

THE MOLECULAR LANDSCAPE OF ALS IN TURKEY:
A MULTIFACETED APPROACH TO THE COMPLEX GENETICS OF ALS

by

Ceren Tunca

B.S., Biological Sciences and Bioengineering, Sabancı University, 2010

M.S., Molecular Biology and Genetics, Boğaziçi University, 2013

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

Graduate Program in Molecular Biology and Genetics
Boğaziçi University

2018

To all our patients and their families...

ACKNOWLEDGEMENTS

I express my deepest gratitude to my supervisor Prof. A. Nazlı Başak for her invaluable guidance, support and strong encouragement throughout this thesis. I thank her for offering endless opportunities and for sharing her scientific and personal experiences.

I would like to further thank Prof. Hande Çağlayan, Prof. Esra Battaloğlu, Prof. Piraye Oflazer, Assoc. Prof. Sibel Uğur İşeri and Assist. Prof. Öznur Taştan for devoting their time to evaluate my thesis.

I humbly am grateful to all patients and family members for their cooperation and would like to specially offer my thanks to the esteemed clinicians throughout Turkey who have been collaborating with us. I respectfully acknowledge the Suna and İnan Kıraç Foundation for financially supporting me and my research.

I deeply thank all current and former NDAL members, for a truly fruitful working environment and for their sincere friendship that will last for many more years. I extend my thanks to Suna for all our creative accomplishments and Fulya for supporting me through my first steps in bioinformatics. I further thank Tuncay for his help in bioinformatic data analysis and Dr. Atay Vural for sharing his thoughts on data presentation. I am also especially grateful to Cemre for her excellent assistance in all experiments.

Last but not least, I heartily would like to express my appreciation and admiration to my parents Ece and Cenk İskender for their endless support and unconditional love and genuinely thank my husband Can Tunca for collaborating with me in the greatest experiment ever, life.

ABSTRACT

THE MOLECULAR LANDSCAPE OF ALS IN TURKEY: A MULTIFACETED APPROACH TO THE COMPLEX GENETICS OF ALS

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease (NDD), characterized by degeneration of both upper and lower motor neurons leading to muscle wasting. With an incidence of two and a prevalence of four in 100.000/year, it is the third most common NDD after Alzheimer's and Parkinson's diseases. The last decade has proven that the 'condition' called ALS is both clinically and genetically heterogenous, and that the genetic component in 90% of the cases that are considered as sporadic, might be stronger than expected. In this thesis, we investigated the complex genetics of ALS in the ethnically heterogeneous yet inbred Turkish population. Familial ALS account for 22% of our cases (159/722) and 563 cases were classified as sporadic (78%). Consanguinity is calculated as 25% among familial and 17% in sporadic cases. Conventional screening of the most common ALS genes (*C9orf72*, *SOD1*, *TARDBP* and *FUS*) in patients with a family history of disease, explained the disease cause in only 33%, pointing towards marked locus heterogeneity within the population. *C9orf72* hexanucleotide repeat expansion was further detected in 3% of 'apparently sporadic' patients. Application of whole exome sequencing (WES) in dominant and recessive Turkish pedigrees presenting with ALS or non-ALS motor neuron diseases (MNDs), revealed distinct rare or novel mutations in 20 out of 39 families and enabled differential diagnosis in cases with atypical ALS features. Common current themes in ALS, like oligogenic inheritance and possible genetic modifiers are addressed in families with incomplete penetrance and in *C9orf72* expansion carriers. Oligogenic inheritance of known ALS genes was not prominent in the Turkish *C9orf72*-positive cases. Hypermethylation in the promoter region of *C9orf72* was confirmed in Turkish patients, however it did not seem to modify age of onset. Finally, analyses of whole genome sequencing data of 625 Turkish ALS patients and 152 healthy controls in the framework of Project MinE showed population-specific aspects and once more substantiated the concept of low penetrance of rare ALS genes among sporadic patients.

ÖZET

TÜRKİYE’DE ALS’NİN MOLEKÜLER YAPISI: KARMAŞIK ALS GENETİĞİNE ÇOK YÖNLÜ BİR YAKLAŞIM

Amiyotrofik lateral skleroz (ALS), alt ve üst motor nöron dejenerasyonun kas zayıflığı ve felce yol açtığı ilerleyici bir nörodejeneratif hastalıktır (NDH). ALS, yıllık 2/100.000 insidans ve 4/100.000 prevalans ile Alzheimer ve Parkinson hastalıklarından sonra en yaygın görülen NDH’dir. Geçtiğimiz on yılda ‘ALS’ olarak adlandırılan fenotipin klinik ve genetik olarak çok heterojen olduğu, ayrıca sporadik olarak kabul edilen ve hastaların %90’ını oluşturan olgulardaki genetik bileşenin tahmin edilenden çok daha fazla olduğu kanıtlandı. Bu tez çerçevesinde, bir yandan etnik olarak oldukça heterojen, diğer taraftan yüksek akraba evliliği oranının etkisinde olan Türk toplumunda ALS’nin karmaşık genetiği araştırıldı. İncelenen tüm hastaların %22’sini (159/722) ailesel olgular oluşturmaktadır, 563 olgu ise sporadik olarak sınıflandırılmıştır (%78). Anne-baba akrabalığı ailesel olgularda %25, sporadik olgularda ise %17 olarak hesaplanmıştır. Ailesel vakalarda yaygın ALS genlerinin (*C9orf72*, *SOD1*, *TARDBP* ve *FUS*) PCR-tabanlı yöntemler ile incelenmesi olguların ancak %33’ünü açıklamıştır, bu bulgu toplumdaki lokus heterojenitesine işaret etmektedir. *C9orf72* tekrar artışı mutasyonu ‘sporadik’ olguların da %3’lük bir kısmını açıklamaktadır. Tezde tüm ekzom dizileme yöntemi dominant ve resesif ALS ve ALS-dışı motor nöron hastalarında 39 ailenin 20’sinde nadir ve yeni mutasyonlar tanımlandı. Oligogenik kalıtım ve modifiye edici genetik faktörler gibi ALS genetiğinin güncel konuları penetrans düşüklüğü gösteren ailelerde ve *C9orf72*-pozitif olgularda incelendi. *C9orf72*-pozitif bulgularda oligogenik kalıtımı öne çıkmadı. *C9orf72* geninin promotor bölgesindeki hipermetilasyon kohortumuzda da gösterilmekle birlikte, hastalık başlangıç yaşı ile arasında bir ilişki gözlemlenmedi. Tezin son aşamasında, MinE Projesi kapsamında tüm genom dizilemeyle incelenen 625 ALS hastası ve 152 sağlıklı kontrolün analizi, topluma özgü özellikleri ortaya çıkardığı gibi, sporadik olgularda da nadir ALS genlerinin etkin ama düşük penetranslı olduğunu tekrar vurguladı.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iv
ABSTRACT.....	v
ÖZET	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES	xi
LIST OF TABLES	xiv
LIST OF SYMBOLS	xvii
LIST OF ACRONYMS/ABREVIATIONS.....	xviii
1. INTRODUCTION	1
1.1. Neurodegenerative Diseases: The Umbrella Term	1
1.2. Motor Neuron Disorders and ALS	1
1.3. Gene Discovery in ALS and Impact of Next Generation Sequencing	5
1.4. Overlapping Features of ALS and Other MNDs.....	9
1.5. Autosomal Recessive Cerebellar Ataxias and Their Overlaps with ALS	10
1.6. Mechanisms Involved in ALS	11
1.7. C9orf72 as the Main Player in ALS	15
1.7.1. Discovery and Clinical Characteristics	15
1.7.2. Function and Suggested Mechanisms Leading to C9orf72 Pathology....	17
1.7.3. Promoter Hypermethylation	20
1.7.4. Modifiers of C9orf72-based Disease.....	22
1.8. The Big Project of ALS: Project MinE	23
2. PURPOSE.....	25
3. MATERIALS	26
3.1. Subjects	26
3.1.1. The ALS Cohort Under Study	26
3.1.2. Cases subjected to WES	27
3.1.2.1. Individuals Diagnosed with ALS/MND.....	27
3.1.2.2. Families Showing Incomplete Penetrance	27
3.1.3. Project MinE.....	27

3.2.	Equipment, Solutions, Chemicals and Kits	30
3.2.1.	DNA Isolation	30
3.2.2.	Agarose Gel Electrophoresis and Gel Extraction	30
3.2.3.	Genotyping Experiments	30
3.2.4.	Detection of C9orf72 Promoter Methylation Levels	33
3.2.4.1.	Bisulfite Conversion	33
3.2.4.2.	Methylation-specific Restriction Enzyme Digestion	33
3.3.	Next Generation Sequencing Platforms and Enrichment Kits	33
3.4.	Hardware	34
3.5.	Software, Online Databases and Bioinformatic Tools	34
3.6.	General Laboratory Equipment	34
4.	METHODS	37
4.1.	DNA Isolation from Whole Blood	37
4.2.	Agarose Gel Electrophoresis and Extraction from the Gel	37
4.3.	Screening of Common ALS-causing Genes	37
4.4.	Detection of C9orf72 Promoter Methylation Levels	44
4.4.1.	Bisulfite Sequencing	44
4.4.2.	Methylation-specific Restriction Enzyme Digestion	46
4.5.	Whole Exome Sequencing Analyses (WES)	48
4.5.1.	Validation and Segregation Analyses in Families Subjected to WES	49
4.6.	Homozygosity Mapping	50
4.7.	Data Analyses in Samples from Project MinE	51
4.7.1.	Variant Accumulation in sALS Patients	51
4.7.2.	Pathogenic Variations in sALS Patients	52
5.	RESULTS	53
5.1.	Screening of Common ALS Genes by Conventional Approaches	53
5.1.1.	C9orf72 hexanucleotide Repeat Expansion Analyses	53
5.1.2.	SOD1, TARDBP, FUS and UBQLN2 Analyses	55
5.1.3.	Three Rare SOD1 Mutations with Distinct Inheritance Pattern	58
5.1.3.1.	SOD1-Asn86Ser	58
5.1.3.2.	SOD1-Leu117Val	58
5.1.3.3.	SOD1-Glu133Lys	59
5.2.	Promoter Methylation Levels of C9orf72 Expansion Carriers	61

5.3.	Whole Exome Sequencing Analyses for Identification of Mutations in Rare and Novel ALS Genes.....	63
5.3.1.	Families with an Autosomal Dominant Inheritance Pattern.....	66
5.3.1.1.	Kinesin Family Member 5A (KIF5A) (Family 1).....	66
5.3.1.2.	Tank-binding Kinase 1 (TBK1) (Family 2)	67
5.3.1.3.	TRK-Fused Gene (TFG) (Family 19)	67
5.3.1.4.	MORC Family CW-Type Zinc Finger 2 (MORC2) (Family 20).....	68
5.3.2.	Families with Autosomal Recessive Inheritance Pattern	69
5.3.2.1.	Spatacsin (SPG11) (Families 4, 5, 6, 7)	69
5.3.2.2.	Synaptic Nuclear Envelope 1 (SYNE1) (Family 8 and 9)	71
5.3.2.3.	Endoplasmic Reticulum Lipid Raft-Associated Protein 1 (ERLIN1) (Family 10)	73
5.3.2.4.	Zinc Finger FYVE Domain-Containing Protein 26 (ZFYVE26) (Family 11).....	77
5.3.2.5.	DNAJ/HSP40 Homolog, Subfamily B, Member 2 (DNAJB2) (Family 12).....	77
5.3.2.6.	Sigma Nonopioid Intracellular Receptor 1 (SIGMAR1) (Family 13).....	78
5.3.2.7.	Set-Binding Factor 1 (SBF1) (Family 14).....	79
5.3.2.8.	Paraplegin (SPG7) (Family 16 and 17)	80
5.3.2.9.	Vaccinia Related Kinase 1 (VRK1) (Family 18)	81
5.3.3.	A Family with a Dual ALS Phenotype.....	82
5.3.4.	Search for Disease Modifiers and Oligogenic Inheritance in ALS	85
5.3.4.1.	Families Showing Incomplete Penetrance	85
5.3.4.2.	Oligogenic Inheritance and Modifiers in C9orf72-positive Cases.....	88
5.4.	Whole Genome Sequencing Analyses in Sporadic ALS Cases: Project MinE...	89
5.4.1.	Variant Accumulation in sALS Patients.....	89
5.4.2.	Pathogenic Variations in sALS Cases	89
6.	DISCUSSION.....	95
6.1.	Challenging Nature of ALS and Common Disputes	95
6.2.	Clinical Presentation of ALS in Turkey	97

6.3. Impact of Common ALS Genes on ALS in Turkey	99
6.3.1. The Dynamic C9orf72 Expansion Mutation and Promoter Hypermethylation	100
6.3.2. Allelic Heterogeneity in SOD1 and Genotype-Phenotype Correlations	103
6.3.3. de novo FUS Mutations in the Turkish Cohort	106
6.3.4. Two UBQLN2 Variations with Unknown Significance (VUS).....	107
6.4. Whole Exome Sequencing Analyses.....	108
6.4.1. Mutations in Rare ALS Genes.....	109
6.4.2. Differential Diagnosis in non-ALS MNDs.....	111
6.4.3. Phenotypic Heterogeneity Caused by SPG11 Mutations	114
6.4.4. SYNE1 in the Turkish ALS Cohort.....	115
6.4.5. A Novel ALS Gene: ERLIN1.....	116
6.4.6. Clinical Exome Sequencing: Where Are We?	118
6.5. Incomplete Penetrance in ALS, Disease Modifiers and Oligogenic Inheritance	120
6.6. Preliminary Analyses in Project MinE Samples	123
6.7. Conclusion.....	125
REFERENCES	128
APPENDIX A: NUMBER OF VARIANTS REMAINING IN VARIANT FILTRATION STEPS	155
APPENDIX B: COMMANDS EXECUTED IN ANALYSES OF WHOLE EXOME SEQUENCING DATA.....	157
APPENDIX C: PUBLICATIONS DURING THE THESIS	159

LIST OF FIGURES

Figure 1.1. Impact of NGS to gene discovery in ALS.....	8
Figure 1.2. Mechanisms involved in ALS pathogenesis	12
Figure 1.3. <i>C9orf72</i> gene and its transcripts	17
Figure 4.1. Repeat primed PCR design for the amplification of <i>C9orf72</i> hexanucleotide repeat expansion	38
Figure 4.2. Bisulfite conversion reaction diagram.....	46
Figure 4.3. Filtration criteria of variants in Project MinE samples	52
Figure 5.1. Example for <i>C9orf72</i> RP-PCR result	53
Figure 5.2. Distribution of the repeat sizes in the non-expanded allele of the <i>C9orf72</i> gene	54
Figure 5.3. Locations of amino acid changes identified in SOD1, TDP-43, FUS and UBQLN2 proteins	57
Figure 5.4. Pedigree of the family with a heterozygous SOD1-p.Asn86Ser variation.....	59
Figure 5.5. Pedigrees of the families with the SOD1-p.Leu117Val variation.....	60
Figure 5.6. Pedigree of the family with a heterozygous SOD1-p.Glu133Lys variation.....	61

Figure 5.7. Correlation between age of onset and the number of methylated CpG sites	62
Figure 5.8. Correlation between two distinct methylation assays.	63
Figure 5.9. Pedigree of Family 1	66
Figure 5.10. Pedigree of Family 2	67
Figure 5.11. Pedigree of Family 19	68
Figure 5.12. Pedigree of Family 20	69
Figure 5.13. Pedigrees of Families 4 and 5.....	70
Figure 5.14. Pedigrees of Families 6 and 7.....	71
Figure 5.15. Pedigrees of Families 8 and 9.....	72
Figure 5.16. Pedigree of Family 10	75
Figure 5.17. Pedigree of Family 11.	77
Figure 5.18. Pedigree of Family 12	78
Figure 5.19. Pedigree of Family 13	79
Figure 5.20. Pedigree of Family 14	80
Figure 5.21. Pedigrees of Families 16 and 17.....	81
Figure 5.22. Pedigree of Family 18	82
Figure 5.23. Pedigree of Family 37	84

Figure 5.24. Families with low or incomplete penetrance.....	86
Figure 5.25. Number of variants per individuals in controls and patients subjected to Project MinE	91
Figure 6.1. KIF5A protein domains and variations identified in the Turkish cohort	109

LIST OF TABLES

Table 1.1. Causative/implicated genes in ALS	3
Table 3.1. Clinical data of the Turkish cohort under study	26
Table 3.2. Clinical information of ALS/MND cases investigated with WES	28
Table 3.3. Materials used in agarose gel electrophoresis	30
Table 3.4. Primer sequences used in screening of five common ALS genes	31
Table 3.5. Primer sequences used in validation of WES results and segregation analyses	32
Table 3.6. Sequencing platforms and enrichment kits	33
Table 3.7. Software, online databases and bioinformatic tools	35
Table 3.8. General laboratory equipment and instruments	36
Table 4.1. PCR reagents for C9orf72 RP-PCR	39
Table 4.2. PCR reagents for C9orf72 flanking PCR	39
Table 4.3. PCR conditions for C9orf72 flanking PCR	40
Table 4.4. PCR conditions for C9orf72 flanking PCR	41
Table 4.5. PCR reagents for SOD1 Exon1, Exon3	41
Table 4.6. PCR reagents for SOD1 Exon2	41

Table 4.7. PCR reagents for SOD1 Exon4, Exon5	42
Table 4.8. PCR reagents for TARDBP Exon5, FUS Exons 14 and 15	42
Table 4.9. PCR reagents for UBQLN2 primer sets 1 and 4.....	42
Table 4.10. PCR reagents for UBQLN2 primer sets 2, 3 and 5.....	43
Table 4.11. PCR conditions for SOD1 Exons 1,3,4 and 5.....	43
Table 4.12. PCR conditions for SOD1 Exon2	43
Table 4.13. PCR conditions for TARDBP and FUS.....	44
Table 4.14. PCR conditions for UBQLN2 primer sets	44
Table 4.15. PCR reagents for nested PCR reactions of bisulfite sequencing	47
Table 4.16. Nested PCR conditions for bisulfite sequencing	47
Table 4.17. PCR reagents for amplification after RE digestion with HhaI	47
Table 4.18. Conditions for PCR amplification after HhaI restriction enzyme digestion.....	48
Table 4.19. Reagents for validation and segregation PCRs.....	49
Table 4.20. Conditions for validation and segregation PCRs.....	50
Table 4.21. PLINK parameters for homozygosity mapping.....	51
Table 5.1. Clinical data of <i>C9orf72</i> hexanucleotide expansion carriers in our cohort	54
Table 5.2. Clinical data of patients with <i>SOD1</i> , <i>TARDBP</i> , <i>FUS</i> and <i>UBQLN2</i> mutations.	56

Table 5.3. <i>C9orf72</i> promoter methylation levels of expansion carriers versus controls	62
Table 5.4. Variants identified in families subjected to WES and their OMIM associations.....	64
Table 5.5. Information regarding the location, conservation and minor allele frequencies of the variants identified.....	65
Table 5.6. ROH on chromosome 10 of individuals with WES data.....	76
Table 5.7. VarSifter filters in families with incomplete penetrance	88
Table 5.8. Rare ALS causing/associated variants identified via whole genome sequencing in the framework of Project MinE.....	92

LIST OF SYMBOLS

g	Gram
G	G-force
kb	Kilobase
mM	Milimolar
mg	Miligram
ml	Mililiter
M	Molar
ng	Nanogram
V	Volt
μg	Microgram
μl	Microliter
μM	Micromolar
*	Asterix
°C	Centigrade degree
#	Number
%	Percentage

LIST OF ACRONYMS/ABBREVIATIONS

5mC	5 methyl cytosine
5-hmC	5-hydroxymethylcytosine
ABCC2	ATP Binding Cassette Subfamily C Member
ACMG	American College of Medical Genetics and Genomics
AD	Alzheimer's Disease
AD	Autosomal Recessive
AF	Allele Frequency
ALAD	Aminolevulinate Dehydratase
ALS	Amyotrophic Lateral Sclerosis
ALS2	Alsin
ALS-FTD	Amyotrophic Lateral Sclerosis-Frontotemporal Dementia
ALYREF	Aly/REF Export Factor
AMP	Association for Molecular Pathology
AMPK	AMP-activated Protein Kinase
ANG	Angiogenin
ANXA11	Anexin A11
AO	Age of Onset
AOA	Ataxia with Oculomotor Apraxia
AR	Autosomal Recessive
ARCA1	Autosomal Recessive Cerebellar Ataxia Type 1
ARHSP	Autosomal recessive Hereditary Spastic Paraplegia
ARJALS	Autosomal Recessive Juvenile ALS
ASO	Antisense Oligonucleotide
AT	Ataxia Telangiectasia
ATXN2	Ataxin 2
BAM	Binary Alignment Map
BST-PCR	Bisulfite Sequencing
bvFTD	Behavioral Variant Frontotemporal Dementia
BVVL	Brown-Vialetto-Van Laere syndrome
BWA	Burrows-Wheeler Aligner

C19orf12	Chromosome 19 Open Reading Frame 12
C21orf2	Chromosome 21 Open Reading Frame 2
C9orf72	Chromosome 9 Open Reading Frame 72
CCNF	Cyclin F
CNS	Central Nervous System
CHCHD10	Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10
CHMP2B	Charged Multivesicular Body Protein 2B
CMT2X	Charcot Marie Tooth Type 2X
CMT4B3	Charcot Marie Tooth Type 4B3
Ca ²⁺	Calcium Ion
DANN	Deep Learning Approach for Annotation
DAO	D-amino-acid oxidase
DCTN1	Dynactin 1
DENN	Differentially Expressed in Normal and Neoplasia
DNA	Deoxyribonucleic Acid
DNAJB2	DnaJ Homolog Subfamily B Member 2
Dnmts	DNA methyltransferases
dNTP	Deoxyribonucleotide Triphosphate
DM1	Myotonic Dystrophy 1
DPR	Dipeptide Repeat
DTR	Deep Tendon Reflexes
EDTA	Ethylenediaminetetraacetic Acid
EEC	El Escorial Criteria
EMG	Electromyography
ER	Endoplasmic Reticulum
ERBB4	Erb-B2 Receptor Tyrosine Kinase 4
ERLIN1	Endoplasmic Lipid Raft Associated Protein 1
ExAC	Exome Aggregation Consortium
fALS	Familial Amyotrophic Lateral Sclerosis
FIG4	Polyphosphoinositide phosphatase
FRDA	Friedreich's Ataxia
FTD	Frontotemporal Dementia
FUS	Fused in Sarcoma

GATK	Genome Analysis Toolkit
gDNA	Genomic Deoxyribonucleic Acid
gVCF	Genomic Variant Call Format
GWAS	Genome Wide Association Studies
GOF	Gain of Function
GRN	Granulin
H ₂ O	Water
HGP	Human Genome Project
hnRNP-H	Heterogenous Ribonucleoprotein Particle H
HSMN	Hereditary Sensory Motor Neuropathy
HSP	Hereditary Spastic Paraplegia
IGV	Integrated Genomic Viewer
KIF5A	Kinesin Family 5A
KO	Knock-out
LCD	Low Complexity Domain
LMN	Lower Motor Neuron
LOF	Loss of Function
MAPT	Microtubule-Associated Protein Tau
MATR3	Matrin 3
MEP	Motor Evoked Potential
miRNA	Micro Ribonucleic Acid
MND	Motor Neuron Disease
MOBP	Myelin-Associated Oligodendrocyte Basic Protein
MORC2	MORC Family CW-Type Zinc Finger 2
MRI	Magnetic Resonance Imaging
mRNA	Messenger Ribonucleic Acid
MTMR2	Myotubularin Related Protein 2
NBIA	Neurodegeneration with Brain Iron Accumulation
NDAL	Neurodegeneration Research Laboratory
NDDs	Neurodegenerative diseases
NEFH	Neurofilament, Heavy Polypeptide
NEK1	NIMA-related Kinase 1
NGS	Next Generation Sequencing

NHLBI	National Heart, Lung and Blood Institute
NLS	Nuclear Localization Signal
NMJ	Neuromuscular Junction
OMIM	Online Mendelian Inheritance in Man
OPTN	Optineurin
PD	Parkinson's disease
PCH1A	Pontocerebellar Hypoplasia Type 1A
PCR	Polymerase Chain Reaction
PEG	Percutaneous Endoscopic Gastrostomy
PFN1	Profilin 1
PLS	Primary Lateral Sclerosis
PMA	Progressive Muscle Atrophy
PMP22	Peripheral Myosin Protein 22
PRPH	Peripherin
RAN	Repeat Associated Non-ATG
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
ROH	Runs of Homozygosity
ROS	Reactive Oxygen Species
RVAS	Rare Variant Association Studies
SACS	Spastic Ataxia of Charlevoix-Saguenay
sALS	Sporadic Amyotrophic Lateral Sclerosis
SAM	Sequence Alignment Map
SBF1	Set Binding Factor 1
SBMA	Spino-bulbar Muscular Atrophy
SCA	Spinocerebellar Ataxia
SCFD1	Sec1 Family Domain Containing
SETX	Senataxin
SIGMAR1	Sigma Non-Opioid Intracellular Receptor 1
SMA	Spinal Muscular Atrophy
SOD1	Superoxide Dismutase1
SPG	Spastic Paraplegia
SQSTM1	Sequestosome 1

SYNE1	Synaptic Nuclear Envelope 1
TAF15	TATA-binding protein-associated factor 2N
TARDBP/ TDP-43	TAR DNA binding protein
TBE	Tris/Borate/EDTA
TBK1	TANK Binding Kinase 1
TCC	Thin Corpus Callosum
TET1	ten-eleven translocation methylcytosine dioxygenase 1
TFG	TRK-Fused Gene
TMEM106B	Transmembrane Protein 106B
tRNA	Transfer Ribonucleic Acid
UBQLN2	Ubiquilin 2
UMN	Upper Motor Neurons
UNC13A	Unc-13 Homolog A
UPR	Unfolded Protein Response
UV	Ultraviolet
VABP	Vesicle-associated Membrane Protein-associated Protein B
VCP	Valosin-Containing Protein
VRK1	Vaccinia Related Kinase 1
VUS	Variant of Unknown Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
wt	Wild-Type
XLD	X-linked Dominant
ZFYVE26	Zinc Finger FYVE-type Containing 26

1. INTRODUCTION

1.1. Neurodegenerative Diseases: The Umbrella Term

Neurodegenerative diseases (NDs) are a heterogeneous group of progressive disorders that effect specific populations of cells in the central and peripheral nervous systems. Progressive loss of neuronal structures and function impair motor or cognitive abilities resulting in irreversible symptoms. Although distinct areas are vulnerable in different NDs, there are many overlapping clinical and pathological features suggesting for common affected pathways. Almost 90% of all NDs are complex disorders where multiple gene mutations may result in the same phenotype and/or even environmental and epigenetic factors may play role in the disease pathogenesis. The remaining 10% is inherited according to Mendelian laws, but even in this small group the genetic background is highly heterogeneous. The era of high-throughput sequencing enabled more detailed and sophisticated research in understanding the genetic deformities behind these complex diseases. Many rare disease-causing genes are identified that were missed previously with linkage analysis and population screens. With the identification of novel causative genes and mutations, it is now known that the same phenotype can result from different gene mutations, but also one gene can cause two different phenotypes. Thus, diseases that are once accepted as distinct are now being considered as the diseases of the same spectrum and it has become crucial to understand the converging affected pathways.

1.2. Motor Neuron Disorders and ALS

Motor neuron disorders (MNDs) are a group of progressive, fatal neurodegenerative diseases in which motor neurons that are responsible for making the connection between the brain and the muscles for initiating voluntary movements are affected. Normally, the signal initiated from the motor cortex in the brain is transmitted through the upper motor neurons (UMNs) to the lower motor neurons (LMNs) in the brain stem and spinal cord that further carry the impulse to the skeletal muscles. MNDs are traditionally classified according to the affected motor neuron type in this circuit. For example, in primary lateral sclerosis (PLS) or in hereditary spastic paraplegias (HSPs) only the upper motor neurons

are affected, and this leads to spasticity and hyperexcitability; whereas in progressive muscle atrophy (PMA), lower motor neurons are affected leading to fasciculations due to the atrophy and degeneration of the muscles. In amyotrophic lateral sclerosis (ALS) or spinal-bulbar muscular atrophy (SBMA), both of the motor neuron types are affected, leading to a combination of symptoms. SBMA is a monogenic X-linked disease resulting from mutations in the androgen receptor gene and affects the males, whereas in ALS multiple patterns of inheritance may be observed according to the gene mutated and both genders carry the same disease risk.

ALS is a progressive disease of the motor neurons, where the upper motor neurons of the cortex and lower motor neurons in the brainstem and spinal cord are both affected. With an incidence of two and a prevalence of four in 100.000, it is the third most common neurodegenerative disease after Alzheimer's and Parkinson's diseases. The average age of onset of the disease is between 55 and 60, however juvenile cases are also observed (Ghasemi and Brown, 2018). Two-thirds of ALS patients present with spinal onset, including arms and legs, the rest show bulbar onset disease with dysarthria or swallowing problems. In addition to motor symptoms, studies point out to cognitive and behavioral changes in almost 50% of ALS cases (van Es *et al.*, 2017). ALS symptoms emerge focally and spread throughout the body; progressive nature of the disease results in death of the patients within 2-5 years due to respiratory failure (Pasinelli & Brown 2006; Kiernan *et al.*, 2011). The survival can be prolonged for some more years with proper medical care, e.g. tracheostomy and percutaneous endoscopic gastrostomy (PEG), however there is no cure for ALS (Brown and Al-Chalabi, 2017).

The most commonly prescribed drug for ALS is riluzole, a chemical compound that is designed to control hyperexcitability of motor neurons, however patient benefit is at minimum (Sreedharan and Brown, 2013). Two new therapeutic options for ALS patients are edaravone, a free radical scavenger and masitinib, a protein kinase inhibitor that inhibits the neuroinflammatory reactions (Petrov *et al.*, 2017). The efficacy outcomes of these drugs are still being investigated.

As with many neurodegenerative diseases, most ALS cases are sporadic (sALS) and only about 10% show a family history, mostly with autosomal dominant inheritance

(fALS). Both familial and sporadic forms of ALS result in the same phenotype and are clinically indistinguishable. Many of the gene mutations that are known to give rise to familial ALS have also been shown to be instrumental in sporadic disease pathogenesis (Brown and Al-Chalabi, 2017). The major genes linked to ALS pathogenesis are *Chromosome 9 open reading frame 72 (C9orf72)*, *Superoxide dismutase 1 (SOD1)*, *TAR DNA binding protein (TARDBP)*, whose protein product is TDP-43 and *Fused in Sarcoma (FUS)* which together explain about 40-50% of all familial cases. *Ubiquilin-2 (UBQLN2)*, *valosin containing protein (VCP)*, *angiogenin (ANG)*, *profilin 1 (PFN1)* and *optineurin (OPTN)* are among other genes, each implicated in 1-2% of cases (Ghasemi and Brown, 2018; Renton *et al.*, 2013). Although the majority of ALS genes are inherited in an autosomal dominant and *UBQLN2* in an X-linked dominant manner, autosomal recessive inheritance is also observed as in *OPTN*, *alsin (ALS2)*, *spatacsin (SPG11)* and some well-known variations in the *SOD1* gene (Andersen and Al-Chalabi, 2011; Ticozzi *et al.*, 2011). Genes identified as causative in ALS and genes that are associated with the disease are listed in Table 1.1.

Table 1.1. Causative/implicated genes in ALS, adapted from Ghasemi and Brown 2017.

Gene	Locus	Fraction fALS (%)	Phenotype	Inheritance	Possible mechanism
<i>C9ORF72</i>	9p21.3	40-50	ALS, ALS+FTD, FTD	AD	RNA toxicity
<i>SOD1</i>	21q22	20-25	ALS	AD, AR	Protein dyshomeostasis
<i>TARDBP</i>	1p36.2	4-5	ALS, ALS+FTD, FTD	AD	RNA toxicity
<i>FUS</i>	16p11.2	4-5	ALS, ALS+FTD, FTD	AD	RNA toxicity
<i>OPTN</i>	10p13	2-3	ALS, ALS+FTD	AD, AR	Protein dyshomeostasis
<i>PFN1</i>	17p13	1-2	ALS	AD	Cytoskeletal defects
<i>VCP</i>	9p13	1-2	ALS, ALS+FTD, FTD	AD	Protein dyshomeostasis
<i>ANG</i>	14q11.2	1	ALS, ALS+FTD, FTD	AD	RNA toxicity
<i>TUBA4A</i>	2q35	<1	ALS, ALS+FTD	AD	Cytoskeletal defects
<i>UBQLN2</i>	Xp11	<1	ALS, ALS+FTD, FTD	XLD	Protein dyshomeostasis
<i>TAF15</i>	17q11	<1	ALS	AD	RNA toxicity
<i>EWSR1</i>	22q12.2	<1	ALS	AD	RNA toxicity
<i>hnRNPA1</i>	12q13	<1	ALS, ALS+FTD, FTD	AD	RNA toxicity
<i>hnRNPA2B1</i>	7p15	<1	ALS, ALS+FTD, FTD	AD	RNA toxicity
<i>SETX</i>	9q34.13	<1	ALS	AD	Cytoskeletal defects
<i>CREST</i>	20q13.3	<1	ALS	-	Cytoskeletal defects

Table 1.1. Causative/implicated genes in ALS, adapted from Ghasemi and Brown 2017
(cont.).

Gene	Locus	Fraction fALS (%)	Phenotype	Inheritance	Possible mechanism
<i>MATR3</i>	5q31.2	<1	ALS, ALS+FTD	AD	RNA toxicity
<i>ATXN2</i>	12q24	<1	ALS, ALS+FTD	AD	RNA toxicity
<i>ELP3</i>	8p21.1	<1	ALS	-	Cytoskeletal defects
<i>SQSTM1</i>	5q35	<1	ALS, ALS+FTD, FTD	AD	Protein dyshomeostasis
<i>CHMP2B</i>	3p11	<1	ALS, FTD	AD	Protein dyshomeostasis
<i>ALS2</i>	2q33.1	<1	ALS, PLS	AR	Cytoskeletal defects
<i>VAPB</i>	20q13	<1	ALS, PLS	AD	Protein dyshomeostasis
<i>SIGMAR1</i>	9p13.3	<1	ALS, ALS+FTD, FTD	AR	Protein dyshomeostasis
<i>DCTN1</i>	2p13	<1	ALS	AD, AR	Cytoskeletal defects
<i>FIG4</i>	6q21	<1	ALS, PLS	AD	Protein dyshomeostasis
<i>SPG11</i>	15q21.1	<1	ALS, HSP	AR	Cytoskeletal defects
<i>NEFH</i>	22q12.2	<1	ALS	AD, AR	Cytoskeletal defects
<i>PRPH</i>	12q13	<1	ALS	AD, AR	Cytoskeletal defects
<i>NTE</i>	19p13	<1	ALS, HSP	AR	Protein dyshomeostasis
<i>PONI-3</i>	7q21	<1	ALS	-	Oxidative stress
<i>DAO</i>	12q22	<1	ALS	AD	Protein dyshomeostasis
<i>CHRNA3</i> , <i>CHRNA4</i> , <i>CHRNA4</i>	15q24, 20q13, 15q24	<1	ALS	-	Calcium toxicity
<i>ERBB4</i>	2q34	<1	ALS	AD	-
<i>CHCHD10</i>	22q11	<1	ALS+FTD	AD	Mitochondrial dysfunction
<i>C19ORF12</i>	9q12	<1	ALS, NBIA	AR	-
<i>ALS3</i>	18q21	<1	ALS	-	-
<i>ALS7</i>	20p13	<1	ALS	-	-
<i>ALS6-21</i>	6p25, 21q22	<1	ALS	-	-
<i>ALS-FTD</i>	16p12	<1	ALS+FTD	-	RNA toxicity
<i>TBK1</i>	12q14.2	<1	ALS+FTD	AD	Protein dyshomeostasis
<i>CCNF</i>	16p13.3	<1	ALS+FTD	AD	Protein dyshomeostasis

ALS: amyotrophic lateral sclerosis, FTD: frontotemporal dementia, PLS: primary lateral sclerosis, HSP: hereditary spastic paraplegia, NBIA: neurodegeneration with brain iron accumulation, AD: autosomal dominant, AR: autosomal recessive, XLD: X-linked dominant

The evidence of fALS genetics contributing to sALS is very promising for the understanding of sporadic disease, however it also raises questions upon the current classification. It is suggested that cases that harbor mutations in known genes are ‘apparently sporadic’, which in turn will inherit the mutation to the next generation and

that the real sporadic disease emerges from interaction of genetic and environmental/epigenetic factors (Brown and Al-Chalabi, 2017).

Although contribution of environmental factors to ALS pathogenesis is undoubted, it is very hard to study such parameters. The effect of several environmental factors like extensive exercise, smoking, chemical exposure, exposure to heavy metals, etc. have been studied, however there are no replicable or conclusive studies yet (Ingre *et al.*, 2015; van Es *et al.*, 2017). As in all late-onset neurodegenerative diseases, aging is accepted to be one of the main players in disease pathogenesis. Al-Chalabi and Hardiman suggest the GTE model for neurodegeneration, where the genomic load (G) of the individual, together with time (T) and environmental factors (E) build-up to trigger the disease onset (Al-Chalabi and Hardiman, 2013). In this accumulation process, several risk and protective factors that may be either genetic or environmental play role in the onset and progression of the disease. The intermediate length CAG repeat expansions (27-33 repeats) in the *Ataxin-2* (*ATXN2*) gene are the best representation of an inherited risk factor in ALS (Elden *et al.*, 2011). Other than the inherited factors, epigenetic factors that are very much regulated by the environmental effects can add up to the genetic load of the individual without actually making heritable changes. We now know that, advancing technologies and emerging consortium projects are promising to better search for these factors and puzzle out the complexity behind ALS.

1.3. Gene Discovery in ALS and Impact of Next Generation Sequencing

In early 90s linkage analysis in several ALS families detected a region on chromosome 21q that was inherited with the disease among affected family members (Siddique *et al.*, 1991). Linkage analysis is used to trace genetic markers that are inherited together with a frequency higher than expected by chance, due to their close proximity on the genome. The search for a candidate gene in the specified region revealed the *SOD1* gene, which was shown to be mutated in 13 different ALS families with a very high allelic-heterogeneity of 11 different mutations widespread on the small SOD1 protein made up of 153 amino acids (Rosen *et al.*, 1993). After this discovery several loci have been linked to ALS, however there was time gap until the identification of *TARDBP* and *FUS* genes in 2008 and 2009, respectively (Kwiatkowski *et al.*, 2009; Sreedharan *et al.*, 2008).

Linkage studies are successful in identifying strong genetic signals, but they are limited in power when applied in small families with a few number of affected cases and when the genetic defect displays low penetrance (Ghasemi and Brown, 2018). The Human Genome Project (HGP) enabled the application of genome-wide association studies (GWAS), which target the diseases on single nucleotide level by comparing the enrichment of a SNP in cases more than in controls (Hood and Rowen, 2013). GWAS became beneficial in identifying disease-associated alleles even with low power, in large populations harboring individuals with different genetic backgrounds. The loci at chromosome 9p21, previously linked to ALS was replicated using GWAS, pointing to the *C9orf72* gene, which then led to the discovery of a pathologic expansion in the promoter region of the gene (van Es *et al.*, 2009).

In the last decade, GWAS data based on SNP arrays was used to identify novel SNPs associated with several phenotypes, especially in complex disorders affecting the majority of the human population, like cardiovascular diseases, diabetes and cancer; however, GWAS lacked the ability to target rare changes that are mostly implicated in rare diseases (Pranavchand and Reddy, 2016). Next generation sequencing (NGS) technologies that adopt the principal of high-throughput sequencing of DNA fragments, offer time and cost-effective sequencing of the whole genome (WGS), whole exome (WES) or transcriptome (RNA-seq).

Today, population screens of the common genes that cause ALS are made by conventional methods mostly in families with a dominant inheritance pattern and with affected members in the upper generations. Families that do not respond to mutational analyses of common genes are then subjected to WES to identify rare/novel genes implicated in ALS pathogenesis. Association of *VCP*, *PFN1* and *v-erb-a erythroblastic leukemia viral oncogene homolog 4 (ERBB4)* with ALS are among the first examples (Johnson *et al.*, 2010; Takahashi *et al.*, 2013; Wu *et al.*, 2012). In cases where affected individual(s) with healthy parents are observed, consanguinity of the parents is questioned, and WES approach is interiorized to search for recessive mutations in rare or novel genes. In this purpose, parent-child trios are the optimal starting point, where each additional affected individual will decrease the number of candidate variants and increase the chance

of targeting a causative mutation (Bamshad *et al.*, 2011; Ott *et al.*, 2015). Homozygosity mapping, a method that detects homozygous stretches within a defined threshold, can be applied to WES data to narrow down the number of candidate genes. The discovery rate of WES was found to be very high for detecting recessive and *de novo* mutations (Bamshad *et al.*, 2011; Boycott *et al.*, 2013). For the filtration steps data of healthy relatives can be used for the elimination of benign polymorphisms running through the family. Filtering of variants against publicly available sets of polymorphisms in an approach called discrete-filtering, proved to be extremely useful (Bamshad *et al.*, 2011).

In addition to family-based studies using NGS, WES performed on large cohorts is very promising to identify novel disease-causing genes. This approach was applied to 2869 ALS patients and 6405 controls to reveal the gene encoding for *TANK-binding kinase 1* (*TBK1*) as a novel ALS gene although it is very rare (Cirulli *et al.*, 2015). Following studies with similar sample sizes were conducted to associate *NIMA-related kinase* (*NEK1*), *cyclin-F* (*CCNF*), *Annexin A11* (*ANXA11*) and *kinesin family member 5A* (*KIF5A*) to ALS (Brenner *et al.*, 2018; Chia *et al.*, 2018; Smith *et al.*, 2017). Nowadays, the other widely used approach is to couple WGS with GWAS allowing association of rare SNPs with disease phenotype in search for rare genetic factors (Al-Chalabi *et al.*, 2017; Auer and Lettre, 2015). This method confirmed the implication of C-terminal KIF5A mutations in ALS and found that variations in *chromosome 21 open reading frame 2* (*C21orf2*) confers ALS risk (Nicolas *et al.* 2018; van Rheenen *et al.* 2016). Fascinating is that although these gene mutations are very rare, the interaction between them and their protein products suggests a strong connection for functional consequences. It is shown that MATR3 is in physical bond with well-known proteins mutated in ALS, TDP-43 and FUS and NEK1 interacts with ALS2, VAPB and the newly identified C21orf2 (Chia *et al.*, 2018).

Advancing sequencing technologies and candidate gene approaches explained above fueled the identification of causative genes implicated in ALS in the last 20 years and helped researchers to learn clues about common cellular pathways behind the genetic complexity of ALS (Figure 1.1). Furthermore, identification of ALS genes contributing to other motor neuron diseases like PMA, HSP etc., expanded the concept of ALS and suggested possible subtypes (van Blitterswijk *et al.*, 2012a). It is known that pleiotropic genes can result in different clinical phenotypes and thus gather different diseases under

the same roof. The complexity is also increased with the presence of multiple ALS-causing mutations in patients displaying evidence for an oligogenic basis of ALS (van Blitterswijk *et al.*, 2012b). The presence of oligogenic inheritance in ALS pathogenesis might be further addressed with (i) the identification of several low-impact ALS genes that can each trigger small cellular events and have a combinatorial effect or (ii) the incomplete penetrance observed in families carrying mutations in high-impact ALS genes suggesting for multiple genetic factors that effects the disease process (Nguyen *et al.*, 2018).

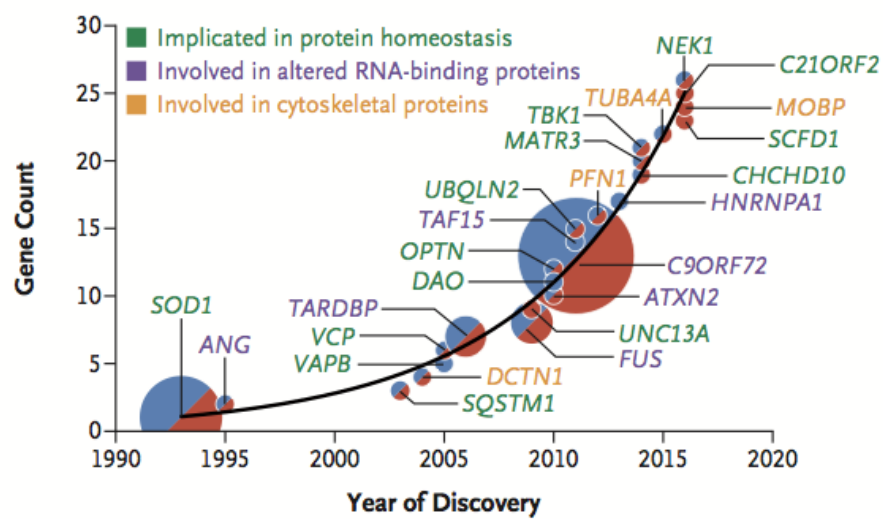


Figure 1.1. Impact of NGS to gene discovery in ALS (Brown and Al-Chalabi, 2017). Color codes of genes show their contribution in different biological processes, explained in detail in Section 1.6. Circle sizes represent the proportion of associated cases (blue: fALS, red: sALS).

NGS enabled novel gene discovery in fALS cases and further helped the differential diagnoses of cases with recessive inheritance patterns, however for the sporadic ALS cases that are not explained by the screening of common gene mutations, further analysis is still limited. The possible involvement of several genetic factors in the sporadic disease makes it harder to target the causative/risk factors in disease pathogenesis. For this purpose, an initiative has been started under the name Project MinE (Section 1.8).

1.4. Overlapping Features of ALS and Other MNDs

The differential diagnosis of MNDs are especially challenging when the disease is in its early stages and patients present with atypical symptoms. ALS diagnosis may often cover several phenotypes that are (i) similar to ALS, like progressive SMA; (ii) more benign like PLS and HSP and (iii) ALS-mimic syndromes (van Es *et al.*, 2017).

Although ALS is a late-onset and fast progressive disease, there are several patients fulfilling the clinical criteria of ALS diagnosis at adolescence or even younger. These cases that usually have consanguineous parents, show slower disease progression and may display additional clinical evidence prominent in other diseases like cognitive impairment, cerebellar atrophy, polyneuropathy and other extrapyramidal features. With the advancement of NGS, it has become possible to perform molecular analysis on these cases in an unbiased way, independent of the initial diagnosis.

One of the atypical forms of ALS is defined as autosomal recessive juvenile ALS (ARJALS). *Alsin* (*ALS2*) and *spatacsin* (*SPG11*) are the most common genes that give rise to ARJALS and it is shown that the disease representation of patients carrying mutations in these genes manifest overlapping clinical features with autosomal recessive hereditary spastic paraplegia (ARHSP) patients (Orlacchio *et al.*, 2010; Yang *et al.*, 2001). ARHSP is a non-pure form of HSP, where the affected UMNs lead to spasticity and muscle weakness. Another phenotype resembling juvenile ALS is a subtype of neurodegeneration with brain iron accumulation (NBIA) caused by defects in the *Chromosome 19 open reading frame 12* (*C19orf12*) (Deschauer *et al.*, 2012). Furthermore, it is reported that patients with distal hereditary motor neuropathy (dHMN) carrying *DNAJB2* mutations display an ALS phenotype at the later stages of the disease (Frasquet and Sevilla 2017). In addition to the autosomal recessive forms, evidence also shows common features between a dominant form of HSP, caused by mutations in the *SPG6* (*spastic paraplegia 6*) gene; where the patient with complicated HSP features and cognitive decline was found positive for TDP-43 inclusions that are the hallmark of ALS (Martinez-Lage *et al.*, 2012).

WES analysis made a considerable progress in the identification of novel causative disease genes, but maybe, its most important contribution was that it removed the

boundaries between several neurological diseases and helped us understand the overlapping factors. Even more, the concept of gene-based disease is now considered; new genetic discoveries represent a new clinical picture including features of previously defined diseases.

1.5. Autosomal Recessive Cerebellar Ataxias and Their Overlaps with ALS

Hereditary ataxias are seen in autosomal dominant, autosomal recessive, X-linked or mitochondrial forms. In this decade, there have been great advances in unraveling the molecular basis of hereditary ataxias, especially autosomal dominant spinocerebellar ataxias (SCAs) (Sailer and Houlden, 2012). However, a few causative genes are identified for the autosomal recessive forms including Friedreich's ataxia (FRDA), autosomal recessive ataxia type 1 (ARCA1), ataxia with Vitamin E deficiency, ataxia telangiectasia (AT), recessive spastic ataxia of Charlevoix-Saguenay (ARSACS) and ataxia with oculomotor apraxia types 1 and 2 (AOA1 and AOA2).

Autosomal recessive cerebellar ataxias (ARCAs) represent a heterogeneous group of rare clinical entities, in which although the cerebellum is primarily affected, additional symptoms often exist including neuropathy, pyramidal or extrapyramidal symptoms, cognitive deficits and oculomotor abnormalities (Anheim *et al.*, 2012). The age of disease onset is mostly below 20 years. ARCAs are distinguished according to their pathobiology and each newly identified causative gene explains a specific clinical subtype very often.

In addition to the broad clinical diversity among different rare recessive ataxias, it has been shown that even mutations in the same gene can result in further complexity by bringing ataxia and other neurodegenerative diseases together. It is known that recessive mutations in the *senataxin* (*SETX*) gene with RNA/DNA helicase activity cause AOA2, whereas individuals carrying a dominant *SETX* mutation develop an autosomal dominant form of juvenile ALS. More complexity is introduced by the recessive truncation mutations in the *spectrin repeat containing, nuclear envelope 1* (*SYNE1*) gene, whose loss of function can cause either a pure ataxia called autosomal recessive cerebellar ataxia type 1 (ARCA1) or ataxia accompanying motor neuron disease (Gros-Louis *et al.*, 2007;

Synofzik *et al.*, 2016). In addition, it was previously known that dominant mutations in the *SYNE1* gene are the cause of Emery-Dreifuss muscular dystrophy (Zhang *et al.*, 2007).

It is now very important to consider and interpret every clinical detail rather than focusing on the initial diagnosis of the patient. Clinical details that do not fit the classical clinical picture of a disease might give clues about the gene responsible for that picture. NGS in that sense, enable us to look for genes responsible of distinct phenotypes segregating in one family and to differentially diagnose the patient. Each new genetic discovery broadens the borders of a clinical entity and provides insights about the possible mechanisms leading to pathogenicity. Overlapping genetic factors in different diseases identified via NGS is the most obvious indicative of the converging cellular pathways in neurodegeneration.

1.6. Mechanisms Involved in ALS

Motor neurons in ALS are suggested to be under a series of cellular stresses, including oxidative stress, axonal transport defects, glutamate excitotoxicity, mitochondrial defects, impaired ubiquitin-proteasome system and neuroinflammation that are mainly caused by protein aggregation and defects in RNA metabolism (Figure 1.2). In addition, literature reveals that the disease is not restricted only to motor neurons but is a result of consecutive damage that occurs in motor system including other glial cell types involved in nourishment and neurotransmission processes (Robberecht and Philips 2013; Taylor, Brown, and Cleveland 2016). Overlaps between these biological mechanisms and pathways lead to disease onset and progression. Growing amount of identified gene mutations is expected to make-up a complete picture of cellular events that trigger one another (Figures 1.1 and 1.2).

It is known that protein aggregates are a hallmark of several neurodegenerative diseases (Ramaswami, 2014). Since the identification of mutations in SOD1, protein misfolding and formation of ubiquitinated aggregates has been the focus of interest in ALS pathogenesis (Leigh *et al.*, 1988). UBQLN2, that binds the proteasome from its N-terminal domain and the ubiquitinated proteins with its C-terminal domain, is also mutated in ALS patients. The mutations are specific to the proline-rich domain of the protein located in the

C-terminus, thus, targeting of the cargo to the proteasome is impaired in the presence of mutations. *VCP*, coding for a ubiquitin-sensitive chaperone, is responsible of targeting misfolded proteins to the ubiquitin-proteasome system, to ER-associated degradation and autophagosomes (Robberecht & Philips 2013). The importance of autophagy has become evident in ALS with the identification of mutations in genes like *sequestosome 1* (*SQSTM1*) and *optineurin* (*OPTN*) that recognize and direct the cargo to lysosomal degradation. The recently identified TBK1 is responsible of phosphorylating *SQSTM1* and *OPTN*, further supporting the defects caused in autophagy (Cirulli *et al.*, 2015).

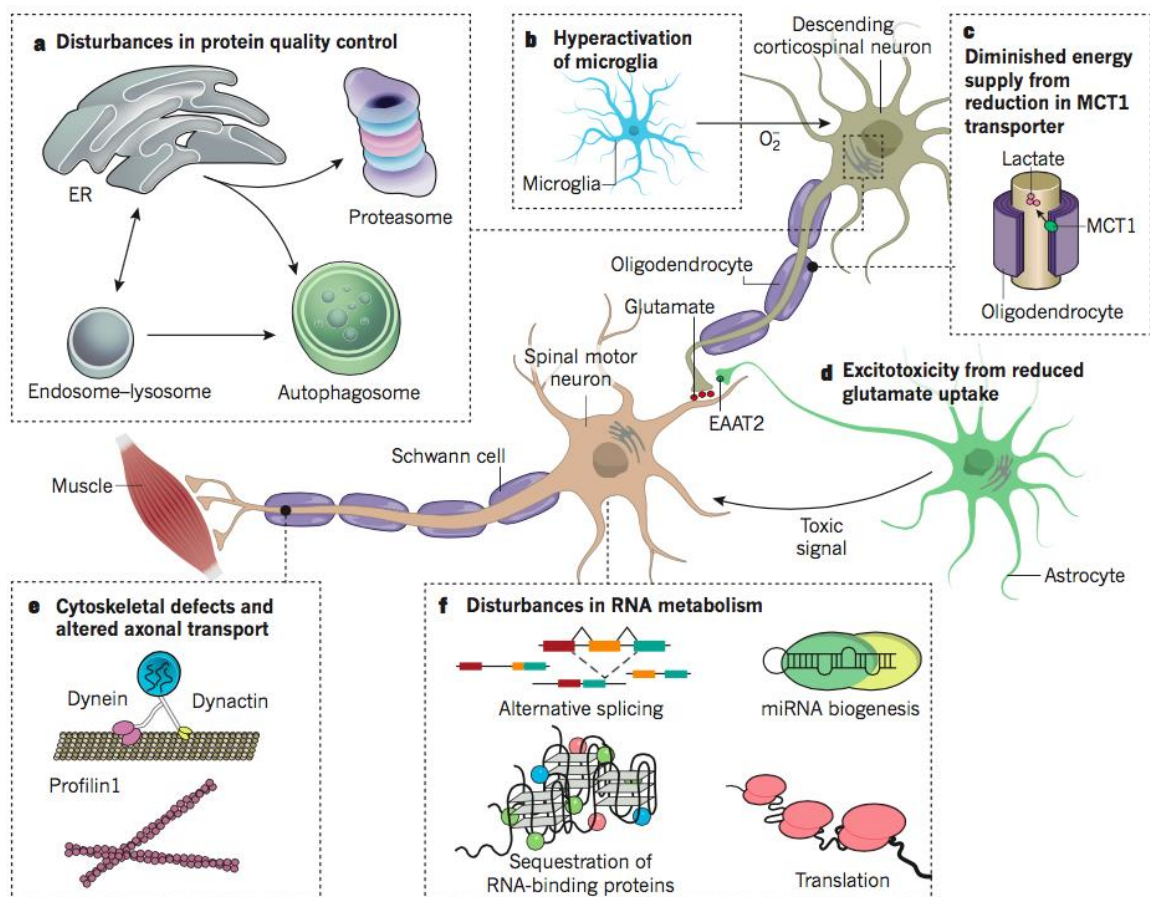


Figure 1.2. Mechanisms involved in ALS pathogenesis (Taylor *et al.*, 2016).

The association of mutations in RNA-binding proteins TDP-43 and FUS with ALS has introduced the defects in RNA processing into ALS pathology (Robberecht and Philips, 2013). TDP-43 protein inclusions that accumulate in the shrinking motor neurons are observed in all ALS cases, with the exception of patients carrying mutations in *SOD1* and *FUS* genes, in whom abnormal *SOD1* and *FUS* inclusions are present, respectively

(Taylor *et al.*, 2016). TDP-43 and FUS, normally located in the nucleus, are exported to cytoplasm in case of disease, form stress granules and cause sequestration of several RNA-binding proteins, silencing mRNA translation. Under normal cellular conditions, the low-complexity domains (LCDs) of TDP-43 and FUS, in addition to some other ribonuclear proteins (RNPs) associated with ALS, are required for phase separation in the cytoplasm for the formation of stress granules. However, mislocalization of mutated RNPs in ALS, increases their cytoplasmic concentration, enhancing formation of stress granules harboring many RNA-binding proteins that are able to form toxic fibrils through binding to each other (Taylor *et al.*, 2016). Furthermore, evidence of protein inclusions observed in sALS pathology point out to possible uptake of these aggregates by neighboring cells, which lead to additional conformational changes in wild-type proteins and cause the spreading of the disease in a prion-like manner (Peters *et al.*, 2015).

Despite growing evidence from TDP-43 and FUS pathology, failure of proteostasis was suggested as the main cause of ALS pathogenesis until the identification of the hexanucleotide repeat in the *C9orf72* gene. The protein coded by the gene is suggested to be involved in endosomal trafficking, however the mechanism of the pathogenic expansion in the non-coding region of the gene is rather distinct (Section 1.7.4). Several reports have shown the formation of nuclear and cytoplasmic RNA foci that harbor RNA-protein aggregates and point out to an altered ribostasis in disease (Cirulli *et al.*, 2015; Ramaswami, 2014). Keeping in mind that RNA metabolism is fundamental in ALS pathogenesis, it is interesting that genes like *TARDBP* and *FUS* have roles in miRNA biogenesis. Recent studies identified several miRNAs to have crucial roles in motor neuron differentiation and suggest changes in expression levels of specific miRNAs (e.g. miR-9 and miR-206) in case of disease, which can be used as potential biomarkers (Hawley *et al.*, 2017; Paez-Colasante *et al.*, 2015).

Another important group of genes mutated in ALS belongs to cytoskeletal proteins that are crucial for axonal integrity (*tubulin 4A*, *PFN1*) and transport of several components along the long axons of motor neurons (*dynactin 1*) (Brown and Al-Chalabi, 2017). Inverse correlation of EPHA4 expression, a receptor that regulates axonal repulsion and motor neuron survival is also reported, strengthening the role of axonal structure (van Hoecke *et al.*, 2012).

Although ALS is considered as a disease of toxic-gain-of function due to the formation of toxic aggregates and stress granules; depletion of nuclear wild-type RNA processing proteins and disorganized axonal structure and transport may also result in the pathogenicity. On top of that, imbalanced energy homeostasis is also reported in patients and animal models; elevated AMP-activated protein kinase (AMPK) levels indicating energy deprivation, disruption of mitochondrial transport through the axon and lowered risk of ALS observed in people with obesity and type II diabetes are among suggested evidence (Vandoorne *et al.*, 2018). Thus, it may be concluded that ALS is a disease in which (i) loss of function of several mutated proteins or (ii) depletion of the wild-type proteins in the areas of normal function lead to disruption of processes, which eventually trigger cytotoxicity, neuroinflammation and cell death.

ALS research is in a boost leading to discovery of genetic factors converging in common pathways, which will hopefully be used for common treatment options, especially required for sALS that makes up the majority of patients. So far, transplantation of neural stem cells is being tested to strengthen cell survival, however, although the method has been proven to be safe, the benefits are yet scarce (Glass *et al.*, 2016). For people with known genetic mutations more personalized gene-therapy approaches are being tested. These therapies aim to deliver wild type genes or DNA or small interfering/micro RNA fragments to silence the mutated gene, using antisense oligo nucleotides (ASOs) and adeno-associated viruses (AAVs) (Brown and Al-Chalabi, 2017; Stoica and Sena-Esteves, 2016). ASO trials in a mouse model of *C9orf72* and the first human Phase I trial for *SOD1* show promising results (Jiang *et al.*, 2016; Miller *et al.*, 2013). Despite this progress, ALS is a late-onset progressive disease and the destruction of the nervous system starts way before the first symptom onset. Thus, the ultimate tools for researchers are the rapidly-developing gene editing techniques like CRISPR/Cas9 that will provide more precise and sustainable solutions for the cure of the disease in embryos and eventually in presymptomatic adults, with known genetic defects (Cai *et al.*, 2016).

1.7. C9orf72 as the Main Player in ALS

1.7.1. Discovery and Clinical Characteristics

Autosomal dominant adult onset ALS accompanying frontotemporal dementia (ALS-FTD) was linked to the chromosome 9p21 locus in 2006. Since then, all known genes in that part of the genome were sequenced, however no mutation was identified as causative, resulting in expectations for a non-coding mutation (Morita *et al.*, 2006). In 2011, two independent groups have identified the heterozygous GGGGCC hexanucleotide repeat expansion, located in the non-coding region of the *C9orf72* gene, as the most common cause of ALS-FTD (DeJesus-Hernandez *et al.*, 2011; Renton *et al.*, 2011). The detection of the expansion was performed by repeat-primed PCR and the repeat sizes were determined by Southern blotting. The pathogenic threshold of the pathogenic expansion is now considered as 30 GGGGCC repeats, however usually the expansion is much longer, ranging from hundreds to thousands without any reports of anticipation (Rohrer *et al.*, 2015). Also, there are cases reported to have frontotemporal deterioration with expansions of 20-22 (Cerami *et al.*, 2013; Gómez-Tortosa *et al.*, 2013). Another recent finding is that there can be extensive size variability in the pathogenic expansion among different tissues like blood and spinal cord (Nordin *et al.*, 2015). The discovery of *C9orf72* repeat expansion has made a drastic change in the etiology of ALS, since the disease has joined the group of repeat expansion disorders that are very common among neurodegenerative diseases.

The GGGGCC expansion mutation is the most common cause of familial ALS with frequencies of 21% in German, 36.2% in North American, 37.8% in Italian and 46.4% in Finnish populations. As opposed to all other common causative ALS genes *C9orf72* is unique in the sense that it explains a high percentage of sporadic cases (~5%) (Majounie *et al.*, 2012). It is shown that the *C9orf72* repeat expansion arose from a single founder haplotype in Europe; the rare occurrence of the mutation in Asian, Indian, Pacific Islander and Native American populations supports this evidence (Liu *et al.*, 2014; Mok *et al.*, 2012; Smith *et al.*, 2013).

In patients carrying the *C9orf72* repeat expansion the percentage of bulbar onset cases is comparable higher being around 40%, the disease onset is earlier and there is a family history of frontotemporal dementia (FTD) in almost 50% of the cases (Cooper-Knock *et al.*, 2015). The most significant clinical difference of the expansion carriers is that the progression of the disease is faster and thus the disease duration is very short, sometimes being below a year.

Incomplete penetrance is reported for sporadic cases, however some of these have a family history of other neurological diseases, especially dementia. Considering the occurrence of the mutation in ALS, ALS-FTD and pure-FTD, it has become very important to obtain detailed family history during counseling. Especially in Europe, ALS is referred as MND, since it is the most common MND and is clinically heterogeneous. Likewise, the term FTD also includes several subtypes, which include behavioral variant FTD (bvFTD), non-fluent progressive aphasia and semantic dementia. Evidence supports that the relation between bvFTD and MND is not a new theme, on the contrary, behavioral symptoms like apathy, greed, mental rigidity, pseudobulbar symptoms and differences in personality were frequently reported among MND patients (Bak, 2010). It is possible to observe involvement of non-CNS systems like FTD in patients with *VCP* or *SQSTM1* mutations; however, *C9orf72* has been the strongest genetic and pathologic link (Renton *et al.*, 2013). Other than bringing ALS and FTD under the same roof, the *C9orf72* repeat expansion, although very rare, has been shown to be implicated in other neurodegenerative diseases including PLS, PMA, Alzheimer's disease and Huntington's disease-like syndrome (Liu *et al.*, 2014). Also, intermediate size expansions (20-30) have been significantly associated with Parkinson's disease (Nuytemans *et al.*, 2013). In addition to this clinical heterogeneity, there is also genetic heterogeneity behind the repeat expansion mutation. There are reports regarding the oligogenicity where expansion can co-occur with mutations in other ALS-associated genes like *granulin (GRN)*, *angiogenin (ANG)* and *microtubule-associated protein tau (MAPT)* (van Blitterswijk *et al.*, 2012b; Williams *et al.*, 2013). Another fact is that, although the mutation is autosomal dominantly inherited, one case affected with frontotemporal dementia was reported to carry homozygous repeat expansion mutation in the *C9orf72* gene (Fratta *et al.*, 2013).

1.7.2. Function and Suggested Mechanisms Leading to C9orf72 Pathology

The C9orf72 protein has three alternative transcripts that are coding for two isoforms, isoforms a and b. The hexanucleotide repeat expansion lies in the promoter region or in the first intron of the gene according to the isoform (Figure 1.3) (Rohrer *et al.*, 2015).

The molecular function of the protein remains to be unknown, but it is widely expressed in different tissues and has been located in the cytoplasm in neurons. Several possible domains are identified through sequence analysis, which include M16 peptidase, cytochrome bcl, glycohydrolase and differentially expressed in normal and neoplasia (DENN) domains (Cooper-Knock *et al.*, 2015). The importance of DENN domain is emphasized due to its role as an exchange factor for Rab proteins and the existence of this domain also in Alsin protein, another protein mutated in ALS (Farg *et al.*, 2014). Rab proteins, small GTPases, are involved in membrane trafficking and alteration in their regulation may lead to defects in the autophagy pathway.

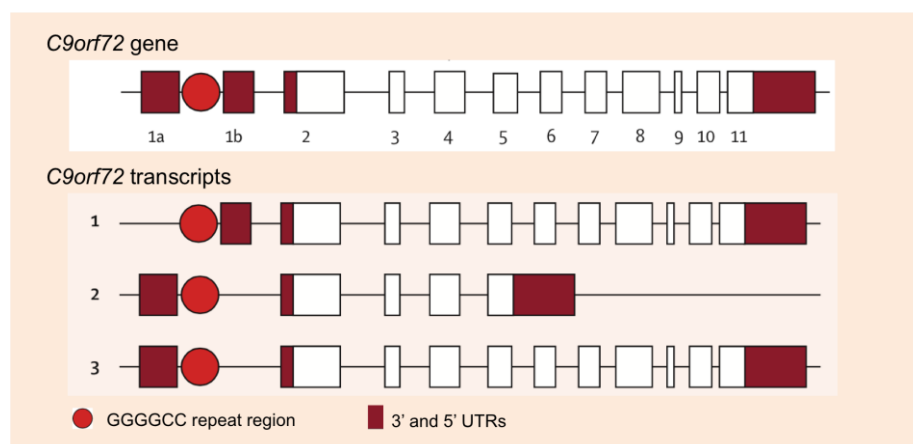


Figure 1.3. *C9orf72* gene and its transcripts (adapted from Rohrer *et al.*, 2015).

Studying the mechanism behind *C9orf72*-based disease has been the hottest topic in the past years and still we lack the knowledge on the protein function and how it causes the pathogenicity. It is suggested that the absence of coding mutations are indicative of a gain of function mechanism (Harms *et al.*, 2013). So far, the field has focused on two mechanisms for a possible gain of function caused by the hexanucleotide repeat expansion in the *C9orf72* gene. Several modes of action have been proposed for the diseases

associated with unstable repetitive elements in DNA, RNA and proteins; these include decreased protein expression, overexpression of wild-type proteins, toxic gain of function and toxic RNA gain of function caused by the expanded transcripts (Pearson, 2011).

Earliest observations supporting a toxic RNA gain of function was the formation of RNA foci from the repeat sequence (DeJesus-Hernandez *et al.*, 2011). As previously reported for other repeat expansion disorders like myotonic dystrophy 1 (DM1) and some SCAs, in *C9orf72*-based ALS and FTD, *C9orf72* RNA transcripts harboring the repeat region form a toxic RNA foci in the nucleus, sequestering several RNA-binding proteins (Gendron *et al.*, 2013; Todd and Paulson, 2010). In case of the *C9orf72* repeat expansion, RNA foci are shown to be build-up of two structures, either hairpins from both sense and anti-sense strands or G-quadruplexes from the sense strand (Fratta *et al.*, 2012; Reddy *et al.*, 2013). Toxic RNA foci caused by the GGGGCC expansion in the *C9orf72* gene trigger cellular stress and apoptosis in a length-dependent manner in cell culture and in vivo studies empowering its role in neurodegeneration (Lee *et al.*, 2013b). It is suggested that longer repeats cause nuclear retention of the mutant transcript preventing its degradation by RNases and allowing sequestration of more nuclear proteins, further disabling their function. One of the target proteins shown to co-localize with the expanded *C9orf72* RNA foci was hnRNP-H, a protein important for the splicing machinery (Lee *et al.*, 2013b). hnRNP-H is known to bind G-rich RNA quadruplexes and intronic repeat regions with high G content to enhance exon skipping. It is shown to inhibit the nuclear export of GUG expansion in DM1. This brings out the possibility of a positive feedback loop where the G-quadruplex recruits proteins including hnRNP-H and since its nuclear export is inhibited, more RNA-binding proteins accumulate (Lee *et al.*, 2013b). In another study, mRNA export adapters, ALYREF, SRSF1, SFRS3 and SFRS7, were also reported as targets of the RNA foci caused by the expanded *C9orf72* transcript. RNA foci were reported also in the cytoplasm and it is suggested that ALYREF may involve in the unwinding of the G-quadruplex for nuclear export, whose size is normally not suitable to cross nuclear pores (Cooper-Knock *et al.*, 2014).

In addition to the accumulation of RNA foci, another possible pathogenic mechanism behind *C9orf72*-mediated toxicity is the repeat associated non-ATG (RAN) translation, which results in transcription of aggregation-prone c9RAN proteins from the sense and

anti-sense strands (Ash *et al.*, 2013; Gendron *et al.*, 2015; Mori *et al.*, 2013). RAN translation was first described by Ranum and colleagues who showed translation of different dipeptide repeat (DPR) proteins from all possible reading frames of the expanded repeat in SCA8 and DM1 (Zu *et al.*, 2010). Studies on CAG expansion disorders suggest that formation of DPRs depends on the size of the expansion, DPRs can differ in size according to the starting point of translation, and although simultaneous translation of different polypeptides is possible, the amounts of each polypeptide can vary (Pearson, 2011; Zu *et al.*, 2010). In case of viruses, internal ribosomal entry sites resemble the tRNA structure and recruit ribosome for initiation of translation. It is thought that in RAN translation hairpins may be the alternative of tRNA structure and thus short repeat sizes may lack the ability to form these structures and may not initiate RAN translation (Zu *et al.*, 2010). RAN translation further contributes to the disease pathogenesis by the formation of insoluble cytoplasmic aggregates that will disrupt the ubiquitin-proteasome system and cause neurotoxicity.

In case of *C9orf72* GGGGCC expansion, six possible DPR proteins are formed; poly-Gly-Ala, poly-Gly-Pro and poly-Gly-Arg from the sense and poly-Ala-Pro, poly-Pro-Gly and poly-Pro-Arg from the anti-sense strands. Antibodies against these proteins detected them in ubiquitinated neuronal cytoplasmic inclusions also positive for p62 (Ash *et al.*, 2013; Gendron *et al.*, 2013; Mori *et al.*, 2013; Zu *et al.*, 2013). In a recent study, it was shown that DPRs containing arginine residues might cause toxicity through a different mechanism. Considering the previous knowledge on the association of the serine:arginine domain-containing proteins with nucleoli, they showed that poly-Gly-Arg and poly-Pro-Arg DPRs bind irreversibly to the nucleolus and result in disruption of mRNA splicing and the synthesis of ribosome itself (Kwon *et al.*, 2014).

Toxic RNA foci have been observed in both neuronal and glial populations, whereas DPRs have been shown to be specific for neurons. Thus, it is speculated that toxic RNA foci and c9RAN positive inclusions are distinct mechanisms and depending on the size of the expansion some transcripts that escape formation of RNA foci are exported out of the nucleus and exposed to RAN translation (Gendron *et al.*, 2013). However, in other studies, it is proposed that RNA foci may unwind, are exported from the nucleus and reform in the cytoplasm, where RAN translation takes place (Cooper-Knock *et al.*, 2014). These

intervening mechanisms cause defects in several crucial mechanisms that altogether converge in neuronal toxicity and death.

Toxic gain-of-function mechanisms are common in pathogenesis of all genetic subtypes of ALS and other late-onset neurodegenerative diseases; however, mechanisms leading to loss of function are also suggested. In case of the hexanucleotide repeat expansion in the *C9orf72* gene, decreased mRNA levels are associated with the pathogenic expansion (Cruts *et al.*, 2013; Gendron *et al.*, 2013). Knock-down *C9orf72* orthologues in zebrafish altered motor neuron morphology and resulted in locomotion defects which were rescued with overexpression of the wild-type protein (Ciura *et al.*, 2013). Finally, it is also suggested that impaired transcription and splicing of the mutant allele due to the expansion in the promoter region may cause haploinsufficiency (Cruts *et al.*, 2013). Despite evidence for loss-of-function, there are many studies with findings against association of low *C9orf72* expression with disease pathogenesis, including mice ASO trials and conditional knock-downs, lowering *C9orf72* expression without observing any motor neuron phenotype (Gitler and Tsuiji, 2016).

1.7.3. Promoter Hypermethylation

The heterogeneous nature of ALS makes the targeting of the exact pathogenic mechanisms underlying the selective death of motor neurons challenging. In the post-genomic era, complex disorders, such as ALS, are studied at different levels for the understanding of the underlying mechanisms involved in the onset and progression of the disease.

Epigenetic mechanisms are known to contribute to the regulation of both coding and noncoding RNA transcription at several levels via many different kinds of DNA, RNA and protein modifications. Positional or chemical changes in these epigenetic marks or in the regulatory machinery would potentially lead to changes at the transcriptome which may lead to diverse diseases. In this manner, epigenetic changes are also considered to be key players in the pathogenesis of ALS and other neurodegenerative diseases (Qureshi and Mehler, 2011). Thus, it is important to track epigenetic changes that are over or above the

genome, which may give insights in to the pathogenesis of complex diseases and has potential therapeutic power since they are reversible (Paez-Colasante *et al.*, 2015).

DNA methylation, especially at the promoter regions of genes, is a well-characterized epigenetic mechanism for gene silencing. The mechanism is based on the transfer of a methyl group from S-adenosyl-methionine to the carbon-5 of a cytosine (5mC), usually located in a CpG island in the promoter regions of the genes, by DNA methyltransferases (Dnmts). Methyl CpG-binding proteins bind methylated cytosines and recruit other elements that, in turn, promote silencing by disabling the translation machinery. Another enzyme, ten-eleven translocation methylcytosine dioxygenase 1 (TET1) is responsible to transform 5mC to 5-hydroxymethylcytosine (5-hmC) as a step of demethylation. Several studies in ALS patients have recently made connections between DNA methylation levels and neuronal cell death (Figueroa-Romero *et al.*, 2012). Also detection of elevated global methylation levels in whole blood has been influential for future biomarker studies, that are expected to give insights about disease stage and progression (Tremolizzo *et al.*, 2014).

As stated previously, the GGGGCC expansion in *C9orf72* gene is not only the most common, but also the first and so far, only repeat expansion mutation linked to both sALS and fALS cases, resulting in a large clinical heterogeneity, ranging from ALS to frontotemporal dementia (FTD). Several studies show that hypermethylation is a common phenomenon for repeat expansion disorders like FRDA, Fragile X and myotonic dystrophy, acting as an epigenetic modifier of the disease (Avitzour *et al.*, 2014; Sandi *et al.*, 2014). It has been recently shown that both *C9orf72* promoter hypermethylation and histone trimethylation (H3K9, H3K27, H3K79, H4K20) are associated with *C9orf72*-based ALS and FTD and result in decreased *C9orf72* mRNA levels (Belzil *et al.*, 2013; Russ *et al.*, 2015; Xi *et al.*, 2013). In these studies, 26 CpGs located in the promoter region of the *C9orf72* gene 5' to the pathogenic expansion were studied in ALS and FTD patients carrying the mutation. The results suggested that hypermethylation was specific to expansion carriers without exceptions for different subtypes like ALS, ALS-FTD or FTD (Russ *et al.*, 2015; Xi *et al.*, 2013). DNA methylation levels were similar in different tissues and methylation was shown to be stable over time (taking new samples from the patients at different times). Familial heterogeneity was not observed with conserved

percent methylation levels among the members of the same family. In addition, in one of the studies increased hypermethylation was associated with shorter repeat size (Russ *et al.*, 2015). Both studies reported longer disease duration in case of *C9orf72* promoter hypermethylation in FTD patients. These studies also speculated that reduced expression levels observed in case of aberrant CpG methylation may lower toxic *C9orf72* RNA accumulation and thus result in prolonged disease duration (Xi *et al.*, 2013 & Russ *et al.*, 2014). Based on the idea that age is the highest risk factor in both ALS and FTD, the DNA methylation age was calculated in patients with the expansion using 353 CpG sites throughout the genome; statistical analysis revealed that DNA methylation accumulated in time leads to earlier disease onset and shorter disease duration (Zhang *et al.*, 2017). The impact of epigenetic modifications on the *C9orf72* gene has become evident; however, whether these modifications result in pathogenicity or whether they act as disease modifiers is not clear. In a follow-up study by Belzil *et al.*, the *C9orf72*-positive ALS-FTD patients who were previously reported to have decreased *C9orf72* mRNA levels due to aberrant histone methylation were assessed for *C9orf72* promoter methylation levels, and hypermethylation of the promoter was not present (Belzil *et al.*, 2014). These findings may exhibit *C9orf72* promoter methylation as a strong disease modifier, however its impact on the expression levels is yet to be clarified.

1.7.4. Modifiers of *C9orf72*-based Disease

The pathogenic repeat expansion in *C9orf72* has been the focus of interest in the ALS field since its identification in 2011. One of the burning questions are how the two distinct phenotypes, ALS and FTD, are caused by this same mutation and how this non-coding expansion results in considerable clinical diversity among the carriers. Several studies have been conducted for the identification of possible genetic modifiers of the disease. In one of these studies it was shown that carrying the minor allele of the *transmembrane 106 B (TMEM106B)* homozygously protects individuals against FTD, but not ALS in the co-presence of *C9orf72* expansion (van Blitterswijk *et al.*, 2014a). Intermediate expansions in the *ATXN2* gene have also been suggested as modifiers in familial ALS and ALS-FTD cases carrying the *C9orf72* expansion (Lattante *et al.*, 2014). So far, the size of the expansion is rarely found to be a modifier of disease severity, one study showing patients with short expansion sizes (<80) with later disease onset; but

repetitive studies might change this fact (DeJesus-Hernandez *et al.*, 2011; Gijselinck *et al.*, 2016; van Blitterswijk *et al.*, 2013). In a recent study, variants that are previously implicated in FTD and MND were examined in a large cohort of 330 *C9orf72* expansion carriers with different phenotypes and 374 controls. As a result, three variants were associated with age of onset and six were associated with survival (van Blitterswijk *et al.*, 2014b). Identification of new modifiers of disease duration, age of onset, site of onset and disease subtype will lead to additional involving pathways in disease pathogenesis, which may in turn fill in the gaps of mechanisms in *C9orf72*-based disease. It may then be possible to have an understanding of the differential vulnerability of different cell types in different regions of the brain in case of ALS and FTD.

1.8. The Big Project of ALS: Project MinE

Project MinE is a world-wide consortium project, which aims to perform WGS in 15,000 ALS patients plus 7,500 controls. The intention is to compare the coding and non-coding parts of the genome of patients and controls to identify rare variants (MAF<0.01) that are associated with the disease. By sequencing this vast number of genomes, the ultimate goal is to discover several novel causative and/or risk loci and fine-map current knowledge (van Rheenen *et al.*, 2017). This is expected pave the way for understanding novel disease mechanisms which eventually will lead to therapeutic interventions.

Complex diseases like ALS need to be targeted from a collective perspective independent of small families that follow Mendelian inheritance or even single populations with homogenous genetic background. Thus, a global effort has been taken in the framework of Project MinE to obtain as much patients as possible, suffering from sporadic ALS, to work on complex genetic, epigenetic and environmental factors effective in disease pathogenesis/progression. Project MinE's working model is based on different working groups (WGs) that are active at different stages like sample collection, data storage and analysis.

ALS research centers from 19 countries are collaborating in Project MinE, including our laboratory from Turkey. As of 2018, a total of 10,485 genomes have been sequenced, which makes up 47% of the ultimate goal of the project. The variant frequency information

and gene burden test results conducted on these samples were gathered together and made recently public in a database for the use of external researchers. Pilot association studies performed with WGS data revealed new risk loci (*C21orf2*) and a novel disease gene (*KIF5A*) for ALS (Nicolas *et al.*, 2018; van Rheenen *et al.*, 2016).

2. PURPOSE

The completion of the human genome project along with the advent of high-throughput technologies has allowed for more detailed research on the mechanisms behind complex diseases. Here, we aim to utilize some of these technologies in order to shed light on a complex neurodegenerative disease, ALS.

With the progress in next generation sequencing technologies there has been a drastic change in our understanding of the epidemiology of ALS. With the identification of new causative genes and modifiers, we now know that ALS is not any more, a single (monolithic) clinical entity, but rather a group of diseases unified by the common theme of motor neuron degeneration, that harbor heterogeneous clinical phenotypes overlapping with many other neurodegenerative processes and diseases.

The main focus of this thesis is to understand the molecular basis of ALS in Turkey by analyzing a large and well-defined cohort of fALS and sALS patients.

For this purpose;

- ALS patients are analyzed for mutations in common and rare ALS genes
- ALS and its overlapping genetic features with other MNDs and similar neurological disorders is investigated to provide differential diagnosis
- Phenotypic heterogeneity within ALS is studied with an emphasis on the most common mutation, the hexanucleotide repeat expansion in the *C9orf72* gene, for the investigation of possible modifiers of disease
- Whole genome sequences of sALS patients are explored to dissect the complexity of sporadic disease.

Looking through the keyhole of genetics, this thesis aims to unravel the genetic factors that are either disease-causing or disease-modifying, using conventional and next generation sequencing approaches, in familial ALS as well as in sporadic disease.

3. MATERIALS

3.1. Subjects

3.1.1. The ALS Cohort Under Study

The Neurodegeneration Research Laboratory (NDAL) hosts a total of 1201 Turkish ALS patients. The initial focus of this thesis is the analysis of four common ALS genes, *SOD1*, *TARDBP*, *FUS* and *UBQLN2*, in 771 out of 1201 patients (208 fALS patients from 159 families and 563 isolated cases). The clinical data of the probands (n=722) are summarized in Table 3.1. Another main focus of the thesis is the more recently defined *C9orf72* expansion, a real breakthrough finding in ALS research. Being by far the most common ALS and FTD mutation worldwide (except in Japan), the *C9orf72* expansion mutation is investigated from multiple aspects in all 1201 patients referred to our laboratory.

Table 3.1. Clinical data of the Turkish cohort under study.

		Total ALS	fALS	sALS
#	probands	722	159	563
	family members	49	49	-
	male/female	1.5	1.1	1.5
	consanguinity	139	40	99
	dementia	26	11	15
Site of onset	limb	515	121	394
	bulbar	131	29	102
	limb+bulbar	51	6	45
	unknown	25	3	22
Age of onset	juvenile (<25yrs)	61	20	41
	middle (25-45 yrs)	165	44	121
	late (>45 yrs)	480	94	386
	unkonwn	16	1	15
Mean age of onset	total \pm SD	51 \pm 14,9	48 \pm 16,2	52 \pm 14,5

Sample collection from patients was approved by Boğaziçi University Ethics Committee. Genetic counseling was given to patients and signed informed consent was

obtained from all subjects included in this study. DNA samples from healthy relatives were obtained for research purposes only with their approved informed consent.

3.1.2. Cases subjected to WES

3.1.2.1. Individuals Diagnosed with ALS/MND. In the scope of this thesis, WES is applied to 54 individuals from 39 pedigrees for patients diagnosed with ALS or MND. This number includes, four families with HSP diagnosis apart from the ALS cohort under study. Selection criteria for WES included first-degree consanguinity in parents of the affected individual or atypical clinical features and early/juvenile disease-onset. In addition, WES is applied in some families with autosomal dominant inheritance pattern, in whom screening in common genes did not reveal any mutations. Clinical information of the cases subjected to WES, their suspected inheritance pattern and initial clinical diagnosis are compiled in Table 3.2.

3.1.2.2. Families Showing Incomplete Penetrance. WES was applied to two families with mutations in known and penetrant ALS genes, to understand the genetic background of non-manifesting mutation carriers and to look for possible effects of genomic background on disease pathogenesis. The index cases in these two families carry heterozygous mutations in *SOD1* and *TARDBP* genes, p.His71Tyr and p.Gly348Cys, respectively.

3.1.3. Project MinE

Six hundred and twenty-five Turkish ALS cases and 152 neurologically healthy controls were subjected to WGS in the scope of Project MinE. Samples selected for Project MinE were chosen on the basis of definitive, mostly late-onset ALS, without a family history.

Family No	NDAL ID	# of Samples Subjected to WES	Gender	Age of Onset	Consanguinity	Family History	Inheritance	Clinical Diagnosis
Family 1	ALS1539	3	M	50	no	yes	AD	ALS
	ALS1584		M	45	no	yes		ALS
Family 2	ALS1381	1	M	20	no	no	inconclusive	ALS
Family 3	ALS1610	1	M	17	no	no	inconclusive	ALS
Family 4	ALS46	2	M	23	yes	yes	AR	ALS
Family 5	ALS132	1	M	14	yes	yes	AR	ALS
Family 6	HSP29	1	M	13	yes	no	AR	HSP
Family 7	HSP33	1	M	23	yes	no	AR	HSP
Family 8	ALS158	1	M	21	yes	yes	AR	ALS
Family 9	ALS314	1	F	17	yes	yes	AR	ALS
Family 10	ALS518	6	F	15	yes	yes	AR	ALS
	ALS869		M	49	yes	yes		ALS
	ALS856		M	32	yes	yes		ALS
Family 11	ALS1164	1	F	17	yes	no	AR	ALS
Family 12	ALS764	1	M	22	yes	no	AR	ALS
Family 13	ALS1594	1	M	17	yes	no	AR	ALS
Family 14	ALS445	1	M	childhood	yes	no	AR	ALS/CMT
Family 15	ALS1716	1	F	37	yes	no	AR	ALS
Family 16	HSP34	1	M	36	yes	no	AR	HSP
Family 17	ALS1813	1	M	37	yes	no	AR	ALS/HSP
Family 18	ALS1704	1	F	22	yes	yes	AR	ALS
Family 19	ALS731	1	F	47	no	yes	AD	MND/Sensory Neuropathy
M: male, F: female, AD: autosomal dominant, AR: autosomal recessive								

Table 3.2. Clinical information of ALS/MND cases investigated with WES (cont.).

[illegible]

3.2. Equipment, Solutions, Chemicals and Kits

3.2.1. DNA Isolation

DNA was isolated using the MagNa Pure Compact Instrument and MagNA Pure Compact Isolation Kit I (Roche, Switzerland), followed by quantification and a quality check in NanoDrop 2000c Spectrophotometer (Thermo Scientific, USA).

3.2.2. Agarose Gel Electrophoresis and Gel Extraction

Equipment, solutions and chemicals used in agarose gel electrophoresis are listed in Table 3.3. The PCR products were extracted from the gel by QIAQuick Gel Extraction Kit (Qiagen, Germany).

Table 3.3. Materials used in agarose gel electrophoresis.

Product		Company
10X TBE Buffer	0.89 M Tris-Base (pH 8.3)	Sigma, USA
	0.89 M Boric acid	MP Biomedicals, USA
	0.5M EDTA	Invitrogen, USA
Agarose		Prona, USA
Electrophoresis Tank, OWL EasyCast B1		Thermo Scientific, USA
Power Supply, EC300XL2		Thermo Scientific, USA
GelRed		Olerup SSP, Sweden
6X Loading Dye		Fermentas, USA

3.2.3. Genotyping Experiments

Genotyping of patients was performed with DNA isolated from whole blood. GoTaq® Flexi DNA Polymerase (Promega, USA), MyTaq™ DNA Polymerase (Bioline, USA), TaKaRa ExTaq™ (Takara Bio Inc., Japan), FastStart Universal Master Mix (Roche, Switzerland) and One Taq® 2X Master Mix (New England Biolabs, USA) were used for all genotyping experiments. The sequences of primers are listed in Tables 3.4 and 3.5.

Table 3.4. Primer sequences used in screening of five common ALS genes.

Primer Name	Primer Sequence (5'-3')	Annealing Temperature (°C)
C9orf72-PRPCR-F1	FAM -AGTCGCTAGAGGCGAAAGC	Touchdown
C9orf72-PRPCR-R	TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG	
C9orf72-PRPCR-anchor	TACGCATCCCAGTTTGAGACG	
C9orf72-Flanking-F	FAM -CAAGGAGGGAAACAACCGCAGCC	Touchdown
C9orf72-Flanking-R	GCAGGCACCGCAACCGCAG	
SOD1e1F	TTCCGTTGCAGTCCTCGGAA	60,5
SOD1e1R	GACCCGCTCCTAGCAAAGGT	
SOD1e2F	TTCAGAAACTCTCTCGAACTT	52,2
SOD1e2R	CGTTTAGGGGCTACTCTACTGT	
SOD1e3F	TGGGAACTTTAATTCATAATT	54,5
SOD1e3R	AGTATACCATATGAACTCCA	
SOD1e4F	CATCAGCCCTAATCCATCTGA	55
SOD1e4R	CCGACTAACAATCAAAGTGA	
SOD1e5F	AGTGATTTACTTGACAGCCCA	57,3
SOD1e5R	TTCTACAGCTAGCAGGATAACA	
TDP43e5F	TGCTTGTAATCTAAGTTTGTGTC	66,1
TDP43e5R	TGCTGAATATACTCCACACTGAAC	
FUSE14F	CAGATGGGTAAGAAAGGCAG	66,1
FUSE14R	TCTCAACAAAACCCTGTTATCC	
FUSE15F	TACTCGCTGGGTTAGGTAGG	66,1
FUSE15R	TTCCAGGAAAGTGAAAAGGG	
UBQLN2-1F	CTTCATCACAGAGGTACCGTG	56
UBQLN2-1R	GTGTGGAGTTACTCCTGGGAG	
UBQLN2-2F	CATGATGGGCTGACTGTTTAC	56
UBQLN2-2R	CTCTTGTGCGGCATTCAGCATC	
UBQLN2-4F	CACAGATGATGCTGAATAGCC	56
UBQLN2-4R	GCTGAATGAACTGCTGGTTGG	
UBQLN2-3F	GACCTGGCTCTTAGCAATCTAG	56
UBQLN2-3R	GTGTCTGGATTCTGCATCTGC	
UBQLN2-5F	CTGCACCTAGTGAAACCACGAG	56
UBQLN2-5R	AACAGCATTGATTCCCACCAC	

Table 3.5. Primer sequences used in validation of WES results and segregation analyses.

Primer Name	Primer Sequence (5'-3')	Annealing Temperature (°C)
ERLIN1e4F	GGAAGCGAACTCTTTCTGTG	60,9
ERLIN1e4R	GCAGAAAGTCAACTCTGTCCA	
SIGMAR1e3F	GAGGCTTGGGAGTGCAGACT	60,3
SIGMAR1e3R	CCGAGCATAGGAGCGAAGAG	
SYNE1e124F	CCACTACGGTAAATGCTTGC	Touchdown
SYNE1e124R	AGAGAATCAAAATACTGCAGGG	
SYNE1e128F	ATACTTGTGCACATGGGCTG	Touchdown
SYNE1e128R	TTTTGAAAATGGTGAGATAACCC	
KIF5Ae14F	CCATAGAAATGGTCACTGGC	60,8
KIF5Ae14R	ATGCACTGAACAGCAGTGC	
KIF5Ae27F	AGAGGGTTTGCGCAAAGT	62
KIF5Ae27R	GCACTGAGGCAGGGGAAT	
SPG11e12F	AGGGTGTCTCTGTATCTAAC	53,8
SPG11e12R	TCTTCCAAGGTTTTCTTC	
SPG11e33F	CAATAGGCCAAGGGTTTCAA	65
SPG11e33R	TATAACTCCTGCTGGAGGGC	
SPG11e39F	AAGGGTTTAAGATAATTTGGGGA	56
SPG11e39R	GGATTCTTGATACTGCTTTGCC	
SPG11e40F	AATTAGCCAGGGTGGTGACA	65
SPG11e40R	CCCACAAAGGACTGATATGG	
SBF1e25F	CTAGGCCCTGTGCTCATCT	64
SBF1e25R	GGGCACACGTGGTCATCA	
ANXA11e13F	GAACAGAAGGAGGAGGAGAAAAG	62
ANXA11e13R	AGGTCCCGAGTGTTCTCTGAT	
ZFYVE26e14F	GCTTTCTGACCCTACAAAGCA	61
ZFYVE26e14R	GTTAGCAGGGAAGCCTGACC	
DNAJB2e9F	GCAGTAATACCCCTGGCTCA	60,2
DNAJB2e9R	CTTCCACAGTGAGTCAGACC	

3.2.4. Detection of *C9orf72* Promoter Methylation Levels

3.2.4.1. Bisulfite Conversion. Bisulfite conversion was achieved with EZ DNA Methylation-Gold™ Kit (Zymo Research, USA). Zymo Taq™ PreMix was chosen for the amplification of the bisulfite converted DNA (Zymo Research, USA).

The primers for the nested PCR were:

1st PCR: BSP_1F: 5'-TATGTAAAGGTAAAATTGTGATGATTT-3'
 BSP_1R: 5'-AAATCTTTTCTTATTCACCCTCAAC-3'
 2nd PCR: BSP_2F: 5'-TGTAAGGTAAAATTGTGATGATTT-3'
 BSP_2R: 5'-CCACACCTACTCTTACTAAACCC-3'

3.2.4.2. Methylation-specific Restriction Enzyme Digestion. *HhaI* enzyme was used for digestion of gDNA (New England Biolabs, USA) and PCR amplification with digested and undigested gDNA was carried out using Mytaq™ DNA Polymerase (Bioline, USA).

Primer sequences for the flanking PCR of the promoter region were:

- C9ORF72_F1_F: 5'-CCCTACCAGGGTTTGCAGT-3'
- C9ORF72_F1_R: 5'-CGACTCCTGAGTTCCAGAGC-3'

3.3. Next Generation Sequencing Platforms and Enrichment Kits

Whole exome sequencing platforms and exome enrichment kits are listed in Table 3.6.

Table 3.6. Sequencing platforms and enrichment kits.

Sequencing Platform	Exome Enrichment Kit
HiSeq 2000	Roche SeqCap EZ Whole Exome V2
HiSeq2000	Roche SeqCap EZ Whole Exome V3
HiSeq 2000	Agilent SureSelect Human All Exon V5
NextSeq 500	Nextera Rapid Capture Exome
HiSeq 2000, 2500, 4000	Agilent SureSelect Human All Exon V5

3.4. Hardware

Bioinformatic pipeline of WES data was run on a workstation computer (12X Intel® Core™ i7-4930K CPU @ 3.40GHz, Samsung SSD hard disk (256GB), 32GB RAM, Seagate HDD 3TB).

The WES and WGS data were stored on a network-attached storage system (DSM 5.2-5644 Update 5, Synology Inc., Taiwan).

3.5. Software, Online Databases and Bioinformatic Tools

Publicly available and commercially purchased software, online databases and bioinformatic tools used in this thesis and their descriptions are listed in Table 3.7.

3.6. General Laboratory Equipment

General laboratory equipment and instruments are listed in Table 3.8.

Table 3.7. Software, online databases and bioinformatic tools.

Software/ Database	Description
1000 Genomes	Human genome variant database
Annovar (Wang <i>et al.</i> , 2010)	Software tool for functional annotation of genetic variations
Burrows-Wheeler Aligner (BWA)	Software package for mapping sequences against a reference genome
CLC Main Workbench (QIAGEN Bioinformatics, Germany)	Software for DNA, RNA and protein sequence analyses
ClinVar (Landrum <i>et al.</i> , 2014)	An archive of relationships among variations and human phenotype
dbSNP (Sherry, 2001)	Database of SNPs and multiple small-scale variations
ExAC (Lek <i>et al.</i> , 2016)	Exome Aggregation Consortium database
GeneCards (Weizmann Institute of Science, 2016)	A database for clinical/functional information of human genes
Genome Analysis Toolkit (GATK) (McKenna <i>et al.</i> , 2010)	A toolkit for variant discovery from NGS data
Integrative Genomics Viewer (IGV)	Visualization tool for interactive exploration of integrated genomic datasets
NHLBI GO Exome Sequencing Project	A database of 6500 human exomes
Online Mendelian Inheritance in Man (OMIM) (McKusick-Nathans Institute of Genetic Medicine)	An online catalog of human genes and disorders
PLINK (Purcell <i>et al.</i> , 2007)	Genome data analysis toolkit
Polymorphism Phenotyping v2 (PolyPhen2) (Adzhubei <i>et al.</i> , 2013)	A web server for prediction of possible impact of amino acid substitutions
R (R Core Team, 2017)	Software for statistical computing and presentation
RFFlow	A software for drawing pedigrees and flowcharts
SamTools (Li <i>et al.</i> , 2009)	A set of tools for post-processing of read alignments in SAM/BAM format
SIFT (Ng and Henikoff, 2003)	A web server for prediction of possible impact of amino acid substitutions
The Reference Sequence Database	A database for vertebrate reference genome
Ubuntu 14.04 operating system / Biolinux	Operating system with built-in bioinformatics packages
UCSC database (Kent <i>et al.</i> , 2002)	WEB browser of University of California Santa Cruz
Varsome	Human genomics search engine
Varsifter (Teer <i>et al.</i> , 2012)	A Java program for the display, sorting or filtering of high-throughput data
Vcftools (Danecek <i>et al.</i> , 2011)	A package designed to work with VCF files

Table 3.8. General laboratory equipment and instruments.

Equipment	Model/ Company
Autoclave	ASB620T, Astell, UK
Balance	TE612, Sartorius, Germany
Centrifuges	C1301, Labnet, USA
	Microfuge16, Beckman Coulter, USA
Deep Freezers	2021D (-20 °C), Arçelik, Turkey
	HT5786-A (-86 °C), Hettich, Germany
DNA extraction	MagNA Pure Compact Instrument, Roche, Switzerland
Documentation System	GelDoc Documentation System, BIO-RAD, USA
Eppendorf Tubes	0.5mL, 1.5mL Biol-proof Tubes, Axygen, USA
Falcon Tubes	EasyOpen 50-ml Centrifuge Tubes, JETBIOFIL, USA
Glassware	Isolab, Germany
Heat Block	BBA1, Grant-Boekel, USA
Heater	MR3001, Heidolph, Germany
Microcentrifuge tubes	0.5mL Thin-wall Flat Cap Tubes, Axygen, USA
	1.5mL Biol-proof Tubes, Axygen, USA
	0.2mL Flat Cap PCR Tubes, Axygen, USA
	0.2mL 8 Strip PCR Strip Tubes, Starlab, belgium
Microwave oven	Intellowave MD554, Arçelik, Turkey
Micropipettes	Rainin Pipet-Lite XLS, Mettler-Toledo International Inc., USA
Multichannel pipettes	Finnpipette F2, Thermo Scientific, USA
Pipetor	Pipetus, Hirschmann Laborgeraete
Pipettes	Finnpipette, Thermo Scientific, USA
Refrigerator	4250T, Arçelik, Turkey
Spectrophotometer	NanoDrop ND-2000c, Thermo Scientific, USA
Thermal Cyclers	T100, Biorad, USA
	TC-512, Techne, UK
Tips/ Filter Tips	Axygen
Vortex	FisonsWhirlimixer, Heidolph, Germany
Water Purification System	Arium® 611UV Ultrapure Water System, Sartorius, Germany

4. METHODS

4.1. DNA Isolation from Whole Blood

Genomic DNA (gDNA) was isolated from 1ml whole blood of patients, family members and healthy controls. Spectrophotometric measurement of 2µl of gDNA was performed with NanoDrop. Samples with absorbance values of A260/280 and A260/230 between 1.8 and 2.0 passed the quality criteria.

4.2. Agarose Gel Electrophoresis and Extraction from the Gel

Agarose gel electrophoresis was performed for the quality check of gDNA samples and to visualize PCR products in genotyping experiments. 1% and 2% agarose gels were prepared for gDNA and PCR products, respectively. 0.5X TBE was used in the preparation of the gels and as running buffer. Loading dye:sample ratio was 1:1 for the genomic DNA and 2:5 for the PCR products.

PCR products were extracted from the agarose gel for further analysis, using QIAQuick Gel Extraction Kit according to the manufacturer's protocol.

4.3. Screening of Common ALS-causing Genes

Screening of mutations in the five most common ALS-causing genes *C9orf72*, *SOD1*, *TARDBP*, *FUS* and *UBQLN2* was performed according to several criteria. Since the hexanucleotide repeat expansion in the *C9orf72* gene are the most common cause of both sporadic and familial forms of ALS, all patients with or without family history of ALS, were analyzed for the expansion mutation in the *C9orf72* gene. This thesis comprises the results of *C9orf72* repeat expansion screening for all ALS cases at NDAL (n=1201). The thesis further

includes the analysis of *SOD1*, *TARDBP*, *FUS* and *UBQLN2* genes in 208 fALS cases, that were referred to the laboratory after 2014.

The GGGGCC hexanucleotide repeat expansion in the promoter region of *C9orf72* was tested using the repeat-primed PCR (RP-PCR) protocol (Figure 4.1., Tables 4.1 and 4.3) and a regular flanking PCR was performed to identify the zygosity and the size of the repeats within the normal range using control samples with known expansion sizes (Tables 4.2. and 4.4). Both reactions were performed with fluorescently labeled (5'-FAM) primers for fragment length analysis. PeakScanner software is used for the visualization of the results.

Conventional screening for *SOD1*, *TARDBP*, *FUS* and *UBQLN2* was performed in all familial ALS cases. Mutational analyses in *TARDBP* and *FUS* genes were restricted to their hotspots, which are exon 5 for *TARDBP* and exons 14 and 15 for *FUS*. Since *UBQLN2* gene mutations are inherited in an X-linked dominant manner, *UBQLN2* screening was restricted to maternal inheritance. Exon 4 of the *SOD1* gene was screened in all cases with consanguineous parents, independent of family history, to identify/exclude some rare recessive mutations reported to cause ALS. The tables below compile the PCR reagents and conditions for mutational screening of *SOD1* (Tables 4.5, 4.6, 4.7, 4.11 and 4.12), *TARDBP* (Tables 4.8 and 4.13), *FUS* (Tables 4.8 and 4.13) and *UBQLN2* (Tables 4.9, 4.10 and 4.14) genes. Primers were either obtained from literature or they were designed using Primer3web and tested using in-silico PCR tool of UCSC database. Sanger sequencing was outsourced to MacroGen (MacroGen Inc., Korea) and CLC Main software was used for analysis.

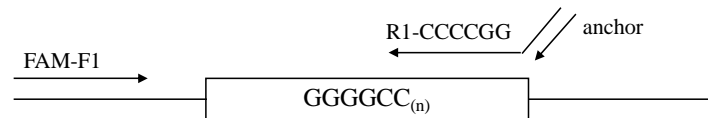


Figure 4.1. Repeat primed PCR design for the amplification of *C9orf72* hexanucleotide repeat expansion.

Table 4.1. PCR reagents for C9orf72 RP-PCR.

Reagent	Volume (µl)	[Stock]	[Final]
Roche FastStart Mix	14	2X	1X
Primer Mix*	2	-	-
DMSO	2	100%	7%
Q-Solution (QIAGEN)	5	5X	1X
7-deaza-dGTP	0,5	5mM	0,9mM
MgCl ₂	1	25mM	0,9mM
DNA	3,5	30-150ng/µl	100-500ng
Total	28		
*Primer Mix	Volume (µl)	[Stock]	[Final]
C9orf72-PRPCR-F1	20	100µM	20µM
C9orf72-PRPCR-R	10	100µM	10µM
C9orf72-PRPCR-anchor	20	100µM	20µM
dH ₂ O	50	-	-
Total	100		

Table 4.2. PCR reagents for C9orf72 flanking PCR.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	5,5	-	-
NEB OneTaq 2X Master Mix with GC Buffer	12,5	2X	1X
Forward Primer	1	10µM	0,4µM
Reverse Primer	1	10µM	0,4µM
DNA	5	10ng/µl	50ng
Total	25		

Table 4.3 PCR conditions for C9orf72 flanking PCR.

Step	Temperature (°C)	Duration (minutes)	# of cycles
Initial denaturation	95	15	1
Denaturation	94	1	2
Annealing	70	1	
Extension	72	3	
Denaturation	94	1	3
Annealing	68	1	
Extension	72	3	
Denaturation	94	1	4
Annealing	66	1	
Extension	72	3	
Denaturation	94	1	5
Annealing	64	1	
Extension	72	3	
Denaturation	94	1	6
Annealing	62	1	
Extension	72	3	
Denaturation	94	1	7
Annealing	60	1	
Extension	72	3	
Denaturation	94	1	8
Annealing	58	1	
Extension	72	3	
Denaturation	94	1	5
Annealing	56	1	
Extension	72	3	
Final extension	72	10	1
Hold	20	∞	1

Table 4.4. PCR conditions for C9orf72 flanking PCR.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	98	5 min	1
Denaturation	97	30 sec	11
Annealing	65 (per cycle -1°C)	30 sec	
Extension	68	90 sec	
Denaturation	97	30 sec	24
Annealing	55	30 sec	
Extension	68	90 sec	
Final extension	68	10 min	1
Hold	20	∞	1

Table 4.5. PCR reagents for SOD1 Exon1, Exon3.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	16,8	-	-
Buffer	5	5X	1X
Forward Primer	1	10µM	0,4µM
Reverse Primer	1	10µM	0,4µM
Mytaq Polymerase	0,2	5U/µl	1U
DNA	1	20 ng/µl	20ng
Total	25		

Table 4.6. PCR reagents for SOD1 Exon2.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	15,6	-	-
Buffer	5	5X	1X
MgCl ₂	2	25mM	2mM
dNTP	0,2	25mM	0,2mM
Forward Primer	0,5	20	0,4µM
Reverse Primer	0,5	20	0,4µM
GoTaq Polymerase	0,2	5U/µl	1U
DNA	1	20 ng/µl	20ng
Total	25		

Table 4.7. PCR reagents for SOD1 Exon4, Exon5.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	16,8	-	-
Buffer	5	5X	1X
Forward Primer	1	20µM	0,8µM
Reverse Primer	1	20µM	0,8µM
Mytaq Polymerase	0,2	5U/µl	1U
DNA	1	20 ng/µl	20ng
Total	25		

Table 4.8. PCR reagents for TARDBP Exon5, FUS Exons 14 and 15.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	5,5	-	-
Buffer	3	5X	1X
MgCl ₂	1,5	25mM	2,5mM
dNTP	1,8	5mM	600µM
Forward Primer	1	10µM	0,6µM
Reverse Primer	1	10µM	0,6µM
GoTaq Polymerase	0,2	5 U/µl	1U
DNA	1,5	20 ng/µl	30ng
Total	15		

Table 4.9. PCR reagents for UBQLN2 primer sets 1 and 4.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	17,9	-	-
Buffer (20mM Mg ⁺²)	2,5	10X	1X
dNTP	2	2,5mM	0,2mM
DMSO	1	100%	4%
Forward Primer	0,25	20	0,2µM
Reverse Primer	0,25	20	0,2µM
GoTaq Polymerase	0,1	5U/µl	0,5U
DNA	1	20 ng/µl	20ng
Total	25		

Table 4.10. PCR reagents for UBQLN2 primer sets 2, 3 and 5.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	16,8	-	-
Buffer (20mM Mg ⁺²)	2,5	10X	1X
dNTP	2	2,5mM	0,2mM
DMSO	2	100%	8%
Forward Primer	0,3	20	0,24µM
Reverse Primer	0,3	20	0,24µM
ExTaq Polymerase	0,1	5U/µl	0,5U
DNA	1	20 ng/µl	20ng
Total	25		

Table 4.11. PCR conditions for SOD1 Exons 1,3,4 and 5.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	95	1 min	1
Denaturation	95	15 sec	35
Annealing	variable*	15 sec	
Extension	72	20 sec	
Final extension	72	8 min	1
Hold	20	∞	1
*Annealing temperatures specific for each primer pair are listed in Table 3.4 (Section 3.4)			

Table 4.12. PCR conditions for SOD1 Exon2.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	94	4 min	1
Denaturation	94	30 sec	32
Annealing	52,2	30 sec	
Extension	72	30 sec	
Final extension	72	8 min	1
Hold	20	∞	1

Table 4.13. PCR conditions for TARDBP and FUS.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	94	5 min	1
Denaturation	94	30 sec	35
Annealing	66,1	30 sec	
Extension	72	75 sec	
Final extension	72	10 min	1
Hold	20	∞	1

Table 4.14. PCR conditions for UBQLN2 primer sets.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	96	2 min	1
Denaturation	95	30 sec	30
Annealing	56	30 sec	
Extension	72	60 sec	
Final extension	72	5 min	1
Hold	20	∞	1

4.4. Detection of *C9orf72* Promoter Methylation Levels

4.4.1. Bisulfite Sequencing

The 5mC levels of the *C9orf72* promoter regions harboring 26 CpG sites were detected using direct bisulfite sequencing assay (BST-PCR). Sodium bisulfite reaction enables the discrimination of methylated and unmethylated cytosines in the genome by converting unmethylated cytosines (C) to uracils (U) and then to thymines (T), while the methylated cytosines (mC) remain intact (Figure 4.2). EZ DNA Methylation-Gold™ Kit (Zymo Research, USA) was used for bisulfite conversion of gDNA.

The bisulfite conversion protocol includes the following steps:

- Prepare fresh CT Conversion Reagent by dissolving the powder in 900µl water, 300µl M-Dilution Buffer and 50µl M-Dissolving Buffer and mixing for 10 minutes.
- Prepare gDNA dilution containing 400ng DNA in a total volume of 20µl.
- Add 130µl CT Conversion Reagent to 20µl DNA sample and mix.
- Perform following incubations: 98°C for 10 minutes, 64°C for 2.5 hours and keep at 4°C.
- Add 600µl of M-Binding Buffer to a Zymo-Spin™ IC Column, then add the sample and mix by inverting several times.
- Centrifuge at full speed (>10,000xG) for 30 seconds and discard flow-through.
- Add 100µl of M-Wash Buffer to the column, spin 300 seconds.
- Add 200µl of M-Desulphonation Buffer to the column and wait for 15-20 minutes. After incubation spin at full speed for 30 seconds.
- Add 200µl of M-Wash Buffer to the column and spin at full speed for 30 seconds. Repeat the step.
- Place the column into a 1.5ml Eppendorf tube, add 15µl of M-Elution Buffer to the column matrix and spin down to elute the converted gDNA.

When the reaction was complete, regions of interest were amplified using nested PCR with specially designed primers that are able to bind the converted sequence (Tables 4.15 and 4.16). ZymoTaq™ Premix (Zymo Research) was used for these consecutive PCR amplifications. Methylation levels were detected by direct evaluation of Sanger sequencing results. Commercially available human methylated (100%) and non-methylated (0%) standards were used as controls, 50% control was prepared by mixing equal amounts of commercial standards (Zymo Research, USA). Number of methylated CpG sites was calculated for each individual and two-tailed Fisher's exact test was used to calculate the significance. The maximum number of methylated CpG sites among controls (2/26) was considered as the threshold for hypermethylation.

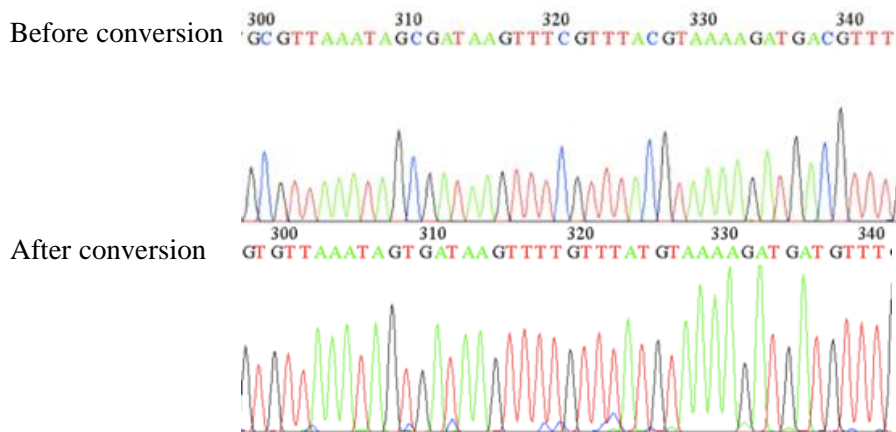
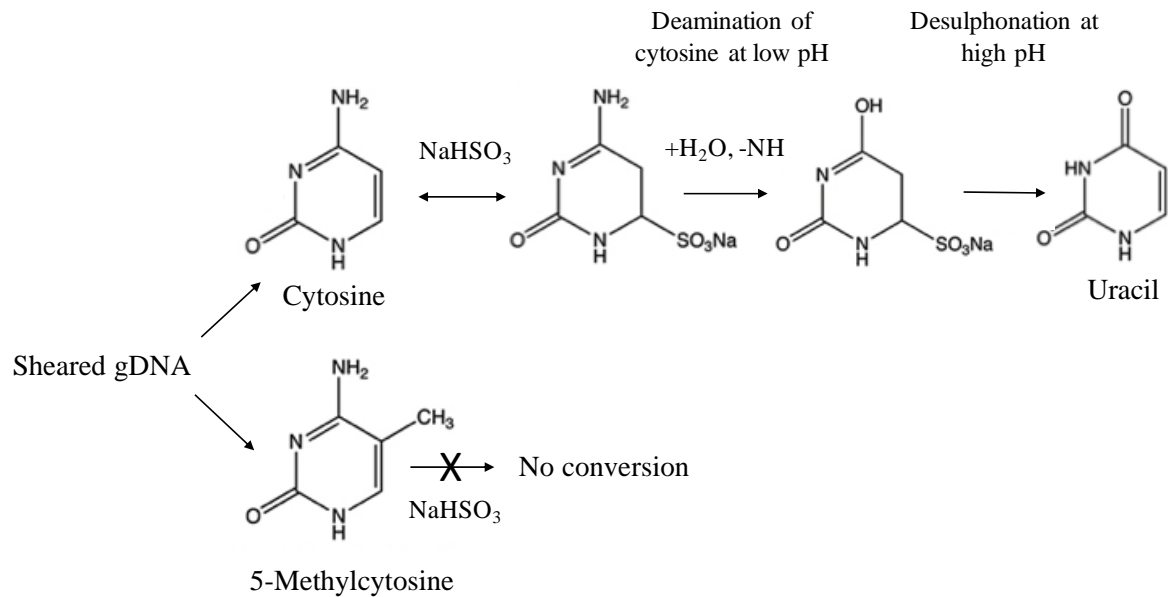


Figure 4.2. Bisulfite conversion reaction diagram.

4.4.2. Methylation-specific Restriction Enzyme Digestion

Methylation-specific restriction enzyme digestion is based on digestion of unmethylated DNA while sparing the methylated sequence due to the methylated C (mC) at the site of restriction. Two hundred nanograms of gDNA was treated with 2 units of HhaI (G-CGC/CGC-G) restriction enzyme using 10X NEB CutSmart® Buffer at a final concentration of 1X (New England Biolabs, USA). HhaI-treated and -untreated genomic DNA samples (200ng) were

subjected to PCR, flanking the restriction site in the promoter region of the *C9orf72* gene (Table 4.17 and 4.18). PCR products were run on a 1.5% agarose gel and band intensities were calculated using ImageJ software. Percent methylation was calculated from the ratio of intensity of digested and undigested samples.

Table 4.15. PCR reagents for nested PCR reactions of bisulfite sequencing.

Nested PCR 1				Nested PCR 2			
Reagent	Volume (µl)	[Stock]	[Final]	Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	9	-	-	dH ₂ O	8	-	-
Zymo Mix	12,5	2X	1X	Zymo Mix	12,5	2X	1X
Forward Primer	1	10µM	0,4µM	Forward Primer	1	10µM	0,4µM
Reverse Primer	1	10µM	0,4µM	Reverse Primer	1	10µM	0,4µM
DNA	1,5	40 ng/µl	60ng	PCR Product from 1 st PCR	2,5	-	-
Total	25			Total	25		

Table 4.16. Nested PCR conditions for bisulfite sequencing.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	95	10 min	1
Denaturation	95	30 sec	35
Annealing	variable*	40 sec	
Extension	72	45 sec	
Final extension	72	7 min	1
Hold	20	∞	1
*1 st PCR: 58 & 2 nd PCR: 58			

Table 4.17. PCR reagents for amplification after RE digestion with HhaI.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	15,8	-	-
Buffer	5	5X	1X
Forward Primer	1	10µM	0,4µM
Reverse Primer	1	10µM	0,4µM
Mytaq Polymerase	0,2	5U/µl	1U
DNA	2	20 ng/µl	40ng
Total	25		

Table 4.18. Conditions for PCR amplification after HhaI restriction enzyme digestion.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	95	3 min	1
Denaturation	95	15 sec	35
Annealing	60,5	15 sec	
Extension	72	15 sec	
Final extension	72	8 min	1
Hold	20	∞	1

4.5. Whole Exome Sequencing Analyses (WES)

In this thesis, WES was outsourced to four different centers, either in the framework of collaborations (Yale University (USA), University of Massachusetts Medical School (UMASS, USA)) or on a commercial basis (Scientific and Technological Research Council of Turkey (TUBITAK, Turkey), Macrogen Inc. (South Korea)). Three μg of gDNA per individual were used as starting material for fragmentation and exome enrichment steps. Bioinformatic analysis was performed using in-house scripts previously prepared by former NDAL students. Paired-end sequencing reads obtained from sequencing platforms were aligned to the human reference genome GRCh37 plus the decoy using Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Quality control and variant calling was performed with HaplotypeCaller tool of Genome Analysis Toolkit (GATK) (McKenna *et al.*, 2010). Genomic variant call format (gvcf) files generated by GATK, containing all nucleotide information of an individual (both variant and reference), were joined together for family members by GenotypeGVCF tool of GATK.

The ANNOVAR Software was used for structural and functional annotation of variants (Wang *et al.*, 2010). MAF values were recruited from 1000 Genomes (1000G) Project and National Heart, Lung and Blood Institute Exome Sequencing Project (NHLBI-ESP6500). Functional consequences of variants were predicted by several sources (e.g. SIFT, PolyPhen2 and GERP ++). DANN scores were also calculated. Association of the candidate gene with

any known human phenotype was obtained from the OMIM database. Annotated variants were filtered using VarSifter Software (version 1.7) according to the mode of inheritance and the minor allele frequency (MAF) (greater than 0.01) (Teer *et al.*, 2012). Functional predictions were used for evaluation, but not for filtration of variants. The absence of the identified candidate variations was confirmed in the Exome Aggregation Consortium (ExAC) and in the in-house exome database of NDAL consisting of 420 Turkish patients and healthy family members. ACMG verdict is determined for each candidate variant for further evaluation.

4.5.1. Validation and Segregation Analyses in Families Subjected to WES

Candidate disease-causing variations identified after WES analyses were validated in the index cases by PCR amplifications of the specific regions of interest, followed by Sanger sequencing. The segregation of the variant with the disease was then shown in all available affected and healthy family members. Reagents and conditions of validation PCRs and segregation experiments are listed in Tables 4.19 and 4.20.

Table 4.19. Reagents for validation and segregation PCRs.

Reagent	Volume (µl)	[Stock]	[Final]
dH ₂ O	16,8	-	-
Buffer	5	5X	1X
Forward Primer	1	10µM	0,4µM
Reverse Primer	1	10µM	0,4µM
Mytaq Polymerase	0,2	5U/µl	1U
DNA	1	20 ng/µl	20ng
Total	25		

Table 4.20. Conditions for validation and segregation PCRs.

Step	Temperature (°C)	Duration	# of cycles
Initial denaturation	95	1 min	1
Denaturation	95	15 sec	35
Annealing	variable*	15 sec	
Extension	72	20 sec	
Final extension	72	8 min	1
Hold	20	∞	1
*Annealing temperatures specific for each primer pair are listed in Table 3.5 (Section 3.5)			

4.6. Homozygosity Mapping

Homozygosity mapping was applied to detect the runs of homozygosity (ROH) that are expected to include the trait loci in affected cases, offspring of consanguineous parents with an autosomal recessive inheritance pattern. The input of homozygosity mapping can either be a SNP array or WES/WGS data. Homozygosity mapping from whole exome sequencing data was performed for two families in this thesis (Family 10 and TARDBP-Gly348Cys Family), using `--homozyg` command in PLINK v1.09. Text files were generated containing family, gender and phenotype information as an input in addition to the family vcf containing SNPs. Variants in linkage disequilibrium were not pruned. The parameters used for detection of ROHs are summarized in Table 4.21 (Purcell *et al.*, 2007). ROHs detected with PLINK were plotted using the R programming language. The squared genotype correlation coefficient (r^2) for linkage disequilibrium between two variants identified on the same chromosome in Family 10 was calculated using the `--ld` command in PLINK v1.09 among our in-house exome database of 420 individuals.

Table 4.21. PLINK parameters for homozygosity mapping.

PLINK Parameters	Value	Definition
--minimum GQ	20	minimum genotype quality
--homozygous kb	500	minimum length
--homozygous SNP	10	minimum SNP count
--homozygous window SNP	20	scanning window size
--homozygous window SNP missing	10	maximum missing calls in scanning window hit
--homozygous window threshold	0.05	minimum scanning window hit rate
--homozygous density	500	maximum inverse density (kb/SNP)
--homozygous gap	2000	maximum internal gap kb length
--homozygous window het	2	maximum heterozygotes

4.7. Data Analyses in Samples from Project MinE

4.7.1. Variant Accumulation in sALS Patients

Accumulation of rare or common variants per individual was questioned in 625 sALS cases and 152 neurologically healthy controls subjected to WGS. Number of variants per individual was calculated and plotted for five frequency thresholds (1, 0.5, 0.1, 0.05, 0.01) and for four categories, being allelic, heterozygous (het), homozygous (hom), and heterozygous or homozygous (het_or_hom). In order not to lose population specific variants, that would be eliminated in online-databases, we have used the internal allele frequency of each variant calculated among 777 Turkish MinE samples. An in-house Python script was used to count all protein-coding and noncoding variants from pre-prepared vcf files. Variant numbers were plotted using violin plot type to observe the current and expected distributions. Student's t-test was applied to detect any significant differences between patients and control groups for all different categories and frequency thresholds.

4.7.2. Pathogenic Variations in sALS Patients

Project MinE samples were screened for protein coding variations in all ALS-causing and -associated genes listed in Table 1.1. Previously generated vcf files were used to scoop out the variants in genes of interest. These annotated vcf files contained missense and disruptive variants called in all 777 individuals were are pre-filtered, based on certain criteria (Figure 4.3). Variant lists for each gene were generated using -tabix in Python command tool. Candidate variants identified in sALS patients were further analyzed for pathogenicity using prediction tools, VarSome Software and the in-house exome database.

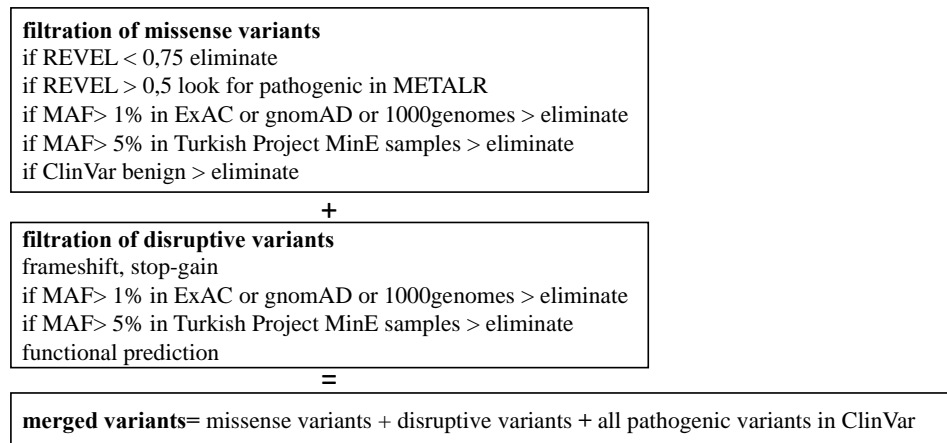


Figure 4.3. Filtration criteria of variants in Project MinE samples.

5. RESULTS

Diverse strategies are adopted throughout this thesis to define disease-causing and/or modifying factors in our cohort of Turkish ALS patients. Results of mutational analyses include conventional screening of common ALS genes. Further, application of NGS technologies to identify ALS-causing mutations in rare and novel genes in both familial and sporadic disease is aimed. *C9orf72*-based ALS is studied in detail with WES analyses and promoter methylation level experiments on *C9orf72* expansion carriers. The work on Project MinE samples exemplifies variant/mutation analyses in a large ALS cohort.

5.1. Screening of Common ALS Genes by Conventional Approaches

5.1.1. *C9orf72* hexanucleotide Repeat Expansion Analyses

The GGGGCC hexanucleotide repeat expansion in the *C9orf72* gene was detected in 33 families and in 32 sporadic cases out of 1201 ALS patients (Figure 5.1., Table 5.1). The distribution of the repeat sizes in the non-expanded allele of *C9orf72* gene in ALS cases with or without the expansion and in controls revealed two, five and eight repeats as the predominant allelic variants in the Turkish population (Figure 5.2).

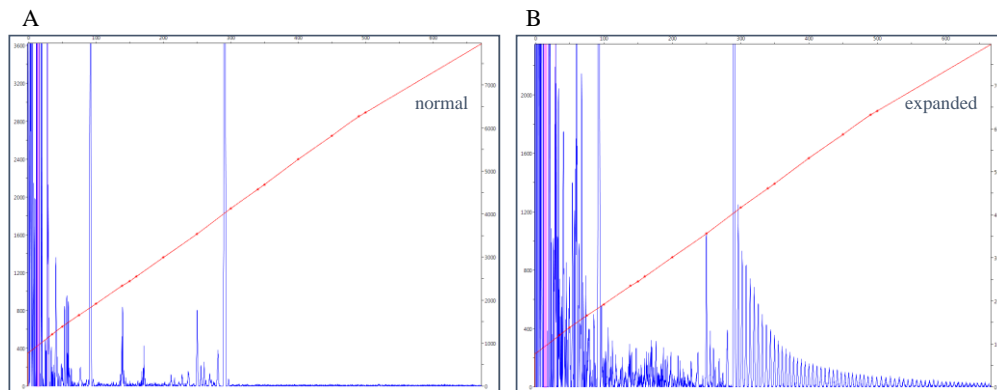


Figure 5.1. Example for *C9orf72* RP-PCR result. Normal (A) and expanded (B).

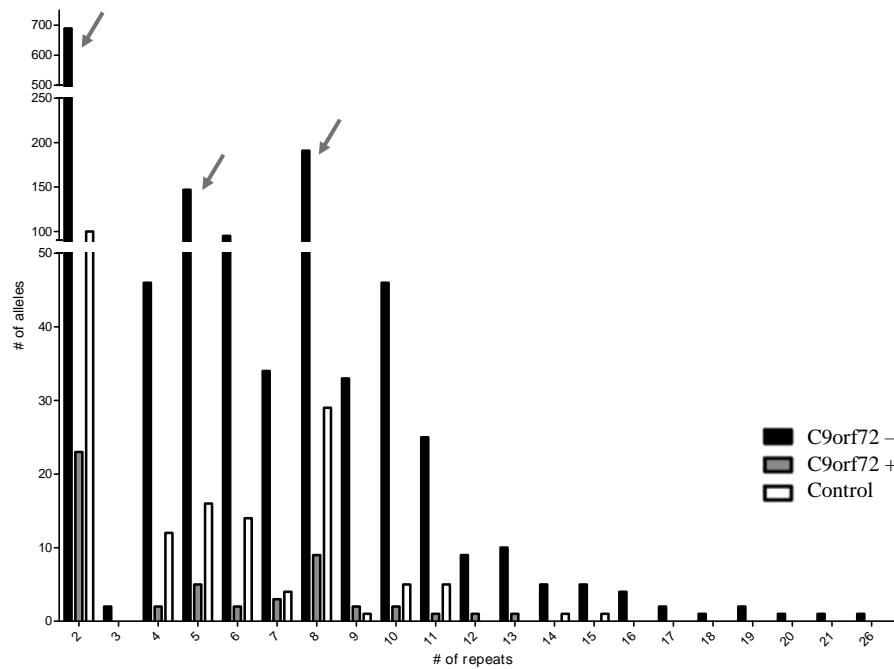


Figure 5.2. Distribution of the repeat sizes in the non-expanded allele of the *C9orf72* gene. Arrows indicate the most frequently observed alleles.

Table 5.1. Clinical data of *C9orf72* hexanucleotide expansion carriers in our cohort.

		Total ALS	fALS	sALS	FTD
#	probands	65	33	32	-
	family members	6+ 5 asymptomatic	6	-	2
	male:female		1,3	1,7	
SO	limb	39	22	21	-
	bulbar	18	13	5	-
	limb+bulbar	5	1	3	-
	unknown	6	3	3	-
dementia		14	6	6	2
AO	juvenile (<25yrs)	1	1	-	-
	middle (25-45 yrs)	9	5	4	-
	late (>45 yrs)	61	33	28	2
	range (years)	32-80	32-80	40-71	57-59
MAO	total \pm SD	54 \pm 10,9	55 \pm 9,6	53 \pm 8,4	-
#: numbers; AO: age of onset; MAO: mean age of onset; SO: site of onset.					

5.1.2. SOD1, TARDBP, FUS and UBQLN2 Analyses

Screening of *SOD1*, *TARDBP*, *FUS* and *UBQLN2* in 159 fALS and in a subgroup of 563 sALS cases revealed mutations/variations in a total of 54 probands. Mutations identified and the clinical data of cases and affected family members are presented in Table 5.2. Figure 5.3 displays the locations of amino acid changes in the domains of SOD1, TDP-43, FUS and UBQLN2 proteins.

Fifteen distinct mutations in the *SOD1* gene were identified in 44 patients from 35 families. The SOD1-p.Leu144Phe mutation, the most common mutation in the Balkan region, is observed in 16 patients from nine families, and it is also the most common SOD1 mutation in the cohort under study (Battistini *et al.*, 2013). The SOD1-p.Asp90Ala change located in the exon 4 of the gene is a highly characterized mutation with dual inheritance pattern (both autosomal dominant and recessive). The recessive haplotype is very common among Scandinavian populations (Andersen *et al.*, 1995) and explains the disease cause in six consanguineous Turkish families investigated in this thesis.

Screening of the hotspots of *TARDBP* and *FUS* genes revealed nine different mutations in ten families; three additional cases were solved with WGS (Project MinE). The heterozygous FUS-p.Pro525Leu and FUS-p.Tyr526Cys mutations were detected in four isolated juvenile cases without a family history. Segregation analyses in the unaffected parents of three patients revealed the *de novo* occurrence of both mutations, analysis is ongoing in the fourth patient (ALS-377).

Two variations located outside the hotspot of the *UBQLN2* gene, p.Thr134Ile and p.Ser340Ile, are observed in four families. The p.Ser340Ile variation does not segregate with the disease in one family with additional affected and unaffected members. The pathogenicity of the novel p.Thr134Ile variation could not be further questioned, since there are no available family members of the index case.

Table 5.2. Clinical data of patients with *SOD1*, *TARDBP*, *FUS* and *UBQLN2* mutations.

Gene	ALS ID	Protein Change	Gender	Age of Onset	Site of Onset	Inheritance	ALS Type
SOD1	1398	p.Ala4Thr	F	25	B	AD	familial
	1167†	P.Val14Met	M	42	L	AD	sporadic
	960	p.Gln22Leu	F	30	L	AD	familial
	1327	p.Val31Ala	M	45	B	AD	familial
	1547†		F	64	L		sporadic
	1453	p.Gly37Arg	M	41	L	AD	familial
	802	p.Glu40Gly	F	39	L	AD	familial
	816		F	32	L		familial
	1450	p.Ser68Pro ⁿ	M	54	L	AD	familial
	623†	p.Asn86Ser	F	42	L	AD	sporadic
	1207		M	48	L		familial
	102†	p.Val87Met	F	29	L	AD	sporadic
	741	p.Asp90Ala	F	32	L	AR	familial
	810		F	29	L		familial
	1256		M	44	L		familial
	1359		M	35	L+B		familial
	1545		F	64	L		sporadic
	1579		M	51	L		sporadic
	561	p.Leu117Val	F	62	L	AD	familial
	1527		F	38	L		familial
	1396		F	62	L		familial
	1412		F	40	L		familial
	1472†		F	36	L		sporadic
	1439		F	24	L	AR	familial
	355†	p.Asp125Asn	M	50	L	AD	sporadic
	1716*	p.Glu133Lys	F	37	L	AR	sporadic
	1064†		M	34	L	AD	sporadic
	281†	p.Leu144Phe	M	57	L	AD	sporadic
	607		F	45	L		familial
	713		F	53	L		familial
	724		M	52	L		familial
	727		F	NA	L		familial
	1773		M	60	L		familial
	635		F	54	L		familial
	772		M	49	L		familial
	1059		F	51	L		familial
	1063		F	56	L		familial
	1235		F	59	L		familial
	935		M	37	L		familial
	1036		F	60	L		familial
	1633		M	64	L		familial
	1691		M	60	L		familial
	1715		F	34	L		familial
	97	p.Val148Gly	F	46	L	AD	familial

Table 5.2. Clinical data of patients with *SOD1*, *TARDBP*, *FUS* and *UBQLN2* mutations (cont.).

Gene	ALS ID	Protein Change	Gender	Age of Onset	Site of Onset	Inheritance	ALS Type
TARDBP	1082	p.Gly298Val ⁿ	F	66	L	AD	familial
	600	p.Ala315Thr	F	57	L	AD	familial
	1448		M	62	L		familial
	408†		M	48	L		sporadic
	910	p.Gly348Cys	M	37	L	AD	familial
	919		M	42	L		familial
	911		M	NA	L		familial
	660	p.Gln354Glu ⁿ	F	42	L	AD	sporadic
	976	p.Ala382Thr	M	39	L	AD	familial
FUS	1034	c.1394-1 G>T ⁿ	F	47	L	AD	familial
	1208	p.Gln519Ter	F	42	L	AD	familial
	485†	p.Pro521Leu	M	39	L	-	sporadic
	581	p.Arg524Met	M	53	L	AD	familial
	549	p.Pro525Leu	M	14	L	<i>de novo</i>	sporadic
	1610*		M	17	L	<i>de novo</i>	sporadic
	377†		F	16	L	-	sporadic
	1423*	p.Tyr526Cys	M	12	L	<i>de novo</i>	sporadic
UBQLN2	637	P.Thr134Ile ⁿ	M	58	L	XD	sporadic
	683	p.Ser340Ile	M	20	NA	XD	sporadic
	498		F	60	L + B		familial
	422		F	60	L + B		familial
	608		M	54	L		familial

Bold indicates index cases, L: limb, B: bulbar, AD: autosomal dominant, AR: autosomal recessive, ⁿ novel, * identified with WES, † identified in Project MinE, NA: not available

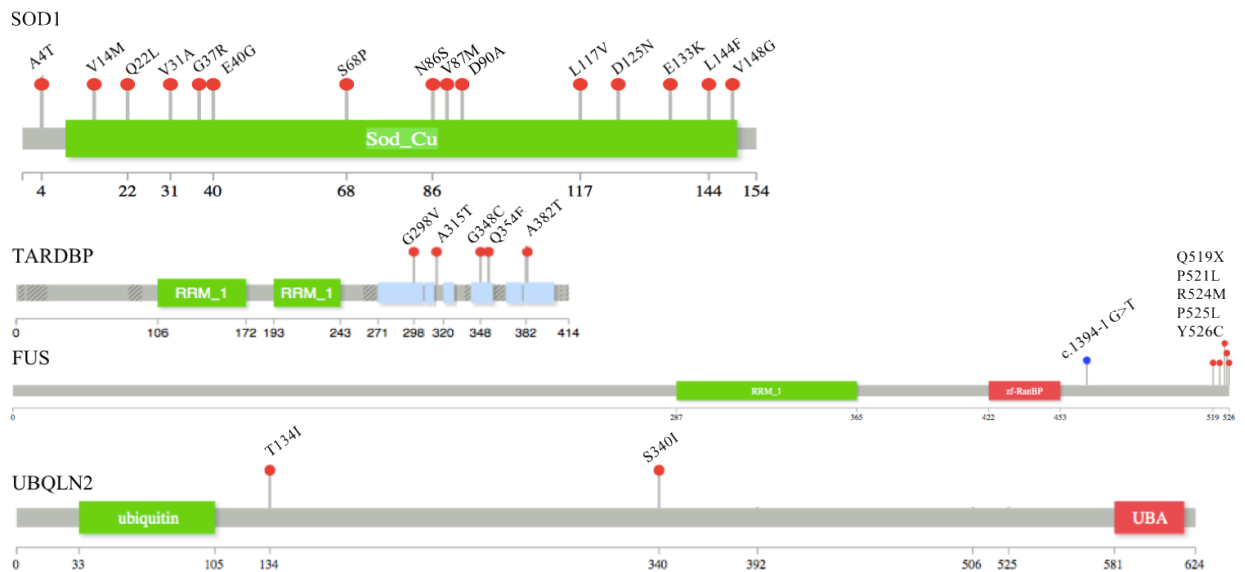


Figure 5.3. Locations of amino acid changes identified in *SOD1*, *TDP-43*, *FUS* and *UBQLN2* proteins.

5.1.3. Three Rare SOD1 Mutations with Distinct Inheritance Pattern

The complex inheritance pattern acknowledged for the SOD1-Asp90Ala mutation, which causes ALS in both autosomal dominant and in autosomal recessive forms, is observed in our cohort in three additional rare changes in SOD1; p.Asn86Ser, p.Leu117Val, and p.Glu133Lys. These variations are identified either through conventional screening or NGS analyses in eight different pedigrees with or without family history of ALS. Ongoing segregation analyses will help to better understand their penetrance and pathogenicity.

5.1.3.1. SOD1-Asn86Ser. The SOD1-p.Asn86Ser mutation previously reported in homozygous form in Turkish ALS patients (Özoğuz *et al.*, 2015), is also detected in heterozygous form in two apparently sporadic ALS cases, in one male and a female in this thesis (identified through Project MinE). This rare mutation has been reported to cause the disease in both heterozygous and homozygous forms in familial and apparently sporadic cases with variable penetrance among family members (Hayward *et al.*, 1998; Khani *et al.*, 2015). The pathogenicity of the heterozygous SOD1-p.Asn86Ser mutation could be investigated only in one patient (SOD1-Asn86Ser#1), whose core family is very large. None of the healthy sibs (seven of them older than the index) or the healthy mother carried the heterozygous p.Asn86Ser change. The father, who had died at the age 50 in a car accident, could not be tested (Figure 5.4). DNA samples from the family members of the second SOD1-p.Asn86Ser case is currently not available.

5.1.3.2. SOD1-Leu117Val. The SOD1-p.Leu117Val variation is also reported to show intra- and inter-familial variability in inheritance pattern (Synofzik *et al.*, 2012). This variation is identified in four independent families in our cohort (Figure 5.5). The heterozygous variant is observed in two families (SOD1-Leu117Val#1 and #2) with four affected members, some having prolonged disease over 10 years with ages of onset varying from 38 to 62. The sporadic patient from consanguineous parents (SOD1-Leu117Val#3) carrying the homozygous p.Leu117Val change has an early age of onset of 24 and displays a faster disease progression

(Figure 5.5). Yet a fourth patient carrying the SOD1-p.Leu117Val change in heterozygous form is an apparently sporadic patient without any family history of ALS (pedigree not shown). Analysis to confirm the segregation and to understand the penetrance of the variation is ongoing.

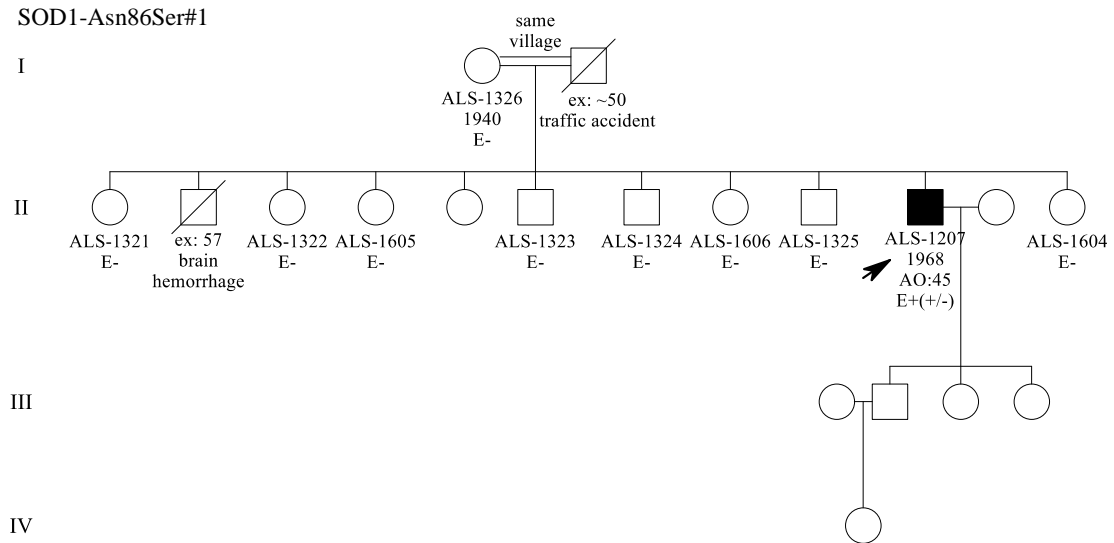


Figure 5.4. Pedigree of the family with a heterozygous SOD1-p.Asn86Ser variation. Segregation of the identified variation E(SOD1-p.Asn86Ser) is designated below pedigree symbols. AO: age of onset, ex: exitus.

5.1.3.3. SOD1-Glu133Lys. The homozygous p.Glu133Lys mutation is identified in a female patient with a relatively young age of onset of 37 and a progressive disease through WES analysis (Figure 5.6). Sanger sequencing in the neurologically healthy parents showed the heterozygous change in both. The p.Glu133Lys change has been previously shown to cause ALS in autosomal dominant manner in a late-onset male patient (AO:62) (Lange *et al.*, 2017). The autosomal dominant form of the variation is also observed in a male patient (ALS-1064) in our cohort with no family history of ALS; analysis is ongoing in additional family members to assess the pathogenicity of the variant. Altogether, these findings suggest a low penetrance for the variation and a severe disease phenotype in the presence of two mutated copies.

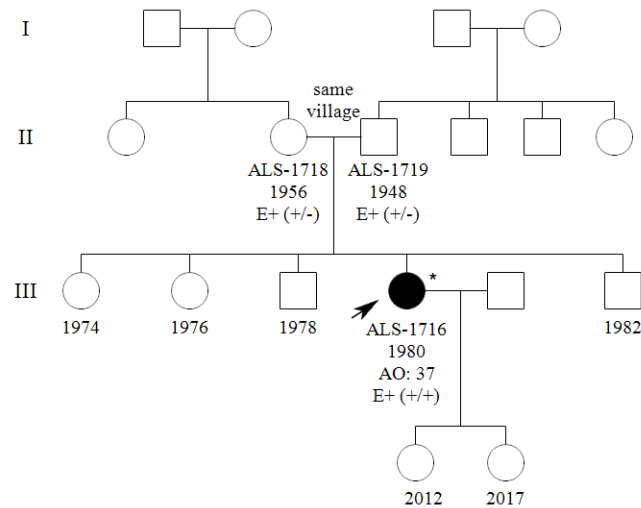


Figure 5.6. Pedigree of the family with a heterozygous SOD1-p.Glu133Lys variation. Segregation of the identified variation E(SOD1-p.Glu133Lys) is designated below pedigree symbols. AO: age of onset.

5.2. Promoter Methylation Levels of *C9orf72* Expansion Carriers

A total of 52 *C9orf72* expansion carriers and 31 neurologically healthy age- and sex-matched controls were subjected to bisulfite conversion and sequencing. The maximum number of methylated CpG sites among controls is 2/26, so individuals with more than two methylated sites are considered as hypermethylated. The results of bisulfite sequencing assay revealed a significant increase in promoter hypermethylation for the expansion carriers ($p < 0.0023$) (Table 5.3). There is no significant correlation between number of CpGs methylated in the promoter region of *C9orf72* and age of onset of patients (Figure 5.7). In a separate study, conducted with this data, we have found that *C9orf72* promoter hypermethylation positively correlates (although non-significant) with whole-blood methylation levels of the expansion carriers (Hamzeiy *et al.*, 2018).

Table 5.3. *C9orf72* promoter methylation levels of expansion carriers versus controls.

Methylation level	ALS expanded		Controls	
	# of Individuals	Frequency	# of Individuals	Frequency
No Methylation (0)	34	0,65	29	0,94
Low Methylation (1-2)	5	0,10	2	0,06
High Methylation (3-26)	13	0,25	0	0
Total	52	-	31	-
p-value (high methylation)	0.0023			

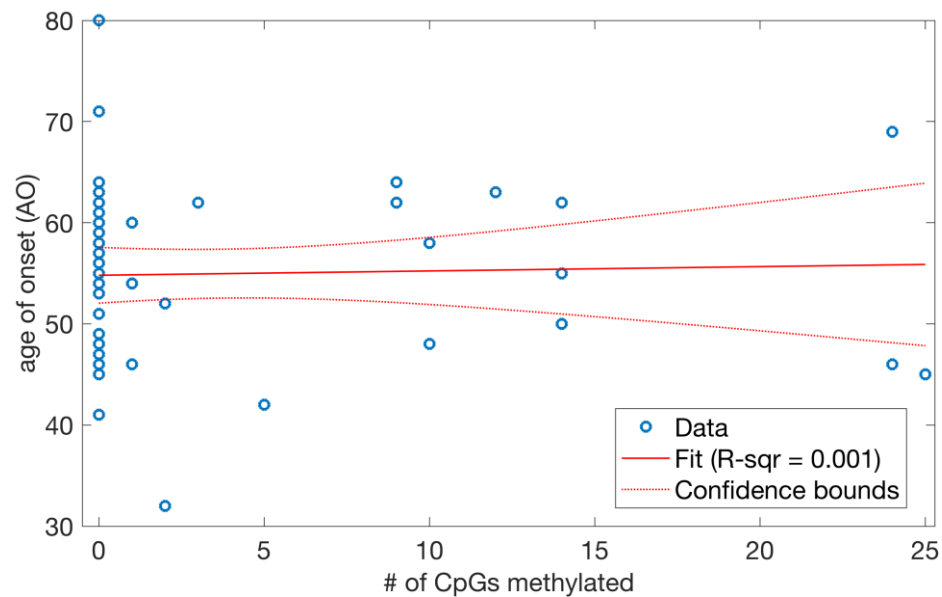


Figure 5.7. Correlation between age of onset and the number of methylated CpG sites.

Methylation-specific restriction enzyme digestion is performed in 30 expansion carriers. Percent methylation levels detected in the samples do not represent a correlation with the results of the bisulfite sequencing assay (Figure 5.8).

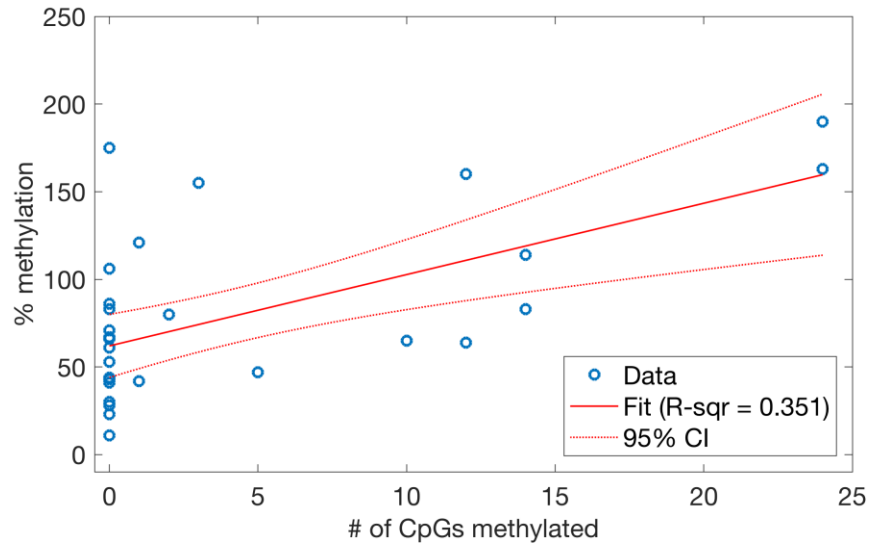


Figure 5.8. Correlation between two distinct methylation assays. Number of methylated sites out of 26 detected by bisulfite sequencing are denoted in the x-axis, percent methylation detected by restriction enzyme digestion assay is shown on y-axis.

5.3. Whole Exome Sequencing Analyses for Identification of Mutations in Rare and Novel ALS Genes

Whole exome sequencing in 39 pedigrees, with a dominant inheritance pattern or first-degree consanguinity revealed distinct rare or novel mutations in 20 families. Among the cases solved, two had autosomal dominant inheritance, one had a *de novo* mutation and 13 carried homozygous disease-causing variations, adding up to a discovery rate of 43% for WES in this cohort. In four families phenotype-genotype correlations and segregation analyses are ongoing to assess the pathogenicity of the variations. Details of the candidate variations and their OMIM associations are presented in Table 5.4. Further information regarding the MAF and functional predictions of the variants are displayed in Table 5.5. In most cases in which consanguinity and juvenile disease-onset are present, mutations in genes that are not primarily associated with typical ALS were identified, reflecting the clinical and pathogenic complexity behind neurodegenerative diseases.

Table 5.4. Variants identified in families subjected to WES and their OMIM associations.

Family No	NDAL ID	Gene	Variation			Zygoty	OMIM Association
			DNA Change	Protein Change	Transcript ID		
1	ALS1539	KIF5A	c.3005A>G	p.Asp1002Gly	NM_004984.2	Het	ALS, SPG10
	ALS1584						
2	ALS1381	TBK1	c.1434-1435delTG	p.Val479Glufs*4	NM_013254	Het	ALS
3	ALS1610	FUS	c.1562C>T	p.Pro525Leu	NM_001170937	<i>de novo</i>	ALS
4	ALS46	SPG11	c.6224A>G	p.Asn2075Ser	NM_025137	Hom	ALS, SPG11, CMT2X
			c.7155T>G	p.Tyr2385Ter	NM_025137	Hom	
5	ALS132	SPG11	c.7132T>C	p.Phe2378Leu	NM_025137	Hom	
			c.2250delT	p.Phe750Leufs*3	NM_025137	Hom	
6	HSP29	SPG11	c.7076delG	p.Gly2359Glufs*14	NM_0125137	Hom	
7	HSP33	SPG11	c.1203delA	p.Lys401Lysfs*14	NM_0125137	Hom	
		SACS	c.932C>T	p.Thr311Ile	NM_001278055	Hom	SACS
8	ALS158	SYNE1	c.22930C>T	p.Gln7644Ter	NM_182961.3	Hom	ARCA1
9	ALS314	SYNE1	c.23524C>T	p.Arg7842Ter	NM_182961.3	Hom	ARCA1
10	ALS518	ERLIN1	c.281T>C	p.Val94Ala	NM_006459	Hom	SPG62
11	ALS1164	ZFYVE26	c.2615-2617delGCTinsTGAA	p.Arg872Hisfs*17	NM_015346.3	Hom	SPG15
12	ALS764	DNAJB2	c.14A>G	p.Tyr5Cys	NM_001039550	Hom	Distal SMA
13	ALS1594	SIGMAR1	c.358A>G	p.Thr120Ala	NM_147157	Hom	ALS, SMA
14	ALS445	SBF1	c.3157C>T	p.Arg1053Trp	NM_002972	Hom	CMT4B3
15	ALS1716	SOD1	c.400G>A	p.Glu133Lys	NM_00454	Hom	ALS
16	HSP34	SPG7	c.1715C>T	p.Ala572Val	NM_003119.3	Hom	SPG7
17	ALS1813	SPG7	c.771-772delTG	p.Val258Glyfs*30	NM_003119.3	Hom	SPG7
18	ALS1704	VRK1	c.961C>T	p.Arg321Cys	NM_003384	Hom	PCH1A
19	ALS731	TFG	c.854C>T	p.Pro285Leu	NM_001007565	Het	HMSN Okinawa type
20	ALS1708	MORC2	c.489G>T	p.Glu163Asp	NM_001303256	Het	CMT2Z

5.3.1. Families with an Autosomal Dominant Inheritance Pattern

5.3.1.1. Kinesin Family Member 5A (KIF5A) (Family 1). The male index case (ALS-1539) was referred to our laboratory with ALS diagnosis. He initially had weakness in his upper limbs at the age of 50, followed by bulbar involvement. Clinical examination stated amyotrophy, muscle atrophy, fasciculations, spasticity and extensor plantar reflex. He displayed motor deficit, his reflexes in the lower limbs were hyperactive and he had positive Hoffman's sign. Family history showed several affected members who have succumbed to the disease. An affected older step brother (ALS-1584), living with the disease for six years since he was 45 years old, had similar clinical symptoms with the addition of dysarthria. Three individuals including an unaffected aunt were subjected to WES. The novel heterozygous c.3005A>G, p.Asp1002Gly change in the *KIF5A* gene is identified in this large ALS kindred with autosomal dominant inheritance. The results of the segregation analysis are depicted in the pedigree of Family 1 (Figure 5.9). A younger sister of the index case (ALS-1586), positive for the heterozygous variation, is currently unaffected.

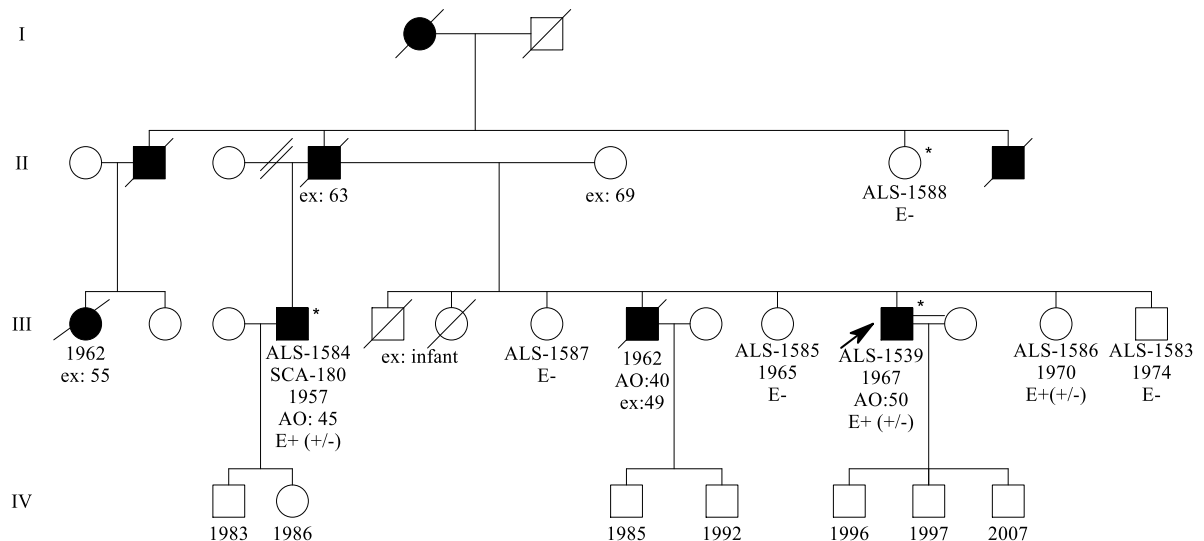


Figure 5.9. Pedigree of Family 1. E (*KIF5A*, c.3005A>G, p.(Asp1002Gly)), * indicates individuals with WES results, AO: age of onset.

5.3.1.2. Tank-binding Kinase 1 (TBK1) (Family 2). A previously identified heterozygous c.1434-1435delTG, p.Val479Glufs*4 truncation mutation in the *TBK1* gene is identified by WES in our cohort in a male index case with an AO of 20 (Figure 5.10). He was a professional wrestler when he started to suffer from tremor in hands and mild difficulties in his speech. He developed weakness in shoulders that spread throughout the body. ALS diagnosis was firmly established after electroneuromyography (EMG) and motor evoked potential (MEP) findings. The pathogenic mutation, (GERP: 5,25) leads to the deletion of 248 amino acids in the C-terminal coiled-coil regions of the protein resulting in haploinsufficiency. Both parents of the index case are reported to be healthy, segregation analysis in this family is in progress to observe the genotypes of the unaffected parents.

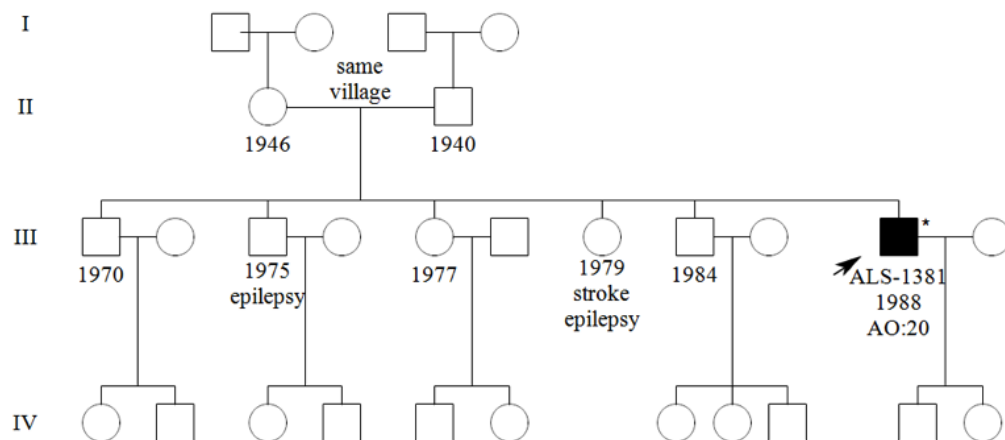


Figure 5.10. Pedigree of Family 2. * indicates the individual with WES results, AO: age of onset.

5.3.1.3. TRK-Fused Gene (TFG) (Family 19). The female index case (ALS-731) with an apparent autosomal dominant disease history was initially diagnosed with ALS with accompanying sensory neuropathy symptoms (Figure 5.11). Affected members of the family had disease survival of 10-15 years. Screening of common ALS genes did not reveal any pathogenic variation. WES detected the well-characterized heterozygous c.854C>T, p.Pro285Leu variation in the *TFG* gene, associated with hereditary sensory motor neuropathy, Okinawa type (HSMNO) (MIM#604484) (Lee *et al.*, 2013a). The p.Pro285Leu variation,

absent both in the in-house exome database and in ExAC, has a top GERP score of 6,16 and is predicted to be pathogenic with DANN algorithm (0,9986). A secondary variation in the *VABP* gene, associated with ALS (MIM#608627) and SMA (MIM#182980) is also present among the candidate variations, which is an asparagine to glutamic acid change at position 130 of the protein (NM_004738) (GERP:4,38, DANN:0,95). This variant is also absent in the in-house database, but listed in ExAC (MAF:0,0014) and is registered as likely benign by ClinVar. Segregation analysis in the family is likely to confirm the pathogenic variation in the *TFG* gene, that is in correlation with the phenotype of the index case.

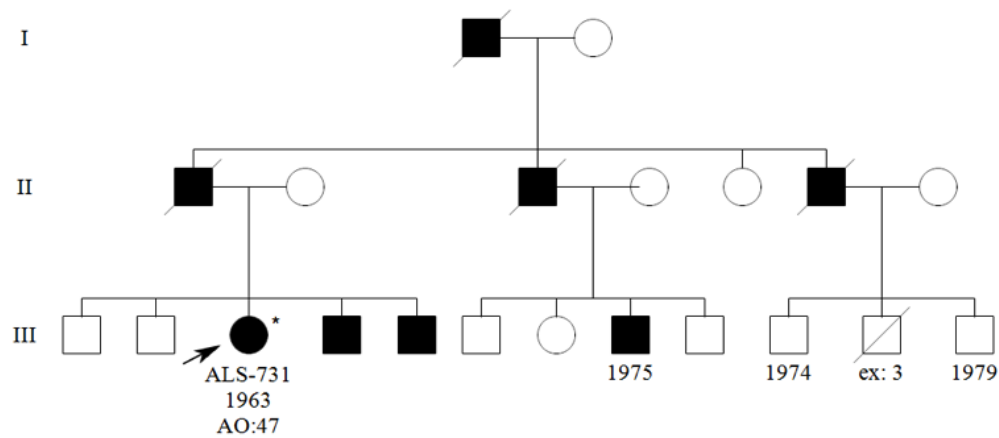


Figure 5.11. Pedigree of Family 19. * indicates the individual with WES results, AO: age of onset.

5.3.1.4. MORC Family CW-Type Zinc Finger 2 (MORC2) (Family 20). The 21-year-old male patient had progressive clinical features characterized by muscle spasms in the thumb and weakness and muscle atrophy in hands (Figure 5.12). His LMN symptoms were prominent. Multifocal motor neuropathy and brachial plexus neuropathy were excluded before his referral to NDAL with anterior horn involvement. WES in the index case detected a heterozygous variation in the *MORC2* gene (c.489G>T, p.Glu163Asp), associated with CMT type 2Z (MIM#616688). Detailed clinical evaluation by the clinician and the ongoing segregation analyses will confirm the differential diagnosis of CMT2Z in the patient.

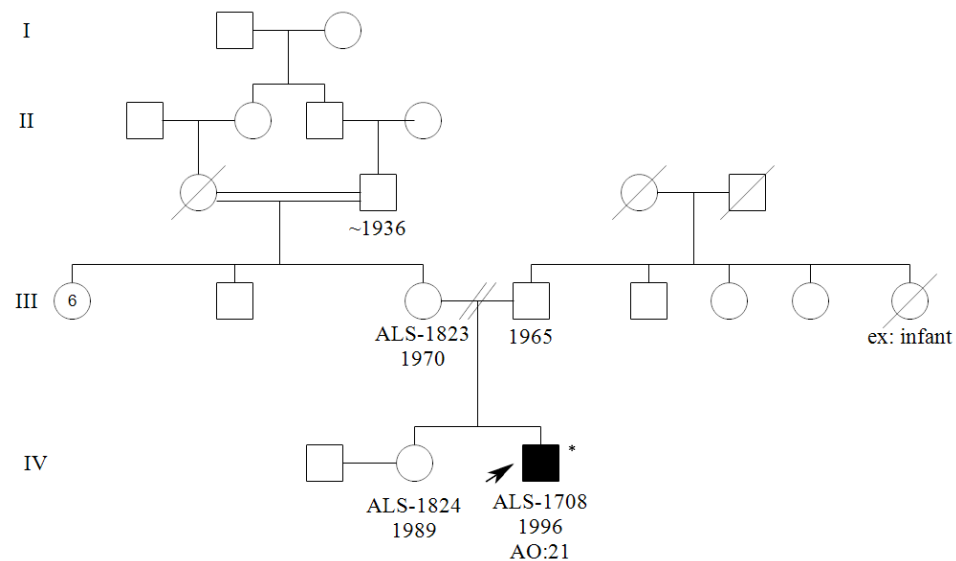


Figure 5.12. Pedigree of Family 20. * indicates the individual with WES results, AO: age of onset.

5.3.2. Families with Autosomal Recessive Inheritance Pattern

5.3.2.1. Spatacsin (SPG11) (Families 4, 5, 6, 7). Four homozygous truncation mutations in the *SPG11* gene are identified using exome sequencing as disease-causing in five patients from four distinct families diagnosed with ARJALS or ARHSP (Iskender et al. 2015). In Family 4, the patient's father developed classical ALS at the age of 50 and died within five years (Figure 5.13). In contrast, the proband, offspring of first-degree consanguineous parents, has a long-standing neurologic history dating to early childhood. At the age of four, he was assessed for moderate psychomotor delay, including fine motor and language abilities. At the age of 23, his clinical situation was re-evaluated due to frequent falling and imbalance with bilateral leg weakness followed by upper extremity weakness. In Family 5, the probands are siblings from first-degree consanguineous parents (Figure 5.13). They had limb-onset disease at the ages of 14 (male) and 16 (female), with walking difficulties, twitching, spasticity, and fasciculations in legs. Both UMN and LMN involvement and dysarthria were present. There is no cognitive impairment and no optical complaints. The patients, now in their 30s, share a similar

phenotype and have a slowly progressive motor neuron disease without any mental disabilities. In addition to the disease-causing truncation mutations identified in Family 4 and 5, affected individuals in each family also carry secondary homozygous missense mutations p.Asn2075Ser (in Family 4) and p.Phe2378Leu (in Family 5) in their *SPG11* genes. The segregation analysis is performed by Sanger sequencing in the available members of both families (Figure 5.13).

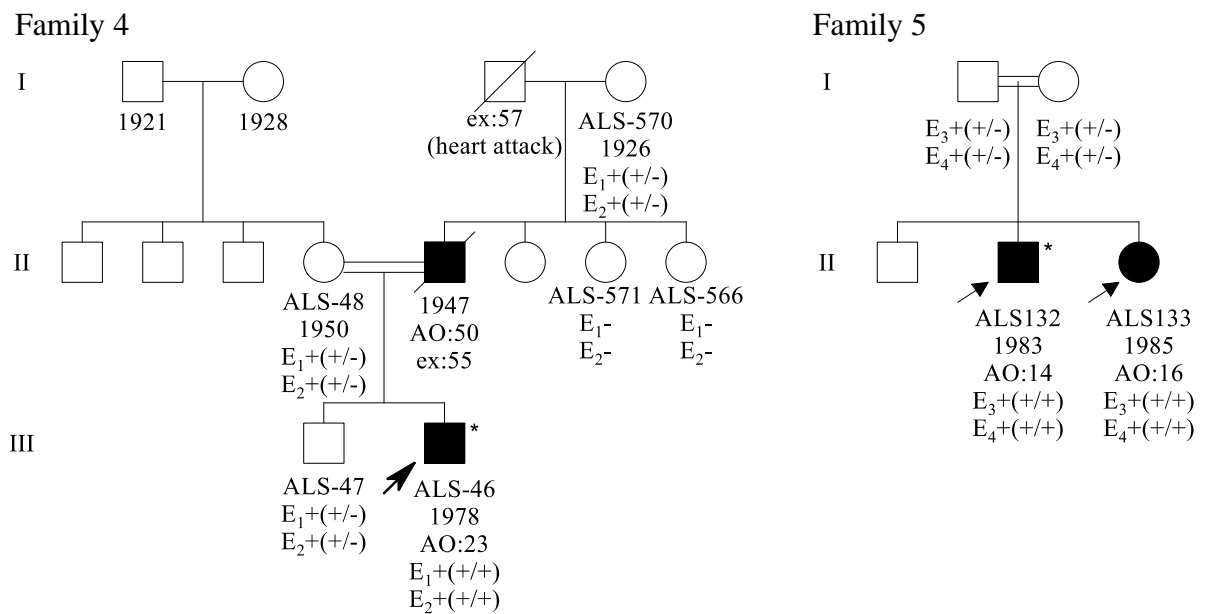


Figure 5.13. Pedigrees of Families 4 and 5. E_1 (*SPG11*, c.7155T>G, p.(Tyr2385Ter)), E_2 (*SPG11*, c.6224A>G, p.(Asn2075Ser)), E_3 (*SPG11*, c.2250delT, p.(Phe750Leufs*3)) and E_4 (*SPG11*, c.7132T>C, p.(Phe2378Leu)), * indicates WES results, AO: age of onset.

Two male patients with consanguineous parents in Families 6 and 7 were diagnosed with ARHSP. The index case in Family 6 (HSP-29) had difficulty in walking and imbalance at the age of 13. His clinical examination showed bilateral nystagmus, spastic paraparesis and thin corpus callosum (TCC). He was positive for Babinski reflex and EMG supported MN involvement. WES identified the novel homozygous c.7076delG, p.Gly2359Glufs*14

truncation mutation, which was validated in the parents, who are heterozygous (Figure 5.14). The male offspring from consanguineous parents in Family 7 (HSP-33) is found to carry two homozygous variations in *SPG11* and *SACS* genes (Figure 5.14). The patient had limping in his left leg, which then progressed to the right leg. Clinical evaluation reports UMN involvement and spastic paraparesis in addition to bilateral resting tremor and magnetic resonance imaging (MRI) shows TCC. The genotype-phenotype correlation by the clinician confirmed a novel homozygous frameshift mutation (p.Lys401Lysfs*14) as disease-causative in the patient. The relevance of the secondary homozygous variation in the *SACS* gene (p.Thr311Ile), which is not present in homozygous state in the NDAL exome database, remains to be elucidated with clinical reassessment of the patient.

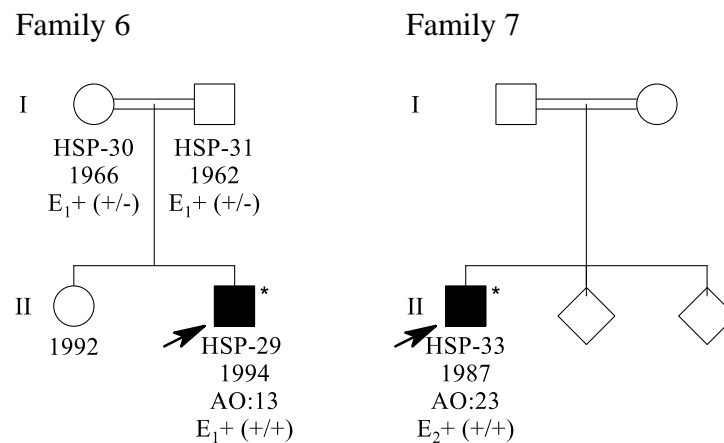


Figure 5.14. Pedigrees of Families 6 and 7. E₁(*SPG11*, c.7076delG, p.(Gly2359Glu fs*14)), E₂(*SPG11*, c.1203delA, p.(Lys401Lys fs*14)), * indicates individuals with WES results, AO: age of onset.

5.3.2.2. Synaptic Nuclear Envelope 1 (SYNE1) (Family 8 and 9). Homozygous nonsense mutations in the *SYNE1* gene are identified in five patients from two families, whose parents are first cousins. Affected members of the two independent families, Family 8 and Family 9, were referred to our center with young-onset ALS and displayed additional ataxic features. Both nonsense mutations identified in the *SYNE1* gene, lead to a premature stop-codon, thus to a truncated protein. The mutations are located in the spectrin repeats of the protein.

In Family 8, two brothers with young-onset ALS symptoms at the age of 20 and 21 showed features of upper and lower motor neuron involvement with dominance in LMN, amyotrophy and muscle atrophy, accompanied by pyramidal cerebellar syndrome. In addition, spastic ataxic gait, dysarthria and dysmetria were present. WES in the index case (ALS-157) revealed a novel homozygous c.22930C>T, p.Gln7644Ter mutation in the *SYNE1* gene, which was also shown to be present in the affected brother (ALS-158) (Figure 5.15). The disease progression was fast for both patients. The father and healthy sister are found to be heterozygous carriers of the mutation.

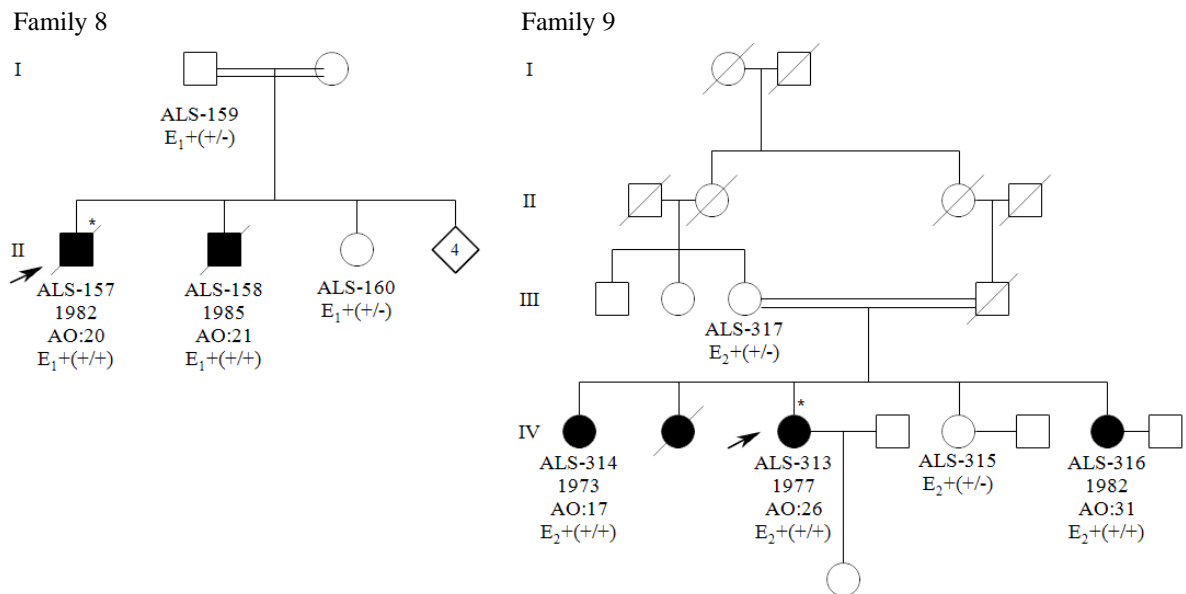


Figure 5.15. Pedigrees of Families 8 and 9. E₁ (*SYNE1*, c.22930C>T, p.(Gln7644Ter)), E₂ (*SYNE1*, c.23524C>T, p.(Arg7842Ter)) * indicates individuals with WES results, AO: age of onset.

In Family 9, two sisters with young-onset ALS (AO:17 and 21) displayed pyramidal cerebellar syndrome, upper and lower motor neuron involvement with dominance in UMN and amyotrophy. Spastic ataxic gait, dysarthria and dysmetria were present. Additionally, oculomotor abnormalities were also reported. Patients did not carry mutations in common ALS genes. Family history revealed a deceased sister with similar symptoms and a younger sister with recently emerging symptoms at the age of 21. The homozygous c.23524C>T,

p.Arg7842Ter (rs775935265) mutation was detected in the index case (ALS-313) via exome sequencing and was validated in all living affected sibs in homozygous form and in the mother and an unaffected sister in heterozygous state by Sanger sequencing (Figure 5.15).

5.3.2.3. Endoplasmic Reticulum Lipid Raft-Associated Protein 1 (ERLIN1) (Family 10).

Whole exome sequencing in three affected and three unaffected members of the large and highly inbred family, revealed a novel homozygous c.281T>C, p.Val94Ala mutation in the *ERLIN1* gene as the cause of the disease in Family 10 (Tunca *et al.*, 2018). The female index patient (ALS-518) was referred to NDAL, at the age of 25, with young-onset ALS and a family history of ALS in the deceased father (IV.24) as well as in the paternal aunt (IV.22) and grandfather (III.18) (Figure 5.16). The grandfather was reported to have walking difficulties after the age of 35; he had then developed swallowing problems and died at 54. The father limped mildly after returning from military service at the age of 20. He underwent surgery with spinal stenosis diagnosis; however, symptoms worsened. EMG showed fibrillation and fasciculation in genioglossus, right first dorsal interossei, biceps, quadriceps, gastrocnemius and tibialis anterior muscles. Upon these findings, he was diagnosed with ALS and succumbed to the disease at the age of 42. The daughter, our index case, had complaints of abnormal gait during her late teens. Her neurological examination at the age of 25 depicted increased deep tendon reflexes (DTRs) in both lower extremities and atrophy in left interosseous muscles, electrophysiological evaluation revealed asymmetrical upper and lower motor neuron signs with fibrillation and fasciculation in genioglossus, right first interossei, biceps and both tibialis anterior muscles, confirming definite ALS. Her brother (ALS-1077) has quadriparesis with mild atrophy in both intrinsic hand muscles and spasticity in both legs. He had slowed MEP in asymmetrically bilateral lower limbs. The electrophysiological examination findings supported LMN involvement. Two additional male individuals in the pedigree also had mild gait abnormalities (ALS-869 and ALS-856). Neurological examination of ALS-869, 52 years old, showed hyperreflexia in the legs, positive Babinski sign on the right and unresponsive plantar reflex on the left foot with bilateral clonus. ALS-856 has difficulty in step climbing.

Variant filtration for autosomal recessive inheritance pattern remained with 21 gene variants, only two of which had a MAF lower than 0.01. The homozygous presence of the *ERLIN1* gene variation was validated by Sanger sequencing in four affected family members (ALS-869, ALS-856, ALS-518 and ALS-1077); additional 22 family members were shown to be either heterozygous or wild-type for the variation, are symptom-free at present and not expected to develop the disease (Figure 5.16). Homozygosity mapping, performed from WES data, detected the *ERLIN1* locus in all affected individuals, but not in the controls (Table 5.6). The valine to alanine substitution at position 94 of *ERLIN1*, located in a highly conserved region, was predicted to be disease-causing, deleterious, and possibly damaging by MutationTaster, SIFT and PolyPhen2. No other gene associated with ARHSP or ALS was mutated in the affected individuals.

The second homozygous change remaining after filtration was the c.2714G>T, p.Arg905Ile variation in the ATP Binding Cassette Subfamily C Member 2 (*ABCC2*) gene reported to be associated with Dubin-Johnson Syndrome and characterized by hyperbilirubinemia (ExAC MAF:0.000049) (Kajihara *et al.*, 1998). The *ABCC2* gene resides in the same homozygous region with *ERLIN1*. Segregation analysis for this variation showed that the allelic states (genotypes) for *ERLIN1* and *ABCC2* genes were the same for all tested family members (Figure 5.16). Pair-wise linkage disequilibrium for *ERLIN1* and *ABCC2* variations was estimated as $r^2=1$ among 420 individuals in our in-house exome database, indicating that the two variations are in complete linkage disequilibrium. Also, according to the HapMap data *ABCC2* and *ERLIN1* reside in 0.2 cM distance, mapped to 127.3 cM and 127.5 cM of chromosome 10, respectively.

Table 5.6. ROH on chromosome 10 of individuals with WES data. Intervals marked with asterisk harbor the ERLIN1 and ABCC2 genes and are only present in affected individuals.

PHOM: Proportion of homozygous calls.

ALS ID	Phenotype	CHR	Start Position	End Position	Length (kb)	# of SNPs	PHOM
ALS518	affected	10	101571528*	102120255	548.728	70	0.471
			105738868	106256915	518.048	177	0.508
ALS856	affected	10	21682974	22198352	515.379	45	0.444
			74452931	75335161	882.231	201	0.642
			90982081	91738825	756.745	127	0.606
			96011865	96577509	565.645	112	0.545
			97302787	98220462	917.676	240	0.504
			101567426*	102379297	811.872	120	0.55
			103436852	104191800	754.949	127	0.528
			104552865	105510081	957.217	167	0.563
			105607966	106274197	666.232	209	0.656
			113674195	114381762	707.568	90	0.567
ALS869	affected	10	21805466	22826093	1,020.628	63	0.46
			35226069	35872130	646.062	58	0.517
			48557340	49201898	644.559	21	0.524
			60489808	61112161	622.354	55	0.418
			74584314	75643885	1,059.572	280	0.614
			90903791	91567961	664.171	122	0.623
			95210173	95780656	570.484	151	0.55
			95826281	96609568	783.288	128	0.57
			97287754	98481733	1193.98	300	0.513
			101409994*	102385066	975.073	166	0.536
ALS859	unaffected	10	74695287	75374735	679.449	185	0.6
ALS862	unaffected	10	22193422	22822744	629.323	33	0.303
			50828506	51464656	636.151	96	0.5
ALS864	unaffected	10	22193336	22826093	632.758	36	0.444
			96011865	96540536	528.672	109	0.514
			104552865	105176179	623.315	108	0.5

5.3.2.4. Zinc Finger FYVE Domain-Containing Protein 26 (ZFYVE26) (Family 11). The female index case (ALS-1164) with first degree consanguineous parents had teenage onset complaints of progressive walking difficulties. Initial diagnosis was restricted to motor neuron disease. Clinical examination showed definitive upper motor neuron signs in cervical, thoracic and lumbar regions. WES performed in the index case revealed the novel homozygous c.2615-2617delGCTinsTGAA, p.Arg872Hisfs*17 frameshift mutation in the *ZFYVE26* gene associated with a form of ARHSP, spastic paraplegia 15 (SPG15) (MIM#270700). The p.Arg872Hisfs*17 variation leads to a truncated protein, which is 2539 amino acids long in the wild-type form. Sanger sequencing in both parents revealed, as expected, the variation in heterozygous state. (Figure 5.17).

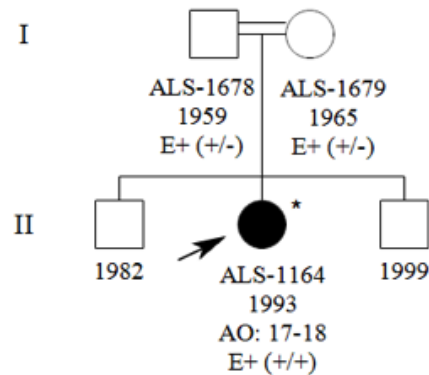


Figure 5.17. Pedigree of Family 11. E (*ZFYVE26*, c.2615-2617delGCTinsTGAA, p.(Arg872Hisfs*17)), *indicates individuals with WES results, AO: age of onset.

5.3.2.5. DNAJ/HSP40 Homolog, Subfamily B, Member 2 (DNAJB2) (Family 12). The male index case (ALS-764) was reported to have walking difficulties at the age of 22, coupled with cramps and aching of muscles in lower extremities. WES was performed only in the index patient with the initial diagnosis of juvenile ALS. After filtration against mode of inheritance and MAF, the previously reported c.14A>G, p.Tyr5Cys variant in the *DNAJB2* gene, associated with distal SMA (MIM#614881), explained the phenotype in the patient. The autosomal recessive inheritance pattern of the missense change is confirmed in both parents (Figure 5.18). Genotyping could not be executed for the two cousins of the index case with

walking difficulties, since their DNA was not available; however, family members stated that they had different clinical properties.

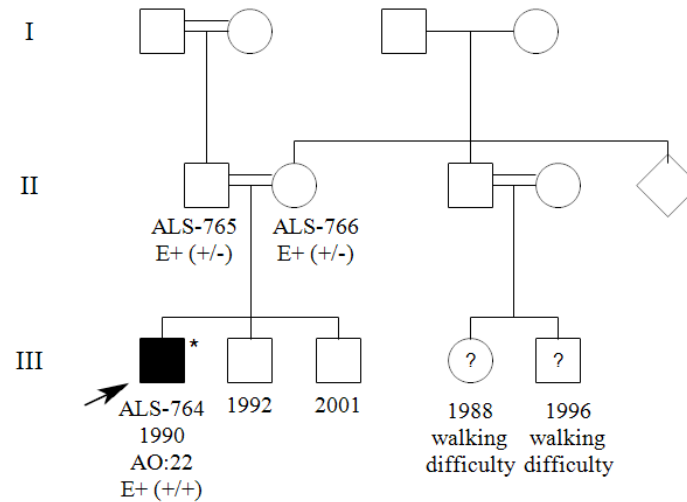


Figure 5.18. Pedigree of Family 12. E (*DNAJB2*, c.14A>G, p.(Tyr5Cys)), * indicates individuals with WES results, AO: age of onset.

5.3.2.6. Sigma Nonopioid Intracellular Receptor 1 (*SIGMAR1*) (Family 13). A novel homozygous c.358A>G, p.Thr120Ala variation in the *SIGMAR1* gene, previously associated with ARJALS (MIM#614373) is observed in a juvenile patient (ALS-1594), offspring of consanguineous parents. Currently 26 years old, he suffered from frequent falling during his late teens and was diagnosed with a progressive form of juvenile ALS. He was positive for both UMN and LMN signs. WES data was available only for the index case, Sanger sequencing in this family revealed the heterozygous presence of the variation in the parents (Figure 5.19). The older brother (ALS-1695) did not carry the variation in both alleles. The prediction algorithms gave high conservation (GERP: 4,67) and pathogenicity scores (DANN: 0,9871) for the novel p.Thr120Ala change. Segregation analysis in the remaining family members is expected to confirm the pathogenicity.

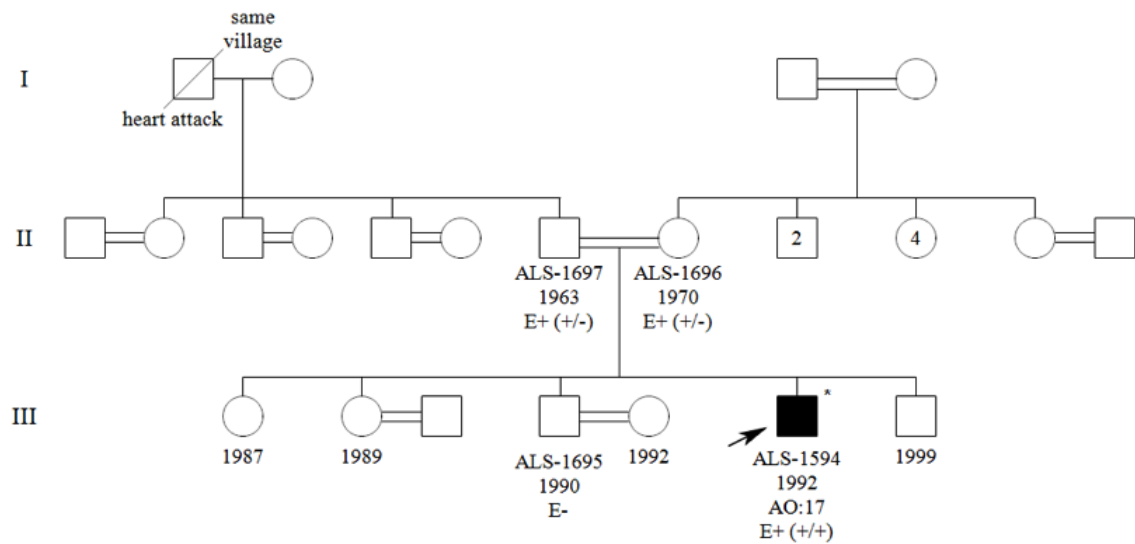


Figure 5.19. Pedigree of Family 13. E (*SIGMAR1*, c.358A>G, p.(Thr120Ala)), * indicates individuals with WES results, AO: age of onset.

5.3.2.7. Set-Binding Factor 1 (SBF1) (Family 14). The juvenile male index case (ALS-445) had predominant features of muscle weakness in all four extremities, pes cavus and swallowing problems. He also developed imbalance, tremor and fasciculations in the hand and occasional cramps in legs. EMG showed axonal and demyelinating polyneuropathy. WES performed in the index case detected a homozygous variant in the *SBF1* gene associated with CMT Type 4B3 (MIM#615284). No other variant associated with any other MND or motor neuropathy was among the candidates. The c.3157C>T, p.Arg1053Trp change is reported in the literature without any information regarding the clinical status of the carrier and the mode of inheritance (rs202156491). The variation is absent in our in-house exome database. The pathogenicity regarding this candidate variant with a high DANN score of 0,9985, a relatively low conservation score (GERP:1,58), still remains to be elusive. Sanger sequencing showed that both parents are carriers for the p.Arg1053Trp variation; ongoing genotype-phenotype correlations with the assistance of the clinician will affirm the causality of the variation (Figure 5.20).

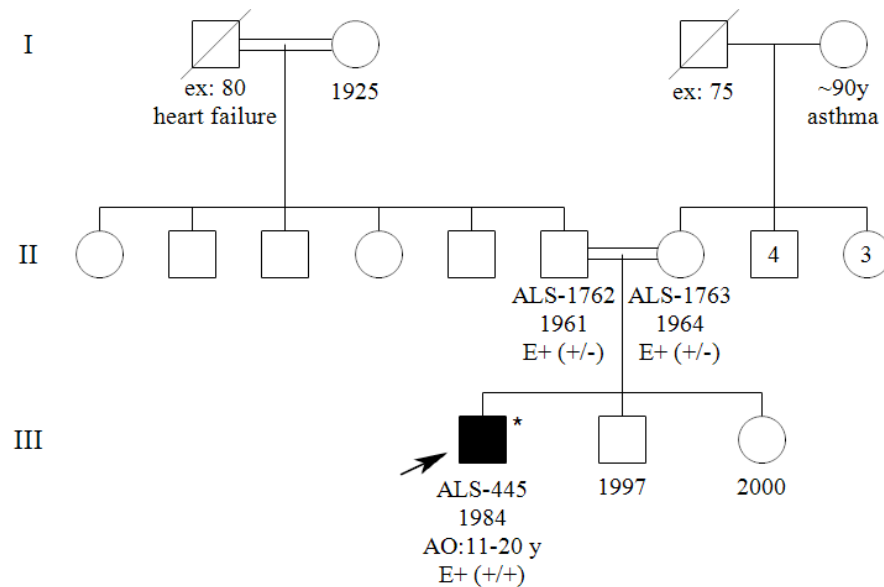


Figure 5.20. Pedigree of Family 14. E (*SBF1*, c.3157C>T, p.(Arg1053Trp)), * indicates individuals with WES results, AO: age of onset, ex: exitus.

5.3.2.8. Paraplegin (SPG7) (Family 16 and 17). WES revealed two distinct homozygous variations, c.1715C>T (p.Ala572Val) and c.771-772delTG (p.Val258Glyfs*30) in Families 16 and 17, respectively (Figure 5.21). The male index case in Family 16 developed progressive pain in his legs and difficulty in walking at the age of 36. Clinical examination detected live DTRs in the lower extremities, bilateral Babinski sign and bilateral cerebellar (intention) tremor in upper extremities. The previously reported homozygous variation in the *SPG7* gene causing spastic paraplegia 7 (MIM#607259) was concordant with the initial diagnosis of ARHSP in the patient. In Family 17, the index case from first degree consanguineous parents had a slow progressive disease for five years after the onset and was referred to our center for differential diagnosis of overlapping HSP and ALS symptoms. He was operated for spinal hernia, but symptoms did not disappear. Clinical features include hyperactive DTRs and LMN involvement is seen in EMG. The novel homozygous truncation mutation (p.Val258Glyfs*30) identified in the patient explained the phenotype. Validation of the *SPG7* variations identified in the index cases and in available family members is ongoing.

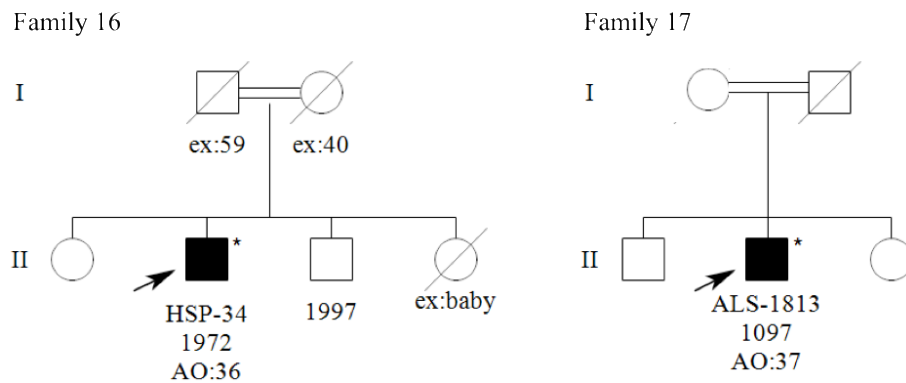


Figure 5.21. Pedigrees of Families 16 and 17. * indicates individuals with WES results, AO: age of onset.

5.3.2.8. Vaccinia Related Kinase 1 (VRK1) (Family 18). Three affected sibs from consanguineous parents in Family 18 suffered from walking disability and muscle weakness (Figure 5.22). The index case (ALS-1704), currently in her 30s, lost ability to walk long distances in the last 10 years, her walk is ataxic, and she cannot walk on her toes. Clinical examination showed LMN and UMN signs, distal and proximal weakness in muscles and a positive Romberg sign indicative of sensory ataxia. Initial diagnosis was not certain, but distal hereditary motor neuropathy or MND was suspected. After filtration against MAF and mode of inheritance, a homozygous variation in the *VRK1* gene, associated with pontocerebellar hypoplasia type 1A (PCH1A) (MIM#607596) characterized by anterior horn cell degeneration, is identified as a candidate variant. In the literature, homozygous and compound heterozygous variants in the *VRK1* gene have been reported in patients with complex motor and sensory neuropathy phenotypes (Gonzaga-Jauregui *et al.*, 2013). Segregation of the variation with the disease in the family is shown in three unaffected and one affected sibling(s) (Figure 5.22). The p.Arg321Cys change is predicted to be damaging and disease-causing by SIFT and MutationTaster, respectively, and has high GERP (5,13) and DANN (0,9892) scores. MRIs of the index patient and her sister do not support the PCH1A phenotype. Deep phenotyping of all affected individuals in the family and additional genotyping are expected to relate the homozygous variation identified in *VRK1* to the previously defined cases with neuropathy phenotypes.

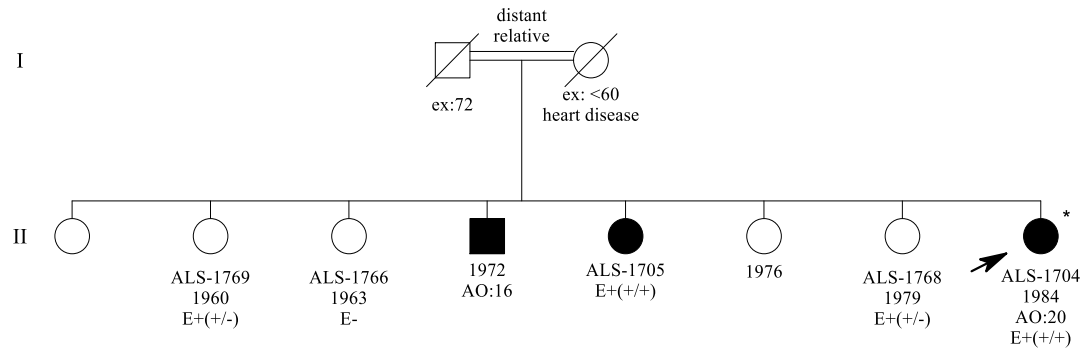


Figure 5.22. Pedigree of Family 18. E (*VRK1*, c.961C>T, p.(Arg321Cys)), * indicates individuals with WES results, AO: age of onset, ex:exitus.

5.3.3. A Family with a Dual ALS Phenotype

A female patient (ALS-20) with an age of onset of 52 years was referred to our laboratory with ALS diagnosis (Family 37, Figure 5.23). She had weakness in her right leg, which then spread to other extremities. Fasciculations were present in both arms and legs. Her neurological examination showed asymmetric quadriparesis, live DTRs, bilateral positive Babinski and Hoffman-Tromner reflexes and atrophy in thenar, hypothenar and frontal arm muscles. EMG suggested LMN involvement. She had progressive disease with the involvement of bulbar signs including respiratory stress and she succumbed to the disease in four years. Her parents were free of disease except that her mother had parkinsonism, however a maternal aunt and a distant cousin were reported to have died from adult-onset motor neuron disease (Figure 5.23). Molecular analyses did not detect mutations in all five common ALS genes (*C9orf72*, *SOD1*, *TARDBP*, *FUS* and *UBQLN2*). Ten years later a female nephew of the index case developed ALS symptoms at the age of 32 (ALS-742). She had fasciculations in her tongue and her left arm and had steppage gait. She was positive for Babinski and Hoffman-Tromner reflexes, DTRs were live for the upper extremities and atrophy was evident in the thenar-hypothenar-interossei and frontal arm muscles. EMG favored anterior horn involvement.

Considering the close consanguinity in the parents of all affected individuals in the pedigree, WES is applied to five individuals to look for an ALS gene with AR inheritance. DNA samples from the index case, her mother, the affected nephew and the parents of the nephew were subjected to WES. Exome analyses revealed the common homozygous SOD1-p.Asp90Ala mutation in the nephew, however the index case did not carry this mutation, as previously shown by conventional screening. Figure 5.23 represents the segregation of the SOD1-p.Asp90Ala mutation among 31 family members. After consulting the clinician, we have confirmed the slow progressive stereotyped homozygous Asp90Ala phenotype in patient ALS-742 with a younger disease onset and prolonged disease survival, now over six years. Her 30-year-old brother (ALS-1271), who has a mild speaking difficulty, was also shown to carry the homozygous mutation.

So far, additional candidate variant analysis to identify the disease-causing variation in the index patient (ALS-20) did not offer a homozygous variation in relevant genes. Although not compatible with the pedigree, the possibility of autosomal dominant inheritance in the family was not overruled, since the clinical picture of the index case and the affected members in the upper generations present with a classical form of ALS with AD inheritance. A heterozygous variant in the *ANXA11* gene (NM_145869, c.1335 G>T, p. Arg445Ser), associated with classical ALS (MIM#617839), is identified as a candidate variant in WES analysis, however the segregation analysis in 31 members of the family was not concordant with the disease phenotype. Analysis in Family 38 is ongoing to explain the genetic cause of the second phenotype observed in the family; clinical follow-ups are regularly made by the family clinician among offspring of affected individuals.

5.3.4. Search for Disease Modifiers and Oligogenic Inheritance in ALS

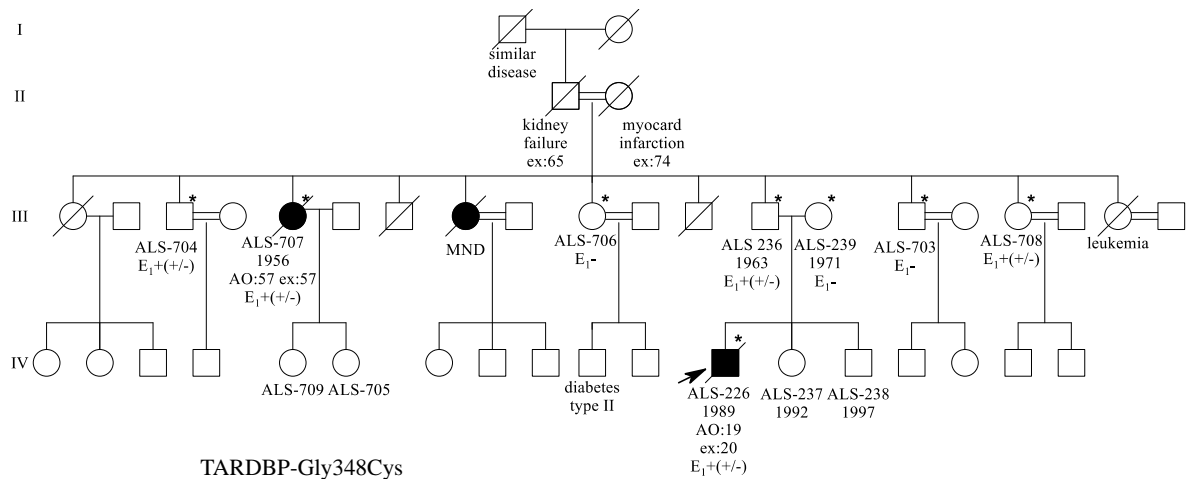
5.3.4.1. Families Showing Incomplete Penetrance. Reduced penetrance of a disease-associated allele has been recognized for ALS as genetic testing became widespread. A considerable number of apparently sporadic cases have mutations in common ALS genes, suggesting low penetrance in asymptomatic parents. In our cohort, we have detected four heterozygous variants in four families, showing incomplete penetrance. These variants leads to p.Gln22Leu, p.Glu40Gly and p.His71Tyr amino acid changes in the SOD1 and the p.Gly348Cys change in the TDP-43 proteins. In the scope of this thesis, the genetic backgrounds of symptomatic and asymptomatic mutation carriers in two of these families are investigated to look for possible modifiers of ALS pathogenesis that either protect the mutation carriers or make them more susceptible to disease (pedigrees are shown in Figure 5.24).

In the family with the novel heterozygous SOD1-p.His71Tyr mutation, the male index case (ALS-226) showed his first disease symptoms at the age of 19 and with a very fast disease progression, he succumbed to ALS in one year. The parents were reported to be healthy, however family history showed a paternal aunt who died with MND. After the identification of the SOD1-p.His71Tyr change in the index patient, analysis showed the presence of the same mutation in the unaffected father (ALS-236), who was 50 years old at that time. Six years later, another paternal aunt (ALS-707) of the index case was diagnosed with ALS at the age of 57; she was shown to be positive for the SOD1-His71Tyr mutation. Her disease course was as progressive as her nephew and she died within eight months. Extended segregation analysis in the available family members showed the presence of the p.His71Tyr mutation in two additional asymptomatic individuals (ALS-704 and ALS-708).

The heterozygous TARDBP-p.Gly348Cys mutation present in the male index case (ALS-910) and in his affected father (ALS-911) was also confirmed in the affected male cousin. Although there was a heavy disease load in the family, none of the female individuals

was affected with ALS. After mutational analysis in several female members of the family, three mutation positive females (ALS-914, ALS-991 and ALS-992) were observed. The two female cousins of the index patient who carry the p.Gly348Cys are currently 55 and 53 years old and are still free of disease.

SOD1-His71Tyr



TARDBP-Gly348Cys

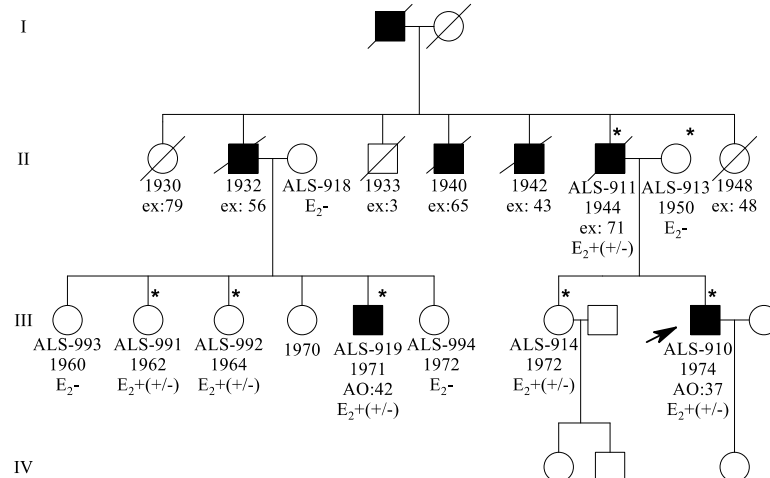


Figure 5.24. Families with low or incomplete penetrance. E₁ (*SOD1*, p.(His71Tyr)), E₂ (*TARDBP*, p.(Gly348Cys)), AO: age of onset, ex: age at death, MND: motor neuron disease.

To assess the differences in the genetic backgrounds of the mutation carriers in these two families, variant comparisons among individuals are carried out using the VarSifter

Software. To exclude the presence of any other known ALS-causing mutation in the additional affected family members, the vcf files of all affected cases were once more individually evaluated. The patient ALS707 from the SOD1-His71Tyr family (age of onset: 67), was found to carry a heterozygous missense variation in the *SETX* gene that is known to cause juvenile onset ALS (Figure 5.24) (Chen *et al.*, 2004). This variation (rs143661911) did not overlap with the patient's phenotype and was observed in two of the unaffected family members. In addition, compound heterozygosity in the non-consanguineous index (ALS226) was questioned and no variation in known neurological disease genes was detected.

After excluding known genes in all affected family members in an unbiased approach, several filters are applied using the VarSifter Software in the SOD1-His71Tyr and TARDBP-Gly348Cys families, to narrow down possible modifying variants in secondary genes. The filters applied, based on incomplete penetrance among family members, are stated in Table 5.7. Disease modifiers may be SNPs that are common in the population, thus the MAF threshold was not applied while filtering. The candidate variants identified in different filtration approaches have to be assessed. In the family with the TARDBP-p.Gly348Cys change, only two exonic nonsynonymous X-linked variants were identified in the asymptomatic carriers that are not present in the affected individuals. These variants are in *SHROOM2* (NM_001649, c.G4821C, p.Leu1607Phe) and *CIGALT1C1* (NM_001011551, c.T393A, p.Asp131Glu) genes (Table 5.7).

In an alternative approach, homozygosity mapping was performed in the SOD1-His71Tyr family, to search for any protective loci among the asymptomatic carriers in the third generation (ALS236, ALS704 and ALS708). The exonic SNPs filtered in the asymptomatic carriers were compared with ROHs identified in homozygosity mapping, however no common ROH was observed that harbors a homozygous candidate protective variant.

Table 5.7. VarSifter filters in families with incomplete penetrance.

Filtering Purpose	Filter Criteria	Family	# of variants
Additional heterozygote risk variant in affected individuals	ALS226 ALS707 het ALS704 ALS236 ALS708 homref	SOD1-His71Tyr	77
Protective variant in asymptomatic carriers	ALS236 ALS704 ALS708 het ALS226 ALS707 homref	SOD1-His71Tyr	113
Additional homozygote risk variant in affected individuals	ALS707hom ALS704 ALS706 ALS236 ALS703 ALS708 nothom	SOD1-His71Tyr	18
Heterozygous protective factor in asymptomatic carriers	ALS914 ALS991 ALS992 het ALS910 ALS911 ALS919 reference	TARDBP-Gly348Cys	38
X-linked protective variant in asymptomatic females	ALS914 ALS991 ALS992 het or homvar ALS910 ALS911 ALS919 reference	TARDBP-Gly348Cys	2
het: heterozygous, homref: homozygous reference, homvar: homozygous variant, nothom: not homozygous			

5.3.4.2. Oligogenic Inheritance and Modifiers in *C9orf72*-positive Cases. Exome sequencing was performed in 33 *C9orf72* expansion carriers to identify additional ALS-causing mutations and to question additional disease modifiers. A variant (rs143144050) in the *VABP* gene, previously shown as a variant of unknown significance in a PD patient, was identified in a female patient and her yet asymptomatic daughter (Kun-rodrigues *et al.*, 2015). In addition, three additional SNPs in genes associated with ALS were also detected in four patients. A sporadic female patient was found to carry the rs75087725 SNP in the *C21orf2* gene, reported as a risk factor in a large GWAS study conducted in the framework of Project MinE (van Rheenen *et al.*, 2016). A splice site variant (NM_002087.3:c.835+7G>A, rs72824736) with a high frequency in the population (ExAC MAF: 0.04) was observed in the *GRN* gene of a male patient. Two affected brothers carried the rs75795663 change in their *UNC13A* genes, which was absent in two other affected family members. Screening of published modifiers of *C9orf72*-based disease revealed the rs1800435 in the *ALAD* gene, previously associated with prolonged survival (van Blitterswijk *et al.*, 2014b). This SNP was found in three out of four affected relatives in the large family also carrying the *UNC13A* SNP. All three represented different survival times.

5.4. Whole Genome Sequencing Analyses in Sporadic ALS Cases: Project MinE

So far NDAL has contributed to Project MinE with 625 Turkish ALS samples and 152 neurologically healthy controls. In this thesis, data from Project MinE was used to investigate a possible accumulation of rare or common variants in patients compared to controls. Further, search for mutations in rare ALS genes that would explain the disease in a portion of apparently sporadic cases was undertaken. In addition a subset of samples were used as a replication study cohort for the identification of *C21orf2*, *MOBP* and *SCFD1* genes as novel risk loci in ALS (van Rheenen *et al.*, 2016).

5.4.1. Variant Accumulation in sALS Patients

Based on the nature of sporadic ALS, the hypothesis that rare and/or common variants accumulated in patients, but not in controls might confer to increased disease risk. The number of homozygous variants detected in patients was significantly higher than controls for all five frequency thresholds (green p-values, Figure 5.25). The significance in homozygous variant accumulation in patients increased with higher threshold. In contrast, heterozygous variants tended to accumulate more in controls compared to patients (blue p-values, Figure 5.25). Increase in significance with higher frequency threshold was also observed in this group, in fact significance was lost for variants seen in less than 1% of samples (MAF<0.01). No difference was observed in allelic and het_or_hom categories.

5.4.2. Pathogenic Variations in sALS Cases

Six hundred and twenty-five Turkish ALS cases and 152 neurologically healthy controls whole genome sequenced within Project MinE are screened for mutations in known ALS genes using bioinformatics tools. Data analysis revealed six *SOD1*, one *TARDBP* and two *FUS* mutations in sALS patients (Table 5.2). Variants with low or no reported MAF are identified in rare autosomal dominant and recessive familial ALS genes, like *OPTN*, *TBK1*,

VCP, *ANG* and *Erb-B2 Receptor Tyrosine Kinase 4 (ERBB4)*, together with several variants in genes that were previously associated with ALS (Table 5.8).

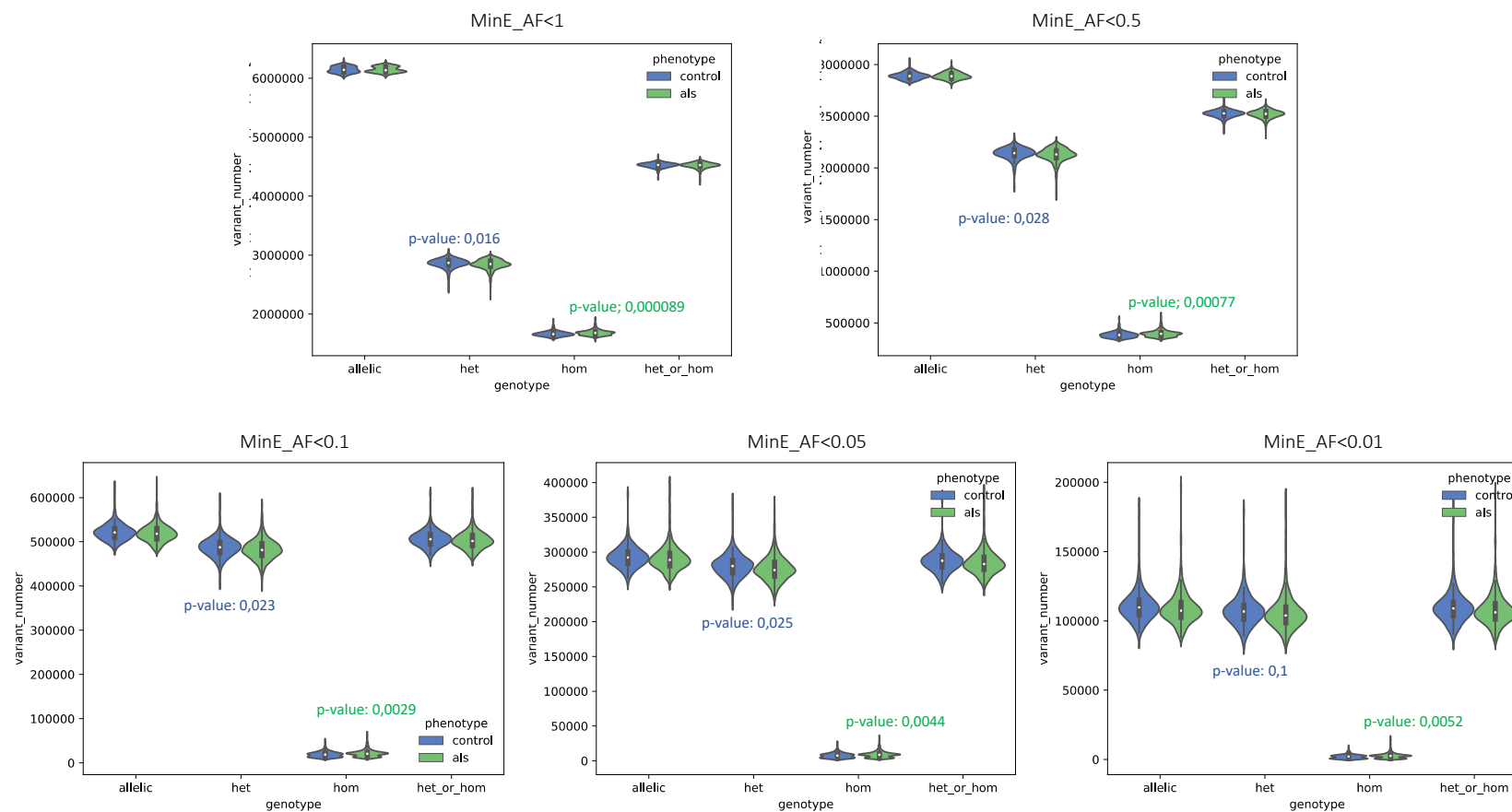


Figure 5.25. Number of variants per individuals in controls and patients subjected to Project MinE. Data is plotted for five allele frequency (AF) thresholds of 1, 0.5, 0.1, 0.05 and 0.01. p-values for higher significance in controls and patients are depicted in the same color with the category.

Table 5.8. Rare ALS causing/associated variants identified via whole genome sequencing in the framework of Project MinE.

Chromosome location	Gene	DNA change	Protein change	Zygosity	rsID	# of Patients	# of Controls
chr12:64875731	TBK1	c.922C>T	p.Arg308Ter	het	-	1	0
chr10:13151191	OPTN	c.76del	p.His26Thrfs*19	hom	rs766608795	1	0
chr10:13167493	OPTN	c.1078_1079del	p.Lys360Valfs*18	hom	-	2	0
chr10:13168013	OPTN	c.1217del	p.Thr406Lysfs*5	hom	-	1	0
chr9:35065349	VCP	c.475C>T	p.Arg159Cys	het	rs387906789	1	0
chr9:35065361	VCP	c.463C>T	p.Arg155Cys	het	rs121909330	1	0
chr9:35067907	VCP	c.283C>T	p.Arg95Cys	het	rs121909332	1	0
chr9:35068336	VCP	c.41C>T	p.Thr14Ile	het	-	0	1
chr2:212295795	ERBB4	c.2518G>A	p.Val840Ile	het	rs369248674	1	0
chr2:212989553	ERBB4	c.157A>G	p.Tyr53Cys	het	rs756650586	1	0
chr2:212589887	ERBB4	c.654G>A	p.Gly219Ser	het	rs757597004	1	0
chr2:212812268	ERBB4	c.307G>A	p.Arg103His	het	rs754487821	1	0
chr22:24109646	CHCHD10	c.176C>T	p.Ser59Leu	het	-	1	0
chr14:21161931	ANG	c.208A>G	p.Ile70Val	het	rs121909541	2	1
chr14:21162046	ANG	c.324dup	p.Gly109Trpfs*24	het	-	1	0
chr9:135163699	SETX	c.6248G>T	p.Arg2083Ile	het	rs751252138	1	0
chr9:135172398	SETX	c.5825T>C	p.Ile1942Thr	het	rs773379832	0	1
chr4:170428877	NEK1	c.1816G>T	p.Glu606Ter	het	-	1	0
chr5:179251221	SQSTM1	c.319G>A	p.Gly107Arg	het	rs781478225	1	0
chr17:4851653	PFN1	c.37G>A	p.Ala13Thr	hom	rs763837842	2	0
chr12:109278828	DAO	c.46G>A	p.Ala16Thr	het	rs778735604	1	0
chr12:109281243	DAO	c.212C>T	p.Thr71Ile	het	rs138277420	1	0

Table 5.8. Rare ALS causing/associated variants identified via whole genome sequencing in the framework of Project MinE (cont.).

Chromosome location	Gene	DNA change	Protein change	Zygosity	rsID	# of Patients	# of Controls
chr12:109286795	DAO	c.490G>T	p.Val164Leu	het	-	0	1
chr12:109294229	DAO	c.962G>A	p.Gly321Glu	het	-	1	0
chr2:74593112	DCTN1	c.2794C>T	p.Arg932Cys	het	rs373818927	1	1
chr2:74593597	DCTN1	c.2617G>A	p.Ala873Thr	het	rs764492372	0	1
chr2:74594488	DCTN1	c.2244C>G	p.Asp748Glu	het	rs751069902	1	0
chr2:74594495	DCTN1	c.2237T>C	p.Leu746Pro	het	-	1	0
chr2:74605312	DCTN1	c.94C>T	p.Arg32Cys	het	rs751177222	1	0
chr8:28017873	ELP3	c.1385G>A	p.Arg462His	hom	rs190129217	1	0
chr6:110036336	FIG4	c.122T>C	p.Ile41Thr	het	rs121908287	1	0
chr6:110056402	FIG4	c.547C>T	p.Arg183Ter	het	rs121908288	1	0
chr6:110086229	FIG4	c.1448G>A	p.Arg483Gln	het	rs749233172	1	0
chr6:110106223	FIG4	c.1940A>G	p.Tyr647Cys	het	rs150301327	0	1
chr6:110106873	FIG4	c.366-1G>C	-	het	rs545579302	1	0
chr12:54674858	HNRNPA1	c.58_59del	p.Leu20Lysfs*7	het	-	1	0
chr17:34171734	TAF15	c.423+2T>C	-	het	rs144917137	1	0
chr19:7605129	PNPLA6	c.532G>T	p.Gly178Cys	het	-	1	0
chr19:7620584	PNPLA6	c.2914G>A	p.Gly972Arg	het	rs768107851	1	0
chr19:7626175	PNPLA6	c.365del	p.Ala123Glnfs*67	het	-	0	1
chr7:94935670	PON1	c.707A>G	p.Tyr236Cys	het	rs755475189	2	0
chr7:95024007	PON3	c.94C>T	p.Arg32Ter	het	rs147006695	5	0
chr12:49689173	PRPH	c.190C>T	p.Arg64Ter	het	-	2	0

Table 5.8. Rare ALS causing/associated variants identified via whole genome sequencing in the framework of Project MinE (cont.).

Chromosome location	Gene	DNA change	Protein change	Zygosity	rsID	# of Patients	# of Controls
chr15:78894232	CHRNA3	c.752C>G	p.Pro251Arg	het	-	1	0
chr15:78894258	CHRNA3	c.725del	p.Leu242Cysfs*32	het	-	2	1
chr15:78894275	CHRNA3	c.708_709insG	p.Ile237Aspfs*35	het	-	2	1
chr15:78910978	CHRNA3	c.247_248insG	p.Thr83Serfs*11	het	-	0	1
chr15:78911138	CHRNA3	c.1A>G	p.Met1?	het	rs745905590	3	0
chr20:61981766	CHRNA4	c.997C>T	p.Arg333Cys	het	rs761631713	0	1
chr20:61981784	CHRNA4	c.979G>A	p.Val327Met	het	rs201841018	1	0
chr20:61982107	CHRNA4	c.656A>C	p.Asn219Thr	het	rs201645533	1	0
chr20:61982321	CHRNA4	c.442C>T	p.Arg148Trp	het	rs121912243	1	0
chr15:78921343	CHRNA4	c.1304C>T	p.Ala435Val	het	rs56317523	1	1
chr15:78922149	CHRNA4	c.498C>G	p.Asn166Lys	het	rs148540431	1	1
chr15:78927869	CHRNA4	c.116G>T	p.Arg39Leu	het	-	1	0
chr20:60734000	SS18L1	c.155+2T>C	-	het	rs771943598	1	0

6. DISCUSSION

ALS is characterized by selective death of upper and lower motor neurons that leads to paralysis of skeletal muscles; patients are generally lost with respiratory failure. Nevertheless, the mechanism behind catastrophic events resulting in motor neuron death is not fully understood. The ameliorating technological advances in sequencing in the last decade enabled genetic research on large cohorts and identification of several novel genes associated with ALS. These genes are expected to explain why motor neurons go under cellular stress through distinct but converging/interconnected mechanisms. In the 21st century, genetics and molecular biology became inseparable partners of medicine (Hood and Rowen, 2013). Accumulating knowledge on genetics of rare and common diseases affecting the human population is thriving cell-, tissue- and animal-based research which is inevitable for developing novel and robust therapeutic targets.

6.1. Challenging Nature of ALS and Common Disputes

ALS is the most common neurodegenerative disease with restricted therapy options and no treatment. Diagnosis of ALS is based on exclusion of other syndromes with laboratory testing, electrodiagnostic, imaging and restriction of symptoms to specific body regions. There are two diagnostic criteria used by the clinicians for diagnosis of ALS. El Escorial criteria (EEC), harbors five detailed categories for diagnosis, which are, clinically definite, clinically probable, clinically probable with laboratory support, clinically possible and clinically suspected ALS (Brooks and Miller, 2000). Awaji-shima criteria is another diagnostic method with more simplified classifications of clinically definite, probable and possible (de Carvalho *et al.*, 2008). Although more commonly used, EEC is condemned for its restrictive nature. It is extremely challenging to identify the mild symptoms at the initial stage of a progressive disease like ALS, thus the diagnostic delay is usually more than a year (Salameh *et al.*, 2015). This diagnostic odyssey involves clinicians with different backgrounds ranging from

orthopedics to neurosurgery and neurology and requires differential diagnosis of other neurological diseases that mimic ALS and physical syndromes like post-polio syndrome. A comprehensive follow-up study including almost 500 patients with suspected ALS, resulted in re-diagnosis of 5% of patients with a non-MND, and only 38% of the rest of the MND cases had received a confirmed ALS diagnosis (Pupillo *et al.*, 2017). Moreover, the evidence regarding to cognitive dysfunction in ALS, recognized with the identification of *C9orf72* repeat expansion, has introduced a new phenotypic complexity and promoted the importance of detailed clinical examination and family history in the diagnostic process.

The next important question for the researcher, the patient and family members, after the diagnosis of the patient, is that if the disease is hereditary or not? As in most of the neurodegenerative diseases, only a small percentage of ALS is considered to be familial, however, literature nowadays appreciates the limitation and bias in classification of ALS as familial and sporadic. An example from Ireland draws attention to the increasing percentage of familial ALS in recent years and suggests that although *C9orf72* has made a great impact in explaining the disease cause in a considerable number of familial cases, the increase in the fALS percentage is actually correlates with collecting detailed family history in the clinic (Hardiman *et al.*, 2017). Accompanying clinical features and related diseases like behavioral disorders or Parkinson's disease may not be recognized as positive disease history in the family and should be paid attention (Turner *et al.*, 2017). Low penetrance reported in several ALS genes is also another issue that hides an apparent segregation through generations and leads to a false classification. Talbot suggests that variants with low penetrance might follow Mendelian segregation and act as disease-causing mutations or might only increase disease risk depending on the genetic background of the individual, thus fALS and sALS cannot be easily divided and it is hard to define the heritability of the disease, while counseling an at-risk individual (Talbot, 2011).

Genetic testing in sALS patients is a matter of debate, since it can reveal a penetrant variation (e.g. the pathogenic expansion in the *C9orf72* gene or a *de novo* variation in the *FUS*

gene), which would help the diagnosis of the patient, however, lead to changes in the lives of the relatives. Furthermore, it is not clear how to counsel a sALS patient carrying a variation with low penetrance in the population or in the same family, since his/her disease might not be solely explained by that variant and would not be informative for his/her children. In addition, although the heritability is not properly understood, studies suggest that relatives of sALS patients carry the same life time risk of getting the disease (1:400) as the normal population (Talbot, 2011). Altogether, it is proposed that sALS patients should not be offered genetic testing, but Turner and colleagues suggest that this opinion might be reevaluated taking factors into consideration like false classification due to incomplete family history, advancements in preimplantation genetics and initiatives in clinical trials for personalized/gene-based therapies (Turner *et al.*, 2017).

6.2. Clinical Presentation of ALS in Turkey

For almost two decades, well-established patient registers operating in European countries are gathering organized patient data to understand the epidemiology of ALS (Hardiman *et al.*, 2017). These registers work countrywide and are unbiased in terms of origin, socio-economic status and the disease stage of the patient in contrast to a local clinic. The incidence reported for ALS world-wide is accepted to be around 2-3/100.000, yet this number is argued to be misleading in the absence of long-running patient registers, that may be more efficient in recognizing family history and hidden symptoms like FTD (Hardiman *et al.*, 2017). In contrast to the higher incidences expected in Europe, a recent study on ALS in Asia proposed lower incidences especially for China and India (Shahrizaila *et al.*, 2016). One of the reasons for different incidences among populations is thought to be the diversity of the gene pool in ancestrally mixed populations, which lowers the risk of disease-associated variant accumulation and thus disease incidence (Zaldivar *et al.*, 2009). Unfortunately, Turkey does not have an ALS patient registry, hence reliable information regarding the incidence, prevalence and survival rates of patients are not available. NDAL, being the only reference center for the molecular analysis of ALS in Turkey, recruits patients from across the country

and gathers as much information as possible to learn the clinical and molecular basis of a complex neurodegenerative disease in an admixed population living in the crossroads of many civilizations.

This thesis offers the clinical characterization of 722 Turkish probands diagnosed with ALS. Familial ALS accounts for 22% of the cases (159/722) in the cohort under study compared to 563 cases classified as sporadic; this makes up 78% of all cases. The picture does not fit with the numbers accepted for familial ALS in North American and European populations, however, it can be explained by several population-specific factors like (i) large kindreds with higher number of offspring resulting in less isolated cases, or (ii) close relationships within families leading to more accurate family history. Another unique aspect in Turkey is the high proportion of close consanguinity in the population, exceeding almost 50% of all marriages in the eastern parts of the country. Consanguinity in the ALS cohort under study is calculated as 25% in fALS and 17% in sALS, suggesting an even higher percentage for Mendelian inheritance in yet unexplained cases that are classified originally as sALS due to singlet patients in the family. Thus, familial disease or, as we actually want to refer as, ALS harboring a genetic component, seems to be exceeding 22% among Turkish cases. On a related note, in a study specific for the Sardinian population, cases with a family history of ALS and/or FTD were calculated as 26.7%, which shows that homogeneity within populations also increases the density of disease-specific alleles and thus, results in elevated frequencies of fALS (Borghero *et al.*, 2014). These findings further support the inefficient clinical classification of the genetically undiagnosed patients, however, to the best of our knowledge, this is the only approach available at the moment. So, the statistical evidence and results in this thesis are calculated and discussed based on this approach.

Sixty-six per cent of all cases in the Turkish cohort investigated in this thesis had an age of disease onset over 45 years and the mean age of onset is 48 for fALS and 52 for sALS patients. This number for sporadic disease is in correlation with other European and North American populations, in addition to South Korea (58) and China (52) (Ghasemi and Brown,

2018; Shahrizaila *et al.*, 2016). The younger age of onset of fALS cases in our cohort is most probably due to the high rate of consanguinity resulting in atypical juvenile phenotypes.

In 71% of the cases spinal symptoms were detected as the initial clinical features and 18% reported to suffer from bulbar symptoms, which usually is also true for one third of ALS patients worldwide. The gender ratio (male to female) of the Turkish population is 1.5, which decreases to 1.1 in familial cases. The overall high gender bias for men is a known fact about ALS worldwide. There are several factors that may be acting on this finding, which may include environmental factors like trauma, working injuries, heavy sports or chemical exposure in field, which are suggested to be playing a role in sALS pathogenesis (Ingre *et al.*, 2015). These factors are expected to affect men more than women, especially in the earlier societies and in the developing countries; or simply, better records might be available for working-men. However, an interesting scientific study suggests that this bias may originate from uneven number of age-matched individuals in population-based studies. Manjaly and colleagues observed that incidence of ALS is higher in post-menstrual women and women tend to have a higher risk of ALS in their fifth or sixth decade of life, which leads to a gender bias for men when the ratios are calculated for all age ranges (Manjaly *et al.*, 2010). In accordance with this observation, we have noticed that female patients make up 34% of cases with an AO lower than 50, in contrast to 44% with an AO equal to or higher than 50. Thus, the male:female ratio in the Turkish population decreases to 1.2 in older patients. More data from patient registers may strengthen this observation and offer modifying factors to be investigated, that postpone the onset of disease in female individuals.

6.3. Impact of Common ALS Genes on ALS in Turkey

The *C9orf72* hexanucleotide repeat expansion is detected in 16% of familial and 3% of sporadic cases in all Turkish ALS cases at NDAL. *SOD1*, *TARDBP* and *FUS* are observed in 15%, 3% and 2% of familial ALS investigated in this thesis, respectively.

WES and WGS performed in apparently sporadic/isolated cases identified mutations in common ALS genes that explained approximately 3% of cases referred as sporadic (*SOD1*: 2%, *TARDBP*: 0.3% and *FUS*: 0.8%). These results support the evidence of incomplete penetrance and *de novo* occurrence of mutations in these genes, leading to misclassification of patients. We believe that the coarse classification of familial and sporadic disease is still needed for strategic genetic screening, however, should be handled with care during genetic counseling.

In the cohort under study, the total percentage of common ALS genes implicated in the familial form of the disease is calculated as 33%. This number increases to 39% when we include familial cases carrying mutations in rare ALS genes solved by WES. According to this picture, higher proportion of fALS cases remains to be solved, which point towards locus heterogeneity in the population. Turkey, geographically located at the intersection of many civilizations has a heterogeneous ethnic and genetic background. This genetic complexity in the population leads to dilution of common mutations like *C9orf72* hexanucleotide repeat expansion or *SOD1* mutations. In this sense, the frequencies observed in Turkey are concordant with the common notion of decreasing north-south gradient for *C9orf72* and *SOD1* (Andersen, 2006; Smith *et al.*, 2013).

6.3.1. The Dynamic C9orf72 Expansion Mutation and Promoter Hypermethylation

The hexanucleotide repeat expansion in the *C9orf72* gene is the most common genetic cause of ALS worldwide, with the exception of Japan (Ogaki *et al.*, 2012). Although less in frequency, compared with Northern European countries, in which the frequency of this expansion explains almost 50% of familial ALS, *C9orf72* expansion is found to be the most abundant mutation both in familial and sporadic ALS also in the Turkish cohort.

Mean age of disease onset among the expansion carriers was 54, seven years later than fALS mean AO and very similar to the mean AO of sALS patients under study, indicating that

C9orf72 expansion mutations result in a typical ALS prognosis. Sixty per cent of the probands had spinal/limb-onset disease, whereas 27% had bulbar symptoms first emerging. A higher frequency of bulbar-onset ALS is observed among *C9orf72*-positive cases, compared to 18% in all ALS cases. The higher tendency of bulbar disease-onset for *C9orf72*-based ALS suggested previously in the literature, seems to be true for the Turkish cohort under study.

Dementia was reported in 14 patients (20%) with the *C9orf72* expansion, in two of whom, this was the only symptom. *C9orf72* hexanucleotide repeat expansions, being the most frequent cause of both ALS and FTD, result in high inter and intra-familial phenotypic variability. In this thesis, several families were detected with members presenting either with ALS, ALS-FTD or solely FTD symptoms. Intra-individual variability in the GGGGCC repeat size in different tissues was previously shown in ALS and/or FTD patients, however no difference was observed in patients with a cognitive deficit (Nordin *et al.*, 2015). The phenotypic variability among individuals may arise from the presence of modifying SNPs, epigenetic mechanisms or environmental factors and life style.

In this thesis, three ALS patients were identified with 30-35 GGGGCC repeats in one of their alleles. (i) The dominant inheritance of the allele with 30 repeats from an unaffected father to his son was shown in one case. (ii) In addition, the affected mother of the fALS case with 35 repeats had only two repeats, showing that the disease does not segregate with the GGGGCC₍₃₅₎ repeats. (iii) DNA from family members was not available in the third family. These findings point out to non-pathogenic nature of short (<30-35) GGGGCC expansions in the *C9orf72* in our cohort. So far, no other genetic mutation is identified as the cause of the disease in these patients, which are considered as negative for the *C9orf72* expansion mutation.

One study performed in 5000 individuals (ALS, FTD, ALS-FTD and controls) also demonstrates that there is no correlation between the long-normal-range GGGGCC repeats and disease risk or manifestation (Rutherford *et al.*, 2012). However, genomic instability is

proved for the GGGGCC repeat region of the *C9orf72* gene with a phylogenetic analysis, that was able to separate 139 haplotypes into two ancestral branches, only one of them harboring the expansion (Smith *et al.*, 2013). Individuals rising from this branch, carrying the risk haplotype, happen to have longer non-expanded repeats (most prevalent 8, 26 being the highest), compared to the non-risk haplotype carriers. How the genomic instability works in the repeat region is not known, but secondary structures that will be formed from the long GC-rich sequence may affect the replication machinery or a faulty repair process may act in the emergence of an expanded repeat. Apart from these findings, there is no evidence for a correlation between longer intermediate alleles (>30) of *C9orf72* and ALS, as in the case of ATXN2. The frequency of the intermediate expansion in the population or the role as a risk factor for sporadic ALS pathogenesis can be questioned. Another question about the non-coding hexanucleotide repeat in the *C9orf72* gene is if there is a sharp, one or two repeats, threshold between a patient and a healthy individual, as in the case of trinucleotide repeat disorders with exonic mutations. Or does the *C9orf72* pre-mutation suddenly extend to over 100-1000 repeats? A detailed study in one family favored the sudden extension of the pre-mutation of 70 repeats (Xi *et al.*, 2015). The healthy male with 70 repeats had fathered five children three of whom inherited an elongated expansion ranging from 600 to 3000, that suggested a possible failure in spermatogenesis, resulting in the sudden expansion of the repeat region in one generation. To be able to understand the nature of this dynamic repeat, more studies should be conducted in large families, which will allow us to look for possible genetic factors that lead to failure in replication or repair mechanisms. In addition, the missing heritability of *C9orf72* can be explained by detecting the exact repeat sizes of the parents of sporadic patients who are expected to carry the pre-mutation.

Our results on the suspected promoter hypermethylation in *C9orf72* expansion carriers confirmed significant hypermethylation in 26 CpG sites located 5' of the promoter region in the Turkish cohort. High methylation levels were not detected in any of the 31 control samples, whereas 13 out of 52 *C9orf72* expanded cases had high methylation levels. In accordance with the previous studies, promoter hypermethylation did not show correlation

with ages of onset of the patients (Russ *et al.*, 2015; Xi *et al.*, 2013). Reduction in 50% of the C9orf72 expression reported in expansion carriers suggests a LOF mechanism caused by haploinsufficiency. Promoter hypermethylation could be the mechanism behind this hypothesis, based on the finding of Xi *et al.*, that hypermethylation of the promoter region is correlated with shorter disease duration among ALS cases (Xi *et al.*, 2013). Nonetheless, in another study, increased promoter methylation was found to be concordant with longer disease duration in FTD but not ALS, suggesting for a GOF mechanism in FTD cases, where decreased transcription of mRNA containing the expansion leads to less toxicity (Russ *et al.*, 2015). The discordancy between these findings may be related with different mechanisms implicated in ALS and FTD or may be due to survival information of the different cohorts in two different studies. Unfortunately, the survival information in our patients is not sufficient to conduct such comparison, thus these observations could not be tested in our samples. Methylation-specific restriction enzyme digestion was performed for a subset of expansion carriers; however, results were not in agreement with the bisulfite sequencing assay. Thus, has to be further investigated.

6.3.2. Allelic Heterogeneity in SOD1 and Genotype-Phenotype Correlations

SOD1 mutations are the second most common genetic cause in the Turkish ALS cohort investigated in this thesis. The allelic heterogeneity accepted for *SOD1* was observed also in our cohort, where 15 different mutations explained the disease-cause in 34 families (Table 5.2). Evidence for the reduced penetrance of *SOD1* mutations was seen in the Turkish population in several situations, including (i) the mutations detected in sporadic patients through Project MinE, (ii) the families with asymptomatic carriers, (iii) dual inheritance patterns observed for the Asn86Ser, Asp90Ala, Leu117Val and Glu133Lys mutations.

There is an ample number of Turkish people with a Balkan origin migrated to Turkey many generations ago. This fact rationalizes the predominance of the dominant SOD1-Leu144Phe variation in our cohort. The age of onset average for Leu144Phe mutation carriers

is 52 years without any gender bias and all had limb-onset disease. This variation results in a classical form of ALS. In our cohort the mutation appears to be highly penetrant, observed in large families consisting of several branches. Only one exception is patient ALS-281, who is a sporadic male patient, solved through Project MinE; however, this result may also arise from inadequate information about the deceased parents of the index patient (Table 5.2).

The recessive SOD1-Asp90Ala variation is the second most common mutation in our cohort. Up to date, no affected individuals in the Turkish population have been detected carrying the heterozygous variant. The presence of the Scandinavian founder haplotype in Turkish recessive Asp90Ala cases was previously shown in our laboratory (Özoğuz *et al.*, 2015). Except one patient with mixed site of onset, limb-onset disease predominates in Asp90Ala patients and the disease progression is very slow. The average age of onset of the recessive Asp90Ala carriers is 10 years earlier than the SOD1-Leu144Phe mutation carriers and is in accordance with the stereotyped Scandinavian phenotype.

Apart from the highly penetrant and frequent Leu144Phe and Asp90Ala mutations, other SOD1 mutations identified in our cohort are present in relatively small families with a few numbers of affected children, usually in the same generation. This may be the result of incomplete penetrance enabling the disappearance and reappearance of the disease phenotype in small branches of large families. Examples of SOD1 mutations with evidence of reduced penetrance in our cohort are the p.Glu22Leu, p.Glu40Gly and the novel p.His71Tyr (identified at NDAL) variations with asymptomatic mutation carriers in the family. Speculatively, if we were able to obtain detailed generation-wide pedigree information, we might see the same phenotype in their distant relatives.

The SOD1-Ala4Val mutation is known to be a very severe mutation with 1.2 years of average survival time (Rosen *et al.*, 1994). This mutation, the most common mutation in North Americans and is not present in our Turkish ALS cohort, however a female patient (ALS-1398) from a previously reported large Turkish Cypriot pedigree, was shown to carry the rare

heterozygous Ala4Thr variation reported in her extended family (Aksoy *et al.*, 2003). Different ages of onset are reported in the literature for this variation ranging from 25 to 40. Our patient had an early disease-onset of 25, with bulbar symptoms, known to be associated with faster disease progression. The average survival of patients in her family is reported to be 14 months, yet the average age of onset is 46 years (Aksoy *et al.*, 2003). The alanine to threonine change in the fourth amino acid also leads to a severe phenotype, comparable with Ala4Val, thus the position of the change rather than the amino acid change might be of importance in this case.

SOD1 is a ubiquitously expressed protein that acts as a superoxide radical scavenger in the cell, however its pathogenicity in ALS is explained by the misfolding of the product from the mutant allele, that accumulates within the cell in aggregates. Unfortunately, there is not enough evidence in the literature to comment on the mechanism behind the phenotypic heterogeneity of different SOD1 mutations as well as the acting mechanisms of mutations that show both AD and AR inheritance. Although most evidence supports the gain-of-function mechanism behind SOD1 mutations, loss-of-function of the mutant allele may still have a role in the presentation of the disease. Reduction in overall activity of mutant SOD1 compared to wild-type enzyme has been shown in blood and fibroblast samples, but the two specific mutations, Asp90Ala and the heterozygous Leu117Val, had only slight reductions in enzymatic activity, which might explain the milder phenotype in patients carrying these mutations and the low penetrance observed in parents of homozygous mutation carriers (Saccon *et al.*, 2013). On the other hand, the homozygous Leu117Val mutation was reported to have more reduction in enzymatic activity than the heterozygous variant, which is also concordant with the early age of onset and the fast disease progression of the female patient reported in this thesis (Synofzik *et al.*, 2012). Although the function of the mutant protein is not completely lost, the activity may be reduced by aggregation, thus different SOD1 mutations with different aggregation propensities may have variable enzymatic activity, which may act on the phenotypic representation of the disease. On top of all, reduction in activity of

such an abundant and important enzyme might result in accumulation of toxic radicals within the cell, leading to susceptibility especially in large and demanding cells like motor neurons.

6.3.3. *de novo* FUS Mutations in the Turkish Cohort

Juvenile ALS (AO<25) was observed in 41 isolated/sporadic cases in our cohort, 17 of them having consanguineous parents. Mutational analyses in these patients revealed four cases with non-consanguineous parents carrying two distinct mutations (p.Pro525Leu and p.Tyr526Cys) in the *FUS* gene. Genetic testing in parents of three cases revealed the *de novo* occurrence of the mutations in all three patients (ALS-549, ALS-1610, ALS-1423, Table 5.2). Juvenile ALS in Turkey most frequently occurs due to consanguineous marriages and has a rather slow disease progression compared to classical ALS. However, these *de novo* cases with ages of onset ranging from 12 to 17 have an aggressive disease progression, resulting in retirement of the children from all daily activities and causing severe bulbar symptoms in addition to initial limb-onset disease, eventually leading to death almost within a year. In addition to four juvenile cases another apparently sporadic male patient, with an age of onset of 39, was found to carry the heterozygous FUS-p.Pro521Leu mutation. His asymptomatic parents are currently not available to us, yet the short survival of the patient increases the chance of the mutation being *de novo*.

ALS-causing *FUS* gene mutations are restricted to the C-terminal Nuclear Localization Signal (NLS) domain of the protein and explain almost 4% of familial disease. In addition, *de novo* FUS mutations located in this region are also repeatedly reported, explaining a considerable percentage of juvenile cases in populations where consanguinity is not common (Hübers *et al.*, 2015; Leblond *et al.*, 2016). All cases in the literature with *de novo* FUS mutations suffer from a severe early-onset disease with fast progression and this observation fits the reduced reproducibility generally suspected for *de novo* mutations, that do not allow the patients to reach the reproductive age (Therrien *et al.*, 2016).

Sporadic ALS is expected to result from the combinatorial effect of genetic, epigenetic and environmental factors, however a good proportion might be explained also by *de novo* mutations. Effect of *de novo* mutations in ALS is not yet fully investigated, due to high cost of trio-sequencing and inefficient calculation methods for *de novo* mutations in the absence of the data of the parents. Based on our findings, we are now confident to say that *de novo* FUS mutations should be considered as the initial step in isolated juvenile cases with a fast disease progression and asymptomatic parents. In addition, further *de novo* events should be carefully investigated in isolated patients during family/trio WES analyses. Even in some familial cases, where there is a chance of phenotypic difference among family members, *de novo* mutations should not be ignored.

6.3.4. Two UBQLN2 Variations with Unknown Significance (VUS)

In our laboratory, mutations located in the C-terminal hotspot of the UBQLN2 protein were shown to result in ALS pathogenesis in Turkish ALS patients (Özoğuz *et al.*, 2015). UBQLN2 plays role in targeting misfolded proteins to the proteasome by binding them to the UBA domain located in the C-terminal of the protein. The p.Thr134Ile and p.Ser340Ile variants identified in the scope of this thesis are outside the hotspot of the gene. Moreover, additional mutations implicated in ALS (e.g. *VCP* and *C9orf72*) were shown to be present in these individuals and both variations were shown in independent healthy individuals (Akçimen, 2017). Thus, Thr134Ile and Ser340Ile variations should be considered as not pathogenic rather than VUS. Although *UBQLN2* mutations cause ALS in an X-linked dominant manner, it is possible that the second allele in females may have a compensatory effect for its role. However, this hypothesis is rejected by the fact that there are also male cases carrying the hemizygous variant without showing the disease.

6.4. Whole Exome Sequencing Analyses

In this thesis, WES is performed in 39 Turkish families consisting of 54 individuals (42 affected individuals and 12 healthy family members) with evident motor neuron symptoms in patients to identify disease-causing mutations. Consanguinity for the cohort under study is calculated as 25% for familial and 17% for sporadic ALS cases. Due to this high consanguinity in Turkey, juvenile cases with homozygous mutations are very common. These cases often present with additional (novel) clinical features that are not common in ALS and they show slower disease progression. Making a firm diagnosis in such cases is very difficult for the clinicians, since the boundaries between diseases get blurred with overlapping symptoms. Whole exome sequencing allows to investigate these cases in an unbiased manner according to the inheritance pattern and to focus on the additional clinical features rather than the initial diagnosis of the patient.

Our subjects in WES can be divided into three groups; the most crowded group is the familial and isolated ALS and HSP cases with suspected autosomal recessive inheritance. In addition, we have performed WES in large ALS families where the autosomal dominant inheritance is definite and the common ALS genes were excluded. Finally, as the third group a few subjects with an inconclusive inheritance pattern or atypical/interesting symptoms were investigated with WES.

Results showed a 41% (16/39) discovery rate for WES in the overall cohort, which is in accordance with the literature (Trujillano *et al.*, 2017). Confirmation of the pathogenicity of the identified candidate variations in four families described earlier (Family 14, 18, 19 and 20) is expected to increase this percentage. Causative mutations in two AD and 16 AR families plus two families with inconclusive inheritance remains to be solved.

6.4.1. Mutations in Rare ALS Genes

Two families with AD inheritance were shown to carry heterozygous pathogenic variants in the *KIF5A* and *TBK1* genes, both recently associated as rare genetic causes of familial ALS.

KIF5A: Autosomal dominant missense mutations in the N-terminal motor domains of *KIF5A* are associated with hereditary spastic paraplegia 10 (SPG10) or CMT2 (Crimella *et al.*, 2012). Phenotypic heterogeneity ranges from pure to complex forms of spastic paraplegia with additional features like axonal sensorimotor peripheral neuropathy or cognitive impairment. Recently, a hotspot region in the C-terminal of *KIF5A* was shown to be responsible for classical ALS with a severe phenotype (Figure 6.1) (Brenner *et al.*, 2018; Nicolas *et al.*, 2018). The mutations in the cargo binding domain of the protein possibly result in dysfunction of the intracellular organelle transport.

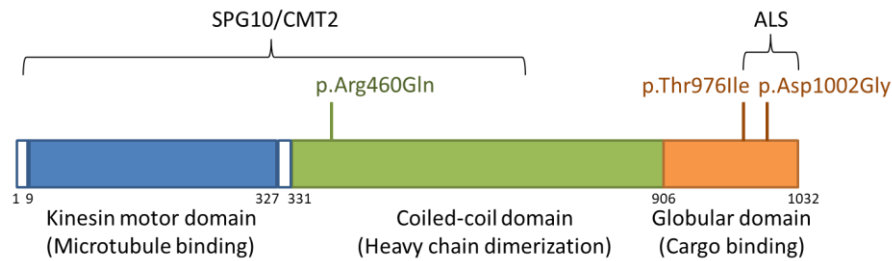


Figure 6.1. KIF5A protein domains and variations identified in the Turkish cohort.

The novel p.(Asp1002Gly) change identified in two step-brothers with a classical ALS phenotype and a strong family history of ALS, is also located in the C-terminal of the protein (Figure 6.1). Three additional *KIF5A* variants with unknown significance were identified within the scope of this thesis. The c.2927C>T, p.Thr976Ile variant (rs139801016, ExAC MAF: 0.0002), previously reported as VUS, is observed in an isolated ALS patient with consanguineous parents and in one of the juvenile-ALS patients with a *de novo* FUS mutation (ALS-1610, Table 5.2). Further questioning of *KIF5A* variations in our in-house exome

database revealed a rare heterozygous variation c.1379G>A, p.Arg460Gln in an ataxia patient with accompanying symptoms of spasticity and cognitive deficit. The Arg460Gln variant is located further N-terminal of the *KIF5A* gene (Figure 6.1). The previously reported p.Pro986Leu variant (ExAC MAF: 0,01), is present in a neurologically healthy individual in our cohort. This SNV is reported to be enriched among fALS cases and is suggested as a common risk allele with low penetrance (Brenner *et al.*, 2018; Nicolas *et al.*, 2018). Pathogenic *KIF5A* mutations associated with ALS are loss-of-function mutations that are either splice site mutations or non-synonymous missense variants, causing aberrant pre-mRNA splicing, and thus haploinsufficiency. Further functional analysis in these families to observe the consequences of the mutations on splicing might help to better understand the pathogenic nature of *KIF5A* variants.

TBK1: The gene encodes for a 729 amino acid long protein that colocalizes with optineurin and p62 in autophagosomes and phosphorylates them, enabling the autophagic turnover of several substances including cytosolic protein aggregates (Cirulli *et al.*, 2015). Haploinsufficiency of TBK1 caused by LOF mutations in the C-terminal of the protein was shown to be the cause of both familial ALS and FTD. The TBK1-Val479Glufs*4 truncation mutation identified in our cohort was previously shown to result in misbinding to optineurin and inability to phosphorylate downstream targets (Freischmidt *et al.*, 2015). Questioning this variant in the unaffected parents will reveal whether it is a *de novo* mutation or if its penetrance is low. Low penetrance is common in many rare genes causing ALS.

SIGMAR1: The transmembrane receptor protein, SIGMAR1, is involved in a broad range of activities including Ca^{+2} and K^{+} signaling, lipid transport in ER (acting as a chaperone in high-density membranes) and also in unfolded protein response. shRNA knock-down of SIGMAR1 increased apoptosis rate in NSC-34 cells and upregulation of SIGMAR1 expression was observed in case of ER stress (Al-Saif *et al.*, 2011). This evidence is indicative of SIGMAR1's role in fighting accumulation of protein aggregates and is consistent with the roles of other ALS-associated genes. The Glu102Gln mutation reported in the literature is

located in a region important for ligand binding. The novel Thr120Ala variant identified in this thesis, is not located in a significant region of the protein, however shows a proper family segregation thus far. In addition, another variation (Glu119Lys), only one amino acid prior to this variant, was previously reported in another juvenile ALS patient in our Turkish samples (Akçimen, 2017).

6.4.2. Differential Diagnosis in non-ALS MNDs

SPG15: The homozygous truncation mutation identified in the spastizin (ZFYVE26) gene, known to cause SPG15, explains the predominant upper motor neuron phenotype in the juvenile female patient. Mental retardation, hearing and visual defects and TCC are common symptoms of SPG15 (Goizet, 2009). These specific symptoms were not present in our patient, whose initial diagnosis could only be restricted to MND. Spastizin, shown to be involved in endosomal trafficking and LOF, results in axonal degeneration and loss of cortical motor neurons and Purkinje cells in mice (Khundadze *et al.*, 2013).

SPG7: Homozygous mutations in the *SPG7* gene (paraplegin protein) are known to be associated with an autosomal recessive form of HSP (SPG7); they were recently associated also with spastic ataxia (Choquet *et al.*, 2016). The already reported SPG7-Ala572Val mutation, explains the HSP diagnosis in our Family 16. Also, the index patient in Family 17 with an uncertain diagnosis of ALS/HSP was found to carry a novel truncation mutation in the *SPG7* gene. The disease onset ages of both patients are concordant with the literature. Additional LMN symptoms in the patient carrying the truncation mutation might further broaden the spectrum of *SPG7* mutations. Paraplegin is a metalloprotease functioning in quality control of mitochondrial proteins, however the mechanism behind LOF mutations in motor neuron disease is not yet understood (Fink, 2013).

DNAJB2: The tyrosine to cysteine change at the fifth amino acid of the *DNAJB2* gene detected in the male patient diagnosed with ALS was previously reported in a family with

CMT type 2; however, *DNAJB2* mutations are reported to have a broad phenotypic range, including CMT, dHMN and SMA. *DNAJB2*, also known as heat shock protein J1, is a chaperone protein and thought to be involved in the ubiquitin-proteasome pathway through its interaction with ubiquitin chains and is suggested to be involved in the clearance of TDP-43 (Lupo *et al.*, 2016). An interesting finding is reported by Frasquet *et al.*, pointing to a change in disease progression of the patients initially diagnosed with CMT2 or dHSMN resembling ALS in clinical follow-ups (Frasquet *et al.*, 2017). TDP-43 is a common hallmark of ALS. Considering the relationship with *DNAJB2* function and TDP-43, a LOF mechanism in *DNAJB2* mutations can be suggested for a slow progressive ALS starting with a mild phenotype. However, more evidence and clinical examination of the patients over the years would be crucial to jump to a conclusion.

SBF1: SBF1 is a catalytically inactive phosphatase that is thought to cause neuropathy through its interaction with MTMR2 (Myotubularin Related Protein 2) which is crucial for axonal development and integrity (Nakhro *et al.*, 2013). The juvenile patient with a CMT plus phenotype, including some bulbar symptoms, was shown to carry a homozygous variant in the *SBF1* gene, that might explain the polyneuropathy observed in the patient. Although consanguinity is evident in the pedigree, upon request of the clinician, duplication in the PMP22 gene, the most common mutation in CMT was excluded in the laboratory of Dr. Esra Battaloğlu. Filtered variants with AD or a possible *de novo* inheritance were not associated with the phenotype of the patient. *SBF1* is a very rare CMT gene and is associated with CMT4B3; phenotypic heterogeneity between patients in two reports includes additional developmental symptoms like microcephaly and syndactyly in the latter (Griggs *et al.*, 2014; Nakhro *et al.*, 2013). We propose that this variant, yet with unknown significance, is responsible from the CMT plus phenotype in our index patient and brings an additional complexity to the phenotype caused by the *SBF1* gene, however detailed phenotyping and additional segregation analysis will prove its pathogenicity.

MORC2: Another CMT-associated gene *MORC*, causing CMT2Z in autosomal dominant manner, is identified as candidate in the 21-year-old index patient with ALS diagnosis. *de novo* C>T conversions located in CpG-sequences, caused asymmetrically progressive cramps and weakness in individuals with AOs ranging from 6 months to 14 years with additional SMA-like features (Sevilla *et al.*, 2016). Furthermore, a following study reported another *de novo* mutation in a 5 year-old patient with a severe SMA phenotype, without CMT features (Schottmann *et al.*, 2016). The MORC2-p.Glu163Asp variant, identified in our index patient, is a G to T change and is not located in CpG-site, however muscle cramps and LMN involvement along with a highly possible *de novo* occurrence of the mutation in the sporadic/isolated patient make the *MORC2* variant a good candidate for further analysis in available family members. Mechanism of *MORC2* pathogenesis is not yet understood and there is no experimental finding to relate the mutations located in the CpG-sequences with the associated phenotype. In addition, the previously identified mutations (p.Ser25Leu, p.R190W and p.Thr362Arg) do not reside in a specific region/domain of the protein and are located far from each other.

TFG: The protein is suggested to play role in ER trafficking. Heterozygous mutations in the *TFG* gene are associated with a type of sensory motor neuropathy with proximal dominant involvement (HMSN-P) and are common in Okinawa, Japan. The female index patient in a large Turkish ALS kindred (Family 19) carry the common Japanese variant TFG-Pro285Leu (Ishiura *et al.*, 2012). Reported HMSN-P cases displayed fasciculations, proximal muscle weakness, atrophy and sensory involvement (Ishiura *et al.*, 2012). The Turkish patient diagnosed with ALS clinically displays similar symptoms as reported by the clinician. Another study conducted in Korean patients reported more severe and progressive disease in the affected individuals carrying the same mutation, compared to the original report (Lee *et al.*, 2013a). This observation supports the ALS phenotype in the Turkish family with relatively short survival times in affected individuals in the upper generations compared to a normal HMSN phenotype. Haplotype analysis in the family may unravel the same Japanese/Korean haplotype in this Turkish family.

VRK1: Homozygous *VRK1* mutations are associated with pontocerebellar hypoplasia type 1A, however two recent reports present patients with suspected ALS, SBMA, CMT or distal SMA diagnoses without pontocerebellar hypoplasia (Nguyen *et al.*, 2015; Stoll *et al.*, 2016). The clinical representation of patients, with ages of onset between 15-20, is compatible with the patients in the family reported in this thesis. More cases identified with *VRK1* mutations will enable a more robust clinical description of the non-PCH1A patients and will help its recognition as an MND gene.

6.4.3. Phenotypic Heterogeneity Caused by SPG11 Mutations

SPG11 is one of the most common genes implicated in ARHSP, which explains 34% of cases together with *SPG7*, *SPG5A* and *ZFYVE26 (SPG15)* (Parodi *et al.*, 2017). *SPG11* mutations results in HSP phenotypes ranging from pure-to-complex with the addition of weakness in limbs, peripheral neuropathy, cognitive deficit and cerebellar symptoms. More than half of ARHSP patients have reported TCC in MRI (Parodi *et al.*, 2017). *SPG11* mutations are also frequently observed among ARJALS cases with the emergence of LMN symptoms that are not common in HSP.

In all four juvenile ALS patients from our Families 4 and 5, both UMN and LMN signs were present, with LMN signs predominating. The presence of bulbar signs and the absence of sensory involvement and ocular anomalies, excluded the diagnosis of ARHSP. Moreover, TCC was not detected in our patients and cognitive decline was not reported. The vertical inheritance pattern initially suspected in Family 4, could not be further questioned, due to the absence of DNA material from the deceased father of the index case. However, the age of onset and fast disease progression coupled with non-consanguineous parents of the father, increases the possibility of other genetic or sporadic factors playing role in his pathogenesis (Figure 5.12). The effect of secondary rare missense variations identified in both cases is unknown, yet in Family 5 the truncating mutation is prior to the missense variant, which is

probably ineffective. Two male cases in Families 6 and 7 diagnosed with ARHSP also were found to carry two different truncation mutations in their *SPG11* genes. Their clinical features included evident TCC. Both patients had other clinical symptoms like bilateral nystagmus or resting tremor, indicative of ocular anomalies and extrapyramidal signs concordant with complex HSP.

Previous reports also suggest phenotypic differences between two siblings one presenting with HSP and the other showing ALS symptoms (Daoud *et al.*, 2012). Genetic heterogeneity of HSP and ALS is reflected by the presence of more than 70 and 40 associated gene loci, respectively (Parodi *et al.*, 2017). Altogether, these four Turkish families once more point to the complex and broad phenotypic spectrum that are caused by mutations in one single gene.

6.4.4. SYNE1 in the Turkish ALS Cohort

Mutations in the *SYNE1* gene, located on chromosome 6p25, were first identified in the French-Canadian population and the clinical phenotype was referred to as autosomal recessive cerebellar ataxia type I (ARCA1). ARCA1 is a slow progressing middle-age-onset ataxia with predominating features of gait ataxia, dysarthria, cerebellar atrophy and mild oculomotor abnormalities (Gros-Louis *et al.*, 2007). However, recent reports outside the French-Canadian population showed that *SYNE1* mutations cause a type of ataxia that commonly presents with multisystemic neurodegenerative disease, where non-cerebellar features accompanying ataxia are frequent (Synofzik *et al.*, 2016). Patients from families SYNE1#1 and #2 described here are in accordance with the reported phenotypes. Three additional families in our ataxia cohort are shown to carry distinct homozygous truncation mutations in the *SYNE1* gene; patients in two of these families present with motor neuron symptoms in addition to prominent cerebellar ataxia. Altogether, considering the identification of truncating *SYNE1* variants in five cases in our laboratory, it can be concluded that *SYNE1* mutations are common causes of ataxia

accompanied by motor neuron symptoms in the Turkish population. These specific clinical characteristics should be kept in mind by the clinicians during deep phenotyping.

SYNE1 is among the largest genes in the human genome encoding a 8797 amino acid-long protein, nesprin-1, from its longest isoform of 147 exons. Almost 50 truncation or nonsense mutations have been reported, all scattered on the large protein, which relate to ataxia phenotype either in homozygous or compound heterozygous forms. Nesprin-1 belongs to the family of spectrin repeat containing proteins and is involved in the maintenance of nuclear organization and structural integrity by binding the nuclear envelope to the cytoplasm by F-actin. It anchors specialized myonuclei to the neuromuscular junctions (NMJs). Spectrin repeat containing proteins like alsin and pleckstrin were previously associated with ALS (Maystadt *et al.*, 2007; Yang *et al.*, 2001). Mouse brain tissue showed expression levels higher in the Purkinje cells and neurons of olivary region of the brain stem and the protein was not detected in the non-neuronal cells (Gros-Louis *et al.*, 2007). *SYNE1* gene, one of the largest genes in the human genome, harbors many missense mutations with high MAFs. However, relevance of non-truncating missense mutations with ALS and/or cerebellar ataxia remains to be elusive (Synofzik *et al.*, 2016). Strong segregation in as many samples as possible is needed to make an assumption for these variants.

6.4.5. A Novel ALS Gene: ERLIN1

ERLIN1 encodes for a prohibitin-domain containing protein located in the ER membrane that forms a ring-shaped complex with ERLIN2. ERLIN1/2 complex is implicated in endoplasmic reticulum-associated degradation (ERAD) control and is responsible for forming a bridge between the substrates recognized by the ER lumen and the E3 ligases located in the ER membrane (Olzmann *et al.*, 2013). ER stress is one of the major pathogenic events in ALS, as protein products of ALS-causing genes like optineurin, ubiquilin-2, valosin-containing protein and TANK-binding kinase-1 are involved in several steps of the ERAD pathway which includes the recognition of misfolded proteins, ubiquitination and transportation of the

cargo, finally degradation through proteasome or autophagosome (Cirulli *et al.*, 2015; Taylor *et al.*, 2016).

Three different ERLIN1 mutations have been previously identified in seven patients with pure ARHSP in three independent families, all having infantile-onset symptoms, except one patient with an age of onset of 13 (Novarino *et al.*, 2014). Two of these variations, are located in the low complexity domain of the protein, whereas one is in the prohibitin domain, similar to our variation. Mutations in the ERLIN1 gene, associated with spastic paraplegia 62 (SPG62), were reported to cause a pure form of HSP that primarily affects the upper motor neurons (Parodi *et al.*, 2017). As a novel observation, the ERLIN1 gene variant in our index case and her father have led to definitive ALS according to El Escorial criteria. The ages of onset and shortened survival times due to lower motor neuron involvement later in the disease resemble typical adult-onset ALS in all patients investigated within this family.

Very often, it is hard to distinguish diseases from each other, when there are overlapping clinical features. *SPG11*-based disease is a good example, where clinical diagnosis may differ even among sibs as either ARHSP or autosomal recessive juvenile ALS (Daoud *et al.*, 2012; Iskender *et al.*, 2015). There are key clinical features that are present in one case, but not in others, which draw a line between two clinical diagnoses, but may change even during the course of the disease. We therefore strongly suggest that the ERLIN1 variation in our Family 10 results in a clinical presentation, which starts as a mild form of HSP in the patients and progresses to ALS.

Our knowledge today is not comprehensive enough to hold *ABCC2* responsible for the motor neuron phenotype seen in our family, although it segregates with the disease gene. More research is needed to understand the effect of the *ABCC2* variant on manifestation and progression of the phenotype in this family, if there is any.

6.4.6. Clinical Exome Sequencing: Where Are We?

The results depicted in this thesis once more demonstrate that, phenotypic heterogeneity within ALS has risen to a point that the disease is no more considered as a single clinical entity, but a group of phenotypes caused by different gene mutations. This allelic heterogeneity recently recognized with the help of WES, introduced the concept of gene-based diseases, each with distinct clinical features. Identification of new genetic players implicated in diseases open new opportunities for understanding the converging mechanisms for broader concepts like neurodegeneration or motor neuron loss; moreover, they also contribute to the development of more specific, even personalized, therapeutic targets like gene-specific antisense oligonucleotides. Thus, today, it has become more important to define the genetic causes of even yet untreatable diseases, to drive pharmaceutical/gene-editing research and to offer hope to patients and their asymptomatic relatives (Fogel *et al.*, 2016).

Developing NGS technologies and decreasing costs boosted genetic testing both for clinical and research purposes in the last decade. An exponential increase in the number of genes leading especially to the familial form of ALS is a courtesy of NGS and reflects its power. It is speculated that the majority of inherited human diseases are caused by mutations in protein coding regions. In this perspective, WES is expected to reveal the causative mutations in a considerable number of cases allowing data interpretation, less dependent of the initial clinical diagnosis as in the case of gene panels, that are used to screen disease-specific genes.

This disease-independent candidate variant prioritization approach allowed to differentially diagnose cases with uncertain clinical diagnosis due to overlapping phenotypic features between diseases. Beneficial outcomes of clinical exome sequencing include; (i) treatment of several patients with syndromic diseases like enzyme deficiencies, especially in the case of ataxias, (ii) using genetic information for reproductive genetic counselling and family planning and (iii) recruitment of patients with specific mutations to clinical trials

(Fogel *et al.*, 2014, 2016). Although its contribution to the literature is undeniable, Schwarze and colleagues suggest that current data is still not sufficient to say whether WES and WGS are cost-effective for frequent clinical use (Schwarze *et al.*, 2018). On the other hand, WES shortens the diagnostic delay, which may include several invasive and expensive procedures, and these should be taken into account, while considering its cost-effectiveness.

Like in the case of all rapidly evolving technological advances, we are now running in the front line with WES, which together brings issues to be addressed. Interpretation of an excessive number of variants detected and ethical issues concerning secondary findings in patients irrelevant to the disease of interest are some of the current problems to be solved. A workgroup representing The American College of Medical Genetics and Genomics (ACMG), the Association for Molecular Pathology (AMP), and College of American Pathologists created guidelines for the interpretation of sequence variants in a more standardized manner (Richards *et al.*, 2015). According to these guidelines evidence regarding frequency, conservation, functional relevance and segregation analyses are used to categorize a variant as pathogenic, likely pathogenic, variant of uncertain significance, likely benign, or benign. A study showed that laboratories using internal guidelines based on self-experience, leads to interpretation differences among groups and thus it is important to interiorize a more standard approach (Amendola *et al.*, 2016). In general, to overcome misinterpretation we suggest deep phenotyping of the patient, comprehensive analyses on the candidate variants using available bioinformatic tools and guidelines, and most importantly a tight researcher-clinician relationship that will result in a correct genotype-phenotype correlation.

WES has proven itself to be more than useful for research and clinical purposes. However, as our results also show, it has failed to detect pathogenic variants in almost 60% of cases worldwide and also in our cohort (Trujillano *et al.*, 2017). This situation can derive from several factors with either clinical or technical bases. Insufficient clinical and pedigree information in the patient can result in misleading candidate gene prioritization, which uses the mode of inheritance and filtration of variants within family members. On the other hand,

existence of a novel disease-causing gene in the patient/family can be hard to detect in small families with a few number of affected individuals. Finally, inability of WES to detect mutations in intronic and intergenic regions, which may have important regulatory effects on protein coding sequences and technical incapacity of NGS technologies to map large structural variations to the genome, might explain the missing success rate in clinical exome sequencing. Altogether, today, WES is accepted to be the gold standard in clinical diagnosis of complex genetic disorders. We believe that although WGS technology will be replacing WES for research purposes in the future, WES will probably secure its pioneering place in the clinic due to easier bioinformatic handling, interpretation and storage of the data.

6.5. Incomplete Penetrance in ALS, Disease Modifiers and Oligogenic Inheritance

The concept of modifier genes was introduced in mid-1900s as secondary genes that contribute to differences in age of onset or disease progression, where a primary disease gene is present (Haldane, 1941). Today, we are confident that the genetic background of a disease-causing mutation is very important for clinical variability and penetrance, but few studies have managed to identify modifiers, probably due to their complex mode of actions. The term modifier can alter according to the disease of interest; (i) in a polygenic disease, the genetic (locus) heterogeneity can lead to different disease manifestations, (ii) in case of a single gene, secondary variations that are in cis or trans positions of the same gene can have an extra effect on the phenotype and (iii) variations in other genes may result in variable phenotypes among cases carrying a common primary mutation (Génin *et al.*, 2008). Existing modifier studies mostly rely on animal model studies in which candidates are selected using gene expression screens or rescue/suppressor screens (van Hoecke *et al.*, 2012; Zhan *et al.*, 2013). With the increasing number of studies based on next-generation sequencing approaches we expect more modifying genes/variants to be detected. Current studies with these approaches are mostly association analyses that look for the secondary variants hidden in the genomes of the patients with the same clinical variability and thus require large sample cohorts as in the example of a genome-wide association study performed with more than 4000 Huntington patients to define

two regions on chromosome 15, modifying age of onset (Lee *et al.*, 2015). In addition to the biological data driven and association approaches, linkage analysis is an alternative approach for searching for disease modifiers within families (Génin *et al.*, 2008). In this case, candidate modifier loci are identified by the co-segregation of the markers and the desired phenotype, which may be earlier age of onset, dementia, fast progression or even non-penetrance.

As seen in almost all neurodegenerative disorders, ALS harbors a broad phenotypic spectrum, including phenotypic variability among affected individuals carrying mutations in the same gene. This variability may even be present among the members of the same family. Search for modifiers/disease-triggering factors using families is advantageous considering the reduced genetic variability among family members compared to other people. Our attempt was to investigate two families with repetitive incomplete penetrance of highly penetrant ALS genes using WES. VarSifter filtrations revealed numerous exonic variants, whose interpretation is ongoing.

Incomplete penetrance, in general, may be associated with the existence of a factor modifying the age of onset that delays the phenotype into an old age, where the individual might die with other diseases. In this sense, ALS-707 (the paternal aunt) in the SOD1-His71Tyr family, whose disease onset is 38 years later than her nephew (the index patient), might also carry the same modifier with all other asymptomatic carriers (ALS-236, ALS-704 and ALS-708), who will eventually manifest the disease if they live long enough. Thus, VarSifter filtering and homozygosity mapping data can be further analyzed based on this hypothesis to look for a common risk allele or an additional rare variant segregating among family members. Interestingly, in the TARDBP-Gly348Cys family, only two variants with possible X-linked protective effect remained in the non-manifesting females. Unfortunately, it is not possible to make an assumption on the relevance of these variants on disease phenotype at this point.

Searching for disease modifiers is challenging and our findings are too preliminary, yet we believe that large families with distinct properties are valuable materials to understand possible genetic and/or epigenetic events in disease pathogenesis. In addition, although working with families has its benefits, we agree in the requirement of large-scale association tests to detect genome-wide common SNPs that act as modifiers or risk alleles. With the increasing number of mutation carriers in our laboratory and our international collaborations, a future modifier study with *SOD1*- and *TARDBP*-positive cases would enable performing linear association tests of different phenotypic features. Candidate variations identified through filtration steps within this study, can be questioned in these future investigations. Furthermore, as successfully shown in a recent study, the modifying effect of possible regulatory variants in *cis* with the primary mutation can be assessed using WGS to focus on the variable expressivity among affected and unaffected family members (Castel *et al.*, 2018).

Oligogenic inheritance was not prominent among *C9orf72* expansion carriers in our cohort and the secondary effects of the identified variants in the *C9orf72* expansion-positive patients are not known. Oligogenic inheritance is shown to be evident in ALS, especially in cases carrying the *C9orf72* expansion, however drawing conclusions about the pathogenic effects of individual SNPs might be wrong. As can be seen from Table 5.8, some of the SNPs identified in ALS-associated genes in Project MinE samples are observed in the control population and might be just polymorphisms without any adverse effect. In addition, one important point is to consider the *cis* or *trans* location of double mutations to be able to suggest a combinatorial effect. Due to the small cohort size, regression analysis could not be performed to search for novel modifier SNPs that may affect several characteristics of the disease, like age of onset, site of onset and survival. In order to find novel disease-modifying factors and to confirm the already proposed ones, large sample cohorts with detailed phenotype information is needed. We are partners in an international collaboration project that aims to investigate the modifiers of *C9orf72*-based disease by sequencing the whole genomes of expansion carriers. Our contribution to the project was initiated with 70 *C9orf72*-positive samples with diverse clinical manifestations.

6.6. Preliminary Analyses in Project MinE Samples

Sample collection and stratification is one of the most important steps in consortium projects. In the framework of this thesis, DNA samples from 625 sporadic ALS patients and 152 healthy controls were carefully chosen and prepared for genome sequencing in the framework of Project MinE. Detailed clinical data was collected from all patients. *C9orf72* repeat expansion was excluded in all patients, to be able to choose as much patients as possible representing true sporadic ALS with complex inheritance. Although super controls (individuals over 80-90 years) are preferred as control subjects for adult-onset neurodegenerative diseases, we included several age-matched controls for the future analysis of epigenetic data that is generated within the scope of the project. Only individuals without family history of neurodegenerative diseases were included in the study as healthy controls. The mean AO for the patients included in Project Mine was 51 years, in agreement with 52 years-old mean AO of the total sALS cohort; control subjects had a mean AO of 55 years.

Causative mutations in genes associated with ALS, explain so far 60-70% of familial cases in Caucasian populations (40% in Turkey), and this number decreases to less than 10% in sporadic disease. In the genomic era, next generation sequencing technologies offer promising results for the identification of novel rare genetic causes of diseases and common SNPs inferring disease risk, using large cohorts. Project MinE is expected to shed light on the effect of oligogenic/complex inheritance suspected in sporadic ALS. In addition, the project will complete a missing piece in the disease with simplex inheritance (fALS) through possible identification of rare or noncoding disease-causing mutations, which could not be detected so far due to small cohorts or technological limitations of GWAS and WES.

Analyses based on the hypothesis of expected accumulation of rare or common variants in patients compared to controls revealed significant increase of variants in patients only for homozygous variants. For our dataset, we believe that this observation roots from population-based properties rather than its relationship with ALS pathogenesis. Considering the high rates

of consanguinity in rural parts of Turkey, increased number of homozygous variants in the overall population is expected. However, this notion does not explain the difference between two groups. We can speculate that the accumulation of homozygous variants in patients might be due to a bias in patients' origins compared to controls that are collected mostly from Istanbul province and are better educated, in whom consanguinity is not common. In order to make a conclusion, the origins and the rate of consanguinity in families of both patients and controls should be assessed. It is expected that as the number of controls with diverse origins increases, this bias would disappear. On the other hand, the gradual increase in p-value significance in higher thresholds is acceptable, since the statistical test yields more conclusive results with increased number of variants added with less strict thresholds. Finally, the tendency of increased number of heterozygous variants in controls compared to patients is probably a reflection of decreased number of homozygous variants in controls, which again, might be related to their more heterogenous genetic background. At this point, our analyses do not allow to make a correlation between variant accumulation and disease risk. However, our pilot study presents some population-specific aspects and can be further expanded to the data of different populations participating at Project MinE. Furthermore, we hope that with increased number of controls or with the comparison of data from other populations, we might see some disease-specific features that are independent of the genetic background of the population.

Computational screening of exonic variants in our WGS cohort revealed variants in either ALS-causing or ALS-associated genes that are predicted to be pathogenic (Table 5.8). The truncation mutations detected in *TBKI* and *OPTN* genes are concordant with the LOF mechanism suggested for these genes. The pathogenicity of all other variants detected should be further investigated considering the low penetrance observed in rare ALS-causing variations and the positional effect of the variants, especially in *VCP* and *ERBB4* genes, the ALS-causing mutations of which, are restricted to their C-terminals. As discussed frequently in the literature, a high deleterious score of functional prediction or novelty of a candidate variant should not be considered sufficient for classifying a variant as disease-causing in a

particular individual. The most powerful proof of pathogenicity is the segregation of the variant with the phenotype of interest that is preferably coupled with functional experimental evidence. However, especially for variants/genes with low penetrance and risk alleles segregation analyses can be misleading. Large-scale cohort studies and expanding literature can give insight into these variants, but all this information should be interpreted with care. Variants identified through bioinformatic analyses need further assessment to be considered as pathogenic. Our current observations once more point out that the disease cause in a fair portion of sporadic ALS cases is explained by rare ALS-causing genes and this arises from reduced penetrance. Computational screening of exonic ALS gene mutations in our WGS cohort is expected to explain 3-4% of apparently sporadic ALS cases, which adds up to almost 10% of sALS cases with the inclusion of mutations in common ALS genes implicated in sporadic cases. This percentage is concordant with all other studies and compels attention to the yet unidentified factors acting on ALS pathogenesis that will eventually be understood through large collaboration projects investigating regulatory non-coding regions and structural variations and addressing environmental and epigenetic factors.

6.7. Conclusion

The completion of the Human Genome Project and advances in massive parallel sequencing taught us that genetics is instrumental in the understanding of the molecular mechanisms that underlie complex diseases like ALS, which is characterized by a large contribution of rare mutations. In the framework of this thesis the genetic bases of ALS and non-ALS MNDs were investigated in Turkey in a large and well-studied cohort nationwide-collected, using state-of-the-art genetic technologies.

In the first part of this thesis screening for pathogenic variants in known ALS genes was exploited as a diagnostic tool in familial ALS using conventional PCR-based DNA methods and Sanger sequencing. The results obtained indicate that ALS in Turkey has features in accordance with the pattern predominating in Caucasian populations, however with also a

distinct character. Whereas in European populations the most common four ALS genes account for 70% of the population under investigation, in Turkey this figure is less than 50%, pointing to a more heterogeneous genetic make-up.

ALS is a heterogeneous condition, it presents with several pathogenic mechanisms and different clinical manifestations. Thus, many ALS-linked genes are associated with a wide spectrum of diseases. In recent years WES provided rapid progress in systematically unraveling the complete genetic landscape of ALS, and this technique was applied in the second part of this thesis to all families with a distinct inheritance pattern, which could not be solved by conventional screening. This high-throughput and unbiased technology is very effective in identifying not only single ALS-associated genes, but also multiple gene mutations. The allelic heterogeneity in ALS, with distinct clinical/phenotypical features, recognized by WES holds also true for the Turkish population. The results obtained in this thesis are unique examples of a gene-based disease concept, explaining the factors, which render the differential diagnosis of neurodegenerative diseases difficult.

Not all genetic variations can be captured by WES. In the final part of this thesis, genetic analysis was employed in the identification of rare and common variants, which confer risk for sporadic ALS, by using WGS. The effect of genetic factors is expected to be somewhat milder in sporadic ALS as compared to familial disease, but very often the same genes are involved in both. Furthermore, the frequency of ALS patients carrying two or more mutations in ALS-associated genes is found to be higher than expected. A very important topic in the investigation of heterogeneous diseases is understanding oligogenic inheritance and disease modifiers, which either result in increased risk of developing the disease or can affect the course of disease, the age of onset and/or disease survival. Sporadic ALS may even be an oligogenic disease caused by multiple rare variants with synergistic effects on disease development. In collaboration with Project MinE, in the framework of which WGS data from several thousand individuals worldwide will be released, ~800 Turkish samples were whole-genome sequenced. With the corresponding methylome data this is a huge information. Data

in this size is available for the first time for the Turkish population, the preliminary results of which are presented in this thesis for the first time. Most importantly, the data obtained in this thesis initiated the basis for a Turkish population-specific database. The interpretation of this data will occupy us for many years to come.

ALS provides an excellent model to study and understand neurodegenerative diseases. To the best of our knowledge this is the most comprehensive study investigating the complex genetics of ALS and non-ALS MNDs in an ethnically heterogeneous, partly inbred and largely unexplored population like Turkey. Emerging novel themes in ALS research and changing paradigms in the understanding of disease pathology, e.g. convergence of familial and sporadic disease, questioning penetrance, oligogenic inheritance and modifier genes, de novo mutations, the *C9orf72* expansion introducing ALS and FTD as the extremes of the same disease spectrum, have all been dealt within its scope. The results produced in the framework of this thesis will hopefully contribute to the diversity of genetic and mechanistic factors underlying ALS, further driving ALS research. Moreover, they are expected to shed a new light on this multilevel heterogeneity, as an important factor in a precision medicine approach towards the development of long-awaited molecular therapies, tailor-made for stratified patient subgroups.

REFERENCES

- Adzhubei, I., D. M. Jordan, and S. R. Sunyaev, 2013, “Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2”, *Current Protocols in Human Genetics*, Vol. 76, No. 1, p. 7.20.1-7.20.41.
- Akçimen, F., 2017, *Investigation of Motor Neuron Diseases By WES: Genetic Dissection of a Turkish ALS Cohort*, M.S., Boğaziçi University.
- Aksoy, H., G. Dean, M. Elian *et al.*, 2003, “A4T Mutation in the SOD1 Gene Causing Familial Amyotrophic Lateral Sclerosis”, *Neuroepidemiology*, Vol. 22, No. 4, pp. 235–238.
- Al-Chalabi, A., and O. Hardiman, 2013, “The Epidemiology of ALS: A Conspiracy of Genes, Environment and Time”, *Nature Reviews Neurology*, Vol. 9, No. 11, pp. 617–628.
- Al-Chalabi, A., L. H. Van Den Berg, and J. Veldink, 2017, “Gene Discovery in Amyotrophic Lateral Sclerosis: Implications for Clinical Management”, *Nature Reviews Neurology*, Vol. 13, No. 2, pp. 96–104.
- Al-Saif, A., F. Al-Mohanna, and S. Bohlega, 2011, “A Mutation in Sigma-1 Receptor Causes Juvenile Amyotrophic Lateral Sclerosis”, *Annals of Neurology*, Vol. 70, No. 6, pp. 913–919.
- Amendola, L. M., G. P. Jarvik, M. C. Leo *et al.*, 2016, “Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium”, *American Journal of Human Genetics*, Vol. 99, No.

1, p. 247.

Andersen, P. M., P. Nilsson, V. Ala-Hurula *et al.*, 1995, “Amyotrophic Lateral Sclerosis Associated with Homozygosity for an Asp90Ala Mutation in CuZn-Superoxide Dismutase.”, *Nature Genetics*, Vol. 10, No. 1, pp. 61–6.

Andersen, P. M., 2006, “Amyotrophic Lateral Sclerosis Associated with Mutations in the CuZn Superoxide Dismutase Gene”, *Current Neurology and Neuroscience Reports*, Vol. 6, No. 1, pp. 37–46.

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

Anheim, M., C. Tranchant, and M. Koenig, 2012, “The Autosomal Recessive Cerebellar Ataxias”, *New England Journal of Medicine*, Vol. 366, No. 7, pp. 636–646.

Ash, P. E. A., K. F. Bieniek, T. F. Gendron *et al.*, 2013, “Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to C9FTD/ALS”, *Neuron*, Vol. 77, No. 4, pp. 639–646.

Auer, P. L., and G. Lettre, 2015, “Rare Variant Association Studies: Considerations, Challenges and Opportunities”, *Genome Medicine*, Vol. 7, No. 1, p. 16.

Avitzour, M., H. Mor-Shaked, S. Yanovsky-Dagan *et al.*, 2014, “FMR1 Epigenetic Silencing Commonly Occurs in Undifferentiated Fragile X-Affected Embryonic Stem Cells.”, *Stem Cell Reports*, Vol. 3, No. 5, pp. 699–706.

- Bak, T., 2010, “Motor Neuron Disease and Frontotemporal Dementia: One, Two, or Three Diseases?”, *Annals of Indian Academy of Neurology*, Vol. 13, No. 6, p. 81.
- Bamshad, M. J., S. B. Ng, A. W. Bigham, H. K. Tabor, M. J. Emond, D. A. Nickerson, and J. Shendure, 2011, “Exome Sequencing as a Tool for Mendelian Disease Gene Discovery”, *Nature Reviews Genetics*, Vol. 12, No. 11, pp. 745–755.
- Battistini, S., M. Benigni, C. Ricci, and A. Rossi, 2013, “SOD1 Mutations in Amyotrophic Lateral Sclerosis”, *European Neurological Journal*, Vol. 252, No. 7, pp. 782–788.
- Belzil, V. V., P. O. Bauer, M. Prudencio *et al.*, 2013, “Reduced C9orf72 Gene Expression in C9FTD/ALS Is Caused by Histone Trimethylation, an Epigenetic Event Detectable in Blood”, *Acta Neuropathologica*, Vol. 126, No. 6, pp. 895–905.
- Belzil, V. V., P. O. Bauer, T. F. Gendron, M. E. Murray, D. Dickson, and L. Petrucelli, 2014, “Characterization of DNA Hypermethylation in the Cerebellum of C9FTD/ALS Patients”, *Brain Research*, Vol. 1584, pp. 15–21.
- Borghero, G., M. Pugliatti, F. Marrosu *et al.*, 2014, “Genetic Architecture of ALS in Sardinia”, *Neurobiology of Aging*, Vol. 35, No. 12, p. 2882.e7-2882.e12.
- Boycott, K. M., M. R. Vanstone, D. E. Bulman, and A. E. MacKenzie, 2013, “Rare-Disease Genetics in the Era of next-Generation Sequencing: Discovery to Translation”, *Nature Reviews Genetics*, Vol. 14, No. 10, pp. 681–691.
- Brenner, D., R. Yilmaz, K. Müller *et al.*, 2018, “Hot-Spot KIF5A Mutations Cause Familial ALS”, *Brain*, Vol. 141, No. 3, pp. 688–697.

- Brooks, B., and R. Miller, 2000, “El Escorial Revisited: Revised Criteria for the Diagnosis of Amyotrophic Lateral Sclerosis”, *Amyotrophic Lateral Sclerosis*, Vol. 1, No. 5, pp. 293–9.
- Brown, R. H., and A. Al-Chalabi, 2017, “Amyotrophic Lateral Sclerosis”, *New England Journal of Medicine*, Vol. 377, No. 2, pp. 162–172.
- Cai, L., A. L. Fisher, H. Huang, and Z. Xie, 2016, “CRISPR-Mediated Genome Editing and Human Diseases”, *Genes and Diseases*, Vol. 3, No. 4, pp. 244–251.
- Castel, S. E., A. Cervera, P. Mohammadi *et al.*, 2018, “Modified Penetrance of Coding Variants by Cis-Regulatory Variation Contributes to Disease Risk”, *Nature Genetics*, Vol. 50, pp. 1327–1334.
- Cerami, C., A. Marcone, D. Galimberti, M. Zamboni, C. Fenoglio, M. Serpente, E. Scarpini, and S. F. Cappa, 2013, “Novel Evidence of Phenotypical Variability in the Hexanucleotide Repeat Expansion in Chromosome 9”, *Journal of Alzheimer’s Disease*, Vol. 35, pp. 455–462.
- Chen, Y.-Z., C. L. Bennett, H. M. Huynh *et al.*, 2004, “DNA/RNA Helicase Gene Mutations in a Form of Juvenile Amyotrophic Lateral Sclerosis (ALS4)”, *The American Journal of Human Genetics*, Vol. 74, No. 6, pp. 1128–1135.
- Chia, R., A. Chiò, and B. J. Traynor, 2018, “Novel Genes Associated with Amyotrophic Lateral Sclerosis: Diagnostic and Clinical Implications”, *The Lancet Neurology*, Vol. 17, No. 1, pp. 94–102.
- Choquet, K., M. Tétreault, S. Yang *et al.*, 2016, “SPG7 Mutations Explain a Significant Proportion of French Canadian Spastic Ataxia Cases”, *European Journal of Human*

Genetics, Vol. 24, No. 7, pp. 1016–1021.

Cirulli, E. T., B. N. Lasseigne, S. Petrovski *et al.*, 2015, “Exome Sequencing in Amyotrophic Lateral Sclerosis Identifies Risk Genes and Pathways”, *Science*, Vol. 347, No. 6229, pp. 1436–1441.

Ciura, S., S. Lattante, I. Le Ber, M. Latouche, H. Tostivint, A. Brice, and E. Kabashi, 2013, “Loss of Function of C9orf72 Causes Motor Deficits in a Zebrafish Model of Amyotrophic Lateral Sclerosis”, *Annals of Neurology*, Vol. 74, No. 2, pp. 180–187.

Cooper-Knock, J., M. J. Walsh, A. Higginbottom *et al.*, 2014, “Sequestration of Multiple RNA Recognition Motif-Containing Proteins by C9orf72 Repeat Expansions”, *Brain*, Vol. 137, No. 7, pp. 2040–2051.

Cooper-Knock, J., J. Kirby, R. Highley, and P. J. Shaw, 2015, “The Spectrum of C9orf72-Mediated Neurodegeneration and Amyotrophic Lateral Sclerosis”, *Neurotherapeutics*, Vol. 12, No. 2, pp. 326–339.

Crimella, C., C. Baschiroto, A. Arnoldi *et al.*, 2012, “Mutations in the Motor and Stalk Domains of KIF5A in Spastic Paraplegia Type 10 and in Axonal Charcot-Marie-Tooth Type 2”, *Clinical Genetics*, Vol. 82, No. 2, pp. 157–164.

Cruts, M., I. Gijselinck, T. Van Langenhove, J. van der Zee, and C. Van Broeckhoven, 2013, “Current Insights into the C9orf72 Repeat Expansion Diseases of the FTL/ALS Spectrum”, *Trends in Neurosciences*, Vol. 36, No. 8, pp. 450–459.

Danecek, P., A. Auton, G. Abecasis *et al.*, 2011, “The Variant Call Format and VCFtools”, *Bioinformatics*, Vol. 27, No. 15, pp. 2156–2158.

- Daoud, H., S. Zhou, A. Noreau, M. Sabbagh, V. Belzil, A. Dionne-Laporte, C. Tranchant, P. Dion, and G. a. Rouleau, 2012, “Exome Sequencing Reveals SPG11 Mutations Causing Juvenile ALS”, *Neurobiology of Aging*, Vol. 33, No. 4, pp. 5–9.
- de Carvalho, M., R. Dengler, A. Eisen *et al.*, 2008, “Electrodiagnostic Criteria for Diagnosis of ALS”, *Clinical Neurophysiology*, Vol. 119, No. 3, pp. 497–503.
- DeJesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve *et al.*, 2011, “Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS.”, *Neuron*, Vol. 72, No. 2, pp. 245–56.
- Deschauer, M., C. Gaul, C. Behrmann, H. Prokisch, S. Zierz, and T. B. Haack, 2012, “C19orf12 Mutations in Neurodegeneration with Brain Iron Accumulation Mimicking Juvenile Amyotrophic Lateral Sclerosis”, pp. 2434–2439.
- Elden, A. C., H. Kim, M. P. Hart *et al.*, 2011, “Ataxin-2 Intermediate-Length Polyglutamine Expansions Are Associated with Increased Risk for ALS”, *Nature*, Vol. 466, No. 7310, pp. 1069–1075.
- Farg, M. a., V. Sundaramoorthy, J. M. Sultana *et al.*, 2014, “C9ORF72, Implicated in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia, Regulates Endosomal Trafficking”, *Human Molecular Genetics*, Vol. 23, No. 13, pp. 3579–3595.
- Figuerola-Romero, C., J. Hur, D. E. Bender *et al.*, 2012, “Identification of Epigenetically Altered Genes in Sporadic Amyotrophic Lateral Sclerosis”, *PLoS ONE*, Vol. 7, No. 12, p. e52672.
- Fink, J. K., 2013, “Hereditary Spastic Paraplegia: Clinico-Pathologic Features and Emerging

Molecular Mechanisms”, *Acta Neuropathologica*, Vol. 126, No. 3, pp. 307–328.

Fogel, B. L., H. Lee, J. L. Deignan *et al.*, 2014, “Exome Sequencing in the Clinical Diagnosis of Sporadic or Familial Cerebellar Ataxia.”, *JAMA Neurology*, Vol. 71, No. 10, pp. 1–10.

Fogel, B. L., S. Satya-Murti, and B. H. Cohen, 2016, “Clinical Exome Sequencing in Neurologic Disease: AAN Model Coverage Policy.”, *Neurology: Clinical Practice*, Vol. 6, No. 2, pp. 164–176.

Frasquet, M., J. F. Va, and T. Sevilla, 2017, “The Role of DNAJB2 in Amyotrophic Lateral Sclerosis”, Vol. 139, No. 10, p. e57.

Fratta, P., S. Mizielinska, A. J. Nicoll, M. Zloh, E. M. C. Fisher, G. Parkinson, and A. M. Isaacs, 2012, “C9orf72 Hexanucleotide Repeat Associated with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Forms RNA G-Quadruplexes”, *Scientific Reports*, Vol. 2, pp. 1–6.

Fratta, P., M. Poulter, T. Lashley *et al.*, 2013, “Homozygosity for the C9orf72 GGGGCC Repeat Expansion in Frontotemporal Dementia”, *Acta Neuropathologica*, Vol. 126, No. 3, pp. 401–409.

Freischmidt, A., T. Wieland, B. Richter *et al.*, 2015, “Haploinsufficiency of TBK1 Causes Familial ALS and Fronto-Temporal Dementia”, *Nature Neuroscience*, Vol. 18, No. 5, pp. 631–636.

Gendron, T. F., K. F. Bieniek, Y.-J. Zhang *et al.*, 2013, “Antisense Transcripts of the Expanded C9ORF72 Hexanucleotide Repeat Form Nuclear RNA Foci and Undergo Repeat-Associated Non-ATG Translation in C9FTD/ALS”, *Acta Neuropathologica*, Vol.

126, No. 6, pp. 829–844.

Gendron, T. F., M. van Blitterswijk, K. F. Bieniek *et al.*, 2015, “Cerebellar C9RAN Proteins Associate with Clinical and Neuropathological Characteristics of C9ORF72 Repeat Expansion Carriers”, *Acta Neuropathologica*, Vol. 130, No. 4, pp. 559–573.

Génin, E., J. Feingold, and F. Clerget-Darpoux, 2008, “Identifying Modifier Genes of Monogenic Disease: Strategies and Difficulties”, *Human Genetics*, Vol. 124, No. 4, pp. 357–368.

Ghasemi, M., and R. H. Brown, 2018, “Genetics of Amyotrophic Lateral Sclerosis”, Vol. 8, No. 5, pp. 1–38.

Gijssels, I., S. Van Mossevelde, J. Van Der Zee, A. Sieben, S. Engelborghs, J. De Bleecker, A. Ivanoiu, O. Deryck, and D. Edbauer, 2016, “The C9orf72 Repeat Size Correlates with Onset Age of Disease, DNA Methylation and Transcriptional Downregulation of the Promoter”, *Molecular Psychiatry*, Vol. 21, No. 8, pp. 1112–1124.

Gitler, A. D., and H. Tsuiji, 2016, “There Has Been an Awakening: Emerging Mechanisms of C9orf72 Mutations in FTD/ALS”, *Brain Research*, Vol. 15, No. 1647, pp. 19–29.

Glass, J. D., V. S. Hertzberg, N. M. Boulis *et al.*, 2016, “Transplantation of Spinal Cord-Derived Neural Stem Cells for ALS: Analysis of Phase 1 and 2 Trials”, *Neurology*, Vol. 87, No. 4, pp. 392–400.

Goizet, C. E. A., 2009, “SPG15 Is the Second Commonest Cause of Hereditary Spastic Paraplegia with Thin Corpus Callosum.”, *Neurology*, Vol. 73(14), pp. 1111–1119.

- Gómez-Tortosa, E., J. Gallego, R. Guerrero-López *et al.*, 2013, “C9ORF72 Hexanucleotide Expansions of 20-22 Repeats Are Associated with Frontotemporal Deterioration”, *Neurology*, Vol. 80, No. 4, pp. 366–370.
- Gonzaga-Jauregui, C., T. Lotze, L. Jamal *et al.*, 2013, “Mutations in VRK1 Associated with Complex Motor and Sensory Axonal Neuropathy plus Microcephaly”, *JAMA Neurology*, Vol. 70, No. 12, pp. 1491–1498.
- Griggs, R. C., C. Furberg, C. Kawas, and M. Carlson, 2014, “Write Click : Editor ’ s Choice”, *Neurology*, Vol. 82, No. 1, pp. 1192–1193.
- 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.
- Haldane, J. B. S., 1941, “The Relative Importance of Principal and Modifying Genes in Determining Some Human Diseases”, *Journal of Genetics*, Vol. 41, No. 2, pp. 149–157.
- Hamzeiy, H., D. Savaş, C. Tunca *et al.*, 2018, “Elevated Global DNA Methylation Is Not Exclusive to Amyotrophic Lateral Sclerosis and Is Also Observed in Spinocerebellar Ataxia Types 1 and 2”, *Neurodegenerative Diseases*, Vol. 18, No. 1, pp. 38–48.
- Hardiman, O., A. Al-Chalabi, C. Brayne, E. Beghi, L. H. Van Den Berg, A. Chio, S. Martin, G. Logroscino, and J. Rooney, 2017, “The Changing Picture of Amyotrophic Lateral Sclerosis: Lessons from European Registers”, *Journal of Neurology, Neurosurgery and Psychiatry*, Vol. 88, No. 7, pp. 557–563.

- Harms, M. B., J. Cady, C. Zaidman *et al.*, 2013, “Lack of C9ORF72 Coding Mutations Supports a Gain of Function for Repeat Expansions in Amyotrophic Lateral Sclerosis”, *Neurobiology of Aging*, Vol. 34, No. 9, p. 2234.e13-2234.e19.
- Hawley, Z. C. E., D. Campos-Melo, C. A. Droppelmann, and M. J. Strong, 2017, “MotomiRs: MiRNAs in Motor Neuron Function and Disease”, *Frontiers in Molecular Neuroscience*, Vol. 10, No. 127, pp. 1–19.
- Hayward, C., D. J. Brock, R. A. Minns, and R. J. Swingler, 1998, “Homozygosity for Asn86Ser Mutation in the CuZn-Superoxide Dismutase Gene Produces a Severe Clinical Phenotype in a Juvenile Onset Case of Familial Amyotrophic Lateral Sclerosis.”, *Journal of Medical Genetics*, Vol. 35, No. 2, p. 174.
- Hood, L., and L. Rowen, 2013, “The Human Genome Project: Big Science Transforms Biology and Medicine”, *Genome Medicine*, Vol. 5, No. 9, p. 1.
- Hübers, A., W. Just, A. Rosenbohm *et al.*, 2015, “De Novo FUS Mutations Are the Most Frequent Genetic Cause in Early-Onset German ALS Patients”, *Neurobiology of Aging*, Vol. 36, No. 11, p. 3117.e1-3117.e6.
- Ingre, C., P. M. Roos, F. Piehl, F. Kamel, and F. Fang, 2015, “Risk Factors for Amyotrophic Lateral Sclerosis”, *Clinical Epidemiology*, Vol. 7, pp. 181–193.
- Ishiura, H., W. Sako, M. Yoshida *et al.*, 2012, “The TRK-Fused Gene Is Mutated in Hereditary Motor and Sensory Neuropathy with Proximal Dominant Involvement”, *American Journal of Human Genetics*, Vol. 91, No. 2, pp. 320–329.
- Iskender, C., E. Kartal, F. Akcimen, C. Kocoglu, D. Kotan, M. Eraksoy, Y. Parman, and A. N.

- Başak, 2015, “Turkish Families with Juvenile Motor Neuron Disease Broaden the Phenotypic Spectrum of SPG11”, *Neurology: Genetics*, Vol. 1, No. 3, p. e25.
- Jiang, J., Q. Zhu, T. F. Gendron *et al.*, 2016, “Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs”, *Neuron*, Vol. 90, No. 3, pp. 535–550.
- Johnson, J. O., J. Mandrioli, M. Benatar *et al.*, 2010, “Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS.”, *Neuron*, Vol. 68, No. 5, pp. 857–64.
- Kajihara, S., A. Hisatomi, T. Mizuta, T. Hara, I. Ozaki, I. Wada, and K. Yamamoto, 1998, “A Splice Mutation in the Human Canalicular Multispecific Organic Anion Transporter Gene Causes Dubin-Johnson Syndrome”, *Biochemical and Biophysical Research Communications*, Vol. 253, No. 2, pp. 454–457.
- Kent, J. W., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, Haussler, and D., 2002, “The Human Genome Browser at UCSC”, *Genome Research*, Vol. 12, No. 6, pp. 996–1006.
- Khani, M., A. Alavi, S. Nafissi, and E. Elahi, 2015, “Observation of c.260A > G Mutation in Superoxide Dismutase 1 That Causes p.Asn86Ser in Iranian Amyotrophic Lateral Sclerosis Patient and Absence of Genotype/Phenotype Correlation.”, *Iranian Journal of Neurology*, Vol. 14, No. 3, pp. 152–157.
- Khundadze, M., K. Kollmann, N. Koch *et al.*, 2013, “A Hereditary Spastic Paraplegia Mouse Model Supports a Role of ZFYVE26/SPASTIZIN for the Endolysosomal System”, *PLoS Genetics*, Vol. 9, No. 12, p. e1003988.

- Kiernan, M. C., S. Vucic, B. C. Cheah, M. R. Turner, A. Eisen, O. Hardiman, J. R. Burrell, and M. C. Zoing, 2011, “Amyotrophic Lateral Sclerosis”, *The Lancet*, Vol. 377, No. 9769, pp. 942–955.
- Kun-rodrigues, C., C. Ganos, R. Guerreiro *et al.*, 2015, “A Systematic Screening to Identify de Novo Mutations Causing Sporadic Early-Onset Parkinson’s Disease”, *Human Molecular Genetics*, Vol. 24, No. 23, pp. 6711–6720.
- Kwiatkowski, T. J., D. A. Bosco, A. L. Leclerc *et al.*, 2009, “Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis.”, *Science*, Vol. 323, No. 5918, pp. 1205–1208.
- Kwon, I., S. Xiang, M. Kato *et al.*, 2014, “Poly-Dipeptides Encoded by the C9ORF72 Repeats Bind Nucleoli, Impede RNA Biogenesis, and Kill Cells.”, *Science*, Vol. 345, No. 6201, p. 1254917.
- Landrum, M. J., J. M. Lee, G. R. Riley, W. Jang, W. S. Rubinstein, D. M. Church, and D. R. Maglott, 2014, “ClinVar: Public Archive of Relationships among Sequence Variation and Human Phenotype”, *Nucleic Acids Research*, Vol. 42, pp. D980-5.
- Lange, D. J., M. Shahbazi, V. Silani *et al.*, 2017, “Pyrimethamine Significantly Lowers Cerebrospinal Fluid Cu/Zn Superoxide Dismutase in Amyotrophic Lateral Sclerosis Patients with SOD1 Mutations”, *Annals of Neurology*, Vol. 81, No. 6, pp. 837–848.
- Lattante, S., S. Millecamps, G. Stevanin *et al.*, 2014, “Contribution of ATXN2 Intermediary PolyQ Expansions in a Spectrum of Neurodegenerative Disorders.”, *Neurology*, Vol. 83, No. 11, pp. 990–5.

- Leblond, C. S., A. Webber, Z. Gan-Or, F. Moore, A. Dagher, P. A. Dion, and G. A. Rouleau, 2016, “De Novo *FUS* P525L Mutation in Juvenile Amyotrophic Lateral Sclerosis with Dysphonia and Diplopia”, *Neurology Genetics*, Vol. 2, No. 2, p. e63.
- Lee, J. M., V. C. Wheeler, M. J. Chao *et al.*, 2015, “Identification of Genetic Factors That Modify Clinical Onset of Huntington’s Disease”, *Cell*, Vol. 162, No. 3, pp. 516–526.
- Lee, S. S., H. J. Lee, J. M. Park *et al.*, 2013a, “Proximal Dominant Hereditary Motor and Sensory Neuropathy with Proximal Dominance Association with Mutation in the TRK-Fused Gene”, *JAMA Neurology*, Vol. 70, No. 5, pp. 607–615.
- Lee, Y., H. Chen, J. N. Peres *et al.*, 2013b, “Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic”, *Cell Reports*, Vol. 5, No. 5, pp. 1178–1186.
- Leigh, P. N., B. H. Anderton, A. Dodson, J. M. Gallo, M. Swash, and D. M. Power, 1988, “Ubiquitin Deposits in Anterior Horn Cells in Motor Neurone Disease”, *Neuroscience Letters*, Vol. 93, No. 2–3, pp. 197–203.
- Lek, M., K. J. Karczewski, E. V. Minikel *et al.*, 2016, “Analysis of Protein-Coding Genetic Variation in 60,706 Humans”, *Nature*, Vol. 536, No. 7616, pp. 285–291.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, 2009, “The Sequence Alignment/Map Format and SAMtools”, *Bioinformatics*, Vol. 25, No. 16, pp. 2078–2079.
- Li, H., and R. Durbin, 2009, “Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform”, *Bioinformatics*, Vol. 25, No. 14, pp. 1754–1760.

- Liu, Y., J. T. Yu, Y. Zong, J. Zhou, and L. Tan, 2014, “C9ORF72 Mutations in Neurodegenerative Diseases”, *Molecular Neurobiology*, Vol. 49, No. 1, pp. 386–398.
- Lupo, V., C. Aguado, E. Knecht, and C. Espinós, 2016, “Chaperonopathies: Spotlight on Hereditary Motor Neuropathies”, *Frontiers in Molecular Biosciences*, Vol. 3, No. December, pp. 1–7.
- Majounie, E., A. E. Renton, K. Mok *et al.*, 2012, “Frequency of the C9orf72 Hexanucleotide Repeat Expansion in Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia: A Cross-Sectional Study”, *The Lancet Neurology*, Vol. 11, No. 4, pp. 323–330.
- Manjaly, Z. R., K. M. Scott, K. Abhinav *et al.*, 2010, “The Sex Ratio in Amyotrophic Lateral Sclerosis: A Population Based Study”, *Amyotrophic Lateral Sclerosis*, Vol. 11, No. 5, pp. 439–442.
- Martinez-Lage, M., L. Molina-Porcel, D. Falcone, L. Mccluskey, V. M. Lee, V. M. Van Deerlin, and J. Q. Trojanowski, 2012, “TDP-43 Pathology in a Case of Hereditary Spastic Paraplegia with a NIPA1 / SPG6 Mutation”, Vol. 124, No. 2, pp. 285–291.
- Maystadt, I., R. Rezsöhazy, M. Barkats *et al.*, 2007, “The Nuclear Factor KB–Activator Gene PLEKHG5 Is Mutated in a Form of Autosomal Recessive Lower Motor Neuron Disease with Childhood Onset”, *The American Journal of Human Genetics*, Vol. 81, No. 1, pp. 67–76.
- McKenna, A., M. Hanna, E. Banks *et al.*, 2010, “The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data”, *Genome Research*, Vol. 20, No. 9, pp. 1297–1303.

- Miller, T., A. Pestronk, W. David *et al.*, 2013, “A Phase I, Randomised, First-in-Human Study of an Antisense Oligonucleotide Directed Against SOD1 Delivered Intrathecally in SOD1-Familial ALS Patients”, *Lancet Neurol*, Vol. 12, No. 5, pp. 434–442.
- Mok, K., B. J. Traynor, J. Schymick *et al.*, 2012, “The Chromosome 9 ALS and FTD Locus Is Probably Derived from a Single Founder”, *Neurobiology of Aging*, Vol. 33, No. 1, p. 209.e3.
- Mori, K., S.-M. Weng, T. Arzberger *et al.*, 2013, “The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLN/ALS.”, *Science*, Vol. 339, No. 6125, pp. 1–13.
- Morita, M., a Al-Chalabi, P. M. Andersen *et al.*, 2006, “A Locus on Chromosome 9p Confers Susceptibility to ALS and Frontotemporal Dementia.”, *Neurology*, Vol. 66, No. 6, pp. 839–44.
- Nakhro, K., J. M. Park, Y. Bin Hong *et al.*, 2013, “SET Binding Factor 1 (SBF1) Mutation Causes Charcot-Marie-Tooth Disease Type 4B3”, *Neurology*, Vol. 81, No. 2, pp. 165–173.
- 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.
- Nguyen, H. P., C. Van Broeckhoven, and J. van der Zee, 2018, “ALS Genes in the Genomic Era and Their Implications for FTD”, *Trends in Genetics*, Vol. 34, No. 6, pp. 404–423.
- Nguyen, T. P., S. Biliciler, W. Wiszniewski, and K. Sheikh, 2015, “Expanding Phenotype of VRK1 Mutations in Motor Neuron Disease”, *Journal of Clinical Neuromuscular Disease*,

Vol. 17, No. 2, pp. 69–71.

- Nicolas, A., K. P. Kenna, A. E. Renton *et al.*, 2018, “Genome-Wide Analyses Identify KIF5A as a Novel ALS Gene”, *Neuron*, Vol. 97, No. 6, p. 1268–1283.e6.
- Nordin, a., C. Akimoto, a. Wuolikainen *et al.*, 2015, “Extensive Size Variability of the GGGGCC Expansion in C9orf72 in Both Neuronal and Non-Neuronal Tissues in 18 Patients with ALS or FTD”, *Human Molecular Genetics*, Vol. 24, No. 11, pp. 3133–3142.
- Novarino, G., A. G. Fenstermaker, M. S. Zaki *et al.*, 2014, “Exome Sequencing Links Corticospinal Motor Neuron Disease to Common Neurodegenerative Disorders.”, *Science*, Vol. 343, No. 6170, pp. 506–11.
- Nuytemans, K., G. Bademci, M. M. Kohli *et al.*, 2013, “C9orf72 Intermediate Repeat Copies Are a Significant Risk Factor for Parkinson Disease”, *Annals of Human Genetics*, Vol. 77, No. 5, pp. 351–363.
- Ogaki, K., Y. Li, N. Atsuta *et al.*, 2012, “Analysis of C9orf72 Repeat Expansion in 563 Japanese Patients with Amyotrophic Lateral Sclerosis”, *Neurobiology of Aging*, Vol. 33, No. 10, pp. 11–16.
- Olzmann, J. A., R. R. Kopito, and J. C. Christianson, 2013, “The Mammalian Endoplasmic Reticulum-Associated Degradation System”, *Cold Spring Harbor Perspective Biology*, Vol. 5, No. 9, p. a013185.
- Orlacchio, A., C. Babalini, A. Borreca *et al.*, 2010, “SPATACSIN Mutations Cause Autosomal Recessive Juvenile Amyotrophic Lateral Sclerosis”, *Brain*, Vol. 133, No. 2,

pp. 591–598.

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 *et al.*, 2015, “The Distinct Genetic Pattern of ALS in Turkey and Novel Mutations”, *Neurobiology of Aging*, Vol. 36, No. 4, p. 1764.e9-1764.e18.

Paez-Colasante, X., C. Figueroa-Romero, S. A. Sakowski, S. A. Goutman, and E. L. Feldman, 2015, “Amyotrophic Lateral Sclerosis: Mechanisms and Therapeutics in the Epigenomic Era”, *Nature Reviews Neurology*, Vol. 11, No. 5, pp. 266–279.

Parodi, L., S. Fenu, G. Stevanin, and A. Durr, 2017, “Hereditary Spastic Paraplegia : More than an Upper Motor Neuron Disease”, *Revue Neurologique*, pp. 1–9.

Pasinelli, P., and R. H. Brown, 2006, “Molecular Biology of Amyotrophic Lateral Sclerosis: Insights from Genetics.”, *Nature Reviews. Neuroscience*, Vol. 7, No. 9, pp. 710–23.

Pearson, C. E., 2011, “Repeat Associated Non-ATG Translation Initiation: One DNA, Two Transcripts, Seven Reading Frames, Potentially Nine Toxic Entities!”, *PLoS Genetics*, Vol. 7, No. 3, pp. 1–5.

Peters, O. M., M. Ghasemi, and R. H. Brown, 2015, “Emerging Mechanisms of Molecular Pathology in ALS.”, *The Journal of Clinical Investigation*, Vol. 125, No. 5, pp. 1767–79.

Petrov, D., C. Mansfield, A. Moussy, and O. Hermine, 2017, “ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment?”, *Frontiers in Aging Neuroscience*, Vol. 9, No. 68, pp. 1–11.

- Pranavchand, R., and B. M. Reddy, 2016, “Genomics Era and Complex Disorders: Implications of GWAS with Special Reference to Coronary Artery Disease, Type 2 Diabetes Mellitus, and Cancers.”, *Journal of Postgraduate Medicine*, Vol. 62, No. 3, pp. 188–198.
- Pupillo, E., E. Bianchi, M. Poloni *et al.*, 2017, “Is Firstly Diagnosed ALS Really ALS? Results of a Population-Based Study with Long-Term Follow-Up”, *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, Vol. 18, No. 3–4, pp. 221–226.
- Purcell, S., B. Neale, K. Todd-Brown *et al.*, 2007, “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses”, *The American Journal of Human Genetics*, Vol. 81, No. 3, pp. 559–575.
- Qureshi, I. a., and M. F. Mehler, 2011, “Advances in Epigenetics and Epigenomics for Neurodegenerative Diseases”, *Current Neurology and Neuroscience Reports*, Vol. 11, No. 5, pp. 464–473.
- R Core Team, 2017, “R: A Language and Environment for Statistical Computing.”, *R Foundation for Statistical Computing, Vienna, Austria.*, Vol. 3, No. 1, p. 201.
- Ramaswami, T. J. M., 2014, “Persistence in the Development of Degenerative Disorders”, *Cell*, Vol. 154, No. 4, pp. 1–17.
- Reddy, K., B. Zamiri, S. Y. R. Stanley, R. B. Macgregor, and C. E. Pearson, 2013, “The Disease-Associated r(GGGGCC)_n Repeat from the C9orf72 Gene Forms Tract Length-Dependent Uni- and Multimolecular RNA G-Quadruplex Structures”, *Journal of Biological Chemistry*, Vol. 288, No. 14, pp. 9860–9866.

- Renton, A. E., E. Majounie, A. Waite *et al.*, 2011, “A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD.”, *Neuron*, Vol. 72, No. 2, pp. 257–68.
- Renton, A. E., A. Chiò, and B. J. Traynor, 2013, “State of Play in Amyotrophic Lateral Sclerosis Genetics”, *Nature Neuroscience*, Vol. 17, No. 1, pp. 17–23.
- Richards, S., N. Aziz, S. Bale *et al.*, 2015, “Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology”, *Genetics in Medicine*, Vol. 17, No. 5, pp. 405–424.
- Robberecht, W., and T. Philips, 2013, “The Changing Scene of Amyotrophic Lateral Sclerosis”, *Nature Reviews Neuroscience*, Vol. 14, No. 4, pp. 248–264.
- Rohrer, J. D., A. M. Isaacs, S. Mizielinska *et al.*, 2015, “C9orf72 Expansions in Frontotemporal Dementia and Amyotrophic Lateral Sclerosis”, *The Lancet Neurology*, Vol. 14, No. 3, pp. 291–301.
- Rosen, D. R., T. Siddique, D. Patterson *et al.*, 1993, “Mutations in Cu/Zn Superoxide Dismutase Gene Are Associated with Familial Amyotrophic Lateral Sclerosis.”, *Nature*, Vol. 362, No. 6415, pp. 59–62.
- Rosen, D. R., A. C. Bowling, D. Patterson *et al.*, 1994, “A Frequent Ala-4 To Val Superoxide Dismutase-1 Mutation Is Associated With a Rapidly Progressive Familial Amyotrophic-Lateral-Sclerosis”, *Human Molecular Genetics*, Vol. 3, No. 6, pp. 981–987.
- Russ, J., E. Y. Liu, K. Wu *et al.*, 2015, “Hypermethylation of Repeat Expanded C9orf72 Is a

- Clinical and Molecular Disease Modifier”, *Acta Neuropathologica*, Vol. 129, No. 1, pp. 39–52.
- Rutherford, N. J., M. G. Heckman, M. DeJesus-Hernandez *et al.*, 2012, “Length of Normal Alleles of C9ORF72 GGGGCC Repeat Do Not Influence Disease Phenotype”, *Neurobiology of Aging*, Vol. 33, No. 12, p. 2950.e5-2950.e7.
- Saccon, R. A., R. K. A. Bunton-Stasyshyn, E. M. C. Fisher, and P. Fratta, 2013, “Is SOD1 Loss of Function Involved in Amyotrophic Lateral Sclerosis?”, *Brain*, Vol. 136, No. 8, pp. 2342–2358.
- Sailer, A., and H. Houlden, 2012, “Recent Advances in the Genetics of Cerebellar Ataxias”, *Current Neurology and Neuroscience Reports*, Vol. 12, pp. 227–236.
- Salameh, J. S., R. H. Brown, and J. D. Berry, 2015, “Amyotrophic Lateral Sclerosis: Review”, *Seminars in Neurology*, Vol. 35, No. 4, pp. 469–476.
- Sandi, C., M. Sandi, S. Anjomani Virmouni, S. Al-Mahdawi, and M. a Pook, 2014, “Epigenetic-Based Therapies for Friedreich Ataxia”, *Frontiers in Genetics*, Vol. 5, No. 165, pp. 1–12.
- Schottmann, G., C. Wagner, F. Seifert, W. Stenzel, and M. Schuelke, 2016, “MORC2 Mutation Causes Severe Spinal Muscular Atrophy-Phenotype, Cerebellar Atrophy, and Diaphragmatic Paralysis”, *Brain*, Vol. 139, No. 12, p. e70.
- Schwarze, K., J. Buchanan, J. C. Taylor, and S. Wordsworth, 2018, “Are Whole-Exome and Whole-Genome Sequencing Approaches Cost-Effective? A Systematic Review of the Literature”, *Genetics in Medicine*, Vol. 2, No. 7, pp. 32–43.

- Sevilla, T., V. Lupo, D. Martínez-Rubio *et al.*, 2016, “Mutations in the MORC2 Gene Cause Axonal Charcot-Marie-Tooth Disease”, *Brain*, Vol. 139, No. 1, pp. 62–72.
- Shahrizaila, N., G. Sobue, S. Kuwabara *et al.*, 2016, “Amyotrophic Lateral Sclerosis and Motor Neuron Syndromes in Asia”, *Journal of Neurology, Neurosurgery and Psychiatry*, Vol. 87, No. 8, pp. 821–830.
- Sherry, S. T., 2001, “DbSNP: The NCBI Database of Genetic Variation”, *Nucleic Acids Research*, Vol. 29, No. 1, pp. 308–311.
- Siddique, T., D. A. Figlewicz, M. A. Pericak-Vance *et al.*, 1991, “Linkage of a Gene Causing Familial Amyotrophic Lateral Sclerosis to Chromosome 21 and Evidence of Genetic-Locus Heterogeneity.”, *The New England Journal of Medicine*, Vol. 324, No. 20, pp. 1381–1384.
- Smith, B. N., S. Newhouse, A. Shatunov *et al.*, 2013, “The C9ORF72 Expansion Mutation Is a Common Cause of ALS+/-FTD in Europe and Has a Single Founder.”, *European Journal of Human Genetics*, Vol. 21, No. 1, pp. 102–8.
- Smith, B. N., S. D. Topp, C. Fallini *et al.*, 2017, “Mutations in the Vesicular Trafficking Protein Annexin A11 Are Associated with Amyotrophic Lateral Sclerosis”, *Science Translational Medicine*, Vol. 9, No. 388, p. eaad9157.
- Sreedharan, J., I. P. Blair, V. B. Tripathi *et al.*, 2008, “TDP-43 Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis”, *Science*, Vol. 319, No. 5870, pp. 1668–1672.
- Sreedharan, J., and R. H. Brown, 2013, “Amyotrophic Lateral Sclerosis: Problems and Prospects”, *Annals of Neurology*, Vol. 74, No. 3, pp. 309–316.

- Stoica, L., and M. Sena-Esteves, 2016, “Adeno Associated Viral Vector Delivered RNAi for Gene Therapy of SOD1 Amyotrophic Lateral Sclerosis”, *Frontiers in Molecular Neuroscience*, Vol. 9, No. 56, pp. 1–7.
- Stoll, M., H. Teoh, J. Lee *et al.*, 2016, “Novel Motor Phenotypes in Patients with VRK1 Mutations without Pontocerebellar Hypoplasia”, *Neurology*, Vol. 87, No. 1, pp. 65–70.
- Synofzik, M., D. Ronchi, I. Keskin *et al.*, 2012, “Mutant Superoxide Dismutase-1 Indistinguishable from Wild-Type Causes ALS”, *Human Molecular Genetics*, Vol. 21, No. 16, pp. 3568–3574.
- Synofzik, M., K. Smets, M. Mallaret *et al.*, 2016, “SYNE1 Ataxia Is a Common Recessive Ataxia with Major Non-Cerebellar Features: A Large Scale Multi-Centre Study”, *Brain*, Vol. 138, No. 5, pp. 1378–1393.
- Takahashi, Y., Y. Fukuda, J. Yoshimura *et al.*, 2013, “ERBB4 Mutations That Disrupt the Neuregulin-ErbB4 Pathway Cause Amyotrophic Lateral Sclerosis Type 19”, *American Journal of Human Genetics*, Vol. 93, No. 5, pp. 900–905.
- Talbot, K., 2011, “Familial versus Sporadic Amyotrophic Lateral Sclerosis--a False Dichotomy?”, *Brain*, Vol. 134, No. 12, pp. 3429–3434.
- Taylor, J. P., R. H. Brown, and D. W. Cleveland, 2016, “Decoding ALS: From Genes to Mechanism”, *Nature*, Vol. 539, No. 7628, pp. 197–206.
- Teer, J. K., E. D. Green, J. C. Mullikin, and L. G. Biesecker, 2012, “VarSifter: Visualizing and Analyzing Exome-Scale Sequence Variation Data on a Desktop Computer”, *Bioinformatics*, Vol. 28, No. 4, pp. 599–600.

- Therrien, M., P. A. Dion, and G. A. Rouleau, 2016, “ALS: Recent Developments from Genetics Studies”, *Current Neurology and Neuroscience Reports*, Vol. 16, No. 6.
- Ticozzi, N., C. Tiloca, C. Morelli *et al.*, 2011, “Genetics of Familial Amyotrophic Lateral Sclerosis.”, *Archives Italiennes de Biologie*, Vol. 149, No. 1, pp. 65–82.
- Todd, P. K., and H. L. Paulson, 2010, “RNA Mediated Neurodegeneration in Repeat Expansion Disorders”, *Annals of Neurology*, Vol. 67, No. 3, pp. 291–300.
- Tremolizzo, L., P. Messina, E. Conti *et al.*, 2014, “Whole-Blood Global DNA Methylation Is Increased in Amyotrophic Lateral Sclerosis Independently of Age of Onset.”, *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, Vol. 15, No. 1–2, pp. 98–105.
- Trujillano, D., A. M. Bertoli-Avella, K. Kumar Kandaswamy *et al.*, 2017, “Clinical Exome Sequencing: Results from 2819 Samples Reflecting 1000 Families”, *European Journal of Human Genetics*, Vol. 25, No. 2, pp. 176–182.
- Tunca, C., F. Akçimen, C. Coşkun, A. Gündoğdu-Eken, C. Kocoglu, B. Çevik, C. E. Bekircan-Kurt, E. Tan, and A. N. Başak, 2018, “ERLIN1 Mutations Cause Teenage-Onset Slowly Progressive ALS in a Large Turkish Pedigree”, *European Journal of Human Genetics*, Vol. 26, No. 5, pp. 745–748.
- Turner, M. R., A. Al-Chalabi, A. Chio, O. Hardiman, M. C. Kiernan, J. D. Rohrer, J. Rowe, W. Seeley, and K. Talbot, 2017, “Genetic Screening in Sporadic ALS and FTD”, *Journal of Neurology, Neurosurgery and Psychiatry*, Vol. 88, No. 12, pp. 1042–1044.
- van Blitterswijk, M., L. Vlam, M. van Es *et al.*, 2012a, “Genetic Overlap between Apparently

- Sporadic Motor Neuron Diseases.”, *PloS One*, Vol. 7, No. 11, p. e48983.
- van Blitterswijk, M., M. van Es, E. M. Hennekam *et al.*, 2012b, “Evidence for an Oligogenic Basis of Amyotrophic Lateral Sclerosis”, *Human Molecular Genetics*, Vol. 21, No. 17, pp. 3776–3784.
- van Blitterswijk, M., M. DeJesus-Hernandez, E. Niemantsverdriet *et al.*, 2013, “Association between Repeat Sizes and Clinical and Pathological Characteristics in Carriers of C9ORF72 Repeat Expansions (Xpansize-72): A Cross-Sectional Cohort Study”, *The Lancet Neurology*, Vol. 12, No. 10, pp. 978–988.
- van Blitterswijk, M., B. Mullen, A. M. Nicholson *et al.*, 2014a, “TMEM106B Protects C9ORF72 Expansion Carriers against Frontotemporal Dementia”, *Acta Neuropathologica*, Vol. 127, No. 3, pp. 397–406.
- van Blitterswijk, M., B. Mullen, A. Wojtas *et al.*, 2014b, “Genetic Modifiers in Carriers of Repeat Expansions in the C9ORF72 Gene.”, *Molecular Neurodegeneration*, Vol. 9, No. 1, p. 38.
- van Es, M. A., J. H. Veldink, C. G. J. Saris *et al.*, 2009, “Genome-Wide Association Study Identifies 19p13.3 (UNC13A) and 9p21.2 as Susceptibility Loci for Sporadic Amyotrophic Lateral Sclerosis”, *Nature Genetics*, Vol. 41, No. 10, pp. 1083–1087.
- van Es, M. A., O. Hardiman, A. Chio, A. Al-Chalabi, R. J. Pasterkamp, J. H. Veldink, and L. H. van den Berg, 2017, “Amyotrophic Lateral Sclerosis”, *The Lancet*, Vol. 4, No. 390, pp. 2084–2098.
- van Hoecke, A., L. Schoonaert, R. Lemmens *et al.*, 2012, “EPHA4 Is a Disease Modifier of

- Amyotrophic Lateral Sclerosis in Animal Models and in Humans”, *Nature Medicine*, Vol. 18, No. 9, pp. 1418–1422.
- van Rheenen, W., A. Shatunov, A. M. Dekker *et al.*, 2016, “Genome-Wide Association Analyses Identify New Risk Variants and the Genetic Architecture of Amyotrophic Lateral Sclerosis”, *Nature Genetics*, Vol. 48, No. 9, pp. 1043–1048.
- van Rheenen, W., S. L. Pulit, A. M. Dekker *et al.*, 2017, “Project MinE: Study Design and Pilot Analyses of a Large-Scale Whole-Genome Sequencing Study in Amyotrophic Lateral Sclerosis”, *BioRxiv*, Vol. 26, No. 10, pp. 1537–1546.
- Vandoorne, T., K. De Bock, and L. Van Den Bosch, 2018, “Energy Metabolism in ALS: An Underappreciated Opportunity?”, *Acta Neuropathologica*, Vol. 135, No. 4, pp. 489–509.
- 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. 1–7.
- Williams, K. L., J. Fifita, S. Vucic, J. C. Durnall, M. C. Kiernan, I. P. Blair, and G. a Nicholson, 2013, “Pathophysiological Insights into ALS with C9ORF72 Expansions.”, *Journal of Neurology, Neurosurgery & Psychiatry*, Vol. 84, No. 8, pp. 931–935.
- Wu, C., C. Fallini, N. Ticozzi *et al.*, 2012, “Mutations in the Profilin 1 Gene Cause Familial Amyotrophic Lateral Sclerosis”, *Nature*, Vol. 488, No. 7412, pp. 499–503.
- Xi, Z., L. Zinman, D. Moreno *et al.*, 2013, “Hypermethylation of the CpG Island Near the G4C2 Repeat in ALS with a C9orf72 Expansion”, *The American Journal of Human Genetics*, Vol. 92, No. 6, pp. 981–989.

- Xi, Z., M. van Blitterswijk, M. Zhang *et al.*, 2015, “Jump from Pre-Mutation to Pathologic Expansion in C9orf72”, *American Journal of Human Genetics*, Vol. 96, No. 6, pp. 962–970.
- Yang, Y., A. Hentati, H. X. Deng *et al.*, 2001, “The Gene Encoding Alsin, a Protein with Three Guanine-Nucleotide Exchange Factor Domains, Is Mutated in a Form of Recessive Amyotrophic Lateral Sclerosis.”, *Nature Genetics*, Vol. 29, No. 2, pp. 160–165.
- Zaldivar, T., J. Gutierrez, G. Lara, M. Carbonara, G. Logroscino, and O. Hardiman, 2009, “Reduced Frequency of ALS in an Ethnically Mixed Population: A Population-Based Mortality Study”, *Neurology*, Vol. 72, No. 19, pp. 1640–1645.
- Zhan, L., K. A. Hanson, S. H. Kim, A. Tare, and R. S. Tibbetts, 2013, “Identification of Genetic Modifiers of TDP-43 Neurotoxicity in *Drosophila*”, *PLoS ONE*, Vol. 8, No. 2, p. e57214.
- Zhang, M., M. Carmela, D. Moreno *et al.*, 2017, “DNA Methylation Age - Acceleration Is Associated with Disease Duration and Age at Onset in C9orf72 Patients”, *Acta Neuropathologica*, Vol. 134, No. 2, pp. 271–279.
- Zhang, Q., C. Bethmann, N. F. Worth *et al.*, 2007, “Nesprin-1 and -2 Are Involved in the Pathogenesis of Emery - Dreifuss Muscular Dystrophy and Are Critical for Nuclear Envelope Integrity”, *Human Molecular Genetics*, Vol. 16, No. 23, pp. 2816–2833.
- Zu, T., B. Gibbens, N. S. Doty, M. Gomes-pereira, A. Huguet, and M. D. Stone, 2010, “Non-ATG – Initiated Translation Directed by Microsatellite Expansions”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 1, pp. 260–265.

Zu, T., Y. Liu, M. Bañez-Coronel *et al.*, 2013, “RAN Proteins and RNA Foci from Antisense Transcripts in C9ORF72 ALS and Frontotemporal Dementia.”, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 110, No. 51, pp. E4968-77.

APPENDIX A: NUMBER OF VARIANTS REMAINING IN VARIANT FILTRATION STEPS

Table A1. Remaining variations after each filtration step.

Family #	Index ID	Total Variants	Exonic Variants	Inheritance Pattern	Family Members	MAF 0.01 1000G and ESP6500	# of samples
Family 1	ALS1539	350042	11372	6907	751	133	3
Family 2	ALS1381	342046	11335	7044	-	1029	1
Family 3	ALS1610	332650	11464	7045	-	1075	1
Family 6	HSP29	281078	10856	4773	-	283	1
Family 7	HSP33	248737	11241	4600	-	269	1
Family 10	ALS518	182462	10026	3888	21	2	6
Family 11	ALS1164	334630	11383	4529	-	254	1
Family 13	ALS1594	327113	11146	4759	-	269	1
Family 14	ALS445	329037	11135	4472	-	256	1
Family 15	ALS1716	339418	11542	4337	-	238	1
Family 16	HSP34	139991	10333	4628	-	242	1
Family 17	ALS1813	110161	4544	1860	-	133	1
Family 18	ALS1704	303501	11263	4308	-	217	1
Family 19	ALS731	312632	11702	7298	-	1080	1
Family 20	ALS1708	337187	11505	7076	-	1053	1
Family 21	ALS1538	237531	7281	2765	-	171	1
Family 22	ALS1481	332461	11136	4708	-	335	1
Family 23	ALS1446	292355	11548	4575	-	227	1
Family 24	ALS1560	323672	10958	4306	-	221	1
Family 25	ALS1482	197757	9763	4450	688	29	3
Family 26	ALS1549	306082	10916	4397	-	238	1
Family 27	ALS1574	301391	11166	4730	-	270	1
Family 28	ALS1292	296912	11261	4521	-	264	1
Family 29	ALS711	269509	11031	4459	-	247	1
Family 30	ALS1132	264661	11291	7021	-	1014	1
Family 31	ALS1683	274734	10951	4605	-	308	1

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

Table B1. List of alignment commands.

bwa aln -t 32 -f sampleID.R1.sai referencegenome sampleID.R1.fastq.gz
bwa aln -t 32 -f sampleID.R2.sai referencegenome sampleID.R2.fastq.gz
bwa sampe -f sampleID.sam -r "@RG" referencegenome sampleID.R1.sai sampleID.R2.sai sampleID.R1.fastq.gz sampleID.R2.fastq.gz
samtools view -bS sampleID.sam > sampleID.bam
samtools sort sampleID.bam sampleID.sorted
samtools rmdup -sS sampleID.sorted.bam sampleID.rmdup.bam
samtools index sampleID.rmdup.bam

Table B2. List of variant calling commands.

java -jar GenomeAnalysisTK -T RealignerTargetCreator -R referencegenome -I sampleID.rmdup.bam -o sampleID.rmdup.bam.intervals -nt 3 -known Mills_and_1000G_gold_standard.indels.b37.vcf -known 1000G_phase1.indels.b37.vcf
java -jar GenomeAnalysisTK -T IndelRealigner -targetIntervals sampleID.rmdup.bam.intervals -R referencegenome -I sampleID.rmdup.bam -known Mills_and_1000G_gold_standard.indels.b37.vcf -known 1000G_phase1.indels.b37.vcf -o \$sampleID.realigned.bam
java -jar GenomeAnalysisTK -T BaseRecalibrator -I sampleID.realigned.bam -R referencegenome -knownSites dbsnp_138.b37.vcf -nct 4 -o sampleID.report.grp -lqt 2 - mdq -l
java -jar GenomeAnalysisTK -T PrintReads -R reference -I sampleID.realigned.bam -nct 4 -BQSR sampleID.report.grp -o sampleID.final.bam
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
java -jar GenomeAnalysisTK -T GenotypeGVCFs -R referencegenome --variant sampleID.raw.snps.indels.g.vcf -o sampleID.raw.snps.indels.vcf

Table B2. List of variant calling commands (cont.).

java -jar GenomeAnalysisTK -T VariantAnnotator -R referencegenome -o sampleID.ann.snp.indel.vcf -A Coverage -A InbreedingCoeff --variant sampleID.raw.snps.indels.vcf -L sampleID.raw.snps.indels.vcf --dbsnp dbsnp_138.b37.vcf
java -jar GenomeAnalysisTK -T VariantRecalibrator -R 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.
java -jar GenomeAnalysisTK -T ApplyRecalibration -R referencegenome -input sampleID.ann.snp.indel.vcf -- ts_filter_level 99.0 -recalFile sampleID.snp.recal - tranchesFile sampleID.snp.tranches -mode SNP -o sampleID.snp.vqsr.vcf
java -jar GenomeAnalysisTK -T VariantRecalibrator -R 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
java -jar GenomeAnalysisTK -T ApplyRecalibration -R referencegenome --input sampleID.snp.vqsr.vcf -mode INDEL --ts_filter_level 99.0 -recalFile sampleID.indel.recal -tranchesFile sampleID.indel.tranches -o sampleID.snp.indel.vqsr.vcf

Table B3. List of annotation commands.

perl convert2annovar.pl sampleID.vcf -format vcf4old -includeinfo > sampleID.anno
perl summarize_annoar.pl -out sampleID.annotated -buildver hg19 -verdb SNP 138 -ver1000g 1000g2012apr - veresp 6500 -remove -alltranscript sampleID.anno humandb
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
python wannotate_pinar.py sampleID.final.vcf sampleID.annotated.genome_summary.txt sampleID.final

APPENDIX C: PUBLICATIONS DURING THE THESIS

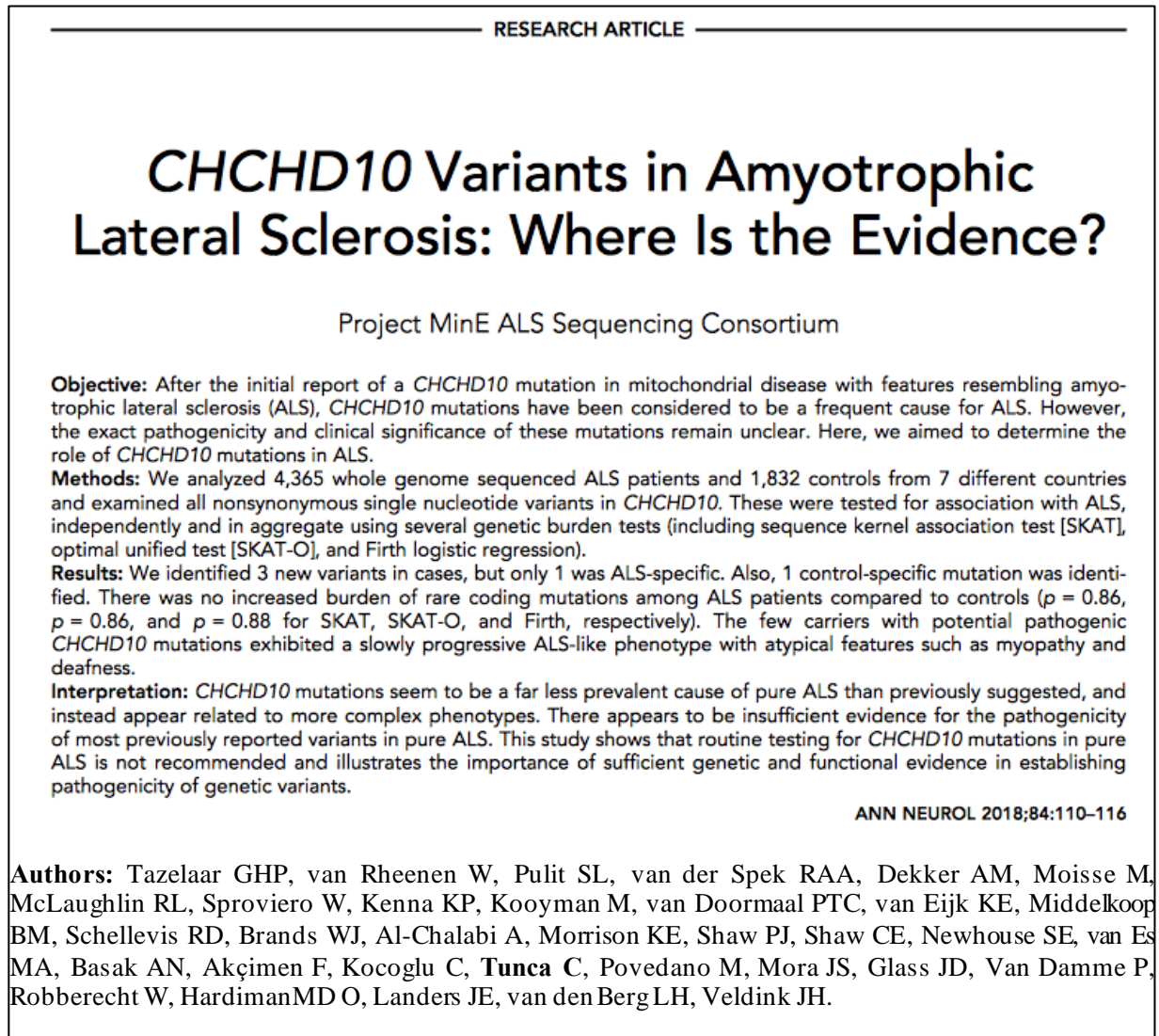


Figure C.1. Publication 1.

A new splice-site mutation in *SLC12A6* causing Andermann syndrome with motor neuropathy

INTRODUCTION

Solute carrier family 12 member 6 (*SLC12A6*) mutations have been found to be responsible for the phenotype that causes agenesis of the corpus callosum (CC) with peripheral neuropathy, a condition which is also known as Andermann syndrome (AS). AS is an autosomal-recessive disease with syndromic features that has rarely been reported anywhere in the world other than Quebec, Canada.¹ Hereditary neuropathies (HN) and complicated hereditary spastic paraparesis (HSP) call for differential diagnosis, since AS has both neurodevelopmental and neurodegenerative aspects affecting both sensory and motor fibres.^{1 2} This study presents an AS family who has motor neuropathy without sensory involvement due to a new *SLC12A6* mutation.

Nihan Hande Akçakaya,^{1,2} Zühal Yapıcı,³
Ceren İskender Tunca,⁴ Pınar Tektürk,³
Fulya Akçimen,⁴ Ayşe Nazlı Başak⁴

¹Department of Genetics, Istanbul University, Aziz Sancar Institute of Experimental Medicine, Istanbul, Turkey

²Spastic Children's Foundation of Turkey, Istanbul, Turkey

³Division of Child Neurology, Department of Neurology, Istanbul University, Istanbul Faculty of Medicine, Istanbul, Turkey

⁴MBG Department, Suna and İnan Kırac Foundation, Bogazici University, NDAL, Istanbul, Turkey

Correspondence to Dr Nihan Hande Akçakaya, Department of Genetics, Istanbul University, Aziz Sancar Institute of Experimental Medicine, Istanbul 34093, Turkey; nhakcakaya@gmail.com

Acknowledgements The authors thank Suna and İnan Kırac Foundation for their generous gift and Bogazici University for its support.

Contributors NHA and ZY: substantially contributed to design of the manuscript. NHA, ZY and ANB gave

J Neurol Neurosurg Psychiatry October 2018 Vol 89 No 10

Figure C.2. Publication 2.

Elevated Global DNA Methylation Is Not Exclusive to Amyotrophic Lateral Sclerosis and Is Also Observed in Spinocerebellar Ataxia Types 1 and 2

Hamid Hamzeiy^a Doruk Savaş^a Ceren Tunca^a Nesli Ece Şen^a
Aslı Gündoğdu Eken^a Irmak Şahbaz^a Daniela Calini^b Cinzia Tiloca^b
Nicola Ticozzi^{b, c} Antonia Ratti^{b, c} Vincenzo Silani^{b, c} A. Nazlı Başak^a

^aSuna and İnan Kırac Foundation, Neurodegeneration Research Laboratory (NDAL), Molecular Biology and Genetics Department, Boğaziçi University, Istanbul, Turkey; ^bDepartment of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan, Italy; ^cDepartment of Pathophysiology and Transplantation, “Dino Ferrari” Center, Università degli Studi di Milano, Milan, Italy

Keywords

Amyotrophic lateral sclerosis · Trinucleotide repeat disorder · Spinocerebellar ataxia · ELISA · Global DNA methylation · 5-Methylcytosine

Abstract

Adult-onset neurological disorders are caused and influenced by a multitude of different factors, including epigenetic modifications. Here, using an ELISA kit selected upon careful testing, we investigated global 5-methylcytosine (5-mC) levels in sporadic and familial amyotrophic lateral sclerosis (sALS and fALS), spinocerebellar ataxia types 1 and 2 (SCA1 and SCA2), Huntington's disease, Friedreich's ataxia, and myotonic dystrophy type 1. We report a significant elevation in global 5-mC levels of about 2–7% on average for sALS ($p < 0.01$ [$F(1, 243) = 9.159$, $p = 0.0027$]) and various forms of fALS along with SCA1 ($p < 0.01$ [$F(1, 83) = 11.285$, $p = 0.0012$]) and SCA2 ($p < 0.001$ [$F(1, 122) = 29.996$, $p = 0.0001$]) when compared to age- and sex-matched healthy controls. *C9orf72* expansion carrier ALS patients exhibit the highest global 5-mC levels along with *C9orf72* promoter hy-

permethylation. We failed to measure global 5-hydroxymethylcytosine (5-hmC) levels in blood, probably due to the very low levels of 5-hmC and the limitations of the commercially available ELISA kits. Our results point towards a role for epigenetics modification in ALS, SCA1, and SCA2, and help conclude a dispute on the global 5-mC levels in sALS blood.

© 2018 S. Karger AG, Basel

Introduction

Epigenetic mechanisms that regulate both coding and noncoding RNA transcription range from DNA methylation to various histone modifications, differential nucleosomal positioning and modifications at the RNA level [1]. One such epigenetic modification is the well-characterized promoter methylation at CpG sites, carried out by DNA methyltransferases (DNMTs) [2]. In addition, another group of enzymes, namely the ten-eleven translocation methylcytosine dioxygenase 1, are responsible for converting 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), a relatively newly identified DNA

Figure C.3. Publication 3.

BRIEF COMMUNICATION



ERLIN1 mutations cause teenage-onset slowly progressive ALS in a large Turkish pedigree

Ceren Tunca¹ · Fulya Akçimen¹ · Cemre Coşkun¹ · Aslı Gündoğdu-Eken¹ · Cemile Kocoglu¹ · Betül Çevik² · Can Ebru Bekircan-Kurt³ · Ersin Tan³ · A. Nazlı Başak¹

Received: 5 October 2017 / Revised: 26 December 2017 / Accepted: 5 January 2018
 © European Society of Human Genetics 2018



Abstract

Amyotrophic lateral sclerosis (ALS) is a late-onset motor neuron disease with mostly dominant inheritance and a life expectancy of 2–5 years; however, a quite common occurrence of atypical forms of the disease, due to recessive inheritance, has become evident with the use of NGS technologies. In this paper, we describe a family with close consanguinity for at least four generations, suffering from a slowly progressive form of ALS. Spastic walking is observed since teenage years, while bulbar symptoms start much later, at the fifth or sixth decade of life. Patients usually die because of respiratory failure. Using whole-exome sequencing, we identified a novel homozygous p.(Val94Ala) (c.281T>C) (NG_052910.1) (NM_006459) variation in the *endoplasmic reticulum lipid raft associated protein 1* (*ERLIN1*) gene, which segregates with the disease in the family. Here we suggest that *ERLIN1* variants, previously shown in juvenile hereditary spastic paraplegia cases, may also be the cause of a slowly progressive early-onset ALS, starting with upper motor neuron features and developing into classical ALS with the addition of lower motor neuron dysfunction. We also demonstrate that *ATP-binding cassette subfamily C member 2* (*ABCC2*) gene, responsible for hyperbilirubinemia, is linked to *ERLIN1*.

Figure C.4. Publication 4.

Medical Genetics

Clinical and molecular characterization and response to acitretin in three families with Sjögren-Larsson syndrome

Seçil Vural^{1,2}, MD , Atay Vural^{2,3}, MD, PhD, Fulya Akçimen⁴, BSc , Işın S. Bağcı¹, MD, Ceren Tunca⁴, MSc, Asli Gündoğdu Eken⁴, MSc, Thomas Ruzicka¹, MD, PhD, and A. Nazli Başak⁴, PhD

¹Ludwig Maximilians University, Dermatology and Allergology, Munich, Germany, ²Adıyaman Besni State Hospital, Adıyaman, Turkey, ³Department of Neurology, Koç University, Istanbul, Turkey, and ⁴Suna and İnan Kırac Foundation, Neurodegeneration Research Laboratory (NDAL), Boğaziçi University, Istanbul, Turkey

Correspondence

Seçil Vural, MD
Department of Dermatology and Allergology, Ludwig Maximilians University, Munich
Germany
E-mails: secil.vural@med.uni-muenchen.de; secilsaral@gmail.com

Conflict of interest: None

Abstract

Introduction Sjögren-Larsson syndrome (SLS) is a rare congenital disorder characterized by the triad of ichthyosis, spasticity, and mental retardation. Patients are usually referred to dermatology clinics during infancy. As paraplegia becomes the most debilitating symptom of the disease within a few years, ichthyosis, although a major burden for the patient, takes a back seat. Optimum treatment of ichthyosis in these children and the effect of treatment on different aspects such as severity of the ichthyosis, pruritus, or quality of life of the patients' and their caregivers is not well established.

Materials and Methods Genetic background of eight patients from three families diagnosed clinically with SLS was determined with whole-exome and Sanger sequencing. Clinical phenotypes, laboratory findings, magnetic resonance imaging (MRI), and treatment of the ichthyosis with acitretin were assessed.

Results All patients had the classical triad of Sjögren-Larsson syndrome. Genetic analysis revealed that one patient had a novel c.799-1 (+/-) homozygous splicing mutation in the ALDH3A2 gene. Other patients had the c.683G>A p.R228H (NM_000382.2) mutation in the same gene. Other manifestations included skeletal anomalies, enamel hypoplasia, bilateral T2-hyperintensities in white matter, and moderate-severe pruritus. Acitretin treatment in a maintenance dose of 0.25 mg/kg/day decreased the severity of ichthyosis in all children. It increased quality of life significantly in all of the children and their caregivers.

Conclusion We conclude that ichthyosis can be treated effectively with low-dose acitretin in children with Sjögren-Larsson syndrome, and this treatment is associated with a significant improvement in the quality of life.

Figure C.5. Publication 5.



Figure C.6. Publication 6.

Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis

To elucidate the genetic architecture of amyotrophic lateral sclerosis (ALS) and find associated loci, we assembled a custom imputation reference panel from whole-genome-sequenced patients with ALS and matched controls ($n = 1,861$). Through imputation and mixed-model association analysis in 12,577 cases and 23,475 controls, combined with 2,579 cases and 2,767 controls in an independent replication cohort, we fine-mapped a new risk locus on chromosome 21 and identified *C21orf2* as a gene associated with ALS risk. In addition, we identified *MOBP* and *SCFD1* as new associated risk loci. We established evidence of ALS being a complex genetic trait with a polygenic architecture. Furthermore, we estimated the SNP-based heritability at 8.5%, with a distinct and important role for low-frequency variants (frequency 1–10%). This study motivates the interrogation of larger samples with full genome coverage to identify rare causal variants that underpin ALS risk.

18,741,510 single-nucleotide variants (SNVs). Imputing this custom reference panel into Dutch ALS cases considerably increased the imputation accuracy for low-frequency variants (minor allele frequency (MAF) = 0.5–10%) in comparison to commonly used reference panels from 1000 Genomes Project Phase 1 (ref. 9) and Genome of the Netherlands¹⁰ (Fig. 1a). Improvement was also observed when imputing into ALS cases from the UK (Fig. 1b). To benefit from the global diversity of haplotypes, the custom and 1000 Genomes Project panels were combined, which further improved imputation. Given these results, we used the merged reference panel to impute all strata in our study.

In total, we imputed 8,697,640 variants passing quality control into the 27 strata and tested the strata separately for association with ALS risk by logistic regression. We then included the results in an inverse-variance-weighted, fixed-effects meta-analysis, which identified four loci associated at genome-wide significance ($P < 5 \times 10^{-8}$) (Fig. 2a). The previously reported *C9orf72* (rs3849943)^{3–5,8}, *UNC13A*

Authors: van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, van der Spek RA, Vösa U, de Jong S, Robinson MR, Yang J, Fogh I, van Doormaal PT, Tazelaar GH, Koppers M, Blokhuis AM, Sproviero W, Jones AR, Kenna KP, van Eijk KR, Harschnitz O, Schellevis RD, Brands WJ, Medic J, Menelaou A, Vajda A, Ticozzi N, Lin K, Rogelj B, Vrabec K, Ravnik-Glavac M, Koritnik B, Zidar J, Leonardis L, Grošelj LD, Millicamps S, Salachas F, Meininger V, de Carvalho M, Pinto S, Mora JS, Rojas-García R, Polak M, Chandran S, Colville S, Swingle R, Morrison KE, Shaw PJ, Hardy J, Orrell RW, Pittman A, Sidle K, Fratta P, Malaspina A, Topp S, Petri S, Abdulla S, Drepper C, Sendtner M, Meyer T, Ophoff RA, Staats KA, Wiedau-Pazos M, Lomen-Hoerth C, Van Deerlin VM, Trojanowski JQ, Elman L, McCluskey L, Basak AN, **Tunca C**, Hamzeiy H, Parman Y, Meitinger T, Lichtner P, Radivojkovic-Blagojevic M, Andres CR, Maurel C, Bensimon G, Landwehrmeyer B, Brice A, Payan CA, Saker-Delye S, Dürr A, Wood NW, Tittmann L, Lieb W, Franke A, Rietschel M, Cichon S, Nöthen MM, Amouyel P, Tzourio C, Dartigues JF, Uitterlinden AG, Rivadeneira F, Estrada K, Hofman A, Curtis C, Blauw HM, van der Kooij AJ, de Visser M, Goris A, Weber M, Shaw CE, Smith BN, Pansarasa O, Cereda C, Del Bo R, Comi GP, D'Alfonso S, Bertolin C, Sorarù G, Mazzini L, Pensato V, Gellera C, Tiloca C, Ratti A, Calvo A, Moglia C, Brunetti M, Arcuti S, Capozzo R, Zecca C, Lunetta C, Penco S, Riva N, Padovani A, Filosto M, Müller B, Stuit RJ; PARALS Registry; SLALOM Group; SLAP Registry; FALS Sequencing Consortium; SLAGEN Consortium; NNIPPS Study Group, Blair I, Zhang K, McCann EP, Fifita JA, Nicholson GA, Rowe DB, Pamphlett R, Kiernan MC, Grosskreutz J, Witte OW, Ringer T, Prell T, Stubendorff B, Kurth I, Hübner CA, Leigh PN, Casale F, Chio A, Beghi E, Pupillo E, Tortelli R, Logroscino G, Powell J, Ludolph AC, Weishaupt JH, Robberecht W, Van Damme P, Franke L, Pers TH, Brown RH, Glass JD, Landers JE, Hardiman O, Andersen PM, Corcia P, Vourc'h P, Silani V, Wray NR, Visscher PM, de Bakker P, van Es MA1, Pasterkamp RJ, Lewis CM, Breen G, Al-Chalabi A, van den Berg LH, Veldink JH.

Figure C.7. Publication 7.

Turkish families with juvenile motor neuron disease broaden the phenotypic spectrum of *SPG11*

OPEN

Ceren Iskender, MSc*
Ece Kartal, MSc*
Fulya Akcimen, BSc
Cemile Kocoglu, BSc
Aslihan Ozoguz, PhD
Dilcan Kotan, MD
Mefkure Eraksoy, MD
Yesim G. Parman, MD
Ayse Nazli Basak, PhD

Correspondence to
Dr. Basak:
basak@boun.edu.tr

ABSTRACT

Objective: Identification of causative mutations in 3 consanguineous families (with 4 affected members) referred to our center with young-onset motor neuron disease and overlapping phenotypes resembling autosomal recessive juvenile amyotrophic lateral sclerosis (ARJALS) and autosomal recessive hereditary spastic paraplegia (ARHSP).

Methods: Patients have a slowly progressive motor neuron disease with upper and lower motor neuron dysfunction. There is distal muscle weakness and atrophy associated with pyramidal signs. Whole-exome sequencing was performed on the patients and the unaffected parent samples to identify disease-causing mutations. Variants were prioritized according to their predicted pathogenicity and their relevance to the clinical phenotypes.

Results: Five distinct homozygous mutations within the *SPG11* gene were identified, 3 of which were novel and truncating: c.7155T>G/p.Tyr2385Ter, c.2250delT/p.Phe750Leufs*3, and c.1966_1967delAA/p.Lys656Valfs*11. The copresence of 2 distinct homozygous missense variations was observed in 2 families: c.6224A>G/p.Asn2075Ser and c.7132T>C/p.Phe2378Leu. The segregation of these variations in the family members was validated by Sanger sequencing.

Conclusions: Four patients with juvenile-onset motor neuron disease with consanguineous parents were found to carry homozygous mutations in the *SPG11* gene. Our findings confirm the overlapping phenotypes of *SPG11*-based ARJALS and ARHSP, indicating that these 2 entities may be the extreme phenotypes of the same disease continuum with many common features. This, in turn, confirms the difficult differential diagnosis of these 2 diseases in the clinic. *Neurol Genet* 2015;1:e25; doi: 10.1212/NXG.0000000000000025

Figure C.8. Publication 8.

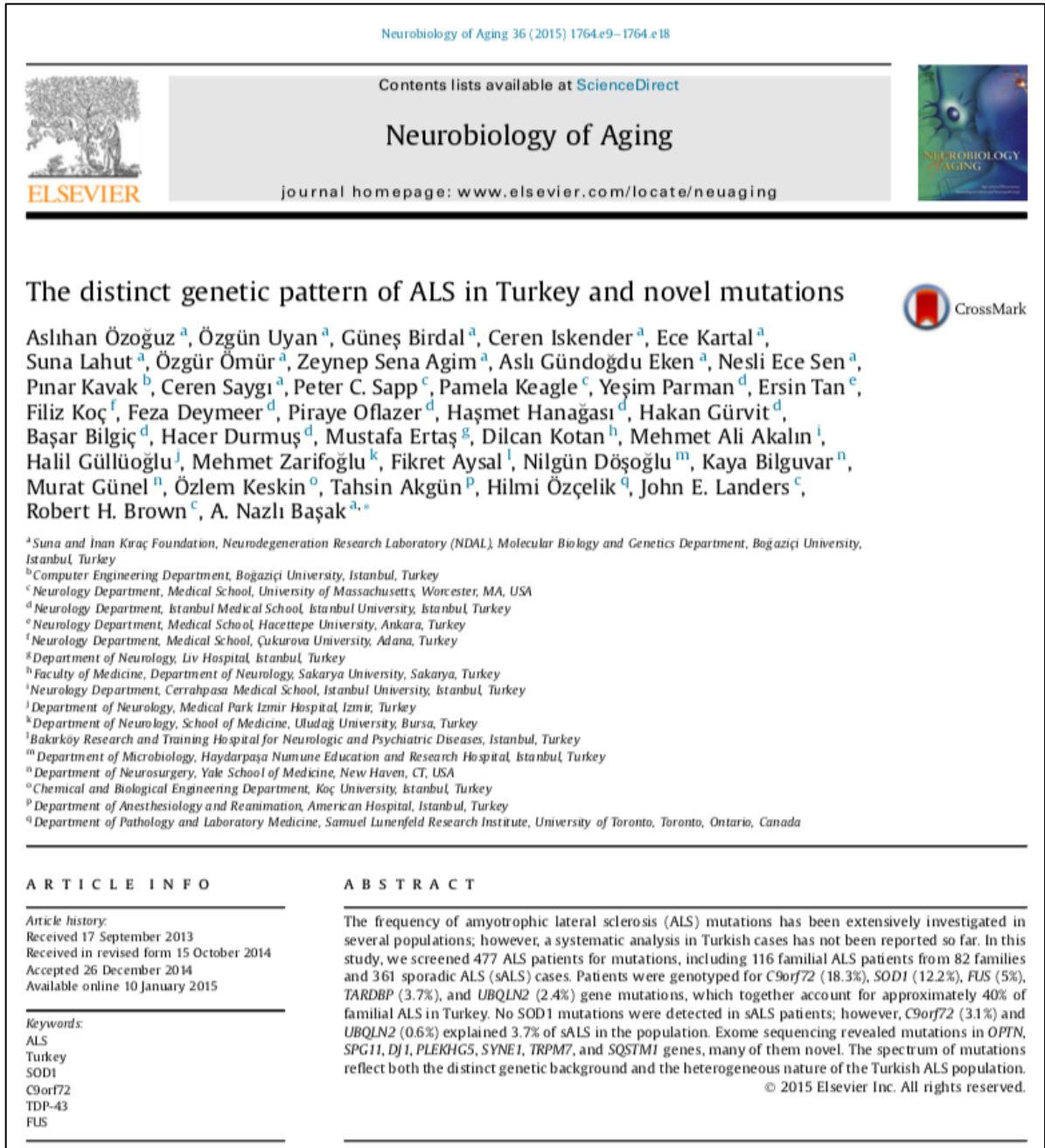


Figure C.9. Publication 9.