THE GENETICS OF SPORADIC ALS: THE FIRST TURKISH GWAS AND NOVEL SNP AND CNV ASSOCIATION

by

Özgün Uyan

B.S., Molecular Biology and Genetics, Izmir Institute of Technology, 2009

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

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2012

To my beloved grandmothers

and

To my family

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my thesis advisor Prof. A. Nazlı Başak for her valuable criticisms and continuous support. Her encouragement and guidance led me to finish this thesis. I am also very grateful to my thesis co-advisor and our collaborator Assoc. Prof. Hilmi Özçelik for his continuous guidance and sharing his experience, bright ideas during this thesis and for his great hospitality during my stay at Özçelik Lab, University of Toronto.

I would like to express my special thanks to Prof. Kuyaş Buğra, Prof. Özlem Keskin, and Cemalettin Bekpen, Ph.D. for devoting their time to evaluate this thesis.

I am endlessly grateful to my girlfriend Zeynep İtah for her invaluable support, motivation and scientific comments. Her assistance made me feel confident and comfortable in my scientific and social life since our undergraduate years.

I would like to thank my colleague, Zeynep Sena Ağım for her friendship, endless support and always being next to me in my anxious times. We were like twins throughout this three year M.Sc. program.

I would like to extend my thanks to NDAL team and my friends, Özgür Ömür, Gönenç Çobanoğlu, Pınar Aksoy, Suna Lahut, Ceren İskender, Kadir Özkan, Aslıhan Özoğuz, Arman Aksoy, Güneş Birdal, Mehmet Ozansoy, İzzet Enünlü, Selda Dağdeviren, İlknur Aydın, Irmak Şahbaz, Aslı Gündoğdu, Ece Kartal, Ceren Saygı and Ali Reza Khodadadi. I cannot forget the wonderful friendship and supports.

I would also like to acknowledge Suna and İnan Kıraç Foundation for the financial support it provided throughout my study.

Last but not least, I would like to express my deepest gratitude to my whole family. Their presence has been a constant source of support, love, regard and strength from the beginning of my life.

ABSTRACT

THE GENETICS OF SPORADIC ALS: THE FIRST TURKISH GWAS AND NOVEL SNP AND CNV ASSOCIATION

Selective degeneration of both lower and upper motor neurons located in the brain cortex, brainstem and ventral horn of the spinal cord result in amyotrophic lateral sclerosis (ALS). ALS is the most frequent late-onset motor neuron disorder and so far no cure or treatment for ALS has been developed. The vast majority of ALS cases are sporadic, approximately 10% of patients have a family history. At least 25 genes and several loci have been identified so far in familial cases. The genetic factors that contribute to ALS correspond to atmost 10% in sporadic cases so there is a huge genetic gap in sALS. The missing part of genetic factors needs to be revealed to fill the lacking players in the mechanisms underlying disease pathogenesis. GWAS enables large scale cohort studies between cases and controls to pinpoint new candidate genomic regions that may be associated with disease phenotype. In the framework of this thesis, we aimed to investigate sporadic ALS patients to identify novel Single Nucleotide Polymorphisms (SNPs), Copy Number Variations (CNVs) and genes, associated with ALS, using GWAS and computational biology approaches. First, we analyzed single marker associations of SNPs in the regions of fALS-causing genes in two published datasets. The ACA haplotype in TARDBP locus was found to be associated with ALS. Next, we performed GWAS on Turkish ALS samples, which included 116 sALS patients and 109 controls. We identified two SNP clusters in TMEM90B and CPNE5 genes associated with ALS, yet all SNPs failed to keep the significance after Bonferroni correction. Third, we performed genome-wide CNV analysis using raw data of GWAS and the PennCNV tool. We identified several novel and reported CNVs in gene and intergenic regions. To the best of our knowledge, this is the first ALS-GWAS of sALS patients in Turkey. We hope that the findings of this thesis in first-stage GWAS and genome-wide CNV analyses will shed light to the unknown genetics of sALS and help to pave the ways in the identification of novel genes associated with sALS pathogenesis.

ÖZET

SPORADİK ALS GENETİĞİ: İLK TÜRK GWAS ÇALIŞMASI VE YENİ SNP VE CNV İLİŞKİLERİ

Amiyotrofik lateral skleroz (ALS), beyin korteksi, beyin sapı ve omuriliğin ön boynuzundaki üst ve alt motor nöronların selektif dejenerasyonu sonucunda oluşur. ALS, en yaygın, geç başlangıçlı motor nöron hastalığıdır ve halen tedavisi yoktur. ALS olgularının büyük çoğunluğu sporadikken yüzde onluk kısmında aile öyküsü vardır. Ailesel olgularda, bugüne kadar en az yirmibeş gen ve birkaç lokus tanımlanmıştır. Sporadik ALS olgularının yüzde onluk kısmı genetik faktörlerle açıklanabilirken, sporadik olguların büyük bir kısmının genetiği bilinmemektedir. Bu genetik faktörlerin bulunması, ALS patogenezine neden olan yeni mekanizmaların ortaya çıkmasına yardımcı olacaktır. GWAS, genetik faktörleri belli olmayan hasta olgularını, sağlıklı birey gruplarıyla kıyaslayarak hastalık fenotipine neden olabilecek aday bölgelerin tanımlanmasını sağlamaktadır. Bu tez çerçevesinde, GWAS yöntemi ve ileri biyoinformatik yöntemleri kullanılarak sporadik ALS hastalarında yeni varyantların (SNP ve CNV) ve gen bölgelerinin bulunması hedeflendi. İlk olarak daha önceden yayınlanmış iki GWAS verisi kullanılarak ailesel ALS'ye neden olan genlerdeki SNP'ler araştırıldı. Sonuç olarak, TARDBP gen bölgesinde ACA haplotipinin ALS ile ilişkili olduğu görüldü. İkinci aşamada, 116 sporadik olgu ve 109 sağlıklı birey kullanılarak Türk toplumunda ilk defa ALS-GWAS çalışması gerçekleştirildi. TMEM90B ve CPNE5 gen bölgelerinde bulunan iki SNP kümesi ALS ile ilişkilendirildi ancak Bonferroni düzeltmesi sonucunda SNP'ler anlamlarını yitirdi. Üçüncü aşamada, GWAS ham verisi kullanılarak, PennCNV programı ile, genom-çapında CNV analizi yapıldı. Genomun intergenik ve gen lokuslarında birçok yeni ve daha önce tanımlanmış CNV bulundu. Tez çerçevesinde yapılan bu ilk-aşama GWAS çalışması bilgimiz dahilinde olan ilk Türk ALS-GWAS'ıdır. Bundan ve CNV analizlerinden çıkan ön bulguların, genetiği büyük oranda çözülmemiş olan sporadik ALS'ye ışık tutması beklenmektedir; ayrıca bu bulgular ALS patogenezinde etkin yeni genlerin bulunması ve dolayısıyla ALS'nin karmaşık patogenezinin anlaşılması yönünde de umut vericidir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

A	Adenine
AD	Autosomal dominant
ALS	Amyotrophic Lateral Sclerosis
ALS1/SOD1	Superoxide dysmutase 1
ALS10/TARDBP	TAR DNA-binding protein
ALS11/FIG4	Polyphosphoinositide phosphatase
ALS12/OPTN	Optineurin
ALS13/ATXN2	Ataxin 2
ALS14/VCP	Valosin-containing protein
ALS15/UBQLN2	Ubiquilin 2
ALS16/ SIGMAR1	σ Non-opioid receptor 1
ALS2	Alsin
ALS4/SETX	Senataxin
ALS5/SPG11	Spatacsin
ALS6 /FUS	Fused in sarcoma
ALS8/ VAPB	Vesicle-associated membrane protein-associated protein B
ALS9/ANG	Angiogenin
ALS-FTD2/C9ORF72	Chromosome 9 open reading frame 72
ALSoD	ALS Online Genetic Database
AMPA	Alpha-amino-3-hydroxi-5-methyl-4-isoxazolepropionic acid receptor
ApoE-ε4	Apolipo protein E
AR	Autosomal recessive
ATP6V1B1	V-type proton ATPase subunit B
ATXN1	Ataxin 1

B4GALT6	Beta-1,4-galactosyl transferase 6
BAF	B allele frequency
C	Cytosine
CA	California
CDCV	Common disease, common variant
CDRV	Common disease rare variant
CEU	Caucasian European
CFH	Complement factor H
Ctnna2	Catenin (cadherin-associated protein)
CNTN4	Contactin 4
CNV	Copy number variation
conf	confidence threshold
CPNE5	Copine-5
CR1	CR1 binding domain
CSNK1G3	Casein kinase I isoform gamma-3
Cu/Zn	Cupper/Zinc
dbGAP	Database of Genotypes and Phenotypes
dbSNP	Single Nucleotide Polymorphism Database
DCTN1	Dynactin 1
DGV	Database of Genomic Variants
DNA	Deoxyribonucleic acid
DPP6	Dipeptidyl-peptidase 6
EFEMP1	EGF-containing fibulin-like extracellular matrix protein 1
ER	Endoplasmic reticulum
fALS	Familial Amyotrophic Lateral Sclerosis
FDR	False discovery rate

FTD	Frontotemporal Dementia
G	Guanine
GALNT6	Galactosyl transferase 6
GWAS	Genome-wide association studies
HapMap	International Haplotype Map Project
HMM	Hidden Markov model
HWE	Hardy-Weinberg disequilibrium
IFNK	Interferone kappa
ITPR2	Inositol 1,4,5-triphosphate receptor type 2
Kb	Kilo base
KIAA1727	FH2 domain-containing protein 1
KIFAP3	Kinesine-associated protein 3
LD	Linkage disequilibrium
LIPC	Lipoprotein C
LMN	Lower motor neuron
LOD	Logarithm of odds
LRR	Log R Ratio
MAF	Minor allele frequency
MAPT	Microtubule-associated protein tau
MASP2	Mannan-binding lectin serine protease 2
Mb	Mega base
MGH	Massachusetts General Hospital
MND	Motor Neuron Disorder
MOBKL2B	MOB kinase activator 2B
NCBI	National Center for Biotechnology Information
NINDS	National Institute of Neurological Disorders and Stroke

NIPA1	Non-imprinted in Prader-Willi/Angelman syndrome region protein 1
NJ	New Jersey
NK	Neuromedin-K receptor
NRG1	Neuroglin 1
OR	Odds ratios
PARK16	Parkinson's disease 16
PDIA5	Protein disulfide isomerase family A, member 5
polyQ	Polyglutamine
PTGER3	Prostaglandin E2 receptor EP3 subtype
PXX	Proline rich domain
RNA	Ribonucleic acid
sALS	Sporadic Amyotrophic Lateral Sclerosis
SCA2	Spinocerebellar ataxia type 2
SCN7A	Sodium channel protein type 7 subunit alpha
SELL	L-selectin
Sen1p	Hypothetical
SMN	Survival motor neuron protein
SNAP	SNP Annotation and Proxy Search
SNCA	Alpha synuclein
SNP	Single nucleotide polymorphism
STI1	Heat shock chaperonin-binding motif
SUSD1	Sushi domain-containing protein 1
T	Thymine
TDP-43	TAR DNA binding protein 43
TMEM90B/C20ORF39	Synapse differentiation-inducing gene protein 1
UBA	Ubiquitin-associated domain

UBL	Ubiquitin-like domain
UCSC	University of California Santa Cruz
UK	United Kingdom
UMN	Upper motor neuron
UNC13A	Protein unc-13 homolog A
USA	United States of America
UTR	Untranslated region
ZFP64	Zinc finger protein 64
ZNF746	Zinc finger protein 746

1. INTRODUCTION

Neurodegenerative disorders, including Alzheimer's disease, Parkinson's disease and Motor Neuron Disorders (MNDs), all result from the death of neurons in different parts of the brain (Bertram and Tanzi, 2005). Their common features are: (i) they are represented as familial or sporadic, (ii) they are mostly late-onset where multiple genetic and environmental factors exhibit effects and (iii) protein aggregates are observed (Gandhi and Wood, 2010).

Familial forms allow understanding the pathogenesis of disease and unraveling the underlying mechanisms causing the disease phenotype. Effects of multiple genetic and environmental factors make the sporadic forms complex diseases. The molecular pathogeneses of neurodegenerative diseases are associated with inner and outer accumulation of misfolded protein aggregates in neuronal and non-neuronal cells, for instance amyloid-beta-positive inclusions in Alzheimer's disease, alfa synuclein-positive inclusions in Parkinson's disease (Polymenidou and Cleveland, 2011) and SOD1- or TDP-43-positive inclusions in Amyotrophic Lateral Sclerosis (ALS).

Selective degeneration of motor neurons causes MNDs. ALS, also known as Lou Gehrig's disease or Motor Neuron Disease, is the most frequent late-onset MND. ALS was first recognized by the French neurologist François Aran in 1848 and was named progressive muscular atrophy with family history (Aran, 1848). Another French neurologist, J. M. Charcot, reported it as non-hereditary in 1869 (Charcot, 1869). ALS affects both lower and upper motor neurons (LMN, UMN), located in the brainstem, ventral horn of the spinal cord (LMN) and cortex of the brain (UMN) (Figure 1.1). Degeneration of these particular neurons results in muscle atrophy and weakness, spasticity and hyperreflexia (Rowland & Shneider, 2001). The muscle weakness turns progressively into paralysis until death occurs, due to respiratory failure within 3-5 years after the first symptoms appear. Only half of ALS cases can survive lower than 5 years. Up to now, no cure or treatment has been developed for ALS (Bento-Abreu *et al.*, 2010). The only drug, riluzole that inhibits the release of glutamate and blocks AMPA receptors,

shows slight effects as up to 3 months of survival (Bellingham, 2011). The devastating effects of ALS make the life quality of patients even worse than in other neurodegenerative diseases.

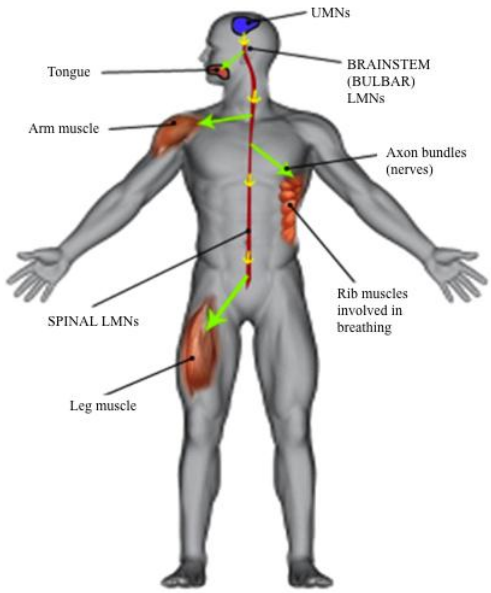


Figure 1.1. The affected neurons of the central nervous system in ALS.

ALS is the most common adult-onset motor neuron disease with a prevalence of 4-6 individuals per 100000 people and an incidence of 2-3 cases per 100000 per year. There is a slight variation of incidence within genders, men to women ratio is 1.2-1.5:1. The mean age of onset is between 55-60, however juvenile cases are also present (Kiernan *et al.*, 2011).

The vast majority of ALS cases are sporadic, approximately 10% of patients have a family history. Familial ALS (fALS) cases are inherited in autosomal dominant with complete/incomplete penetrance, autosomal recessive and X-linked dominant manners. Although the genetic cause of sporadic cases is not known, familial and sporadic ALS (sALS) patients share similar clinical and neuropathological features. There is a heterogeneous distribution in terms of types of symptoms, age of onset, disease duration among ALS patients, even within family members of the same family (Cozzolino *et al.*, 2012).

1.1. Genetics of ALS

Although the exact etiology of ALS has not been clarified yet, identification of fALS-causing genes shed light onto the underlying pathogenic mechanisms, such as disease initiation and progression (Table 1.1). The discovery of Cu/Zn Superoxide Dismutase (*SOD1)* gene (Rosen *et al.*, 1993) has dramatically promoted acceleration in genetic research of ALS. To date at least 20 genes and several loci have been identified (Andersen and Al-Chalabi, 2011). Genetic variations have also been associated with sporadic ALS using genome-wide association studies (GWAS). Moreover, in fALS, even though the genetic factors, which cause familial ALS are discovered to a great extent, the pathogenic mechanisms of mutated proteins are not well understood. (Ferraiuolo *et al.*, 2011).

Table 1.1. Genes associated with familial ALS.

ALS symbol / Gene Name	**Chromosome Locus**	**Type of Inheritance/ Onset**	**Protein name**	**Reference**
ALS1 /SOD1	21q22.1	AD, AR, de novo mutations / Adult	Superoxide dismutase 1	Rosen *et al.* (1993)
ALS2 / ALS2	2q33.2	AR/ Juvenile	Alsin	Yang *et al.* (2001)
ALS3 /Unknown	18q21	AD/Adults	Unknown	Hand *et al.* (2002)
ALS4/SETX	9q34	AD/Juvenile	Senataxin	Chen *et al.* (2004)
ALS5/SPG11	15q21.1	AR/Juvenile	Spatacsin	Orlacchio *et al.* (2010)
ALS6 /FUS	16q11.2	AD, AR, de novo mutations/Adult	Fused in sarcoma	Vance *et al.* (2009)
ALS7 /Unknown	20p13	AD/ Adult	Unknown	Sapp *et al.* (2003)
ALS8/ VAPB	20q13.3	AD/Adult	Vesicle-associated membrane protein-associated protein B (VAPB)	Nishimura *et al.* (2004)
ALS9/ANG	14q11.2	AD/Adult	Angiogenin	Greenway *et al.* (2006
ALS10/ TARDBP	1p36.2	AD, AR/Adult	TAR DNA-binding protein (TARDBP)	Gitcho *et al.* (2008) Sreedharan *et al.* (2008)
ALS11/FIG4	6q21	AD/Adult	Polyphosphoinositide phosphatase (FIG4)	Chow *et al.* (2009)
ALS12/OPTN	10p15–p14	AD, AR/ Adult	Optineurin (OPTN)	van Es *et al.* (2009)
ALS13/ATXN2	12q24	AD/Adult	Ataxin 2	Elden *et al.* (2010)
ALS14/VCP	9p13–p12	AD/ Adult	Valosin-containing protein (VCP)	Johnson *et al.* (2010)

Table 1.1. Genes associated with familial ALS (cont.).

ALS15/UBQLN2	Xp11	X-linked/ Adult	Ubiquilin 2 (UBQLN2)	Deng *et al.* (2011)
ALS16/ SIGMAR1	9p13.3	AD/ Adult AR/ Juvenile	σ Non-opioid receptor 1 (SIGMAR1)	Luty *et al.* (2010) Al-Saif *et al.* (2011)
ALS-FTD2/C9ORF72	9p21.2	AD/Adult	Chromosome 9 open reading frame 72	Renton *et al.*, 2011, DeJesus-Hernandez *et al.*, 2011

AD: Autosomal dominant
AR: Autosomal recessive

1.1.1. Genes associated with fALS

10 loci and genes, summarized below, comprise the majority of known fALS cases.

1.1.1.1. ALS1: Cu/Zn Superoxide Dismutase (SOD1). SOD1 is ubiquitously expressed and functions as a catalyst for removing free radicals in cells. It has five exons and encodes 153 amino acids, which are evolutionarily conserved. Mutations in SOD1 are responsible for 20% of familial ALS and for 2% of all ALS cases (Rosen *et al.*, 1993). Up to now, more than 160 disease-causing mutations, distributed over all five exons of the gene, have been described (http://alsod.iop.kcl.ac.uk/). Most of these are missense mutations, moreover several insertions/deletions have been reported that cause truncation of the protein.

Mutations in SOD1 can alter the metal binding sites resulting in inefficient binding of copper and zinc residues or affect the tertiary structure of the protein, resulting in improper folding and formation of unstable homodimers. *In vitro* and *in vivo* studies reliably showed that the enzymatic activity of the mutated sod1 is retained; the enzyme gains a toxic function in the cell when mutated, and sequesters the native SOD1 to form accumulations in the cell body (Polymenidou and Cleveland, 2011). Animal studies also confirmed that SOD1 mutated mice models did not show an ALS phenotype (Reaume *et al.*, 1996). How, the toxic gain of function of the mutated SOD1 protein leads to ALS still remains a mystery.

1.1.1.2. ALS2: Alsin (ALS2). Alsin is composed of 34 exons and has two isoforms of 1657 and 396 amino acids. Alsin is suggested to play a role in endosomal motility and

degradation, vesicular trafficking and cytoskeletal organization, because of its localization in the cell (Hadano *et al.*, 2010). Mutations in exon 34 of ALS2 cause sometimes a juvenile onset, slowly progressive ALS phenotype or infantile hereditary spastic paraparesis (Yang *et al.*, 2001; Hadano *et al.*, 2001). Until today, only a few mutations have been identified in juvenile cases, no adult patient has been detected with an Alsin mutation. Like SOD1, animal models of alsin did not show any upper or lower motor neuron loss (Deng *et al.*, 2007).

1.1.1.3. ALS4: Senataxin (SETX). Senataxin is composed of 26 exons and has three isoforms. The SETX gene encodes 2677 amino acids. Mutations in SETX can cause several phenotypes, such as cerebellar ataxia, oculomotor apraxia, peripheral neuropathy immunodeficiency and ALS (Moreira *et al.*, 2004; Chen *et al.*, 2004). Only dominantly inherited mutations in the SETX gene are observed in few fALS and some sALS cases. Patients with SETX mutations show juvenile onset and slow progression (Hirano *et al.*, 2011). The function of the senataxin protein is not known, however, it shows homology to Sen1p of fungi, which functions as an RNA helicase. It is suggested that it may play roles in RNA maturation and is involved in double-strand DNA break damage response (Suraweera *et al.*, 2009; De Amicis, 2011).

1.1.1.4. ALS10: TAR DNA-binding protein (TDP-43). The TDP-43 gene contains six exons encoding for a 414 aa protein. It functions as a regulator of transcription and splicing (Tollervey *et al.*, 2011). Like alsin and senataxin, TDP-43 is also associated with another neurological disease which is Frontotemporal Dementia (FTD). In 2006, TDP-43 inclusions were found in affected neurons of sALS and FTD patients (Arai *et al.*, 2006; Neuman *et al.*, 2006). In normal cells, TDP-43 is mostly localized in the nucleus, but when mutated, TDP-43 loses its function and the ubiquitinated form is exported to the cytoplasm. This mislocalization raises the possibility of a gain of toxic function of TDP-43 like SOD1 (Giordana *et al.*, 2009). Until now, 44 different mutations with mostly autosomal dominant, but also autosomal recessive (rare) inheritance have been reported in fALS and sALS patients (http://alsod.iop.kcl.ac.uk/), accounting for 4% and 1%, respectively (Daoud *et al.*, 2009; Da Cruz and Cleveland, 2011).

1.1.1.5. ALS6: Fused in Sarcoma (FUS). Mutations found in the TDP-43 gene promoted the discovery of a new candidate, FUS, which has a similar function. FUS is located on chromosome 16q11.2 which was since long associated with ALS. The mutation screening of the FUS gene revealed that approximately 4% of fALS and 1% of sALS cases are affected due to FUS mutations (Millecamps *et al.*, 2010; Tsai *et al.*, 2011). Up to now, 44 mutations have been described in ALS patients.

1.1.1.6. ALS8: Vesicle-associated membrane protein (VAMP)-associated protein type B (VAPB). VAPB encodes for a 99 amino acid long protein, which is localized mainly in the endoplasmic reticulum (ER). It plays role in intracellular membrane transportation, from ER to Golgi and from ER along to axon. Mutations in VAPB gene cause late onset muscular atrophy Finkel type and juvenile onset, slowly progressive ALS (Nishimura *et al.*, 2004; Chen *et al.*, 2010). Mutations found in patients are rare and the mutated protein promotes insoluble aggregates in the cell. Animal models of the VAPB gene in Drosophila show that when its mutated form is overexpressed, mislocalization of the protein and aggregate formation are present in the neurons, resulting in neuronal death (Ratnaparkhi *et al.*, 2008).

1.1.1.7. ALS9: Angiogenin (ANG). Angiogenin located on chromosome 14, is a hypoxia-regulated gene and encodes 147 amino acids. It was first associated with ALS since single nucleotide polymorphisms observed in ANG in Irish and Scottish families resulted in ALS (Greenway *et al.*, 2004). Further studies in different populations (US, German and French) showed that missense mutations, found in the ANG gene of fALS and sALS patients, are associated with the disease phenotype (Greenway *et al.*, 2006; Paubel *et al.*, 2008; Fernandez-Santiago *et al.*, 2009). The ANG protein has ribonuclease activity in the cell and is highly expressed in the nucleus and cytoplasm of spinal cord ventral horn neurons. Expression studies showed that the mutated form is unable to enter the nucleus, therefore it can not perform its RNase activity (Wu *et al.*, 2007).

1.1.1.8. ALS13: Ataxin-2 (ATXN2). Ataxin-2 encodes for a 1313 amino acid protein, which includes a polyglutamine (polyQ) repeat in exon-1. The CAG repeats in normal alleles of ATXN2 gene vary, however the most prominent length of repeats of the normal allele is 22. CAG repeats with a threshold of 34 repeats or more are associated with

spinocerebellar ataxia type 2 (SCA2) (Elden *et al.*, 2010). In addition, due to two SNPs (rs695871 and rs695872), located upstream of the CAG repeat region, distinct genotypes are observed in controls (GT) versus patients (CC). Recently, intermediate-length (27-33) polyQ repeats were reported to be an increased risk factor for ALS in different populations (Yu *et al.*, 2011; Van Damme *et al.*, 2011; Gellera *et al.*, 2012).

1.1.1.9. ALS15: Ubiquilin-2 (UBQLN2). Ubiquilin-2, ubiquitin-like protein 2, consists of one exon and encodes for 624 amino acids. Ubiquilin-2 is a member of the ubiquitin-like protein family. In humans, this family contains four ubiquilin members that are characterized by two domains, located at the N- and C- terminals, a ubiquitin-like domain (UBL) and a ubiquitin-associated domain (UBA). The function of ubiquilins is to deliver ubiquitinated proteins to the proteasome for degradation. The UBL and UBA domains enable the protein to bind to the subunit of the proteasome and to the poly-ubiquitinated proteins. According to the predicted structure of ubiquilin-2, the protein contains a UBL, four heat shock chaperonin-binding motifs (STI1) and 12 PXX repeats. The mid part of the protein varies among the other members of the ubiquitin-like protein family due to characteristic roles. Only ubiquilin-2 contains a unique 12 PXX repeat. Recently, five different UBQLN2 mutations in the PXX domain were reported to be associated with ALS. The inheritance pattern is X-linked dominant. Immunohistochemistry analysis of patients showed that UBQLN2-positive and skein-like inclusions are present in the spinal cord and hippocampal sections (Deng *et al.*, 2011). In cell culture studies, it has been also shown that ubiquilin-2 is co-localized with the C-terminal of TDP-43 and forms aggregates in the cell cytoplasm (Brettschneider *et al.*, 2012).

1.1.1.10. ALS-FTD2: C9ORF72. The product of the C9ORF72 gene is a hypothetical protein that is composed of 481 amino acids. The function of the protein has not been characterized yet. One key point is that the gene contains hexanucleotide repeats $(GGGGCC)_n$ in intron 1. C9ORF72 is located on 9p21.2, which was previously associated with ALS by linkage analysis (Morita *et al.*, 2006; Vance *et al.*, 2006). The locus remained as a mystery until 2011 though several GWAS indicated this locus as an ALS-associated region (van Es *et al.*, 2009; Laaksovirta *et al.*, 2010; Shatunov *et al.*, 2010). Finally, further focusing on this region, where especially three genes, MOBKL2B, IFNK and C9ORF72 were of prime importance, are located, C9ORF72 was defined as the

responsible gene. Resequencing analysis showed that expansion of a hexanucleotide repeat $(GGGGCC)_n$ is observed in ALS-FTD patients. Normally the number of repeats is lower than 23, however patients can have up to 1600 repeats (Renton *et al.*, 2011; DeJesus-Hernandez *et al.*, 2011). The frequency of the C9ORF72 hexanucleotide repeat expansion is found to be 37.6% in fALS and 6.3% in sALS patients in a large scale study population (Majounie *et al.*, 2012). In a recent functional study, ALS and FTLD-TDP cases with the C9ORF72 expansion revealed UBQLN-positive cytoplasmic inclusions and extensive UBQLN-positive aggregates. Thus, not only UBQLN2 mutations, but also C9ORF72 expansions cause UBQLN2 pathology in both fALS and sALS cases (Brettschneider *et al.*, 2012).

1.2. Molecular Mechanisms Involved In ALS Pathology

The etiology of ALS is still elusive and the pathological mechanisms are not fully determined yet, although many genetic factors and clinical features have been identified. The genetic factors associated with ALS take roles in different parts of cellular mechanisms. Therefore, many potential pathogenic processes underlying ALS have been proposed. Common clinical features of sALS and fALS patients and identification of fALS associated genes in some sALS patients interrogate sharing common mechanisms in disease pathology (Kiernan *et al.*, 2011).

As a result of all findings, multiple and partly overlapping mechanisms interplay roles in ALS pathogenesis. These mechanisms are involved in excitotoxicity, protein aggregation, oxidative stress and impairments in RNA processing, mitochondrial function and axonal transport. Beside intracellular mechanisms, extracellular non-cell autonomous events are also considered, such as neuro-inflammation, glutamate excitotoxicity and reduced lactate release, caused by neighboring cells (Ferraiuolo *et al.*, 2011) (Figure 1.2).

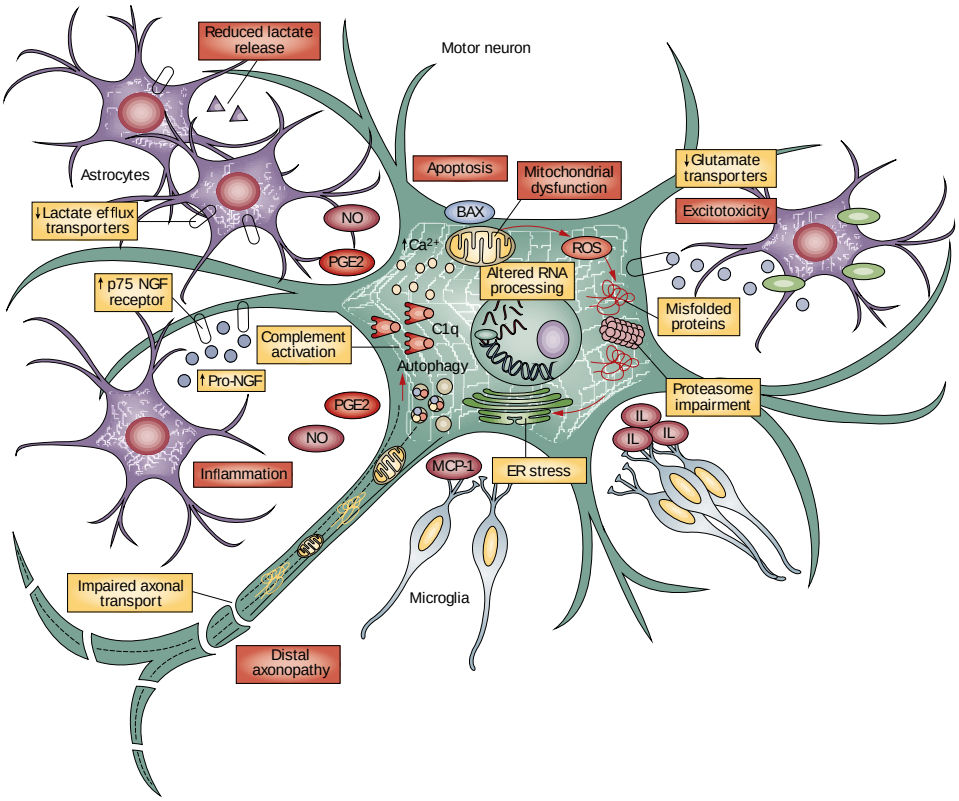


Figure 1.2. Pathological mechanisms in ALS (Ferraiuolo *et al.*, 2011).

All the mechanisms above have effects on motor neuron vulnerability, they are complex and multifactorial. Although at least 20 genes have so far been associated with fALS, there are still more than 50% of fALS cases whose defects need to be determined. The important point is whether these mechanisms work together, or not. It is shown in tissue samples from SOD1 mutated fALS patients that the inclusions in the cellular matrix are SOD1-positive and TPD43-negative. This indicates that different players in different mechanisms can cause motor neuron degeneration. SOD1-related mechanisms such as oxidative stress can also trigger other mechanisms in the cell, for example, excitotoxicity, mitochondrial impairment, protein aggregation leading to neurodegeneration. Another important key point is that different types of protein aggregations are found in sALS and fALS cases. Several recent histopathology studies of sALS cases showed that without any mutations in TDP-43 or SOD1, misfolded SOD1 and loss of nuclear-localized TDP-43 are observed in cytoplasmic inclusions (Andersen and Al-Chalabi *et al.*, 2011).

In conclusion, ALS is a complex and multifactorial disease when its etiology, genetic heterogeneity and pathologic mechanisms are considered. The effects of environmental factors are still under investigation especially in sALS cases.

1.3. Genome-wide Association Studies

Genome-wide association studies (GWAS) are robust, unbiased and high throughput techniques which can test SNPs, located in the human genome, for association with a disease between a large case and control population (Figure 1.3).

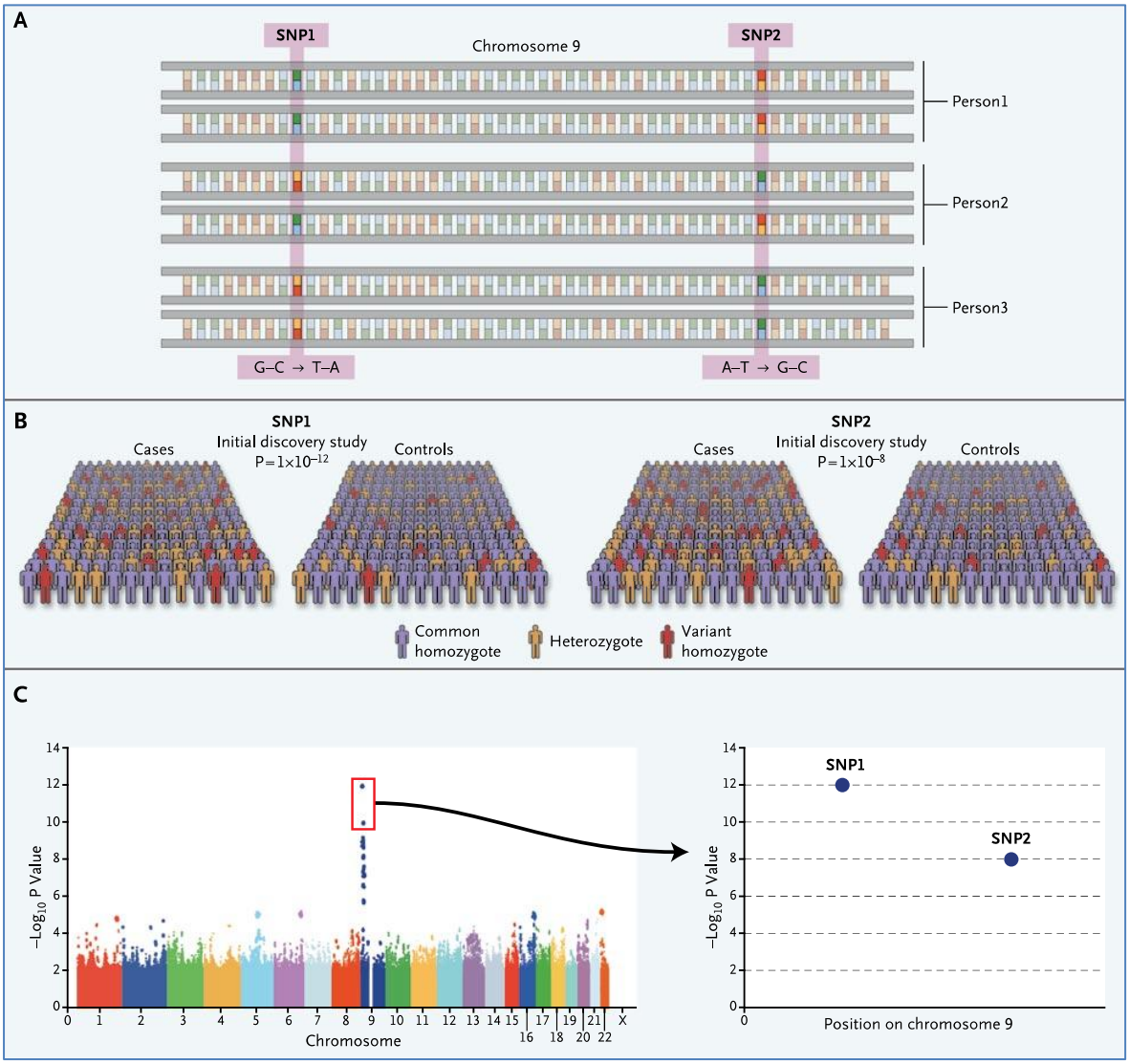


Figure 1.3. Design and analysis of a Genome-wide Association Study (Manolio, 2010).

Before the completion of the human genome project and identification of hundreds of thousands SNPs in the HapMap project, researchers performed candidate gene association studies which were small-scale, low-resolution and most importantly time consuming. With the technological advances in microarray platforms, statistical tools and the relevance of SNPs, located all throughout the human genome and in numbers higher than expected (to date >10000000) powerful genome-wide association approaches, became technically possible and feasible (International HapMap Consortium, 2007; Manolio, 2010).

A disease with complex traits, involving many genes and environmental factors contributing to disease phenotype, is hard to explain by old fashion methodologies. Two hypotheses have been proposed to complex disease susceptibility: (i) the common disease, common variant (CDCV) hypothesis, suggesting a high frequency and low penetrance genetic variations contributing to disease susceptibility, (ii) common disease rare variant (CDRV) hypothesis claiming that common disease is also caused by multiple rare variations with high penetrance. According to these first hypotheses, GWAS using SNPs would be a powerful tool to identify novel variants associated with complex disease, however, this was not the case (Schork *et al.*, 2009).

Genome-wide association studies are based on SNPs which are the most frequent genomic variations on the genome. A SNP can only be defined as SNP if it is observed at least in 1% of the population. To date the human genome has been shown to contain more than 10 million common SNPs with a minor allele frequency lower than 5% (Manolio, 2010). SNPs that are close to each other segregate together resulting in reduction in genetic diversity. SNPs that are transmitted to the next generation within a block allow us to tag lots of SNPs as a tagged-SNP and capture maximum information by using a minimum number of SNPs. Today, DNA microarrays containing 1 million SNPs, can cover approximately 90% of the 10 million SNPs on the human genome (Wellcome Trust Case Control Consortium, 2007; McCarthy *et al.*, 2008).

The result of a huge GWAS of a case-control population, contains an enormous output of data. To analyze such big data, powerful association tests and statistical tools are needed for avoiding false positive results to start with. In such a study, more than 500000

comparisons are made to test the associations, and it is possible to obtain false positive results (Hunter *et al.*, 2007). To eliminate false positive outcomes from GWAS, a large set of sample size is required. This is also required for stringent statistical thresholds of multiple testing, such as Bonferroni correction. To achieve genetic variation, associated with a particular disease phenotype, without having false positive and negatives, a serial of experiments are needed (Gandhi and Wood, 2010).

In the first stage of a large GWAS, significant SNPs are defined; this is called the discovery set. In the second stage, SNPs that are significantly associated, are genotyped in a smaller population group and so on in the third stage (Pearson and Manolio, 2008) (Figure 1.4).

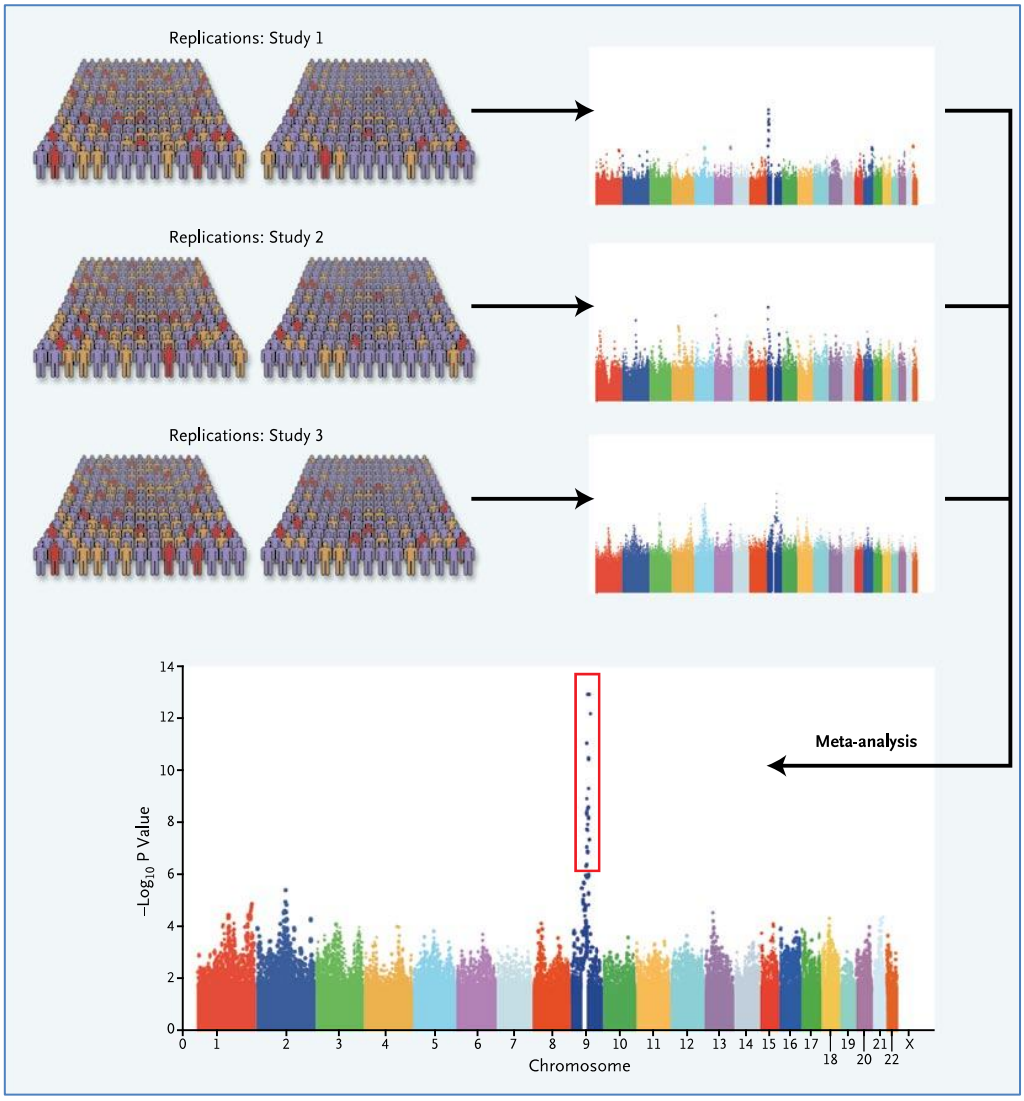


Figure 1.4. Stages of a GWAS and meta-analysis (Manolio, 2010).

These processes basically help to identify false positives and validate true positive associations. Sometimes, pooling results of independently performed GWAS, also called meta-analysis, can improve the power of associated SNPs. Independent studies may give low signals before combining the results in meta-analysis (Mok *et al.*, 2011) (Figure 1.4).

Identification of associated SNPs simply by considering the absolute p-values is not enough as a significant evidence. To obtain stronger associations, they need to be evaluated by multiple testing and odds ratio (OR) (Barrett *et al.*, 2009). Finally, confirmation should be performed by genetic and functional analyses (Jakobsdottir *et al.*, 2009; Hindorff *et al.*, 2009).

To date 1223 GWAS have been published in 237 traits and more than 1400 trait-associated SNPs have been reported as significant ($p< 5x\ 10^{-5}$) (Figure 1.5). The number of studies performed on complex traits using high-throughput GWA methodology is increasing every day (http://www.genome.gov/gwastudies) (Figure 1.6). This promising methodology enables a genome-wide scan to discover genetic variants for common diseases. This idea has been proven in several diseases such as age-related macular degeneration and Crohn's disease (Maller *et al.*, 2006; Barrett *et al.*, 2008).

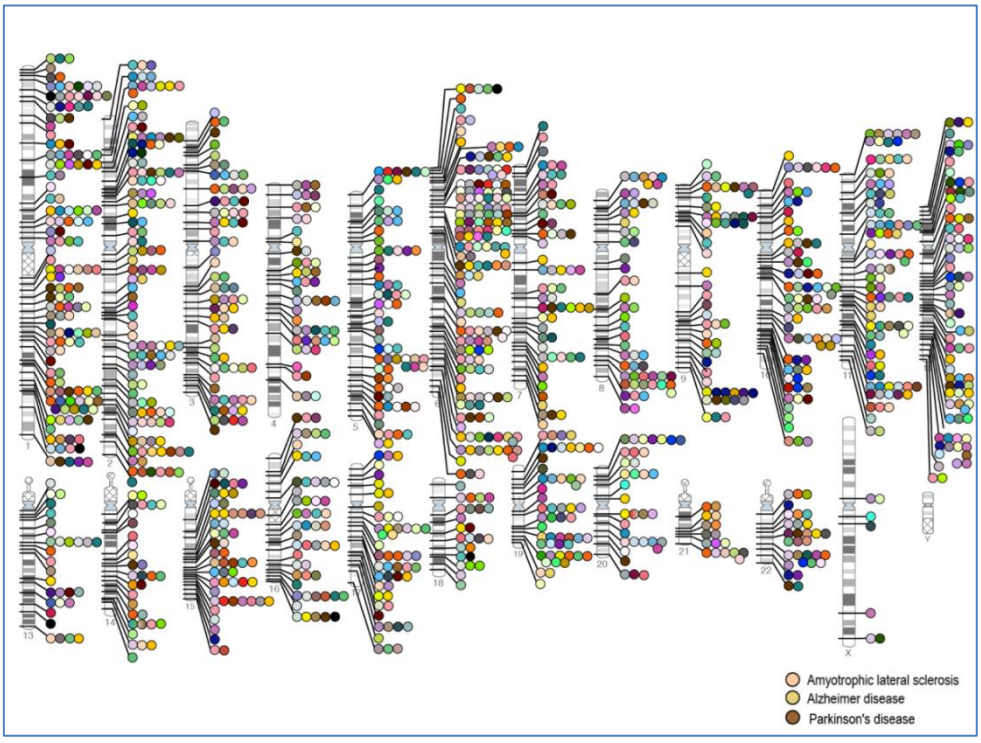


Figure 1.5. Schematic distribution of significantly associated SNPs on human chromosomes.

Many variants have been associated with those diseases. Individuals carrying one or more of these variants, have an increased risk for developing the disease.

The identification of variants in the CFH gene, associated with age-related macular degeneration, suggests that the complement mediated inflammation pathway possibly plays a role in disease pathogenesis (Klein *et al.*, 2005). This findings lead to generate animal models with the aim of developing new therapies against pathogenesis of macular degeneration (Rohrer *et al.*, 2009).

On the other hand, GWAS are not as accurate as it was thought previously. Although GWAS have achieved their goal to unravel new disease-associated genes, loci and variations, there are several studies without any replicated results, such as GWAS of schizophrenia (Sullivan *et al.*, 2008; Shi *et al.*, 2009). In general, all GWAS results of complex diseases, e.g. neurodegenerative diseases, including Parkinson's disease, Alzheimer's disease and ALS, have shown that all associations can explain only a small portion of the disease pathogenesis, the majority of unknown genetic factors still remain unsolved (Gandhi and Wood, 2010). This phenomenon is called "missing heritability".

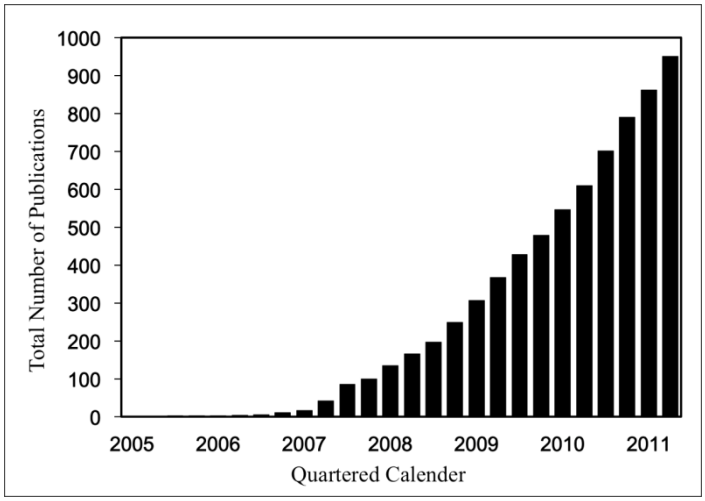


Figure 1.6. Distribution of increasing number of GWAS done up to June 2011.

Although GWAS identified more than thousand variants in complex diseases and traits, the presence of the high missing heritability in those diseases fails to satisfy the expectations. This can be explained due to overestimation of the total heritability of the

complex phenotypes or underestimation of the effect sizes of the common variants. The associated SNP might not directly be the cause of a disease but can represent a SNP cluster that has a larger effect size. In another case, common variants cannot cover the rare variants (minor allele frequency (MAF) <0.01), resulting in loss of association, which cannot be captured by GWAS (Durbin *et al,.* 2010). It requires other techniques, such as Whole Exome Sequencing and Whole Genome Sequencing. In addition to these, epistasis (gene x gene and gene x environment) interactions and epigenetics can explain a part of the missing heritability (Zuk *et al.*, 2012). All in all, unraveling the missing heritability of disease will allow us to understand the contribution of all DNA sequence variants (common or rare) to molecular mechanisms of complex disease development (Marian, 2012).

1.3.1. GWAS in Neurodegenerative Diseases

Up to now, more than 10 GWAS studies have been performed in Parkinson's disease, Alzheimer's disease and ALS (http://www.genome.gov/gwastudies). The majority of these studies had inconsistent results. Replication stages failed to validate initial stage (discovery) or another independent study (Schymick *et al.*, 2007; Dunckley *et al.*, 2007). Only a minority is able to confirm results found in an independent study (Cronin *et al.*, 2007; van Es *et al.*, 2007). In Alzheimer's disease, the only genetic risk factor, which has been consistently identified and confirmed in GWAS, is ApoE-ε4 (Lambert *et al.*, 2009; Harold *et al.*, 2009). Meanwhile, two groups showed associated variants, located in the CLU and CR1 gene loci, however the associated SNPs were not overlapping. Despite the identification of these, previously unknown two genes by GWAS, it is not able to understand the disease pathogenesis in sporadic forms. In Parkinson's disease, different variants in the MAPT, SNCA and PARK16 genes were found to be associated with disease susceptibility. Furthermore, other susceptibility regions have been found as associated, yet they have not been replicated in any other study (Simón-Sánchez *et al.*, 2009; Satake *et al.*, 2009; Edwards *et al.*, 2010). Additional studies are required for non-replicated, associated variants, such as meta-analysis or functional studies, in order to understand whether these are false positives or population- specific results (Gandhi and Wood, 2010; Manolio, 2010).

1.3.2. GWAS in ALS

As mentioned earlier, the genetics of sALS is still poorly understood. With the advances in GWAS, it was thought that GWAS would provide effective information to understand the fundamental mechanisms that cause the disease pathology. (Dunckley *et al.*, 2007). From 2007 to 2011, comprehensive, unbiased high-throughput GWAS technology was not able to identify novel variants or new candidate genes, which would give insights into the disease pathogenesis. In addition, the pathways and mechanisms already known in disease development have not been validated. Eleven GWAS have been performed in ALS between 2007 and 2011 (Table 2). A high number of samples has been used from different countries, including Ireland, USA, UK, Finland, Germany, Sweden, Netherlands and France. Although several candidate genes, survival factors, and a candidate locus have been associated with ALS in these studies (van Es *et al.*, 2009; Shatunov *et al.*, 2010), yet only one gene (DPP6) and one locus (9p21) have been replicated in other studies. The DPP6 gene was validated by two different groups who used the same samples in their study design, but has not been identified in other populations (van Es *et al.*, 2007; Cronin *et al.*, 2008).

The locus on chromosome 9p21 has been first associated with ALS in studies of Finland and UK populations in 2010 (Laaksovirta *et al.*, 2010; Shatunov *et al.*, 2010). In contrast to other studies, the Finnish study also included fALS patients in the study population. Shatunov *et al.* also identified significant association in the same locus. A combined meta-analysis study, including five different populations, defined a twenty-SNP haplotype block located in this locus. As a result of all these series of GWAS analyses, C9ORF72 was identified as a candidate gene for ALS development. Resequencing analysis of this locus revealed that GGGGCC hexanucleotide expansions in the first intron of the C9ORF72 gene were present in ALS patients, but not in controls (Renton *et al.*, 2011; DeJesus-Hernandez *et al.*, 2011).

Table 1.2. GWAS in ALS.

First Author (year published)	Sample size	Ethnicity	Replication Sample size	Ethnicity	Region	Genes	Strongest SNPs	P-value	Platform (# of SNPs)
Schymick JC (2007)	276 cases, 276 controls	US	-	-	10q26.2 2p24.2 8q24.23 20q13.2 9q31.3 4q31.3	Intergenic Intergenic Intergenic ZFP64 SUSD1 KIAA1727	rs4363506 rs16984239 rs12680546 rs6013382 rs2782931 rs11099864	7 x 10-7 2 x 10-6 3 x 10-6 5 x 10-6 6 x 10-6 9 x 10-6	Illumina (549,062)
Dunckley T (2007)	386 cases, 542 controls	White	901 cases, 1025 controls	Whites non-whites	-	-	-	-	Affymetrix and Illumina (776955)
Van Es MA (2007)	737 cases, 721 controls	Dutch	1030 cases, 1195 controls	Dutch Belgian Swedish	12p11.23	ITPR2	rs2306677	3 x 10-6	Illumina (311946)
Cronin S (2007)	221 cases, 211 controls	Irish	737 cases, 737 controls	Irish Dutch US	7q36.2	DPP6	rs10260404	3 x 10-6	Illumina (497917)
Van Es MA (2007)	737 cases, 721 controls	Dutch	1030 cases, 1195 controls	Dutch Belgian Swedish	7q36.2 15q21.3 2p24.2	DPP6 LIPC Intergenic	rs10260404 rs3825776 rs7580332	5 x 10-8 9 x 10-6 9 x 10-6	Illumina (311946)
Cronin S (2008)	958 cases, 932 controls	Irish Polish	309 cases, 404 controls	Irish US Dutch	7q36.2	DPP6	rs10260404	3 x 10-6	Illumina (287522)
Chio A (2009)	553 cases, 2338 controls	European	3149 cases, 3335 controls	European	-	-	-	-	Illumina (545066)

Table 1.2. GWAS in ALS (cont.).

First Author (year published)	Sample size	Ethnicity	Replication Sample size	Ethnicity	Region	Genes	Strongest SNPs	P-value	Platform (# of SNPs)
Landers JE (2009)	1821 cases, 2258 controls	US European	538 cases, 556 controls	US European	1q24.2 7q36.1 18q12.1 2p16.1 5q23.2 1q24.2 6p22.3 2q24.3 3p26.3	KIFAP3 ZNF746 B4GALT6 EFEMP1 CSNK1G3 SELL ATXN1 SCN7A CNTN4	rs1541160 rs855913 rs10438933 rs7577894 rs11241713 rs3177980 rs697739 rs13015447 rs2619566	2 x 10-8 (survival) 4 x 10-8 1 x 10-6 1 x 10-6 3 x 10-6 4 x 10-6 4 x 10-6 7 x 10-6 7 x 10-6	Illumina (288357)
Van Es MA (2009)	2323 cases, 9013 controls	Dutch US Belgian Irish Swedish	2532 cases, 5940 controls	Dutch US UK Irish Swedish Polish German	19p13.1 19p21.2 9p21.2 Xq13.3 12p12.3	UNC13A MOBKL2B, IFNK, C9orf72 MOBKL2B, IFNK, C9orf72 Intergenic Intergenic	rs12608932 rs2814707 rs3849942 rs5937496 rs9971637	3 x 10-14 7 x 10-9 1 x 10-8 6 x 10-7 2 x 10-6	Illumina (292768)
Laaksovirta H (2010)	405 cases, 497 controls	Finnish	-	-	9p21.2 21q22.11	MOBKL2B,IF NK,C9orf72 SOD1	rs3849942 rs13048019	9 x 10-11 3 x 10-8	Illumina (318167)
Shatunov A (2010)	4857 cases, 8987 controls	UK US Dutch Irish Italian French Swedish Belgian	-	-	9p21.2 9p21.2 11q14.3 18q23	Intergenic Intergenic Intergenic Intergenic	2-SNP haplotype 2-SNP haplotype rs1488902 rs4799088	5 x 10-11 8 x 10-10 3 x 10-6 9 x 10-6	Illumina (227475)

1.4. Copy Number Variations

CNVs are the changes in number of copies of a particular fragment of DNA. This can be either deletion or duplication of the fragment. A CNV can range from 1-kilo base up to mega bases. The genome-wide presence (spreads) of CNVs were detected in several studies, including healthy populations (Conrad *et al.*, 2010; Redon *et al.*, 2006). Thus widespread deletions and duplications of CNVs in the human genome were reported in several different studies (Ku *et al.*, 2010). The identification and characterization of CNVs allow the understanding of their importance in the human genome defining all genetic components of the genome (Feuk *et al.*, 2006), yet the significance of CNVs has not been distinguished due to lack of information of their rare presence and absence in healthy controls. In contrast with SNPs, the frequency of CNVs are mostly lower than 5% throughout the genome, and they are composed of more than 1000 nucleotides. In fact, structural variants, including insertions, deletions and CNVs, are rare throughout the genome; their presence affects cellular metabolisms in the cell (Stranger *et al.*, 2007). Losses (deletion) or gains (duplication), caused by chromosome rearrangements, can alter the expression levels and/or lead to truncation of proteins, which influence the risk of complex diseases especially in sporadic cases. Several recent studies reported CNVs to be associated with autoimmune disorders and Crohn's disease, schizophrenia, and epilepsy (Fellermann *et al.*, 2006; Stefansson *et al.*, 2008; Brunetti-Pierri *et al.*, 2011).

1.4.1. Copy Number Variation Studies in ALS

There are only four genome-wide CNV studies on ALS in literature. Unlike other neurological diseases, such as schizophrenia and autism, no significant associations with CNVs have been identified in ALS (Cronin *et al.*, 2008; Wain *et al.*, 2009). Only the DPP6 and NIPA1 loci are candidates for further investigation (Blauw *et al.*, 2010). All of these studies use genome-wide SNP intensity data to analyze CNVs. The analysis of raw intensity data requires sophisticated algorithms such as QuantiSNP and PennCNV (Wang *et al.*, 2007). In a recent study, duplication of the SMN1 gene, which is responsible for spinal muscular atrophy, has been associated with sporadic ALS as a risk factor (Blauw *et al.*, 2012); furthermore homozygous SMN2 deletions are associated with ALS as a protective factor in the Swedish population (Corcia *et al.*, 2012).

All in all, multiple rare CNVs, including both deletions and duplications, may play roles in disease susceptibility and pathology of sporadic ALS, may be even more than common variants.

2. PURPOSE

The array technology and GWAS opened a new era in understanding complex diseases. Although it was initially thought that common variants, captured in GWAS, would largely unravel the mechanisms underlying complex disease pathogeneses, efforts and great expenses spent on GWAS did not fulfill the expectations of many scientists. The main reasons behind this are the drawbacks of GWAS due to its SNP coverage.

To date, 11 GWAS have been completed in ALS and the platforms (arrays) used in study design are mostly 300k and 500k SNP arrays; this indicates that at most 500000 SNPs capture 90% of the three billion base pairs of the human genome. The captured tagged-SNPs represent mainly common SNPs. Although these tag SNPs also cover some rare variants, these are usually not captured in association tests.

Several genes and SNPs were shown to be associated with ALS, however none of them was found to be an ALS-causing variant, with the only exception of C9ORF72. This can be either due to small effect size of common alleles or due to underestimation of effect size of rare variants. In addition, these data contain more information than we can retrieve and interpret so far. The data generated are still expected to explain the missing heritability using different methodology approaches and how common or rare variants direct us to solve the missing heritability.

In this respect, this study has several aims:

- to analyze raw data obtained in previously published GWAS and find new (novel) variants which might have larger effect sizes.

- to apply GWAS for the first time to a Turkish case and control cohort.

- to analyze other common and rare variants such as CNVs to understand their roles in disease phenotype.

3. MATERIALS

3.1. Subjects: Study Groups

Three subject groups used in this study and their features are summarized in Table 3.1: (i) the Turkish dataset contains Turkish patients and Turkish control samples, (ii) the other two datasets selected include samples of the US and Irish populations and are obtained from dbGap.

3.1.1. The Turkish Genome-wide Association Study

ALS patients were referred to us from different hospitals and neurology clinics throughout Turkey. El Escorial criteria were applied for clinical diagnosis (Brooks *et al.*, 2000). All patients gave written informed consent to participate in the study and the approval of the use of patient samples was obtained from The Ethics Committee of Boğaziçi University. Control samples were collected anonymously from the Microbiology Department of Haydarpaşa State Hospital (İstanbul).

For this study, 130 sALS patients (75 males and 55 females) and age- and sex-matched healthy controls were selected for SNP genotyping. The mean age of onset of sALS patients was 48.1, ranging from 17 to 79 years, and the mean age of controls was 55.1, ranging from 23 to 84 years.

3.1.2. The US Genome-wide Association Study (Schymick *et al.*, 2007)

The study was carried out by Bryan J. Traynor and co-workers. Samples were obtained from the NINDS Neurogenetics Repository at the Coriell Institute for Medical Research, NJ, USA. 276 white, non-Hispanic US individuals diagnosed with sporadic ALS were selected for the study. The case and controls groups, selected from the same ethnic background, were, however, not age- and sex-matched. The mean age of patients at onset was 54,8, ranging from 26 to 87 years. Two hundred seventy five neurologically normal,

white, non- Hispanic controls were selected from many different sites of the US. Controls consisted of 131 men and 144 women, the mean age was 68 years (range 55–88).

3.1.3. The Irish Genome-wide Association Study (Cronin *et al.*, 2007)

The study was carried out by Orla Hardiman and co-workers. The Irish dataset included 221 sALS patients and 211 controls. DNA samples were collected at the ALS clinic in Beaumont Hospital, Dublin. All patients met the criteria of sALS, according to the El Escorial criteria. 221 sALS patients included 120 males and 101 females and the mean age of patients at onset was 61 years; 211 controls consisted of 112 males and 99 females with mean age of 58 years.

Table 3.1. Characteristics of the Turkish, US and Irish populations used in this study.

Population	Type	Total	Male	Female	Mean age at onset (years)	Mean age at sample collection (years)
Turkish	sALS	130	75	55	48	55
	Controls	130	75	55	-	55
US	sALS	276	-	-	54	-
	Controls	275	131	144	-	68
Irish	sALS	221	120	101	61	-
	Controls	211	112	99	-	58

3.2. Genotyping

3.2.1. The US Population

DNA samples of all ALS patients in this study were assayed with the Illumina Infinium II HumanHap550 SNP chips (Illumina, San Diego, CA, USA). 555352 SNPs from phase I and II of the HapMap Project were genotyped. Out of 275, 227 control samples were assayed with the Illumina Infinium II 240S SNP chip according to the manufacturer's protocol. These data were combined with previously generated genotype information of the same 227 control samples containing 317511 SNPs. In addition, 48 additional control samples were genotyped with HumanHap550 SNP chips. In total, the

control dataset with 275 genotyped samples was obtained, containing the same 555352 SNPs as the ALS samples (Schymick *et al.,* 2007).

3.2.2. The Irish Population

The DNA samples of the Irish individuals were genotyped using Illumina Infinium II 550K SNP chips (Illumina, San Diego, CA, USA) at the Laboratory of Neurogenetics, Bethesda, according to manufacturer protocols. 561466 SNPs of each individual were assayed in this study (Cronin *et al.,* 2007).

3.3. Equipment

Table 3.2. Devices used for bioinformatic analyses.

Device	Manufacturer
Laptop Computer	R522, Samsung, Korea
Desktop Computer	HP Pavillion Elite, USA
Laptop Computer	MacBook Pro, Apple, USA
Desktop Computer	iMac, Apple, USA

3.4. Databases and Bioinformatic Tools

Several databases and softwares were used during dataset arrangements and statistical analyses.

3.4.1. Databases

3.4.1.1. Database of Genotypes and Phenotypes (dbGAP). dbGAP is a public repository that is hosted by the National Center for Biotechnology Information (NCBI). It stores and distributes data and results of studies that have investigated genotype-phenotype correlations, such as genome-wide association studies, medical sequencing and molecular diagnostic assays (http://www.ncbi.nlm.nih.gov/gap).

3.4.1.2. Single Nucleotide Polymorphism Database (dbSNP). dbSNP is the online database of short genomic variants including single nucleotide polymorphisms and short deletions/insertions and microsatellite markers in various species. The database is hosted by NCBI in collaboration with the National Human Genome Research Institute (http://www.ncbi.nlm.nih.gov/sites/entrez?db=snp).

3.4.1.3. GeneCards. GeneCards is an online searchable database that contains all human genes with their available features in topics of transcriptomic, genetic, proteomic, functional and disease information from various integrated sources. It provides user-friendly access and hyperlinks to other databases to achieve more information (http://www.genecards.org/).

3.4.1.4. International Haplotype Map Project (HapMap). The HapMap project data is a free, online database that contains common patterns in human genetic variations (SNPs, haplotypes) of four different populations (African, Asian and European ancestry). The project is funded by agencies from Japan, the United Kingdom, Canada, China, Nigeria, and the United States (http://hapmap.ncbi.nlm.nih.gov/).

3.4.1.5. University of California Santa Cruz (UCSC) Genome Browser. The UCSC Genome Browser is an online, massive database that contains more than 30 genomes of different species. Genome browser allows to explore any region interested on the chromosomes and to visualize any information belonging to the region of interest. Visualization and analysis tools are present under tracks, such as genes and gene prediction, expression, regulation and variation and repeats. The browser is hosted by the UCSC (http://genome.ucsc.edu).

3.4.1.6. Database of Genomic Variants (DGV). The Database of Genomic Variants is another online database that contains comprehensive summary of structural variations of the human genome. The Database includes only the structural variations identified in healthy control samples. DGV is supported by the Genome Canada/Ontario Genomics Institute, the McLaughlin Centre for Molecular Medicine, the Wellcome Trust and the Canadian Institutes for Health Research (http://projects.tcag.ca/variation/).

3.4.1.7. ALS Online Genetic Database (ALSoD). ALSoD is the only database that contains all the published information about ALS. It is funded by ALS Association, ALS Society of Canada, Ingenuity Systems, MND Association in Iceland and coordinated by Prof. Ammar Al-Chalabi. Several well-known laboratories throughout the world are involved in gathering information to the database (http://alsod.iop.kcl.ac.uk/misc/labs.aspx).

3.4.2. Bioinformatic Tools

3.4.2.1. PLINK. This software is a free, open source whole genome association analysis toolset that performs basic and large-scale computational genotype/phenotype correlation analysis and rearrangements. The software was developed by Shaun Purcell at the Center for Human Genetic Research, Massachusetts General Hospital (MGH) and the Broad Institute of Harvard and MIT (http://pngu.mgh.harvard.edu/purcell/plink/).

3.4.2.2. Haploview 4.2 version. The Haploview software performs LD and haplotype block analysis by using SNP information. Additional tests are also included for association and multiple testing. The software was developed at Mark Daly's lab at the Broad Institute by Jeffrey Barrett, David Bender, Julian Maller and Jesse Whitworth (http://www.broad.mit.edu/haploview/haploview).

3.4.2.3. SNP Annotation and Proxy Search (SNAP). SNAP is an online software that finds the proxy SNPs according to linkage disequilibrium and the distance between them. It correlates linkage between SNPs presence in different commercial genotyping arrays. Pair-wise linkage disequilibrium is established based on genotype data from the International HapMap Project and 1000 Genome Projects. The software was developed by Robert Handsaker and Andrew Johnson from Broad Institute (http://www.broadinstitute.org/mpg/snap/).

3.4.2.4. PennCNV: Copy Number Variation Detection. PennCNV is a free tool for detecting Copy Number Variations from SNP genotyping arrays. The software uses signal intensity data from Illumina and Affymetrix arrays. It detects CNV calls using a hidden

Markov model (HMM) algorithm for genotyped individuals. The software was developed by Wang Kai from the University of Pennsylvania (http://www.openbioinformatics.org/penncnv/).

4. METHODS

The experimental strategy of this study is summarized in Figure 4.1.

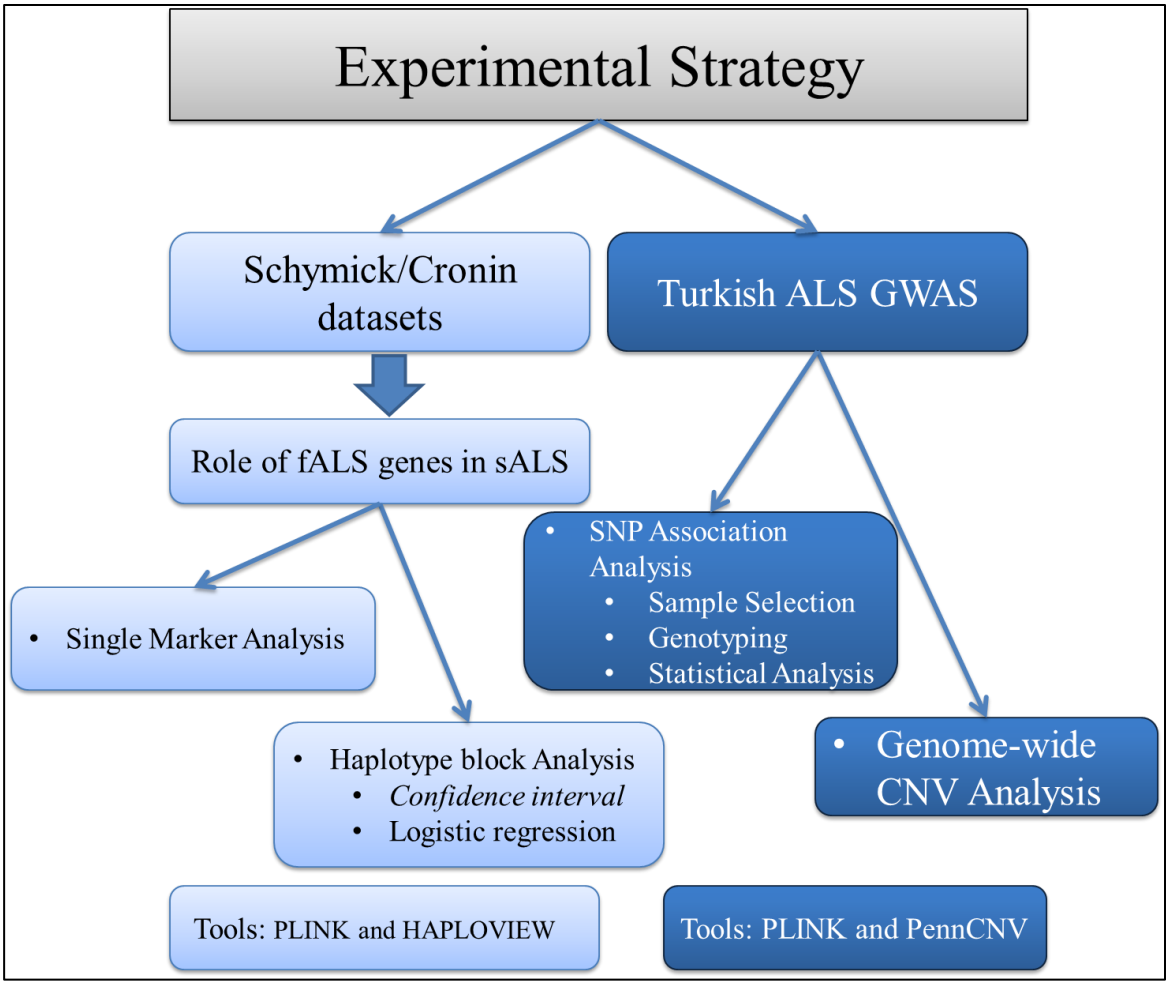


Figure 4.1. The experimental strategy followed in this work. In the first part of this study, the experimental flow indicated with light blue was followed. Then the dark blue part was performed.

4.1. Study Design of the US and Irish Datasets

Raw genotype data of the Cronin study (Cronin *et al.*, 2007) consisted of 221 sALS cases and 211 neurologically normal controls from the Irish population. The Schymick study (Schymick *et al.*, 2007) included 276 sALS cases and 268 neurologically normal controls from the US population. Any individuals that have genotype rate lower than 99%

in the Schymick study, and 99% in the Cronin study were excluded from datasets. Also, SNPs that have failed in missingness test, with a call rate lower than 99%, were excluded from both data. SNPs in each data were tested by HWE and were required to have a p-value equal or higher than 0.05 and a MAF equal or greater than 0.01. As a result, a total of 17 individuals in Schymick and 9 individuals in Cronin datasets were excluded from this study due to low genotyping rates. 26685 and 23324 SNPs failed in missingness test, respectively. A total of 527801 SNPs in the Schymick and 538142 SNPs in the Cronin datasets were able to pass the criteria to be used in single marker and haplotype analyses applied in this study.

4.2. Single Marker and Haplotype Association Analyses

SNPs present in both datasets were synchronized and included in analysis. After synchronization, allelic and genotypic association tests were computed for each SNP using PLINK, and summary statistics were generated.

All the SNPs located within the genomic regions of the candidate genes and the 10kb flanking regions from both ends were retrieved using PLINK version 1.07 (Purcell *et al.*, 2007). SNPs were used for haplotype-based case-control association tests, using PLINK and HAPLOVIEW version 4.2 (Barrett *et al.*, 2005) . Haplotypes within each gene region were defined by using confidence intervals (Gabriel *et al.*, 2002). To test blocks, logistic regression and OMNIBUS were used to perform haplotype-based association tests. Frequency in samples, ORs, and asymptotic p-values of each block were calculated. Different combinations of blocks generated were tested in both datasets and the tests resulted in relatively precise and consistent p-values. To validate results obtained from the Schymick and Cronin studies, the procedure above was applied to Turkish samples.

4.3. Fine Mapping of SNPs and Haplotypes

The USCS Genome Browser was used to examine SNPs within gene loci. Location of each SNP and region of haplotypes, covering certain SNPs, were also determined (5’ and 3’ UTRs, intergenic, intronic and exonic regions).

4.4. The Turkish Genome-wide Association Study

All DNA samples of patient and control individuals were SNP-genotyped at the Ozcelik Lab, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada. Among a total of 260 Turkish individuals, 225 were included in the final analyses (116 sALS patients and 109 Turkish controls). Genotyping of 15 samples had to be replicated. DNA samples which did not pass quality control were excluded from the study. All samples were assayed using the Illumina HumanOmniExpress SNP array (Illumina, CA, USA) for 733202 SNPs, selected from the HapMap Project. The raw genotype data were rearranged by the bioinformatics team of the Ozcelik lab to use appropriate input format for the PLINK software.

4.5. Statistical Analysis of Turkish GWAS

Standard quality control procedures were applied to all genotyped individuals and SNPs. Individuals with a call rate below 95% were excluded from the study population. Also SNPs that have call rates less than 95%, MAFs less than 0.01 and Hardy-Weinberg disequilibrium (HWE) p-values less than 0.05 were excluded using the PLINK toolset. For SNP association tests, each SNP was tested using allelic and genotypic models.

4.6. CNV Analysis

Genome-wide data of Turkish GWAS was used in CNV analysis. The SNP information of each individual was extracted from GWAS data. 733202 SNPs were included with Log R Ratio (LRR) and B allele frequency (BAF) values. The PennCNV software was used for quality control and CNV analysis. To remove artificial genomic waves, the genomic waviness adjustment procedure of PennCNV was applied. The generated CNV calls were filtered using confidence value to discard false positives.

5. RESULTS

5.1. Summary Statistics of the US (Schymick) and Irish (Cronin) Datasets

In this study we have taken advantage of the publicly available GWA data from two other populations reported by Schymick *et al.* (276 cases, 268 controls) (Schymick *et al.*, 2007) and Cronin *et al.* (221 cases, 211 controls) (Cronin *et al.*, 2007). Individuals with genotype rate lower than 99% in the Schymick study, and 99% in the Cronin study were excluded from the analysis. As a result, a total of 17 individuals in Schymick and 9 individuals in Cronin datasets were excluded from this study due to low genotyping rates. Also, 45136 SNPs in Schymick and 23324 SNPs in Cronin datasets that failed to have a call rate lower than 99% were also excluded. Based on the deviation from HWE p-value and MAFs, 43541 and 43425 SNPs failed, and thus excluded from the study. A total of 477045 SNPs in the Schymick and 494717 SNPs in the Cronin datasets matched the criteria, and therefore included in the analysis (Table 5.1).

Table 5.1. Summary statistics of Schymick and Cronin datasets.

	Schymick	**Cronin**
Total individuals	544	432
Individuals with 99% call rate (individuals with call rates lower than 99%)	527 (17)	423 (9)
Total SNPs	565722	561466
SNPs with call rates higher than 99% (SNPs with call rates lower than 99%)	520586 (45136)	538142 (23324)
HWE ($p< 0.05$)	20815	19025
MAF (< 0.01)	22726	24400
After cut-offs applied (SNPs passed quality control)	477045	494717

5.2. Genetic Association of fALS-causing Genes in sALS

In this part of the study, SNPs located in the region of fALS-causing genes were investigated. Eight fALS genes were included in the analysis. Genes and locations were obtained from HapMap (http://hapmap.ncbi.nlm.nih.gov/) (Table 5.2).

Table 5.2. fALS-causing genes that are included in the analysis.

No	Gene Symbol	Chr	Gene region (Hapmap) Start	Gene region (Hapmap) End	Gene region (10kb extended) Start	Gene region (10kb extended) End
1	SOD1	21	31943806	31973112	31933806	31983112
2	ALS2	2	202263522	202363983	202253522	202373983
3	SETX	9	134116649	134230193	134106649	134240193
4	FUS	16	31088954	31120598	31078954	31130598
5	VAPB	20	56387651	56465367	56377651	56475367
6	ANG	14	20212609	20242183	20202609	20252183
7	TARDBP	1	10985266	11018135	10975266	11028135
8	DCTN1	2	74431790	74471472	74421790	74481472

In eight gene loci, a total of 106 SNPs from 8 fALS genes that were common in both Schymick and Cronin datasets were extracted (Figure 5.3). A total of four SNPs in Schymick and two SNPs in Cronin datasets were eliminated since they either failed the HWE test ($p< 0.05$) or had a low MAF (< 0.01).

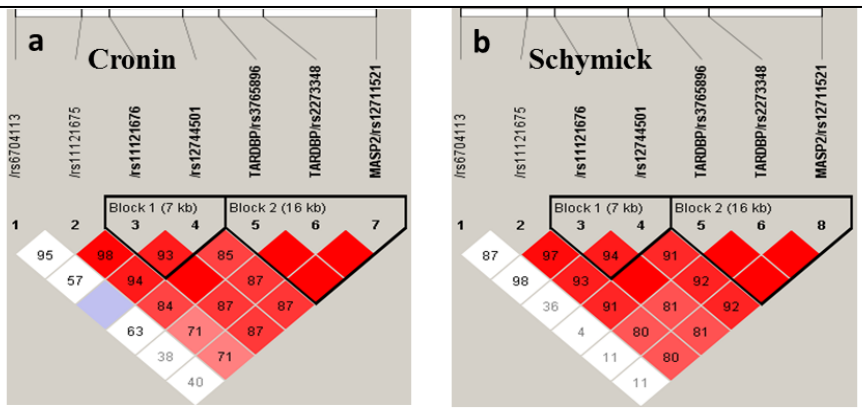
Table 5.3. fALS genes and their common SNPs in Schymick and Cronin datasets.

			Number of SNPs		After MAF (< 0.01) and HWE (< 0.05) Number of SNPs		
Gene	**Chr Location**	**Genomic size (bp)**	**Schymick**	**Cronin**	**Schymick**	**Cronin**	**Common SNPs**
SOD1	21q22.11	9310	4	6	4	6	4
ALS2	2q33.2	80927	29	32	26	32	23
SETX	9q34.13	9363	30	30	30	30	28
FUS	16p11.2	11683	3	3	2	1	1
VAPB	20q13.33	61980	24	25	24	25	24
ANG	14q11.1	10010	15	16	15	16	14
TARDBP	1p36.22	13383	8	7	8	7	7
DCTN1	2p13	30934	5	5	5	5	5
Total	-	-	118	124	114	122	106

5.2.1. Single marker analysis

Among 106 SNPs investigated, only 11 SNPs in three genes (TARDBP, DCTN1, VAPB) in the Schymick dataset, and eight SNPs in five genes (TARDBP, DCTN1, VAPB,

ANG, SOD1) in the Cronin dataset have significant associations with a p-value <0.05, whereas both were observed to be as non-significant in the second dataset. However, we have identified two SNPs, rs2273348, p_s=0.0098; p_c=0.0215 and rs12711521, p_s=0.0098; p_c=0.0278, within TARDBP region which were significantly associated with sALS risk in both Schymick and Cronin datasets, respectively (Figure 5.1c and 5.1d).



c

SNP	Assoc. Allele	Case, Control Ratios	Chi Square	p-value
rs6704113	A	0.986, 0.981	0.32	0.5715
rs11121675	T	0.685, 0.664	0.432	0.5112
rs11121676	C	0.843, 0.792	3.658	0.0558
rs12744501	C	0.347, 0.345	0.002	0.9644
rs3765896	G	0.164, 0.124	2.812	0.0935
rs2273348	T	0.197, 0.138	5.286	0.0215
rs12711521	C	0.197, 0.140	4.842	0.0278

d

SNP	Assoc. Allele	Case, Control Ratios	Chi Square	p-value
rs6704113	A	0.974, 0.964	0.898	0.3433
rs11121675	T	0.700, 0.638	4.463	0.0346
rs11121676	C	0.852, 0.838	0.408	0.5228
rs12744501	T	0.688, 0.605	7.987	0.0047
rs3765896	A	0.847, 0.804	3.291	0.0696
rs2273348	C	0.836, 0.773	6.673	0.0098
rs12711521	A	0.836, 0.773	6.673	0.0098

e

Haplotype	Freq.	Case, Control Ratios	Chi Square	p-value
Block 1				
CT	0.647	0.650, 0.644	0.042	0.8368
TC	0.175	0.154, 0.197	2.77	0.0961
CC	0.17	0.192, 0.148	2.943	0.0863
TT	0.007	0.003, 0.011	1.688	0.1938
Block 2				
ACA	0.831	0.803, 0.860	4.842	0.0278
GTC	0.144	0.164, 0.124	2.812	0.0935
ATC	0.024	0.033, 0.014	3.162	0.0754
ACC	0.001	0.000, 0.002	1.015	0.3136

f

Haplotype	Freq.	Case, Control Ratios	Chi Square	p-value
Block 1				
CT	0.642	0.679, 0.601	6.962	0.0083
CC	0.203	0.173, 0.237	6.599	0.0102
TC	0.149	0.139, 0.159	0.788	0.3748
TT	0.006	0.009, 0.003	1.199	0.2735
Block 2				
ACA	0.806	0.836, 0.773	6.673	0.0098
GTC	0.174	0.153, 0.196	3.291	0.0696
ATC	0.021	0.011, 0.032	5.501	0.019

Figure 5.1. Haplotype block analysis of the TARDP region using confidence intervals. Haplotype blocks of TARDBP in (a) Cronin and (b) Schymick. Single marker associations of SNPs of TARDBP in (c) Cronin and (d) Schymick datasets, haplotypes observed using Haploview in (e) Cronin and (f) Schymick datasets. Red indicates significant haplotypes (p<0.05).

5.2.2. Haplotype block analysis

Since the single marker analysis has not shown informative and consistent associations between the datasets, we have applied haplotype analysis to find more appropriate genetic profiles associated with sALS. Haplotypes within each gene region were defined using confidence intervals (Gabriel *et al.*) by Haploview tool (Figure 5.1e and 5.1f). To test defined blocks, logistic regression and omnibus tests were used to perform haplotye-based association analysis for all genes studied via the PLINK tool. Frequency in samples, ORs and asymptotic p-values of each block were calculated in both datasets.

Among 72 different haplotypes and 19 different blocks in the Cronin dataset, only four haplotypes in TARDBP, ANG and VAPB genes and two of those haplotype blocks showed significance with $p<0.05$ (Table 5.4a). Among 72 different haplotypes and 17 different blocks in the Schymick dataset, a total of nine haplotypes in TARDBP, DCTN1, ALS2 and VAPB genes and four of those haplotype blocks showed significance with a $p<0.05$ (Table 5.4b). In the region of TARDBP, VAPB, ANG, DCTN1 and ALS2 genes, few haplotypes showed high ORs with $p<0.05$ with no replication/validation in both datasets.

Remarkably, a 16kb haplotype block covering TARDP region between chromosome 1, positions between 10997007-11013503 (flanked by SNPs rs3765896 and rs12711521 found to be significantly associated with sALS in both datasets (Table 5.4a and 5.4b). Furthermore the haplotype ACA showed a significantly increased risk of ALS (OR=1.5 and p=0.00862) in the Schymick dataset, however, the same haplotype was found to have a significantly decreased risk of ALS (OR=0.649 and p=0.0247) in the Cronin dataset.

Table 5.4. Haplotype-based association test using logistic regression and omnibus, (a) Cronin and (b) Shymick datasets.

a

Gene region	Number of SNPs	Chr	Start position	End position	Start SNP	End SNP	Haplotype	F	Logistic Regression			Omnibus	
									OR	STAT	p-value	STAT	p-value
TARDBP	3	1	10997007	11013503	rs3765896	rs12711521	ACA	0.831	0.649	5.04	0.0247	6.05	0.0485
ANG	5	14	20219909	20226288	rs17277571	rs8008440	GAAGG	0.483	0.741	5.01	0.0252	5.88	0.208
VAPB	2	20	56388714	56389876	rs6015260	rs713406	CT	0.398	0.711	5.55	0.0185	7.09	0.0289
VAPB	12	20	56394865	56456968	rs6092640	rs4549163	AGGCTCCAGTCA	0.0993	1.74	5.18	0.0228	8.22	0.222

b

Gene region	Number of SNPs	Chr	Start position	End position	Start SNP	End SNP	Haplotype	F	Logistic Regression			Omnibus	
									OR	STAT	p-value	STAT	p-value
TARDBP	2	1	10985630	10993239	rs11121676	rs12744501	CC	0.203	0.689	6.12	0.0134	7.62	0.0222
TARDBP	2	1	10985630	10993239	rs11121676	rs12744501	CT	0.642	1.4	6.78	0.00923	7.62	0.0222
TARDBP	3	1	10997007	11013503	rs3765896	rs12711521	ATC	0.0209	0.332	5.14	0.0234	8.74	0.0126
TARDBP	3	1	10997007	11013503	rs3765896	rs12711521	ACA	0.806	1.49	6.42	0.0113	8.74	0.0126
DCTN1	2	2	74432044	74448522	rs3771744	rs909177	AA	0.129	0.658	4.81	0.0283	4.63	0.0987
ALS2	2	2	202273815	202275009	rs11300	rs3731707	GT	0.155	0.679	4.86	0.0275	4.86	0.0881
ALS2	9	2	202329233	202365438	rs10211216	rs7597686	CCACAATGG	0.146	0.686	4.47	0.0344	6.51	0.26
VAPB	2	20	56379407	56382611	rs6064647	rs6015258	GC	0.261	0.73	4.81	0.0283	4.89	0.0866
VAPB	6	20	56456968	56465163	rs4549163	rs5007291	GCGCCG	0.134	0.644	5.6	0.018	9.86	0.197

5.2.3. Fine mapping of TARDP region

Contradictory OR values in the two datasets led to further investigation of the TARDBP region. We have observed that the frequency of the ACA haplotype, as well as the individual SNPs within that region, differed in cases and controls in both datasets. To understand the differences, the frequency of each single SNP was examined, and the frequencies of minor alleles of SNPs and fold changes between cases and controls of datasets were compared and investigated (Figure 5.2).

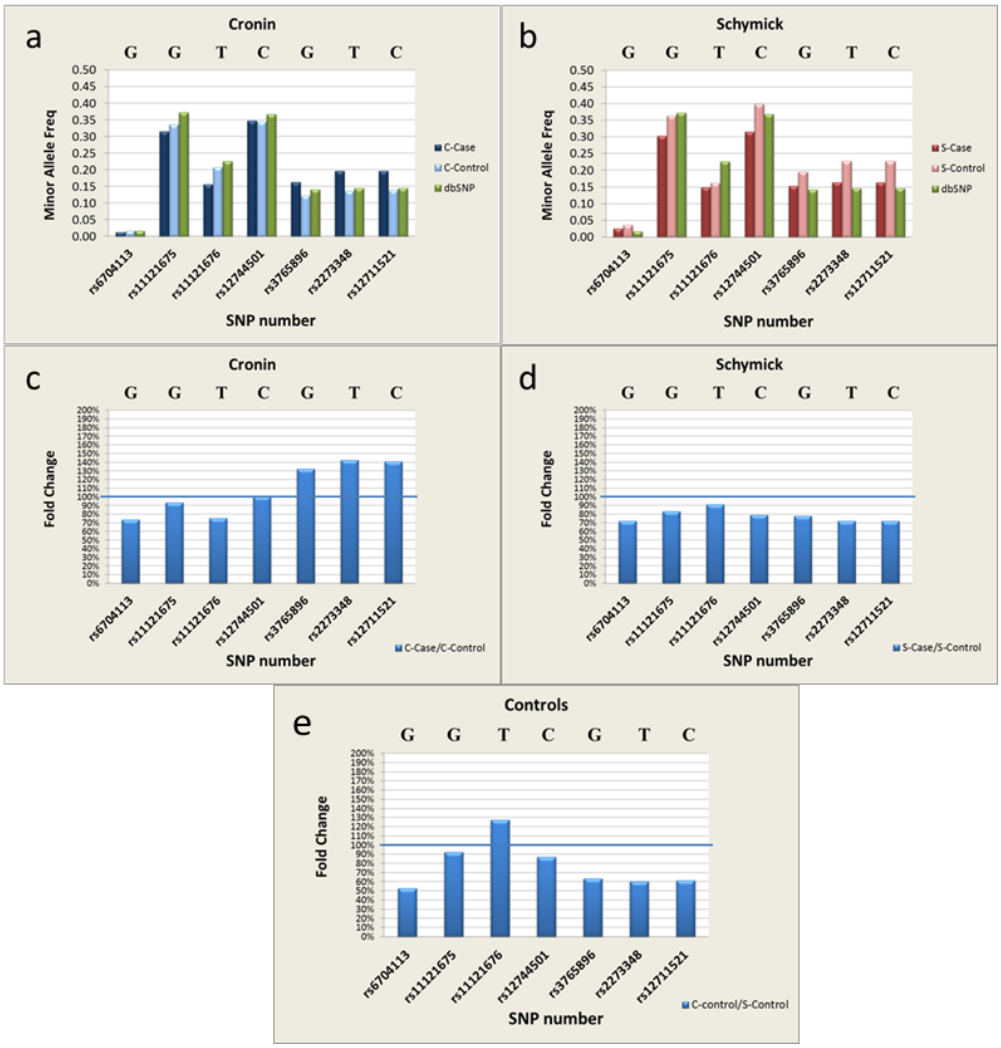


Figure 5.2. Minor allele frequencies of SNPs located in the TARDBP region from cases and controls of Cronin and Schymick datasets and fold changes between cases and controls. MAFs of SNPs located in TARDBP region of cases, controls and CEU population (dbSNP) (a) in Cronin (b) in Schymick datasets. and fold changes between cases and controls of (c) Cronin and (d) Schymick datasets. (e) Fold change between controls of Cronin and Schymick datasets.

The fold changes (MAF of cases observed/MAF of controls observed) of cases and controls among rs6704113, rs11121675, rs11121676 in both datasets displayed a slight decrease. However, fold changes among rs3765896, rs2273348, rs12711521 indicated opposite results (Figure 5.2a-d). Frequencies of controls of datasets significantly differed in those SNPs that were involved in the ACA haplotype block (Figure 5.2e). Decreased presence of haplotype frequency was observed in sALS cases of Cronin dataset. On the other hand, the same haplotype block had frequencies of 0.836 in cases and 0.773 in controls which showed an increase in cases of Schymick dataset (Figure 5.1c-d). To find the contradictory frequencies of certain SNPs in the TARDBP locus, whether it was caused by loci of interest or population variety, a larger locus including the TARDBP locus was investigated. The locus between positions 10900kb and 11150kb was plotted using the Hapmap Genome Browser (release 27). Both datasets showed similar patterns with LD of CEU population as expected (Figure 5.3).

In total, datasets contained 37 common SNPs. Their compact LD views are shown in Figure 5.4a and 5.5a. The observed minor allele frequency of each SNP in cases versus controls and the MAF of each SNP in the Hapmap Project of CEU population were plotted (Figure 5.4b and 5.5b) and also fold changes between cases and controls were calculated and plotted (Figure 5.4c and 5.5c). The fold changes in the Cronin study demonstrated significant alterations starting from TARDBP locus and ending at 11150kb. An increase of ~40% observed in rs3765896, rs2273348, rs12711521 which were included in the ACA haplotype and the 40% increased fold change was observed till end of the locus. The fold change was not altered dramatically at first half of the whole locus.

Like Cronin, the Schymick dataset showed significant alterations (decrease) starting at the TARDBP locus and continued to the end of rs17036350. The fold changes of the first half of the locus demonstrated slight increases and decreases around 10%, however, the second half showed a dramatic decrease up to 50%. It seemed that first one half of the large locus in chromosome 1, the maintenance was apparently preserved and the SNPs frequencies were relatively conserved in both datasets. At second half of the locus, starting from TARDBP promoter region, significant alterations, decreases and increases were observed which support unstable chromosome loci in that region of chromosome 1.

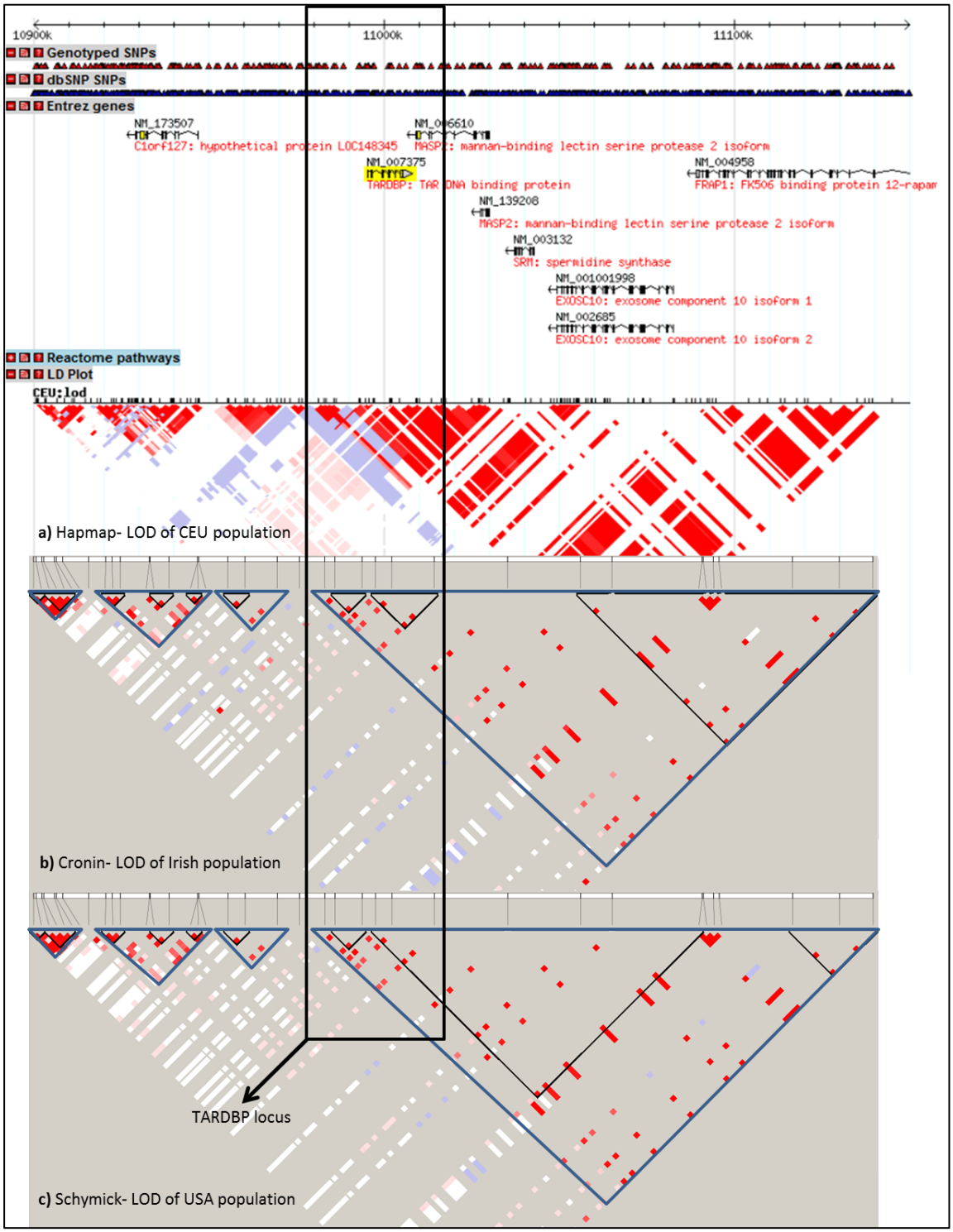


Figure 5.3. LD plots of genotyped SNPs in the extended region of TARDBP in three populations. (a) LD of CEU population from Hapmap, (b) LD of Cronin dataset and (c) LD of Schymick dataset. Blue triangles indicate similar LOD patterns in Hapmap, Cronin and Schymick. Black triangle displays the region of TARDBP gene.

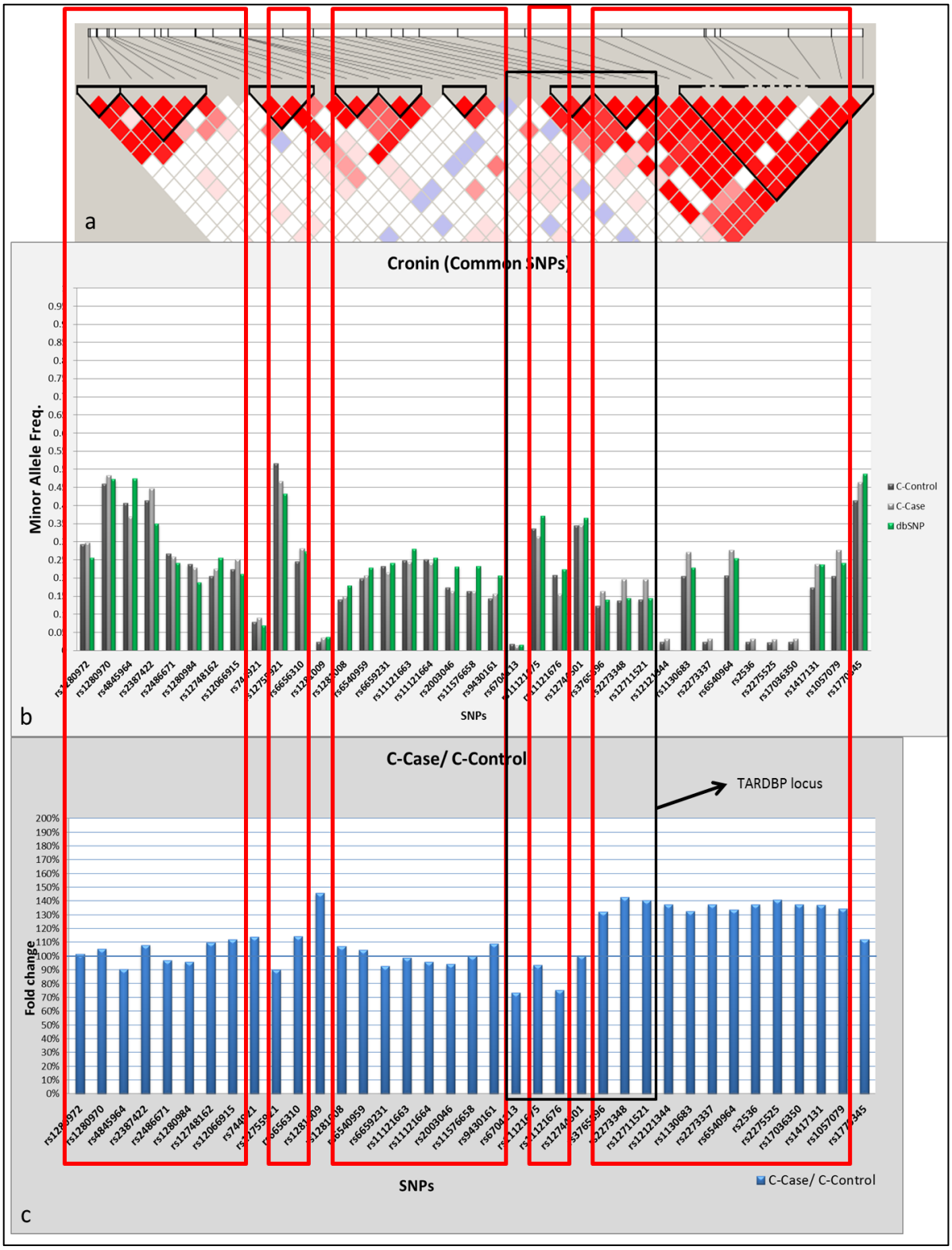


Figure 5.4. Fine mapping analysis of a 250kb region covering TARDBP locus in Cronin dataset. (a) LD plots and haplotypes, (b) MAFs of 37 common SNPs of cases, controls and CEU population, (c) fold changes between cases and controls. Red triangles indicate altered MAFs observed between cases and controls.

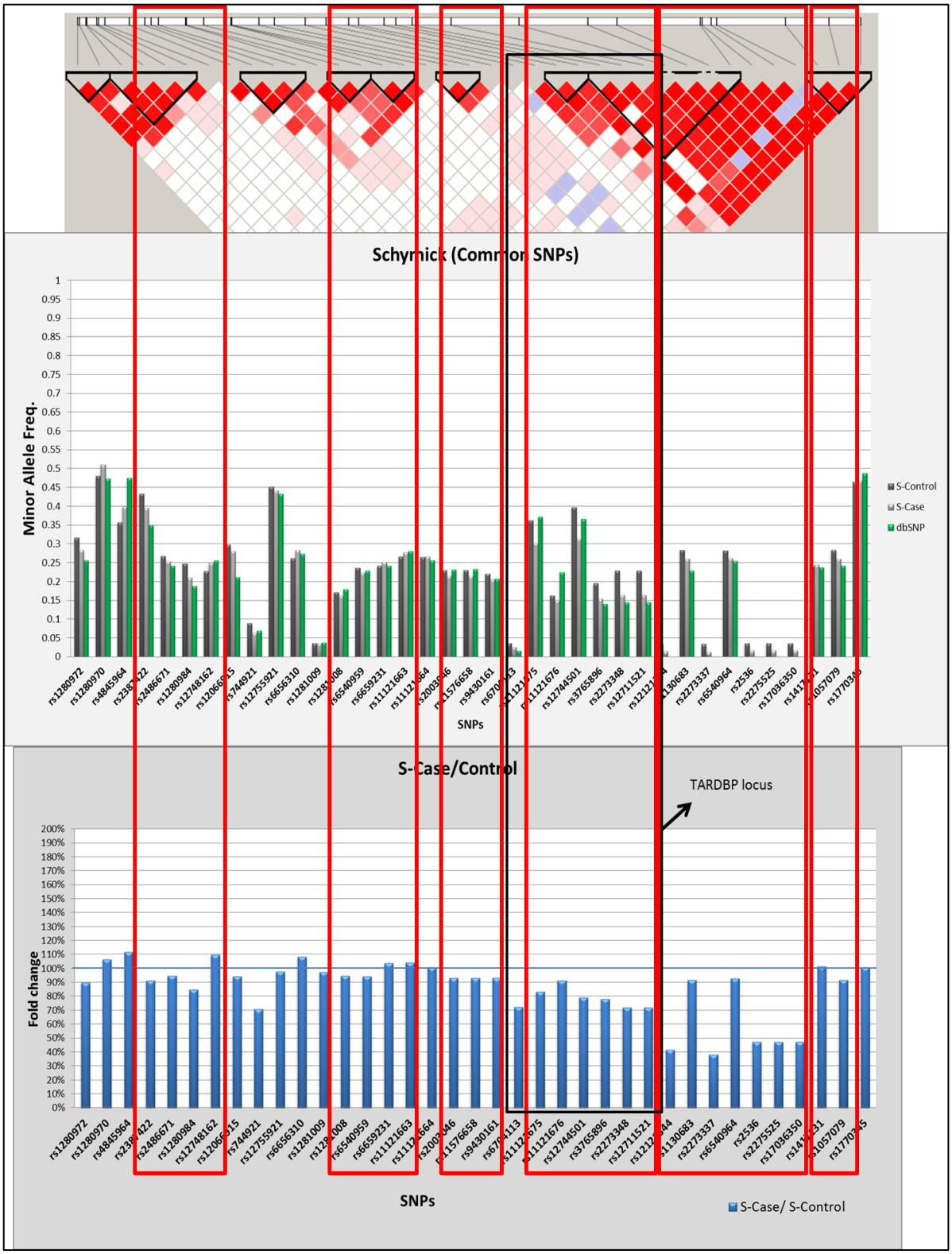


Figure 5.5. Fine mapping analysis of a 250kb region covering TARDBP locus in Schymick dataset. (a) LD plots and haplotypes, (b) MAFs of 37 common SNPs of cases, controls and CEU population, (c) fold changes between cases and controls. Red triangles indicate altered MAFs observed between cases and controls.

5.3. Statistical Analysis of the Turkish GWAS

5.3.1. Summary statistics of the Turkish GWAS

The Turkish GWAS study was performed with 225 individuals including 116 ALS patients and 109 healthy controls. A total of 733202 SNPs from each individual were genotyped using the Illumina HumanOmniExpress SNP array platform. Standard quality control parameters were applied to all individuals and SNPs using PLINK toolset. Two individuals with a call rate lower than 95% were excluded from the study. Total genotyping rate of the remaining individuals was 0.9977. Also, 4137 SNPs with call rates less than 95% were removed from the study. In addition, 97931 SNPs were eliminated according to HWE ($p < 0.05$) and MAF (< 0.01) criteria. In total, 631134 SNPs passed the standard quality control procedure and were included in association analysis (Table 5.5).

Table 5.5. Summary statistics of the Turkish GWAS.

Total individuals	225
Individuals with 99% call rate (Individuals with call rates lower than 99%)	223 (2)
Total SNPs	733,202
SNPs with call rates higher than 95% (SNPs with call rates lower than 95%)	729,065 (4,137)
HWE ($p < 0.05$)	27,195
MAF (< 0.01)	70,736
After cut-offs applied (SNPs passed quality control)	631,134

5.3.2. Allelic and genotypic association test results of the Turkish GWAS

In total, 631134 out of 733202 SNPs passed the quality control tests. Allelic and genotypic association tests were performed with these qualified SNPs. The Manhattan plot distribution of SNP associations was displayed (Figure 5.6). Also top significant (lowest p-value) SNPs found in both allelic and genotypic association tests were summarized (Table 5.2 and 5.3). The SNP with lowest p-value 4.95×10^{-7} in allelic test was rs6616711 and in genotypic test with $p=3.84 \times 10^{-7}$ was rs4863387. Although those SNPs were highly

significant, they failed to keep the significance ($p<0.05$) after applying multiple testing such as Bonferroni correction, Benjamini & Hochberg (1995) step-up FDR control (FDR_BH) and Benjamini & Yekutieli (2001) step-up FDR control (FDR_BY) procedures. At SNP level, according to multiple corrections, none of the SNPs were found to be associated with ALS, yet by looking at association tests, SNPs located in two gene regions, CPNE5 and C20ORF39 (TMEM90B), were shown to have the strongest, although non-significant, association with ALS (Table 5.6 and 5.7).

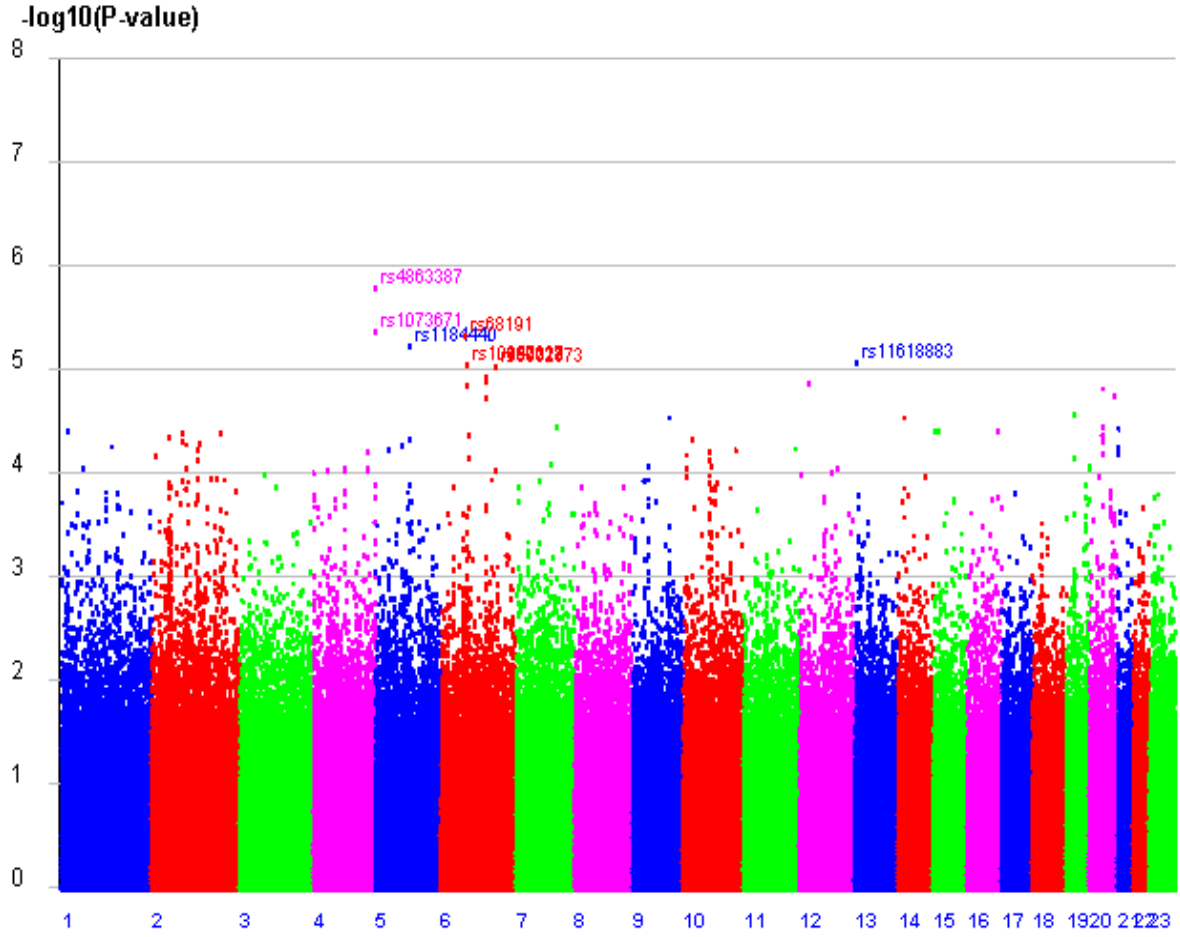


Figure 5.6. p-values from the Turkish genome-wide association study. 114 Turkish ALS patients and 109 healthy controls were included in this study. Each dot represents a SNP on the chromosomes. The SNPs with rs-numbers had the lowest p-values. Horizontal axis indicates chromosomes and positions, 23 being the X chromosome. Vertical axis shows -$\log_{10}$ p-values of SNPs.

Table 5.6. SNPs with lowest p-values and their adjusted p-values in allelic test.

#	CHR	SNP	UNADJ	GC	BONF	FDR_BH	FDR_BY
1	23	/rs6616711	4.95E-07	1.28E-06	0.3121	0.3121	1
2	4	/rs4863387	1.33E-06	3.22E-06	0.8406	0.4203	1
3	6	/rs68191	2.74E-06	6.30E-06	1	0.557	1
4	4	/rs1073671	3.53E-06	7.97E-06	1	0.557	1
5	5	/rs1184440	5.66E-06	1.24E-05	1	0.6093	1
6	13	/rs11618883	6.99E-06	1.51E-05	1	0.6093	1
7	23	/rs5913336	8.17E-06	1.74E-05	1	0.6093	1
8	6	CPNE5/rs10947628	9.11E-06	1.93E-05	1	0.6093	1
9	6	CPNE5/rs13193217	9.48E-06	2.00E-05	1	0.6093	1
10	6	/rs6932373	1.02E-05	2.13E-05	1	0.6093	1
11	23	/rs5918294	1.06E-05	2.22E-05	1	0.6093	1
12	6	CPNE5/rs916295	1.46E-05	2.99E-05	1	0.6704	1
13	20	TMEM90B/rs4815285	1.54E-05	3.14E-05	1	0.6704	1
14	12	/rs12580191	1.71E-05	3.47E-05	1	0.6704	1
15	20	/rs6123735	1.88E-05	3.78E-05	1	0.6704	1
16	19	/rs7247623	2.23E-05	4.42E-05	1	0.6704	1
17	23	NSBP1/rs1954611	2.83E-05	5.52E-05	1	0.6704	1
18	23	/rs17328555	2.83E-05	5.52E-05	1	0.6704	1
19	2	CD207/rs3755348	2.94E-05	5.73E-05	1	0.6704	1
20	20	TMEM90B/rs6138332	3.27E-05	6.33E-05	1	0.6704	1
21	2	/rs1562256	3.35E-05	6.46E-05	1	0.6704	1
22	1	/rs4113453	3.44E-05	6.64E-05	1	0.6704	1
23	9	CTNNAL1/rs866311	3.44E-05	6.64E-05	1	0.6704	1
24	20	TMEM90B/rs1883924	3.46E-05	6.66E-05	1	0.6704	1
25	15	RYR3/rs12911184	3.77E-05	7.21E-05	1	0.6704	1
26	2	CD207/rs4852708	3.88E-05	7.41E-05	1	0.6704	1
27	21	/rs2823121	3.89E-05	7.42E-05	1	0.6704	1
28	5	/rs2967099	4.21E-05	8.00E-05	1	0.6704	1
29	2	/rs42799	4.22E-05	8.01E-05	1	0.6704	1
30	20	TMEM90B/rs6049781	4.25E-05	8.08E-05	1	0.6704	1

* Red indicates SNPs located in the same gene region
* UNADJ: Unadjusted p-value
* GC: Genomic-control corrected p-values
* BONF: Bonferroni single-step adjusted p-values
* FDR_BH: Benjamini & Hochberg (1995) step-up FDR control
* FDR_BY: Benjamini & Yekutieli (2001) step-up FDR control

Table 5.7. SNPs with the lowest p-values in genotypic test.

#	CHR	SNP	Allele 1	Allele 2	Chi Square	P-value
1	4	/rs4863387	G	A	24.94	3.84E-06
2	20	TMEM90B/rs4815285	A	G	24.72	4.30E-06
3	4	/rs1073671	A	G	21.75	1.90E-05
4	20	TMEM90B/rs4815278	A	G	21.48	2.17E-05
5	5	/rs1184440	G	A	21.27	2.41E-05
6	20	TMEM90B/rs6049781	A	G	21.16	2.54E-05
7	4	/rs2877520	A	G	21.04	2.71E-05
8	20	TMEM90B/rs1883924	A	G	20.87	2.94E-05
9	6	CPNE5/rs10947628	G	A	19.85	4.90E-05
10	20	TMEM90B/rs6138330	A	G	19.58	5.59E-05
11	6	CPNE5/rs916295	A	C	19.42	6.07E-05
12	2	CTNNA2/rs11677290	G	A	19.06	7.26E-05
13	20	TMEM90B/rs4815286	A	G	18.76	8.44E-05
14	20	TMEM90B/rs6138332	A	G	18.67	8.82E-05
15	2	CTNNA2/rs11695685	A	G	18.59	9.21E-05
16	19	/rs7247623	C	A	18.57	9.29E-05
17	6	CPNE5/rs13193217	G	A	18.48	9.68E-05
18	14	/rs1205103	G	A	9.558	0.008403
19	4	/rs4588426	G	A	9.556	0.008414
20	4	/rs3797042	G	A	9.556	0.008414
21	3	PDIA5/rs1969262	G	A	9.555	0.008415
22	1	PTGER3/rs2182325	A	G	9.555	0.008419
23	21	/rs2850030	C	G	9.551	0.008434
24	14	/rs1956299	A	G	9.551	0.008435
25	6	/rs10214692	A	G	9.544	0.008463
26	16	/rs1465452	A	G	9.543	0.008467
27	12	GALNT6/rs2288369	G	A	9.543	0.008467
28	2	ATP6V1B1/rs11695103	G	A	9.54	0.008482
29	6	FYN/rs9372311	G	A	9.538	0.008489
30	8	NRG1/rs7002063	G	A	9.538	0.008489

* Red indicates SNPs located in the same gene region

5.3.3. Haplotype block analysis of C20ORF39 (TMEM90B) and CPNE5 gene regions

According to single marker association tests, many of the SNPs in the TMEM90B and CPNE5 gene regions were shown to be among the top 30 significant (Table 5.6 and 5.7). The SNPs genotyped in the region of these genes and their flanking regions (200kb from each side) were extracted. All these SNPs with p-values (obtained in allelic test) were plotted using LocusZoom tool (http://csg.sph.umich.edu/locuszoom/). In addition, recombination rates and linkage disequilibrium of those regions were displayed by information from CEU (Caucasian European) population of 1000 genomes project (Figure 5.7a and 7b).

In the TMEM90B gene, a certain SNP cluster located between position 24.4 Mb and 24.6 Mb on chromosome 20 gave higher p-value signals. The p-value signals of the neighboring regions of the gene were drastically low (Figure 5.7a). Like TMEM90B, in the CPNE5 gene, a certain region between 36.8 Mb and 36.9 Mb on chromosome 6 was shown to have higher p-value signals (Figure 5.7b). These genes represent significant difference when cases and controls were compared. These regions were further investigated via haplotype block analysis using Haploview and PLINK tools.

5.3.3.1. Haploview results using confidence intervals (Gabriel *et al.*, 2002). In haplotype block analysis, SNPs located in TMEM90B and CPNE5 were extracted via PLINK and analyzed by Haploview. 55 SNPs in TMEM90B and 77 SNPs in CPNE5 gene were included in the analyses. 11 and 14 blocks were defined via confidence intervals in TMEM90B and CPNE5, respectively (Figure 5.8a and 8b).

In the region of TMEM90B, among 11 blocks, seven of them had significant haplotypes (Figure 5.9). In block 4, the AAGG haplotype showed a highly significant p-value: 7.71×10^{-5} and retained its significance after 10000 permutations ($p=0.0014$). The GA haplotype in block 5, GAAA in block 6, CGGCAG in block 7, GGGA and AAGG in block 8, AGAA and AGAG in block 9 and GCGG in block 10 were found to show high significance, and they were still significant ($p<0.05$) even after 10000 permutations applied (Figure 5.9a, 9b and 9c). The significant blocks were present between the middle part and

the end of the gene where SNPs with lowest p-values were located and gave a peak in that region.

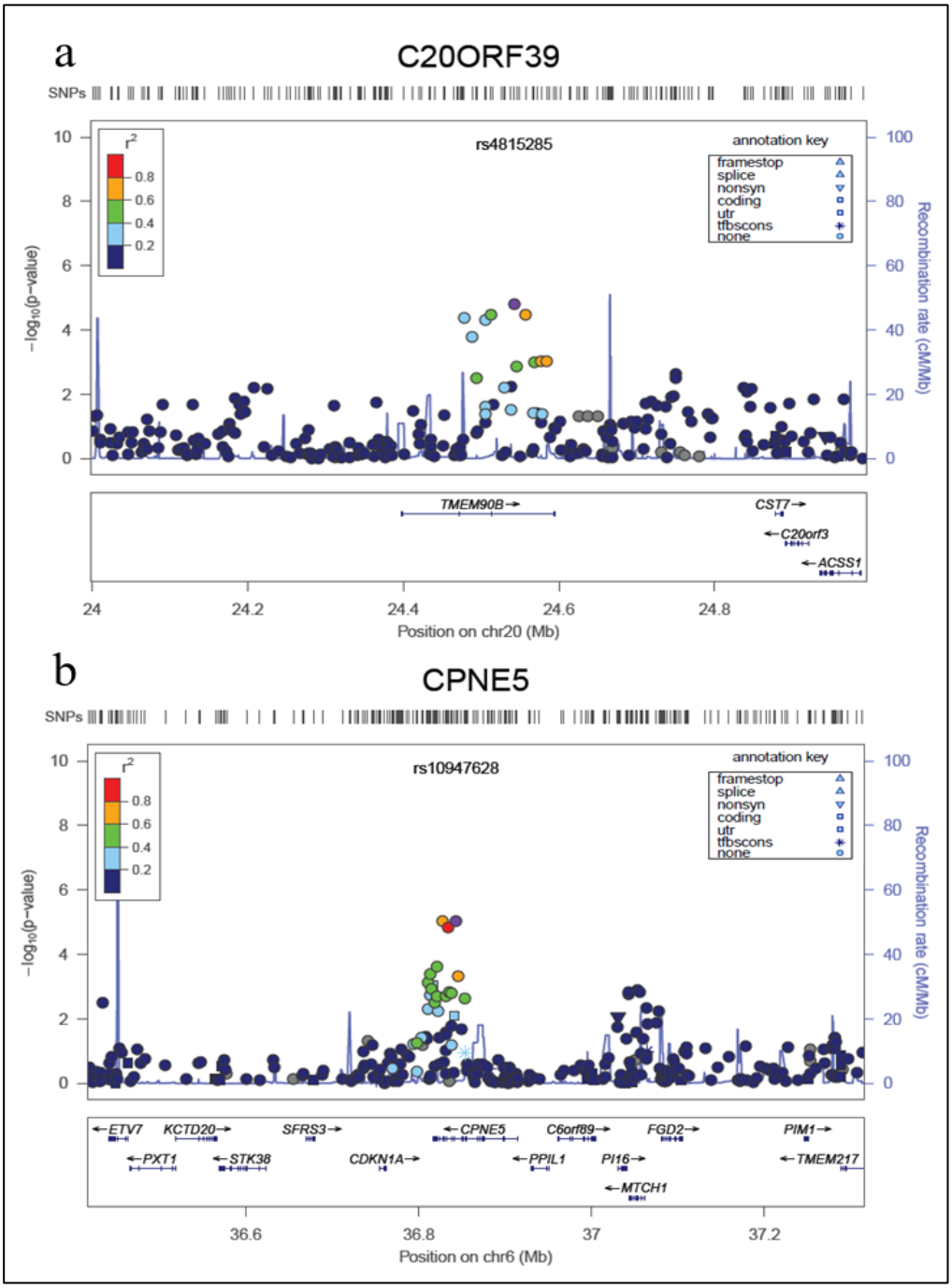


Figure 5.7. Allelic association test results using LocusZoom. (a)Association results of SNPs located in the TMEM90B gene region. (b) Association results of SNPs located in the CPNE5 gene region. Horizontal axis demonstrates chromosome positions and genes located in the region. Left vertical axis shows -$\log_{10}$p-values of SNPs. Right vertical axis, blue line in graph, indicates recombination rates in the region of chromosome in CEU population.

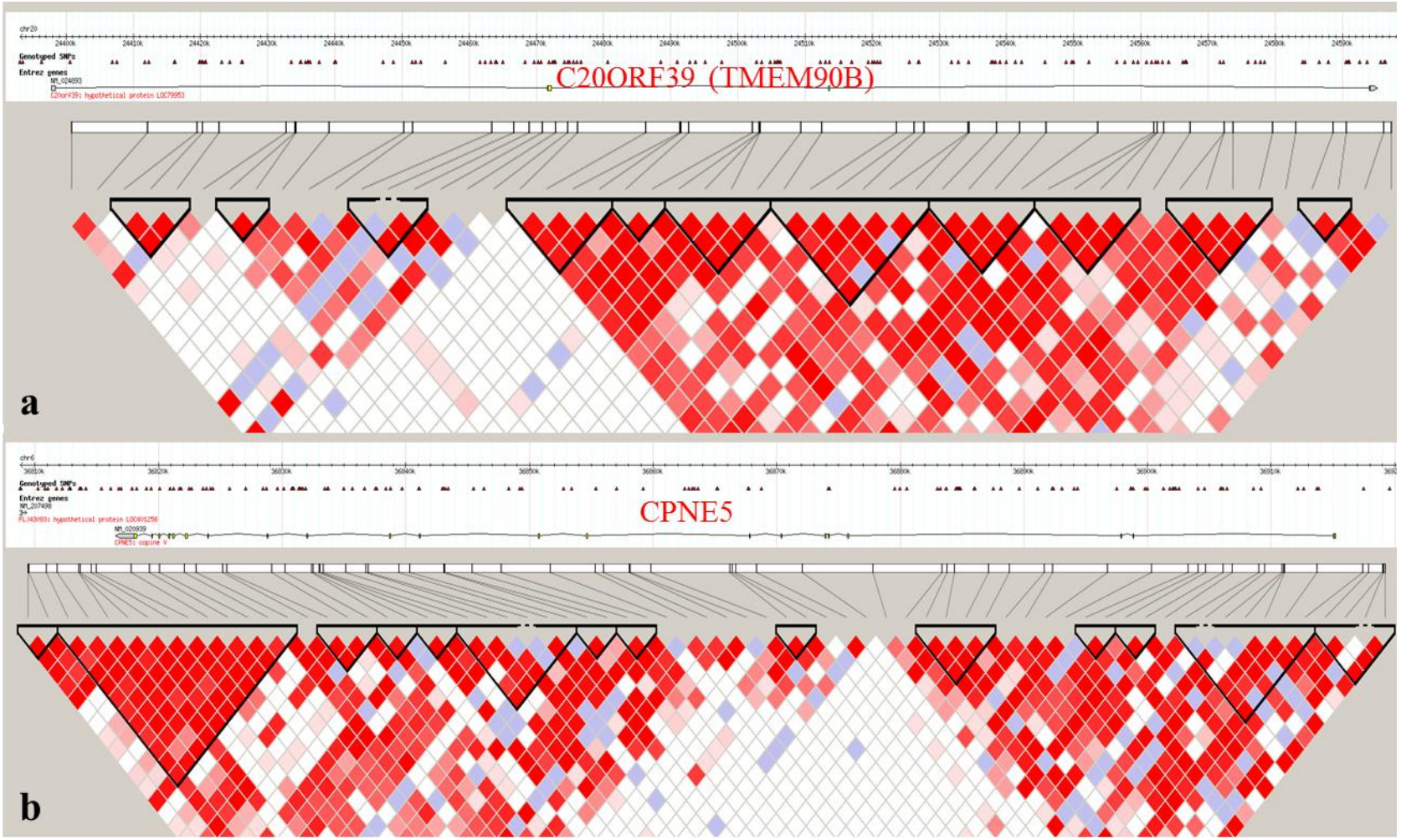
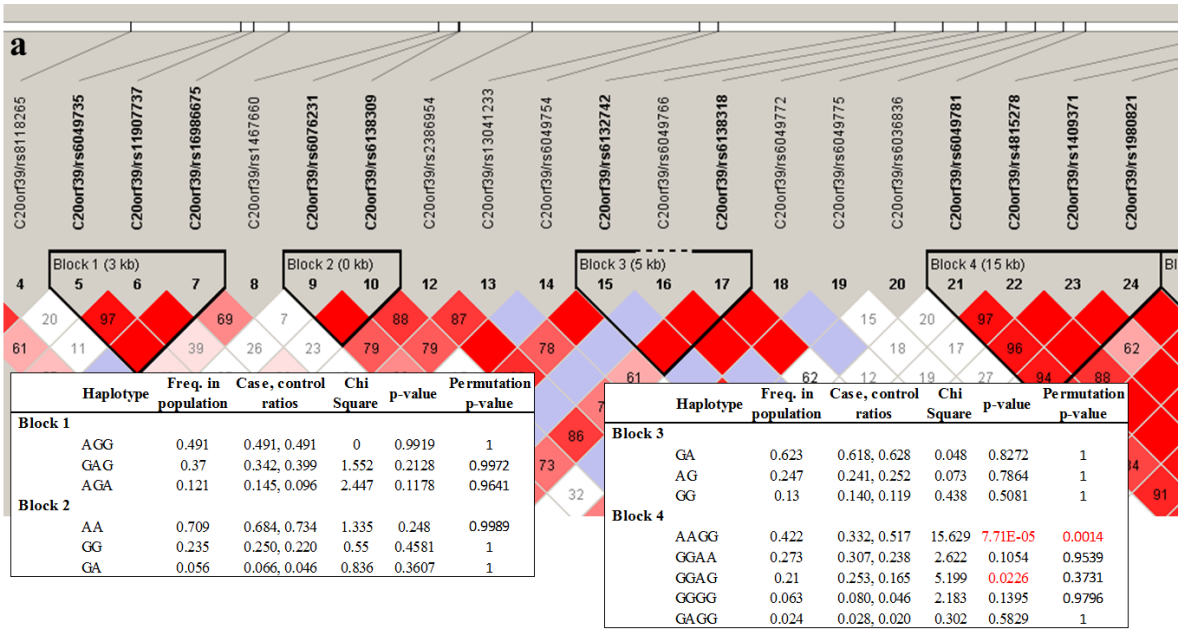
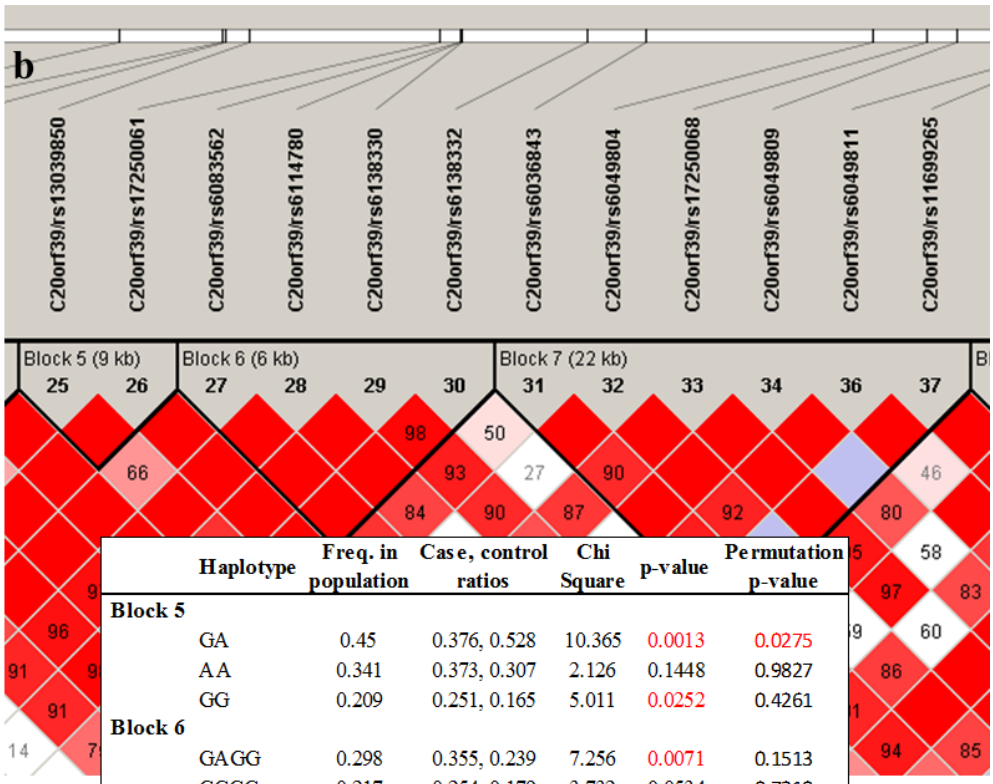


Figure 5.8. Overview of haplotype block analyses using confidence intervals (Gabriel *et al.*, 2002) and LD plots via Haploview tool. (a) 11 haplotype blocks were defined along the TMEM90B gene on chromosome 20, positions between 24.4-24.6 Mb (b) 14 haplotype blocks were defined along the CPNE5 gene on chromosome 6 position between 36.8-36.93 Mb.

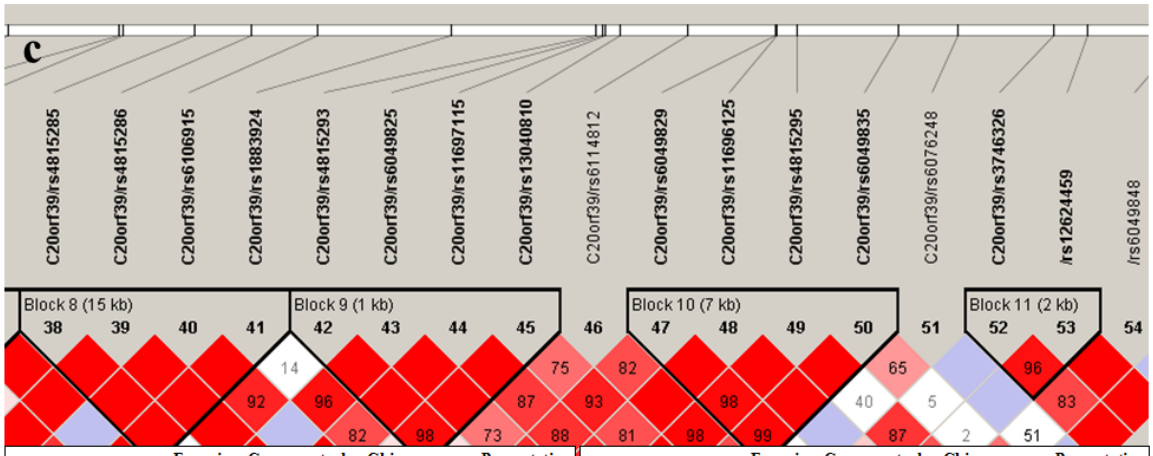


	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 1						
	AGG	0.491	0.491, 0.491	0	0.9919	1
	GAG	0.37	0.342, 0.399	1.552	0.2128	0.9972
	AGA	0.121	0.145, 0.096	2.447	0.1178	0.9641
Block 2						
	AA	0.709	0.684, 0.734	1.335	0.248	0.9989
	GG	0.235	0.250, 0.220	0.55	0.4581	1
	GA	0.056	0.066, 0.046	0.836	0.3607	1

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 3						
	GA	0.623	0.618, 0.628	0.048	0.8272	1
	AG	0.247	0.241, 0.252	0.073	0.7864	1
	GG	0.13	0.140, 0.119	0.438	0.5081	1
Block 4						
	AAGG	0.422	0.332, 0.517	15.629	7.71E-05	0.0014
	GGAA	0.273	0.307, 0.238	2.622	0.1054	0.9539
	GGAG	0.21	0.253, 0.165	5.199	0.0226	0.3731
	GGGG	0.063	0.080, 0.046	2.183	0.1395	0.9796
	GAGG	0.024	0.028, 0.020	0.302	0.5829	1



	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 5						
	GA	0.45	0.376, 0.528	10.365	0.0013	0.0275
	AA	0.341	0.373, 0.307	2.126	0.1448	0.9827
	GG	0.209	0.251, 0.165	5.011	0.0252	0.4261
Block 6						
	GAGG	0.298	0.355, 0.239	7.256	0.0071	0.1513
	GGGG	0.217	0.254, 0.179	3.732	0.0534	0.7318
	GAAA	0.205	0.144, 0.268	10.472	0.0012	0.0264
	AAAA	0.187	0.154, 0.222	3.473	0.0624	0.7755
	GAGA	0.09	0.088, 0.092	0.022	0.8826	1
Block 7						
	CGGCAG	0.457	0.363, 0.555	16.422	5.07E-05	0.0009
	AGACCG	0.208	0.258, 0.156	7.115	0.0076	0.1553
	CGGCAA	0.173	0.211, 0.133	4.686	0.0304	0.489
	AAGACG	0.108	0.114, 0.101	0.2	0.655	1
	AGGCAG	0.025	0.022, 0.028	0.132	0.7165	1
	CGACCG	0.012	0.014, 0.009	0.173	0.6777	1



	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 8						
	GGGA	0.446	0.351, 0.546	17.148	3.46E-05	0.0008
	AAGG	0.233	0.294, 0.170	9.59	0.002	0.0447
	AGGG	0.175	0.211, 0.138	4.106	0.0427	0.6435
	GGAG	0.11	0.105, 0.115	0.103	0.7478	1
	GGGG	0.034	0.035, 0.032	0.033	0.8569	1
Block 9						
	AGAA	0.313	0.241, 0.389	11.392	7.00E-04	0.0156
	AGAG	0.22	0.281, 0.157	10.027	0.0015	0.0344
	GGGG	0.191	0.228, 0.151	4.249	0.0393	0.5878
	GAAG	0.177	0.162, 0.193	0.752	0.3859	1
	GGAG	0.096	0.083, 0.109	0.83	0.3623	1

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 10						
	GCGG	0.478	0.403, 0.555	10.258	0.0014	0.0298
	ACAA	0.309	0.355, 0.261	4.588	0.0322	0.5239
	AAGA	0.204	0.237, 0.170	3.091	0.0787	0.8723
Block 11						
	AG	0.706	0.671, 0.743	2.789	0.0949	0.944
	GA	0.255	0.285, 0.225	2.132	0.1443	0.9822
	GG	0.032	0.035, 0.028	0.208	0.648	1

Figure 5.9. Haplotype block analysis of TMEM90B gene region via Haploview tool. Each block and their features are demonstrated. Haplotypes in each block, frequencies in the population, case-control ratios, chi-squares, p-values and permutated p-values are explained in integrated tables. Red indicates significant haplotypes which were lower than p=0.05 in both tests. (a) block 1-4, (b) block 5-7 and (c) block 8-11.

In the CPNE5 gene region, out of 14, eight blocks included significant haplotypes. Many haplotypes were found to be as significant ($p<0.05$), however, they failed to keep their significance after 10000 permutations. The AAGGAGACAAGC haplotype in block 2, CGG in block 3, GCACG in block 6, AG in block 7 and AA and AG haplotypes in block 8 preserve their significance even after 10000 permutations. The haplotypes with lowest p-values (~$9.5x10^{-6}$) were observed in block 3 and 6. Their adjusted p-values after 10000 permutations were still highly significant ($p=5x10^{-4}$) (Figure 5.10a and 5.10b). Like TMEM90B, significant haplotypes were defined where SNPs with lowest p-values obtained from allelic test were located in the CPNE5 gene region.

5.3.3.2. PLINK results using logistic regression. To validate the results obtained from Haploview tool and calculate the OR, the same analyses were performed using haplotype-based logistic regression test by PLINK tool. Same haplotype blocks were

generated for logistic regression test. Logistic regression test provides OR which enables to interpret whether associated haplotypes were a risk or a protective haplotype block. Haplotypes in each block and their generated p-values were almost consistent with the results of the Haploview tool. Some haplotypes failed to keep their significance after Bonferroni threshold specific for that block.

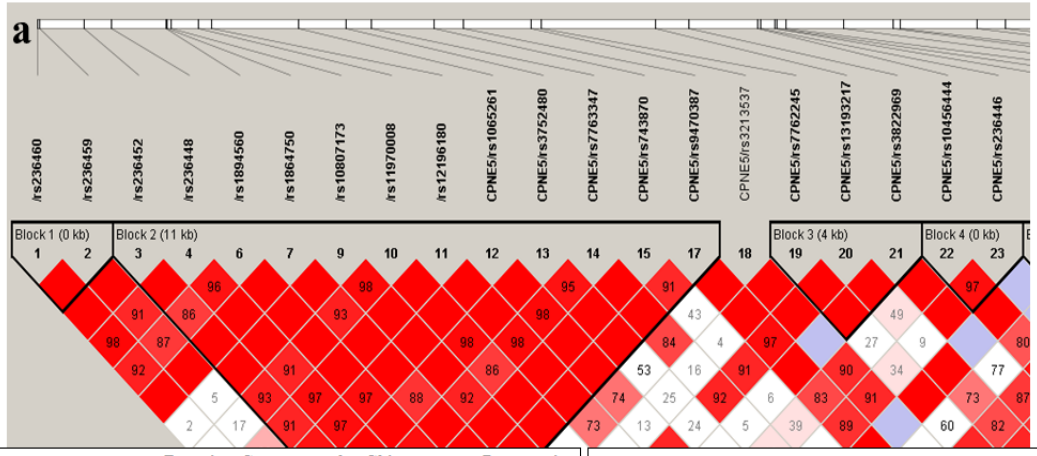
In the TMEM90B gene, none of haplotypes in block 1, 2 and 3 were significant as they were found in Haploview. In block 4, GGAG and AAGG haplotypes were found to be significant. According to ORs, GGAG (OR=1.76) was found to be a risk haplotype and AAGG (0.44) was found to be protective in the population. Furthermore GGAG haplotype did not keep the significance after Bonferroni correction ($p<0.01=0.05/5$). Like in block 4, two haplotypes were defined as significant in each block from block 5-10. Significant haplotypes, GG in block 5, GAGG in block 6, AGACCG in block 7, AAGG in block 8, AGAG in block 9 and ACAA in block 10, were found as risk haplotypes. On the other hand, significant haplotypes, GA in block 5, GAAA in block 6, CGGCAG in block 7, GGGA in block 8, AGAA in block 9 and GCGG in block 10 were found as protective when cases and controls were analyzed. When p-values of risk and protective haplotypes were defined in each block, p-values of protective haplotypes were found to be more significant than the risk haplotypes (Table 5.8).

In the CPNE5 gene region, haplotypes in block 9, 10, 11, 12, 13 and 14 were not found significant. The results of these haplotypes were also not shown as significant in Haploview analysis. However, in blocks 1-8, several haplotypes were found to be significantly associated with ALS, although the significance of some could not retained after Bonferroni correction. The haplotypes with lowest p-values were CGG in block 3, GCACG in block 6 and GG in block 8. Haplotypes in block 3 and 8 had high ORs, 2.37 and 2.47 which indicated them as risk haplotypes, however, haplotype in block 6 had reduced OR=0.396, indicating a protective haplotype. Unlike haplotypes of TMEM90B, haplotypes with lowest p-value in each block can be either risk-conferring or protective haplotype (Table 5.5).

In block 1, GC and AA haplotypes were found to be significant, yet they lost their significance after Bonferroni corrections ($p<0.0166$). In block 2, GCAAAGGAAGAA

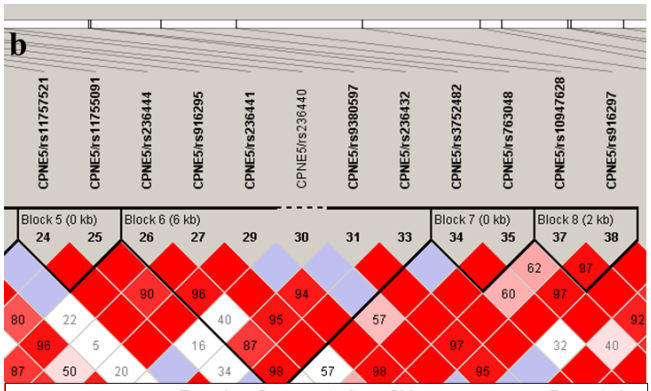
and AAGGAGACAAGC haplotypes had significant p-values even when Bonferroni threshold was applied ($p<0.007$). The AAGGAGACAAGC haplotype has an OR=2.12 and GCAAAGGAAGAA haplotype has an OR=0.54. Block 2 represents two significant haplotypes, one was observed with a higher frequency in controls and the other was higher in cases (Figure 5.5a). CAG haplotype in block 3, GA and AC in block 5, and AAGAG in block 6 did not keep their significance after applying Bonferroni correction.

AG and AA in block 7 and GG and AA in block 8 had significant p-values. AG in block 7 and GG in block 8 were considered as risk haplotypes according to their ORs, whereas the other two haplotypes were protective. When these two neighboring haplotypes were considered as a 4-SNP haplotype block, two distinct haplotypes were present, one risk (AGGG) and one protective (AAAA) haplotype (Table 5.9).

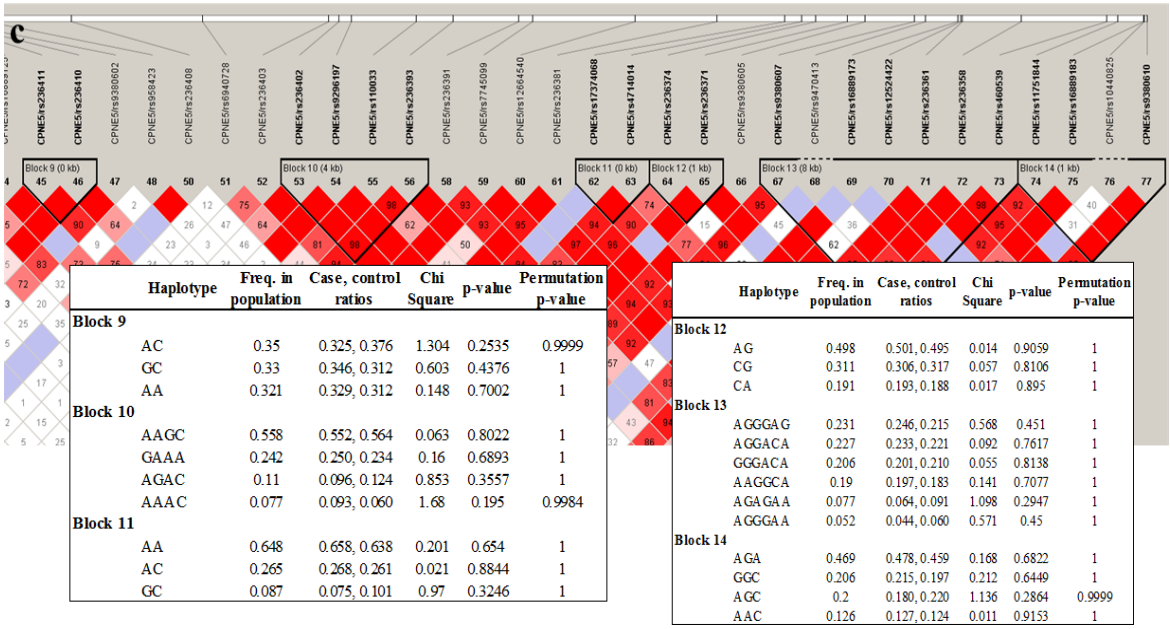


	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 1						
	AA	0.496	0.544, 0.445	4.361	0.0368	0.6627
	GC	0.457	0.412, 0.505	3.826	0.0505	0.8059
	GA	0.047	0.044, 0.050	0.108	0.7422	1
Block 2						
	GCAAAGGAAGAA	0.298	0.232, 0.367	9.599	0.0019	0.0628
	AAGGAGACAAGC	0.269	0.338, 0.197	11.182	8.00E-04	0.0263
	GAGAGAACGAGC	0.159	0.171, 0.147	0.47	0.4931	1
	GAGAAGGAAGAC	0.135	0.123, 0.147	0.551	0.4581	1
	GAGAAAACAAGC	0.041	0.031, 0.050	1.049	0.3058	1
	GCAAAGGAAGGC	0.031	0.022, 0.041	1.342	0.2466	0.9999
	GCAAAGGAAGGA	0.016	0.022, 0.010	1.116	0.2907	1

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 3						
	CGG	0.374	0.474, 0.271	19.615	9.47E-06	0.0005
	CAG	0.327	0.272, 0.385	6.507	0.0107	0.2903
	AAG	0.222	0.193, 0.252	2.27	0.1319	0.9864
	AAA	0.076	0.061, 0.092	1.457	0.2275	0.9995
Block 4						
	GG	0.47	0.429, 0.513	3.125	0.0771	0.9174
	AG	0.3	0.354, 0.244	6.402	0.0114	0.3027
	AA	0.227	0.212, 0.242	0.585	0.4445	1



	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 5						
	AC	0.895	0.864, 0.927	4.628	0.0315	0.6066
	GA	0.085	0.114, 0.055	4.975	0.0257	0.5598
	AA	0.02	0.022, 0.018	0.072	0.788	1
Block 6						
	GCACG	0.587	0.486, 0.692	19.589	9.60E-06	0.0005
	AAGAA	0.184	0.215, 0.151	2.992	0.0837	0.9274
	GAACG	0.083	0.110, 0.055	4.4	0.0359	0.6594
	AAGAG	0.081	0.109, 0.050	5.226	0.0223	0.5003
	AAACG	0.029	0.031, 0.028	0.044	0.8344	1
	GAAAG	0.013	0.017, 0.009	0.543	0.4611	1
Block 7						
	AA	0.57	0.509, 0.633	7.018	0.0081	0.2257
	AG	0.3	0.373, 0.225	11.62	7.00E-04	0.0214
	GG	0.13	0.118, 0.142	0.557	0.4554	1
Block 8						
	AA	0.488	0.403, 0.577	13.585	2.00E-04	0.0091
	GG	0.399	0.495, 0.298	18.116	2.08E-05	0.0011
	AG	0.108	0.093, 0.124	1.161	0.2814	0.9999



	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 9						
	AC	0.35	0.325, 0.376	1.304	0.2535	0.9999
	GC	0.33	0.346, 0.312	0.603	0.4376	1
	AA	0.321	0.329, 0.312	0.148	0.7002	1
Block 10						
	AAGC	0.558	0.552, 0.564	0.063	0.8022	1
	GAAA	0.242	0.250, 0.234	0.16	0.6893	1
	AGAC	0.11	0.096, 0.124	0.853	0.3557	1
	AAAC	0.077	0.093, 0.060	1.68	0.195	0.9984
Block 11						
	AA	0.648	0.658, 0.638	0.201	0.654	1
	AC	0.265	0.268, 0.261	0.021	0.8844	1
	GC	0.087	0.075, 0.101	0.97	0.3246	1

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value	Permutation p-value
Block 12						
	AG	0.498	0.501, 0.495	0.014	0.9059	1
	CG	0.311	0.306, 0.317	0.057	0.8106	1
	CA	0.191	0.193, 0.188	0.017	0.895	1
Block 13						
	AGGGAG	0.231	0.246, 0.215	0.568	0.451	1
	AGGACA	0.227	0.233, 0.221	0.092	0.7617	1
	GGGACA	0.206	0.201, 0.210	0.055	0.8138	1
	AAGGCA	0.19	0.197, 0.183	0.141	0.7077	1
	AGAGAA	0.077	0.064, 0.091	1.098	0.2947	1
	AGGGAA	0.052	0.044, 0.060	0.571	0.45	1
Block 14						
	AGA	0.469	0.478, 0.459	0.168	0.6822	1
	GGC	0.206	0.215, 0.197	0.212	0.6449	1
	AGC	0.2	0.180, 0.220	1.136	0.2864	0.9999
	AAC	0.126	0.127, 0.124	0.011	0.9153	1

Figure 5.10. Haplotype block analysis of CPNE5 gene region via Haploview tool. Each block and their features are demonstrated. Haplotypes in each block, frequencies in the population, case-control ratios, chi-squares, p-values and permutated p-values are explained in integrated tables. Red indicates significant haplotypes which were lower than p=0.05 in both tests. (a) block 1-4, (b) block 5-8 and (c) block 9-14.

Table 5.8. Haplotype-based case-control logistic regression test for haplotype block associations of TMEM90B gene region via PLINK.

Block number	Number of SNPs	Number of haplotypes	Chr	Start SNP	End SNP	Haplotype	Frequency	Odds Ratio	p-value
Block1	3	3	20	TMEM90B/rs6049735	TMEM90B/rs16986675	AGA	0.121	1.61	0.117
Block1	3	3	20	TMEM90B/rs6049735	TMEM90B/rs16986675	GAG	0.37	0.798	0.233
Block1	3	3	20	TMEM90B/rs6049735	TMEM90B/rs16986675	AGG	0.491	1	0.992
Block2	2	3	20	TMEM90B/rs6076231	TMEM90B/rs6138309	GG	0.236	1.18	0.447
Block2	2	3	20	TMEM90B/rs6076231	TMEM90B/rs6138309	GA	0.0563	1.47	0.358
Block2	2	3	20	TMEM90B/rs6076231	TMEM90B/rs6138309	AA	0.707	0.779	0.23
Block3	2	3	20	TMEM90B/rs6132742	TMEM90B/rs6138318	AG	0.247	0.943	0.788
Block3	2	3	20	TMEM90B/rs6132742	TMEM90B/rs6138318	GG	0.13	1.23	0.484
Block3	2	3	20	TMEM90B/rs6132742	TMEM90B/rs6138318	GA	0.623	0.954	0.819
Block4	4	5	20	TMEM90B/rs6049781	TMEM90B/rs1980821	GGAA	0.272	1.39	0.122
Block4	4	5	20	TMEM90B/rs6049781	TMEM90B/rs1980821	GGAG	0.209	1.76	0.0226
Block4	4	5	20	TMEM90B/rs6049781	TMEM90B/rs1980821	AAGG	0.424	0.44	0.000114
Block4	4	5	20	TMEM90B/rs6049781	TMEM90B/rs1980821	GAGG	0.0239	1.47	0.559
Block4	4	5	20	TMEM90B/rs6049781	TMEM90B/rs1980821	GGGG	0.0631	1.88	0.134
Block5	2	3	20	TMEM90B/rs13039850	TMEM90B/rs17250061	GG	0.209	1.76	0.0224
Block5	2	3	20	TMEM90B/rs13039850	TMEM90B/rs17250061	AA	0.34	1.31	0.167
Block5	2	3	20	TMEM90B/rs13039850	TMEM90B/rs17250061	GA	0.45	0.518	0.00133
Block6	4	5	20	TMEM90B/rs6083562	TMEM90B/rs6138332	AAAA	0.187	0.641	0.0688
Block6	4	5	20	TMEM90B/rs6083562	TMEM90B/rs6138332	GAAA	0.205	0.476	0.00224
Block6	4	5	20	TMEM90B/rs6083562	TMEM90B/rs6138332	GAGA	0.0899	0.952	0.882
Block6	4	5	20	TMEM90B/rs6083562	TMEM90B/rs6138332	GGGG	0.217	1.6	0.0504

Table 5.8. Haplotype-based case-control logistic regression test for haplotype block associations of TMEM90B gene region via PLINK (cont.).

Block number	Number of SNPs	Number of haplotypes	Chr	Start SNP	End SNP	Haplotype	Frequency	Odds Ratio	p-value
Block6	4	5	20	TMEM90B/rs6083562	TMEM90B/rs6138332	GAGG	0.298	1.71	0.01
Block7	6	6	20	TMEM90B/rs6036843	TMEM90B/rs11699265	CGGCAA	0.172	1.76	0.0321
Block7	6	6	20	TMEM90B/rs6036843	TMEM90B/rs11699265	AAGACG	0.108	1.14	0.658
Block7	6	6	20	TMEM90B/rs6036843	TMEM90B/rs11699265	AGACCG	0.208	1.93	0.00791
Block7	6	6	20	TMEM90B/rs6036843	TMEM90B/rs11699265	CGACCG	0.0111	1.37	0.74
Block7	6	6	20	TMEM90B/rs6036843	TMEM90B/rs11699265	AGGCAG	0.025	0.787	0.7
Block7	6	6	20	TMEM90B/rs6036843	TMEM90B/rs11699265	CGGCAG	0.457	0.425	5.95E-05
Block8	4	5	20	TMEM90B/rs4815285	TMEM90B/rs1883924	GGGA	0.446	0.392	2.59E-05
Block8	4	5	20	TMEM90B/rs4815285	TMEM90B/rs1883924	GGAG	0.11	0.913	0.758
Block8	4	5	20	TMEM90B/rs4815285	TMEM90B/rs1883924	AAGG	0.233	2.11	0.00204
Block8	4	5	20	TMEM90B/rs4815285	TMEM90B/rs1883924	AGGG	0.175	1.7	0.0422
Block8	4	5	20	TMEM90B/rs4815285	TMEM90B/rs1883924	GGGG	0.0337	1.1	0.859
Block9	4	5	20	TMEM90B/rs4815293	TMEM90B/rs13040810	AGAA	0.313	0.488	0.000936
Block9	4	5	20	TMEM90B/rs4815293	TMEM90B/rs13040810	GGGG	0.191	1.67	0.041
Block9	4	5	20	TMEM90B/rs4815293	TMEM90B/rs13040810	GAAG	0.177	0.816	0.401
Block9	4	5	20	TMEM90B/rs4815293	TMEM90B/rs13040810	GGAG	0.0957	0.731	0.349
Block9	4	5	20	TMEM90B/rs4815293	TMEM90B/rs13040810	AGAG	0.22	2.14	0.00193
Block10	4	3	20	TMEM90B/rs6049829	TMEM90B/rs6049835	GCGG	0.478	0.535	0.00165

***** Red indicates significant haplotypes which were lower than p=0.05

Table 5.9. Haplotype-based case-control logistic regression test for haplotype block associations of CPNE5 gene region via PLINK.

Block number	Number of SNPs	Number of haplotypes	Chr	Start SNP	End SNP	Haplotype	Frequency	Odds Ratio	p-value
Block1	2	3	6	/rs236460	/rs236459	GC	0.459	0.669	0.0402
Block1	2	3	6	/rs236460	/rs236459	AA	0.493	1.51	0.031
Block1	2	3	6	/rs236460	/rs236459	GA	0.0473	0.861	0.733
Block2	12	7	6	/rs236452	CPNE5/rs9470387	GCAAAGGAAGAA	0.3	0.542	0.00383
Block2	12	7	6	/rs236452	CPNE5/rs9470387	GAGAAGGAAGAC	0.132	0.781	0.377
Block2	12	7	6	/rs236452	CPNE5/rs9470387	GCAAAGGAAGGC	0.0313	0.51	0.241
Block2	12	7	6	/rs236452	CPNE5/rs9470387	GCAAAGGAAGGA	0.0136	1.65	0.515
Block2	12	7	6	/rs236452	CPNE5/rs9470387	AAGGAGACAAGC	0.269	2.12	0.00104
Block2	12	7	6	/rs236452	CPNE5/rs9470387	GAGAGAACGAGC	0.159	1.19	0.496
Block2	12	7	6	/rs236452	CPNE5/rs9470387	GAGAAAACAAGC	0.0406	0.592	0.296
Block3	3	4	6	CPNE5/rs7762245	CPNE5/rs3822969	AAA	0.0762	0.689	0.269
Block3	3	4	6	CPNE5/rs7762245	CPNE5/rs3822969	CGG	0.374	2.37	3.29E-05
Block3	3	4	6	CPNE5/rs7762245	CPNE5/rs3822969	AAG	0.222	0.694	0.124
Block3	3	4	6	CPNE5/rs7762245	CPNE5/rs3822969	CAG	0.327	0.604	0.0134
Block4	2	3	6	CPNE5/rs10456444	CPNE5/rs236446	AA	0.227	0.841	0.444
Block4	2	3	6	CPNE5/rs10456444	CPNE5/rs236446	GG	0.47	0.73	0.0883
Block4	2	3	6	CPNE5/rs10456444	CPNE5/rs236446	AG	0.3	1.7	0.0127
Block5	2	3	6	CPNE5/rs11757521	CPNE5/rs11755091	GA	0.0852	2.11	0.0365
Block5	2	3	6	CPNE5/rs11757521	CPNE5/rs11755091	AA	0.0202	1.2	0.786
Block5	2	3	6	CPNE5/rs11757521	CPNE5/rs11755091	AC	0.895	0.523	0.0416

Table 5.9. Haplotype-based case-control logistic regression test for haplotype block associations of CPNE5 gene region via PLINK (cont.).

Block number	Number of SNPs	Number of haplotypes	Chr	Start SNP	End SNP	Haplotype	Frequency	Odds Ratio	p-value
Block6	5	6	6	CPNE5/rs236444	CPNE5/rs236432	AAGAA	0.184	1.53	0.0878
Block6	5	6	6	CPNE5/rs236444	CPNE5/rs236432	AAGAG	0.0805	2.03	0.0442
Block6	5	6	6	CPNE5/rs236444	CPNE5/rs236432	GAAAG	0.0128	2	0.452
Block6	5	6	6	CPNE5/rs236444	CPNE5/rs236432	AAACG	0.0293	1.12	0.84
Block6	5	6	6	CPNE5/rs236444	CPNE5/rs236432	GAACG	0.0835	2.03	0.0479
Block6	5	6	6	CPNE5/rs236444	CPNE5/rs236432	GCACG	0.587	0.396	1.60E-05
Block7	2	3	6	CPNE5/rs3752482	CPNE5/rs763048	GG	0.13	0.83	0.483
Block7	2	3	6	CPNE5/rs3752482	CPNE5/rs763048	AG	0.3	2.09	0.000826
Block7	2	3	6	CPNE5/rs3752482	CPNE5/rs763048	AA	0.57	0.599	0.0092
Block8	2	3	6	CPNE5/rs10947628	CPNE5/rs916297	AA	0.488	0.487	0.000339
Block8	2	3	6	CPNE5/rs10947628	CPNE5/rs916297	GG	0.399	2.47	2.89E-05
Block8	2	3	6	CPNE5/rs10947628	CPNE5/rs916297	AG	0.108	0.741	0.308
Block9	2	3	6	CPNE5/rs236411	CPNE5/rs236410	AA	0.321	1.08	0.709
Block9	2	3	6	CPNE5/rs236411	CPNE5/rs236410	GC	0.33	1.17	0.432
Block9	2	3	6	CPNE5/rs236411	CPNE5/rs236410	AC	0.35	0.815	0.28
Block10	4	4	6	CPNE5/rs236402	CPNE5/rs236393	GAAA	0.242	1.09	0.689
Block10	4	4	6	CPNE5/rs236402	CPNE5/rs236393	AGAC	0.11	0.783	0.391
Block10	4	4	6	CPNE5/rs236402	CPNE5/rs236393	AAAC	0.0767	1.58	0.208

***** Red indicates significant haplotypes which were lower than p=0.05

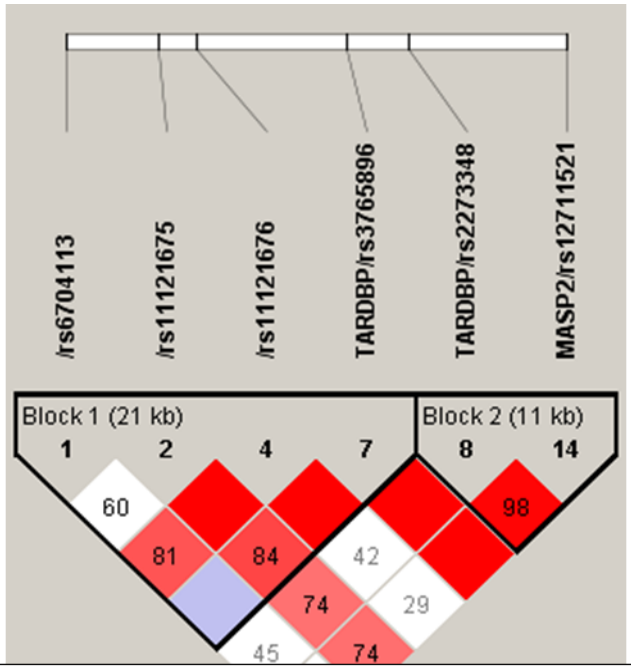
5.3.4. Single marker association and haplotype block analysis of TARDBP locus in Turkish GWAS

The locus that was found to be significant in both datasets, was investigated in the Turkish population after the completion of the Turkish ALS-GWAS. 17 SNPs in the locus of TARDBP gene were obtained. The single marker association demonstrated that none of these 17 SNPs were associated with ALS (Table 5.10).

Table 5.10. Single marker associations of SNPs located in the TARDBP region.

SNP	Associated Allele	Case, control ratios	Chi Square	p-value
/rs6704113	A	0.969, 0.958	0.384	0.5355
/rs11121675	A	0.596, 0.573	0.245	0.6207
/rs7517230	A	0.704, 0.670	0.59	0.4424
/rs11121676	G	0.811, 0.798	0.124	0.7243
TARDBP/rs9430335	A	0.792, 0.773	0.232	0.6302
TARDBP/rs3765895	G	0.798, 0.771	0.503	0.4783
TARDBP/rs3765896	A	0.794, 0.771	0.353	0.5524
TARDBP/rs2273348	G	0.746, 0.736	0.052	0.8193
MASP2/rs2273347	G	0.752, 0.739	0.109	0.7408
MASP2/rs1033638	G	0.783, 0.766	0.187	0.6657
MASP2/rs1782455	A	0.785, 0.766	0.232	0.63
MASP2/rs2273346	A	0.965, 0.963	0.008	0.9272
MASP2/rs12711521	A	0.750, 0.729	0.247	0.6194
MASP2/rs3765901	A	0.702, 0.697	0.011	0.9173
MASP2/rs2273344	A	0.746, 0.729	0.152	0.6965

At haplotype level, two different SNP clusters were analyzed by Haploview. The first cluster included 17 SNPs genotyped in the Turkish GWAS, however, none of the blocks were shown to be statistically significant. Moreover seven SNPs, commonly found in Cronin and Schymick datasets, were investigated. In the Turkish dataset, six of seven SNPs were common. When haplotype block analysis was performed with these SNPs, the results were the same as in analysis of 17 SNPs (Figure 5.11).



	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
Block 1					
	AAGA	0.558	0.575, 0.540	0.579	0.4468
	ACGG	0.195	0.188, 0.201	0.125	0.7236
	ACAA	0.165	0.163, 0.166	0.008	0.9287
	GCAA	0.029	0.023, 0.035	0.595	0.4406
	ACGA	0.025	0.027, 0.023	0.052	0.8188
	AAGG	0.022	0.016, 0.028	0.772	0.3797
Block 2					
	GA	0.738	0.746, 0.729	0.153	0.696
	AC	0.256	0.250, 0.261	0.077	0.7814

Figure 5.11. Haplotype block analysis of the TARDP region using confidence intervals in Turkish GWAS. Haplotype blocks observed using Haploview and integrated table showing the haplotypes in each block and their p-values.

5.3.5. Single marker and haplotype block association analyses of the TMEM90B and CPNE5 gene regions in Schymick and Cronin datasets

TMEM90B and CPNE5 gene regions identified in the Turkish GWAS were tested for their presence in the Schymick and Cronin datasets. Single marker associations of the TMEM90B gene region showed only one significant SNP (rs6076248, p=0.0080) in Cronin dataset (Figure 5.12a). In addition, none of the SNPs in Schymick dataset was found to be statistically significant. In the CPNE5 gene region, rs236427, rs1864750, rs1010791, rs236419, rs236444, rs236452 and rs9296197 in Schymick and rs236410 and rs958423 in Cronin were not as significant as they are in the Turkish dataset (unadjusted $p<0.05$) (Figure 5.12b).

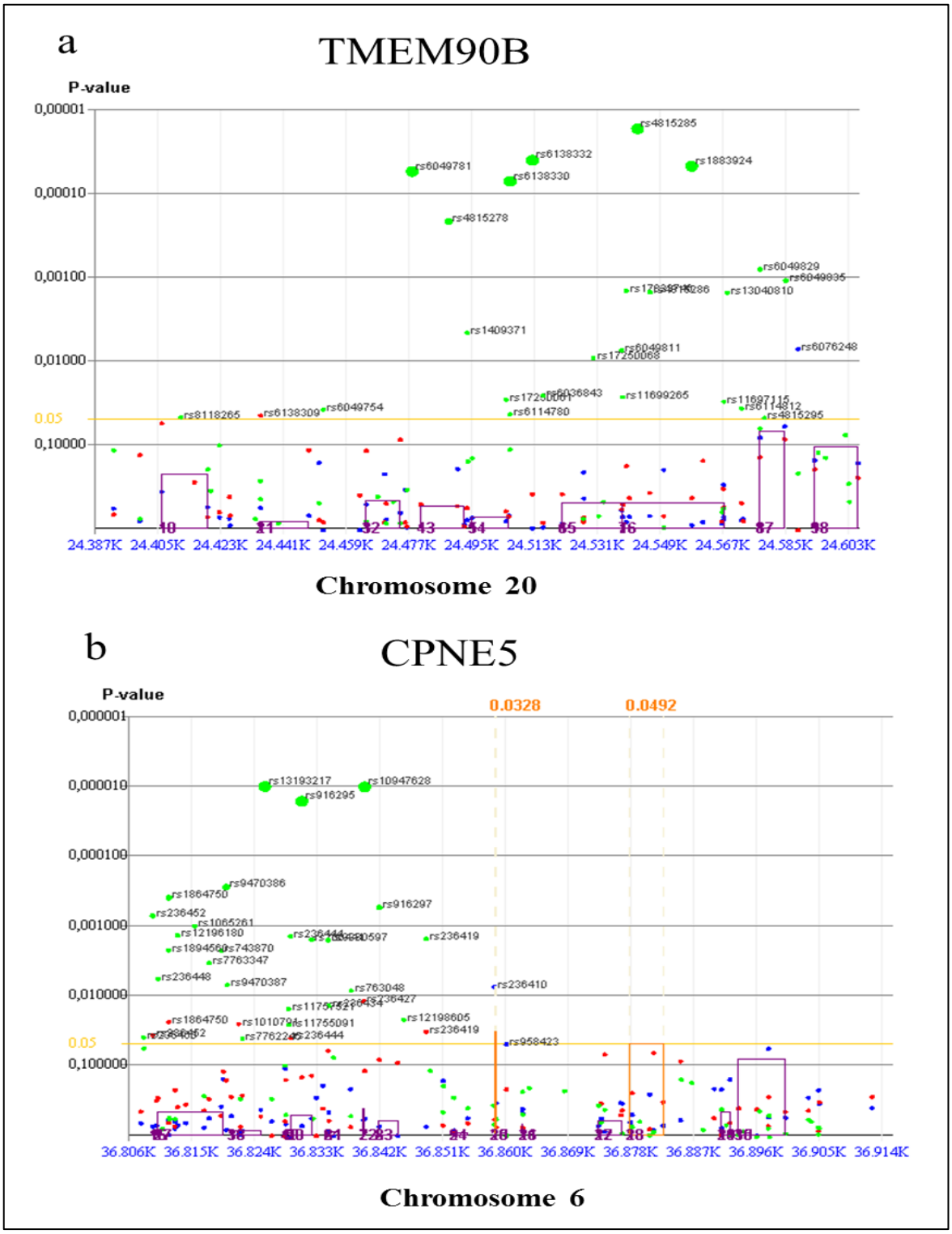


Figure 5.12. Allelic association test results using GWADView. (a)Association results of SNPs located in the TMEM90B gene region. (b) Association results of SNPs located in the CPNE5 gene region. Horizontal axis demonstrates chromosome positions. Left vertical axis shows -$\log_{10}$p-values of SNPs. Each dot represents a SNP. Green dots: Turkish GWAS, blue dots: Cronin dataset and red dots: Schymick dataset.

After Bonferroni correction, all SNPs failed to pass the significance threshold. Furthermore haplotype blocks of those regions were also checked by Haploview. In the TMEM90B gene, almost no haplotypes showed any association. The TA haplotype in block 2 in Schymick and one CT haplotype in block 9 in Cronin had border-line significance ($p<0.05$). Those haplotypes did not overlap with the associated regions found in the Turkish population. On the other hand, in the CPNE5 gene, several significant haplotypes (most of them were border-line significance) were identified in both datasets. Few associated blocks overlapped in those datasets, for example the block 10 in Schymick and the block 13 in Cronin, however, they were not found to be associated in Turkish dataset (Appendices A-D). In summary, the comparison of SNPs located at TMEM90B and CPNE5 genes in three GWAS datasets did not show any correlation in terms of single marker and haplotype block associations.

5.4. Genome-wide CNV Analysis of Turkish ALS Population

CNV analyses were carried out for 116 ALS patients and 109 neurologically healthy individuals using the PennCNV software to identify potential gene deletions or duplications impacting ALS in the Turkish population. Analysis of the 733K markers with default parameters of the program yielded ~25000 CNV calls. These calls were computed using Log R Ratio (LRR) and B allele frequency (BAF) values obtained from the array. To eliminate false positives, the first waviness factor of each individual was adjusted. According to GC content of the genome, signal intensities coming from different parts of the genome vary. GC model signal adjustment would reduce false positive calls in individuals with high fluctuation of signal waviness. After signal adjustment, 18200 CNV calls were generated. Second, individuals with high number of CNV calls and CNV calls with low confidence value were eliminated. Two ALS patients and three healthy controls were discarded due to excess amount of CNV presence. Approximately 10000 CNV calls in five individuals were excluded.

Confidence value is another important threshold for eliminating false positive CNV calls. This value is calculated by PennCNV according to likelihood of a CNV state for a given region. 3000 calls were excluded from CNV analysis due to low confidence threshold ($conf>10.00$).

5.4.1. Summary statistics of genome-wide CNV analysis

After exclusion of CNVs calls based on several criteria, 5000 CNVs were taken into count and 115 ALS patients and 106 controls were included in the analysis. Average CNV number per individual in cases was 23.8 and in controls 20.7 (Table 5.11).

Table 5.11. Summary statistics of CNV analysis.

	ALS patients	Control
Number of Samples (n)	115	106
Average CNV number per individual	23.8	20.7
Range of CNVs observed in individuals	9-85	7-76
Average SNP number per CNV	14	15
Average CNV length (kb)	87.1	85.2
Range of CNV lengths	97 bp-8438 kb	33 bp-3538 kb

5.4.2. Overlapping and discrete CNVs

Among 5000 CNV calls, many overlapping and discrete regions were detected. Some regions also showed discrete and overlapping CNV patterns in individuals, with common and rare CNVs. Many of the common CNVs observed in the Turkish population were located in previously identified CNV regions as expected (see CNV column: "located in CNV region", Tables 5.12 and 5.13). Most of the CNV calls observed in cases and controls were in balance. The highest frequency difference observed in cases and controls are summarized in Tables 5.12 and 5.13.

All CNV calls were visualized by plotting signal intensities of each SNP (Log R Ratio value) to examine whether CNV calls were true or false-positive. Two CNV calls of individuals, with deletions on chromosome 11, were plotted between positions 50.4 Mb and 51.1 Mb which is indicated at the top of Table 5.12. According to the plotted region in red, intensity captured from the chip was below the average line, indicating a deletion occurred in that region (Figure 5.13). The region was searched in DGV and HapMap databases and this novel and intergenic CNV is found to be located in the vicinity of the centromere.

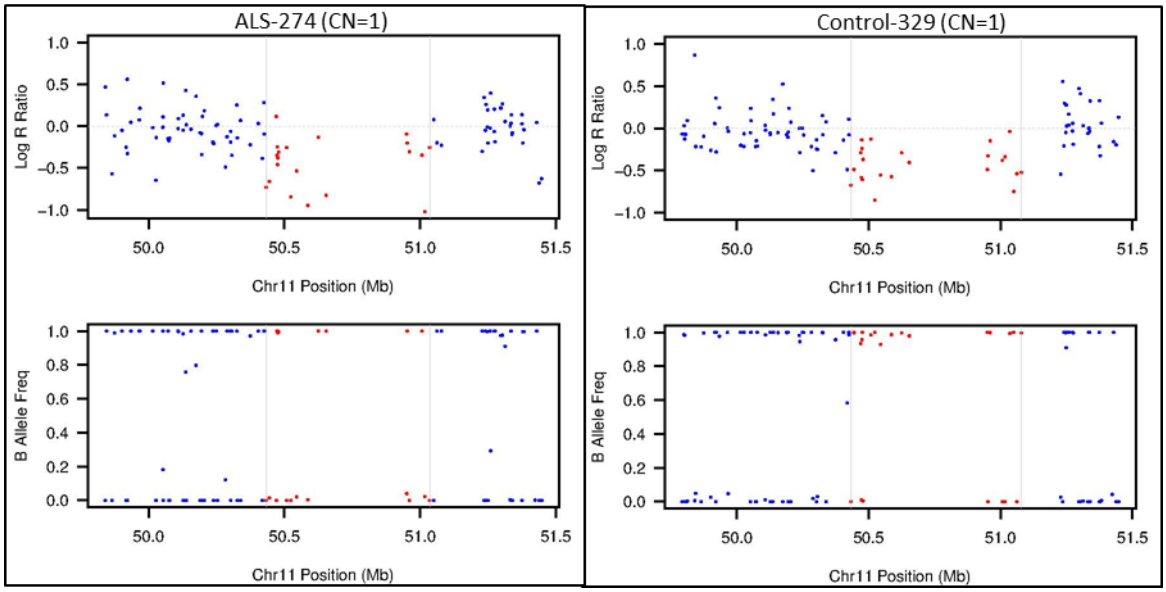


Figure 5.13. Plotting CNV calls of ALS-274 and Control-329 using signal intensities by PennCNV. Both individuals have one copy deletion (CN=1) on chromosome 11, between positions 50.4 Mb and 51.1 Mb. Each dot represents its Log R Ratio and B Allele Frequency of a SNP. Red indicates SNPs in the CNV site, blue represents SNPs neighboring the CNV region.

Two CNV calls of individuals with duplications on chromosome 1, between positions 97.83 Mb and 97.85 Mb were also plotted. As indicated in Table 5.12, it was in the DPYD gene region. According to the plotted area in red, there was an increased intensity above average intensity line, resulting in one copy duplication in that particular region (Figure 5.14). When its location was examined, this CNV was found to be reported previously in the territory of the DPYD gene. Duplication is detected in a small region of the gene containing exonic and intronic parts.

In addition, several novel CNV regions were found and those regions might be specific to the Turkish population. Also, approximately 500 CNV calls were observed only in one individual. There are several CNVs that were seen in only ALS patients (Table 5.14). In general they were not frequent, but some of them were seen in gene regions. When they were searched in CNV databases, they were found to be absent in all databases, which make them candidates for ALS susceptibility. One example for rare CNVs observed in only cases is deletion in the ACYP2 gene (Table 5.14). Two of the four CNV calls of individuals on chromosome 2, positions between 54.2 Mb and 54.37 Mb were plotted. One copy deletion was seen in both individuals. This novel CNV is detected in the intronic part of the gene (Figure 5.15).

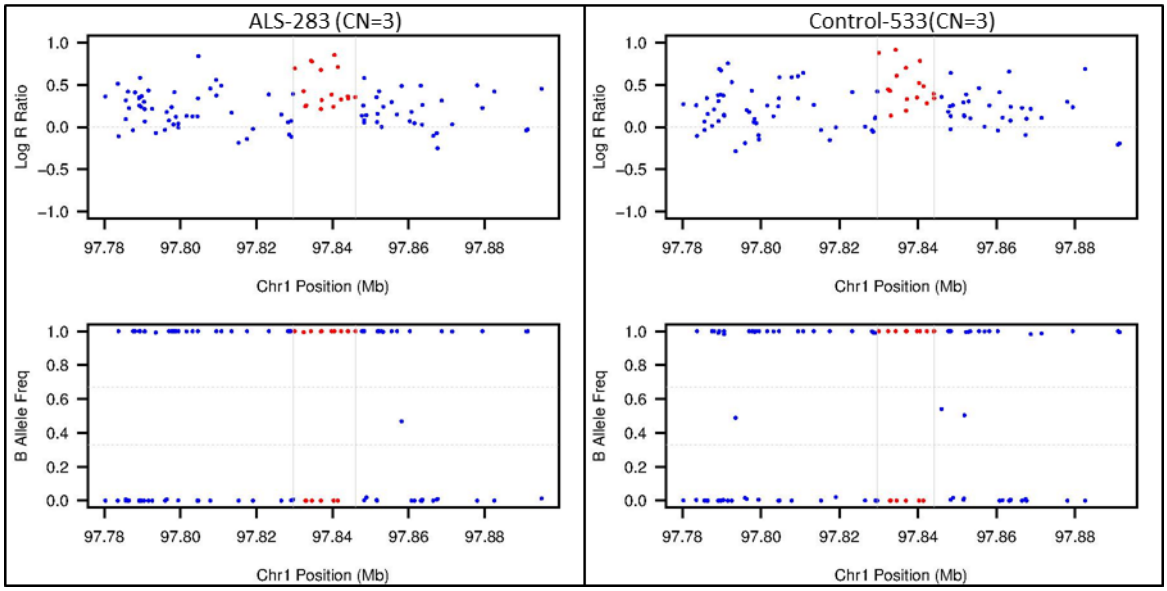


Figure 5.14. Plotting CNV calls of ALS-283 and Control-533 using signal intensities by PennCNV. Both individuals have one copy duplication (CN=3) on chromosome 1, between positions 97.83 Mb and 97.85 Mb. Each dot represents its Log R Ratio and B Allele Frequency of a SNP. Red indicates SNPs in the CNV site, blue represents SNPs neighboring the CNV region.

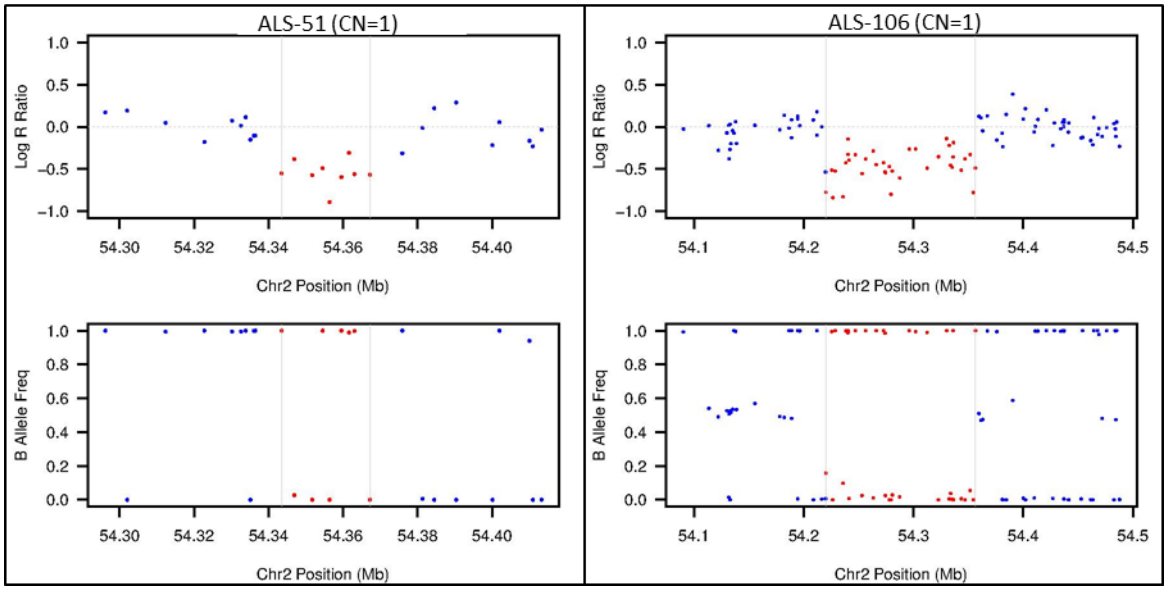


Figure 5.15. Plotting CNV calls of ALS-51 and ALS-106 using signal intensities by PennCNV. Both individuals have one copy deletion (CN=1) on chromosome 2, between positions 54.2 Mb and 54.37 Mb. Each dot represents its Log R Ratio and B Allele Frequency of a SNP. Red indicates SNPs in the CNV site, blue represents SNPs neighboring the CNV region.

Table 5.12. Overlapping CNVs observed in analysis.

CHR	start Position	end position	length	CNV state	genes	CNV	% cases	% controls	p-value	Fisher's Exact Test
11	50545009	50586426	41417	state2,cn=1	intergenic	novel	18.26%	1.89%	6.82E-05	3.23E-05
19	20860930	20875787	14857	state2,cn=1	intergenic	novel, near ZNF85	13.04%	1.89%	0.001874	0.001403
12	36404411	36411114	6703	state2,cn=1	centromeric	located in CNV region	21.74%	7.55%	0.003103	0.002411
2	39372016	39428488	56472	state5,cn=3	MAP4K3	novel	12.17%	1.89%	0.003196	0.002535
3	84486776	84510027	23251	state5,cn=3	intergenic	novel	9.57%	0.94%	0.004717	0.003902
6	31389749	31393270	3521	state1,cn=0	intergenic	located in CNV region	6.96%	0.00%	0.0007	0.004764
3	89485137	89499861	14724	state2,cn=1	EPHA3	located in CNV region	1.74%	10.38%	0.006399	0.0062208
5	151496845	151499002	2157	state2,cn=1	intergenic	located in CNV region	4.35%	15.09%	0.006495	0.0058109
1	97830032	97841389	11357	state5,cn=3	DPYD	located in CNV region	10.43%	1.89%	0.009154	0.007979
4	153010030	153012241	2211	state2,cn=1	intergenic	located in CNV region	10.43%	1.89%	0.009154	0.007979
3	163644310	163671428	27118	state5,cn=3	intergenic	near CNV region	0.87%	7.55%	0.012102	0.0129209
2	89731562	89757456	25894	state2,cn=1	centromeric	located in CNV region	5.22%	0.00%	0.003431	0.0186203
2	208064053	208066082	2029	state2,cn=1	intergenic	located in CNV region	1.74%	10.38%	0.006399	0.0062208
1	150519809	150526366	6557	state2,cn=1	intergenic	near CNV region	0.00%	3.77%	0.010781	0.051363
11	107166452	107175438	8986	state5,cn=3	intergenic	novel	6.96%	0.94%	0.023853	0.0237312
13	63241820	63285508	43688	state5,cn=3	intergenic	located in CNV region	6.96%	0.94%	0.023853	0.0237312
7	61792309	61797361	5052	state2,cn=1	centromeric	located in CNV region	10.43%	2.83%	0.024742	0.0217869
13	62890418	62903462	13044	state5,cn=3	intergenic	near CNV region	4.35%	0.00%	0.00765	0.0365634
11	48820745	48835439	14694	state2,cn=1	intergenic	located in CNV region	20.87%	10.38%	0.032811	0.0247327
9	138380284	138416305	36021	state5,cn=3	CARD9	located in CNV region	0.00%	3.77%	0.010781	0.0247327
10	58574865	58606945	32080	state2,cn=1	intergenic	located in CNV region	0.00%	3.77%	0.010781	0.0247327
6	62237262	62247872	10610	state2,cn=1	intergenic	located in CNV region	11.30%	3.77%	0.035823	0.030501
15	54580082	54588851	8769	state2,cn=1	intergenic	located in CNV region	0.87%	5.66%	0.042188	0.0475191
12	58222193	58228389	6196	state2,cn=1	intergenic	located in CNV region	20.00%	10.38%	0.047618	0.0355077

Table 5.13. Discrete CNVs observed in analysis.

CHR	start Position	end position	length	CNV state	genes	CNV	% cases	% controls	p-value	Fisher's Exact Test
19	32455280	32620453	165,173	state2,cn=1	intergenic	located in CNV region	11.30%	0.94%	0.001583	0.0011082
13	20626134	20630348	4,214	state5,cn=3	C13ORF3	novel, near CNV region	9.57%	0.94%	0.004717	0.0039018
6	31389749	31393270	3,521	state1,cn=0	intergenic	located in CNV region	6.96%	0.00%	0.0007	0.0047641
3	89485137	89499861	14,724	state2,cn=1	EPHA3	located in CNV region	1.74%	10.38%	0.006399	0.0062208
6	62237262	62281216	43,954	state2,cn=1	intergenic	located in CNV region	8.70%	0.94%	0.008111	0.0072081
3	84408677	84534756	126,079	state5,cn=3	intergenic	novel	6.09%	0.00%	0.001548	0.00944
11	50432844	50654074	221,230	state2,cn=1	intergenic	novel	6.09%	0.00%	0.001548	0.00944
2	208064053	208066082	2,029	state2,cn=1	intergenic	located in CNV region	1.74%	9.43%	0.011671	0.0116196
1	25470862	25515182	44,320	state2,cn=1	RHD	located in CNV region	7.83%	0.94%	0.01392	0.0131629
2	39365941	39442301	76,360	state5,cn=3	MAP4K3	novel	7.83%	0.94%	0.01392	0.0131629
11	48704187	48845702	141,515	state2,cn=1	intergenic	located in CNV region	7.83%	0.94%	0.01392	0.0131629
12	58221292	58228389	7,097	state2,cn=1	intergenic	located in CNV region	15.65%	5.66%	0.017075	0.0138436
2	89731562	89757456	25,894	state2,cn=1	intergenic	located in CNV region	5.22%	0.00%	0.003431	0.0186203
1	97830032	97844111	14,079	state5,cn=3	DPYD	located in CNV region	4.35%	0.00%	0.00765	0.0365634
13	62890418	62903462	13,044	state5,cn=3	intergenic	novel	4.35%	0.00%	0.00765	0.0365634
19	20860930	20880696	19,766	state2,cn=1	intergenic	novel, near ZNF85	4.35%	0.00%	0.00765	0.0365634

Table 5.14. Rare CNVs (including overlapping and discrete CNVs) observed in analysis (p>0.05 according to Fisher's exeact test).

CHR	start Position	end position	length	CNV state	genes	CNV	% cases	% controls	p-value	Fisher's Exact Test
2	54343530	54356415	12885	state2,cn=1	ACYP2	novel	3.48%	0.00%	0.017223	0.0714798
4	62506258	62566026	59768	state5,cn=3	LPHN3	novel	3.48%	0.00%	0.017223	0.0714798
6	19153539	19156752	3213	state1,cn=0	intergenic	located in CNV region	3.48%	0.00%	0.017223	0.0714798
8	2798648	2815064	16416	state5,cn=3	CSMD1	novel	3.48%	0.00%	0.017223	0.0714798
9	10394336	10455615	61279	state5,cn=3	PTPRD	novel	3.48%	0.00%	0.017223	0.0714798
20	61791411	61844885	53474	state5,cn=3	RTEL1, TNFRSF6B, LIME1, SLC2A4RG, ARFRP1, ZGPAT	located in CNV region	3.48%	0.00%	0.017223	0.0714798
1	1128775	1239049	110274	state5,cn=3	TNFRSF18, TNFRSF4, SDF4, UBE2J2, C10DC2, SCNN1D, CENTB5, PUSL1, CPSF3L	located in CNV region	5.22%	0.94%	0.06991	0.0737016
8	47647579	47654762	7183	state5,cn=3	intergenic	located in CNV region	5.22%	0.94%	0.06991	0.0737016
9	7999428	8005061	5633	state2,cn=1	intergenic	novel	5.22%	0.94%	0.06991	0.0737016
12	7891602	7993653	102051	state5,cn=3	SLC2A14	located in CNV region	5.22%	0.94%	0.06991	0.0737016
11	17254239	17295703	41464	state5,cn=3	NUCB2	novel, near CNV region	6.96%	1.89%	0.070061	0.0660045

6. DISCUSSION

Many neurological diseases, mental and neurodegenerative, have unknown or poorly-known etiologies. These are commonly known as complex disorders. To understand the etiology, the mechanisms involved in disease development and pathology have to be fully determined. Genetics and genetic studies have been crucial to identify genetic causes and factors leading to neurological diseases in recent years, which permit insights into disease pathology.

Until 2006, many Mendelian genes had been discovered in neurological diseases by candidate gene or linkage approaches. In the last five years, improvements in technology, especially in chip array platforms, enhanced the discovery of novel genes and loci. The advances in array technologies allow high throughput screening of the whole genome and genotyping of many samples simultaneously sparing substantial time. The unsolved part of complex diseases, mainly the sporadic forms, account for more than 90% of diseases. GWAS enable to perform large scale cohort studies between cases and controls to discover new candidate genomic regions that may be associated with disease phenotype. In the last five years, a tremendous amount of GWAS have been performed and thousands of variants and regions have been identified as candidate loci (Figure 1.6).

In ALS, like in most neurodegenerative and complex diseases, the unknown part of the pie is again 90%, which consists of the sporadic forms. On the other hand, more than 25 genes were identified in familial form of ALS, which account for only 50% of fALS. The missing part of genetic factors is still needed to be deciphered to complete the lacking players in the mechanisms underlying ALS pathogenesis.

In literature, 11 large scale GWAS on ALS have been reported so far. These have identified several SNPs and loci associated with ALS. However, none of those have been replicated in an independent study until today. The only recent exception is a chromosome 9p21 locus which was found to be associated with ALS in several populations (van Es *et al.*, 2009; Laaksovirta *et al.*, 2010; Shatunov *et al.*, 2010), several neighboring SNPs located in that loci displayed high signals. Unlike chromosome 9p21, the top significant

SNPs associated with ALS in other studies, did not gather a SNP cluster in a specific gene region or loci. The studies of Schymick and Cronin were the first trials in ALS-GWAS. Their results demonstrated significant associations of ZFP64, SUSD1, KIAA1727 and DPP6 (Table 1.2). Only DPP6 was replicated in another GWAS, however the same study population was used in both studies.

This thesis compiles the results of the first Turkish GWAS in ALS.

6.1. Genetic Association of fALS-causing Genes in sALS

In this part of this thesis, we aimed to investigate genetic association of SNPs, located in fALS-causing genes using sALS-GWAS data. For this purpose, two previously published GWAS from Schymick and Cronin were used. Besides identification of common low-risk loci in GWAS, we investigated Mendelian genes that result in disease phenotype. Identifying rare but high-risk markers would allow us new insights into sALS genetics.

In this respect, we analyzed single marker associations of SNPs in the regions of fALS-causing genes. Only two SNPs within TARDBP gene were shown to have border line significance in both datasets. After Bonferroni correction, both SNPs failed to keep significant association. Next, we investigated haplotypes that might be seen in both populations. Interestingly, only one haplotype, ACA, located in TARDBP gene (16kb) was co-inherited significantly in both datasets. However, when we calculated their ORs and considered case-control frequencies, a controversial correlation between both datasets, was found. Further, we investigated MAFs of SNPs in cases and controls in Schymick and Cronin datasets; in addition, MAFs of controls were also compared to data from the HapMap European population.

In fine-mapping analysis of a larger region (250kb), in order to see whether the controversial relationship between both datasets is caused by an experimental error, the same pattern of difference was observed in a more than 100kb region. The controls of two datasets gave also controversial results at TARDBP region. Next, we questioned whether CNVs are located in this very region and found previously reported CNVs nearby (Redon *et al.*, 2006). When we investigated the same region for a haplotype block and for CNVs,

neither the same haplotype block nor any CNVs were found in the same region of TARDBP locus in our Turkish cohort. These findings show us that population-specific results should be interpreted carefully, even if they are significant.

6.2. The Turkish ALS-GWAS

We have performed GWAS with 116 ALS patients and 109 controls on a 733k SNP platform, to reveal novel SNPs associating with the disease phenotype. Genotyping of 15 samples were replicated due to low coverage. Out of the total of 225 individuals (116+109), two did not pass the quality control (QC) due to low genotyping rate. Among 733k SNPs, approximately 100k SNPs were eliminated according to genotyping rates, HWE p-values and MAFs. We ended up with 223 individuals, including 114 cases and 109 controls and 631134 SNPs for association analyses.

In allelic and genotypic association tests, several markers (SNPs) were associated with ALS. Although many of those SNPs have low p-values, none of the SNPs could reach a genome-wide significance after Bonferroni corrections. When we analyzed the association results of the genotypic and allelic test, among the top 30 SNPs, several SNPs located in the region of two genes, TMEM90B (chromosome 20) and CPNE5 (chromosome 6), were found to be dominating. Associations of these two genes with ALS had not been previously shown and seem to be a novel finding.

Focusing to those genes by narrowing down to their gene-coding regions, the 3’ end of TMEM90B and CPNE5 had the highest signals within gene and neighboring intergenic regions. In haplotype block analysis, several blocks were in LD and showed significant association. In TMEM90B, among 11 haplotype block regions, seven blocks contained significant haplotypes. According to their p-values and OR, haplotypes with lowest p-values and highest frequencies, had low OR=~0.5 that represent protective haplotypes. In other words, those haplotypes located between block 4-10 had higher frequencies in controls compared to cases (Figure 5.10). On the other hand, in CPNE5, haplotypes with lowest p-values had high ORs (>2.0), pointing to higher susceptibility to ALS development, whereas there were also haplotypes with low ORs representing protective effects.

The TMEM90B gene, also known as C20ORF39 or SYNG1, encodes for the synapse differentiation inducing 1 protein. Its proposed function is regulation of AMPA receptor content at synapses and involvement in postsynaptic development and maturation. The protein is potentially a cell membrane protein, by comparing its similarity to known proteins. It has a long cytoplasmic domain, two helical and two extracellular matrix domains and a poly-glu domain with five glutamic acids. Its expression is observed mainly in parts of central nervous system, such as brain, cortex, cerebellum and the spinal cord. The critical role of this protein in the nervous system justifies the results obtained in our GWAS study.

The CPNE5 gene, also known as Copine-5, encodes for a calcium-dependent phospholipid binding protein. It may have a role in regulation of membrane trafficking; it is also expressed in several tissues, such as brain, spinal cord and kidney, also in neural progenitor cells and in the differentiated neurons during development of murine (Ding *et al.*, 2008).

According to their functional properties, both TMEM90B and CPNE5 can be promising candidates for investigating their potential roles in ALS development. In the literature, there are only very few reports about their functions and functional structures. Further investigation of expression profiles and protein functions of those proteins are required to gain a deeper knowledge about their role in ALS.

In conclusion, in our ALS-GWAS, we identified two novel gene loci on chromosomes 6 and 20. Being the first GWAS performed in Turkish sALS patients, however, this study needs replication in another and larger cohort. Increasing the sample size would favourably affect the genome-wide significance of SNPs analyzed. Further, the regions identified have to be investigated in more detail and risk-conferring and protective haplotypes defined should be validated in new sALS cases. Finally, these regions should be re-sequenced to define causative variants. The preliminary findings of this GWAS study are expected to shed some more light on the unknown and very complex genetics of sporadic ALS patients.

6.3. Genome-wide CNV Analysis of the Turkish ALS population

So far only four genotyping studies in ALS worldwide have investigated CNVs. The latest study by Blauw *et al.,2010,* reported 2 genes; DPP6 previously found to be associated with ALS in a SNP GWAS (Van Es *et al.*, 2007) and NIPA1 known to be causative for hereditary spastic paraparesis type 6 (Rainier *et al.*, 2003). However, independent replication of the DPP6 study by the above group did not validate the same association (Blauw *et al.*, 2010).

In this last part of our GWAS study, results of 116 ALS patients and 109 controls were subjected to CNV analysis. After the QC analysis, two ALS patients and three controls were eliminated. To reduce false positive CNV calls, GC model adjustment procedure was applied. First, we executed the program with default parameters to see the total CNV call outcomes, thus were able to determine how many false CNV calls we eliminated after QC. In default parameters, we identified 25000 CNVs, however, with the adjustment and QC analysis we only obtained 8000 CNV calls. Another important parameter was to check confidence value of each CNV which defines the probability of likelihood of CNVs present in a given state. 3000 CNVs were eliminated due to low probability of CNVs given in that state (confidence value<10).

Among 5000 CNV calls, several types of CNVs were observed: deletions (state1, cn=0), partial deletions (state2, cn=1), partial duplications (state5, cn=3) and duplications (state6, cn=4). Some CNVs were observed in both case and control samples. Most of the CNVs identified in our Turkish cohort were, as expected, already reported in population databases (HapMap and DGV databases). Some CNVs, present in both cases and controls, were tested by Chi-square and Fischer's exact test and were found to be still disease-associated.

When we compared CNV analysis results with *versus* without signal adjustment, the total amount of CNV calls decreased drastically per individual. Without signal adjustment, the average was approximately 85 CNV calls per individual, whereas, after adjustment, the average became 24 and 21 CNV calls in cases and controls, respectively. In the literature, several studies on healthy controls use the same strategy. It has been shown in a report that

in CEU (Caucasian European) population, the mean number of CNVs detected per individual is 21 (Wang *et al.*, 2007). In southeastern populations, the mean was approximately 45 CNVs per individual (Ku *et al.*, 2010). CNV analysis with signal adjustment showed, on the other hand, that the vast majority of false positive CNV calls were eliminated. Our data, with an average number of 21 CNVs per individual, is in concordance with the CEU population.

When CNVs were categorized, common and rare CNV calls, and overlapping and discrete CNV calls were obtained. Most of the CNVs categorized were found to be located in CNV regions (previously reported in databases). Some of them, on the other hand, are located in gene regions, such as MAP4K3, DPYD, ACYP2, LPHN3, CSMD1 and PTPRD.

As described in the results section, one of our top genes, MAP4K3, was found to have a duplication. Since it was shown in only two controls but not in any other populations, this CNV is novel in our cohort and its presence is high in ALS patients. According to literature, only one study reported a male Korean individual with an InDel in the MAP4K3 gene (Ahn *et al.*, 2009) and no association with ALS has been reported so far. MAP4K3 is ubiquitously expressed in all tissues examined with high levels in heart, brain, placenta, skeletal muscle, kidney and pancreas and lower levels in lung and liver. It has multiple functions in signal transduction of mammalian cells. One of them is to activate JNK pathway in response to stress. The second function is to induce the activation of the TORC1 complex to trigger cellular responses against starvation. Also, MAP4K3 is reported as a proapoptotic kinase that enhances mitochondrial apoptosis (Lam *et al.*, 2010). Mutations in MAP4K3 have been shown to cause pancreatic cancer (Jones *et al.*, 2008). Moreover, loss of function and over-expression of MAP4K3 protein were reported to affect cell growth and viability in the Drosophila wing (Resnik-Docampo and de Celis, 2011). These findings suggest that MAP4K3 is an important player in cell growth and cell survival. The alterations in gene expressions or protein levels might affect the cell viability in the central nervous system and we know that motor neuron survival is a key point in ALS.

Overlapping and discrete CNVs which are mainly common CNVs, did not reveal any novel regions in our study. Although they were observed in the control population in

related databases, they were also found to be associated with ALS when statistical tests were applied. In addition, rare CNVs have been observed in some patients. In the region of the ACYP2, LPHN3, CSMD1 and PTPRD genes, deletions and duplications have been observed. These regions are novel and so far not reported in databases, thus need to be further investigated in terms of their functions in the cell, involvement in cellular pathways and association with other diseases.

In this part of the study, we used the advantage of SNP raw data to investigate potential CNVs in ALS patients and controls. The potential CNVs were interpreted by PennCNV which is a powerful open source tool. We used this tool because it has a sophisticated hidden Markov model and several parameters for reduction of false positive CNV calls. Unfortunately, PennCNV predicts CNVs according to data, coming from the SNP array, so its accuracy is not 100%, although more than 10000 false positive CNV calls were eliminated. Thus, the results obtained by PennCNV need further validation analyses. One option is to use another CNV prediction tool to confirm whether the CNVs detected can be observed by the other tool. Another option is the construction of a custom CNV microarray, which would include the candidate regions predicted by PennCNV tool. Screening candidate loci between cases and controls would not only test the precision and the power of CNV detection with PennCNV, but it would also more reliably reveal the ALS-related CNVs.

To sum up, in the framework of this study, we were not able identify a common CNV in ALS patients. This may have several reasons:

- heterogeneity between ALS cases
  (i) unique CNVs among patients
  (ii) genetic diversity seen in the Turkish population
- rare CNVs which would not show significance in GWAS

On the other hand, we detected many rare CNVs in our ALS cohort. Those rare variants may be present in patients with similar symptoms, so these CNVs need fine-mapping experiments which would give more accurate and informative results on the association of CNVs with ALS.

APPENDIX A: HAPLOTYPE BLOCK ANALYSIS OF TMEM90B GENE REGION IN SCHYMCIK STUDY USING HAPLOVIEW

Table A.1. Haploview results of TMEM90B gene region in Schymick dataset.

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
Block 1					
	AC	0.494	0.494, 0.495	0.002	0.9679
	GC	0.4	0.391, 0.408	0.313	0.5761
	AT	0.105	0.114, 0.094	1.121	0.2897
Block 2					
	TA	0.678	0.706, 0.648	4.058	0.044
	CG	0.257	0.239, 0.276	1.947	0.1629
	CA	0.057	0.050, 0.066	1.245	0.2646
Block 3					
	ACCACC	0.585	0.592, 0.577	0.218	0.6409
	GTCGTC	0.258	0.236, 0.282	2.837	0.0921
	ACTGCT	0.092	0.095, 0.089	0.111	0.7393
	GCCGCC	0.037	0.048, 0.026	3.491	0.0617
Block 4					
	AA	0.446	0.442, 0.451	0.086	0.7693
	AG	0.443	0.453, 0.433	0.416	0.519
	GG	0.111	0.106, 0.117	0.309	0.5785
Block 5					
	CA	0.445	0.434, 0.457	0.526	0.4684
	TA	0.362	0.365, 0.360	0.032	0.8586
	CG	0.193	0.201, 0.184	0.485	0.4861
Block 6					
	TTCCC	0.236	0.243, 0.228	0.36	0.5483
	GCCCA	0.235	0.226, 0.245	0.516	0.4725
	GTCCA	0.189	0.195, 0.182	0.286	0.593
	GCCTA	0.183	0.180, 0.185	0.037	0.8475
	TTTCC	0.104	0.105, 0.104	0.002	0.9655
	TTCCA	0.038	0.036, 0.040	0.162	0.6869
Block 7					
	CACAGGTTCA	0.234	0.254, 0.214	2.294	0.1298
	CGCGGACTTA	0.226	0.217, 0.237	0.635	0.4256
	TGTGGGCCCG	0.189	0.197, 0.182	0.387	0.5339
	CGCGGACCCA	0.16	0.157, 0.164	0.099	0.7525
	CGCGAGCTTA	0.125	0.124, 0.128	0.038	0.8457
	CGCGGGCCCA	0.041	0.037, 0.046	0.549	0.4589
Block 8					
	TA	0.518	0.542, 0.492	2.619	0.1056
	CG	0.477	0.456, 0.500	2.018	0.1554
Block 9					
	AAG	0.662	0.679, 0.644	1.407	0.2356
	AGA	0.253	0.237, 0.271	1.561	0.2115
	GAG	0.062	0.062, 0.061	0.003	0.9581
	AGG	0.016	0.015, 0.018	0.17	0.6801

*Red indicates significant haplotypes which were lower than p=0.05

APPENDIX B: HAPLOTYPE BLOCK ANALYSIS OF TMEM90B GENE REGION IN CRONIN STUDY USING HAPLOVIEW

Table B.1. Haploview results of TMEM90B gene region in Cronin dataset.

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
Block 1					
	TA	0.597	0.588, 0.606	0.273	0.6012
	TG	0.223	0.243, 0.204	1.863	0.1722
	CG	0.175	0.166, 0.185	0.514	0.4734
Block 2					
	AAT	0.45	0.425, 0.476	2.249	0.1337
	GAC	0.268	0.272, 0.264	0.069	0.7923
	AGT	0.248	0.270, 0.226	2.17	0.1407
	AAC	0.03	0.026, 0.033	0.416	0.5192
Block 3					
	AGG	0.612	0.617, 0.607	0.093	0.7601
	GAG	0.275	0.284, 0.267	0.32	0.5717
	GGA	0.07	0.059, 0.081	1.618	0.2034
	GGG	0.042	0.038, 0.045	0.313	0.5757
Block 4					
	CTA	0.439	0.439, 0.440	0.002	0.9662
	TTG	0.221	0.225, 0.217	0.093	0.7608
	CTG	0.202	0.186, 0.219	1.485	0.2231
	CCG	0.136	0.148, 0.123	1.052	0.3052
Block 5					
	CTC	0.405	0.403, 0.406	0.009	0.9228
	TTT	0.295	0.293, 0.297	0.017	0.8966
	CCC	0.223	0.230, 0.217	0.218	0.6404
	TTC	0.067	0.059, 0.075	0.787	0.3749
	CTT	0.01	0.015, 0.006	1.786	0.1814
Block 6					
	AAGCGCGGG	0.252	0.251, 0.253	0.002	0.9675
	CGGCTCGGA	0.214	0.204, 0.224	0.482	0.4874
	CAGCTTAGG	0.167	0.160, 0.174	0.307	0.5798
	CGGTTCGGA	0.16	0.169, 0.150	0.57	0.4502
	AAACGCGAG	0.138	0.148, 0.128	0.662	0.416
	AAGCTCGGG	0.023	0.016, 0.031	1.95	0.1626
	CGGTTCGGG	0.017	0.016, 0.017	0.001	0.9787
	CGGTTCGAG	0.013	0.016, 0.010	0.786	0.3752
	CAGCGCGGG	0.012	0.012, 0.012	0	0.9825
Block 7					
	ATCA	0.381	0.385, 0.376	0.069	0.7924
	ACCA	0.267	0.270, 0.264	0.034	0.853
	GCCG	0.167	0.160, 0.174	0.308	0.5791
	GCTA	0.139	0.153, 0.126	1.227	0.2679
	GCCA	0.045	0.033, 0.057	2.896	0.0888

Table B.1. Haploview results of TMEM90B gene region in Cronin dataset (cont.).

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
Block 8					
	AA	0.538	0.507, 0.569	3.271	0.0705
	GG	0.46	0.491, 0.429	3.277	0.0703
Block 9					
	TC	0.734	0.749, 0.719	0.958	0.3278
	CT	0.242	0.213, 0.271	3.864	0.0493
	CC	0.014	0.021, 0.007	2.941	0.0863

*Red indicates significant haplotypes which were lower than p=0.05

APPENDIX C: HAPLOTYPE BLOCK ANALYSIS OF CPNE5 GENE REGION IN SCHYMICK STUDY USING HAPLOVIEW

Table C.1. Haploview results of CPNE5 gene region in Schymick dataset.

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
Block 1					
	CA	0.485	0.499, 0.471	0.823	0.3644
	CG	0.273	0.286, 0.259	1.001	0.3171
	TG	0.237	0.209, 0.267	5.007	0.0252
Block 2					
	CATCGAACTC	0.332	0.336, 0.331	0.025	0.8737
	AGTTAGGTCC	0.203	0.212, 0.195	0.428	0.5128
	AGCCAGATCT	0.174	0.153, 0.199	3.947	0.047
	AGTCGAACTC	0.128	0.135, 0.122	0.372	0.5418
	AGCCAGATCC	0.05	0.041, 0.060	2.021	0.1552
	CATCGAACCC	0.043	0.046, 0.040	0.246	0.6202
	AGTTAGATCC	0.043	0.044, 0.042	0.019	0.8894
Block 3					
	CCCCTC	0.342	0.352, 0.332	0.474	0.4914
	CTCCCC	0.227	0.197, 0.259	5.723	0.0167
	TCTACC	0.146	0.151, 0.140	0.262	0.6088
	TCCACC	0.11	0.119, 0.101	0.85	0.3566
	TCCCCC	0.094	0.095, 0.093	0.013	0.9094
	TCCACT	0.073	0.075, 0.071	0.052	0.819
Block 4					
	CAG	0.511	0.520, 0.502	0.345	0.5568
	TAA	0.276	0.258, 0.297	2.02	0.1553
	TAG	0.157	0.167, 0.146	0.894	0.3443
	TGG	0.054	0.053, 0.055	0.019	0.8892
Block 5					
	GCAGGAGAT	0.563	0.589, 0.538	2.741	0.0978
	AAGGTAAGC	0.148	0.123, 0.176	5.887	0.0153
	GCAGGAGAC	0.098	0.096, 0.101	0.07	0.7906
	GAAAGAGGC	0.093	0.092, 0.095	0.031	0.8597
	AAGGTGGGC	0.046	0.049, 0.042	0.374	0.541
	AAAGGAGGC	0.013	0.011, 0.016	0.468	0.494
Block 6					
	TGC	0.449	0.457, 0.440	0.305	0.581
	CGC	0.363	0.348, 0.378	1.008	0.3154
	CAT	0.107	0.106, 0.108	0.005	0.9411
	CGT	0.078	0.086, 0.070	0.936	0.3333
Block 7					
	CG	0.477	0.469, 0.486	0.312	0.5767
	AG	0.406	0.418, 0.393	0.66	0.4165
	CA	0.117	0.113, 0.121	0.14	0.708
Block 8					
	GG	0.465	0.471, 0.458	0.175	0.6755
	AT	0.348	0.345, 0.352	0.055	0.8146
	AG	0.187	0.184, 0.190	0.062	0.8035
Block 9					
	TTTAC	0.525	0.540, 0.510	0.976	0.3232
	CCTGT	0.221	0.231, 0.211	0.568	0.4512

Table C.1. Haploview results of CPNE5 gene region in Schymick dataset (cont.).

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
	CTCGT	0.124	0.104, 0.146	4.207	0.0402
	TTTGT	0.111	0.105, 0.119	0.536	0.4643
Block 10					
	GCTTT	0.253	0.266, 0.239	1.047	0.3061
	GCTGT	0.216	0.191, 0.243	4.217	0.04
	GCTTC	0.205	0.220, 0.188	1.561	0.2116
	AACTT	0.188	0.184, 0.192	0.113	0.7372
	GCCTT	0.075	0.060, 0.091	3.563	0.0591
	GACTT	0.059	0.073, 0.043	4.04	0.0444
Block 11					
	TT	0.514	0.504, 0.526	0.512	0.4744
	TG	0.375	0.387, 0.362	0.713	0.3984
	CG	0.111	0.109, 0.113	0.027	0.8704
Block 12					
	CAG	0.361	0.367, 0.354	0.194	0.6597
	CGG	0.287	0.274, 0.302	1.054	0.3046
	TAG	0.209	0.223, 0.194	1.335	0.2479
	CAA	0.143	0.137, 0.150	0.381	0.5369
Block 13					
	GGT	0.371	0.377, 0.365	0.164	0.6857
	AGT	0.269	0.252, 0.289	1.801	0.1796
	GTC	0.229	0.237, 0.221	0.372	0.5417
	GTT	0.127	0.128, 0.125	0.026	0.8722
Block 14					
	AA	0.452	0.473, 0.431	1.895	0.1686
	CA	0.293	0.290, 0.297	0.058	0.8101
	CG	0.247	0.231, 0.265	1.54	0.2146

*Red indicates significant haplotypes which were lower than p=0.05

APPENDIX D: HAPLOTYPE BLOCK ANALYSIS OF CPNE5 GENE REGION IN CRONIN STUDY USING HAPLOVIEW

Table D.1. Haploview results of CPNE5 gene region in Cronin dataset.

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
Block 1					
	CT	0.447	0.453, 0.442	0.089	0.7649
	CC	0.317	0.317, 0.317	0	0.9949
	TC	0.234	0.230, 0.238	0.076	0.7829
Block 2					
	GAACCAAGG	0.344	0.347, 0.342	0.03	0.8623
	TGATTGGAG	0.181	0.176, 0.186	0.149	0.6995
	TGACCAAGG	0.18	0.181, 0.180	0.002	0.9664
	TGGCTGAAA	0.167	0.152, 0.183	1.482	0.2234
	TGGCTGAAG	0.063	0.078, 0.048	3.283	0.07
	TGATTGAAG	0.036	0.035, 0.037	0.03	0.8628
Block 3					
	GCCCTC	0.306	0.318, 0.294	0.551	0.4578
	GTCCCC	0.236	0.239, 0.233	0.044	0.8345
	ACTACC	0.19	0.195, 0.186	0.114	0.7354
	ACCACC	0.105	0.106, 0.104	0.007	0.9321
	ACCCCC	0.077	0.068, 0.086	0.99	0.3197
	ACCACT	0.074	0.070, 0.078	0.221	0.6385
Block 4					
	CG	0.458	0.453, 0.463	0.081	0.7756
	TA	0.333	0.342, 0.324	0.34	0.5599
	TG	0.208	0.202, 0.214	0.17	0.6804
Block 5					
	CCA	0.677	0.683, 0.671	0.132	0.7166
	TAG	0.214	0.209, 0.219	0.128	0.7204
	CAA	0.086	0.082, 0.090	0.184	0.6678
	TAA	0.018	0.024, 0.012	1.609	0.2047
Block 6					
	CC	0.78	0.782, 0.779	0.012	0.9128
	AT	0.173	0.174, 0.171	0.008	0.9301
	AC	0.047	0.045, 0.050	0.137	0.7114
Block 7					
	GA	0.686	0.695, 0.676	0.341	0.5592
	AG	0.158	0.166, 0.151	0.325	0.5689
	GG	0.156	0.140, 0.173	1.744	0.1867
Block 8					
	AT	0.618	0.627, 0.609	0.275	0.6003
	GC	0.302	0.305, 0.300	0.027	0.8694
	GT	0.077	0.068, 0.086	0.953	0.329
Block 9					
	CG	0.762	0.769, 0.754	0.274	0.6007
	CA	0.119	0.116, 0.122	0.087	0.7687
	TA	0.115	0.108, 0.123	0.51	0.4749
Block 10					
	AG	0.439	0.486, 0.390	7.824	0.0052
	CG	0.422	0.387, 0.457	4.227	0.0398
	CA	0.139	0.127, 0.152	1.157	0.2822

Table D.1. Haploview results of CPNE5 gene region in Cronin dataset (cont.).

	Haplotype	Freq. in population	Case, control ratios	Chi Square	p-value
Block 11					
	GC	0.51	0.524, 0.497	0.615	0.433
	AA	0.342	0.329, 0.355	0.599	0.4388
	AC	0.148	0.147, 0.149	0.005	0.9446
Block 12					
	ATATG	0.525	0.542, 0.507	1.045	0.3066
	GCACA	0.232	0.225, 0.238	0.193	0.6605
	GTGCA	0.141	0.143, 0.138	0.045	0.8312
	ATACA	0.087	0.075, 0.100	1.64	0.2003
Block 13					
	GCAAA	0.238	0.221, 0.255	1.346	0.2459
	GCAAG	0.21	0.211, 0.209	0.003	0.9541
	GCACA	0.208	0.202, 0.214	0.198	0.6566
	AAGAA	0.196	0.191, 0.202	0.146	0.7029
	GCGAA	0.086	0.117, 0.055	10.501	0.0012
	GAGAA	0.06	0.055, 0.065	0.357	0.5501
Block 14					
	GTG	0.382	0.401, 0.364	1.235	0.2664
	GTT	0.277	0.268, 0.286	0.347	0.5558
	ATT	0.234	0.220, 0.247	0.858	0.3544
	GCG	0.105	0.110, 0.100	0.24	0.6244
Block 15					
	GGGT	0.289	0.276, 0.302	0.657	0.4177
	GGTC	0.226	0.218, 0.235	0.361	0.5481
	AGGT	0.212	0.211, 0.212	0.001	0.9819
	GAGT	0.148	0.171, 0.124	3.798	0.0513
	GGTT	0.12	0.123, 0.117	0.056	0.8133

*Red indicates significant haplotypes which were lower than p=0.05

REFERENCES

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.

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.

Arai, T., M. Hasegawa, H. Akiyama, K. Ikeda, T. Nonaka, H. Mori, D. Mann, K. Tsuchiya, M. Yoshida, Y. Hashizume, and T. Oda, 2006, "Tdp-43 Is a Component of Ubiquitin-Positive Tau-Negative Inclusions in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis", *Biochemical and Biophysical Research Communication,* Vol. 351, No. 3, pp. 602-611.

Barrett, J. C., D. G. Clayton, P. Concannon, B. Akolkar, J. D. Cooper, H. A. Erlich, C. Julier, G. Morahan, J. Nerup, C. Nierras, V. Plagnol, F. Pociot, H. Schuilenburg, D. J. Smyth, H. Stevens, J. A. Todd, N. M. Walker, and S. S. Rich, 2009, "Genome-Wide Association Study and Meta-Analysis Find That over 40 Loci Affect Risk of Type 1 Diabetes", *Nature Genetics*, Vol. 41, No. 6, pp. 703-707.

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly, 2005, "Haploview: Analysis and Visualization of Ld and Haplotype Maps", *Bioinformatics,* Vol. 21, No. 2, pp. 263-265.

Barrett, J. C., S. Hansoul, D. L. Nicolae, J. H. Cho, R. H. Duerr, J. D. Rioux, S. R. Brant, M. S. Silverberg, K. D. Taylor, M. M. Barmada, A. Bitton, T. Dassopoulos, L. W. Datta, T. Green, A. M. Griffiths, E. O. Kistner, M. T. Murtha, M. D. Regueiro, J. I. Rotter, L. P. Schumm, A. H. Steinhart, S. R. Targan, R. J. Xavier, C. Libioulle, C. Sandor, M. Lathrop, J. Belaiche, O. Dewit, I. Gut, S. Heath, D. Laukens, M. Mni, P. Rutgeerts, A. Van Gossum, D. Zelenika, D. Franchimont, J. P. Hugot, M. de Vos, S. Vermeire, E. Louis, L. R. Cardon, C. A. Anderson, H. Drummond, E. Nimmo, T. Ahmad, N. J. Prescott, C. M. Onnie, S. A. Fisher, J. Marchini, J. Ghori,

S. Bumpstead, R. Gwilliam, M. Tremelling, P. Deloukas, J. Mansfield, D. Jewell, J. Satsangi, C. G. Mathew, M. Parkes, M. Georges, and M. J. Daly, 2008, "Genome-Wide Association Defines More Than 30 Distinct Susceptibility Loci for Crohn's Disease", *Nature Genetics,* Vol. 40, No. 8, pp. 955-962.

Bellingham, M. C., 2011, "A Review of the Neural Mechanisms of Action and Clinical Efficiency of Riluzole in Treating Amyotrophic Lateral Sclerosis: What Have We Learned in the Last Decade?", *CNS Neuroscience Therapeutics,* Vol. 17, No. 1, pp. 4-31.

Bento-Abreu, A., P. Van Damme, L. Van Den Bosch, and W. Robberecht, 2010, "The Neurobiology of Amyotrophic Lateral Sclerosis", *European Journal of Neuroscience,* Vol. 31, No. 12, pp. 2247-2265.

Bertram, L., and R. E. Tanzi, 2005, "The Genetic Epidemiology of Neurodegenerative Disease", *The Journal of Clinical Investigation,* Vol. 115, No. 6, pp. 1449-1457.

Blauw, H. M., A. Al-Chalabi, P. M. Andersen, P. W. van Vught, F. P. Diekstra, M. A. van Es, C. G. Saris, E. J. Groen, W. van Rheenen, M. Koppers, R. Van't Slot, E. Strengman, K. Estrada, F. Rivadeneira, A. Hofman, A. G. Uitterlinden, L. A. Kiemeney, S. H. Vermeulen, A. Birve, S. Waibel, T. Meyer, S. Cronin, R. L. McLaughlin, O. Hardiman, P. C. Sapp, M. D. Tobin, L. V. Wain, B. Tomik, A. Slowik, R. Lemmens, D. Rujescu, C. Schulte, T. Gasser, R. H. Brown, Jr., J. E. Landers, W. Robberecht, A. C. Ludolph, R. A. Ophoff, J. H. Veldink, and L. H. van den Berg, 2010, "A Large Genome Scan for Rare Cnvs in Amyotrophic Lateral Sclerosis", *Human Molecular Genetics,* Vol. 19, No. 20, pp. 4091-4099.

Blauw, H. M., C. P. Barnes, P. W. van Vught, W. van Rheenen, M. Verheul, E. Cuppen, J. H. Veldink, and L. H. van den Berg, 2012, "SMN1 Gene Duplications Are Associated with Sporadic ALS", *Neurology,* Vol. 78, No. 11, pp. 776-780.

Blauw, H. M., J. H. Veldink, M. A. van Es, P. W. van Vught, C. G. Saris, B. van der Zwaag, L. Franke, J. P. Burbach, J. H. Wokke, R. A. Ophoff, and L. H. van den Berg, 2008, "Copy-Number Variation in Sporadic Amyotrophic Lateral Sclerosis: A Genome-Wide Screen", *The Lancet Neurology,* Vol. 7, No. 4, pp. 319-326.

Brettschneider, J., V. M. Van Deerlin, J. L. Robinson, L. Kwong, E. B. Lee, Y. O. Ali, N. Safren, M. J. Monteiro, J. B. Toledo, L. Elman, L. McCluskey, D. J. Irwin, M. Grossman, L. Molina-Porcel, V. M. Lee, and J. Q. Trojanowski, 2012, "Pattern of Ubiquilin Pathology in Als and Ftld Indicates Presence of C9orf72 Hexanucleotide Expansion", *Acta Neuropathologica*, Vol. 123, No. 6, pp. 825-839.

Brooks, B. R., R. G. Miller, M. Swash, and T. L. Munsat, 2000, "El Escorial Revisited: Revised Criteria for the Diagnosis of Amyotrophic Lateral Sclerosis", *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, Vol. 1, No. 5, pp. 293-299.

Brunetti-Pierri, N., A. R. Paciorkowski, R. Ciccone, E. Della Mina, M. C. Bonaglia, R. Borgatti, C. P. Schaaf, V. R. Sutton, Z. Xia, N. Jelluma, C. Ruivenkamp, M. Bertrand, T. J. de Ravel, P. Jayakar, S. Belli, K. Rocchetti, C. Pantaleoni, S. D'Arrigo, J. Hughes, S. W. Cheung, O. Zuffardi, and P. Stankiewicz, 2011, "Duplications of Foxg1 in 14q12 Are Associated with Developmental Epilepsy, Mental Retardation, and Severe Speech Impairment", *European Journal of Human Genetics*, Vol. 19, No. 1, pp. 102-107.

Charcot, J. M. and A. Joffory, 1869, “Deux Cas D’atrophie Musculaire Progressive avec Lesions de la Substance Grise et des Faisceaux Antero-lateraux de la Moelle Epiniere”, *Archives of Physiology Neurology Pathology*, Vol. 2, pp. 744-754.

Chen, H. J., G. Anagnostou, A. Chai, J. Withers, A. Morris, J. Adhikaree, G. Pennetta, and J. S. de Belleroche, 2010, "Characterization of the Properties of a Novel Mutation in Vapb in Familial Amyotrophic Lateral Sclerosis", *The Journal of Biological Chemistry*, Vol. 285, No. 51, pp. 40266-40281.

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

Chio, A., J. C. Schymick, G. Restagno, S. W. Scholz, F. Lombardo, S. L. Lai, G. Mora, H. C. Fung, A. Britton, S. Arepalli, J. R. Gibbs, M. Nalls, S. Berger, L. C. Kwee, E. Z. Oddone, J. Ding, C. Crews, I. Rafferty, N. Washecka, D. Hernandez, L. Ferrucci, S. Bandinelli, J. Guralnik, F. Macciardi, F. Torri, S. Lupoli, S. J. Chanock, G. Thomas, D. J. Hunter, C. Gieger, H. E. Wichmann, A. Calvo, R. Mutani, S. Battistini, F. Giannini, C. Caponnetto, G. L. Mancardi, V. La Bella, F. Valentino, M. R. Monsurro, G. Tedeschi, K. Marinou, M. Sabatelli, A. Conte, J. Mandrioli, P. Sola, F. Salvi, I. Bartolomei, G. Siciliano, C. Carlesi, R. W. Orrell, K. Talbot, Z. Simmons, J. Connor, E. P. Pioro, T. Dunkley, D. A. Stephan, D. Kasperaviciute, E. M. Fisher, S. Jabonka, M. Sendtner, M. Beck, L. Bruijn, J. Rothstein, S. Schmidt, A. Singleton, J. Hardy, and B. J. Traynor, 2009, "A Two-Stage Genome-Wide Association Study of Sporadic Amyotrophic Lateral Sclerosis", *Human Molecular Genetics*, Vol. 18, No. 8, pp. 1524-1532.

Chow, C. Y., J. E. Landers, S. K. Bergren, P. C. Sapp, A. E. Grant, J. M. Jones, L. Everett, G. M. Lenk, D. M. McKenna-Yasek, L. S. Weisman, D. Figlewicz, R. H. Brown, and M. H. Meisler, 2009, "Deleterious Variants of Fig4, a Phosphoinositide Phosphatase, in Patients with ALS", *The American Journal of Human Genetics*, Vol. 84, No. 1, pp. 85-88.

Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C. H. Ihm, K. Kristiansson, D. G. Macarthur, J. R. Macdonald, I. Onyiah, A. W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Tyler-Smith, N. P. Carter, C. Lee, S. W. Scherer, and M. E. Hurles, 2010, "Origins and Functional Impact of Copy Number Variation in the Human Genome", *Nature*, Vol. 464, No. 7289, pp. 704-712.

Corcia, P., C. Ingre, H. Blasco, R. Press, J. Praline, C. Antar, C. Veyrat-Durebex, Y. O. Guettard, W. Camu, P. M. Andersen, P. Vourc'h, and C. R. Andres, 2012, "Homozygous SMN2 Deletion Is a Protective Factor in the Swedish ALS Population", *European Journal of Human Genetics*, Vol. 20, No. 5, pp. 588-591.

Cozzolino, M., M. G. Pesaresi, V. Gerbino, J. Grosskreutz, and M. T. Carri, 2012, "Amyotrophic Lateral Sclerosis: New Insights into Underlying Molecular

Mechanisms and Opportunities for Therapeutic Intervention", *Antioxidant and Redox Signalling*.

Cronin, S., S. Berger, J. Ding, J. C. Schymick, N. Washecka, D. G. Hernandez, M. J. Greenway, D. G. Bradley, B. J. Traynor, and O. Hardiman, 2007, "A Genome-Wide Association Study of Sporadic ALS in a Homogenous Irish Population", *Human Molecular Genetics,* Vol. 17, No. 5, pp. 768-774.

Cronin, S., H. M. Blauw, J. H. Veldink, M. A. van Es, R. A. Ophoff, D. G. Bradley, L. H. van den Berg, and O. Hardiman, 2008, "Analysis of Genome-Wide Copy Number Variation in Irish and Dutch ALS Populations", *Human Molecular Genetics,* Vol. 17, No. 21, pp. 3392-3398.

Cronin, S., B. Tomik, D. G. Bradley, A. Slowik, and O. Hardiman, 2009, "Screening for Replication of Genome-Wide Snp Associations in Sporadic ALS", *European Journal of Human Genetics,* Vol. 17, No. 2, pp. 213-218.

Da Cruz, S., and D. W. Cleveland, 2011, "Understanding the Role of Tdp-43 and Fus/Tls in ALS and Beyond", *Current Opinion in Neurobiology,* Vol. 21, No. 6, pp. 904-919.

Daoud, H., P. N. Valdmanis, E. Kabashi, P. Dion, N. Dupre, W. Camu, V. Meininger, and G. A. Rouleau, 2009, "Contribution of Tardbp Mutations to Sporadic Amyotrophic Lateral Sclerosis", *Journal of Medical Genetics,* Vol. 46, No. 2, pp. 112-114.

De Amicis, A., M. Piane, F. Ferrari, M. Fanciulli, D. Delia and L. Chessa, 2011, “Role of senataxin in DNA damage and telomeric stability”, *DNA Repair (Amst),*Vol. 10, pp. 199-209.

DeJesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, A. M. Nicholson, N. A. Finch, H. Flynn, J. Adamson, N. Kouri, A. Wojtas, P. Sengdy, G. Y. Hsiung, A. Karydas, W. W. Seeley, K. A. Josephs, G. Coppola, D. H. Geschwind, Z. K. Wszolek, H. Feldman, D. S. Knopman, R. C. Petersen, B. L. Miller, D. W. Dickson, K. B. Boylan, N. R. Graff-Radford, and R. Rademakers, 2011, "Expanded GGGGCC Hexanucleotide Repeat in Noncoding

Region of C9orf72 Causes Chromosome 9p-Linked FTD and ALS", *Neuron,* Vol. 72, No. 2, pp. 245-256.

Deng, H. X., W. Chen, S. T. Hong, K. M. Boycott, G. H. Gorrie, N. Siddique, Y. Yang, F. Fecto, Y. Shi, H. Zhai, H. Jiang, M. Hirano, E. Rampersaud, G. H. Jansen, S. Donkervoort, E. H. Bigio, B. R. Brooks, K. Ajroud, R. L. Sufit, J. L. Haines, E. Mugnaini, M. A. Pericak-Vance, and T. Siddique, 2011, "Mutations in Ubqln2 Cause Dominant X-Linked Juvenile and Adult-Onset ALS and ALS/Dementia", *Nature,* Vol. 477, No. 7363, pp. 211-215.

Deng, H. X., H. Zhai, R. Fu, Y. Shi, G. H. Gorrie, Y. Yang, E. Liu, M. C. Dal Canto, E. Mugnaini, and T. Siddique, 2007, "Distal Axonopathy in an Alsin-Deficient Mouse Model", *Human Molecular Genetics,* Vol. 16, No. 23, pp. 2911-2920.

Ding, X., Y. Jin, Y. Wu, H. Wu, L. Xiong, X. Song, S. Liu, W. Fan, and M. Fan, 2008, "Localization and Cellular Distribution of Cpne5 in Embryonic Mouse Brain", *Brain Research,* Vol. 1224, No. pp. 20-28.

Dunckley, T., M. J. Huentelman, D. W. Craig, J. V. Pearson, S. Szelinger, K. Joshipura, R. F. Halperin, C. Stamper, K. R. Jensen, D. Letizia, S. E. Hesterlee, A. Pestronk, T. Levine, T. Bertorini, M. C. Graves, T. Mozaffar, C. E. Jackson, P. Bosch, A. McVey, A. Dick, R. Barohn, C. Lomen-Hoerth, J. Rosenfeld, T. O'Connor D, K. Zhang, R. Crook, H. Ryberg, M. Hutton, J. Katz, E. P. Simpson, H. Mitsumoto, R. Bowser, R. G. Miller, S. H. Appel, and D. A. Stephan, 2007, "Whole-Genome Analysis of Sporadic Amyotrophic Lateral Sclerosis", *The New England Journal of Medicine,* Vol. 357, No. 8, pp. 775-788.

Durbin, R.M., G.R.Abecasis, and D.L. Altshuler, 2010, "A map of human genome variation from population-scale sequencing" *Nature*, Vol. 467, pp. 1061–1073.

Edwards, T. L., W. K. Scott, C. Almonte, A. Burt, E. H. Powell, G. W. Beecham, L. Wang, S. Zuchner, I. Konidari, G. Wang, C. Singer, F. Nahab, B. Scott, J. M. Stajich, M. Pericak-Vance, J. Haines, J. M. Vance, and E. R. Martin, 2010, "Genome-Wide Association Study Confirms Snps in Snca and the Mapt Region as Common Risk

Factors for Parkinson Disease", *Annals of Human Genetics,* Vol. 74, No. 2, pp. 97-109.

Elden, A. C., H. J. Kim, M. P. Hart, A. S. Chen-Plotkin, B. S. Johnson, X. Fang, M. Armakola, F. Geser, R. Greene, M. M. Lu, A. Padmanabhan, D. Clay-Falcone, L. McCluskey, L. Elman, D. Juhr, P. J. Gruber, U. Rub, G. Auburger, J. Q. Trojanowski, V. M. Lee, V. M. Van Deerlin, N. M. Bonini, and A. D. Gitler, 2010, "Ataxin-2 Intermediate-Length Polyglutamine Expansions Are Associated with Increased Risk for ALS", *Nature,* Vol. 466, No. 7310, pp. 1069-1075.

Fellermann, K., D. E. Stange, E. Schaeffeler, H. Schmalzl, J. Wehkamp, C. L. Bevins, W. Reinisch, A. Teml, M. Schwab, P. Lichter, B. Radlwimmer, and E. F. Stange, 2006, "A Chromosome 8 Gene-Cluster Polymorphism with Low Human Beta-Defensin 2 Gene Copy Number Predisposes to Crohn Disease of the Colon", *The American Journal of Human Genetics,* Vol. 79, No. 3, pp. 439-448.

Fernandez-Santiago, R., S. Hoenig, P. Lichtner, A. D. Sperfeld, M. Sharma, D. Berg, O. Weichenrieder, T. Illig, K. Eger, T. Meyer, J. Anneser, C. Munch, S. Zierz, T. Gasser, and A. Ludolph, 2009, "Identification of Novel Angiogenin (ANG) Gene Missense Variants in German Patients with Amyotrophic Lateral Sclerosis", *Journal of Neurology,* Vol. 256, No. 8, pp. 1337-1342.

Ferraiuolo, L., J. Kirby, A. J. Grierson, M. Sendtner, and P. J. Shaw, 2011, "Molecular Pathways of Motor Neuron Injury in Amyotrophic Lateral Sclerosis", *Nature Reviews Neurology,* Vol. 7, No. 11, pp. 616-630.

Feuk, L., A. R. Carson, and S. W. Scherer, 2006, "Structural Variation in the Human Genome", *Nature Reviews Genetics,* Vol. 7, No. 2, pp. 85-97.

Gandhi, S., and N. W. Wood, 2010, "Genome-Wide Association Studies: The Key to Unlocking Neurodegeneration?", *Nature Neuroscience,* Vol. 13, No. 7, pp. 789-794.

Gellera, C., N. Ticozzi, V. Pensato, L. Nanetti, A. Castucci, B. Castellotti, G. Lauria, F. Taroni, V. Silani, and C. Mariotti, 2012, "Ataxin2 CAG-Repeat Length in Italian Patients with Amyotrophic Lateral Sclerosis: Risk Factor or Variant Phenotype?

Implication for Genetic Testing and Counseling", *Neurobiology of Aging*, Vol. 33, No. 8, pp. 1847.

Giordana, M. T., M. Piccinini, S. Grifoni, G. De Marco, M. Vercellino, M. Magistrello, A. Pellerino, B. Buccinna, E. Lupino, and M. T. Rinaudo, 2010, "TDP-43 Redistribution Is an Early Event in Sporadic Amyotrophic Lateral Sclerosis", *Brain Pathology*, Vol. 20, No. 2, pp. 351-360.

Gitcho, M. A., R. H. Baloh, S. Chakraverty, K. Mayo, J. B. Norton, D. Levitch, K. J. Hatanpaa, C. L. White, 3rd, E. H. Bigio, R. Caselli, M. Baker, M. T. Al-Lozi, J. C. Morris, A. Pestronk, R. Rademakers, A. M. Goate, and N. J. Cairns, 2008, "TDP-43 A315T Mutation in Familial Motor Neuron Disease", *Annals of Neurology*, Vol. 63, No. 4, pp. 535-538.

Greenway, M. J., M. D. Alexander, S. Ennis, B. J. Traynor, B. Corr, E. Frost, A. Green, and O. Hardiman, 2004, "A Novel Candidate Region for ALS on Chromosome 14q11.2", *Neurology*, Vol. 63, No. 10, pp. 1936-1938.

Greenway, M. J., P. M. Andersen, C. Russ, S. Ennis, S. Cashman, C. Donaghy, V. Patterson, R. Swingler, D. Kieran, J. Prehn, K. E. Morrison, A. Green, K. R. Acharya, R. H. Brown, Jr., and O. Hardiman, 2006, "ANG Mutations Segregate with Familial and 'Sporadic' Amyotrophic Lateral Sclerosis", *Nature Genetics*, Vol. 38, No. 4, pp. 411-413.

Hadano, S., C. K. Hand, H. Osuga, Y. Yanagisawa, A. Otomo, R. S. Devon, N. Miyamoto, J. Showguchi-Miyata, Y. Okada, R. Singaraja, D. A. Figlewicz, T. Kwiatkowski, B. A. Hosler, T. Sagie, J. Skaug, J. Nasir, R. H. Brown, Jr., S. W. Scherer, G. A. Rouleau, M. R. Hayden, and J. E. Ikeda, 2001, "A Gene Encoding a Putative GTPase Regulator Is Mutated in Familial Amyotrophic Lateral Sclerosis 2", *Nature Genetics*, Vol. 29, No. 2, pp. 166-173.

Hadano, S., A. Otomo, R. Kunita, K. Suzuki-Utsunomiya, A. Akatsuka, M. Koike, M. Aoki, Y. Uchiyama, Y. Itoyama, and J. E. Ikeda, 2010, "Loss of ALS2/Alsin Exacerbates Motor Dysfunction in a SOD1-Expressing Mouse ALS Model by Disturbing Endolysosomal Trafficking", *PLoS One*, Vol. 5, No. 3.

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

Harold, D., R. Abraham, P. Hollingworth, R. Sims, A. Gerrish, M. L. Hamshere, J. S. Pahwa, V. Moskvina, K. Dowzell, A. Williams, N. Jones, C. Thomas, A. Stretton, A. R. Morgan, S. Lovestone, J. Powell, P. Proitsi, M. K. Lupton, C. Brayne, D. C. Rubinsztein, M. Gill, B. Lawlor, A. Lynch, K. Morgan, K. S. Brown, P. A. Passmore, D. Craig, B. McGuinness, S. Todd, C. Holmes, D. Mann, A. D. Smith, S. Love, P. G. Kehoe, J. Hardy, S. Mead, N. Fox, M. Rossor, J. Collinge, W. Maier, F. Jessen, B. Schurmann, H. van den Bussche, I. Heuser, J. Kornhuber, J. Wiltfang, M. Dichgans, L. Frolich, H. Hampel, M. Hull, D. Rujescu, A. M. Goate, J. S. Kauwe, C. Cruchaga, P. Nowotny, J. C. Morris, K. Mayo, K. Sleegers, K. Bettens, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, G. Livingston, N. J. Bass, H. Gurling, A. McQuillin, R. Gwilliam, P. Deloukas, A. Al-Chalabi, C. E. Shaw, M. Tsolaki, A. B. Singleton, R. Guerreiro, T. W. Muhleisen, M. M. Nothen, S. Moebus, K. H. Jockel, N. Klopp, H. E. Wichmann, M. M. Carrasquillo, V. S. Pankratz, S. G. Younkin, P. A. Holmans, M. O'Donovan, M. J. Owen, and J. Williams, 2009, "Genome-Wide Association Study Identifies Variants at CLU and PICALM Associated with Alzheimer's Disease", *Nature Genetics,* Vol. 41, No. 10, pp. 1088-1093.

Hindorff, L. A., P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, and T. A. Manolio, 2009, "Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits", *Proceedings of The National Academy Sciences,* Vol. 106, No. 23, pp. 9362-9367.

Hirano, M., C. M. Quinzii, H. Mitsumoto, A. P. Hays, J. K. Roberts, P. Richard, and L. P. Rowland, 2011, "Senataxin Mutations and Amyotrophic Lateral Sclerosis", *Amyotrophic Lateral Sclerosis,* Vol. 12, No. 3, pp. 223-227.

Hunter, D. J., G. Thomas, R. N. Hoover, and S. J. Chanock, 2007, "Scanning the Horizon: What Is the Future of Genome-Wide Association Studies in Accelerating

Discoveries in Cancer Etiology and Prevention?", *Cancer Causes and Control,* Vol. 18, No. 5, pp. 479-484.

Jakobsdottir, J., M. B. Gorin, Y. P. Conley, R. E. Ferrell, and D. E. Weeks, 2009, "Interpretation of Genetic Association Studies: Markers with Replicated Highly Significant Odds Ratios May Be Poor Classifiers", *PLoS Genetics,* Vol. 5, No. 2.

Johnson, J. O., J. Mandrioli, M. Benatar, Y. Abramzon, V. M. Van Deerlin, J. Q. Trojanowski, J. R. Gibbs, M. Brunetti, S. Gronka, J. Wuu, J. Ding, L. McCluskey, M. Martinez-Lage, D. Falcone, D. G. Hernandez, S. Arepalli, S. Chong, J. C. Schymick, J. Rothstein, F. Landi, Y. D. Wang, A. Calvo, G. Mora, M. Sabatelli, M. R. Monsurro, S. Battistini, F. Salvi, R. Spataro, P. Sola, G. Borghero, G. Galassi, S. W. Scholz, J. P. Taylor, G. Restagno, A. Chio, and B. J. Traynor, 2010, "Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS", *Neuron,* Vol. 68, No. 5, pp. 857-864.

Jones, S., X. Zhang, D. W. Parsons, J. C. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, H. Kamiyama, A. Jimeno, S. M. Hong, B. Fu, M. T. Lin, E. S. Calhoun, M. Kamiyama, K. Walter, T. Nikolskaya, Y. Nikolsky, J. Hartigan, D. R. Smith, M. Hidalgo, S. D. Leach, A. P. Klein, E. M. Jaffee, M. Goggins, A. Maitra, C. Iacobuzio-Donahue, J. R. Eshleman, S. E. Kern, R. H. Hruban, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu, and K. W. Kinzler, 2008, "Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses", *Science,* Vol. 321, No. 5897, pp. 1801-1806.

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", *Lancet,* Vol. 377, No. 9769, pp. 942-955.

Klein, R. J., C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, and J. Hoh, 2005, "Complement Factor H Polymorphism in Age-Related Macular Degeneration", *Science,* Vol. 308, No. 5720, pp. 385-389.

Ku, C. S., E. Y. Loy, Y. Pawitan, and K. S. Chia, 2010, "The Pursuit of Genome-Wide Association Studies: Where Are We Now?", *Journal of Human Genetics*, Vol. 55, No. 4, pp. 195-206.

Laaksovirta, H., T. Peuralinna, J. C. Schymick, S. W. Scholz, S. L. Lai, L. Myllykangas, R. Sulkava, L. Jansson, D. G. Hernandez, J. R. Gibbs, M. A. Nalls, D. Heckerman, P. J. Tienari, and B. J. Traynor, 2010, "Chromosome 9p21 in Amyotrophic Lateral Sclerosis in Finland: A Genome-Wide Association Study", *The Lancet Neurology*, Vol. 9, No. 10, pp. 978-985.

Lam, D., S. Shah, I. P. de Castro, S. H. Loh, and L. M. Martins, 2010, "Drosophila Happyhour Modulates Jnk-Dependent Apoptosis", *Cell Death and Disease*, Vol. 1.

Lambert, J. C., S. Heath, G. Even, D. Campion, K. Sleegers, M. Hiltunen, O. Combarros, D. Zelenika, M. J. Bullido, B. Tavernier, L. Letenneur, K. Bettens, C. Berr, F. Pasquier, N. Fievet, P. Barberger-Gateau, S. Engelborghs, P. De Deyn, I. Mateo, A. Franck, S. Helisalmi, E. Porcellini, O. Hanon, M. M. de Pancorbo, C. Lendon, C. Dufouil, C. Jaillard, T. Leveillard, V. Alvarez, P. Bosco, M. Mancuso, F. Panza, B. Nacmias, P. Bossu, P. Piccardi, G. Annoni, D. Seripa, D. Galimberti, D. Hannequin, F. Licastro, H. Soininen, K. Ritchie, H. Blanche, J. F. Dartigues, C. Tzourio, I. Gut, C. Van Broeckhoven, A. Alperovitch, M. Lathrop, and P. Amouyel, 2009, "Genome-Wide Association Study Identifies Variants at CLU and CR1 Associated with Alzheimer's Disease", *Nature Genetics*, Vol. 41, No. 10, pp. 1094-1099.

Landers, J. E., J. Melki, V. Meininger, J. D. Glass, L. H. van den Berg, M. A. van Es, P. C. Sapp, P. W. van Vught, D. M. McKenna-Yasek, H. M. Blauw, T. J. Cho, M. Polak, L. Shi, A. M. Wills, W. J. Broom, N. Ticozzi, V. Silani, A. Ozoguz, I. Rodriguez-Leyva, J. H. Veldink, A. J. Ivinson, C. G. Saris, B. A. Hosler, A. Barnes-Nessa, N. Couture, J. H. Wokke, T. J. Kwiatkowski, Jr., R. A. Ophoff, S. Cronin, O. Hardiman, F. P. Diekstra, P. N. Leigh, C. E. Shaw, C. L. Simpson, V. K. Hansen, J. F. Powell, P. Corcia, F. Salachas, S. Heath, P. Galan, F. Georges, H. R. Horvitz, M. Lathrop, S. Purcell, A. Al-Chalabi, and R. H. Brown, Jr., 2009, "Reduced Expression of the Kinesin-Associated Protein 3 (Kifap3) Gene Increases Survival

in Sporadic Amyotrophic Lateral Sclerosis", *Proceedings of The National Academy Sciences*, Vol. 106, No. 22, pp. 9004-9009.

Luty, A. A., J. B. Kwok, C. Dobson-Stone, C. T. Loy, K. G. Coupland, H. Karlstrom, T. Sobow, J. Tchorzewska, A. Maruszak, M. Barcikowska, P. K. Panegyres, C. Zekanowski, W. S. Brooks, K. L. Williams, I. P. Blair, K. A. Mather, P. S. Sachdev, G. M. Halliday, and P. R. Schofield, 2010, "Sigma Nonopioid Intracellular Receptor 1 Mutations Cause Frontotemporal Lobar Degeneration-Motor Neuron Disease", *Annals of Neurology*, Vol. 68, No. 5, pp. 639-649.

Majounie, E., A. E. Renton, K. Mok, E. G. Dopper, A. Waite, S. Rollinson, A. Chio, G. Restagno, N. Nicolaou, J. Simon-Sanchez, J. C. van Swieten, Y. Abramzon, J. O. Johnson, M. Sendtner, R. Pamphlett, R. W. Orrell, S. Mead, K. C. Sidle, H. Houlden, J. D. Rohrer, K. E. Morrison, H. Pall, K. Talbot, O. Ansorge, D. G. Hernandez, S. Arepalli, M. Sabatelli, G. Mora, M. Corbo, F. Giannini, A. Calvo, E. Englund, G. Borghero, G. L. Floris, A. M. Remes, H. Laaksovirta, L. McCluskey, J. Q. Trojanowski, V. M. Van Deerlin, G. D. Schellenberg, M. A. Nalls, V. E. Drory, C. S. Lu, T. H. Yeh, H. Ishiura, Y. Takahashi, S. Tsuji, I. Le Ber, A. Brice, C. Drepper, N. Williams, J. Kirby, P. Shaw, J. Hardy, P. J. Tienari, P. Heutink, H. R. Morris, S. Pickering-Brown, and B. J. Traynor, 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.

Maller, J., S. George, S. Purcell, J. Fagerness, D. Altshuler, M. J. Daly, and J. M. Seddon, 2006, "Common Variation in Three Genes, Including a Noncoding Variant in Cfh, Strongly Influences Risk of Age-Related Macular Degeneration", *Nature Genetics*, Vol. 38, No. 9, pp. 1055-1059.

Manolio, T. A., 2010, "Genomewide Association Studies and Assessment of the Risk of Disease", *The New England Journal of Medicine*, Vol. 363, No. 2, pp. 166-176.

Marian, A. J., 2012, "Molecular Genetic Studies of Complex Phenotypes", *Translational Research*, Vol. 159, No. 2, pp. 64-79.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis, and J. N. Hirschhorn, 2008, "Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges", *Nature Reviews Genetics*, Vol. 9, No. 5, pp. 356-369.

McCarthy, M. I., and J. N. Hirschhorn, 2008, "Genome-Wide Association Studies: Past, Present and Future", *Human Molecular Genetics*, Vol. 17, No. 2, pp. 100-101.

Millecamps, S., F. Salachas, C. Cazeneuve, P. Gordon, B. Bricka, A. Camuzat, L. Guillot-Noel, O. Russaouen, G. Bruneteau, P. F. Pradat, N. Le Forestier, N. Vandenberghe, V. Danel-Brunaud, N. Guy, C. Thauvin-Robinet, L. Lacomblez, P. Couratier, D. Hannequin, D. Seilhean, I. Le Ber, P. Corcia, W. Camu, A. Brice, G. Rouleau, E. LeGuern, and V. Meininger, 2010, "SOD1, ANG, VAPB, TARDBP, and FUS Mutations in Familial Amyotrophic Lateral Sclerosis: Genotype-Phenotype Correlations", *Journal of Medical Genetics*, Vol. 47, No. 8, pp. 554-560.

Mok, K., B. J. Traynor, J. Schymick, P. J. Tienari, H. Laaksovirta, T. Peuralinna, L. Myllykangas, A. Chio, A. Shatunov, B. F. Boeve, A. L. Boxer, M. DeJesus-Hernandez, I. R. Mackenzie, A. Waite, N. Williams, H. R. Morris, J. Simon-Sanchez, J. C. van Swieten, P. Heutink, G. Restagno, G. Mora, K. E. Morrison, P. J. Shaw, P. S. Rollinson, A. Al-Chalabi, R. Rademakers, S. Pickering-Brown, R. W. Orrell, M. A. Nalls, and J. Hardy, 2012, "Chromosome 9 ALS and FTD Locus Is Probably Derived from a Single Founder", *Neurobiology of Aging*, Vol. 33, No. 1, pp. 203-208.

Moreira, M. C., S. Klur, M. Watanabe, A. H. Nemeth, I. Le Ber, J. C. Moniz, C. Tranchant, P. Aubourg, M. Tazir, L. Schols, M. Pandolfo, J. B. Schulz, J. Pouget, P. Calvas, M. Shizuka-Ikeda, M. Shoji, M. Tanaka, L. Izatt, C. E. Shaw, A. M'Zahem, E. Dunne, P. Bomont, T. Benhassine, N. Bouslam, G. Stevanin, A. Brice, J. Guimaraes, P. Mendonca, C. Barbot, P. Coutinho, J. Sequeiros, A. Durr, J. M. Warter, and M. Koenig, 2004, "Senataxin, the Ortholog of a Yeast RNA Helicase, Is Mutant in Ataxia-Ocular Apraxia 2", *Nature Genetics*, Vol. 36, No. 3, pp. 225-227.

Morita, M., A. Al-Chalabi, P. M. Andersen, B. Hosler, P. Sapp, E. Englund, J. E. Mitchell, J. J. Habgood, J. de Belleroche, J. Xi, W. Jongjaroenprasert, H. R. Horvitz, L. G. Gunnarsson, and R. H. Brown, Jr., 2006, "A Locus on Chromosome 9p Confers Susceptibility to ALS and Frontotemporal Dementia", *Neurology,* Vol. 66, No. 6, pp. 839-844.

Neumann, M., D. M. Sampathu, L. K. Kwong, A. C. Truax, M. C. Micsenyi, T. T. Chou, J. Bruce, T. Schuck, M. Grossman, C. M. Clark, L. F. McCluskey, B. L. Miller, E. Masliah, I. R. Mackenzie, H. Feldman, W. Feiden, H. A. Kretzschmar, J. Q. Trojanowski, and V. M. Lee, 2006, "Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis", *Science,* Vol. 314, No. 5796, pp. 130-133.

Nishimura, A. L., M. Mitne-Neto, H. C. Silva, A. Richieri-Costa, S. Middleton, D. Cascio, F. Kok, J. R. Oliveira, T. Gillingwater, J. Webb, P. Skehel, and M. Zatz, 2004, "A Mutation in the Vesicle-Trafficking Protein VAPB Causes Late-Onset Spinal Muscular Atrophy and Amyotrophic Lateral Sclerosis", *The American Journal of Human Genetics,* Vol. 75, No. 5, pp. 822-831.

Orlacchio, A., C. Babalini, A. Borreca, C. Patrono, R. Massa, S. Basaran, R. P. Munhoz, E. A. Rogaeva, P. H. St George-Hyslop, G. Bernardi, and T. Kawarai, 2010, "Spatacsin Mutations Cause Autosomal Recessive Juvenile Amyotrophic Lateral Sclerosis", *Brain,* Vol. 133, No. 2, pp. 591-598.

Paubel, A., J. Violette, M. Amy, J. Praline, V. Meininger, W. Camu, P. Corcia, C. R. Andres, and P. Vourc'h, 2008, "Mutations of the Ang Gene in French Patients with Sporadic Amyotrophic Lateral Sclerosis", *Archives Neurology,* Vol. 65, No. 10, pp. 1333-1336.

Pearson, T. A., and T. A. Manolio, 2008, "How to Interpret a Genome-Wide Association Study", *The Journal of The American Medical Association,* Vol. 299, No. 11, pp. 1335-1344.

Polymenidou, M., and D. W. Cleveland, 2011, "The Seeds of Neurodegeneration: Prion-Like Spreading in ALS", *Cell,* Vol. 147, No. 3, pp. 498-508.

Rainier, S., J. H. Chai, D. Tokarz, R. D. Nicholls, and J. K. Fink, 2003, "NIPA1 Gene Mutations Cause Autosomal Dominant Hereditary Spastic Paraplegia (SPG6)", *The American Journal of Human Genetics,* Vol. 73, No. 4, pp. 967-971.

Ratnaparkhi, A., G. M. Lawless, F. E. Schweizer, P. Golshani, and G. R. Jackson, 2008, "A Drosophila Model of ALS: Human ALS-Associated Mutation in VAP33A Suggests a Dominant Negative Mechanism", *PLoS One,* Vol. 3, No. 6.

Reaume, A. G., J. L. Elliott, E. K. Hoffman, N. W. Kowall, R. J. Ferrante, D. F. Siwek, H. M. Wilcox, D. G. Flood, M. F. Beal, R. H. Brown, Jr., R. W. Scott, and W. D. Snider, 1996, "Motor Neurons in Cu/Zn Superoxide Dismutase-Deficient Mice Develop Normally but Exhibit Enhanced Cell Death after Axonal Injury", *Nature Genetics,* Vol. 13, No. 1, pp. 43-47.

Redon, R., S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles, 2006, "Global Variation in Copy Number in the Human Genome", *Nature,* Vol. 444, No. 7118, pp. 444-454.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas, H. Kalimo, A. Paetau, Y. Abramzon, A. M. Remes, A. Kaganovich, S. W. Scholz, J. Duckworth, J. Ding, D. W. Harmer, D. G. Hernandez, J. O. Johnson, K. Mok, M. Ryten, D. Trabzuni, R. J. Guerreiro, R. W. Orrell, J. Neal, A. Murray, J. Pearson, I. E. Jansen, D. Sondervan, H. Seelaar, D. Blake, K. Young, N. Halliwell, J. B. Callister, G. Toulson, A. Richardson, A. Gerhard, J. Snowden, D. Mann, D. Neary, M. A. Nalls, T. Peuralinna, L. Jansson, V. M. Isoviita, A. L. Kaivorinne, M. Holtta-Vuori, E. Ikonen, R. Sulkava, M. Benatar, J. Wuu, A. Chio, G. Restagno, G. Borghero, M. Sabatelli, D. Heckerman, E. Rogaeva, L. Zinman, J. D. Rothstein, M. Sendtner, C. Drepper, E. E. Eichler, C. Alkan, Z. Abdullaev, S. D. Pack, A. Dutra, E. Pak, J. Hardy, A. Singleton, N. M. Williams, P. Heutink, S. Pickering-Brown, H. R.

Morris, P. J. Tienari, and B. J. Traynor, 2011, "A Hexanucleotide Repeat Expansion in C9orf72 Is the Cause of Chromosome 9p21-Linked ALS-FTD", *Neuron,* Vol. 72, No. 2, pp. 257-268.

Resnik-Docampo, M., and J. F. de Celis, 2011, "Map4k3 Is a Component of the Torc1 Signalling Complex That Modulates Cell Growth and Viability in Drosophila Melanogaster", *PLoS One,* Vol. 6, No. 1.

Rohrer, B., Q. Long, B. Coughlin, R. B. Wilson, Y. Huang, F. Qiao, P. H. Tang, K. Kunchithapautham, G. S. Gilkeson, and S. Tomlinson, 2009, "A Targeted Inhibitor of the Alternative Complement Pathway Reduces Angiogenesis in a Mouse Model of Age-Related Macular Degeneration", *Investigative Ophthalmology & Visual Science,* Vol. 50, No. 7, pp. 3056-3064.

Rosen, D. R., T. Siddique, D. Patterson, D. A. Figlewicz, P. Sapp, A. Hentati, D. Donaldson, J. Goto, J. P. O'Regan, H. X. Deng *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.

Rowland, L. P., and N. A. Shneider, 2001, "Amyotrophic Lateral Sclerosis", *The New England Journal of Medicine,* Vol. 344, No. 22, pp. 1688-1700.

Sapp, P. C., B. A. Hosler, D. McKenna-Yasek, W. Chin, A. Gann, H. Genise, J. Gorenstein, M. Huang, W. Sailer, M. Scheffler, M. Valesky, J. L. Haines, M. Pericak-Vance, T. Siddique, H. R. Horvitz, and R. H. Brown, Jr., 2003, "Identification of Two Novel Loci for Dominantly Inherited Familial Amyotrophic Lateral Sclerosis", *The American Journal of Human Genetics,* Vol. 73, No. 2, pp. 397-403.

Satake, W., Y. Nakabayashi, I. Mizuta, Y. Hirota, C. Ito, M. Kubo, T. Kawaguchi, T. Tsunoda, M. Watanabe, A. Takeda, H. Tomiyama, K. Nakashima, K. Hasegawa, F. Obata, T. Yoshikawa, H. Kawakami, S. Sakoda, M. Yamamoto, N. Hattori, M. Murata, Y. Nakamura, and T. Toda, 2009, "Genome-Wide Association Study Identifies Common Variants at Four Loci as Genetic Risk Factors for Parkinson's Disease", *Nature Genetics,* Vol. 41, No. 12, pp. 1303-1307.

Schork, N. J., S. S. Murray, K. A. Frazer, and E. J. Topol, 2009, "Common Vs. Rare Allele Hypotheses for Complex Diseases", *Current Opinion in Genetics & Development,* Vol. 19, No. 3, pp. 212-219.

Schymick, J. C., S. W. Scholz, H. C. Fung, A. Britton, S. Arepalli, J. R. Gibbs, F. Lombardo, M. Matarin, D. Kasperaviciute, D. G. Hernandez, C. Crews, L. Bruijn, J. Rothstein, G. Mora, G. Restagno, A. Chio, A. Singleton, J. Hardy, and B. J. Traynor, 2007, "Genome-Wide Genotyping in Amyotrophic Lateral Sclerosis and Neurologically Normal Controls: First Stage Analysis and Public Release of Data", *The Lancet Neurology,* Vol. 6, No. 4, pp. 322-328.

Shatunov, A., K. Mok, S. Newhouse, M. E. Weale, B. Smith, C. Vance, L. Johnson, J. H. Veldink, M. A. van Es, L. H. van den Berg, W. Robberecht, P. Van Damme, O. Hardiman, A. E. Farmer, C. M. Lewis, A. W. Butler, O. Abel, P. M. Andersen, I. Fogh, V. Silani, A. Chio, B. J. Traynor, J. Melki, V. Meininger, J. E. Landers, P. McGuffin, J. D. Glass, H. Pall, P. N. Leigh, J. Hardy, R. H. Brown, Jr., J. F. Powell, R. W. Orrell, K. E. Morrison, P. J. Shaw, C. E. Shaw, and A. Al-Chalabi, 2010, "Chromosome 9p21 in Sporadic Amyotrophic Lateral Sclerosis in the Uk and Seven Other Countries: A Genome-Wide Association Study", *The Lancet Neurology,* Vol. 9, No. 10, pp. 986-994.

Shi, J., D. F. Levinson, J. Duan, A. R. Sanders, Y. Zheng, I. Pe'er, F. Dudbridge, P. A. Holmans, A. S. Whittemore, B. J. Mowry, A. Olincy, F. Amin, C. R. Cloninger, J. M. Silverman, N. G. Buccola, W. F. Byerley, D. W. Black, R. R. Crowe, J. R. Oksenberg, D. B. Mirel, K. S. Kendler, R. Freedman, and P. V. Gejman, 2009, "Common Variants on Chromosome 6p22.1 Are Associated with Schizophrenia", *Nature,* Vol. 460, No. 7256, pp. 753-757.

Simon-Sanchez, J., C. Schulte, J. M. Bras, M. Sharma, J. R. Gibbs, D. Berg, C. Paisan-Ruiz, P. Lichtner, S. W. Scholz, D. G. Hernandez, R. Kruger, M. Federoff, C. Klein, A. Goate, J. Perlmutter, M. Bonin, M. A. Nalls, T. Illig, C. Gieger, H. Houlden, M. Steffens, M. S. Okun, B. A. Racette, M. R. Cookson, K. D. Foote, H. H. Fernandez, B. J. Traynor, S. Schreiber, S. Arepalli, R. Zonozi, K. Gwinn, M. van der Brug, G. Lopez, S. J. Chanock, A. Schatzkin, Y. Park, A. Hollenbeck, J. Gao, X. Huang, N. W. Wood, D. Lorenz, G. Deuschl, H. Chen, O. Riess, J. A.

Hardy, A. B. Singleton, and T. Gasser, 2009, "Genome-Wide Association Study Reveals Genetic Risk Underlying Parkinson's Disease", *Nature Genetics*, Vol. 41, No. 12, pp. 1308-1312.

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

Stefansson, H., D. Rujescu, S. Cichon, O. P. Pietilainen, A. Ingason, S. Steinberg, R. Fossdal, E. Sigurdsson, T. Sigmundsson, J. E. Buizer-Voskamp, T. Hansen, K. D. Jakobsen, P. Muglia, C. Francks, P. M. Matthews, A. Gylfason, B. V. Halldorsson, D. Gudbjartsson, T. E. Thorgeirsson, A. Sigurdsson, A. Jonasdottir, A. Bjornsson, S. Mattiasdottir, T. Blondal, M. Haraldsson, B. B. Magnusdottir, I. Giegling, H. J. Moller, A. Hartmann, K. V. Shianna, D. Ge, A. C. Need, C. Crombie, G. Fraser, N. Walker, J. Lonnqvist, J. Suvisaari, A. Tuulio-Henriksson, T. Paunio, T. Toulopoulou, E. Bramon, M. Di Forti, R. Murray, M. Ruggeri, E. Vassos, S. Tosato, M. Walshe, T. Li, C. Vasilescu, T. W. Muhleisen, A. G. Wang, H. Ullum, S. Djurovic, I. Melle, J. Olesen, L. A. Kiemeney, B. Franke, C. Sabatti, N. B. Freimer, J. R. Gulcher, U. Thorsteinsdottir, A. Kong, O. A. Andreassen, R. A. Ophoff, A. Georgi, M. Rietschel, T. Werge, H. Petursson, D. B. Goldstein, M. M. Nothen, L. Peltonen, D. A. Collier, D. St Clair, and K. Stefansson, 2008, "Large Recurrent Microdeletions Associated with Schizophrenia", *Nature*, Vol. 455, No. 7210, pp. 232-236.

Stranger, B. E., M. S. Forrest, M. Dunning, C. E. Ingle, C. Beazley, N. Thorne, R. Redon, C. P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S. W. Scherer, S. Tavare, P. Deloukas, M. E. Hurles, and E. T. Dermitzakis, 2007, "Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes", *Science*, Vol. 315, No. 5813, pp. 848-853.

Sullivan, P. F., D. Lin, J. Y. Tzeng, E. van den Oord, D. Perkins, T. S. Stroup, M. Wagner, S. Lee, F. A. Wright, F. Zou, W. Liu, A. M. Downing, J. Lieberman, and S. L.

Close, 2008, "Genomewide Association for Schizophrenia in the Catie Study: Results of Stage 1", *Molecular Psychiatry,* Vol. 13, No. 6, pp. 570-584.

Suraweera, A., Y. Lim, R. Woods, G. W. Birrell, T. Nasim, O. J. Becherel, and M. F. Lavin, 2009, "Functional Role for Senataxin, Defective in Ataxia Oculomotor Apraxia Type 2, in Transcriptional Regulation", *Human Molecular Genetics,* Vol. 18, No. 18, pp. 3384-3396.

Tollervey, J. R., T. Curk, B. Rogelj, M. Briese, M. Cereda, M. Kayikci, J. Konig, T. Hortobagyi, A. L. Nishimura, V. Zupunski, R. Patani, S. Chandran, G. Rot, B. Zupan, C. E. Shaw, and J. Ule, 2011, "Characterizing the Rna Targets and Position-Dependent Splicing Regulation by TDP-43", *Nature Neuroscience,* Vol. 14, No. 4, pp. 452-458.

Tsai, C. P., B. W. Soong, K. P. Lin, P. H. Tu, J. L. Lin, and Y. C. Lee, 2011, "FUS, TARDBP, and SOD1 Mutations in a Taiwanese Cohort with Familial ALS", *Neurobiology of Aging,* Vol. 32, No. 3, pp. 513-521.

Van Damme, P., J. H. Veldink, M. van Blitterswijk, A. Corveleyn, P. W. van Vught, V. Thijs, B. Dubois, G. Matthijs, L. H. van den Berg, and W. Robberecht, 2011, "Expanded Atxn2 CAG Repeat Size in ALS Identifies Genetic Overlap between ALS and SCA2", *Neurology,* Vol. 76, No. 24, pp. 2066-2072.

van Es, M. A., F. P. Diekstra, J. H. Veldink, F. Baas, P. R. Bourque, H. J. Schelhaas, E. Strengman, E. A. Hennekam, D. Lindhout, R. A. Ophoff, and L. H. van den Berg, 2009, "A Case of ALS-FTD in a Large FALS Pedigree with a K17I ANG Mutation", *Neurology,* Vol. 72, No. 3, pp. 287-288.

van Es, M. A., P. W. Van Vught, H. M. Blauw, L. Franke, C. G. Saris, P. M. Andersen, L. Van Den Bosch, S. W. de Jong, R. van 't Slot, A. Birve, R. Lemmens, V. de Jong, F. Baas, H. J. Schelhaas, K. Sleegers, C. Van Broeckhoven, J. H. Wokke, C. Wijmenga, W. Robberecht, J. H. Veldink, R. A. Ophoff, and L. H. van den Berg, 2007, "Itpr2 as a Susceptibility Gene in Sporadic Amyotrophic Lateral Sclerosis: A Genome-Wide Association Study", *The Lancet Neurology,* Vol. 6, No. 10, pp. 869-877.

van Es, M. A., J. H. Veldink, C. G. Saris, H. M. Blauw, P. W. van Vught, A. Birve, R. Lemmens, H. J. Schelhaas, E. J. Groen, M. H. Huisman, A. J. van der Kooi, M. de Visser, C. Dahlberg, K. Estrada, F. Rivadeneira, A. Hofman, M. J. Zwarts, P. T. van Doormaal, D. Rujescu, E. Strengman, I. Giegling, P. Muglia, B. Tomik, A. Slowik, A. G. Uitterlinden, C. Hendrich, S. Waibel, T. Meyer, A. C. Ludolph, J. D. Glass, S. Purcell, S. Cichon, M. M. Nothen, H. E. Wichmann, S. Schreiber, S. H. Vermeulen, L. A. Kiemeney, J. H. Wokke, S. Cronin, R. L. McLaughlin, O. Hardiman, K. Fumoto, R. J. Pasterkamp, V. Meininger, J. Melki, P. N. Leigh, C. E. Shaw, J. E. Landers, A. Al-Chalabi, R. H. Brown, Jr., W. Robberecht, P. M. Andersen, R. A. Ophoff, and L. H. van den Berg, 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.

Vance, C., A. Al-Chalabi, D. Ruddy, B. N. Smith, X. Hu, J. Sreedharan, T. Siddique, H. J. Schelhaas, B. Kusters, D. Troost, F. Baas, V. de Jong, and C. E. Shaw, 2006, "Familial Amyotrophic Lateral Sclerosis with Frontotemporal Dementia Is Linked to a Locus on Chromosome 9p13.2-21.3", *Brain,* Vol. 129, No. 4, pp. 868-876.

Vance, C., B. Rogelj, T. Hortobagyi, K. J. De Vos, A. L. Nishimura, J. Sreedharan, X. Hu, B. Smith, D. Ruddy, P. Wright, J. Ganesalingam, K. L. Williams, V. Tripathi, S. Al-Saraj, A. Al-Chalabi, P. N. Leigh, I. P. Blair, G. Nicholson, J. de Belleroche, J. M. Gallo, C. C. Miller, and C. E. Shaw, 2009, "Mutations in Fus, an Rna Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6", *Science,* Vol. 323, No. 5918, pp. 1208-1211.

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

Wang, K., M. Li, D. Hadley, R. Liu, J. Glessner, S. F. Grant, H. Hakonarson, and M. Bucan, 2007, "Penncnv: An Integrated Hidden Markov Model Designed for High-Resolution Copy Number Variation Detection in Whole-Genome SNP Genotyping Data", *Genome Research,* Vol. 17, No. 11, pp. 1665-1674.

Wu, D., W. Yu, H. Kishikawa, R. D. Folkerth, A. J. Iafrate, Y. Shen, W. Xin, K. Sims, and G. F. Hu, 2007, "Angiogenin Loss-of-Function Mutations in Amyotrophic Lateral Sclerosis", *Annals of Neurology,* Vol. 62, No. 6, pp. 609-617.

Yang, Y., A. Hentati, H. X. Deng, O. Dabbagh, T. Sasaki, M. Hirano, W. Y. Hung, K. Ouahchi, J. Yan, A. C. Azim, N. Cole, G. Gascon, A. Yagmour, M. Ben-Hamida, M. Pericak-Vance, F. Hentati, and T. Siddique, 2001, "The Gene Encoding Alsin, a Protein with Three Guanine-Nucleotide Exchange Factor Domains, Is Mutated in a Form of Recessive Amyotrophic Lateral Sclerosis", *Nature Genetics,* Vol. 29, No. 2, pp. 160-165.

Yu, Z., Y. Zhu, A. S. Chen-Plotkin, D. Clay-Falcone, L. McCluskey, L. Elman, R. G. Kalb, J. Q. Trojanowski, V. M. Lee, V. M. Van Deerlin, A. D. Gitler, and N. M. Bonini, 2011, "PolyQ Repeat Expansions in ATXN2 Associated with ALS Are CAA Interrupted Repeats", *PLoS One,* Vol. 6, No. 3.

Zuk, O., E. Hechter, S. R. Sunyaev, and E. S. Lander, 2012, "The Mystery of Missing Heritability: Genetic Interactions Create Phantom Heritability", *Proceedings of The National Academy Sciences,* Vol. 109, No. 4, pp. 1193-1198.