INVESTIGATION OF MOTOR NEURON DISEASES BY WES: GENETIC DISSECTION OF A TURKISH ALS COHORT

by

Fulya Akçimen

B.S., Molecular Biology and Genetics, Izmir Institute of Technology, 2013

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

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2017

INVESTIGATION OF MOTOR NEURON DISEASES BY WES:
GENETIC DISSECTION OF A TURKISH ALS COHORT

APPROVED BY:

Prof. Esra Battaloğlu (Thesis Supervisor)	……..
Prof. A. Nazlı Başak (Thesis Co-advisor)	……
Prof. S. Hande Çağlayan	……
Prof. Sibel Ertan	……

DATE OF APPROVAL: 27.07.2017

To my beloved grandparents Semine and Mehmet Küpeli,
for their love and encouragement.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis supervisor Prof. A. Nazlı Başak for her guidance and valuable criticism throughout this work. I am very grateful for her endless support.

I would like to extend my thanks to Prof. Esra Battaloğlu, Prof. Hande Çağlayan, and Prof. Sibel Ertan for devoting their time to evaluate this thesis.

I would further like to express my thanks to Prof. Jan H. Veldink for his mentorship during my stay at UMC Utrecht and for his encouragement to pursue the genetics of complex neurological disease. I am grateful for my stay at UMC Utrecht. I also cordially thank to Sara Pulit and Kristel Kool van Eijk for their valuable guidance in data analysis and for sharing their scientific knowledge.

I deeply thank all members of NDAL, Cemile, İlknur, Selda, Aslı, Irmak and Suna and Dr. Atay Vural (Koç University) for their valuable support. I also would like to especially thank Ceren for her friendship and for being a great research partner.

I thankfully acknowledge Suna-İnan Kıraç Foundation and Boğaziçi University Research Funds for financial support.

Last but not least, I deeply thank my mother Gülcan Akçimen, my brother Can Akçimen, my beloved sister Funda Akçimen Hatipoğlu for supporting me in all my decisions and my beloved Can for his endless support an unconditioned love during my graduate education. Nothing would have been possible without them.

ABSTRACT

INVESTIGATION OF MOTOR NEURON DISEASES BY WES: GENETIC DISSECTION OF A TURKISH ALS COHORT

Amyotrophic lateral sclerosis (ALS), the most common motor neuron disease, is characterized by muscle weakness and atrophy due to the degeneration of motor neurons in the motor cortex, brain stem and spinal cord. Both conventional gene discovery methods and association studies helped identify the genetic variants causing several ALS phenotypes. Recently, with the advent of whole exome sequencing (WES), it became possible to sequence the coding regions of the genome for a low cost and in a short time, changing the landscape of genetic disease research, including ALS. Thus, there are more than 40 genes with Mendelian inheritance identified in ALS. However, a significant portion of ALS cases is still genetically unexplained due to the complex genetic background of the disease.

In this study, WES was applied to investigate disease-causing variants in a cohort of 57 cases with ALS or other motor neuron diseases. *In silico* workflow was performed in our laboratory from the raw sequencing data to the final candidate variant lists. Homozygosity mapping was applied to recessively inherited pedigrees. Mutations in 19 distinct genes were identified as the genetic cause in 20 families. Identification of genes causing distal spinal muscular atrophy and neurodegeneration with brain iron accumulation in some cases, suggested controversies between the initial and the final diagnosis of the patients. These findings allowed us to draw two main facts: (i) the complex and heterogeneous nature of ALS and other motor-neuron diseases due to phenotypic overlaps, and (ii) the great success of WES as a current trend in rare disease genetics and differential diagnosis.

ÖZET

TÜM EKZOM DİZİLEME İLE MOTOR NÖRON HASTALIKLARININ ANALİZİ: TÜRK ALS KOHORTUNUN GENETİK İNCELENMESİ

En yaygın motor nöron hastalığı olan amiyotrofik lateral skleroz (ALS), motor korteks, beyin sapı ve omurilikteki motor nöronların dejenerasyonunun yol açtığı kas zayıflığı ve atrofi ile karakterize edilir. Geleneksel gen bulma yöntemleri ve ilişkilendirme çalışmaları ALS fenotipine yol açan birçok genetik varyasyonunun tanımlanmasında etkili olmuştur. Günümüzde, tüm ekzom dizilemedeki hızlı gelişmeler ile, genom üzerinde protein kodlayan bölgelerin düşük maliyetle ve kısa sürede dizilenmesi mümkün olmuş, bu yolla ALS de dahil olmak üzere hastalık genetiği araştırmaları yeni bir boyut kazanmış ve ALS'de bugün Mendel türü kalıtım gösteren 40'dan fazla mutasyonun tanımlanmasını sağlamıştır. Buna rağmen, hastalığın karmaşık genetik altyapısı nedeniyle olguların büyük bir kısmı genetik olarak hala açıklanamamıştır.

Bu tez çerçevesinde, ALS ve diğer motor nöron hastalarından oluşan 57 kişilik bir kohortta ekzom dizileme uygulanarak hastalık nedeni olabilecek varyasyonlar incelendi. Ham veriden başlayarak aday varyasyon listesi ile sonuçlanan biyoinformatik analizlerin bütünü laboratuvarımızda gerçekleştirildi. Resesif geçişli olgularda homozigotluk haritalaması da uygulandı. Bunların sonucunda, 19 birbirinden farklı gende tanımlanan mutasyonlar 20 ailedeki hastalığın genetik nedeni olarak tanımlandı. Olguların bazılarında gösterilen beyinde demir birikimi ya da distal spinal müsküler atrofiye neden olduğu bilinen genlerdeki değişimler, hastaların öncül ve ayırıcı tanılarında olası uyuşmazlıkların olabileceğine işaret etmektedir. Bu bulgular; (i) Fenotiplerindeki örtüşmeler dolayısıyla ALS ve diğer motor nöron hastalıklarının kompleks ve heterojen doğalarını ve (ii) tüm ekzom dizilemenin nadir hastalıkların genetiği ve ayırıcı tanısındaki etkin başarısını anlamamıza yardımcı olmuştur.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

kb	Kilobase
°C	Centigrade degree
µl	Microliter
*	Asterisk
#	Number
%	Percentage

LIST OF ACRONYMS/ABBREVIATIONS

ACADS	Acyl-CoA Dehydrogenase, C-2 to C-3 Short Chain
ACCPN	Agenesis of the Corpus Callosum with Peripheral Neuropathy
AD	Alzheimer's Disease
ALS	Amyotrophic Lateral Sclerosis
ALS2	Alsin2
ANG	Angiogenin
AO	Age of Onset
AR	Autosomal Recessive
AR	Autosomal Recessive Hereditary Spastic Paraplegia
ARJALS	Autosomal Recessive Juvenile ALS
BAM	Binary Alignment Map
BVVL	Brown-Vialetto-Van Laere syndrome
BWA	Burrows-Wheeler Aligner
C19ORF12	Chromosome 19 Open Reading Frame 12
C21ORF2	Chromosome 21 Open Reading Frame 2
C9ORF72	Chromosome 9 Open Reading Frame 72
ChIP-seq	Chromatin Immunoprecipitation
CMT2	Charcot-Marie-Tooth type 2
CNV	Copy Number Variation
dHMN	Distal Hereditary Motor Neuropathy
DJ1	Parkinson Protein 7
DNA	Deoxyribonucleic Acid
DNAJB2	DnaJ Heat Shock Protein Family (Hsp40) Member B2
ERBB4	Erb-B2 Receptor Tyrosine Kinase 4

ExaC	Exome Aggregation Consortium
F	Female
fALS	Familial ALS
FTD	Frontotemporal Dementia
FTDALS3	ALS with or without FTD
FUS	Fused in Sarcoma
GATK	Genome Analysis Toolkit
GVCF	Genomic Variant Call Format
GWAS	Genome Wide Association Studies
HGP	Human Genome Project
HMN	Hereditary Motor Neuropathy
HSJ1	Heat Shock Protein 1
HSP	Hereditary Spastic Paraplegia
IGHMBP2	Immunoglobulin Mu Binding Protein 2
IMBPFD	Inclusion Body Myopathy with Paget's Disease
INDEL	Insertion-Deletion
LMN	Lower Motor Neuron
LRSAM1	Leucine Rich Repeat And Sterile Alpha Motif Containing 1
M	Male
MAF	Minor Allele Frequency
MMND	Madras type Motor Neuron Disease
MND	Motor Neuron Disease
MOBP	Myelin-associated Oligodendrocyte Basic Protein
MPAN	Mitochondrial Membrane Protein Associated Neurodegeneration
MPZ	Myelin Protein Zero
NA	Not Available

NAS	Network-attached Storage System
NBIA	Neurodegeneration with Brain Iron Accumulation
ND	Neurodegenerative Disorders
NEK1	NIMA-related Kinase 1
NGS	Next Generation Sequencing
OPTN	Optineurin
P	Patient
PANK2	Pantothenate Kinase 2
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PD	Parkinson’s Disease
PDB	Paget Disease of Bone
PKAN	Pantothenate Kinase Associated Neurodegeneration
PLA2G6	Phospolipases A2 Group 6
PLEKHG5	Pleckstrin Homology and RhoGEF Domain Containing G5
PLS	Primary Lateral Sclerosis
PFN1	Profilin 1
RFVT3	Riboflavin Transporter protein 3
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
ROH	Runs of Homozygosity
rRNA	Ribosomal RNA
RVAS	Rare Variant Association Studies
sALS	Sporadic ALS
SAM	Sequence Alignment Map
SBMA	Spinal and Bulbar Muscular Atrophy
SCAD	Short Chain Acly-Coa Dehydrogenase

SCFD1	Sec1 Family Domain Containing
SCNA	Alpha-Synuclein
SIGMAR1	Sigma Non-opioid Intracellular Receptor
SLC12A6	Solute Carrier Family 12 Member 6
SLC52A3	Solute Carrier Family 52 Member 3
SMA	Spinal Muscular Atrophy
SMARD1	Spinal Muscular Atrophy with Respiratory Distress
SMN1	Survival of Motor Neuron 1
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
SOD1	Superoxide Dismutase 1
SPG11	Spastic Paraplegia 11
SQSTM1	Sequestosome 1
SV	Structural Variation
SYNE1	Spectrin Repeat Containing, Nuclear Envelope 1
TARDBP	Transactive Response DNA Binding Protein
TBK1	Tank-binding Kinase 1
TCC	Thin Corpus Collasum
TRMP7	Transient Receptor Potential Melastatin 7
TRPV4	Transient Receptor Potential Cation Channel Subfamily Member 4
Ts	Transition
Tv	Transversion
UBQLN1	Ubiquilin 1
UBQLN2	Ubiquilin 2
UMN	Upper Motor Neuron
USD	United States Dollar

VCF	Variant Call Format
VCP	Valosin Containing Protein
VEGF	Vascular Endothelial Cell Growth Factor
VUS	Variant of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
XLD	X Linked Dominant
ZFYVE26	Zinc Finger FYVE-type Containing 26

1. INTRODUCTION

Neurodegenerative disorders (NDs) are a heterogeneous group of neurological diseases characterized by neuronal loss in the central and peripheral nervous systems. The most common NDs are Alzheimer's (AD) and Parkinson's diseases (PD), followed by amyotrophic lateral sclerosis (ALS) (Przedborski *et al.*, 2003). While the affected regions are primarily the cerebral cortex in AD and extrapyramidal system in PD, in ALS neurodegeneration occurs predominantly in the spinal cord (Tsuji *et al.*, 2010). The main characteristics of AD are age-related dementia and cognitive decline, while PD is characterized by tremor, bradykinesia and rigidity. ALS is a rapidly progressive degeneration of motor neurons leading to paralysis and premature death (Bertram *et al.*, 2005). Although most ND cases are sporadic, there are some strictly Mendelian hereditary forms, the genetic mutations in which have shed light on the pathogenesis of these diseases

1.1. Introduction to Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis is a fatal neurodegenerative disorder that is characterized by the degeneration of upper and lower motor neurons. In the 1930s it became well known after the famous baseball player Lou Gehrig was diagnosed with the disease in the United States (Taylor *et al*., 2016).

ALS was first described by the neurologist Jean-Martin Charcot, known as the founder of modern neurology. In 1860s, he and his colleague Joffroy discovered that the lesions within the different regions of the spinal cord are associated with their distinct clinical presentations: (i) lesions within the lateral column of the spinal cord resulted in progressive paralysis and contractures of muscles without atrophy, (ii) lesions in the anterior horn of the spinal cord caused paralysis and muscle atrophy without any contractures. This discovery led Charcot to understand the motor component of the spinal cord. In 1874, the name of the disease as amyotrophic lateral sclerosis was offered by Charcot in the publication of the complete collection of his works (Kumar *et al.*, 2011).

ALS symptoms start focally as cramping or weakness in the limb or bulbar muscles and spread, ultimately causing paralysis (Taylor *et al.,* 2016). ALS is diagnosed with the combination of both upper and lower motor neuron (UMN and LMN) signs. UMN disturbance involves spasticity and brisk deep tendon reflexes, and LMN disturbance leads to fasciculations, wasting and weakness. The clinical presentations of the disease may be varying: (i) limb onset ALS; (ii) bulbar onset ALS with speech and swallowing difficulties followed by limb features as the disease progresses; (iii) primary lateral sclerosis defined by pure UMN involvement; and (iv) progressive muscular atrophy characterized by pure LMN involvement. Limb-onset form of the disease constitutes 70%, bulbar-onset 25% and initial respiratory or trunk involvement about 5% among patients.

The average age of onset in ALS is 55, however it may affect people at any age, even in the first or second decade, as well as in later life. Although some forms of ALS present a longer survival, half of the patients die within the first 30 months and 20% of patients survive less than 10 years after the symptom onset. While older age of onset and bulbar-onset are associated with reduced survival, younger age of onset and the limb-onset disease are marks of a protracted survival (Kiernan *et al*., 2011).

Although ALS was considered a motor neuron-specific disease for a long time, frontotemporal dementia (FTD) and cognitive impairment is present among several ALS patients. In fact, ALS and FTD are two diverse ends of the same disease, as well as a mixture of both. Hence, ALS and FTD might share a common pathogenic mechanisms (Therrien *et al.,* 2016).

ALS is classified as an orphan disease, with less than 200,000 affected cases worldwide; the prevalence is approximately five cases per 100,000. However, ALS is still responsible for about one in 500 adult deaths (Ghasemi and Brown, 2017). There is no effective treatment yet, except for riluzole which has a modest benefit (Therrien *et al.,* 2016).

1.2. Genetic Basis of ALS

About 90 % of ALS cases are sporadic (sALS), while the remaining 10 % are referred as familial (fALS) and have a classical genetic inheritance pattern. There is no clinical difference between fALS and sALS, aside from the lower mean age of onset of fALS cases. The genes mutated in fALS patients have also been found mutated in cases diagnosed with sALS, thus familial ALS made possible the identification of novel genes and mutations and shed light into the genetics of the disease (Andersen and Al-Chalabi, 2011, Therrien *et al.*, 2016).

1.2.1. Genes Implicated in ALS

Superoxide dismutase 1 (*SOD1*) is the first ALS gene discovered by linkage analysis (1993) using fALS cases. Eleven different *SOD1* mutations were shown to segregate in several fALS and sALS families (Rosen *et al.*, 1993). Today, more than 170 mutations have been seen in the *SOD1* gene which explain about 20 % of fALS and 1-3 % of sALS (Taylor *et al.*, 2016). These disease-causing mutations are found in either heterozygous or in homozygous state. Similar to other genes with allelic heterogeneity, each mutation has its own signature; e.g., while the Ala4Val substitution results in an aggressive form of ALS, the homozygous Asp90Ala substitution leads to milder symptoms with a slower progression (Therrien *et al.*, 2016).

Transactive response DNA binding protein (TARDBP) and *fused in sarcoma* (*FUS)* are the two subsequently identified ALS genes (Sreedharan *et al.*, 2008; Kwiatkowski *et al.*, 2009). *TARDBP* and *FUS* mutations are thought to cause a toxic gain of function, since their products form cytoplasmic aggregates which are common in motor-neuron diseases (MND) (Therrien *et al.*, 2016).

To date, the most common known cause of ALS and FTD is a repeat expansion mutation in the first intron of the *chromosome 9 open reading frame 72* (*C9ORF72*). The locus was discovered by two independent groups via the combination of association and linkage studies. The size of the hexanucleotide repeat (G_4C_2) is 2-23 in healthy persons, while it may be up to hundreds or thousands in affected individuals (Dejesus-Hernandez *et al.,* 2011; Renton *et al.,* 2011). The *C9ORF72* repeat expansion mutation explains 10 % of sALS and 30 % of fALS cases (Al-Chalabi *et al.,* 2016) with a recognizable amount of bulbar tendency (Ghasemi and Brown, 2017). Since it is hard to examine the precise number of repeats and because the clinical findings are contradictory, the anticipation pattern of the *C9ORF72* mutation could not be determined yet (Therrien *et al.,* 2016).

With the advent of whole exome and genome sequencing techniques, the number of ALS genes and mutations, including single nucleotide variations (SNVs), insertions and deletions (INDELs); has drastically increased in the last few years. Today, there are 41 genes shown to cause the ALS phenotype (Table 1.1).

Although most of the mutations in fALS genes appear with autosomal dominant form of inheritance, some of them are inherited autosomal recessively such as *alsin2 (ALS2), spastic paraplegia 11 (SPG11)* and *optineurin (OPTN)* (Ghasemi and Brown, 2017). Moreover, several de novo mutations and oligogenic inheritance (mutations in more than one ALS gene or the presence of modifier genes) are reported (Therrien *et al.,* 2016). To date, it has proved challenging to determine how mutations in all these divergent genes converge into the same clinical phenotype of ALS.

Table 1.1. Gene mutations that cause ALS, adapted from Ghasemi and Brown, 2017.

Gene	Locus	Fraction fALS (%)	Inheritance	Associated phenotype	Reference
C9ORF72	9p21.3	40-50	AD	ALS, ALS+FTD, FTD	Renton *et al.*, 2011, Dejesus-Hernandez *et al*., 2011
SOD1	21q22	20-25	AD, AR	ALS	Rosen *et al*., 1993
TARDBP	1p36.2	4-5	AD	ALS, ALS+FTD, FTD	Sreedharan *et al.*, 2008
FUS	16p11.2	4-5	AD	ALS, ALS+FTD, FTD	Kwiatkowski *et al*., 2009
OPTN	10p13	2-3	AD, AR	ALS, ALS+FTD	Maruyama *et al*., 2010
PFN1	17p13	1-2	AD	ALS	Wu *et al*., 2012
VCP	9p13	1-2	AD	ALS, ALS+FTD, FTD	Johnson *et al*., 2010
ANG	14q11.2	1	AD	ALS, ALS+FTD, FTD	Greenway *et al*., 2006
TUBA4A	2q35	<1	AD	ALS, ALS+FTD	Smith *et al*., 2014
UBQLN2	Xp11	<1	XLD	ALS, ALS+FTD, FTD	Deng *et al.*, 2011
TAF15	17q11	<1	AD	ALS	Couthouis *et al.*, 2011
EWSR1	22q12.2	<1	AD	ALS	Couthouis *et al.*, 2012
hnRNPA1	12q13	<1	AD	ALS, ALS+FTD, FTD	Kim *et al.*, 2013
hnRNPA2B1	7p15	<1	AD	ALS, ALS+FTD, FTD	Kim *et al.*, 2013
SETX	9q34.13	<1	AD	ALS	Chen *et al*., 2004
CREST	20q13.3	<1	-	ALS	Chesi *et al*., 2013
MATR3	5q31.2	<1	AD	ALS, ALS+FTD	Johnson *et al*., 2014
ATXN2	12q24	<1	AD	ALS, ALS+FTD,	Elden *et al*., 2010
ELP3	8p21.1	<1	-	ALS	Simpson *et al*., 2009
FIG4	6q21	<1	AD	ALS, PLS	Zhang *et al*., 2008

Table 1.1. Gene mutations that cause ALS, adapted from Ghasemi and Brown, 2017 (cont.).

Gene	Locus	Fraction fALS (%)	Inheritance	Associated phenotype	Reference
SQSTM1	5q35	<1	AD	ALS, ALS+FTD, FTD	Gal *et al*., 2009, Fecto *et al*., 2010
CHMP2B	3p11	<1	AD	ALS, FTD	Cox *et al*., 2010,
ALS2	2q33.1	<1	AR	ALS, PLS	Ben Hamida *et al*., 1990, Yang *et al*., 2001
VAPB	20q13	<1	AD	ALS, PLS	Nishimura *et al*., 2004
SIGMAR1	9p13.3	<1	AR	ALS, ALS+FTD, FTD	Al-Saif *et al*., 2011
DCTN1	2p13	<1	AD, AR	ALS	Munch *et al*., 2004
SPG11	15q21.1	<1	AR	ALS, HSP	Orlacchio *et al*., 2010
NEFH	22q12.2	<1	AD, AR	ALS	Figlewicz *et al*., 1994
PRPH	12q13	<1	AD, AR	ALS	Gros-Louis *et al*., 2004
PNPLA6	19p13	<1	AR	ALS, HSP	Rainier *et al*., 2008
PON1-3	7q21	<1	-	ALS	Slowik *et al*., 2006
DAO	12q22	<1	AD	ALS	Mitchell *et al*., 2010
CHRNA3, CHRNA4, CHRNB4	15q24, 20q13, 15q24	<1	-	ALS	Sabatelli *et al*., 2009, 2012
ERBB4	2q34	<1	AD	ALS	Takahashi *et al*.,, 2013
CHCHD10	22q11	<1	AD	ALS+FTD	Bannwarth *et al*., 2014
C19ORF12	9q12	<1	AR	ALS, MPAN	Deschauer *et al*., 2012
ALS3	18q21	<1	-	ALS	Hand *et al*., 2002
ALS7	20p13	<1	-	ALS	Hand *et al*., 2002
ALS6-21	6p25, 21q22	<1	-	ALS	Butterfield *et al*., 2009
ALS-FTD	16p12	<1	-	ALS+FTD	Dobson-Stone *et al*., 2013
TBK1	12q14.2	<1	AD	ALS+FTD	Cirulli *et al*., 2015
CCNF	16p13.3	<1	AD	ALS+FTD	Williams *et al*., 2015

1.2.2. Overview of ALS in the Turkish Cohort

The investigation of disease-causing mutations in our Turkish ALS cohort, performed via both conventional (PCR-based) and next generation techniques, reveals the presence of mutations in *C9ORF72*, *SOD1*, *TARDBP*, *FUS* and *UBQLN2*, explaining approximately 41 % of fALS (Figure 1.1) and 4 % of sALS cases (Figure 1.2). Moreover, mutations in *OPTN, SPG11, DJ1, PLEKHG5, SYNE1, TRPM7, and SQSTM1* have been identified via whole exome sequencing in fALS cases, which unravel another 11 % of the Turkish fALS cases (Ozoguz *et al*., 2015).

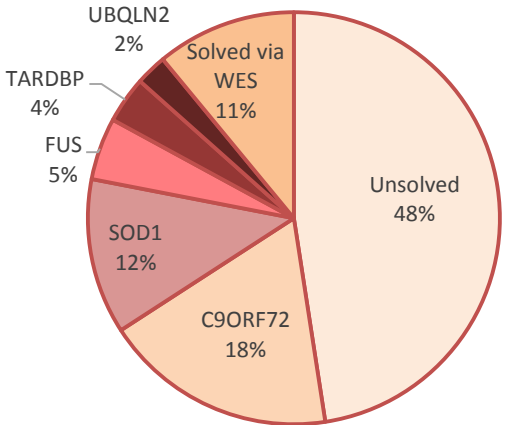


Figure 1.1. The proportion of ALS genes in Turkish fALS cases (Ozoguz *et al*., 2015).

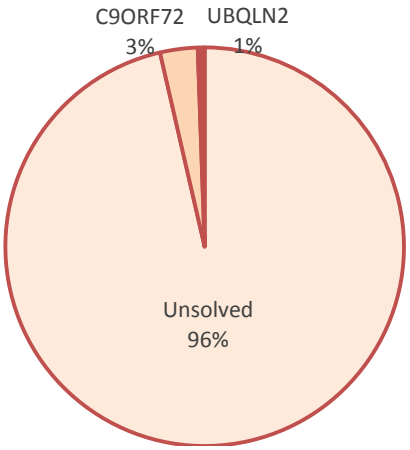


Figure 1.2. The proportion of ALS genes in Turkish sALS cases (Ozoguz *et al*., 2015).

1.3. Overlapping Phenotypes of ALS and Other Motor Neuron Diseases

Although the term motor neuron disease (MND) is often used to describe ALS, it involves a group of disorders characterized by selective loss of specialized neurons. The differences in clinical presentation provide distinct nomenclatures and diagnostic classification among ALS and other non-ALS motor neuron diseases: spinal muscular atrophy (SMA), spinal and bulbar muscular atrophy (SBMA), hereditary motor neuropathy (HMN), hereditary spastic paraplegia (HSP), Charcot–Marie–Tooth type 2 (CMT2) or neurodegeneration with brain iron accumulation (NBIA) (James & Talbot, 2006). Even though each MND has its own causative genes and specific diagnostic features, there are both genetic and phenotypic overlaps among MNDs leading to misdiagnosis.

The pleiotropy of motor neuron diseases is a proof of their common genetic mechanisms. Homozygous mutations in the *SPG11* gene are shown to cause SPG11-based ALS and/or HSP. Overlapping phenotypes of SPG11-based ALS and HSP confirm their difficult clinical differential diagnosis. Indeed, this phenotypic overlap may help to unravel the common mechanistic levels of these diseases (Iskender *et al.,* 2015). Similarly, Neurodegeneration with Brain Iron Accumulation Type 4 (NBIA4) caused by *C19ORF12* mutations, mimics juvenile onset ALS, since iron accumulation may not be apparent during the first decade of disease (Kim *et al.,* 2016).

1.4. Methodologies to Identify Causative Genes/Mutations in ALS

1.4.1. Linkage Analysis

Linkage analysis is a family-based genetic method that involves (i) identifying a genetic marker of known chromosomal location which is linked to an unknown gene and (ii) testing every neighboring gene to identify the phenotype causing ones. Linkage analysis is based on the transmission of specific alleles from affected parents to affected offsprings

more often than expected by chance. Linkage studies are useful for identifying variants predominantly in Mendelian diseases (Ott *et al.,* 2011; Al-Chalabi *et al.,* 2016).

To date, the biochemical mechanisms underlying many neurological diseases remain elusive. The identification of the chromosomal location of a disease-causing gene is a useful initial step for understanding the molecular pathology of the disease (Pulst *et al*., 1999). In 1983, the location of Huntington disease gene was mapped to chromosome 4 via linkage analysis using recombinant DNA technology, making it the first disease gene identified with linkage (Gusella *et al*., 1983). The first locus associated with ALS was identified in 1991 by the same approach and two years later *SOD1* (*ALS1*) was discovered using linkage followed by a conventional genotyping method, single-strand conformational polymorphism analysis. Several different variations were found segregating in both fALS and sALS cases, explaining a significant proportion of the disease genetics (Siddique *et al*., 1991; Rosen *et al.*, 1993).

1.4.2. Homozygosity Mapping

In consanguineous families, the coefficient of inbreeding increases, which in turn amplifies the possibility of the presence of disease-causing mutations within homozygous blocks (Alkuraya *et al*., 2010). Homozygosity mapping is based on the inheritance of the same mutation from a common ancestor to consanguineous parents on the same chromosomal stretch, and transmission of the mutation to offspring in homozygous state (Kancheva *et al.*, 2015). It is a positional cloning method which allows the detection of runs of homozygosity (ROH) as a measure of homozygous stretches.

Identification of the locus harboring the disease-causing mutations via homozygosity mapping is a strong gene discovery method for rare disease genetics, especially in the case of isolated populations. Identification of *OPTN* was a result of such a study in which three ALS cases from consanguineous marriages were subjected to homozygosity mapping; their

overlapping ROH made the detection of the candidate region possible, followed by the discovery of the gene (Maruyama *et al*., 2010).

1.4.3. Genome-Wide Association Studies

The completion of the Human Genome Project (HGP) was a major breakthrough in human genetics that provided the first map of the 3 billion bases in the human genome. With the map, it became possible to identify genetic variants in an individual, which did not match the reference sequence (Wheeler *et al*., 2008). Common variants with more than 1 % minor allele frequency (MAF) were defined as single nucleotide polymorphisms (SNPs); such variations were reported in the International HapMap Project, an extension of the HGP (International HapMap Consortium, 2003). With the completion of Phase III, the database contains more than three million SNPs, and the information of the genetic location of variants contributed to the development of SNP arrays, paving the way to the era of genome-wide association studies (GWAS) (International HapMap 3 Consortium, 2010).

Genome-wide association studies (GWAS) search for whether a SNP is observed in individuals with a disease significantly more or less often than expected by chance, which would mean that this variant is associated with the disease (Mullen *et al*., 2009). While linkage analysis examines the relationship of loci, association studies focus on the relationship of alleles (Pulst *et al*., 1999).

In 2011, a significant genetic association was identified in chromosome 9p21, in which the *C9ORF72* repeat (G_4C_2) expansion mutation was subsequently found (Dejesus-Hernandez *et al*., 2011; Renton *et al*., 2011). In addition to *C9ORF72*, there are several other associated loci which were identified and replicated in ALS GWAS (Table 1.2) (Al-Chalabi *et al*., 2016).

Table 1.2. ALS associated loci identified in GWA & replication studies, adapted from Al-Chalabi, 2016.

Locus	Single nucleotide polymorphism	Gene	Reference
9p21.3	-	*C9ORF72*	Renton *et al*., 2011, Dejesus-Hernandez *et al*., 2011
17q11.2	rs35714695	*SARM1*	Fogh *et al*., 2014
19p13	rs12608932	*UNC13A*	van Es *et al*., 2009
21q22.3	rs75087725	*C21ORF2*	van Rheenan *et al*., 2016
12q14.2	rs74654358	*TBK1*	Cirulli *et al*., 2015
3p22.1	rs616147	*MOBP, RPSA, SNORA6, SNORA62*	Hoglinger *et al*., 2011
14q12	rs10139154	*SCFD1, G2E3*	van Rheenan *et al*., 2016

1.4.4. Structural Variations

Structural variation in the human genome comprising deletions, duplications, insertions, inversions, translocations and copy-number variations (CNV) are less studied genetic contributors of late-onset human diseases. Nevertheless, there are a few studies investigating CNVs in ALS. Abnormal copy-number of *survival of motor neuron 1 (SMN1)* gene which is known to cause spinal muscular atrophy was shown to be associated with sALS (Corcia *et al.,* 2002), as well as the number and median-size of duplications in the *SMN1* were found higher in sALS compared to controls (Wain *et al.,* 2009). Another CNV analysis showed that the deletions of the *SMN1* associate with shortened survival in ALS (Veldink *et al.,* 2005). Since subsequent studies have failed to replicate these findings, there is no evidence supporting the contribution of CNVs to ALS pathogenesis (Leblond *et al.,* 2014; Ghasemi and Brown, 2017).

1.4.5. Next Generation Sequencing

Next generation sequencing (NGS) is a parallel DNA sequencing method that produces millions of short reads from 25 to 500 base pairs (Boycott *et al*., 2013). Unlike the capillary-based first generation sequencing (Sanger sequencing) which may take several years and would cost millions of dollars to sequence an entire genome, an NGS platform can produce the same genome sequence within a few weeks for about $1000 USD (Foo *et al*., 2012). It is possible to sequence whole genome (WGS), whole exome (WES) as well as transcriptome (RNA-seq) and DNA-protein interaction by chromatin immunoprecipitation-sequencing (ChIP-seq) via NGS technology, depending on the type of variation to be detected.

WGS and WES are unbiased approaches for rapid detection of SNVs, as well as short INDELs within the genome (Jiang *et al*., 2014). Based on the knowledge from previous studies, explaining the role of mutations in diseases, locus heterogeneity, availability of only a small number of samples/families and the required labour were critical limitations of conventional methods that have been overcome by NGS which changed the landscape of disease genetics (Boycott *et al*., 2013).

Both WGS and WES have their own challenges by producing vast amount of variations making it difficult to catch the disease-causing one(s) among them. However, with the decreasing cost and increased use of NGS, it became possible to combine linkage analysis and WGS, providing a statistical evidence for the involvement of a variant/gene in disease etiology. Similarly, homozygosity mapping is an approach which can also be performed in combination with WES to narrow down the list of the candidate variants in consanguineous cases. Today, with the advancements in NGS technologies, linkage analysis and homozygosity mapping can be directly applied to WES and WGS data in a single step, without the need of prior SNP genotyping (Ott *et al*., 2015; Kancheva *et al.*, 2015).

Protein coding regions (exomes) constitute approximately 1% of the human genome and are shown to harbor 85 % of disease-causing variations. Besides, due to its low cost and less complexity compared to WGS, today WES is a more preferred platform in the discovery of novel disease genes and mutations (Boycott *et al*., 2013).

1.4.5.1. General Workflow of Exome Sequencing. WES is a multistep process consisting of wet-lab and *in silico*-lab workflows. In each of these workflows, there are pipelines common for all types of studies, as well as parameters which users are able to interfere and optimize based on the purpose of the study. The wet-lab is the step where the actual sequencing occurs, consisting of (i). DNA isolation and fragmentation, (ii). Addition of adaptors to the fragments, (iii). Exome enrichment via capturing and washing out uncaptured DNA, (iv). Cluster generation and (v). sequencing and base calling (Figure 1.4.3) (Jiang *et al*., 2014).

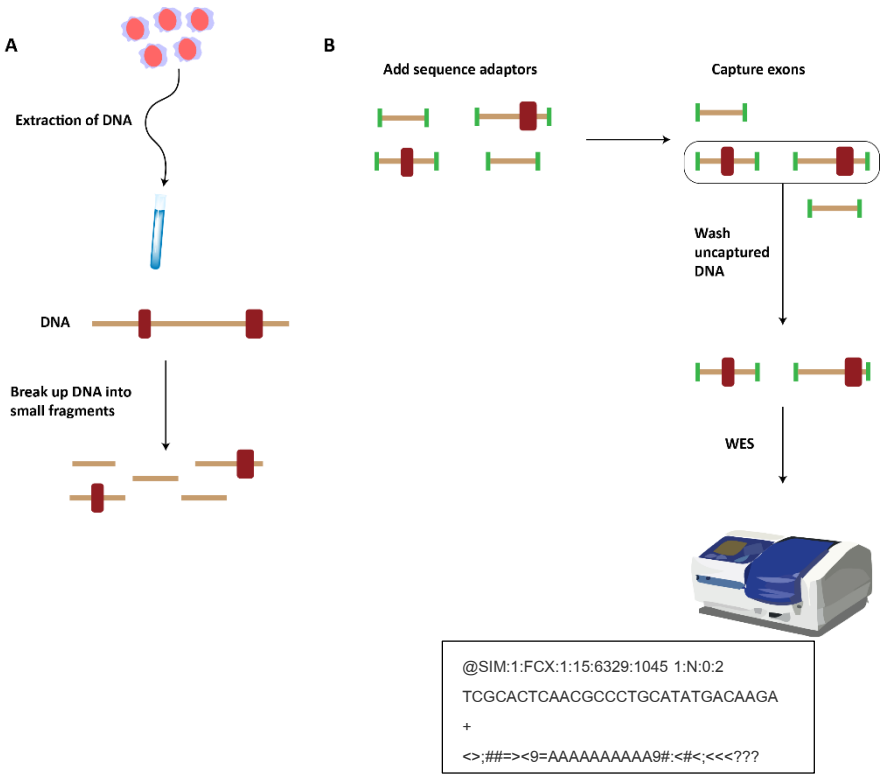


Figure 1.3. Wet-lab workflow of WES.

The *in silico* step consists of the computational pipeline to generate a meaningful information from raw sequencing data. This includes the alignment of raw reads to the reference genome, variant calling, functional annotation and priorization of variations (Foo *et al*., 2012). The choice of the algorithm to be used in the pipeline is a crucial step. Indexing the genome via an exact algorithm is an exhaustive process for large sequences of genomes, thus generally, heuristic algorithms such as Burrows Wheeler Transform are preferred, even though they do not guarantee to find all local hits (Li and Durbin *et al*., 2009). There are several different tools based on the different algorithms for identification of SNVs and INDELs. The Genome Analysis Toolkit (GATK) is one of the most popular variant calling software among both researchers and clinicians, which was created for Illumina reads by the Broad Institute (McKenna *et al.*, 2010).

With the development of public databases which catalogue alleles and variants systemically, the interpretation of thousands of variations and determination of their association to diseases became a computational step within the workflow rather than being an exhaustive manual approach. Previous publicly available databases, the Exome Variant Server and 1000 Genomes Project contain smaller amount of samples; 6503 exomes and 2504 individuals, respectively. After HapMap Project, the second revolutionary breakthrough is the creation of a dataset which consists of approximately seven million high-quality protein-coding variations from 60,706 individuals by the Exome Aggregation Consortium (ExAC). The application of this data set to the bioinformatic analysis provides the discovery of widespread mutational recurrence and a respectable increase in the resolution of very low-frequency variations (Lek *et al.*, 2016).

Like other rare disease cases, Mendelian inheritance with a family segregation, where affected and healthy samples are available, is the best model for WES analysis. The inheritance pattern helps to narrow down the number of susceptible variations in a family, getting us one step closer to the identification of disease causative gene(s).

1.4.5.2. Application of Whole Genome and Exome Sequencing to ALS. NGS is a highly effective approach in the discovery of novel ALS genes. Several different mutations in *valosin-containing protein (VCP)* and *profilin1 (PFN1)* in five and seven familial cases, respectively, were identified by family-based WES analyses, leading to the discovery of these genes in ALS phenotype (Johnson *et al*., 2010; Wu *et al*., 2012). Furthermore, WES can be applied to the identification of novel mutations in known disease-causing genes like *OPTN*, *SPG11* and *SQSTM1* which are too large and complex to be investigated by conventional PCR-based methods.

Besides family-based WES and WGS studies, large-scale genome-wide sequencing analyses have been performed to unravel various ALS genes and risk variations. While GWAS is a good approach to identify common variants, rare variant association tests (RVAS) are more suitable strategies to unravel the association of rare variants with ALS. Since it is hard to catch the rare variants among a limited number of samples, in RVAS, variants are grouped based on gene, location or functional characterization to compensate for the low statistical power. Burden test is a gene-based analysis, which basically asks, whether individuals carrying a rare variant in a gene are phenotypically similar to individuals which do not (Auer *et al.,* 2015).

A burden analysis of 2,874 ALS patients and 6,405 control samples led to the identification of *TANK-binding kinase 1 (TBK1)* with significant enrichment of rare loss-of-function mutations (Cirulli *et al.,* 2015). *TBK1* is responsible for the phosphorylation of the ALS gene *OPTN* in the autophagy pathway. It has been shown that mutant *TBK1* alleles cause the loss of interaction with its adaptor protein *OPTN*, which pinpointed the role of autophagic pathway in ALS. With the detection of eight loss of function *TBK1* mutations in 13 fALS pedigrees among 252 fALS cases, it was confirmed that haploinsufficiency of *TBK1* causes ALS (Freischmidt *et al.,* 2015). Another gene burden analysis with 1,022 index fALS cases and 7,312 control samples revealed an association between *NIMA related kinase 1 (NEK1)* loss of mutations and fALS, and replication studies showed that *NEK1* is a risk factor in ALS with 3 % frequency among 10,589 fALS and sALS samples (Kenna *et al.,* 2016).

1.4.5.3. Project MinE. The largest multi-national whole-genome consortium of ALS aims to sequence 15,000 patients with ALS and 7,500 controls to uncover associations between specific variations/genes and ALS. In the pilot study of the project, three loci harboring the genes c*hromosome 21 open reading frame 2* (*C210RF2)*, *myelin-associated oligodendrocyte basic protein* (*MOBP)* and s*ec1 family domain containing 1* (*SCFD1)* were associated with ALS risk at genome-wide significance (van Rheenen *et al*., 2016). As the number of samples from the participating countries increases, the quality of the studies will get better with higher amount of data.

2. PURPOSE

ALS is the most common motor-neuron disease and has a complex genetic background. Up to date, more than 40 genes were identified as pathogenic, however the genetic components of this progressively degenerative neurological disease have not been understood completely yet. Considering the overlap between ALS and other MNDs including HSP, SMA, BVVL, this thesis focuses on the identification of genetic mutations leading to several distinct phenotypes in MND patients.

Turkey is a large country with a high birth rate and a high degree of consanguinity on one hand and a large ethnic heterogeneity on the other. Thus, Turkey harbors potential mutations in several genes which might be involved in ALS pathogenesis. Hence, in this study, our cohort consists of typical late-onset and dominant forms of ALS as well as juvenile-onset recessive ALS which is due to consanguinity.

This thesis aims to;

- Establish an efficient *in-silico* workflow to process the WES data.
- Characterize novel genotype-phenotype associations in MNDs by
  (i) identifying both known and novel mutations in known ALS-MND genes.
  (ii) describing mutations in novel genes associated with an MND phenotype.

3. MATERIALS

3.1. Subjects

In the framework of this thesis 57 families including 81 patients referred to our laboratory with an initial diagnosis of motor neuron disease were examined. In 35 out of these families consanguinity was observed; hence in first line an autosomal recessive mode of inheritance was expected. For the remaining families, all transmission modes were considered including autosomal recessive (true homozygosity and compound heterozygosity), autosomal dominant, and X-linked (Figures 3.1 – 3.8). The initial clinical diagnoses of the families were ALS and/or other motor-neuron diseases, phenotypically similar to ALS: SBMA, HSP, CMT, SMA, SMARD1[1], MMND[2], and BVVL[3].

All patients were screened for four common ALS genes: *SOD1, C9ORF72, TDP-43* and *FUS*. After exclusion of these genes, the families were selected for WES, based on the presence of sufficient clinical data and/or number of available family members (Table 3.1).

The study content was approved by the Ethics Committee on Research with Human Participants (INAREK) at Boğaziçi University. Clinical evaluations of the index cases were performed in collaboration with expert neurologists from several hospitals throughout Turkey. Blood samples were collected into EDTA-containing tubes with written consent.

---

[1] spinal muscular atrophy with respiratory distress type 1

[2] madras motor neuron disease

[3] brown-vialetto-van laere syndrome

Table 3.1. Families investigated in this study.

	ID	Gender	AO	Consanguinity	# of samples subjected to WES	Clinics
Family 1	P1	F	31	+	3	distal motor neuropathy
Family 2	P2	M	9	+	4	Atypical ALS
Family 3	P3	F	10	+	5	Atypical ALS
Family 4	P4	M	24	+	3	ALS
Family 5	P5	F	13	+	1	HSP
Family 6	P6	F	1	+	4	MND
Family 7	P7	F	20	+	5	ALS
	P8	M	13			
	P9	F	20			
Family 8	P10	M	3	+	4	HSP
	P11	M	3			
Family 9	P12	F	25	+	4	ALS
Family 10	P13	F	NA	+	1	MMND-BVVL
Family 11	P14	M	17	+	1	MND
Family 12	P15	M	20	+	1	MND
Family 13	P16	M	2	+	1	MND
Family 14	P17	F	childhood	-	4	CMT
	P18	F				Scapuloperoneal SMA
Family 15	P19	M	52	-	1	ALS
Family 16	P20	M	43	-	4	CMT
	P21	M	11			
	P22	F	11			
Family 17	P23	F	60	-	2	ALS/FTD
	P24	F	60			

Table 3.1. Families investigated in this study (cont.).

	ID	Gender	AO	Consanguinity	# of samples subjected to WES	Clinics
Family 18	P25	F	48	-	5	ALS
	P26	F	48			
	P27	M	47			
Family 19	P28	F	21	-	1	ALS
Family 20	P29	F	16	-	1	MMND
Family 21	P30	M	17	+	3	ALS
Family 22	P31	F	10	+	3	ALS
	P32	F				
Family 23	P33	M	19	+	3	ALS
Family 24	P34	M	12	+	4	ALS
Family 25	P35	M	35	+	3	ALS
Family 26	P36	M	25	+	4	ALS
Family 27	P37	F	~3 months	+	2	SMARD1
	P38	F				
Family 28	P39	M	25	-	4	ALS/PLS
Family 29	P40	F	9	+	6	ALS
Family 30	P41	F	57	+	2	ALS
	P42	M	44			
Family 31	P43	M	20	+	1	ALS
Family 32	P44	F	52	+	6	ALS
	P45	M	40			
Family 33	P46	F	58	-	1	ALS
Family 34	P47	F	76	-	1	ALS
Family 35	P48	M	51	-	2	ALS
	P49	F	NA			
Family 36	P50	F	40	-	4	ALS
	P51	M	NA			
	P52	F	NA			

Table 3.1. Families investigated in this study (cont.).

	ID	Gender	AO	Consanguinity	# of samples subjected to WES	Clinics
Family 37	P53	M	46	-	1	ALS
Family 38	P54	M	40	-	2	ALS
	P55	F	67			
Family 39	P56	M	52	-	1	ALS
Family 40	P57	M	46	-	1	ALS
Family 41	P58	M	65	-	1	ALS
Family 42	P59	M	41	-	1	ALS
Family 43	P60	M	39	-	2	ALS
	P61	F	24			
Family 44	P62	F	54	-	3	ALS
Family 45	P63	M	52	-	2	ALS
Family 46	P64	M	38	+	1	ALS
Family 47	P65	M	24	+	1	ALS
Family 48	P66	M	6	+	1	ALS
Family 49	P67	M	14	+	1	ALS
Family 50	P68	F	22	+	1	ALS
Family 51	P69	M	childhood	+	7	BVVL
	P70	M				
	P71	M				
	P72	F				
Family 52	P73	M	3	+	3	BVVL
Family 53	P74	M	NA	+	1	BVVL
Family 54	P75	F	childhood	-	6	HSP
	P76	F				
	P77	M				
	P78	F	55			
Family 55	P79	F	NA	+	1	HSP
Family 56	P80	M	NA	-	4	ALS
Family 57	P81	F	20	+	1	ALS

3.1.1. Family Trees

3.1.1.1. Pedigrees with an Autosomal Recessive (AR) Inheritance

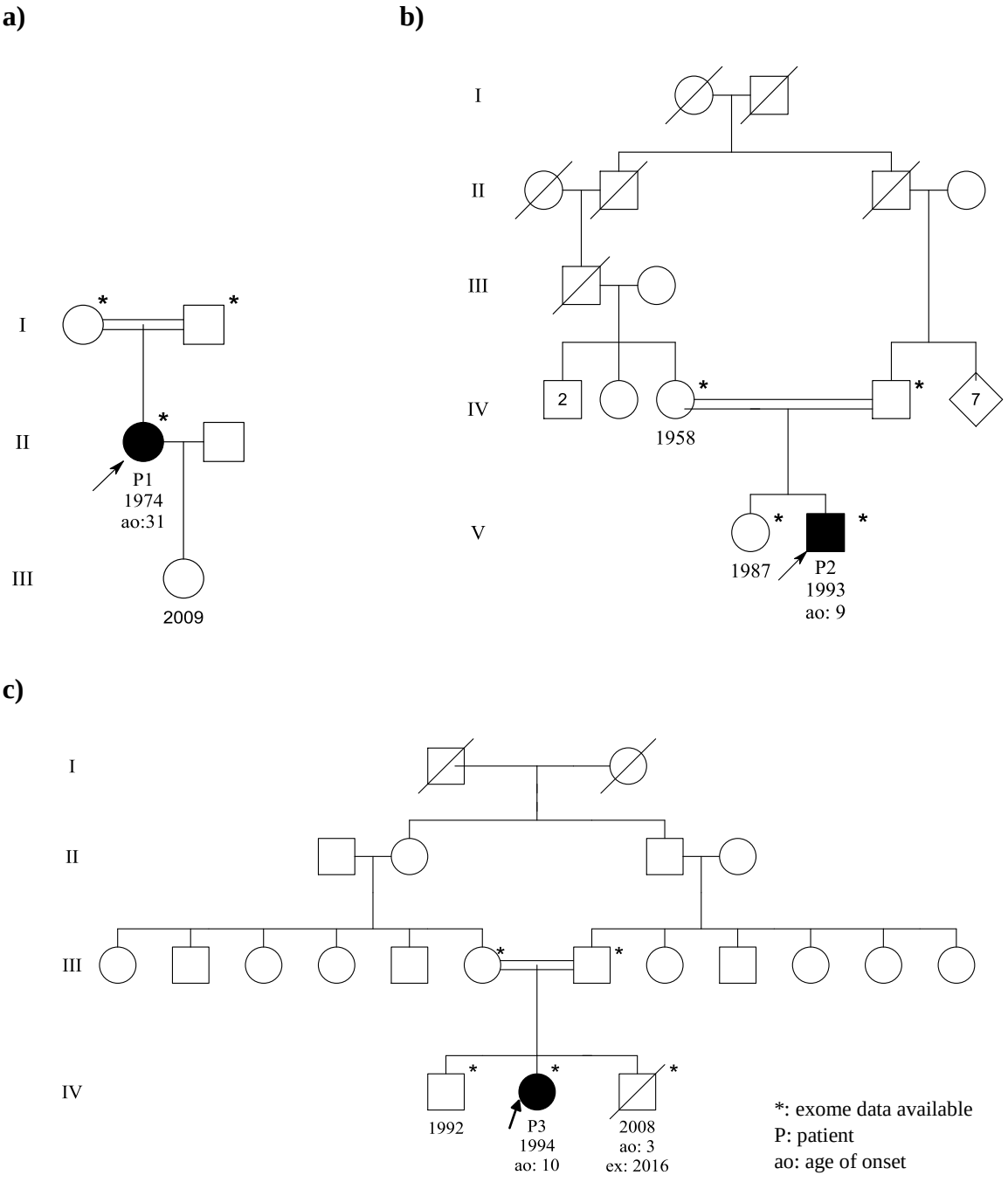


Figure 3.1. Pedigrees of families with an AR inheritance. A) Family 1 (Patient P1), b) Family 2 (Patient P2) and c) Family 3 (Patient P3).

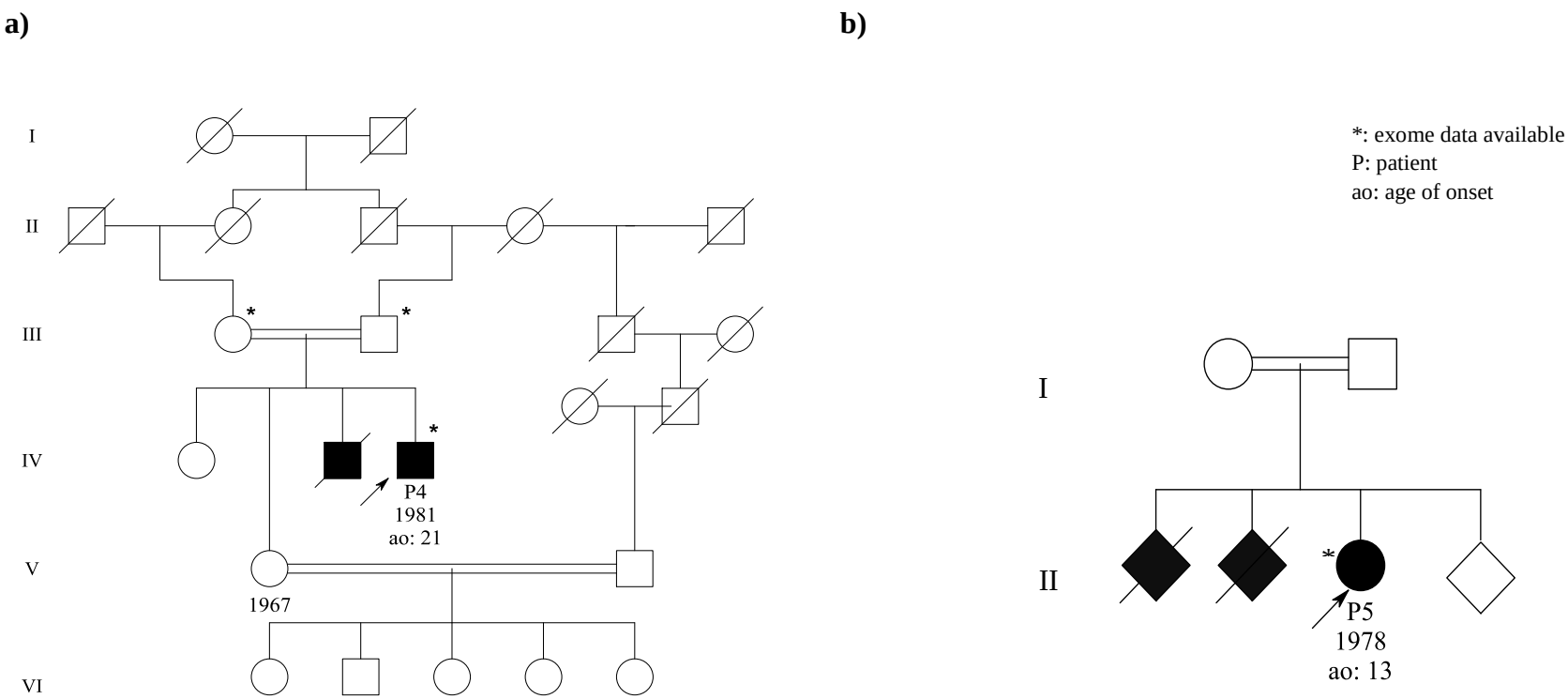


Figure 3.2. Pedigrees of families with an AR inheritance. A) Family 4 (Patient P4) and b) Family 5 (Patient P5).

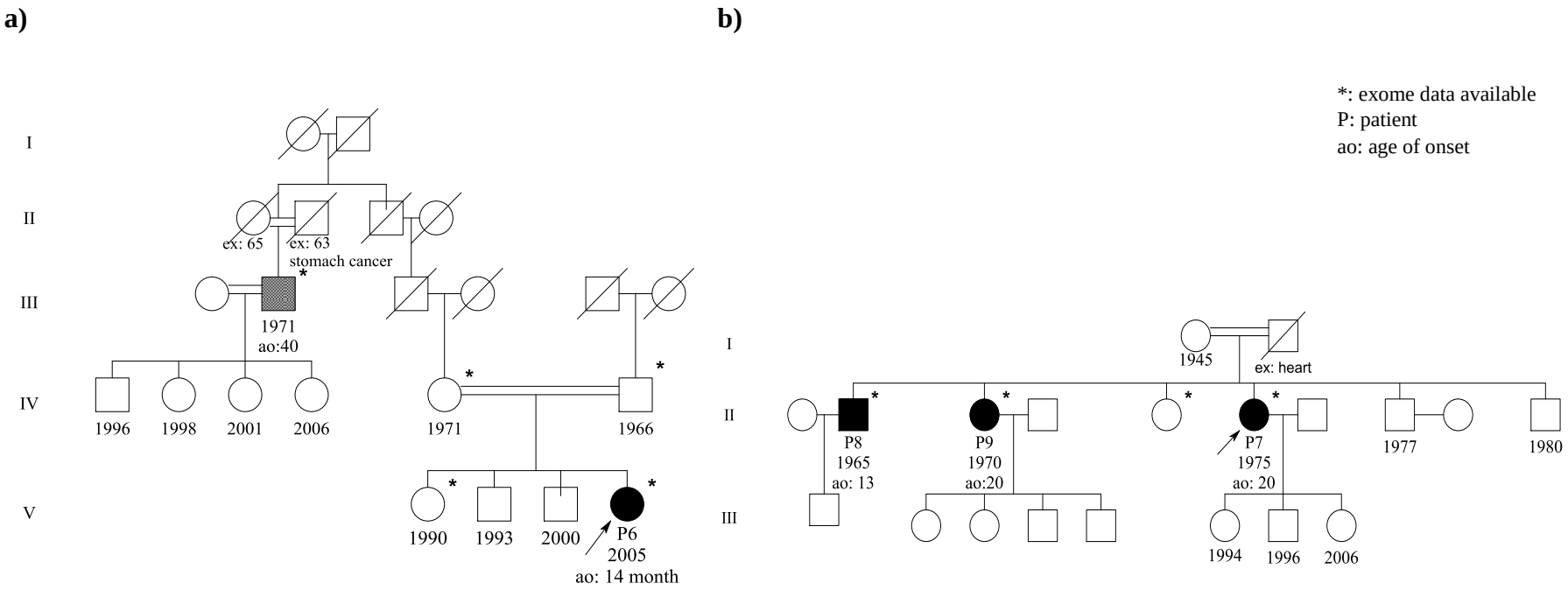


Figure 3.3. Pedigrees of families with an AR inheritance. A) Family 6 (Patient P6) and b) Family 7 (Patient P7-P9).

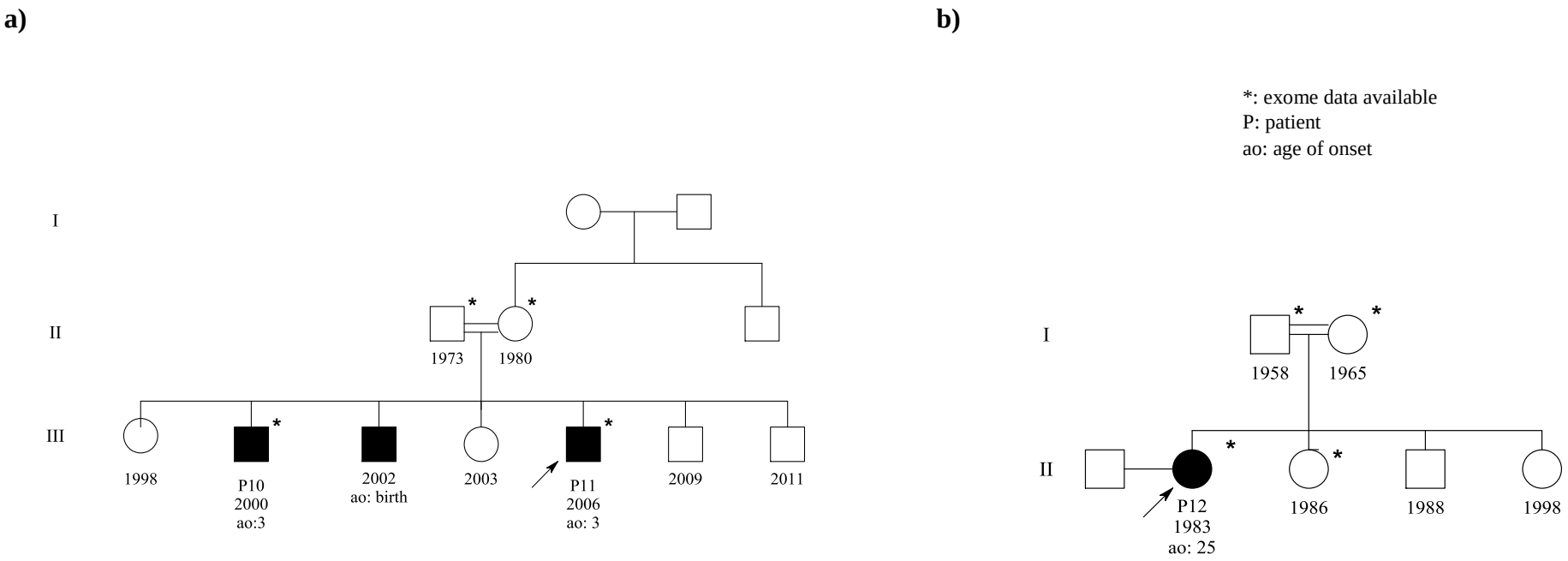


Figure 3.4. Pedigrees of families with an AR inheritance. A) Family 8 (Patient P10 and P11), b) Family 9 (Patient P12)

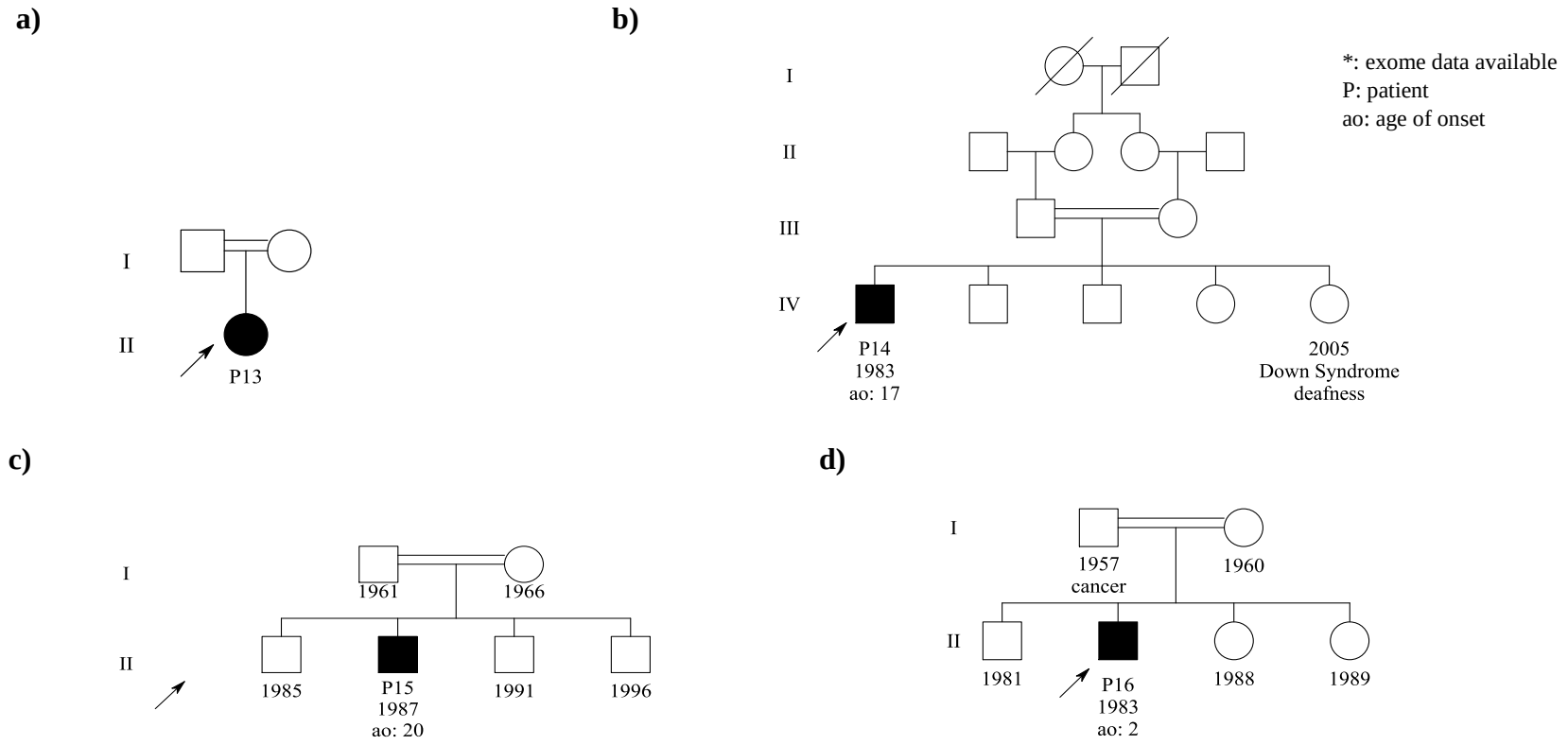


Figure 3.5. Pedigrees of families with an AR inheritance. A) Family 10 (Patient P13), b) Family 11 (Patient P14), c) Family 12 (Patient P15), d) Family 13 (Patient P16)

3.1.1.1. Pedigrees with Autosomal Dominant (AD) Inheritance

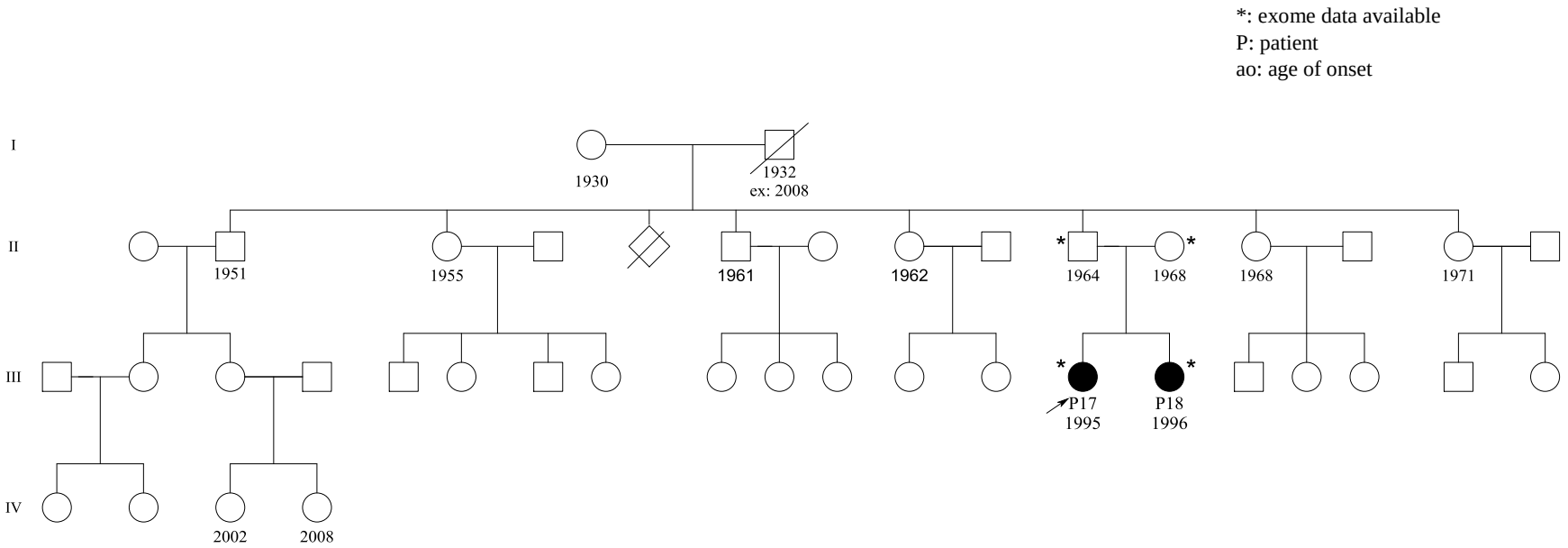


Figure 3.6. Pedigree of the family 14 (Patient P17 and P18).

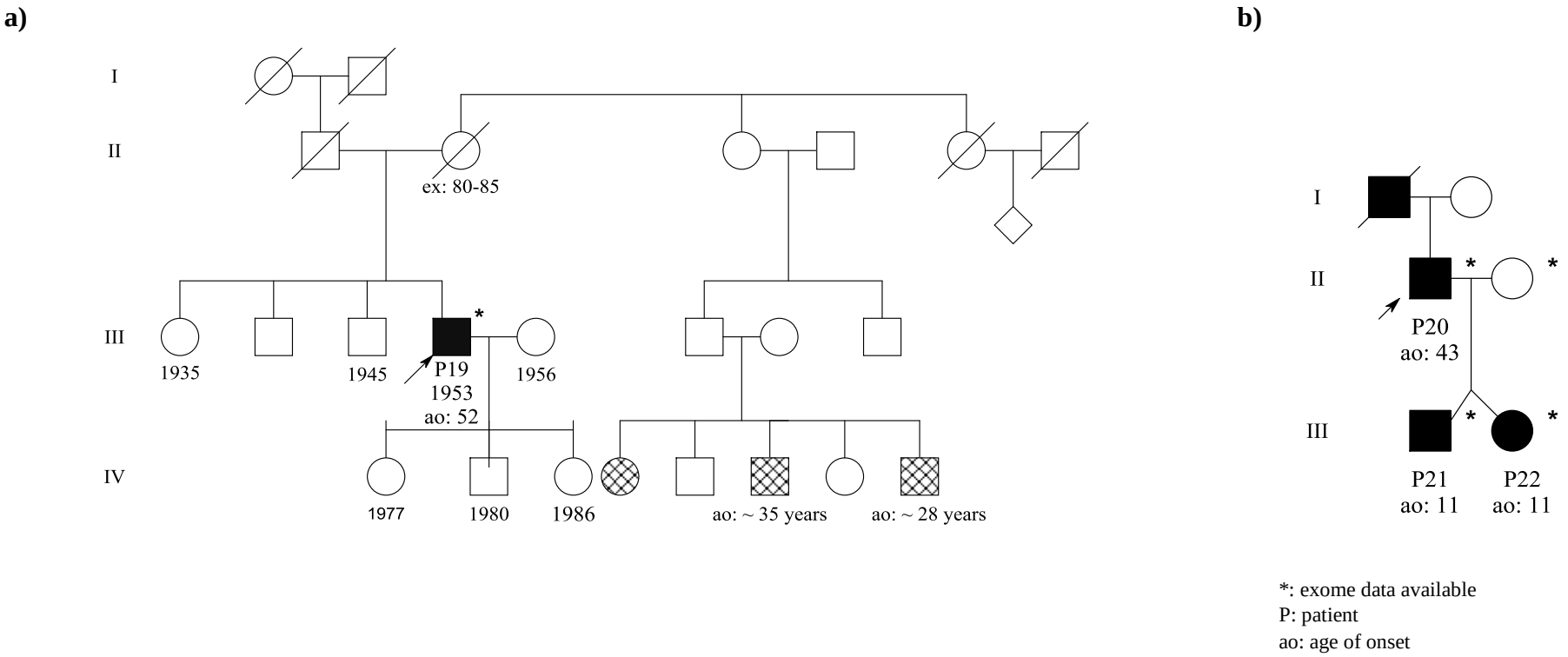


Figure 3.7. Pedigrees of families with an AD inheritance a) Family 15 (Patient P19) and a) Family 16 (Patient P20-22).

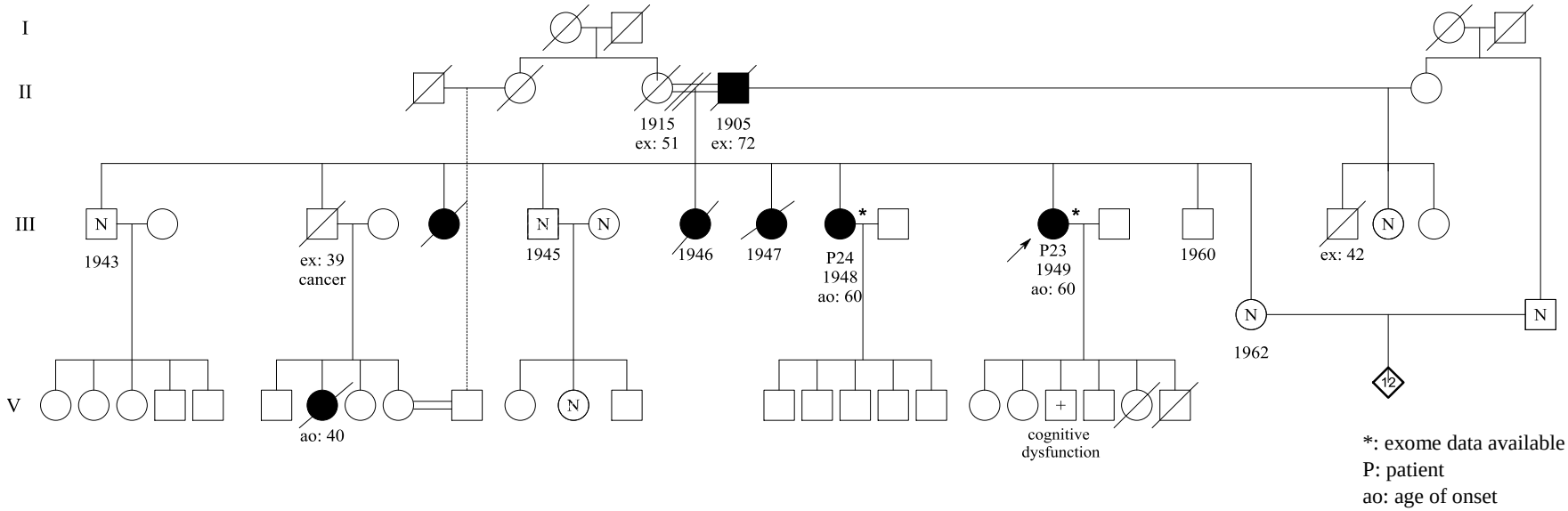


Figure 3.8. Pedigree of the family 17 with an AD inheritance (Patient P23 and Patient P24).

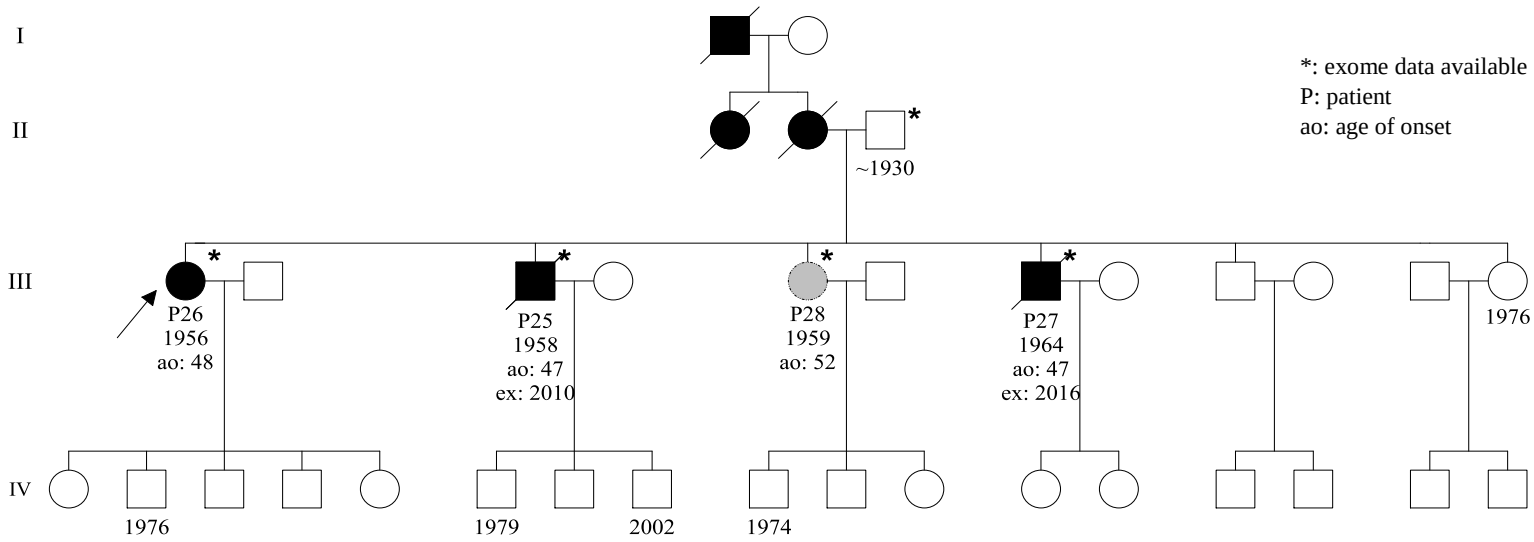


Figure 3.9. Pedigree of the family 18 (Patient P25, Patient P26 and Patient P27) showing an AD inheritance pattern.

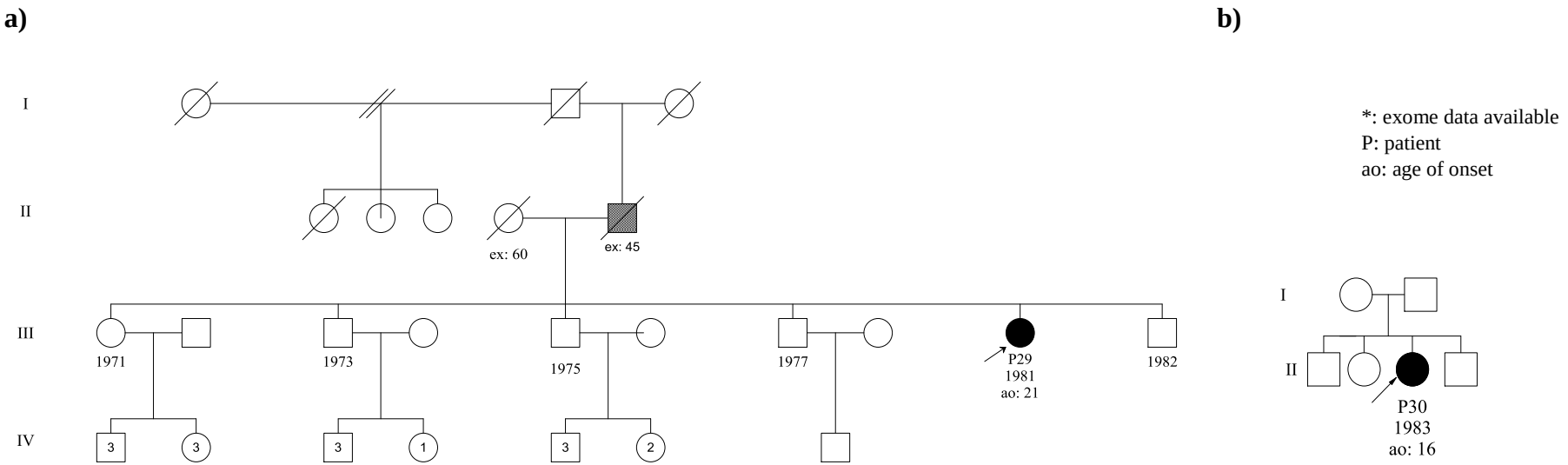


Figure 3.10. Pedigrees of the family 19 (Patient 28) (a), family 20 (Patient 29) showing AD inheritance pattern.

3.2. Whole Exome Sequencing Platforms and Enrichment Kits

Whole exome sequencing was outsourced to different institutions and companies, either in the framework of a collaboration or commercially. These were University of Massachusetts Medical School (UMASS), Scientific and Technological Research Council of Turkey (TUBITAK), Macrogen Inc., DNA Laboratories, Medipol University and The Center of Applied Genomics (TCAG). Sequencing was performed by NextSeq 500, Illumina HiSeq 2000, HiSeq 2500 and HiSeq 4000 using exome enrichment kits listed in Table 3.2.

Table 3.2. Whole exome sequencing platforms and enrichment kits.

Sequencing platform	**Kit**	**Company/ Institution**
HiSeq 2000	Roche SeqCap EZ Whole Exome V2, MedExome	UMASS
HiSeq 2000	Roche SeqCap EZ Whole Exome V3, TruSeq Exome Library Prep Kit	TUBITAK
HiSeq 2000	Roche SeCap EZ Whole Exome V2	Medipol University
HiSeq 2000	Agilent SureSelect Human All Exon V5	TCAG
NextSeq 500	Nextera Rapid Capture Exome	DNA Laboratories
HiSeq 2000 HiSeq 2500, HiSeq 4000	Agilent SureSelect Human All Exon V5, V5-post,	Macrogen Inc.

3.3. Hardware

Hardware features of computers and the network-attached storage system (NAS) used in the framework of this thesis, are listed in Table 3.3.

Table 3.3. Features of the computers and the network-attached storage system

Type	**Features**	**Manufacturer**
Computer	Intel I Core I i7-4930K CPU @3.40GHz 3.40 GHz, 12 core, SSD hard disk, 32GB RAM	Hewlet-Packard (HP), USA
	XPS L412Z Intel I Core I i7-2640M CPU @ 2.80GHz 2.80 GHz	Dell, USA
Network-attached storage system (NAS)	DSM 5.2-5644 Update 5	Synology Inc.

3.4. Software, Online Databases and Bioinformatics Tools

Computational workflow of WES data analysis was executed on the Ubuntu 14.04 operating system. Bioinformatics analysis and evaluation were performed both on Ubuntu 14.04 and Windows 8 operating systems. Open-source bioinformatics software, tools and online databases used in this thesis are listed in Table 3.4.

Table 3.4. Software, bioinformatics tools and databases

Software / Database	**Description**
Ubuntu 14.04 operating system / Biolinux	Operating system in which bioinformatics packages are installed
Teamviewer	A package for remote control
Burrows-Wheeler Aligner (BWA)	Software package for mapping sequences against a reference genome
Genome Analysis Toolkit (GATK) (McKenna et al., 2010)	A toolkit for variant discovery in high-throughput sequencing data
SamTools (H. Li et al., 2009)	A package for alignment, manipulating the reads in the SAM / BAM format
Annovar (K. Wang et al., 2010)	Functional annotation of genetic variations
Vcftools (Danecek et al., 2011)	A package to summarize and filter the variations on VCF files
R (R Development Core Team, 2011)	Software for statistical computing and presentation
Varsifter (Teer et al., 2012)	A Java program designed to parse and filter the high throughput data
PLINK (Purcell et al., 2007)	Genome data analysis toolset
Rfflow (Rfflow, 1989)	Tool for drawing flowcharts and pedigrees
Integrative Genomics Viewer (IGV) (IGV (Integrative Genomic Viewer), 2013)	Visualization tool for interactive exploration of integrated genomic datasets
The Reference Sequence Database	A reference genome database for vertebrates
ExAC (Lek et al., 2016)	Exome Aggregation Consortium
Online Mendelian Inheritance in Man (OMIM) (McKusick-Nathans Institute of Genetic Medicine)	An online catalog of human genes and disorders
ClinVar (Landrum et al., 2014)	A public archive of relationships among sequence variation and human phenotype
NHLBI GO Exome Sequencing Project	A database of 6500 human exome
1000 Genomes	A comprehensive resource of human genetic variation

Table 3.4. Software, bioinformatics tools and databases (cont.).

Software / Database	**Description**
GeneCards (Weizmann Institute of Science, 2016)	A human gene database including clinical and functional information
dbSNP (Sherry et al., 2001)	A catalog of SNVs and small indels
BioMart/ Ensembl (Smedley et al., 2015)	A web-based tool for comparative genomics
Polymorphism Phenotyping v2 (PolyPhen2) (Adzhubei et al., 2010)	A web server that predicts the possible impact of amino acid substitutions
SIFT (P. C. Ng and Henikoff, 2003)	A web server that predicts the possible impact of amino acid substitutions
UCSC in silico (UCSC, 2002)	WEB browser of University of California Santa Cruz

4. METHODS

4.1. Sample Preparation and Whole Exome Sequencing

DNA was extracted from whole blood (1000 µl) of subjects using the MagNA Pure Compact Instrument (Serial Number: MPCB 511, Roche) and the MagNA Pure Compact Nucleic Acid Isolation Kit I. Whole exome sequencing was outsourced to institutions and companies stated in section 3.1. Sequencing in these institutions was performed on different platforms of NextSeq 500, Illumina HiSeq 2000, HiSeq 2500 and HiSeq 4000.

4.2. Alignment and Variant Calling

Bioinformatic analysis of raw paired-end reads generated by Illumina was performed in an in-house computational pipeline. The main steps of the pipeline are the alignment and variant calling followed by the annotation of the candidate variations. Raw sequence reads stored in the FASTQ files were aligned to human reference genome GRCh37 plus the decoy via Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Aligner basically map the FASTQ reads to the given version of the human genome generating sequence alignment map (SAM) files. Using SAMtools package, the mapped reads stored in SAM files were converted into the binary aligned map (BAM) format, which has exactly the same information, but in a more compact form. In the final step of the alignment, false duplicates were removed and cleaned sequences were sorted and indexed using SAMtools (H. Li *et al*., 2009). Recommended indel realignment and base score recalibration were the pre-processing steps of the data prior to variant calling by Genome Analysis Toolkit (GATK) of Broad Institute (McKenna *et al.*, 2010). Single nucleotide variations (SNV) and small indels were called for each individual from their separate bam files by the HaplotypeCaller tool of GATK. At the end of this step, genomic variant call format (gvcf) files containing the information of both variant and reference sites were obtained. Vcf files for each family were generated from gvcfs of the family members at the same joint genotyping step via GenotypeGVCFs tool of GATK; this reduces the false positives. SNV and indel recalibration

of the raw vcf files were performed based on GATK Best Practices recommendations by Broad Institute (Appendix A).

4.3. Quality Check Metrics

Quality check was undertaken for each sample to detect the presence of any outlier sample or site. For this approach, VCFtools was applied to obtain the depth of coverage, the rate of transition and transversion (Ts/Tv) and missing genotype rate of individuals (Danecek *et al.,* 2011).

4.4. Principal Component Analysis and Inference of Relationships

Principal component analysis (PCA) was applied to identify population clusters, heterogeneity and to detect the outliers in the cohort. Identity-by-Descent (IBD) estimation was performed on the family vcf samples to confirm the relationships among individuals. Pi-hat scores were calculated by PLINK v1.9 to check the degree of relatedness among the family members (Purcell *et al.*, 2007).

4.5. Homozygosity Mapping

Homozygosity mapping was performed in consanguineous families by PLINKv1.9. New files were created including family, gender and phenotype information to be used as input for PLINK. Family vcf and the newly generated files were converted into binary PLINK hard calls with a genotype quality filter of 30 (as minimum 30 reads were needed per SNP to be included in the analysis). If there were any additional family members in the vcf file, the variants in linkage disequilibrium were pruned with r2 threshold 0.2 (Purcell *et al.*, 2007). Runs of homozygosity (ROHs) were detected for each case with optimized parameters for WES data (Table 4.1). The distribution of homozygous stretches were displayed based on their length using R plotting.

Table 4.1. Parameters of runs of homozygosity detection in PLINK.

Parameter	Threshold value
Size threshold (kb) to call on ROH	500
SNP number threshold to call an ROH	10
Sliding window size in SNPs	20
Allowed missing SNPs in a window	10
Proportion of homozygous window threshold	0.05
Minimum SNP density to call an ROH	200, 400
Maximum allowed gap between two SNPs	2000
Allowed heterozygous SNPs in a window	1,2

4.6. Generation of In-house Cohort

An in-house data-set was generated including 330 individuals with several neurological diseases and 100 healthy family members. The variants were called and stored for each chromosome by joint genotyping of the GVCFs of individuals, generating 25 chromosomal (22 autosomal, X, Y and mitochondrial) vcfs of 430 samples. These in-house data-set is currently being used for the screening of candidate genes/variants in our cohort for a more sensitive variant filtration which would consider population-specific common variations.

4.7. Annotation and Prioritization of Variations

Structural and functional annotation of the variations called was performed using ANNOVAR (Wang *et al*., 2010). Minor allele frequencies (MAF) of the variants were obtained from several data-sets consisting of dbSNP138, 1000 Genomes (October 2014 release), Lung and Blood Institute (NIHLBI) Exome Sequencing Project (ESP) 6500 exome, The Exome Aggregation Consortium (ExAC). Functional effects and evolutionary conservation rate of the variants were predicted based on their SIFT, PolyPhen-2, MutationTaster, GERP and PhyloP scores. Clinical information of variations and genes were acquired from the Online Mendelian Inheritance in Man (OMIM) and ClinVar databases to check the presence of any association to previously defined phenotypes. Variant filtration

was performed based on the MAF values; variations present in the population with a frequency greater than 1% were considered as polymorphisms and excluded from the analysis. However, the information on functional effects and evolutionary conservation rates of the variants were not used in the filtration step as they are likely to give false positive results. For the priorization of variations, a java-based software VarSifter was applied (Teer *et al.*, 2012). Vcf files were parsed based on their annotation terms and variations were prioritized according to the inheritance pattern on the pedigrees (Figure 4.1).

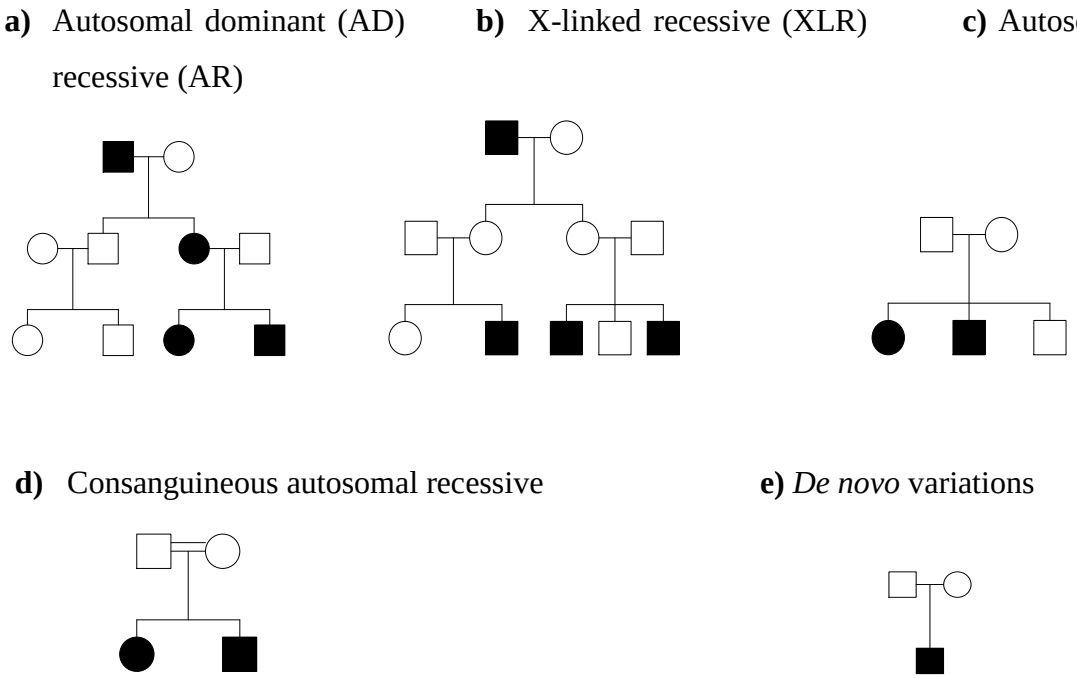


Figure 4.1. Example pedigrees with different inheritance patterns. Autosomal dominant inheritance: heterozygous variations in affected individuals & wild type in unaffected individuals (a), X-linked recessive inheritance: X chromosome variations in affected males & heterozygous in carriers (b), Autosomal recessive inheritance: compound heterozygous variations in affected siblings & heterozygous variations in unaffected individuals (c), Consanguineous autosomal recessive inheritance: homozygous variations in affected siblings & heterozygous variations in unaffected individuals (d), *De novo* variations: Heterozygous variations in affected individual & wild type in unaffected individuals (e).

4.8. Validation of WES Results by Sanger Analysis and Family Segregation

The presence and segregation of the candidate variations obtained from bioinformatic analysis were validated by PCR-based Sanger sequencing in our laboratory. Primers to amplify the regions containing the variation were retrieved from the literature and confirmed via UCSC in silico PCR tool (see Appendix B).

5. RESULTS

In this study, whole exome sequencing data of 57 Turkish patients, in majority with MND, and unaffected family members were evaluated. Analyses, consisting of sequence quality control metrics and family-based variant prioritization, is presented in the following sections.

5.1. Sequencing Quality Metrics

Sample-based quality control was performed by calculating mean depth of coverage, missing genotype rate and Ts/Tv ratio. Missingness and Ts/Tv ratio are reported for each individual, and mean depth of coverage was compiled for calibrated family-vcf files. (Figure 5.1-5.3). The values can be found in Appendix C.

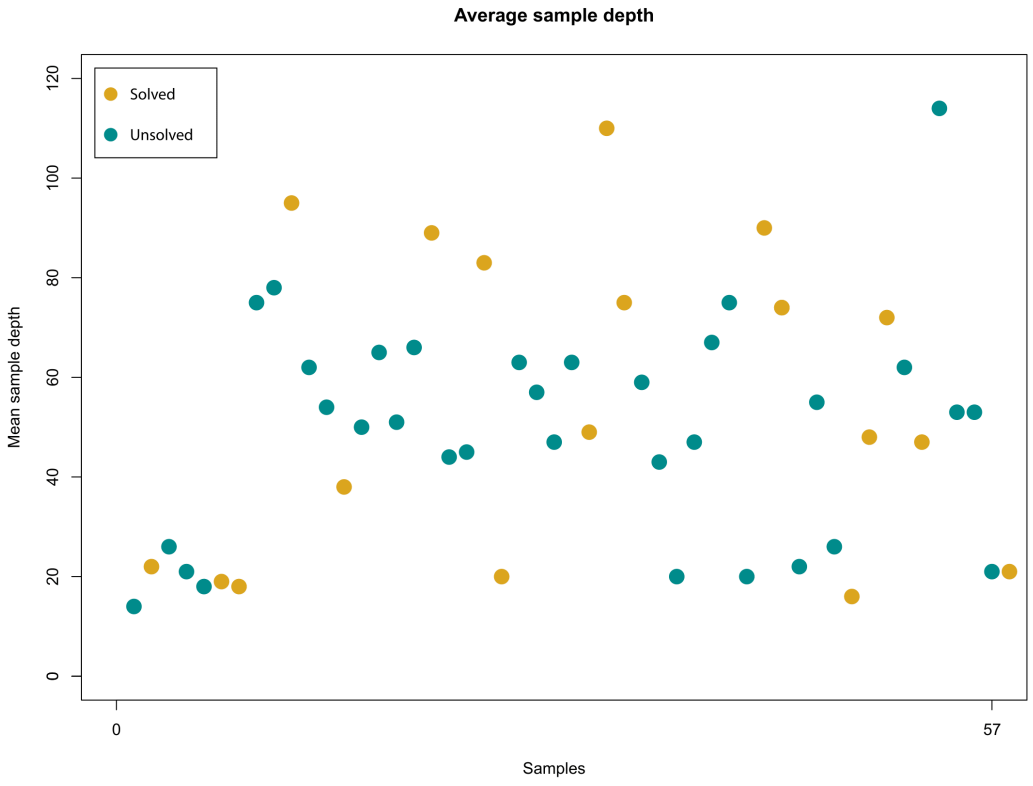


Figure 5.1. Mean depth of coverage for samples.

The mean depth of coverage for samples ranged from 20-120 X with an average of 63.8 (Figure 5.1). The irregular distribution of samples solved and unsolved in the graph shows no association between coverage and the success rate of mutation identification.

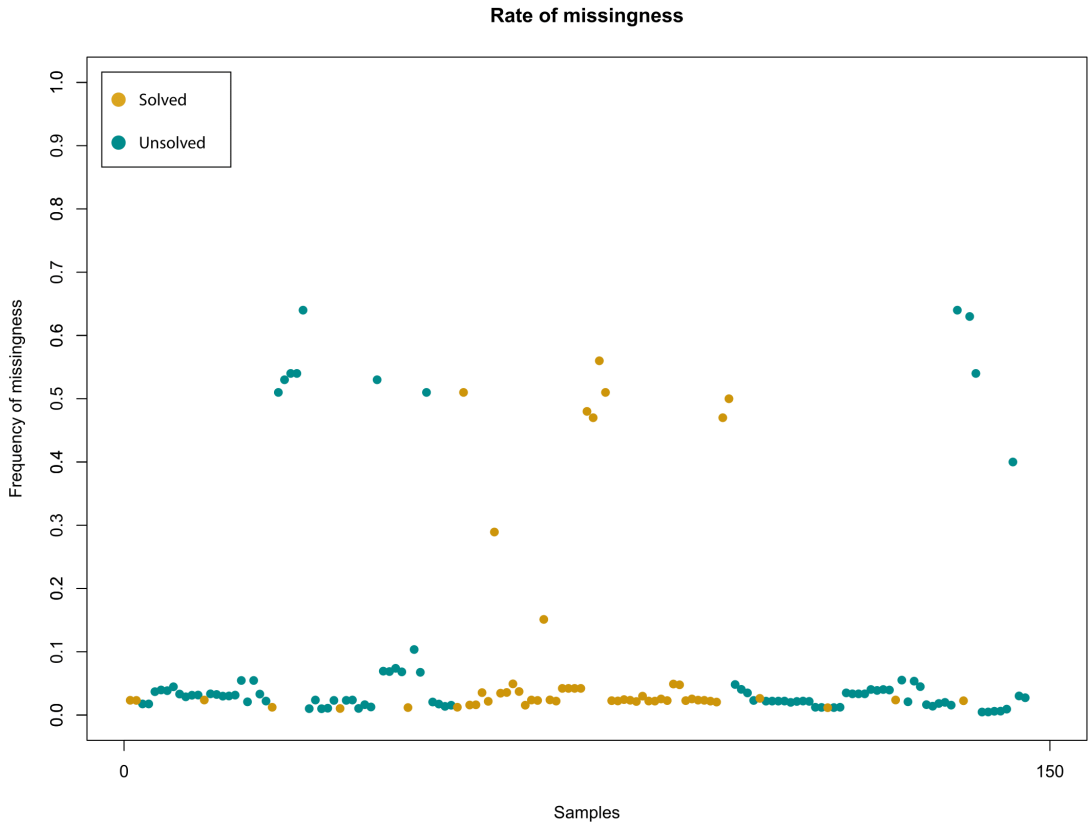


Figure 5.2. Frequency of missingness for all individuals.

The majority of the individuals had a ratio of missingness less than 0.01. The average of missingness among individuals was 0.0925 with a standard deviation of 0.1677. Some individuals had significantly higher missingess, however, these were not excluded from the study. There were some cases in which the disease-causing mutation could be identified, even at the high missing ratio of nearly 0.6. The mean of Ts/Tv ratio was 2.218 with a standard deviation of 0.079, ranging from 2.041 to 2.448. No outliers were detected based on this quality metric.

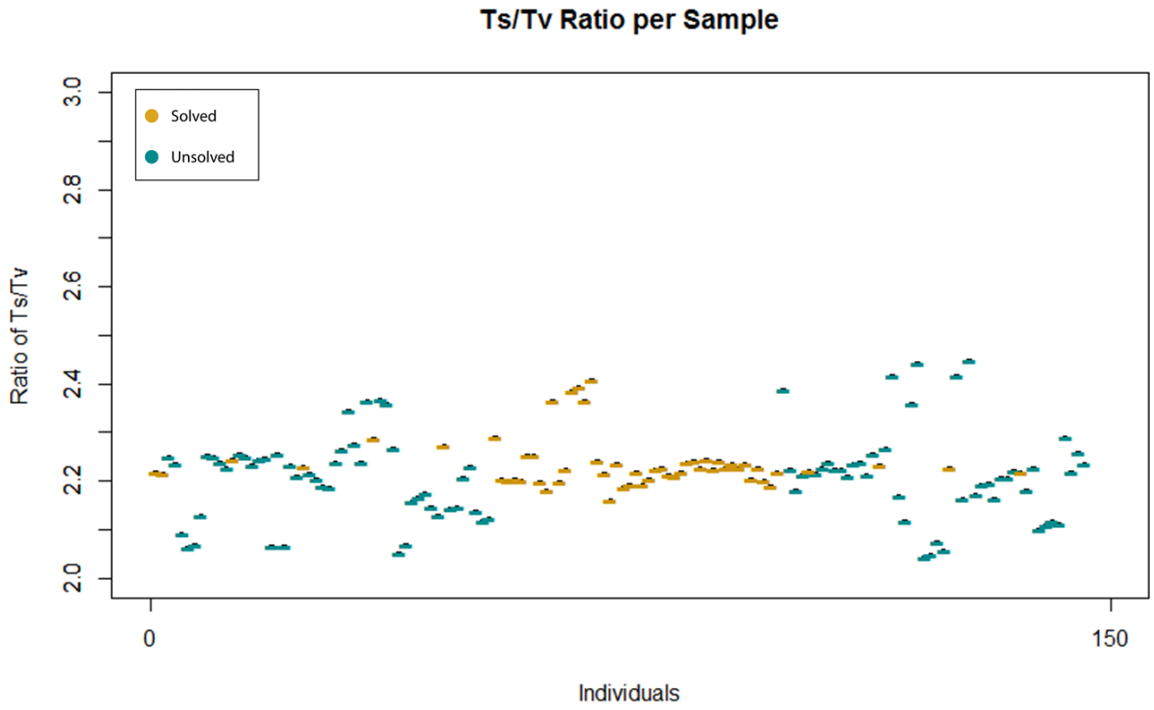


Figure 5.3. Ratio of Ts/Tv for all individuals.

5.2. Population Stratification

Principal component analysis was performed to identify and distinguish the population clusters in the study cohort. Participants were divided into three main clusters using the first four principal components (Figure 5.4).

5.3. Whole Exome Data Analysis

In this study, 19 different mutations in 21 distinct genes were detected. Thus, we were able to identify the genetic cause in 20 out of 57 families (35%). The step-by-step procedure of the bioinformatic evaluation of the samples solved is compiled in Table 5.1. The pathogenic variations identified, the inheritance pattern, initial referral and final diagnosis via deep phenotyping and OMIM associations of the genes are listed in Table 5.2. Depth of coverage, minor allele frequencies (MAF) and conservation scores retrieved from prediction

tools for all variations identified are presented in Table 5.3. The preliminary evaluation of the samples not solved in the framework of this study is presented in Table 5.4.

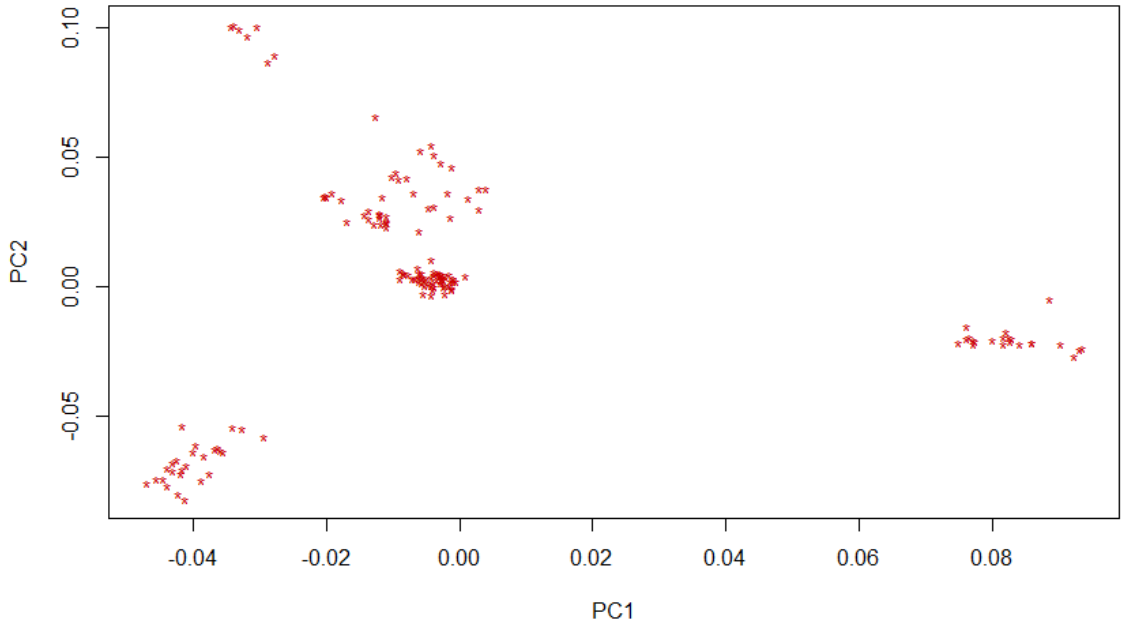


Figure 5.4. Multi-dimensional scaling plot of study cohort.

A total of 11 homozygous mutations in the genes *DNAJB2, C19ORF12, PANK2, IGHMBP2, PLEKHG5, SLC12A6, ACADS, SLC52A3, ZFVYE26, SPG11 and SIGMAR1* with an AR inheritance were detected. Homozygosity mapping was performed to narrow down the region of interest in the families with an expected autosomal recessive inheritance pattern due to consanguinity.

Seven heterozygous mutations in *TRPV4, ANG, MPZ, VCP, ERBB4, LRSAM1, SQSTM1* and one X-linked *UBQLN2* mutation were detected with an AD inheritance pattern.

Table 5.1. The number of remaining variations per family after each filtration step.

	# of total variants	type of variation	pedigree info	Minor allele frequency		# of samples
				1000G+ESP6500	ExAC	
Family 1	146639	10125	389	15	6	3
Family 2	149112	10296	393	21	14	4
Family 3	158855	10171	505	28	16	5
Family 4	193799	10994	584	35	14	3
Family 5	254765	105222	4499	181	33	1
Family 6	416684	10691	434	30	13	5
Family 7	546063	11130	141	10	8	5
Family 8	106175	11030	131	9	7	4
Family 9	435337	10944	487	25	9	4
Family 10	334970	11193	4317	198	22	1
Family 11	342520	10896	4398	203	33	1
Family 12	245611	10984	4379	202	26	1
Family 13	294055	11121	4577	212	25	1
Family 14	158511	10443	3501	503	210	4
Family 15	145660	10088	6251	902	351	1
Family 16	121566	10304	777	162	111	4
Family 17	141791	10734	3918	637	201	2
Family 18	155416	10418	855	85	36	5
	128595	10330	2275	390	260	2
Family 19	434505	11055	6744	1664	1247	1
Family 20	307233	11307	7170	1027	558	1

Table 5.2. List of all variations and genes in this thesis and their OMIM associations.

	Inheritance	**Initial diagnosis**	**Variation**			**OMIM Association**
			Gene	**Coding sequence**	**Protein sequence**	
Family 1	AR	distal motor neuropathy	*DNAJB2*	c.757G>A	p.Glu253Lys	distal spinal muscular atrophy
Family 2	AR	Atypical ALS	*C19ORF12*	c.194G>T	p.Gly65Val	NBIA4
Family 3	AR	Atypical ALS	*C19ORF12*	c.194G>T	p.Gly65Val	NBIA4
Family 4	AR	ALS	*C19ORF12*	c.32C>T	p.Thr11Met	NBIA4
Family 5	AR	HSP	*PANK2*	c.427G>A	p.Ala143Thr	NBIA1
Family 6	AR	MND	*IGHMBP2*	c.638A>G	p.His213Arg	SMARD1
Family 7	AR	ALS	*PLEKHG5*	c.1648C>T	p.Gln550Ter	distal spinal muscular atrophy
Family 8	AR	HSP	*SLC12A6*	c.1073+G>A	-	Andermann syndrome
Family 9	AR	MND	*ACADS*	c.1108A>G	p.Met370Val	(SCAD) deficiency
Family 10	AR	BVVL/MMND	*SLC52A3*	c.802C>T	p.Arg268Trp	BVVL1
Family 11	AR	MND	*ZFYVE26*	c.2074delC	p.Lys692fs	SPG15
Family 12	AR	MND	*SPG11*	c.1423C>T	p.Gln478Ter	SPG11, ARJALS

Table 5.2. List of all variations and genes in this thesis and their OMIM associations (cont.).

	Inheritance	Initial diagnosis	Variation			OMIM Association
			Gene	Coding sequence	Protein sequence	
Family 13	AR	MND	*SIGMAR1*	c.355G>A	p.Glu119Lys	ALS-16
Family 14	AD	Scapuloperoneal SMA/CMT	*TRPV4*	c.943C>T	p.Arg315Trp	scapuloperoneal SMA / hereditary motor and sensory neuropathy type 2
Family 15	AD	ALS	*ANG*	c.208A>G	p.Ile70Val	ALS-9
Family 16	AD	CMT	*MPZ*	c.293G>A	p.Arg98His	CMT1B
Family 17	AD	ALS/FTD	*VCP*	c.572G>C	p.Arg191Pro	ALS-14 w/wo FTD
Family 18	AD	ALS	*ERBB4*	c.3334C>T	p.Arg1112Cys	ALS-19
			LRSAM1	c.578G>A	p.Cys193Tyr	CMT2P
Family 19	AD	ALS	*SQSTM1*	c.374A>G	p.Asn125Ser	ALS/FTD/Paget disease of bone
Family 20	XLD	ALS/MMND	*UBQLN2*	c.374A>G	p.Met391Ile	ALS-15 w/wo FTD

Table 5.3. Minor allele frequencies and conservation scores of the mutations described in this thesis.

	Position	Gene	Variation	dbSNP ID	1000G MAF	ExAC MAF	PolyPhen2	SIFT	GERP ++
Family 1	chr2:220149491	*DNAJB2*	p.Glu253Lys	-	-	-	0.28	0.98	4.48
Family 2	chr19:30193884	*C19ORF12*	p.Gly65Val						
Family 3	chr19:30193884	*C19ORF12*	p.Gly65Val	-	-	1.65e-05	0.981	1	4.57
Family 4	chr19:30199322	*C19ORF12*	p.Thr11Met	rs397514477	-	8.31e-06	0.54	0.77	-11.2
Family 5	chr20:3893169	*PANK2*	p.Ala143Thr	-	-	-	0.512	0.98	4.6
Family 6	chr11:68678998	*IGHMBP2*	p.His213Arg	rs137852666	-	-	1	1	4.7
Family 7	chr1:6530920	*PLEKHG5*	p.Gln550Ter	-	-	-	0.74	0.90	4.1
Family 8	chr15:34546548	*SLC12A6*	-	-	-	8.26e-06	-	-	-
Family 9	chr12:121177120	*ACADS*	p.Met370Val	rs566325901	-	0.002223	0.99	0.77	4.39
Family 10	chr20:744413	*SLC52A3*	p.Arg268Trp	rs145498634	-	0.00004945	0.51	0.98	4.6

Table 5.3. Minor allele frequencies and conservation scores of the mutations described in this thesis (cont.).

	Position	**Gene**	**Variation**	**dbSNP ID**	**1000G MAF**	**ExAC MAF**	**PolyPhen2**	**SIFT**	**GERP ++**
Family 11	chr14:68264904	*ZFYVE26*	p.Lys692fs	-	-	-	-	-	-
Family 12	chr15:44943713	*SPG11*	p.Gln478Ter	-	-	-	0.73	0.90	5.71
Family 13	chr9:34635853	*SIGMAR1*	p.Glu355Lys	-	-	-	0.06	0.94	4.32
Family 14	chr12:110236628	*TRPV4*	p.Arg315Trp	rs267607143	-	-	0.99	1	0.22
Family 15	chr14:21161931	*ANG*	p.Ile70Val	rs121909541	-	0.0006095	0.05	0.58	-4.2
Family 16	chr1:161276653	*MPZ*	p.Arg98His	rs121913589	-	-	0.73	0.9	4.26
Family 17	chr9:35065252	*VCP*	p.Arg191Pro	-	-	-	1	1	5.64
Family 18	chr2:212251725	*ERBB4*	p.Arg1112Cys	rs144311212	-	0.00004942	0	1	5.25
	chr9:130230068	*LRSAM1*	p.Cys193Tyr	-	-	0.00004782	0.99	0.99	4.79
Family 19	chr5:179250930	*SQSTM1*	p.Asn125Ser	-	-	0.00001658	0.45	0.77	2.51
Family 20	chrX:56591482	*UBQLN2*	p.Met391Ile	-	-	-	0.99	0.82	2.95

5.3.1. DNAJB2: DnaJ Heat Shock Protein Family (Hsp40) Member B2 (AR)

5.3.1.1. Family 1. Variant filtration and prioritization analysis based on the recessive inheritance pattern resulted in a total of 389 exonic variations and was decreased to six after filtering for MAF. Runs of homozygosity revealed five homozygous regions in the chromosomes 2, 7, 12 and X, harboring the six variations remained from the filtration step (Figure 5.1a). Among these homozygous variations, a novel missense mutation in the *DNAJB2* gene (chr2:220149491, G>A; Glu253Lys) was detected. This gene was previously associated with distal spinal muscular atrophy (MIM #614881). The variant has not been reported in population polymorphism databases including dbSNP and 1000 Genomes Project and is absent in ExAC and our in-house database. Sanger sequencing confirmed the presence and segregation of the mutation in homozygous state in the index case and in heterozygous form in the unaffected parents (Figure 5.1b).

5.3.2. C19ORF12: Chromosome 19 Open Reading Frame 12 (AR)

Two distinct homozygous mutations were identified in the *C19ORF12* gene in three families with consanguinity. The missense Gly65Val mutation was detected in two patients referred to our laboratory with a clinical diagnosis of juvenile onset atypical ALS (with an early age of onset and slow progression, with uneven involvement of UMN and LMN) and the Thr11Met mutation was found in a patient with an initial diagnosis of early onset ALS. Mutations in the *C19ORF12* gene have been previously associated with neurodegeneration with brain iron accumulation (NBIA) type 4 and spastic paraplegia 43 (SPG43) (MIM #614298, #615043).

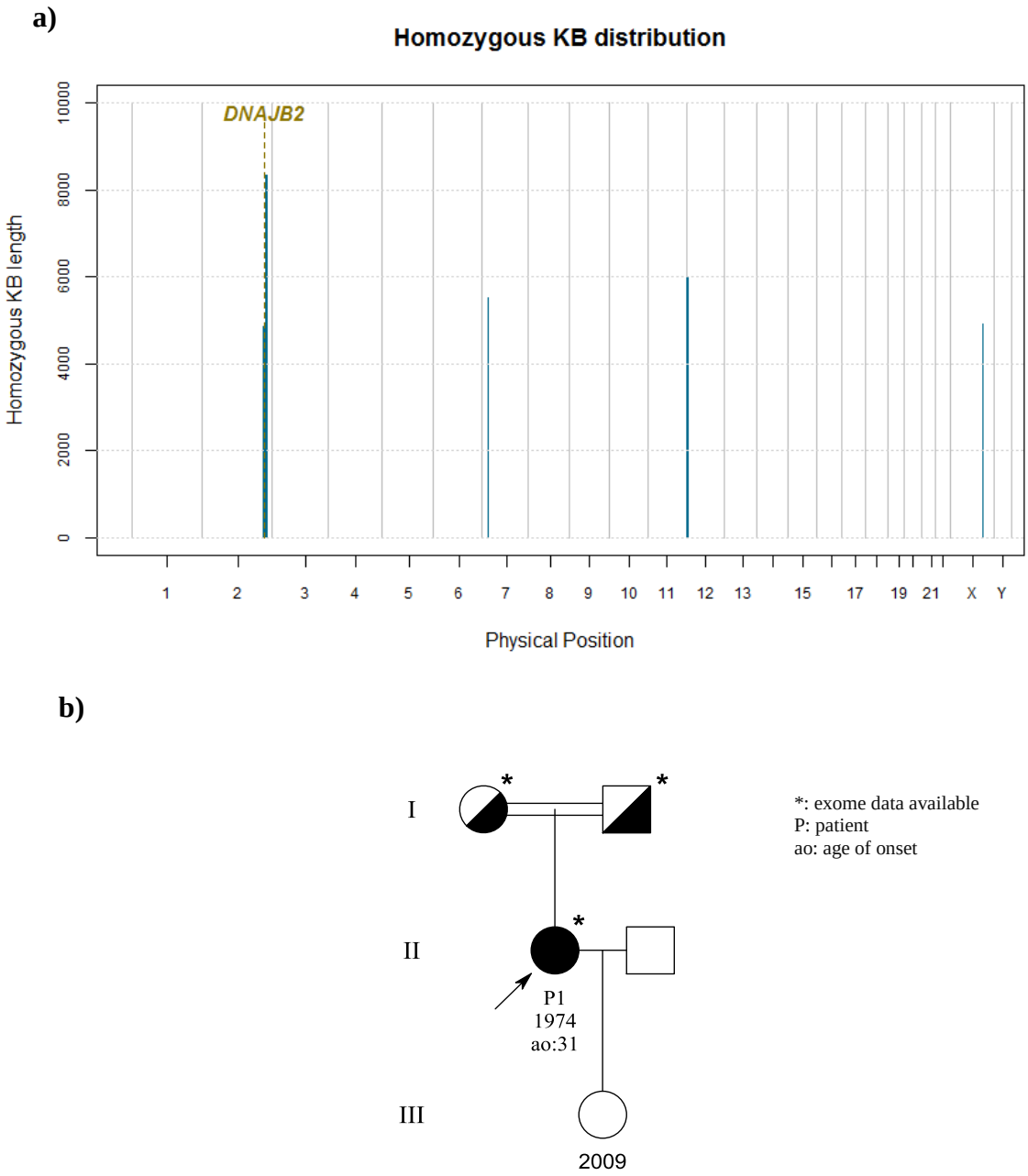


Figure 5.5. Homozygosity mapping plot of the patient (P1) (a) and the segregation of the *DNAJB2* variation in Family 1 (b).

5.3.2.1. Family 2. Evaluation of family 2, including four samples with WES data, resulted in 14 nonsynonymous rare variations which were in homozygous state in the index case and heterozygously present in the unaffected parents. Homozygosity mapping revealed various stretches throughout the genome in which the remaining variations after filtration were located (Figure 5.2a). Among these candidate variants, the missense mutation in the

C19ORF12 gene (chr19:30193884, G>T; Gly65Val) was found as the causative mutation. Other homozygous regions were not harboring any mutations associated with a neurological disease. The candidate mutation is not present in dbSNP and 1000 Genomes Project, but reported in two individuals in ExAC database as heterozygous with a frequency of 1.65e-05. Exome analysis was validated by Sanger sequencing (Figure 5.2b).

5.3.2.2. Family 3. The missense mutation Gly65Val in the *C19ORF12* gene was detected in homozygous state in the affected individual and heterozygously in the parents and in the younger brother. The unaffected elder brother was found to carry the reference sequence in both alleles. Based on the runs of homozygosity, the mutation was located within one of the homozygous segments, the remaining homozygous regions were not harboring any mutation associated with a neurological disease (Figure 5.3a). The younger brother presenting with different and more severe neurological problems, did not carry the *C19ORF12* mutation. No other disease-causative variation(s) was (were) identified in his exome data, although he had passed away at the age of 15. The variation was validated by Sanger sequencing (Figure 5.3b).

5.3.2.3. Family 4. According to runs of homozygosity in the family, 27 homozygous regions were detected in the chromosomes 2, 3, 8, 13, 15, 19 and 21 (Figure 5.4a). These regions harbored 14 rare coding mutations which were homozygous in the affected individual and heterozygous in his unaffected parents. The missense mutation in the *C19ORF12* gene (chr19:30199322, C>T; Thr11Met) was present among the variations. This variant has been reported in heterozygous state in dbSNP (rs397514477) and in the ExAC database with an allele frequency of 0.00000827. Sanger sequencing was performed to the trio subjected to WES, including the unaffected elder sister whose DNA was also available. Segregation among the family was confirmed and the elder sister was shown to carry the mutation in heterozygous state (Figure 5.4b).

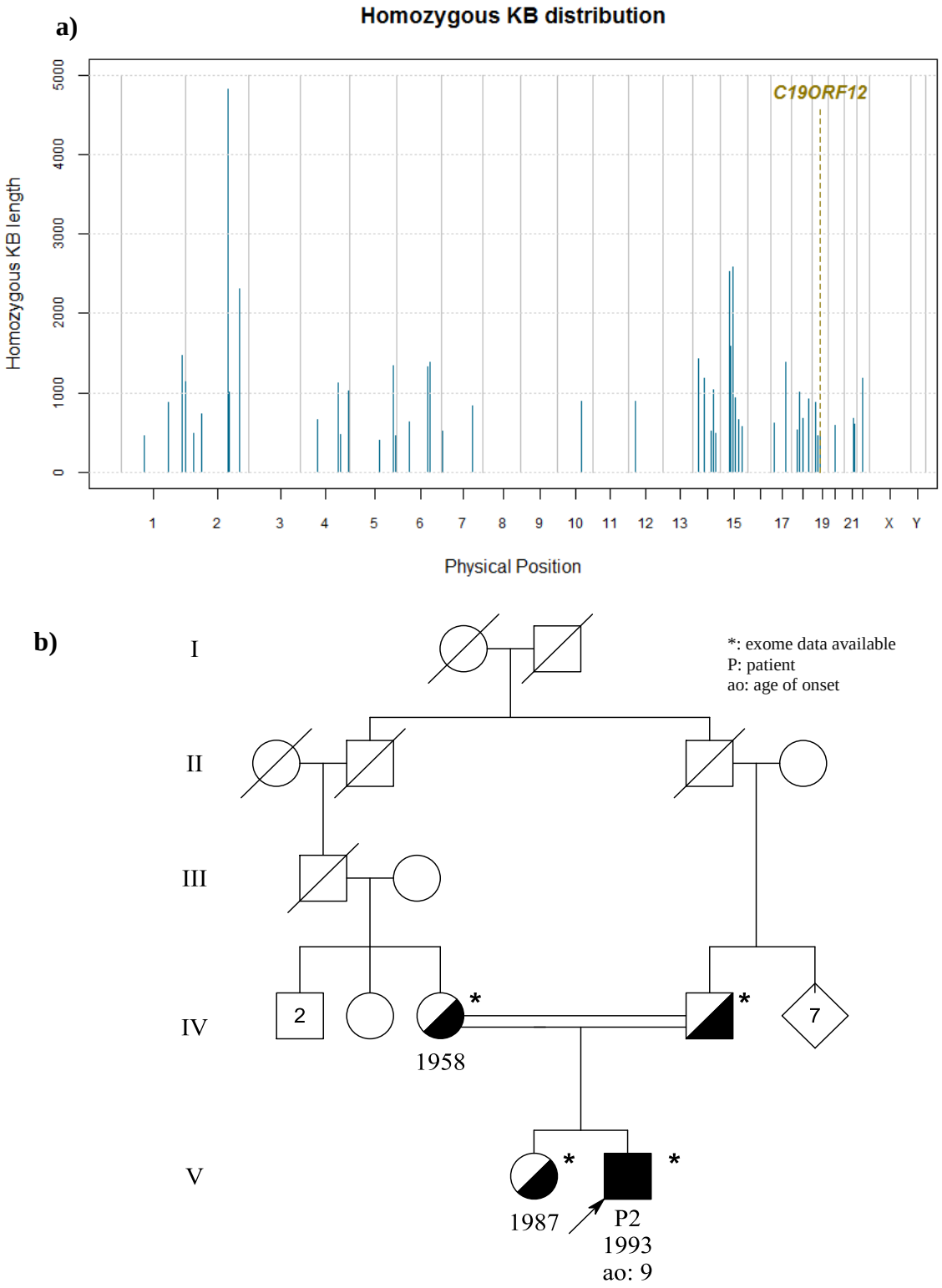
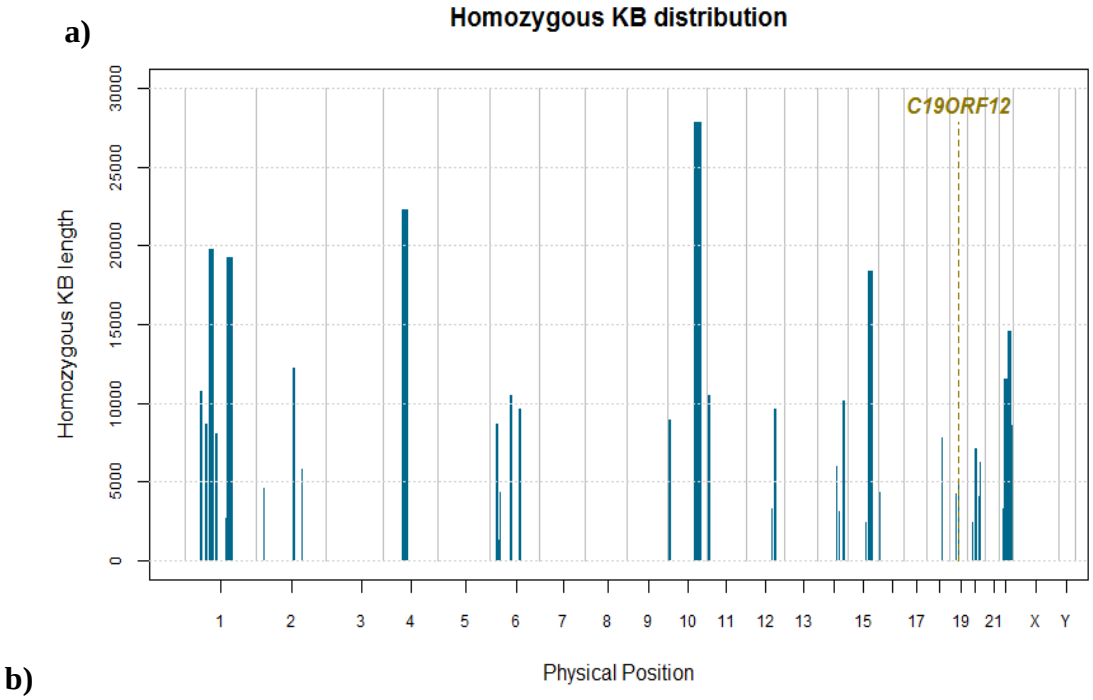


Figure 5.6. Homozygosity mapping plot of the patient (P2) (a) and segregation of the *C19ORF12* variation in Family 2 (b).



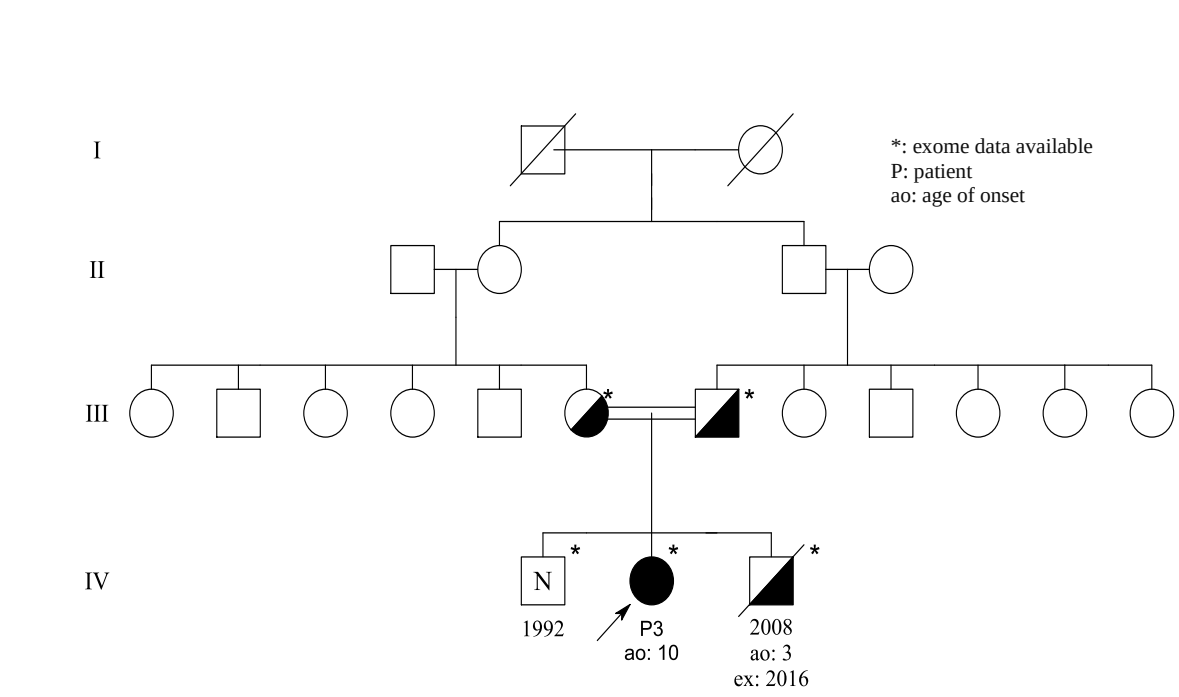


Figure 5.7. Homozygosity mapping plot of the patient (P3) (a), segregation of the *C19ORF12* variation in Family 3 (b)

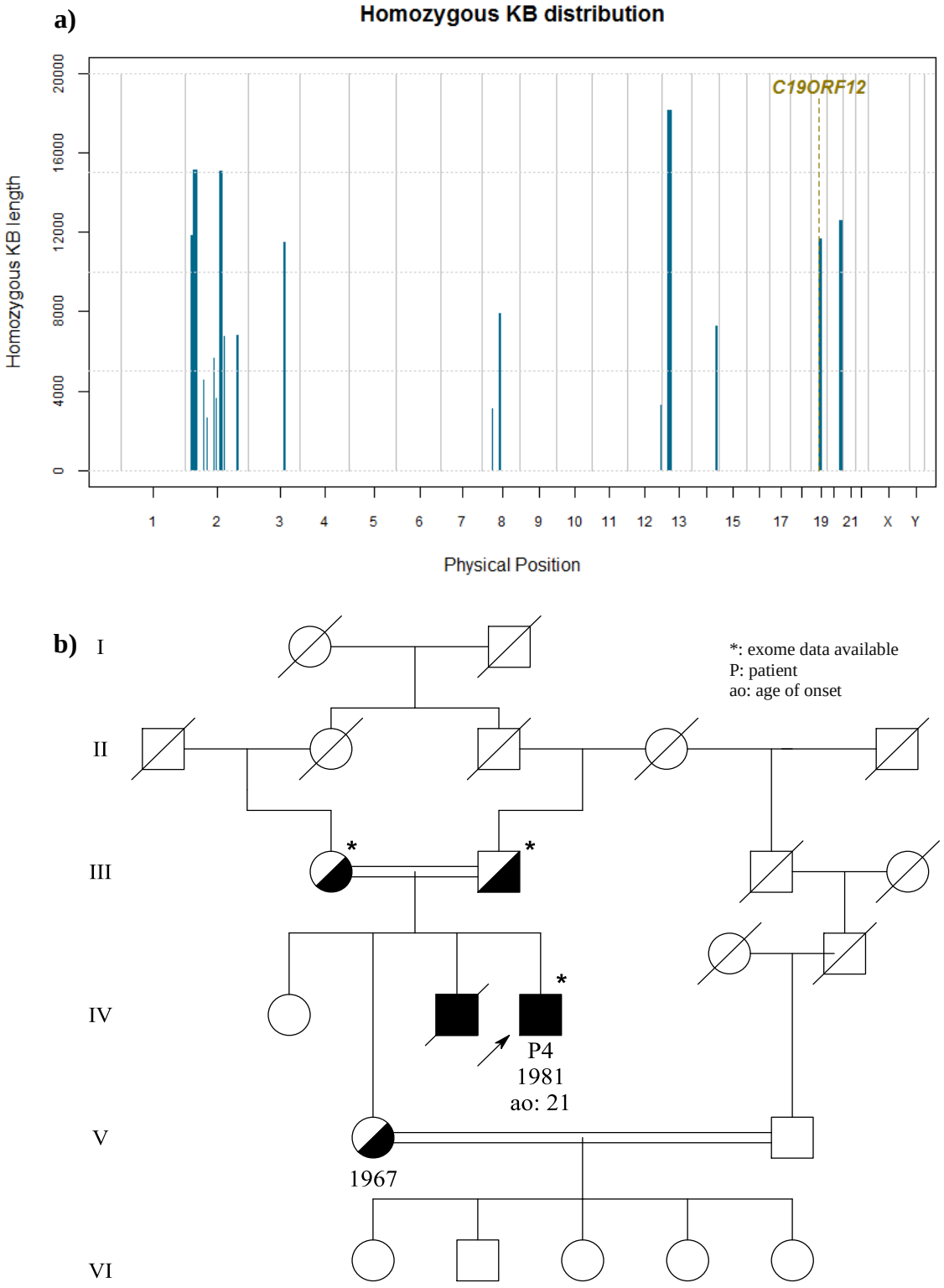


Figure 5.8. Homozygosity mapping plot of the patient (P4) (a) and segregation of the *C19ORF12* variation in Family 4 (b).

5.3.3. PANK2: Pantothenate Kinase 2 (AR)

5.3.3.1. Family 5. Several homozygous segments were detected via runs of homozygosity (Figure 5.5a). Within these regions, a homozygous missense mutation in the *PANK2* gene (chr20:3893169, G>A; Ala143Thr) was identified. The mutation is not present in dbSNP and ExAC, however several mutations in this gene were shown to cause NBIA type1 (#MIM 234200). The presence and segregation of the mutation will be confirmed with Sanger sequencing when the blood samples of the family members are available to us (Figure 5.5b).

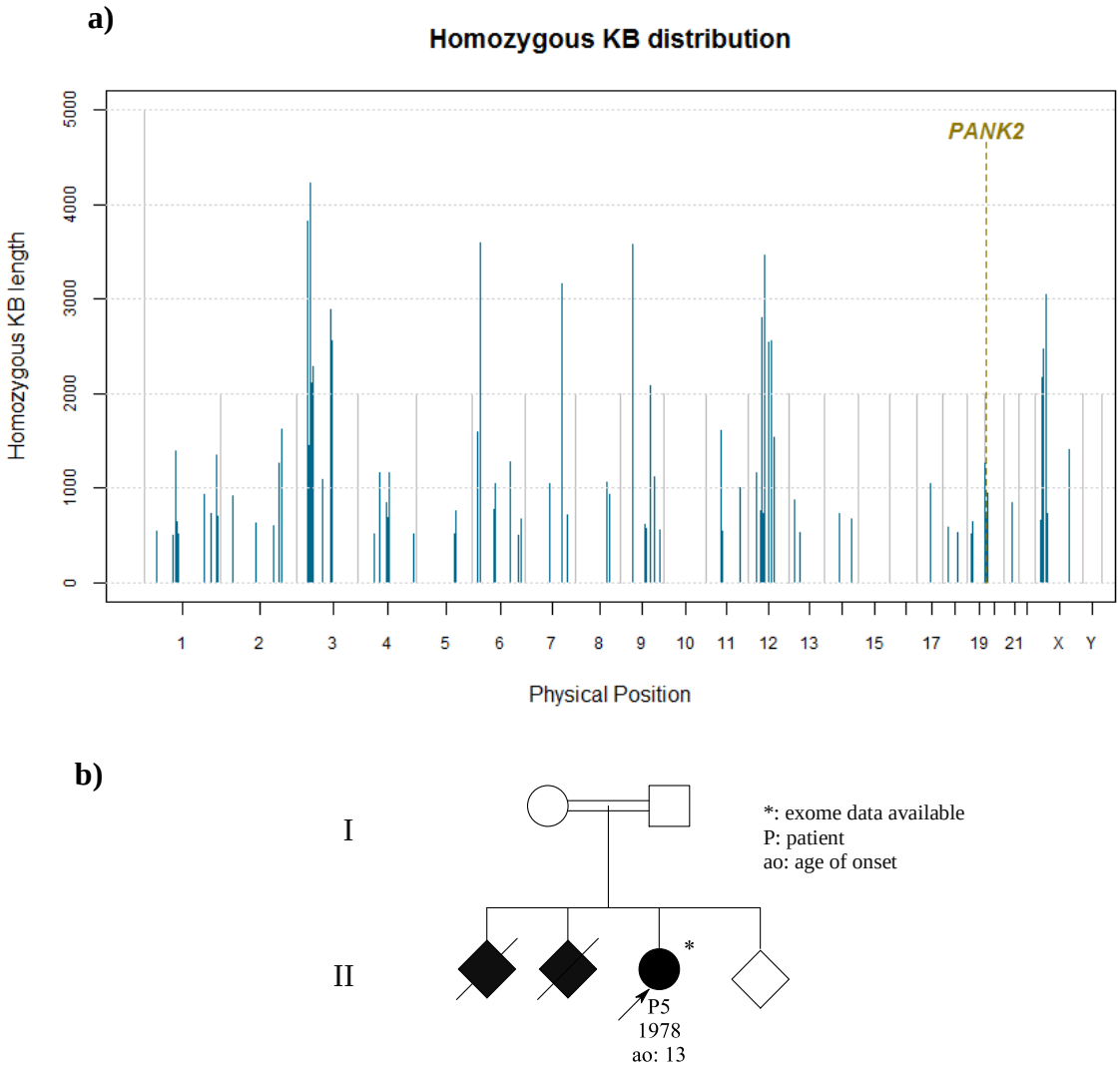


Figure 5.9. Homozygosity mapping of the patient (P5) (a) and the pedigree of Family 5 (b).

5.3.4. IGHMBP2: Immunoglobulin Mu Binding Protein 2 (AR)

5.3.4.1. Family 6. The index case with an initial diagnosis of motor neuron disease was subjected to WES together with her unaffected parents, a sister and a third-degree relative diagnosed with classical ALS. The missense mutation in the *IGHMBP2* gene (chr11:68678998, A>G; His213Arg) was found within one of the homozygous regions detected by homozygosity mapping (Figure 5.6a). The mutation was associated with spinal muscular atrophy with respiratory stress 1 (SMARD1) (MIM #604320) and submitted to dbSNP (rs137852666). The unaffected parents were heterozygous while the unaffected sister and the relative with ALS were wild type for the mutation. No mutation was found to explain the phenotype of the family member with classical ALS. Sanger sequencing confirmed the presence and segregation of the *IGHMBP2* mutation among family members (Figure 5.6b).

5.3.5. PLEKHG5: Pleckstrin Homology and RhoGEF Domain Containing G5 (AR)

5.3.5.1. Family 7. The index case with an initial diagnosis of ALS was referred to our laboratory together with her unaffected mother, an unaffected sister and a brother with an initial clinical diagnosis of SBMA. Numerous homozygous regions were detected by runs of homozygosity, and eight variations within the homozygous segments remained after filtration (Figure 5.7a). The stop-gain mutation in the *PLEKHG5* gene (chr1:6530920, C>T; Gln550Ter) was detected in homozygous state in all three affected siblings and in heterozygous state in the unaffected mother and sister. The mutation was not reported in any of the polymorphism databases and ExAC. Mutations in *PLEKHG5* gene are reported to be associated with distal spinal muscular atrophy (MIM #611067). Sanger sequencing confirmed the segregation of the mutation in our five samples (Figure 5.7b).

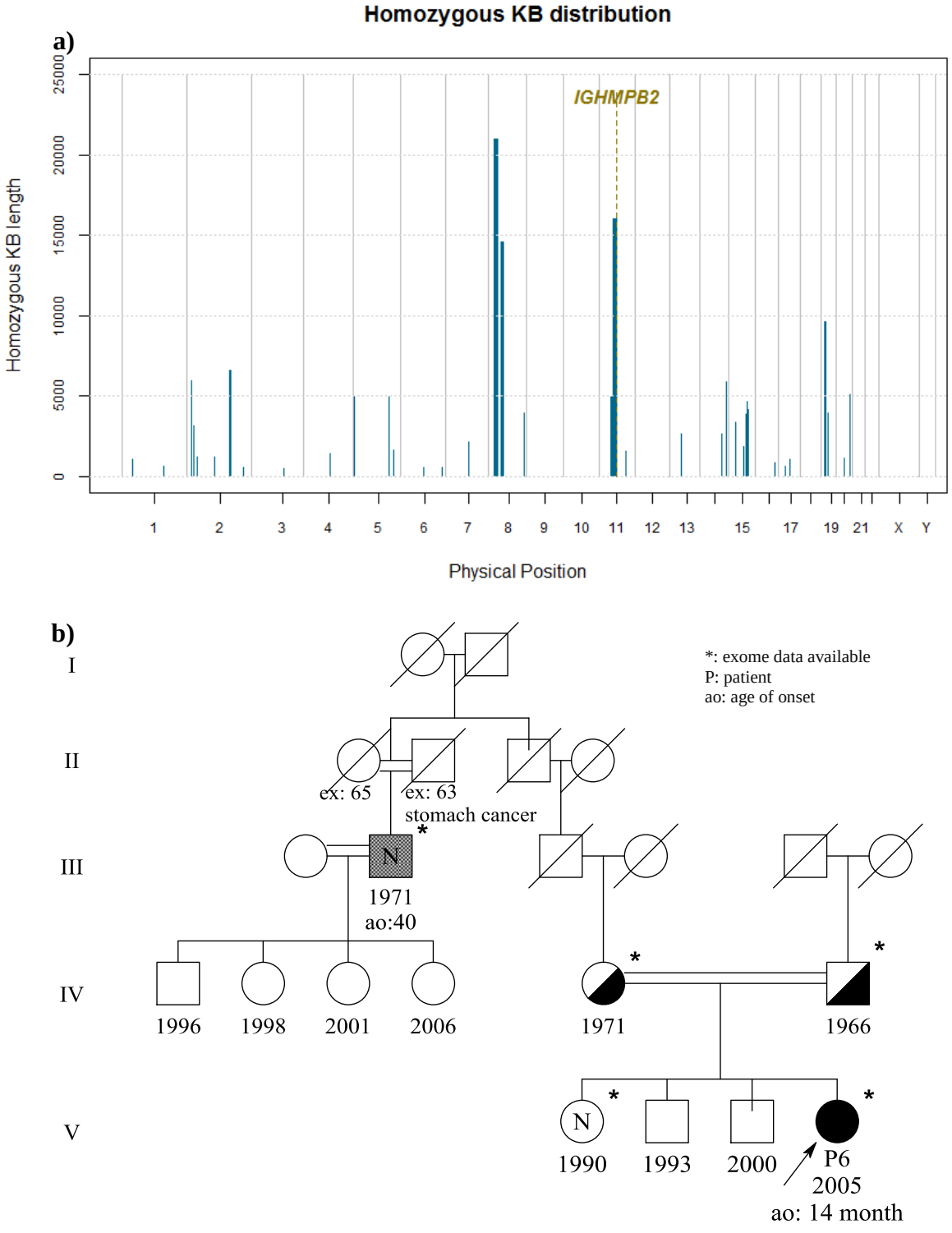


Figure 5.10. Homozygosity mapping plot of the patient (P6) (a) and the segregation of the *IGHMBP2* mutation in Family 6 (b).

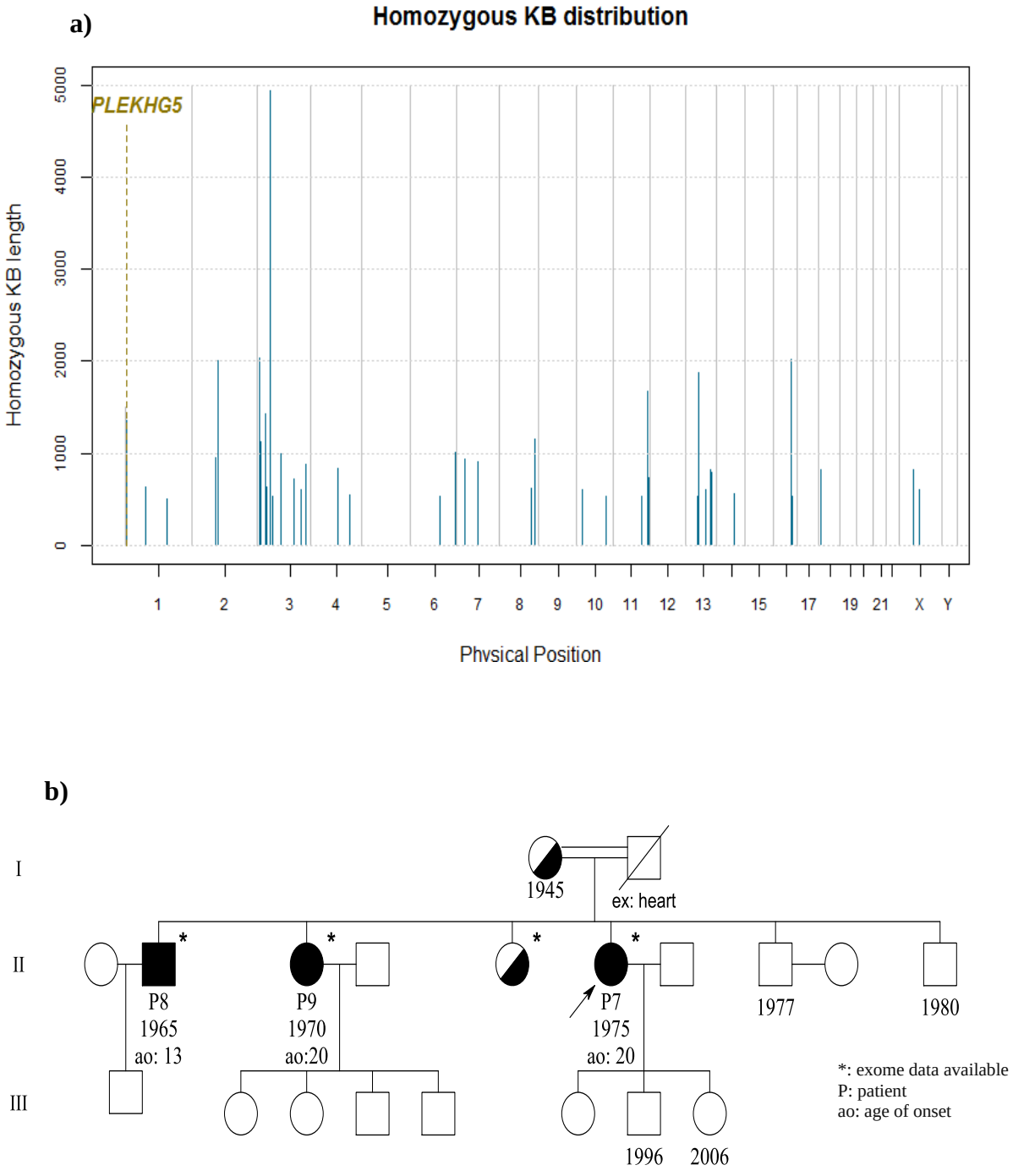


Figure 5.11. Homozygosity mapping plot of the patient (P7) (a) and the segregation of the *PLEKHG5* mutation in Family 7 (b).

5.3.6. SLC12A6: Solute Carrier Family 12 Member 6 (AR)

5.3.6.1. Family 8. In this family, evaluation of the exome data of four samples, including two affected siblings with an initial diagnosis of HSP, and their asymptomatic parents resulted in seven rare homozygous variations. Among these, a splice site mutation c.1073+1G>A (chr15:34546548, G>A) in the *SLC12A6* gene was detected in homozygous state in the affected cases and in heterozygous state in the parents (Figure 5.8a).

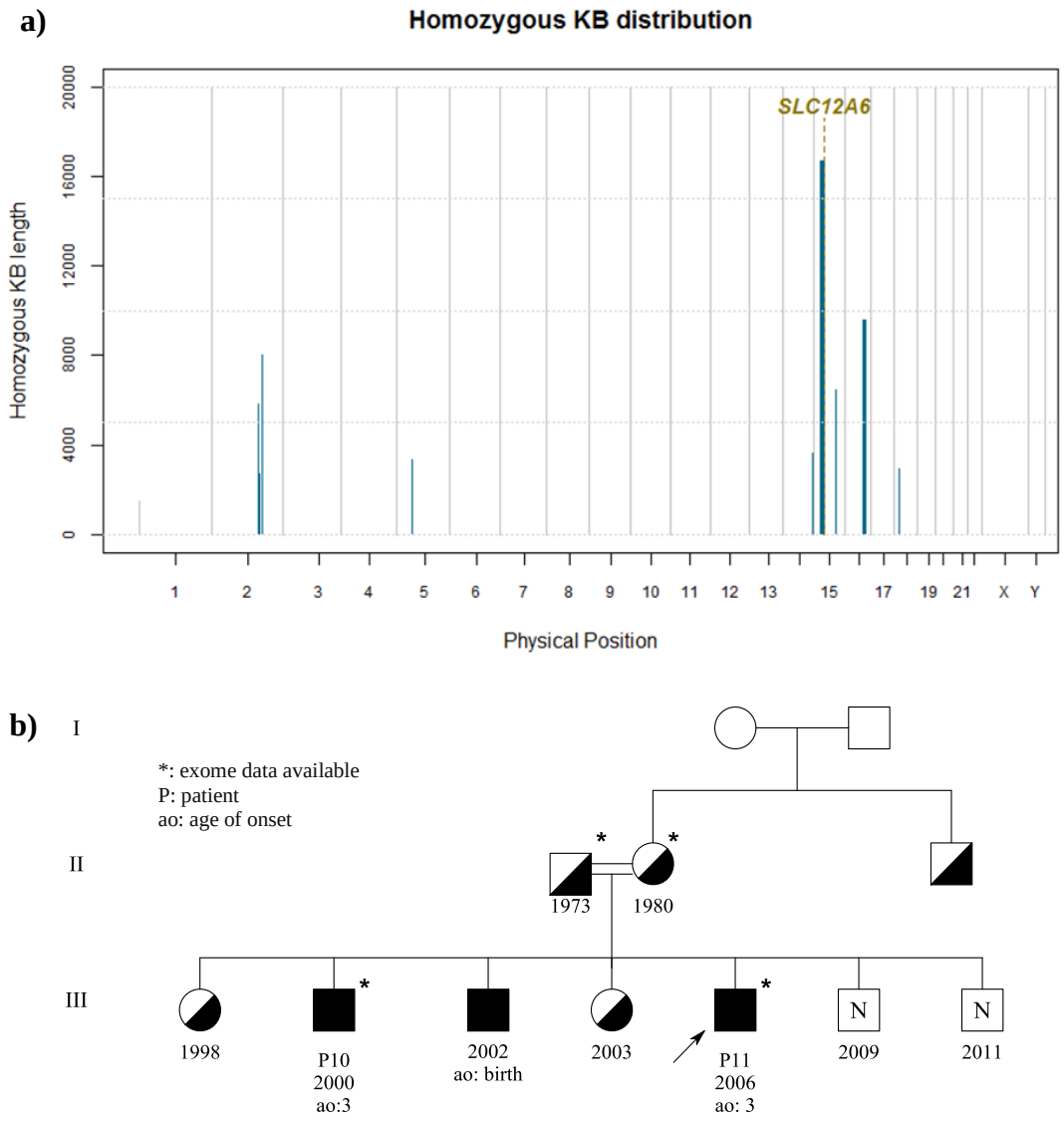


Figure 5.12. Homozygosity mapping plot of the patient (P10) (a) and the segregation of the *SLC12A6* mutation in Family 8 (b).

Mutations in the *SLC12A6* gene are known to be associated with Andermann syndrome (MIM #218000). The variation was not reported in dbSNP and ExAC in homozygous state. Sanger sequencing revealed that two unaffected siblings are wild-type and the other two unaffected siblings and the uncle are heterozygous for the mutation (Figure 5.8b).

5.3.7. ACADS: Acyl-CoA Dehydrogenase, C-2 to C-3 Short Chain (AR)

5.3.7.1. Family 9. Although runs of homozygosity resulted in homozygous segments in several chromosomes, only nine variations in chromosomes 12 and 17 remained after the filtration step. The missense mutation in the *ACADS* gene (chr12:121177120, A>G; Met370Val) was found in homozygous state in the index, in heterozygous state in the parents and wild-type in the unaffected sister (Figure 5.9). The *ACADS* gene has been associated with short-chain acyl-CoA dehydrogenase (SCAD) deficiency (MIM# 201470). The mutation found in the family was present in dbSNP (rs566325901), ExAC (0.0022) and Clinvar with an uncertain clinical significance.

5.3.8. SLC52A3: Solute Carrier Family 52 Member 3 (AR)

5.3.8.1. Family 10. Runs of homozygosity revealed a few homozygous regions in the index. However, homozygosity mapping failed to cover the homozygous missense mutation identified in the *SLC52A3* (chr20:744413, C>T; Arg268Trp) (Figure 5.10a). The mutation was present in dbSNP and ExAC (in heterozygous state) with a frequency of 4.95e-05. The *SLC52A3* gene was shown to cause BVVL1 when mutated (MIM# 211530). Validation and segregation analysis will be performed when the blood samples of the family members are available to us (Figure 5.10b).

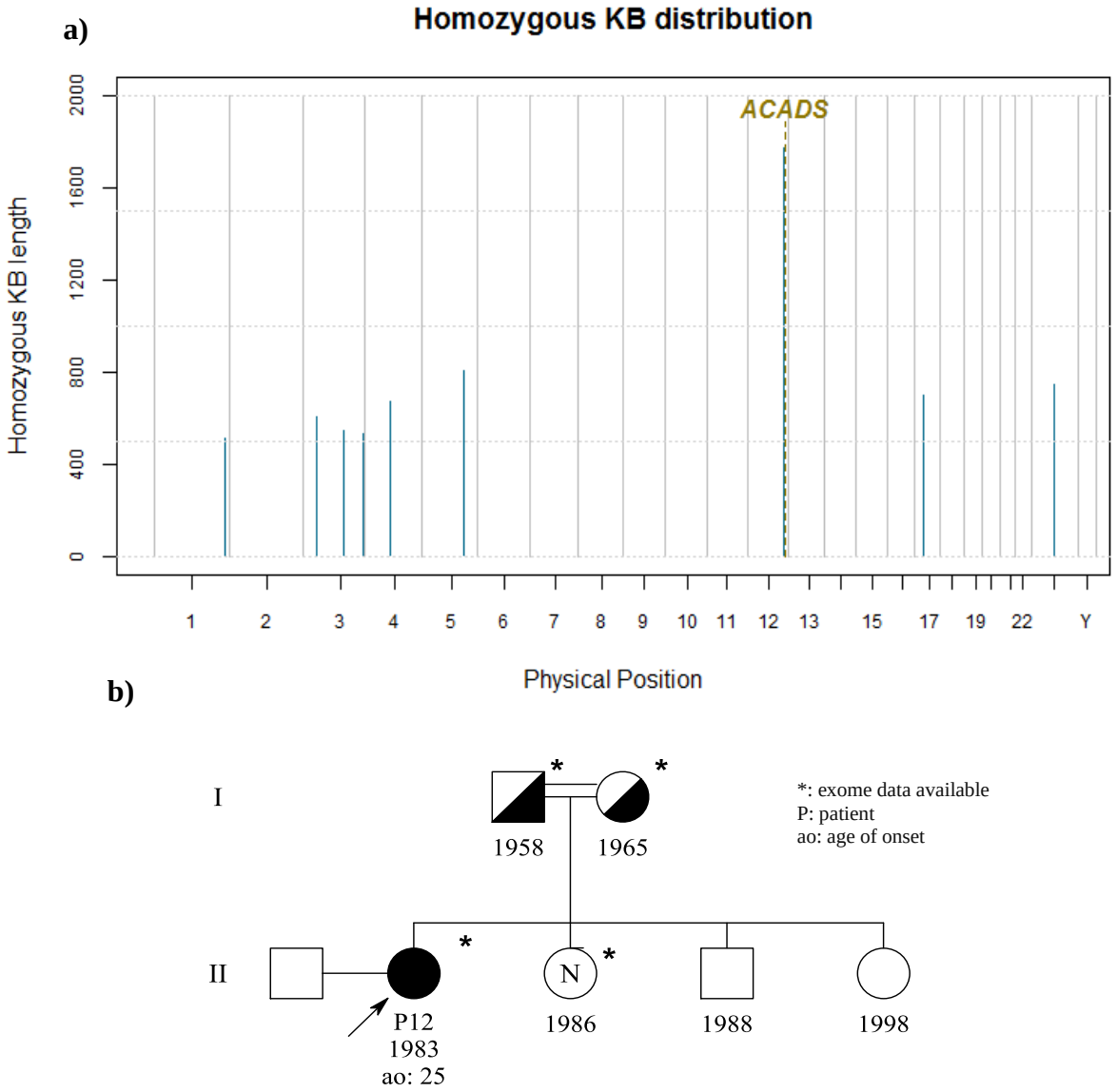


Figure 5.13. Homozygosity mapping plot of the patient (P12) (a) and segregation of the *ACADS* mutation in Family 9 (b).

5.3.9. ZFYVE26: Zinc Finger FYVE-type Containing 26 (AR)

5.3.9.1. Family 11. Among the rare homozygous mutations present in the index case, a nucleotide deletion in the *ZFYVE26* gene (chr14:68264904, delG) was detected, resulting in a frameshift mutation at position 692 and leading to a premature stop codon after 52 amino acids. The locus harboring the mutation was also found to be homozygous based on the runs of homozygosity (Figure 5.10a). Several mutations in the *ZFYVE26* gene were shown to

cause autosomal recessive spastic paraplegia 15 (#MIM 27077), but the frameshift mutation we describe in this family was not reported before. Validation and segregation analysis is pending.

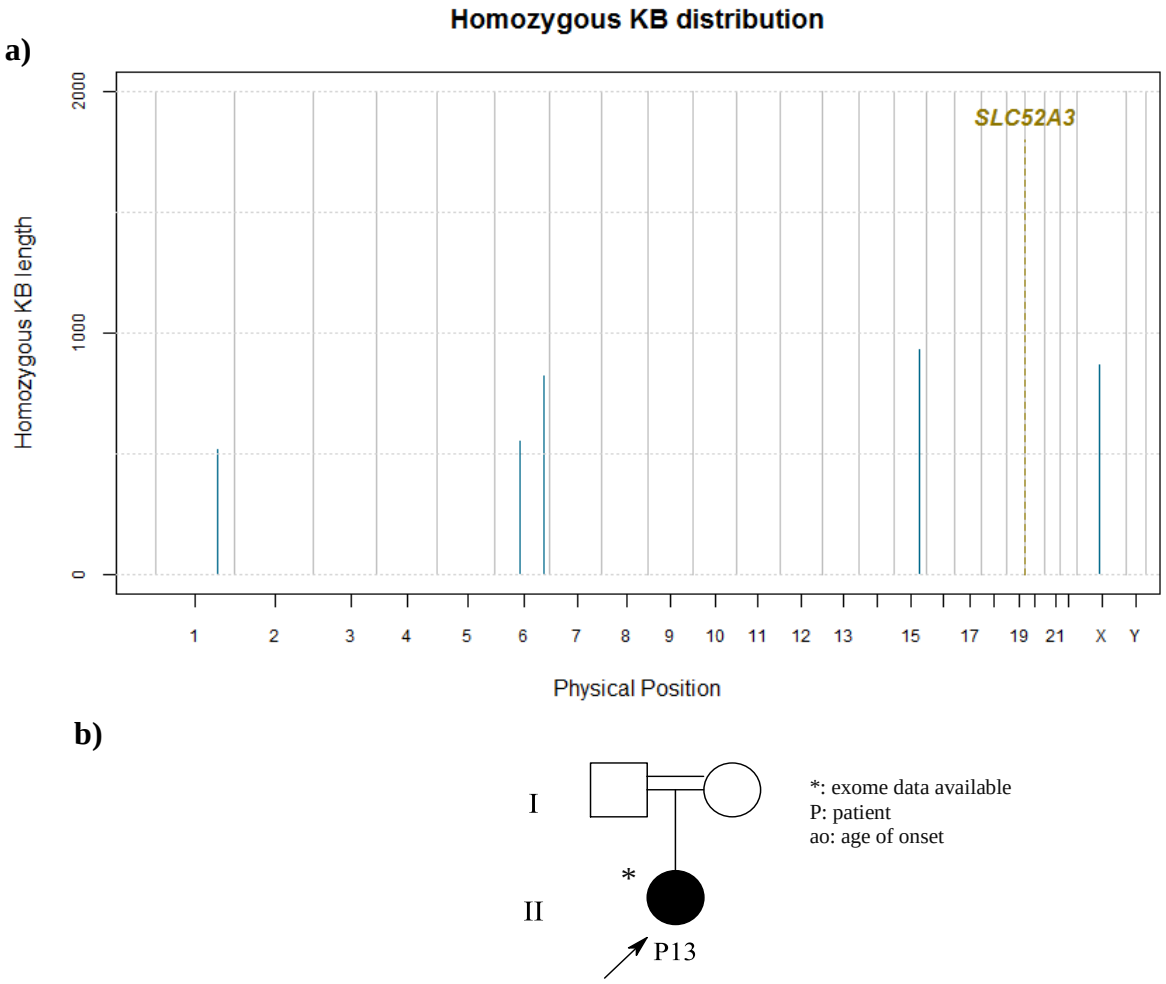


Figure 5.14. Homozygosity mapping plot of the patient (P13) (a) and the pedigree of Family 10 (b).

5.3.10. SPG11: Spatacsin Vesicle Trafficking Associated (AR)

5.3.10.1. Family 12. A homozygous stop-gain mutation in the *SPG11* gene (chr15:44943713, C>T; Gln478Ter) was present in the index patient, falling into a well identified region in runs of homozygosity (Figure 5.12). The mutation was not reported in dbSNP or ExAC before. The *SPG11* gene was earlier associated with autosomal recessive

juvenile ALS (ALS-5) and spastic paraplegia 11 (#MIM 602099, #MIM 604360). The presence and segregation of the variant will be confirmed with Sanger sequencing.

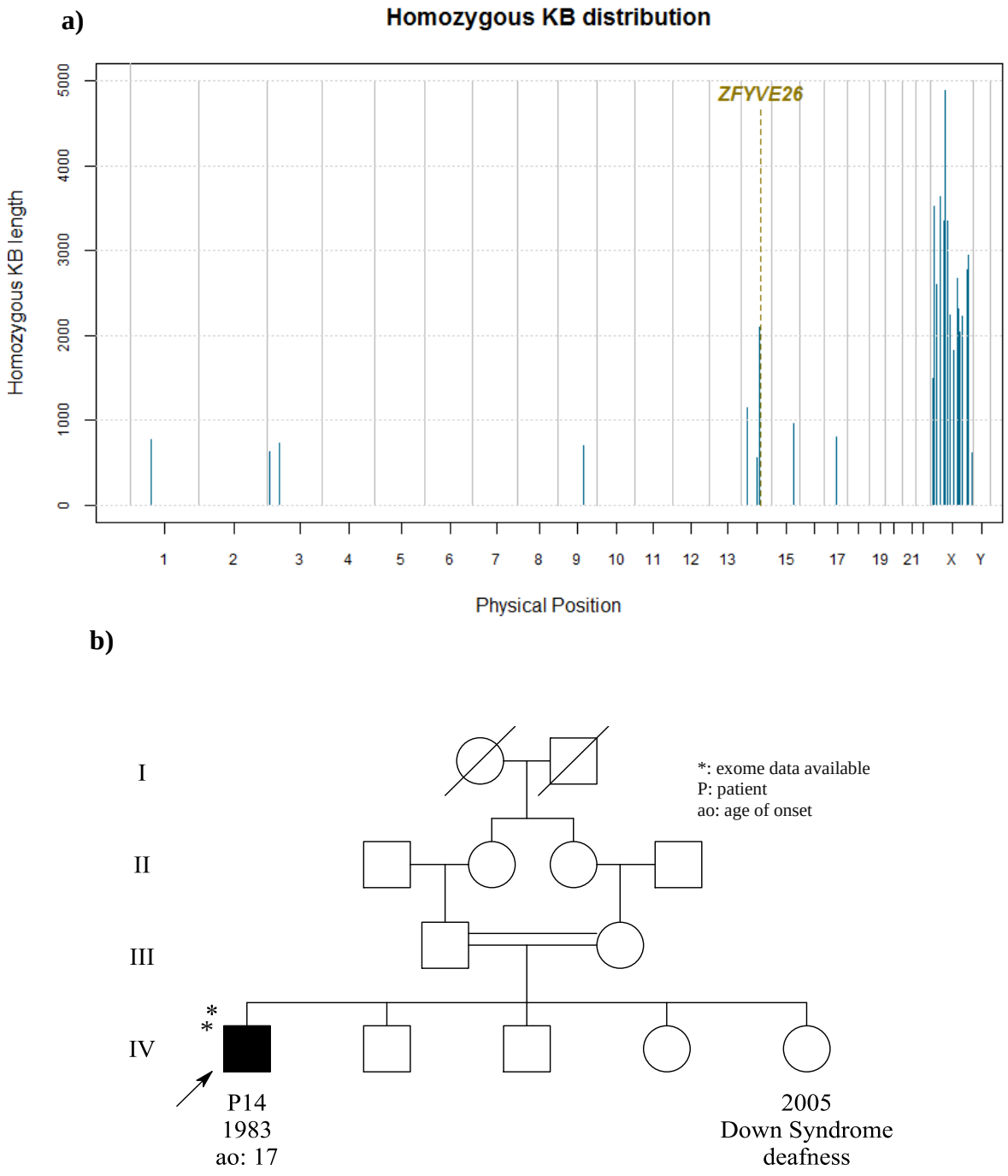


Figure 5.15. Homozygosity mapping plot of the patient (P14) (a) and the pedigree of Family 11 (b).

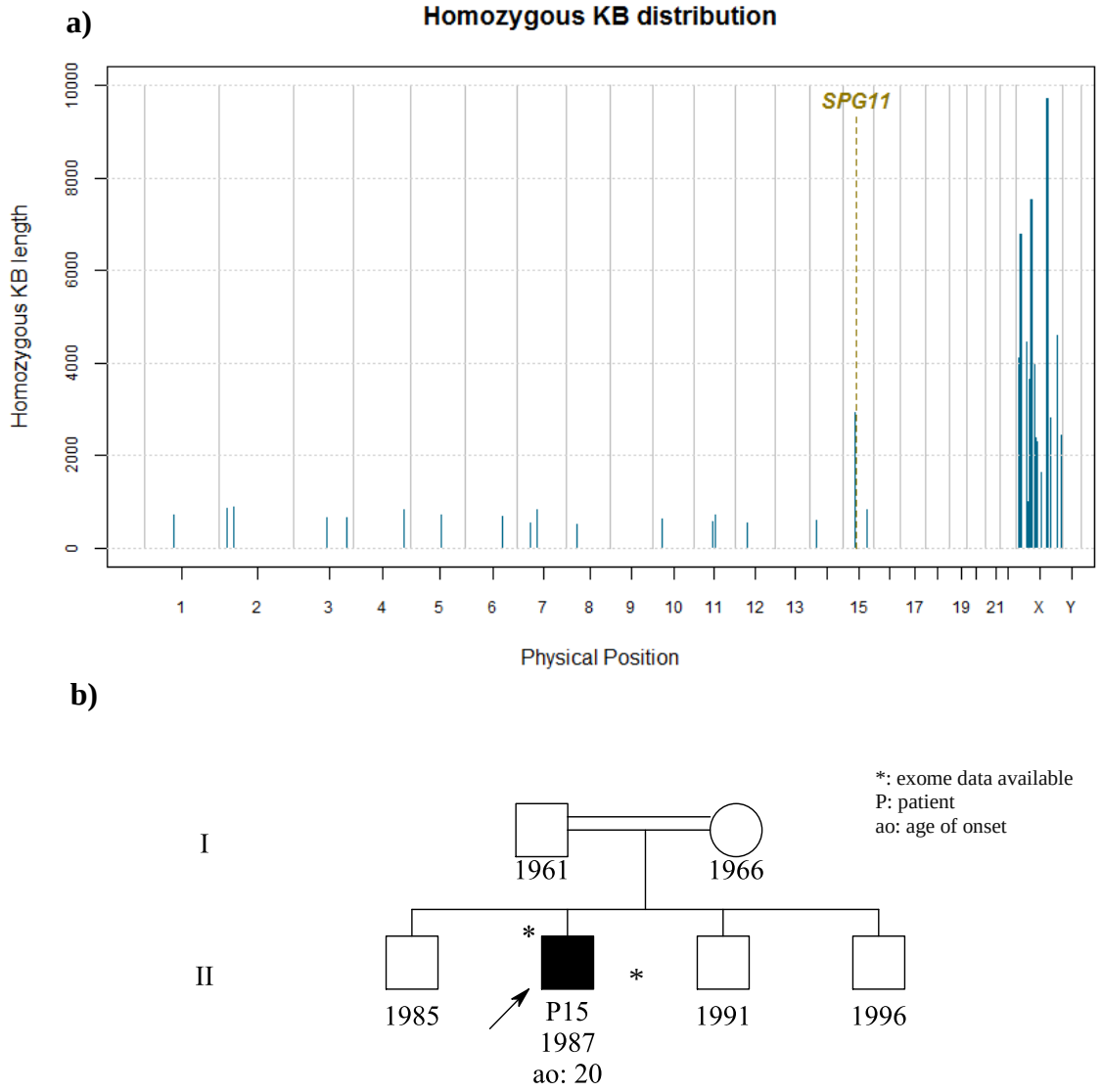


Figure 5.16. Homozygosity mapping plot of the patient (P15) (a) and the pedigree of Family 12 (b).

5.3.11. SIGMAR1: Sigma Non-opioid Intracellular Receptor (AR)

5.3.11.1. Family 13. Numerous shared homozygous regions were revealed throughout the chromosomes as a result of homozygosity mapping. The missense mutation in the *SIGMAR1* gene (chr9:34635853, G>A; Glu119Lys) was found within one of the homozygous regions detected (ALS-16) (Figure 5.13). The mutation was novel; it was not present in any

polymorphism database or Clinvar. It was also absent in our in-house control samples. Validation and segregation analysis is pending.

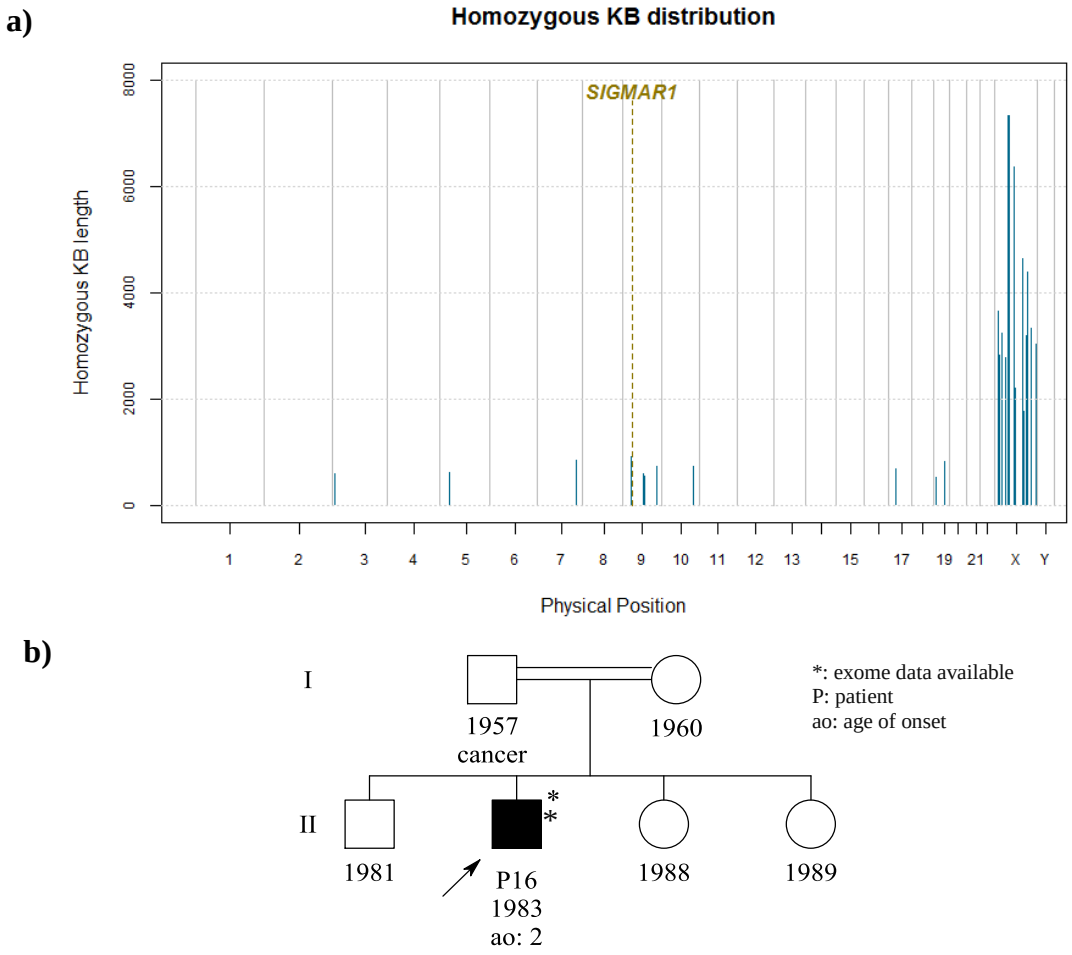


Figure 5.17. Homozygosity mapping plot of the patient (P16) (a) and the pedigree of Family 13(b).

5.3.12. TRPV4: Transient Receptor Potential Cation Channel Subfamily V Member 4 (AD)

5.3.12.1. Family 14. A total of 210 rare variations, shared between two siblings with young-onset motor neuron disease, remained after computational filtration to be evaluated. Deep phenotyping revealed a similar phenotype, sloping shoulders and scapular winging, in the

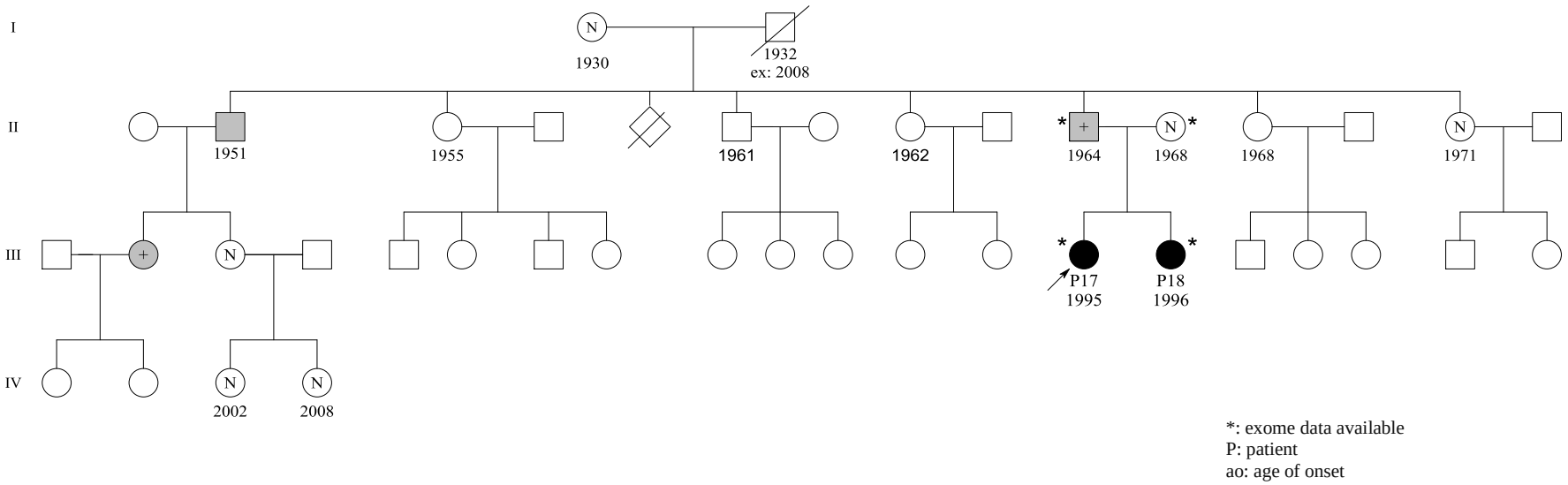
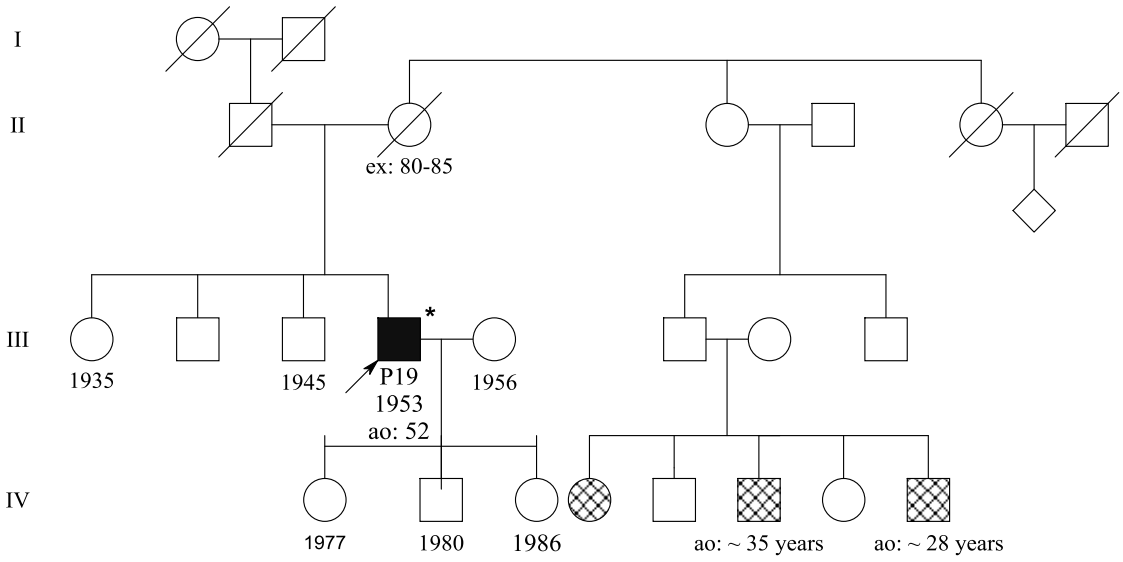


Figure 5.18. The segregation of the *TRPV4* variation in Family 14. The sisters presented with two different phenotypes (SPSMA and CMT2C).

asymptomatic father and several members of the family (subclinical penetrance). Among the mutations, the missense substitution in the *TRPV4* gene (chr12:110236628, C>T; Arg315Trp) was detected. The mutation is present in dbSNP (rs267607143) and has been reported to be associated with autosomal dominant scapuloperoneal spinal muscular atrophy (SPSMA) and hereditary motor and sensory neuropathy type 2 (MIM# 181405, #MIM 606071). The young sisters presented with two different phenotypes (SPSMA and CMT2C). Sanger sequencing confirmed the presence of the mutation in the father, paternal uncle and a cousin of the patients, while their mother, grandmother, aunt and the other cousins were found to carry the wild-type sequence (Figure 5.14).

5.3.13. ANG: Angiogenin (AD)

5.3.13.1. Family 15. The index case was referred with an initial diagnosis of motor neuron disease. Bioinformatic analysis resulted in a total of 351 rare variants. Among these, the heterozygous missense mutation in the *ANG* gene (chr14:21161931, A>G; Ile70Val) was detected. Several mutations in *ANG* have been associated with ALS in the literature (#MIM 611895) (ALS-9). The above mutation was not present in our in-house control samples, but in dbSNP (rs121909541) and ExAC with a frequency of 0.0006095. (Figure 5.19).



*: exome data available
P: patient
ao: age of onset

Figure 5.19. Pedigree of Family 15.

5.3.14. MPZ: Myelin Protein Zero (AD)

5.3.14.1. Family 16. Four samples, including the index case, with a clinical diagnosis of CMT, his affected twin sons and unaffected wife were subjected to WES. Considering an autosomal dominant inheritance pattern, the heterozygous variations common in the index patient and his sons were selected and polymorphisms were filtered out. Among the remaining 111 rare coding variations, the missense mutation in the *MPZ* gene (chr1:161276653, G>A; Arg98His) was found to be heterozygous in the affected individuals, and wild-type in the unaffected mother of the twins (Figure 5.16). The mutation was not present in our in-house control samples, but is reported in dbSNP (rs121913589) and associated with autosomal dominant CMT type 1B (MIM# 118200).

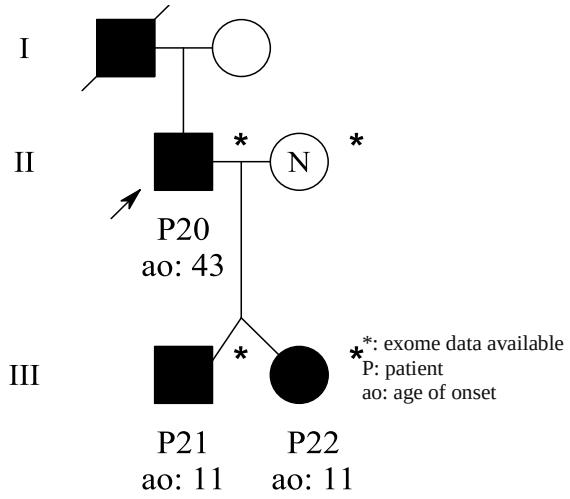


Figure 5.20. Pedigree of the Family 16.

5.3.15. *VCP: Valosin Containing Protein* (AD)

5.3.15.1. Family 17. Two sisters were referred to our laboratory with ALS. An autosomal dominant inheritance pattern was observed: the father, three older sisters and one of their nephews presented with a similar phenotype. With the selection of heterozygous mutations

shared by the two affected individuals and through filtering out the polymorphisms, 201 mutations remained. A novel missense mutation in the *VCP* gene (chr9:35065252, G>C; Arg191Pro) was suspected as the candidate. The *VCP* gene has been associated with autosomal dominant ALS with or without FTD (ALS-14, MIM# 613954). The mutation was not present in our in-house control samples, dbSNP and ExAC database. Sanger sequencing confirmed the presence and segregation of the mutation in the family; the sister with cognitive dysfunction had the mutation, whereas three unaffected siblings and a nephew were found to be wild-type for the mutation (Figure 5.17).

5.3.16. *ERBB4: Erb-B2 Receptor Tyrosine Kinase 4* (AD)

5.3.16.1. Family 18. Four siblings were reported to suffer from ALS. The initial analysis aimed to find shared heterozygous mutations among these affected individuals. This analysis failed to detect any causative variations. Individual-based analysis in each patient revealed a heterozygous missense mutation in the *ERBB4* gene (chr2:212251725, C>T; Arg1112Cys) (#MIM 615515, ALS-19) in P25, P26 and P27. The father and one of the affected siblings (P28) were wild type for the mutation. The mutation was not present in dbSNP, and reported in ExAC with a frequency of 4.942e-05. Deep phenotyping revealed that the clinical symptoms of individual P28 (shaded in grey) resembled a CMT phenotype, rather than ALS, which was later explained by a missense mutation in the *LRSAM1* gene (chr9:130230068, G>A; Cys193Tyr) (#MIM 614436). The mutation has a frequency of 4.782e-05 in ExAC database. Sanger sequencing confirmed the presence of the variations among all family members. Furthermore, the *LRSAM1* mutation was also found to be coexisting in one of the siblings with ALS, P26 (Figure 5.22).

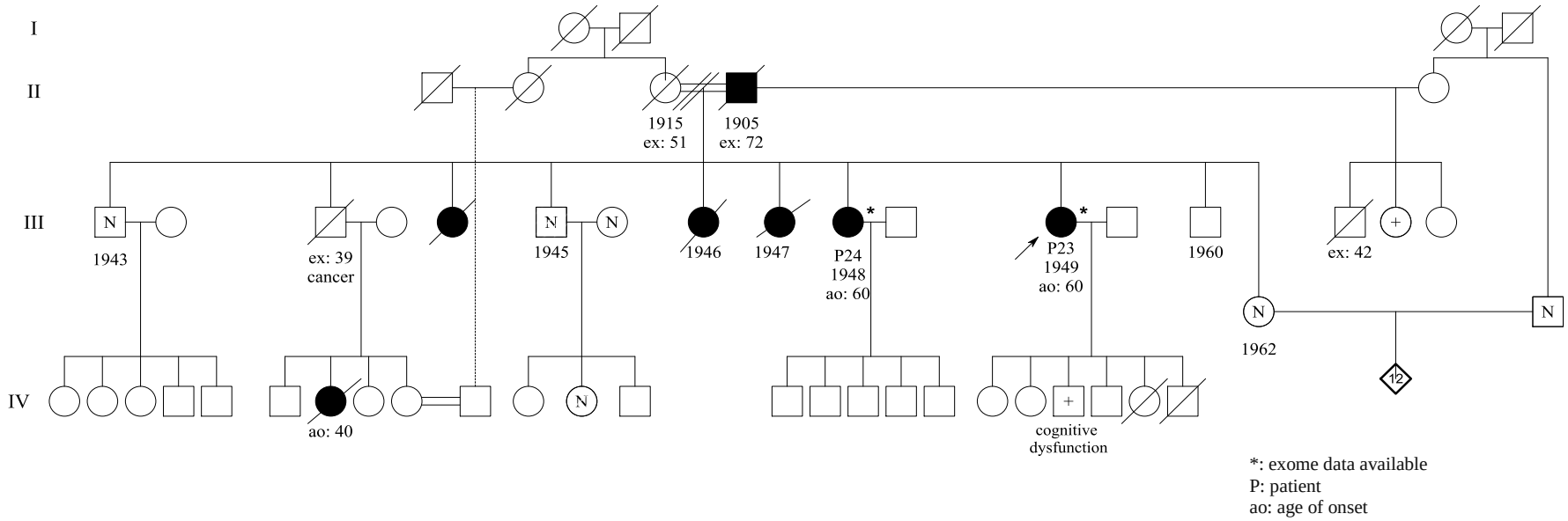


Figure 5.21. The segregation of the *VCP* mutation in Family 17.

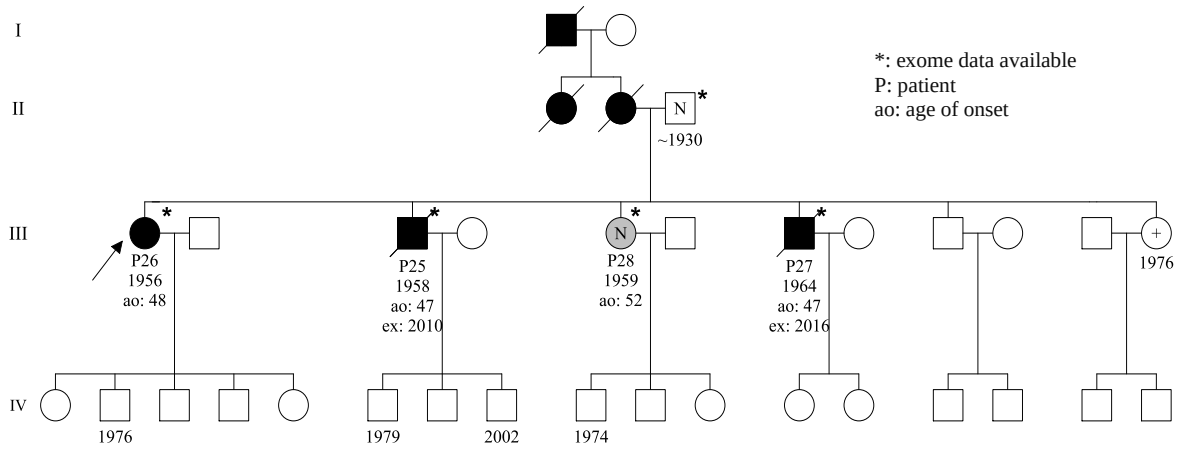


Figure 5.22. The segregation of the *ERBB4* mutation in Family 18.

5.3.17. *SQSTM1: Sequestosome 1* (AD)

5.3.17.1. Family 19. The index case was referred to our laboratory with an initial diagnosis of motor neuron disease. A total of 1247 heterozygous mutations remained after filtration. When screening for ALS genes, a missense mutation (chr5:179250930, A>G; Asn125Ser) was detected in the *SQSTM1* gene (#MIM 616437, #MIM 167250). The mutation was not present in our in-house control samples, but in ExAC database with a frequency of 1.658e-05. Validation and segregation analysis is pending (Figure 5.19).

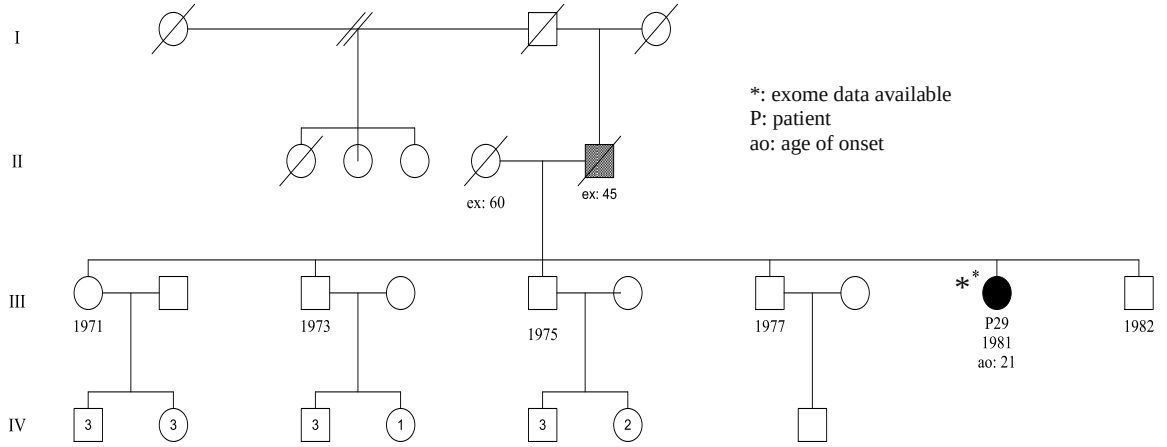


Figure 5.23. Pedigree of Family 19.

5.3.18. *UBQLN2: Ubiquilin 2* (XLD)

5.3.18.1. Family 20. The index patient was referred to us with a MMND phenotype. Conventional PCR-based Sanger sequencing revealed a mutation in the *UBQLN2* gene (chrX:56591482, G>A; Met391Ile) (ALS-15, #MIM 300857) (Figure 5.20). On the search for another variation to be the cause for the phenotype described as MMND, we performed exome analysis. No additional variation was detected and the presence of the above *UBQLN2* mutation was confirmed, which was not reported in ExAC and Clinvar databases.

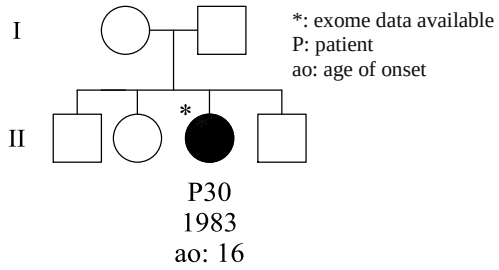


Figure 5.24. Pedigree of Family 20.

Table 5.4. Remaining variations after each filtration step in families without a confirmed causative mutation.

	# of total variants	type of variation	pedigree info	Minor allele frequency		# of samples
				1000G+ESP6500	ExAC	
Family 21	21973	10663	614	41	29	3
Family 22	153012	10406	423	13	4	3
Family 23	149003	10004	525	34	27	3
Family 24	155057	9958	469	43	30	4
Family 25	245138	11047	380	10	5	3
Family 26	696396	11198	629	31	16	4
Family 27	294649	9911	286	28	23	4
Family 28	90216	10720	2186	430	299	4
Family 29	145872	10385	17	3	3	6
Family 30	10823	10823	4503	224	52	2
Family 31	277003	10855	4394	195	23	1
Family 32	490021	11398	1583	356	238	6
Family 33	147321	10243	10243	1053	876	1
Family 34	150950	10079	6203	987	635	1
Family 35	119569	10011	3124	480	203	2
Family 36	299245	10867	327	53	39	4
Family 37	132779	9964	6037	872	590	1
Family 38	146163	10205	3119	489	218	2
Family 39	141234	10082	6087	906	520	1
Family 40	132277	9922	9922	1084	757	1
Family 41	134921	9876	9876	1849	1345	1
Family 42	146471	10307	10307	1181	758	1
Family 43	147542	10110	6154	493	187	2
Family 44	126429	10619	2768	325	117	3
Family 45	109844	9733	3520	486	239	2
Family 46	340686	10803	4649	234	44	1
Family 47	282736	10851	4370	224	39	1
Family 48	269117	11017	4331	202	28	`1
Family 49	256465	10776	4230	184	31	1
Family 50	174487	10619	4394	201	33	1
Family 51	436103	11323	32	2	2	7
Family 52	278877	10743	567	37	20	3
Family 53	256918	11121	4419	210	34	1
Family 54	210582	8463	241	43	26	6
Family 55	292849	10750	4489	207	49	1
Family 56	145387	10166	542	147	81	4
Family 57	259226	11098	4427	234	9	1

6. DISCUSSION

In this thesis, whole exome sequencing analysis of 57 Turkish families which included 81 MND patients and their 66 unaffected family members was performed. Pathogenic variants in 20 families were identified so far and 37 remained genetically undefined. In 13 out of 35 AR families (37%), the causative homozygous variants were successfully identified. In seven cases out of 22 dominantly inherited families (21 AD and one XLD) the pathogenic mutations explaining the phenotype were described (32%). Our overall success rate is 35%, which is in agreement with the previous clinical exome sequencing studies (Figure 6.1) (Trujillano *et al.*, 2017).

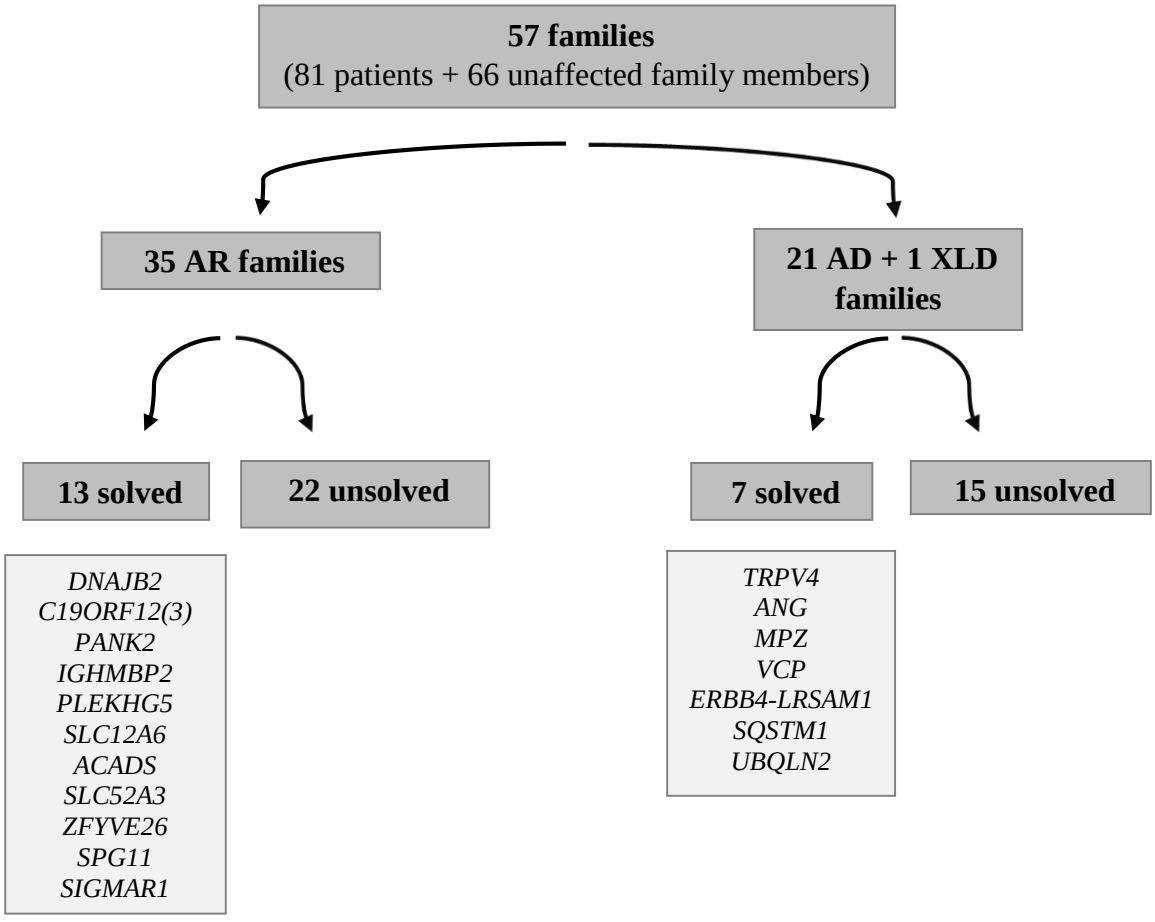


Figure 6.1. An overview of the Turkish MND cohort

We identified 21 distinct mutations in our patients with the initial diagnosis of either ALS or other MNDs. In seven families mutations in known ALS genes; *VCP, ANG, SIGMAR1, ERBB4, SPG11, SQSTM1* and *UBQLN2* were identified. Further, mutations defined in *DNAJB2, TRPV4, SLC52A3, IGHMBP2, PLEKHG5, MPZ, SLC12A6, LRSAM1* and *ZFYVE26* implicated a non-ALS MND phenotype in these patients. The final diagnoses of these non-ALS MND patients are a group of disorders, which can be phenotypically overlapping, including distal and scapuloperoneal SMA, BVVL, HSP, SMARD1, Andermann syndrome and CMT, emphasizing the role of whole exome sequencing in differential diagnosis. Mutations in the two NBIA genes, *C19ORF12* and *PANK2* were described in patients with a phenotype mimicking ALS and HSP, suggesting an overlap between NBIA, HPS and ALS, expanding the phenotypic spectrum of these diseases. Finally, a mutation with a so far uncertain significance in the *ACADS* gene was identified, since this variation was not sufficient to explain the MND phenotype in the index case.

6.1. Mutations in Known ALS Genes

A heterozygous missense mutation in the *VCP* gene was identified in two sisters with ALS accompanied by cognitive dysfunction. Mutations in the *VCP* gene had previously been shown to cause FTD and inclusion body myopathy with Paget's disease (IBMPFD) (Watts *et al.,* 2004). Soon after, with the advent of exome sequencing, additional *VCP* gene mutations were described in adult-onset ALS with or without dementia (ALS-14). The *VCP* gene encodes for valosin-containing protein that is a ubiquitously expressed multifunctional protein implicated in the maturation of ubiquitin-containing autophagosomes. It has been shown that mutant VCP toxicity results in ubiquitin-positive TDP-43 inclusions, the key pathological hallmark of ALS (Johnson *et al.,* 2010).

The heterozygous *ANG* Ile46Val mutation, which was identified in one of our patients, was previously shown to be the cause of adult onset ALS (ALS-9) (Greenway et al., 2006). The *ANG* gene encodes for angiogenin, a 147-residue protein belonging to pancreatic ribonuclease superfamily. Functional studies showed that ANG-mediated rRNA transcription is required for angiogenesis, induced by vascular endothelial cell growth factor

(VEGF) which has also been implicated in ALS. Since mutant ANG lacks angiogenic activity, it was suggested that *ANG* is the first gene in which typical loss-of-function mutations were reported in ALS (Wu *et al.,* 2007).

A mutation in the Glu102 position of the *SIGMAR1* (*ALS-16*) was previously shown to cause slow progressive ALS. The *SIGMAR1* gene has four exons and two isoforms, one long isoform including exon-3 and one short isoform excluding exon-3. We identified the homozygous p.Glu119Lys mutation residing in the fourth exon based on the longer isoform. This variation was located in the neighborhood of a previously identified mutation in ALS. The encoded protein sigma-receptor 1 is a transmembrane receptor for ion channels and is involved in lipid transport and neuronal cell differentiation. Based on cell culture studies, aberrant distribution of the protein was reported in neuron-like cell lines, indicating the role of SIGMAR1 in neural function and neurodegenerative diseases (Al-Saif *et al.*, 2011).

Two heterozygous *Erb-B2 Receptor Tyrosine Kinase 4 (ERBB4)* gene mutations, p.Met831Leu and p.Met1059Val, had been previously described in adult-onset ALS (ALS-19) in Japanese and Canadian families. As a transmembrane protein, ErbB4 phosphorylates its C-terminal domain upon neuregulin stimulation. It was shown that ErbB4 mutations specifically within the tyrosine kinase and C-terminal domains reduce autophosphorylation, which in turn disrupts the neuregulin-ErbB4 pathway involved in the pathogenesis of ALS (Takahashi *et al.,* 2013). The heterozygous missense p.Arg1112Cys mutation explaining the ALS phenotype in our patients also resides in the C terminal domain, and to our knowledge it is only the third mutation identified in the *ERBB4* gene/protein (Figure 6.2).

Homozygous mutations in the *SPG11,* encoding for the spatacsin gene, were described as the predominant cause of ARHSP with thin corpus callosum (TCC) (Stevanin *et al*., 2007) and soon after, were reported to give rise to autosomal recessive juvenile ALS (ARJALS) (Orlacchio *et al.,* 2010). The spatacsin dysfunction leads to axonal pathology and vesicle trafficking defects. The axonal involvement in both ARJALS and ARHSP suggests the presence of a common pathway contributing to these diseases (Branguli *et al.,* 2014). In the framework of this study, we found a homozygous mutation in the *SPG11* gene, causing

MND. Four additional *SPG11* mutations were previously reported in Turkish families with MND in our laboratory, highlighting the considerable prevalence of *SPG11* mutations in Turkish MND patients (Iskender *et al.*, 2015).

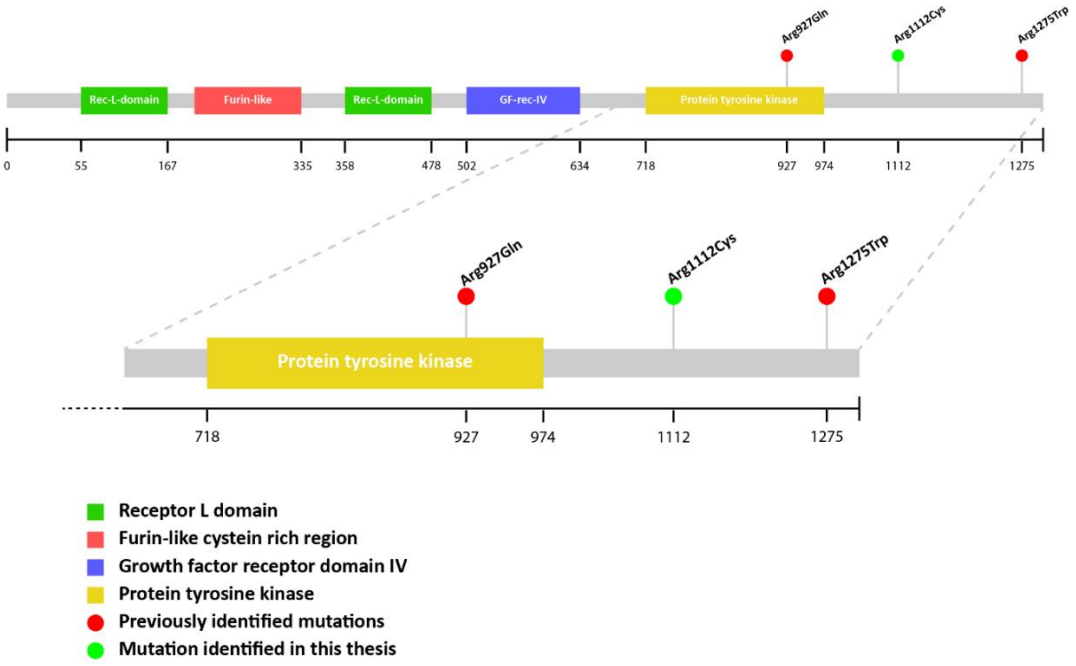


Figure 6.2. Mutations described in the *ERBB4* gene.

A heterozygous mutation in the *SQSTM1* gene was identified in an individual with ALS whose father had a skeletal disease. Mutations in *SQSTM1* were previously shown to cause Paget disease of bone (PDB) and ALS with or without FTD (FTDALS3) (Laurin *et al.*, 2002, Fecto *et al.*, 2011). The large phenotypic spectrum the *SQSTM1* gene gives rise to, is once again supported by the clinical heterogeneity of the Turkish family in question, with both PDB and ALS phenotypes. *SQSTM1* encodes for p62 which has several roles in protein homeostasis, as well as in the autophagic degradataion of the ubiquitin-positive protein aggregates (Kwok *et al.*, 2014).

Earlier, an X-linked dominant *UBQLN2* (*ALS-15*) mutation Met391Ile had been identified in our laboratory with the Madras type of MND (MMND). The mutation had been detected by PCR-based Sanger sequencing. In this study we examined the existence of any additional variation in this patient associated with MMND. Since no other pathogenic mutation aside the one in *UBQLN2* was detected, we conclude that the *UBQLN2* variant

(Met391Ile) is responsible for the phenotype of the patient. This again expands the clinical spectrum of *UBQLN2* mutations.

6.2. Genes Implicated in non-ALS MNDs

In the framework of this thesis a homozygous missense mutation in the *DNAJB2* gene was identified. Mutations in *DNAJB2* (also known as *HSJ1*, *heat-shock protein J1*) are known to cause distal hereditary motor neuron disease (dHMN), and it was shown that the heat shock protein encoded by the *DNAJB2* has an important role in TDP-43 clearance (Gess *et al.*, 2014). Since TDP-43 aggregates are the major hallmark of ALS pathology, loss of function mutations in the *DNAJB2* may cause failure in the resolving of aggregates, thus leading to an ALS phenotype. Also, two Spanish families with the *DNAJB2* mutation have been reported in the literature. In the Spanish study, the patients were followed for 30 years and the phenotype of one of the patients was shown to evoke the final stage of ALS (Frasquet *et al.*, 2016). This scenario points to the importance of long-term follow-up of patients. It would be useful to determine whether these two diseases converge.

The mutation in the *TRPV4* gene described in one of our families shows a remarkable intra-familial clinical variation, ranging from a subclinical phenotype in the asymptomatic father of the probands to a relatively mild phenotype of CMT2C in the younger sister and a more severe scapuloperoneal SMA in the older. While scapuloperoneal SMA is characterized by a congenital reduction of muscles in the peroneus and scapula (shoulder blade), resulting in the typical appearance of ‘scapular winging’ CMT2C is described by a slow progressive muscle weakness and atrophy of the distal muscles (Nilius and Voets, 2013). The phenotypic variability among the reported family members in this thesis, combined with similar examples in the literature, bring together the distinct phenotypes of CMT2C, scapuloperoneal and distal SMA under the same spectrum of *TRPV4* channelopathies.

The BVVL1 syndrome, caused by a mutation in the *SLC52A3* gene, was reported in one of our patients. The *SLC52A3* gene encodes for the riboflavin transporter protein 3 (RFVT3) which is responsible for the transport of riboflavin (commonly known as vitamin B2) across the cell membrane. It has been shown that riboflavin supplementation is an effective treatment for this syndrome. BVVL is characterized by a progressive pontobulbar palsy associated with sensorineural deafness and has phenotypic similarities to ALS with bulbar and LMN involvement. In the literature, a mutation in *UBQLN1*, a gene which belongs to the same family as *UBQLN2 (ALS-15)*, was reported in a patient with BVVL and an atypical early-onset ALS with bulbar palsy and hearing loss, highlighting the overlap of BVVL and ALS (Manole and Houlden, 2015). In the light of these findings, BVVL is considered as the only ALS-like disease which can be treated.

The missense His213Arg mutation in the *IGHMBP2* gene was reported in this thesis in a one-year old infant with spinal muscular atrophy with respiratory distress (SMARD1). The clinical diagnosis of SMARD1 is referred to as “non-5q” or “unusual variant” of SMA. Aside from genetic testing, SMARD1 can be distinguished from SMA1 by the predominance of distal muscle weakness, early involvement of the diaphragm and manifestation of all symptoms in reverse order (Grohmann *et al.*, 2003).

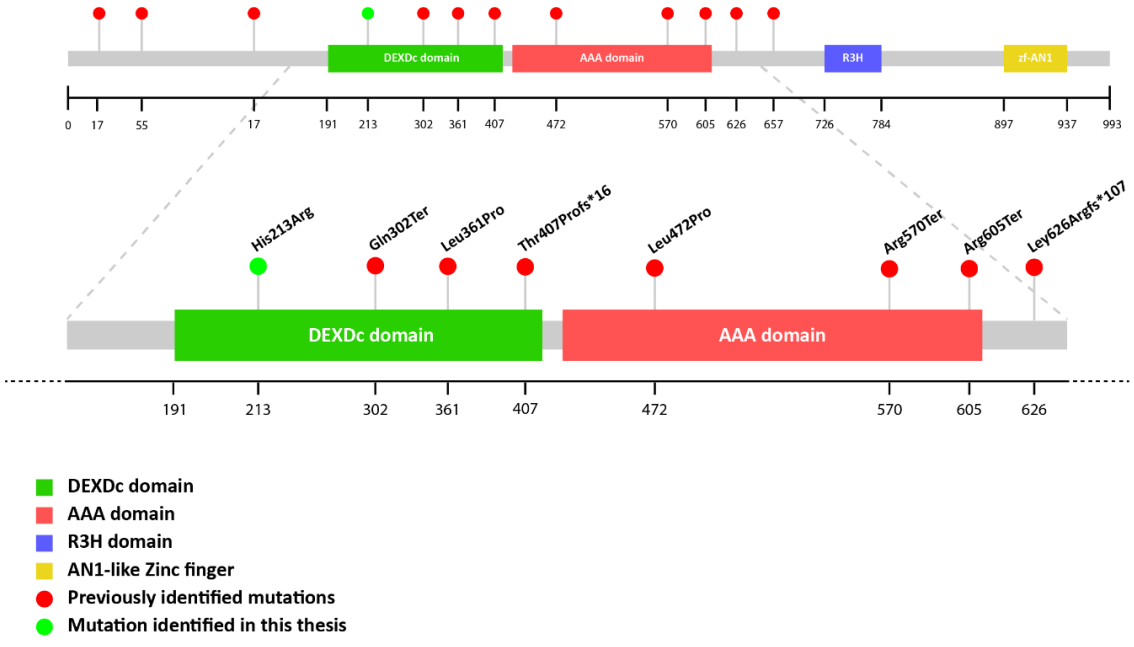


Figure 6.3. Mutations residing on the DEXDc and AAA domains of the *IGHMBP2* gene.

The IGHMBP2 is a multi-domain protein consisting of the following four domains: DNA/RNA-helicase (DEXDc), ATPases associated (AAA), putative single-stranded nucleic acids binding (RH3) and zinc finger motif (zf-AN1). Most of the mutations, including His213Arg were found within or adjacent to the DEXDc and AAA domains, affecting the helicase and ATPase activities of the IGHMBP2 protein (Figure 6.3). Although the precise cellular function and mechanism of IGHMBP2 are still unknown, loss of function mutations in the helicase and ATPase domains seem to be involved in the major pathogenesis of SMARD1. However, rarely, mutations outside the catalytic domains were also shown to cause the SMARD1 phenotype through a reduction in protein level or disruption of protein stability (Guenther *et al.,* 2008).

A homozygous stop-gain mutation in the *PLEKHG5* gene was identified in our cohort in two sisters with an initial diagnosis of ALS and in their brother with a clinical diagnosis of SBMA. The *PLEKHG5* mutations were previously shown to cause juvenile-onset lower motor neuropathy (LMN), leading to muscle wasting of both upper and lower limbs, with an impaired respiration (Maystadt *et al.,* 2006). However, clinical reports suggested an overlap between lower motor neuron diseases and ALS, since some forms of LMN with a rapid progression mimic ALS as well as some forms of ALS, characterized by predominant LMN involvement (Vos and Van den Berg *et al.,* 2001).

The heterozygous Arg98His mutation in the *MPZ* gene was identified in a family with a CMT phenotype. This locus was previously associated with the CMT1B phenotype, harboring the most frequent mutations (Arg98His, Arg98Pro and Arg98Cys) in the *MPZ* gene in the European populations. The *MPZ* gene encodes for myelin protein zero, the most abundant protein in myelin, providing the transmission of nerve impulses; their disruption may cause either demyelinating or axonal CMT (Lagueny *et al.,* 1999).

A splice-site mutation in the *SLC12A6* was identified in two siblings with an initial referral diagnosis of HSP. Mutations in the *SLC12A6* gene, encoding for the ion-transporter protein KCC3, lead to agenesis of the corpus callosum with peripheral neuropathy (ACCPN); this phenotype, also known as Andermann syndrome is present in the Charlevoix

and Saguenay–Lac-St-Jean regions of the province of Quebec with high incidence. The disease is characterized by peripheral neuropathy with partial or complete agenesis of the corpus callosum, several dysmorphic features, mental retardation, and psychosis (Howard *et al.,* 2002). The differential diagnosis of Andermann syndrome may be difficult due to its phenotypic similarities to other forms of HSP as in our case (Schwartzman, 2006).

SPG15 (also known as Kjellin syndrome) is the second most common cause of ARHSP with TCC after SPG11. It is characterized by mental impairment, pigmented maculopathy, dysarthria, cerebellar signs, and distal amyotrophy. Mutations in the *ZFYVE26* gene which encodes for spastizin (spasticity due to the ZFYVE26 protein) are reported to cause the SPG15 phenotype. Spastizin has been shown to localize to the endoplasmic reticulum and endosomes, pointing to a possible role in intracellular trafficking. This might help to understand the mechanism leading to axonal degeneration in SPG15 (Hanein *et al.,* 2008).

The missense mutation in the *LRSAM* gene was found to cause, in addition to ALS, a CMT phenotype in our *ERRB4* family described in 6.1. *LRSAM1* encodes for an E3-ubiquitin protein ligase that has roles in membrane vesicle fusion and proper adhesion of neuronal cells (Guernsey *et al.,* 2010). The *LRSAM1* and *ERBB4* mutations in our patients with ALS and/or CMT2P may explain the phenotypic heterogeneity in our family under investigation.

6.3. Mutations in NBIA Genes Causing ALS and HSP-like Phenotypes

We observed the role of *C19ORF12* mutations in three Turkish patients who were diagnosed with early onset ALS. Mutations in this gene have been associated with autosomal recessive NBIA type 4 called mitochondrial membrane protein-associated neurodegeneration (MPAN). *C19ORF12* is a small gene with less than 17 kb genomic sequence and codes for a transmembrane protein with two alternative isoforms. The first exon of the shorter isoform is not protein-coding, while the longer isoform has a start codon in exon 1 making it eleven amino acids longer. The Gly65Val mutation, which was identified as pathogenic in two of our patients, is located within the predicted transmembrane domain

(Figure 6.4). The third *C19ORF12* mutation is the Thr11Met substitution, the only pathogenic mutation located at the N-terminal of the protein. The Thr11Met mutation affects only the longer isoform of the protein, since it is located upstream of the coding region of the shorter isoform (Hartig *et al.*, 2011). Similar to our cases, two patients with *C19ORF12* mutations have been reported, presenting upper and lower motor neuron dysfunction, mimicking juvenile-onset ALS (Deschauer *et al.*, 2012). Thus, *C19ORF12* is considered as one of the genes causing the juvenile ALS phenotype (Ghasemi and Brown, 2017).

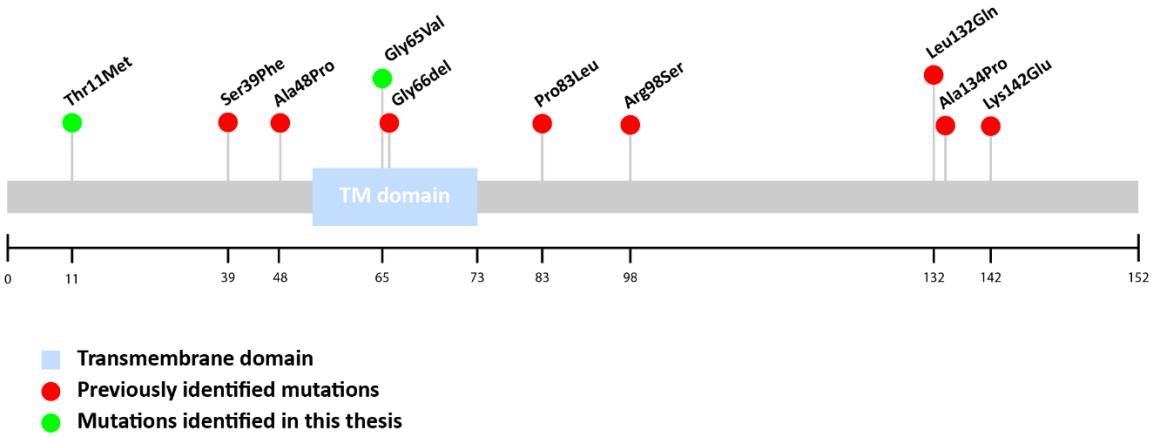


Figure 6.4. Mutations described in the *C19ORF12* gene.

One patient in this study, with a clinical diagnosis of HSP was found to carry a homozygous missense mutation in the *PANK2* gene which is known to cause the most prevalent NBIA type PKAN (pantothenate kinase-associated neurodegeneration). Furthermore, recently a study was reported showing the pathogenic role of the *phospholipases A2 group 6* (*PLA2G6)* gene in HSP patients. The *PLA2G6* is known to cause NBIA type 2, however in this particular study it was shown to be implicated in HSP (Ozes *et al.*, 2017). Our findings combined with the knowledge from the literature review suggest that the genes known to cause NBIA may also be responsible for HSP and ALS, broadening the genotypic spectrum of these diseases.

6.4. Variants with Uncertain Significance

In the framework of this study, a missense mutation in the *ACADS* gene was shown to cause short-chain acyl-CoA dehydrogenase (SCAD) deficiency in a patient with motor neuron disease. SCAD deficiency is a disorder that is characterized by neuromuscular symptoms such as developmental delay, hypotonia, and seizures (Pedersen *et al.*, 2008).To the best of our knowledge, motor neuron involvement in SCAD deficiency has not been reported in the literature. Thus, the mutation in the *ACADS* gene is not sufficient to explain the phenotype of our case by itself. On the other hand, the pathogenicity of the *ACADS* mutation can be tested by measuring the short-chain acyl-CoA dehydrogenase enzyme activity in muscle biopsy and this should be anticipated. Since, no other variation(s) was (were) found in the index case to be associated with motor neuron involvement, the *ACADS* mutation identified was classified as a variant of uncertain significance (VUS) until further validation.

6.5. Remaining Cases to be Solved

In the framework of this thesis, we were able to describe pathogenic mutations in 20 families diagnosed with ALS and/or MND, but we failed to identify the genetic causes in 37 cases (65 %). This result is in a good accordance with recent exome analysis studies in the literature (Iglesias *et al;* 2014; Trujillano *et al.*, 2017). There is still a considerable piece of the puzzle waiting to be solved as classical familial WES approach was not sufficient to uncover all disease-causing factors. The challenges observed in this study can be categorized into the following four major groups, which are presented below.

6.5.1. Technical Limitations of WES in ALS

One of the major drawbacks of the WES is its inability to detect structural variations (SVs) including CNVs, large deletions, insertions and translocations, due to the short-read

sequencing approach in NGS. However, these SVs may lead to an abnormal phenotype, as well as they may represent benign and polymorphic changes (Stankiewiczl and Lupski, 2010). Keeping in mind the SVs in other NDs such as the *SCNA (alpha-synuclein)* duplication in PD and the *SMN1 (survival motor neuron 1)* deletion in SMA, possible roles of SVs in ALS have been questioned. Indeed, the recent discovery of the intronic *C9ORF72* hexanucleotide repeat expansion mutation in ALS and FTD have clarified this point well. The repeat expansions are stretches of satellite DNA sequences and the expansion range is in between hundreds and thousands. Both being such large and residing in the intronic region of the genome, the *C9ORF72* repeat expansion mutation is neither detected by WES nor by WGS. Today, several approaches have been developed to call SVs including read-depth, read-pair, split-read and de-novo sequence assembly (Alkan *et al.,* 2011). However, even the combination of all of these existing algorithms are not yet sufficient to interrogate the SVs and repeat sequences efficiently.

The two major ALS genes *TARDBP* and *FUS* were shown to regulate RNA-splicing by binding to intronic sites (Tourenne *et al.,* 2012). Moreover, mutation analysis of the *OPTN* and *VCP* genes revealed the presence of intronic mutations having role in ALS pathology (Del Bo *et al.,* 2011; Miller *et al.,* 2012). However, WES is designed to capture the exons, thus does not screen intronic regions and regulatory elements, including promoter regions, enhancers and some cryptic splice sites. That means any mutation that occurs at the targeted intronic regions in the above genes and possibly several others are not detected by WES.

WES promises to capture the whole protein-coding region of the genome. However, there are still gaps in the human genome sequence and uncertainties about which sequences are protein-coding and which are not; because the annotation of the approximately 1% of the exome has not been completed yet (Coffey *et al.,* 2011). This incomplete annotation results in the missingness in exome sequencing kits, the region captured by WES. Another technical limitation of WES is the low-coverage problem. This is an even greater problematic situation when the causative variant is in heterozygous state. Since only a few reads are obtained for

a sequence, the causative heterozygous variant would be easily missed due to the low coverage of a particular region.

Aside from sequencing, data processing is the other major step in disease gene discovery. In the framework of this thesis, BWA-GATK best practices with the HaplotypeCaller tool, which is the most widely preferred pipeline was applied. There are several other WES pipelines (BWA-GATK with UnifiedGenotyper tool, Freebayes and BWA-SAMtools with mpileup tool) generating different sets of variations from the same datasets (Hwang *et al.,* 2015). These pipelines may yield lots of false positive mutations and it might be difficult to determine whether a variant is a true false positive or if it is indeed a variant, but covered by only a specific pipeline. We have greatly overcome this limitation as we have the data of family members together with our index cases, which provided us a better calibration reducing the number of false positive variations.

6.5.2. Small Sample Sizes

In the framework of this study, variant lists obtained from the bioinformatic evaluation, were screened for the genes associated with neurological diseases. For some cases, we ended up with a variant list not associated to any neurological dysfunction and failed to pick up the culprit gene(s). Therefore, it means several novel ALS-MND genes are waiting to be unraveled within these lists, except for the cases in which the exact variation is not captured by WES. Since it would be tedious to perform functional analysis for each of those variants, the discovery of the genes that underlie complex disease are possible in two ways: linkage analysis and association studies. However, these analyses require an adequate statistical power, thus larger sample sizes (Glazier *et al.,* 2002). Especially in late-onset diseases, linkage analysis is very limited due to the lack of sufficient family members to examine the cosegregation of disease markers. Besides, linkage analysis of the genes contributing to ALS pathogenesis may be challenging due to locus heterogeneity or low penetrance in ALS. Association is the other approach to uncover the genetic markers of a disease, especially in complex diseases which do not obey a Mendelian pattern of inheritance. Association studies are also based on the statistical significance and to reach a sufficient statistical power is dependent on the size of samples (Baron, 2001; Kiezun *et al.,* 2013). With the increasing number of our ALS cohort and healthy individuals, we would be able to classify our samples

into subgroups based on their phenotypic expressions, such as age of onset, site of onset or a characteristic symptom and associate the genetic information to these different phenotypic expressions.

6.5.3. Importance of a Detailed and Correct Pedigree Information

It is assumed that 10% of ALS patients have a family history of ALS (fALS), and the remaining 90% of patients with no evident family history of ALS are defined as sporadic (sALS). The term sALS might be the result of a misleading pedigree information due to a reduced penetrance, incorrect diagnosis of ancestors, or death from other causes prior to onset of ALS. Today, an apparently sALS patient with a family history of FTD or AD should be considered as fALS due to overlapping genetic backgrounds of ALS and FTD (Boylan, 2015).

The *TRPV4* family in this study is a good example of missing clinical and misleading pedigree information. The sisters have two distinct phenotypic features of a TRPV4-channelopathy, scapuloperoneal SMA and CMT2C. In the initial step, we had evaluated the family, based on the recessive inheritance pattern, since there was no additional family history and their parents were reported to be unaffected. However, a deeper phenotyping revealed that the asymptomatic father and several other members of the family in the upper generations present with scapular winging, moving the direction of our attention towards an autosomal dominant inheritance pattern. Consequently, we have identified the pathogenic *TRPV4* mutation in several individuals among the family with an intra-familial clinical heterogeneity, ranging from asymptomatic to a severe phenotype, emphasizing the importance of deep phenotyping on the pedigree information. In an opposite manner, patients may be misdiagnosed with a phenotype that in fact they do not have. In one of our families with four apparently affected siblings, we had failed to identify the causative mutation. Later, we recognized that three of them (the ones with a severe phenotype) were sharing an *ERRB4* mutation causing ALS, while the remaining affected individual had a CMT mutation explaining her milder phenotype.

6.5.4. The Challenging Epidemiology of ALS

Oligogenic inheritance and pleiotropy are genetic effects of complex diseases, compounding the challenge of interpreting newly identified mutations in ALS. Pleiotropy, the ability of the genetic variations in a particular gene to cause different phenotypes in different individuals is a challenging factor in correlating phenotype and genotype. While pleiotropy in ALS is most frequently associated with FTD, mutations in several other ALS genes were shown to cause different diseases (Andersen and Al-Chalabi, 2011). In the framework of this thesis, we described a *VCP* gene mutation in two patients diagnosed with ALS with a cognitive dysfunction. Using segregation analysis, we detected the very mutation in one of the family members having a cognitive dysfunction without ALS. Since the *VCP* gene mutations were earlier associated both with ALS and FTD, this intra-familial clinical heterogeneity is probably due to the pleiotropy of the *VCP* gene (Watts *et al.,* 2004).

An oligogenic inheritance pattern is defined by the fact that multiple genes or risk variants can be implicated in disease pathogenesis. It refers to the insufficiency of a single gene mutation to cause the disease, hence other risk variants including epigenetic modifications and environmental risk factors might be required to develop the disease (Al-Chalabi and Hardiman, 2013; Al-Chalabi *et al.,* 2016). Considering the cases in whom we could not identify the causative genetic factors so far, two or more mutations could be responsible for their ALS pathology, making these cases much more intriguing.

6.6. WES is Still the Gold Standard to Uncover the Genetics of MND

WES provides the whole protein-coding profile of individuals in an unbiased manner, unlike the conventional methods or targeted NGS-sequencing. Moreover, conventional screening of larger genes such as *SPG11* and *ZFYVE26*, harboring mutations identified in this thesis, would be highly exhaustive and neither time- nor cost-effective. Thus, WES enables us to identify novel variations and novel genotype-phenotype associations. It is possible to screen all suspected genes through WES data at once.

ALS and other MNDs, in fact most neurodegenerative diseases, overlap clinically and may be mimicking each other, e.g. we identified mutations in the distal SMA and NBIA genes in patients with a referral diagnosis of ALS. Our findings support not only the overlapping pathological mechanisms of these diseases, but also the value of WES in differential diagnosis. The genetic background of the patients unraveled allowed us to get the whole picture and especially helps in differential diagnosis of these diseases.

This thesis is a pilot study in a Turkish ALS-MND cohort demonstrating the power of WES approach with a significant success rate. The unbiased nature of exome sequencing was highly effective in unravelling the genetic causes of ALS and other MND patients with a complex genetic and phenotypic heterogeneity. Despite the limitations and challenges both in the technical work and bioinformatic evaluations discussed above, today WES is still the gold standard in investigating complex genetic diseases.

7. CONCLUSION

ALS is the most common motor neuron disease in which the complex genetic background has not been fully described. Keeping in mind the overlap between ALS and other MNDs and the large genotypic spectrum these disease span, complete genetic and environmental factors must be identified first to enlighten the pathogenesis of MNDs. In this study, we aimed to unravel disease-causing mutations in ALS and other MNDs. By using whole exome sequencing we were able to identify pathogenic mutations in several different genes, providing the differential diagnoses of clinically and genetically overlapping MND families. Our results point to a great heterogeneity which, on one hand, stems from the genetic complexity of ALS and, on the other, the ethnic admixture of the Turkish population.

Over the past decade, WES has been proven to be highly efficient in the identification of genes implicated in disease pathogenicity. Since the analysis of high-throughput sequencing data requires a standardized computational pipeline, this thesis is comprised of the establishment of an efficient *in-silico* workflow to process the WES data and the investigation of the MND cases to dissect the genetic components implicated in their phenotype.

This thesis is to the best of our knowledge the most comprehensive study, if not the only, comprised of the bioinformatic evaluation of the WES data of a reasonably large Turkish ALS-MND cohort. We hope that, the results presented in this thesis will not only pave the ways for a more accurate diagnosis of ALS and MND in future, but will eventually also open the avenues for the molecular therapies in motor neuron diseases and ALS in the era of translational medicine.

REFERENCES

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev, 2010, "A Method and Server for Predicting Damaging Missense Mutations.", *Nature Methods*, Vol. 7, No. 4, pp. 248-249.

Al-Chalabi, A., and O. Hardiman, 2013, "The Epidemiology of ALS: A Conspiracy of Genes, Environment and Time", *Nature Reviews Neurology*, Vol. 9, No. 11, pp. 617-628.

Al-Chalabi, A., L. H. van den Berg, and J. Veldink, 2016, "Gene Discovery in Amyotrophic Lateral Sclerosis: Implications for Clinical Management", *Nature Reviews Neurology*, Vol. 13, No. 2, pp. 96-104.

Al-Saif, A., F. Al-Mohanna, and S. Bohlega, 2011, "A Mutation in Sigma-1 Receptor Causes Juvenile Amyotrophic Lateral Sclerosis", *Annals of Neurology*, Vol. 70, No. 6, pp. 913-919.

Alkan, C., B. P. Coe, and E. E. Eichler, 2011, "Genome Structural Variation Discovery and Genotyping", *Nature Reviews Genetics*, Vol. 12, No. 5, pp. 363-376.

Alkuraya, F. S., 2010, "Homozygosity Mapping: One More Tool in the Clinical Geneticists Toolbox", *Genetics Medicine*, Vol. 12, No. 4, pp. 236-239.

Andersen, P. M., and A. Al-Chalabi, 2011, "Clinical Genetics of Amyotrophic Lateral Sclerosis: What Do We Really Know?", *Nature Reviews Neurology*, Vol. 7, No. 11, pp. 603-615.

Auer, P. L., and G. Lettre, 2015, "Rare Variant Association Studies: Considerations, Challenges and Opportunities", *Genome Medicine*, Vol. 7, No.1, pp. 16-26.

Bannwarth, S., S. Ait-El-Mkadem, A. Chaussenot, E. C. Genin, S. Lacas-Gervais, K. Fragaki, L. Berg-Alonso, Y. Kageyama, V. Serre, D. G. Moore, A. Verschueren, C.

Rouzier, I. Le Ber, G. Auge, C. Cochaud, F. Lespinasse, K. NGuyen, A. de Septenville, A. Brice, P. Yu-Wai-Man, H. Sesaki, J. Pouget, and V. Paquis-Flucklinger, 2014, "A Mitochondrial Origin for Frontotemporal Dementia and Amyotrophic Lateral Sclerosis through CHCHD10 Involvement", *Brain*, Vol. 137, No. 8, pp. 2329-2345.

Baron M., 2001, "The Search for Complex Disease Genes: Fault by Linkage or Fault by Association?", *Molecular Psychiatry*,, Vol. 6, No. 2, pp. 143-149.

Hamida Ben M., F. Hentati, and C. B. Hamida, 1990, "Hereditary Motor System Diseases (Chronic Juvenile Amyotrophic Lateral Sclerosis)", *Brain*, Vol. 113, No. 2, pp. 347-363.

Boycott, K. M., M. R. Vanstone, D. E. Bulman, and A. E. MacKenzie, 2013, "Rare-Disease Genetics in the Era of Next-generation Sequencing: Discovery to Translation", *Nature Reviews Genetics*, Vol. 14, No. 10, pp. 681-691.

Boylan, K., 2015, "Familial Amyotrophic Lateral Sclerosis", *Neurologic Clinics*, Vol. 33, No. 4, pp. 807-830.

Butterfield, R. J., D. Ramachandran, S. J. Hasstedt, B. E. Otterud, M. F. Leppert, K. J. Swoboda, and K. M. Flanigan, 2009, "A Novel Form of Juvenile Recessive ALS Maps to Loci on 6p25 and 21q22", *Neuromuscular Disorders*, Vol. 19, No. 4, pp. 279-287.

Chen, Y. Z., C. L. Bennett, H. M. Huynh, I. P. Blair, I. Puls, J. Irobi, I. Dierick, A. Abel, M. L. Kennerson, B. A. Rabin, G. A. Nicholson, M. Auer-Grumbach, K. Wagner, P. De Jonghe, J. W. Griffin, K. H. Fischbeck, V. Timmerman, D. R. Cornblath and P. F. Chance, 2004, "DNA/RNA Helicase Gene Mutations in a Form of Juvenile Amyotrophic Lateral Sclerosis (ALS4)", *The American Journal of Human Genetics*, Vol. 74, No. 6, pp. 1128–1135.

Chesi, A., B. T. Staahl, A. Jovicic, J. Couthouis, M. Fasolino, A. R. Raphael, T. Yamazaki, L. Elias, M. Polak, C. Kelly, K. L. Williams, J. A. Fifita, N. J. Maragakis, G. A. Nicholson, O. D. King, R. Reed, G. R. Crabtree, I. P. Blair, J. D. Glass, and A. D. Gitler, 2013, "Exome Sequencing to Identify De Novo Mutations in Sporadic ALS Trios", *Nature Neuroscience*, Vol. 16, No. 7, pp. 851-855.

Cirulli, E. T., B. N. Lasseigne, S. Petrovski et al., 2015, "Exome Sequencing in Amyotrophic Lateral Sclerosis Identifies Risk Genes and Pathways", *Science*, Vol. 347, No. 6339, pp. 1436-1441.

Coffey, A. J., F. Kokocinski, M. S. Calafato, C. E. Scott, P. Palta, E. Drury, C. J. Joyce, E. M. Leproust, J. Harrow, S. Hunt, A. E. Lehesjoki, D. J. Turner, T. J. Hubbard, and A. Paloti, 2011, "The GENCODE Exome: Sequencing the Complete Human Exome", *Europenan Journal of Human Genetics*, Vol. 19, No. 7, pp. 827-831.

Corcia, P., J. Khoris, P. Couratier, V. Mayeux-Portas, M. H. Meisler, E. Bieth, B. Toffol, A. Autret, J.P. Müh, C. Andres and W. Camu, 2002, "SMN1 Gene Study in Three Families in Which ALS and Spinal Muscular Atrophy Co-Exist", *Neurology*, Vol. 59, No. 9, pp. 1464-1466.

Couthouis, J., M. P. Hart, R. Erion et al., 2012, "Evaluating the Role of the FUS/TLS-Related Gene EWSR1 in Amyotrophic Lateral Sclerosis", *Human Molecular Genetics*, Vol. 21, No. 13, pp. 2899-2911.

Couthouis, J., M. P. Hart, J. Shorter et al., 2011, "A Yeast Functional Screen Predicts New Candidate ALS Disease Genes", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 52, pp. 20881-20890.

Cox, L. E., L. Ferraiuolo, E. F. Goodall, P. R. Heath, A. Higginbottom, H. Mortiboys, H. C. Hollinger, J. A. Hartley, A. Brockington, C. E. Burness, K. E. Morrison, S. B. Wharton, A. J. Grierson, P. G. Ince, J. Kirby, and P. J. Shaw, 2010, "Mutations in CHMP2B in Lower Motor Neuron Predominant Amyotrophic Lateral Sclerosis (ALS)", *PLoS One*, Vol. 5, No. 3, pp. e9872-e9872.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and Group Genomes Project Analysis, 2011, "The Variant Call Format and VCFtools", *Bioinformatics*, Vol. 27, No. 15, pp. 2156-2158.

DeJesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, A. M. Nicholson, N. A. Finch, H. Flynn, J. Adamson, N. Kouri, A. Wojtas, P. Sengdy, G. Y. Hsiung, A. Karydas, W. W. Seeley, K. A. Josephs, G. Coppola, D. H. Geschwind, Z. K. Wszolek, H. Feldman, D. S. Knopman, R. C. Petersen, B. L. Miller, D. W. Dickson, K. B. Boylan, N. R. Graff-Radford, and R. Rademakers, 2011, "Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS", *Neuron*, Vol. 72, No. 2, pp. 245-256.

Del Bo, R., C. Tiloca, V. Pensato, L. Corrado, A. Ratti, N. Ticozzi, S. Corti, B. Castellotti, L. Mazzini, G. Soraru, C. Cereda, S. D"Alfonso, C. Gellera, G. P. Comi, V. Silani, and Slagen Consortium, 2011, "Novel Optineurin Mutations in Patients with Familial and Sporadic Amyotrophic Lateral Sclerosis", *The Journal of Neurology, Neurosurgery, and Psychiatry*, Vol. 82, No. 11, pp. 1239-43.

Deng, H. X., W. Chen, S. T. Hong, K. M. Boycott, et al., 2011, "Mutations in UBQLN2 Cause Dominant X-Linked Juvenile and Adult-Onset ALS and ALS/dementia", *Nature*, Vol. 477, No. 7363, pp. 211-215.

Deschauer, M., C. Gaul, C. Behrmann, H. Prokisch, S. Zierz, and T. B. Haack, 2012, "C19orf12 Mutations in Neurodegeneration with Brain Iron Accumulation Mimicking Juvenile Amyotrophic Lateral Sclerosis", *Journal of Neurology,* Vol. 259, No. 11, pp. 2434-2439.

Dobson-Stone, C., A. A. Luty, E. M. Thompson, P. Blumbergs, W. S. Brooks, C. L. Short, C. D. Field, P. K. Panegyres, J. Hecker, J. A. Solski, I. P. Blair, J. M. Fullerton, G.

M. Halliday, P. R. Schofield, and J. B. Kwok, 2013, "Frontotemporal Dementia-Amyotrophic Lateral Sclerosis Syndrome Locus on Chromosome 16p12.1-q12.2: Genetic, Clinical and Neuropathological Analysis", *Acta Neuropathologica*, Vol. 125, No. 4, pp. 523-33.

Elden, A. C., H. J. Kim, M. P. Hart, A. S. Chen-Plotkin, B. S. Johnson, X. Fang, M. Armakola, F. Geser, R. Greene, M. M. Lu, A. Padmanabhan, D. Clay-Falcone, L. McCluskey, L. Elman, D. Juhr, P. J. Gruber, U. Rub, G. Auburger, J. Q. Trojanowski, V. M. Lee, V. M. Van Deerlin, N. M. Bonini, and A. D. Gitler, 2010, "Ataxin-2 Intermediate-Length Polyglutamine Expansions are Associated with Increased Risk for ALS", *Nature*, Vol. 466, No. 7310, pp. 1069-75.

Figlewicz, D. A., A. Krizus, M. G. Martinoli, V. Meininger, M. Dib, G. A. Rouleau, and J. P. Julien, 1994, "Variants of the Heavy Neurofilament Subunit are Associated with the Development of Amyotrophic Lateral Sclerosis", *Human Molecular Genetics*, Vol. 3, No. 10, pp. 1757-1761.

Fogh, I., A. Ratti, C. Gellera et al., 2014, "A Genome-Wide Association Meta-Analysis Identifies a Novel Locus at 17q11.2 Associated with Sporadic Amyotrophic Lateral Sclerosis", *Human Molecular Genetics*, Vol. 23, No. 8, pp. 2220-31.

Foo, J. N., J. J. Liu, and E. K. Tan, 2012, "Whole-genome and Whole-Exome Sequencing in Neurological Diseases", *Nature Reviews Neurology*, Vol. 8, No. 9, pp. 508-17.

Frasquet, M., J. F. Vazquez-Costa, and T. Sevilla, 2016, "The Role of DNAJB2 in Amyotrophic Lateral Sclerosis", *Brain*, Vol. 139, No. 10, pp. e57-e57.

Freischmidt, A., T. Wieland, B. Richter et al., 2015, "Haploinsufficiency of TBK1 Causes Familial ALS and Fronto-Temporal Dementia", *Nature Neuroscience*, Vol. 18, No. 5, pp. 631-6.

Gal, J., A. L. Strom, D. M. Kwinter, R. Kilty, J. Zhang, P. Shi, W. Fu, M. W. Wooten, and H. Zhu, 2009, "Sequestosome 1/p62 Links Familial ALS Mutant SOD1 to LC3 via an Ubiquitin-Independent Mechanism", *Journal of Neurochemistry*, Vol. 111, No. 4 pp. 1062-73.

Gess, B., M. Auer-Grumbach, A. Schirmacher, T. Strom, M. Zitzelsberger, S. Rudnik-Schoneborn, D. Rohr, H. Halfter, P. Young, and J. Senderek, 2014, "HSJ1-related Hereditary Neuropathies: Novel Mutations and Extended Clinical Spectrum", *Neurology*, Vol. 83, No. 19, pp. 1726-32.

Ghasemi, M., and R. H. Brown, Jr., 2017, "Genetics of Amyotrophic Lateral Sclerosis", *Cold Spring Harbour Perspect Medicine*, Vol. 7, No. 3, pp. a024125- a024125.

Glazier A. M., J. H. Nadeau and T. J. Aitman, 2002, "Finding Genes that Underlie Complex Traits.", *Science*, Vol. 298, No. 5602, pp. 2345-2349.

Greenway, M. J., P. M. Andersen, C. Russ, S. Ennis, S. Cashman, C. Donaghy, V. Patterson, R. Swingler, D. Kieran, J. Prehn, K. E. Morrison, A. Green, K. R. Acharya, R. H. Brown, Jr., and O. Hardiman, 2006, "ANG Mutations Segregate with Familial and Sporadic Amyotrophic Lateral Sclerosis", *Nature Genetics*, Vol. 38, No. 4, pp. 411-3.

Grohmann K., R. Varon, P. Stolz, M. Schuelke, C. Janetzki, E. Bertini, K. Bushby, F. Muntoni, R. Ouvrier, L. Van Maldergem, N. M. L. A. Goemans, H. Lochmüller, S. Eichholz, C. Adams, F. Bosch, P. Grattan-Smith, C. Navarro, H. Neitzel, T. Polster, H. Topaloğlu, C. Steglich, U. P. Guenther, K. Zerres, S. Rudnik-Schöneborn and C. Hübner, 2003, "Infantile Spinal Muscular Atrophy with Respiratory Distress Type 1 (SMARD1)", *Annals of Neurology*, Vol. 54, No. 6, pp. 719-724.

Gros-Louis, F., R. Lariviere, G. Gowing, S. Laurent, W. Camu, J. P. Bouchard, V. Meininger, G. A. Rouleau, and J. P. Julien, 2004, "A Frameshift Deletion in

Peripherin Gene Associated with Amyotrophic Lateral Sclerosis", *The Journal of Biological Chemistry*, Vol. 279, No. 44, pp. 45951-6.

Guenther, U. P., R. Varon, M. Schlicke, V. Dutrannoy, A. Volk, C. Hubner, K. von Au, and M. Schuelke, 2007, "Clinical and Mutational Profile in Spinal Muscular Atrophy with Respiratory Distress (SMARD): Defining Novel Phenotypes through Hierarchical Cluster Analysis", *Human Mutation*, Vol. 28, No. 8, pp. 808-15.

Guernsey, D. L., H. Jiang, K. Bedard, S. C. Evans, M. Ferguson, M. Matsuoka, C. Macgillivray, M. Nightingale, S. Perry, A. L. Rideout, A. Orr, M. Ludman, D. L. Skidmore, T. Benstead, and M. E. Samuels, 2010, "Mutation in the Gene Encoding Ubiquitin Ligase LRSAM1 in Patients with Charcot-Marie-Tooth Disease", *PLoS Genetics*, Vol. 6, No. 8, pp. e1001081- e1001081

Gusella J. F., N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, A. Y. Sakaguchi, A. B. Young, I. Shoulson, E. Bonilla and J. B. Martin, 1983, "A Polymorphic DNA Marker Genetically Linked to Huntington's Disease", *Nature*, Vol. 306, No. 5940, pp. 234-238.

Hand, C. K., J. Khoris, F. Salachas, F. Gros-Louis, A. A. Lopes, V. Mayeux-Portas, C. G. Brewer, R. H. Brown, V. Meininger, W. Camu and G. A. Rouleau, 2002, "A Novel Locus for Familial Amyotrophic Lateral Sclerosis, on Chromosome 18q", *The American Journal of Human Genetics*, Vol. 70, No. 1, pp. 251-256.

Hanein, S., E. Martin, A. Boukhris, P. Byrne, C. Goizet, A. Hamri, A. Benomar, A. Lossos, P. Denora, J. Fernandez, N. Elleuch, S. Forlani, A. Durr, I. Feki, M. Hutchinson, F. M. Santorelli, C. Mhiri, A. Brice, and G. Stevanin, 2008, "Identification of the SPG15 Gene, Encoding Spastizin, As a Frequent Cause of Complicated Autosomal-Recessive Spastic Paraplegia, Including Kjellin Syndrome", *The American Journal of Human Genetics*, Vol. 82, No. 4, pp. 992-1002.

Hartig, M., H. Prokisch, T. Meitinger, and T. Klopstock, 2013, "Mitochondrial Membrane Protein-Associated Neurodegeneration (MPAN)", *International Review of Neurobiology*, Vol. 110, No. 1, p. 73-84.

Hoglinger, G. U., N. M. Melhem, D. W. Dickson et al., 2011, "Identification of Common Variants Influencing Risk of the Tauopathy Progressive Supranuclear Palsy", *Nature Genetics*, Vol. 43, No. 7, pp. 699-705.

Hwang, S., E. Kim, I. Lee, and E. M. Marcotte, 2015, "Systematic Comparison of Variant Calling Pipelines Using Gold Standard Personal Exome Variants", *Scientific Reports*, Vol. 5, No. 17875.

Iglesias, A., K. Anyane-Yeboa, J. Wynn, A. Wilson, M. Truitt Cho, E. Guzman, R. Sisson, C. Egan, and W. K. Chung, 2014, "The Usefulness of Whole-Exome Sequencing in Routine Clinical Practice", *Genetics Medicine*, Vol. 16, No. 12, pp. 922-31.

Iskender, C., E. Kartal, F. Akcimen, C. Kocoglu, A. Ozoguz, D. Kotan, M. Eraksoy, Y. G. Parman, and A. N. Basak, 2015, "Turkish Families with Juvenile Motor Neuron Disease Broaden the Phenotypic Spectrum of SPG11", *Neurology Genetics*, Vol. 1, No. 3, pp. e25-e25.

James, P. A., and K. Talbot, 2006, "The Molecular Genetics of Non-ALS Motor Neuron Diseases", *Biochimica et Biophysica Acta*, Vol. 1762, No. 11, pp. 986-1000.

Jiang, T., M. S. Tan, L. Tan, and J. T. Yu, 2014, "Application of Next-Generation Sequencing Technologies in Neurology", *Annals of Translational Medicine*, Vol. 2, No. 12, pp. 125.

Johnson, J. O., J. Mandrioli, M. Benatar et al., 2010, "Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS", *Neuron*, Vol. 68, No. 5, pp. 857-64.

Johnson, J. O., E. P. Pioro, A. Boehringer et al., 2014, "Mutations in the Matrin 3 Gene Cause Familial Amyotrophic Lateral sclerosis", *Nature Neuroscience*, Vol. 17, No. 5, pp. 664-66.

Kancheva, D., D. Atkinson, P. De Rijk, M. Zimon, T. Chamova, V. Mitev, A. Yaramis, G. Maria Fabrizi, H. Topaloglu, I. Tournev, Y. Parman, E. Battaloglu, A. Estrada-Cuzcano, and A. Jordanova, 2016, "Novel mutations in Genes Causing Hereditary Spastic Paraplegia and Charcot-Marie-Tooth Neuropathy Identified by an Optimized Protocol for Homozygosity Mapping Based on Whole-Exome Sequencing", *Genetics Medicine*, Vol. 18, No. 6, pp. 600-7.

Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Hausler, 2002, "The human genome browser at UCSC", *Genome Resourcest*, Vol. 12, No. 6, pp. 996-1006. https://genome.ucsc.edu, accessed at July 2017.

Kiernan, M. C., Steve Vucic, B. C. Cheah, M. R. Turner, A. Eisen, O. Hardiman, J. R. Burrell, and M. C. Zoing, 2011, "Amyotrophic lateral sclerosis", *The Lancet*, Vol. 377, No. 9769, pp. 942-55.

Kiezun, A., K. Garimella, R. Do, N. O. Stitziel, B. M. Neale, P. J. McLaren, N. Gupta, P. Sklar, P. F. Sullivan, J. L. Moran, C. M. Hultman, P. Lichtenstein, P. Magnusson, T. Lehner, Y. Y. Shugart, A. L. Price, P. I. de Bakker, S. M. Purcell, and S. R. Sunyaev, 2012, "Exome Sequencing and the Genetic Basis of Complex Traits", *Nature Genetics*, Vol. 44, No. 6, pp. 623-30.

Kim, H. J., N. C. Kim, Y. D. Wang et al., 2013, "Mutations in Prion-Like Domains in HnRNPA2B1 and HnRNPA1 Cause Multisystem Proteinopathy and ALS", *Nature*, Vol. 495, No. 7442, pp. 467-73.

Kim, J., Y. H. Liao, C. Ionita, A. E. Bale, B. Darras, and G. Acsadi, 2016, "Mitochondrial Membrane Protein-Associated Neurodegeneration Mimicking Juvenile Amyotrophic Lateral Sclerosis", *Pediatric Neurology*, Vol. 64, , p. 83-86.

Kumar, D. R., F. Aslinia, S. H. Yale, and J. J. Mazza, 2011, "Jean-Martin Charcot: The Father of Neurology", *Clinical Medical Resources*, Vol. 9, No. 1, pp. 46-49.

Kwiatkowski TJ Jr, Bosco DA, Leclerc AL, Tamrazian E,Vanderburg CR, Russ C, Davis A, Gilchrist J, Kasarskis EJ,Munsat T, P. Valdmanis, G. A. Rouleau, B. A. Hosler, P. Cortelli, P. J. de Jong, Y. Yoshinaga, J. L. Haines, M. A. Pericak-Vance, J. Yan, N. Ticozzi, T. Siddique, D. McKenna-Yasek, P. C. Sapp, H. R. Horvitz, J. E. Landers, R. H. Brown Jr., 2009, "Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis", *Science* Vol. 323, No. 5918, pp. 1205–1208.

Kwok, C. T., A. Morris, and J. S. de Belleroche, 2014, "Sequestosome-1 (SQSTM1) Sequence Variants in ALS Cases in the UK: Prevalence and Coexistence of SQSTM1 Mutations in ALS Kindred with PDB", *The European Journal of Human Genetics*, Vol. 22, No. 4, pp. 492-6.

Lagier-Tourenne, C., M. Polymenidou, K. R. Hutt, A. Q. Vu, M. Baughn, S. C. Huelga, K. M. Clutario, S. C. Ling, T. Y. Liang, C. Mazur, E. Wancewicz, A. S. Kim, A. Watt, S. Freier, G. G. Hicks, J. P. Donohue, L. Shiue, C. F. Bennett, J. Ravits, D. W. Cleveland, and G. W. Yeo, 2012, "Divergent Roles of ALS-Linked Proteins FUS/TLS and TDP-43 Intersect in Processing Long Pre-mRNAs", *Nature Neuroscience*, Vol. 15, No. 11, pp. 1488-97.

Laugeny A., L.P. Latour, A. Vital, Y. Rajabally, G. Le Masson, X. Ferrer, I. Bernard, J. Julien, C. Vital and A. Vandenberghe, 1999, "Peripheral Myelin Modification in CMT1B Correlates with MPZ Gene Mutations.", *Neuromuscular Disordors,* Vol. 9, No. 6, pp. 361-367.

Laurin, N., J. P. Brown, J. Morissette, and V. Raymond, 2002, "Recurrent Mutation of the Gene Encoding Sequestosome 1 (SQSTM1/p62) in Paget Disease of Bone", *The American Journal of Human Genetics*, Vol. 70, No. 6, pp. 1582-8.

Leblond, C. S., H. M. Kaneb, P. A. Dion, and G. A. Rouleau, 2014, "Dissection of Genetic Factors Associated with Amyotrophic Lateral Sclerosis", *Experimental Neurology*, Vol. 262, , pp.91-101.

Lek, M., K. J. Karczewski, E. V. Minikel et al., 2016. "Analysis of Protein-Coding Genetic Variation in 60,706 Humans", *Nature*, 536: 285-91.

Li, H., and R. Durbin, 2009, "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform", *Bioinformatics*, Vol. 25, No. 14, pp. 1754-60.

Manole A. and H. Houlden, 2015, "Riboflavin Transporter Deficiency Neuronopathy", *Gene Reviews.*

Maruyama, H., H. Morino, H. Ito, Y. Izumi, H. Kato, Y. Watanabe, Y. Kinoshita, M. Kamada, H. Nodera, H. Suzuki, O. Komure, S. Matsuura, K. Kobatake, N. Morimoto, K. Abe, N. Suzuki, M. Aoki, A. Kawata, T. Hirai, T. Kato, K. Ogasawara, A. Hirano, T. Takumi, H. Kusaka, K. Hagiwara, R. Kaji, and H. Kawakami, 2010, "Mutations of Optineurin in Amyotrophic Lateral Sclerosis", *Nature*, Vol. 465, No. 7295, pp. 223-6.

Maystad I., M. Zarhrate, D. Leclair-Richard, B. Estournet, A. Barois, F. Renault, M. C. Routon, M. C. Durand, S. Lefebvre, A. Munnich, C. Verellen-Dumoulin and L. Viollet, 2006, "A Gene for an Autosomal Recessive Lower Motor Neuron Disease with Childhood Onset Maps to 1p36", *Neurology*, Vol. 67, No. 1, pp. 120-124.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, 2010, "The Genome Analysis Toolkit: A Map Reduce Framework for Analyzing Next-Generation DNA Sequencing Data", *Genome Resources*, Vol. 20, No. 9, pp. 1297-303.

Mitchell, J., P. Paul, H. J. Chen, A. Morris, M. Payling, M. Falchi, J. Habgood, S. Panoutsou, S. Winkler, V. Tisato, A. Hajitou, B. Smith, C. Vance, C. Shaw, N. D. Mazarakis, and J. de Belleroche, 2010, "Familial Amyotrophic Lateral Sclerosis is Associated With a Mutation in D-Amino Acid Oxidase", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 16, pp. 7556-61.

Mullen S. A., D. E. Crompton, P. W. Carney, I. Helbig and S. F. Berkovic, 2009, "A Neurologist's Guide to Genome-Wide Association Studies", *Neurology*, Vol. 72, No. 6, pp. 558-565.

Munch, C., R. Sedlmeier, T. Meyer, V. Homberg, A. D. Sperfeld, A. Kurt, J. Prudlo, G. Peraus, C. O. Hanemann, G. Stumm, and A. C. Ludolph, 2004, "Point Mutations of the p150 Subunit of Dynactin (DCTN1) Gene in ALS", *Neurology*, Vol. 63, No. 4, pp. 724-26.

NG P. C., and S. Henikoff, 2003, "SIFT: Predicting Amino Acid Changes That Affect Protein Function", *Nucleic Acid Resources*, Vol. 13, No. 13, pp. 3812–3814.

Nilius, B., and T. Voets, 2013, "The Puzzle of TRPV4 Channelopathies", *EMBO Reports*, Vol. 14, No. 2, pp. 152-63.

Nishimura, A. L., M. Mitne-Neto, H. C. Silva, A. Richieri-Costa, S. Middleton, D. Cascio, F. Kok, J. R. Oliveira, T. Gillingwater, J. Webb, P. Skehel and M. Zatz. 2004. "A Mutation in the Vesicle-Trafficking Protein VAPB Causes Late-Onset Spinal Muscular Atrophy and Amyotrophic Lateral Sclerosis", *Am J Human Genetics*, 75: 822–831.

Orlacchio, A., C. Babalini, A. Borreca, C. Patrono, R. Massa, S. Basaran, R. P. Munhoz, E. A. Rogaeva, P. H. St George-Hyslop, G. Bernardi, and T. Kawarai, 2010, "SPATACSIN Mutations Cause Autosomal Recessive Juvenile Amyotrophic Lateral Sclerosis", *Brain*, Vol. 133, No. 2, pp. 591-8.

Ott, J., Y. Kamatani, and M. Lathrop, 2011, "Family-Based Designs for Genome-Wide Association Studies", *Nature Review Genetics*, Vol. 12, No. 7, pp. 465-74.

Ott, J., J. Wang, and S. M. Leal, 2015, "Genetic Linkage Analysis in the Age of Whole-Genome Sequencing", *Nature Review Genetics*, Vol. 16, No. 5, pp. 275-84.

Ozes, B., N. Karagoz, R. Schule, A. Rebelo, M. J. Sobrido, F. Harmuth, M. Synofzik, S. I. P. Pascual, M. Colak, B. Ciftci-Kavaklioglu, B. Kara, A. Ordonez-Ugalde, B. Quintans, M. A. Gonzalez, A. Soysal, S. Zuchner, and E. Battaloglu, 2017, "PLA2G6 Mutations Associated With a Continuous Clinical Spectrum From Neuroaxonal Dystrophy to Hereditary Spastic Paraplegia", *Clinical Genetics*, epub: DOI: 10.1111/cge.13008.

Ozoguz, A., O. Uyan, G. Birdal et al., 2015, "The Distinct Genetic Pattern of ALS in Turkey and Novel Mutations", *Neurobiology of Aging*, Vol. 36, No. 4, pp. 1764 e9-18.

Pedersen, C. B., S. Kolvraa, A. Kolvraa, V. Stenbroen, M. Kjeldsen, R. Ensenauer, I. Tein, D. Matern, P. Rinaldo, C. Vianey-Saban, A. Ribes, W. Lehnert, E. Christensen, T. J. Corydon, B. S. Andresen, S. Vang, L. Bolund, J. Vockley, P. Bross, and N. Gregersen, 2008, "The ACADS Gene Variation Spectrum in 114 Patients with Short-Chain Acyl-CoA Dehydrogenase (SCAD) Deficiency is Dominated by Missense Variations Leading to Protein Misfolding at the Cellular Level", *Human Genetics*, Vol. 124, No. 1, pp. 43-56.

Perez-Branguli, F., H. K. Mishra, I. Prots, S. Havlicek, Z. Kohl, D. Saul, C. Rummel, J. Dorca-Arevalo, M. Regensburger, D. Graef, E. Sock, J. Blasi, T. W. Groemer, U. Schlotzer-Schrehardt, J. Winkler, and B. Winner, 2014, "Dysfunction of Spatacsin Leads to Axonal Pathology in SPG11-Linked Hereditary Spastic Paraplegia", *Human Molecular Genetics*, Vol. 23, No. 18, pp. 4859-74.

Przedborski, S., M. Vila, V. Jackson-Lewis, 2003, "Series Introduction: Neurodegeneration: What Is It and Where Are We?", *Journal of Clinical Investigation*, Vol. 111, No. 1, pp. 3-10.

Pulst, S. M, 1999, "Genetic Linkage Analysis", *Archieves of Neurolology*, Vol. 56, No. 6, pp. 667-72.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham, 2007, "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses", *The American Journal of Human Genetics*, Vol. 81, No. 3, pp. 559-75.

Rainier, S., M. Bui, E. Mark, D. Thomas, D. Tokarz, L. Ming, C. Delaney, R. J. Richardson, J. W. Albers, N. Matsunami, J. Stevens, H. Coon, M. Leppert, and J. K. Fink, 2008, "Neuropathy Target Esterase Gene Mutations Cause Motor Neuron Disease", *The American Journal of Human Genetics*, Vol. 82, No. 3, pp. 780-5.

Renton, A. E., E. Majounie, A. Waite et al., 2011, "A Hexanucleotide Repeat Expansion in C9ORF72 is the Cause of Chromosome 9p21-Linked ALS-FTD", *Neuron*, Vol. 72, No. 2, pp. 257-68.

RFFlow, http://www.rff.com, accessed at July 2017.

Rosen, R., T. Siddique, D. Patterson et al., 1993, "Mutations in Cu/Zn Superoxide Dismutase Gene are Associated With Familial Amyotrophic Lateral Sclerosis", *Nature*, Vol. 362, No. 6415, pp. 59-62.

Sabatelli, M., F. Eusebi, A. Al-Chalabi et al., 2009, "Rare Missense Variants of Neuronal Nicotinic Acetylcholine Receptor Altering Receptor Function are Associated With Sporadic Amyotrophic Lateral Sclerosis", *Human Molecular Genetics*, Vol. 18, No. 20, pp. 3997-4006.

Schwartzmann R. J., 2006, "Spastic Paraparesis", *Differential Diagnosis in Neurology*, Vol. 1, , pp. 225.

Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin, 2001, "dbSNP: The NCBI Database of Genetic Variation", *Nucleic Acids Research,* Vol. 29, No. 1, pp. 308-311.

Siddique T., D. A. Figlewicz, M. A. Pericak-Vance, J. L. Haines, G. A. Rouleau, A. J. Jeffers, P. Sapp, W. Y. Hung, J. Bebout, D. McKenna-Yasek, G. Deng, H. R. Horvitz, J. F. Gusella, R. H. Brown and A. D. Roses, 1991, "Linkage of a Gene Causing Familial Amyotrophic Lateral Sclerosis to Chromosome 21 and Evidence of Genetic-Locus Heterogeneity.", *The New England Journal of Medicine,* Vol. 324, No. 20, pp. 1381-1384.

Simpson, C. L., R. Lemmens, K. Miskiewicz et al., 2009, "Variants of the Elongator Protein 3 (ELP3) Gene are Associated With Motor Neuron Degeneration", *Human Molecular Genetics*, Vol. 18, No. 3, pp. 472-81.

Slowik, A., B. Tomik, P. P. Wolkow, D. Partyka, W. Turaj, M. T. Malecki, J. Pera, T. Dziedzic, A. Szczudlik, and D. A. Figlewicz, 2006, "Paraoxonase Gene Polymorphisms and Sporadic ALS", *Neurology*, Vol. 67, No. 5, pp. 766-70.

Smedley, D., S. Haider, S. Durinck et al., 2015, "The BioMart Community Portal: An Innovative Alternative to Large, Centralized Data Repositories", *Nucleic Acids Res*, Vol. 43, No. 1, pp. W589-98.

Smith, B. N., N. Ticozzi, C. Fallini et al., 2014, "Exome-Wide Rare Variant Analysis Identifies TUBA4A Mutations Associated With Familial ALS", *Neuron*, Vol. 84, No. 2, pp. 324-31.

Sreedharan J1., I. Blair, V. B. Tripathi, X. Hu, C. Vance, B. Rogelj, S. Ackerley, J. C. Durnall, K. L. Williams, E. Buratti, F. Baralle, J. de Belleroche, J. D. Mitchell, P. N. Leigh, A. Al-Chalabi, C. C. Miller, G. Nicholson and C. E. Shaw, 2008, "TDP-43 Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis", *Science*, Vol. 319, No. 5870, pp. 1668-1672.

Stankiewicz, P., and J. R. Lupski, 2010, "Structural Variation in the Human Genome and its Role in Disease", *Annual Review of Medicine*, Vol. 61, No. 1, pp. 437-55.

Stevanin, G., F. M. Santorelli, H. Azzedine, P. Coutinho, J. Chomilier, P. S. Denora, E. Martin, A. M. Ouvrard-Hernandez, A. Tessa, N. Bouslam, A. Lossos, P. Charles, J. L. Loureiro, N. Elleuch, C. Confavreux, V. T. Cruz, M. Ruberg, E. Leguern, D. Grid, M. Tazir, B. Fontaine, A. Filla, E. Bertini, A. Durr, and A. Brice, 2007, "Mutations in SPG11, Encoding Spatacsin, Are A Major Cause of Spastic Paraplegia With Thin Corpus Callosum", *Nature Genetics*, Vol. 39, No. 3, pp. 366-72.

Takahashi, Y., Y. Fukuda, J. Yoshimura et al., 2013, "ERBB4 Mutations That Disrupt the Neuregulin-ErbB4 Pathway Cause Amyotrophic Lateral Sclerosis Type 19", *The American Journal of Human Genetics*, Vol. 93, No. 5, pp. 900-5.

Taylor, J. P., R. H. Brown, Jr., and D. W. Cleveland, 2016, "Decoding ALS: From Genes To Mechanism", *Nature*, Vol. 539, No. 7628, pp. 197-206.

Teer, J. K., E. D. Green, J. C. Mullikin, and L. G. Biesecker, 2012, "VarSifter: Visualizing and Analyzing Exome-Scale Sequence Variation Data on A Desktop Computer", *Bioinformatics*, Vol. 28, No. 4, pp. 599-600.

Therrien, M., P. A. Dion, and G. A. Rouleau, 2016, "ALS: Recent Developments From Genetics Studies", *Current Neurology and Neuroscience Reports*, Vol. 16, No. 6, pp. 59.

Trujillano, D., A. M. Bertoli-Avella, K. Kumar Kandaswamy, M. E. Weiss, J. Koster, A. Marais, O. Paknia, R. Schroder, J. M. Garcia-Aznar, M. Werber, O. Brandau, M. Calvo Del Castillo, C. Baldi, K. Wessel, S. Kishore, N. Nahavandi, W. Eyaid, M. T. Al Rifai, A. Al-Rumayyan, W. Al-Twaijri, A. Alothaim, A. Alhashem, N. Al-Sannaa, M. Al-Balwi, M. Alfadhel, A. Rolfs, and R. Abou Jamra, 2017, "Clinical Exome Sequencing: Results From 2819 Samples Reflecting 1000 Families", *The European Journal of Human Genetics*, Vol. 25, No. 2, pp. 176-82.

Tsuji, S., 2010, "Genetics of Neurodegenerative Diseases: Insights From High-Throughput Resequencing", *Human Molecular Genetics*, Vol. 19, No. 1, pp. 65-70.

Van den Berg-Vos R. M., L. H. Van den Berg, G. H. Jansen, M. Parton, C. E. Shaw, A. L. Hesseling-Janssen and J. H. Wokke, 2001, "Hereditary Pure Lower Motor Neuron Disease With Adult Onset and Rapid Progression", *Neurology*, Vol. 67, No. 4, pp. 120-124.

van Es, M. A., J. H. Veldink, C. G. Saris et al., 2009, "Genome-Wide Association Study Identifies 19p13.3 (UNC13A) and 9p21.2 as Susceptibility Loci for Sporadic Amyotrophic Lateral Sclerosis", *Nature Genetics*, Vol. 41, No. 10, pp. 1083-7.

Veldink JH, Kalmijn S, Van der Hout AH, Lemmink HH, Groeneveld GJ, Lummen C, Scheffer H, Wokke JH, Van den Berg LH, 2005, "SMN Genotypes Producing Less SMN Protein Increase Susceptibility to and Severity of Sporadic ALS", *Neurology*, Vol. 65, No. 6, pp. 820-825.

Wain, L. V., I. Pedroso, J. E. Landers, G. Breen, C. E. Shaw, P. N. Leigh, R. H. Brown, M. D. Tobin, and A. Al-Chalabi, 2009, "The Role of Copy Number Variation in Susceptibility to Amyotrophic Lateral Sclerosis: Genome-Wide Association Study and Comparison With Published Loci", *PLoS One*, Vol. 4, No. 12, pp. e8175-e8175.

Wang, K., M. Li, and H. Hakonarson, 2010, "ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data", *Nucleic Acids Resources*, Vol. 38, No. 16, pp. e164-e164.

Watts, G. D., J. Wymer, M. J. Kovach, S. G. Mehta, S. Mumm, D. Darvish, A. Pestronk, M. P. Whyte, and V. E. Kimonis, 2004, "Inclusion Body Myopathy Associated with Paget Disease of Bone and Frontotemporal Dementia is Caused by Mutant Valosin-Containing Protein", *Nature Genetics*, Vol. 36, No. 4, pp. 377-81.

Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y. J. Chen, V. Makhijani, G. T. Roth, X. Gomes, K. Tartaro, F. Niazi, C. L. Turcotte, G. P. Irzyk, J. R. Lupski, C. Chinault, X. Z. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D. M. Muzny, M. Margulies, G. M. Weinstock, R. A. Gibbs, and J. M. Rothberg, 2008, "The Complete Genome of an Individual by Massively Parallel DNA Sequencing", *Nature*, Vol. 452, No. 7189, pp. 872-6.

Williams, K. L., S. Topp, S. Yang et al., 2016, "CCNF Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia", *Nature Communiactions*, Vol. 7, No. 11253, pp. 5-8.

Wu, C. H., C. Fallini, N. Ticozzi et al., 2012, "Mutations in the Profilin 1 Gene Cause Familial Amyotrophic Lateral Sclerosis", *Nature*, Vol. 488, No. 7412, pp. 499-503.

Yang, Y., A. Hentati, H. X. Deng, O. Dabbagh, T. Sasaki, M. Hirano, W. Y. Hung, K. Ouahchi, J. Yan, A. C. Azim, N. Cole, G. Gascon, A. Yagmour, M. Ben-Hamida, M. Pericak-Vance, F. Hentati and T. Siddique, 2001, "The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis", *Nature Genetics*, Vol. 29, No. 2, pp. 160-5.

Zhang, X., C. Y. Chow, Z. Sahenk, M. E. Shy, M. H. Meisler, and J. Li., 2008, "Mutation of FIG4 causes a rapidly progressive, asymmetric neuronal degeneration.", *Brain*, Vol. 131, No. 2, pp. 1990-2001.

APPENDIX A: COMMANDS EXECUTED IN ANALYSES OF WHOLE EXOME SEQUENCING DATA

Table A1. List of alignment commands.

Command List for Alignment
bwa aln -t 200 -f $sampleID_R1.sai $referencegenome $sampleID_R1.fastq.gz
bwa sampe -r "$RG" $referencegenome $sampleID_R1.sai $sampleID_R2.sai $sampleID_R1.fastq.gz $sampleID_R2.fastq.gz
samtools view -bS - > $sampleID.bam
samtools sort $sampleID.bam $sampleID.sorted
samtools rmdup -sS $sampleID.sorted.bam $sampleID.rmdup.bam
samtools index $sampleID.rmdup.bam

Table A2. List of variant calling commands.

Command List for Variant Calling
java –jar GenomeAnalysisTK -T RealignerTargetCreator -R $referencegenome -I $sampleID.rmdup.bam -o $sampleID.rmdup.bam.intervals -nt 3 -known Mills_and_1000G_gold_standard.indels.b37.vcf -known 1000G_phase1.indels.b37.vcf
java –jar GenomeAnalysisTK -T IndelRealigner -targetIntervals $sampleID.rmdup.bam.intervals -R $referencegenome -I $sampleID.rmdup.bam -known Mills_and_1000G_gold_standard.indels.b37.vcf -known 1000G_phase1.indels.b37.vcf -o $sampleID.realigned.bam
java –jar GenomeAnalysisTK -T BaseRecalibrator -I $sampleID.realigned.bam -R $referencegenome -knownSites dbsnp_138.b37.vcf -nct 4 -o $sampleID.report.grp -lqt 2 -mdq -1
java –jar GenomeAnalysisTK -T PrintReads -R $referencegenome -I $sampleID.realigned.bam -nct 4 -BQSR $sampleID.report.grp -o $sampleID.final.bam
java –jar GenomeAnalysisTK -T HaplotypeCaller -R $referencegenome -I $sampleID.final.bam --doNotRunPhysicalPhasing --emitRefConfidence GVCF --dbsnp dbsnp_138.b37.vcf -stand_call_conf 30 -stand_emit_conf 10 -gt_mode DISCOVERY -nct 4 -mbq 20 -G Standard -A AlleleBalance -o $sampleID.raw.snps.indels.g.vcf

Table A2. List of variant calling commands (cont).

Command List for Variant Calling
java –jar GenomeAnalysisTK -T GenotypeGVCFs -R $referencegenome --variant $sampleID.raw.snps.indels.g.vcf -o $sampleID.raw.snps.indels.vcf
java –jar GenomeAnalysisTK -T VariantAnnotator -R $referencegenome -o $sampleID.ann.snp.indel.vcf -A Coverage -A InbreedingCoeff --variant $sampleID.raw.snps.indels.vcf -L $sampleID.raw.snps.indels.vcf --dbsnp dbsnp_138.b37.vcf
java –jar GenomeAnalysisTK -T VariantRecalibrator -R $referencegenome -input $sampleID.ann.snp.indel.vcf -resource:hapmap,VCF,known=true,training=true,truth=true,prior=15.0 hapmap3.3.b37.vcf -resource:omni,VCF,known=true,training=true,truth=true,prior=12.0 1000Gomni2.5.b37.vcf -resource:dbsnp,VCF,known=true,training=true,truth=true,prior=6.0 dbsnp_138.b37.vcf -an QD -an MQRankSum -an ReadPosRankSum -an FS -an MQ -mode SNP -recalFile $sampleID.snp.recal -tranchesFile $sampleID.snp.tranches -rscriptFile $sampleID.snp.plots.R -nt 6 --maxGaussians 4 --TStranche 100.0 --TStranche 99.9 --TStranche 99.5 --TStranche 99.0 --TStranche 98.0 --TStranche 97.0 --TStranche 95.
java –jar GenomeAnalysisTK -T ApplyRecalibration -R $referencegenome -input $sampleID.ann.snp.indel.vcf --ts_filter_level 99.0 -recalFile $sampleID.snp.recal -tranchesFile $sampleID.snp.tranches -mode SNP -o $sampleID.snp.vqsr.vcf
java –jar GenomeAnalysisTK -T VariantRecalibrator -R $referencegenome -input $sampleID.snp.vqsr.vcf -resource:mills,known=true,training=true,truth=true,prior=12.0 Mills_and_1000G_gold_standard.indels.b37.vcf -resource:dbsnp,VCF,known=true,training=true,truth=true,prior=6.0 dbsnp_138.b37.vcf -an QD -an DP -an FS -an SOR -an MQRankSum -an ReadPosRankSum -mode INDEL -recalFile $sampleID.indel.recal -tranchesFile $sampleID.indel.tranches -rscriptFile $sampleID.indel.R
java –jar GenomeAnalysisTK -T ApplyRecalibration -R $referencegenome --input $sampleID.snp.vqsr.vcf -mode INDEL --ts_filter_level 99.0 -recalFile $sampleID.indel.recal -tranchesFile $sampleID.indel.tranches -o $sampleID.snp.indel.vqsr.vcf

APPENDIX B: PRIMER SEQUENCES USED IN VALIDATION EXPERIMENTS

Table B.1. List of primer sequences.

Primer Name	Melting Temperature Tm (°C)	Sequence (5' -> 3')
DNAJB2 E9F	55.0	GCAGTAATACCCCTGGCTCA
DNAJB2 E9R	57.1	CTTCCCACAGTGAGTCAGACC
C19ORF12 E3F	61.0	GTGGTGTGCACTCAGTGGG
C19ORF12 E3R	59.4	AACTCCCAAGCCACCTCTTC
C19ORF12 E2F	58.5	GGAAATACTCTTATGCTCATTGAAAC
C19ORF12 E2R	55.3	GTTTCAACGGCCCTTTTATG
IGHMBP2 E5F	67.8	GAGGAACACCCACAGCTCCCC
IGHMBP2 E5R	57.4	CTCTGACAGGGAAGTGGCAT
PLEKHG5 E15F	62.8	GAGGACGGGACCCTGGAC
PLEKHG5 E15R	59.4	AGCTTCAGGTCCAGGGTCAT
SLC12A6 E8F	53.3	TGCAAACGAATACAGCCTTT
SLC12A6 E8R	57.9	GGGCTTATCTGAGAGGGAAAA
TRPV4 E6F	60	CCAGAGAAACGTGCAGTTCA
TRPV4 E6R	59	TTCTTGAGCTGGGACATCTG
VCP E5F	57.9	GGGCAATATCTAATGAAGGGC
VCP E5R	59.8	ACTGGGATTACAGGTGTCAGC
ERBB4 E11F	59.7	ACAACGCCTTCTCTCCACAT
ERBB4 E11R	59.5	AATGGCGATCGTTTCTGAAT
LRSAM1 E9F	59	AAGGAAATCGTGTGGTCTCC
LRSAM1 E9R	59.8	TGTGGCCATTTCTGTCTCTTG
SQSTM1 F	63.2	CTCACCTAAGTGGCTGAATTTTGTG
SQSTM1 R	65.4	GGTGGGGGGTATCCTGAATTCTT

APPENDIX C: SEQUENCING ANALYSIS METRICS

Table C.1. Quality check metrics for all individuals.

Individual	Mean Depth of Coverage	FMISS	Ts/Tv
Individual 1	75	0.023322667	2,218
Individual 2		0.021693113	2,251
Individual 3		0.022062951	2,241
Individual 4	72	0.024421103	2,213
Individual 5		0.025474465	2,227
Individual 6		0.026355548	2,222
Individual 7		0.011645838	2,231
Individual 8	83	0.023816988	2,226
Individual 9		0.022609708	2,218
Individual 10		0.023157280	2,216
Individual 11		0.023766185	2,245
Individual 12		0.012300930	2,230
Individual 13	74	0.010362764	2,288
Individual 14		0.011821875	2,273
Individual 15		0.012331974	2,289
Individual 16	21	0.510000000	2,203
Individual 17	47	0.015864390	2,200
Individual 18		0.016211380	2,204
Individual 19		0.035441796	2,200
Individual 20		0.289311178	2,252
Individual 21	48	0.034528135	2,197
Individual 22		0.035702826	2,180
Individual 23		0.049291287	2,364
Individual 24		0.037206117	2,199
Individual 25		0.015557482	2,224
Individual 26	271	0.023767717	2,386
Individual 27		0.023076196	2,393
Individual 28		0.151164489	2,365
Individual 29		0.023964313	2,409
Individual 30	38	0.042149457	2,215
Individual 31		0.042149457	2,161
Individual 32		0.042149457	2,236
Individual 33		0.042149457	2,185
Individual 34	18	0.480000000	2,192
Individual 35	16	0.470000000	2,217
Individual 36	22	0.560000000	2,191
Individual 37	20	0.510000000	2,203

Table C.1. Quality check metrics for all individuals (cont).

Individual	Mean Depth of Coverage	FMISS	Ts/Tv
Individual 37	20	0.510000000	2,203
Individual 38	89	0.022716593	2,223
Individual 39		0.022302021	2,227
Individual 40		0.023369500	2,208
Individual 41		0.021414634	2,217
Individual 42	49	0.029874848	2,238
Individual 43	90	0.022169696	2,242
Individual 44		0.022052091	2,227
Individual 45		0.025430864	2,244
Individual 46		0.022851686	2,224
Individual 47	110	0.049034691	2,242
Individual 48		0.047815600	2,225
Individual 49	95	0.022796968	2,235
Individual 50		0.023590348	2,236
Individual 51		0.023287461	2,202
Individual 52		0.022005109	2,227
Individual 53		0.020540537	2,200
Individual 54	24	0.470000000	2,190
Individual 55	19	0.500000000	2,217
Individual 56	55	0.048298922	2,387
Individual 57		0.040624935	2,224
Individual 58		0.034993300	2,180
Individual 59	75	0.023084460	2,212
Individual 60		0.021986747	2,216
Individual 61		0.022062951	2,225
Individual 62	75	0.022042617	2,237
Individual 63		0.022104987	2,223
Individual 64		0.020033475	2,223
Individual 65	78	0.021497894	2,210
Individual 66		0.022127551	2,235
Individual 67		0.021617640	2,239
Individual 68		0.012316465	2,211
Individual 69	63	0.012076924	2,255
Individual 70		0.011821875	2,267
Individual 71		0.012331974	2,417

Table C.1. Quality check metrics for all individuals (cont).

Individual	Mean Depth of Coverage	FMISS	Ts/Tv
Individual 72	26	0.035051730	2,169
Individual 73		0.033538591	2,118
Individual 74		0.033538591	2,358
Individual 75		0.033538591	2,441
Individual 76	77	0.040415696	2,041
Individual 77		0.038921874	2,049
Individual 78		0.040385639	2,074
Individual 79		0.039553670	2,057
Individual 80	62	0.055214570	2,416
Individual 81		0.021021088	2,163
Individual 82		0.053610878	2,448
Individual 83		0.044922061	2,173
Individual 84	66	0.016428263	2,192
Individual 85		0.013993863	2,196
Individual 86		0.018289244	2,162
Individual 87		0.020029004	2,206
Individual 88		0.015672604	2,205
Individual 89	35	0.640000000	2,221
Individual 90		0.630000000	2,181
Individual 91	45	0.540000000	2,226
Individual 92	54	0.004785456	2,100
Individual 93		0.004860300	2,108
Individual 94		0.005989453	2,118
Individual 95		0.006164314	2,110
Individual 96		0.009358660	2,289
Individual 97		0.400000000	2,219
Individual 98	51	0.030192626	2,258
Individual 99	50	0.027379413	2,234
Individual 100	67	0.017573924	2,248
Individual 101		0.017573924	2,234
Individual 102	26	0.036951848	2,092
Individual 103		0.039639639	2,062
Individual 104		0.038560922	2,067
Individual 105		0.044698200	2,128
Individual 106	43	0.033201387	2,251
Individual 107	63	0.028831161	2,250
Individual 108		0.031320687	2,237
Individual 109	47	0.031608144	2,225
Individual 110	44	0.033544287	2,255
Individual 111	45	0.032516574	2,250

Table C.1. Quality check metrics for all individuals (cont).

Individual	Mean Depth of Coverage	FMISS	Ts/Tv
Individual 112	57	0.029876839	2,233
Individual 113	45	0.030151239	2,245
Individual 114		0.031501896	2,247
Individual 115	62	0.054659961	2,064
Individual 116		0.020909526	2,254
Individual 117		0.054659961	2,064
Individual 118	49	0.033152926	2,233
Individual 119		0.022103252	2,208
Individual 120	18	0.510000000	2,214
Individual 121	20	0.530000000	2,204
Individual 122	20	0.540000000	2,189
Individual 123	14	0.540000000	2,187
Individual 124	17	0.640000000	2,237
Individual 125	114	0.010142287	2,264
Individual 126		0.023752690	2,345
Individual 127		0.009994826	2,275
Individual 128		0.010824323	2,239
Individual 129		0.023173358	2,365
Individual 130		0.023246496	2,366
Individual 131		0.023757646	2,359
Individual 132	53	0.010480070	2,267
Individual 133		0.016407491	2,052
Individual 134		0.012776628	2,068
Individual 135	17	0.530000000	2,157
Individual 136	53	0.069388374	2,165
Individual 137		0.068770248	2,174
Individual 138		0.073888248	2,146
Individual 139		0.068272965	2,128
Individual 140		0.103599935	2,142
Individual 141		0.067593396	2,146
Individual 142	21	0.510000000	2,206
Individual 143	59	0.020621342	2,229
Individual 144		0.017396120	2,136
Individual 145		0.013716851	2,116
Individual 146		0.015516538	2,124
Individual 147	22	0.0147287	2,232