Complex Genetics and Disease Mechanisms in a Turkish Ataxia Cohort

by

Gülşah Şimşir

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Molecular Biology and Genetics

KOÇ
ÜNİVERSİTESİ

Complex Genetics and Disease Mechanisms in a Turkish Ataxia Cohort

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Gülşah Şimşir

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

_____

Prof. Dr. A. Nazlı Başak (Advisor)

_____

Prof. Dr. Sibel Ertan

_____

Prof. Dr. Sibel Aylin Uğur İşeri

Date:

To my family…

ABSTRACT

Complex Genetics and Disease Mechanisms in a Turkish Ataxia Cohort

Gülşah Şimşir

Master of Science in Molecular Biology and Genetics

Ataxias are a clinically, genetically, and mechanistically heterogeneous group of disorders, characterized by degeneration of the cerebellum. There are significant subtypes of ataxias which also overlap with other neurological disorders, making the precise diagnosis challenging. The developments in next generation sequencing in recent years contributed to the molecular diagnosis of ataxias. It became possible to sequence the coding regions of the genome in a time-saving and cost-effective manner by whole exome sequencing (WES). WES evolved to a powerful tool to identify genetic causes of complex ataxias. In the framework of this thesis, 83 index patients with complex ataxia phenotypes were investigated by WES. Mutations in 18 different genes were identified as the genetic cause in 26 families, which corresponds to a diagnostic yield of 31%. Identification of mutations in hereditary spastic paraplegias and other neurodegenerative disease genes showed that there is a significant overlap among different neurodegenerative disorders. The adaptation of rapidly improving sequencing approaches and detailed clinical information of the remaining patients will help to identify the disease genes in unsolved families. The results presented here complement the picture of the molecular basis of ataxias in Turkey. This will hopefully pave the ways for more precise diagnosis in future and will also contribute to the development of therapeutic approaches in ataxias.

ÖZETÇE

Türk Ataksi Kohortunda Komplex Genetik ve Hastalık Mekanizmaları

Gülşah Şimşir

Moleküler Biyoloji ve Genetik, Yüksek Lisans

Beyincikte dejenerasyon ile karakterize olan ataksiler, klinik, genetik ve mekanistik açıdan heterojen bir hastalık grubudur. Ataksi alt-türlerinin çok farklı olmasının yanında, ataksilerin diğer nörodejeneratif hastalıklarla da örtüşmesi ayırıcı tanıyı zorlaştırmaktadır. Yeni nesil dizileme tekniklerinde son yıllarda olan gelişmeler ataksilerin moleküler tanısına büyük katkıda bulunmuştur. Tüm ekzom dizileme ile genomun kodlayıcı bölgelerini düşük maliyetle ve kısa sürede analiz etmek mümkün hale gelmiştir. Bu tez çerçevesinde, kompleks ataksi fenotipi olan 83 indeks hasta tüm ekzom dizileme ile incelendi. Onsekiz farklı gende bulunan mutasyonlar, 26 ailenin hastalık nedeni olarak tanımlandı. Tez çerçevesinde tüm ekzom dizilemeden %31 oranında verim elde edildi. Literatüre uygun olan bu sonuç, tüm ekzom dizilemenin ayırıcı tanıda ne kadar kullanışlı bir araç olduğunu göstermektedir. Herediter spastik parapareziye ve diğer nörodejeneratif hastalıklara neden olan genlerde bulunan mutasyonlar, farklı nörodejeneratif hastalıklar arasındaki örtüşmeye işaret etti. Hızla gelişen dizileme tekniklerini benimsemek ve geri kalan hastaların daha detaylı klinik bilgilerine ulaşmak, çözülemeyen ailelerde hastalık genlerinin tanımlanmasına yardımcı olacaktır. Bu tezin sonuçlarının Türkiye'de ataksilerin moleküler temeline ışık tutmasının yanında, ayırıcı tanyı kolaylaştırması ve gelecekte ataksilerde tedavi yöntemlerinin geliştirilmesine katkıda bulunması umulmaktadır.

ACKNOWLEDGEMENTS

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

I would like to extend my thanks to Prof. Sibel Ertan and Prof. Sibel İşeri for devoting their time to evaluate my thesis.

I deeply thank all members of NDAL, especially Şeyma, for her friendship and for being a great research partner.

I thankfully acknowledge Suna and Inan Kıraç Foundation and Koç University-KUTTAM for their invaluable support.

Last but not least, I deeply thank my parents İpek and Mehmet Şimşir, my brother Tarık Şimşir and my sister-in-law İlknur Şimşir for their encouragement and for being supportive in every decision I had. Nothing would have been possible without them.

TABLE OF CONTENTS

Abstract iv

Özetçe v

Acknowledgements vi

List of Tables x

List of Figures xi

Abbreviations xii

Chapter 1: Introduction 1

1.1 Ataxias 1

1.1.1 Autosomal Dominant Cerebellar Ataxias 2

1.1.2 Autosomal Recessive Cerebellar Ataxias 4

Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay 5

Ataxias with Oculomotor Apraxia 6

1.2 Differential Diagnosis of ND with Ataxic Features 8

1.2.1 Next Generation Sequencing 8

1.2.2 Whole Exome Sequencing 9

Data Analysis and Bioinformatics 9

Limitations 10

1.2.3 Molecular Analysis of Ataxias by WES: Ataxias in Turkey 11

Chapter 2: Purpose 12

Chapter 3: Materials 13

3.1 Subjects 13

3.1.1 Family Trees 17

Families with AD Inheritance 17

Families with Autosomal Recessive (AR) Inheritance 19

Families with Sporadic Inheritance 21

3.2 DNA Isolation 22
3.3 Whole Exome Sequencing Platforms 22
3.4 Validation Experiments 22
3.4.1 Agarose Gel Electrophoresis and Extraction from the Gel 25
3.4.2 Laboratory Equipment and Kits 25
3.5 Online Databases and Bioinformatic Tools 27
Chapter 4: Methods 29
4.1 DNA Isolation 29
4.2 Whole Exome Sequencing 29
4.3 Whole Exome Sequencing Data Analysis 31
4.3.1 Variant Prioritization 31
4.4 PCR Experiments for Validation of Variants 31
4.4.1 PCR Conditions 31
4.4.2 Agarose Gel Electrophoresis 32
4.4.3 Purification of PCR Products before Sanger Sequencing 32
4.5 Restriction Enzyme Digestion 32
Chapter 5: Results 33
5.1 Whole Exome Sequencing Analysis 33
Chapter 6: Discussion 49
6.1 Whole Exome Sequencing and the Complex Genetic Architecture of the Cohort…… 49
6.1.1 Genes in Mitochondrial Metabolism 49
6.1.2 Genes in DNA Repair Pathway 50
6.1.3 Genes in Lipid Metabolism 51
6.1.4 Genes in Autophagy and Lysosomal Activity 51
6.1.5 Other Genes with Different Mechanisms 52
6.2 Variants of Uncertain Significance 53

6.3 Novel Genotype-Phenotype Associations 54
6.4 Limitations of Whole Exome Sequencing 54
6.5 Remaining Cases to be Solved 56
6.6 WES is the Golden Standard in Diagnosis of Ataxias 56
Chapter 7: Conclusion 58
BIBLIOGRAPHY 59
Appendix A 65

LIST OF TABLES

Table 1.1 Autosomal Dominant Cerebellar Ataxias. 3
Table 1.2 Examples of Autosomal Recessive Cerebellar Ataxias. 5
Table 3.1 Characteristics of families investigated in this thesis. 14
Table 3.2 Primers used in validation experiments. 22
Table 3.3 PCR reagents for amplification. 24
Table 3.4 Conditions for PCR amplification. 25
Table 3.5 Chemicals used in gel electrophoresis. 25
Table 3.6 Laboratory equipment. 26
Table 3.7 Commercially available kits, other than exome capture kits. 27
Table 3.8 Open-source bioinformatics tools and online databases 27
Table 5.1 List of all causative variants and associated diseases in OMIM in the families solved. 35
Table 5.2 Online prediction tool scores and frequencies of the variants. 38

LIST OF FIGURES

Figure 3.1 Family 30, Family 59, Family 67, and Family 41 showing an autosomal dominant inheritance pattern with several affected individuals 18

Figure 3.2 Pedigrees of families with an AR inheritance 20

Figure 3.3 Family 50, Family 51, Family 52 showing sporadic inheritance pattern with only one affected individual and no consanguinity. 21

Figure 4.1 Steps of whole exome sequencing 30

Figure 4.2 The workflow for whole exome sequencing 30

Figure 5.1 Genes identified in this study giving rise to ataxias. 34

Figure 5.2 31% of patients was solved as opposed to 69% unsolved in the cohort under study. 34

Figure 5.3 CAPN1 families 42

Figure 5.4 ARSACS families 43

Figure 5.5 SPG7 and SPG11 families. 44

Figure 5.6 HSD17B4-fam 3, HSD17B4-fam 20. 45

Figure 5.7 ATM-fam 28, KIF1C-fam 9, VPS13D-fam 7, PNPLA6-fam 2, SYNE1-fam 1, L2HGDH-fam 22. 46

Figure 5.8 MCOLN1-fam 11, MFSD8-fam 4, NKX6-2-fam 25, NPC1-fam 6, SYT14-fam 24, CACNA1A-fam 19, STUB1-fam 21. 47

Figure 5.9 Agarose gel electrophoresis of the PCR products of each exon 48

Figure 6.1 Mutations in the sacsin protein are shown in red. UBL 50

Figure 6.2 Mutations described in calpain-1 in the cohort shown in red 52

ABBREVIATIONS

AD	Autosomal Dominant
ADCAs	Autosomal Dominant Cerebellar Ataxias
AFG3L2	ATPase Family Gene 3-like 2
AFP	Alpha-Fetoprotein
ALS	Amyotrophic Lateral Sclerosis
AO	Age of Onset
AOA	Ataxias with Oculomotor Apraxia
APTX	Aprataxin
AR	Autosomal Recessive
ARCAs	Autosomal Recessive Cerebellar Ataxias
ARSACS	Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay
AT	Ataxia Telangiectasia
ATLD	AT like Disorder
ATM	Ataxia-Telangiectasia Mutated Gene
ATN1	Atrophin 1
ATX	Ataxia
ATXN	Ataxin
B	Benign
BEAN1	Brain-Expressed, Associated with Nedd4, 1
CACNA1A	Calcium Channel, Voltage-Dependent, P/Q Type, Alpha-1a Subunit
CADD	Combined Annotation-Dependent Depletion
CAPN1	Calpain 1
CHIP-seq	Chromatin Immunoprecipitation Sequencing
CLN7	Ceroid Lipofuscinosis, Neuronal, 7
CMT	Charcot-Marie-Tooth
CNS	Central Nervous System
CNVs	Copy Number Variations
Conc	Concentration
D	Damaging
DNA	Deoxyribonucleic Acid

EA	Episodic Ataxia
ExAC	Exome Aggregation Consortium
F	Female
FA	Friedreich's Ataxia
Fam	Family
FGF14	Fibroblast Growth Factor 14
FXN	Frataxin
GERP	Genomic Evolutionary Rate Profiling
gnomAD	The Genome Aggregation Database
het	Heterozygous
HGP	Human Genome Project
hom	Homozygous
HSD17B4	17-Beta-Hydroxysteroid Dehydrogenase Iv
HSP	Hereditary Spastic Paraplegias
IGV	Integrative Genomics Viewer
ITPR1	Inositol 1,4,5-Triphosphate Receptor, Type 1
KCNC3	Potassium Channel, Voltage-Gated, Shaw-Related Subfamily, Member 3
KCND3	Potassium Voltage-Gated Channel, Shal-Related Subfamily, Member 3
KIF1C	Kinesin Family Member 1c
L2HGA	L-2-Hydroxyglutaric Aciduria
L2HGDH	L-2-Hydroxyglutarate Dehydrogenase
LP	Likely Pathogenic
M	Male
MCOLN1	Mucolipin 1
MFSD8	Major Facilitator Superfamily Domain-Containing Protein 8
ML4	Mucolipidosis IV
MRIs	Magnetic Resonance Images
n	Novel
n.a	Not Available
ND	Neurodegenerative Diseases
NGS	Next Generation Sequencing
NKX6-2	NK6 Homeobox 2
NOP56	NOP56 Ribonuclear Protein

NPC1	NPC Intracellular Cholesterol Transporter 1
OMA	Oculomotor Apraxia
P	Patient
Pat	Pathogenic
PCR	Polymerase Chain Reaction
PD	Parkinson’s Disease
PNPLA6	Patatin-Like Phospholipase Domain-Containing Protein 6
PNS	Peripheral Nervous System
PPP2R2B	Protein Phosphatase 2, Regulatory Subunit B, Beta
PRKCG	Protein Kinase C, Gamma
PRLTS1	Perrault Syndrome 1
REVEL	Rare Exome Variant Ensemble Learner
RFC1	Replication Factor C, Subunit 1
RNA	Ribonucleic Acid
SACS	Sacsin
SCAs	Spinocerebellar Ataxias
SCAR	Spinocerebellar Ataxia, Autosomal Recessive
SETX	Senataxin
SIFT	Sorting Intolerant from Tolerant
SMA	Spinal Muscular Atrophy
SPAX	Spastic Ataxia
SPG	Spastic Paraplegia
SPG7	Paraplegin
SPG11	Spatacsin
SPTBN2	Spectrin, Beta, Nonerythrocytic, 2
STUB1	Stip1 Homologous and U Box-Containing Protein 1
SYNE1	Spectrin Repeat-Containing Nuclear Envelope Protein 1
SYT14	Synaptotagmin 14
SV	Structural Variations
T	Tolerated
TBP	TATA Box-Binding Protein-Like Protein 1
TGM6	Transglutaminase 6
Tm	Melting Temperature

TNR	Trinucleotide Repeat
TTBK2	Tau Tubulin Kinase 2
VPS13D	Vacuolar Protein Sorting 13 Homolog D
VUS	Variants of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Chapter 1:
INTRODUCTION

Neurodegenerative diseases (ND) are a heterogeneous group of disorders characterized by progressive neuronal loss in the central nervous system (CNS) or peripheral nervous systems (PNS). The most prevalent among ND are Alzheimer's disease and Parkinson's Disease (PD). ND are classified based on predominant clinical phenotype and topography of predominant lesions. ND in the CNS can be categorized into four groups according to the affected region: i) cerebral cortex, ii) basal ganglia, iii) spinal cord, and iv) cerebellum. Cerebral cortex diseases are associated with dementia such as Alzheimer's disease, whereas basal ganglia diseases are delineated by movement disorders including PD. The spinal cord is affected in diseases such as amyotrophic lateral sclerosis (ALS) and spinal muscular atrophy (SMA). Cerebellar disorders, the topic of this thesis, are complex because of the overlap among the pathological conditions in which the cerebellum is damaged (Przedborski et al., 2003). All cerebellar diseases present with ataxia and are sometimes accompanied by various other neuropathological conditions (Koeppen, 2018). Complex cases in which more than one CNS region is affected may occur because of the patients' genetic background or environmental factors, which make the border among different diseases blurry.

1.1 Ataxias

Ataxia means "loss of order" in the Greek language. Clinically, the term ataxia is used to characterize a group of ND in which the predominant clinical phenotype is ataxia. Cerebellar degeneration results in a variety of symptoms depending on the location of neurodegeneration. These symptoms usually include broad-based gait, incoordination, lack of balance, dysarthria, and impaired eye movement.

Ataxias are a heterogenous group of disorders which may stem from both genetic and non-genetic origins. According to current etiology-based classifications, ataxias can be i) acquired, ii) sporadic or iii) hereditary. Cerebellum damage in acquired ataxias can be caused by excessive alcohol consumption, infections, surgery, vitamin deficiencies, tumors, and exposure to toxins. In sporadic ataxias, no acquired or genetic cause can be determined; these are usually adult-onset ataxias with no apparent family history and a complex inheritance. On the other hand, in hereditary ataxias, early-onset and positive

family history are considered typical properties; however, hereditary ataxias may occur as sporadic disorders with adult-onset. Genetically they can be divided into four groups: Autosomal dominant (AD), autosomal recessive (AR), X-linked, and mitochondrial disorders (Klockgether, 2010) (Jones et al., 2014).

1.1.1 Autosomal Dominant Cerebellar Ataxias

The autosomal dominant cerebellar ataxias (ADCAs), currently known as spinocerebellar ataxias (SCAs), are a heterogeneous group of degenerative disorders that are characterized by dysfunction of the cerebellum, the brainstem, and their associated pathways. Age of onset (AO) is often in adulthood. All SCA patients show cerebellar ataxia. Other symptoms may include extrapyramidal features, seizures, neuropathy, or cognitive impairment although these are disease-specific. The original classification of SCAs was proposed by Harding in 1993 who divided them into three groups. The first group, ADCA type I, includes SCAs with dysfunction of CNS and/or PNS. SCAs 1,2, and 3 are in this group. The second group, ADCA type II, is associated with pigmentary retinopathy and only SCA7 belongs to this group. The third group, ADCA type III, is identified as pure ataxia of which SCA6 is the most frequent (Mundwiler and Shakkottai, 2018).

There are more than 40 genetically distinct subtypes of SCAs, and they are numbered in chronological order when the causative gene of the subtype is identified. SCAs fall into two major groups: repeat expansion SCAs and non-repeat mutations as seen in Table 1.1, adapted from (Durr, 2010); (Jayadev and Bird, 2013), and (Klockgether et al., 2019). The most common types of repeat expansion SCAs are caused by CAG repeats, thus belonging to the “polyglutamine disorders” group. Expansion sizes causing the disease phenotype among SCAs are different from each other and the size affects the severity of the phenotype. Longer repeats cause a more severe phenotype with earlier AO. The size of the repeat can change between generations (Klockgether et al., 2019).

Table 1.1 Autosomal Dominant Cerebellar Ataxias.

Disease	**Gene**	**Average AO (decade)**	**Clinical phenotype**
Polyglutamine expansions			
SCA1	ATXN1	3-4	Spasticity, ophthalmoplegia, bulbar and sensory symptoms
SCA2	ATXN2	3 – 4	Slow saccades and sensory symptoms
SCA3	ATXN3	4	Spasticity, basal ganglia symptoms, sensory symptoms, amyotrophy including facial atrophy, and fasciculations
SCA6	CACNA1A	5 – 6	Pure cerebellar ataxia and downbeat nystagmus
SCA7	ATXN7	3 – 4	Visual loss, ophthalmoplegia, and spasticity
SCA17	TBP	4	Spasticity, basal ganglia symptoms, psychiatric disorders, and dementia
DRPLA	ATN1	4	chorea, seizures, dementia, myoclonus; often confused with Huntington disease
Non-coding expansions			
SCA8	ATXN8	4	Spasticity, sensory symptoms, cognitive and mood changes
SCA10	ATXN10	4	Epilepsy
SCA12	PPP2R2B	4	Tremor
SCA31	BEAN1	5 – 6	Pure cerebellar ataxia
SCA36	NOP56	5	Amyotrophy and hearing loss
Conventional SCAs			
SCA5	SPTBN2	3-4	Pure cerebellar ataxia
SCA11	TTBK2	3	Mild, remain ambulatory
SCA13	KCNC3	Childhood or adulthood	Intellectual disability
SCA14	PRKCG	3-4	Myoclonus
SCA15/16	ITPR1	4	Pure cerebellar ataxia

SCA19/22	KCND3	4	Slowly progressive, rare cognitive impairment, myoclonus, hyperreflexia
SCA27	FGF14	2	Early-onset tremor; dyskinesia, cognitive deficits
SCA28	AFG3L2	2	Spasticity, ophthalmoplegia, and ptosis
SCA35	TGM6	4	Hyperreflexia, Babinski responses, spasmodic torticollis

1.1.2 Autosomal Recessive Cerebellar Ataxias

Autosomal recessive cerebellar ataxias (ARCAs) comprise a heterogenous group of rare ND with a prevalence of 3-6/100.000. In contrast to ADCAs, they usually start early in life. The hallmark of ARCAs is progressive damage to the cerebellum and/or its associated tracts. ARCAs usually present as sporadic cases and include various other neurologic or non-neurologic systems (Synofzik and Németh, 2018). Other than cerebellar ataxia, patients may have peripheral neuropathy, chorea, dystonia, oculomotor abnormalities, spasticity, cognitive impairment or epilepsy (Anheim et al., 2012). This clinical heterogeneity among ARCAs make their clinical recognition and diagnosis challenging (Vermeer et al., 2011).

ARCAs can be classified based on their clinical phenotype, neuropathology, or mechanism of pathogenesis. Possible disease mechanisms are mitochondrial dysfunction, DNA repair defects, protein misfolding and chaperone dysfunction, mislocalization of synaptic myonuclei, calcium mediated chloride channel dysfunction, and complex lipid metabolism (Vermeer et al., 2011) (Synofzik et al., 2019). Even though specific subtypes have certain causative genes, several genes play a role in the same mechanism, causing overlapping phenotypes. Some examples of ARCAs can be seen in Table 1.2. adapted from (Synofzik et al., 2019); (Cortese et al., 2019), (Jayadev and Bird, 2013), and (Anheim et al., 2012).

Table 1.2 Examples of Autosomal Recessive Cerebellar Ataxias.

Disease	**Gene**	**Average AO (decade)**	**Clinical phenotype**
Repeat expansions			
Friedreich's ataxia (FA)	FXN	1-2	Sensory ataxia, sensory axonal neuropathy, pyramidal weakness, cardiomyopathy, diabetes
CANVAS	RFC1	5	Cerebellar ataxia, neuropathy, vestibular areflexia
Non-repeat ARCAs			
ARSACS	SACS	childhood	Cerebellar ataxia, pyramidal tract damage, axonal demyelinating sensorimotor neuropathy
ARCA1	SYNE1	3	Pure cerebellar ataxia
AOA2	SETX	childhood	Oculomotor apraxia, chorea, dystonia, cognitive impairment, axonal sensorimotor neuropathy
AOA1	APTX	childhood	Oculomotor apraxia, chorea, dystonia, cognitive impairment, axonal sensorimotor neuropathy
AT	ATM	childhood	Oculomotor apraxia, choreo-athetosis, dystonia, cognitive impairment, sensorimotor neuropathy, telangiectasias, immunodeficiency, cancer susceptibility

Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay

Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) is a distinct type of early-onset spastic ataxia, which was identified first in Quebec, Canada (Bouchard et al., 1978). In this region, the carrier rate is 1/22 with a prevalence of 1/484. Initially, it was thought that ARSACS is restricted to the Quebec region, however, after identifying the *SACS* gene, it became evident that ARSACS is not limited to Quebec. ARSACS patients usually show a triad of lower limb spasticity, early-onset cerebellar ataxia and peripheral neuropathy. However, there are patients showing atypical phenotypes such as

Charcot–Marie–Tooth-like disorders or epilepsy; therefore, ARSACS may be underdiagnosed (Synofzik and Németh, 2018).

ARSACS is caused by homozygous or compound heterozygous mutations in the *SACS* gene which encodes the "sacsin" protein (Engert et al., 2000). Sacsin is most highly expressed in brain motor systems, including cerebellar Purkinje cells. The protein has a ubiquitin-like (UbL) domain which interacts with proteasome, and a J domain, suggesting that it has a role in the ubiquitin-proteasome pathway. It is also contained to the region homologous to the Hsp90 protein, suggesting a role in protein folding mechanisms (Parfitt et al., 2009). Loss of function of the protein is considered to be a disease mechanism because of its chaperon-mediated protein folding role. Most of the mutations in the *SACS* gene, causing the disease phenotype, are frameshift or truncating mutations, disrupting the protein's working mechanism (Bereznyakova and Dupré, 2018).

Ataxias with Oculomotor Apraxia

Oculomotor apraxia is a condition which is described as the absence or defect of controlled, voluntary, and purposeful eye movement, caused by difficulties in saccade initiation and impaired cancellation of the vestibule-ocular reflex. This results in saccadic hypometria with a typical staircase pattern (Subramony et al., 2015). Ataxias with oculomotor apraxia (AOA) are a group of ARCAs occurring due to DNA repair damage, RNA termination or maturation problems. They include ataxia with oculomotor apraxia type 1 (AOA1) due to mutations in *APTX*, ataxia with oculomotor apraxia type 2 (AOA2) caused by mutations in *SETX*, ataxia telangiectasia (AT) due to mutations in *ATM*, and the rarer AT-like disorder (ATLD) linked to mutations in *MRE11* (Mariani et al., 2017).

Ataxia telangiectasia is an early-onset cerebellar ataxia with oculomotor apraxia, immunodeficiency, choreoathetosis, myoclonic jerks, dystonia, sensitivity to ionizing radiation, and an increased risk for malignancy. Telangiectasia is a condition characterized by small, dilated blood vessels near the surface of the skin. The classic AT phenotype usually starts before the age of 5, but older ages of onset have been observed. Serum concentration of alpha-fetoprotein (AFP) is elevated above 10 ng/mL in about 95% of individuals with AT, hence providing a blood-based biomarker for identifying AT. Cerebellar atrophy of the frontal and posterior vermis and both hemispheres are seen in magnetic resonance images (MRIs) of AT patients. AT accounts for approximately 3-5% of all ARCAs. It is caused by the *ATM* gene which has a role in the double-strand DNA repair mechanism (Synofzik and Németh, 2018).

Ataxia telangiectasia-like disorder is a very similar disorder to AT, however, unlike AT, ATLD patients do not have telangiectasia. ATLD is characterized by later onset neurological symptoms with slower progression when compared to AT. Cancer inclination is not observed in ATLD patients (Taylor et al., 2004). Mutations in the *MRE11* gene, which is a part of hMre11/hRad50/Nbs1 complex that acts as a sensor of DNA damage, are the causes of this disease (Stewart et al., 1999). MRE11 deficiency, together with oxidative stress causing DNA damage, is the suspected mechanism of selective cerebellar damage and ATLD disease pathology (Oba et al., 2010).

AOA1 is an early-onset cerebellar ataxia with hypometric horizontal saccades, sensorimotor neuropathy, and common intellectual disability. Not all AOA1 patients have oculomotor apraxia (OMA) (Mariani et al., 2017). It has a frequency of approximately 2-5% in all ARCAs. It is caused by biallelic mutations in the *APTX* gene which codes for the "aprataxin" protein. Aprataxin plays a role in facilitating the repair of damage in both single and double-strand DNA. AOA1 patients with truncating *APTX* mutations are more prone to have OMA and cognitive impairment than those with missense mutations. The mean age of onset is 6.8 years; however, AOA1 can also occur later in life. The onset may be dependent on the type of mutation. Patients with truncating mutations have earlier onset than those with missense mutations (Synofzik and Németh, 2018).

AOA2 has a frequency of about 3% in all ARCAs. It presents with cerebellar ataxia, sensorimotor neuropathy, OMA and additional hyperkinetic movement disorders like chorea or dystonia. AFP is elevated in AOA2 patients, and it is a convenient biomarker. AOA2 starts early with onset ages between 7 and 25, but there are cases with AO of 40. It is caused by biallelic mutations in the *SETX* gene. It encodes for the protein "senataxin" which act as DNA/RNA helicase in single-strand DNA repair mechanism. In contrast to AT, there is no evidence of increased sensitivity to ionizing radiation or susceptibility to cancer in AOA2 (Synofzik and Németh, 2018).

Heterozygous mutations in *SETX* cause an autosomal dominant juvenile form of amyotrophic lateral sclerosis type 4 (ALS4). It was shown that distinct types of mutations change the gene expression differentially in both AOA2 and ALS4, suggesting alternate pathways to disease pathogenesis (Fogel et al., 2014).

1.2 Differential Diagnosis of ND with Ataxic Features

Neurologic disorders may have a significant number of overlapping symptoms even though they originate from different sites of the nervous system. This makes accurate clinical diagnoses difficult. The classification and diagnoses of genetic ataxias are also difficult because cerebellar ataxia may be a symptom of many neurological diseases. Motor neuron disorders, neuropathies or metabolic disorders may have ataxic features. Although ataxia is a presenting symptom in degenerative disorders that target the cerebellum, it may be seen in hereditary spastic paraplegias (HSP), Charcot-Marie-Tooth (CMT) disease or various encephalopathies (Beaudin et al., 2019). Clinicians have difficulty diagnosing the disease when ataxia and pyramidal signs/spasticity are present. Both ataxia and spasticity share a genetic background which results in "spastic ataxia" phenotypes such as ARSACS, hereditary spastic paraplegia type 7 (SPG7), spastic ataxias type 1-5, FA and SCA3 (de Bot et al., 2012). Among these, SPG7 seems to be the most common gene in patients with undiagnosed ataxia (Choquet et al., 2016). Therefore, before making the final diagnosis, genetic testing and continuous follow-up of the patients are required to distinguish between the different clinical phenotypes.

1.2.1 Next Generation Sequencing

The Human Genome Project (HGP) ushered in a new age in biological sciences. HGP took more than 10 years and was completed in 2003, paving the ways to technological developments in sequencing technologies. New massively parallel sequencing technologies have been devised by many companies to decrease cost and time while increasing accuracy. These so-called next generation sequencing (NGS) techniques have altered the understanding of genetics drastically (Slatko et al., 2018).

NGS is a method in which millions of DNA fragments are simultaneously sequenced and it has been used in clinical laboratories because of its ability to analyze several genes or genomic regions with a single technique compared to traditional methods. NGS technology has various applications depending on the purpose: Targeted panel sequencing, whole exome sequencing, whole genome sequencing (WGS), transcriptome sequencing, and CHIP-seq (chromatin immunoprecipitation sequencing) (Slatko et al., 2018).

1.2.2 Whole Exome Sequencing

Exons are the protein coding regions in the genome, and they constitute approximately 1% of the whole genome. There are 180.000 exons which are collectively called the exome. The rest of the genome contains repetitive regions, introns, and intergenic regions which do not have any direct effect on protein sequence, but they may affect the gene expression. The exome contains mutations of 85% of all known disorders. Therefore, it is logical to use WES for diagnostic purposes, because it focuses on the exome with a higher probability to cause a disease, and it means working with less complex data as compared to the whole genome (Marian, 2014; Rabbani et al., 2014).

WES is an NGS application technique used to identify the variations in all coding regions by enriching them. WES covers 95% of the exons, and more than 150 genes were discovered by WES (Rabbani et al., 2014). WES is a promising tool in the identification of single-gene disorders due to its high-throughput technology and low cost. The genetic cause of Miller syndrome was discovered by WES and this was the first successful application of exome sequencing to identify a gene for a Mendelian disorder (Ng et al., 2010). This study shows the power of the technology in discovering the genes related to Mendelian disorders by using very few affected individuals.

WES consists of two parts: Bench-work and *in silico* lab. In bench-work, wet-lab, library preparation to enrich the exons and sequencing take place. *In silico* lab is where the data analysis is performed. Data analysis includes aligning of the reads to a reference genome and variant calling followed by annotation of the variants. The annotated variants are filtered by the defined strategy for the study (Yohe and Thyagarajan, 2017).

Data Analysis and Bioinformatics

WES requires a proper design starting from the sample selection to identifying the causative variant behind the disease phenotype. All family pedigrees must be examined carefully. An individual may seem unaffected, but this may be misleading for adult-onset dominant disorders. Thus, it is important to have a significant amount of information about all individuals in the family. The family-based approach is helpful when a number of affected and unaffected family members are available. The affected individuals are likely to have the same mutation which decreases the number of candidate genes and unaffected individuals are useful to validate those mutations.

The choice of the platform on which WES is performed is important because it defines the limits of sensitivity and accuracy of any type of the variations received from the data (Yohe and Thyagarajan, 2017). There are commercially available kits that cover different regions and have different efficacy levels. The depth of coverage, which means how many times one nucleotide is sequenced, should be around 100X or higher for medical sequencing. While increasing coverage raises the cost, it is important to increase coverage because doing so decreases the number of false positives (Marian, 2014). Other parameters among different platforms that must be considered are read length and methods. For example, the long-read platforms should be used for highly repetitive regions, and short read platforms should be used for large scale genome variation studies and clinical applications (van Dijk et al., 2018).

Many bioinformatic analysis tools and software are used to analyze NGS data, and each have their own advantages and limitations depending on the purpose. It is important to carefully select which program and reference genome version to use. The computational pipeline includes: the alignment of raw reads to the reference genome, variant calling, functional annotation and prioritization of the variants (Foo et al., 2012). In our lab, these steps are performed by SEQ Platform of Genomize, Turkey (https://genomize.com/seq/). SEQ is an organizational platform on genomic data management with features such as storage, sharing, and usage.

Interpretation of the variants generated by SEQ takes place by using public databases and pathogenicity scoring tools. The public databases are The Exome Variant Server and 1000 Genomes Project, which include 6503 exomes and 2504 individuals, respectively. A revolutionary breakthrough created by the Exome Aggregation Consortium (ExAC) is a dataset which contains 60,706 individuals. ExAC provides increased resolution of low-frequency variations and is a public resource for clinical interpretations of genetic variants (Lek et al., 2016). ExAc data is now available in the The Genome Aggregation Database (gnomAD) browser (https://gnomad.broadinstitute.org/). As for the pathogenicity scoring tools, DANN, GERP++, SIFT, MutationTaster, REVEL, and MetalR are used.

Limitations

Despite its great advantages, WES also has some technical limitations both in wet-lab and *in silico* steps. One major drawback is that WES has short reads while genomes often contain numerous repeated sequences longer than WES reads, which may cause mis-assemblies and gaps. In addition, larger structural variations (SV) including copy

number variations (CNVs), large deletions and translocations are more difficult to analyze. Several approaches have been developed to call SV; however, they are not yet efficient (van Dijk et al., 2018).

Implementing WES is promising because it covers all coding regions in the genome, but there are still gaps and uncertainties in human genome sequences because the annotation of the exome has not yet been completed (Coffey et al., 2011). This incomplete annotation results in missing parts in exome sequencing kits, which, in turn, lead to missing regions.

1.2.3 Molecular Analysis of Ataxias by WES: Ataxias in Turkey

Genetics of ND, especially hereditary ataxia, has been evolving with the developments in NGS technology and the number of genes giving rise to ataxias has increased drastically. The introduction of NGS techniques, especially WES, into research and diagnostic laboratories has helped in the understanding of complex molecular bases of ataxias (Fogel et al., 2016).

Turkey is a very dynamic country with its young population and high birth rate (15/1000). Consanguineous marriage is very common in Turkey, especially in eastern provinces. Until very recently, the Turkish population was underrepresented in current databases, however, a recent publication from our laboratory showed the frequency and distribution of ataxias in Turkey. (Vural, et al., 2021).

Chapter 2**:**

PURPOSE

Ataxias constitute the majority of cerebellum diseases, and they overlap not only with each other, but also with other neurological conditions making their diagnosis challenging. Although most ataxia types do not have any available treatment yet, there are some subtypes with possible treatments, providing patients with a better life quality. Hence, having a definitive genetic diagnosis is crucial for the patients.

Turkey is a large country with high birth and consanguinity rates and a large ethnic heterogeneity. Thus, Turkish ataxia patients harbor potential mutations in several genes involving ataxia pathogenesis. Early-onset recessive ataxias are common in Turkey in addition to the late-onset dominant forms.

The aim of this thesis is to investigate the complex genetic structure of ataxias in a Turkish cohort comprising 83 index patients by using WES and bioinformatic analysis tools. Understanding the complicated genetic mechanisms behind ARCAs will help to pave the ways to successful therapies.

Chapter 3:
MATERIALS

3.1 Subjects

In the framework of this thesis, 83 families (index patients) harboring 124 patients referred to our laboratory with the initial clinical diagnosis of ataxia were investigated. Inheritance was categorized into three classes: If there were patients in upper generations, autosomal dominant (AD) inheritance pattern was considered. Autosomal recessive (AR) inheritance pattern was assumed in case of consanguinity between parents and/or if they were from same or closely neighboring villages, and if there were family members affected in the same generation. If there was no consanguinity between parents and no other affected family members were known, sporadic inheritance pattern was considered. AD inheritance was observed in 11 out of 83 families, 57 had AR inheritance pattern, and 15 have sporadic inheritance pattern. Consanguinity was present in 59 of 83 families (Figures 3.1-3.3). The mean age at onset of all patients with reliable information (88 patients) was 27.7 with a standard deviation of 21.5, ranging from infancy to 74 years of age (Table 3.1).

Clinical evaluation of the index cases was performed by expert clinicians in several different hospitals throughout Turkey. Informed consent was obtained from all participants in this study. Peripheral blood samples were collected into EDTA-containing tubes.

The patients were screened for trinucleotide repeat (TNR) disorders according to the referring clinicians' initial diagnosis and the inheritance pattern obtained from the pedigree. After excluding TNR diseases, the disease history of the family was further questioned, and extensive clinical phenotypes of the patients were obtained from clinicians.

Table 3.1 Characteristics of families investigated in this thesis.

Family ID	ID	Gender	Consanguinity	AO	Phenotype	Inheritance
Family 1	P1	F	+	37	SPAX	AR
Family 2	P2	F	+	1	SPAX	AR
	P3	F	+	1	SPAX	
	P4	F	+	1	SPAX	
Family 3	P5	M	+	17	ATX	AR
	P6	F	+	n.a	ATX	
Family 4	P7	M	+	9	ATX	AR
Family 5	P8	M	+	43	ATX	AR
Family 6	P9	F	+	9	ATX	AR
Family 7	P10	M	+	49	SPAX	AR
	P11	M	+	57	SPAX	
Family 8	P12	M	+	congenital	SPAX	AR
Family 9	P13	M	+	7	SPAX	AR
	P14	M	+	n.a	SPAX	
	P15	M	+	n.a	SPAX	
Family 10	P16	F	+	21	SPAX	AR
Family 11	P17	M	+	2	ATX	AR
	P18	F	+	1	ATX	
Family 12	P19	F	+	7	SPAX	AR
Family 13	P20	M	+	31	SPAX	AR
	P21	M	+	n.a	SPAX	
Family 14	P22	M	+	39	ATX	AR
Family 15	P23	M	+	14	SPAX	AR
Family 16	P24	F	+	1	SPAX	AR
	P25	F	+	n.a	SPAX	
	P26	M	+	n.a	SPAX	
	P27	F	+	n.a	SPAX	
Family 17	P28	M	+	childhood	SPAX	AR
Family 18	P29	M	+	41	SPAX	AR
Family 19	P30	M	-	67	EA	AD
Family 20	P31	M	+	6	SPAX	AR
Family 21	P32	M	-	36	ATX	Sporadic
Family 22	P33	M	+	7	ATX	AR
	P34	F	+	n.a	ATX	
Family 23	P35	M	+	37	SPAX	AR
Family 24	P36	F	+	childhood	ATX	AR
	P37	F	+	n.a	ATX	
Family 25	P38	M	+	3	SPAX	AR
Family 26	P39	F	+	36	SPAX	AR
	P40	M	+	32	SPAX	
Family 27	P41	M	-	11	ATX	Sporadic
Family 28	P42	M	+	1	ATX	AR
Family 29	P43	F	+	46	SPAX	AR
	P44	F	+	n.a	SPAX	
Family 30	P45	M	+	44	HSP	AD
	P46	M	n.a	59	HSP	

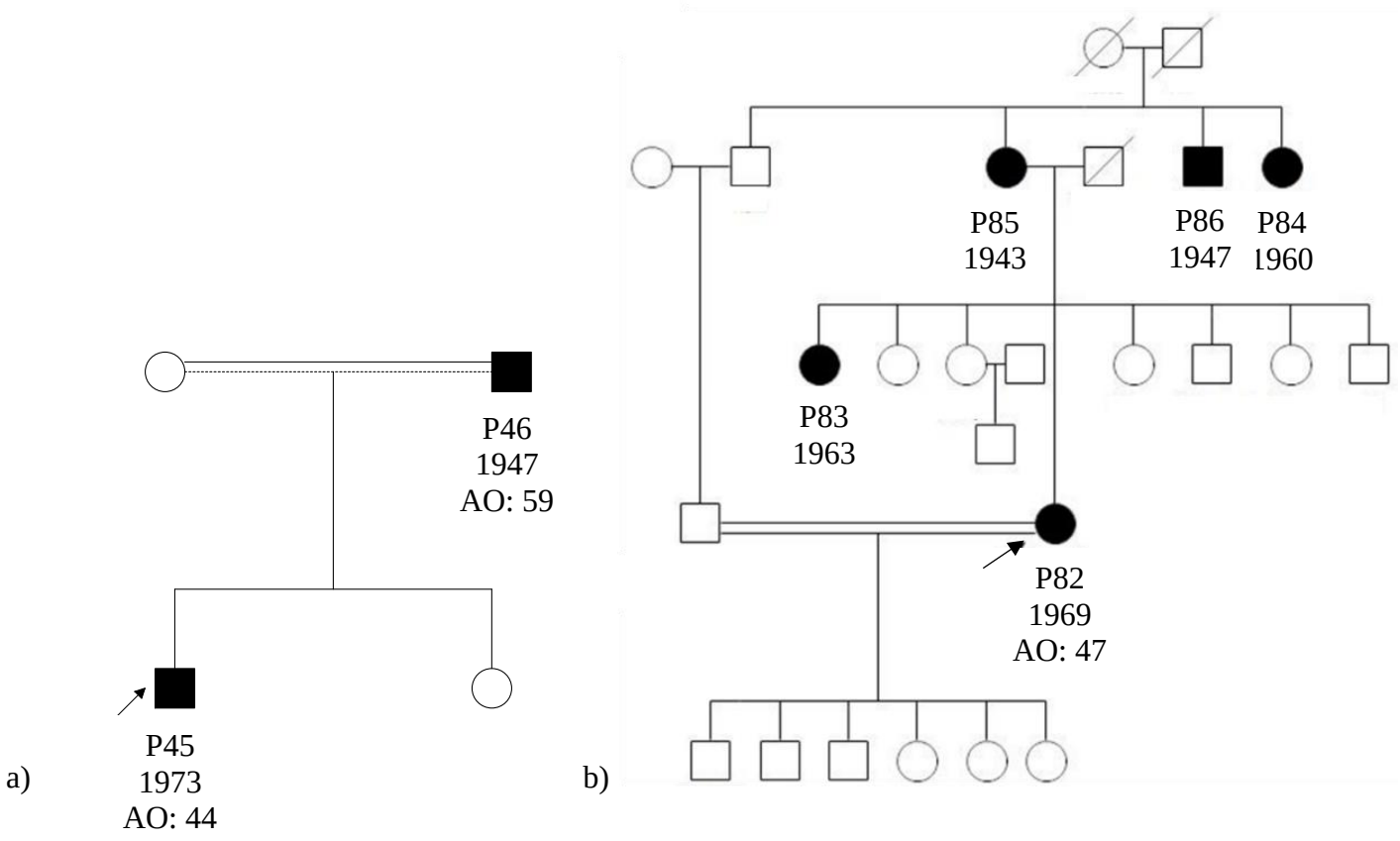
Family 31	P47	M	+	55	EA	AR
Family 32	P48	F	-	babyhood	ATX	AR
Family 33	P49	F	+	24	ATX	AR
	P50	M	+	20	ATX	
Family 34	P51	F	+	55	SPAX	AD
Family 35	P52	M	+	55	SPAX	AR
Family 36	P53	M	+	n.a	ATX	AR
	P54	F	+	n.a	ATX	
Family 37	P55	M	+	35	ATX	AR
Family 38	P56	M	+	42	ATX	AR
	P57	F	+	5	AT	
	P58	M	+	5	AT	
Family 39	P59	F	+	13	ATX	AR
Family 40	P60	M	+	35	ATX	AR
Family 41	P61	F	-	50	ATX	AD
Family 42	P62	M	+	38	SPAX	AR
Family 43	P63	F	+	6	ATX	AR
	P64	M	+	4	ATX	
	P65	M	+	congenital	ATX	
Family 44	P66	M	+	17	SPAX	AR
	P67	M	+	n.a	SPAX	
Family 45	P68	F	+	congenital	ATX	AR
Family 46	P69	F	+	15	ATX	AR
Family 47	P70	F	-	childhood	ATX	Sporadic
Family 48	P71	F	+	10	ATX	AR
Family 49	P72	M	-	n.a	ATX	Sporadic
Family 50	P73	F	-	48	ATX	Sporadic
Family 51	P74	M	-	66	ATX	Sporadic
Family 52	P75	F	-	38	SPAX	Sporadic
Family 53	P76	M	+	26	ATX	AR
Family 54	P77	M	+	13	ATX	AR
Family 55	P78	M	+	52	ATX	AR
Family 56	P79	F	-	67	ATX	AD
Family 57	P80	M	+	51	ATX	AR
Family 58	P81	M	+	6	EA	AR
Family 59	P82	F	+	47	ATX	AD
	P83	F	+	n.a	ATX	
	P84	F	-	n.a	ATX	
	P85	F	-	n.a	ATX	
	P86	M	-	n.a	ATX	
Family 60	P87	M	+	69	ATX	AR
Family 61	P88	F	+	23	ATX	AR
Family 62	P89	M	-	congenital	ATX	Sporadic
Family 63	P90	M	+	22	SPAX	AR
Family 64	P91	F	+	56	SPAX	AR
Family 65	P92	F	-	68	ATX	AD
Family 66	P93	M	+	38	ATX	AD
Family 67	P94	M	+	30	ATX	AD

Family 68	P95	F	-	5	EA	Sporadic
Family 69	P96	M	+	40	ATX	AR
	P97	M	+	n.a	ATX	
	P98	M	+	n.a	ATX	
Family 70	P99	M	-	49	ATX	Sporadic
Family 71	P100	M	+	50	SPAX	AR
Family 72	P101	M	-	74	ATX	Sporadic
Family 73	P102	M	+	4	EA	AR
Family 74	P103	F	-	17	ATX	Sporadic
Family 75	P104	F	+	12	ATX	AR
Family 76	P105	F	+	43	ATX	AR
Family 77	P106	F	-	16	ATX	Sporadic
Family 78	P107	M	+	2	ATX	AR
Family 79	P108	M	-	9	SPAX	Sporadic
Family 80	P109	F	+	44	ATX	AD
	P110	M	+	n.a	ATX	
Family 81	P111	M	+	n.a	ATX	AR
	P112	M	+	58	ATX	
Family 82	P113	M	-	27	ATX	Sporadic
Family 83	P114	M	+	10	ATX	AD
	P115	M	+	n.a	ATX	
	P116	M	+	n.a	ATX	
	P117	M	+	n.a	ATX	
	P118	M	+	n.a	ATX	
	P119	M	+	n.a	ATX	
	P120	F	+	n.a	ATX	
	P121	F	+	n.a	ATX	
	P122	F	+	n.a	ATX	
	P123	M	+	n.a	ATX	
	P124	M	+	n.a	ATX	

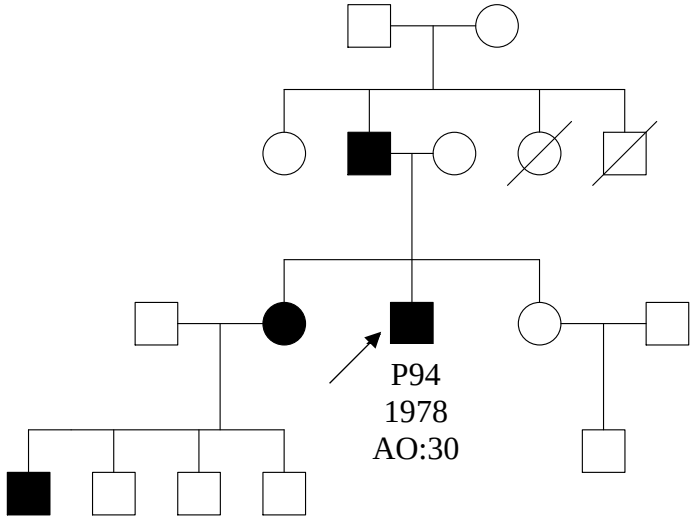
P: Patient, M: Male, F: Female, AO: Age of onset, ATX: Ataxia, SPAX: Spastic ataxia, EA: Episodic ataxia, HSP: Hereditary spastic paraplegia, AT: Ataxia telangiectasia, n.a: not available, AR: Autosomal recessive, AD: Autosomal dominant

3.1.1 Family Trees

Families with AD Inheritance



c)



d)

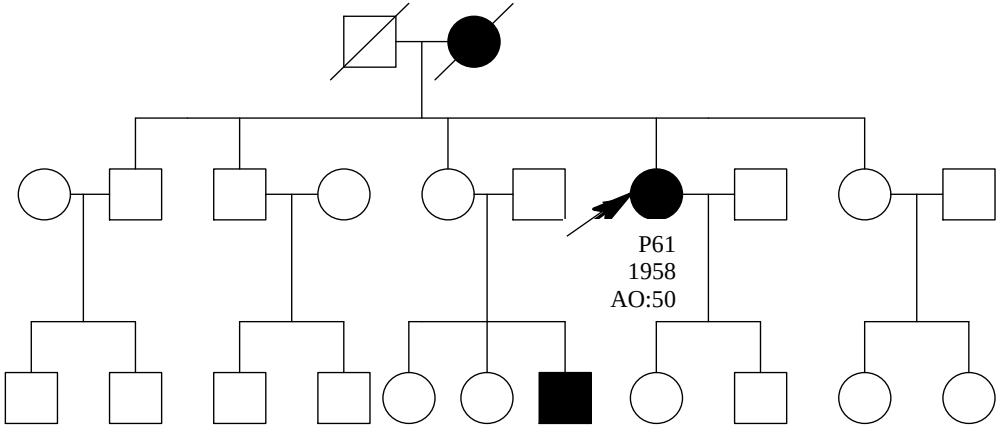
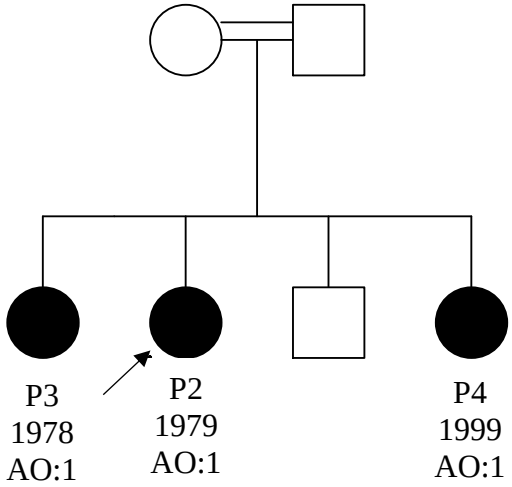


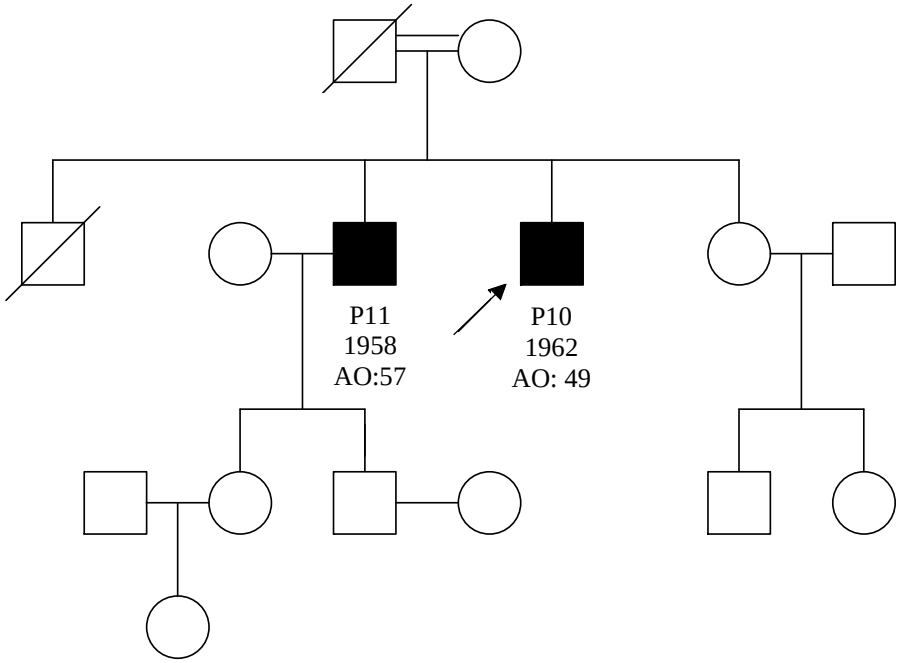
Figure 3.1 a) Family 30 (Patients P45-P46), b) Family 59 (Patients P82-P86), c) Family 67 (Patient P94) and d) Family 41 (Patient P61) showing an autosomal dominant inheritance pattern with several affected individuals.

Families with Autosomal Recessive (AR) Inheritance

a)



b)



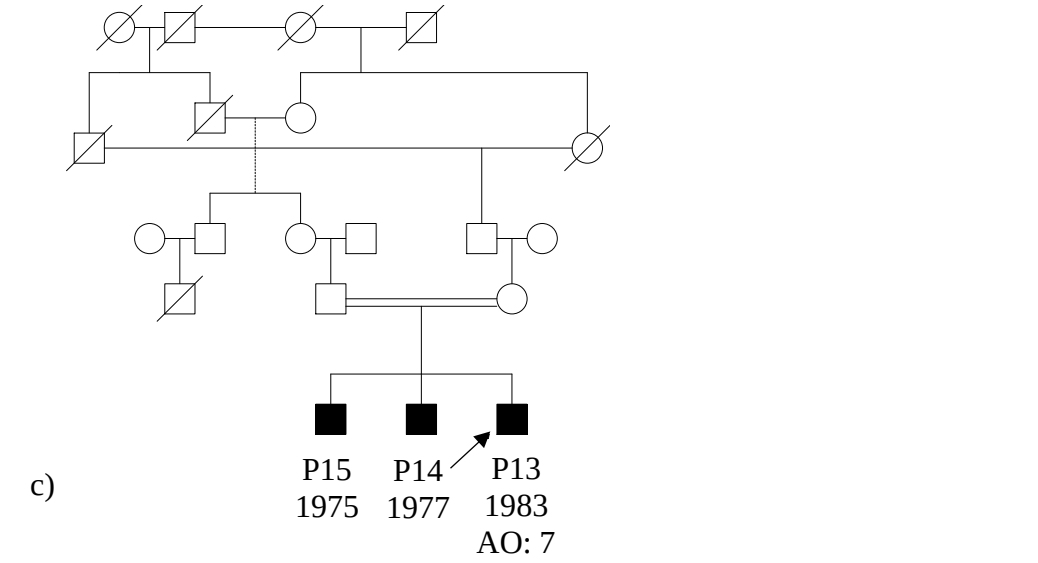


Figure 3.2 Pedigrees of families with an AR inheritance. a) Family 2 (Patients P2-P4), b) Family 7 (Patients P10-P11), c) Family 9 (Patients P13-P15) and d) Family 13 (Patients P20-P21) showing AR inheritance pattern with consanguinity

Families with Sporadic Inheritance

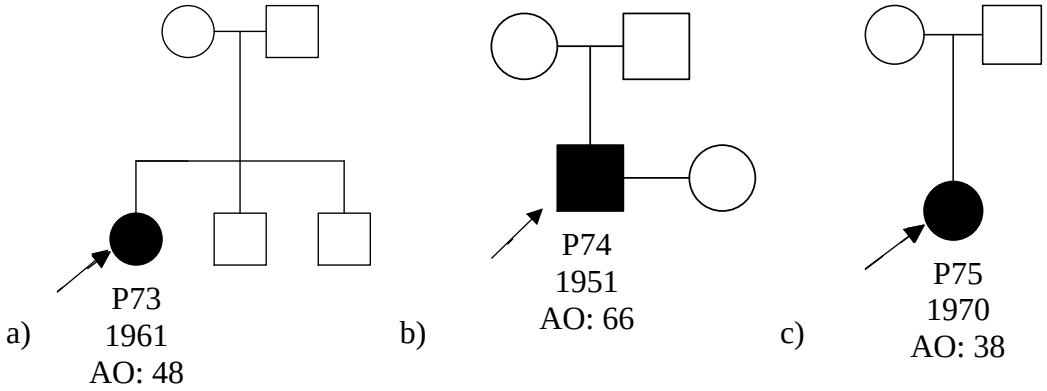


Figure 3.3 a) Family 50 (Patient P73), b) Family 51 (Patient P74) and c) Family 52 (Patient P75) showing sporadic inheritance pattern with only one affected individual and no consanguinity.

3.2 DNA Isolation

MagNA Pure Compact Nucleic Acid Isolation Kit I and MagNA Pure Compact Instrument from Roche were used to extract DNA from peripheral blood samples. Thermo Scientific NanoDrop 2000c Spectrophotometer was used for concentration and quality measurements.

3.3 Whole Exome Sequencing Platforms

Whole exome sequencing was outsourced to different companies like QIAGEN, Germany and Macrogen Inc, Korea and Amsterdam. The sequencing platform used was Illumina NovaSeq 6000. Agilent SureSelect Human All Exon V6 and IDT xGen Exome Research Panel V2 were used as exome enrichment kits.

3.4 Validation Experiments

The MyTaq^{TM} DNA Polymerase (Bioline) with a concentration of 5u/µL and 5X MyTaq Reaction Buffer were used in PCR to amplify 10 ng genomic DNA.
The sequences of the primers used in validation experiments are listed in Table 3.2. The PCR reagents and conditions for amplification are shown in Tables 3.3 and 3.4, respectively.

Table 3.2 Primers used in validation experiments.

Primer Name	Tm (°C)	Sequence (5’ -> 3’)
MFSD8-T160N-F	58.4	AGT GAC TGT ACT GTT GAC CAA G
MFSD8-T160N-R2	57.3	CCT TTC CCT TCA GCT ATG AG
SYNE1-D1124G-F	59.4	GCG TGG GAA TCT TCC TTT GG
SYNE1-D1124G-R	60.3	GCC TCA CCC TGA TTC TGA CTA A
PNPLA6-R314W-F	59.4	GTT CAT CCG TTA TGC TGC CG
PNPLA6-R314W-R	61.4	CAG GCC GAT AAC CAC AGC TC
HSD17B4-p.D142G-F	57.3	ACT CAT TGG GGT GTG ACA GT
HSD17B4-p.D142G-R	57.3	AAG AAC CCC CAG ACA CAA CA
NPC1-ex8-F	59.8	CAT TAC TGC GTG TTC GTC AGG
NPC1-ex8-R	59.8	GCC AGC AAA CCA CAA GGT CAT
VPS13D-e67-F	57.3	TCG ACA GTG GAA GGT GTG AA

VPS13D-e67-R	57.9	ATG CCA GGT TTC TGT ACC TCA
ATM-ex22-F	58.4	TCT CTG TAA GAA TGG CCC TAG T
ATM-ex22-R	55.9	TTG CAA CTG TGA GCT GTT ACT
ATM-ex24-F	58.4	ACT GAC GTT GAT AGC TGT GGT T
ATM-ex24-R	59.7	TGG AAT GTA AGA AAC AGT GCA TAC C
KIF1C-e11-F	59.4	AAG AGG GTC TCA TCC CCA CA
KIF1C-e11-R	59.4	GCA GGA CAG TCA CAG AAC CA
MCOLN1-e4-F	58.8	CGT TGC CTG ACG TGT CAC T
MCOLN1-e4-R	59.4	GAC AAG GTC TTC CTA CGT GG
SPG11-e22-F	61.4	GGG AGA GGA GGC CAC AAA TC
SPG11-e22-R	61.4	GCC TTA GAC CTC GTC ACA CC
CAPN1-e8-F	59.4	TCC TGC CCT GAC TGA CTG TA
CAPN1-e8-R	59.4	TGG TAG AGT CCC GAG CAG AA
CAPN1_I6_F	59.4	TCC TGC TTG ATG CCA GAG TG
CAPN1_I6_R	61.4	GGC AGG GTA TGC ATC CAG AG
SACS_E9_F	59.4	CCC CAC AGC AGA AGT GAA AC
SACS_E9_R	57.3	TGG TTG GTG TAG AAG GAG CA
SPG7_E14_F	59.4	CCT TGT GCC AGG TCT CCA TA
SPG7_E14_R	59.4	CAG AAG GAG TCA TGC AGG GA
CAPN1_E4_F	61.4	GAG GGG AGT TAA GTG GCA CC
CAPN1_E4_R	61.4	CGA GAT GGC ACT GAG AGG TC
SACS-E9-Leu2402-F	59.4	GAG GCG GCA CCA TAC CTT TA
SACS-E9-Leu2402-R	59.4	CTT CCC TGG GTA TGT CAG CA
SACS_Glu1162Ter_F	55.3	AAA GGA TGT TGT GCA AGT GG
SACS_Glu1162Ter_R	53.2	ACC CAT GGA AAT TTT AAG GC
L2HGDH_E7_F	59.4	GCA GGA CTT TAC TCA GAC CG
L2HGDH_E7_R	59.8	GCA CTC ATG CTT TCT GGA GAC
CAPN1_E18-19_F	57.3	TGG CTT TTG CTG CTT CTC CT
CAPN1_E18-19_R	59.4	GTG CTA GTA ACC GTG GGG AA
CAPN1_E20_F	59.4	GGT GCT AGT GGT GAC ATG GT
CAPN1_E20_R	61.0	CTC TCT GGC ACA CAC CCA C
CAPN1_E21_F	61.4	CTT CTC AGG CAG CAG GAA GG
CAPN1_E21_R	61.0	GGC TTC AGG GTT GGC TCT G

CAPN1_E22_F	61.4	CAG GGG TGG AGG AGG TAA GA
CAPN1_E22_R	59.4	AGA TGA ACG ATG GTG GCT GG
NKX6-2_E2_F	59.4	GGA CAA GGA CGG GAA GAA GA
NKX6-2_E2_R	58.8	TTC TTG GCC GAC GCC ATC T
SYT14-E14-F	59.3	GCT GCT GAA TGA ATT GAA CAA CTG
SYT14-E14-R	57.9	TCC TCA AGT GTA TCT GCT CCT
CACNA1A-i29-F	61.4	CCT GAA CTG TGC CTC CTA CC
CACNA1A-i29-R	59.4	TTG AGG TTG GGG TTC TTG GG
SPTAN1_E2_F	59.4	CGT GTG TGG GTT AGG CAT CT
SPTAN1_E2_R	59.4	CCT GCA AGT TGG TTG GGT CT
SPTAN1_E48_F	59.4	AGC AAG TTC CAG TCC TGT GG
SPTAN1_E48_R	59.4	CCC TCT CAT TTG CCC TGG TT
SACS_E9_F	59.4	GTT GCG AGG GGT TGC TTT TG
SACS_E9_R	59.4	GGA GGT AAA GCG CAA GGT CT
ATM_E42_F	59.4	TTG TAT GGC TGT GGT GGA GG
ATM_E42_R	59.4	CAT CCC TCT GCT TCA GGA GT
ATM_E37_F	55.3	TAT GTC AAC GGG GCA TGA AA
ATM_E37_R	58.9	CCT GTA AAG TGC TTT TAG TGG GA

Table 3.3 PCR reagents for amplification.

Reagent	**Volume (µL)**	**Stock conc.**	**Final conc.**
dH_2O	16.8	-	-
Buffer	5	5X	1X
Forward Primer	1	10 µM	0.4 µM
Reverse Primer	1	10 µM	0.4 µM
MyTaq Polymerase	0.2	5u/µL	1u
DNA	1	10 ng/µL	10 ng/µL
Total	25		

Table 3.4 Conditions for PCR amplification.

Step	**Temperature (ºC)**	**Duration**	**# of cycles**
Initial denaturation	95	1 min	1
Denaturation	95	15 sec	35
Annealing	variable*	15 sec	
Extension	72	10 sec	
Final extension	72	10 min	1
Hold	24	∞	1
*Annealing temperature is specific for each primer pair			

3.4.1 Agarose Gel Electrophoresis and Extraction from the Gel

All chemicals and equipment used in agarose gel electrophoresis are listed in Table 3.5. Gel extraction of the PCR products was performed with QIAQuick Gel Extraction Kit, Qiagen.

Table 3.5 Chemicals used in gel electrophoresis.

Solution/Chemical/Equipment		**Catalog No.**	**Company**
10X TBE Electrophoresis Buffer	89 mM Tris-base	01000258	Thermo Scientific, USA
	89 mM Boric acid		
	2 mM EDTA		
Agarose Molecular Biology Grade		17J174122	GeneON, Germany
GelRed™ Dropper Bottle		103.302-05	Olerup SSP, Sweeden
GeneRuler 100 bp DNA Ladder		00767368	Thermo Scientific, USA
6X TriTrack DNA Loading Dye		00747349	Thermo Scientific, USA
Electrophoresis Tank, OWL EasyCast B1		349900	Thermo Scientific, USA
Power Supply, EC250-90		25090 ECA-LVD	Thermo Scientific, USA
ChemiDoc™ MP Imaging System		17001402	Bio-Rad, USA

3.4.2 Laboratory Equipment and Kits

All laboratory equipment and kits used in the framework of this thesis are listed in Tables 3.6 and 3.7, respectively.

Table 3.6 Laboratory equipment.

Equipment	**Model**	**Catalog No.**	**Company**
Autoclave	HV-110L	30515011496	HMC HIRAYAMA
Balance	TE612	-	Sartorius, Germany
Centrifuge	Microfuge 16	A46473	Beckman Coulter, USA
DNA Extraction System	MagNa Pure Compact Instrument	03731146001	Roche, Switzerland
Electrophoretic Equipment	OWL EasyCast B1 Tank	349900	Thermo Scientific, USA
Falcon Tubes	50-ml Screw Cap Tubes	100619-02903	Axygen, USA
Gel Documentation System	ChemiDoc™ MP Imaging System	17001402	Bio-Rad, USA
Glassware	-	-	Isolab, Germany
Heat Block	ThermoMixer F1.5	5384000012	Eppendorf, Dubai
Microcentrifuge tubes	0.5 mL, 1.5 mL, 2.0 mL	-	Axygen, USA
Micropipettes	0.2 uL, 10 uL, 20uL, 200uL, 1000 uL	-	RaininPipet-Lite XLS, Mettler-Toledo International Inc., USA
Microwave oven	Intellowave MD554	7885270100	Arçelik, Turkey
Parafilm	-	PM-996	Bemis, USA
Power Supply	EC250-90	25090 ECA-LVD	Thermo Scientific, USA
Refrigerator	391640 EI	-	Arçelik, Turkey
Spectrophotometer	NanoDrop 2000c	E112352	Thermo Scientific, USA
Thermal Cycler	T100™	1861096	BioRad, USA
Tips/Filter tips	10 uL, 100 uL, 200 uL, 1000 uL	-	Axygen, USA
Vortex	REAX top	541-10000-00-0	Heidolph, Germany

Table 3.7 Commercially available kits, other than exome capture kits.

Kit	**Catalog No.**	**Company**
MagNA Pure Compact Nucleic Acid Isolation Kit I	03730972001	3730964001
QIAquick Gel Extraction Kit	163033035	Qiagen, Germany
QIAquick PCR Purification Kit (250)	163018180	Qiagen, Germany
Genomic DNA Clean & Concentrator (25 preps)	ZRC175442	Zymo Research, USA

3.5 Online Databases and Bioinformatic Tools

The open-source NGS and bioinformatics tools and online databases used in WES analysis are listed in Table 3.8.

Table 3.8 Open-source bioinformatics tools and online databases.

Software/Tool/Database	**Description**
CADD (Rentzsch et al., 2021)	A tool to score the deleteriousness of single nucleotide and insertion/deletions variants in the human genome
CLC Main Workbench (Qiagen, Germany)	Software for DNA, RNA, and protein sequence analyses
ClinVar (Landrum et al., 2020)	An archive of the relationships among human variations and phenotypes, with supporting evidence
DANN (Quang et al., 2015)	Pathogenicity score ranging is 0 to 1, 1 being the most damaging
dbSNP (Sherry et al., 2001)	Database for SNPs and multiple small-scale variations
FinchTV	An application to view Sanger sequencing data
Franklin by Genoox	A community-based variant interpretation platform
GeneCards	A database for information on all annotated and predicted human genes
GERP++ (Davydov et al., 2010)	Conservation score calculated by quantifying substitution deficits

GnomAD	Database of 76,156 genomes from unrelated individuals
Integrative Genomics Viewer (IGV)	A tool for the visual exploration of genomic data
MetalR	A prediction tool that integrates nine independent variant scores for missense variants to classify them as tolerated or damaging
MutationTaster	A prediction tool for the pathogenicity of a variant based on evolutionary conservation, splice-site, mRNA, protein, and regulatory features.
NCBI Primer Blast (Ye et al., 2012)	A tool to make primers that are specific to intended PCR target
NetPrimer (Premier Biosoft)	A tool to analyze primers
OligoCalc (Kibbe, 2007)	A tool to analyze primers
Online Mendelian Inheritance in Man (OMIM) (McKusick-Nathans Institute of Genetic Medicine)	An online collection of human genes and genetic phenotypes
REVEL (Ioannidis et al., 2016)	A prediction tool for pathogenicity of missense variants based on a combination of scores from 13 individual tools
RFFlow	A software for drawing pedigrees, flowcharts, and other diagrams
SEQ Platform by Genomize	An organizational platform on storage, sharing, and usage of genomic data
SIFT (Ng and Henikoff, 2003)	A prediction tool to analyze impact of amino acid substitutions
UCSC database (Kent et al., 2002)	WEB browser of University of California Santa Cruz
Varsome (Kopanos et al., 2018)	A search engine for human genomics

Chapter 4:
METHODS

4.1 DNA Isolation

Peripheral blood in EDTA tubes was collected from all participants. MagNA Pure Compact Instrument was used to extract genomic DNA. Concentration of the extracted DNA was measured by NanoDrop spectrophotometer at 260 nm optical density. The absorption values at 260/230 nm and 260/280 nm were evaluated and documented to detect chemical and protein contaminations in isolated DNA.

4.2 Whole Exome Sequencing

Whole exome sequencing and CNV analysis from WES data were performed on DNA samples of all index patients. In some cases, affected family members or parents were also subjected to WES. All samples were outsourced to companies: Macrogen (Korea and Amsterdam) and Qiagen (Germany).

Whole exome sequencing involves three basic steps: Library preparation and amplification, sequencing, and data analysis (Figure 4.1). The workflow of WES is shown in Figure 4.2.

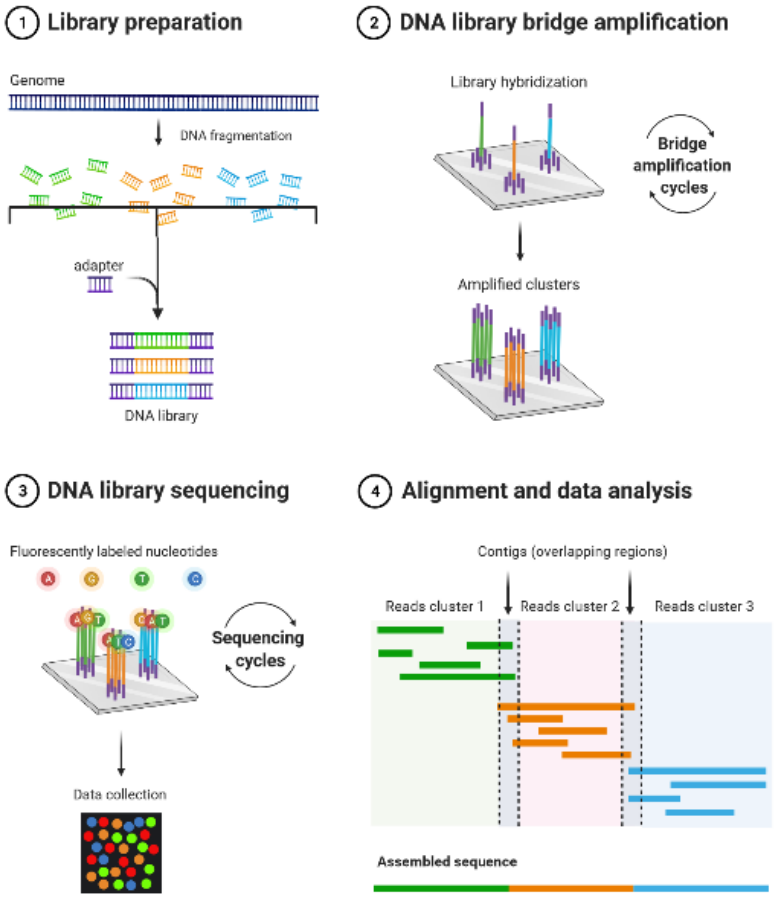


Figure 4.1 Steps of whole exome sequencing. Reprinted from "Next Generation Sequencing (Illumina)", by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates.

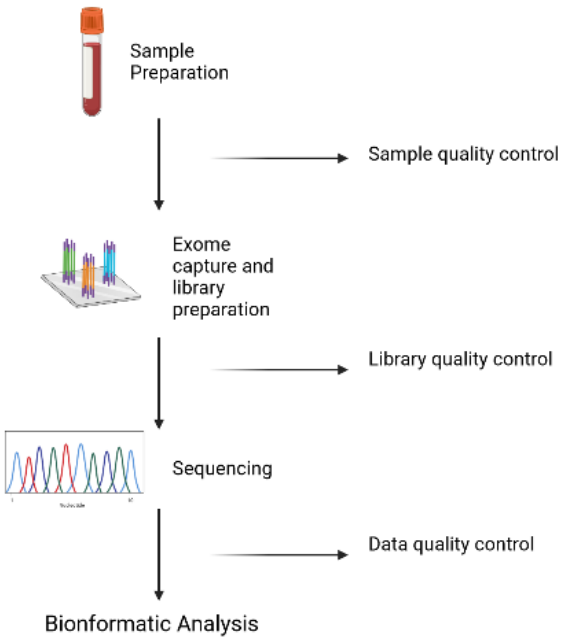


Figure 4.2 The workflow for whole exome sequencing, created with BioRender.com.

4.3 Whole Exome Sequencing Data Analysis

SEQ Platform by Genomize (Istanbul, Turkey) is employed for bioinformatic data analysis starting from cleaning of the raw reads, alignment to reference genome, variant calling and annotation.

4.3.1 Variant Prioritization

Variants were classified into different categories: Confirmed pathogenic variants, possible pathogenic novel variants, and variants of uncertain significance (VUS). Confirmed pathogenic variants are the ones reported to be pathogenic in previous studies. Possible pathogenic novel variants are those detected in genes associated to ataxias, affecting protein structure or function and which segregate with family pedigree. Variants of uncertain significance are possible pathogenic variants found in genes that are not associated with ataxias directly but associated to other neurological disorders. DANN, CADD, MutationTaster, MetalR, REVEL, and SIFT scores are used to interpret the pathogenicity of the variants and the effect of the variant on the protein structure and/or function. GERP++ score is used to evaluate the evolutionary conservation rate of the variant site.

4.4 PCR Experiments for Validation of Variants

4.4.1 PCR Conditions

PCR-based Sanger sequencing is employed to validate the presence and segregation of candidate variants from WES data analysis. Primers specific to the variant region were either designed by NCBI Primer Blast tool (Ye et al., 2012) or retrieved from literature. OligoCalc and NetPrimer tools were used to check the possibility of self-dimer and cross-dimer between forward and reverse primers. UCSC in silico PCR was used to see the specificity of the region in the genome. Lyophilized primers were synthesized and purchased from Sentromer (Istanbul, Turkey). 100 uM primers were prepared by adding the corresponding amount of dH_2O and 10 uM dilutions were prepared for further use in PCR experiments.

The PCR products were run on 2% agarose gel. GelRed fluorescent DNA dye was used as intercalating agent for visualization of the products under UV light.

4.4.2 Agarose Gel Electrophoresis

2% agarose gel was prepared using 0.5X TBE buffer and GelRed fluorescent DNA dye and placed into an electrophoresis chamber which contains 0.5X TBE. Three volumes of PCR product were mixed with two volumes of 6X loading dye and loaded onto the gel. The gel was run for an hour at 120 A and visualized under UV light using Bio-Rad ChemiDoc MP Imaging System. PCR products were outsourced to Macrogen (Amsterdam) for Sanger sequencing, the results were analyzed using FinchTV and CLC Main Workbench programs.

4.4.3 Purification of PCR Products before Sanger Sequencing

PCR products were purified either from agarose gel or directly from the product itself to avoid primer dimers and unspecific bands so that Sanger sequencing results would not have background noise. QIAquick Gel Extraction Kit (Qiagen, Germany) was used for agarose gel extraction and QIAquick PCR Purification Kit (Qiagen, Germany) for PCR product purification.

4.5 Restriction Enzyme Digestion

Endonuclease digestion was carried out in a 15 µL volume containing 12 µL of each PCR product and 1-5 units of the restriction enzyme in the presence of the buffer required. After incubation at 37 ^{o}C for 2 hours the samples were run on a 2% agarose gel at 120 A for one hour and visualized under UV light. The undigested PCR products were also applied on the gel and served as controls.

Chapter 5:
RESULTS

In this study, WES comprising also CNV analysis and followed by Sanger sequencing was applied to 83 index patients with a complex ataxia phenotype and their families

5.1 Whole Exome Sequencing Analysis

27 different variants in 18 genes were identified in 26 out of 83 distinct families (Figure 5.1). Fifty-seven of these families presented with a recessive and 11 with a dominant inheritance pattern, 15 family trees were sporadic. Eleven variants were novel. The variants are listed in Table 5.1 along with the associated disease in OMIM. *CAPN1* is the most common gene in the cohort followed by *SACS*, *SPG7*, *SPG11*, and *HSD17B4*. Family pedigrees are shown in Figure 5.3-Figure 5.8.

The diagnostic yield in recessive families was 42% (24/57). Disease-causing variants in AD and sporadic families were identified in only one family each, corresponding to diagnostic yields of 9% (1/11) and %7 (1/15), respectively. The overall diagnostic yield corresponds to 31% (Figure 5.2).

Sanger sequencing was applied to confirm the presence of the variants and to validate the segregation of the variant in the family. Family segregation, online prediction tool scores, and gnomAD frequencies of the variants are shown in Table 5.2.

CNV analysis from WES data was also performed. In two families, homozygous exonic deletions were found in *CAPN1* and *SYT14* genes. Their validation by PCR and agarose gel electrophoresis is shown in Figure 5.9.

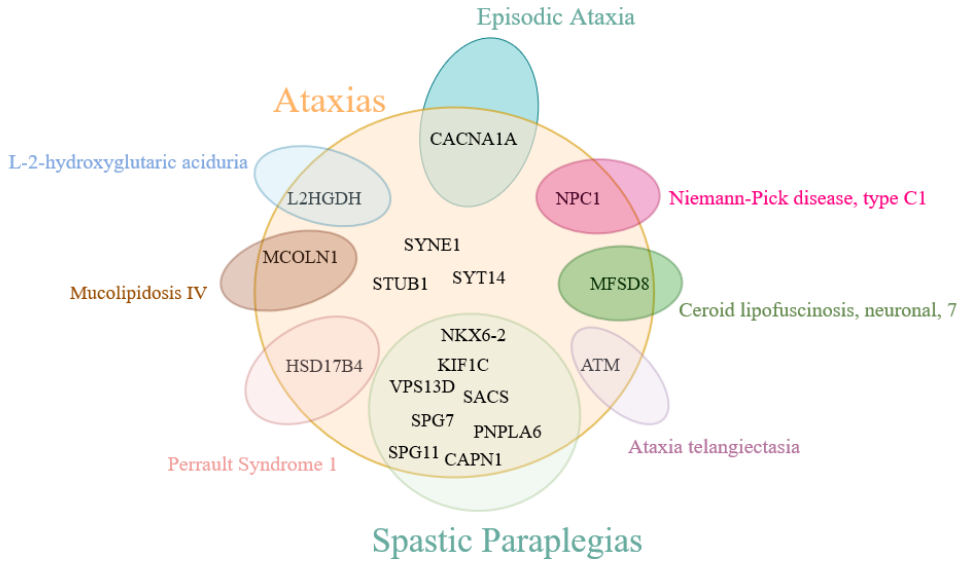


Figure 5.1 Genes identified in this study giving rise to ataxias.

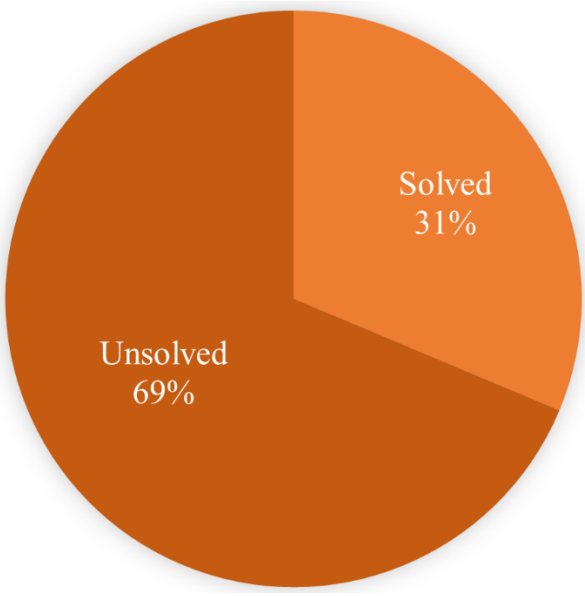


Figure 5.2 31% of patients was solved as opposed to 69% unsolved in the cohort under study.

Table 5.1 List of all causative variants and associated diseases in OMIM in the families solved.

Family ID	**Gene**	**Variant**		**Transcript ID**	**Zygosity**	**OMIM**
		Coding sequence	**Protein sequence**			
Family 1	SYNE1	c.3371A>G	p.Asp1124Gly[n]	ENST00000367255.5	hom	SCAR8
Family 2	PNPLA6	c.940C>T	p.Arg314Trp	ENST00000414982.3	hom	SPG39
Family 3	HSD17B4	c.425A>G	p.Asp142Gly[n]	ENST00000504811.1	hom	PRLTS1
Family 4	MFSD8	c.479C>A	p.Thr160Asn	ENST00000296468.3	hom	CLN7
Family 6	NPC1	c.1165C>T	p.Arg389Cys	ENST00000269228.5	hom	NPC1
Family 7	VPS13D	c.12629C>T	p.Ala4210Val	ENST00000358136.3	hom	SCAR4
Family 8	SPG11	c.7158_7161dupACAA	p.His2388ThrfsTer6[n]	ENST00000261866.7.5	hom	SPG11
Family 9	KIF1C	c.866A>C	p.Gln289Pro[n]	ENST00000320785.5	hom	SPAX2
		c.940+1G>A[n]	-		hom	

Family 10	CAPN1	c.907T>G	p.Trp303Gly[n]	ENST00000533820.1	hom	SPG76
Family 11	MCOLN1	c.534_536delCAA	p.Asn179del[n]	ENST00000264079.6	hom	ML4
Family 12	SPG11	c.3809T>A	p.Val1270Asp	ENST00000261866.7	hom	SPG11
Family 13	CAPN1	c.759+1G>A	-	ENST00000533820.1	hom	SPG76
Family 14	SPG7	c.1861C>T	p.Gln621Ter[n]	ENST00000268704.2	hom	SPG7
Family 15	CAPN1	c.397C>T	p.Arg133Ter	ENST00000533820.1	hom	SPG76
Family 16	SACS	c.7205_7206delTT	p.Leu2402ArgfsTer6	ENST00000382292.3	hom	ARSACS
Family 17	SACS	c.3484G>T	p.Glu1162Ter	ENST00000382292.3	hom	ARSACS
Family 18	SPG7	c.1715C>T	p.Ala572Val	ENST00000268704.2	hom	SPG7
Family 20	HSD17B4	c.425A>G	p.Asp142Gly[n]	ENST00000504811.1	hom	PRLTS1

Family 22	L2HGDH	c.905C>T	p.Pro302Leu	ENST00000421284.3	hom	L2HGA
Family 23	CAPN1	Exon18-22 deletion[n]	-	ENST00000533820	hom	SPG76
Family 24	SYT14	Exon4 deletion[n]	-	ENST00000422431	hom	SCAR11
Family 25	NKX6-2	c.545C>T	p.Ala182Val	ENST00000368592.5	hom	SPAX8
Family 26	SACS	c.11542_11544del	p.Ile3848del	ENST00000382298.3	hom	ARSACS
Family 28	ATM	c.5610delC	p.Phe1870LeufsTer47	ENST00000278616.4	het	AT
		c.6154G>A	p.Glu2052Lys		het	
Family 19	CACNA1A	c.5998G>A	p.Gly2000Arg[n]	ENST00000360228.5	het	EA2
Family 21	STUB1	c.146A>G	p.Tyr49Cys	ENST00000219548.4	het	SCA48

n: novel, hom: homozygous, het: heterozygous

Table 5.2 Online prediction tool scores and frequencies of the variants.

Chr. Location	Gene	Variant	ACMG (Varsome)	CADD	DANN	GERP ++	Mutation Taster	REVEL	MetalR	SIFT	gnomAD	Fam. Seg.
6:152771784	SYNE1	p.Asp1124Gly	VUS: PM2, PP3	25.1	0.99	5.78	Disease causing	B	D	D	-	+
19:7605926	PNPLA6	p.Arg314Trp	LP: PM1, PM2, PP5, PP2, PP3	29.1	0.99	5.38	Disease causing	B	T	D	-	+
5:118813112	HSD17B4	p.Asp142Gly	LP: PP3, PM1, PM2	32	0.99	6.02	Disease causing	Pat	D	D	-	+
4:128863274	MFSD8	p.Thr160Asn	LP: PM2, PM5, PP2, PP3	24.6	0.99	4.86	Disease causing	B	T	D	-	+
18:21136368	NPC1	p.Arg389Cys	LP: PM2, PP5, PP2, PP3	30	0.99	5.97	Disease causing	Pat	D	D	0.000003981	n.a
1:12520418	VPS13D	p.Ala4210Val	LP: PM2, PP5, PP2, PP3	32	0.99	6.16	Disease causing	B	T	-	0.00003959	+

15:44855490	SPG11	p.His2388ThrfsTer6	Pat: PVS1, PM2, PP3, PP5	-	-	5.62	-	-	-	-	0.000003979	n.a
17:4907294	KIF1C	p.Gln289Pro	LP: PP3, PM1, PM2, BP1	31	0.99	5.66	Disease causing	Pat	D	D	0.001153	+
17:4907369		c.940+1G>A	Pat: PVS1, PM2, PP3	34	0.99	5.13	Disease causing	-	-	-	-	+
11:64955489	CAPN1	p.Trp303Gly	LP: PM1, PM2, PP2, PP3	33	0.98	4.63	Disease causing	Pat	D	D	-	n.a
19:7591775	MCOLN1	p.Asn179del	VUS: PM2, PM4, PP3	-	-	5.39	-	-	-	-	-	+
15:44890912	SPG11	p.Val1270Asp	VUS: PM2, PP3	28.2	0.99	6.03	Disease causing	Pat	D	D	-	+
11:64953810	CAPN1	c.759+1G>A	Pat: PVS1, PM2, PP3, PP5	34	0.99	4.29	Disease causing	-	-	-	0.000004139	+

16:89619468	SPG7	p.Gln621Ter	Pat: PVS1, PM2, PP3	42	0.99	5.93	Disease causing automatic	-	-	-	-	+
11:64951004	CAPN1	p.Arg133Ter	Pat: PVS1, PM2, PP3	37	0.99	5.26	Disease causing automatic	-	-	-	-	n.a
13:23910809	SACS	p.Leu2402ArgfsTer6	Pat: PVS1, PM2, PP5	-	-	5.73	-	-	-	-	0.00001995	+
13:23914531	SACS	p.Glu1162Ter	Pat: PVS1, PM2, PP3	35	0.99	6.16	Disease causing	-	-	-	0.000003980	+
16:89616953	SPG7	p.Ala572Val	LP: PP5, PM2, PP2, PP3	24.3	0.99	5.84	Disease causing	Pat	D	-	0.00003537	n.a
14:50735882	L2HGDH	p.Pro302Leu	Pat: PP5, PP3, PM2, PP2	33	0.99	5.57	Disease causing automatic	Pat	D	D	0.00001194	+
11:64977116	CAPN1	Exon18-22 deletion	LP (Franklin by Genoox)	-	-	-	-	-	-	-	-	n.a

1:210186668	SYT14	Exon4 deletion	VUS (Franklin by Genoox)	-	-	-	-	-	-	-	-	n.a
10:134598818	NKX6-2	p.Ala182Val	LP: PM1, PM2, PP5, PP2, PP3	27.7	0.99	3.77	Disease causing	Pat	D	D	0.00003195	+
13:23906471	SACS	p.Ile3848del	VUS: PM2, PM4, PP3	-	-	5.93	-	-	-	-	-	+
11:108175515	ATM	p.Phe1870LeufsTer47[a]	LP: PVS1, PM2	-	-	5.23	-	-	-	-	-	-
11:108186796		p.Glu2052Lys	VUS: PM1, PM2, PP3	25.1	0.99	5.29	Disease causing	B	T	T	0.00005658	+
19:13323497	CACNA1A	p.Gly2000Arg	VUS: PM2, PP2	22.3	0.96	4.55	Disease causing	B	D	T	0.00001561	n.a
16: 730671	STUB1	p.Tyr49Cys	VUS: PM2, PP2, PP3	24.4	0.99	3.66	Disease causing	Pat	D	D	-	n.a

Pat: Pathogenic, LP: Likely pathogenic, VUS: Variant of uncertain significance, D: Damaging, T: Tolerated, B: Benign, a: de novo variant, n.a: not available

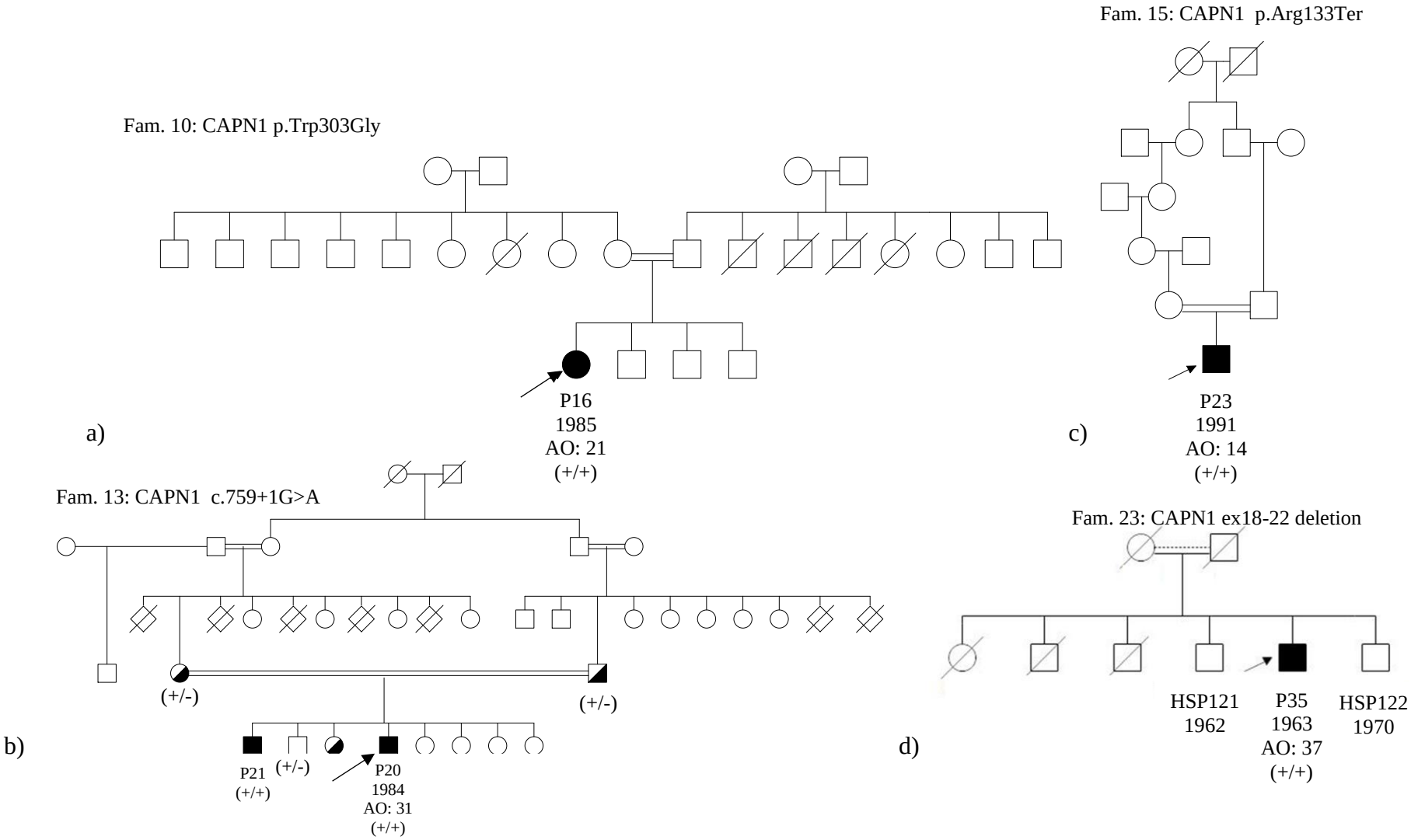


Figure 5.3 CAPN1 families. a) fam 10, b) fam 13, c) fam 15, d) fam 23.

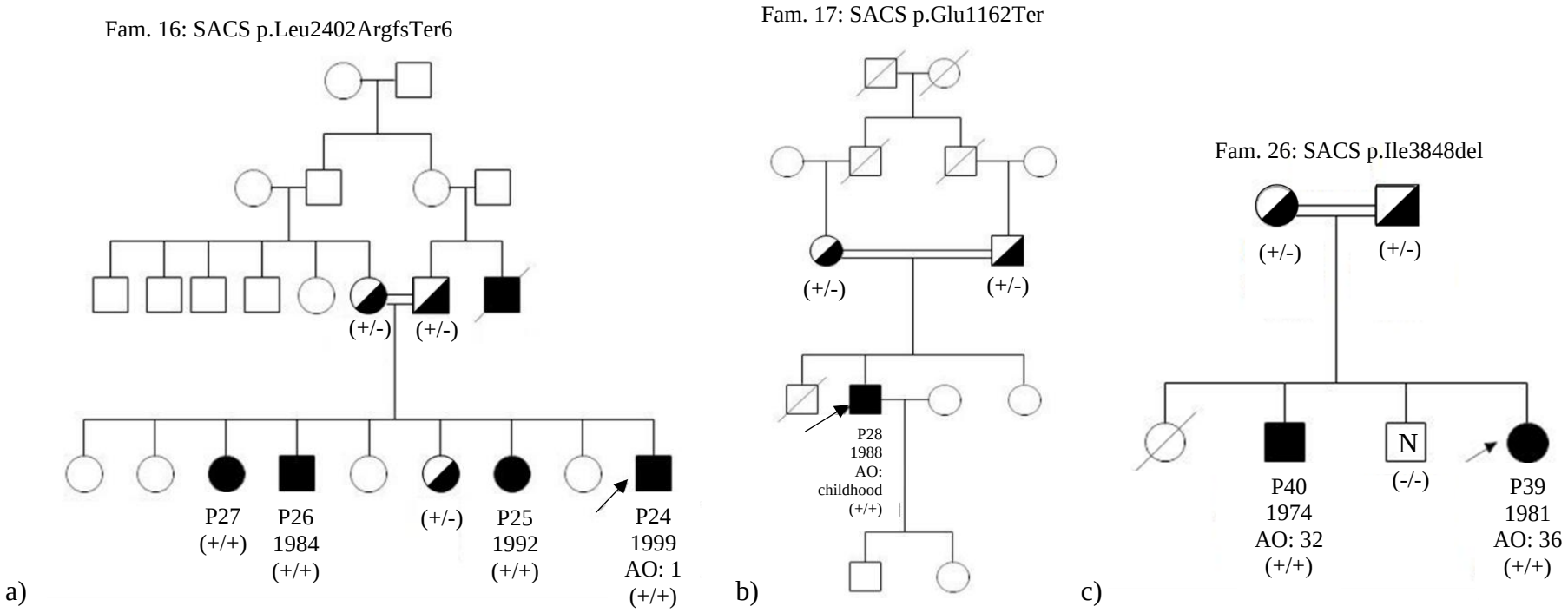


Figure 5.4 ARSACS families. a) fam 16, b) fam 17, c) fam 26.

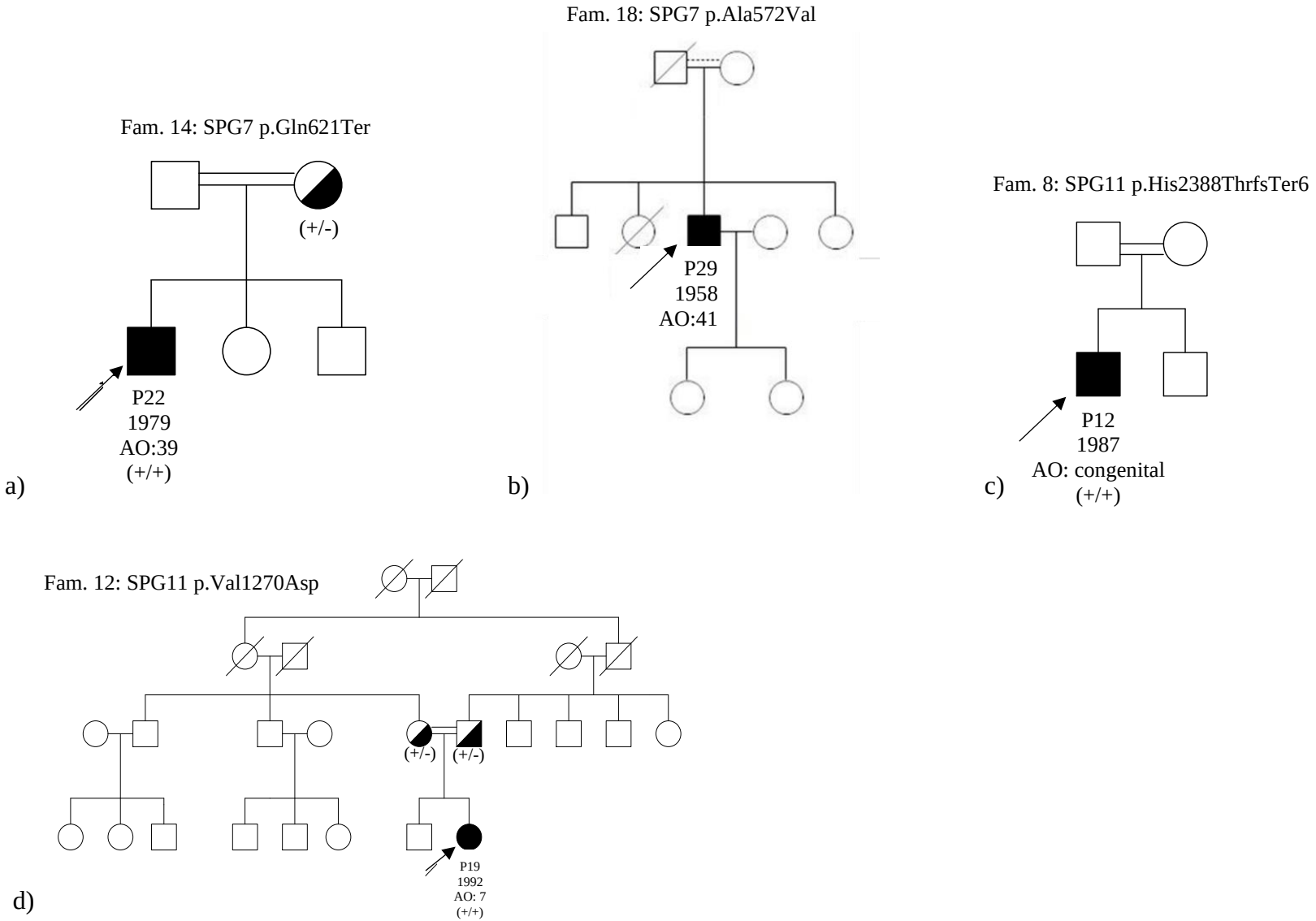
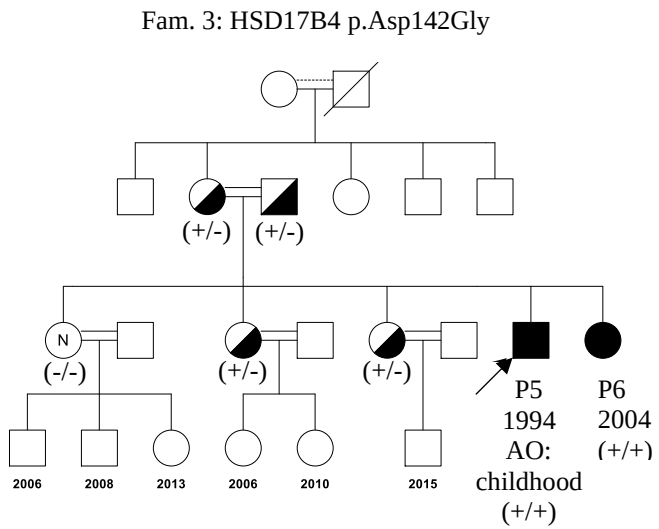


Figure 5.5 SPG7 and SPG11 families. a) SPG7-fam 14, b) SPG7- fam 18, c) SPG11-fam 8, d) SPG11-fam 12.

a)



b)

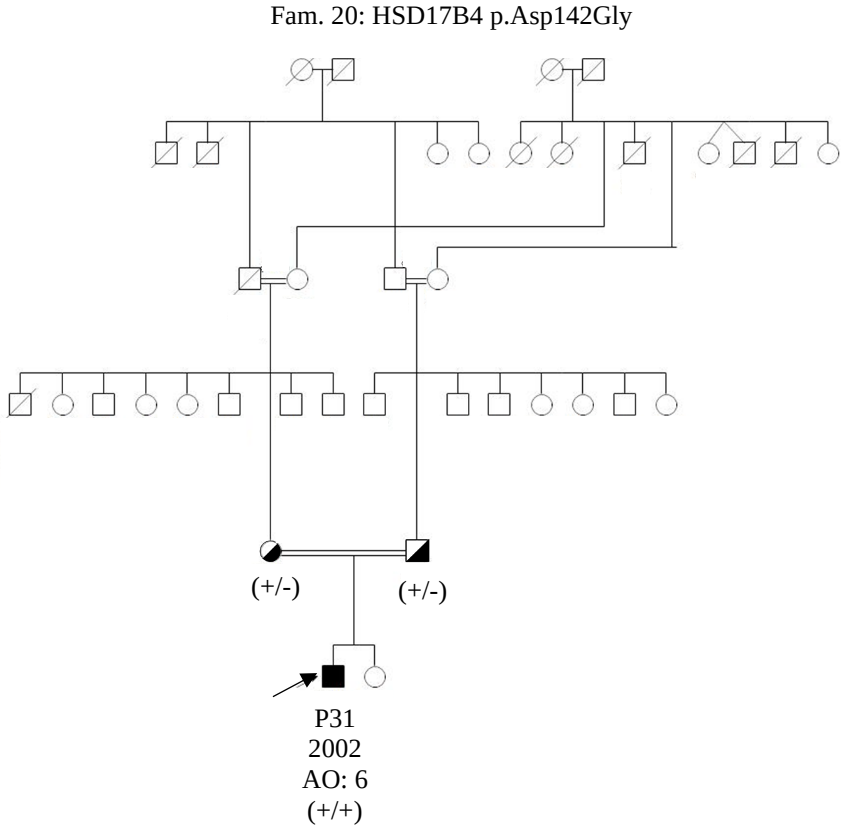


Figure 5.6 a) HSD17B4-fam 3, b) HSD17B4-fam 20.

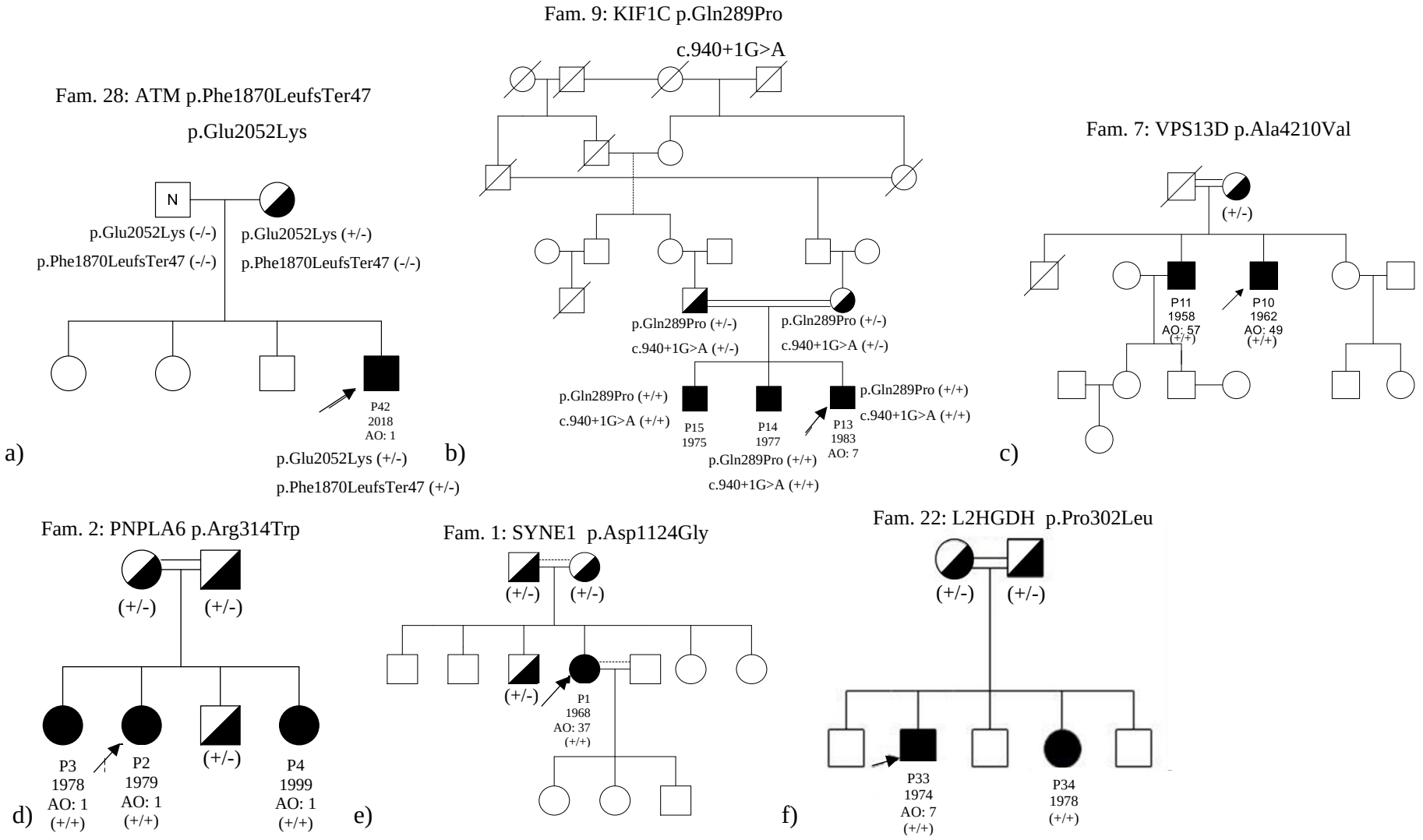


Figure 5.7 a) ATM-fam 28, b) KIF1C-fam 9, c) VPS13D-fam 7 d) PNPLA6-fam 2, e) SYNE1-fam 1, f) L2HGDH-fam 22.

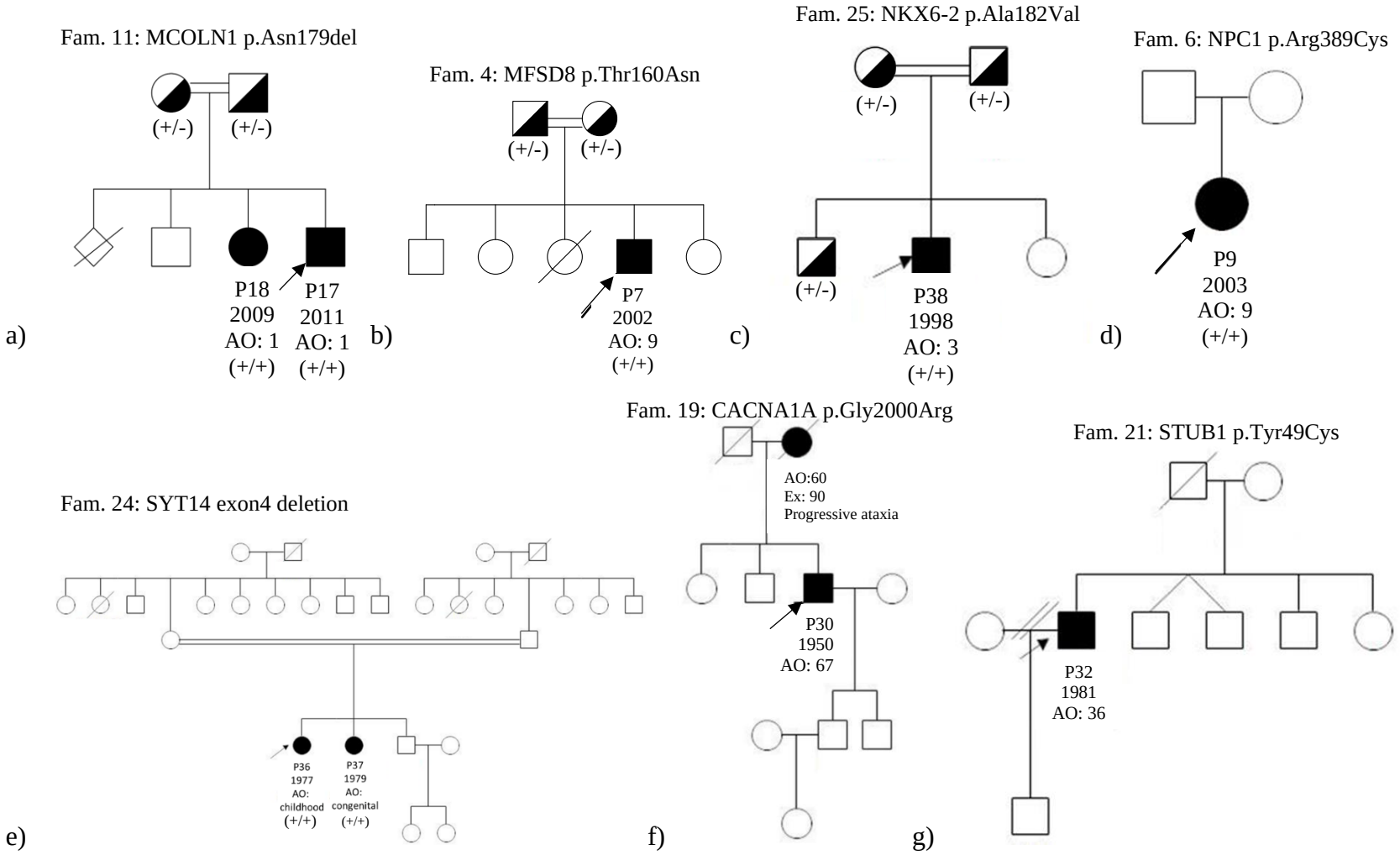
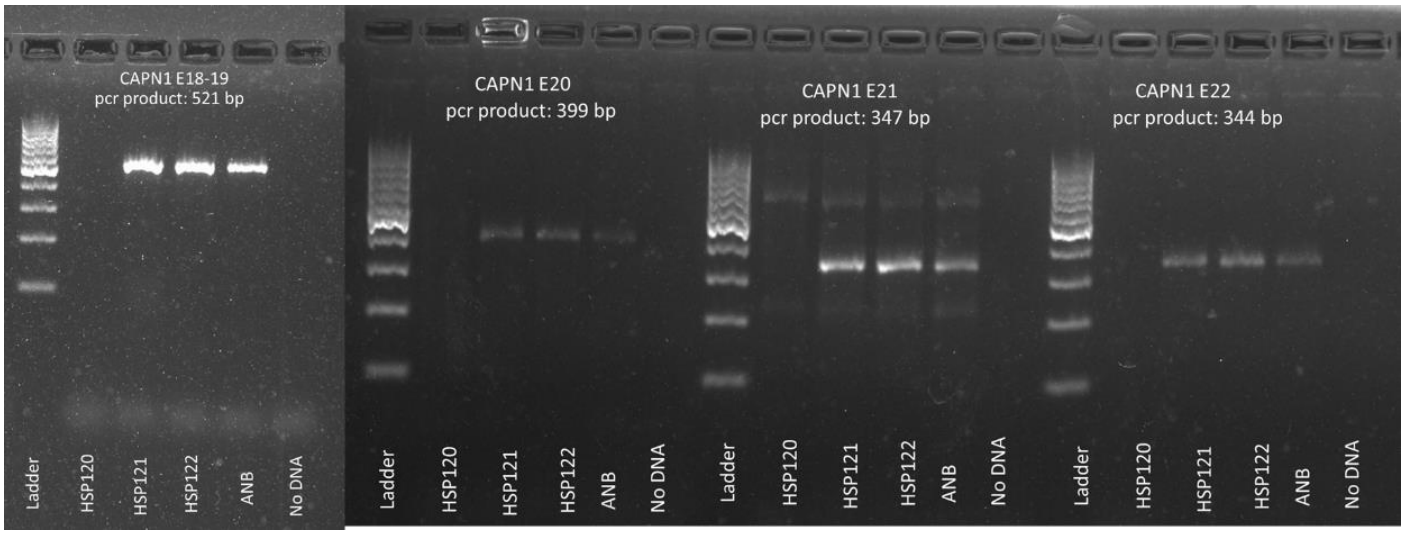


Figure 5.8 a) MCOLN1-fam 11, b) MFSD8-fam 4, c) NKX6-2-fam 25, d) NPC1-fam 6, e) SYT14-fam 24, f) CACNA1A-fam 19, g) STUB1-fam 21.

a)



b)

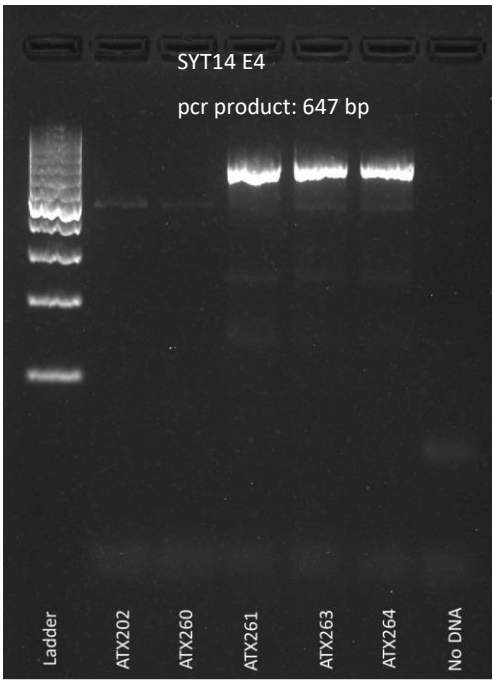


Figure 5.9 Agarose gel electrophoresis of the PCR products of each exon. a) Family 23, CAPN1. HSP120: index patient (P35). HSP121 and HSP122: unaffected siblings. ANB: negative control, b) Family 24, SYT14. ATX202: index case, ATX260: affected sibling, ATX261:asymptomatic sibling, ATX263 and ATX264: parents.

Chapter 6:
DISCUSSION

In this thesis, whole exome sequencing was applied to 83 index patients. Family history suggested AD inheritance in 13% (11/83) and AR inheritance in 69% (57/83). The remaining 18% (15/83) were apparently sporadic. Consanguinity was present in 71% of the cohort under study. Causative variants were identified in 25 out of 68 familial cases studied, which corresponds to a diagnostic yield of 37%. This number is in accordance with the literature (Retterer et al., 2016) (Trujillano et al., 2017).

Heterozygous pathogenic variations were found in a family with autosomal dominant inheritance pattern and in one apparently sporadic patient. Biallelic variants were identified in 24 families.

6.1 Whole Exome Sequencing and the Complex Genetic Architecture of the Cohort

In the cohort consisting of 83 families, 27 causative variants were identified in 18 genes (Figure 5.1, Table 5.1). In a family with an AD inheritance pattern, *CACNA1A* was found while in an apparently sporadic family the *STUB1* gene was identified. In recessive cases, the most common gene was *CAPN1* which was homozygously present in four different families, followed by *SACS* in three families, and *SPG7*, *SPG11*, and *HSD17B4* in two families for each. Variants in *ATM*, *KIF1C*, *VPS13D*, *PNPLA6*, *SYNE1*, *L2HGDH*, *MCOLN1*, *MFSD8*, *NKX6-2*, *NPC1*, and *SYT14* were found once in the remaining families.

6.1.1 Genes in Mitochondrial Metabolism

SACS, *SPG7* and *L2HGDH* are the genes that are known to have a role in mitochondrial metabolism. Biallelic mutations in the *SACS* gene are associated with autosomal recessive spastic ataxia of Charlevoix-Saguenay. Mitochondrial function is secondarily impaired in ARSACS patients. Sacsin protein which is encoded by the *SACS* gene acts as an intermediate filament regulator, thus when sacsin is mutated, mitochondrial fission is impaired (Synofzik et al., 2019). In this thesis, three distinct families with SPAX phenotype were shown to have homozygous variants in the *SACS* gene: one in-frame, one frameshift and one stop gained variant (Figure 6.1). Phenotypes and ages of onset of all three families were in accordance with the ARSACS phenotype.

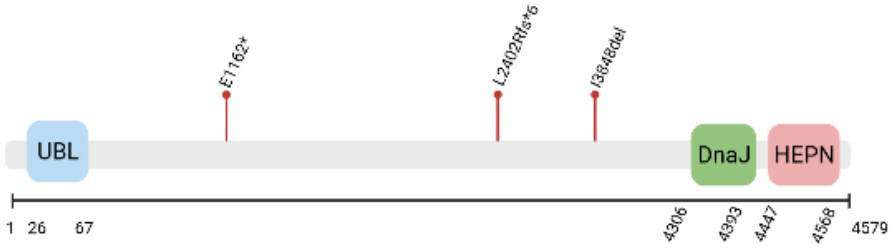


Figure 6.1 Mutations in the sacsin protein are shown in red. UBL: Ubiquitin like domain, HEPN: Higher eukaryotes and prokaryotes nucleotide-binding domain. Positions were taken from InterPro (https://www.ebi.ac.uk/interpro/protein/UniProt/Q9NZJ4/).

SPG7 encodes for the mitochondrial protein paraplegin. It is part of the hetero-oligomeric proteolytic complex so when it is mutated, mitochondrial protein control is impaired (Synofzik et al., 2019). Two homozygous variants in two unrelated families were discovered in the *SPG7* gene in our cohort. Family 18 had a typical SPAX phenotype with a missense variant, whereas family 14 had an ataxia phenotype with a stop gained variant. (Table 3.1). This points to the phenotypic heterogeneity among hereditary spastic paraplegias and ataxias. The SPG7 disease phenotype is variable, some cases present with pure HSP, and some are more complex with other neurological symptoms. Even nearly pure ataxia forms have been seen in SPG7 (Yahikozawa et al., 2015). In our patients, the variant types probably have roles in disease manifestation with or without spasticity.

L-2-hydroxyglutaric aciduria is caused by biallelic variants in the *L2HGDH* gene which encodes for an enzyme called L-2-hydroxyglutarate dehydrogenase, located in mitochondria and participating in reactions for energy production (Rzem et al., 2006). Our family with an *L2HGDH* variant harbored two affected siblings with symptoms like head tremor, gait disturbance, mental retardation, basal ganglion involvement and white and grey matter involvement. The causative variant was previously reported, and the phenotypes of our patients were in line with the literature (Topçu et al., 2004).

6.1.2 Genes in DNA Repair Pathway

The *ATM* gene takes a role in the double-strand break repair mechanism (Synofzik et al., 2019). A family with a compound heterozygous *ATM* variant was identified in our cohort. One of the variants (p.Phe1870LeufsTer47) was shown to have a *de novo* occurrence in the patient. Biallelic variants in the *ATM* gene cause ataxia telangiectasia with cerebellar ataxia, oculocutaneous telangiectasia, variable immunodeficiency, and susceptibility to malignancies. AT has a wide clinical spectrum (Amirifar et al., 2019). In

our patient, there was ataxia, no pathologic eye movement findings and the alpha-fetoprotein level was only slightly higher than normal. The absence of telangiectasia might be due to the young age of the patient. The compound heterozygosity may have a role in the relatively low alpha-fetoprotein levels and in the absence of pathologic movements.

6.1.3 Genes in Lipid Metabolism

HSD17B4, *PNPLA6*, and *NPC1* genes belong to this group. Two independent families with the same homozygous and pathogenic *HSD17B4* variant were identified in the cohort. *HSD17B4* gene is an important player in β-oxidation of very long chain fatty acids (Synofzik et al., 2019). Biallelic mutations in *HSD17B4* are known to cause D-bifunctional protein deficiency and Perrault syndrome I. All patients have hearing loss and ataxia; hence, their phenotype is in accordance with the Perrault syndrome I.

PNPLA6 gene participates in de-esterification of phosphatidylcholine into its constituent fatty acids and glycerophosphocholine (Synofzik et al., 2019). In our cohort, a causative variant was found in the *PNPLA6* gene in three affected siblings (Family 2). Biallelic variants in the *PNPLA6* gene cause Boucher-Neuhäuser syndrome, Gordon Holmes syndrome, Oliver-McFarlane syndrome, Laurence-Moon syndrome, and spastic paraplegia type 39 (SPG39) (Synofzik et al., 1993). In our family, there was spastic ataxia and Holmes tremor with a variant previously reported in our center (Emekli et al., 2021). The genetic finding is in accordance with the phenotype of the patients.

The NPC1 patient in our cohort was referred to our laboratory with ataxia and mental retardation. NPC primarily affects the trafficking of intracellular cholesterol in the brain and peripheral organs (Synofzik et al., 2019). Biallelic variants in the *NPC1* gene cause Niemann-Pick Disease type C. The genetic finding explains the disease phenotype of the patient.

6.1.4 Genes in Autophagy and Lysosomal Activity

SPG11 encodes for the spatacsin protein which was detected in autophagic lysosome reformation (Fraiberg and Elazar, 2020). Biallelic mutations in *SPG11* are known to cause three different phenotypes, SPG11, juvenile ALS or CMT disease (Branchu et al., 2017). Two homozygous variants were found in *SPG11* in two distinct families. One of the variants was a novel frameshift, the other one was a reported missense

variant. In our cohort, the variants resulted in SPG11. Both patients in the cohort have a SPAX phenotype (Table 3.1), which is in accordance with the *SPG11* finding.

6.1.5 Other Genes with Different Mechanisms

CAPN1 acts in proteolysis of substrates involved in cytoskeletal remodeling and signal transduction (Hsu et al., 2011). In four independent families, homozygous mutations in *CAPN1* were present. A missense, a stop-gain, an intronic splice region variant, and a copy number variation of a deletion which encompassed five exons of *CAPN1* (Exon18-exon22 del) were described (Figure 6.2). Except for the large deletion, all mutations are in the calpain catalytic domain. The large exon deletion caused loss of the C-end of the protein. All mutations were either pathogenic or likely pathogenic with high scores of pathogenicity (Table 5.2). Biallelic variants in *CAPN1* are associated with autosomal recessive spastic paraplegia (SPG76) (Gan-Or et al., 2016). Patients in our cohort have both spasticity and ataxia as their main phenotypes, thus, our *CAPN1* findings correlate well with the phenotype. These patients are recently compiled under the rare group of spastic ataxia (SPAX) (Table 3.1) (Shetty et al., 2019).

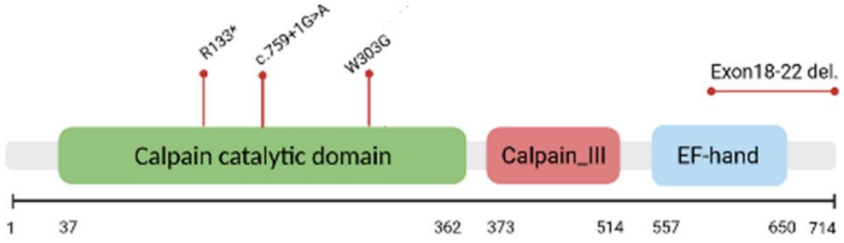


Figure 6.2 Mutations described in calpain-1 in the cohort shown in red. Domains were taken from InterPro (https://www.ebi.ac.uk/interpro/protein/UniProt/P07384/).

Biallelic mutations in *KIF1C* cause spastic ataxia type 2. *KIF1C* functions in retrograde transport of Golgi vesicles to the ER (Dorner et al., 1998). There was one family with two homozygous variants in the *KIF1C* gene. Although both variants were pathogenic or likely pathogenic with high scores (Table 5.2), it is not clear if only one causes the disease phenotype or if both contribute to disease. Since the p.Gln289Pro variant was relatively common in heterozygous state in our in-house database, the rarer c.940+1G>A splice variant may have the major role in disease formation. All three affected siblings had a SPAX phenotype which is in line with the genetic finding.

VPS13D acts in autophagy and mitochondrial clearance (Anding et al., 2018). The variant identified in *VPS13D* was previously reported in compound heterozygosity with a termination variant (Seong et al., 2018). In our family this very variant (Ala4210Val) is

reported for the first time in homozygous state, and it is one of the two rare missense homozygous variants in *VPS13D* reported to date. In the literature, the patients with ataxia are mostly children or young adults. Our patients, two brothers, offspring of consanguineous parents, have AO of 57 and 49. In addition to the late age of onset, the disease phenotype is milder than in the previously reported patient with the Ala4210Val mutation; the variant being homozygous missense may lead to a milder disease phenotype.

The patient from Family 25 was referred to our laboratory with recessive ataxia. He had cerebellar and spinal atrophy. A previously reported biallelic variant in *NKX6-2* gene was found in the patient. Biallelic variants in *NKX6-2* cause spastic ataxia type 8. The protein is involved in the genesis and development of oligodendrocytes (Dorboz et al., 2017). In the literature, a patient with the same variant had dystonia, mild intellectual disability, cerebellar and pyramidal symptoms (Almatrafi et al., 2020). Our patient's phenotype is milder than the previously reported case.

The *SYT14* gene functions in mediating membrane trafficking in synaptic transmission (Sheng et al., 2018). Biallelic variants in *SYT14* cause SCAR11. Our patients were referred to our laboratory with psychomotor retardation and recessive ataxia. After the identification of the biallelic *SYT14* deletion by CNV analysis, the clinician's further investigation unraveled cognitive impairment in the patients. In literature, the ataxia onset is in fifties while our patients' disease onset was in childhood (Doi et al., 2011). This might be due to the deletion. This is the first report of a homozygous exon deletion in the *SYT14* gene. The biallelic CNV may cause a more severe phenotype in our patients than reported in the literature

6.2 Variants of Uncertain Significance

WES in disease diagnosis introduced the question of an ever-increasing number of variants of uncertain significance. Interpretation of these variants are of prime importance for the clinical management of patients. Genotype-phenotype correlations and intrafamilial allele segregation are crucial to define the pathogenicity of VUS. Precise clinical information, deep and reverse phenotyping are significantly important to evaluate these variants.

In the framework of this study, eight out of 27 variants identified, presented with VUS criteria according to Varsome (Table 5.2), four of them being novel. All VUS variants

were extensively discussed with the respective clinicians to put the final and correct diagnosis.

6.3 Novel Genotype-Phenotype Associations

Next generation sequencing techniques have not only greatly contributed to the number of genes associated with ataxia or other ND, but they have also blurred the line among these diseases. In our cohort, whole exome sequencing in 83 families revealed novel genotype-phenotype associations and novel variants in known ataxia genes. These include variants in HSP-related or other genes, as well as identification of novel phenotypes in ataxia genes.

Variants in HSP-related genes were identified in 15 families, which represent more than half of the families solved. Fourteen out of these 15 families had a SPAX phenotype. Only one family (Family 14) with SPG7 had an ATX phenotype without spasticity. Although SPG7 was initially introduced as an HSP gene, with the advancement of NGS technologies today, SPG7 is now known as one of the most common causes of recessive ataxias; in some cohorts, cerebellar features are more prominent than spasticity (Synofzik and Schüle, 2017). We further support this knowledge with our Family 14.

Novel phenotypic features in known ataxia genes were identified in this study. For example, in our Family 24, the *SYT14* gene, which is associated with SCAR11, was the cause of a childhood ataxia, although it was previously reported to cause ataxia in the sixth decade (Doi et al., 2011). The exon deletion present may well be the cause of a more severe ataxia phenotype in this family. Furthermore, our patients (Family 7) are the oldest reported cases with spastic ataxia based on *VPS13D*, and they present with a milder phenotype than the phenotypes reported in the literature (Gauthier et al., 2018; Koh et al., 2020; Seong et al., 2018).

6.4 Limitations of Whole Exome Sequencing

In the framework of this thesis, a diagnostic yield of 37% was achieved by WES. In the remaining 63% of cases, WES has failed to identify mutations in disease-associated genes. Sequencing other members of the families may help to identify some candidate variants, but the problem can also be technical. WES not only lacks the non-coding regions of the genome, but it has also a limited ability to detect structural variations and repeat expansions. Furthermore, the disease may be caused by a variant in a novel gene

that has not been associated with a related disease phenotype. Therefore, new strategies to discover novel genes must be designed.

Its lack of non-coding regions is a major drawback of WES. Mutations in intronic regions of known disease genes may cause the disease phenotype by changing their expression. This may be overcome by applying whole genome sequencing, although it is not yet as cost-efficient as WES is. WGS also brings up additional and huge computational problems. Interpretation of data is more challenging and requires more sophisticated bioinformatic tools. Until WGS becomes more cost-effective and time-saving, expanding the targeted regions of exome capture kits may be a solution.

Low-coverage problem is another technical limitation of WES. When coverage is low for a heterozygous variant, it might be missed due to a few reads of a particular region.

The problem of false positives in WES should also not be underestimated. The reasons might be biases in library construction, erroneous polymerase reactions, difficulty in making genotype calls at the end of short reads, loss of synchrony among DNA sequencing reactions within a cluster or platform-specific problems (Fuentes Fajardo et al., 2012). In the framework of this thesis, there was only one false-positive occurrence. A pathogenic *CACNA1A* variant was identified in an episodic ataxia patient by WES, however, Sanger sequencing devalidated the variant. Since it was a strong mutation which would have explained the disease, we performed restriction enzyme digestion which also refuted the WES result. This shows the importance of validation and family segregation analysis by Sanger sequencing or other relevant techniques for validating the WES results. This will avoid wrong results.

Due to its read length limitation, detecting structural variants is another shortcoming of WES. New algorithms are being developed to detect structural variants from WES data as the read length increases with new technologies. There are several tools to detect CNVs from WES data. However, they result in a high rate of false-positives and because of non-uniform read-depth among captured regions, CNV detection from WES data is not always reliable. In the framework of this thesis, we were able to solve two families with WES-based CNV analysis.

6.5 Remaining Cases to be Solved

In this study, pathogenic mutations were described in 26 families, we failed to identify the genetic causes in the remaining 57 cases. This shows that the classical WES approach is not sufficient to unravel all disease-causing elements.

Limitations and challenges of WES, discussed above, is one factor that is in the way to solve these families.

Another factor may be novel genes. The remaining unsolved cases may uncover novel genes. For this purpose, a deep phenotyping and using the last version of data analysis tools are crucial.

Oligogenic inheritance may be present in these unsolved families, meaning that more than one gene or other risk factors may be contributing to disease phenotype. Thus, identifying epigenetic risk factors in ataxias become another important topic to unravel the causes in some of these families.

6.6 WES is the Golden Standard in Diagnosis of Ataxias

Ataxias are complex and heterogeneous disorders. Their mode of inheritance and clinical presentation vary widely. There are subtypes of ataxias, in terms of involvement of the systems, such as eye involvement, pyramidal involvement or neuropathy. This complexity may confuse the clinicians during neurological examination and may lead to misdiagnosis. Although imaging and metabolic tests are useful to exclude or suspect some subtypes, the exact diagnosis is only possible by molecular diagnosis.

WES is the sequencing of the whole protein-coding area in the genome. It is more time-saving and cost effective than all other conventional methods and targeted sequencing.

Our findings in the cohort under study show a large genetic heterogeneity of ataxias. Some genes are related to neurological conditions other than ataxia. Most ND overlap clinically and genetically, for example, in this study, variants were identified in genes like *L2HGDH* and *NPC1* in patients referred to us with an ataxia clinical diagnosis (Figure 5.1). This implies the power and importance of whole exome sequencing and bioinformatic analysis. WES helps us to identify novel variations and novel genotype-phenotype associations.

The genetic findings in this cohort allowed us to get the big picture on the genetic architecture of ataxias in Turkey and also helped in their differential diagnosis. Exome

sequencing is an unbiased and highly effective technique in identifying the genetic causes of complex and heterogeneous ataxias. WES is the golden standard in examining complex genetic disorders in spite of its limitations and challenges.

Chapter 7:
CONCLUSION

Hereditary ataxias are rare and heterogeneous disorders, the complex genetic backgrounds of which have not been fully understood yet. Genetic and environmental factors must be identified to understand the complete pathogenesis of ataxias, which overlap with other ND and span a large genotypic spectrum.

The aim of this thesis was to investigate the complex genetics of ataxias in Turkey by using whole exome sequencing and bioinformatic analysis. We were able to identify the causative mutations in 26 families and showed the effect of WES in differential diagnosis. Our results indicate the great heterogeneity stemming from the genetic complexity of ataxias on one hand and the ethnic admixture of the Turkish population on the other.

Advances in massive parallel sequencing have greatly contributed to our knowledge of the genetic pathology of ataxias. Although the discovery of new genes and variants grows in an exponential rate, there are still undefined types of ataxias. They are waiting to be discovered by further developments and innovations in sequencing technologies.

The mechanisms leading to cerebellar dysfunction will be exhaustively uncovered and understood once all genes and mutations causing ataxias are known. This will also help to improve personalized medicine approaches.

This thesis is a comprehensive study of the bioinformatic evaluation of the WES data of a Turkish ataxia cohort. We showed that recessive ataxias may occur more frequently because of common consanguineous marriages in Turkey and that WES is an efficient tool to investigate inherited ataxias.

We hope that the results presented here will pave the ways for more precise diagnosis of ataxias in Turkey, eventually contributing to the field of molecular therapies in the era of translational medicine.

BIBLIOGRAPHY

Almatrafi, A., Umair, M., Eldardear, A., Al-Luqmani, M., Hashmi, J. A., Albalawi, A. M., . . . Basit, S. (2020). A homozygous missense variant in the homeobox domain of the NKX6-2 results in progressive spastic ataxia type 8 associated with lower limb weakness and neurological manifestations. *J Gene Med, 22*(8), e3196. doi:10.1002/jgm.3196

Amirifar, P., Ranjouri, M. R., Yazdani, R., Abolhassani, H., & Aghamohammadi, A. (2019). Ataxia-telangiectasia: A review of clinical features and molecular pathology. *Pediatr Allergy Immunol, 30*(3), 277-288. doi:10.1111/pai.13020

Anding, A. L., Wang, C., Chang, T.-K., Sliter, D. A., Powers, C. M., Hofmann, K., . . . Baehrecke, E. H. (2018). Vps13D Encodes a Ubiquitin-Binding Protein that Is Required for the Regulation of Mitochondrial Size and Clearance. *Current Biology, 28*(2), 287-295.e286. doi:10.1016/j.cub.2017.11.064

Anheim, M., Tranchant, C., & Koenig, M. (2012). The autosomal recessive cerebellar ataxias. *The New England journal of medicine, 366*(7), 636-646. doi:10.1056/NEJMra1006610

Beaudin, M., Matilla-Dueñas, A., Soong, B.-W., Pedroso, J. L., Barsottini, O. G., Mitoma, H., . . . Dupre, N. (2019). The Classification of Autosomal Recessive Cerebellar Ataxias: a Consensus Statement from the Society for Research on the Cerebellum and Ataxias Task Force. *Cerebellum (London, England), 18*(6), 1098-1125. doi:10.1007/s12311-019-01052-2

Bereznyakova, O., & Dupré, N. (2018). Spastic ataxias. *Handbook of clinical neurology, 155*, 191-203. doi:10.1016/B978-0-444-64189-2.00012-3

Bouchard, J. P., Barbeau, A., Bouchard, R., & Bouchard, R. W. (1978). Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay. *Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques, 5*(1), 61-69. doi:10.1017/S0317167100024793

Branchu, J., Boutry, M., Sourd, L., Depp, M., Leone, C., Corriger, A., . . . Darios, F. (2017). Loss of spatacsin function alters lysosomal lipid clearance leading to upper and lower motor neuron degeneration. *Neurobiol Dis, 102*, 21-37. doi:10.1016/j.nbd.2017.02.007

Choquet, K., Tétreault, M., Yang, S., La Piana, R., Dicaire, M.-J., Vanstone, M. R., . . . Care4Rare Canada, C. (2016). SPG7 mutations explain a significant proportion of French Canadian spastic ataxia cases. *European Journal of Human Genetics, 24*(7), 1016-1021. doi:10.1038/ejhg.2015.240

Coffey, A. J., Kokocinski, F., Calafato, M. S., Scott, C. E., Palta, P., Drury, E., . . . Palotie, A. (2011). The GENCODE exome: sequencing the complete human exome. *European Journal of Human Genetics, 19*(7), 827-831. doi:10.1038/ejhg.2011.28

Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W. Y., . . . Houlden, H. (2019). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat Genet, 51*(4), 649-658. doi:10.1038/s41588-019-0372-4

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol, 6*(12), e1001025. doi:10.1371/journal.pcbi.1001025

de Bot, S. T., Willemsen, M. A. A. P., Vermeer, S., Kremer, H. P. H., & van de Warrenburg, B. P. C. (2012). Reviewing the genetic causes of spastic-ataxias. *Neurology, 79*(14), 1507-1514. doi:10.1212/WNL.0b013e31826d5fb0

Doi, H., Yoshida, K., Yasuda, T., Fukuda, M., Fukuda, Y., Morita, H., . . . Matsumoto, N. (2011). Exome Sequencing Reveals a Homozygous SYT14 Mutation in Adult-Onset, Autosomal-Recessive Spinocerebellar Ataxia with Psychomotor Retardation. *The American Journal of Human Genetics, 89*(2), 320-327. doi:10.1016/j.ajhg.2011.07.012

Dorboz, I., Aiello, C., Simons, C., Stone, R. T., Niceta, M., Elmaleh, M., . . . Bertini, E. (2017). Biallelic mutations in the homeodomain of NKX6-2 underlie a severe hypomyelinating leukodystrophy. *Brain, 140*(10), 2550-2556. doi:10.1093/brain/awx207 %J Brain

Dorner, C., Ciossek, T., Müller, S., Møller, P. H., Ullrich, A., & Lammers, R. (1998). Characterization of KIF1C, a new kinesin-like protein involved in vesicle transport from the Golgi apparatus to the endoplasmic reticulum. *J Biol Chem, 273*(32), 20267-20275. doi:10.1074/jbc.273.32.20267

Durr, A. (2010). Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *The Lancet. Neurology, 9*(9), 885-894. doi:10.1016/S1474-4422(10)70183-6

Emekli, A. S., Samanci, B., Şimşir, G., Hanagasi, H. A., Gürvit, H., Bilgiç, B., & Başak, A. N. (2021). A novel PNPLA6 mutation in a Turkish family with intractable Holmes tremor and spastic ataxia. *Neurol Sci, 42*(4), 1535-1539. doi:10.1007/s10072-020-04869-6

Engert, J. C., Bérubé, P., Mercier, J., Doré, C., Lepage, P., Ge, B., . . . Richter, A. (2000). ARSACS, a spastic ataxia common in northeastern Québec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nature Genetics, 24*(2), 120-125. doi:10.1038/72769

Fogel, B. L., Cho, E., Wahnich, A., Gao, F., Becherel, O. J., Wang, X., . . . Coppola, G. (2014). Mutation of senataxin alters disease-specific transcriptional networks in patients with ataxia with oculomotor apraxia type 2. *Human Molecular Genetics, 23*(18), 4758-4769. doi:10.1093/hmg/ddu190 %J Human Molecular Genetics

Fogel, B. L., Satya-Murti, S., & Cohen, B. H. (2016). Clinical exome sequencing in neurologic disease. *Neurology. Clinical practice, 6*(2), 164-176. doi:10.1212/CPJ.0000000000000239

Foo, J.-N., Liu, J.-J., & Tan, E.-K. (2012). Whole-genome and whole-exome sequencing in neurological diseases. *Nature reviews. Neurology, 8*(9), 508-517. doi:10.1038/nrneurol.2012.148

Fraiberg, M., & Elazar, Z. (2020). Chapter Thirteen - Genetic defects of autophagy linked to disease. In A. B. Martinez & L. Galluzzi (Eds.), *Progress in Molecular Biology and Translational Science* (Vol. 172, pp. 293-323): Academic Press.

Fuentes Fajardo, K. V., Adams, D., Mason, C. E., Sincan, M., Tifft, C., Toro, C., . . . Markello, T. (2012). Detecting false-positive signals in exome sequencing. *Hum Mutat, 33*(4), 609-613. doi:10.1002/humu.22033

Gan-Or, Z., Bouslam, N., Birouk, N., Lissouba, A., Chambers, Daniel B., Vérièpe, J., . . . Ouled Amar Bencheikh, B. (2016). Mutations in CAPN1 Cause Autosomal-Recessive Hereditary Spastic Paraplegia. *The American Journal of Human Genetics, 98*(5), 1038-1046. doi:10.1016/j.ajhg.2016.04.002

Gauthier, J., Meijer, I. A., Lessel, D., Mencacci, N. E., Krainc, D., Hempel, M., . . . Campeau, P. M. (2018). Recessive mutations in VPS13D cause childhood onset movement disorders. *Ann Neurol, 83*(6), 1089-1095. doi:10.1002/ana.25204

Hsu, C. Y., Henry, J., Raymond, A. A., Méchin, M. C., Pendaries, V., Nassar, D., . . . Simon, M. (2011). Deimination of human filaggrin-2 promotes its proteolysis by calpain 1. *J Biol Chem, 286*(26), 23222-23233. doi:10.1074/jbc.M110.197400

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., . . . Sieh, W. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *The American Journal of Human Genetics, 99*(4), 877-885. doi:10.1016/j.ajhg.2016.08.016

Jayadev, S., & Bird, T. D. (2013). Hereditary ataxias: overview. *Genet Med, 15*(9), 673-683. doi:10.1038/gim.2013.28

Jones, T. M., Shaw, J. D., Sullivan, K., & Zesiewicz, T. A. (2014). Treatment of cerebellar ataxia. *Neurodegenerative disease management, 4*(5), 379-392. doi:10.2217/nmt.14.27

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Res, 12*(6), 996-1006. doi:10.1101/gr.229102

Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res, 35*(Web Server issue), W43-46. doi:10.1093/nar/gkm234

Klockgether, T. (2010). Sporadic ataxia with adult onset: classification and diagnostic criteria. *The Lancet. Neurology, 9*(1), 94-104. doi:10.1016/S1474-4422(09)70305-9

Klockgether, T., Mariotti, C., & Paulson, H. L. (2019). Spinocerebellar ataxia. *Nat Rev Dis Primers, 5*(1), 24. doi:10.1038/s41572-019-0074-3

Koeppen, A. H. (2018). The neuropathology of the adult cerebellum. *Handbook of clinical neurology, 154*, 129-149. doi:10.1016/B978-0-444-63956-1.00008-4

Koh, K., Ishiura, H., Shimazaki, H., Tsutsumiuchi, M., Ichinose, Y., Nan, H., . . . Takiyama, Y. (2020). VPS13D-related disorders presenting as a pure and complicated form of hereditary spastic paraplegia. *8*(3), e1108. doi:https://doi.org/10.1002/mgg3.1108

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., & Massouras, A. (2018). VarSome: the human genomic variant search engine. *Bioinformatics, 35*(11), 1978-1980. doi:10.1093/bioinformatics/bty897 %J Bioinformatics

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., . . . Kattman, B. L. (2020). ClinVar: improvements to accessing data. *Nucleic Acids Res, 48*(D1), D835-d844. doi:10.1093/nar/gkz972

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., . . . Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285-291. doi:10.1038/nature19057

Marian, A. J. (2014). Sequencing your genome: what does it mean? *Methodist DeBakey cardiovascular journal, 10*(1), 3-6. doi:10.14797/mdcj-10-1-3

Mariani, L. L., Rivaud-Péchoux, S., Charles, P., Ewenczyk, C., Meneret, A., Monga, B. B., . . . Anheim, M. (2017). Comparing ataxias with oculomotor apraxia: a multimodal study of AOA1, AOA2 and AT focusing on video-oculography and alpha-fetoprotein. *Scientific Reports, 7*(1), 15284. doi:10.1038/s41598-017-15127-9

Mundwiler, A., & Shakkottai, V. G. (2018). Autosomal-dominant cerebellar ataxias. *Handb Clin Neurol, 147*, 173-185. doi:10.1016/b978-0-444-63233-3.00012-9

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res, 31*(13), 3812-3814. doi:10.1093/nar/gkg509

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., . . . Bamshad, M. J. (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics, 42*(1), 30-35. doi:10.1038/ng.499

Oba, D., Hayashi, M., Minamitani, M., Hamano, S., Uchisaka, N., Kikuchi, A., . . . Mizutani, S. (2010). Autopsy study of cerebellar degeneration in siblings with ataxia-telangiectasia-like disorder. *Acta neuropathologica, 119*(4), 513-520. doi:10.1007/s00401-010-0639-4

Parfitt, D. A., Michael, G. J., Vermeulen, E. G. M., Prodromou, N. V., Webb, T. R., Gallo, J.-M., . . . Chapple, J. P. (2009). The ataxia protein sacsin is a functional co-chaperone that protects against polyglutamine-expanded ataxin-1. *Human molecular genetics, 18*(9), 1556-1565. doi:10.1093/hmg/ddp067

Przedborski, S., Vila, M., & Jackson-Lewis, V. (2003). Neurodegeneration: what is it and where are we? *J Clin Invest, 111*(1), 3-10. doi:10.1172/jci17522

Quang, D., Chen, Y., & Xie, X. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics, 31*(5), 761-763. doi:10.1093/bioinformatics/btu703

Rabbani, B., Tekin, M., & Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics, 59*(1), 5-15. doi:10.1038/jhg.2013.114

Rentzsch, P., Schubach, M., Shendure, J., & Kircher, M. (2021). CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med, 13*(1), 31. doi:10.1186/s13073-021-00835-9

Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., . . . Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet Med, 18*(7), 696-704. doi:10.1038/gim.2015.148

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . on behalf of the, A. L. Q. A. C. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine, 17*(5), 405-423. doi:10.1038/gim.2015.30

Rzem, R., Van Schaftingen, E., & Veiga-da-Cunha, M. (2006). The gene mutated in l-2-hydroxyglutaric aciduria encodes l-2-hydroxyglutarate dehydrogenase. *Biochimie, 88*(1), 113-116. doi:10.1016/j.biochi.2005.06.005

Seong, E., Insolera, R., Dulovic, M., Kamsteeg, E. J., Trinh, J., Brüggemann, N., . . . Burmeister, M. (2018). Mutations in VPS13D lead to a new recessive ataxia with spasticity and mitochondrial defects. *Ann Neurol, 83*(6), 1075-1088. doi:10.1002/ana.25220

Sheng, B., Jiang, Y., Wu, D., Lai, N., Ye, Z., Zhang, B., . . . Xu, S. (2018). RNAi-mediated SYT14 knockdown inhibits the growth of human glioma cell line U87MG. *Brain Research Bulletin, 140*, 60-64. doi:https://doi.org/10.1016/j.brainresbull.2018.04.002

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res, 29*(1), 308-311. doi:10.1093/nar/29.1.308

Shetty, A., Gan-Or, Z., Ashtiani, S., Ruskey, J. A., van de Warrenburg, B., Wassenberg, T., . . . Suchowersky, O. (2019). CAPN1 mutations: Expanding the CAPN1-related phenotype: From hereditary spastic paraparesis to spastic ataxia. *Eur J Med Genet, 62*(12), 103605. doi:10.1016/j.ejmg.2018.12.010

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current protocols in molecular biology, 122*(1), e59-e59. doi:10.1002/cpmb.59

Stewart, G. S., Maser, R. S., Stankovic, T., Bressan, D. A., Kaplan, M. I., Jaspers, N. G., . . . Taylor, A. M. (1999). The DNA double-strand break repair gene hMRE11 is mutated in individuals with an ataxia-telangiectasia-like disorder. *Cell, 99*(6), 577-587. doi:10.1016/s0092-8674(00)81547-0

Subramony, S. H., Moscovich, M., & Ashizawa, T. (2015). Chapter 62 - Genetics and Clinical Features of Inherited Ataxias. In M. S. LeDoux (Ed.), *Movement Disorders (Second Edition)* (pp. 939-978). Boston: Academic Press.

Synofzik, M., Hufnagel, R. B., & Züchner, S. (1993). PNPLA6 Disorders. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, G. Mirzaa, & A. Amemiya (Eds.), *GeneReviews(®)*. Seattle (WA): University of Washington, Seattle

Copyright © 1993-2021, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.

Synofzik, M., & Németh, A. H. (2018). Recessive ataxias. *Handbook of clinical neurology, 155*, 73-89. doi:10.1016/B978-0-444-64189-2.00005-6

Synofzik, M., Puccio, H., Mochel, F., & Schöls, L. (2019). Autosomal Recessive Cerebellar Ataxias: Paving the Way toward Targeted Molecular Therapies. *Neuron, 101*(4), 560-583. doi:10.1016/j.neuron.2019.01.049

Synofzik, M., & Schüle, R. (2017). Overcoming the divide between ataxias and spastic paraplegias: Shared phenotypes, genes, and pathways. *32*(3), 332-345. doi:https://doi.org/10.1002/mds.26944

Taylor, A. M. R., Groom, A., & Byrd, P. J. (2004). Ataxia-telangiectasia-like disorder (ATLD)—its clinical presentation and molecular basis. *DNA Repair, 3*(8), 1219-1225. doi:https://doi.org/10.1016/j.dnarep.2004.04.009

Topçu, M., Jobard, F., Halliez, S., Coskun, T., Yalçinkayal, C., Gerceker, F. O., . . . Fischer, J. (2004). L-2-Hydroxyglutaric aciduria: identification of a mutant gene C14orf160, localized on chromosome 14q22.1. *Hum Mol Genet, 13*(22), 2803-2811. doi:10.1093/hmg/ddh300

Trujillano, D., Bertoli-Avella, A. M., Kumar Kandaswamy, K., Weiss, M. E., Köster, J., Marais, A., . . . Abou Jamra, R. (2017). Clinical exome sequencing: results from 2819 samples reflecting 1000 families. *Eur J Hum Genet, 25*(2), 176-182. doi:10.1038/ejhg.2016.146

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in genetics : TIG, 34*(9), 666-681. doi:10.1016/j.tig.2018.05.008

Vermeer, S., van de Warrenburg, B. P. C., Willemsen, M. A. A. P., Cluitmans, M., Scheffer, H., Kremer, B. P., & Knoers, N. V. A. M. (2011). Autosomal recessive cerebellar ataxias: the current state of affairs. *Journal of medical genetics, 48*(10), 651-659. doi:10.1136/jmedgenet-2011-100210

Vural, A., Şimşir, G., Tekgül, Ş., Koçoğlu, C., Akçimen, F., Kartal, E., . . . Başak, A. N. (2021). The Complex Genetic Landscape of Hereditary Ataxias in Turkey and Implications in Clinical Practice. *Mov Disord*. doi:10.1002/mds.28518

Yahikozawa, H., Yoshida, K., Sato, S., Hanyu, N., Doi, H., Miyatake, S., & Matsumoto, N. (2015). Predominant cerebellar phenotype in spastic paraplegia 7 (SPG7). *Hum Genome Var, 2*, 15012. doi:10.1038/hgv.2015.12

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics, 13*, 134. doi:10.1186/1471-2105-13-134

Yohe, S., & Thyagarajan, B. (2017). Review of Clinical Next-Generation Sequencing. *Archives of pathology & laboratory medicine, 141*(11), 1544-1557. doi:10.5858/arpa.2016-0501-RA

APPENDIX A

The papers published and prepared throughout this thesis:

1. Palvadeau, R., Kaya-Güleç, Z.E., **Şimşir, G**. et al. Cerebellar cognitive-affective syndrome preceding ataxia associated with complex extrapyramidal features in a Turkish SCA48 family. Neurogenetics 21, 51–58 (2020). https://doi.org/10.1007/s10048-019-00595-0

2. Emekli, A.S., Samanci, B., **Şimşir, G**. et al. A novel PNPLA6 mutation in a Turkish family with intractable Holmes tremor and spastic ataxia. Neurological Sciences (2020). https://doi.org/10.1007/s10072-020-04869-6

3. Vural, A., **Şimşir, G.**, Tekgül, Ş. et al. The Complex Genetic Landscape of Hereditary Ataxias in Turkey and Implications in Clinical Practice. Movement Disorders (2021). https://doi.org/10.1002/mds.28518

4. Çakar, A., İnci, M., Özdağ-Acarlı, A.N., Çomu, S., Candayan, A., Battaloğlu, E., **Şimşir, G.** et al. Expanding the phenotypical spectrum of SACS variants: Neuromuscular perspective of a complex neurodegenerative disorder (submitted to Neurological Sciences).

5. Baydemir, R., Yılmaz, G., **Şimşir, G**. et al. Varied phenotypic spectrum presenting of paroxysmal exercise-induced dyskinesia: A Turkish family with SLC2A1 mutation (submitted to Neurological Sciences).

6. Öztop-Çakmak, Ö., **Şimşir, G**., Aygün, M.S. et al. A homozygous VPS13D missense mutation in two brothers causes pre-senile-onset cerebellar ataxia with spasticity and hearing loss: DTI reveals preferential affection of fronto-pontine fibers (in preparation)