	88355248	KRCC1, 5'UTR	1	8	104	0.0769231

The exons with no or incomplete coverage were 5' UTR sequences of *TMCC1*, *PHF20L1* and *KRCC1*. These regions were also found to be incompletely covered in some other exome sequencing analyses performed in our laboratory. Thus, these regions are probably not captured sufficiently due to the experimental method rather than containing deletions, perhaps because they are GC-rich in sequence.

5.1.4 CNV Analysis

DNA sample of affected boy 501 was subjected to CNVs analysis using program Combined_CNV. Homozygous deletions that were not reported in Database of Genomic Variants were extracted, and no such deletions were located within the candidate loci. None was in a region with shared homozygosity with the brother. The novel homozygous deletions detected are presented in Table 5.6.

Table 5.6. Novel homozygous deletions in ataxic brother 501. The locations are according to NBCI36/hg18 assembly.

Chromosome	Start	Stop	Size (bp)	# of probes	Type	Genes at site	Shared homozygosity
2	76627301	76628901	1600	16	Intergenic	-	No
3	180050366	180050823	457	12	Intronic	-	No
11	11832323	11832825	502	11	Intronic	-	No
12	58256951	58257000	49	2	Intergenic	-	No

In summary, four candidate loci were found for ataxia. All were analyzed with exome sequencing. *OSBPL11* c.511C→T (p.R171W) and *COL6A6* g.101976G→A stood out as the best candidate mutations. Population screening for the latter revealed a frequency of 2 chromosomes in 240; therefore, the mutation was excluded as the disease mutation. All relevant genes within the candidate loci had been covered. No homozygous deletions were found within the candidate loci. *OSBPL11* c.511C→T (p.R171W) was assessed as the strongest candidate mutation, since it was novel, lying in an ultra conserved region of the protein and predicted to be damaging by all three online tools employed.

5.2 Azoospermia

Linkage analysis was performed using the data generated by the genome scans of four affected brothers with 610000 SNP markers, and two candidate disease loci were identified. Exome sequencing revealed the causative mutation, and thus the gene responsible for the disease was identified. All coding exons of the disease gene were scanned in a group of 45 azoospermic and 15 oligospermic individuals.

5.2.1 Multipoint Linkage Analysis

Genome scan data obtained by SNP genotyping was analyzed with program Allegro in easyLINKAGE package. Multipoint LOD scores were calculated assuming an autosomal recessive inheritance model with full penetrance. The highest LOD score of 3.0057 was obtained in a few loci. Multipoint LOD score results for all chromosomes are

given in Figure 5.14. The largest of such locus was found at 4p16.2-p16.1. The locus was flanked by rs4689888 (6.41 cM, 4569460 bp) and rs12500979 (12.12 cM, 6454012 bp) and was 1884552 bp in size. An investigation via HCiE (Cetinkaya, 2010) revealed that three of the affected brothers shared a homozygous genotype in a region flanked by rs4689888 (6.41 cM, 4569460 bp) and rs10939485 (26.22 cM, 13538956 bp) and was 8969496 bp in size. In the remaining affected brother, a crossover narrowed down the locus centromerically. An analysis for homozygosity via HCiE showed a larger locus at 18p11.21-q21.1, where three affected siblings were found to share a homozygous genotype in a 31-Mb region flanked by rs12959318 (41.48 cM, 11643985 bp) and rs719820 (68.34 cM, 42709494 bp). In the other affected brother two crossovers, one in the paternal and the other in the maternal allele, narrowed down the locus to 229,566 bp at 18q12.1, flanked by rs2339102 (52.69 cM, 23797853 bp) and rs1609839 (53.10 cM, 24027419 bp). LOD score graphics of the candidate loci at 18p11.21-q21.1 and 4p16.2-p16.1 are shown in Figure 5.15.

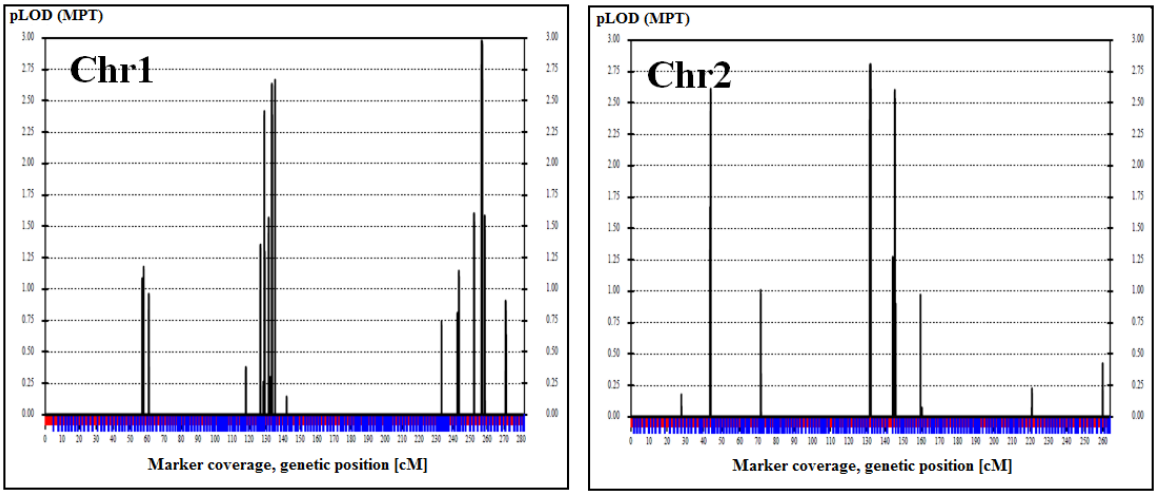


Figure 5.14. Multipoint LOD score results for azoospermia for all autosomal chromosomes and PAR regions.

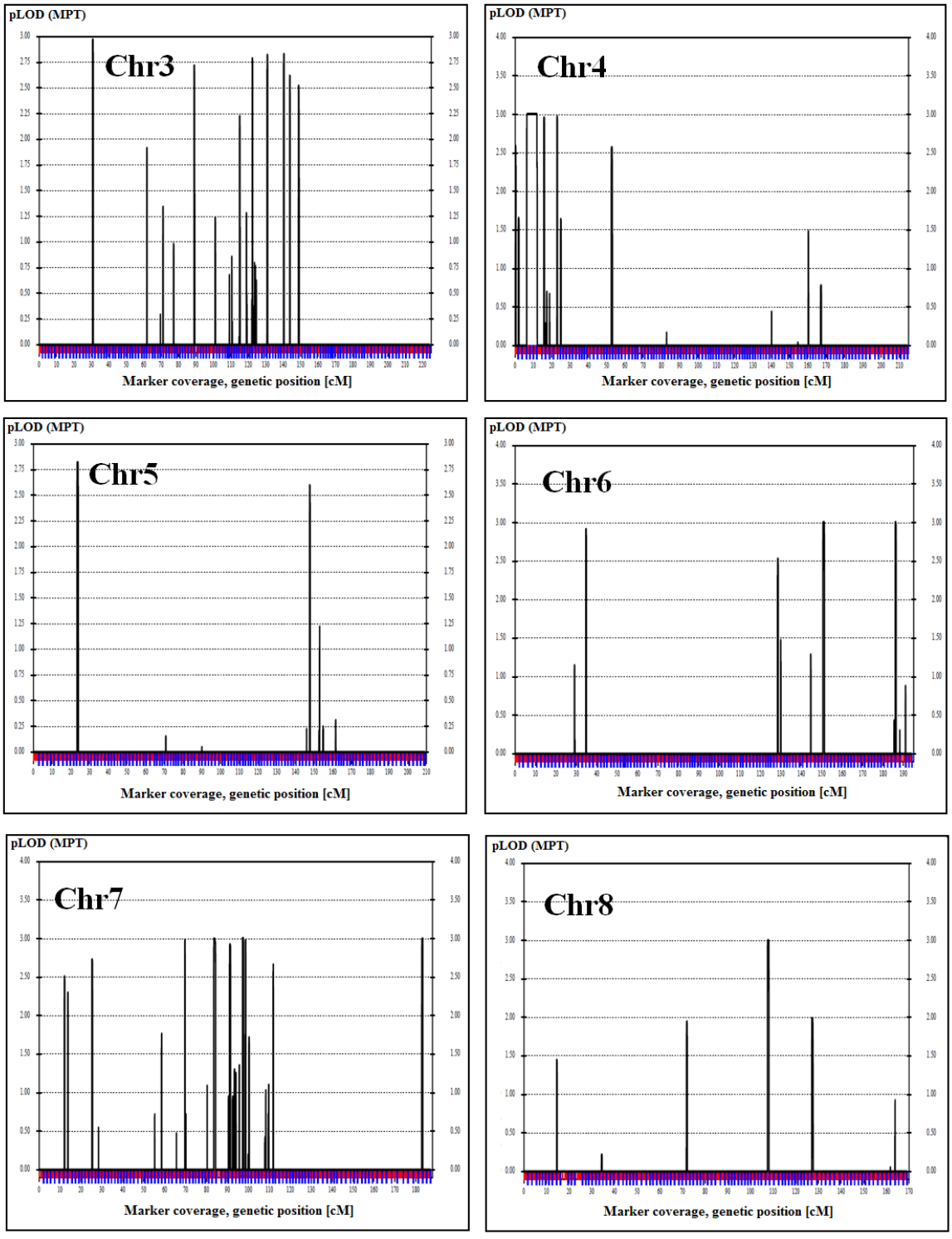


Figure 5.14. Multipoint LOD score results for azoospermia for all autosomal chromosomes and PAR regions (cont.).

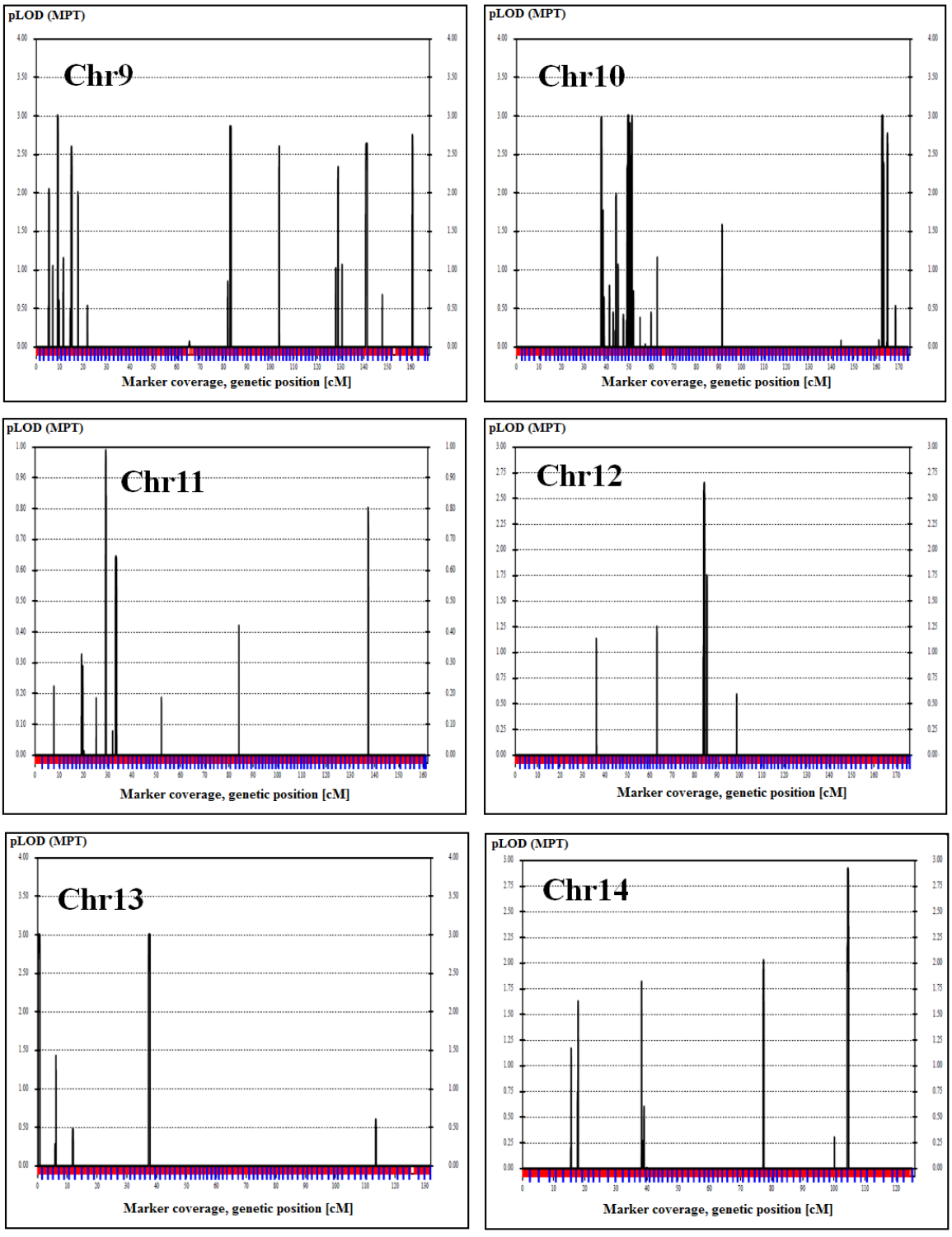


Figure 5.14. Multipoint LOD score results for azoospermia for all autosomal chromosomes and PAR regions (cont.).

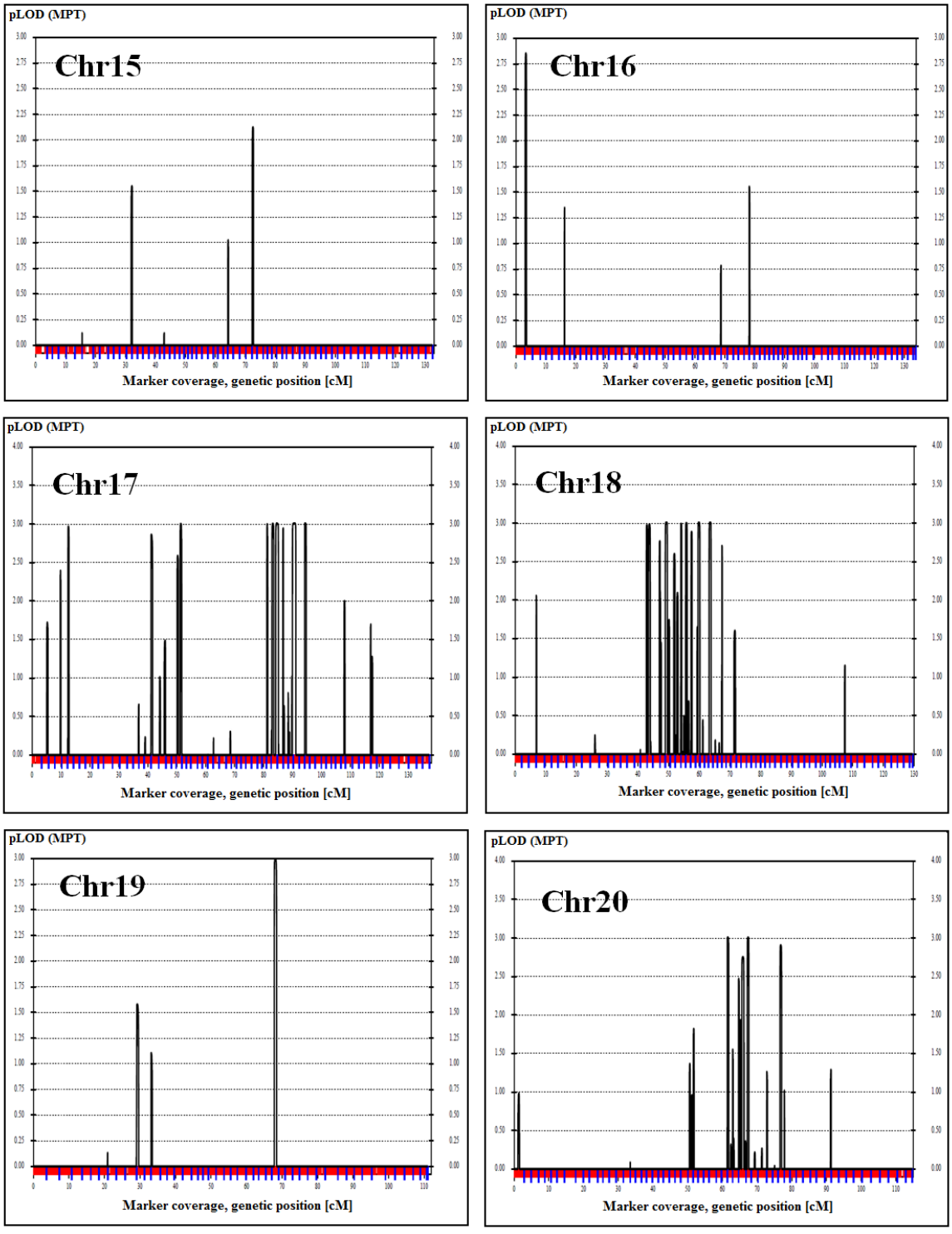


Figure 5.14. Multipoint LOD score results for azoospermia for all autosomal chromosomes and PAR regions (cont.).

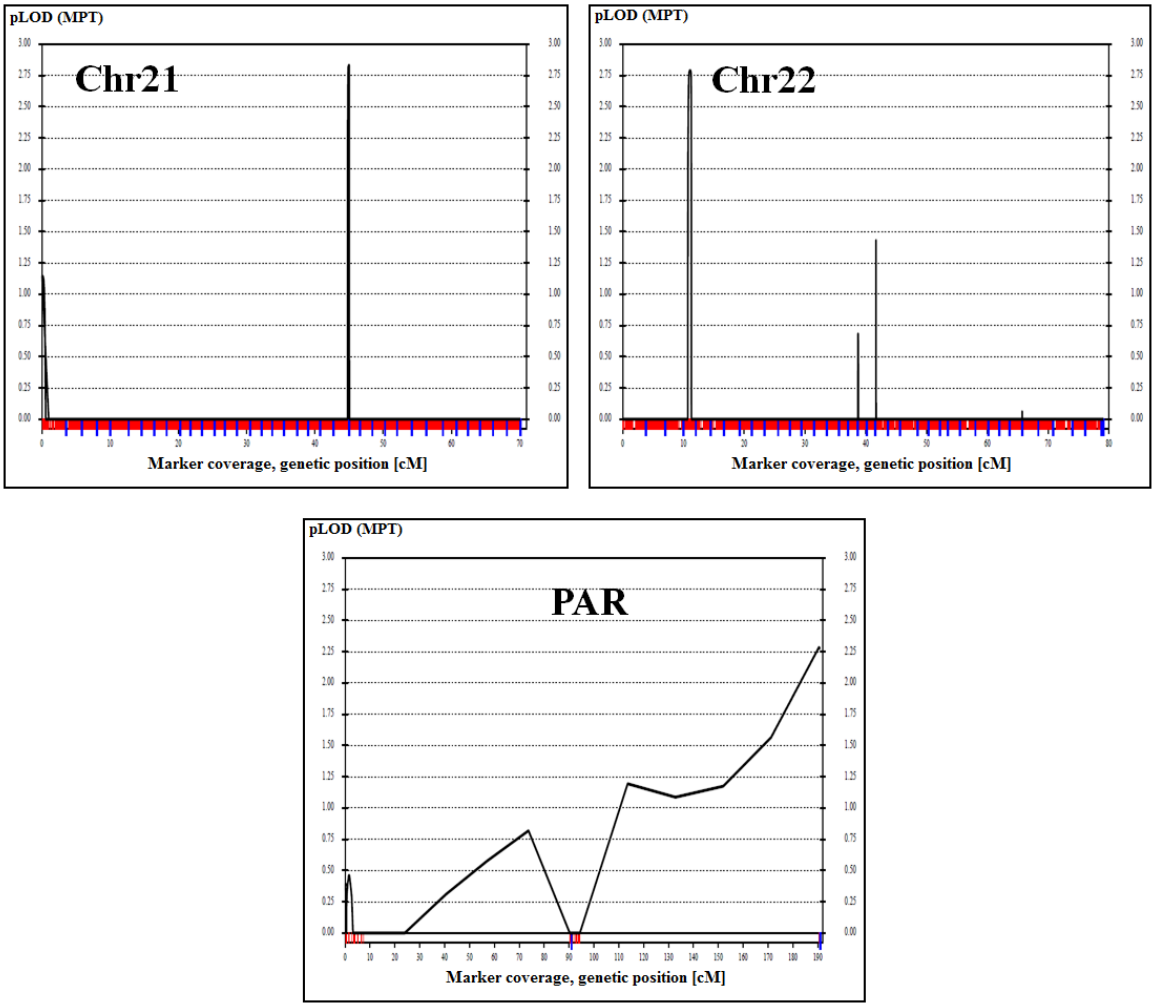


Figure 5.14. Multipoint LOD score results for azoospermia for all autosomal chromosomes and PAR regions (cont.).

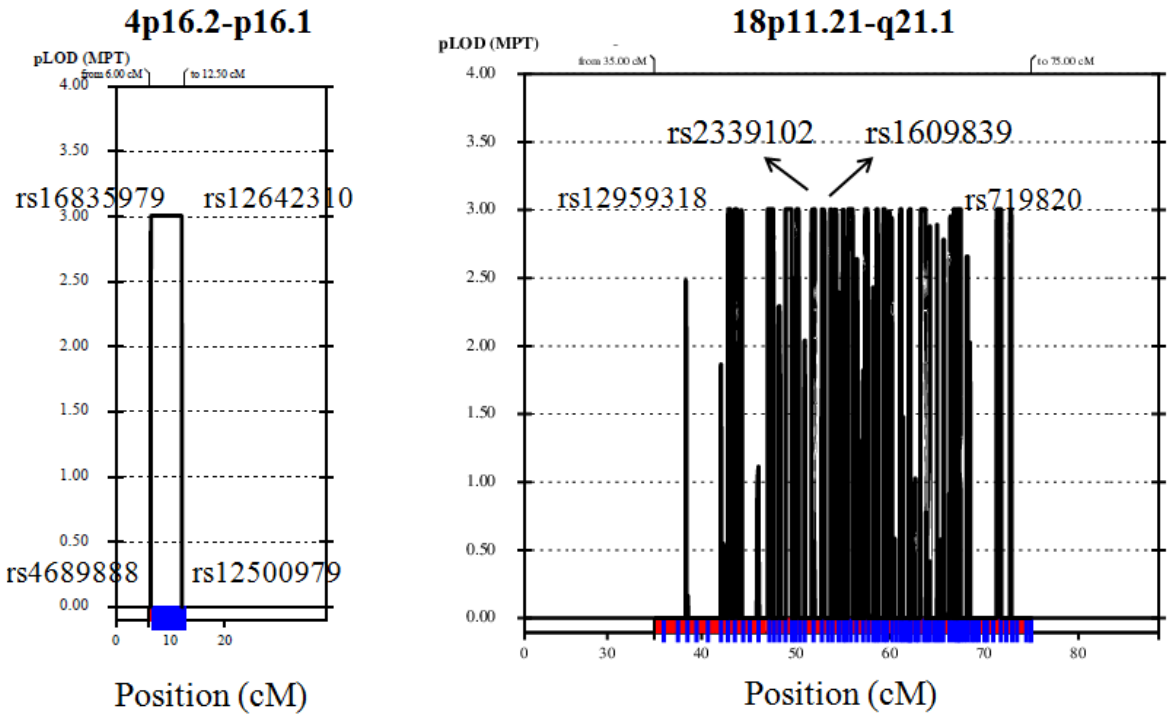


Figure 5.15. Multipoint LOD score curves at 4p16.2-p16.1 and 18p11.21-q21.1, using all markers included in the region.

Haplotypes were constructed using program GeneHunter in the easyLINKAGE package. The haplotypes of the four affected siblings at 4p16.2-p16.1 and 18q12.1 are given in Figure 5.16 and 5.17.

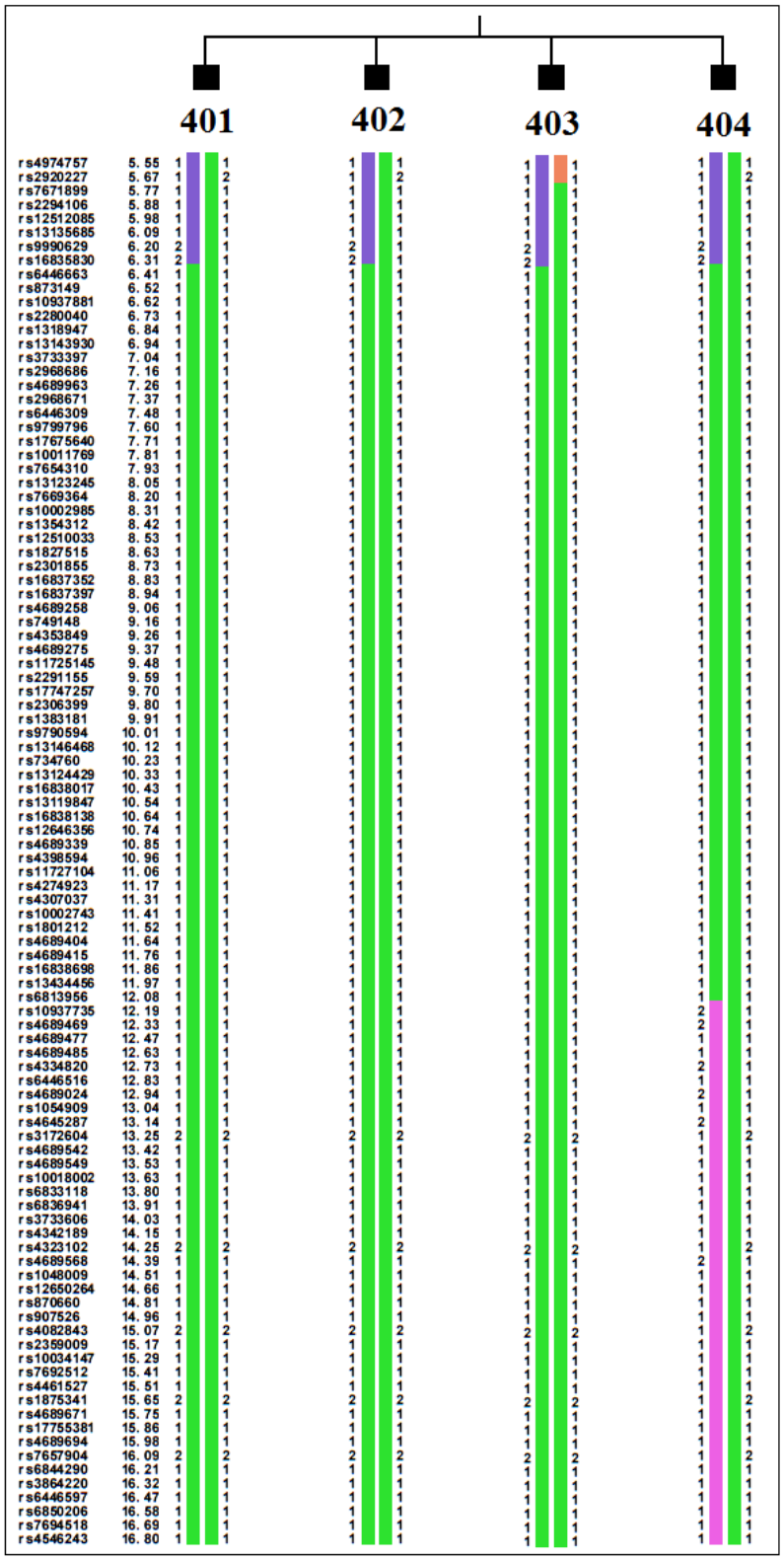


Figure 5.16. Haplotypes of four affected azoospermic brothers at 4p16.2-p16.1.

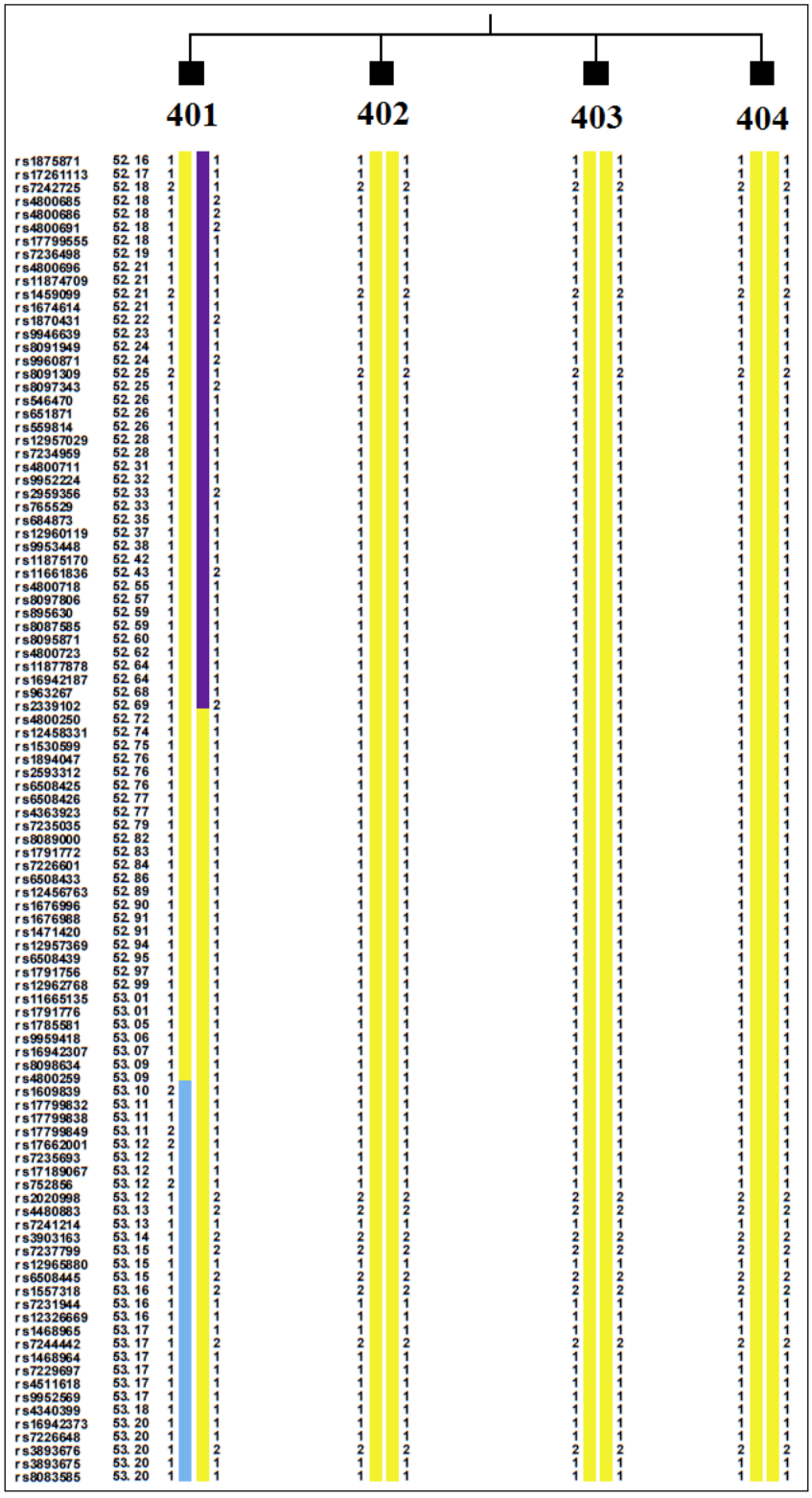


Figure 5.17. Haplotypes of four affected azoospermic brothers at 18q12.1.

5.2.2 Exome sequencing and evaluation of the variants

DNA sample of affected brother 501 was subjected to exome sequencing with Illumina TruSeq capture kit at MacroGen Inc. (South Korea). Alignment of the reads, variant calling and annotation of the variants were performed at the company with standard parameters. Sequencing statistics are given in Table 5.7 and Appendix A.

Table 5.7. Exome sequencing statistics in the azoospermia patient

Target regions (bp)	62085286
% Coverage of target regions (more than 1X)	92.9%
% Coverage of target regions (more than 10X)	78.7%
Median read depth of target regions	36X
Mean read depth of target regions	44.7X
Number of total SNPs found	61877
Number of coding SNPs found	17625
Number of indels found	12671
Number of coding indels found	491

Novel mutations in the candidate loci were selected. *TAF4B* c.1831C→T was the only novel mutation detected in the candidate region at 18q12.1 (Table 5.8). The candidate locus at 4p16.2-p16.1 was also investigated, but no variants that could possibly be associated with disease phenotype were observed (Table 5.8).

Table 5.8. Novel variants within the candidate loci in the azoospermia patient.

Location	Reference base	Observed base	Hom /Het	Quality Score	Total Depth	Observed Base's Depth	Region	Gene
Chromosome 18, rs2339102 (23797853 bp) - rs1609839 (24027419 bp)								
23873494	C	T	hom	194	85	84	Exonic, Splicing	TAF4B , Stopgain SNV, c.1831C→T, p.R611X
Chromosome 4, rs4689888 (4569460 bp) - rs12500979 (6454012 bp)								
4865686	TTTTAT	-	hom	567	12	8	Intergenic	MSX1
5755565	G	A	hom	196	31	31	Exonic	EVC, Nonsynonymous SNV, c.1369G →A
5813555	C	T	hom	101	29	23	UTR3	EVC

Raw reads generated by sequencing were aligned to the reference genome using BWA, and variant calling was performed with SAMTools. Next, reads covering the mutation *TAF4B* c.1831C→T were visualized with program BamView, and the mutation was presumed to be real, since it was not residing near the end of the reads, where sequencing error rates are high. Variant calling was also repeated with program GATK in order to avoid the exclusion of variants escaping from the algorithm of SAMTools, and no additional novel variants were detected. DNA samples of affected brothers, the healthy sister and a control individual were subjected to Sanger sequencing to validate the mutation. Affected brothers were homozygous for the mutation whereas the sister was a heterozygous carrier (Figure 5.18). The mutation is in the second last nucleotide of the exon 9, and results in the substitution of the cytosine with thymine. This in turn changes the arginine codon (CGA) to a stop codon (UGA). The protein is predicted to be truncated by 251 amino acids (native has 862).

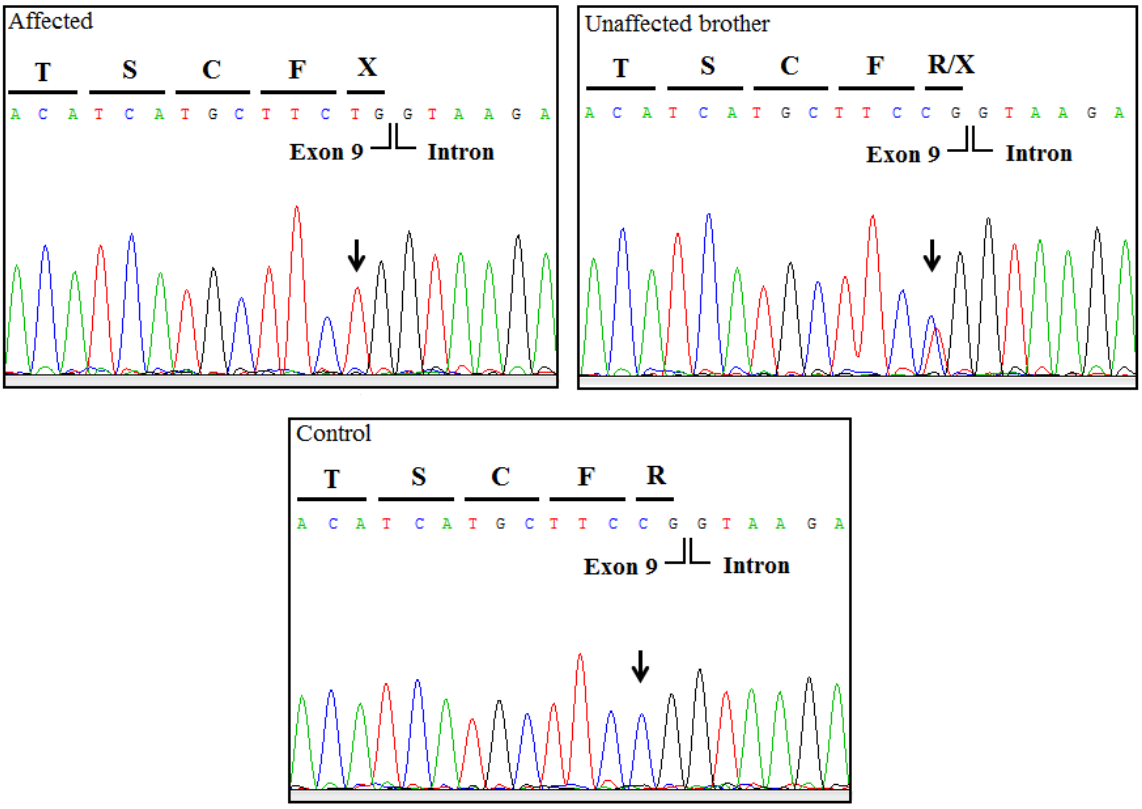


Figure 5.18. Chromatograms showing mutation *TAF4B* c.1831C→T in an azoospermic brother, the unaffected brother and the reference sequence.

Since c.1831C→T could possibly affect also a consensus splice site, Human Splicing Finder program was used to check whether the mutation has an influence on splicing mechanism. The program predicted that the mutation did not interfere with the splicing process.

Genes that were not included in the capture kit were extracted, and all relevant genes in the candidate loci were found to be included in the kit. The extent of coverage for each exon in the candidate loci was calculated via program BEDTools, and coverage in all exons of relevant genes in the candidate regions was found sufficient. In total, two exons at 4p16.2-p16.1 were found with no or incomplete coverage (Table 5.9). The exons with insufficient coverage were also found to be incompletely covered in other exome sequencing results in our laboratory.

Table 5.9. Exons with incomplete coverage within azoospermia candidate locus 4p16.2-p16.1.

Chromosome	Start	Stop	Gene and exon	Coverage	# of covered bases	Length (bp)	Fraction of covered bases
4	5894315	5894785	CRMP1, 5'UTR	0	0	470	0.0000000
4	6201979	6202318	JAKMIP1, 5'UTR	14	254	339	0.7492625

5.2.3 CNV Analysis

CNV analysis was performed via program Combined_CNV using the SNP genotyping data of an azoospermic brother. Novel homozygous deletions that are absent in Database of Genomic Variants (DGV) were extracted, and none was within the candidate loci. The two novel homozygous deletions detected using program Combined_CNV are presented in Table 5.10.

Table 5.10. Novel homozygous deletions in an azoospermic brother. The locations are according to NBCI36/hg18 assembly.

Chromosome	Start	Stop	Size (bp)	Number of Probes	Type	Genes at site	Shared homozygosity
4	19130834	19144341	13507	4	Intergenic	-	No
5	27462485	27462654	169	3	Intergenic	-	No

5.2.4 Population screening

The family members and 120 healthy individuals were screened for *TAF4B* c.1831C→T with SSCP analysis (Figure 5.19). The mutation segregated with the trait in the family and was not found in a healthy population of 120 individuals.

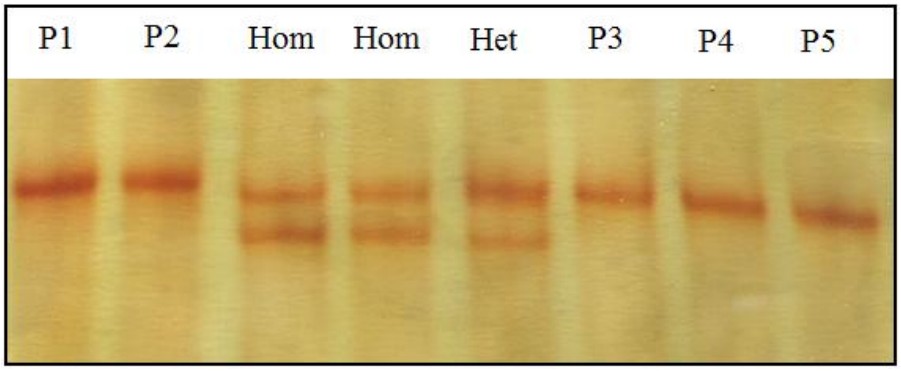


Figure 5.19. An example of SSCP results of population screening for *TAF4B* c.1831C→T. P1-5 denote individuals from the population.

In a group of 45 azoospermia and 15 oligospermia patients, all coding exons of *TAF4B* were screened for mutations via HRM analysis using primers with sequences presented in Table 4.6. The samples were spiked with control DNA prior to analysis, and samples showing aberrant melting behavior were subjected to Sanger sequencing, at least 25 nucleotides into the introns. Two regions TAF4B_EX1A and TAF4B_EX10 could not be amplified with HRM, and thus, were analyzed with SSCP. The alleles of TAF4B_EX1A could not be resolved on SSCP gels likely due to the high GC-content of the region, and the analysis of this region is inconclusive. However, the alleles of TAF4B_EX10 region

was resolved on SSCP gels, and none of the individuals screened displayed an aberrant migration pattern.

No homozygous *TAF4B* mutations were observed in the group, but two novel heterozygous variants were detected in individuals AZO35 and OLI8. The results of HRM analysis, the list of the variants found, the effect of nonsynonymous variants predicted by Polyphen and minor allele frequencies (MAFs) for known SNPs are given in Table 5.11.

Table 5.11. HRM analysis results for *TAF4B* in 45 azoospermia and 15 oligospermia patients and the variants found.

Region	**HRM analysis results**	**Samples subjected to sequencing**	**Sequencing results**
TAF4B_EX1A	-	-	-
TAF4B_EX1B	3 groups with different melting behavior	AZO9, AZO25, AZO27	No sequence difference
TAF4B_EX2	1 group with different melting behavior	AZO10, AZO12, AZO24, AZO25, OLI12	No sequence difference
TAF4B_EX3	2 groups with different melting behavior	Group 1: AZO9, AZO10	No sequence difference.
		Group 2: AZO35	Novel heterozygous missense c.544G>T, p.Ala182Ser. Polyphen score: 0.001 (Benign)
TAF4B_EX4	1 group with different melting behavior	AZO1, AZO14, AZO40	AZO1 and AZO14: Heterozygous rs17224558 (synonymous), MAF: 0.118
			AZO40: No sequence difference
TAF4B_EX5	2 groups with different melting behavior	Group 1: AZO24	AZO24: Heterozygous rs147920680 (intronic), MAF: unknown
		Group 2: AZO31, AZO21	No sequence difference
TAF4B_EX6	3 groups with different melting behavior	Group 1: AZO1, AZO12	No sequence difference
		Group 2: AZO6, AZO26, AZO43	
		Group 3: OLI5	

Table 5.11. HRM analysis results for *TAF4B* in 45 azoospermia and 15 oligospermia patients and the variants found (cont.).

Region	**HRM analysis results**	**Samples subjected to sequencing**	**Sequencing results**
TAF4B_EX7A	3 groups with different melting behavior	Group 1: AZO1,AZO3, AZO35, AZO36	No sequence difference
		Group 2: AZO12, AZO22, AZO29, AZO31,	
		Group 3: AZO42, AZO46	
TAF4B_EX7B	3 groups with different melting behavior	AZO1, AZO4, AZO7, AZO31, OLI5	AZO1, AZO4, AZO7, OLI5: Heterozygous rs74947492 (missense) p.Val438Leu. Polyphen score: 0.278 (Benign), MAF: 0.053
			AZO31: No sequence difference
TAF4B_EX7C	1 group with different melting behavior	AZO14, AZO18, AZO23	Heterozygous rs74947492 (missense) p.Val438Leu. Polyphen score: 0.278 (Benign) MAF: 0.053
TAF4B_EX7D	1 group with different melting behavior	AZO24, AZO31, AZO43, OLI1, OLI10	Heterozygous rs3744961 (synonymous) MAF: 0.125
TAF4B_EX8	4 groups with different melting behavior	Group 1: AZO1, AZO9	No sequence difference
		Group2: AZO6, OLI7	Homozygous rs12963653 (missense) p.Asn539Ser. Polyphen score 0.000 (Benign), MAF: 0.149
		Group 3: AZO8	Heterozygous rs12963653 (missense) p.Asn539Ser. Polyphen score 0.000 (Benign), MAF: 0.149
		Group 4: AZO26, AZO31, AZO46, OLI11	No sequence difference

Table 5.11. HRM analysis results for *TAF4B* in 45 azoospermia and 15 oligospermia patients and the variants found (cont.).

Region	**HRM analysis results**	**Samples subjected to sequencing**	**Sequencing results**
TAF4B_EX9	5 groups with different melting behavior	Group 1: AZO3, AZO19	Homozygous rs1677016 (synonymous), MAF: 0.466
		Group 2: AZO31, OLI8	AZO31: No sequence difference
			OLI8: Novel heterozygous missense c.1814A>G, p.Asn605Ser. Polyphen score: 0.495 (Possibly damaging) and Heterozygous rs1677016 (synonymous), MAF: 0.466
		Group 3: AZO46, OLI17	AZO46 and OLI17: Heterozygous rs1677016 (synonymous), MAF: 0.466
		Group 4: OLI2	Homozygous rs1677016 (synonymous) MAF: 0.466
		Group 5: OLI7	No sequence difference
TAF4B_EX10		-	-
TAF4B_EX11	No different melting behavior	-	-
TAF4B_EX12	4 groups with different melting behavior	Group 1: AZO6, AZO35, AZO44, AZO45.	No sequence difference
		Group 2: AZO33, AZO36, AZO37, AZO38, AZO39, AZO40, AZO42, AZO43.	No sequence difference
		Group 3: AZO27, OLI14, OLI15, OLI17.	No sequence difference
		Group 4: AZO31	No sequence difference
TAF4B_EX13	No different melting behavior	-	-

Table 5.11. HRM analysis results for *TAF4B* in 45 azoospermia and 15 oligospermia patients and the variants found (cont.).

Region	**HRM analysis results**	**Samples subjected to sequencing**	**Sequencing results**
TAF4B_EX14	3 groups with different melting behavior	Group 1: AZO3	No sequence difference
		Group 2: AZO21	
		Group 3: OLI2	
TAF4B_EX15	2 groups with different melting behavior	Group 1: AZO7	No sequence difference
		Group 2: OLI6	

In summary, two candidate loci for azoospermia 4p16.2-p16.1 and 18q12.1 were found. Exome sequencing in one affected individual facilitated the identification of *TAF4B* c.1831C→T, a stopgain mutation, at 18q12.1. No other novel exonic variants and no homozygous deletions were observed within the candidate loci. All relevant genes were found to be sufficiently covered by exome sequencing. *TAF4B* c.1831C→T (p.R611X) was predicted to truncate the protein and was not found in a population of 120 individuals; therefore, *TAF4B* was assessed as the gene responsible for azoospermia. In addition, all coding exons of the gene except the first exon were screened for mutations in a cohort of 45 azoospermia and 15 oligospermia patients, but no mutations were found.

5.3 PMED

Previously in our laboratory, the disease locus was mapped to 16pter-p13.3 (Duru, 2006; Duru *et al.*, 2010). In this study, the causal mutation was identified in a gene responsible from a similar disease. The mutation was found to affect only one of the two known transcript isoforms. In addition, two novel transcript isoforms of the disease gene were identified. Relative expression of three transcript isoforms was investigated in seven different brain regions.

5.3.1 Mutation Screening

While this study was in progress, *TBC1D24* was reported to be mutated in two families having familial infantile myoclonic epilepsy (FIME, MIM 605021) (Corbett *et al*., 2010, Falace *et al*., 2010). The phenotypes of those families were much milder than that of PMED family studied in our laboratory. Nonetheless, the gene was considered a strong candidate because it was at the gene locus and all families had the common clinical feature of myoclonus with onset at 2 months. All known exons and flanking sequences of the gene were Sanger sequenced in an affected individual for mutations. The extent of sequencing is presented in Table 5.12.

While sequencing was in progress, a new transcript isoform of *TBC1D24* (NM_001199107.1) was reported in CCDS database (09/07/2011, CCDS ID: 55980). The isoform contained an 18-bp exon that was absent in the previously reported isoform NM_020705.2. Then the 18-bp exon (Exon 3 in Table 5.12) was sequenced in an affected individual, and a homozygous c.969_970delGT (p.S324TfsX3) mutation was identified (Figure 5.20).

Table 5.12. Extent of sequencing in *TBC1D24* in a PMED patient.

Region	Sequenced part
Exon 1	-191 and +142
Exon 2	-279 and +48
Exon 3	-48 and +62
Exon 4	-48 and +156
Exon 5	-102 and +68
Exon 6	-26 and +66
Exon 7	-151 and +94
Exon 8	-187and +312

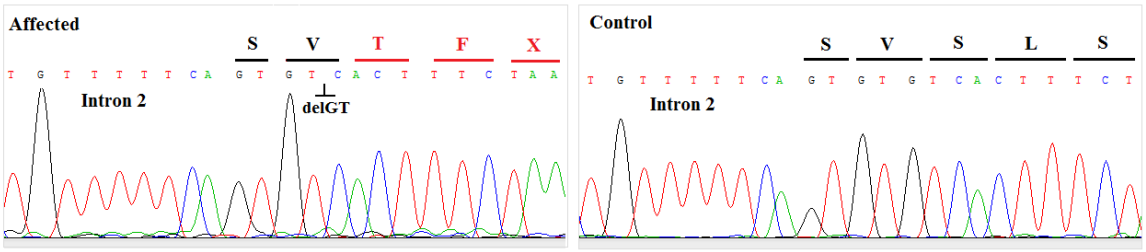


Figure 5.20. Chromatograms showing c.969_970delGT (p.S324TfsX3) mutation in a PMED patient and the reference sequence.

DNA samples of all family members were subjected to SSCP analysis in order to screen for the mutation, and the mutation was found to segregate with the disease. The same method was used also to screen a population of 120 unrelated individuals from the population, and the mutation was not found in any.

5.3.2 Isoform Identification

RNA levels of different isoforms of *TBC1D24* were compared in a cDNA panel of seven brain regions via quantitative real-time PCR. In the first attempt, primers specific to isoform 1 (NM_001199107.1) (spanning exon 3 and 4) and isoform 2 (NM_020705.2) (spanning exon 2 and 4) were designed and the products of quantitative real-time PCR were subjected to agarose gel electrophoresis. Two aberrant bands were observed (Figure 5.21).

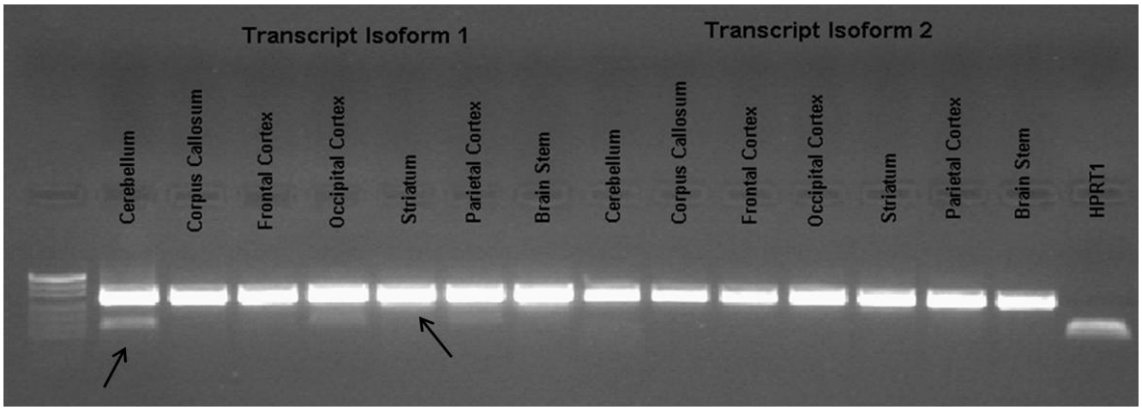


Figure 5.21. Amplification products of the initial real-time PCR reactions. The aberrant bands are indicated with arrows.

The aberrant bands were extracted from the gel and subjected to DNA sequencing. The results showed that the bands were specific to novel transcript isoforms (Figures 5.22 and 5.23), which we designated as isoform 3 and isoform 4. Their structures were investigated further by direct sequencing of cDNAs amplified with primers specific to each one and extending until the 3'UTR sequences. The transcripts and the deduced encoded proteins are given in Figure 5.24.

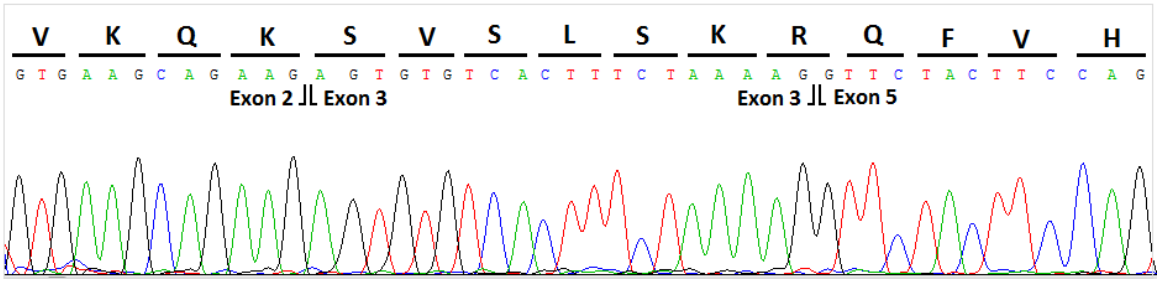


Figure 5.22. Chromatogram showing *TBC1D24* novel transcript isoform 3.

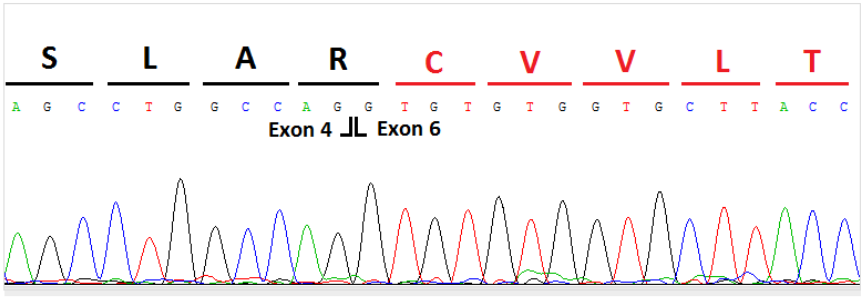


Figure 5.23. Chromatogram showing *TBC1D24* novel transcript isoform 4. Residues coded in a different frame than in other isoforms are indicated in red.

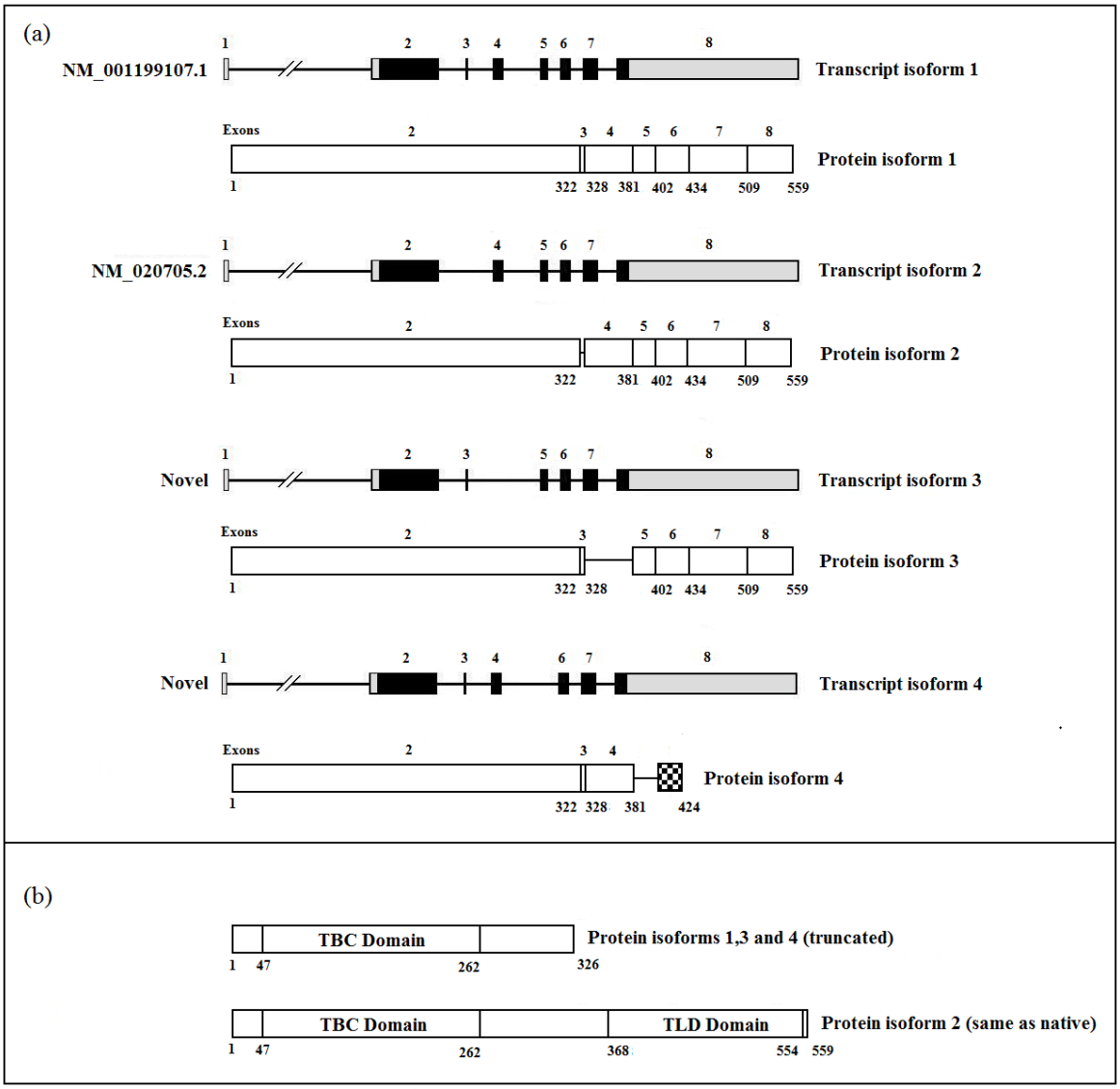


Figure 5.24. The deduced structures of *TBC1D24* transcript and protein isoforms.

(a) Coding sequences in transcripts are shown in black and noncoding in grey Last amino acids in each exon are given and excluded residues are indicated with lines in protein isoforms. Carboxyl terminus of protein isoform 4 that is coded in a different frame than in the other isoforms is shown as a dotted box.

(b) Effects of p.S324TfsX3 mutation on protein isoforms.

5.3.3 Relative Quantification of *TBC1D24* transcript isoforms

The levels of *TBC1D24* transcript isoforms 1 and 3 relative to transcript isoform 2 were determined via quantitative real time PCR (Figure 5.25). In all brain regions, isoform 1 was found more abundant than isoform 2. The expression of isoform 1 was highest in cerebellum and moderately high in frontal cortex and occipital cortex. The expression pattern of novel isoform 3 was similar to isoform 1, being highest in cerebellum and moderately high in frontal cortex and occipital cortex.

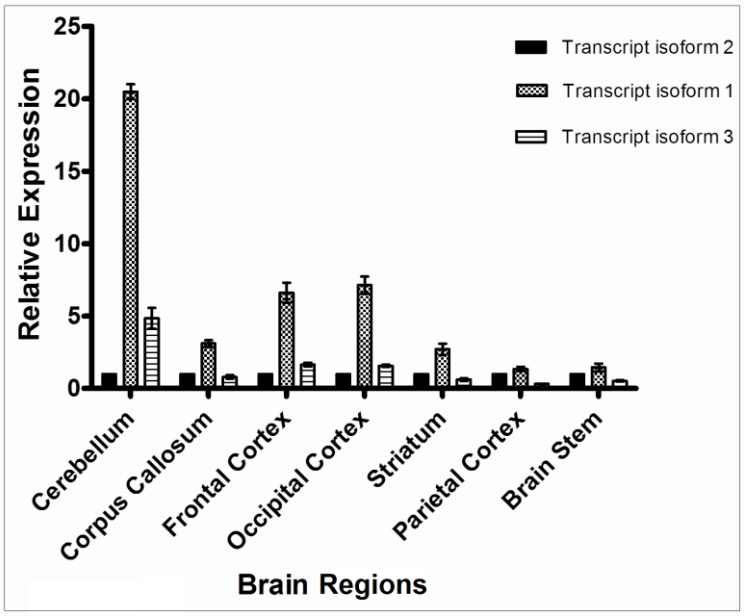


Figure 5.25. The relative expression levels of three of the *TBC1D24* transcript isoforms. All assays were normalized to *HPRT1,* and isoform 2 was set as the calibrator. The relative abundances of isoform 1 and isoform 3 are represented with means ± SEM.

In summary, *TBC1D24* was sequenced in an affected individual, and mutation c.969_970delGT (p.S324TfsX3) was identified. The mutation is predicted to affect only one of the reported isoforms. In addition, the expression of different *TBC1D24* isoforms were investigated in cDNAs of seven brain regions, and meanwhile two novel isoforms were identified. The expression of isoform 1 and 3 were found higher than the expression of isoform 2.

6. DISCUSSION

With the progress in the last few years in technologies of genotyping arrays and next-generation sequencing, identification of novel monogenic disease genes became more efficient. High density SNP arrays facilitate the detection of a large number of CNVs at a very low cost. Subsequent linkage analysis results in the localization of a disease gene in a large family. Since the disease causing mutations mostly reside in the coding regions of the genome, exome sequencing is a very suitable method for identifying disease genes, once the locus is known.

In this study, genes responsible for two diseases are identified and a mutation for another disease was found. In ataxia and azoospermia, the disease loci were found by linkage analysis and the disease genes were identified via exome sequencing. The variants generated by exome sequencing were evaluated on the bases of their locations with respect to coding sequences, effects on protein function, the reported functions of the genes and the tissues the genes are expressed in. In PMED, the mutation was found by direct sequencing of a candidate gene and relative RNA quantification is performed to further assess gene function.

6.1 Ataxia

The genomes of the two affected brothers and their healthy sister were scanned with one million SNPs, and multipoint linkage analysis was performed assuming autosomal recessive inheritance with full penetrance. LOD scores did not reach 3 because of the small size of the family. Simulation programs such as Allegro and Slink can calculate the power of such families; however, the presence of consanguinity loops in the family creates computational limitations for such an analysis. Also, a maximum LOD score of 2.5076 was obtained at numerous very small regions (< 2 cM) throughout the genome, as presented in Figure 5.7. Those small regions were investigated via HCiE and

homozygosities were found to be identical in state and not by descent. Generally, a large number of shared haplotypes in an inbred community generate high LOD scores in many small regions that are not identical by descent from a recent common ancestor.

As for the larger loci, linkage analysis revealed four loci >2 cM with a maximum LOD score of 2.5076. Simulation studies have shown that generally multiple homozygous loci with a maximum achievable LOD score are expected in a linkage analysis performed with SNP genotyping results in inbred families (Yang *et al.*, 2008). Also, the identical by descent (IBD) region harboring the disease mutation is generally larger than the IBD loci arising due to random chance (Yang *et al.*, 2008). Therefore, loci 3q21.1-22.1 and 8q24.13-24.22 were assessed as the strongest candidates for ataxia in the study family. Nevertheless, all four loci were investigated with exome sequencing. Locus 2p12-11.2 was considered as the weakest candidate because it was spanning the centromere where recombination frequency is expected to be low, but its size of 8.67 cM was larger than the size of locus 15q12; therefore, it was considered a candidate locus.

DNA sample of one of the affected brothers was subjected to exome sequencing, and candidate loci were investigated for a novel, homozygous mutation that could be associated with the disease phenotype. In all candidate loci, all genes that we assessed as strong candidates had been included in the capture kit and covered to a sufficient extent. Candidate loci 8q24.13-24.22 and 15q12 did not contain any novel exonic mutations. At candidate locus 3q21.1-22.1 a c.511C→T transition in *OSBPL11* was identified. Another nonsynonymous variant was in *KIAA1257,* a gene of unknown function. Online tool PolyPhen predicted the variant as damaging with a score of 1.00/1.00. In future studies, the variant could be validated by Sanger sequencing and, if found not to be a false positive call, screened for in a healthy population. At candidate locus 2p12-11.2 two nonsynonymous mutations, *LRRTM1* c.1531A→C (p.T511P) and *POLR1A* c.2801C→T (p.S934L) were found. *LRRTM1* (Leucine rich repeat transmembrane neuronal 1) is involved in neuronal cell adhesion network and has a role in synaptic functioning (Linhoff *et al.*, 2009; Soler-Llavina *et al.*, 2011). Furthermore, sequence variations in *LRRTM1* are associated with handedness and schizophrenia (Francks *et al.*, 2007). So, the gene seemed like a plausible gene for the disease due to its expression pattern and suggested functions.

However, the variant identified was predicted to be benign with a score of 0.337/1.00 by online tool PolyPhen. Also, the quality score of the variant was 29, which suggests that the variant could be a false positive call. *POLR1A* (polymerase RNA I polypeptide A) encodes the largest subunit of RNA polymerase I complex, which is also the catalytic subunit. Online tool PolyPhen predicted this variant also as benign with a score of 0.213/1.00. Nevertheless, both *LRRTM1* c.1531A→C and *POLR1A* c.2801C→T need to be validated via Sanger sequencing in future studies. Since *OSBPL11* stands out as a very strong candidate, it is unlikely that *KIAA1257*, *LRRTM1* or *POLR1A* have roles in disease manifestation.

c.511C→T in *OSBPL11* substitutes a positively charged, hydrophilic arginine residue with a non-polar, hydrophobic tryptophan. This change is predicted as damaging with high reliability via all three online tools employed, MMB, PolyPhen and SIFT. A population of 120 individuals was screened for the mutation to reach at least 80 per cent power to distinguish a sequence variant, and the mutation was not found in any (Collins and Schwartz, 2002). In the OSBPL11 protein, R171 lies between PH (pleckstrin homology) domain (residues 61-153) and ORD (oxysterol binding protein-related ligand binding) domain (residues 368-741). The mutation likely decreases protein function either by altering the three dimensional structure or by the loss of a hydrophilic functional residue. It has been shown that serine at position 172, the residue adjacent to R171, is phosphorylated (Cantin *et al.,* 2008). How the mutation affects the protein function and conformation and whether it affects the phosphorylation at serine 172 remain to be found.

Oxysterols are derivatives of cholesterol and by-products of cholesterol biosynthesis (Schroepfer, 2000). Protein families with sequence homology to oxysterol binding proteins (OSBPs) have been identified in all eukaryotic species. The members of such families are conserved from yeast to humans and designated OSBP-related (ORP) or OSBP-Like (OSBPL) proteins (Yan and Olkkonen, 2008). OSBPL11 (also known as ORP11) is a member of this protein family and has been suggested to act as either an intracellular lipid sensor or a transporter (Zhou *et al.,* 2010). Pathways of cholesterol trafficking and the exact mechanism of OSBP proteins in lipid metabolism at the molecular level are still unclear.

Niemann-Pick type C is an autosomal recessive, fatal disorder that mostly results from mutations in *NPC1* (95 per cent of the cases), while mutations in *NPC2* (5 per cent of the cases) also give rise to the disease (Vanier, 2010). Niemann-Pick type C was considered as a plausible diagnosis in the initial clinical examination of the ataxia patients; however, linkage analysis excluded the loci of both of the genes.

The most prominent cellular characteristics of Niemann-Pick Disease type C is cholesterol accumulation in cytoplasm of patient fibroblasts. Filipin staining is the routinely used method for the diagnosis of the disease. Filipin dye selectively binds unesterified cholesterol, and occlusions in patient skin fibroblasts can be observed under a fluorescence microscope. Since the phenotype of the study patients resembled Niemann-Pick type C phenotype and, thus, their cells could be defective in cholesterol trafficking due to deficient OSBPL11, cholesterol occlusions in lysosomes was expected. Skin fibroblasts from the patients were collected by Dr. Bülent Kara. Sections from fibroblasts stained with filipin were visualized under confocal microscope, and cholesterol inclusions were observed in the late endosomes/lysosomes (LE/L) (performed by Dr. Seyhun Solakoğlu at Istanbul University). This finding further validated that *OSBPL11* c.511C→T causes impairment in cholesterol metabolism, leading to a phenotype in our patients, that is similar to Niemann-Pick disease type C.

Cells obtain cholesterol through uptake from exogenous sources of cholesterol, mainly low density lipoproteins (LDLs), or through endogenous synthesis. LDLs are endocytosed and brought to late endosomes/lysosomes (LE/L) where cholesterol is unesterified by acid lipase (Brown and Goldstein, 1986). Unesterified cholesterol is distributed to plasma membrane and endoplasmic reticulum, which acts as a cholesterol sensor and regulates cholesterol synthesis and uptake (Peake and Vance, 2010). NPC1 and NPC2 are cholesterol binding proteins, and defects in those proteins are thought to impair the exit of unesterified cholesterol from LE/L. *OSBPL11* c.511C→T, giving rise to a defective OSBPL11 protein, may account for the accumulation of cholesterol in LE/L of affected patients in a similar fashion.

6.2 Azoospermia

In this study, the disease locus and the responsible mutation for azoospermia was identified through linkage analysis and exome sequencing in a consanguineous family with four affected brothers. The azoospermic brothers were genotyped with 610000 SNPs, and multipoint LOD scores were calculated assuming autosomal recessive inheritance with full penetrance. A maximum LOD score of 3.0057 was obtained at several loci. Initially 4p16.2-16.1 was considered as the strongest candidate, since it was the largest locus having the maximum LOD score. The homozygous genotype was shared by all affected brothers. However, homozygosity comparison via HCiE revealed an even larger locus at 18p11.21-21.1, which was shared by three affected brothers only, with a maximum LOD score of 3.0057. The genotype of the fourth brother narrowed down the locus to 229566 bp by two crossovers, one in the maternal and the other in the paternal allele.

A maximum LOD score of 3.0057 was achieved also in very small regions at 6q24.2, 6q27, 7q11.23, 7q21.11, 8q22.2, 9p24.2, 10p12.1, 13q13.3, 17q22 and 20q12, as presented in Figure 5.1. Fine LOD score calculations were also performed at the four loci, but the LOD score of 3.0057 was again obtained. Those regions were investigated via HCiE, and in all of them not all four affected siblings were found to share the same paternal and maternal alleles. Instead, a homozygous genotype was present in one or both parents due to the noninformativeness at the region. Therefore, the regions were assessed as homozygous by state and excluded as candidate loci.

Exome sequencing results in one brother were used to search for a novel mutation at 4p16.2-p16.1 and 18p11.21-q21.1, possibly underlying the disease. Locus 4p16.2-p16.1 harbored *EVC* c.1369G→A, a novel missense mutation. *EVC* (Ellis van Creveld syndrome) encodes a protein involved in Hedgehog signaling pathway, and mutations in *EVC* are associated with Ellis-van Creveld syndrome (EVC, MIM 225500) and Weyers acrodental dysostosis (MIM 193530), diseases in which impairments in skeletal and acrofacial development are observed. Online tool PolyPhen predicted *EVC* c.1369G→A as benign with a score of 0.123/1.00. Therefore, *EVC* was assessed as not associated with the disease.

Locus 18p11.21-q21.1 harbored *TAF4B* c.1831C→T (p.R611X), a stopgain mutation. The mutation was validated by Sanger sequencing, and the healthy brother was found heterozygous for the mutation. *TAF4B* c.1831C→T converts the arginine codon (CGA) at position 611 to a stop codon (UGA), truncating the protein by 252 amino acids. The truncated protein is 610 residues long and lacks the histone fold domain (residues 653-702) and TAF12 interaction domain (residues 830-862).

TFIID is the major factor in polymerase II machinery of transcription with the role of core promoter binding (Smale and Kadonaga, 2003). TFIID is composed of TATA-binding protein (TBP) and approximately 14 TBP-associated factors (TAFs). TBP recognizes and binds the TATA-box, whereas TAFs interact with elements residing upstream or downstream of TATA-box. Histone fold domains are present in 9 of the 14 TAFs, and it has been shown that the presence of histone fold domains increases the DNA binding activity of TAFs (Shao *et al.*, 2005). Dikstein *et al*. (1996) reported TAF4B as a cell type-specific subunit of TFIID, and Gazit *et al.* (2009) showed that TAF4B interacts with TAF12, another TAF protein containing a histone fold domain. The interaction of TAF4B with TAF12 is essential for DNA binding at core promoter (Gazit *et al.,* 2009). c.1831C→T transition in TAF4B removes both the histone fold domain and the TAF12 interaction domain, and presumably abolishes cell type specific DNA binding activity of TAF4B.

Spermatogenesis requires specific transcription factors with precise temporal and spatial expression (Sassone-Corsi, 1997). Studies on polymerase II transcription machinery suggest that its components have essential roles in gonad specific gene expression patterns (Hochheimer and Tjian, 2003). For instance, sequence alterations in *TAF7L*, encoding another testis-specific TFIID factor component, has been associated with infertility in humans (Akinloye *et al.,* 2007). Taf4b was shown to be enriched in mouse gonadal tissues and to be essential for ovarian follicle development (Freiman *et al.,* 2001). *Taf4b*-null mice had impaired spermatogenesis with reduced expression of spermatogonial stem cell markers and subsequently became infertile (Falender *et al.,* 2005). The exact role of *TAF4B* in human spermatogenesis remains elusive.

All coding exons of *TAF4B*, except the first 281 nucleotides of the first exon, was screened in a population of 45 azoospermia and 15 oligospermia patient samples, collected from the same region of Turkey. No homozygous mutations or variants were found, except for homozygous rs1677016 in exon 9 in two subjects, a synonymous SNP with a frequency of 0.466 in 1000 Genome phase I population. In addition, novel heterozygous missense mutations c.544G→T (p.A182S) in exon 3 of an azoospermia patient and c.1814A→G (p.N605S) in exon 9 of an oligospermia patient were observed. c.544G→T (p.A182S) was predicted to be benign by online program PolyPhen with a score of 0.001/1.00, whereas c.1814A→G (p.N605S) was assessed as probably damaging with a score of 0.495/1.00. The initial 281 nucleotides of exon 1 (denoted as TAF4B_EX1A in section 4.5.6), had a GC content of 75 per cent and was refractory to amplification with the HRM kit used. The amplification of the region was achieved with a PCR system specifically designed for GC-rich sequences; however, the alleles could not be resolved on SSCP gels. The conditions of SSCP analysis will be optimized in future studies. Since the population screen of this exon is inconclusive, a mutation residing in the first exon of *TAF4B* could have escaped detection. In addition, the subjects can harbor mutations in non-coding parts of the gene such as introns and UTRs or intergenic regulatory regions.

6.3 Progressive Myoclonic Epilepsy with Dystonia (PMED)

The disease locus was 16pter-p13.3 (Duru, 2006; Duru *et al.,* 2010). The aim in this study was to identify the defective gene residing at that locus. The region contained 276 genes (NCBI, GRCh37), and many of them seemed as good candidates, such as *CACNA1H*, associated with childhood absence epilepsy and idiopathic generalized epilepsy (MIM 611942), *ATP6V0C* (ATPase, H+ transporting, lysosomal, V0 subunit c), *TSC2* (tuberous sclerosis 2), *SSTR5* (somatostatin receptor 5) and *A2BP1*(Ataxin-2 binding protein 1). Since at the disease locus there were too many candidate genes to screen with a reasonable effort, the first approach was to analyze the region via exome sequencing. A DNA sample of an affected subject was sent to Yale University (USA) for the analysis. Exome capture was performed with 2.1M NimbleGen Human Exome Array v1.0 (Roche, Germany), and sequencing was carried out via single-read chemistry in Illumina Genome Analyzer IIx. The initial bioinformatics analysis, alignment of the reads and variant calling

were performed by Dr. Murat Günel's team at Yale University. The list of variants generated by the analysis was investigated for a novel mutation that could possibly be associated with disease phenotype. Interestingly, no novel exonic mutations residing at the disease locus were observed. Subsequently, the raw data was obtained from Yale University and bioinformatics analyses were repeated in our laboratory, as described in Section 4.2.2. Again, no novel exonic mutations were found.

A possible explanation for not finding the disease mutation was that the mutation was residing in a gene not covered by the capture kit used. The genes that were not included in the kit were extracted, and *TBC1D24* was found not to be included. Meanwhile *TBC1D24* was reported as responsible for familial infantile myoclonic epilepsy (FIME, MIM 605021), a much milder disease, but with similar clinical features (Corbett *et al.,* 2010; Falace *et al.,* 2010). Sequencing of all coding regions and UTRs of *TBC1D24* revealed a 2-bp deletion (c.969_970delGT, p.S324TfsX3) in exon 3. The mutation was deduced to affect only one of the two reported transcript isoforms (NM_001199107.1) and predicted to truncate the protein by 234 residues.

It was surprising to find a mutation in *TBC1D24* since the two previously reported families had a benign form of epilepsy (Corbett *et al.,* 2010; Falace *et al.,* 2010). Most of the patients in those families were adults, and seizures could be controlled by antiepileptic medication. On the other hand, the patients we studied never walked or talked, were severely deteriorated by the age of 2 years and died before the age of 8. Episodic phenomena such as dystonias, post-ictal enduring hemipareses, autonomic involvements and periods of obtundation and lethargy were also observed in addition to epilepsy and various types of myoclonic seizures. Also, seizures could not be controlled by medication. Thus, the manifestations of the disease in our research family were much more severe than those in the two previously reported families.

All three previously reported mutations in *TBC1D24* were missense mutations. In the Italian family, the patients were compound heterozygous for p.Ala509Val (NP_065756.1) and p.Asp147His (NP_065756.1), whereas in the Arab family the patients

were carrying a homozygous p.Phe251Leu (NP_065756.1) mutation (Corbett *et al.,* 2010; Falace *et al.,* 2010). Mutation c.969_970delGT (p.S324TfsX3) in our research family is predicted to result in a frameshift in the translational reading frame that would truncate the protein after residue 325. The mutant isoform 1 (NM_001199107.1) is predicted to encode a protein that is shorter by 234 amino acids (normal has 559), whereas isoform 2 (NM_020705.2) is not affected by the mutation. The mutation is also predicted to truncate the putative proteins encoded by the two novel isoforms identified in this thesis. The mutant protein isoform 3 is deduced to be shorter by 181 residues (normal has 506) and isoform 4 shorter by 99 residues (normal has 424).

In the previous two studies, the severity of all three missense mutations was assessed by investigating the neuronal arborization in primary cortical or hippocampal neurons prepared from mouse embryos (Corbett *et al.,* 2010; Falace *et al.,* 2010). The assays showed that the mutations decreased the length and the number of neurite termini whereas the overexpression of the wild type *TBC1D24* significantly increased the length of axons (Corbett *et al.,* 2010; Falace *et al.,* 2010). In addition, TBC1D24 protein was found to interact with ARF6 (ADP ribosylation factor 6), a small G protein that connects the actin cytoskeleton to membrane trafficking in developing nervous system (Jaworski, 2007). All mutations were assessed as loss of function mutations. Although those assays clarify the function of the gene, they do not represent the absolute measure of the protein activity. How the three reported mutations affect the precise spatial and temporal activity of TBC1D24 protein in developmental processes remains unknown. Therefore, it is possible that some TBC1D24 protein activity could have persisted in the patients in both families. The truncating mutation identified in this study led to a much more severe phenotype than the missense mutations that had been also assessed as loss of function; hence, the more severe mutation in our family is hypothesized to result in a more severe phenotype.

In humans, *TBC1D24* was reported to be expressed highest in brain, followed by testis, skeletal muscle, heart, kidney, lung and liver (Falace *et al*. 2010). Hirosawa *et al*. (1999) found moderate expression in amygdala and cerebellum and low expression in all other human brain regions. However, the sequences of the primers used in those assays are not available, and so which of the transcript isoforms the assays covered could not be

deduced. Furthermore, *TBC1D24* was found highly expressed in the cortex and hippocampus in developing mouse brain and the expression increased during cortical development (Falace *et al.,* 2010). In the present study, the expression of three of the *TBC1D24* transcript isoforms in various regions of adult human brain was investigated. Relative quantification showed that the expression of isoform 1 was higher than those of isoforms 2 and 3 in all the human brain regions we assayed. Therefore, isoform 1 stands out as the main brain isoform of *TBC1D24*.

It is necessary to mention that in the relative quantification assay isoform 1 was not amplified alone but together with isoform 4. However, this shortcoming in the assay is expected not to affect the results significantly, since the expression of isoform 4 was very low as compared to isoform 1, as judged on agarose gels (Figure 5.21).

cDNAs specific to various isoforms of *TBC1D24* were used for amplification of blood RNA samples. Interestingly, only the amplification of isoform 2 was observed and none of the other three isoforms was present in the blood RNA samples (data not shown). This suggests that isoform 2 may be the more ubiquitously expressed isoform and the other isoforms, all containing exon 3, are more specific to the central nervous system. The functions of different *TBC1D24* isoforms in developing brain remain to be studied.

In the relative quantification analysis, two novel transcript isoforms were detected, and their nucleotide structures from exon 2 (the first coding exon) to 3'UTR were determined. Isoform 3 did not contain exon 4, whereas isoform 4 lacked exon 5. The exclusion of exon 4 in isoform 3 did not change the translational frame, but the absence of exon 5 in isoform 4 is predicted to shift the frame so that 43 residues coded in a different frame than in the other isoforms, encoded in exons 7 and 8, were added after residue 381. Future studies are needed to investigate its function.

In summary, novel disease genes for ataxia and azoospermia were identified. In PMED a novel mutation in a reported gene was found and the expression of the gene in human brain was investigated; additionally, two novel transcript isoforms were identified.

In ataxia, *OSBPL11* c.511C→T stood out as the best candidate among the mutations detected because of the gene's role in cholesterol metabolism and the severity of the mutation. In azoospermia, a stopgain mutation *TAF4B* c.1831C→T that severely truncates the protein was found. The role of *TAF4B* in spermatogenesis was previously shown, where mice models deficient in TAF4B exhibited spermatogenic failure. In progressive myoclonic epilepsy with dystonia (PMED), a 2-bp deletion, c.969_970delGT, was found in *TBC1D24*, the gene responsible for FIME (MIM 605021). The mutation affects one of the reported transcript isoforms. In addition, two novel isoforms of *TBC1D24* were identified. Two of the isoforms that were predicted to be affected by the mutation were found to be more abundant than the unaffected isoform in all seven human brain regions assayed.

APPENDIX A: SEQUENCING DEPTH DISTRIBUTIONS

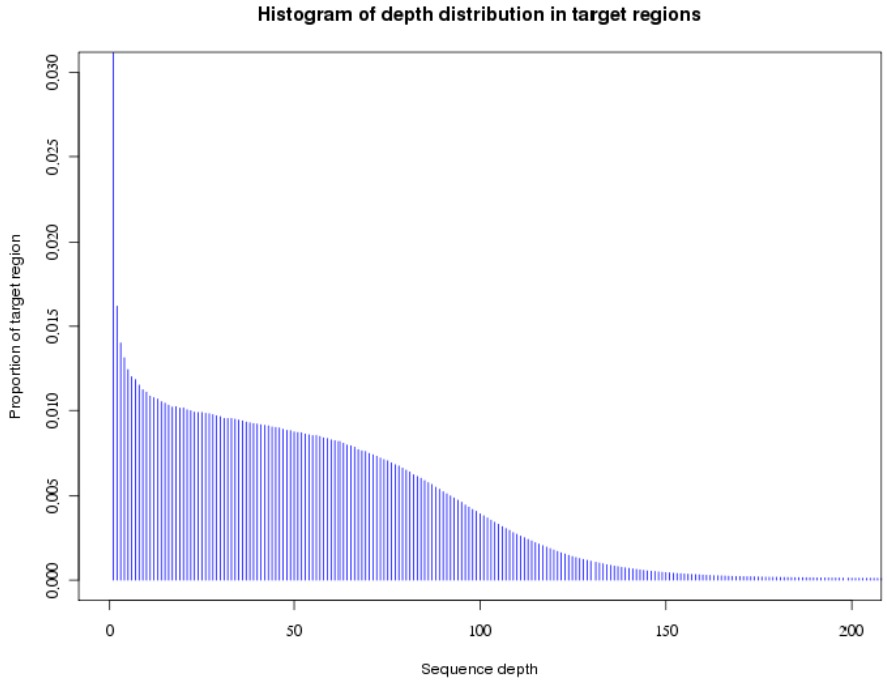


Figure 7.1. Sequencing depth distribution in the individual afflicted with ataxia.

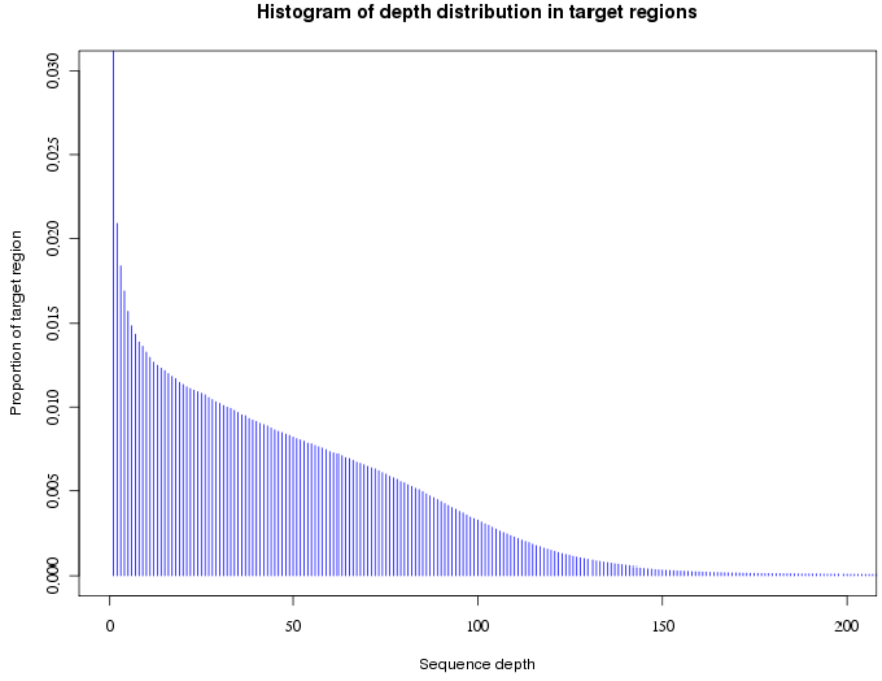


Figure 7.2. Sequencing depth distribution in the individual afflicted with azoospermia.

REFERENCES

Aicardi, J., C. Barbosa, E. Andermann, F. Andermann, R. Morcos, Q. Ghanem, Y. Fukuyama, Y. Awaya and P. Moe, 1988, "Ataxia-Ocular Motor Apraxia: a Syndrome Mimicking Ataxia Telangiectasia", *Annals of Neurology*, Vol. 24, No. 4, pp. 497-502.

Akinloye, O., J. Gromoll, C. Callies, E. Nieschlag and M. Simoni, 2007, "Mutation Analysis of the X-Chromosome Linked, Testis-Specific TAF7L Gene in Spermatogenic Failure", *Andrologia*, Vol. 39, No. 5, pp.190-195.

Al-Mahdawi, S., R. M. Pinto, D. Varshney, L. Lawrence, M. B. Lowrie, S. Hughes, Z. Webster, J. Blake, J.M. Cooper, R. King and M. A. Pook, 2006, "GAA Repeat Expansion Mutation Mouse Models of Friedreich Ataxia Exhibit Oxidative Stress Leading to Progressive Neuronal and Cardiac Pathology", *Genomics*, Vol. 88, No. 5 pp. 580-590.

Anttonen, A. K., I. Mahjneh, R. H. Hamalainen, C. Lagier-Tourenne, O. Kopra, L. Waris, M. Anttonen, T. Joensuu, H. Kalimo, A. Paetau, L. Tranebjaerg, D. Chaigne, M. Koenig, O. Eeg-Olofsson, B. Udd, M. Somer, H. Somer and A. E. Lehesjoki, 2005, "The Gene Disrupted in Marinesco-Sjogren Syndrome Encodes SIL1, an HSPA5 Cochaperone" *Nature Genetics*, Vol. 37, pp. 1309-1311.

Aston, K. I., C. Krausz, I. Laface, E. Ruiz-Castané and D. T. Carrell, 2010, "Evaluation of 172 Candidate Polymorphisms for Association with Oligozoospermia or Azoospermia in a Large Cohort of Men of European Descent", *Human Reproduction*, Vol. 25, No. 6 pp. 1383-1397.

Bashamboo, A., B. Ferraz-de-Souza, D. Lourenco, L. Lin, N. J. Sebire, D. Montjean, J. Bignon-Topalovic, J. Mandelbaum, J.P. Siffroi, S. Christin-Maitre, U. Radhakrishna, H. Rouba, C. Ravel, J. Seeler, J. C. Achermann and K. McElreavey, 2010, "Human Male Infertility Associated with Mutations in NR5A1 Encoding Steroidogenic Factor 1" *American Journal of Human Genetics*, Vol. 87, pp. 505-512.

Berkovic, S.F., S. Carpenter, A. Evans, G. Karpati, E. A. Shoubridge, F. Andermann, E. Meyer, J. L. Tyler, M. Diksic, D. Arnold, L. S. Wolfe, E. Andermann and A. M. Hakim, 1989, "Myoclonus epilepsy and ragged-red fibres (MERRF). 1. A Clinical, Pathological, Biochemical, Magnetic Resonance Spectrographic and Positron Emission Tomographic Study", *Brain*, Vol. 112, No. 5, pp.1231-1260.

Bouchard, J. P., A. Barbeau, R. Bouchard, R. W. Bouchard, 1978," Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay", *Canadian Journal of Neurological Sciences*, Vol. 5, pp. 61-69.

Braverman, N., L. Chen, P. Lin, C. Obie, G. Steel, P. Douglas, P. K. Chakraborty, J. T. R. Clarke, A. Boneh, A. Moser, H. Moser and D. Valle, 2002, "Mutation Analysis of PEX7 in 60 Probands with Rhizomelic Chondrodysplasia Punctata and Functional Correlations of Genotype with Phenotype" *Human Mutation* Vol. 20, pp: 284-297.

Brown, M. S. and J. L. Goldstein, 1986, "A Receptor-Mediated Pathway for Cholesterol Homeostasis", *Science*, Vol. 232, No. 4746, pp.34-47.

Carstea, E. D., J. A. Morris, K. G. Coleman, S. K. Loftus, D. Zhang, C. Cummings, J. Gu, M. A. Rosenfeld, W.J. Pavan, D. B. Krizman, J. Nagle, M. H. Polymeropoulos, S. L. Sturley, Y. A. Ioannou, M. E. Higgins, M. Comly, A. Cooney, A. Brown, C. R. Kaneski, E. J. Blanchette-Mackie, N. K. , E. B. Neufeld, T. Y. Chang, L. Liscum, J. F. Strauss 3rd, K. Ohno, M. Zeigler, R. Carmi, J. Sokol, D. Markie, R. R. O'Neill, O. P. van Diggelen, M. Elleder, M. C. Patterson, R. O. Brady, M. T. Vanier, P. G.

Pentchev and D. A. Tagle, 1997, "Niemann-Pick C1 Disease Gene: Homology to Mediators of Cholesterol Homeostasis", *Science*, Vol. 277, No. 5323 pp.228-231.

Cantin, G. T., W. Yi, B. Lu, S.K. Park, T. Xu, J. D. Lee and J. R. 3rd Yates, 2008, "Combining Protein-Based IMAC, Peptide-Based IMAC, and Mudpit for Efficient Phosphoproteomic Analysis", *Journal of Proteome Research*. Vol. 7, No. 3, pp. 1346-1351.

Chan, E. M., E. J. Young, L. Ianzano, I. Munteanu, X. Zhao, C. C. Christopoulos, G. Avanzini, M. Elia, C. A. Ackerley, N. J. Jovic, S. Bohlega, E. Andermann, G. A. Rouleau, A. V. Delgado-Escueta, B. A. Minassian and S. W. Scherer, 2003, "Mutations in NHLRC1 Cause Progressive Myoclonus Epilepsy", *Nature Genetic*s, Vol. 35, No. 2, pp. 125-127.

Chandley, A. C., 1979, "The Chromosomal Basis of Human Infertility", *British Medical Bulletin*, Vol. 35, No. 2 pp. 181-186.

Choi, M., U.I. Scholl, W. Ji, T. Liu, I. R. Tikhonova, P. Zumbo, A. Nayir, A. Bakkaloğlu, S. Ozen, S. Sanjad, C. Nelson-Williams, A. Farhi, S. Mane and R. P. Lifton, 2009, "Genetic Diagnosis by Whole Exome Capture and Massively Parallel DNA Sequencing" *Proceedings of the National Academy of Sciences*, Vol. 106, No. 45, pp. 19096-19101.

Colella, S., C. Yau, J. M. Taylor, G. Mirza, H. Butler, P. Clouston, A. S. Bassett, A. Seller, C. C. Holmes and J. Ragoussis, 2007, "Quantisnp: an Objective Bayes Hidden-Markov Model to Detect and Accurately Map Copy Number Variation Using SNP Genotyping Data", *Nucleic Acids Research*, Vol. 35, No. 6, pp. 2013-2025.

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

Corbett, M. A., M. Bahlo, L. Jolly, Z. Afawi, A. E. Gardner, K. L. Oliver, S. Tan, A. Coffey, J. C. Mulley, L. M. Dibbens, W. Simri, A. Shalata, S. Kivity, G. D. Jackson, S. F. Berkovic and J. Gecz, 2010, "A Focal Epilepsy and Intellectual Disability Syndrome Is Due to a Mutation in TBC1D24" *American Journal of Human Genetics,* Vol. 87, No. 3, pp.371-375.

Cowan, L. D., 2002, "The Epidemiology of The Epilepsies in Children" *Mental Retardation and Developmental Disabilities Research Reviews*, Vol. 8, No. 3, pp.171-181.

Çetinkaya, M., 2010, *Gene Hunt in Four Inherited Diseases*, PhD Thesis, Boğaziçi University.

Date, H., O. Onodera, H. Tanaka, K. Iwabuchi, K. Uekawa, S. Igarashi, R. Koike, T. Hiroi, T. Yuasa, Y. Awaya, T. Sakai, T. Takahashi, H. Nagatomo, Y. Sekijima, Kawachi I, Y. Takiyama, M. Nishizawa, N. Fukuhara, K. Saito, S. Sugano and S. Tsuji, 2001, "Early-Onset Ataxia with Ocular Motor Apraxia and Hypoalbuminemia Is Caused by Mutations in a New HIT Superfamily Gene", *Nature Genetics*, Vol. 29, pp. 184-188.

De Falco, F. A., L. Majello, R. Santangelo, M. Stabile, F. D. Bricarelli and F. Zara, 2001, "Familial Infantile Myoclonic Epilepsy: Clinical Features in a Large Kindred with Autosomal Recessive Inheritance" *Epilepsia,* Vol. 42, pp. 1541-1548.

Deng, Y., W. Zhang, D. Su, Y. Yang, Y. Ma, H. Zhang, S. Zhang, 2008 , "Some Single Nucleotide Polymorphisms of MSY2 Gene Might Contribute To Susceptibility To Spermatogenic Impairment in Idiopathic Infertile Men", *Urology,* Vol. 71, No. 5, pp. 878-882.

Dikstein, R., S. Zhou and R. Tjian, 1996, "Human TAFII 105 Is a Cell Type-Specific TFIID Subunit Related To HTAFII130", *Cell,* Vol. 87, No. 1, pp.137-146.

Di Donato, I., S. Bianchi and A. Federico, 2010, "Ataxia With Vitamin E Deficiency: Update of Molecular Diagnosis" *Neurological Sciences*, Vol. 31, No. 4, pp. 511-515.

Dohle, G. R., D. J. Halley, J. O. Van Hemel, A. M. van den Ouwel, M. H. Pieters, R. F. Weber and L. C. Govaerts, 2002, "Genetic Risk Factors in Infertile Men with Severe Oligozoospermia and Azoospermia" *Human Reproduction*, Vol. 17, No. 1, pp: 13-16.

Duru, N., 2006, *Gene Localization Studies in Two Neurological Disorders*, MS Thesis, Boğaziçi University.

Duru, N., S. A. Iseri, N. Selçuk and A. Tolun, 2010, "Early-Onset Progressive Myoclonic Epilepsy with Dystonia Mapping To 16pter-P13.3", *Journal of Neurogenetics*, Vol. 24, No. 4, pp. 207-215.

Elmslie, F. V., M. P. Williamson, M. Rees, M. Kerr, M. J. Kjeldsen, K. A. Pang, A. Sundqvist, M. L. Friis, A. Richens, D. Chadwick, W. P. Whitehouse and R. M. Gardiner, 1996, "Linkage Analysis of Juvenile Myoclonic Epilepsy and Microsatellite Loci Spanning 61 Cm of Human Chromosome 6p in 19 Nuclear Pedigrees Provides No Evidence For A Susceptibility Locus In This Region" *American Journal of Human Genetics*, Vol. 59, No. 3, pp. 653-663.

Falace, A., F. Filipello, V. La Padula, N. Vanni, F. Vanni, D. De Pietri Tonelli, F. A. de Falco, P. Striano, F. Dagna Bricarelli, C. Minetti, F. Benfenati, A. Fassio and F. Zara, 2010, "TBC1D24, an ARF6-Interacting Protein, Is Mutated In Familial Infantile Myoclonic Epilepsy" *American Journal of Human Genetics*, Vol. 87, No. 3 pp. 365-370.

Falender, A. E., R. N. Freiman, K. G. Geles, K. C. Lo, K. Hwang, D. J. Lamb, P. L. Morris, R. Tjian and J. S. Richards, 2005, "Maintenance of Spermatogenesis Requires TAF4b, a Gonad-Specific Subunit Of TFIID", *Genes and Development*, Vol. 19, No. 7, pp. 794-803.

Ferlin, A., F. Raicu, V. Gatta, D. Zuccarello, G. Palka and C. Foresta, 2007, "Review – Male Infertility: Role of Genetic Background", *Reproductive Biomedicine Online*, Vol. 14, No. 6, pp. 734-745.

Fogel, B. L. and S. Perlman, 2007, "Clinical Features and Molecular Genetics of Autosomal Recessive Cerebellar Ataxias", *Lancet Neurology*, Vol. 6, No. 3, pp. 245-257.

Freiman, R. N., S. R. Albright, S. Zheng, W. C. Sha, R. E. Hammer, R. Tjian, 2001, "Requirement of Tissue-Selective TBP-Associated Factor TAFII105 in Ovarian Development", *Science*, Vol. 293, No. 5537, pp. 2084-1087.

Francks, C., S. Maegawa, J. Laurén, B. S. Abrahams, A. Velayos-Baeza, S. E. , S. Colella, M. Groszer, E. Z. McAuley, T. M. Caffrey, T. Timmusk, P. Pruunsild, I. Koppel, P. A. Lind, N. Matsumoto-Itaba, J. Nicod, L. Xiong, R. Joober, W. Enard, B. Krinsky, E. Nanba, A. J. Richardson, B. P. Riley, N. G. Martin, S. M. Strittmatter, H. J. Möller, D. Rujescu, D. St Clair, P. Muglia, J. L. Roos, S. E. Fisher, R. Wade-Martins, G. A. Rouleau, J. F. Stein, M. Karayiorgou, D. H. Geschwind, J. Ragoussis, K. S. Kendler, M. S. Airaksinen, M. Oshimura, L. E. DeLisi and A. P. Monaco, 2007, "LRRTM1 on Chromosome 2p12 Is a Maternally Suppressed Gene That Is Associated Paternally with Handedness and Schizophrenia" *Molecular Psychiatry*, Vol. 12, No. 12, pp. 1129-1139.

Galjart, N. J., N. Gillemans, A. Harris, G. T. van der Horst, F. W. Verheijen, H. Galjaard and A. d'Azzo, 1988, "Expression of cDNA Encoding The Human "Protective Protein" Associated With Lysosomal Beta-Galactosidase and Neuraminidase: Homology To Yeast Proteases", *Cell*, Vol. 54, No. 6, pp. 755-764.

Gazit, K., S. Moshonov, R. Elfakess, M. Sharon, G. Mengus, I. Davidson and R. Dikstein, 2009,"TAF4/4b x TAF12 Displays a Unique Mode of DNA Binding and Is Required For Core Promoter Function of a Subset Of Genes", *Journal of Biological Chemistry*, Vol. 284, No. 39, pp. 26286-26296.

Guarducci, E., F., L. Becherini, M. Rotondi, G. Balercia, G. Forti and C. Krausz, 2006, "Estrogen Receptor Alpha Promoter Polymorphism: Stronger Estrogen Action Is Coupled with Lower Sperm Count", *Human Reproduction*, Vol. 21, No. 4 pp. 994-1001.

Hammoud, S., B. R. Emery, D. Dunn, R. B. Weiss and D. T. Carrell, 2009, "Sequence Alterations in the Ybx2 Gene Are Associated with Male Factor Infertility". *Fertility and Sterility*. Vol. 91, No. 4, pp. 1090-1095.

Harris, J. L., B. Jakob, G. Taucher-Scholz, G. L. Dianov, O. J. Becherel and M. F. Lavin, 2009, "Aprataxin, Poly-ADP Ribose Polymerase 1 (PARP-1) and Apurinic Endonuclease 1 (APE1) Function Together To Protect the Genome Against Oxidative Damage", *Human Molecular Genetics*, Vol. 18, No. 21, pp. 4102-4117.

Hirosawa, M., T. Nagase, K. Ishikawa, R. Kikuno, N. Nomura and O. Ohara, 1999, "Characterization of cDNA Clones Selected by the Genemark Analysis From Size-Fractionated cDNA Libraries From Human Brain", *DNA Research*, Vol. 6, No. 5, pp. 329-336.

Hochheimer, A. and R. Tjian, 2003, "Diversified Transcription Initiation Complexes Expand Promoter Selectivity and Tissue-Specific Gene Expression", *Genes and Development* Vol. 17, No. 11, pp. 1309-1320.

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.

Jaworski, J., 2007, "ARF6 in The Nervous System", *European Journal of Cell Biology*, Vol. 86, No. 9, pp. 513-524.

Kimmins, S., N. Kotaja, I. Davidson and P. Sassone-Corsi, 2004, "Testis-Specific Transcription Mechanisms Promoting Male Germ-Cell Differentiation", *Reproduction*, Vol. 128, No. 1, pp. 5-12.

Kuhlenbäumer, G., J. Hullmann and S. Appenzeller, 2011, "Novel Genomic Techniques Open New Avenues in the Analysis of Monogenic Disorders", *Human Mutation*, Vol. 32, No. 2, pp. 144-151

Lafrenière, R. G., D. L. Rochefort, N. Chrétien, J. M. Chrétien, J. I. Cochius, R. Kälviäinen, U. Nousiainen, G. Patry, K. Farrell, B. Söderfeldt, A. Federico, B. R. Hale, O. H. Cossio, T. Sørensen, M. A. Pouliot, T. Kmiec, P. Uldall, J. Janszky, M. R. Pranzatelli, F. Andermann, E. Andermann and G. A. Rouleau, 1997, "Unstable Insertion in the 5' Flanking Region of the Cystatin B Gene Is The Most Common Mutation in Progressive Myoclonus Epilepsy Type 1, EPM1", *Nature Genetics*. Vol. 15, No. 3, pp. 298-302.

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

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

Lazaros, L., N. Xita, A. Kaponis, K. Zikopoulos, N. Sofikitisa and I. Georgiou, 2008, "Evidence for Association of Sex Hormone-Binding Globulin and Androgen Receptor Genes with Semen Quality", *Andrologia*, Vol. 40, No. 3, pp. 186-191.

Linhoff, M. W., J. Laurén, R. M. Cassidy, F. A. Dobie, H. Takahashi, H. B. Nygaard, M. S. Airaksinen, S. M. Strittmatter and A. M. Craig, 2009, "An Unbiased Expression Screen for Synaptogenic Proteins Identifies the LRRTM Protein Family as Synaptic Organizers", *Neuron*, Vol. 61, No. 5, pp. 734-749.

Liu, A. W., A. V. Delgado-Escueta, J. M. Serratosa, M. E. Alonso, M. T. Medina, M. N. Gee, S. Cordova, H. Z. Zhao, J. M. Spellman, J. R. Peek, F. R. Donnadieu and R. S. Sparkes, 1995, "Juvenile Myoclonic Epilepsy Locus in Chromosome 6p21.2-P11: Linkage to Convulsions and Electroencephalography Trait" *American Journal of Human Genetics*, Vol. 57, No. 2, pp. 368-381.

Matzuk, M. M. and D. J. Lamb, 2008, "The Biology of Infertility: Research Advances and Clinical Challenges", *Nature Medicine*, Vol. 14, No. 11, pp. 1197-1213.

Mihalik, S. J., J. C. Morrell, D. Kim, K. A. Sacksteder, P. A. Watkins and S. J. Gould, 1997,"Identification of PAHX, a Refsum Disease Gene", *Nature Genetics*, Vol. 17, pp. 185-189.

Minassian, B. A., J. R. Lee, J. A. Herbrick, J. Huizenga, S. Soder, A. J. Mungall, I. Dunham, R. Gardner, C. Y. Fong, S. Fong, L. Jardim, P. Satishchandra, E. Andermann, O. C. Snead 3rd, I. Lopes-Cendes, L. C. Tsui, A. V. Delgado-Escueta, G. A. Rouleau and S. W. Scherer, 1998, "Mutations in a Gene Encoding a Novel Protein Tyrosine Phosphatase Cause Progressive Myoclonus Epilepsy", *Nature Genetics*, Vol. 20, No. 2, pp. 171-174.

Miyamoto, T., S. Hasuike, L. Yogev, M. R. Maduro, M. Ishikawa, H. Westphal and D. J. Lamb, 2003, "Azoospermia in Patients Heterozygous for a Mutation in SYCP3", *Lancet*, Vol. 362, pp. 1714-1719.

Moreira, M. C., C. Barbot, N. Tachi, N. Kozuka, E. Uchida, T. Gibson, P. Mendonca, M. Costa, J. Barros, T. Yanagisawa, M. Watanabe, Y. Ikeda, M. Aoki, T. Nagata, P. Coutinho, J. Sequeiros and M. Koenig, 2001, "The Gene Mutated in Ataxia-Oculomotor Apraxia 1 Encodes the New HIT/Zn-Finger Protein Aprataxin" *Nature Genetics*, Vol. 29, pp: 189-193.

Moreira, M. C., S. Klur, M. Watanabe, A. H. Nemeth, I. Le Ber, J. C. Moniz, C. Tranchant, P. Aubourg, M. Tazir, L. Schols, P. Pandolfo, J. B. Schulz, J. Pouget, P. Calvas, M. Shizuka-Ikeda, M. Shoji, M. Tanaka, L. Izatt, C. E. Shaw, A. M'Zahem, E. Dunne, P. Bomont, T. Benhassine, N. Bouslam, G. Stevanin, A. Brice, J. Guimarães, P. Mendonça, C. Barbot, P. Coutinho, J. Sequeiros, A. Dürr, J. M. Warter and M. Koenig, 2004, "Senataxin, the Ortholog of a Yeast RNA Helicase, Is Mutant in Ataxia-Ocular Apraxia 2" *Nature Genetics*,Vol. 36, pp. 225-227.

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

Naureckiene, S., D. E. Sleat, H. Lackland, A. Fensom, M. T. Vanier, R. Wattiaux, M. Jadot and P. Lobel, 2000, "Identification of HE1 as the Second Gene of Niemann-Pick C Disease", *Science*, Vol. 290, No. 5500, pp. 2298-2301.

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

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

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

Oates, R.D., 2008, "The Genetic Basis of Male Reproductive Failure", *Urologic Clinics of North America*, Vol. 35, No.2, pp. 257-270.

O'Flynn O'Brien, K. L. , A. C. Varghese and A. Agarwal, 2010, "The Genetic Causes of Male Factor Infertility: A Review", *Fertility and Sterility*, Vol.93, No.1, pp.1-12.

Ouahchi, K., M. Arita, H. Kayden, F. Hentati, M. Ben Hamida, R. Sokol, H. Arai, K. Inoue, J. L. Mandel and M. Koenig, 1995, "Ataxia with Isolated Vitamin E Deficiency Is Caused by Mutations in the Alpha-Tocopherol Transfer Protein", *Nature Genetics*, Vol. 9, No. 2 pp. 141-145.

Pshezhetsky, A.V., C. Richard, L. Michaud, S. Igdoura, S. Wang, M. A. Elsliger, J. Qu, D. Leclerc, R. Gravel, L. Dallaire and M. Potier, 1997, "Cloning, Expression and Mapping of Human Lysosomal Sialidase and Characterization of Mutations in Sialidosis", *Nature Genetics*, Vol. 15, No. 3, pp. 316-320.

Pandolfo M., 2008, "Friedreich Ataxia", *Archives of Neurology*, Vol. 65, No. 10, pp. 1296-1303.

Parfitt, D. A. , G. J. Michael, E. G. Vermeulen, N. V. Prodromou, T. R. Webb, J. M. Gallo, M. E. Cheetham, W. S. Nicoll, G. L. Blatch and J. P. Chapple, 2009, "The Ataxia Protein Sacsin Is a Functional Co-Chaperone That Protects Against Polyglutamine-Expanded Ataxin-1", *Human Molecular Genetics*, Vol. 18, No. 9, pp. 1556-1565.

Peake, K. B. and J. E.Vance, 2010, "Defective Cholesterol Trafficking in Niemann-Pick C-Deficient Cells", *FEBS Letters,* Vol. 584, No. 13, pp. 2731-2739.

Pettersson, E., J. Lundeberg and A. Ahmadian, 2009, "Generations of Sequencing Technologies", *Genomics*. Vol. 93, No. 2, pp. 105-111.

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

Quinzii, C.M. and M. Hirano, 2011, "Primary and Secondary Coq(10) Deficiencies in Humans", *Biofactors*. Vol. 37, No. 5, pp. 361-365.

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

Rosing, H. S., L. C. Hopkins, D. C. Wallace, C. M. Epstein and K. Weidenheim, 1985, "Maternally Inherited Mitochondrial Myopathy and Myoclonic Epilepsy", *Annals of Neurology*, Vol. 17, pp. 228-237.

Sassone-Corsi P., 1997, "Transcriptional Checkpoints Determining the Fate of Male Germ Cells" *Cell*, Vol. 88, No. 2, pp. 163-166.

Savitsky, K., A. Bar-Shira, S. Gilad, G. Rotman, Y. Ziv, L. Vanagaite, D. A. Tagle, S. Smith, T. Uziel, S. Sfez, M. Ashkenazi, I. Pecker, M. Frydman, R. Harnik, S. R. Patanjali, A. Simmons, G. A. Clines, A. Sartiel, R. A. Gatti, L. Chessa, O. Sanal, M. F. Lavin, N. G. Jaspers, A. M. Taylor, C. F. Arlett, T. Miki, S. M. Weissman, M. Lovett, F. S. Collins and Y. Shiloh, 1995, "A Single Ataxia Telangiectasia Gene with a Product Similar to PI-3 Kinase", *Science*, Vol. 268, pp. 1749-1753.

Schroepfer, GJ Jr., 2000, "Oxysterols: Modulators of Cholesterol Metabolism and Other Processes", *Physiological Reviews,* Vol. 80, No. 1, pp. 361-554.

Shahwan, A., M. Farrell and N. Delanty, 2005, "Progressive Myoclonic Epilepsies: a Review of Genetic and Therapeutic Aspects", *Lancet Neurology*, Vol. 4, No. 4, pp. 239-248.

Shiloh, Y.,2003, "ATM and Related Protein Kinases: Safeguarding Genome Integrity", *Nature Reviews Cancer,* Vol. 3, No. 3, pp. 155-168.

Shao, H., M. Revach, S. Moshonov, Y. Tzuman, K. Gazit, S. Albeck, T. Unger and R. Dikstein, 2005, "Core Promoter Binding by Histone-Like TAF Complexes", Molecular and Cellular Biology, Vol. 25, No. 1, pp. 206-219.

Smale, S. T. and J. T. Kadonaga, 2003, "The RNA Polymerase II Core Promoter". *Annual Review of Biochemistry,* Vol. 72, pp. 449-479.

Soler-Llavina G. J., M. V. Fuccillo, J. Ko, T. C. Südhof and R. C. Malenka, 2011, "The Neurexin Ligands, Neuroligins and Leucine-Rich Repeat Transmembrane Proteins, Perform Convergent and Divergent Synaptic Functions in Vivo", *Proceedings of the National Academy of Sciences*, Vol. 108, No. 40, pp. 16502-16509.

Sun, C., H. Skaletsky, B. Birren, K. Devon, Z. Tang, S. Silber, R. Oates and D. C. Page, 1999, "An Azoospermic Man with a De Novo Point Mutation in the Y-Chromosomal Gene USP9Y", *Nature Genetics*, Vol. 23, pp. 429-432.

Suraweera, A., O. J. Becherel, P. Chen, N. Rundle, R. Woods, J. Nakamura, M. Gatei, C. Criscuolo, A. Filla, L. Chessa, M. Fusser, B. Epe, N. Gueven and M. F. Lavin, 2007, "Senataxin, Defective in Ataxia Oculomotor Apraxia Type 2, Is Involved in the Defense Against Oxidative DNA Damage", *Journal of Cell Biology,* Vol. 177, No. 6, pp. 969-979.

Tapanainen, J. S. , K. Aittomäki, J. Min, T. Vaskivuo and I. T. Huhtaniemi, 1997, "Men Homozygous for an Inactivating Mutation of the Follicle-Stimulating Hormone (FSH) Receptor Gene Present Variable Suppression of Spermatogenesis and Fertility", *Nature Genetics*, Vol. 15, No. 2, pp. 205-206.

Taylor A.M. and P. J. Byrd, 2005, "Molecular Pathology of Ataxia Telangiectasia", *Journal of Clinical Pathology*, Vol. 58, No. 10, pp. 1009-1015.

Terwilliger, J. D. and J. Ott, 1994, *Handbook of Human Genetic Linkage*, Johns Hopkins University, Baltimore.

Van Bogaert, P., R. Azizieh, J. Désir, A. Aeby, L. De Meirleir, J. F. Laes, F. Christiaens and M. J. Abramowicz, 2007, "Mutation of a Potassium Channel-Related Gene in Progressive Myoclonic Epilepsy", *Annals of Neurology*, Vol. 61, No. 6, pp. 579-586.

Vanier, M. T., 2010, "Niemann-Pick Disease Type C", *Orphanet Journal of Rare Diseases*, Vol. 5, No.16.

Vogt, P.H., A. Edelmann, S. Kirsch, O. Henegariu, P. Hirschmann, F. Kiesewetter, F. M. Köhn, W. B. Schill, S. Farah, C. Ramos, M. Hartmann, W. Hartschuh, D. Meschede, H. M. Behre, A. Castel, E. Nieschlag, W. Weidner, H. J. Gröne, A. Jung, W. Engel, G. Haidl, 1996, "Human Y Chromosome Azoospermia Factors (AZF) Mapped to Different Subregions in Yq11", *Human Molecular Genetics*, Vol. 5, No. 7, pp. 933-943.

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S. F. Grant, H. Hakonarson and M. Bucan, 2007, "PennCNV: an Integrated Hidden Markov Model Designed for High-Resolution Copy Number Variation Detection in Whole-Genome SNP Genotyping Data", *Genome Research*, Vol. 17, No. 11, pp. 1665-1674.

Wold, B. and R. M. Myers, 2008, "Sequence Census Methods for Functional Genomics". *Nature Methods*.Vol. 5, No. 1, pp. 19-21.

Yan, D. and V. M. Olkkonen, 2008, "Characteristics of Oxysterol Binding Proteins", *International Review of Cytology*, Vol. 265, pp. 253-285.

Yang, F., K. Gell, G. W. van der Heijden, S. Eckardt, N. A. Leu, D. C. Page, R. Benavente, C. Her, C. Hoog, K. J. McLaughlin, P. J. Wang, 2008, "Meiotic Failure in Male Mice Lacking an X-Linked Factor", *Genes and Development*, Vol. 22, pp. 682-691.

Yang, J., S. Medvedev, J. Yu, L. C. Tang, J. E. Agno, M. M. Matzuk, R. M. Schultz and N. B. Hecht, 2005, "Absence of the DNA-/RNA-Binding Protein MSY2 Results in Male and Female Infertility", *Proceedings of the National Academy of Sciences*, Vol. 102, No. 16, pp. 5755-5760.

Yang, W., Z. Wang, L. Wang, P. C. Sham, P. Huang, Y. L. Lau, 2008, "Predicting the Number and Sizes of IBD Regions Among Family Members and Evaluating the Family Size Requirement for Linkage Studies", *European Journal of Human Genetics*, Vol. 16, No. 12, pp. 1535-1543.

Yen, P. H., 1998, "A Long-Range Restriction Map of Deletion Interval 6 Of The Human Y Chromosome: a Region Frequently Deleted in Azoospermic Males", *Genomics*, Vol. 54, pp. 5-12.

Zara, F., E. Gennaro, M. Stabile, I. Carbone, M. Malacarne, L. Majello, R. Santangelo, F. A. de Falco and F. D. Bricarelli, 2000, "Mapping of a Locus for a Familial Autosomal Recessive Idiopathic Myoclonic Epilepsy of Infancy to Chromosome 16p13", *American Journal of Human Genetics*, Vol 66, pp. 1552-1557.

Zhou, Y., S. Li, M. I. Mäyränpää, W. Zhong, N. Bäck, D. Yan, V. M. Olkkonen, 2010, "OSBP-Related Protein 11 (ORP11) Dimerizes with ORP9 and Localizes at the Golgi-Late Endosome Interface", *Experimental Cell Research*, Vol. 316, No. 19, pp. 3304-3316.