The Molecular Architecture of Ataxias in Turkey

by

Şeyma Tekgül

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

The Molecular Architecture of Ataxias in Turkey

Koç University

Graduate School of Sciences and Engineering

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

Şeyma Tekgül

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)

Doç. Dr. Atay Vural

Doç. Dr. Gençer Genç

Date:

to everyone who gives meaning to my life…

ABSTRACT

The Molecular Architecture of Ataxias in Turkey

Şeyma Tekgül

Master of Science in Molecular Biology and Genetics

Ataxias are a clinically, genetically, and mechanistically heterogeneous group of neurological disorders characterized by motor incoordination, resulting from dysfunction of the cerebellum and its connections. Ataxias with different phenotypes overlap and make a precise clinical diagnosis challenging. In recent years, advances in next-generation sequencing (NGS) have drastically contributed to the molecular diagnosis of ataxias. With whole exome sequencing (WES), one of the NGS technologies, it has become possible to analyse the coding regions of the genome in a time-saving and cost-effective manner. WES, which is a state-of-the-art method, has evolved into an efficient tool for identifying genetic causes of complex ataxias. In the framework of this thesis, 95 index patients with complex ataxia phenotypes were investigated by WES. Causative mutations in 28 different genes were identified in 40 families, this corresponds to a diagnostic yield of 42%. A significant overlap among different neurodegenerative disorders has been demonstrated by identifying mutations in hereditary spastic paraplegias and other neurodegenerative disease genes. Rapidly improving sequencing technologies, collaborative projects, and detailed clinical information will help to identify disease genes in remaining unsolved families. The results presented in this thesis will hopefully contribute to pave the ways for more definitive diagnosis of ataxias in the future and to developing molecular therapies.

ÖZETÇE

Türkiye'de Ataksilerin Moleküler Yapısı

Şeyma Tekgül

Moleküler Biyoloji ve Genetik, Yüksek Lisans

Beyincik ve bağlantılarının işlev bozukluğundan kaynaklanan motor koordinasyon hataları ile karakterize olan ataksiler klinik, genetik ve mekanistik olarak heterojen bir nörolojik hastalık grubudur. Farklı fenotipleri olan ataksilerin örtüşen özellikleri, ayırıcı tanıyı zorlaştırmaktadır. Son yıllarda, yeni nesil dizilemedeki gelişmeler, ataksilerin moleküler tanısına büyük ölçüde katkıda bulunmuştur. NGS teknolojilerinden biri olan tüm ekzom dizileme ile genomun kodlayıcı bölgelerini düşük maliyetle ve kısa sürede etkin bir şekilde analiz etmek mümkün hale gelmiştir. Güncel bir yöntem olan tüm ekzom dizileme, karmaşık ataksilerin genetik nedenlerini belirlemek için etkili bir araç haline gelmiştir. Bu tez çerçevesinde, kompleks ataksi fenotipi olan 95 indeks hasta tüm ekzom dizileme ile incelendi. Toplam 40 ailede 28 farklı gende bulunan mutasyonlar hastalık nedeni olarak tanımlandı. Tez çerçevesinde tüm ekzom dizilemeden %42 oranında verim elde edildi. Bulunan genler sadece ataksi değil, kalıtsal spastik paraplejiler ve benzeri hastalıklara da neden olduğu için, tez çerçevesinde farklı nörodejeneratif bozukluklar arasında önemli bir örtüşme gösterilmiş oldu. Hızla gelişen dizileme teknolojileri ile birlikte, uluslararası katılımlı projeler ve ayrıntılı klinik bilgiler, çözülmemiş ailelerdeki hastalık genlerinin belirlenmesine yardımcı olacaktır. Bu çalışmada elde edilen sonuçlar, Türkiye'deki ataksilerin moleküler temeline işaret etmektedir Sonuçların ataksilerin moleküler temeline ışık tutmasının yanında, gelecekte tanıların daha kesinleşmesine ve translasyonel tıp araştırmalarında moleküler tedavilerin geliştirilmesine katkıda bulunması umulmaktadır.

ACKNOWLEDGEMENTS

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

I would like to further extend my thanks to Doç. Dr. Atay Vural and Doç. Dr. Gençer Genç for devoting their time to evaluate my thesis.

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

I deeply thank all current and former NDAL members for a productive working environment and their sincere friendship. I extend my thanks to Gülşah for her friendship and for being an excellent research partner. Being a part of the NDAL team has been a life-changing experience.

I deeply thank Dilara for her sincere support and friendship during my graduate education. I would also like to thank Dorukan for his endless support and unconditional love.

Last but not least, I sincerely thank my parents, Hediye and Hüseyin, and my elder sister Gül for their endless support in all my decisions and unconditional love. Nothing would have been possible without them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AAA	ATPase family associated with various cellular activities domain
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal Dominant
ADCAs	Autosomal Dominant Cerebellar Ataxias
ADCK3	AARF Domain-Containing Kinase 3
ADDH	Ataxia, Delayed Dentition, and Hypomyelination
AFG3L2	ATPase Family Gene 3-like 2
AFP	Alpha-Fetoprotein
ALS	Amyotrophic Lateral Sclerosis
ANO10	Anoctamin 10
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
ATCAY	Caytaxin
ATLD	AT like Disorder
ATM	Ataxia-Telangiectasia Mutated Gene
ATN1	Atrophin 1
ATP13A2	ATPase 13A2
ATX	Ataxia
ATXN	Ataxin
AVED	Ataxia with Isolated Vitamin E Deficiency
B	Benign
BEAN1	Brain-Expressed, Associated with Nedd4, 1
BSCL2	Seipin
C19orf12	Chromosome 19 Open Reading Frame 12
CACNA1A	Calcium Channel, Voltage-Dependent, P/Q Type, Alpha-1a Subunit
CADD	Combined Annotation-Dependent Depletion
CANVAS	Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome

CAPN1	Calpain 1
CCDC88C	Coiled-Coil Domain-Containing Protein 88C
CH	Cerebellar Hypoplasia
CHIP-seq	Chromatin Immunoprecipitation Sequencing
Chr	Chromosome
CMT	Charcot-Marie-Tooth
CNS	Central Nervous System
CNVs	Copy Number Variations
Conc	Concentration
COX20	Cytochrome c Oxidase Assembly Factor COX20
CTX	Cerebrotendinous Xanthomatosis
CYP27A1	Cytochrome P450, Subfamily XXVIIA
D	Damaging
DNA	Deoxyribonucleic Acid
DYS	Dystonia
EA	Episodic Ataxia
ExAC	Exome Aggregation Consortium
F	Female
Fam	Family
FGF14	Fibroblast Growth Factor 14
FRDA	Friedreich's Ataxia
FXN	Frataxin
GAN	Gigaxonin
GAN1	Giant Axonal Neuropathy-1
GERP	Genomic Evolutionary Rate Profiling
gnomAD	The Genome Aggregation Database
HEPN	Higher eukaryotes and prokaryotes nucleotide-binding domain
het	Heterozygous
HGP	Human Genome Project
hom	Homozygous
HSP	Hereditary Spastic Paraplegias
IGV	Integrative Genomics Viewer

IMNEPD1	Neurologic, Endocrine, and Pancreatic Disease, Multisystem, Infantile-Onset 1
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
LP	Likely Pathogenic
M	Male
MC4DN11	Mitochondrial Complex IV Deficiency, Nuclear Type 11
MGCA3	3-Methylglutaconic Aciduria, Type III
MRE11	MRE11 Homolog, Double-Strand Break Repair Nuclease
MRI	Magnetic Resonance Imaging
n	Novel
n.a	Not Available
NADGP	Neurodegeneration with Ataxia, Dystonia, and Gaze Palsy, Childhood-Onset
NBIA4	Neurodegeneration with Brain Iron Accumulation 4
NDs	Neurodegenerative Diseases
NGS	Next Generation Sequencing
NIPA1	Nipa Magnesium Transporter 1
NOP56	NOP56 Ribonuclear Protein
OMA	Oculomotor Apraxia
OPA3	Outer Mitochondrial Membrane Lipid Metabolism Regulator OPA3
P	Patient
Path	Pathogenic
PCH2D	Pontocerebellar Hypoplasia Type 2D
PCR	Polymerase Chain Reaction
PD	Parkinson’s Disease
PNP	Polyneuropathy
PNS	Peripheral Nervous System
POLR3A	Polymerase III, RNA, Subunit A
Polymorp	Polymorphism
PPP2R2B	Protein Phosphatase 2, Regulatory Subunit B, Beta
PRKCG	Protein Kinase C, Gamma

PTRH2	Peptidyl-tRNA Hydrolase 2
RD	Rare Diseases
REVEL	Rare Exome Variant Ensemble Learner
RFC1	Replication Factor C, Subunit 1
RNA	Ribonucleic Acid
SACS	Sacsin
SCAN1	Spinocerebellar Ataxia, Autosomal Recessive, with Axonal Neuropathy 1
SCAR	Spinocerebellar Ataxia, Autosomal Recessive
SCAs	Spinocerebellar Ataxias
Seg	Segregation
SEPSECS	O-Phosphoserine tRNA-Selenocysteine tRNA Synthase
SETX	Senataxin
SIFT	Sorting Intolerant from Tolerant
SMA	Spinal Muscular Atrophy
SPAX	Spastic Ataxia
SPG	Spastic Paraplegia
SPG11	Spatacsin
SPG7	Paraplegin
SPTBN2	Spectrin, Beta, Nonerythrocytic, 2
SQSTM1	Sequestosome 1
STUB1	Stip1 Homologous and U Box-Containing Protein 1
SV	Structural Variations
SYNE1	Spectrin Repeat-Containing Nuclear Envelope Protein 1
T	Tolerated
TBE	Tris/Borate/EDTA
TBP	TATA Box-Binding Protein-Like Protein 1
TDP1	Tyrosyl-DNA Phosphodiesterase 1
TGM6	Transglutaminase 6
Tm	Melting Temperature
TNR	Trinucleotide Repeat
TTBK2	Tau Tubulin Kinase 2
TTPA	Tocopherol Transfer Protein, Alpha
UBL	Ubiquitin like domain

UV	Ultraviolet
VUS	Variants of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Chapter 1:
INTRODUCTION

Neurodegenerative diseases (NDs) are a diverse group of progressive neurological conditions that affect specific neuronal cells in the central and peripheral nervous systems. The loss of structure and functions of neurons that impair motor or cognitive abilities causes irreversible symptoms. NDs are generally late-onset; nevertheless, early-onset forms also exist, especially in families with consanguineous marriages. Even though most ND cases are sporadic, a small group with variable genetic backgrounds presents according to the laws of Mendelian inheritance. The most common NDs are Alzheimer's disease and Parkinson's disease (PD), followed by amyotrophic lateral sclerosis (ALS). The degeneration of specific neuronal cells causes cognitive impairment and/or movement disorders along with various neuropathological changes. Neurodegenerative diseases are subdivided into four main categories based on the affected region of the central nervous system (CNS): diseases of the (i) cerebral cortex, (ii) basal ganglia, (iii) spinal cord or brain stem, and (iv) cerebellum. While movement disorders, including PD, result from basal ganglia degeneration, Alzheimer's disease is associated with cerebral cortex. ALS and spinal muscular atrophy (SMA), also severe nervous system diseases, are characterized by deterioration of the neurons in the brain cortex and spinal cord (ALS) and only spinal cord (SMA) (Przedborski et al., 2003). Ataxia, one of the heterogenous movement disorders, is caused by neurodegeneration in the cerebellum. The subject of this thesis is diseases of the cerebellum, which are complicated, due to clinical overlaps with disorders in which connections in the cerebellum are damaged. Moreover, complicating conditions can also vary depending on the individual's genetic background or environmental factors.

1.1 Ataxias

The term ataxia, originating from the Greek meaning 'loss of order', is clinically used for a group of NDs, characterized by cerebellar degeneration in which the predominant clinical phenotype is ataxia. There are various symptoms depending on the location of the degeneration. These symptoms are broad-based gait, incoordination, lack of balance, dysarthria, and injured eye movement.

Ataxias are a complex group of disorders that can arise from both genetic and non-genetic origins. Ataxias are classified as acquired, sporadic, or hereditary depending on

the background, manifestation, and progression of the disease. Acquired ataxias may occur upon damage to the cerebellum due to extreme alcohol consumption, infections, brain tumors, operation, vitamin deficiencies, congenital abnormalities, and exposure to toxins. Sporadic ataxias are late-onset ataxias without an explicit family history. Hereditary ataxias, which are characterized by early-onset and positive family history, may occur as sporadic diseases with adult-onset. Familial ataxias can be divided into autosomal dominant (AD), autosomal recessive (AR), X-linked, and mitochondrial disorders (Klockgether, 2010; Jones et al., 2014).

1.1.1 Autosomal Dominant Cerebellar Ataxias

Autosomal dominant cerebellar ataxias (ADCAs), also known as spinocerebellar ataxias (SCAs), are a range of progressive NDs that substantially manifest after the third decade. ADCAs with the estimated prevalence of 1-5/100,000 are described by dysfunction and atrophy of the cerebellum and brain stem, which can occur alongside pyramidal or extrapyramidal symptoms or cognitive decline. The primary systematic classification of ADCAs was introduced in 1993 by Harding, who separated them into three major groups. ADCA I (first group) includes SCAs with degeneration of other nervous systems; ADCA II (second group), represented by SCA7, shows retinal dysfunction; and the last group, ADCA III, refers to pure cerebellar ataxia with SCA6 as the most prevalent representative (Mundwiler et al., 2018).

SCAs, which have more than 40 subtypes, are divided in two main groups according to the type of variants giving rise to them: repeat expansion and conventional (Table 1.1). While mainly repeat expansions cause ADCAs, point mutations, are also a genetically detected small group. The most common groups of ADCAs, also known as polyglutamine disorders, are exonic glutamine (CAG) trinucleotide repeat (TNR) expansions. These diseases are observed when the size of the CAG trinucleotide rises above a certain threshold. In this group, there is a reverse correlation between the size of the repeat and the age of disease onset and disease severity (Klockgether et al., 2019).

Table 1.1 Autosomal Dominant Cerebellar Ataxias, adapted from (Klockgether et al., 2019), (Jayadev et al., 2013) and (Durr, 2010).

Disease OMIM	**Gene**	**Mean AO (decade)**	**Clinical symptoms**
Polyglutamine expansions			
SCA1 (# 164400)	ATXN1	3rd-4th	spasticity, ophthalmoplegia, bulbar and sensory symptoms, peripheral neuropathy
SCA2 (# 183090)	ATXN2	3rd-4th	decreased DTR, dementia, slow eye movements
SCA3 (# 109150)	ATXN3	4th	extrapyramidal signs and pyramidal, spasticity, fasciculations, sensory loss
SCA6 (# 183086)	CACNA1A	5th-6th	pure cerebellar ataxia and nystagmus
SCA7 (# 164500)	ATXN7	3rd-4th	visual loss, ophthalmoplegia, and spasticity, dementia
SCA17 (# 607136)	TBP	4th	chorea, dystonia, myoclonus, epilepsy, and psychiatric diseases
DRPLA (# 125370)	ATN1	4th	chorea, seizures, dementia, myoclonus, epilepsy
Non-coding expansions			
SCA8 (# 608768)	ATXN8	4th	spasticity, sensory symptoms, psychiatric findings
SCA10 (# 603516)	ATXN10	4th	epilepsy
SCA12 (# 604326)	PPP2R2B	4th	tremor, dementia
SCA31 (# 117210)	BEAN1	5th-6th	pure cerebellar ataxia
SCA36 (# 614153)	NOP56	5th	hearing loss

Conventional ADCAs			
SCA5 (# 600224)	SPTBN2	3rd-4th	pure cerebellar ataxia
SCA11 (# 604432)	TTBK2	3rd	mild, remain ambulatory
SCA13 (# 605259)	KCNC3	childhood or adulthood	intellectual disability
SCA14 (# 605361)	PRKCG	3rd-4th	myoclonus
SCA15/16 (# 606658)	ITPR1	4th	pure cerebellar ataxia
SCA19/22 (# 607346)	KCND3	4th	myoclonus, hyperreflexia, slowly progressive, cognitive impairment
SCA27 (# 609307)	FGF14	2nd	tremor, dyskinesia, cognitive impairment
SCA28 (# 610246)	AFG3L2	2nd	spasticity, ophthalmoplegia, ptosis
SCA35 (# 613908)	TGM6	4th	hyperreflexia, Babinski signs

1.1.2 Autosomal Recessive Cerebellar Ataxias

Autosomal recessive cerebellar ataxias (ARCAs) with a prevalence of 3-6/100,000 comprise a heterogeneous group of rare degenerative and metabolic genetic diseases. In contrast to ADCAs, these diseases, which vary broadly among clinical subtypes, begin early in life. The hallmark of ARCAs is the progressive deterioration of the cerebellum and its related tracts. ARCAs usually appear as sporadic cases (Synofzik et al., 2018). Patients may present with peripheral neuropathy, chorea, dystonia, oculomotor abnormalities, spasticity, cognitive impairment, or epilepsy as well as cerebellar ataxia (Anheim et al., 2012). These phenotypic heterogeneities complicate the clinical identification and diagnosis of ARCAs (Vermeer et al., 2011).

ARCAs are categorized based on their clinical characteristics, neuropathology, or mechanism of pathogenicity. Mitochondrial dysfunction, impaired DNA repair, and

complex lipid metabolism are the three main mechanisms underlying ARCAs. Protein misfolding and chaperone dysfunction, misplacement synaptic myonuclei, and calcium-mediated chloride channel dysfunction are other possible mechanisms in ARCAs (Vermeer et al., 2011; Synofzik et al., 2019). While there are specific subtypes with certain causative genes, several additional genes play a role in the same mechanism, resulting in overlapping phenotypes (Table 1.2).

Table 1.2 Examples of Primary Autosomal Recessive Cerebellar Ataxias, adapted from (Synofzik et al., 2019), (Cortese et al., 2019), (Jayadev at al., 2013) and (Anheim et al., 2012).

Disease OMIM	**Gene**	**Mean AO (decade)**	**Clinical symptoms**
Repeat expansions			
Friedreich's ataxia (FRDA) (# 229300)	FXN	1st-2nd	sensory ataxia, neuropathy, pyramidal weakness, cardiomyopathy, diabetes
CANVAS (# 614575)	RFC1	5th	cerebellar ataxia, neuropathy, vestibular areflexia
Non-repeat ARCAs			
ARSACS (# 270550)	SACS	childhood	cerebellar ataxia, lower limb spasticity, pyramidal tract damage, peripheral neuropathy
ARCA1 (# 610743)	SYNE1	3rd	pure cerebellar ataxia
AOA1 (# 208920)	APTX	childhood	oculomotor apraxia, chorea, dystonia, axonal sensorimotor neuropathy, cognitive impairment
AOA2 (# 606002)	SETX	childhood	oculomotor apraxia, chorea, dystonia, axonal sensorimotor neuropathy, cognitive impairment

AT (# 208900)	ATM	1st	telangiectasias, oculomotor apraxia, choreo-athetosis, dystonia, immunodeficiency, cancer susceptibility, sensorimotor neuropathy, cognitive impairment
ATLD (# 604391)	MRE11	1st	milder phenotype of AT without telangiectasia

Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay

Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS), which is a different form of early-onset spastic ataxia, was defined first in Quebec, Canada (Bouchard et al., 1978). The carrier ratio is 1/22 with a prevalence of 1/484 in this region. Initially, it was considered that ARSACS is restricted to the Quebec region; after discovery of the *SACS* gene, it became evident that ARSACS is not limited to Quebec. While ARSACS patients usually present with lower extremity spasticity, early-onset cerebellar ataxia, and peripheral neuropathy, there are patients with an atypical phenotype, such as Charcot–Marie–Tooth-like disorders or epilepsy. These atypical phenotypes may result in underdiagnosis of ARSACS (Synofzik et al., 2018).

Ataxias with Oculomotor Apraxia

Oculomotor apraxia, which is characterized as the absence or defect of controlled, voluntary, and purposeful eye movement due to difficulties in saccade initiation and impaired cancellation of the vestibulo-ocular reflex, is a deficit showing saccadic hypometry with a typical ladder pattern (Subramony et al., 2015). Ataxias with oculomotor apraxia (AOA), as a group of ARCAs, have been shown during the past two decades as being due to DNA repair damage, RNA termination or maturation deficiencies or both. These involve ataxia-telangiectasia (AT) due to mutations in *ATM*, the rarer AT-like disorder (ATLD) related to mutations in *MRE11*, ataxia with oculomotor apraxia type 1 (AOA1) due to mutations in *APTX* and ataxia with oculomotor apraxia type 2 (AOA2), caused by mutations in *SETX* genes (Mariani et al., 2017).

Ataxia-telangiectasia, which is a primary immunodeficiency disease, represents a broad clinical spectrum, including progressive cerebellar ataxia, oculocutaneous telangiectasia, variable immunodeficiency, radiosensitivity, susceptibility to

malignancies, and increased metabolic diseases. The manifestation of the disease is typically shown by the presence of cerebellar ataxia and Vermian atrophy on magnetic resonance imaging (MRI) before five years of age. While the AT phenotype is generally characterized as early-onset, it has also been shown in late-onset cases. Telangiectasia is distinguished by small, dilated blood vessels near the surface of the skin, while high levels of alpha-fetoprotein (AFP) are evident at AT. Therefore, the serum concentration of AFP acts as a blood-based biomarker to identify the AT phenotype. AT accounts for about 3-5% of all ARCAs (Amirifar et al., 2019; Synofzik et al., 2018).

Ataxia-telangiectasia-like disease is a rare syndrome similar to the AT phenotype without telangiectasia and raised AFP levels. ATLD is characterized by later-onset neurological signs and progresses more slowly than AT. It is unknown whether the ATLD phenotype, which has been described in very few patients so far, inclines to cancer (Taylor et al., 2004).

Ataxia with oculomotor apraxia type 1 is an early-onset disease described by hypometric horizontal saccades, sensorimotor neuropathy, and widespread intellectual disability. Oculomotor apraxia (OMA) is not seen in all AOA1 patients (Mariani et al., 2017). AOA1 has a frequency of about 2-5% in all recessive ataxias. The susceptibility of AOA1 patients to OMA and cognitive impairment varies according to the mutation type. This susceptibility is higher in patients with truncating than in those with missense mutations. Although the average age of onset is 6.8 in AOA1 phenotype, there are also forms that appear later in life. Furthermore, the type of mutation may also affect the onset of the disease. Patients with truncating mutations show symptoms earlier than those with missense mutations (Synofzik et al., 2018).

Ataxia with oculomotor apraxia type 2, which has a frequency of approximately 3% in all ARCAs, is described by cerebellar ataxia, sensorimotor neuropathy, OMA, and other hyperkinetic movement disorders, such as chorea or dystonia. The disease is usually observed between 7 and 25 years, but there are also rare AOA2 cases at 40 years. AFP is elevated in AOA2 patients and can be used as a biomarker of the disease. Although the high level of AFP in AOA2 is similar to AT, there is no proof of cancer susceptibility or increased sensitivity to ionizing radiation in the AOA2 phenotype (Synofzik et al., 2018).

1.2 Differential Diagnosis of Heterogeneous Ataxia Phenotypes in the Genomic Area

Neurological disorders arising from different parts of the nervous system may have significantly overlapping phenotypes. These overlapping phenotypes complicate the accuracy of clinical diagnoses. Cerebellar ataxia, which can be a symptom of many neurological diseases, is difficult to classify and diagnose. Ataxic features may also be seen in motor neuron disorders, neuropathies, or metabolic disorders. Ataxia, which is a degenerative phenotype targeting the cerebellum, can be observed in hereditary spastic paraplegias (HSP), Charcot-Marie-Tooth (CMT) disease, or various types of encephalopathies. When the ataxia phenotype is accompanied by pyramidal signs or spasticity, it becomes difficult for clinicians to diagnose the disease. Spastic ataxia is a phenotype in which both ataxia and spasticity are seen together. ARSACS, hereditary spastic paraplegia type 7 (SPG7), spastic ataxia types 1-5, FRDA, and SCA3 have a common genetic background with this phenotype (de Bot et al., 2012). Among these, SPG7 is most frequent in unexplained ataxias with mid-adult age onset (Pfeffer et al., 2015). As a result, distinguishing various clinical phenotypes with genetic tests and follow-up of patients become essential for precise diagnosis.

1.2.1 Next Generation Sequencing

The Human Genome Project (HGP) is one of the most ground-breaking discoveries in the biological field. Completed in 2003, HGP paved the way by accelerating innovative developments in sequencing technologies. New sequencing approaches have been developed by many companies to reduce time and cost per mega base while increasing accuracy. The next generation of sequencing (NGS), also known as high throughput, has drastically improved understanding the human genome (Boycott et al., 2013).

NGS, a method that sequences millions of DNA fragments simultaneously, is used in genetic and genomic research worldwide due to its ability to produce the genomic region with a single analysis as compared to traditional technologies. The NGS technology involves performing diverse applications, such as whole genome sequencing (WGS), whole exome sequencing (WES), whole transcriptome analysis (RNA-seq), genome-wide profiling of epigenetic marks (methyl-seq), and chromatin structure (ChIP-seq) (Slatko et al., 2018).

1.2.2 Whole Exome Sequencing

Exons, which constitute about 1% of the whole genome, are protein-coding regions. The exome, also known as the sum of exons in the genome, consists of 180,000 exons. Repetitive regions, introns, and intergenic regions that do not directly affect the protein sequence form the rest of the genome. Mutations in the exome are associated with 85% of all known disorders. Therefore, the exome has been the focus of genetic research and diagnosis as it is more likely to cause disease and is less complex than whole genome analysis (Botstein et al., 2003; Marian, 2014).

WES is an efficient and powerful application of the NGS technology; more than 150 genes have been identified by WES, comprising 95% of the exome. WES is an effective tool in explaining single-gene disorders with its high throughput capacity and relatively low cost. The genetic cause of Miller syndrome was found by WES and this was the first successful application of WES technology to describe a gene for a Mendelian disorder. (Ng, 2010). With this study, the causative gene underlying a rare Mendelian disease was defined in a small number of affected individuals, demonstrating an effective technology strategy.

WES comprises two major parts: wet-lab and *in-silico* workflows. The wet-lab is the step where exonic regions are enriched for library preparation. *In-silico* is the part where data analysis is carried out. The alignment of raw reads to the reference genome, variant calling, functional annotation, and prioritization of variations are performed in data analysis. The variants are filtered according to the purpose of the study (Yohe et al., 2017).

Data Analysis

WES requires the appropriate design to detect the causative variant underlying the disease, starting with sample selection. All pedigrees must be evaluated according to the criteria established for the study. Family-based studies help to reduce candidate variants when affected and unaffected individuals are present. In addition, unaffected individuals are considered as the control group to validate variants detected in the affected individual/s. Individuals who do not manifest the disease may be regarded as unaffected, but this may be misleading in terms of adult-onset dominant disorders. Therefore, detailed family information is critical.

The selection of the platform on which WES is completed is significant since it identifies the sensitivity and accuracy limits of any variations received from the data (Yohe et al., 2017). Kits that target different sites and have different efficacy levels are commercially available. Depth (or coverage) indicates how many times a particular nucleotide is read in DNA sequencing. This depth should be approximately 100X or higher in clinical sequencing. It is necessary to increase the coverage to decrease false-positive results, even if it increases the cost (Marian, 2014). Read length and methods are other parameters to be considered among different platforms. For instance, short-read platforms should be used for large-scale genome variation studies and clinical applications, while long-read platforms should be preferred for multi-repetitive regions (van Dijk et al., 2018).

Evaluation Criteria for Bioinformatics in This Study

There are many bioinformatic analysis tools and software used to analyze NGS data; each has its advantages and limitations depending on the aim. It is significant to choose which program and reference genome version to prefer. In our lab, computational pipeline steps (e.g., alignment, variant calling, annotation, prioritization) are performed by SEQ Platform of Genomize, Turkey (https://genomize.com/seq/). SEQ is a platform on genomic data analysis and management that provides services such as storage, sharing, and usage.

Public databases and pathogenicity scoring tools are used for the interpretation of variants obtained by SEQ. The public databases are the 1000 Genomes Project and the Exome Variant Server, containing 2504 individuals and 6503 exomes, respectively. The Exome Aggregation Consortium (ExAC) database, which includes exome sequencing data from 60,706 individuals, helps to estimate the prevalence of homozygous and heterozygous variants in genes. ExAC, a public resource for clinical interpretations of variants, also demonstrates increased resolution of low-frequency variations (Lek et al., 2016). The Genome Aggregation Database (gnomAD) browser, which currently also includes ExAC data, is another database (https://gnomad.broadinstitute.org/). DANN, GERP++, SIFT, MutationTaster, REVEL, and MetalR are used as pathogenicity scoring tools.

Limitations

WES, which is used worldwide as the gold standard in clinical diagnosis, has excellent advantages but has some technical limitations both in wet-lab and in *in-silico* workflows. Genomes often include numerous repeated sequences that are longer than WES reads, which may cause incorrect assemblies and gaps. This is one of the main disadvantages of WES. Besides, larger structural variations (SV) such as copy number variations (CNV), large deletions and translocations are more challenging to analyze. Although various approaches have been developed to call SV, they are not yet sufficient (van Dijk et al., 2018).

Applying WES, which covers all coding regions in the genome, is promising in medical genetics. In contrast, there are still gaps and uncertainties in human genome sequences, as the capture and annotation of the exome have not yet been completed (Ku et al., 2012). Due to this incompleteness of the genome, missing regions may occur in exome sequencing kits.

1.2.3 Molecular Analysis of Complex Ataxias by WES in the Turkish Cohort

The advances in NGS-based methods have contributed to the development in NDs genetics, including also hereditary ataxias. These developments have significantly accelerated the increase in the number of genes that give rise to ataxias. NGS techniques, especially WES, have helped to explain the underlying causes of the complex molecular basis of ataxias in the research and diagnostic laboratory (Fogel et al., 2016). Thus, the diagnostic rate may be broadly improved by using WES for disease analysis.

Turkey is a very dynamic country with a young population, a high birth rate, and widespread consanguineous marriages on one hand and extreme ethnic heterogeneity in some parts of the country, on the other. Therefore, Turkey harbors potential mutations in many genes that may play a role in the pathogenesis of ataxias. The Turkish population was underrepresented in the literature until recently. This has changed with a recent publication on the distribution of ataxias in Turkey (Vural, Şimşir, Tekgül et al., 2021).

Chapter 2:
PURPOSE

Diseases of the cerebellum, which significantly affect the life quality of patients, cause a broad spectrum of neurodegenerative disorders. Among these, ataxias that constitute the majority of cerebellar disorders, overlap with different phenotypes, and make a precise clinical diagnosis challenging. Most types of ataxias do not have any available treatment yet, but some subtypes with possible treatment options improve the life quality of patients. Consequently, having a precise genetic diagnosis is critical for patients.

Turkey, a country with significant ethnic heterogeneity, has high birth and consanguinity rates. In addition to early-onset recessive ataxias, which are most common in Turkey, late-onset dominant forms are not rare, as well.

The purpose of this thesis is to understand the complex genetic structure behind ataxias and to identify novel variants associated with NDs in a Turkish cohort comprised of 95 index patients. WES, bioinformatic analysis and related tools were applied. Discovering the complex genetic mechanisms behind ataxias will help find ways to successful treatments.

Chapter 3:
MATERIALS

3.1 Subjects

Ninety-five families (index patients) harboring 123 patients, referred to our laboratory with a primary diagnosis of clinical ataxia, were studied in the framework of this thesis. Patients were divided into three main classes according to their inheritance patterns, as autosomal dominant, autosomal recessive, and sporadic. AD inheritance was seen in seven out of 95 families, AR inheritance was present in 71 and 17 had sporadic disease. Consanguinity was present in 69 of 95 families (Figures 3.1-3.3). The mean age at onset of all patients with reliable information (106 patients) was 25.6 with a standard deviation of 17.8, ranging from congenital to 66 years of age (Table 3.1).

Clinical evaluation of index cases was performed by specialist clinicians in different hospitals across the nation. Informed consent was obtained from all participants. Peripheral blood samples were collected into EDTA-containing tubes. Prior to DNA analysis, comprehensive clinical phenotypes of patients were obtained from the clinicians, and the family's disease history was deeply questioned.

Dominant families were first screened for TNR disorders. Furthermore, FRDA was excluded in recessive families.

Table 3.1 Families studied in this thesis.

Family ID	ID	Gender	Cons.	AO	Phenotype	Inheritance
1	P1	M	+	37	ATX	AR
2	P2	M	+	congenital	SPAX	AR
	P3	F	+	congenital	SPAX	
3	P4	F	+	3	AT, CH	AR
4	P5	F	+	33	ATX	AR
	P6	F	+	30	ATX	
5	P7	M	+	3	ATX	AR
	P8	M	+	34	ATX	
	P9	M	+	51	ATX	
6	P10	M	+	43	ATX	AR

7	P11	M	+	53	ATX	AR
8	P12	F	+	17	ATX, DYS	AR
9	P13	M	+	14	HSP	AR
10	P14	F	+	3	AT	AR
	P15	F	+	5 (ex:41)	AT	
11	P16	M	+	25	ATX	AR
12	P17	M	+	33	HSP	AR
13	P18	F	+	3	ATX	AR
	P19	F	+	n.a	ATX	
14	P20	M	+	40	SPAX	AR
	P21	F	+	n.a	SPAX	
15	P22	M	+	47	HSP	AR
16	P23	F	+	11	ATX	AR
17	P24	F	+	17	Complex ATX	AR
18	P25	F	+	61	ATX	AR
19	P26	M	+	congenital	ATX	AR
	P27	F	+	n.a	Complex ATX	
20	P28	M	+	12	ATX, PNP	AR
21	P29	M	-	52	ATX	AR
22	P30	F	+	30	HSP	AR
23	P31	M	+	23	ATX	AR
	P32	F	+	n.a	ATX	
24	P33	M	+	childhood	ATX	AR
25	P34	F	+	21	ATX	AR
26	P35	F	+	7 months	ATX	AR
27	P36	F	+	3	ATX	AR
28	P37	F	+	30	SPAX	AR
29	P38	M	+	15	SPAX	AR
30	P39	M	+	6	AT	AR
31	P40	M	+	35	SPAX	AR
32	P41	F	+	16	ATX	AR
33	P42	M	+	childhood	SPAX	AR
	P43	F	+	congenital	SPAX	

	P44	M	+	18	SPAX	
34	P45	F	+	9	ATX	AR
35	P46	F	+	19	ATLD	AR
	P47	F	+	childhood	ATLD	
36	P48	M	+	14 months	SPAX	AR
37	P49	M	+	congenital	SPAX	AR
	P50	M	+	congenital	SPAX	
38	P51	M	+	7	SPAX	AR
39	P52	F	+	congenital	ATX	AR
40	P53	F	+	17	ATX	AR
	P54	F	+	17	ATX	
41	P55	M	+	40	HSP	AR
42	P56	F	+	11	ATX	AR
43	P57	M	+	3	ATX	AR
44	P58	M	+	51	ATX	AR
45	P59	M	+	18	ATX	AR
46	P60	F	+	24	ATX	AR
	P61	F	+	n.a	ATX	
47	P62	M	+	33	ATX	AR
48	P63	M	+	49	ATX	AR
49	P64	M	+	16 months	ATX	AR
50	P65	M	+	53	ATX	AR
51	P66	F	+	45	ATX	AR
52	P67	F	+	35	ATX	AR
53	P68	M	+	40	ATX	AR
54	P69	M	+	48	ATX	AR
55	P70	M	+	5	ATX	AR
56	P71	M	+	21	ATX	AR
57	P72	M	+	47	ATX	AR
58	P73	F	+	46	ATX	AR
59	P74	M	-	49	Complex ATX	AR
	P75	M	-	25	Complex ATX	
	P76	M	-	35	Complex ATX	

60	P77	M	+	44	ATX	AR
61	P78	M	+	63	ATX	AR
62	P79	M	+	34	ATX	AR
63	P80	M	+	28	ATX	AR
64	P81	F	+	16	ATX	AR
	P82	M	-	7	ATX	
65	P83	F	+	41	ATX	AR
66	P84	M	+	14	SPAX	AR
67	P85	M	+	28	SPAX	AR
	P86	M	+	n.a	SPAX	
68	P87	F	+	19	SPAX	AR
69	P88	M	+	24	ATX	AR
	P89	F	+	15	ATX	
70	P90	F	+	childhood	ATX	AR
	P91	F	+	n.a	ATX	
	P92	F	+	15	ATX	
71	P93	F	+	20	SPAX	AR
	P94	F	+	n.a	SPAX	
	P95	F	+	n.a	SPAX	
	P96	M	+	n.a	SPAX	
72	P97	M	-	56	ATX	AD
73	P98	F	-	36	ATX	AD
74	P99	M	-	8	HSP	AD
	P100	M	-	n.a	HSP	
75	P101	F	-	23	ATX	AD
76	P102	M	-	39	SPAX	AD
77	P103	F	+	6	ATX	AD
	P104	M	-	n.a	ATX	
	P105	F	-	n.a	ATX	
78	P106	M	-	40	ATX	AD
79	P107	M	-	60	ATX	Sporadic
80	P108	M	-	22	HSP	Sporadic
81	P109	F	-	30	SPAX	Sporadic

82	P110	F	-	11	HSP	Sporadic
83	P111	F	-	42	HSP	Sporadic
84	P112	M	-	22	ATX	Sporadic
85	P113	M	-	27	Fabry disease, HSP	Sporadic
86	P114	F	-	59	HSP	Sporadic
87	P115	F	-	19	HSP	Sporadic
88	P116	F	-	55	ATX	Sporadic
89	P117	F	-	38	ATX	Sporadic
90	P118	F	-	38	ATX	Sporadic
91	P119	M	-	22	ATX	Sporadic
92	P120	M	-	66	ATX	Sporadic
93	P121	F	-	7	EA	Sporadic
94	P122	M	-	25	SPAX	Sporadic
95	P123	F	-	22	ATX	Sporadic

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, CH: cerebellar hypoplasia, ATLD: AT-like Disorder, DYS: Dystonia, n.a: not available, AR: Autosomal recessive, AD: Autosomal dominant

3.1.1 Family Trees

Families with AD Inheritance

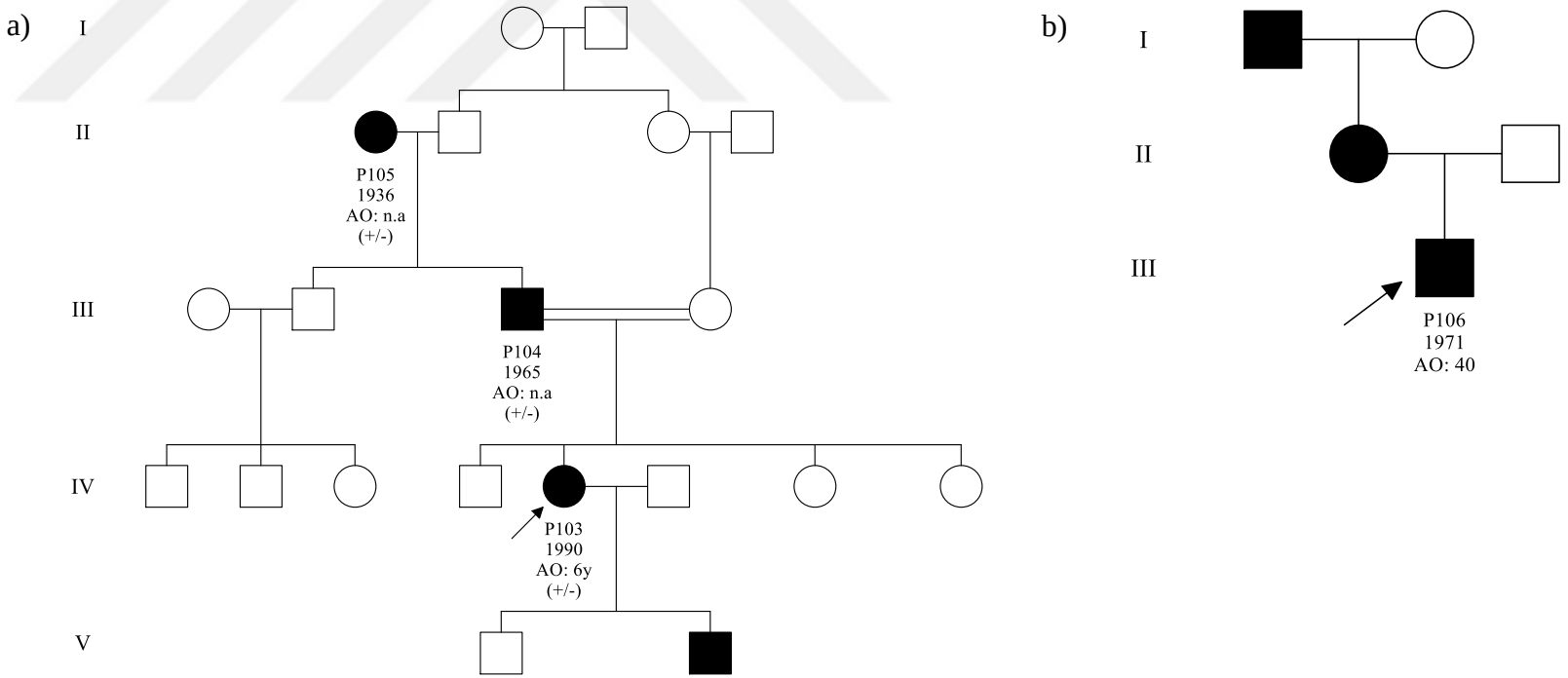
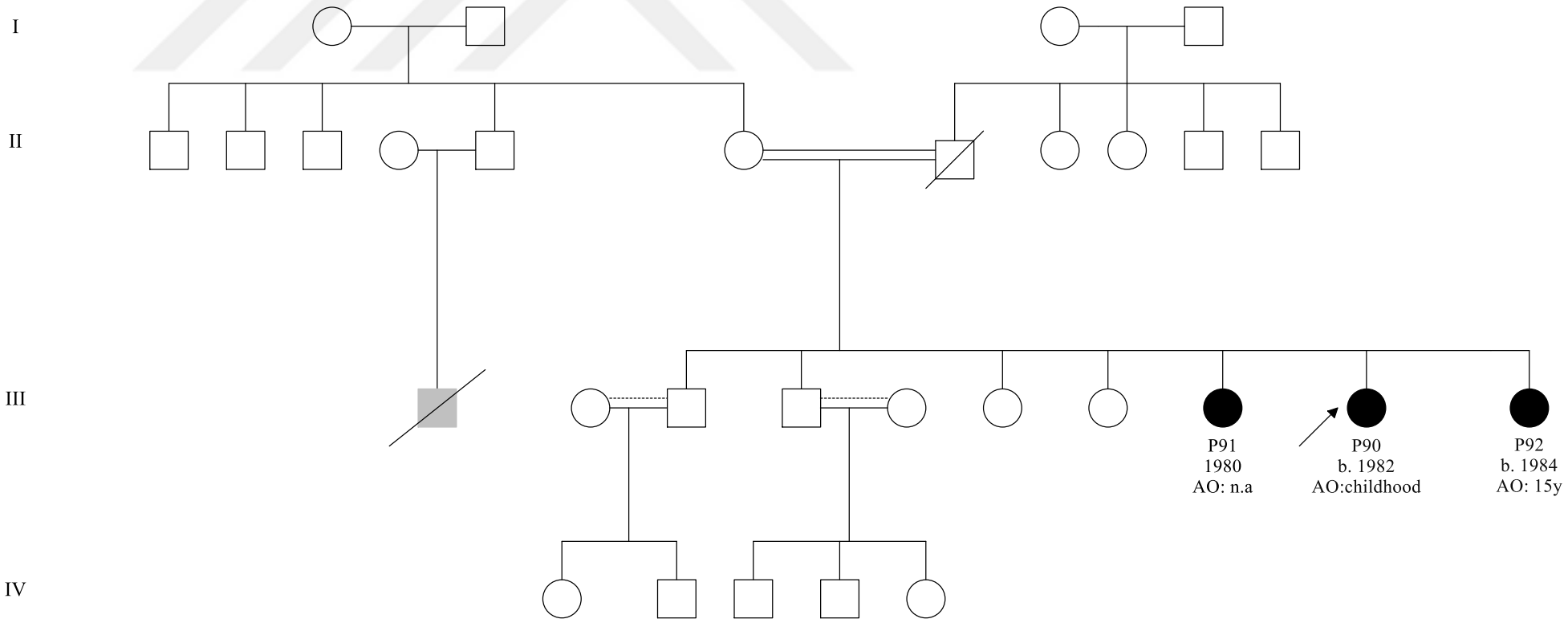


Figure 3.1 a) Family 77 (Patients P103-P105), and b) Family 78 (Patient P106) showing an autosomal dominant inheritance pattern with several affected individuals.

Families with Autosomal Recessive (AR) Inheritance

a)



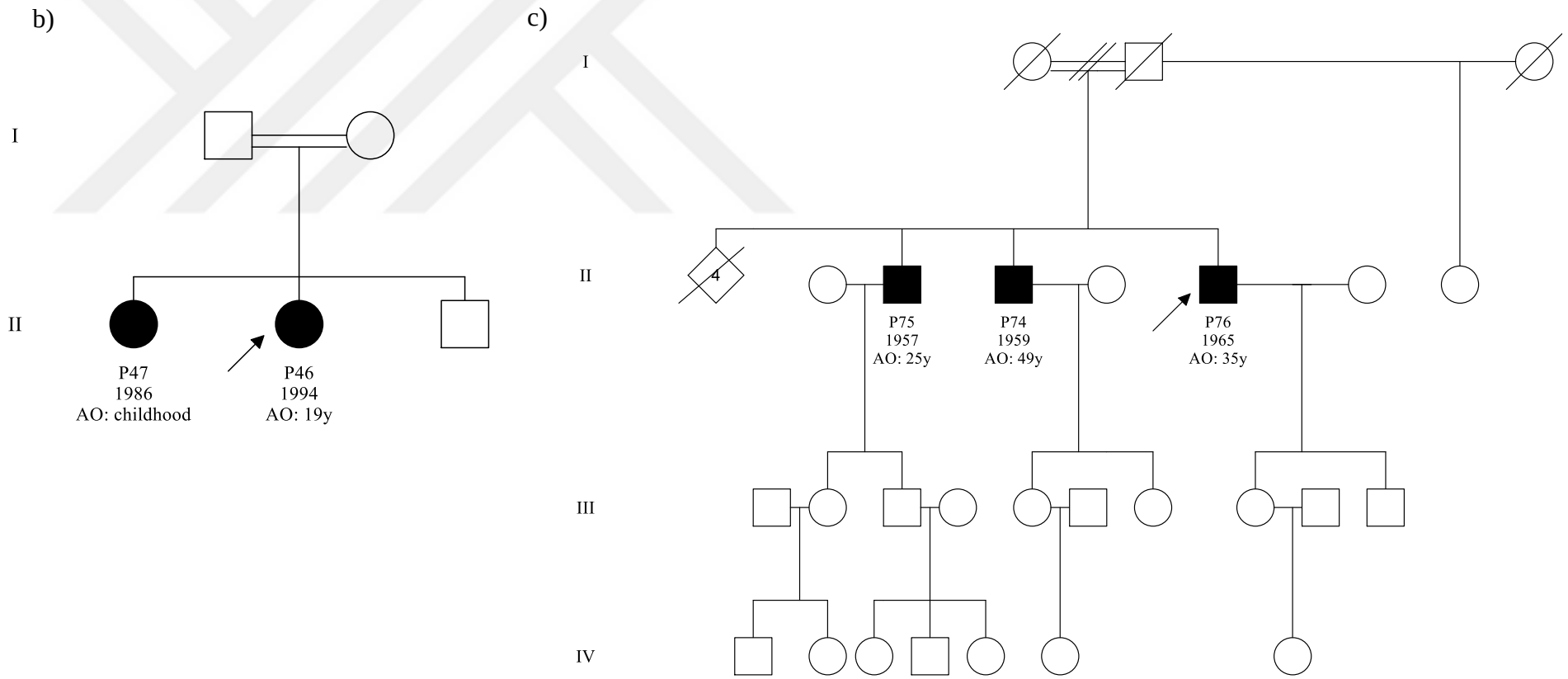


Figure 3.2 Pedigrees of families with an AR inheritance. a) Family 70 (Patients P90-P92), b) Family 35 (Patients P46-P47) and c) Family 59 (Patients P74-P76) showing AR inheritance pattern with consanguinity.

Families with Sporadic Disease

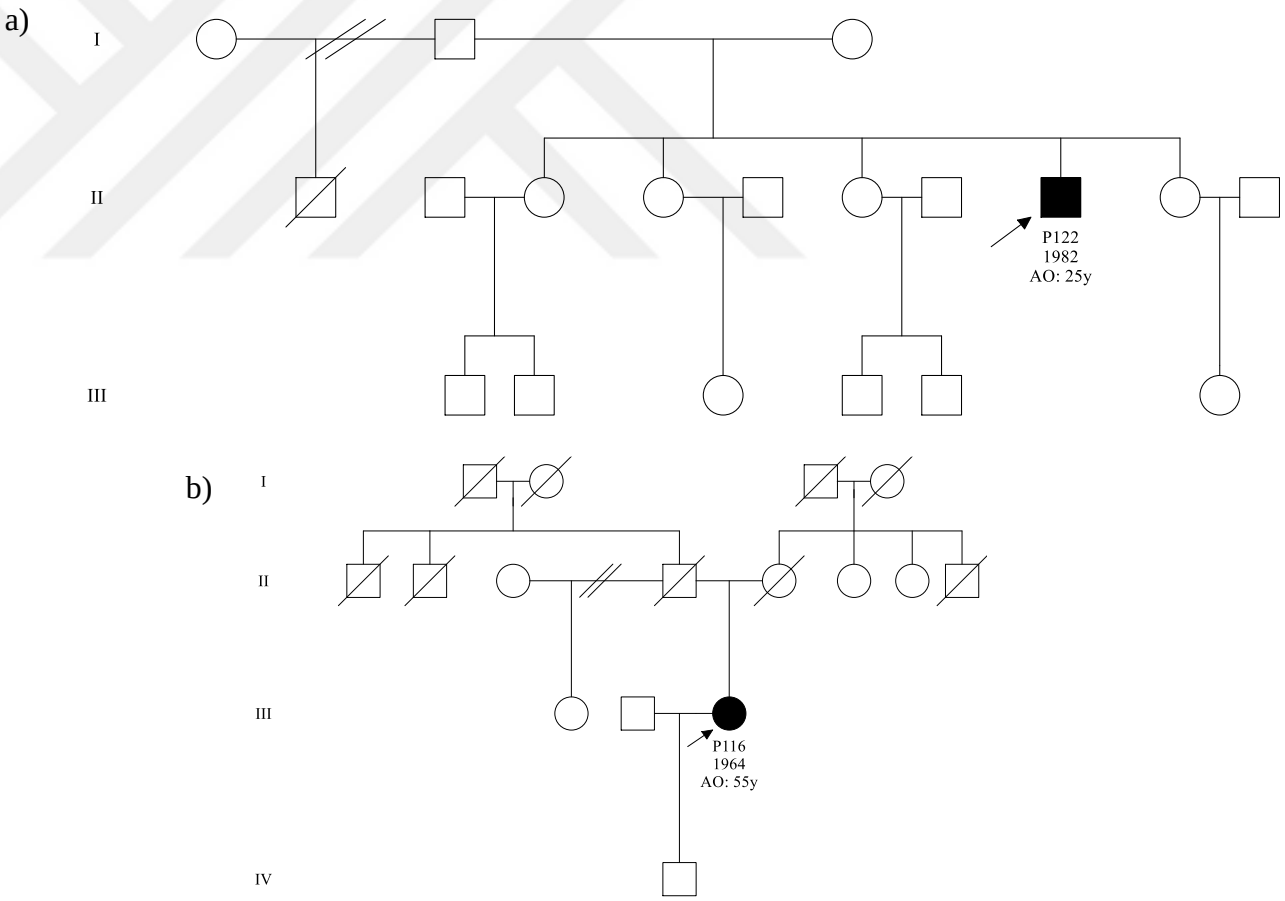


Figure 3.3 a) Family 94 (Patient P122) and b) Family 88 (Patient P116) showing sporadic disease with only one affected individual and no consanguinity.

3.2 DNA Isolation

MagNA Pure Compact Nucleic Acid Isolation Kit I and MagNA Pure Compact Instrument from Roche were used to extract genomic DNA from peripheral blood samples. The quality and concentration of DNA were measured using the Thermo Scientific NanoDrop 2000c Spectrophotometer.

3.3 Whole Exome Sequencing Platforms

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

3.4 Validation Experiments

PCR was performed using MyTaq^{TM} DNA Polymerase (Bioline, 5u/μL) and 5X MyTaq Reaction Buffer to amplify 10 ng of genomic DNA. The sequences of the primers used to validate the experiments are listed in Table A.1. The PCR reagents and conditions for amplification used are shown in Tables 3.2 and 3.3, respectively.

Table 3.2 PCR reagents for amplification.

Reagent	**Volume (μL)**	**Stock conc.**	**Final conc.**
dH2O	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	5 u/μL	1u
DNA	1	10 ng/μL	10 ng/μL
Total	25		

Table 3.3 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

Solutions, chemicals, and equipment used in agarose gel electrophoresis are listed in Table 3.4. The QIAQuick Gel Extraction Kit (Qiagen, Germany) was used for gel extraction of PCR products.

Table 3.4 Chemicals used in gel electrophoresis.

Solution/Chemical/Equipment		Catalog No.	Company
10X 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 B.1 and 3.5, respectively.

Table 3.5 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 databases, tools, and software used in WES analysis are listed in Table C.1.

Chapter 4:
METHODS

4.1 DNA Isolation

Peripheral blood stored in EDTA tubes was collected from all participants. Genomic DNA was extracted using the MagNA Pure Compact Instrument. The 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 assessed and documented to identify chemical and protein contaminations.

4.2 Whole Exome Sequencing

Whole exome sequencing was performed on DNA samples of all index patients. In some conditions, affected family members or parents were also subjected to WES.

Whole exome sequencing involves three basic steps: i) library preparation and amplification, ii) sequencing, and iii) data analysis (Figure 4.1). The workflow of WES is present in Figure 4.2.

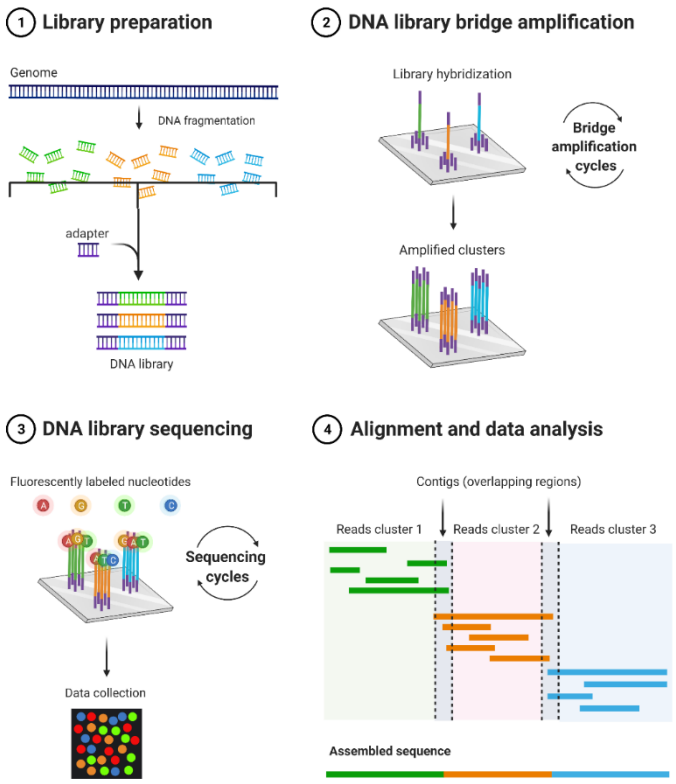


Figure 4.1 Steps of whole exome sequencing. Republished from "Next Generation Sequencing (Illumina)" by BioRender.com (2021). Retrieved from https://app.biorender.com/biorender-templates.

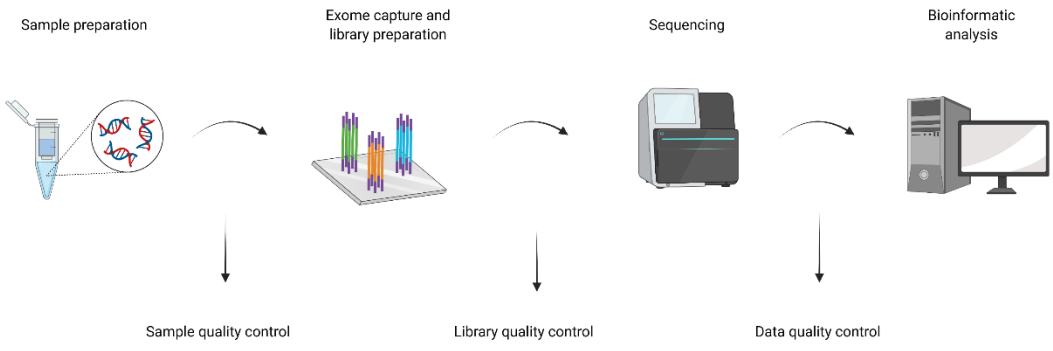


Figure 4.2 The workflow for whole exome sequencing, created with BioRender.com.

4.3 Whole Exome Sequencing Data Analysis

The cleaning of raw reads, alignment to the reference genome, variant calling, and annotation are performed by SEQ Platform, Genomize (Istanbul, Turkey), which is used for bioinformatics data analysis.

4.3.1 Variant Prioritization

Variants were classified as verified pathogenic variants, likely pathogenic novel variants, and variants of uncertain significance (VUS). In the literature, verified pathogenic variants are the ones previously published to be pathogenic. Likely pathogenic novel variants are identified in genes associated with ataxias, which affect protein structure or function and segregate with the family tree. Likely pathogenic variants found in genes that are not directly associated with ataxias but related to other neurological conditions are defined as VUS. MutationTaster, SIFT, MetalR, CADD, DANN, and REVEL scores are used to evaluate the pathogenicity of the variant and its effect on protein structure and/or function. The GERP++ score is used to assess the evolutionary conservation rate of the variable region.

4.4 PCR Experiments for Validation of Variants

4.4.1 PCR Conditions

PCR-based Sanger sequencing is used to confirm the presence and segregation of candidate variants obtained from WES data analysis. Variant-specific primers were either retrieved from the literature or designed by the NCBI Primer Blast tool (Ye et al., 2012). NetPrimer and OligoCalc tools were employed to control the possibilities of hairpins, self-dimers, and cross-dimers in primer pairs. To detect the specificity of the region in the genome, UCSC *in-silico* PCR was used. Lyophilized primers were synthesized and purchased from Sentromer (Istanbul, Turkey). 100µM primers were prepared by adding the recommended amount of dH2O and diluted to working concentration (10µM) for use in PCR experiments. PCR products were run on 2% agarose gel with GelRed fluorescent DNA dye, used as the intercalating agent for visualization under UV light.

4.4.2 Agarose Gel Electrophoresis

2% agarose gel was prepared with 0.5X TBE buffer and GelRed fluorescent DNA dye and placed into an electrophoresis chamber containing 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 from Sanger sequencing were analyzed using the programs FinchTV and CLC Main Workbench.

4.4.3 Purification of PCR Products for Sanger Sequencing

PCR products were purified to avoid background noise in Sanger sequencing results. The purification was carried out either from the agarose gel or directly from the product itself to eliminate possible primer dimers and non-specific bands. In agarose gel extraction and PCR product purification, QIAquick Gel Extraction Kit (Qiagen, Germany) and QIAquick PCR Purification Kit (Qiagen, Germany) were used, respectively.

Chapter 5:
RESULTS

In this study, WES analysis followed by Sanger sequencing was performed on 95 index patients with a complex ataxia phenotype; also, their family members were investigated.

5.1 Whole Exome Sequencing Analysis

Seventy-one of these families presented with a recessive and seven with a dominant inheritance pattern; 17 family trees were apparently sporadic. Forty-three different variants in 28 genes were identified in 40 distinct families, but in the remaining cases, variants in disease-related genes could not be detected due to technical or strategical limitations. Sixteen variants were novel. The variants and their associated diseases in OMIM are listed in Table 5.1. *SACS* is the most common gene in this study, followed by *SPG7*; and *ATM, APTX, SETX,* and *ATP13A2*. Genetic heterogeneity, pleiotropy, and overlapping genes in phenotypes categorized as different diseases are shown in Figure 5.2. Some of family trees are presented in Figure 5.3-Figure 5.9.

The diagnostic yield was 49.3% (35/71) in AR, 29% (2/7) in AD and 18% (3/17) in sporadic families, the overall diagnostic yield corresponding to 42% (Figure 5.1).

The presence of the variants and family segregation were confirmed by Sanger sequencing. Segregation results, prediction tool scores, and gnomAD frequencies of the variants are compiled in Table 5.2.

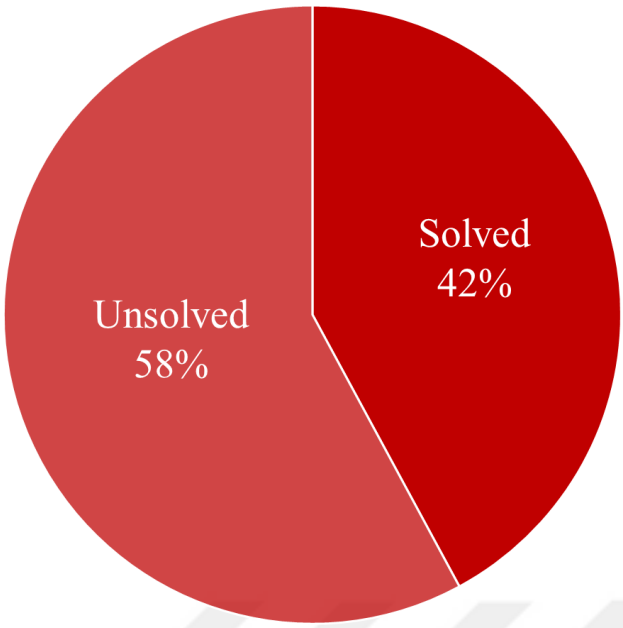


Figure 5.1 42% of patients was solved and 58% could not be solved in the cohort under study.

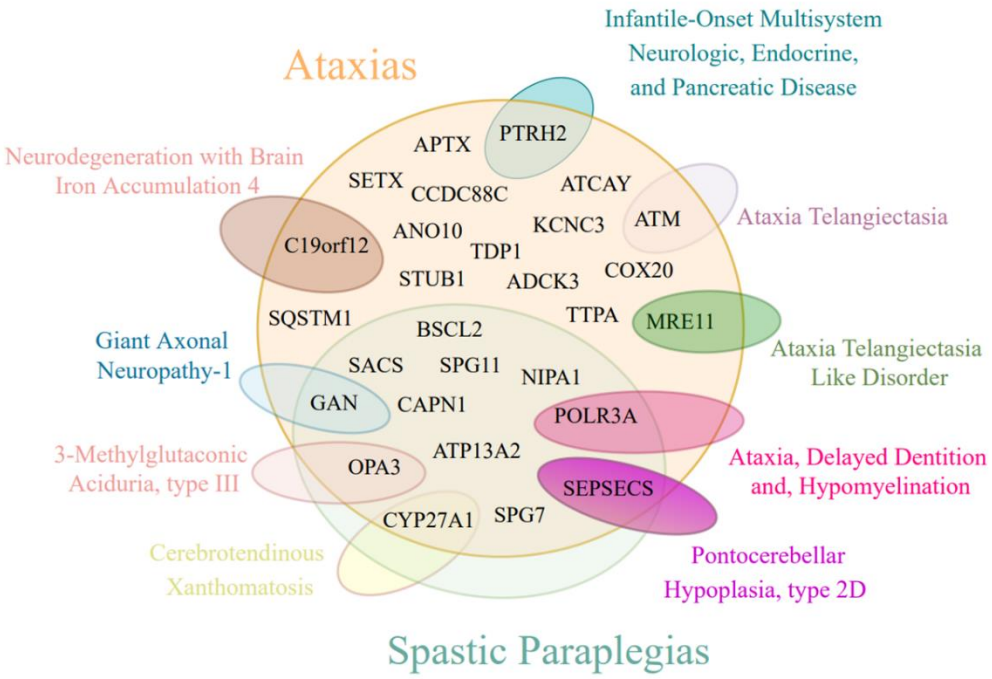


Figure 5.2 Genes identified in this study giving rise to ataxias.

Table 5.1 List of all confirmed causative variants associated with diseases in OMIM.

Family ID	Gene	Variant		Transcript ID	Zygosity	OMIM
		Coding Sequence	Protein Sequence			
2	SACS	c.7276C>T	p.Arg2426Ter	ENST00000382298.3	hom	ARSACS
5	POLR3A	c.1771-6C>G	-	ENST00000372371.3	hom	ADDH
8	ADCK3	c.248dupT	p.His85AlafsTer42	ENST00000366779.1	hom	SCAR9
9	SPG11	c.1203delA	p.Asp402IlefsTer14	ENST00000261866.7	hom	SPG11
10	ATM	c.7865C>T	p.Ala2622Val	ENST00000278616.4	hom	AT
11	C19orf12	c.32C>T	p.Thr11Met	ENST00000392278.2	hom	NBIA4
13	OPA3	c.322_339delCAGCGCCACAAGGAGGAG	p.Gln108_Glu113del	ENST00000263275.4	hom	MGCA3
14	ATP13A2	c.2816T>C	p.Leu939Pro[n]	ENST00000326735.8	hom	SPG78
20	GAN	c.1450G>T	p.Gly484Cys[n]	ENST00000568107.2	hom	GAN1
23	SETX	c.6461C>T	p.Thr2154Met	ENST00000372169.2	hom	AOA2
24	APTX	c.982_998delAAGCTGCCCCTTCGTTG	p.Lys328SerfsTer2	ENST00000379819.1	hom	AOA1
26	ATCAY	c.570C>G	p.Tyr190Ter[n]	ENST00000398448.3	hom	ATCAY
27	PTRH2	c.272_273delCT	p.Ala91GlyfsTer13	ENST00000409433.2	hom	IMNEPD1
28	SACS	c.8867T>C	p.Leu2956Ser[n]	ENST00000382292.3	hom	ARSACS
29	CAPN1	c.397C>T	p.Arg133Ter	ENST00000524773.1	hom	SPG76

30	ATM	c.26delT	p.Ile10SerfsTer6	ENST00000278616.4	hom	AT
31	SPG7	c.1861C>T	p.Gln621Ter[n]	ENST00000268704.2	het	SPG7
		c.1715C>T	p.Ala572Val		het	
32	APTX	c.731T>G	p.Val244Gly	ENST00000379819.1	hom	AOA1
33	SACS	c.12923_12927delAAGAA	p.Lys4308SerfsTer21	ENST00000382292.3	hom	ARSACS
34	SQSTM1	c.819_820delAG	p.Ser275PhefsTer17	ENST00000389805.4	hom	NADGP
35	MRE11	c.1442C>A	p.Thr481Lys	ENST00000323929.3	hom	ATLD
36	SACS	c.3767A>G	p.Tyr1256Cys[n]	ENST00000382292.3	hom	ARSACS
		c.7732_7734delGAT	p.Asp2578del[n]		hom	
37	SACS	c.4192T>C	p.Cys1398Arg	ENST00000382292.3	hom	ARSACS
38	SACS	c.11374C>T	p.Arg3792Ter	ENST00000382292.3	hom	ARSACS
	TDP1	c.92C>G	p.Ser31Cys[n]	ENST00000335725.4	hom	SCAN1
39	SEPSECS	c.1321G>A	p.Gly441Arg	ENST00000382103.2	hom	PCH2D
40	SETX	c.6628C>T	p.Gln2210Ter[n]	ENST00000372169.2	hom	AOA2
41	SPG7	c.233T>A	p.Leu78Ter	ENST00000268704.2	hom	SPG7
43	TTPA	c.553-2A>C[n]	-	ENST00000260116.4	hom	AVED

52	CYP27A1	c.1476+2T>C	-	ENST00000258415.4	hom	CTX
66	SPG7	c.2096T>C	p.Met699Thr[n]	ENST00000268704.2	hom	SPG7
67	ATP13A2	c.289-3C>G[n]	-	ENST00000326735.8	hom	SPG78
68	SPG7	c.1715C>T	p.Ala572Val	ENST00000268704.2	hom	SPG7
69	ANO10	c.1025G>A	p.Trp342Ter	ENST00000292246.3	hom	SCAR10
70	COX20	c.154A>C	p.Thr52Pro	ENST00000411948.2	hom	MC4DN11
71	SACS	c.11766A>T	p.Leu3922Phe[n]	ENST00000382298.3	het	ARSACS
		c.7940T>A	p.Ile2647Asn[n]		het	
76	CCDC88C	c.4120T>G	p.Leu1374Val[n]	ENST00000389857.6	het	SCA40
77	KCNC3	c.1268G>A	p.Arg423His	ENST00000477616.1	het	SCA13
83	BSCL2	c.974G>A	p.Gly325Asp[n]	ENST00000433053.1	het	SPG17
87	NIPA1	c.731A>G	p.Gln244Arg	ENST00000337435.4	het	SPG6
89	STUB1	c.146A>G	p.Tyr49Cys[n]	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.
13:23910739	SACS	p.Arg2426Ter	Path: PVS1, PP5, PM2, PP3	35	0.9951	5.59	Disease causing	n.a	n.a	n.a	-	+
10:79769439	POLR3A	c.1771-6C>G	LP: PP5, PM2, BP4	-	0.6108	5.71	n.a	n.a	n.a	n.a	0.00006024	+
1:227152770	ADCK3	p.His85AlafsTer42	Path: PVS1, PM2, PP5	-	n.a	5.46	n.a	n.a	n.a	n.a	-	+
15:44943941	SPG11	p.Asp402IlefsTer14	Path: PVS1, PM2, PP5	-	n.a	6.07	n.a	n.a	n.a	n.a	0.00001193	+
11:108203565	ATM	p.Ala2622Val	Path: PP5, PM1, PM2	24.2	0.9842	5.09	Disease causing, Polymorp.	B	T	T	-	+
19:30199322	C19orf12	p.Thr11Met	Path: PP5, PM2, PP2, BP4	-	0.8919	6.07	Polymorp.	B	T	T	0.00001069	+
19:46056972	OPA3	p.Gln108_Glu113del	VUS-LP: PM2, PP3, PP5	-	n.a	4.76	n.a	n.a	n.a	n.a	-	+

1:17314676	ATP13A2	p.Leu939Pro	VUS-LP: PM2, PP3	30	0.9991	5.51	Disease causing	Path	D	D	-	+
16:81399031	GAN	p.Gly484Cys	VUS: PM2, BP4	23.2	0.9919	5.69	Disease causing	B	T	D	-	+
9:135158736	SETX	p.Thr2154Met	VUS-LP: PM2, PP3	26.9	0.999	4.94	Disease causing	Path	D	D	0.000003976	+
9:32973568	APTX	p.Lys328SerfsTer2	Path: PVS1, PM2, PP3, PP5	-	n.a	5.42	n.a	n.a	n.a	n.a	-	+
19:3908273	ATCAY	p.Tyr190Ter	Path: PVS1, PM2, PP3	39	0.9952	5.13	Disease causing automatic	n.a	n.a	n.a	-	+
17:57775069	PTRH2	p.Ala91GlyfsTer13	Path: PVS1, PP5, PM2, PP3	-	n.a	5.82	n.a	n.a	n.a	n.a	-	+
13:23909148	SACS	p.Leu2956Ser	VUS-LP: PM2, PP3	26.8	0.9986	5.64	Disease causing	Path	D	D	-	n.a
11:64951004	CAPN1	p.Arg133Ter	Path: PVS1, PM2, PP3	37	0.9975	5.26	Disease causing automatic	n.a	n.a	n.a	-	n.a

11:108098376	ATM	p.Ile10SerfsTer6	Path: PVS1, PP5, PM2, PP3	-	n.a	5.28	n.a	n.a	n.a	n.a	-	+
16:89619468	SPG7	p.Gln621Ter	Path: PVS1, PM2, PP3	42	0.9971	5.93	Disease causing automatic	n.a	n.a	n.a	-	+
16:89616953		p.Ala572Val	LP: PP5, PM2, PP2, PP3	24.3	0.9989	5.84	Disease causing	Path	D	D	0.00003537	
9:32984710	APTX	p.Val244Gly	LP: PM2, PP2, PP3	26.1	0.9970	5.59	Disease causing	Path	D	D	0.00001591	+
13:23905087	SACS	p.Lys4308SerfsTer21	Path: PVS1, PP5, PM2, PP3	-	n.a	5.82	n.a	n.a	n.a	n.a	0.00003185	+
5:179260095	SQSTM1	p.Ser275PhefsTer17	Path: PVS1, PM2, PP3	22.6	n.a	4.89	n.a	n.a	n.a	n.a	0.000007954	+
11:94192632	MRE11	p.Thr481Lys	VUS-LP: PM2, PP3, PP5	28.2	0.9954	5.8	Disease causing automatic	Path	D	D	0.000003979	+

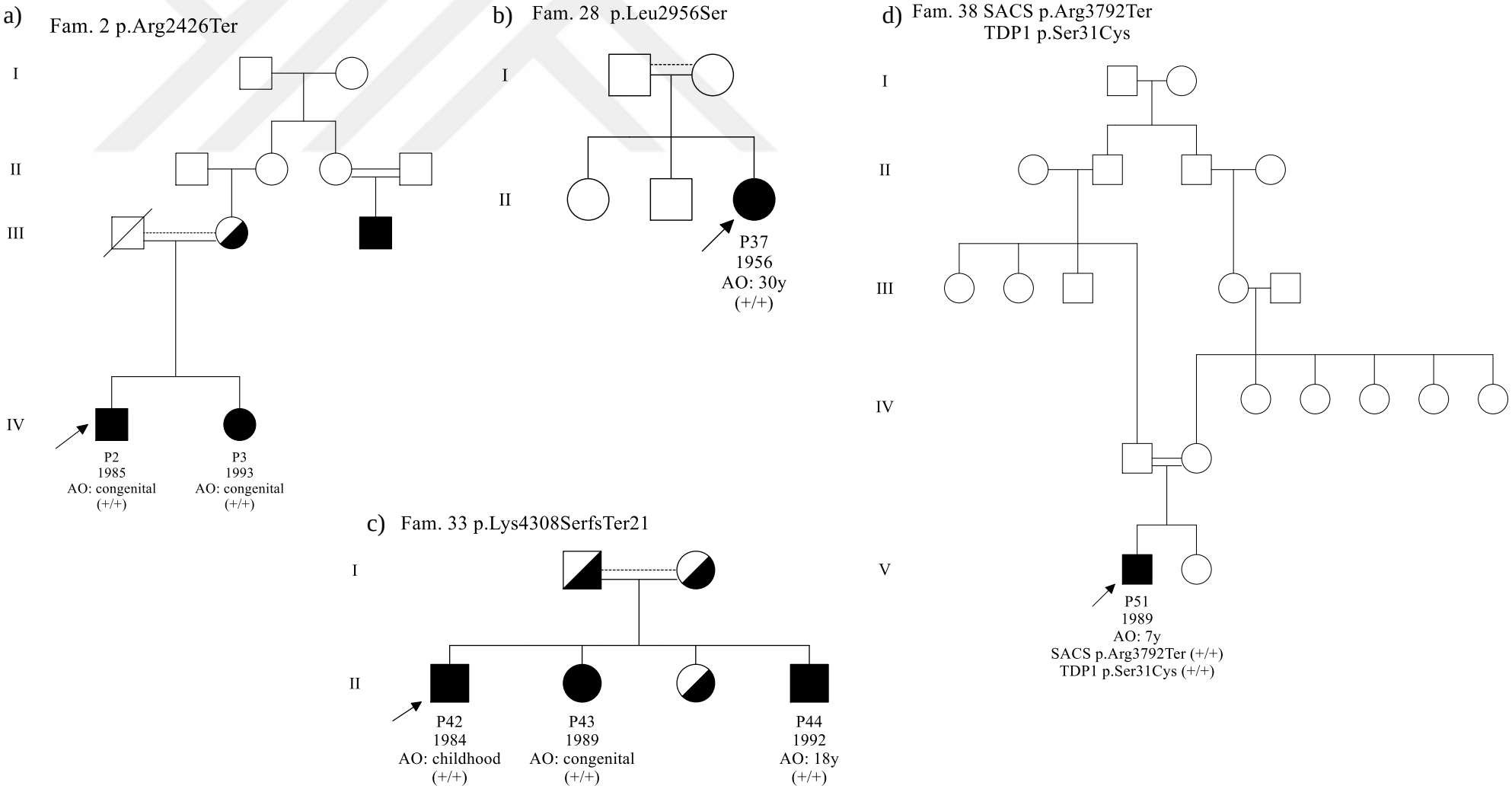
13:23914248	SACS	p.Tyr1256Cys	VUS-LP: PM2, PP3	26.7	0.9973		Disease causing	Path	D	D	-	+
13:23910280		p.Asp2578del	VUS-LP: PM2, PM4, PP3	-	n.a	5.59	n.a	n.a	n.a	n.a	-	
13:23913823	SACS	p.Cys1398Arg	LP: PM2, PP5, PP3	24.1	0.9966	6.06	Disease causing	Path	D	D	-	+
13:23906641	SACS	p.Arg3792Ter	Path: PVS1, PP5, PM2, PP3	40	0.9953	5.67	Disease causing	n.a	n.a	n.a	0.000007078	n.a
14:90429550	TDP1	p.Ser31Cys	VUS-LP: PM2, PP3	23.8	0.9869	5.55	Disease causing	B	T	D	-	
4:25125738	SEPSECS	p.Gly441Arg	LP: PM2, PP5, PP3	29.3	0.9993	5.32	Disease causing	Path	D	D	0.000007970	+
9:135156880	SETX	p.Gln2210Ter	Path: PVS1, PM2, PP3	47	0.9976	5.56	Disease causing automatic	n.a	n.a	n.a	-	+
16:89576947	SPG7	p.Leu78Ter	Path: PVS1, PM2, PP5	33	0.9053	5.31	Disease causing automatic	n.a	n.a	n.a	0.0003964	n.a

8:63976877	TTPA	c.553-2A>C	Path: PVS1, PM2, PP3	-	0.9953	5.48	Disease causing	n.a	n.a	n.a	-	+
2:219679482	CYP27A1	c.1476+2T>C	Path: PVS1, PP5, PM2, PP3	-	0.9923	5.25	Disease causing	n.a	n.a	n.a	-	+
16:89620361	SPG7	p.Met699Thr	VUS-LP: PM2, PP2, PP3	25.5	0.9937	5.21	Disease causing	Path	D	n.a	-	+
1:17331570	ATP13A2	c.289-3C>G	VUS: PM2, BP4	-	0.9213	4.38	n.a	n.a	n.a	n.a	-	+
16:89616953	SPG7	p.Ala572Val	LP: PM2, PP2, PP5, PP3	24.3	0.9989	5.84	Disease causing	Path	D	n.a	0.00003537	n.a
3:43618321	ANO10	p.Trp342Ter	Path: PVS1, PP5, PM2, PP3	40	0.9965	5.65	Disease causing automatic	n.a	n.a	n.a	0.000003980	+
1:245005357	COX20	p.Thr52Pro	Path: PS3, PP5, PP2, PM2, PP3	27.3	0.9969	5.85	Disease causing	Path	D	D	-	+

13:23906249	SACS	p.Leu3922Phe	VUS: PM2, PP3	16.02	0.9851	5.82	Disease causing	Path	D	D	-	+
13:23910075		p.Ile2647Asn	VUS-LP: PM2, PP3	27.2	0.9925	5.35	Disease causing	Path	D	D	-	
14:91757421	CCDC88C	p.Leu1374Val	VUS: PM2, PP3	22.7	0.9971	6.06	Disease causing	B	T	D	-	+
19:50826942	KCNC3	p.Arg423His	Path: PP5, PM1, PM2, PP3	27.4	0.9995	3.04	Disease causing	Path	D	D	-	+
11:62458783	BSCL2	p.Gly325Asp	VUS-LP: PM2, PP3	23.9	0.9977	5.27	Disease causing, Polymorp.	B	D	D	-	n.a
15:23049088	NIPA1	p.Gln244Arg	VUS-LP: PM2, PP3, PP5	-	n.a	5.69	n.a	n.a	n.a	n.a	-	-
16:730671	STUB1	p.Tyr49Cys	VUS-LP: PM2, PP2, PP3	24.4	0.9938	3.66	Disease causing	Path	D	D	-	-

Path: Pathogenic, LP: Likely pathogenic, VUS: Variant of uncertain significance, Polymorp: Polymorphism, D: Damaging, T: Tolerated, B: Benign, n.a: not available

Families with SPAX



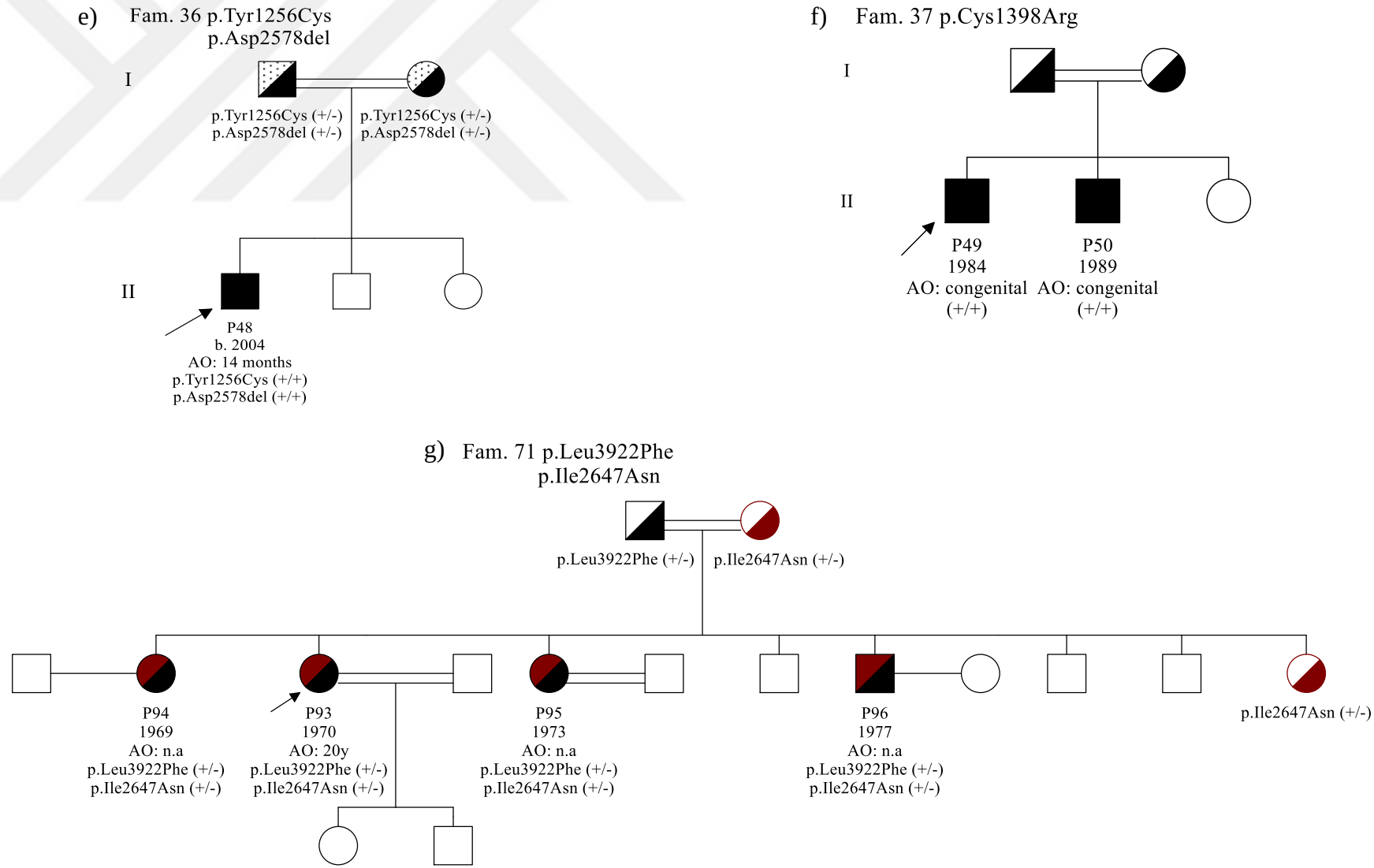
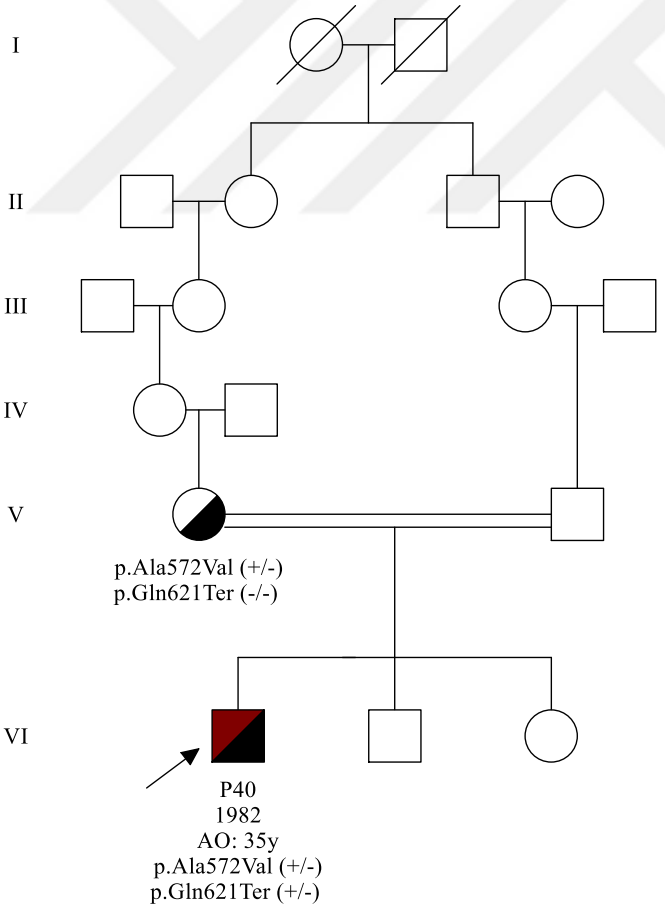
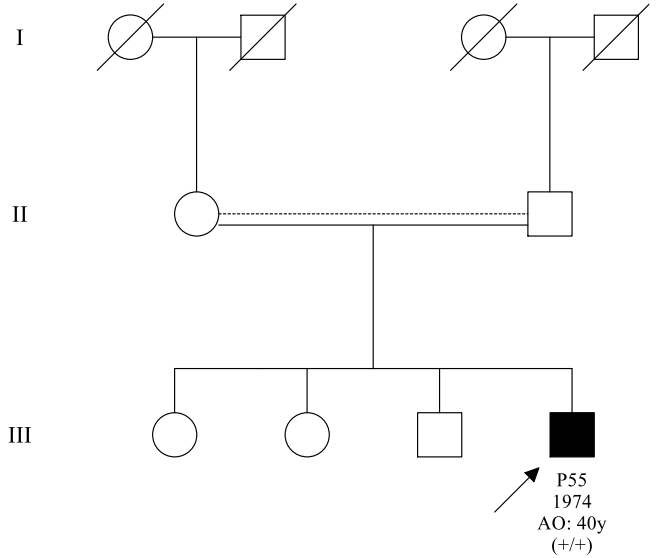


Figure 5.3 SACS families. a) Fam 2, b) Fam 28, c) Fam 33, d) Fam 38 (+ TDP1 variant), e) Fam 36, f) Fam 37, g) Fam 71.

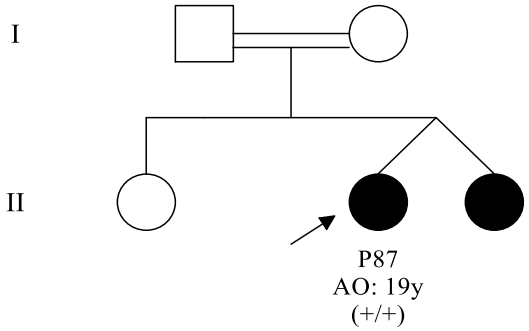
a) Fam. 31 p.Ala572Val
p.Gln621Ter



b) Fam. 41 p.Leu78Ter



c) Fam. 68 p.Ala572Val



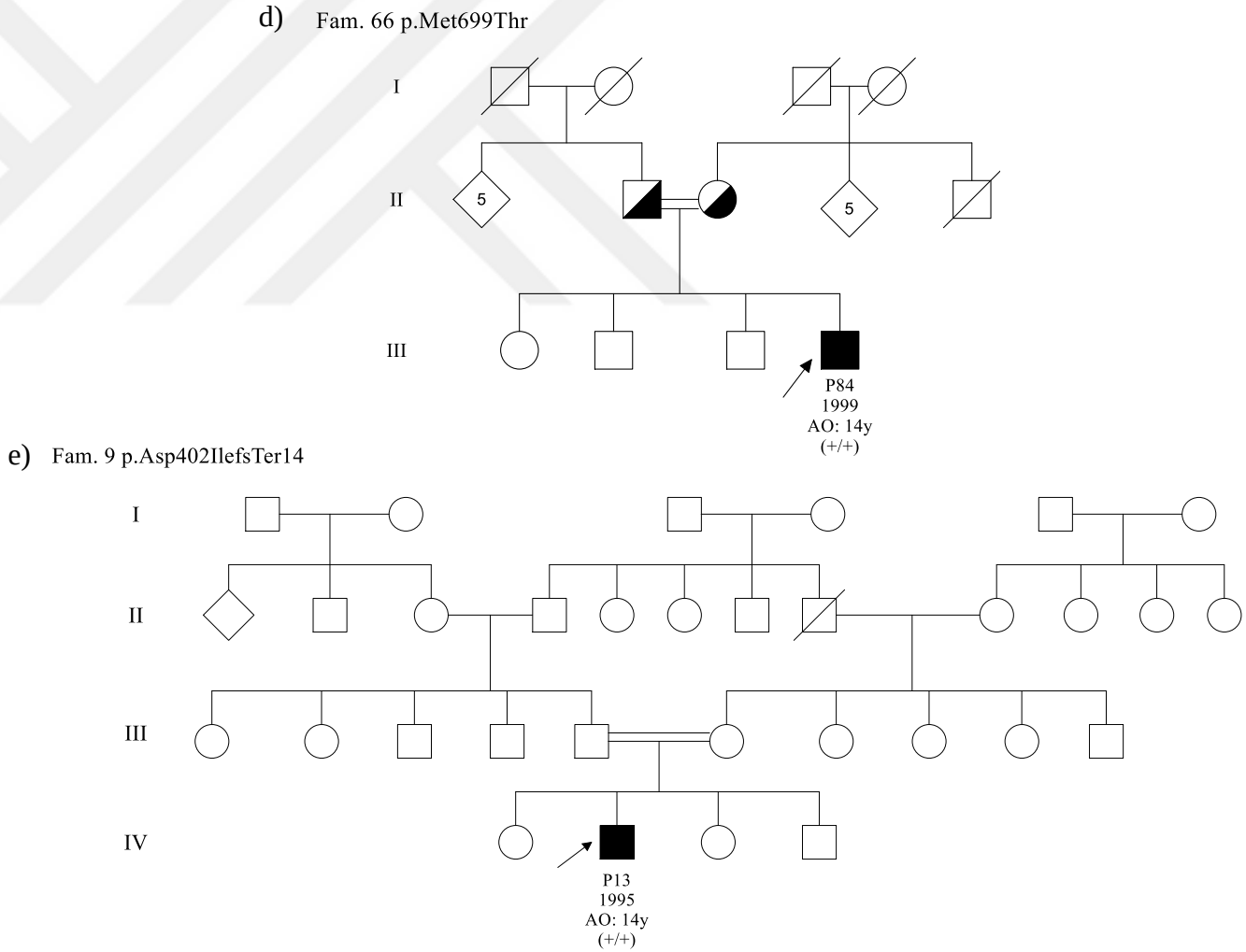


Figure 5.4 a) SPG7-Fam 31, b) SPG7-Fam 41, c) SPG7-Fam 68, d) SPG7-Fam 66, e) SPG11-Fam 9.

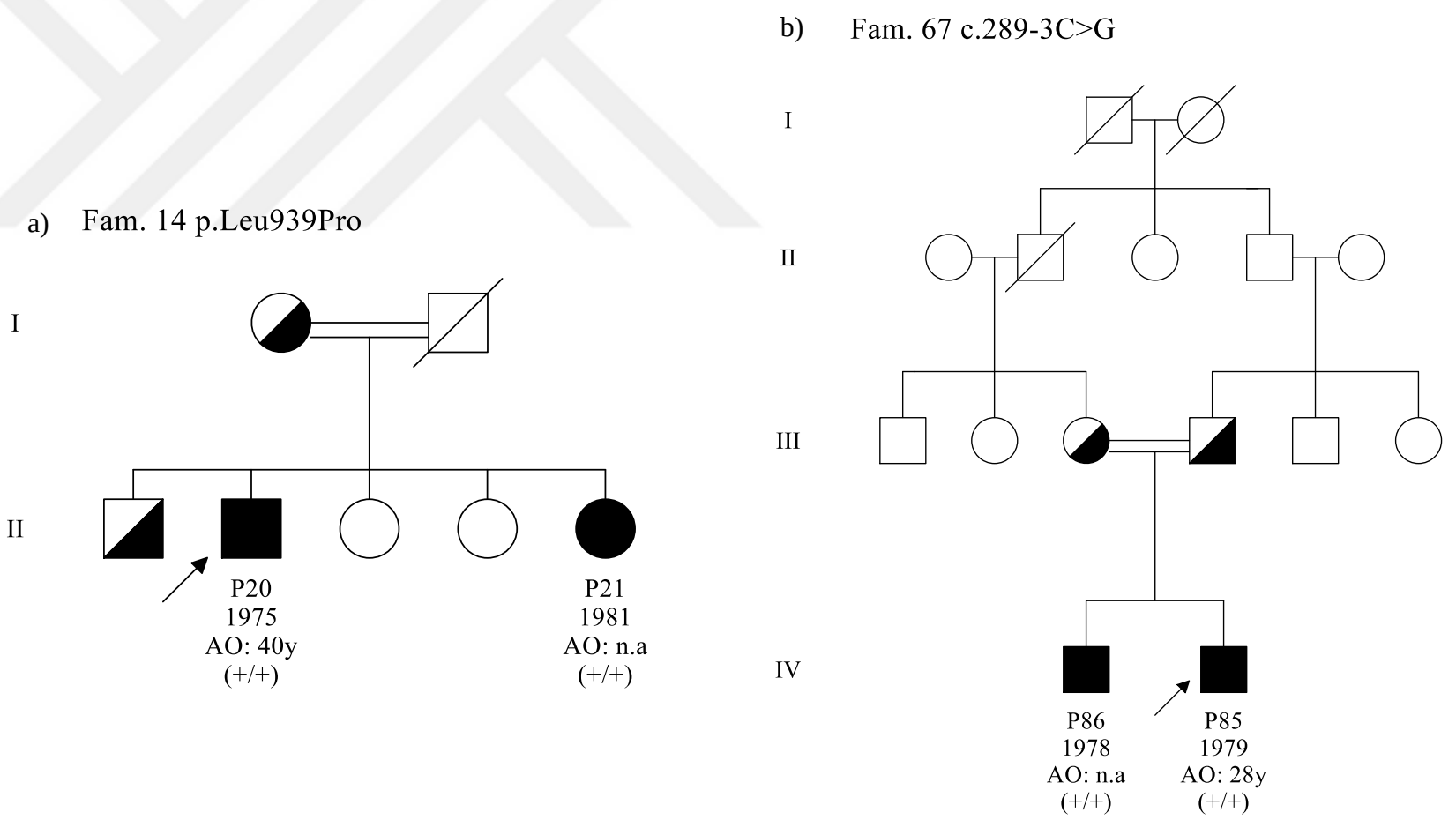


Figure 5.5 ATP13A2 families. a) Fam 14, b) Fam 67.

Families with AT and ATLD

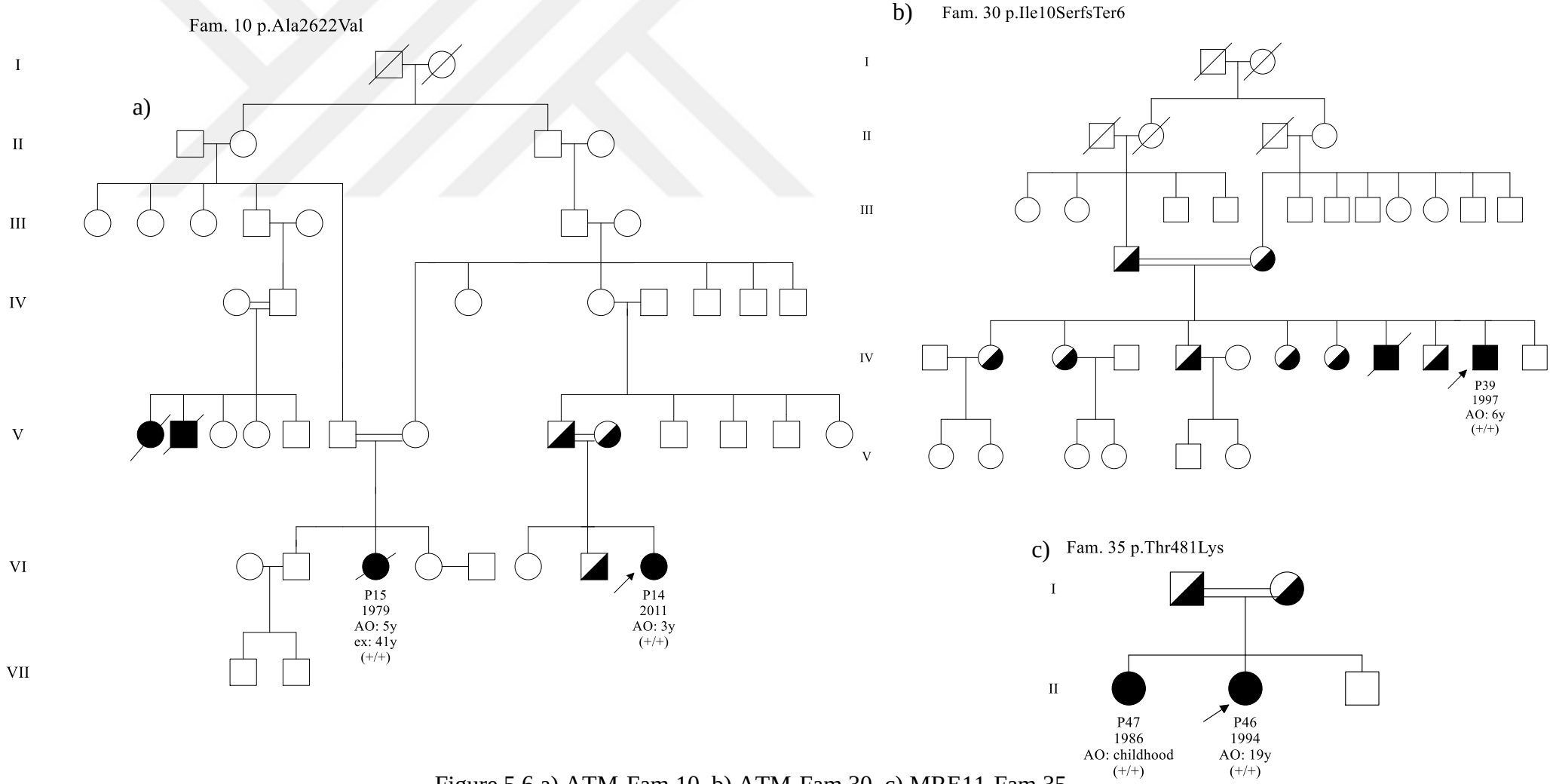


Figure 5.6 a) ATM-Fam 10, b) ATM-Fam 30, c) MRE11-Fam 35.

Families with Oculomotor Apraxia

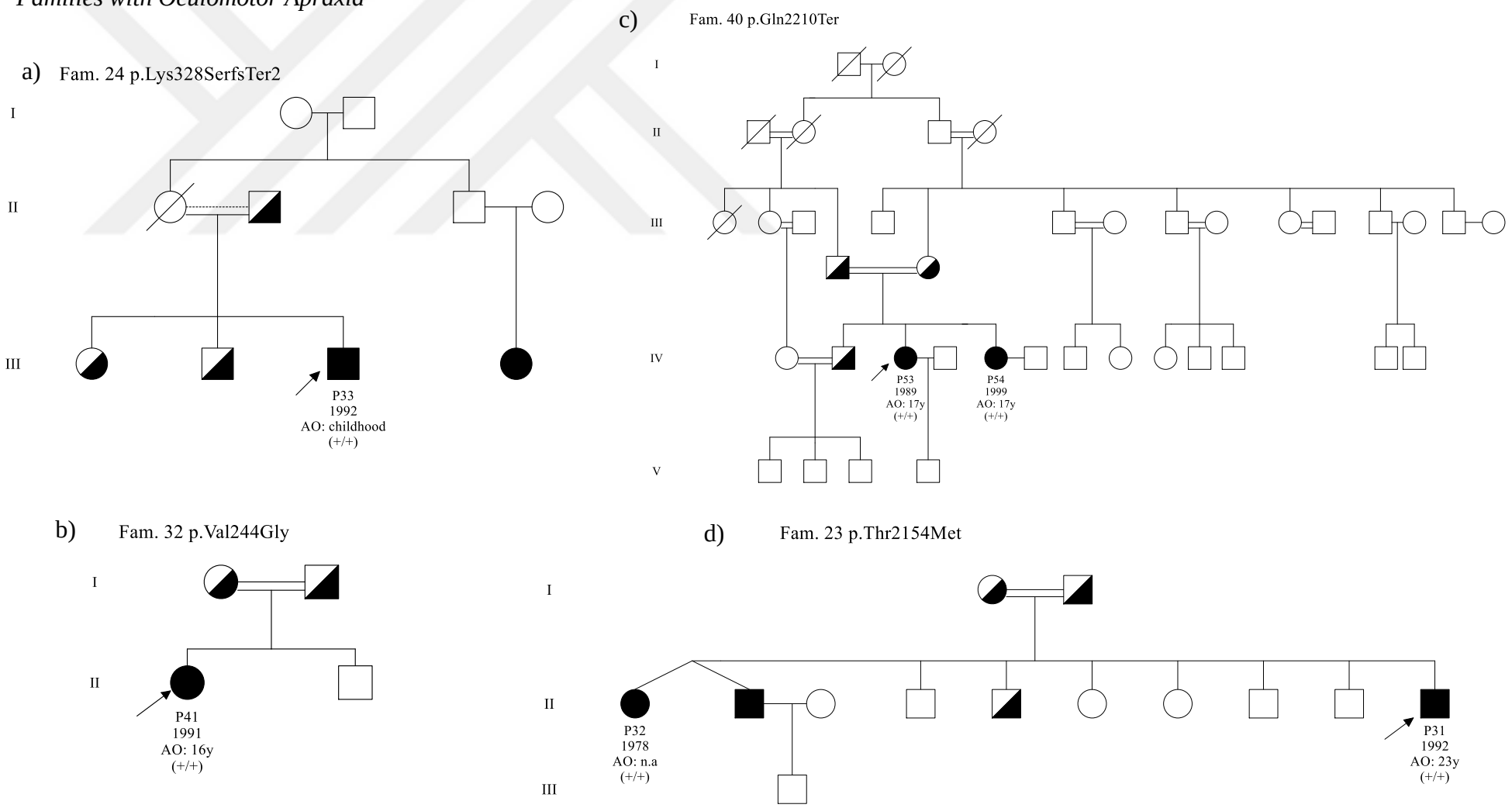


Figure 5.7 a) APTX-Fam 24, b) APTX-Fam 32, c) SETX-Fam 40, d) SETX-Fam 23.

Families with Other Ataxias

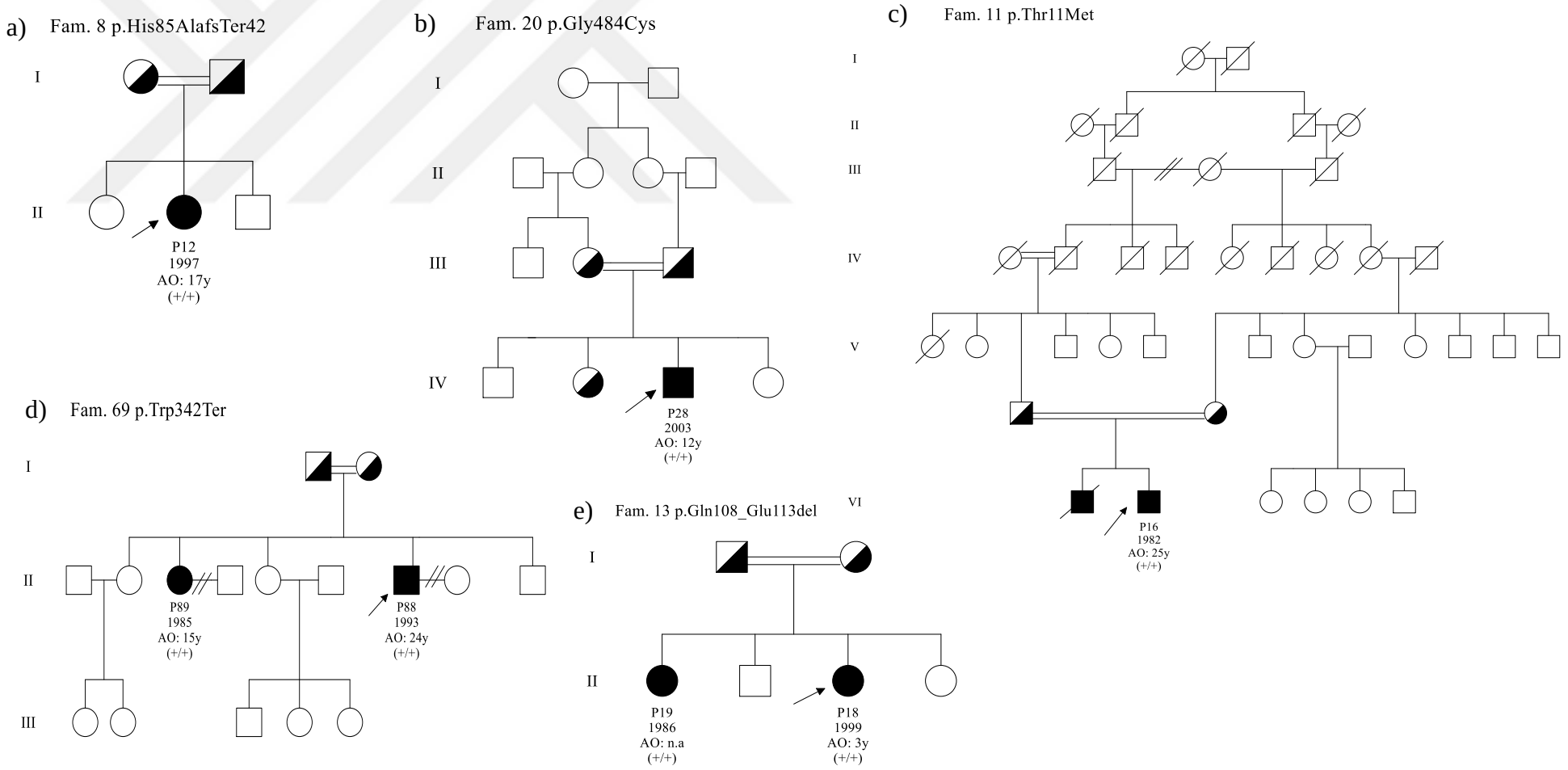


Figure 5.8 a) ADCK3-Fam 8, b) GAN-Fam 20, c) C19orf12-Fam 11, d) ANO10-Fam 69, e) OPA3-Fam 13.

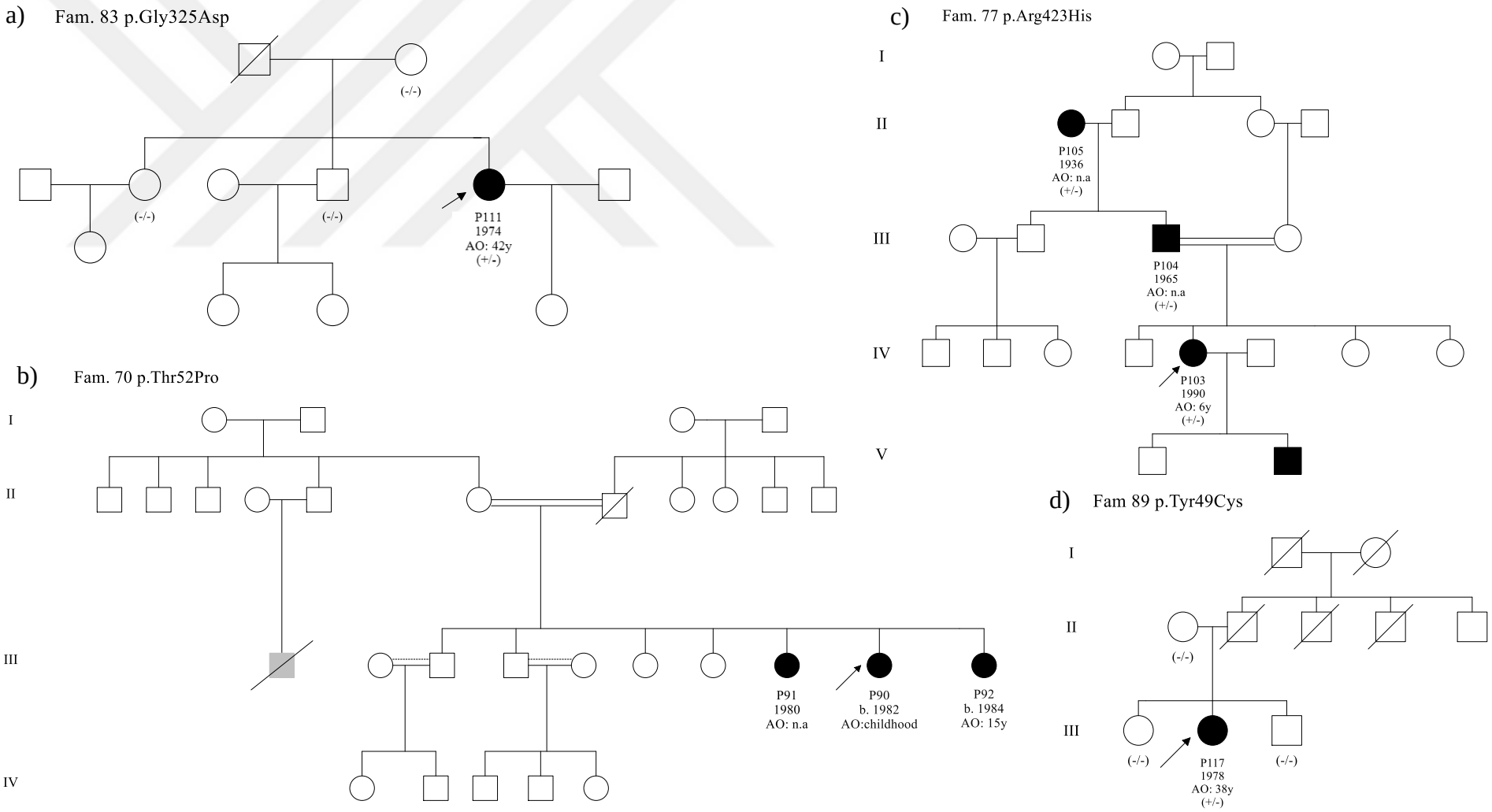


Figure 5.9 a) BSCL2-Fam 83, b) COX20-Fam 70, c) KCNC3-Fam 77, d) STUB1-Fam 89.

Chapter 6:
DISCUSSION

In this thesis, whole exome sequencing was applied to 95 index patients. AD and AR inheritance were present in 7.4% (7/95) and 75% (71/95), respectively. The remaining 18% (17/95) were apparently sporadic. Consanguinity was present in 73% of the cohort under study. Causative variants were identified in 37 out of 78 familial cases studied, which corresponds to a diagnostic yield of 47.4%. This ratio is in accordance with the literature (Retterer et al., 2016; Trujillano et al., 2017).

Biallelic variants were present in 35 families. Heterozygous causative variants were observed in two families with autosomal dominant inheritance pattern and in three apparently sporadic patients.

6.1 Insights from WES: Overlapping Genes, and Variable Mechanisms in a Heterogenous Cohort

In the cohort, which included 95 families, 43 pathogenic variants were detected in 28 genes (Figure 5.1, Table 5.1). In two different families with an AD inheritance pattern, *KCNC3* and *CCDC88C* were shown, while in three apparently sporadic distinct families, *STUB1, NIPA1*, and *BSCL2* were identified. In recessive cases, with FRDA excluded prior to WES, the most common gene was *SACS* which was present in seven different families, followed by *SPG7* in four families, and *ATM, APTX, SETX,* and *ATP13A2* in two families each. Variants in *ADCK3, ANO10, CAPN1, MRE11, SPG11, TTPA, ATCAY, C19orf12, COX20, CYP27A1, GAN, OPA3, POLR3A, PTRH2, SEPSECS, SQSTM1* were found once in the remaining families. Furthermore, a homozygous variant in an additional gene, *TDP1*, was identified in Family 38 which is one of the SACS families.

6.1.1 Genes in Mitochondrial Metabolism

SACS, SPG7, ADCK3, C19orf12, and *OPA3* are the genes that are known to have a role in mitochondrial metabolism. Autosomal recessive spastic ataxia of Charlevoix-Saguenay is caused by biallelic mutations in the *SACS* gene. Mitochondrial function is impaired secondarily in the ARSACS phenotype, which occurs with the loss of function of the sacsin protein. This protein, which is encoded by the *SACS* gene, acts as a regulator in the intermediate filament cytoskeleton organization; therefore, when the sacsin protein

is deficient, it leads to multiple changes, and mitochondrial fission is impaired (Synofzik et al., 2019). In this thesis, seven distinct families with ARSACS phenotypes were shown to have biallelic variants in the *SACS* gene: one homozygous in-frame and one homozygous missense in Family 36; one homozygous frameshift; four missense variants, two of which appear as compound heterozygous in Family 71, and two homozygous stop gained variants. *TDP1* was also present in one of the families (P51) with a homozygous stop gained variant in the *SACS* gene (Figure 6.1). Ages of onset and phenotypes of all seven families were consistent with the ARSACS phenotype.

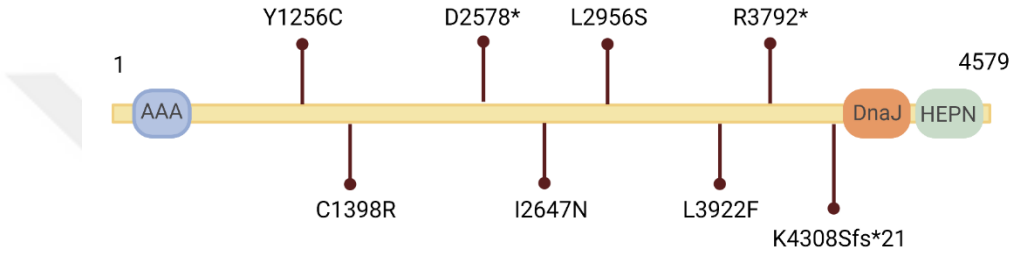


Figure 6.1 Mutations in the sacsin protein are shown in red. UBL: Ubiquitin like domain, HEPN: Higher eukaryotes and prokaryotes nucleotide-binding domain. Positions were taken from InterPro (https://www.ebi.ac.uk/interpro/protein/UniProt/Q9NZJ4/). Created with BioRender.com.

SPG7 encodes the mitochondrial protein paraplegin, which is involved in multiple mitochondrial processes. When this protein, part of the hetero-oligomeric proteolytic complex, is mutated, the mitochondrial protein control mechanism is impaired. In our cohort, SPG7 variants were observed as homozygous in three of the four independent families and as compound heterozygous in one. Family 31, 66, and 68 had a typical SPAX phenotype with missense variants, while Family 41 had a pure HSP phenotype with a stop gained variant (Table 3.1, Figure 6.2). These findings point to the clinical heterogeneity in hereditary spastic paraplegia and ataxia phenotypes. SPG7, which has a variable phenotype, is present in some cases as pure HSP, while some are more complex with other neurological signs (Yahikozawa et al., 2015).

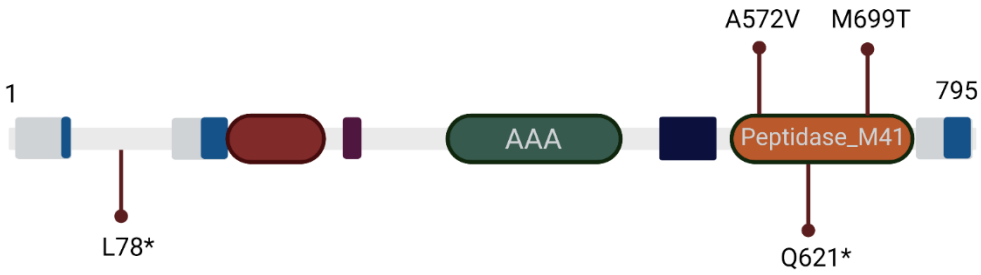


Figure 6.2 Mutations in the paraplegin protein are shown in red. AAA: ATPase family associated with various cellular activities domain, HEPN: Higher eukaryotes and prokaryotes nucleotide-binding domain. Positions were taken from Pfam (http://pfam.xfam.org/protein/Q9UQ90). Created with BioRender.com.

Autosomal recessive spinocerebellar ataxia type 9 is caused by biallelic variants in the *ADCK3* gene which encodes the homolog of yeast Coq8, involved in coenzyme Q10 synthesis (Vazquez-Fonseca et al., 2018). A family with a homozygous *ADCK3* variant was identified in our cohort. Our case had an ataxia clinical phenotype with predominating dystonia. The pathogenic variant was previously reported, and the age of onset of our case was in line with the literature (Traschütz et al., 2020).

Neurodegeneration with brain iron accumulation type 4 is caused by biallelic or heterozygous variants in the *C19orf12* gene. This gene encodes a small mitochondrial membrane-related protein whose function is still unknown (Venco et al., 2015). In our study, a family with a homozygous missense *C19orf12* variant was described. The clinical phenotype of our patient was characterized by gait disturbance, dysphasia, tremor, depression, and mental decline. Pallidal and nigral iron deposition were seen in our patient's MRI. The previously reported causative variant (p.Thr11Met) is common in adult Turkish patients with the NBIA4 phenotype (Olgiati et al., 2017).

3-methylglutaconic aciduria type 3 is caused by biallelic variants in the *OPA3* gene which encodes a protein in the mitochondrial inner membrane. This mitochondrial membrane protein is also a regulator of lipid metabolism and homeostasis (Wells et al., 2012). Our family with an *OPA3* variant harbored two affected siblings with signs of saccadic movement, dysarthria, sensory ataxia, dysdiadochokinesia, and bilateral muscle atrophy in hands and feet. The causative variant was previously published in a Kurdish-Turkish patient (Kleta et al., 2002).

6.1.2 Genes in DNA Repair Pathway

Ataxia-telangiectasia, which has a broad clinical spectrum, is caused by biallelic variants in the *ATM* gene which encodes the ATM protein. This protein is a member of the phosphatidylinositol 3-kinase family of proteins, and it takes a role in the double-strand break repair mechanism (Synofzik et al., 2019). In our cohort, two homozygous variants were found in *ATM* in two distinct families. One of the variants was a reported frameshift (Family 30), the other one was a reported missense variant (Family 10). While both families had the typical AT phenotype, the patient in Family 10 had additional polyneuropathy and developmental cognitive findings.

Ataxia-telangiectasia-like disorder is caused by biallelic variants in the *MRE11* gene which is part of the hMre11/hRad50/Nbs1 complex. This gene encodes double-strand break repair protein MRE11 (Synofzik et al., 2019). Two siblings in a family with a homozygous *MRE11* variant were identified in our cohort. They had oculomotor apraxia, dystonia, parkinsonism, upper motor involvement, and truncal ataxia. The interesting point in this family was the dopa-responsiveness of parkinsonism and dystonia (Charlesworth et al., 2013; Thompson et al., 2014). Dopa-responsive chorea and dystonia were previously reported in the AT. Subsequently, AT was collected under the title of AT dopa-responsive disease look-alikes (Lee et al., 2018). This family, previously reported in our center, suggests that ATLD may also be collected under the same title (Ser et al., 2020).

Biallelic variants in the *APTX* gene, which encodes aprataxin that is involved in the DNA repair mechanism, cause Ataxia-Oculomotor Apraxia 1 (Synofzik et al., 2019). In our cohort, two causative variants were identified in the *APTX* gene in two different families. One of the variants was a reported missense variant, the other one was a reported frameshift. Both families in the cohort had an ATX phenotype (Table 3.1), which is in accordance with the APTX finding.

SETX encodes senataxin, a putative DNA/RNA helicase that plays a significant role in transcription and in the maintenance of genome integrity (Cohen et al., 2018). In two independent families, homozygous mutations in *SETX* were present. Two missense variants were found, one was already reported, and the other is novel. Biallelic variants in the *SETX* gene cause Ataxia-Oculomotor Apraxia 2. The genetic results explain the disease phenotype of the patients and their affected relatives.

Spinocerebellar ataxia with axonal neuropathy is associated with biallelic variants in the *TDP1* gene, a protein that is jointly involved in DNA end processing in single-strand repair (Synofzik et al., 2019). In our study, a homozygous missense variant in the *TDP1* gene was found in addition to a homozygous stop gained variant in the *SACS* gene in Family 38; both were previously reported. Sanger sequencing confirmed the presence of both variants in the index patient, however, family members were not available. Therefore, family segregation could not be validated. The SPAX phenotype of the patient is thought to be explained by the *SACS* variant. To show the effect of the *TDP1* variant on the course of the disease, family segregation analysis should be performed, and if segregation is compatible, further studies should be performed.

6.1.3 Genes in Lipid Metabolism

Heterozygous mutations in the *BSCL2*/*seipin*, which encodes a multi-pass transmembrane protein that is involved in lipid mechanism, causes Silver Spastic Paraplegia Syndrome (Cartwright et al., 2012). A family with a heterozygous *BSCL2* variant was identified in our cohort. The variant was not observed in available family members; the patient's father could not be analyzed, because he was not more available. In this family, which has an apparently sporadic inheritance, the variant stems either from the father or it has a de novo occurrence, which we cannot test any more since the father is not available.

6.1.4 Genes in Autophagy and Lysosomal Activity

SPG11 and *ATP13A2* genes belong to the above group. Biallelic variants in *SPG11* cause three distinct phenotypes, SPG11, CMT, or juvenile ALS disease. Loss of function of spataxin, which is encoded by *SGP11*, causes lysosomal lipid accumulation. The protein also plays a role in autophagic lysosome reformation (Fraiberg et al., 2020). We had one family with the pure HSP phenotype caused by *SPG11*. The pathogenic frameshift variant was previously published. The genetic finding is in accordance with the phenotype of the patient.

ATP13A2 acts in alpha-synuclein accumulation and neurotoxicity. The loss of *ATP13A2* function results in impaired lysosomal function (Usenovic et al., 2012). Biallelic variants in *ATP13A2* cause SPG78. In our cohort, two causative variants were found in the *ATP13A2* gene in two distinct SPAX families. Both variants were novel, one

being a homozygous missense, and the other a splice region mutation. Splice region variants must always be considered, as variants that affect the splicing of precursor mRNAs have a critical function in the development of the disease phenotype (Divina et al., 2009).

6.1.5 Other Genes and More Mechanisms

The *ANO10* gene functions in the calcium-activated chloride channel (Vermeer et al., 2010). Biallelic variants in *ANO10* cause SCAR10. Our patients were referred to our laboratory with recessive ataxia. Our patients, two sibs, offspring of consanguineous parents, have AO of 24 and 15. The homozygous and pathogenic *ANO10* variant explains the disease phenotype of the patients.

COX20 encodes a protein involved in the biogenesis of mitochondrial complex IV which is responsible for reducing molecular oxygen and oxidation of cytochrome C (Otero et al., 2019). Mitochondrial complex IV deficiency nuclear type 11 is caused by homozygous mutations in the *COX20* gene. A previously reported homozygous variant in the *COX20* gene was detected in the patients. All three affected siblings from Family 70 had recessive ataxia and milder psychomotor retardation. The Thr52Pro variant identified in our patients is associated with muscle hypotonia and ataxia in the literature, however, our patients so far do not present with muscle hypotonia (Xu et al., 2019).

Spinocerebellar ataxia 13 is caused by a heterozygous mutation in the *KCNC3* gene which is responsible for the voltage-gated potassium channel. The gene is expressed at high levels in cerebellar Purkinje cells (Bertini et al., 2018; Bushart et al., 2019). The KCNC3 patients in our cohort were referred to our laboratory with dominant ataxia and tremor. The causative variant was previously reported, and the phenotype of the patients is in accordance with the SCA13 disease.

Giant axonal neuropathy-1 is caused by a biallelic mutation in the *GAN* gene which encodes gigaxonin. This protein plays a role in cell survival, organelle motility, and mitochondrial bioenergetics (Opal, 2003 [Updated 2021]). In our cohort, a causative variant was identified in the *GAN* gene in Family 20. The GAN patient in our cohort was referred to our laboratory with ataxia, polyneuropathy, and deep sensory involvement. The genetic result is consistent with the patient's disease phenotype.

Mutations in *STUB1* cause both autosomal dominant (SCA48) spinocerebellar ataxia and autosomal recessive SCAR16. *STUB1* encodes the C-terminus of HSC70-interacting

protein. This protein plays a role in the protein quality control system (Pakdaman et al., 2021). The patient from Family 89 had cerebellar ataxia and cognitive decline. The heterozygous variant (Tyr49Cys) characterized as VUS was present in the patient and absent in available family members. The father's sample was not available since he had passed away. The variant may have been inherited from the father or has occurred *de novo*. The genetic finding explains the patient's phenotype who was earlier classified as apparently sporadic.

6.2 Novel Genotype-Phenotype Associations Identified Throughout This Thesis

Next-generation sequencing technologies, which have significantly contributed to the number of genes associated with ataxia or other NDs, have dramatically blurred the line among these diseases. In our study, WES in 95 families demonstrated 16 novel variants, thus unraveled novel genotype-phenotype associations in known ataxia genes. These consist of variants in HSP-associated or other genes that broaden the spectrum of ataxia genes and identify novel phenotypes.

Variants in genes overlapping with the spastic paraplegia phenotype were identified in 22 families, corresponding to more than half of the solved families. Fourteen out of these 22 families had a SPAX phenotype. In the other families solved, three had a pure HSP phenotype. There were five families with an ataxia phenotype without spasticity, and only one of the ataxia families had a PNP phenotype in addition to ataxia.

In this study, novel phenotypic features were demonstrated in known ataxia genes. For example, muscle hypotonia was not present in Family 70, although the variant in the *COX20* gene that causes mitochondrial complex IV deficiency nuclear type 11 was previously reported with muscle hypotonia in addition to ataxia (Xu et al., 2019). Furthermore, the dopa-responsiveness of parkinsonism and dystonia in our patients (Family 35) supports the notion that the ATLD phenotype may be grouped under the title of AT dopa-responsive disease look-alikes (Ser et al., 2020).

6.3 The Importance of Variants of Uncertain Significance

WES-based molecular analysis in disease diagnosis introduced the major question of an ever-increasing number of variants of uncertain significance. The correct interpretation of these variants is of primary importance for the clinical management of patients. Genotype-phenotype correlations and intrafamilial allele segregation are critical factors

in defining the pathogenicity of VUS. Precise clinical assessments, deep and reverse phenotyping are significantly helpful in evaluating these variants.

In the framework of this study, 17 out of 43 variants determined, presented with VUS or VUS-LP criteria according to Varsome (Table 5.2). Thirteen of these variants were novel. All VUS variants were comprehensively discussed with the relevant expert clinicians to make an accurate final diagnosis in which also family segregation analysis was instrumental.

6.4 Limitations of Whole Exome Sequencing and Challenges of the NGS Era

In the framework of this thesis, a diagnostic yield of 42% was obtained by WES. In the remaining 58%, WES has failed to identify variations in disease-related genes. Sequencing other family members may help to identify more candidate variants or may reduce the number of candidate variants to create a shorter list. There can also be technical or strategical problems. The technical limitations of WES include the lack of non-coding regions of the genome and its limited ability to detect repeat expansion and structural variations. Moreover, the disease may be caused by a variant in a novel gene that has not previously been associated with the related phenotype.

Today, the lack of non-coding regions in the genome is a major drawback of WES. Mutations in the intronic or regulatory regions of the known disease genes may change their expression and can cause the disease phenotype. This drawback may be overcome by applying a more comprehensive analysis using whole genome sequencing. However, data interpretation is more challenging with WGS which brings up additional and huge computational problems. Compared to WES, more sophisticated bioinformatics tools are required for WGS analysis. Extending the targeted areas of exome capture kits may be a solution until WGS becomes more user-friendly.

The low coverage of the gene in some regions is another technical problem of WES-based analysis. If coverage is low for a heterozygous variant, it might be missed due to a few reads of a specific area (false-negative).

On the other hand, the false-positive issue produced by NGS technologies is critical for accurate genetic diagnosis, as well, and should not be underestimated. A false-positive result may result from erroneous polymerase reactions, biases in library construction, 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). To avoid these false-positive results, the validation and family segregation analysis in candidate variants obtained from WES analysis should be performed by Sanger sequencing and/or other methods.

Due to the read length limitation, detecting structural variants such as repeat regions in the genome and large insertions or deletions is another shortcoming of exome sequencing. New algorithms are rapidly being improved to identify structural variants from WES data as the read length increases with recent advances in NGS technologies. Different variant detection tools can identify CNVs from WES data, but they can result in high false-positives due to non-uniform reading depth. Therefore, CNV results obtained from WES are not always reliable.

6.5 Future Prospects in the Remaining Cases to be Solved

In this study, causative variants were identified in 40 distinct families; the challenges and limitations of WES are one factor that is in the way to unravel the remaining families.

Another factor may be novel genes that are not yet associated with a disease. Novel genes may emerge in the remaining unsolved cases. Therefore, deep phenotyping and the last version of data analysis tools are critical. To find novel genes in the remaining unsolved cases, our laboratory collaborates with ARCA research groups in the GENESIS and the Solve-RD Projects. GENESIS, which is a collaborative cloud-based system and a matchmaking software, contains the largest ataxia NGS dataset worldwide (>2,000 ataxia NGS datasets) (Traschütz et al., 2021). The important question for the researchers in this project is whether there is a ‘second family’ similar to the phenotype of unsolved families collected from different centers. Furthermore, Solve-RD, which is a research project to solve rare diseases, is a systematic data sharing and collaborative analysis platform in innovative ways (Zurek et al., 2021). These two important databases may provide a significant advantage in deciphering our unsolved families. A first example is the COX20 family, which was identified via GENESIS.

Oligogenic inheritance, which means that multiple genes or other risk factors may contribute to disease phenotype, may be also present in unsolved ataxia cases. Thus, the identification of epigenetic risk factors and modifier genes in ataxias is another crucial point to reveal the causes in some of these cases.

6.6 Diagnosis of Ataxias: WES is the Golden Standard

WES is a state-of-the-art method that sequences the whole protein-coding region in the genome. With its high throughput capacity and proportionally low cost, WES is a more effective tool today than all other conventional and NGS methods.

Our results in the cohort under study demonstrate that ataxias have a great genetic heterogeneity. Some of the genes are associated with neurological diseases other than ataxia. Most neurodegenerative diseases overlap genetically and clinically e.g., in this study, variants were identified in genes with heterogenous phenotypes like *CYP27A1*, *PTRH2*, and *SEPSECS* in patients with a predominant ataxia. This emphasizes the importance and power of WES-based molecular analysis. WES is instrumental to understand the complex genetic structure behind ataxias and identify novel genes/variations and novel genotype-phenotype associations.

Chapter 7:
CONCLUSION

Hereditary ataxias are rare and heterogeneous movement disorders, the complex genetic mechanisms of which have not been fully understood yet. To unravel the complete pathogenesis of ataxias that have a broad genotypic spectrum overlapping with other NDs, novel genetic and environmental factors have to be defined.

The aim of this thesis was to understand the complex genetic structure behind ataxias in Turkey and to identify novel genes and variants by using WES, bioinformatic analysis, and related tools. We were able to detect causative mutations in 40 families and showed that the more frequent occurrence of recessive ataxias is caused by consanguineous marriages, which are common in Turkey. The genetic findings in this thesis, which allowed us to oversee the comprehensive picture of the genetic architecture of ataxias in Turkey, also helped to determine the differential diagnosis.

WES is an unbiased and highly effective NGS approach to detect genetic variations in complex and heterogeneous ataxias. WES is the gold standard of today's technology in investigating complex genetic diseases, notwithstanding its limitations and challenges.

Once all the genes and mutations that cause ataxias are known, the molecular mechanisms leading to cerebellar degeneration will be more comprehensively unraveled and understood. This promising approach will also help to develop personalized gene-based therapies.

The results presented in the framework of this thesis will hopefully contribute to the diversity of genetic factors underlying ataxias and pave the ways for more definitive diagnosis of ataxias in Turkey.

BIBLIOGRAPHY

Amirifar, P., Ranjouri, M. R., Yazdani, R., Abolhassani, H., Aghamohammadi, A. J. P. A., & Immunology. (2019). Ataxia-telangiectasia: A review of clinical features and molecular pathology. *30*(3), 277-288.

Anheim, M., Tranchant, C., & Koenig, M. J. N. E. J. o. M. (2012). The autosomal recessive cerebellar ataxias. *366*(7), 636-646.

Bertini, E., Zanni, G., & Boltshauser, E. (2018). Nonprogressive congenital ataxias. *Handb Clin Neurol, 155*, 91-103.

Botstein, D., & Risch, N. J. N. g. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *33*(3), 228-237.

Bouchard, J., Barbeau, A., Bouchard, R., & Bouchard, R. J. C. J. o. n. s. (1978). Autosomal recessive spastic ataxia of Charlevoix-Saguenay. *5*(1), 61-69.

Boycott, K. M., Vanstone, M. R., Bulman, D. E., & MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet, 14*(10), 681-691. doi:10.1038/nrg3555

Bushart, D. D., & Shakkottai, V. G. (2019). Ion channel dysfunction in cerebellar ataxia. *Neuroscience letters, 688*, 41-48.

Cartwright, B. R., & Goodman, J. M. (2012). Seipin: from human disease to molecular mechanism. *Journal of lipid research, 53*(6), 1042-1055.

Charlesworth, G., Mohire, M. D., Schneider, S. A., Stamelou, M., Wood, N. W., & Bhatia, K. P. (2013). Ataxia telangiectasia presenting as dopa-responsive cervical dystonia. *Neurology, 81*(13), 1148-1151.

Cohen, S., Puget, N., Lin, Y.-L., Clouaire, T., Aguirrebengoa, M., Rocher, V., . . . Legube, G. (2018). Senataxin resolves RNA: DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nature communications, 9*(1), 1-14.

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

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

Divina, P., Kvitkovicova, A., Buratti, E., & Vorechovsky, I. (2009). Ab initio prediction of mutation-induced cryptic splice-site activation and exon skipping. *European Journal of Human Genetics, 17*(6), 759-765.

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

Fraiberg, M., & Elazar, Z. (2020). Genetic defects of autophagy linked to disease. *Progress in Molecular Biology and Translational Science, 172*, 293-323.

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

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., . . . Karyadi, D. J. T. A. J. o. H. G. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *99*(4), 877-885.

Jones, T. M., Shaw, J. D., Sullivan, K., & Zesiewicz, T. A. (2014). Treatment of cerebellar ataxia. *Neurodegener Dis Manag, 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. J. N. a. r. (2007). OligoCalc: an online oligonucleotide properties calculator. *35*(suppl_2), W43-W46.

Kleta, R., Skovby, F., Christensen, E., Rosenberg, T., Gahl, W. A., & Anikster, Y. (2002). 3-Methylglutaconic aciduria type III in a non-Iraqi-Jewish kindred: clinical and molecular findings. *Molecular genetics and metabolism, 76*(3), 201-206.

Klockgether, T. (2010). Sporadic ataxia with adult onset: classification and diagnostic criteria. *Lancet Neurol, 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

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., & Massouras, A. (2019). VarSome: the human genomic variant search engine. *Bioinformatics, 35*(11), 1978-1980. doi:10.1093/bioinformatics/bty897

Ku, C. S., Wu, M., Cooper, D. N., Naidoo, N., Pawitan, Y., Pang, B., . . . Soong, R. (2012). Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev Mol Diagn, 12*(3), 241-251. doi:10.1586/erm.12.10

Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., . . . Liu, C. J. N. a. r. (2020). ClinVar: improvements to accessing data. *48*(D1), D835-D844.

Lee, W.-W., Jeon, B., & Kim, R. (2018). Expanding the spectrum of dopa-responsive dystonia (DRD) and proposal for new definition: DRD, DRD-plus, and DRD look-alike. *Journal of Korean medical science, 33*(28).

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

Marian, A. J. M. D. c. j. (2014). Sequencing your genome: what does it mean? *, 10*(1), 3.

Mariani, L., Rivaud-Péchoux, S., Charles, P., Ewenczyk, C., Meneret, A., Monga, B., . . . Degos, B. J. S. r. (2017). Comparing ataxias with oculomotor apraxia: a multimodal study of AOA1, AOA2 and AT focusing on video-oculography and alpha-fetoprotein. *7*(1), 1-9.

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. J. E. s. i. t. c. o. a. M. d. N. G. (2010). Buckingham KJ Lee C, etal. *42*, 30-35.

Olgiati, S., Doğu, O., Tufekcioglu, Z., Diler, Y., Saka, E., Gultekin, M., . . . Breedveld, G. J. (2017). The p. Thr11Met mutation in c19orf12 is frequent among adult Turkish patients with MPAN. *Parkinsonism & related disorders, 39*, 64-70.

Opal, P. (2003 [Updated 2021]). GAN-Related Neurodegeneration. *GeneReviews® [Internet]*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK1136/

Otero, M. G., Tiongson, E., Diaz, F., Haude, K., Panzer, K., Collier, A., . . . Cui, H. (2019). Novel pathogenic COX 20 variants causing dysarthria, ataxia, and sensory neuropathy. *Annals of clinical and translational neurology, 6*(1), 154-160.

Pakdaman, Y., Berland, S., Bustad, H. J., Erdal, S., Thompson, B. A., James, P. A., . . . Berge, L. I. (2021). Genetic Dominant Variants in STUB1, Segregating in Families with SCA48, Display In Vitro Functional Impairments Indistinctive from Recessive Variants Associated with SCAR16. *International journal of molecular sciences, 22*(11), 5870.

Pfeffer, G., Pyle, A., Griffin, H., Miller, J., Wilson, V., Turnbull, L., . . . Chinnery, P. F. (2015). SPG7 mutations are a common cause of undiagnosed ataxia. *Neurology, 84*(11), 1174-1176. doi:10.1212/wnl.0000000000001369

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. J. B. (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *31*(5), 761-763.

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

Ser, M. H., Tekgül, Ş., Gündüz, A., Kızıltan, M. E., Kızıltan, G., & Başak, A. N. (2020). Ataxia telangiectasia like disorder: Another dopa-responsive disorder look-alike? *Parkinsonism & related disorders, 74*, 22-24.

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. J. N. a. r. (2001). dbSNP: the NCBI database of genetic variation. *29*(1), 308-311.

Slatko, B. E., Gardner, A. F., & Ausubel, F. M. J. C. p. i. m. b. (2018). Overview of next-generation sequencing technologies. *122*(1), e59.

Subramony, S., Moscovich, M., & Ashizawa, T. (2015). Genetics and Clinical Features of Inherited Ataxias. In *Movement Disorders* (pp. 939-978): Elsevier.

Synofzik, M., & Németh, A. H. (2018). Recessive ataxias. *Handb Clin Neurol, 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

Taylor, A., Groom, A., & Byrd, P. J. D. r. (2004). Ataxia-telangiectasia-like disorder (ATLD)—its clinical presentation and molecular basis. *3*(8-9), 1219-1225.

Thompson, S., Iyer, A., Byrd, P., Taylor, M., & Spinty, S. (2014). Dopa-responsive dystonia and chorea as a presenting feature in ataxia-telangiectasia. *Movement disorders clinical practice, 1*(3), 249.

Traschütz, A., Reich, S., Adarmes, A. D., Anheim, M., Ashrafi, M. R., Baets, J., . . . Gagnon, C. (2021). The ARCA Registry: a collaborative global platform for advancing trial readiness in autosomal recessive cerebellar ataxias. *Frontiers in neurology, 12*.

Traschütz, A., Schirinzi, T., Laugwitz, L., Murray, N. H., Bingman, C. A., Reich, S., . . . Bertini, E. (2020). Clinico-genetic, imaging and molecular delineation of COQ8A-ataxia: A multicenter study of 59 patients. *Annals of neurology, 88*(2), 251-263.

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

Usenovic, M., Tresse, E., Mazzulli, J. R., Taylor, J. P., & Krainc, D. (2012). Deficiency of ATP13A2 leads to lysosomal dysfunction, α-synuclein accumulation, and neurotoxicity. *Journal of Neuroscience, 32*(12), 4240-4246.

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

Vazquez-Fonseca, L., Doimo, M., Calderan, C., Desbats, M., Acosta, M., Cerqua, C., . . . Sartori, G. (2018). Mutations in COQ8B found in patients with steroid-resistant nephrotic syndrome Alter COQ8B function.

Venco, P., Bonora, M., Giorgi, C., Papaleo, E., Iuso, A., Prokisch, H., . . . Tiranti, V. (2015). Mutations of C19orf12, coding for a transmembrane glycine zipper containing mitochondrial protein, cause mis-localization of the protein, inability to respond to oxidative stress and increased mitochondrial Ca2+. *Frontiers in genetics, 6*, 185.

Vermeer, S., Hoischen, A., Meijer, R. P., Gilissen, C., Neveling, K., Wieskamp, N., . . . Knoers, N. (2010). Targeted next-generation sequencing of a 12.5 Mb homozygous region reveals ANO10 mutations in patients with autosomal-recessive cerebellar ataxia. *Am J Hum Genet, 87*(6), 813-819. doi:10.1016/j.ajhg.2010.10.015

Vermeer, S., van de Warrenburg, B., Willemsen, M., Cluitmans, M., Scheffer, H., Kremer, B., & Knoers, N. J. J. o. m. g. (2011). Autosomal recessive cerebellar ataxias: the current state of affairs. *48*(10), 651-659.

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

Wells, T., Davies, J. R., Guschina, I. A., Ball, D. J., Davies, J. S., Davies, V. J., . . . Votruba, M. (2012). Opa3, a novel regulator of mitochondrial function, controls thermogenesis and abdominal fat mass in a mouse model for Costeff syndrome. *Human molecular genetics, 21*(22), 4836-4844.

Xu, H., Ji, T., Lian, Y., Wang, S., Chen, X., Li, S., . . . Dong, X. (2019). Observation of novel COX20 mutations related to autosomal recessive axonal neuropathy and static encephalopathy. *Human genetics, 138*(7), 749-756.

Yahikozawa, H., Yoshida, K., Sato, S., Hanyu, N., Doi, H., Miyatake, S., & Matsumoto, N. J. H. g. v. (2015). Predominant cerebellar phenotype in spastic paraplegia 7 (SPG7). *2*(1), 1-3.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. J. B. b. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *13*(1), 1-11.

Yohe, S., Thyagarajan, B. J. A. o. p., & medicine, l. (2017). Review of clinical next-generation sequencing. *141*(11), 1544-1557.

Zurek, B., Ellwanger, K., Vissers, L. E., Schüle, R., Synofzik, M., Töpf, A., . . . Gilissen, C. (2021). Solve-RD: systematic Pan-European data sharing and collaborative analysis to solve Rare Diseases. *European Journal of Human Genetics*, 1-6.

APPENDIX A

Table A. 1 Primers used in validation experiments.

Primer Name	**Tm (ºC)**	**Sequence (5' -> 3')**
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
POLR3A_I13_F	59.8	CCC ACC GCC TAC AAT CCT AAA
POLR3A_I13_R	59.4	AGT TCC GTC CAC TCA CAG GA
ADCK3_E8_F	59.8	TCT TGC TCT CCT AAT GCT GGC
ADCK3_E8_R	61.4	CAG CAC CCA GGA AAA GTC CC
SPG11_E6_F	51.2	AGT GTT TTA CAT AGT GCC TT
SPG11_E6_R	53.2	CAT ACA TCT CAG CAA TGG AT
ATM_E53_F	57.3	ATA GGA TCG AAC AGA GGC TG
ATM_E53_R	57.9	AGT AAC CAG GGA ATG CTG AAG
C19orf12_E2_F	56.4	GGA AAT ACT CTT ATG CTC ATT GAA A
C19orf12_E2_R	52.4	GTT TCA ACG GCC CTT TTA T
OPA3_E2_F	61.8	TCT CTT CCC CCG CAG TGT ATC
OPA3_E2_R	59.8	CTA TTT CTT GGA CGC AGG CAC
ATP13A2_E25_F	62.4	CCC AGC TGT CAT CAT ATT CTG CC
ATP13A2_E25_R	64.2	CCC ACG TCA TCT ATT CTG GGA CC
GAN_E9_F	59.8	GAA GGC CCA CGT AGT AAT GCT
GAN_E9_R	57.9	CTA GCA CTT GTG AAG TCA CTG
SETX_E19_F	57.3	GGC CTG TCA TAG TCA AGG AT
SETX_E19_R	57.9	TCC ACT TTC ACT ATG GAG GGT
APTX_E8_F	57.3	GCC TGA GGG AAG TCT AAA CT
APTX_E8_R	53.2	TTC ACA AGC AAC CCA GAA TA
ATCAY_E7_F	57.3	TGT TAA GCC GTC AAC TCG CT
ATCAY_E7_R	59.4	GGG CCA CAA TGC AAT CCT TG
PTRH2_E3_F	59.4	CTG TTG GAG TTG CTT GTG GC
PTRH2_E3_R	59.4	GAG CCT GGT GCA ATC TGA GT
SACS_E9_F	59.8	TGG ATT CAG CCA GAA GGA ACC
SACS_E9_R	57.3	AGT GCA AGT CAG AGC CAT CA
CAPN1_E4_F	61.4	GAG GGG AGT TAA GTG GCA CC

CAPN1_E4_R	61.4	CGA GAT GGC ACT GAG AGG TC
ATM_E2_F	57.3	CTT TGA CCA GAA TGT GCC TC
ATM_E2_R	61.4	CAG GAT CTC GAA TCA GGC GC
SPG7_E14_F	59.4	CCT TGT GCC AGG TCT CCA TA
SPG7_E14_R	59.4	CAG AAG GAG TCA TGC AGG GA
SPG7_E13_F	57.3	TGC AGA GAG AGG CTT GCT TT
SPG7_E13_R	59.4	CAT GAC AGT GGC AGG CTT TC
APTX_E6_F	54.0	TGG GAA TTA AGT GAC TTA GTG
APTX_E6_R	57.3	GGG TCT CAG TGC AAT ATG TG
SACS_E10_2F	56.5	CCT GCT GAA ATT CAT TAC ACT C
SACS_E10_2R	55.3	CCA TTT CAA ATA CAA CCG CC
SQSTM1_E6_F	59.4	TCT GTA GTC TCC ACA GGC CA
SQSTM1_E6_R	61.4	CTG CAG AGG TGC TGA GGA TG
MRE11_E13_F	57.3	AAT GTG CAG CTC TCA CTG CT
MRE11_E13_R	61.3	CTC AGA GCT ACT CTT AAA GAC AGA C
SACS_p.Tyr1256Cys_F	59.4	TTA CAG GTC GGT GCT TGT CC
SACS_p.Tyr1256Cys_R	57.3	TGG TTT AAT CAC AGC CTG GG
SACS_p.Asp2578del_F	59.4	GTC CCA AAG CGA CAC AAA GC
SACS_p.Asp2578del_R	59.4	TCT CTA AAC ATG CGT CCG GG
TDP1_E3_F	61.8	GTT AGG TGG TTC AGC CTC TGG
TDP1_E3_R	59.8	AGC TCA TCA TCA CTG CTG GAC
SACS_E10_24F	58.5	GAT GAA GAA ATG GTA AAA ACT AGA GC
SACS_E10_24R	54.7	AAC ATT TGC ACA CCA ATA TTC C
SCN9A_E27_F	57.3	TGT CCC TTC CTG CGT TGT TT
SCN9A_E27_R	55.3	CTC AGA GGT TCA GTA CTT TC
SETX_E20_F	57.3	GGT GCA GGA GAA GTG TGA TA
SETX_E20_R	59.4	GCA CGT CAT CCA TCT AGG CT
TTPA_I3_F	57.3	CTG TAA GCC AGG CTT TTG TG
TTPA_I3_R	59.4	GGG CAG GTA CGA GGA AAG AT
CYP27A1_I8_F	59.4	GAG AAA CAG CCA GCC TGC TA
CYP27A1_I8_R	57.3	TGT GGT TGG GTG GAT TGT GT
SPG7_E15_F	61.4	GCT GAG GAT GCC TCT GTC TC
SPG7_E15_R	61.4	CTA CAG CAG AGC CCA GAA GG

SPG7_E13_F	58.8	CTC GAA CTC CTG TCC TCA G
SPG7_E13_R	59.4	CCC AGT CAG CTA CAG ACA CA
ANO10_E6_F	57.3	GCC AGG ATT TCA TGG TGT CT
ANO10_E6_R	55.3	ATG ATG CTG GGC ACA TAC AA
COX20_E2_F	58.4	GAT GTT GAA AAT ACT CCC TGC G
COX20_E2_R	58.4	CTT AGC AAA CAT ACC AGC ATC C
SACS_E10_F	57.3	GCC AGC ACT TTG TGT GTA CA
SACS_E10_R	59.4	CAT CTG AGC GCA GTT TGT CC
SACS_E10_F_	57.3	GTT GAA GTG TTG AGC CGC AT
SACS_E10_R_	57.3	TGA AGA GAA CAC AAC GCT CC
BSCL2_E8_F	59.4	TCC TGC CTC AGA TAC CTG CT
BSCL2_E8_R	61.4	CAA CAT ACC CCT GAC CAC CC
CCDC88C_E24_F	59.4	GCA TGC CAC CAT TCC CGT AT
CCDC88C_E24_R	59.4	CTC TGA TGT GGG AGG CAT CT
GLA_E5_F	57.9	TCA AGA GAA GGC TAC AAG TGC
GLA_E5_R	59.8	ACT TTC CAC AGC ATC CTG CTC
NIPA1_E5_F	59.4	TCA GCA TCT GCT CCT TGC TG
NIPA1_E5_R	61.4	GCT ATC CAC CAG ACC ACC TG
STUB1_E1_F	57.3	ATG AAG GGC AAG GAG GAG AA
STUB1_E1_R	58.8	AGA ACC ACG GCT GGG CTT T
KCNC3_E2_F	59.4	TAG CAA CAA GAC GGT GAC CC
KCNC3_E2_R	59.4	AAG TAG GTG TGG TTG GAG CC
ATP13A2_E4_E5_F	60.4	CTG GAC CTC AGC TTT CCC ATC
ATP13A2_E4_E5_R	59.2	GAG GGG CAA GGA CTT TTT CAG

APPENDIX B

Table B. 1 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

APPENDIX C

Table C. 1 Open-source databases, bioinformatics software and tools.

Database/Software/Tool	**Definition**
CADD (Rentzsch, Schubach, Shendure, & Kircher, 2021)	A tool that assesses the deleteriousness insertion/deletion as well as single-nucleotide variants in the human genome
CLC Main Workbench (Qiagen, Germany)	A platform for DNA, RNA, and protein sequence analyses
ClinVar (Landrum et al., 2020)	An archive of the connections among human phenotypes and variations, with supporting evidence
DANN (Quang, Chen, & Xie, 2015)	Functional prediction score ranging is 0 to 1, 1 being the most destructive
dbSNP (Sherry et al., 2001)	A database that contains SNPs, small-scale insertions/deletions, and microsatellites
FinchTV	An application that displays data from Sanger DNA sequencing
Franklin by Genoox	A software for variant interpretation based on community
GeneCards	An integrated human gene database that provides information on all known and predicted human genes.
GERP++ (Davydov et al.)	Conservation score estimated by quantifying substitution deficits
GnomAD	A database that includes data from 76,156 genomes of unrelated individuals
Integrative Genomics Viewer (IGV)	A tool that provides the visual exploration of various genomic data
MetalR	A prediction tool for missense variants, which integrates nine independent variant scores
MutationTaster	A prediction tool which depends on mRNA, protein, evolutionary conservation, splice-site, and regulatory points for the pathogenicity of a variant

NCBI Primer Blast (Ye et al., 2012)	A tool to search primers which are specific to the PCR template
NetPrimer	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 database that contains human genes and diseases
REVEL (Ioannidis et al., 2016)	A prediction tool which depends on a combination of scores from 13 individual tools for pathogenicity of missense variants
RFFlow	A software that draws pedigrees, flowcharts, and other diagrams
SEQ Platform by Genomize	A platform on storage, sharing, and usage of genomic data
SIFT (P. C. Ng & Henikoff, 2003)	A tool that predicts whether amino acid substitutions affect protein function
UCSC database (Kent et al., 2002)	A software of University of California Santa Cruz
Varsome (Kopanos et al., 2019)	A search engine with data collector and impact analysis tool for human genomics

APPENDIX D

The papers published and prepared throughout this thesis:

1. Ser, M. H., **Tekgül, Ş.** et al. Ataxia telangiectasia like disorder: Another dopa-responsive disorder look-alike?. Parkinsonism & Related Disorders (2020). https://doi.org/10.1016/j.parkreldis.2020.03.016
2. Şenel, B., G., Tezen, D., **Tekgül, Ş.** et al. Co-existence of Mutations in PRRT2 and ABCC6 Genes in a Turkish Family. Türkiye Klinikleri Journal of Case Reports (2020). doi:10.5336/caserep.2020-75076
3. Sahin, T., Karaarslan T., F., Yılmaz, R., **Tekgül, Ş.** et al. Two cases of early-onset autosomal recessive spastic ataxia of Charlevoix-Saguenay diagnosed in adulthood. Clinical Neurology and Neurosurgery (2020). doi: 10.1016/j.clineuro.2020.106423.s
4. 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
5. Çakar, A., İnci, M., Özdağ-Acarlı, A.N., Çomu, S., Candayan, A., Battaloğlu, E., **Tekgül, Ş.** et al. Phenotypical spectrum of SACS variants: Neuromuscular perspective of a complex neurodegenerative disorder. Acta Neurologica Scandinavica (accepted in January 2022).
6. Şenel-Benbir, G., Abbaszade, H., **Tekgül, Ş.** et al. A case of cerebrotendinous xanthomatosis diagnosed at adulthood (submitted to Turkish J Neurology)
7. Öztop-Çakmak,Ö., Şimşir, G., **Tekgül, Ş.** et al. VPS13D-based disease in the Turkish population: Expansion of the clinical phenotype and mutation diversity (submitted to Revue Neurologique)