

THE MOLECULAR DISSECTION OF ATAXIAS IN TURKEY AND THE IMPACT OF
WHOLE EXOME SEQUENCING IN THEIR PRECISE DIFFERENTIAL DIAGNOSIS

by

Cemile Koçođlu

B.S., Molecular Biology and Genetics, Istanbul Technical University, 2013

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

Graduate Program in Molecular Biology and Genetics
Bođaziçi University
2016

To being ambitious...

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Prof. A. Nazlı Başak for her support, guidance, and encouragement during this thesis. I am very grateful for her valuable criticism and the opportunities she has offered me during my entire master study.

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

I am grateful to Prof. Jan H. Veldink for his hospitality and guidance during my stay at UMC Utrecht. I also appreciate Kristel Kool van Eijk and Sara Pulit for their valuable guidance in data analysis and sharing their scientific knowledge. Their support helped me to grow myself up as a data scientist.

I would like to thank all family members for their cooperation and valuable clinicians throughout the Turkey who provided the blood samples. At this point, I would like to acknowledge Suna and İnan Kırarç Foundation for their generous financial support throughout the study that renders this research possible.

I deeply thank all members of NDAL and the bioinformatics team for making the lab a great place to come and study. I am grateful every single one of them.

I am very grateful to all my friends for their sincere friendship and support. At this point, I would like to express my special thanks to my beloved Eray for his patience, love and for always being there for me.

Last but not least, I am heartily grateful to my devoted parents, Gönül and Mustafa Koçoğlu, and little brother, Ahmet Koçoğlu, for their endless emotional support. They have been always encouraging and supportive in every decision I had. Nothing would have been possible without them.

ABSTRACT

THE MOLECULAR DISSECTION OF ATAXIAS IN TURKEY AND THE IMPACT OF WHOLE EXOME SEQUENCING IN THEIR PRECISE DIFFERENTIAL DIAGNOSIS

Ataxias are a diverse group of genetically, clinically and mechanistically heterogeneous neurodegenerative disorders. There is significant overlap in the phenotype of different subtypes of ataxia, which can also manifest as additional symptoms with other neurological diseases; thus making a precise clinical diagnosis becomes challenging. After the elimination of common late-onset spinocerebellar ataxias, still a high number of ataxia patients remain genetically undiagnosed.

In recent years, next generation sequencing (NGS) technologies have proven to be very powerful to detect the genetic causes of both monogenic and complex disorders. Among these, whole exome sequencing is especially effective in the identification of disease genes, particularly in pedigrees manifesting a Mendelian inheritance pattern. In this study, 47 pedigrees were investigated by whole exome sequencing to dissect the cause of disease. This pipeline was coupled with homozygosity mapping in recessive cases to detect homozygous stretches likely to contain mutated disease genes.

In the framework of this thesis, we identified and validated the genetic causes of ataxia in almost half (43%) of the pedigrees investigated. This diagnostic yield, in accordance with the literature, is a demonstration of the usefulness of the approach in definitive diagnosis. Detailed phenotyping of the remaining individuals in our pedigrees, combined with the adaptation of rapidly evolving approaches to analyze the data, are expected to pave the ways to identification of novel disease genes in the yet unsolved families.

ÖZET

TÜRKİYE'DEKİ ATAKSİLERİN MOLEKÜLER İNCELENMESİ VE TÜM EKZOM DİZİLEMENİN KESİN AYIRICI TANIDAKİ ÖNEMİ

Ataksiler klinik, genetik ve mekanistik açıdan heterojen bir nörodejeneratif hastalık grubudur. Farklı ataksi tipleri arasındaki klinik örtüşmelerin yanında, ataksinin benzer nörolojik hastalıklarda ek semptom olarak görülmesi de kesin tanıyı zor hale getirmektedir. Sık görülen ataksi tiplerinin dışlanması ardından genetik tanı almamış göz ardı edilemeyecek miktarda ataksi hastası mevcuttur.

Son yıllarda yeni nesil dizileme teknolojileri hem monogenik, hem de kompleks kalıtım gösteren hastalıkların genetik nedeninin tanımlanmasında kendini ispatlamış sağlayan güçlü bir araçtır. Farklı uygulamalar arasında ekzom dizileme yöntemi, özellikle Mendel türü kalıtım gösteren hastalıkların genetik nedenlerinin bulunmasında etkili ve kullanışlı olmuştur. Bu tez kapsamında, 47 ataksi ailesi, hastalık nedenini anlama amacıyla ekzom dizilemeye tabi tutuldu. Resesif kalıtım gösteren ailelerde, mutasyonu barındırması muhtemel olan ortak homozigot segmentleri saptamak için homozigotluk haritalama yöntemi kullanıldı.

Bu çalışma çerçevesinde incelenen ailelerin yarısına yakınında (%43) ataksi fenotipine neden olan gen / mutasyon bulundu ve doğrulandı. Bu oran literatürle uyumludur ve ekzom dizileme yönteminin ayırıcı tanındaki önemini göstermektedir. Sonuç alamadığımız ailelerde daha detaylı fenotip bilgisine ulaşılması ve hızla evrimleşen veri analiz yöntemlerinin benimsenmesi yeni hastalık genleri tanımlamanın önünü açacaktır.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
ÖZET	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	x
LIST OF TABLES.....	xiii
LIST OF SYMBOLS	xv
LIST OF ACRONYMS/ABBREVIATIONS.....	xvi
1. INTRODUCTION.....	1
1.1. Ataxias	1
1.1.1. Autosomal Dominant Cerebellar Ataxias.....	2
1.1.2. Autosomal Recessive Cerebellar Ataxias.....	4
1.1.2.1. Autosomal recessive cerebellar ataxia type 1.....	5
1.1.2.2. Autosomal recessive spastic ataxia of Charlevoix-Saguenay.....	6
1.1.2.3. Ataxias with oculomotor apraxia.....	6
1.2. Differential diagnosis of ataxias with diseases having ataxic features	9
1.3. Next Generation Sequencing Technologies	9
1.3.1. Whole Exome Sequencing.....	10
1.3.1.1. Study design for exome sequencing.....	11
1.3.1.2. Homozygosity mapping.....	13
1.3.1.3. Linkage analysis.....	13
1.3.1.4. Limitations.....	14
1.3.2. Molecular analysis of neurodegenerative disorders with NGS	15
2. PURPOSE.....	16
3. MATERIALS	17
3.1. Subjects	17
3.1.1 Family trees.....	24

3.1.1.1. Families with an autosomal dominant inheritance pattern.....	24
3.1.1.2. Families with an autosomal recessive inheritance pattern.....	25
3.2. Equipment and Solutions used in DNA Isolation	30
3.3. Whole Exome Sequencing Platforms and Exome Enrichment Kits	30
3.4. Equipment and Solutions used in Validation Experiments.....	30
3.5. Equipment and Solutions for Agarose Gel Electrophoresis and Extraction from the Gel	32
3.6. General Laboratory Equipment and Kits	33
3.7. Software, Online Databases and Bioinformatics Tools	35
4. METHODS.....	37
4.1. DNA Isolation from peripheral blood samples	37
4.2. Whole Exome Sequencing	37
4.3. Whole Exome Sequencing Data Analysis.....	39
4.3.1. Alignment and Variant Calling.....	39
4.3.2. Homozygosity Mapping	39
4.3.3. Confirmation of relationships and principal component analysis	40
4.3.4. Variant Annotation and Prioritization.....	40
4.4. PCR to validate exome analysis results	41
4.4.1. PCR conditions	41
4.4.2. Agarose gel electrophoresis	41
5. RESULTS.....	42
5.1. Sequencing analysis metrics	42
5.2. Population structure	44
5.3. Whole exome sequencing data analysis.....	44
5.3.1. Families with an autosomal dominant inheritance pattern	46
5.3.1.1. Spinocerebellar ataxia type 28: AFG3L2.....	46
5.3.1.2. Charcot-Marie-Tooth disease type 2A: KIF1B.....	47
5.3.2. Families with autosomal recessive inheritance pattern	47
5.3.2.1. Autosomal recessive cerebellar ataxia type 1: SYNE1.....	48

5.3.2.2. Autosomal recessive spastic ataxia of Charlevoix-Saguenay: SACS.....	49
5.3.2.3. Ataxia with oculomotor apraxia type 1: APTX.....	51
5.3.2.4. Ataxia with oculomotor apraxia type 2: SETX.....	52
5.3.2.5. Hereditary spastic paraplegia type 76: CAPN1.....	56
5.3.2.6. Autosomal recessive hereditary spastic paraplegia with thin corpus callosum: SPG11.....	59
5.3.2.7. Autosomal recessive hereditary spastic paraplegia type 7: SPG7.....	60
5.3.2.8. Autosomal recessive hereditary spastic ataxia type 2: KIF1C.....	62
5.3.2.9. Autosomal recessive cerebellar ataxia type 2: ADCK3.....	63
5.3.2.10. Ataxia with vitamin E deficiency: TTPA.....	65
5.3.3. Families without confirmed pathogenic mutations.....	69
6. DISCUSSION.....	72
6.1. Whole exome sequencing and genetic landscape of ataxias.....	72
6.1.1. Autosomal dominant families.....	72
6.1.2. Autosomal recessive families.....	74
6.1.3. Advantages of whole exome sequencing in differential diagnosis of ataxias.....	77
6.1.4. Novel genotype - phenotype associations.....	78
6.2. Challenges and limitations of clinical exome sequencing: Why did we fail in finding the genes in approximately half of our families?.....	79
6.3. Path to discovery of novel disease genes.....	81
7. CONCLUSION.....	82
REFERENCES.....	83
APPENDIX A: COMMANDS EXECUTED IN ANALYSIS OF WHOLE EXOME SEQUENCING DATA.....	102
APPENDIX B: SEQUENCING QUALITY METRICS.....	105

LIST OF FIGURES

Figure 1.1. Strategies for analysis of whole exome sequencing data.....	11
Figure 3.1. Family 1 and Family 2 showing an autosomal dominant inheritance pattern with several affected individuals.....	24
Figure 3.2. Family 3, Family 4, Family 5, Family 6.....	25
Figure 3.3. Family 7, Family 8, Family 9, Family 10.....	26
Figure 3.4. Family 11, Family 12, Family 13.....	27
Figure 3.5. Family 14, Family 15, Family 16, Family 17.....	28
Figure 3.6. Family 18, Family 19, Family 20.....	29
Figure 4.1. Experimental steps of whole exome sequencing.....	38
Figure 5.1. Mean depth of coverage per sample.....	42
Figure 5.2. Rate of missingness per sample.....	43
Figure 5.3. Transition / Transversion ratios per sample.....	44
Figure 5.4. Multidimensional scaling plot of the participants.....	45
Figure 5.5. Circos representation of families and the mutated genes.....	45
Figure 5.6. Family 1: segregation of the missense variant in thr AFG3L2 gene.....	46
Figure 5.7. Family 2 with the segregation status of the KIF1B heterozygous variant.....	47

Figure 5.8. Homozygosity mapping plot and the segregation of the SYNE1 variant in the pedigree.....	48
Figure 5.9. Runs of homozygosity in affected siblings of Family 4, family tree with the segregation of the SACS variant.....	49
Figure 5.10. Homozygosity mapping plot of shared regions between three affected siblings and the segregation of the SACS variant within the pedigree.....	50
Figure 5.11. Runs of homozygosity plot and the segregation of the missense APTX variant in Family 9.....	51
Figure 5.12. Homozygosity mapping plot of juvenile patient and the segregation of the SETX variant in Family 7.....	52
Figure 5.13. Homozygosity mapping in affected individuals and the pedigree of Family 8 with the segregation of the SETX variant in available individuals.....	53
Figure 5.14. Runs of homozygosity in Family 9, the family tree with the segregation of the SETX variant.....	54
Figure 5.15. Homozygosity mapping plot of shared regions in affected siblings and the segregation of the SETX variant in Family 10.....	55
Figure 5.16. Homozygosity mapping plot of the proband of Family 11, the pedigree of P25.....	56
Figure 5.17. Runs of homozygosity in the index case and the segregation of the missense CAPN1 variant in Family 12.....	57
Figure 5.18. Homozygosity mapping plot of the affected siblings of Family 13 and the segregation of the CAPN1 variant in available individuals.....	58

Figure 5.19. Runs of homozygosity and pedigree of Family 14 with segregation of the SPG11 variant.....	59
Figure 5.20. Homozygosity mapping of the index case and the segregation of the SPG7 variant in available individuals.....	60
Figure 5.21. Homozygosity mapping of P33 and the pedigree of Family 16.....	61
Figure 5.22. Runs of homozygosity in affected cousins and pedigree of Family 17 with the segregation of the missense KIF1C variant.....	62
Figure 5.23. Homozygosity mapping plot of the index case and the segregation of the ADCK3 variant in Family 17.....	63
Figure 5.24. Distribution of runs of homozygosity of the index case and the segregation of the ADCK3 variant in Family 18.....	64
Figure 5.25. Runs of homozygosity of the index case and pedigree of patient P40.....	65
Figure 6.1. Lollipop diagram of the AFG3L2 protein.....	73
Figure 6.2. KIF1B gene structure and mutations.....	73
Figure 6.3. Mutations described in the CAPN1 gene.....	75
Figure 6.4. Structure of the paraplegin protein and mutations identified in this study.....	76

LIST OF TABLES

Table 1.1. Autosomal dominant ataxias.....	3
Table 1.2. Examples of autosomal recessive ataxias.....	4
Table 1.3. Examples of neurodegenerative disease genes identified by whole exome sequencing analysis.....	15
Table 3.1. Clinical characteristics of families investigated in this study.....	18
Table 3.2. Exome enrichment kits.....	30
Table 3.3. Sequences of primers used in validation experiments.....	31
Table 3.4. Chemicals used in gel electrophoresis.....	32
Table 3.5. General laboratory equipment.....	33
Table 3.6. Commercially available kits, other than exome capture kits.....	34
Table 3.7. Open-source bioinformatics software, tools and electronic databases.....	35
Table 4.1. Homozygosity mapping parameters of PLINK.....	40
Table 5.1. Total number of variants per family after each filtering step.....	66
Table 5.2. List of all confirmed pathogenic variants and associated diseases in the OMIM database.....	67
Table 5.3. The quality scores and minor allele frequencies of the confirmed pathogenic variants.....	68

Table 5.4. Online prediction tool scores of the confirmed pathogenic variants.....69

Table 5.5. The inheritance pattern and number of variants in families without
confirmed pathogenic variants.....70

LIST OF SYMBOLS

g	Gram
kb	Kilobase
M	Molar
Mb	Million bases
mg	Miligram
ml	Mililiter
mM	Milimolar
ng	Nanogram
V	Volt
μ l	Microliter
μ M	Micromolar
*	Asterisk
%	Percentage
$^{\circ}$ C	Centigrade degree
#	Number

LIST OF ACRONYMS/ABBREVIATIONS

AD	Autosomal Dominant
ADCA	Autosomal Dominant Cerebellar Ataxia
ADCK3	AarF Domain Containing Kinase 3
ALS	Amyotrophic Lateral Sclerosis
AFG3L2	ATPase Family Gene 3-Like 2
AFP	Alpha Feto Protein
ao	Age of Onset
AOA	Autosomal Recessive Ataxia with Oculomotor Apraxia
APTX	Aprataxin
AR	Autosomal Recessive
ARCA	Autosomal Recessive Cerebellar Ataxias
ARSACS	Autosomal Recessive Spastic Ataxia Of Charlevoix Saguenay
A-T	Ataxia–Telangiectasia
A-TLD	Ataxia-Telangiectasia like Disorder
ATN	Athrophin
ATM	Ataxia-Telangiectasia Mutated
ATP	Adenosine TriPhosphate
ATXN	Ataxin
AVED	Ataxia with vitamin E deficiency
BEAN-TK2	Brain-Expressed Associated with NEDD4-Thymidine Kinase
BWA	Burrows-Wheeler Aligner
CACNA1A	Calcim Channel Voltage-Dependent P/Q type Alpha-1A Subunit
CADD	Combined Annotation Dependent Depletion
CAPN1	Calpain-1
ChIP-seq	Chromatin Immunoprecipitation-sequencing
CLIP-seq	Cross-linking Immunoprecipitation-sequencing
CMT	Charcot-Marie-Tooth
Chr	Chromosome
CNS	Central Nervous System
CNV	Copy Number Variation

DNA	Deoxyribonucleic Acid
dNTP	Deoxyribonucleotide Triphosphate
DRPLA	Dentatorubral-pallidoluysian Atrophy
DSB	Double-strand Break
DSBR	Double-strand Break Repair
DTR	Deep Tendon Reflexes
EDTA	Ethylenediaminetetraacetic Acid
ESP	Exome Sequencing Project
ExAC	Exome Aggregation Consortium
F	Female
FA	Friedreich's ataxia
FGF14	Fibroblast Growth Factor 14
FXN	Frataxin
GATK	Genome Analysis Toolkit
GERP	Genomic Evolutionary Rate Profiling
HGP	Human Genome Project
HIT	Histidine-triad
Hsp	Heat Shock Protein
HSP	Hereditary Spastic Paraplegia
IBS	Identity-by-Segment
IGV	Integrative Genomics Viewer
INAREK	Ethics Committee on Research with Human Participants
ITPR1	Inositol 1,4,5-triphosphate receptor type 1
KCNC3	Potassium Voltage-Gated Channel Subfamily C Member 3
KIF1B	Kinesin Family Member 1B
KIF1C	Kinesin Family Member 1C
LRT	Likelihood Ratio Test
M	Male
MAF	Minor allele Frequency
MIM	Mendelian Inheritance in Man
MRE11	Meiotic Recombination 11 Homolog A
MRN	Mre11-Rad50-Nbs1
MRI	Magnetic Resonance Imaging

mRNA	Messenger Ribonucleic Acid
MTTP	Microsomal Triglyceride Transfer Protein
NA	Not Available
ND	Neurodegenerative Diseases
NGS	Next Generation Sequencing
NOP56	Nucleolar Protein 5A
NPC1	Niemann-Pick Disease, Type C1
OMIM	Online Mendelian Inheritance in Man
P	Patient
PCR	Polymerase Chain Reaction
PEX7	Peroxisomal Biogenesis Factor 7
PHYH	Phytanoyl-CoA 2-Hydroxylase
PNKP	Polynucleotide Kinase 3'-Phosphatase
PolyPhen2	Polymorphism Phenotyping v2
PPP2R2B	Protein Phosphatase 2 Regulatory Subunit Beta
PRKCG	Protein Kinase C Gamma
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
ROH	Runs of homozygosity
RTN2	Reticulon 2
SACS	Sacsin
SCA	Spinocerebellar Ataxia
SETX	Senataxin
SIFT	Sorts Intolerant from Tolerant
SMA	Spinal Muscular Atrophy
SBMA	Spinal Bulbar Muscular Atrophy
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
SPG7	Paraplegin
SPG11	Spatacsin
SPTBN2	Spectrin Beta Non-Erythrocytic
SRR	Sacsin Repeat Region
SSBR	Single Strand Break Repair

SYNE1	Spectrin Repeat Containing, Nuclear Envelope 1
SYT14	Synaptotagmin 14
TBE	Tris/Borate/EDTA
TBK1	Tank-binding Kinase 1
TBP	TATA-box Binding protein
TCC	Thin Corpus Callosum
TGM6	Transglutaminase 6
TM	Melting Temperature
Ts	Transition
TTBK2	Tau Tubulin Kinase 2
TTPA	Tocopherol (Alpha) Transfer Protein
Tv	Transversion
UbL	Ubiquitin-like
VPS35	Vacuolar Protein Sorting 35
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

1. INTRODUCTION

Neurodegenerative disorders (NDs) are a group of neurological diseases distinguished by loss or malfunctioning of a specific regional group of neurons. They are commonly seen in the population above 65 years of age; however, early-onset forms are also present and especially observed in inbred populations. The death of specific neuronal types results in movement disorders and / or cognitive dysfunction and various neuropathological conditions. According to the affected region of the central nervous system (CNS), the NDs can be divided into four general groups: diseases of the (i) cerebral cortex, (ii) basal ganglia, (iii) spinal cord or the brainstem and (iv) cerebellum (Przedborski *et al.*, 2003). The diseases of the cerebral cortex are associated with dementia such as Alzheimer's disease, while diseases affecting the basal ganglia are characterized by movement abnormalities like Parkinson's disease. Amyotrophic lateral sclerosis (ALS) and spinal muscular atrophy (SMA) constitute the group where the spinal cord is affected. Diseases of the cerebellum, the topic of this thesis, are complex due to clinical overlaps with disorders in which connections in the cerebellum are damaged. All cerebellar diseases present ataxia in addition to various other neuropathological conditions. Moreover, complicated cases, where more than one region of the CNS is affected, can also occur depending on the genetic background of the individual or the environmental factors blurring the border between different types of disorders.

1.1. Ataxias

Ataxias are a group of neurodegenerative / neuromuscular disorders characterized by loss of control over muscle movements resulting in impairment of voluntary movement, speech, swallowing and eye movements. These symptoms are manifested upon damage to the cerebellum, located under the occipital and the temporal lobes of the cerebral cortex. The cerebellum is responsible for control of motor movements by taking input from sensory systems and spinal cord. Malfunctioning of this system causes incoordination of motor movements and appearance of ataxia with different subtypes according to the damaged component within the signal process system of the cerebellum. Ataxias can be acquired, sporadic or inherited according to the development of the disease. Acquired ataxias may

develop upon a damage to the cerebellum due to infections, surgery, brain tumor, vitamin deficiencies, congenital abnormalities and exposure to toxins. On the other hand, sporadic ataxias can be defined as adult-onset ataxias without an apparent family history. A significant number of ataxia cases, constituted by inherited ataxias, can be seen in autosomal dominant (AD), autosomal recessive (AR) and X-linked manner. Each group has several defined subtypes where usually a gene is uniquely linked to one specific subtype. Although the underlying genetic cause differs, all types of ataxias have overlapping symptoms, making the clinical diagnosis challenging to the clinicians.

1.1.1. Autosomal Dominant Cerebellar Ataxias

Autosomal dominant cerebellar ataxias (ADCAs) or spinocerebellar ataxias (SCAs) are progressive neurodegenerative disorders mainly manifesting after the age of 35. The estimated prevalence is 3/100,000 (Schöls *et al.*, 2004) with a possibility of being much higher in isolated populations due to a founder effect. SCAs are characterized by atrophy of the cerebellum and brainstem resulting in ataxia of gait and dysarthria, which can be associated with pyramidal or extrapyramidal signs or cognitive impairment (Durr, 2010).

The initial classification of SCAs for clinical diagnosis, introduced by Harding, is composed of three main groups. The first group (ADCA I) includes ataxias with degeneration of other neuronal systems; the second group (ADCA II) shows a similar phenotype and additional retinal degeneration while the third group (ADCA III) reflects pure cerebellar ataxia (Harding, 1982). However, currently the nomenclature has changed and the classification is done according to the order of loci identification (Table 1.1).

SCAs are commonly caused by repeat expansions while conventional point mutations are also observed at genetic level. However, there is still a number of SCA subtypes of which only the suspected loci are identified (Durr, 2010). The most common forms of SCAs are developed as a result of exonic CAG trinucleotide expansions above a specific threshold, belonging to a large group of neurodegenerative disorders called 'polyglutamine disorders'. In this group, the repeat size correlates with the disease severity, being also in reverse correlation with the age at onset.

Table 1.1. Autosomal Dominant Ataxias, adapted from Durr, 2010; Jayadev and Bird, 2013.

Disease	Gene	Average onset	Key clinical findings
<i>Polyglutamine expansions</i>			
SCA1	ATXN1	3 rd -4 th decade	pyramidal signs, peripheral neuropathy
SCA2	ATXN2	3 rd -4 th decade	slowed eye movements, decreased dtr, dementia
SCA3	ATXN3	4 th decade	pyramidal and extrapyramidal signs, fasciculations, sensory loss
SCA6	CACNA1A	5 th -6 th decade	slow progression
SCA7	ATXN7	3 rd -4 th decade	visual loss
SCA17	TBP	4 th decade	mental retardation, chorea, dystonia, myoclonus, epilepsy
DRPLA	ATN1	52.8 years	chorea, seizures, dementia, myoclonus
<i>Noncoding expansions</i>			
SCA8	ATXN8	4 th decade	slow progression, decreased vibration
SCA10	ATXN10	4 th decade	seizures
SCA12	PPP2R2B	4 th decade	action tremor, hyperreflexia, dementia
SCA31	BEAN-TK2	5 th -6 th decade	normal sensation
SCA36	NOP56	53 years*	fasciculation, tongue atrophy, hyperreflexia
<i>Conventional mutations</i>			
SCA5	SPTBN2	3 rd -4 th decade	early onset, slow progression
SCA11	TTBK2	30 years	mild, remain ambulatory
SCA13	KCNC3	childhood	mild intellectual disability, short stature
SCA14	PRKCG	3 rd -4 th decade	early axial myoclonus
SCA15/16	ITPR1	4 th decade	pure ataxia / head tremor
SCA27	FGF14	11 years	early onset tremor, dyskinesia, cognitive decline
SCA28	AFG3L2	19.5 years	nystagmus, ptosis, increased tendon reflexes, ophthalmoparesis
SCA35	TGM6	43.7 years	hyperreflexia

*: These subtypes are very rare, thus the average ages of onset are shown, instead of decades.

1.1.2. Autosomal Recessive Cerebellar Ataxias

Autosomal recessive cerebellar ataxias (ARCAs) comprise a large and heterogeneous group of rare neurodegenerative disorders where a cerebellar syndrome is the key symptom. In contrast to ADCAs, the age at onset is early, usually before the third decade of life and highly variable among clinical subtypes. Patients are characterized by progressive cerebellar ataxia often accompanied by impairment of gait and balance, cerebellar dysarthria, cognitive deficits and oculomotor abnormalities (Anheim, 2012). The overlap in subtypes of disease phenotype makes the accurate clinical diagnosis challenging and implies the importance of clear definition of the genetic etiology. The overall worldwide prevalence is estimated to be around 7 per 100,000 with alterations between populations (Vermeer *et al.*, 2011).

ARCAs can be classified by their underlying mechanism of pathogenesis. The identified genes are shown to take role in mitochondrial dysfunction, altered chaperone activity, DNA repair defects and metabolic functioning (Vermeer *et al.*, 2011). Although each causative gene explains a specific subtype, several genes play role in the same mechanism, resulting in overlapping symptoms, (Table 1.2).

Table 1.2. Examples of autosomal recessive ataxias, adapted from Hershenson, Haworth, and Houlden, 2012; Jayadev and Bird, 2013.

Disease	Gene	Average onset	Key clinical findings
Friedreich's ataxia	FXN	1 st -2 nd decade	neuropathy, cardiac involvement
Autosomal recessive ataxia of Charlevoix-Saguenay	SACS	childhood	lower limb spasticity
Autosomal recessive cerebellar ataxia type 1	SYNE1	2 nd -3 rd decade	pure cerebellar ataxia
Ataxia-telangiectasia	ATM	1 st decade	telangiectasia, oculomotor apraxia, increased cancer risk
Ataxia-telangiectasia-like disorder	MRE11	1 st decade	milder phenotype of A-T, no telangiectasia

Table 1.2. Examples of autosomal recessive ataxias, adapted from Hersheson, Haworth, and Houlden, 2012; Jayadev and Bird, 2013 (cont.).

Disease	Gene	Average onset	Key clinical findings
Ataxia with oculomotor apraxia type 1	APTX	childhood	oculomotor apraxia, peripheral neuropathy
Ataxia with oculomotor apraxia type 2	SETX	2 nd decade	peripheral neuropathy, elevated AFP levels
Abetalipoproteinemia	MTTP	infancy	polyneuropathy, primary retinal degeneration
Refsum disease	PHYH, PEX7	2 nd decade	retinitis pigmentosa, demyelinating neuropathy
Nieman-Pick Type C disease	NPC1	1 st -2 nd decade	extrapyramidal features, dementia, seizures

1.1.2.1. Autosomal recessive cerebellar ataxia type 1. In 2007, an adult onset form of recessive ataxia was described in the French-Canadian population and named as autosomal recessive cerebellar ataxia type I (ARCA-1; MIM#610743) or recessive ataxia of Beauce. The disease is characterized by a relatively pure cerebellar ataxia with gait ataxia, dysarthria, dysmetria, mild oculomotor abnormalities and presence of cerebellar atrophy (Gros-Louis *et al.*, 2007; Noreau *et al.*, 2013; Synofzik *et al.*, 2016). Although it was reported to be mainly restricted to Quebec region of Canada, recent observations show that it is also common in Japanese (Izumi *et al.*, 2013) and European populations (Synofzik *et al.*, 2016; Wiethoff *et al.*, 2016).

ARCA-1 is caused by homozygous or compound heterozygous mutations in *SYNE1*, one of the largest genes in the human genome, located on chromosome 6p25, and encoding the synaptic nuclear envelope protein 1. Most of the mutations were shown to result in truncation of the protein (Gros-Louis *et al.*, 2007; Izumi *et al.*, 2013; Noreau *et al.*, 2013; Synofzik *et al.*, 2016) and occur throughout the gene, implying that there is no hotspot. Moreover, recent observations show that the *SYNE1* mutations result in a large spectrum of disease from pure cerebellar ataxia to motor neuron disease (Izumi *et al.*, 2013; Özoğuz *et al.*, 2015) and mental retardation (Synofzik *et al.*, 2016).

1.1.2.2. Autosomal recessive spastic ataxia of Charlevoix-Saguenay. Autosomal recessive spastic ataxia of Charlevoix Saguenay (ARSACS; MIM #270550), is a distinct form of early-onset spastic ataxia identified in the Quebec province of Canada (Bouchard *et al.*, 1978). The carrier prevalence in the region of Charlevoix-Saguenay is estimated to be 1/22, whereas the worldwide prevalence remains unknown (De Braekeleer *et al.*, 1993). The disease is characterized by early-onset spasticity, dysarthria, nystagmus, muscle denervation, foot deformities, and axonal degeneration in the peripheral nervous system, cerebellar vermis atrophy and a positive Babinski sign (Bouchard *et al.*, 1978; García *et al.*, 2008). The age at onset is generally in the first two years of life, with infant spasticity being unique to ARSACS (Fogel and Perlman, 2007).

ARSACS is caused by homozygous or compound heterozygous mutations in the *SACS* gene encoding for a protein named “sacsin”, highly present in the brain, with the highest expression in Purkinje cells (Engert *et al.*, 1999). The protein contains an ubiquitin-like (UbL) domain which can interact with the proteasome, suggesting a role in protein degradation (Parfitt *et al.*, 2009). Moreover, the presence of the sacsins repeat region (SRR) with ATPase activity (Anderson *et al.*, 2010), which is also homologous to the Hsp90 chaperone protein (Anderson *et al.*, 2011), indicates a role in protein folding mechanism. A loss of function mechanism is hypothesized for the disease pathology since stop gain and frameshift mutations are reported. Additionally, missense mutations are shown to prevent dimerization, resulting in a non-functioning protein (Kozlov *et al.*, 2011). Studies outside of Quebec imply a broad clinical spectrum of *SACS* mutations such as ataxia without spasticity (Synofzik *et al.*, 2013) and/or spasticity without ataxic features (Gregianin *et al.*, 2013).

1.1.2.3. Ataxias with oculomotor apraxia. Oculomotor apraxia is a condition described as the inability to make voluntary eye movements in children, also observed in distinct patients with cerebellar ataxia (Cogan, 1953). Autosomal recessive ataxias with oculomotor apraxia (AOAs) are classified in four different subtypes: Ataxia-telangiectasia (Boder & Sedgwick, 1958), ataxia-telangiectasia-like disorder (Stewart *et al.*, 1999), ataxia with oculomotor apraxia type 1 (Moreira *et al.*, 2001) and ataxia with oculomotor apraxia type 2 (Moreira *et al.*, 2004). All conditions have overlapping phenotypes and share defects in a common mechanism due to mutations in genes taking role in DNA repair.

Ataxia-telangiectasia (A-T, MIM #208900) is the first identified genetic form of AOA caused by mutations in the *ATM* gene encoding a phosphatidylinositol-3 kinase (Savitsky *et al.*, 1995). The disease manifestation is typically before five years of age with cerebellar ataxia and presence of vermian atrophy on MRI (Le Ber *et al.*, 2011). Telangiectasia is a condition distinguished by small dilated blood vessels near the surface of the skin. In A-T, it is the hallmark of the disease together with elevated levels of alpha-fetoprotein (AFP). The progression is severe and the patients become wheelchair-bound by the age of 10, death occurring around the age of 20 (Swift *et al.*, 1993). The ATM protein kinase is implicated in the double-strand break repair (DSBR) mechanism at the initial step of DNA damage response (Le Ber *et al.*, 2011). Therefore, the patients have immune deficiency and are prone to malignancies.

Ataxia-telangiectasia-like disorder (A-TLD, MIM #604391) is a rare condition similar to the phenotype of A-T without the telangiectasia and elevated AFP levels (Klein *et al.*, 1996; Taylor, Groom, and Byrd, 2004). The disease progression is slower than A-T and cancer predisposition is not observed (Taylor *et al.*, 2004). Mutations in the *MRE11* gene which forms the core of MRN complex in DSB detection are identified as the genetic cause of the disease (Stewart *et al.*, 1999). Oxidative stress leading to DNA damage together with *MRE11* deficiency is the suspected mechanism of selective degeneration of the cerebellum and disease pathology (Oba *et al.*, 2010). This finding explains the decreased severity of the disease compared to A-T.

Ataxia with oculomotor apraxia type 1 (AOA1, MIM #208920), first reported in Japanese families, is a subtype of AOA which is different from A-T and described by early-onset cerebellar ataxia, oculomotor apraxia, hypoalbuminemia, hypercholesterolemia and mental retardation (Aicardi *et al.*, 1988). The disease is usually manifested before age of 10 with cerebellar ataxia, chorea or dystonic posture and oculomotor apraxia which is observed approximately nine years later than onset (Le Ber *et al.*, 2003). However, oculomotor apraxia is not observed in all patients (Ferrarini *et al.*, 2007; Le Ber *et al.*, 2003). Progressive severe axonal sensorimotor neuropathy leading to motor deficit, distal amyotrophy and foot deformities result in confinement to a wheelchair after 10 years of onset.

Genetic analyses in patients with AOA1 phenotype led to identification of the *APTX* gene encoding for aprataxin in 2001 (Date *et al.*, 2001; Moreira *et al.*, 2001). The protein has three domains all sharing homology with the histidine-triad (HIT) proteins, the amino terminal of polynucleotide kinase 3'-phosphatase (PNKP) and zinc-finger proteins indicating that it is a DNA-binding protein (Moreira *et al.*, 2001). The HIT proteins and PNKP are known to be involved in single strand break repair (SSBR) mechanism (Clements *et al.*, 2004). This finding suggests that aprataxin plays a role in SSBR and gives rise to cerebellar damage in AOA1 patients through faulty DNA repair.

In 2000, two independent families were reported with a novel form of AOA showing additional clinical features to A-T with elevated AFP levels (Bomont *et al.*, 2000; Németh *et al.*, 2000). The age at onset was shown to be later than AOA1 patients, together with less severe axonal sensorimotor neuropathy and slower disease progression. The disease was later described as ataxia with oculomotor apraxia type 2 (AOA2, MIM #606002) and was shown to be caused by homozygous or compound heterozygous mutations in the *SETX* gene coding for a 2677 amino acid long protein called "senataxin" (Moreira *et al.*, 2004). Similar to A-T, elevated AFP levels in most of the patients is a hallmark of the disease and enables exclusion of other types of AOA in the absence of telangiectasia (Anheim *et al.*, 2009). The patients become wheelchair-bound approximately 16 years after onset, later than AOA1 (Le Ber *et al.*, 2004). Similar to AOA1, oculomotor apraxia is not present in all AOA2 patients.

Interestingly, heterozygous mutations in *SETX* are associated with a rare and slowly progressive autosomal dominant form of juvenile ALS (ALS4), widening the spectrum of *SETX* mutations (Chen *et al.*, 2004). Recently, it was shown that distinct mutations alter the gene expression profiles differentially in both AOA2 and ALS4 implying alternate pathways to disease pathogenesis (Fogel *et al.*, 2014). Although certain mechanisms of pathogenesis are unknown, senataxin is shown to be involved in transcription and mRNA processing and in double-strand break repair caused by oxidative stress (Suraweera *et al.*, 2007, 2009). Therefore, similar to AOA1, cerebellar degeneration through DNA damage is a possible cause of the AOA2 phenotype.

1.2. Differential diagnosis of ataxias with diseases having ataxic features

Although they originate from different affected regions of the CNS, neurodegenerative disorders share many overlapping symptoms, making clinical diagnoses challenging. Motor neuron disorders or neuropathies outside of the CNS may have ataxic features as additional manifestation and mislead the clinician by marking the actual disease. Despite having defined clinical or pathological characterizations, hereditary spastic paraplegias (HSPs) and Charcot-Marie-Tooth (CMT) disease constitute a group of such disorders difficult for differential diagnosis. Presence of ataxia and pyramidal signs usually confuse the clinicians to make the final diagnosis. Both types of disorders share a genetic background resulting in a phenomenon called ‘spastic ataxias’ such as ARSACS, hereditary spastic paraplegia 7 (SPG7, MIM #607259), spastic ataxias type 1-5, late onset Friedreich’s ataxia (MIM #229300) and SCA3 (MIM #607047) (De Bot *et al.*, 2012). Among these, SPG7 is the only HSP gene, however, reported to be frequently found in patients with undiagnosed ataxia as a result of whole exome sequencing (Choquet *et al.*, 2015; Pyle *et al.*, 2015). Moreover, there are reports showing that CMT and ataxias might also overlap despite being distinct phenotypes (Pedroso *et al.*, 2015; Salisachs *et al.*, 1982). Genetic testing and a continuous follow-up of the patients to observe disease progression are required in order to distinguish between different clinical entities before making the final diagnosis.

1.3. Next Generation Sequencing Technologies

The completion of the human genome project (HGP) in 2003, which took more than 10 years, brought a wealth of technological developments to sequencing technologies. Since then new massively parallel sequencing approaches have been developed by different companies to decrease the cost and time per megabase while increasing the accuracy. These so-called next generation sequencing (NGS) technologies have drastically changed the understanding of the human genome and genetics. Following the first publication of the individual genome based on NGS, the number of publications on various applications of NGS has exponentially increased, showing the effectiveness of the technology (Wheeler *et al.*, 2008).

Depending on the purpose, the NGS technology allows performing various applications such as whole genome (WGS), whole exome (WES), transcriptome (RNA-seq) sequencing, genome-wide profiling of protein-DNA (ChIP-seq) and protein-RNA interactions (CLIP-seq). Novel applications and algorithms to analyze the data have also been developed to make data analysis less challenging and more accurate. This technology has begun to be a part of routine research methods in biology laboratories, thereby creating new challenges and strategies in both computational and molecular biology.

1.3.1. Whole Exome Sequencing

Protein coding regions of the human genome include 180,000 exons and constitute approximately 1% of the total DNA sequence which is approximately 30 megabases. The rest of the genome contains repetitive regions, introns and intergenic regions which have high rate of variations among individuals without a direct effect on the protein sequence. Moreover, it has been shown that up to 85% of the disorders are linked to the variations in the exome (Botstein and Risch, 2003). The exome is crucial to focus on while studying the genetic layout of disorders due to its higher probability to cause a disease and lesser complexity compared to the genome.

Whole exome sequencing is an application of the NGS technology by enriching the protein coding regions of the genome. WES is an efficient tool in the identification of single-gene disorders with its high-throughput capacity and relatively low cost. The first use of this technology to discover novel disease genes was shown in the identification of the genetic cause of Miller syndrome (S. B. Ng et al., 2010). This was also a demonstration of the power of the technology to identify a causative gene for a Mendelian disorder with only a small number of affected individuals.

WES is composed of two main parts: the wet-lab where library preparation to enrich the exons and sequencing are performed, and the *in silico* lab which is the data analysis step. Data analysis of WES includes the alignment of the reads to the reference genome, followed by variant calling and functional annotation of the variants. The variants are then filtered and prioritized according to the defined strategy for the particular study.

1.3.1.1. Study design for exome sequencing. Identification of a causative variant from WES requires a proper design starting from the sample selection. All pedigrees must be investigated individually with certain criteria, such as having more than one affected individual and selecting certain unaffected individuals as controls to reduce the amount of variants. Nevertheless, being in unaffected state may be misleading for age-dependent or dominant disorders. Therefore, it is important to have sufficient knowledge on all contributors' phenotypes and to consider the age when age-related traits are investigated.

Several strategies have been described for the analysis of whole exome data such as family-based studies and population studies by sequencing independent individuals with the same phenotype observed in a particular population (Figure 1.1). The family-based approach is useful when a number of affected and unaffected individuals are available within the same pedigree. The affected individuals are likely to have the same mutation decreasing the number of candidate genes. For population studies, which is the case when there are a number of people having the same rare trait, the overlap strategy may help assuming that

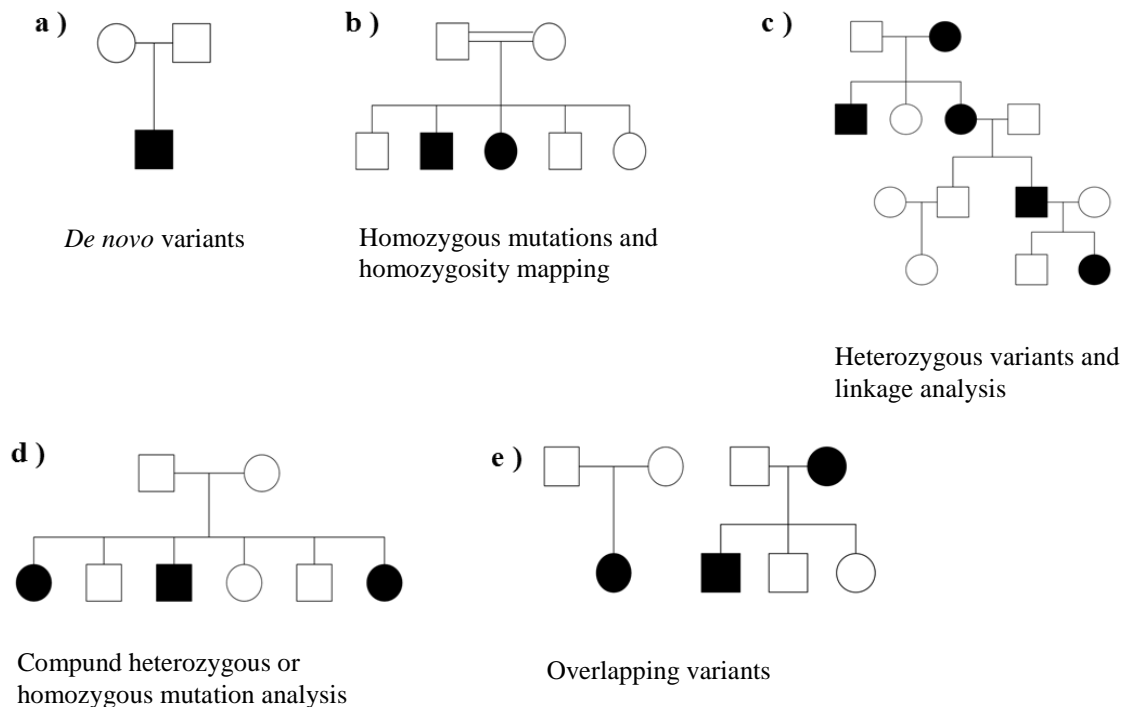


Figure 1.1. Strategies for analysis of whole exome sequencing data. *De novo* variant analysis (a) family-based approaches (b, c, and d) and overlapping variants for population based strategy (e)

there is no genetic heterogeneity and all individuals have the trait because of a variant in the same gene (Rabbani *et al.*, 2012). However, common disorders have a heterogeneous background causing the overlap strategy fail since individuals have their own genetic cause resulting in the trait. *De novo* strategy has been shown to be successful in this scenario, where unaffected parents and affected child trio are sequenced and variants present only in the affected child are identified (Gilissen *et al.*, 2012). Each case must be examined through these criteria and sample choice must be done accordingly prior to the selection of the sequencing platform.

The choice of the sequencing protocol and platform is another crucial step since it defines the limits of accuracy and sensitivity as well as the type of variations that can be received from the data. There are various commercially available kits differing in the regions they cover and the efficiency of the enrichment. The read depth which refers to the number of reads at a certain position, must be around 50-70X for each individual in order to have reliable results. Moreover, the read length and method are differing parameters across platforms which must be carefully taken into account according to the purpose. The long-read platforms must be chosen for *de novo* assembly and investigation of repetitive regions, whereas the short-read platforms must be preferred for comparative assembly and clinical applications.

The number of bioinformatics software on NGS data analysis is increasing day-by-day with their own limitations and advantages depending on the goal. The final step of the study design is to choose which programs / tools to be used for data analysis including the choice of the reference genome. There are two main algorithms for read alignment: indexing of the genome and the Burrows-Wheeler transform. The latter is computationally less heavy and faster, being successful for identification of single nucleotide variations (SNVs) and small insertion or deletions (indels). However, the first algorithm works better in detection of larger indels, repeats and structural variations while bringing computational cost and requiring more time making it not useful for routine applications. The choice of the variant calling software is also critical, since every variant caller ends up with different sets of variants from the same dataset. Variant sets obtained from different variant calling software, have around 40% non-overlapping variations, thereby altering the outcome of the analysis.

Similar to the alignment step, structural variation, and large indel calling should be performed using specialized alternate algorithms to have results that are more accurate.

1.3.1.2. Homozygosity mapping. Geographical isolation, small population size or high rates of consanguinity result in formation of inbred populations characterized by long homozygous stretches within the genome of individuals due to increased level of parental relatedness (McQuillan *et al.*, 2008). Inbreeding influences the prevalence of monogenic disorders by increasing homozygosity which results in elevated probability of deleterious mutation reunion (Wright, 1977). Therefore, the affected individual is likely to have inherited the same mutation from a recent common ancestor, through both parents on the same genomic region, which is referred to as identical by descent. Homozygosity mapping is the method for detecting these regions in affected individuals from inbred populations or consanguineous marriages to reduce the search area for identification of genetic defects.

Runs of homozygosity (ROH) can be defined as the measure of the homozygous stretches between defined polymorphic positions by homozygosity mapping (Nothnagel *et al.*, 2010). Traditionally, detection of ROHs is performed from genotyping of SNPs or other multiallelic markers. Recently, NGS technologies have changed the strategy to detect ROHs from genotyping arrays to WES data. New bioinformatics tools have been developed for detection of ROHs from WES data while the existing ones have been adapted. H3M2 is one such program to detect ROHs from WES data by using Hidden Markov Models (Magi *et al.*, 2014). HomozygosityMapper is an online and user-friendly tool which has been adapted to detect ROHs from WES data using sliding blocks between predefined reported variations (Seelow *et al.*, 2009). PLINK is a state-of-art software for detection of ROHs from genome-wide association data (Purcell *et al.*, 2007). Nonetheless, there are studies reporting the utility of PLINK for runs of homozygosity detection with WES data (Kancheva *et al.*, 2015; Pippucci *et al.*, 2014). Comparison and optimization of the tools, having approximately the same results from all, are needed for more reliability and sensitivity since all tools use different algorithms.

1.3.1.3. Linkage analysis. The first disease gene, the genetic cause of Huntington's disease was achieved by linkage analysis followed by positional cloning (Gusella *et al.*, 1983). Genetic linkage analysis is based on the observation that physically close genes are

transmitted together to the offspring since they still remain linked after the crossing over process during meiosis. Linkage is a powerful tool to detect the chromosomal location of disease genes. However, it requires a number of individuals and is also not cost-effective. The developments in the NGS technology have enabled combining linkage analysis and positional cloning to identify the gene in the linked region at one single step making it faster, more cost-effective and capable of working with smaller sample size (Ott et al., 2015). Although genomic markers are preferred in linkage analysis, there are reports showing its effectiveness with WES data (Gazal *et al.*, 2015).

1.3.1.4. Limitations. Despite its plenty of advantages, WES has some technical drawbacks in both wet-lab and data analysis steps. The most evident drawback is the lacking ability of detecting variants outside of the coding region of the genome. Variations within the intronic or regulatory regions, which are missed in WES, may alter the expression of the genes contributing to disease phenotype. Furthermore, the enrichment of the exons does not always give 100% yield due to problems in target-probe hybridization. The GC-rich regions may fail in capturing and sequencing. Moreover, there is an amplification bias due to the sequencing by the synthesis method causing false positives or false negatives. The error rate of NGS platforms are around 2%, ending up with approximately 6000 errors per exome (Clark *et al.*, 2011). High coverage is crucial to reduce the number of false-negatives or positives, however, errors due amplification bias might still be missed. Therefore, the presence of mutations must be confirmed by Sanger sequencing.

Data analysis of WES has also its drawbacks. There are some widely used programs preferred for being the most effective ones, thereby acknowledging the fact that the result will contain false negatives. Detection of structural variations such as expansions, large indels, copy number variations (CNVs) and satellite regions are major weak points of WES. A majority of these variations occur within intronic and intergenic regions of the genome. Several algorithms focusing on one type of structural variation are still far from being mature.

Public databases containing minor allele frequencies (MAF) do not always reflect every population. Therefore, controlling the MAF of a suspected variant in a database reflecting the population of interest or in-house controls are required. Additionally, there are

online tools to predict the effect of a variation on protein structure and function. PolyPhen2, SIFT, MutationTaster and LRT are such prediction tools, however disease-causing mutations can be predicted as benign or may be tolerated by some of these tools (Foo *et al.*, 2012). Therefore, the scores of these tools must not be primarily used for variant filtration.

1.3.2. Molecular analysis of neurodegenerative disorders with NGS

Neurodegenerative disease genetics evolved with the improvements in NGS technology. The number of causative genes is drastically increasing especially in familial forms of the disease (Table 1.3). Although the rate of success is lower in sporadic cases, this progression in genetics helped the development of disease models which makes the elucidation of the pathogenesis also possible in sporadic disease (Tsuji, 2013). Thus, the distinction between sporadic and familial disease starts to disappear.

Table 1.3. Examples of neurodegenerative disease genes identified by whole exome sequencing analysis

Disease	Inheritance	Gene	Reference
Spinocerebellar ataxia 35	AD	TGM6	(J. L. Wang <i>et al.</i> , 2010)
Spinocerebellar ataxia 28	AD	AFG3L2	(Pierson <i>et al.</i> , 2011)
Spinocerebellar ataxia with psychomotor retardation	AR	SYT14	(Doi <i>et al.</i> , 2011)
Parkinson's disease	AD	VPS35	(Vilariño-Güell <i>et al.</i> , 2011; Zimprich <i>et al.</i> , 2011)
Hereditary spastic paraplegia 12	AD	RTN2	(Montenegro <i>et al.</i> , 2012)
Amyotrophic lateral sclerosis	AD	TBK1	(Cirulli <i>et al.</i> , 2015; Freischmidt <i>et al.</i> , 2015)

2. PURPOSE

Diseases of the cerebellum constitute a broad range of neurodegenerative disorders and substantially affect the living quality of the patients. Among these, ataxias appear as the majority of cerebellar disorders and their presenting with overlapping phenotypes makes an exact clinical diagnosis challenging. Although most ataxias do not have an available treatment, there are subtypes of cerebellar disease with treatment options, providing the patients with a better life quality. Therefore, a definitive genetic diagnosis is important. The genetics of ataxias are well studied in the inbred Japanese and French-Canadian populations having led to identification of novel disease genes in recent years. Also commonly studied in Europe, the epidemiology of ataxias is well defined in European populations. However, Turkey is largely unexplored.

Turkey is a large country with a high birth rate and a high degree of consanguinity among couples, especially in the rural parts of the country. Thus, early-onset recessive forms of ataxia are very frequent in Turkey in addition to late-onset dominant forms.

The aim of this study is to investigate the epidemiology and molecular basis of dominant and recessive ataxias. This is achieved by the application of WES in this thesis. To the best of our knowledge, this is the first and largest study in Turkey applied to ataxia families and reveals the molecular background of several families, which remained unidentified by using conventional methods.

3. MATERIALS

3.1. Subjects

In the framework of this thesis 47 patients (index cases), referred to our laboratory with an initial diagnosis of dominant or recessive ataxia, were investigated along with their affected and unaffected family members. An autosomal dominant inheritance pattern was observed in 12 out of 47 families, thirty-four had an autosomal recessive mode of inheritance, and consanguinity was observed in 34 pedigrees (Figures 3.1 – 3.5). Both transmission modes were considered in one family. The mean age at onset of the index patients was 22.37, with a standard deviation of 15.77, ranging from infancy to 56 years of age; 39 probands had an age of onset below 40 years of age.

Informed consent was obtained from all participants in this study. The study protocol was approved by the Ethics Committee on Research with Human Participants (INAREK) at Boğaziçi University. Expert neurologists in several hospitals throughout Turkey performed clinical evaluation of the index cases. Peripheral blood samples were collected into EDTA-containing tubes for analysis.

Upon referral, the patients were first screened for trinucleotide repeat expansions in several genes for: SCA 1, 2, 3, 6, 7, 17, DRPLA, and FRDA genes (*ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *ATN1*, and *FXN*). Following exclusion of the above disorders, the disease history in the family was further questioned and extensive clinical phenotypes of the patients were requested (Table 3.1).

Table 3.1. Clinical characteristics of families investigated in this study.

	ID / Gender	Age at onset	Cerebellar ataxia	Cerebellar atrophy	Gait ataxia	Dysarthria	Nystagmus	Tremor	Spasticity	Other clinical findings
Family 1	P1, M	18	+	+	+	+	+	-	-	-
	P2, M	20	+	NA	+	+	+	-	-	bilateral ptosis
	P3, M	NA	+	NA	+	+	+	-	-	-
	P4, M	NA	+	NA	+	+	+	-	-	-
Family 2	P5, F	37	+	NA	+	+	-	-	-	pes cavus, loss of vibration sense
	P6, F	46	+	NA	+	-	-	-	+	DTR hyperactive, bilateral hammer toe, loss of vibration sense
	P7, M	46	+	NA	+	+	-	-	+	pes cavus, loss of vibration sense
Family 3	P8, F	17	+	+	+	+	+	-	+	bilateral Babinski sign, bulbar involvement, decreased vibration sense
	P9, M	10	+	+	+	+	-	+	+	loss of vibration sense, dysphagia
	P10, M	5	+	+	+	+	+	-	+	bilateral Babinski sign, pes cavus, amyotrophy
Family 4	P11, M	infant	+	+	+	-	+	-	-	loss of vibration sense, amyotrophy, DTR abolic
	P12, M	20	+	+	+	+	-	-	-	Babinski sign, pes cavus, DTR live in upper extremities; abolic in achilleus
Family 5	P13, F	infant	+	+	+	-	-	-	-	mental retardation, social problems
	P14, M	NA	+	+	+	-	-	-	-	
	P15, M	NA	+	+	+	-	-	-	-	

Table 3.1. Clinical characteristics of families investigated in this study (cont.).

	ID / Gender	Age at onset	Cerebellar ataxia	Cerebellar atrophy	Gait ataxia	Dysarthria	Nystagmus	Tremor	Spasticity	Other clinical findings
Family 6	P16, F	13	+	+	+	+	-	-	-	Babinski sign, slowed saccadic movements, loss of vibration sense
Family 7	P17, M	infant	+	-	+	+	-	-	-	DTR bilateral hypoactive, sensory polyneuropathy
Family 8	P18, F	18	+	NA	+	-	-	-	-	small cerebellum
	P19, M	NA	+	NA	+	-	-	-	-	-
Family 9	P20, F	20	+	NA	+	+	+	+	-	loss of vibration sense
	P21, F	25	+	NA	+	+	+	+	-	
Family 10	P22, F	27	+	NA	+	+	+	-	-	dysmetria, DTR absent in achilleus
	P23, M	25	+	NA	+	+	+	-	-	
	P24, F	28	+	NA	+	+	+	-	-	
Family 11	P25, M	15	+	NA	+		-	-	-	-
Family 12	P26, F	23	+	+	+	+	-	-	+	keratoclonus, dysmetria, bilateral achilleus
Family 13	P27, F	15	+	NA	+	+	-	-	+	mild dysmetria, tendon reflexes live
	P28, F	33	+	NA	+	+	-	-	+	
Family 14	P29, F	17	+	+	+	+	-	-	+	cognitive dysfunction, thin corpus callosum

Table 3.1. Clinical characteristics of families investigated in this study (cont.).

	ID / Gender	Age at onset	Cerebellar ataxia	Cerebellar atrophy	Gait ataxia	Dysarthria	Nystagmus	Tremor	Spasticity	Other clinical findings
Family 15	P30, M	23	+	+	+	+	+	-	-	cognitive decline, decreased DTR
	P31, F	NA	NA	NA	+	+	+	-	-	-
	P32, M	NA	NA	NA	+	+	+	-	-	-
Family 16	P33, M	33	+	+	+	+	-	-	+	vestibular symptoms, DTR increased
Family 17	P34, M	3	+	NA	+	+	-	+	-	mental retardation
	P35, F	NA	NA	NA	+	+	-	+	-	
	P36, M	3	+	NA	+	+	-	+	-	
Family 18	P37, M	6	+	NA	+	+	-	-	-	-
	P38, M	4	NA	NA	+	+	-	-	-	-
Family 19	P39, F	14	+	+	+	+	-	+	-	myoclonus in upper extremities
Family 20	P40, F	23	-	-	+	+	-	+	-	perception difficulty
Family 21	P41, M	32	+	-	+	+	-	-	-	dysphagia
	P42, M	30	+	NA	+	+	-	-	-	
	P43, F	50	+	NA	+	+	-	-	-	
Family 22	P44, M	54	+	NA	+	+	-	-	-	-

Table 3.1. Clinical characteristics of families investigated in this study (cont.).

	ID / Gender	Age at onset	Cerebellar ataxia	Cerebellar atrophy	Gait ataxia	Dysarthria	Nystagmus	Tremor	Spasticity	Other clinical findings
Family 23	P45, M	49	+	+	+	+	+	-	-	Babinski positive, mild cognitive dysfunction
	P46, F	NA	+	NA	+	+	+	-	-	-
	P47, M	NA	+	NA	+	+	+	-	-	-
Family 24	P48, F	39	+	NA	+	+	+	+	-	disability in saccadic movements, Babinski sign
	P49, M	18	+	+	+	+	-	+	-	-
	P50, M	NA	NA	NA	-	-	-	+	-	-
	P51, M	NA	NA	NA	-	-	-	+	-	-
Family 25	P52, F	36	+	NA	-	+	-	-	-	-
	P53, F	53	+	NA	-	+	-	-	-	-
Family 26	P54, F	51	+	NA	+	-	-	-	-	-
Family 27	P55, F	47	+	NA	+	-	+	-	-	Babinski sign, deep sensory impairment, impediment, fatigue
	P56, M	33	+	NA	+	-	-	-	-	-
	P57, M	31	+	NA	+	-	-	-	-	-
Family 28	P58, M	47	+	NA	+	+	-	-	+	DTR increased, gaze palsy, dysmetria
	P59, F	50	+	NA	+	+	-	-	+	DTR increased, gaze palsy, dysmetria, chorea
	P60, F	55	+	NA	+	+	-	-	-	-

Table 3.1. Clinical characteristics of families investigated in this study (cont.).

	ID / Gender	Age at onset	Cerebellar ataxia	Cerebellar atrophy	Gait ataxia	Dysarthria	Nystagmus	Tremor	Spasticity	Other clinical findings
Family 29	P61, F	NA	+	+		+	+	-	-	reflexes live
Family 30	P62, F	11	+	NA	+	-	-	-	-	-
	P63, M	25	+	NA	+	-	-	-	-	-
	P64, M	50	+	NA	+	-	-	-	-	-
	P65, M	57	+	NA	+	-	-	-	-	-
Family 31	P66, M	34	+	+	-	-	-	-	-	psychotic signs
Family 32	P67, M	32	+	NA	+	+	-	+	-	dementia, dysphagia, REM sleep disorder, polyneuropathy, hearing loss, white matter on MRI
Family 33	P68, M	18	+	NA	+	+	-	-	-	sensory polyneuropathy
	P69, M	30	+	NA	+	+	-	-	-	-
	P70, F	NA	+	NA	+	+	-	-	-	-
Family 34	P71, F	20	+	NA	+	-	-	-	-	-
	P72, F	NA	+	NA	+	-	-	-	-	-
	P73, M	NA	+	NA	+	-	-	-	-	-
	P74, M	NA	+	NA	+	-	-	-	-	-
Family 35	P75, M	56	-	NA	+	-	-	-	-	slowly progressive disease
Family 36	P76, M	43	+	NA	+	+	-	-	-	-

Table 3.1. Clinical characteristics of families investigated in this study (cont.).

	ID / Gender	Age at onset	Cerebellar ataxia	Cerebellar atrophy	Gait ataxia	Dysarthria	Nystagmus	Tremor	Spasticity	Other clinical findings
Family 37	P77, F	infant	-	NA		-	-	-	-	unable to walk, mental retardation
Family 38	P78, F	18	+	NA		-	-	-	-	-
Family 39	P79, F	15	+	NA	+	-	+	-	-	dysmetria, loss of vibration sense in lower extremities, DTR lowered
	P80, F	NA	+	NA	+	-	+	-	-	fatigue
Family 40	P81, M	35	+	NA	+	-	-	-	+	-
Family 41	P82, M	39	+	NA		-	-	-	-	-
	P83, F	44	+	NA		-	-	-	-	-
Family 42	P84, M	childhood	+	+	+	+	-	-	+	-
	P85, M	childhood	+	+	+	+	-	-	+	-
Family 43	P86, M	12	+	NA		-	-	-	-	-
Family 44	P87, M	14	+	+	+	-	-	+	-	dysmetria
Family 45	P88, M	2	+	+	+	-	-	-	-	-
Family 46	P89, F	19	+	NA	+	+	+	-	+	dysmetria, DTR increase, Babinski sign, loss of vibration sense
Family 47	P90, M	2	+	NA	+	+	+	+	-	dysphagia, dysmetria, loss of vibration sense
	P91, F	2	+	NA	+	+	+	-	-	-

NA: not available, DTR: deep tendon reflexes, M: male, F: female, P: patient

3.1.1. Family trees

3.1.1.1. Families with an autosomal dominant inheritance pattern

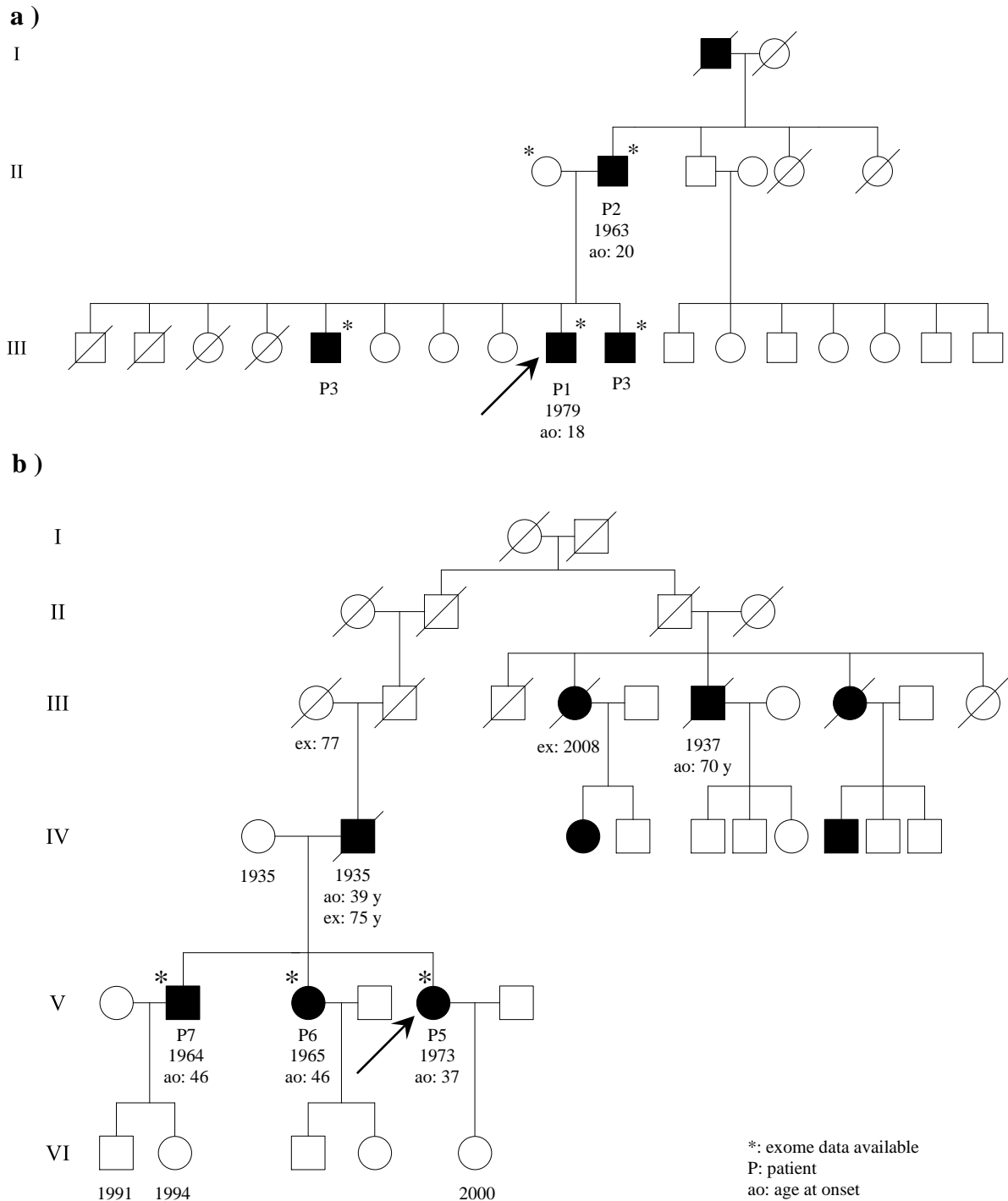
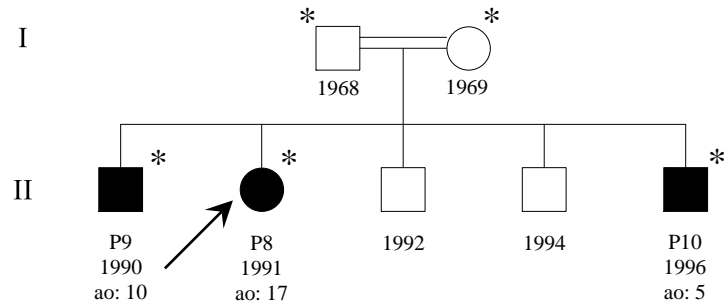


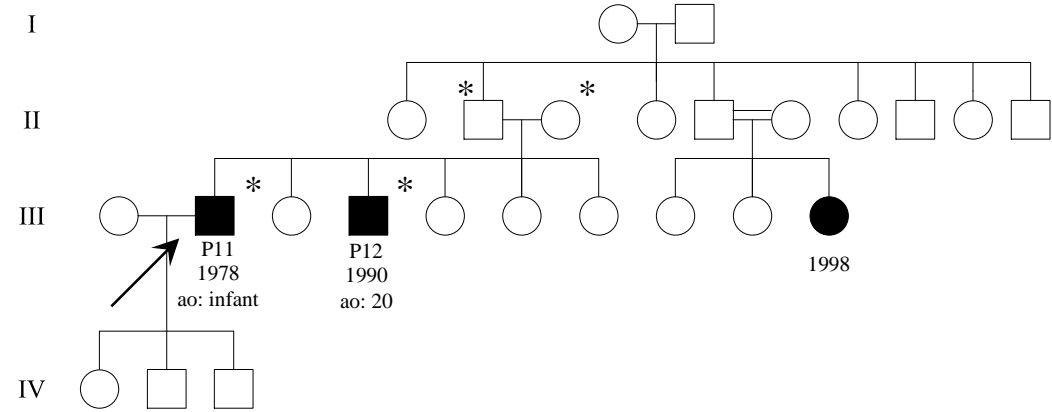
Figure 3.1. a) Family 1 (Patients P1-P4) and b) Family 2 (Patients P5-P7) showing an autosomal dominant inheritance pattern with several affected individuals.

3.1.1.2. Families with an autosomal recessive inheritance pattern

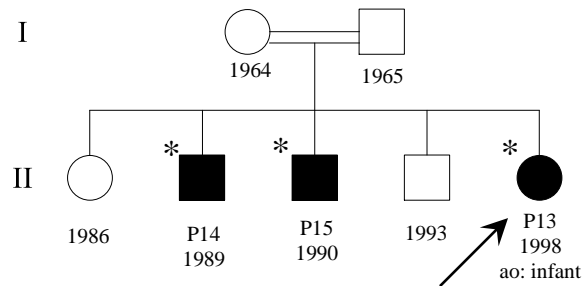
a)



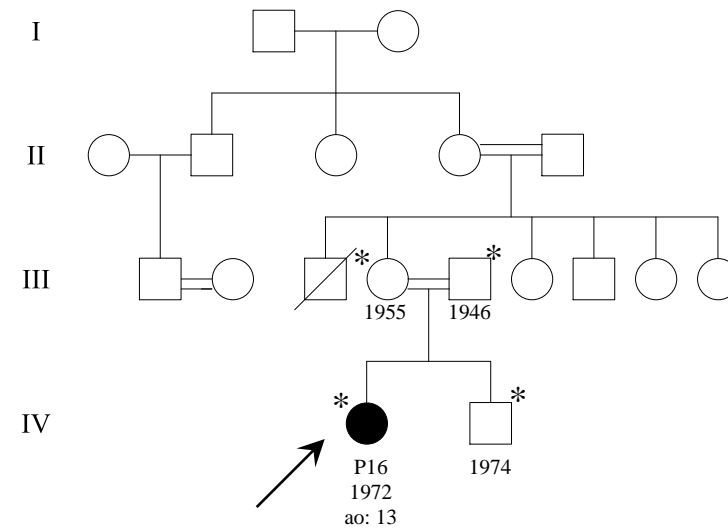
b)



c)



d)



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

Figure 3.2. a) Family 3 (Patients P8-P10) b) Family 4 (Patients P11 and P12) c) Family 5 (Patients P13-P15) d) Family 6 (Patient P16)

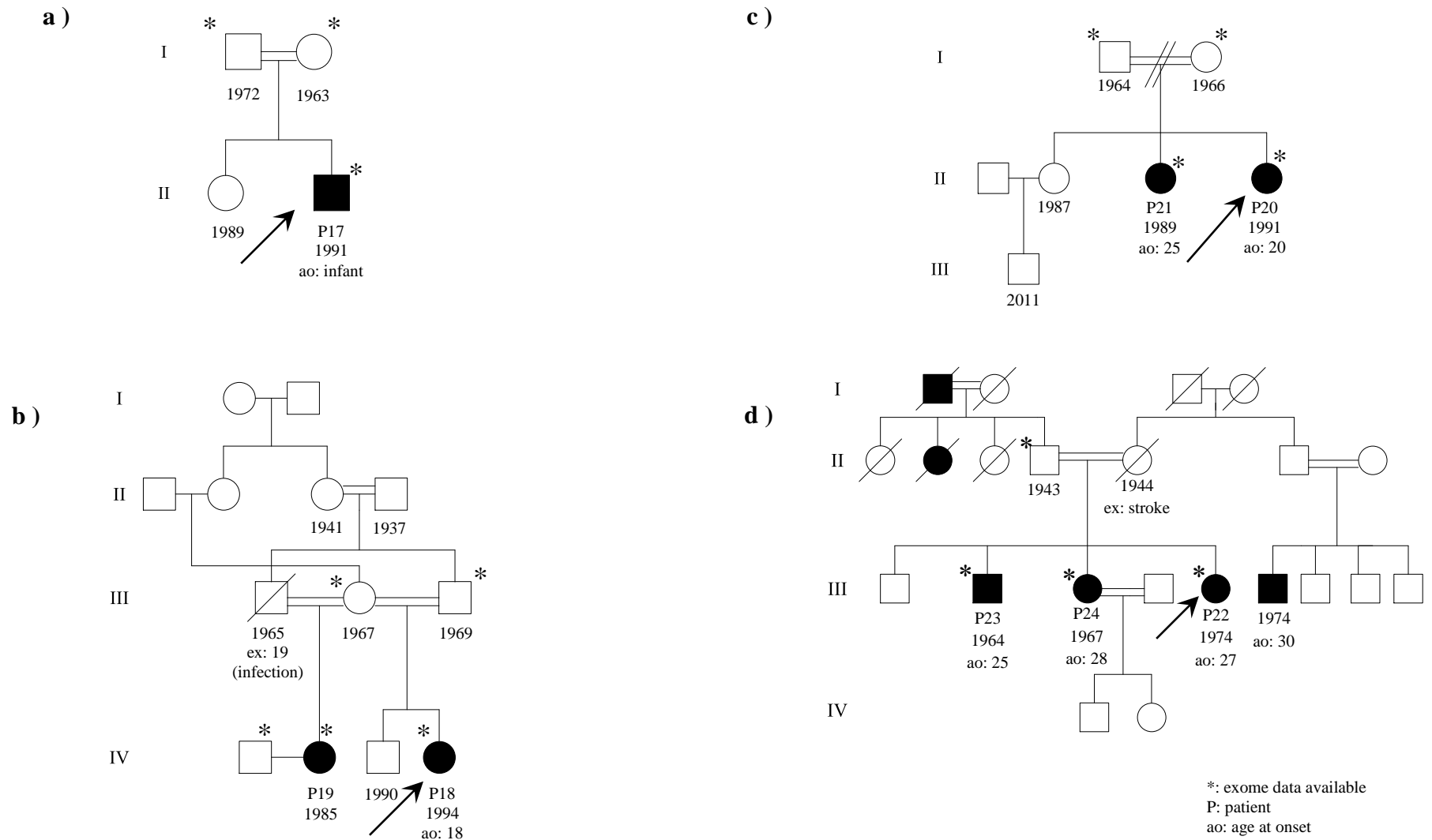


Figure 3.3. . a) Family 7 (Patient P17) b) Family 8 (Patients P18 and P19) c) Family 9 (Patients P20 and P21) d) Family 10 (Patients P22-P24)

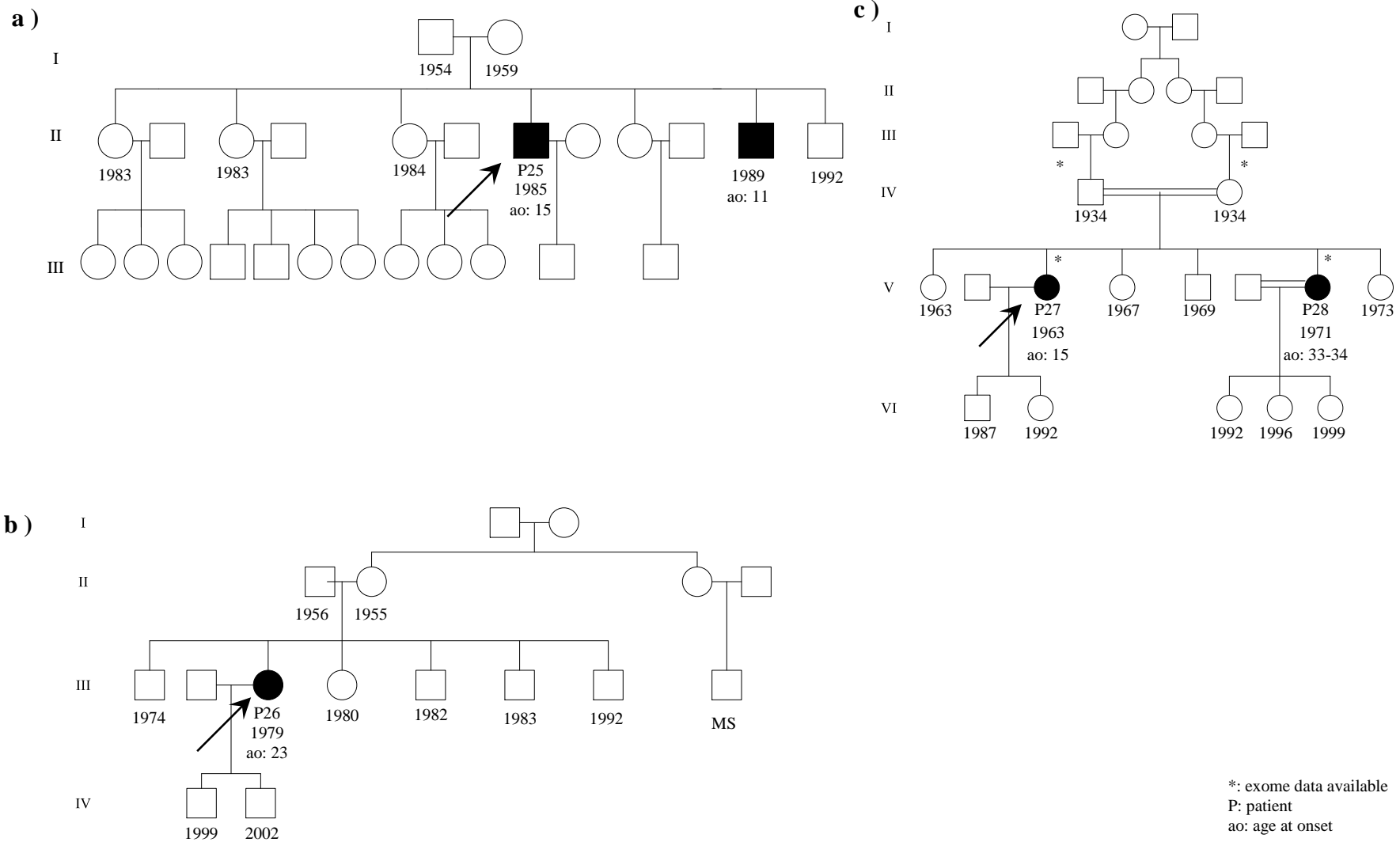


Figure 3.4. a) Family 11 (Patient 25) b) Family 12 (Patient P26) c) Family 13 (Patients P27 and P28)

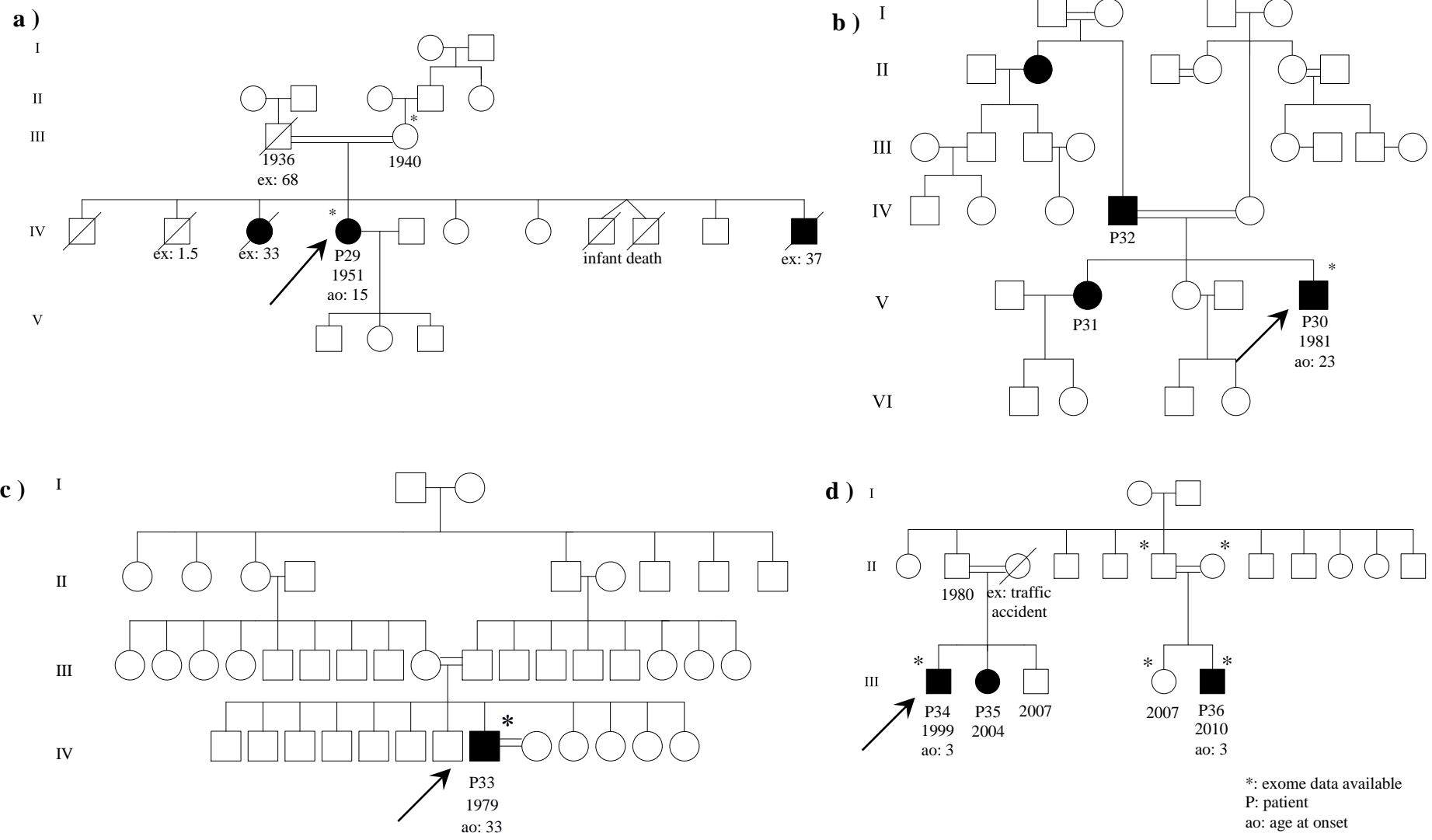
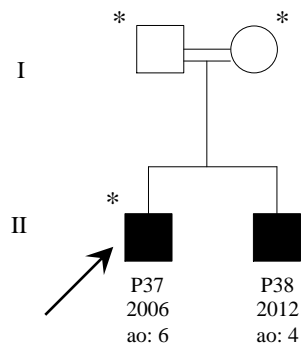
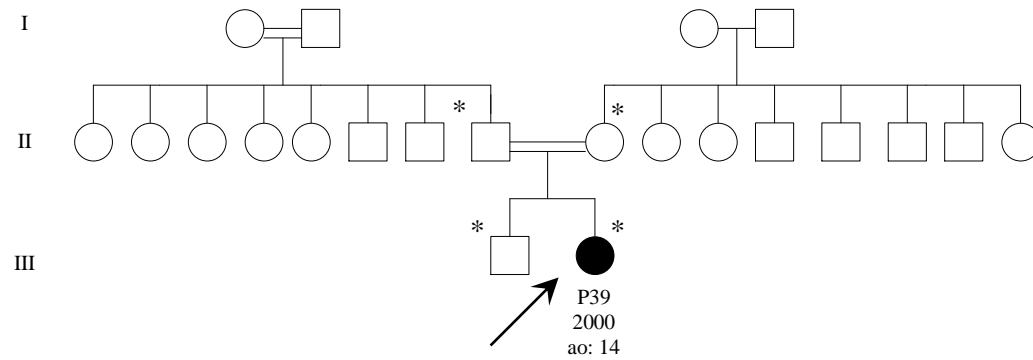


Figure 3.5. a) Family 14 (Patient P29) b) Family 15 (Patients P30-P32) b) Family 16 (Patient P33) c) Family 17 (Patient P34 and P36)

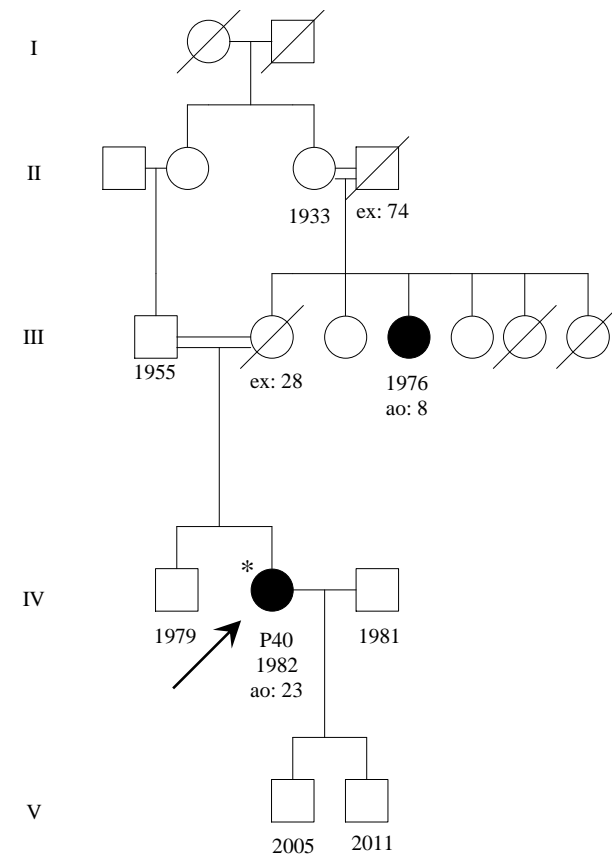
a)



b)



c)



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

Figure 3.6. a) Family 18 (Patient P37 and P38) b) Family 19 (Patient P39) c) Family 20 (Patient P40)

3.2. Equipment and Solutions used in DNA Isolation

MagNA Pure Compact Nucleic Acid Isolation KitI and MagNA Pure Compact Instrument from Roche were used for DNA extraction from peripheral blood samples. Concentration and quality measurements were performed by Thermo Scientific NanoDrop 2000c UV-Vis Spectrophotometer.

3.3. Whole Exome Sequencing Platforms and Exome Enrichment Kits

Samples subjected to whole exome sequencing were outsourced to different institutions and companies (UMASS, Scientific and Technological Research Council of Turkey, Macrogen Inc., and The Center for Applied Genomics). The exome enrichment kits used are listed in Table 3.2. Sequencing platforms were Illumina HiSeq 2000 and Illumina HiSeq 2500.

Table 3.2. Exome enrichment kits.

Kit	Catalog No.	Company
SeqCap EZ Whole Exome V2	05860504001	Roche, Switzerland
SeqCap EZ Whole Exome V3	06465692001	Roche, Switzerland
SureSelect Human All Exon V5	5190-6209	Agilent, USA
SureSelect Human All Exon V5+UTR	5190-6214	Agilent, USA

3.4. Equipment and Solutions used in Validation Experiments

The GoTaq DNA Polymerase (Promega, USA) with a concentration of 5u/μl and Colorless GoTaq® Reaction Buffer (5X) were used in the validation PCR to amplify 50 ng of genomic DNA. The sequences of the primers used in validation experiments are listed in Table 3.3.

Table 3.3. Sequences of primers used in validation experiments.

Primer Name	T _m (°C)	Sequence (5' -> 3')
<i>AFG3L2</i> – E16F	56.9	GATTTTGTCTGGTTAAAGAACAATCA
<i>AFG3L2</i> – E16R	62.2	CCAACCAAAACAGTCTATCTATCACTTC
<i>KIF1B</i> – E43F	56.4	TGTAGTGTTCCGTTTGCTTC
<i>KIF1B</i> – E43R	53.8	ATTGAGCTAGAGGAAGGTTG
<i>SYNE1</i> - E108F	57.0	TTTAAATCCACCTAAAGTCATTACC
<i>SYNE1</i> - E108R	58.4	TGTGCCACCATACCTAGTCC
<i>SACS</i> - E10.1R	60.2	CCATTTCAAATACAACCGCC
<i>SACS</i> - E10.2F	58.2	GATGAAGAAATGGTAAAACTAGAGC
<i>SACS</i> - E10.2R	58.7	AACATTTGCACACCAATATTCC
<i>SETX</i> – E10F	53.2	TCAGTTGGAATCTTTGAGTG
<i>SETX</i> – E10R	53.2	CAAGGTAAAGAGGTAGCTCA
<i>SETX</i> – E13F	55.3	CCTTGATCCCCGCCTCCT
<i>SETX</i> – E13R	55.3	CTGTTACCGTGAACACATT
<i>SETX</i> – E21F	59.1	CCTGAACGTTGCCATATACG
<i>SETX</i> – E21R	52.3	GACAAGACCTAAGACGACAG
<i>CAPN1</i> – E8F	59.10	GGAAGCTAGCACTGACAGGA
<i>CAPN1</i> – E8R	59.02	TGAACTCCCGCATGAAGTCT
<i>CAPN1</i> – E10F	62.7	CTCTGATCCCCGCCTCCT
<i>CAPN1</i> – E10R	58.8	ACCCTCCTGACCTCGTAGA
<i>SPG11</i> – E10F	59.0	CCCAGGACTAATCATGAAGG
<i>SPG11</i> – E10R	59.9	ATCCCCAAACCGATAAAACC
<i>SPG7</i> – E13F	64.8	ACTGCTCTGCGCCTGCAGT
<i>SPG7</i> – E13R	55.1	CCTTGTGTGGTAGACCCA

Table 3.3. Sequences of primers used in validation experiments (cont.).

Primer Name	T _m (°C)	Sequence (5' -> 3')
<i>KIF1C</i> – E3F	60.4	AGGGCAGGAGTGTCTGGAG
<i>KIF1C</i> – E3R	59.8	CGACTAGCACCTGCTCTGTG
<i>ADCK3</i> – E8F	61.4	GACAGACTTGGGGCTTCTCC
<i>ADCK3</i> – E8R	61.0	CCCCATAGGGACAGGCACT
<i>TTPA</i> – E3F	59.7	TCTCCCTTTGGAAATCTTTCTG
<i>TTPA</i> – E3R	57.8	TTGGAAGTATTATGCCTGACAGT

T_m: melting point, F: forward, R: reverse

3.5. Equipment and Solutions for Agarose Gel Electrophoresis and Extraction from the Gel

All chemicals and equipment used in agarose gel electrophoresis are listed in Table 3.4. Gel extraction of the PCR products were performed with QIAQuick Gel Extraction Kit, Qiagen.

Table 3.4. Chemicals used in gel electrophoresis.

Solution/Chemical/ Equipment	Catalog No.	Company	
5X TBE	161-0733	Biorad, USA	
			0.89 M Tris-base
			0.89 M Boric acid
2 mM Na ₂ EDTA			
Agarose Biomax	-	Prona, EU	
DNA Ladder, GeneRuler™ 100bp	SM0241	Thermo Scientific, USA	
Blue/Orange Loading Dye (6X)	G1881	Promega, USA	
Tank, OWL EasyCast B1	349900	Thermo Scientific, USA	
EU GelRed Dropper Bottle	103.302-05	Olerup SSP, Sweden	
Power Supply, EC300XL2	C1596100710705	Thermo Scientific, USA	
GelDoc	-	Biorad, USA	

3.6. General Laboratory Equipment and Kits

All general laboratory equipment and kits used in the framework of this thesis are listed in Tables 3.5 and 3.6, respectively.

Table 3.5. General laboratory equipment

Equipment	Type	Catalog No.	Company
Autoclave	ASB620T	-	Astell, UK
Centrifuges	Microfuge16	A46473	Beckman Coulter, USA
	C1301	C1301B-230V	Labnet, USA
Balance	TE612	-	Sartorius, Germany
Computers	Z820 Workstation (64GB RAM, 12 core, SSD harddisk, NVIDIA quadro, Intel Xeon CPU E5-2640 0 @ 2.5 GHz)	-	Hewlet-Packard (HP), USA
	X550LC (Intel(R) Core (TM) i5-4200U CPU @ 1.60 GHz 2.30 GHz (TM), 4 GB)	-	ASUSTek Computer Inc., Taiwan
Refrigerators	2021D (-20 °C)	-	Arçelik, Turkey
	HT5786-A (-86 °C)	-	Hettich, Germany
DNA extraction system	MagNA Pure Compact Instrument	-	Roche, Switzerland
Thermal Cycler	T100	621BR07123	Biorad, USA
Gel Documentation System	Geldoc	-	Biorad, USA
Electrophoretic Equipment	OWL EasyCast B1 Tank	349900	Thermo Scientific, USA
Power Supply	EC300XL2 Power Supply	C1596100710705	Thermo Scientific, USA

Table 3.5. General laboratory equipment (cont.).

Equipment	Type	Catalog No.	Company
Falcon Tubes	EasyOpen 50-ml Centrifuge Tubes	-	JETBIOFIL, China
Microcentrifuge Tubes	2-ml MaxyClear Snaplock Microcentrifuge Tube	-	Axygen, USA
	1.5-ml Boil-Proof Microtubes	-	
	0.5-ml Thin Wall Flat Cap PCR Tubes	-	
Microwave oven	Intellowave MD554	501283307	Arçelik, Turkey
Micropipettes	Rainin Pipet-Lite XLS, 0,5-2 µl, 2-10 µl, 2-20 µl, 20-200 µl, 200-1000 µl	-	Mettler-Toledo International Inc., USA
Heat Block	BBA1	-	Grant-Boekel, USA
Heater	MR3001	-	Heidolph, Germany
Water Purification System	Arium® 611UV Ultrapure Water System	-	Sartorius, Germany
Aluminum foil	Aluminum foil alupro	A0621	Ecopla, France
Parafilm	-	PM-996	Bemis, USA

Table 3.6. Commercially available kits, other than exome capture kits.

Kit	Catalog No.	Company
MagNA Pure Compact Nucleic Acid Isolation Kit I	3730964001	Roche, Switzerland
QIAquick Gel Extraction Kit	28704	Qiagen, Germany

3.7. Software, Online Databases and Bioinformatics Tools

The open-source NGS and bioinformatics software, tools and online databases used in whole exome sequencing analysis are given in Table 3.7.

Table 3.7. Open-source bioinformatics software, bioinformatics tools and electronic databases.

Software / Tool / Database	Description
The Reference Sequence Database (O'Leary <i>et al.</i> , 2016)	A reference genome database for vertebrates and other eukaryotic species
GeneCards (Weizmann Institute of Science, 2016)	A database of gene-related genomic, transcriptomic, proteomic, clinical, and functional information
Online Mendelian Inheritance in Man (OMIM) (McKusick-Nathans Institute of Genetic Medicine, n.d.)	An online catalog of human genes and associated disorders
ExAC (Lek <i>et al.</i> , 2016)	Exome Aggregation Consortium
1000 Genomes (Auton <i>et al.</i> , 2015)	A human genetic variation catalog
dbSNP (Sherry <i>et al.</i> , 2001)	A catalog of short genetic variation
NHLBI GO Exome Sequencing Project	A database of 6500 human exome results for NHLBI Exome Sequencing Project
Polymorphism Phenotyping v2 (PolyPhen2) (Adzhubei <i>et al.</i> , 2010)	Straightforward physical and comparative considerations prediction
SIFT (P. C. Ng and Henikoff, 2003)	Based on the degree of conservation of amino acid residues
Annovar (K. Wang <i>et al.</i> , 2010)	Functional annotation of genetic variations
FastX Toolkit (HannonLab, n.d.)	Command line tool kit to manipulate fastq files
Integrative Genomics Viewer (IGV) (IGV (Integrative Genomic Viewer), 2013)	Visualization tool for interactive exploration of large, integrated genomic datasets
Genome Analysis Toolkit (GATK) (McKenna <i>et al.</i> , 2010)	Focus on variant discovery and genotyping as well as strong emphasis on data quality assurance

Table 3.7. Open-source bioinformatics software, bioinformatics tools and electronic databases (cont.).

Software / Tool / Database	Description
Picard-tools (Broad Institute, n.d.)	Java-based command-line tools to manipulate SAM / BAM files
SamTools (H. Li <i>et al.</i> , 2009)	Manipulating aligned reads in the SAM / BAM format, including sorting, merging and indexing
Vcftools (Danecek <i>et al.</i> , 2011)	Program package for analysis and manipulation of vcf files
PLINK (Purcell <i>et al.</i> , 2007)	Genome data analysis toolset
R (R Development Core Team, 2011)	Software for statistical computation and graphics
Varsifter (Teer <i>et al.</i> , 2012)	A Java program designed to display, sort, filter of high throughput data
SnEff (Cingolani <i>et al.</i> , n.d.)	Toolset for genetic variant annotation and analysis
Biorad Image-Lab v50	GelDoc visualization software
CLC sequence viewer	Sanger sequencing analysis software
BioMart (Smedley <i>et al.</i> , 2015)	Web-based tool for extraction of data

4. METHODS

4.1. DNA Isolation from peripheral blood samples

Peripheral blood was collected from all participants and stored in EDTA tubes. Genomic DNA was isolated from blood using MagNA Pure Compact Instrument, a system for extracting nucleic acids from tissue or blood samples. The DNA quality was tested by running the sample on a 1% agarose gel with GelRed, an intercalating agent used for visualization of the DNA in gels. The concentration of DNA was measured by using a NanoDrop spectrophotometer at 260 nm optical density. The absorption values at 260/230 nm and 260/280 nm were evaluated and documented to detect chemical and protein contaminations in isolated DNA.

4.2. Whole Exome Sequencing

Whole exome sequencing was performed on the DNA sample of the patients, affected family members or/and parents in each family either in collaboration with UMASS or the samples were outsourced to four different companies, The Scientific and Technological Research Council Of Turkey (Tübitak, Istanbul, Turkey), The Center of Applied Genomics (TCAG, Toronto, Canada) and Macrogen (Korea) and DNA Laboratories (Istanbul, Turkey).

Whole exome sequencing requires three steps prior to the massively parallel sequencing reaction. First, the genomic DNA library, which is obtained by random fragmentation of the genomic DNA, must be prepared. Second, the genomic DNA library is enriched for coding regions by using commercially available kits. The final step is the verification of library quality and enrichment.

An overview of the whole exome sequencing process is shown in Figure 4.1. Initially, the DNA is sheared into fragments of several lengths and the single stranded overhangs are converted into blunt ends. Then, fragments of 250-300 bp in size are selected from gel and adapters are ligated to both ends of the fragments. These serve as primer binding sites for enrichment via PCR to create a genomic DNA library which is then mixed with biotinylated

RNA-based capture probes targeting the coding sequences. Captured fragments are isolated via streptavidin beads and a second PCR is performed to amplify the enriched library. Real-time PCR is used to confirm the success of the exon enrichment.

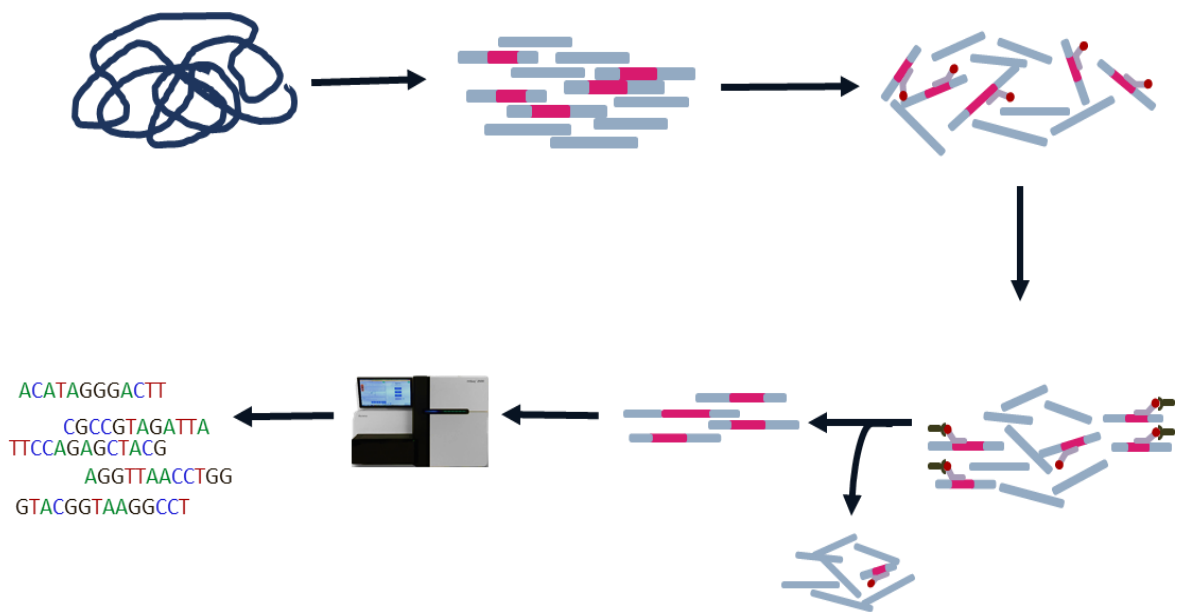


Figure 4.1. Experimental steps of whole exome sequencing

The enriched DNA library is hybridized to a flow cell which is a solid support containing the forward and reverse primers. The templates are amplified by bridge amplification and sequenced via reversible chain termination method. The 3' prime ends of the nucleotides are masked with a reversible chain terminator containing a fluorescent label, specific to each type of nucleotide. The labeled nucleotides are incorporated on the templates per cycle. The reaction is halted upon the addition of a single nucleotide, the unbound nucleotides are washed away and the signal is imaged. Then, the terminator is removed and a new cycle begins with the incorporation of a new nucleotide.

4.3. Whole Exome Sequencing Data Analysis

All steps of bioinformatics data analysis starting from cleaning of the raw reads, alignment to the reference genome, variant calling, and annotation were performed in our laboratory. All analyses were executed on Bio-Linux 7, which is an adapted version of

Ubuntu 14.04 operating system with already installed bioinformatics packages. The commands with preferred parameters of all data processing steps are given in Appendix A.

4.3.1. Alignment and Variant Calling

Raw paired-end sequence reads were subjected to quality control using the FastQC program. Fastq files were cleaned by Fastx toolkit to remove sequence reads with low quality to prevent mapping errors, which would cause false negatives in downstream analyses. The remaining high quality reads were aligned to the human reference genome GRCh37 plus the decoy using sampe algorithm of Burrows-Wheeler Aligner (BWA) software producing sam files (Sequence Alignment Map) (Heng Li and Durbin, 2009). The aligned reads were converted to binary format (bam files) using the Samtools software package (H. Li *et al.*, 2009). Duplicate removal and indexing was performed by Picard tools, followed by realignment around indels and base quality score recalibration by Genome Analysis Toolkit (GATK) of Broad Institute (McKenna *et al.*, 2010). Genomic vcf files (gvcfs) were obtained from fully processed bam files by the HaplotypeCaller tool of GATK and identification of single nucleotide variants (SNV) and small indels was performed using the GenotypeGVCFs. Further processing of raw vcf files was done according to GATK Best Practices recommendations (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013).

Sample quality of the sequencing data was checked by several metrics in the final vcf files. The depth of coverage per sample and missing genotype rate of individuals were obtained with Vcftools (Danecek *et al.*, 2011). The ratio between transitions and transversions (Ts/Tv) per individual was calculated by SnpEff (Cingolani *et al.*, 2012).

4.3.2. Homozygosity Mapping

Homozygosity mapping was performed for the families displaying a recessive inheritance pattern by PLINKv1.9 (Purcell *et al.*, 2007). Initially, fully processed vcf files were converted to PLINK hard calls with setting the genotype quality lower threshold to 35. Then, the variants in linkage disequilibrium were pruned with r^2 threshold 0.2. Runs of homozygosity (ROHs) were detected for each family using the --homozyg option with parameters in Table 4.1, and the distribution graph of ROHs were drawn in R.

Table 4.1. Homozygosity mapping parameters of PLINK

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

4.3.3. Confirmation of relationships and principal component analysis

Pi-hat scores were calculated to check relationships of the family members to ensure further analysis which was based on transmission in the pedigree. This is achieved by --genome parameter of PLINK. Principal component analysis was applied by using multidimensional scaling (--mds-plot) option of PLINK to determine the heterogeneity of population structure.

4.3.4. Variant Annotation and Prioritization

Functional annotation of the variants was done by ANNOVAR (K. Wang et al., 2010), following the quality control of the data. Variations resulting in changes on protein sequences were further filtered using minor allele frequencies from several databases [dbSNP138, 1000 Genomes (October 2014 release), Lung and Blood Institute (NIHLBI) Exome Sequencing Project (ESP) 6500 exomes, The Exome Aggregation Consortium (ExAC)] by excluding all variants present in the population with a frequency greater than 1%. Filtration was done by VarSifter, Java-based software to parse vcf files (Teer et al., 2012). Clinical information to associate suspected genes with the phenotype of each patient was obtained from the Online Mendelian Inheritance in Man (OMIM) database.

Variants were classified into three different categories: (i) a confirmed pathogenic variant, that is reported to be pathogenic in previous studies; (ii) possible novel pathogenic variants detected in a gene, previously reported to be associated with ataxias, predicted to

affect protein function/structure and fully segregate with the pedigree; (iii) variants of unknown significance which are possible pathogenic variants found in a gene that is not associated with ataxias but with another neurological disorder. SIFT, PolyPhen-2, and MutationTaster scores were used to analyze possible effects of suspected variants on the protein although they were not considered to eliminate the variants. Further, GERP and PhyloP scores were used for evolutionary conservation rate of the mutation site.

4.4. PCR to validate exome analysis results

4.4.1. PCR conditions

The presence and segregation of the suspected pathogenic variants from WES data analysis were confirmed by PCR-based Sanger sequencing. Primers specific to the regions containing the mutation were either retrieved from the literature or designed by the Primer3 software (Rozen S., & Skaletsky H., 2000) and were later checked at the UCSC in silico PCR utility for their specificity in the genome. The lyophilized oligos were synthesized and purchased from Sentromer, Turkey. Each primer was dissolved in the suggested amount of dH₂O and 10 μ M dilutions were prepared to be used for further PCR experiments.

The PCR products were analyzed on a 2 % agarose gel using GelRed fluorescent DNA stain as intercalating agent to visualize the products under UV.

4.4.2. Agarose gel electrophoresis

A 2 % agarose gel was prepared using 0.5X TBE buffer and 0.3 % GelRed fluorescent DNA stain and placed into an electrophoresis chamber containing 0.5X TBE. Three volumes of PCR product were mixed with one volume of 6X loading dye, which was then loaded into the gel. The products were run approximately for an hour at 120 V and visualized under UV light using the GelDoc gel documentation system.

PCR products were outsourced to MacroGen Inc., for Sanger sequence analysis and data was then analyzed using the version 6.7 of CLC Main Workbench software.

5. RESULTS

Whole exome sequencing was applied to affected and unaffected members of 47 distinct families. Quality of sequencing data was checked and homozygosity mapping was performed for families displaying an autosomal recessive inheritance. Finally, annotation and filtration of fully processed family vcf files were completed accordingly for both dominant and recessive pedigrees.

5.1. Sequencing analysis metrics

Sample quality control was performed by calculating mean sequencing depth, missingness per sample and Ts/Tv ratios. Distribution of these in families with confirmed mutations are shown in Figures 5.1 – 5.3; members of the same family are represented with the same color. The values can be found in Appendix B. Quality of sequencing data was calculated for the sites manifested in the bed files of the exome enrichment kits. Each patient was evaluated in comparison with his / her family members.

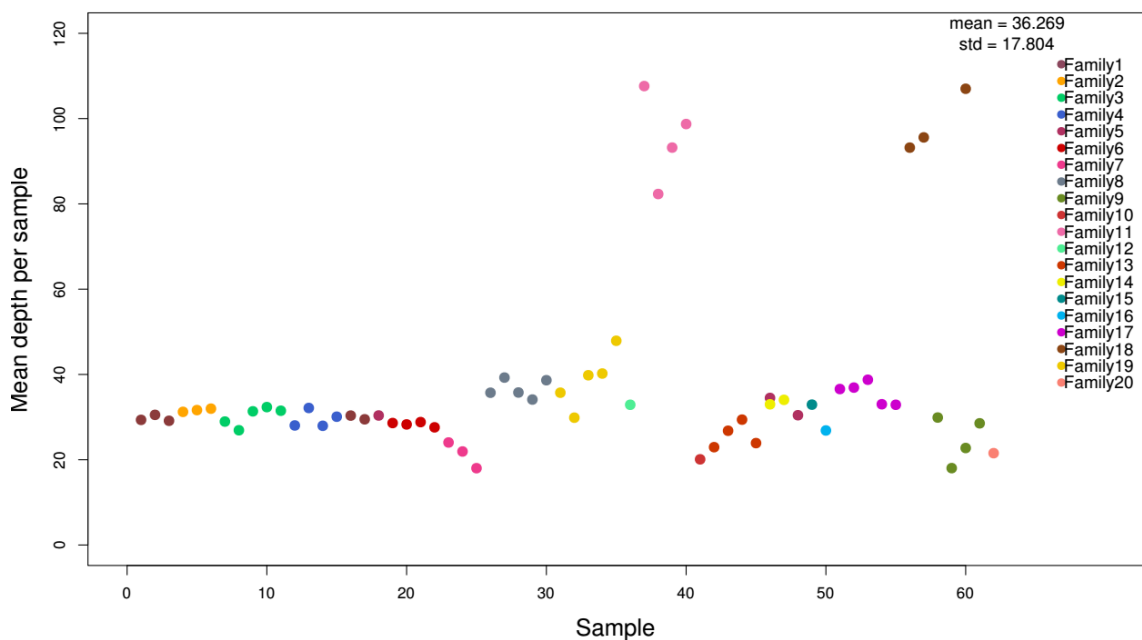


Figure 5.1. Mean depth of coverage per sample (for families solved by WES).

The majority of the samples had a depth of coverage between 30-50 X with a mean of 38.65 (Figure 5.1). Mean depth was higher for all members of Family 10 and Family 17 as compared to the rest of the cohort. However, no individual was excluded from the analysis based on coverage, since there was no difference in mean depth within family members.

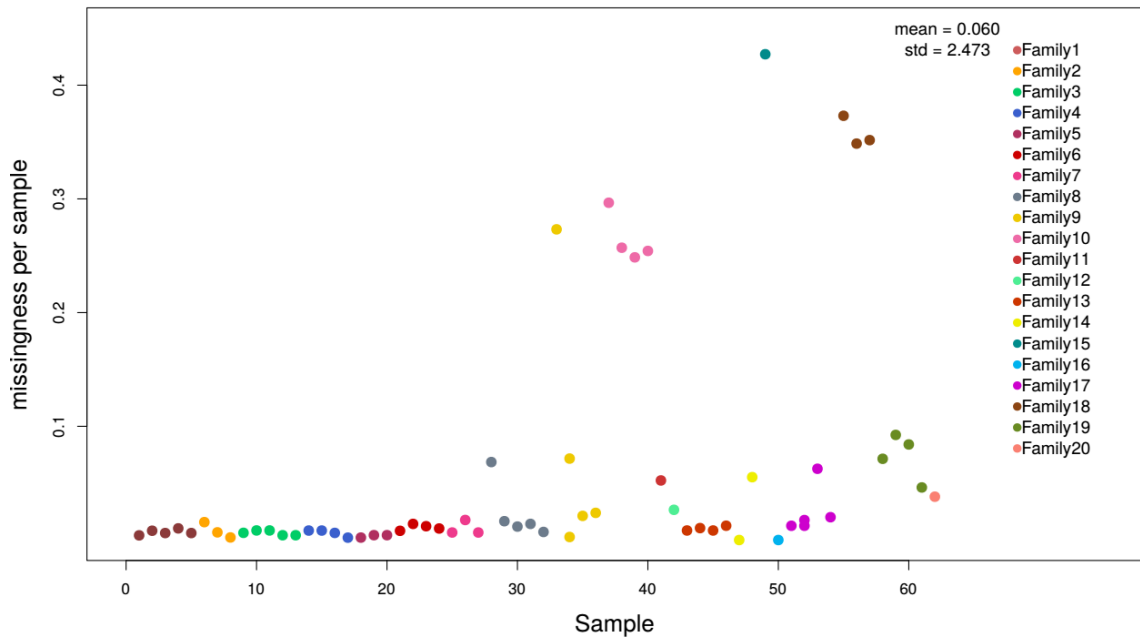


Figure 5.2. Rate of missingness per sample (for families solved by WES).

The mean missingness ratio of individuals was computed as 0.049 with a standard deviation of 0.101. A total of nine among 64 samples, including all members of Family 10 and Family 17, showed high values of missingness considering the standard deviation of the cohort. Additionally, the sample of Family 15 displayed the highest missingness value; however, it was kept for downstream analysis since there were no other available participants.

The average of the T_s/T_v ratio per sample was 2.47 with a standard deviation of 0.268, ranging from 1.701 to 2.723. All members of Family 17 and 18 displayed lower T_s/T_v ratio from the standard deviation of the cohort.

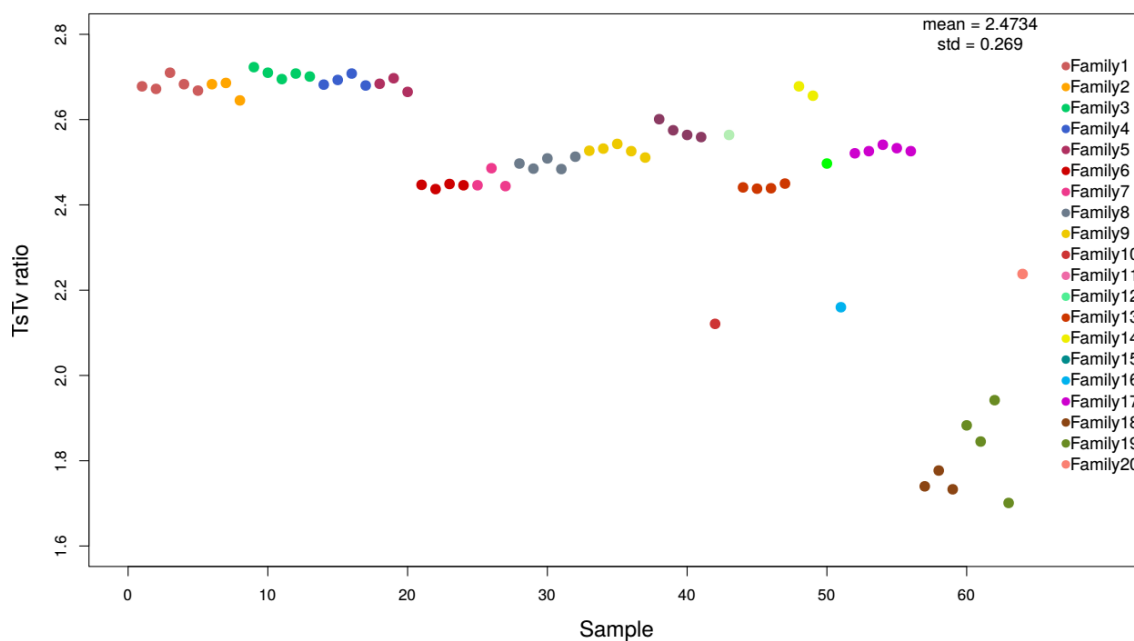


Figure 5.3. Transition / Transversion ratios per sample (for families solved by WES).

5.2. Population structure

Multidimensional scaling of IBS distances were checked for the presence of clusters in the study cohort. First two principal components were used for the analysis. The majority of the participants formed one cluster (Figure 5.4). There are a total of five individuals stayed out of the rest of the cohort.

5.3. Whole exome sequencing data analysis

In this study, 18 different mutations in a total of 12 distinct genes were identified and successfully validated for available members of 20 families. The number of variants per family after each filtering step is reported in Table 5.1. In addition, confirmed pathogenic variants are listed in Table 5.2, together with the inheritance patterns of the families, initial-final diagnoses and the associated diseases that are retrieved from the OMIM database. Quality, read depth, conservation scores, 1000 Genomes and ExAC frequencies for all causative variations are compiled in Table 5.3 and prediction tool scores in Table 5.4.

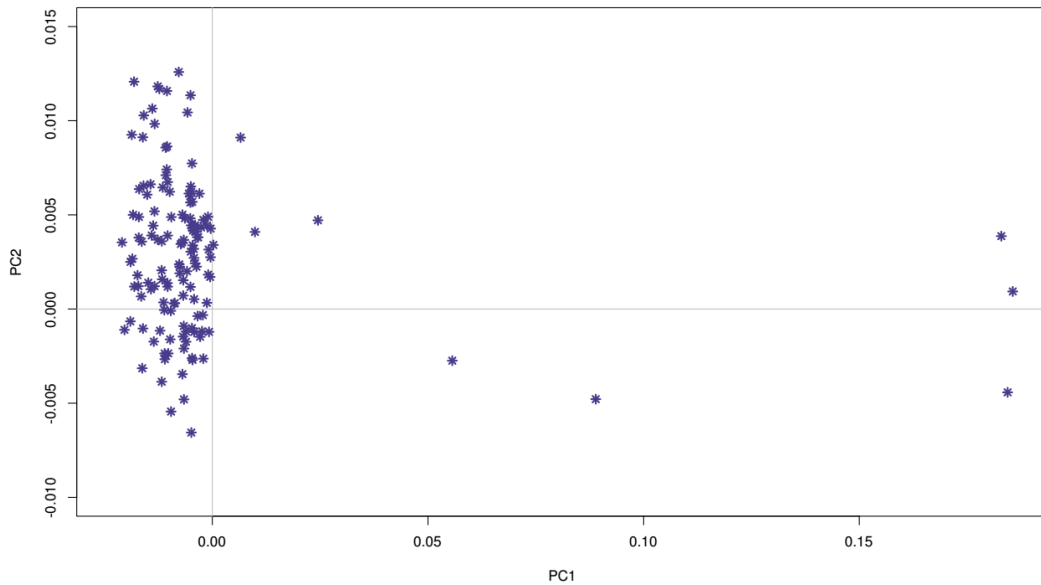


Figure 5.4. Multidimensional scaling plot of the participants

A representation of families with the mutated genes is shown in Figure 5.4. Heterozygous missense mutations in *AFG3L2* and *KIF1B* genes were identified in two families showing an autosomal dominant inheritance pattern (Figures 5.6 and 5.7). The five had homozygous mutations in the *SETX* gene. Additionally, pathogenic variations in *SACS*, *CAPN1*, *SPG7* and *ADCK3* were identified twice in unrelated families for each gene.

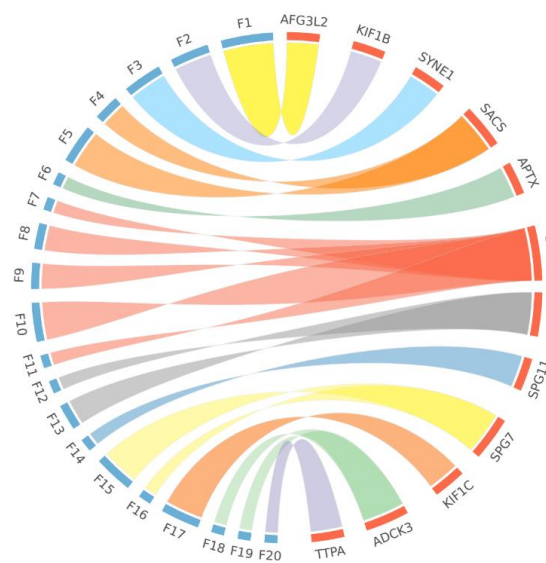


Figure 5.5. Circos representation of families and the mutated genes. Families are shown in blue, genes are indicated in red. (F: family)

Furthermore, mutations in *SYNE1*, *APTX*, *SPG11*, *KIF1C* and *TTPA* were identified in the remaining families.

5.3.1. Families with an autosomal dominant inheritance pattern

5.3.1.1. Spinocerebellar ataxia type 28: *AFG3L2*

Family 1:

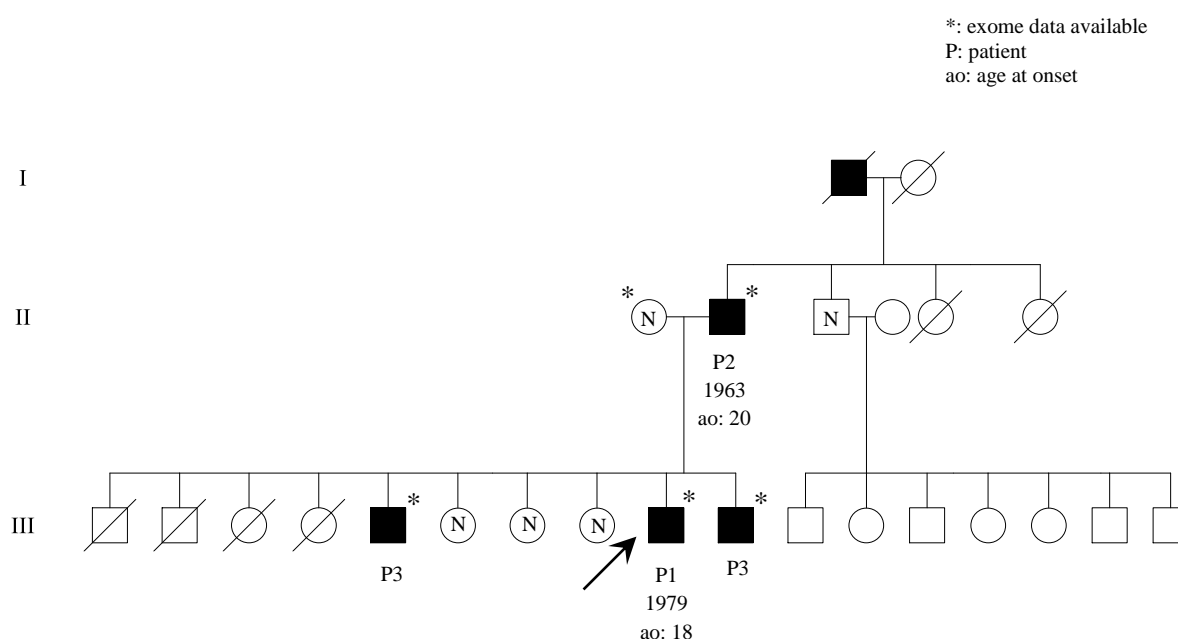


Figure 5.6. Family 1: segregation of the missense variant in the *AFG3L2* gene.

The index case was subjected to WES together with affected father, siblings, and the unaffected mother. First; intronic, intergenic, UTR and noncoding variants were removed from the file, followed by filtration by the inheritance pattern, and novel / rare variants were selected. Among the remaining 81 variants of Family 1, a missense variation in the *AFG3L2* gene was identified. The variant was not present in dbSNP, ExAC databases and our in-house control of neurologically healthy individuals. The segregation of the Pro688Thr missense variant in the *AFG3L2* gene is shown in Figure 5.6. All three affected individuals were found to be heterozygous for the mutation while the unaffected mother, uncle and three sisters of the index case were shown to be free of mutated allele. The affected grandfather of the proband was not available to us.

5.3.1.2. Charcot-Marie-Tooth disease type 2A: *KIF1B*

Family 2:

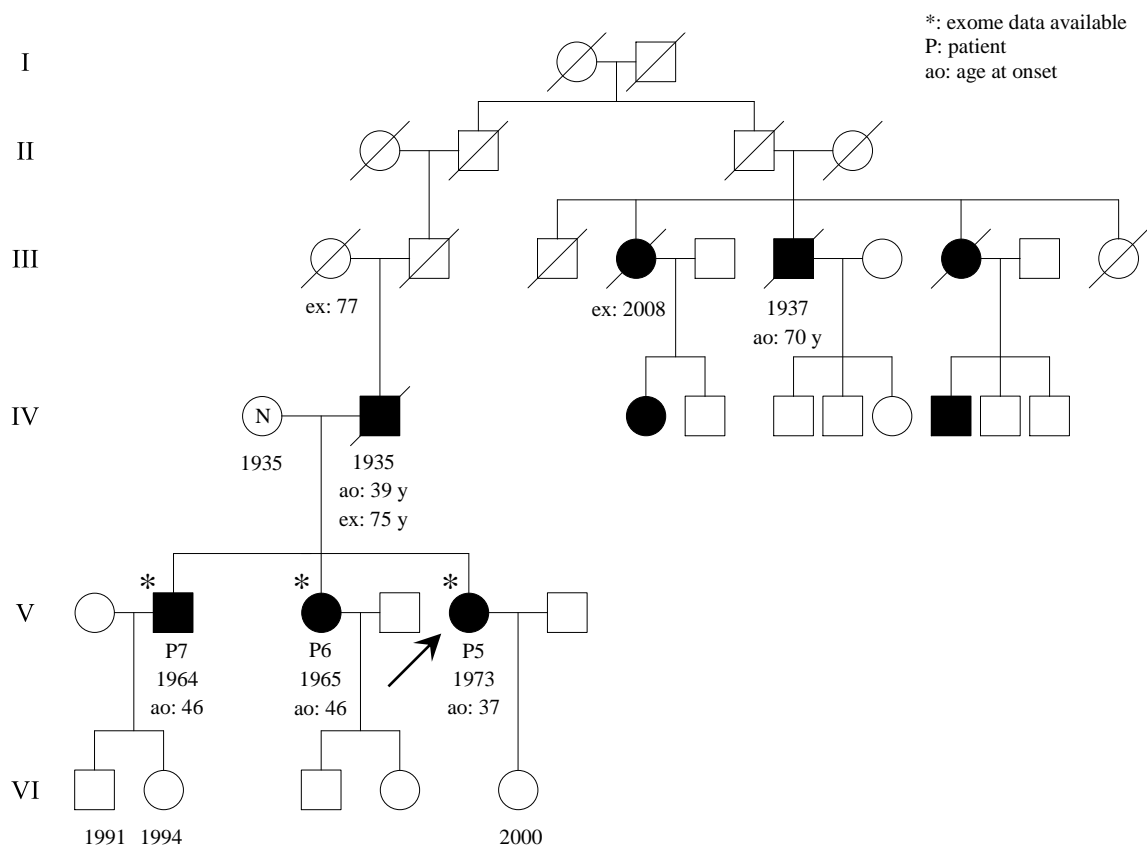


Figure 5.7. Family 2 with the segregation status of the *KIF1B* heterozygous variant.

In family 2, the total number of exonic heterozygous variants shared by siblings with a minor allele frequency of < 1% was 177. The Arg1571Gln missense mutation in the *KIF1B* gene was detected by focusing on variants that were not present in dbSNP, and in our in-house controls. Sanger sequencing confirmed the presence and segregation of the candidate variant in the family (Figure 5.7). Consequently, all affected siblings were shown to carry the mutation in heterozygous state and the unaffected mother was shown to have the wild type sequence.

5.3.2. Families with autosomal recessive inheritance pattern

Homozygosity mapping was performed for families showing an autosomal recessive inheritance pattern in order to reduce the target area to be searched for disease cause. The

outcome is displayed in Figures 5.8 – 5.25. In each plot, the mutated gene is depicted in orange.

5.3.2.1. Autosomal recessive cerebellar ataxia type 1: *SYNE1*

Family 3:

According to the runs of homozygosity in Family 3, 21 homozygous regions were found to be shared by all affected siblings (Figure 5.8a). Among the variations within the homozygous regions, a stop-gain mutation (Arg6755Ter) in the *SYNE1* gene was detected. The variant was absent in our in-house controls and has not been reported in homozygous state in ExAC database. Sanger sequencing confirmed the presence of the variant in homozygous state in all affected individuals and in heterozygous state in the parents and the unaffected siblings (Figure 5.8b).

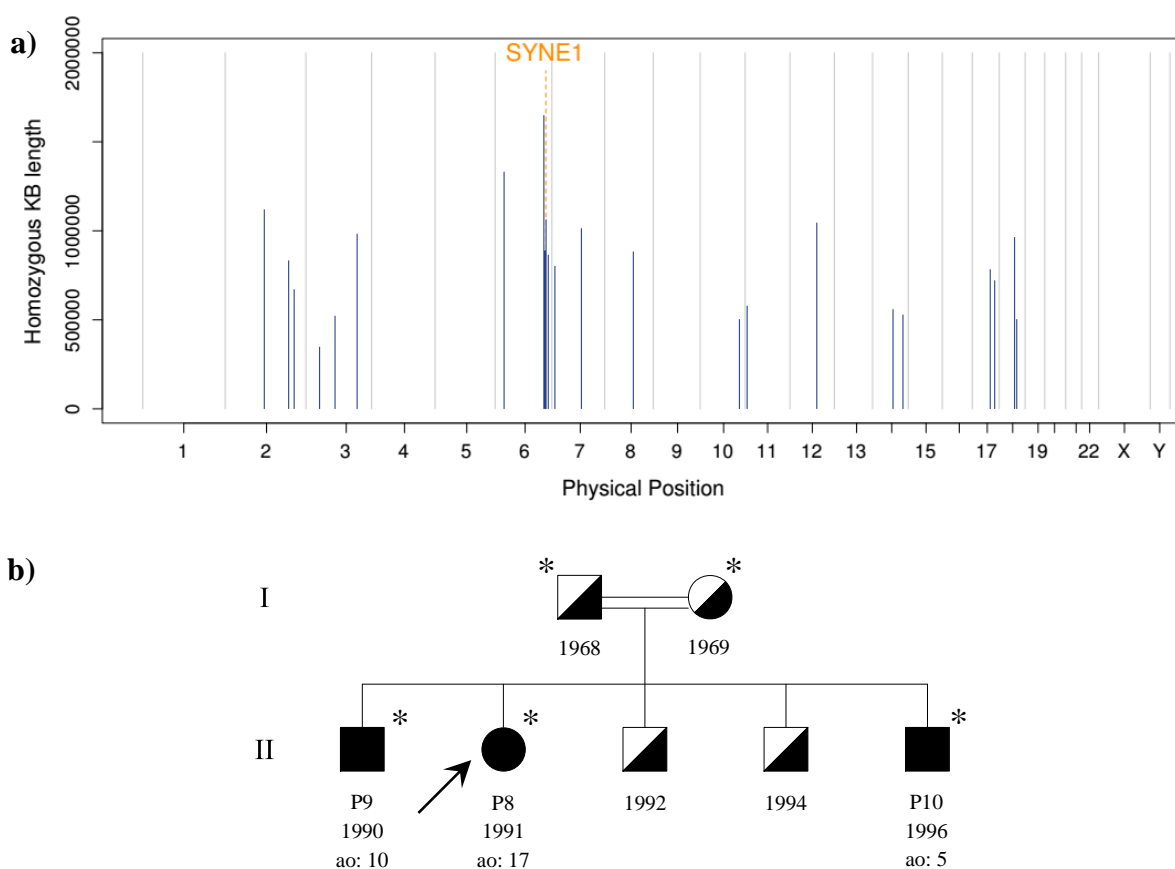


Figure 5.8. Homozygosity mapping plot (a) and the segregation of the *SYNE1* variant in the pedigree (b).

5.3.2.2. Autosomal recessive spastic ataxia of Charlevoix-Saguenay: SACS

One frameshift and one stop-gain mutations in the *SACS* gene, both causing a truncated protein, were identified in two distinct families upon WES and homozygosity mapping (Figure 5.9 and 5.10).

Family 4:

Figure 5.9a represents the runs of homozygosity in Family 4 pointing to only one region in chromosome 13. Homozygosity mapping was performed despite the absence of consanguinity, due to marked recessive inheritance pattern in the family. In addition,

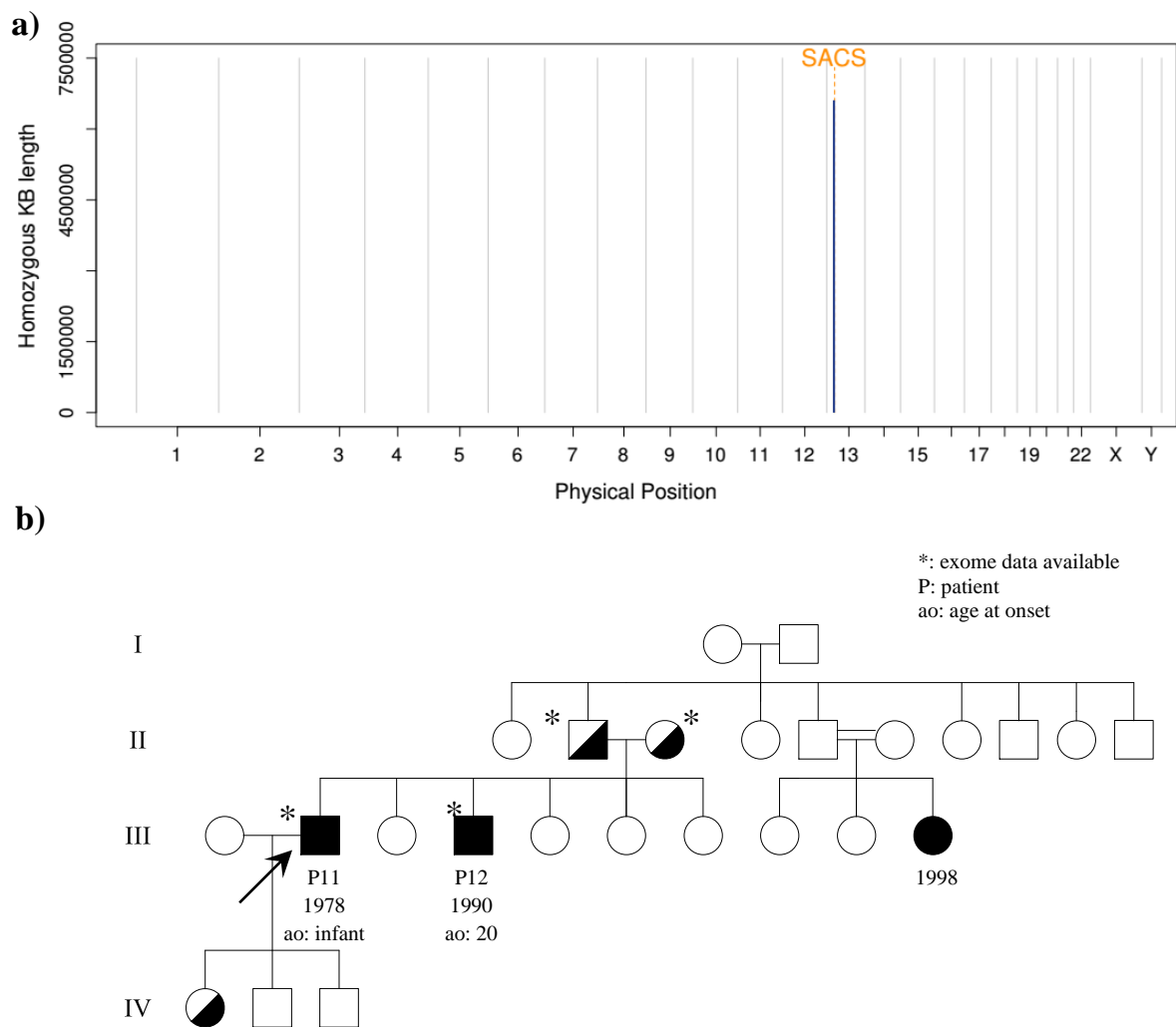


Figure 5.9. Runs of homozygosity in affected siblings of Family 4 (a) family tree with the segregation of the *SACS* variant (b).

compound heterozygous variants were also considered in the analysis. As a result, the homozygous AAGAA deletion was identified within the homozygous stretch in the *SACS* gene. The deletion of five nucleotides caused a frameshift at position 4308 and a premature stop codon after 21 amino acids. The segregation of the variant was validated in available family members (Figure 5.9b).

Family 5:

In Family 5, numerous shared homozygous regions were revealed throughout the genome as a result of homozygosity mapping. However, only four coding variants were present in homozygous state with a minor allele frequency of < 1%. The Arg3792Ter mutation in the *SACS* gene was pinpointed as the candidate variant. Figure 5.10 represents the homozygosity mapping results and the validation status of the candidate. Accordingly,

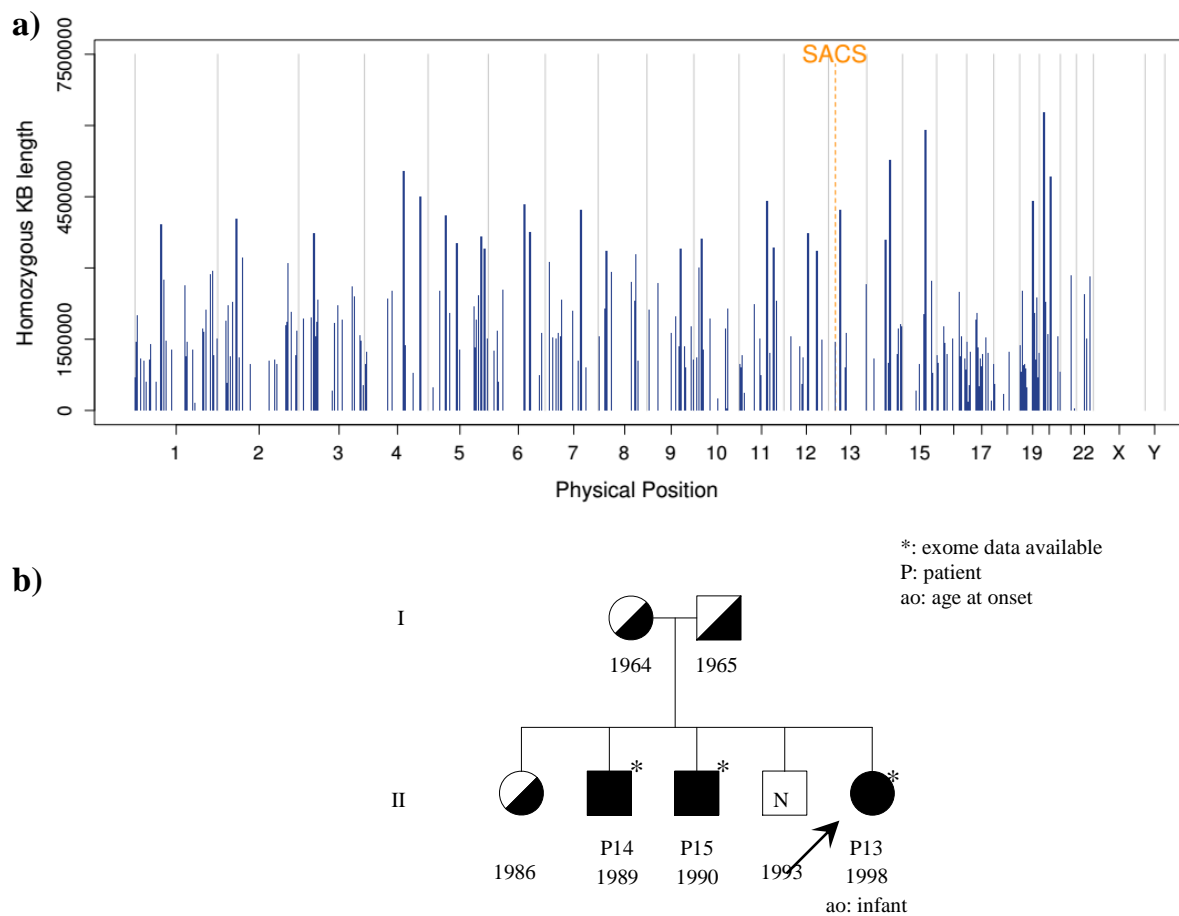


Figure 5.10. Homozygosity mapping plot of shared regions between three affected siblings (a) and the segregation of the *SACS* variant within the pedigree (b).

the affected siblings were homozygous for the mutant allele, whereas the parents and the older daughter were heterozygous; one son does not carry the mutated allele at all.

5.3.2.3. Ataxia with oculomotor apraxia type 1: APTX

Family 6:

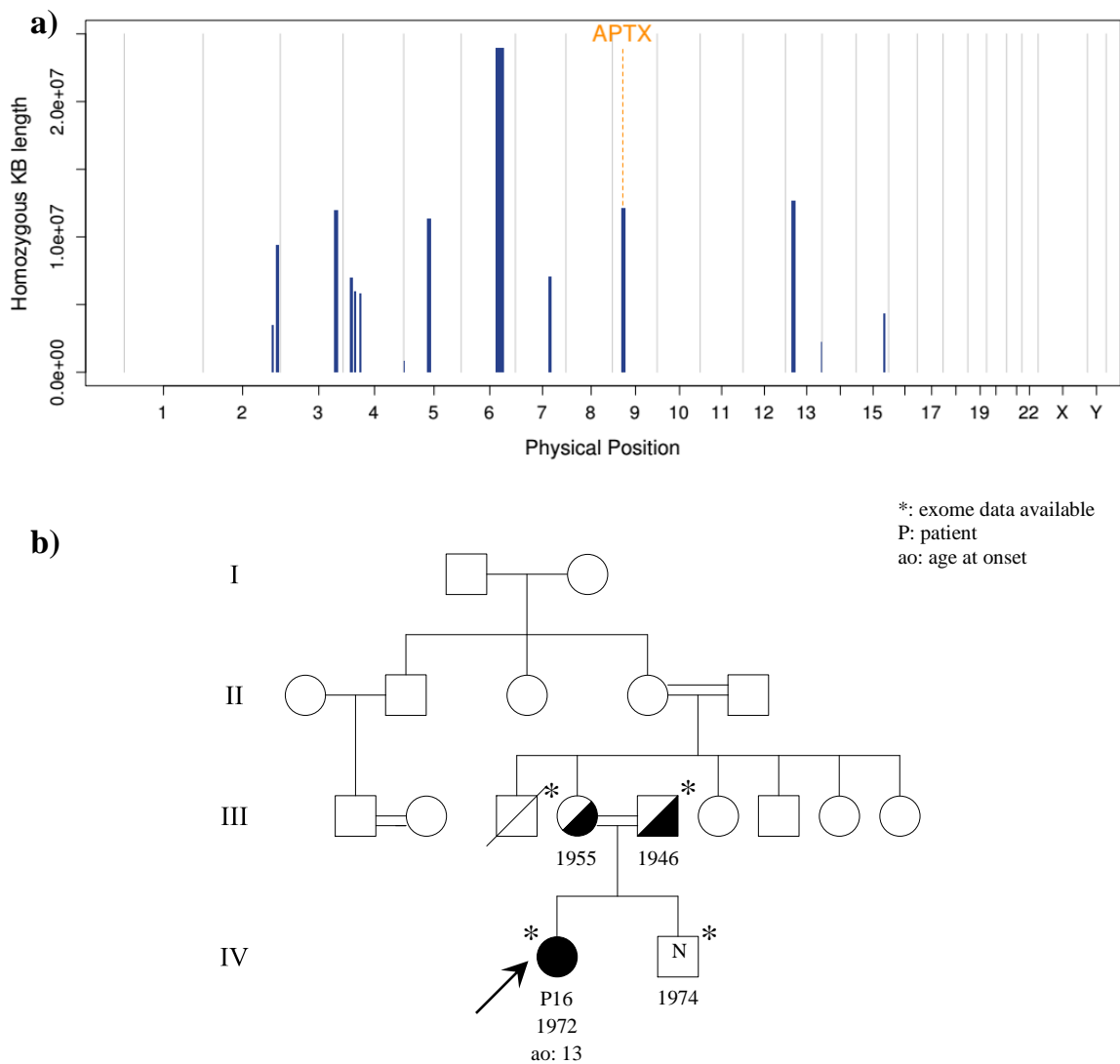


Figure 5.11. Runs of homozygosity plot (a) and the segregation of the missense *APTX* variant in Family 9 (b).

Due to inbreeding in the family, homozygosity mapping was done in Family 6 revealing 13 ROHs in the patient (Figure 5.11a). Within these regions Val190Gly variant in

the *APTX* gene was identified as a candidate. The segregation of this variation was confirmed in the family as homozygous in the patient and heterozygous in the parents, as expected. The unaffected sibling was found to carry the wild-type sequence (Figure 5.11b).

5.3.2.4. Ataxia with oculomotor apraxia type 2: *SETX*

A total of four different homozygous mutations including one splice-site, one stop-gain, one frameshift, and one missense mutations in five families were identified in the *SETX* gene upon homozygosity mapping and WES data analysis (Figures 5.12 – 5.16).

Family 7:

The proband and his parents were subjected to homozygosity mapping after WES. The splice site variant, c.5549-2A>G in the *SETX* gene was identified (Figure 5.12). Sanger

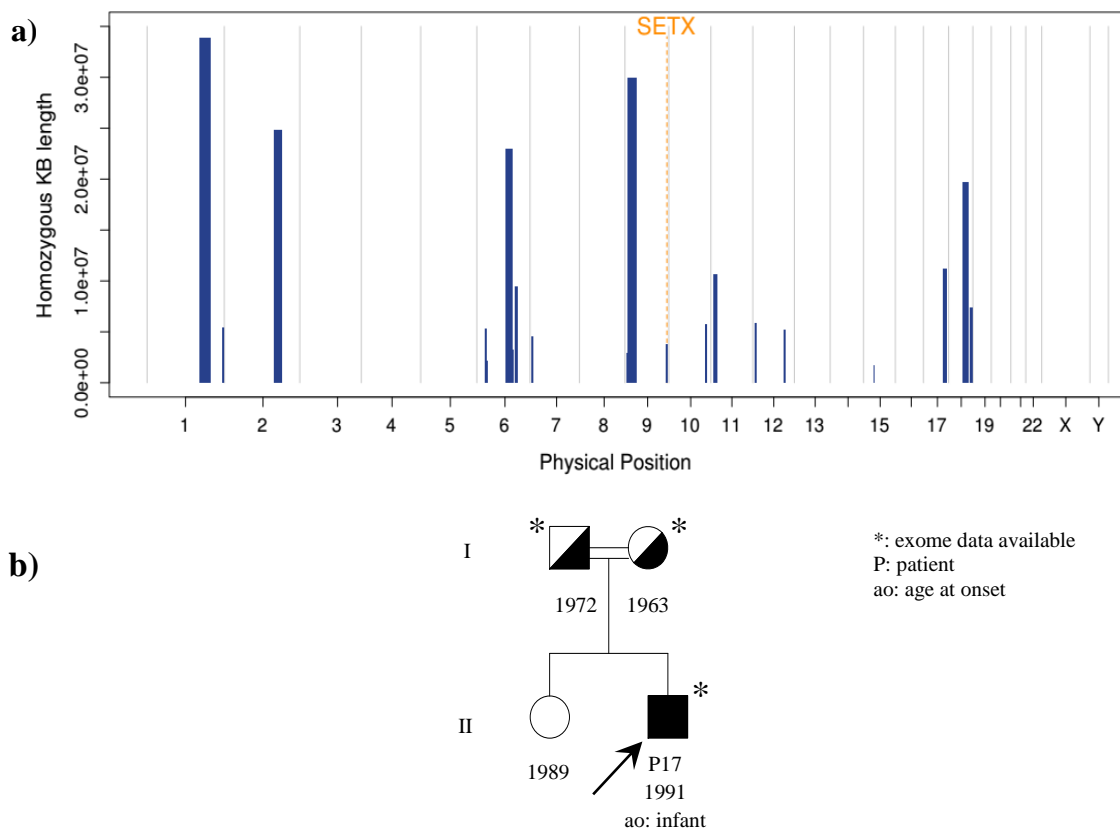


Figure 5.12. Homozygosity mapping plot of juvenile patient (a) and the segregation of the *SETX* variant in Family 7 (b).

sequencing confirmed the presence of the variant in homozygous state in the patient while in heterozygous state in the parents (Figure 5.12b).

Family 8:

Homozygosity mapping was also applied to Family 8 for two half-siblings (common mother, different fathers) with an initial clinical diagnosis of FRDA (Figure 5.13). Runs of homozygosity in the patients revealed two ROHs and eight variants. In these regions, the c.5635delG one nucleotide deletion, causing a stop codon (Val1879Ter), was detected in the

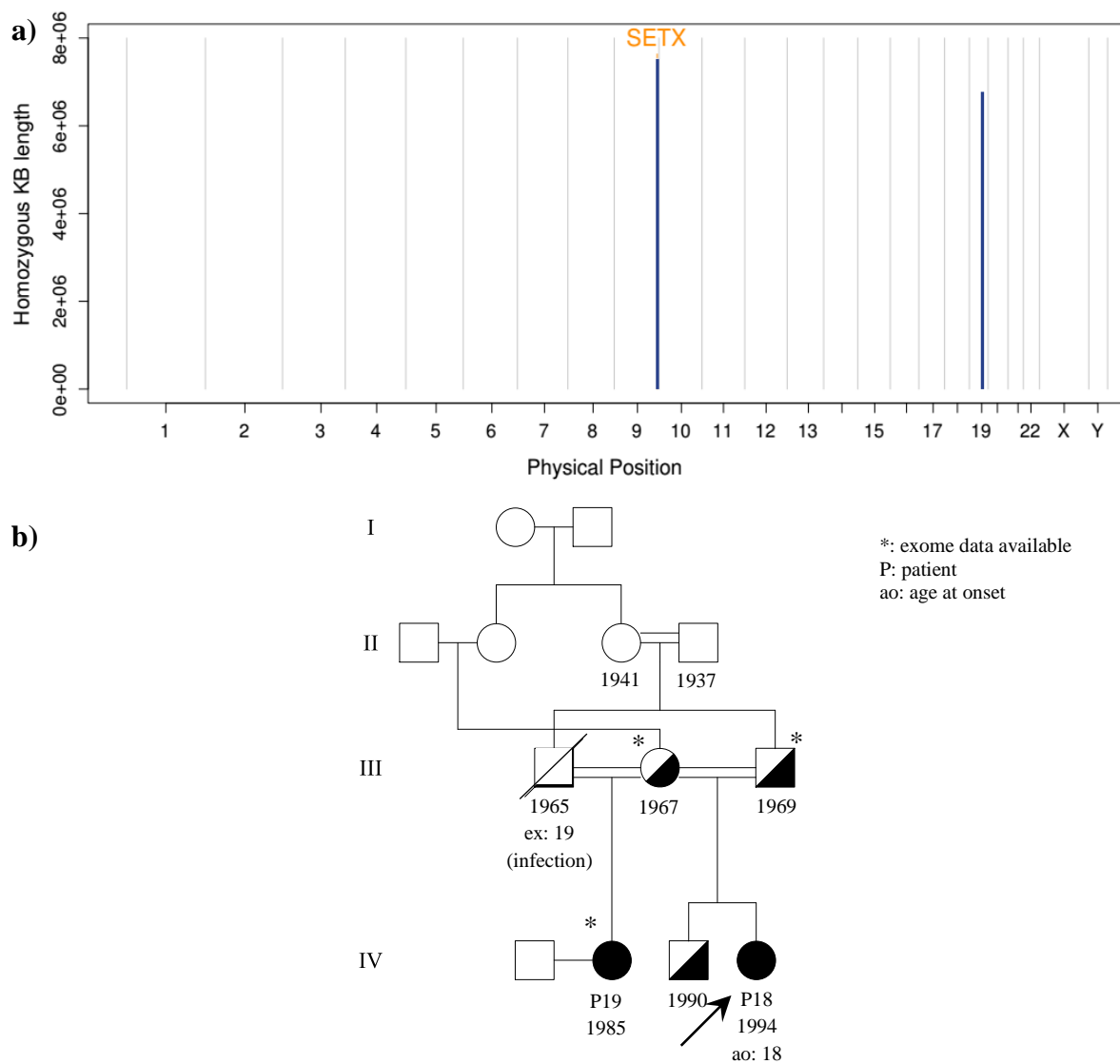


Figure 5.13. Homozygosity mapping in affected individuals (a) and the pedigree of Family 8 with the segregation of the *SETX* variant in available individuals (b).

SETX gene. The unaffected common mother, the unaffected father and the brother of the proband were shown to be heterozygous for the variant (Figure 5.13b).

Family 9:

Two daughters of the consanguineous Family 9 were subjected to WES together with the unaffected parents and older sister. Runs of homozygosity revealed a total of three regions, falling into chromosome 9 and 11 (Figure 5.14a). Among these, one nucleotide insertion in the *SETX* gene (c.5249dupT) was detected resulting in a frameshift mutation at position 1750 and leading to a premature stop codon after seven amino acids. The segregation of the variant in the family was confirmed by Sanger sequencing (Figure 5.14b).

Family 10:

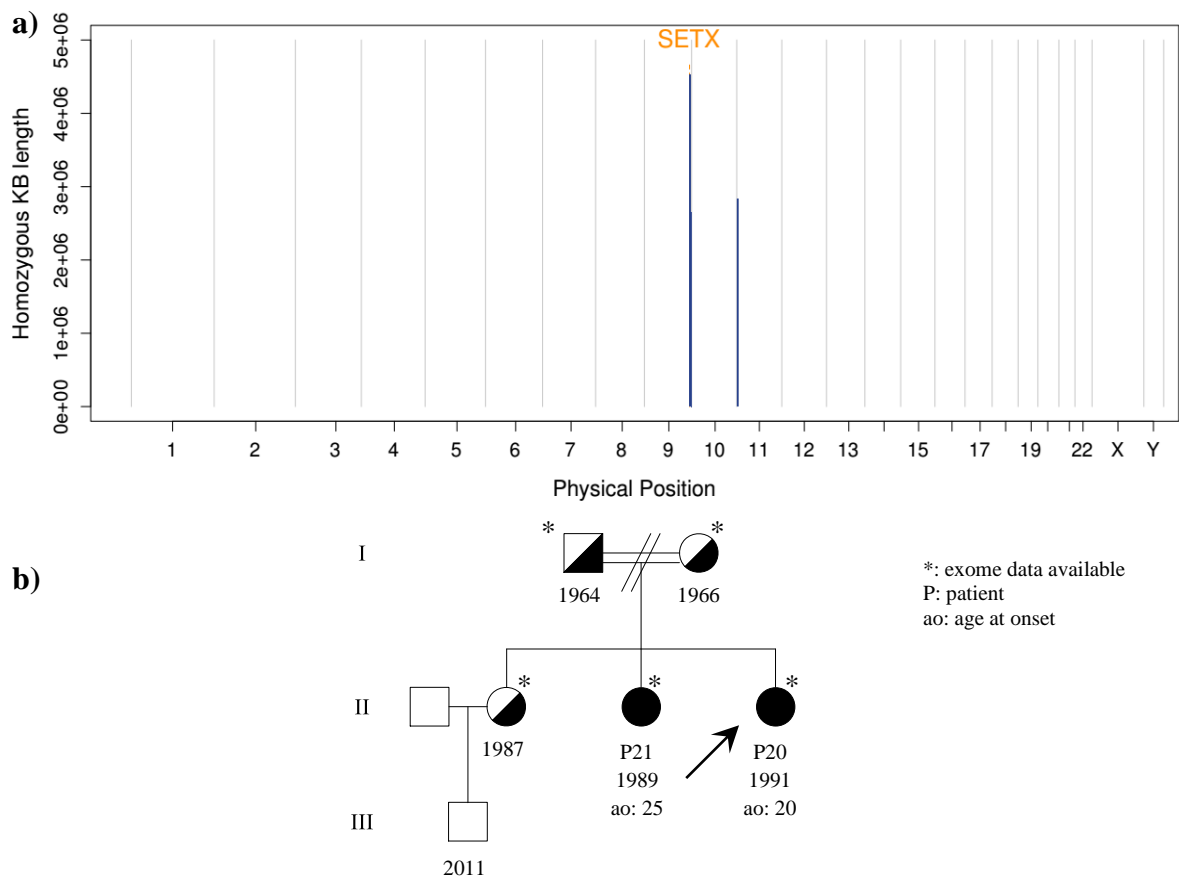


Figure 5.14. Runs of homozygosity in Family 9 (a) and the family tree with the segregation of the *SETX* variant (b).

In Family 10, three affected children were referred to our laboratory with the initial diagnoses of SBMA and ataxia. Because of consanguinity in two generations, homozygosity mapping was performed to find shared homozygous regions between the siblings (Figure 5.15a). From these sites, one nucleotide substitution in exon 19 of the *SETX* gene, replacing the methionine 2229 by threonine, was identified. Sanger sequencing confirmed the presence of the variant homozygously in all affected individuals and heterozygously in the unaffected father and children of P24 (5.15b).

Family11:

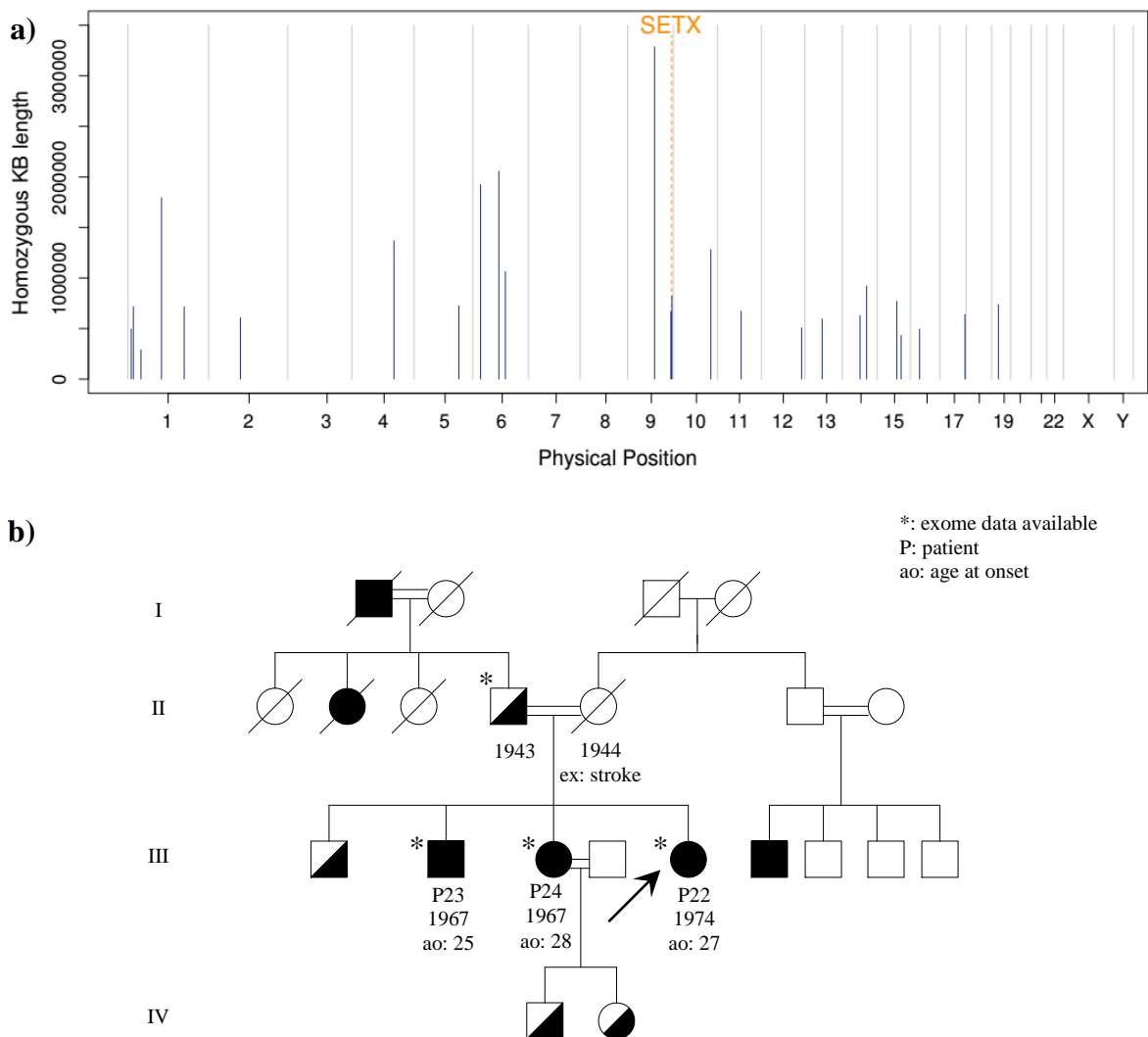


Figure 5.15. Homozygosity mapping plot of shared regions in affected siblings (a) and the segregation of the *SETX* variant in Family 10 (b).

The proband, with early-onset recessive ataxia, was subjected to homozygosity mapping following WES (Figure 5.16a). Consequently, the same frameshift mutation in the *SETX* gene (Leu1750fs*7) described in Family 9 due to a T duplication at position 5249 was detected. The presence of the variant was confirmed in the patient. The segregation will be completed by collecting DNA from affected and unaffected family members.

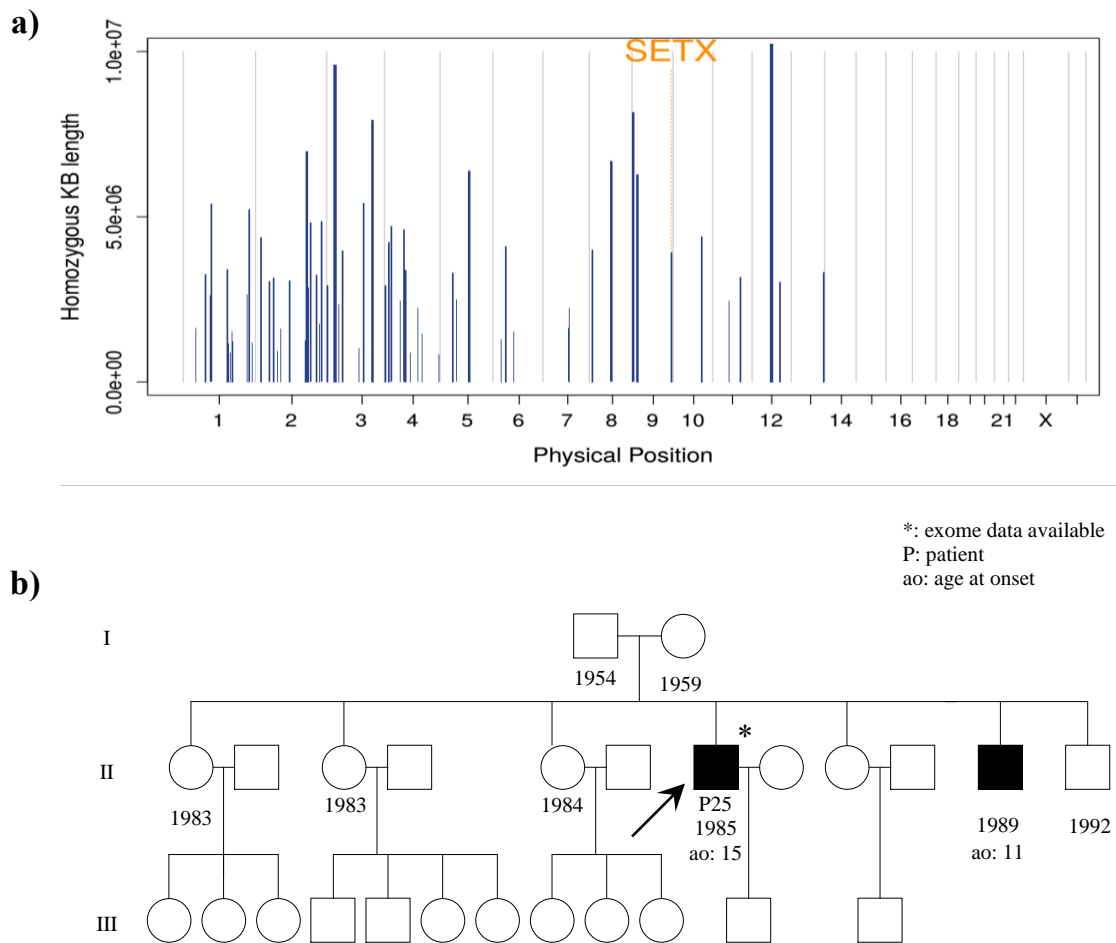


Figure 5.16. Homozygosity mapping plot of the proband of Family 11 (a) pedigree of P25 (b)

5.3.2.5. Autosomal recessive hereditary spastic paraplegia type 76: *CAPN1*

One missense and one stop-gain homozygous mutation in the *CAPN1* gene was identified in two distinct families (Family 12 and 13) by WES (Figures 5.17 and 5.18).

Family 12:

The oldest daughter of the non-consanguineous family, was referred to our laboratory with ataxia and spasticity. Figure 5.17a represents the distribution of the runs of homozygosity of the patient throughout the genome. From the intervals obtained by these regions, a homozygous missense variation in the *CAPN1* gene was identified. The variation was absent in ExAC, dbSNP and our in-house controls. Sanger sequencing confirmed the expected segregation of Gly332Arg mutation in available family members (Figure 5.17b).

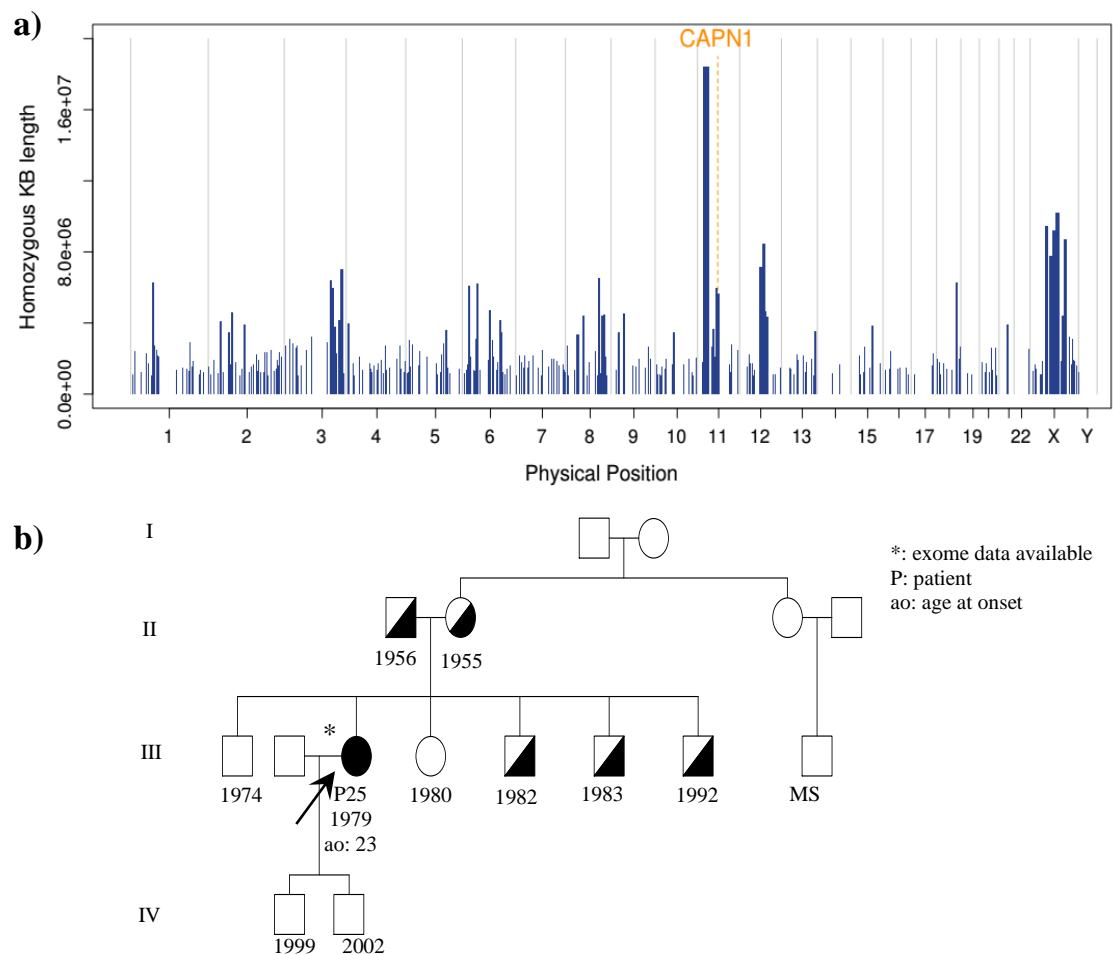


Figure 5.17. Runs of homozygosity in the index case (a) and the segregation of the missense *CAPN1* variant in Family 12 (b).

Family 13:

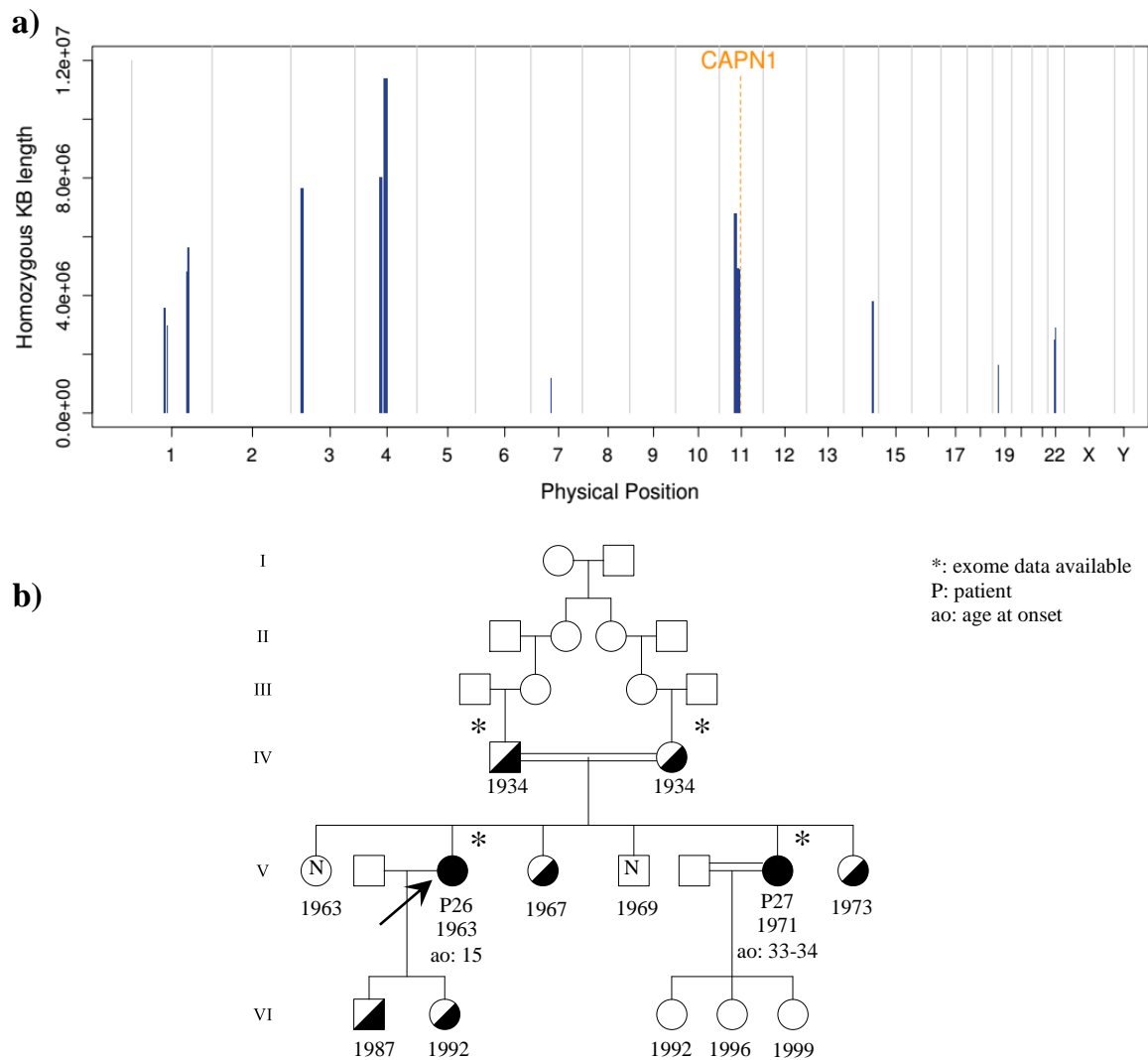


Figure 5.18. Homozygosity mapping plot of the affected siblings of Family 13 (a) and the segregation of the *CAPN1* variant in available individuals (b).

Runs of homozygosity revealed 12 homozygous regions common in two affected siblings of Family 13 (Figure 5.18a). This, led to the detection of the homozygous Trp392Ter mutation in the *CAPN1* gene. The presence of the variant and its segregation in the family with disease phenotype were confirmed by Sanger sequencing. Accordingly, the patients were homozygous for the mutation while unaffected parents and two sisters were heterozygous (Figure 5.18b).

5.3.2.6. Autosomal recessive hereditary spastic paraplegia with thin corpus callosum:

SPG11

Family 14:

The index case of Family 14 was referred to our laboratory with an initial diagnosis of ataxia, and was subjected to WES together with her available unaffected mother. A two nucleotide deletion in the *SPG11* gene causing a frameshift of the protein sequence (K656fs*11) was detected (Figure 5.19a). Sanger sequencing confirmed the presence of the variant in homozygous state in the patient and in heterozygous state in the unaffected mother

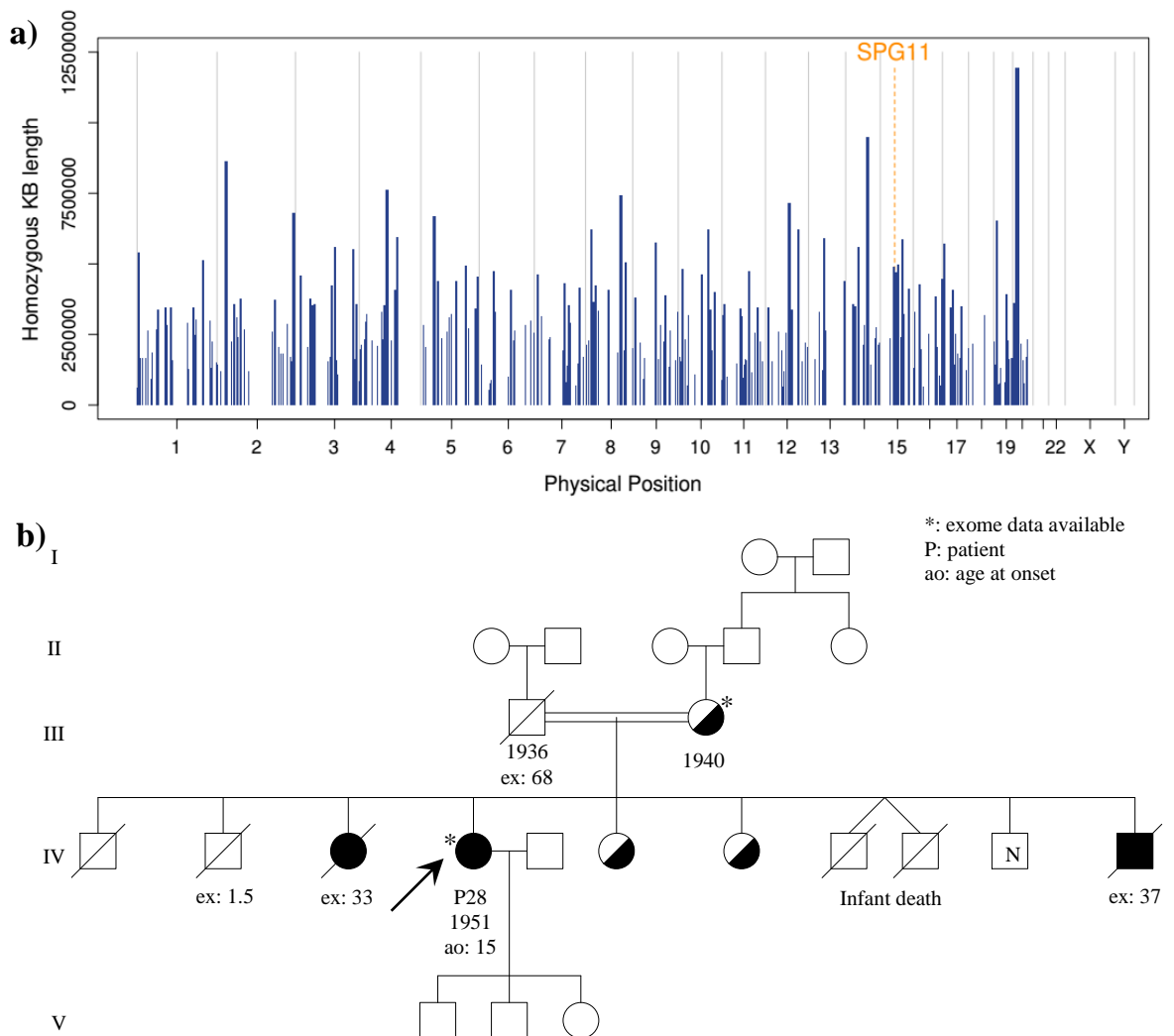


Figure 5.19. Runs of homozygosity (a) and pedigree of Family 14 with segregation of the *SPG11* variant (b).

and two sisters. The unaffected brother was found to carry the wild-type sequence (Figure 5.19b).

5.3.2.7. Autosomal recessive hereditary spastic paraplegia type 7: SPG7

Two missense homozygous mutations in the *SPG7* gene was identified in two distinct families (Family 15 and 16) by WES (Figures 5.20 and 5.21).

Family 15:

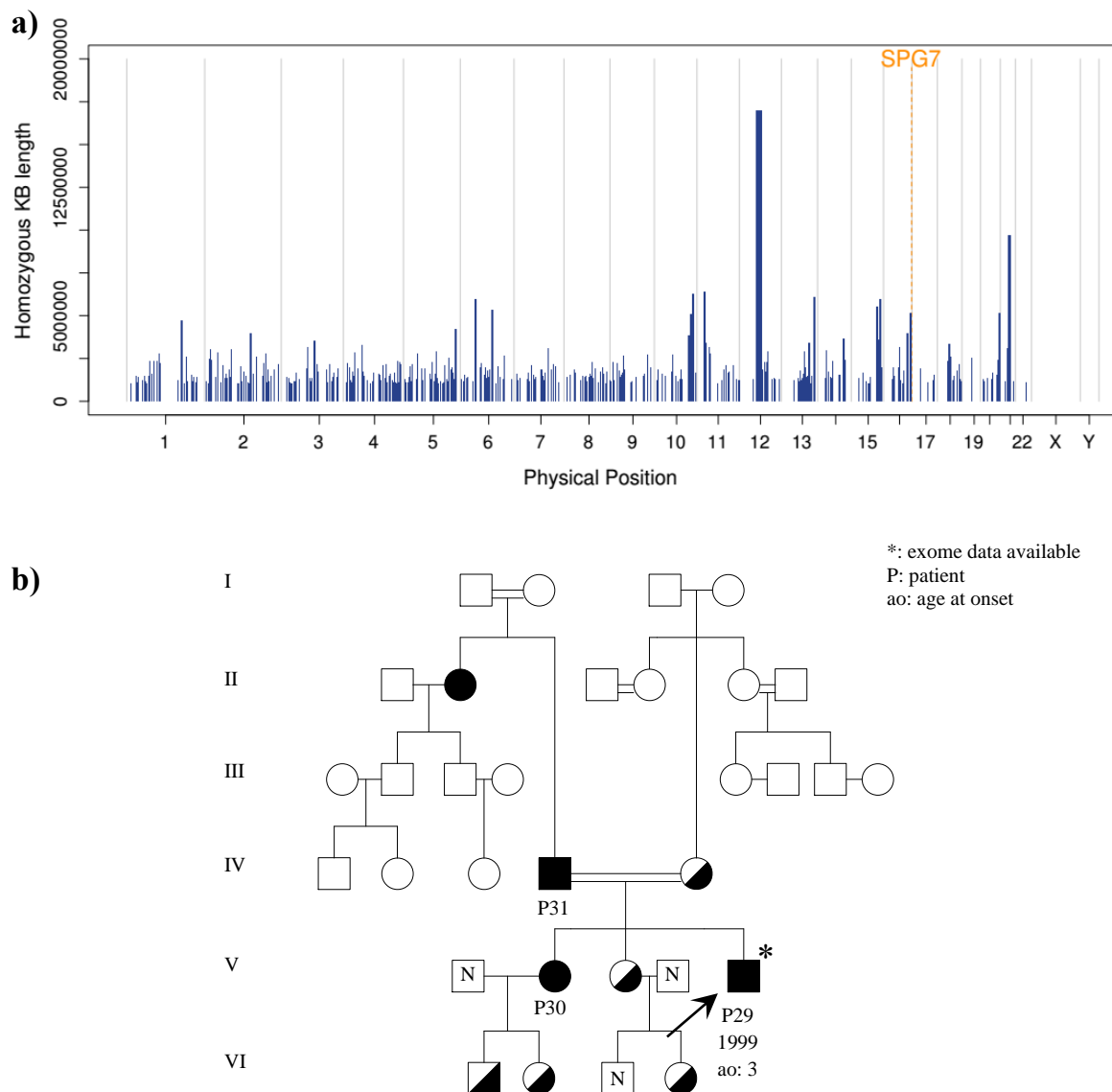


Figure 5.20. Homozygosity mapping of the index case (a) and the segregation of the *SPG7* variant in available individuals (b).

The youngest son of the inbred Family 15 was analyzed by homozygosity mapping. There was a pseudo-dominant appearance of the pedigree due to multigenerational consanguinity. The Ala658Thr missense variant was identified in the *SPG7* gene (Figure 5.20) which was confirmed to be in homozygous state in all affected individuals. The unaffected mother, older sister, nieces and one nephew were heterozygous for the variant. The son of the unaffected sister (VI.3) was found to carry the wild-type sequence (Figure 5.20b).

Family 16:

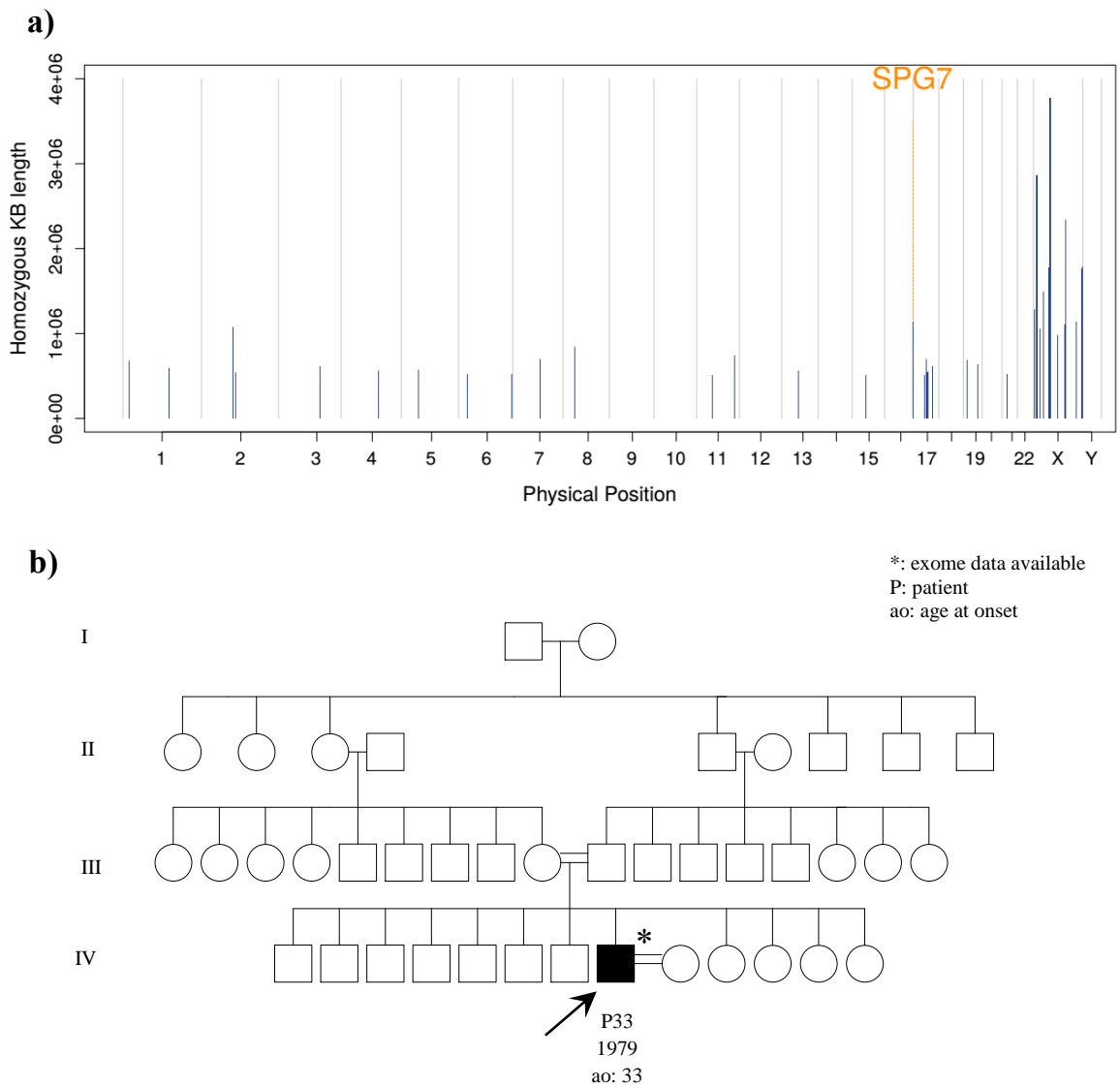


Figure 5.21. Homozygosity mapping of P33 (a) and the pedigree of Family 16 (b).

The index case of Family 16 (P33) was subjected to WES followed by homozygosity mapping. Among 41 homozygous regions, Thr588Met missense mutation in the *SPG7* gene was detected (Figure 5.21a). Sanger sequencing validated the presence of the variant. The segregation will be confirmed by obtaining DNA from additional individuals of the family.

5.3.2.8. Autosomal recessive hereditary spastic ataxia type 2: *KIF1C*

Family 17:

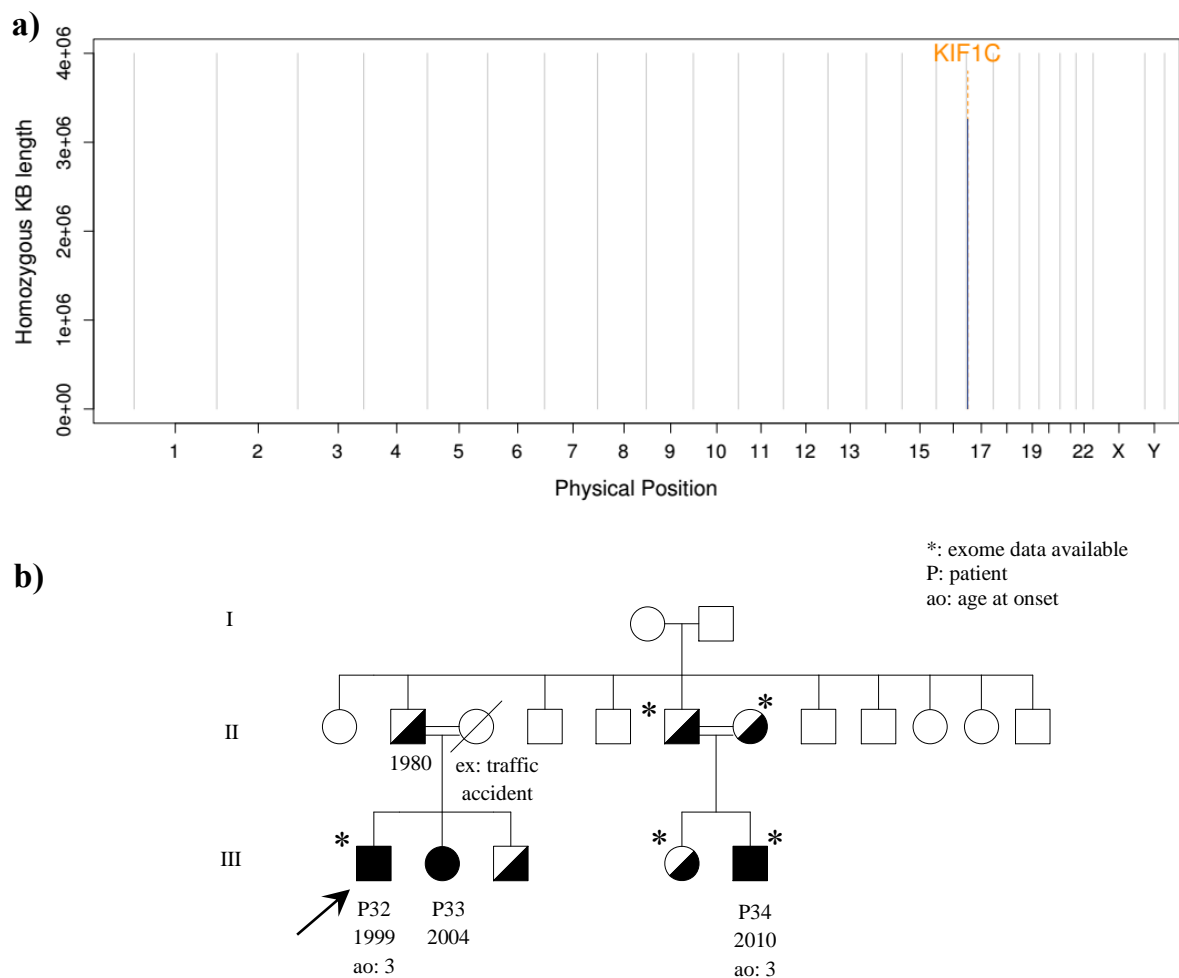


Figure 5.22. Runs of homozygosity in affected cousins (a) and the pedigree of Family 17 with the segregation of the missense *KIF1C* variant (b).

First cousins, both from consanguinity, were subjected to WES together with unaffected sister and parents of P34. Runs of homozygosity, common in P33 and P35, were detected revealing one region. Consequently, the Arg13Gln missense mutation in the *KIF1C*

gene was found within this homozygous region in chromosome 17 of both patients (Figure 5.22a). Presence of this mutation was confirmed in homozygous state in all affected individuals. Unaffected siblings and the parents are heterozygous for the variant (Figure 5.22b).

5.3.2.9. Autosomal recessive cerebellar ataxia type 2: *ADCK3*

Two missense mutations were identified in the *ADCK3* gene by WES data analysis of juvenile patients and their family members (Figures 5.23 – 5.24).

Family 18:

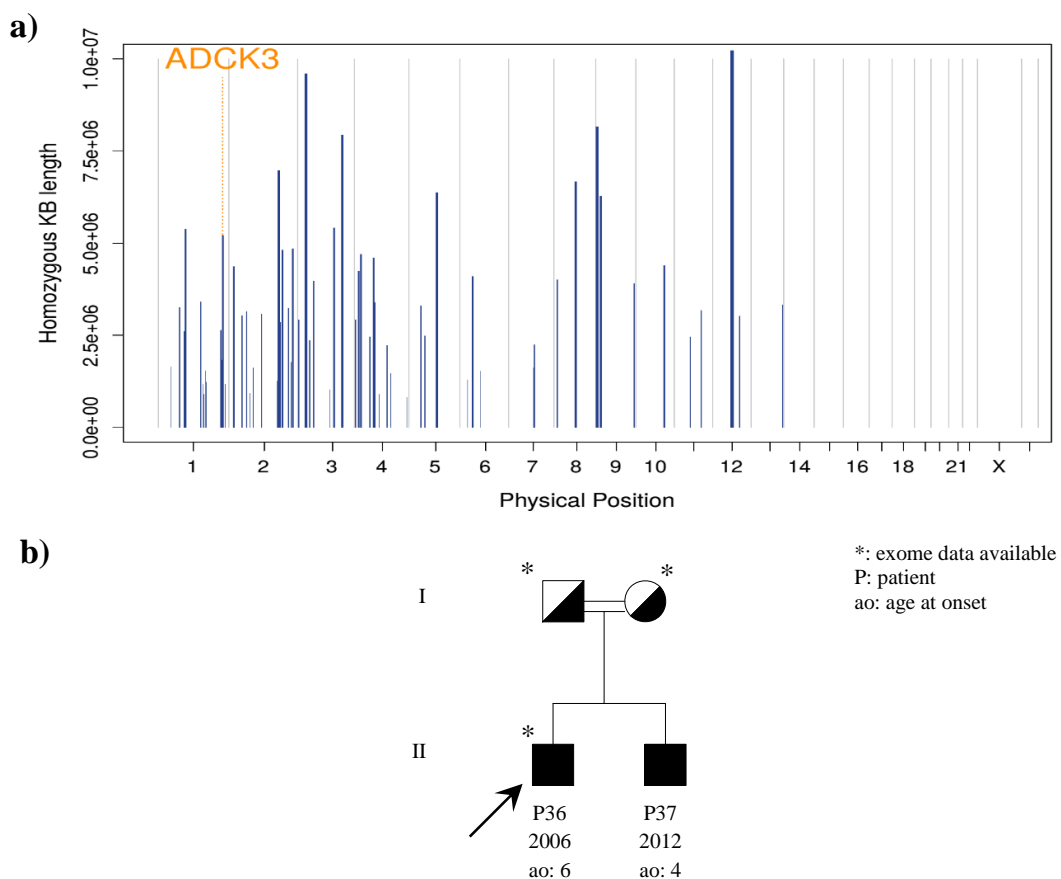


Figure 5.23. Homozygosity mapping plot of the index case (a) and the segregation of the *ADCK3* variant in Family 17 (b).

The juvenile patient in Family 17 was referred to our laboratory with the clinical diagnosis of cerebellar ataxia. He exhibited gait instability and speech problems. In the

course of the analysis, his brother started to develop similar complaints. WES revealed a homozygous Ala337Thr variant in the *ADCK3* gene. The variant was absent in our cohort and had a minor allele frequency of 4.29×10^{-5} in ExAC database without presence of homozygous genotype. Sanger sequencing confirmed the presence and segregation of the variant in the index case and parents. Also, the younger sibling carried the mutation in homozygous state (Figure 5.23).

Family 19:

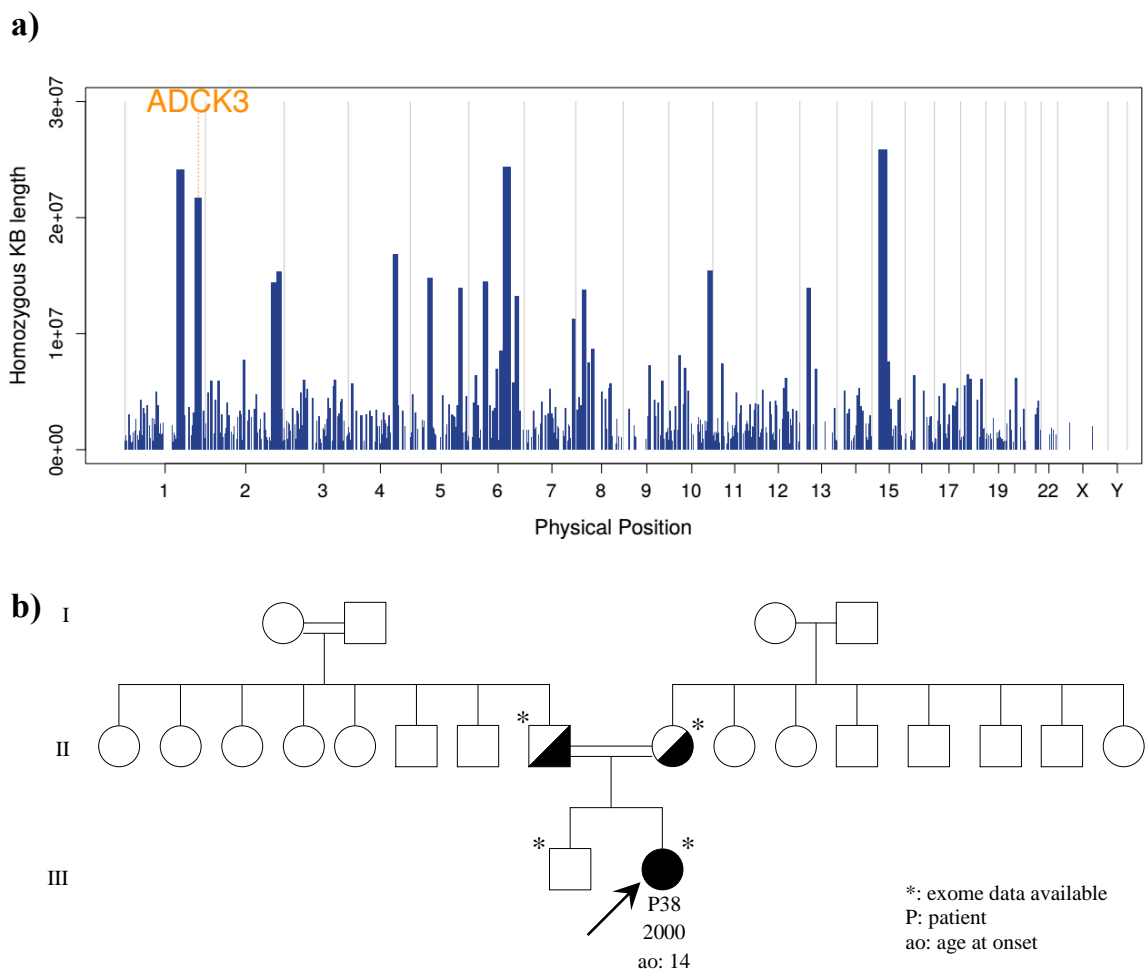


Figure 5.24. Distribution of runs of homozygosity of the index case (a) and the segregation of the *ADCK3* variant in Family 18 (b).

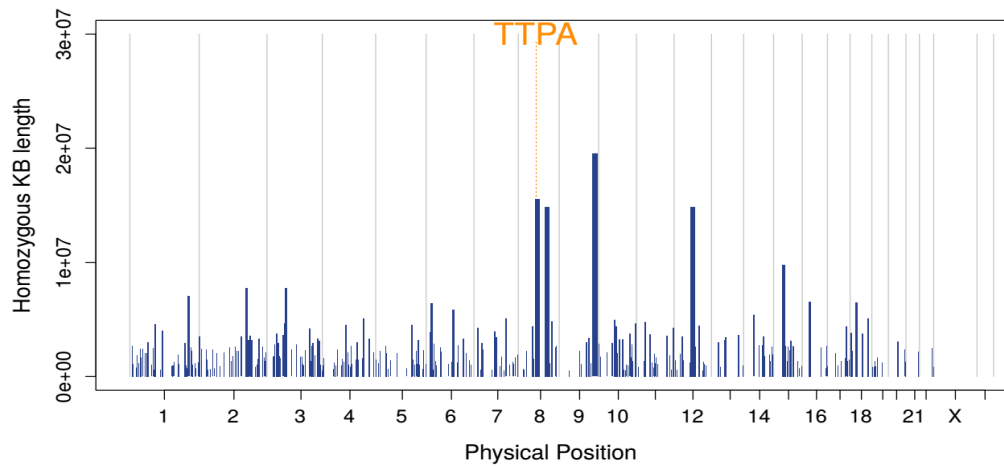
Figure 5.24 represents homozygosity mapping of the index case revealing numerous regions. Among those sites, Ala338Val missense variant in the *ADCK3* gene was identified. The variant was not present in dbSNP and in our cohort. The presence of the variant is

confirmed as homozygously in the patient, heterozygously in the parents while the unaffected brother was free of mutation.

5.3.2.10. Ataxia with vitamin E deficiency: TTPA

Family 20:

a)



b)

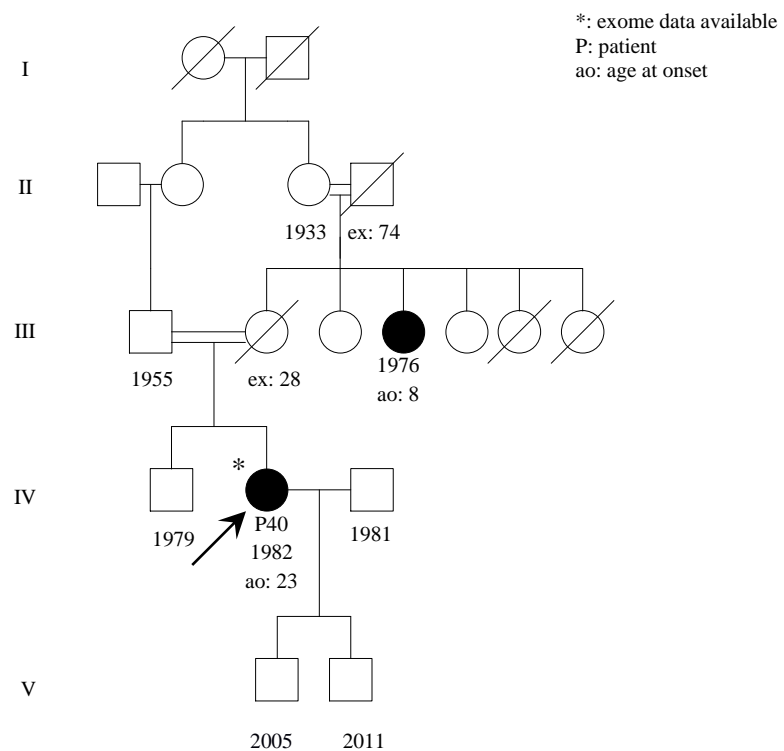


Figure 5.25. Runs of homozygosity of the index case (a) and pedigree of patient P40 (b).

The index case was referred to us with gait ataxia, dysarthria and tremor. She was subjected to WES which pinpointed a single nucleotide deletion in exon 3 of the *TTPA* gene (Figure 5.25). This deletion resulted in a frameshift at position 163 leading to a stop codon after 13 amino acids (Trp163fs*13). The variant had an ExAC MAF of 3.298e-05 and was not found in our cohort. The presence of the variant was validated in the patient.

Table 5.1. Total number of variants per family after each filtering step.

	Total number of variants	Number of exonic variants	Total number of variants by pedigree info	Total number of variants with MAF < 1%
Family 1	264574	30577	921	81
Family 2	209373	28100	4999	177
Family 3	229273	29105	87	2
Family 4	210592	29728	324	1
Family 5	220181	27465	5982	4
Family 6	714732	31649	1299	19
Family 7	467678	29742	1188	29
Family 8	161821	31836	370	8
Family 9	147318	29342	882	32
Family 10	116630	29860	442	3
Family 11	188651	71926	8570	2
Family 12	154668	19462	8005	22
Family 13	747866	31043	413	18
Family 14	177259	25364	1977	22
Family 15	6024432	173181	9224	139
Family 16	209632	23253	10053	43
Family 17	166780	33691	132	5
Family 18	148627	36042	1132	18
Family 19	559829	28581	651	14
Family 20	159059	20034	8073	12

Table 5.2. List of all confirmed pathogenic variants and associated diseases in the OMIM database.

	Inheritance	Initial diagnosis	Mutated gene	Mutation		Status	Associated Disease (OMIM)
				Coding sequence	Protein sequence		
Family 1	dominant	SCA	<i>AFG3L2</i>	c.C2062A	Pro688Thr	het	SCA28
Family 2	dominant	SCA	<i>KIF1B</i>	c.G4712A	Arg1571Gln	het	CMT2A1
Family 3	recessive	SCA	<i>SYNE1</i>	c.C20263T	Arg6755Ter	hom	ARCA1
Family 4	recessive	SCA	<i>SACS</i>	c.19923_12927delAAGAA	Lys4308fs*21	hom	ARSACS
Family 5	recessive	SCA	<i>SACS</i>	c.C11374T	Arg3792Ter	hom	ARSACS
Family 6	recessive	FA	<i>APTX</i>	c.T569G	Val190Gly	hom	AOA1
Family 7	recessive	SCA / FA	<i>SETX</i>	c.5549-2A>G	NA	hom	AOA2
Family 8	recessive	FA	<i>SETX</i>	c.5635delG	Val1879Ter	hom	AOA2
Family 9	recessive	FA	<i>SETX</i>	c.5249dupT	Leu1750fs*7	hom	AOA2
Family 10	recessive	SBMA / FA	<i>SETX</i>	c.T6686C	Met2229Thr	hom	AOA2
Family 11	recessive	FA	<i>SETX</i>	c.5249dupT	Leu1750fs*7	hom	AOA2
Family 12	recessive	SCA / FA	<i>CAPN1</i>	c.G994A	Gly332Arg	hom	ARHSP76
Family 13	recessive	SCA	<i>CAPN1</i>	c.G1176A	Trp392Ter	hom	ARHSP76
Family 14	recessive	SCA	<i>SPG11</i>	c.1966_1967delAA	K656fs*11	hom	ARHSP-TCC
Family 15	recessive	SCA	<i>SPG7</i>	c.G1972A	Ala658Thr	hom	ARHSP-7
Family 16	recessive	SCA	<i>SPG7</i>	c.C1763T	Thr588Met	Hom	ARHSP-7
Family 17	recessive	FA	<i>KIF1C</i>	c.G38A	Arg13Gln	hom	SPAX2
Family 18	recessive	FA / SCA	<i>ADCK3</i>	c.G1009A	Ala337Thr	hom	ARCA-2
Family 19	recessive	SCA	<i>ADCK3</i>	c. C1013T	Ala338Val	hom	ARCA-2
Family 20	recessive	FA	<i>TTPA</i>	c.487delT	Trp163fs*13	hom	AVED

(het: heterozygous, hom: homozygous)

Table 5.3. The quality scores and minor allele frequencies of the confirmed pathogenic variants.

Chr	Position	Mutated gene	Mutation	dbSNP ID	Quality Score	Read depth	Mapping quality	ExAC MAF	1000 Genome MAF
18	12337453	<i>AFG3L2</i>	Pro688Thr	-	4985.5	411	60.00	-	-
1	10431224	<i>KIF1B</i>	Arg1571Gln	-	4004.92	273	60.00	8.237e-06	-
6	152557375	<i>SYNE1</i>	Arg6755Ter	-	14212.17	564	60.00	8.243e-06	-
13	23905087	<i>SACS</i>	Lys4308fs*21	-	206.88	53	60.00	-	-
13	23906641	<i>SACS</i>	Arg3792Ter	-	1101.91	31	60.00	-	-
9	32984710	<i>APTX</i>	Val190Gly	rs536584919	3108.92	188	30.58	3.295e-05	-
9	135173701	<i>SETX</i>	c.5549-2A>G	-	2072.16	87	28.31	-	-
9	135173612	<i>SETX</i>	Val1879Ter	-	54212.88	2083	60.24	-	-
9	135201735	<i>SETX</i>	Leu1750fs*7	-	4874.12	250	60.00	-	-
9	135153613	<i>SETX</i>	Met2229Thr	-	10733.77	1206	60.00	-	-
9	135201735	<i>SETX</i>	Leu1750fs*7	-	1608.38	88	49.27	-	-
11	64955949	<i>CAPN1</i>	Gly332Arg	-	848.4	25	56.28	-	-
11	64972164	<i>CAPN1</i>	Trp392Ter	-	5778.9	230	30.13	8.513e-06	-
15	44920966	<i>SPG11</i>	Lys656fs*11	-	6041.15	190	61.58	-	-
16	89620237	<i>SPG7</i>	Ala658Thr	-	1465.61	1594	58.68	-	-
16	89617001	<i>SPG7</i>	Thr588Met	-	2437.77	102	44.40	-	-
17	4903579	<i>KIF1C</i>	Arg13Gln	-	4506.92	197	60.00	-	-
1	227170664	<i>ADCK3</i>	Ala337Thr	rs143816043	2938.77	240	60.00	4.29e-05	-
1	227170668	<i>ADCK3</i>	Ala338Val	-	3244.92	304	45.12	1.429e-05	-
8	63978527	<i>TTPA</i>	Trp163fs*13	-	118.38	69	49.69	3.298e-05	-

Table 5.4. Online prediction tool scores of the confirmed pathogenic variants.

Chr	Position	Mutated gene	Mutation		SIFT	PolyPhen2	CADD	GERP++
			Coding sequence	Protein sequence				
18	12337453	<i>AFG3L2</i>	c.C2062A	Pro688Thr	0.01	0.994	27.1	5.62
1	10431224	<i>KIF1B</i>	c.G4712A	Arg1571Gln	0.17	1.0	35	5.53
6	152557375	<i>SYNE1</i>	c.C20263T	Arg6755Ter	1	-	44	4.81
13	23905087	<i>SACS</i>	c.19923_12927delAAGAA	Lys4308fs*21	-	-	-	-
13	23906641	<i>SACS</i>	c.C11374T	Arg3792Ter	1	-	58	5.67
9	32984710	<i>APTX</i>	c.T569G	Val190Gly	0.03	1.0	23	5.59
9	135173701	<i>SETX</i>	c.5549-2A>G	NA	-	-	13.67	5.88
9	135173612	<i>SETX</i>	c.5635delG	Val1879Ter	-	-	-	-
9	135201735	<i>SETX</i>	c.5249dupT	Leu1750fs*7	-	-	-	-
9	135153613	<i>SETX</i>	c.T6686C	Met2229Thr	0.00	0.993	23.1	5.69
9	135201735	<i>SETX</i>	c.5249dupT	Leu1750fs*7	-	-	-	-
11	64955949	<i>CAPNI</i>	c.G994A	Gly332Arg	0.00	1.0	27.3	4.63
11	64972164	<i>CAPNI</i>	c.G1176A	Trp392Ter	1	-	39	4.51
15	44920966	<i>SPG11</i>	c.1966_1967delAA	Lys656fs*11	-	-	-	-
16	89620237	<i>SPG7</i>	c.G1972A	Ala658Thr	0.21	1.0	24.6	5.06
16	89617001	<i>SPG7</i>	c.C1763T	Thr588Met	0.00	1.0	23.5	4,88
17	4903579	<i>KIF1C</i>	c.C2062A	Arg13Gln	0.00	0.067	23.6	2.02
1	227170664	<i>ADCK3</i>	c.G1009A	Ala337Thr	0.00	1.0	36	5.66
1	227170668	<i>ADCK3</i>	c.C1013T	Ala338Val	0.00	0.997	31	5.66
8	63978527	<i>TTPA</i>	c.487delT	Trp163fs*13	-	-	-	-

5.3.3. Families without confirmed pathogenic mutations

After the completion of the WES data analysis, 27 pedigrees remained unsolved. One family (Family 31) was later shown to have an increased number of CAG repeats in the *HTT* gene, detected by fragment length analysis. The total number of variants in remaining 26 families are given in Table 5.5. Obviously, much more numerous variants remained in autosomal dominant families as compared to autosomal recessive pedigrees. Family 31 was not included in the table.

Table 5.5. The inheritance pattern, and number of variants in families without confirmed pathogenic variants

	Inheritance	Total number of variants	Number of exonic variants	Total number remaining variants
Family 21	AD	223081	29094	63
Family 22	AD	6130011	102739	498
Family 23	AD	249756	30100	46
Family 24	AD	228536	31182	59
Family 25	AD	148002	12443	247
Family 26	AD	6130011	24626	595
Family 27	AD	550623	29972	71
Family 28	AD	2727424	36142	3
Family 29	AD	67783	23452	515
Family 30	AD / AR	432972	33947	65 / 3
Family 32	AR	322428	34527	17
Family 33	AR	275859	32891	9
Family 34	AR	257857	29734	-
Family 35	AR	199767	14569	12
Family 36	AR	141979	12233	30

Table 5.5. The inheritance pattern, and number of variants in families without confirmed pathogenic variants (cont.)

	Inheritance	Total number of variants	Number of exonic variants	Total number remaining variants
Family 37	AR	190637	14373	-
Family 38	AR	212746	29877	-
Family 39	AR	153324	12645	1
Family 40	AR	191634	28781	6
Family 41	AR	20619	5693	4
Family 42	AR	21561	6332	7
Family 43	AR	201313	29440	8
Family 44	AR	67783	23452	516
Family 45	AR	117344	27957	11
Family 46	AR	133823	31062	7
Family 47	AR	1037974	29311	13

6. DISCUSSION

In this thesis, whole exome sequencing was applied to 150 individuals belonging to 47 distinct families (47 index patients). We successfully identified pathogenic variants in 20 probands / families, while 26 families remained undiagnosed; one patient was found to carry an unexpected CAG repeat expansion in the *HTT* gene. This 43% diagnostic yield in the analysis of ataxias by WES is in accordance with previous studies (Pyle *et al.*, 2015) and obviously more effective than gene panels (Németh *et al.*, 2013).

Heterozygous pathogenic variations were identified in 2/12 (17%) of the families presenting with an autosomal dominant inheritance pattern, as opposed to 18/34 (53%) of the families with an autosomal recessive inheritance. The higher yield in recessive cases implies the power of WES in recessive pedigrees and trios. The pattern of inheritance was not clear in one case, who remained undiagnosed so far.

6.1. Whole exome sequencing and genetic landscape of ataxias

In our cohort of 47 families, we identified mutations in 12 distinct genes encompassing 19 different mutations (Figure 5.5, Table 5.2). In two families with an autosomal dominant inheritance, heterozygous missense mutations in *AFG3L2* and *KIF1B* genes were identified (Figures 5.6 and 5.7). Four different homozygous mutations in the *SETX* gene were detected in five families, accounting for 15% (5/34) of the recessive cohort. Moreover, pathogenic variations in *SACS*, *CAPN1*, *SPG7*, and *ADCK3* were identified twice in unrelated families for each gene. Lastly, mutations in *SYNE1*, *APTX*, *SPG11*, *KIF1C*, and *TTPA* were described in one of each remaining family.

6.1.1. Autosomal dominant families

The number of SCA loci is increasing day-by-day through characterization of new genes causing dominant cerebellar ataxia. Unlike the more commonly seen SCA 1, 2, 3, 6, 7, and 17, the majority of SCA types are caused by point mutations. In this study, a heterozygous Pro688Thr mutation was identified in the *AFG3L2* gene, first shown to cause

SCA28 in a four-generation Italian family (Di Bella *et al.*, 2010). This gene encodes for a mitochondrial metalloprotease, ATPase family gene 3-like 2. It forms a hetero-oligomeric M-AAA protease complex with paraplegin coded by the SPG7 gene, which in turn is associated with hereditary spastic paraplegia (Casari *et al.*, 1998; Koppen *et al.*, 2007). Mutations were shown to disrupt the proteolytic domain. The variant in our family is also located in the active site of the enzyme (Figure 6.1).

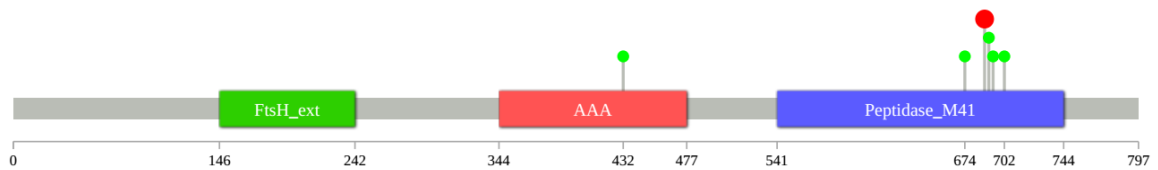


Figure 6.1. Lollipop diagram of the AFG3L2 protein. The Pro688T mutation is shown in red. Mutations reported by Di Bella *et al.* are indicated in green.

Mutations in the KIF1B gene were previously shown to cause Charcot-Marie-Tooth disease type 2A1 in heterozygous state (Zhao *et al.*, 2001). In one of our families, three affected siblings had inherited the Arg1571Gln variant heterozygously while their unaffected mother was wild type for the allele. The samples of the father and other affected members of the family were not available during the study. The siblings present an atypical disease form, dominated by cerebellar ataxia along with pes cavus and claw toe in one sibling, which is common for CMT (Hoyle, Isfort, Roggenbuck, and Arnold, 2015). The variant is not located in any domain of the protein, however is highly conserved (Table 5.5), therefore suggesting a possible effect on protein conformation leading to an extended phenotype.

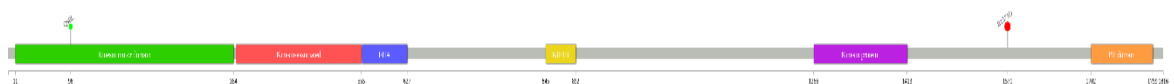


Figure 6.2. *KIF1B* gene structure and mutations. Arg1571Gln shown in red, mutation from Zhao *et al.* indicated in green.

6.1.2. Autosomal recessive families

Loss of function of the SYNE1 gene is known to cause a pure autosomal recessive cerebellar ataxia (Gros-Louis *et al.*, 2007). We identified a homozygous stop-gain mutation in a consanguineous family with three affected siblings. Phenotypic heterogeneity was observed in the family with earlier onset in the youngest brother and presence of a Babinski sign only in the males. Moreover, the youngest brother was being followed with a diagnosis of SMA, a lower motor neuron disorder. This finding is in accordance with the literature reporting involvement of motor neurons in SYNE1 mutations (Izumi *et al.*, 2013). We recently described two other probands carrying homozygous loss of function mutations in the SYNE1 gene in our Turkish cohort (Özoğuz *et al.*, 2015). Here, we present a family with manifestation of both ataxia and motor neuron disorder in the siblings indicating that the same mutation results in a spectrum of SYNE1-based neurodegenerative diseases. Our family contributes to the broadened clinical spectrum of SYNE1 mutations with additional intrafamilial phenotypic heterogeneity.

Following its identification in Quebec, mutations in the SACS gene causing autosomal recessive spastic ataxia of Charlevoix-Saguenay were described worldwide. In this study, we report two distinct families with homozygous mutations in the SACS gene, both resulting in a truncated protein. In one of these families (Family 4), one patient had a later onset of disease at the age of 20 whereas his brother had the disease from infancy. This observation supports the importance of modifiers in neurodegenerative disease. We do not have the detailed information regarding the other two siblings' phenotypes in Family 5; however, presence of the same complaints was reported. This observation can be explained by the relatively more homogeneous genetic background of Family 5 (Figure 5.10) as compared to Family 4 (Figure 5.9) with closer consanguinity and higher number of identity-by-descent segments in Family 5.

Defects in the APTX gene are linked to ataxia with oculomotor apraxia type 1 and hypoalbuminemia (M. M. do C. Moreira *et al.*, 2001). In our cohort, we identified a homozygous missense mutation in the APTX gene in one proband from an inbred family. The variant was located in a conserved site of the aprataxin protein and was predicted to be deleterious with a CADD score of 23 (Table 5.4).

Another type of ataxia with oculomotor apraxia was shown to be the result of homozygous or compound heterozygous mutations in the SETX gene (Moreira *et al.*, 2004). We report here five distinct families with different mutations in the SETX gene. Notably, oculomotor apraxia was not present in any of our patients (Table 3.1). One of the mutations gave rise to a disrupted splice site while two were causing a truncated protein. All patients were referred with the initial clinical diagnosis of ataxia and one with ataxia plus SBMA. In this latter family (Family 10), the Met2229Thr missense mutation was identified and was shown to segregate with the disease (Figure 5.15). Heterozygous missense mutations in the SETX gene are a cause of juvenile ALS, a young-onset motor neuron disease. In Family 10, the variant is inherited in homozygous state (Figure 3.3d, Table 5.2). The Met2229Thr could possibly mimic an SBMA-like phenotype in the patient in addition to ataxia when present homozygously.

Homozygous mutations in the CAPN1 gene were identified in two distinct families in our study cohort. The variants were suspected since the gene was shown to be associated with ataxia in terrier dogs in a GWA study (Forman, De Risio, and Mellersh, 2013). However, it was not reported in humans. Both probands had adult onset disease and spasticity in common (Table 3.1). Soon after, homozygous mutations in this gene were shown to cause a novel form of autosomal recessive hereditary spastic paraplegia (SPG76, MIM#114220) in patients with similar phenotypes (Gan-Or *et al.*, 2016), further supporting our genetic finding. Missense mutations described by Gan-Or *et al.* are located in the protease domain of the protein. This is in accordance with our findings, and we hypothesize that mutations in calpain-1 protein cause spastic ataxia by a loss-of-function type of mechanism (Figure 6.3).

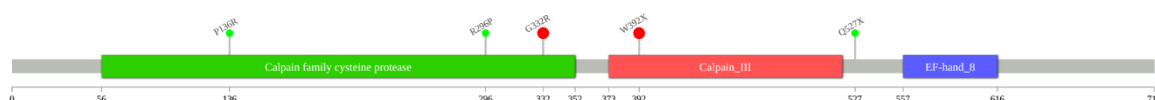


Figure 6.3. Mutations described in the *CAPN1* gene. Red: identified in this thesis, green: previous mutations by Gan-Or *et al.*

Mutations in the *SPG11* gene were reported to be the cause of a wide range of clinical entities such as autosomal recessive hereditary spastic paraplegia with thin corpus callosum (AR-HSP TCC) and autosomal recessive juvenile amyotrophic lateral sclerosis (ARJALS) (Stevanin *et al.*, 2007). The index case of our Family 14 referred to us with an initial diagnosis of cerebellar ataxia. Whole exome sequencing resulted in the identification of a homozygous frameshift mutation in the *SPG11* gene. Reexamination of the patient resulted in a final diagnosis of slowly progressive ALS (Iskender *et al.*, 2015). Thinning of the corpus callosum was also confirmed from the MRI image in addition to the previously reported cerebellar atrophy. This finding supports again the value of WES in clinical diagnosis since large genes like *SPG11* and *SYNE1* are very tedious and also not cost-effective to analyze by conventional methods.

We have detected two homozygous non-synonymous pathologic mutations in the *SPG7* locus (paraplegin), both located in the domain with catalytic activity (Figure 6.4). Although paraplegin mutations were known to cause hereditary spastic paraplegia, recent observations linked the mutations in the *SPG7* gene to a predominantly cerebellar phenotype (Choquet *et al.*, 2015; Yahikozawa *et al.*, 2015). We contribute to this recent association by reporting a family with a homozygous mutation in the *SPG7* gene without spasticity (Family 15). Additionally, another missense mutation (Thr588Met) in the same domain was identified in a patient (P33) with adulthood-onset cerebellar ataxia and very recently developed spasticity (Table 3.1, Figure 6.4). Thus, a follow-up of patients is very useful in order to observe the additional symptoms during disease development. These two families further support the extended phenotype of paraplegin mutations.



Figure 6.4. Structure of the paraplegin protein and mutations identified in this study.

Homozygous mutations in the *KIF1C* gene were identified in the locus that was previously associated with spastic ataxia type-2 (SPAX2, MIM#611302) (Dor *et al.*, 2014). This condition was also called hereditary spastic paraplegia type 58 and classified as

complicated HSP since patients may also exhibit ataxia and dysarthria (Caballero Oteyza *et al.*, 2014; Novarino *et al.*, 2014). Here we report a homozygous Arg13Gln mutation in the KIF1C gene in three individuals from an inbred pedigree (Family 17). The variant was neither present in ExAC and dbSNP databases, nor in our cohort of 303 samples. The phenotype was distinguished by the presence of tremor and dysarthria in addition to ataxia, present in our patients, as opposed to spasticity which was absent in the family (Table 3.1). Although, the onset of spasticity might be age-dependent, the absence of spasticity can as well be explained by the mutation effect on phenotype.

Coenzyme Q10 deficiency (MIM#607426) is associated with seizures, cognitive dysfunction and cerebellar ataxia. Mutations in ADCK3 were identified as a cause of coenzyme Q10 deficiency-related ataxia (Lagier-Tourenne *et al.*, 2008). The disease is characterized by childhood-onset cerebellar ataxia and cerebellar atrophy. Homozygous Ala337Thr and Ala338Val mutations were detected in two of our probands, both offspring of consanguineous couples. Previous reports show that supplement of Coenzyme Q10 masks the disease in patients (Artuch *et al.*, 2006), which emphasizes the value of molecular analysis.

In one of our families, a homozygous frameshift mutation (Trp163fs*13) in the TTPA gene was detected by WES. Homozygous or compound heterozygous mutations in this gene cause ataxia with vitamin E deficiency (MIM#277460) which is a treatable type of ARCA (Doria-Lamba *et al.*, 2006; Hentati *et al.*, 1996). The disease is very similar to Friedreich's ataxia in clinical manifestation. However, it can be distinguished by decreased levels of serum vitamin E. Again, the supplementation of vitamin E would cover the disease symptoms.

6.1.3. Advantages of whole exome sequencing in differential diagnosis of ataxias

Ataxias are complex and heterogeneous in terms of their clinical presentation and their mode of inheritance. They are divided into subtypes by the presence of components of the nervous system involved, such as abnormalities in eye movements, pyramidal tract involvement, or polyneuropathy. Presence of additional features confuses the clinicians

during the neurological examination. Imaging and metabolic tests help to exclude or consider suspected disorders; however, the exact diagnosis is possible only by molecular diagnosis.

The participants of the study cohort clustered in one group as shown in Figure 5.4. However, our findings (12 genes, 19 mutations in 20 families) clearly indicate the immense genetic heterogeneity of ataxias despite their sharing a certain level of genetic background. Of note, we identified mutations in genes, which are related to a wealth of neurological conditions other than ataxia. These findings imply the power of having a greater genome coverage instead of sequencing a limited number of genes. Additionally, some ataxia types are treatable but this requires an exact genetic diagnosis. In the framework of this study, we identified mutations in the *ADCK3* and *TTPA* genes, both of which result in ataxia with deficiency of a metabolic substance. A treatment schedule by coenzyme Q10 will be planned for patients (P36-P38) and patient 39 will receive vitamin E supplementation. These treatments will hopefully help to ameliorate the symptoms and increase the life quality of these patients.

6.1.4. Novel genotype - phenotype associations

Next generation sequencing technologies greatly contributed to our knowledge of disease by identification of novel disease genes thereby increasing the resolution of disease pathogenesis. On the other hand, it also blurred the line between different types of neurodegenerative disorders. In our cohort, whole exome sequencing of 47 families revealed novel genotype-phenotype correlations in addition to novel mutations in known ataxia genes. These include mutations in CMT- or HSP-related genes, as well as description of novel phenotypes in known ataxia genes.

As an example, a heterozygous mutation in the *KIF1B* gene, which was previously associated with CMT, was identified. Although the patients exhibited phenotypic features resembling CMT, cerebellar ataxia was predominant. A similar observation was recently reported in a patient with late-onset Friedreich's ataxia mimicking the CMT phenotype with presence of neuropathy and pes cavus (Salomão *et al.*, 2016). Taken together, these findings point to a clinical overlap between these two groups of disorders. We report an extended phenotype of the *KIF1B* gene which will be tested by further functional studies. Therefore,

we conclude that our finding broadens the *KIF1B* disease spectrum and propose *KIF1B* as a potential cause of dominant ataxia.

Mutations in HSP-associated genes were identified in five distinct families. Among these, the referring clinician later confirmed the phenotype of the index case from Family 14 with an SPG11 mutation. A re-examination of the patient ended up with diagnosis of juvenile ALS. On the other hand, cerebellar syndrome with spasticity was reported in both families with CAPN1 mutations. The disease was called spastic paraplegia 76, and we propose that this is a novel form of complicated HSP. Finally yet importantly, a cerebellar syndrome without spasticity was observed in a family with a homozygous mutation in paraplegin (SPG7). Recent observations showed that paraplegin is also a frequent cause of ataxia (Choquet *et al.*, 2015; Yahikozawa *et al.*, 2015). We further support this argument with our ataxia family and suggest to consider SPG7 mutations as another genetic cause of recessive ataxia.

6.2. Challenges and limitations of clinical exome sequencing: Why did we fail in finding the genes in approximately half of our families?

In our study, whole exome sequencing succeeded in identifying almost half of mutated genes in our families, but, in turn, failed to detect mutations in disease-associated genes in the other half. In families with a high number of remaining variants (Table 5.5), sequencing of additional family members may help to decrease the list of candidates. However, there are also limitations of this workflow in the identification of causal mutations. The problems can be either strategical or technical. Technical limitations include lack of non-coding regions of the genome in WES and limited ability to detect structural variations and repeat expansions. In addition, the disease may be caused by a mutation in a gene that has not been associated with a related phenotype, thus new strategies towards novel gene discovery have to be designed.

A data processing pipeline is crucial in next generation sequencing-based research. Several algorithms and software are being developed in order to reduce the false positive and negative results for both mapping the reads to the reference genome and variant calling. We adapted BWA-GATK best practices pipeline using HaplotypeCaller to our data that is

also the most widely used pipeline in NGS-based research. However, the same data may produce different results when an alternative pipeline is applied. The choice of alignment software is less likely to alter the result and the main difference is caused at the variant calling step (Hwang *et al.*, 2015). In addition to GATK's second variant caller, UnifiedGenotyper, currently available alternative tools are FreeBayes and mpileup by Samtools (Garrison and Marth, 2012; H. Li *et al.*, 2009). Both overlaps and non-overlaps were observed when these alternative algorithms were compared on the same dataset. Therefore, using an additional variant caller may help to uncover pathogenic mutations that may be missed by our regular pipeline.

A major drawback of WES is the lack of non-coding regions of the genome. Mutations may reside in the regulatory or intronic regions of the known disease genes and alter their expression ending up with the disease phenotype. This obstacle can be overcome by whole genome sequencing; however, WGS is not yet as cost-efficient as WES is. In addition, WGS brings additional computational challenges to deal with; interpretation of the data is more challenging since all individuals have private variations in non-coding regions. Prioritization of the variants requires development of additional reliable bioinformatics tools. Therefore, extending the targeted regions of exome capture kits to include known regulatory sites may be applicable until WGS becomes more affordable and routine.

The read-size limitation of NGS brings challenges in the mapping and detection of repetitive regions of the genome and large deletions or insertions. Thus, structural variant discovery from NGS data is practically not applicable. As the read size increases with the developing technology, new algorithms are being designed to detect structural variations from sequencing data. Currently, there is a bulk of tools on this purpose using different approaches such as read depth, read-pair, split-read and de novo assembly (Tattini *et al.*, 2015). However, they end up with a high rate of false-positives and are not very useful in whole exome data due to the lack of 99% of the genome. This problem will be overcome by increasing the read length and development of novel computational algorithms in the near future.

In the majority of cases, a clinical interpretation cannot be made due to the inability to determine the pathogenicity of the variants. These so-called variants of unknown

significance (VUS) are novel or rare variants in a gene not yet shown to be related to the disease of interest. They can be pathogenic and may cause the disease phenotype; however, they need to be confirmed by proper phenotyping and segregation of the variant within a pedigree. Thus, results of whole exome sequencing must be reanalyzed as novel gene-disease associations are being reported.

6.3. Path to discovery of novel disease genes

WES is an effective method for clinical diagnosis of single gene disorders. It is highly effective in discovery of novel gene-disease associations. WES has transformed medical genetics and resulted in viewing the mechanism of pathogenesis both in a wider angle and in more resolution.

New strategies should be designed for disease gene identification in undiagnosed individuals. Interactions or homology between proteins may help to prioritize genes to detect novel disease genes. This strategy helped in identification of AFG3L2 as genetic cause of SCA28 due to its high homology with SPG7 (Figure 6.1 and 6,3) (Di Bella *et al.*, 2010). In addition to family-based variant filtering, combining individuals with similar clinical presentation, such as age at onset and shared common symptoms can be highly effective, especially in definition of population-specific disease genes. Such a study recently resulted in identification of mutations in the *MME* gene in CMT type 2 and was shown to explain a higher proportion of CMT cases as compared to other reported genes involved in autosomal recessive CMT (Higuchi *et al.*, 2016). This very study illustrates the importance of proper and detailed phenotyping towards disease gene discovery.

The unsolved 26 families in this thesis, are expected to unravel new ataxia genes, possibly also Turkey and Mediterranean-specific ones. For this purpose, a deeper phenotyping of the families and a fast adoption of the rapidly evolving data analysis tools become crucial.

7. CONCLUSION

Hereditary ataxias are rare disorders being highly heterogeneous at both clinical, genetic and mechanistic levels. They constitute a significant portion of neurodegenerative disorders. In the framework of this thesis, we aimed to investigate the molecular genetic analyses of ataxia patients referred to us by clinicians throughout the nation. Whole exome sequencing, today the gold standard in investigating single gene disorders, was applied to our pedigrees, the majority of which had a recessive inheritance pattern. Our results unraveled in almost half of the 47 pedigrees the gene mutation responsible of an ataxia phenotype. This high yield points to the impact of WES in precise disease diagnosis.

In the genomic era, our knowledge of the genetic pathology of ataxias clearly shows an exponential increase; however, there are still several subtypes of ataxia not defined yet. This will be accomplished by both advances in the sequencing technologies and development of novel computational algorithms to analyze the wealth of sequencing data. Once all genes and mutations leading to ataxias are known, the mechanisms of pathogenesis leading to cerebellar dysfunction will be enlightened. This will also enable detection of disease modifiers leading to personalized treatment approaches.

To the best of our knowledge, this is the largest genetic study investigating the epidemiology and molecular basis of dominant and recessive ataxias in Turkey. We clearly show that recessive ataxia might occur more frequently in the presence of inbreeding. We also present that WES is very efficient in investigating recessive families. The detailed phenotypic description of the patients gains a dramatic importance in addition to data generation and evaluation. In the era of precision medicine, the successful combination of all these will hopefully pave the ways to understanding the exact ataxia pathogenesis and open the gates to long-awaited therapies.

REFERENCES

- Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev, 2010, "A Method and Server for Predicting Damaging Missense Mutations", *Nature Methods*, Vol. 7, No. 4, pp. 248–249.
- Aicardi, J., C. Barbosa, E. Andermann, F. Andermann, R. Morcos, Q. Ghanem, Y. Fukuyama, Y. Awaya, and P. Moe, 1988, "Ataxia-Ocular Motor Apraxia: A Syndrome Mimicking Ataxia-Telangiectasia.", *Annals of neurology*, Vol. 24, No. 4, pp. 497–502.
- Anderson, J. F., E. Siller, and J. M. Barral, 2010, "The Sacsin Repeating Region (SRR): A Novel Hsp90-Related Supra-Domain Associated with Neurodegeneration", *Journal of Molecular Biology*, Vol. 400, , pp. 665–674.
- Anderson, J. F., E. Siller, and J. M. Barral, 2011, "The Neurodegenerative-Disease-Related Protein Sacsin Is a Molecular Chaperone", *Journal of Molecular Biology*, Vol. 411, No. 4, pp. 870–880.
- Anheim, M., 2012, "Autosomal Recessive Cerebellar Ataxias", *N Engl J Med*, Vol. 366, No. 7, pp. 636–646.
- Anheim, M., B. Monga, M. Fleury, P. Charles, C. Barbot, M. Salih, J. P. Delaunoy, M. Fritsch, L. Arning, M. Synofzik, L. Schöls, J. Sequeiros, C. Goizet, C. Marelli, I. Le Ber, J. Koht, J. Gazulla, J. De Bleecker, M. Mukhtar, N. Drouot, L. Ali-Pacha, T. Benhassine, M. Chbicheb, A. M'Zahem, A. Hamri, B. Chabrol, J. Pouget, R. Murphy, M. Watanabe, P. Coutinho, M. Tazir, A. Durr, A. Brice, C. Tranchant, and M. Koenig, 2009, "Ataxia with Oculomotor Apraxia Type 2: Clinical, Biological and Genotype/phenotype Correlation Study of a Cohort of 90 Patients", *Brain*, Vol. 132, No. 10, pp. 2688–2698.
- Auton, A., L. D. Brooks G. R., R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, Abecasis, 2015: 1000 Genomes Project

- Consortium, "A Global Reference for Human Genetic Variation", *Nature*, Vol. 526, No. 7571, pp. 68–74.
- Boder, E., and R. P. Sedgwick, 1958, "Ataxia-Telangiectasia; a Familial Syndrome of Progressive Cerebellar Ataxia, Oculocutaneous Telangiectasia and Frequent Pulmonary Infection.", *Pediatrics*, Vol. 21, No. 4, pp. 526–54.
- Bomont, P., M. Watanabe, R. Gershoni-Barush, M. Shizuka, M. Tanaka, J. Sugano, C. Guiraud-Chaumeil, and M. Koenig, 2000, "Homozygosity Mapping of Spinocerebellar Ataxia with Cerebellar Atrophy and Peripheral Neuropathy to 9q33–34, and with Hearing Impairment and Optic Atrophy to 6p21–23", *European Journal of Human Genetics*, Vol. 8, No. 12, pp. 986–990.
- Botstein, D., and N. Risch, 2003, "Discovering Genotypes Underlying Human Phenotypes: Past Successes for Mendelian Disease, Future Approaches for Complex Disease.", *Nature genetics*, Vol. 33 Suppl, pp. 228–37.
- Bouchard, J. P., A. Barbeau, R. Bouchard, and R. W. Bouchard, 1978, "Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay.", *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques*, Vol. 5, No. 1, pp. 61–9.
- Broad Institute, n.d., Picard Tools. <http://broadinstitute.github.io/picard>, accessed at November 2016.
- Caballero Oteyza, A., E. Battaloğlu, L. Ocek, T. Lindig, J. Reichbauer, A. P. Rebelo, M. A. Gonzalez, Y. Zorlu, B. Ozes, D. Timmann, B. Bender, G. Woehlke, S. Züchner, L. Schöls, and R. Schüle, 2014, "Motor Protein Mutations Cause a New Form of Hereditary Spastic Paraplegia.", *Neurology*, Vol. 82, No. 22, pp. 2007–16.
- Chen, R., E. Le Rouzic, J. A. Kearney, L. M. Mansky, and S. Benichou, 2004, "Vpr-Mediated Incorporation of UNG2 into HIV-1 Particles Is Required to Modulate the Virus Mutation Rate and for Replication in Macrophages", *JBC Papers in Press*.

Published on April, Vol. 19.

- Choquet, K., M. Tetreault, S. Yang, R. La Piana, M.-J. Dicaire, M. R. Vanstone, J. Mathieu, J.-P. Bouchard, M.-F. Rioux, G. A. Rouleau, K. M. Boycott, J. Majewski, and B. Brais, 2015, "SPG7 Mutations Explain a Significant Proportion of French Canadian Spastic Ataxia Cases.", *European journal of human genetics : EJHG*, No. October, pp. 1–6.
- Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden, n.d., "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of *Drosophila Melanogaster* Strain w1118; Iso-2; Iso-3.", *Fly*, Vol. 6, No. 2, pp. 80–92.
- Cirulli, E. T., B. N. Lasseigne, S. Petrovski, P. C. Sapp, P. A. Dion, C. S. Leblond, J. Couthouis, Y.-F. Lu, Q. Wang, B. J. Krueger, Z. Ren, J. Keebler, ... S. P. Gygi, 2015, "Exome Sequencing in Amyotrophic Lateral Sclerosis Identifies Risk Genes and Pathways.", *Science (New York, N.Y.)*, Vol. 347, No. 6229, pp. 1436–41.
- Clark, M. J., R. R. Chen, H. Y. K. Lam, K. J. Karczewski, R. R. Chen, G. Euskirchen, A. J. Butte, and M. Snyder, 2011, "Performance Comparison of Exome DNA Sequencing Technologies", *Nat Biotechnol*, Vol. 29, No. 10, pp. 908–914.
- Clements, P. M., C. Breslin, E. D. Deeks, P. J. Byrd, L. Ju, P. Bieganowski, C. Brenner, M.-C. Moreira, A. M. R. Taylor, and K. W. Caldecott, 2004, "The Ataxia–oculomotor Apraxia 1 Gene Product Has a Role Distinct from ATM and Interacts with the DNA Strand Break Repair Proteins XRCC1 and XRCC4", *DNA Repair*, Vol. 3, No. 11, pp. 1493–1502.
- Cogan, D. G., 1953, "A Type of Congenital Ocular Motor Apraxia Presenting Jerky Head Movements*", *American Journal of Ophthalmology*, Vol. 36, No. 4, pp. 433–441.
- Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis 1000 Genomes Project Analysis Group, 2011, "The Variant

- Call Format and VCFtools.", *Bioinformatics (Oxford, England)*, Vol. 27, No. 15, pp. 2156–8.
- Date, H., O. Onodera, H. Tanaka, K. Iwabuchi, K. Uekawa, S. Igarashi, R. Koike, T. Hiroi, T. Yuasa, Y. Awaya, T. Sakai, T. Takahashi, H. Nagatomo, Y. Sekijima, I. Kawachi, Y. Takiyama, M. Nishizawa, N. Fukuhara, K. Saito, S. Sugano, and S. Tsuji, 2001, "Early-Onset Ataxia with Ocular Motor Apraxia and Hypoalbuminemia Is Caused by Mutations in a New HIT Superfamily Gene.", *Nature genetics*, Vol. 29, No. 2, pp. 184–188.
- De Bot, S. T., M. A. A. P. Willemsen, S. Vermeer, H. P. H. Kremer, B. P. C. Van De, Warrenburg, and B. P. C. van de Warrenburg, 2012, "Reviewing the Genetic Causes of Spastic-Ataxias", *Neurology*, Vol. 79, No. 14, pp. 1507–1514.
- De Braekeleer, M., F. Giasson, J. Mathieu, M. Roy, J. P. Bouchard, and K. Morgan, 1993, "Genetic Epidemiology of Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay in Northeastern Quebec.", *Genetic epidemiology*, Vol. 10, No. 1, pp. 17–25.
- DePristo, M. A., E. Banks, R. Poplin, K. V Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytzky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly, 2011, "A Framework for Variation Discovery and Genotyping Using next-Generation DNA Sequencing Data.", *Nature genetics*, Vol. 43, No. 5, pp. 491–8.
- Di Bella, D., F. Lazzaro, A. Brusco, M. Plumari, G. Battaglia, A. Pastore, A. Finardi, C. Cagnoli, F. Tempia, M. Frontali, L. Veneziano, T. Sacco, E. Boda, A. Brussino, F. Bonn, B. Castellotti, S. Baratta, C. Mariotti, C. Gellera, V. Fracasso, S. Magri, T. Langer, P. Plevani, S. Di Donato, M. Muzi-Falconi, and F. Taroni, 2010, "Mutations in the Mitochondrial Protease Gene AFG3L2 Cause Dominant Hereditary Ataxia SCA28.", *Nature genetics*, Vol. 42, No. 4, pp. 313–321.
- Doi, H., K. Yoshida, T. Yasuda, M. Fukuda, Y. Fukuda, H. Morita, S. Ikeda, R. Kato, Y. Tsurusaki, N. Miyake, H. Saitsu, H. Sakai, S. Miyatake, M. Shiina, N. Nukina, S,

- Koyano, S. Tsuji, Y. Kuroiwa and N. Matsumoto, 2011, "Exome Sequencing Reveals a Homozygous SYT14 Mutation in Adult-Onset, Autosomal-Recessive Spinocerebellar Ataxia with Psychomotor Retardation.", *American journal of human genetics*, Vol. 89, No. 2, pp. 320–7.
- Dor, T., Y. Cinnamon, L. Raymond, A. Shaag, N. Bouslam, A. Bouhouche, M. Gausson, V. Meyer, A. Durr, A. Brice, A. Benomar, G. Stevanin, M. Schuelke, and S. Edvardson, 2014, "KIF1C Mutations in Two Families with Hereditary Spastic Paraparesis and Cerebellar Dysfunction.", *Journal of medical genetics*, Vol. 51, No. 2, pp. 137–42.
- Durr, A., 2010, "Autosomal Dominant Cerebellar Ataxias: Polyglutamine Expansions and beyond", *The Lancet Neurology*, Vol. 9, No. 9, pp. 885–894.
- Engert, J. C., C. Doré, J. Mercier, B. Ge, C. Bétard, J. D. Rioux, C. Owen, P. Bérubé, K. Devon, B. Birren, S. B. Melançon, K. Morgan, T. J. Hudson, and A. Richter, 1999, "Autosomal Recessive Spastic Ataxia of Charlevoix–Saguenay (ARSACS): High-Resolution Physical and Transcript Map of the Candidate Region in Chromosome Region 13q11", *Genomics*, Vol. 62, No. 2, pp. 156–164.
- Ferrarini, M., G. Squintani, T. Cavallaro, S. Ferrari, N. Rizzuto, and G. M. Fabrizi, 2007, "A Novel Mutation of Aprataxin Associated with Ataxia Ocular Apraxia Type 1: Phenotypical and Genotypical Characterization", *Journal of the Neurological Sciences*, Vol. 260, No. 1–2, pp. 219–224.
- Fogel, B. L., E. Cho, A. Wahnich, F. Gao, O. J. Becherel, X. Wang, F. Fike, L. Chen, C. Criscuolo, G. De Michele, A. Filla, A. Collins, A. F. Hahn, R. A. Gatti, G. Konopka, S. Perlman, M. F. Lavin, D. H. Geschwind, and G. Coppola, 2014, "Mutation of Senataxin Alters Disease-Specific Transcriptional Networks in Patients with Ataxia with Oculomotor Apraxia Type 2", *Human Molecular Genetics*, Vol. 23, No. 18, pp. 4758–4769.
- Fogel, B. L., and S. Perlman, 2007, "Clinical Features and Molecular Genetics of Autosomal Recessive Cerebellar Ataxias", *Lancet Neurology*, Vol. 6, No. 3, pp. 245–257.

- Foo, J., J. Liu, and E. Tan, 2012, "Whole-Genome and Whole-Exome Sequencing in Neurological Diseases", *Nature Rev. Neurol.*, Vol. 8, No. 9, pp. 508–517.
- Freischmidt, A., T. Wieland, B. Richter, W. Ruf, V. Schaeffer, K. Müller, N. Marroquin, F. Nordin, A. Hübers, P. Weydt, S. Pinto, R. Press, S. Millecamps, N. Molko, E. Bernard, C. Desnuelle, M.-H. Soriani, J. Dorst, E. Graf, U. Nordström, M. S. Feiler, S. Putz, T. M. Boeckers, T. Meyer, A. S. Winkler, J. Winkelmann, M. de Carvalho, D. R. Thal, M. Otto, T. Brännström, A. E. Volk, P. Kursula, K. M. Danzer, P. Lichtner, I. Dikic, T. Meitinger, A. C. Ludolph, T. M. Strom, P. M. Andersen, and J. H. Weishaupt, 2015, "Haploinsufficiency of TBK1 Causes Familial ALS and Fronto-Temporal Dementia", *Nature Neuroscience*, Vol. 18, No. 5, pp. 631–636.
- García, A., C. Criscuolo, G. de Michele, and J. Berciano, 2008, "Neurophysiological Study in a Spanish Family with Recessive Spastic Ataxia of Charlevoix-Saguenay.", *Muscle & nerve*, Vol. 37, No. 1, pp. 107–110.
- Garrison, E., and G. Marth, 2012, "Haplotype-Based Variant Detection from Short-Read Sequencing", *arXiv preprint:1207.3907*, pp. 9.
- Gazal, S., S. Gosset, E. Verdura, F. Bergametti, S. Guey, M.-C. Babron, and E. Tournier-Lasserre, 2015, "Can Whole-Exome Sequencing Data Be Used for Linkage Analysis?", *European Journal of Human Genetics*, Vol. 33, No. May, pp. 1–6.
- Gilissen, C., A. Hoischen, H. G. Brunner, and J. A. Veltman, 2012, "Disease Gene Identification Strategies for Exome Sequencing", *Eur J Hum Genet*, Vol. 20, No. 5, pp. 490–497.
- Gregianin, E., G. Vazza, E. Scaramel, F. Boaretto, A. Vettori, E. Leonardi, S. C. E. Tosatto, R. Manara, E. Pegoraro, and M. L. Mostacciuolo, 2013, "A Novel SACS Mutation Results in Non-Ataxic Spastic Paraplegia and Peripheral Neuropathy", *European Journal of Neurology*, Vol. 20, No. 11, pp. 1486-1491.
- Gros-Louis, F., N. Dupré, P. Dion, M. a Fox, S. Laurent, S. Verreault, J. R. Sanes, J.-P.

- Bouchard, and G. a Rouleau, 2007, "Mutations in SYNE1 Lead to a Newly Discovered Form of Autosomal Recessive Cerebellar Ataxia.", *Nature genetics*, Vol. 39, No. 1, pp. 80–85.
- Gusella, J. F., N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, and A. Y. Sakaguchi, 1983, "A Polymorphic DNA Marker Genetically Linked to Huntington's Disease.", *Nature*, Vol. 306, No. 5940, pp. 234–238.
- HannonLab, n.d., FASTX-Toolkit, http://hannonlab.cshl.edu/fastx_toolkit, accessed at November 2016.
- Harding, A. E., 1982, "The Clinical Features and Classification of the Late Onset Autosomal Dominant Cerebellar Ataxias: A STUDY of 11 Families, INCLUDING Descendants OF "the DREW Family OF Walworth"", *Brain*, Vol. 105, No. 1, pp. 1–28.
- Hersheson, J., A. Haworth, and H. Houlden, 2012, "The Inherited Ataxias: Genetic Heterogeneity, Mutation Databases, and Future Directions in Research and Clinical Diagnostics", *Human Mutation*, Vol. 33, No. 9, pp. 1324–1332.
- Higuchi, Y., A. Hashiguchi, J. Yuan, A. Yoshimura, J. Mitsui, H. Ishiura, M. Tanaka, S. Ishihara, H. Tanabe, S. Nozuma, Y. Okamoto, E. Matsuura, R. Ohkubo, S. Inamizu, W. Shiraishi, R. Yamasaki, Y. Ohyagi, J. I. Kira, Y. Oya, H. Yabe, N. Nishikawa, S. Tobisawa, N. Matsuda, M. Masuda, C. Kugimoto, K. Fukushima, S. Yano, J. Yoshimura, K. Doi, M. Nakagawa, S. Morishita, S. Tsuji, and H. Takashima, 2016, "Mutations in MME Cause an Autosomal-Recessive Charcot-Marie-Tooth Disease Type 2", *Annals of Neurology*, Vol. 79, No. 4, pp. 659–672.
- Hwang, S., E. Kim, I. Lee, and E. M. Marcotte, 2015, "Systematic Comparison of Variant Calling Pipelines Using Gold Standard Personal Exome Variants", *Scientific Reports*, Vol. 5, No. December, pp. 17875.
- IGV (Integrative Genomic Viewer), 2013, "Integrative Genomics Viewer", *Broad Institute*,

Vol. 29, No. 1, pp. 24–26.

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

Izumi, Y., R. Miyamoto, H. Morino, A. Yoshizawa, K. Nishinaka, F. Udaka, M. Kameyama, H. Maruyama, and H. Kawakami, 2013, "Cerebellar Ataxia with SYNE1 Mutation Accompanying Motor Neuron Disease (P05.045)", *Neurology*, Vol. 80, No. 7 Supplement, pp. P05.045-P05.045.

Jayadev, S., and T. D. Bird, 2013, "Hereditary Ataxias: Overview.", *Genetics in medicine : official journal of the American College of Medical Genetics*, Vol. 15, No. November 2012, pp. 673–83.

Kancheva, D., D. Atkinson, P. De Rijk, M. Zimon, T. Chamova, V. Mitev, A. Yaramis, G. Maria Fabrizi, H. Topaloglu, I. Tournev, Y. Parma, E. Battaloglu, A. Estrada-Cuzcano, and A. Jordanova, 2015, "Novel Mutations in Genes Causing Hereditary Spastic Paraplegia and Charcot-Marie-Tooth Neuropathy Identified by an Optimized Protocol for Homozygosity Mapping Based on Whole-Exome Sequencing.", *Genetics in medicine : official journal of the American College of Medical Genetics*, No. 18, pp. 600-607.

Klein, C., G. K. Wenning, N. P. Quinn, and C. D. Marsden, 1996, "Ataxia without Telangiectasia Masquerading as Benign Hereditary Chorea", *Movement Disorders*, Vol. 11, No. 2, pp. 217–220.

Kozlov, G., A. Y. Denisov, M. Girard, M.-J. Dicaire, J. Hamlin, P. S. McPherson, B. Brais, and K. Gehring, 2011, "Structural Basis of Defects in the Sacsin HEPN Domain Responsible for Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay (ARSACS)", *Journal of Biological Chemistry*, Vol. 286, No. 23, pp. 20407–20412.

- Le Ber, I., N. Bouslam, S. Rivaud-Péchoux, J. Guimarães, A. Benomar, C. Chamayou, C. Goizet, M.-C. Moreira, S. Klur, M. Yahyaoui, Y. Agid, M. Koenig, G. Stevanin, A. Brice, and A. Dürr, 2004, "Frequency and Phenotypic Spectrum of Ataxia with Oculomotor Apraxia 2: A Clinical and Genetic Study in 18 Patients.", *Brain : a journal of neurology*, Vol. 127, No. Pt 4, pp. 759–67.
- Le Ber, I., A. Dürr, and A. Brice, 2011, "Autosomal recessive cerebellar ataxias with oculomotor apraxia," *Handbook of Clinical Neurology* (1st ed., Vol. 103). Elsevier B.V.
- Le Ber, I., M.-C. Moreira, S. Rivaud-Péchoux, C. Chamayou, F. Ochsner, T. Kuntzer, M. Tardieu, G. Saïd, M.-O. Habert, G. Demarquay, C. Tannier, J.-M. Beis, A. Brice, M. Koenig, and A. Dürr, 2003, "Cerebellar Ataxia with Oculomotor Apraxia Type 1: Clinical and Genetic Studies.", *Brain : a journal of neurology*, Vol. 126, No. Pt 12, pp. 2761–72.
- Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigani A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H. H. Won, D. Yun, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnely, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarrol, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, and D. G. MacArthur, 2016, "Analysis of Protein-Coding Genetic Variation in 60,706 Humans", *Nature*, Vol. 536, No. 7616, pp. 285–291.
- Li, H., and R. Durbin, 2009, "Fast and Accurate Short Read Alignment with Burrows-

- Wheeler Transform.", *Bioinformatics (Oxford, England)*, Vol. 25, No. 14, pp. 1754–60.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup, 2009, "The Sequence Alignment/Map Format and SAMtools", *Bioinformatics*, Vol. 25, No. 16, pp. 2078–2079.
- Magi, A., L. Tattini, F. Palombo, M. Benelli, A. Gialluisi, B. Giusti, R. Abbate, M. Seri, G. F. Gensini, G. Romeo, and T. Pippucci, 2014, "H3M2: Detection of Runs of Homozygosity from Whole-Exome Sequencing Data", *Bioinformatics (Oxford, England)*, Vol. 30, No. 20, pp. 2852–2859.
- Martin, M.-H., J.-P. Bouchard, M. Sylvain, O. St-Onge, and S. Truchon, 2007, "Autosomal Recessive Spastic Ataxia of Charlevoix-Saguenay: A Report of MR Imaging in 5 Patients", *American Journal of Neuroradiology*, Vol. 28, No. 8, pp. 1606–1608.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, 2010, "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data.", *Genome research*, Vol. 20, No. 9, pp. 1297–303.
- McKusick-Nathans Institute of Genetic Medicine, J. H. U. S. of M., www.omim.org, accessed at November 2016.
- McQuillan, R., A. L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. MacLeod, S. M. Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M. G. Dunlop, A. F. Wright, H. Campbell, and J. F. Wilson, 2008, "Runs of Homozygosity in European Populations", *American Journal of Human Genetics*, Vol. 83, No. 3, pp. 359–372.
- Montenegro, G., A. P. Rebelo, J. Connell, R. Allison, C. Babalini, M. D'Aloia, P. Montieri,

- R. Schüle, H. Ishiura, J. Price, A. Strickland, M. A. Gonzalez, L. Baumbach-Reardon, T. Deconinck, J. Huang, G. Bernardi, J. M. Vance, M. T. Rogers, S. Tsuji, P. De Jonghe, M. A. Pericak-Vance, L. Schöls, A. Orlandi, E. Reid, and S. Züchner, 2012, "Mutations in the ER-Shaping Protein Reticulon 2 Cause the Axon-Degenerative Disorder Hereditary Spastic Paraplegia Type 12.", *The Journal of clinical investigation*, Vol. 122, No. 2, pp. 538–44.
- Moreira, M. C., C. Barbot, N. Tachi, N. Kozuka, E. Uchida, T. Gibson, P. Mendonça, M. Costa, J. Barros, T. Yanagisawa, M. Watanabe, Y. Ikeda, M. Aoki, T. Nagata, P. Coutinho, J. Sequeiros, and M. Koenig, 2001, "The Gene Mutated in Ataxia-Ocular Apraxia 1 Encodes the New HIT/Zn-Finger Protein Aprataxin.", *Nature genetics*, Vol. 29, No. 2, pp. 189–93.
- Moreira, M.-C., S. Klur, M. Watanabe, A. H. Németh, I. Le Ber, J.-C. Moniz, C. Tranchant, P. Aubourg, M. Tazir, L. Schöls, M. Pandolfo, J. B. Schulz, J. Pouget, P. Calvas, M. Shizuka-Ikeda, M. Shoji, M. Tanaka, L. Izatt, C. E. Shaw, A. M'Zahem, E. Dunne, P. Bomont, T. Benhassine, N. Bouslam, G. Stevanin, A. Brice, J. Guimarães, P. Mendonça, C. Barbot, P. Coutinho, J. Sequeiros, A. Dürr, J.-M. Warter, and M. Koenig, 2004, "Senataxin, the Ortholog of a Yeast RNA Helicase, Is Mutant in Ataxia-Ocular Apraxia 2.", *Nature genetics*, Vol. 36, No. 3, pp. 225–7.
- Németh, A. H., E. Bochukova, E. Dunne, S. M. Huson, J. Elston, M. A. Hannan, M. Jackson, C. J. Chapman, A. M. R. Taylor, 2000, "Autosomal Recessive Cerebellar Ataxia with Oculomotor Apraxia (Ataxia-Telangiectasia-Like Syndrome) Is Linked to Chromosome 9q34", *The American Journal of Human Genetics*, Vol. 67, No. 5, pp. 1320–1326.
- Ng, P. C., and S. Henikoff, 2003, "SIFT: Predicting Amino Acid Changes That Affect Protein Function", *Nucleic Acids Research*, Vol. 31, No. 13, pp. 3812–3814.
- Ng, S. B., K. J. Buckingham, C. Lee, A. W. Bigham, H. K. Tabor, K. M. Dent, C. D. Huff, P. T. Shannon, E. W. Jabs, D. a Nickerson, J. Shendure, and M. J. Bamshad, 2010, "Exome Sequencing Identifies the Cause of a Mendelian Disorder.", *Nature genetics*,

Vol. 42, No. 1, pp. 30–35.

Noreau, A., C. V Bourassa, A. Szuto, A. Levert, S. Dobrzeniecka, J. Gauthier, S. Forlani, A. Durr, M. Anheim, G. Stevanin, A. Brice, J.-P. Bouchard, P. a Dion, N. Dupré, and G. a Rouleau, 2013, "SYNE1 Mutations in Autosomal Recessive Cerebellar Ataxia.", *JAMA neurology*, Vol. 70, No. 10, pp. 1296–31.

Nothnagel, M., T. T. Lu, M. Kayser, and M. Krawczak, 2010, "Genomic and Geographic Distribution of SNP-Defined Runs of Homozygosity in Europeans.", *Human molecular genetics*, Vol. 19, No. 15, pp. 2927–35.

Novarino, G., a G. Fenstermaker, M. S. Zaki, M. Hofree, J. L. Silhavy, a D. Heiberg, M. Abdellateef, B. Rosti, E. Scott, L. Mansour, A. Masri, H. Kayserili, J. Y. Al-Alama, G. M. Abdel-Salam, A. Karminejad, M. Kara, B. Kara, B. Bozorgmehri, T. Ben-Omran, F. Mojahedi, I. G. Mahmoud, N. Bouslam, A. Bouhouche, A. Benomar, S. Hanein, L. Raymond, S. Forlani, M. Mascaro, L. Selim, N. Shehata, N. Al-Allawi, P. S. Bindu, M. Azam, M. Gunel, A. Caglayan, K. Bilguvar, A. Tolun, M. Y. Issa, J. Schroth, E. G. Spencer, R. O. Rosti, N. Akizu, K. K. Vaux, A. Johansen, A. A. Koh, H. Megahed, A. Durr, A. Brice, G. Stevanin, S. B. Gabriel, T. Ideker, and J. G. Gleeson, 2014, "Exome Sequencing Links Corticospinal Motor Neuron Disease to Common Neurodegenerative Disorders", *Science*, Vol. 343, No. 6170, pp. 506–511.

O’Leary, N. A., M. W. Wright, J. R. Brister, S. Ciufu, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretin, ... K. D. Pruitt, 2016, "Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation", *Nucleic Acids Research*, Vol. 44, No. D1, pp. D733–D745.

Oba, D., M. Hayashi, M. Minamitani, S. Hamano, N. Uchisaka, A. Kikuchi, H. Kishimoto, M. Takagi, T. Morio, and S. Mizutani, 2010, "Autopsy Study of Cerebellar Degeneration in Siblings with Ataxia-Telangiectasia-like Disorder", *Acta Neuropathologica*, Vol. 119, No. 4, pp. 513–520.

- Ott, J., J. Wang, and S. M. Leal, 2015, "Genetic Linkage Analysis in the Age of Whole-Genome Sequencing", *Nature Reviews Genetics*, Vol. 16, No. 5, pp. 275–284.
- Özoğuz, A., Ö. Uyan, G. Birdal, C. Iskender, E. Kartal, S. Lahut, Ö. Ömür, Z. S. Agim, A. G. Eken, N. E. Sen, P. P. Kavak, C. Saygi, P. C. Sapp, P. Keagle, Y. Parman, E. Tan, F. Koç, F. Deymeer, P. Oflazer, H. Hanağası, H. Gürvit, B. Bilgiç, H. Durmuş, M. Ertaş, D. Kotan, M. A. Akalın, H. Güllüoğlu, M. Zarifoğlu, F. Aysal, N. Döşoğlu, K. Bilguvar, M. Günel, Ö. Keskin, T. Akgün, H. Özçelik, J. E. Landers, R. H. Brown, and A. N. Başak , 2015, "The Distinct Genetic Pattern of ALS in Turkey and Novel Mutations", *Neurobiology of Aging*, Vol. 36, No. 4, pp. 1764.e9-1764.e18.
- Parfitt, D. A., G. J. Michael, E. G. M. Vermeulen, N. V Prodrömu, T. R. Webb, J.-M. Gallo, M. E. Cheetham, W. S. Nicoll, G. L. Blatch, and J. P. Chapple, 2009, "The Ataxia Protein Sacsin Is a Functional Co-Chaperone That Protects against Polyglutamine-Expanded Ataxin-1.", *Human molecular genetics*, Vol. 18, No. 9, pp. 1556–65.
- Pedroso, J. L., C. R. R. Rocha, L. I. Macedo-Souza, V. De Mario, W. Marques, O. G. P. Barsottini, A. S. Bulle Oliveira, C. F. M. Menck, and F. Kok, 2015, "Mutation in PNKP Presenting Initially as Axonal Charcot-Marie-Tooth Disease", *Neurology: Genetics*, Vol. 1, No. 4, pp. e30–e30.
- Pierson, T. M., D. Adams, F. Bonn, P. Martinelli, P. F. Cherukuri, J. K. Teer, N. F. Hansen, P. Cruz, J. C. Mullikin For The Nisc Comparative Sequencing Program, R. W. Blakesley, G. Golas, J. Kwan, A. Sandler, K. Fuentes Fajardo, T. Markello, C. Tiffit, C. Blackstone, E. I. Rugarli, T. Langer, W. A. Gahl, C. Toro, J. C. Mullikin, R. W. Blakesley, G. Golas, J. Kwan, A. Sandler, K. Fajardo, T. Markello, C. Tiffit, C. Blackstone, E. I. Rugarli, T. Langer, W. A. Gahl, and C. Toro, 2011, "Whole-Exome Sequencing Identifies Homozygous AFG3L2 Mutations in a Spastic Ataxia-Neuropathy Syndrome Linked to Mitochondrial M-AAA Proteases", *PLoS Genetics*, Vol. 7, No. 10, pp. e1002325.
- Pippucci, T., A. Magi, A. Gialluisi, and G. Romeo, 2014, "Detection of Runs of Homozygosity from Whole Exome Sequencing Data: State of the Art and Perspectives

- for Clinical, Population and Epidemiological Studies", *Human Heredity*, Vol. 77, No. 1–4, pp. 63–72.
- Przedborski, S., M. Vila, V. Jackson-Lewis, 2003, "Series Introduction: Neurodegeneration: What Is It and Where Are We?", *Journal of Clinical Investigation*, Vol. 111, No. 1, pp. 3–10.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham, 2007, "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.", *American journal of human genetics*, Vol. 81, No. 3, pp. 559–75.
- Pyle, A., T. Smertenko, D. Bargiela, H. Griffin, J. Duff, M. Appleton, K. Douroudis, G. Pfeffer, M. Santibanez-Koref, G. Eglon, P. Yu-Wai-Man, V. Ramesh, R. Horvath, and P. F. Chinnery, 2015, "Exome Sequencing in Undiagnosed Inherited and Sporadic Ataxias", *Brain*, Vol. 138, No. 2, pp. 276–283.
- Rabbani, B., N. Mahdieh, K. Hosomichi, H. Nakaoka, and I. Inoue, 2012, "Next-Generation Sequencing: Impact of Exome Sequencing in Characterizing Mendelian Disorders", *Journal of Human Genetics*, Vol. 57, No. 10, pp. 621–632.
- Salisachs, P., L. J. Findley, M. Codina, P. La Torre, and J. M. Martinez-Lage, 1982, "A Case of Charcot-Marie-Tooth Disease Mimicking Friedreich's Ataxia: Is There Any Association between Friedreich's Ataxia and Charcot-Marie-Tooth Disease?", *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques*, Vol. 9, No. 2, pp. 99–103.
- Salomão, R. P. A., M. T. D. Gama, F. M. Rezende Filho, F. Maggi, J. L. Pedroso, and O. G. P. Barsottini, 2016, "Late-Onset Friedreich's Ataxia (LOFA) Mimicking Charcot-Marie-Tooth Disease Type 2: What Is Similar and What Is Different?", *Cerebellum (London, England)*.
- Savitsky, K., A. Bar-Shira, S. Gilad, G. Rotman, Y. Ziv, L. Vanagaite, D. A. Tagle, S. Smith,

- T. Uziel, S. Sfez, M. Ashkenazi, I. Pecker, M. Frydman, R. Harnik, S. R. Patanjali, A. Simmons, G. A. Clines, A. Sartiel, R. A. Gatti, L. Chessa, O. Sanal, M. F. Lavin, N. G. Jaspers, A. M. Taylor, C. F. Arlett, T. Miki, S. M. Weissman, M. Lovett, F. S. Collins, and Y. Shiloh, 1995, "A Single Ataxia Telangiectasia Gene with a Product Similar to PI-3 Kinase.", *Science (New York, N.Y.)*, Vol. 268, No. 5218, pp. 1749–53.
- Schöls, L., P. Bauer, T. Schmidt, T. Schulte, O. Riess, 2004, "Autosomal Dominant Cerebellar Ataxias: Clinical Features, Genetics, and Pathogenesis.", *The Lancet. Neurology*, Vol. 3, No. 5, pp. 291–304.
- Seelow, D., M. Schuelke, F. Hildebrandt, and P. Nürnberg, 2009, "HomozygosityMapper--an Interactive Approach to Homozygosity Mapping.", *Nucleic acids research*, Vol. 37, No. Web Server issue, pp. W593-9.
- Sherry, S. T., M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin, 2001, "dbSNP: The NCBI Database of Genetic Variation.", *Nucleic acids research*, Vol. 29, No. 1, pp. 308–311.
- Smedley, D., S. Haider, S. Durinck, L. Pandini, P. Provero, J. Allen, O. Arnaiz, M. H. Awedh, R. Baldock, G. Barbiera, P. Bardou, T. Beck, A. Blake, M. Bonierbale, A. J. Brookes, G. Bucci, I. Buetti, S. Burge, C. Cabau, J. W. Carlson, C. Chelala, C. Chrysostomou, D. Cittaro, O. Collin, R. Cordova, R. J. Cutts, E. Dassi, A. Di Genova, A. Dijari, A. Esposito, H. Estrella, E. Eyraç, J. Fernandez-Banet, S. Forbes, R. C. Free, T. Fujisawa, E. Gadaleta, J. M. Garcia-Manteiga, D. Goodstein, K. Gray, J. A. Guerra-Assunção, B. Haggarty, D. J. Han, B. W. Han, T. Harris, J. Harshbarger, R. K. Hastings, R. D. Hayes, C. Hoede, S. Hu, Z. L. Zu, L. Hutchins, Z. Kan, H. Kawaji, A. Keliet, A. Kerhornou, S. Kim, R. Kinsella, C. Klopp, L. Kong, D. Lawson, D. Lazarevic, J. H. Lee, T. Letellier, C. Y. Li, P. Lio, C. J. Liu, J. Luo, A. Maass, J. Mariette, T. Maurel, S. Marella, A. M. Mohamed, F. Moreews, I. Nabihoudine, N. Ndegwa, C. Noirot, C. Perez-Llamas, M. Primig, A. Quattrone, H. Quesneville, D. Rambaldi, J. Reecy, M. Riba, S. Rosanoff, A. A. Saddiq, E. Salas, O. Sallou, R. Shepherd, R. Simon, L. Sperling, W. Spooner, D. M. Staines, D. Steinbach, K. Stone, E. Stupka, J. W. Teague, A. Z. Dayem Ullah, J. Wang, D. Ware, M. Wong-Erasmus, K. Youens-Clark, A.

- Zadissa, S. J. Zhang, and A. Kasprzyk, 2015, "The BioMart Community Portal: An Innovative Alternative to Large, Centralized Data Repositories", *Nucleic Acids Research*, Vol. 43, No. W1, pp. W589–W598.
- Stevanin, G., F. M. Santorelli, H. Azzedine, P. Coutinho, J. Chomilier, P. S. Denora, E. Martin, A.-M. Ouvrard-Hernandez, A. Tessa, N. Bouslam, A. Lossos, P. Charles, J. L. Loureiro, N. Elleuch, C. Confavreux, V. T. Cruz, M. Ruberg, E. Leguern, D. Grid, M. Tazir, B. Fontaine, A. Filla, E. Bertini, A. Durr, and A. Brice, 2007, "Mutations in SPG11, Encoding Spatacsin, Are a Major Cause of Spastic Paraplegia with Thin Corpus Callosum.", *Nature genetics*, Vol. 39, No. 3, pp. 366–372.
- Stewart, G. S., R. S. Maser, T. Stankovic, D. A. Bressan, M. I. Kaplan, N. G. Jaspers, A. Raams, P. J. Byrd, J. H. Petrini, and A. M. R. Taylor, 1999, "The DNA Double-Strand Break Repair Gene hMRE11 Is Mutated in Individuals with an Ataxia-Telangiectasia-like Disorder", *Cell*, Vol. 99, No. 6, pp. 577–587.
- Suraweera, A., O. J. Becherel, P. Chen, N. Rundle, R. Woods, J. Nakamura, M. Gatei, C. Criscuolo, A. Filla, L. Chessa, M. Fusser, B. Epe, N. Gueven, and M. F. Lavin, 2007, "Senataxin, Defective in Ataxia Oculomotor Apraxia Type 2, Is Involved in the Defense against Oxidative DNA Damage", *Journal of Cell Biology*, Vol. 177, No. 6, pp. 969–979.
- Suraweera, A., Y. Lim, R. Woods, G. W. Birrell, T. Nasim, O. J. Becherel, and M. F. Lavin, 2009, "Functional Role for Senataxin, Defective in Ataxia Oculomotor Apraxia Type 2, in Transcriptional Regulation.", *Human molecular genetics*, Vol. 18, No. 18, pp. 3384–96.
- Swift, M., R. A. Heim, and N. J. Lench, 1993, "Genetic Aspects of Ataxia Telangiectasia.", *Advances in neurology*, Vol. 61, pp. 115–25.
- Synofzik, M., K. Smets, M. Mallaret, D. Di Bella, C. Gallenmüller, J. Baets, M. Schulze, S. Magri, E. Sarto, M. Mustafa, T. Deconinck, T. Haack, S. Züchner, M. Gonzalez, D. Timmann, C. Stendel, T. Klopstock, A. Durr, C. Tranchant, M. Sturm, W. Hamza, L.

- Nanetti, C. Mariotti, M. Koenig, L. Schöls, R. Schüle, P. de Jonghe, M. Anheim, F. Taroni, and P. Bauer, 2016, "SYNE1 Ataxia Is a Common Recessive Ataxia with Major Non-Cerebellar Features: A Large Scale Multi-Centre Study", *Brain*, pp. 1378-1393.
- Synofzik, M., A. S. Soehn, J. Gburek-Augustat, J. Schicks, K. N. Karle, R. Schüle, T. B. Haack, M. Schöning, S. Biskup, S. Rudnik-Schöneborn, J. Senderek, K.-T. Hoffmann, P. MacLeod, J. Schwarz, J. Bender, S. Krüger, F. Kreutz, P. Bauer, and L. Schöls, 2013, "Autosomal Recessive Spastic Ataxia of Charlevoix Saguenay (ARSACS): Expanding the Genetic, Clinical and Imaging Spectrum", *Orphanet Journal of Rare Diseases*, Vol. 8, No. 1, pp. 41.
- Tattini, L., R. D'Aurizio, and A. Magi, 2015, "Detection of Genomic Structural Variants from Next-Generation Sequencing Data.", *Frontiers in bioengineering and biotechnology*, Vol. 3, No. June, pp. 92.
- Taylor, A. M. R., A. Groom, and P. J. Byrd, 2004, "Ataxia-Telangiectasia-like Disorder (ATLD)—its Clinical Presentation and Molecular Basis", *DNA Repair*, Vol. 3, No. 8, pp. 1219–1225.
- Teer, J. K., E. D. Green, J. C. Mullikin, and L. G. Biesecker, 2012, "VarSifter: Visualizing and Analyzing Exome-Scale Sequence Variation Data on a Desktop Computer", *Bioinformatics*, Vol. 28, No. 4, pp. 599–600.
- Tsuji, S., 2013, "The Neurogenomics View of Neurological Diseases.", *JAMA neurology*, Vol. 70, No. 6, pp. 689–94.
- Uekawa, K., T. Yuasa, S. Kawasaki, T. Makibuchi, and T. Ideta, 1992, "[A Hereditary Ataxia Associated with Hypoalbuminemia and Hyperlipidemia--a Variant Form of Friedreich's Disease or a New Clinical Entity?].", *Rinshō shinkeigaku = Clinical neurology*, Vol. 32, No. 10, pp. 1067–74.
- Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V Garimella, D.

- Altshuler, S. Gabriel, and M. A. DePristo, 2013, "From FastQ Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline.", *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis*, Vol. 43, No. 1110, pp. 11.10.1-33.
- Vermeer, S., B. P. C. van de Warrenburg, M. a. a. P. Willemsen, M. Cluitmans, H. Scheffer, B. P. Kremer, and N. V. a. M. Knoers, 2011, "Autosomal Recessive Cerebellar Ataxias: The Current State of Affairs.", *Journal of medical genetics*, Vol. 48, No. 10, pp. 651–9.
- Vilariño-Güell, C., C. Wider, O. A. Ross, J. C. Dachsel, J. M. Kachergus, S. J. Lincoln, A. I. Soto-Ortolaza, S. A. Cobb, G. J. Wilhoite, J. A. Bacon, Bahareh Behrouz, H. L. Melrose, E. Hentati, A. Puschmann, D. M. Evans, E. Conibear, W. W. Wasserman, J. O. Aasly, P. R. Burkhard, R. Djaldetti, J. Ghika, F. Hentati, A. Krygowska-Wajs, T. Lynch, E. Melamed, A. Rajput, A. H. Rajput, A. Solida, R. M. Wu, R. J. Uitti, Z. K. Wszolek, F. Vingerhoets, and M. J. Farrer, 2011, "VPS35 Mutations in Parkinson Disease", *American Journal of Human Genetics*, Vol. 89, No. 1, pp. 162–167.
- Wang, J. L., X. Yang, K. Xia, Z. M. Hu, L. Weng, X. Jin, H. Jiang, P. Zhang, L. Shen, J. F. Guo, N. Li, Y. R. Li, FL. F. Lei, J. Zhou, J. Du, Y. F. Zhou, Q. Pan, J. Wang, J. Wang, R. Q. Li, and B. S. Tang, 2010, "TGM6 Identified as a Novel Causative Gene of Spinocerebellar Ataxias Using Exome Sequencing.", *Brain : a journal of neurology*, Vol. 133, No. Pt 12, pp. 3510–8.
- Wang, K., M. Li, and H. Hakonarson, 2010, "ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data", *Nucleic acids research*, Vol. 38, No. 16, pp. e164.
- Weizmann Institute of Science, 2016, GeneCards - Human Genes | Gene Database | Gene Search, <http://www.genecards.org>, accessed at November 2016.
- Wheeler, D. A., M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.-J. Chen, V. Makhijani, G. T. Roth, X. Gomes, K. Tartaro, F. Niazi, C. L. Turcotte, G. P.

- Irzyk, J. R. Lupski, C. Chinault, X. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D. M. Muzny, M. Margulies, G. M. Weinstock, R. A. Gibbs, and J. M. Rothberg, 2008, "The Complete Genome of an Individual by Massively Parallel DNA Sequencing.", *Nature*, Vol. 452, No. 7189, pp. 872–6.
- Wiethoff, S., J. Hersheson, C. C. Bettencourt, N. W. Wood, and H. Houlden, 2016, "Heterogeneity in Clinical Features and Disease Severity in Ataxia-Associated SYNE1 Mutations", *J Neurol*, pp. 1–8.
- Wright, S., 1977, "Evolution and the Genetics of Populations, Volume 3: Experimental Results and Evolutionary Deductions."
- Yahikozawa, H., K. Yoshida, S. Sato, N. Hanyu, H. Doi, S. Miyatake, and N. Matsumoto, 2015, "Predominant Cerebellar Phenotype in Spastic Paraplegia 7 (SPG7)", *Human Genome Variation*, Vol. 2, No. November 2014, pp. 15012.
- Zhao, C., J. Takita, Y. Tanaka, M. Setou, T. Nakagawa, S. Takeda, H. W. Yang, S. Terada, T. Nakata, Y. Takei, M. Saito, S. Tsuji, Y. Hayashi, and N. Hirokawa, 2001, "Charcot-Marie-Tooth Disease Type 2A Caused by Mutation in a Microtubule Motor KIF1Bbeta.", *Cell*, Vol. 105, No. 5, pp. 587–97.
- Zimprich, A., A. Benet-Pagès, W. Struhal, E. Graf, S. H. Eck, M. N. Offman, D. Haubenberger, S. Spielberger, E. C. Schulte, P. Lichtner, S. C. Rossle, N. Klopp, E. Wolf, K. Seppi, W. Pirker, S. Presslauer, B. Mollenhauer, R. Katzenschlager, T. Foki, C. Hotzy, E. Reinthaler, A. Harutyunyan, R. Kralovics, A. Peters, F. Zimprich, T. Brücke, W. Poewe, E. Auff, C. Trenkwalder, B. Rost, G. Ransmayr, J. Winkelmann, T. Meitinger, and T. M. Strom, 2011, "A Mutation in VPS35, Encoding a Subunit of the Retromer Complex, Causes Late-Onset Parkinson Disease", *American Journal of Human Genetics*, Vol. 89, No. 1, pp. 168–175.

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

Table A1. List of alignment commands

Command	Function
<code>bwa aln -t 2000 -f \$sampleID_R1.sai \$reference \$sampleID_R1.fastq.gz</code>	actual alignment of the reads to the reference genome
<code>bwa sampe -r "\$RG" \$reference \$sampleID_R1.sai \$sampleID_R2.sai \$sampleID_R1.fastq.gz \$sampleID_R2.fastq.gz</code>	creating aligned reads
<code>samtools view -bS -> \$sampleID.bam</code>	converting aligned reads to binary format
<code>samtools sort \$sampleID.bam \$sampleID.sorted</code>	sorting aligned reads by position
<code>java -jar picard.jar MarkDuplicates INPUT=\$sampleID.sorted.bam OUTPUT=\$sampleID.rmdup.bam ASSUME_SORTED=TRUE METRICS_FILE=/dev/null VALIDATION_STRINGENCY=SILENT REMOVE_DUPLICATES=true CREATE_INDEX=true</code>	marking PCR duplicates, bam file also indexed in this step

Table A2. Arguments used in bam file processing and variant calling

Command	Function
<code>java -jar GenomeAnalysisTK -T RealignerTargetCreator -R \$reference -I \$sampleID.rmdup.bam -o \$sampleID.rmdup.bam.intervals -nt 3 -known Mills_and_1000G_gold_standard.indels.b37.vcf -known 1000G_phase1.indels.b37.vcf</code>	realignment around indels
<code>java -jar GenomeAnalysisTK -T IndelRealigner -targetIntervals \$sampleID.rmdup.bam.intervals -R \$reference -I \$sampleID.rmdup.bam -known Mills_and_1000G_gold_standard.indels.b37.vcf -known 1000G_phase1.indels.b37.vcf -o \$sampleID.realigned.bam</code>	
<code>java -jar GenomeAnalysisTK -T BaseRecalibrator -I \$sampleID.realigned.bam -R \$reference -knownSites dbsnp_138.b37.vcf -nct 4 -o \$sampleID.report.grp -lqt 2 -mdq -1</code>	base quality score recalibration
<code>java -jar GenomeAnalysisTK -T PrintReads -R \$reference -I \$sampleID.realigned.bam -nct 4 -BQSR \$sampleID.report.grp -o \$sampleID.final.bam</code>	

Table A2. Arguments used in bam file processing and variant calling (cont.)

Command	Function
<pre>java -jar GenomeAnalysisTK -T HaplotypeCaller -R \$reference -I \$sampleID.final.bam --doNotRunPhysicalPhasing --emitRefConfidence GVCF -- dbsnp dbsnp_138.b37.vcf -stand_call_conf 30 -stand_emit_conf 10 -gt_mode DISCOVERY -nct 4 -mbq 20 -G Standard -A AlleleBalance -o \$sampleID.raw.snps.indels.g.vcf</pre>	variant calling, a genomic variant call format is obtained
<pre>java -jar GenomeAnalysisTK -T GenotypeGVCFs -R \$reference --variant \$sampleID.raw.snps.indels.g.vcf -o \$sampleID.raw.snps.indels.vcf</pre>	variant calling, only variant sites are reported*
<pre>java -jar GenomeAnalysisTK -T VariantAnnotator -R \$reference -o \$sampleID.ann.snp.indel.vcf -A Coverage -A InbreedingCoeff --variant \$sampleID.raw.snps.indels.vcf -L \$sampleID.raw.snps.indels.vcf --dbsnp dbsnp_138.b37.vcf</pre>	annotation of the variants
<pre>java -jar GenomeAnalysisTK -T VariantRecalibrator -R \$reference -input \$sampleID.ann.snp.indel.vcf - resource:hapmap,VCF,known=true,training=true,truth=true,prior=15.0 hapmap_3.3.b37.vcf - resource:omni,VCF,known=true,training=true,truth=true,prior=12.0 1000G_omni2.5.b37.vcf - resource:dbsnp,VCF,known=true,training=true,truth=true,prior=6.0 dbsnp_138.b37.vcf -an QD -an MQRankSum -an ReadPosRankSum -an FS -an MQ -mode SNP -recalFile \$sampleID.snp.recal -tranchesFile \$sampleID.snp.tranches - rscriptFile \$sampleID.snp.plots.R -nt 6 --maxGaussians 4 --TStranche 100.0 -- TStranche 99.9 --TStranche 99.5 --TStranche 99.0 --TStranche 98.0 --TStranche 97.0 --TStranche 95.</pre>	variant score recalibration by SNP
<pre>java -jar GenomeAnalysisTK -T ApplyRecalibration -R \$reference -input \$sampleID.ann.snp.indel.vcf --ts_filter_level 99.0 -recalFile \$sampleID.snp.recal - tranchesFile \$sampleID.snp.tranches -mode SNP -o \$sampleID.snp.vqsr.vcf</pre>	
<pre>java -jar GenomeAnalysisTK -T VariantRecalibrator -R \$reference -input \$sampleID.snp.vqsr.vcf - resource:mills,known=true,training=true,truth=true,prior=12.0 Mills_and_1000G_gold_standard.indels.b37.vcf - resource:dbsnp,VCF,known=true,training=true,truth=true,prior=6.0 dbsnp_138.b37.vcf -an QD -an DP -an FS -an SOR -an MQRankSum -an ReadPosRankSum -mode INDEL -recalFile \$sampleID.indel.recal -tranchesFile \$sampleID.indel.tranches -rscriptFile \$sampleID.indel.R</pre>	variant score recalibration by INDELS
<pre>java -jar GenomeAnalysisTK -T ApplyRecalibration -R \$reference --input \$sampleID.snp.vqsr.vcf -mode INDEL --ts_filter_level 99.0 -recalFile \$sampleID.indel.recal -tranchesFile \$sampleID.indel.tranches -o \$sampleID.snp.indel.vqsr.vcf</pre>	

* This step is also repeated with more than one gvcf to obtain family vcf files

Table A3. List of commands used in annotation and analysis

Command	Function
perl convert2annovar.pl \$sampleID.vcf -format vcf4old -includeinfo > \$sampleID.anno	conversion of vcf file to annovar input
perl summarize_annovar.pl -out \$sampleID.annotated -buildver hg19 -verdb SNP138 -ver1000g 1000g2012apr -veresp 6500 -remove -alltranscript \$sampleID.anno humandb	actual annotation
perl -pe 'while (s/("[^"]+),\^1<COMMA>/g) {1}; s/"//g; s/,\^t/g; s/<COMMA>/,/g' < \$sampleID.annotated.genome_summary.csv > \$sampleID.annotated.genome_summary.txt	file format conversion: comma-separated to tab-delimited
python wannotate_pinar.py \$sampleID.final.vcf \$sampleID.annotated.genome_summary.txt \$sampleID.final	addition of information from OMIM database
java -jar Varsifter.jar	running VarSifter to analyze annotated vcf files

APPENDIX B: SEQUENCING QUALITY METRICS

Table B1. Sequencing quality metric values of all participants

		Coverage	Missingness	Ts/Tv
Family 1	Individual 1	30.351	0.00409	2.678
	Individual 2	29.481	0.00818	2.672
	Individual 3	29.348	0.00613	2.668
	Individual 4	30.526	0.00818	2.683
	Individual 5	29.136	0.00426	2.700
Family 2	Individual 6	31.229	0.01573	2.683
	Individual 7	31.642	0.00674	2.686
	Individual 8	31.979	0.00225	2.645
Family 3	Individual 9	28.952	0.00632	2.723
	Individual 10	26.916	0.00842	2.71
	Individual 11	31.352	0.00842	2.695
	Individual 12	32.342	0.00421	2.708
	Individual 13	31.4923	0.00421	2.701
Family 4	Individual 14	28.039	0.00833	2.682
	Individual 15	32.138	0.00833	2.693
	Individual 16	27.958	0.00625	2.708
	Individual 17	30.091	0.00208	2.68
Family 5	Individual 18	34.490	0.00213	2.684
	Individual 19	30.439	0.00426	2.665
	Individual 20	30.376	0.00426	2.697
Family 6	Individual 21	28.607	0.00809	2.447
	Individual 22	28.283	0.01417	2.437
	Individual 23	28.826	0.01214	2.449
	Individual 24	27.598	0.01022	2.446
Family 7	Individual 25	24.046	0.00661	2.446
	Individual 26	21.937	0.01762	2.486
	Individual 27	18.017	0.00661	2.444
Family 8	Individual 28	35.715	0.06856	2.497
	Individual 29	39.266	0.01655	2.485
	Individual 30	35.766	0.01182	2.509
	Individual 31	34.114	0.01418	2.484
	Individual 32	38.649	0.00709	2.513
Family 9	Individual 33	35.723	0.00265	2.527
	Individual 34	29.861	0.07162	2.532
	Individual 35	39.811	0.02122	2.543
	Individual 36	40.219	0.02387	2.526
	Individual 37	47.900	0.27321	2.511
Family 10	Individual 38	107.627	0.29661	2.601
	Individual 39	82.312	0.25706	2.575
	Individual 40	93.203	0.24858	2.564
	Individual 41	98.724	0.25424	2.559
Family 11	Individual 42	20.094	0.05238	2.121
Family 12	Individual 43	32.892	0.02655	2.564
Family 13	Individual 44	22.944	0.00839	2.441
	Individual 45	26.798	0.01048	2.438
	Individual 46	29.404	0.00839	2.439
	Individual 47	23.902	0.01258	2.45

Table B1. Sequencing quality metric values of all participants (cont.)

		Coverage	Missingness	Ts/Tv
Family 14	Individual 48	34.044	0.05528	2.656
	Individual 49	33.000	0.00000	2.678
Family 15	Individual 50	32.923	0.42721	2.497
Family 16	Individual 51	26.863	0.00000	2.160
Family 17	Individual 52	36.580	0.01754	2.521
	Individual 53	36.917	0.01253	2.526
	Individual 54	38.762	0.01253	2.541
	Individual 55	33.016	0.06266	2.533
	Individual 56	32.889	0.02005	2.526
Family 18	Individual 57	93.195	0.37309	1.740
	Individual 58	95.591	0.34862	1.777
	Individual 59	107.000	0.35168	1.733
Family 19	Individual 60	29.893	0.07143	1.883
	Individual 61	18.032	0.09244	1.845
	Individual 62	22.750	0.08403	1.942
	Individual 63	28.541	0.04622	1.701
Family 20	Individual 64	21.541	0.03810	2.238
Family 21	Individual 65	33.860	0.00835	2.230
	Individual 66	29.121	0.00835	2.294
	Individual 67	29.605	0.00626	2.190
	Individual 68	29.239	0.00626	2.398
Family 22	Individual 69	23.318	0.47245	2.262
Family 23	Individual 70	33.843	0.00197	2.720
	Individual 71	30.471	0.00789	2.762
	Individual 72	33.215	0.00789	2.465
Family 24	Individual 73	30.306	0.00752	2.795
	Individual 74	29.157	0.00188	2.465
	Individual 75	29.317	0.00376	2.991
	Individual 76	32.026	0.00940	2.367
Family 25	Individual 77	31.0812	0.00000	2.928
	Individual 78	27.463	0.01259	2.636
Family 26	Individual 79	37.769	0.43353	2.283
Family 27	Individual 80	9.839	0.01810	2.378
	Individual 81	5.286	0.06787	2.368
	Individual 82	15.754	0.00226	2.015
	Individual 83	9.839	0.07014	1.790
Family 28	Individual 84	11.102	0.19688	2.217
	Individual 85	45.216	0.00313	1.859
	Individual 86	18.792	0.08750	1.901
	Individual 87	19.837	0.10000	2.014
	Individual 88	18.622	0.08125	2.333
	Individual 89	14.900	0.14375	1.923
	Individual 90	20.487	0.09063	2.205
Family 29	Individual 91	61.118	0.00183	2.208
Family 30	Individual 92	16.768	0.03605	1.936
	Individual 93	24.674	0.00949	1.902
	Individual 94	24.760	0.01708	1.832
	Individual 95	24.847	0.01898	2.000
	Individual 96	25.304	0.00949	1.858
Family 31	Individual 97	58.180	0.18590	2.550

Table B1. Sequencing quality metric values of all participants (cont.)

		Coverage	Missingness	Ts/Tv
Family 32	Individual 98	32.306	0.00177	2.577
	Individual 99	31.302	0.00708	2.571
	Individual 100	29.443	0.00531	2.765
	Individual 101	33.498	0.00177	2.981
	Individual 102	28.304	0.01239	2.927
	Individual 103	29.533	0.00354	2.351
	Individual 104	30.712	0.00354	2.694
Family 33	Individual 105	31.037	0.00762	2.547
	Individual 106	27.193	0.00190	2.343
	Individual 107	30.155	0.00952	2.391
	Individual 108	32.118	0.00762	2.336
Family 34	Individual 109	29.088	0.00816	2.209
	Individual 110	29.387	0.01020	2.39
	Individual 111	27.115	0.00204	2.336
	Individual 112	30.110	0.00612	2.362
	Individual 113	28.739	0.01837	2.277
	Individual 114	26.967	0.01020	2.431
Family 35	Individual 115	30.485	0.00442	2.210
	Individual 116	29.000	0.00000	2.580
	Individual 117	28.168	0.00662	2.139
Family 36	Individual 118	30.175	0.00758	2.351
	Individual 119	26.034	0.01010	2.805
Family 37	Individual 120	31.227	0.00654	2.118
	Individual 121	28.275	0.00871	2.368
	Individual 122	28.264	0.00654	2.192
Family 38	Individual 123	30.214	0.00885	2.475
	Individual 124	30.307	0.00885	2.647
	Individual 125	29.361	0.00664	2.661
Family 39	Individual 126	27.920	0.00469	2.704
	Individual 127	31.918	0.00000	2.672
Family 40	Individual 128	27.683	0.01062	2.383
	Individual 129	29.644	0.00849	2.559
	Individual 130	29.434	0.00212	2.264
Family 41	Individual 131	28.824	0.01033	2.232
	Individual 132	30.439	0.00516	2.828
Family 42	Individual 133	31.051	0.01033	2.473
	Individual 134	32.694	0.00344	2.444
Family 43	Individual 135	29.489	0.00213	2.595
	Individual 136	27.489	0.00426	2.525
	Individual 137	28.066	0.00426	2.678
	Individual 138	28.667	0.00640	2.349
Family 44	Individual 138	51.808	0.00641	2.364
Family 45	Individual 140	29.468	0.00593	2.310
	Individual 141	28.418	0.01780	2.495
	Individual 142	28.500	0.67062	4.944
Family 46	Individual 143	27.723	0.24719	3.053
	Individual 144	35.340	0.01404	2.319
	Individual 145	43.976	0.22753	3.340
	Individual 146	43.330	0.04494	2.455
Family 47	Individual 147	31.492	0.01195	1.915
	Individual 148	22.613	0.04382	2.329
	Individual 149	30.766	0.02788	1.875
	Individual 150	16.989	0.07968	2.229