AN INVESTIGATION OF RECENT AND NOVEL GENETIC VARIANTS THAT ARE ASSOCIATED WITH THE PATHOGENESIS OF AMYOTROPHIC LATERAL SCLEROSIS AND THEIR IMPLICATIONS ON PHENOTYPES OF THE DISEASE

Kwok, Chun Tak

In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy

Division of Brain Sciences
Department of Medicine
Imperial College London
2014

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease caused by loss of motor neurons in the spinal cord, brain stem and cerebral cortex. ALS is characterized by both upper and lower motor neuron symptoms and death usually occurs 3-5 years after onset. Familial histories are found in 5-10% of ALS cases while the rest are sporadic. This study is focused on analysing known and novel candidate genes in ALS and the aims of study are to characterize causal genes and risk factors for Familial ALS (FALS) and Sporadic ALS (SALS) in the Imperial Cohorts, in which genetic causes have been assigned for 64% of FALS cases. Three strategies were pursued and genes involved in proteostasis pathways were emphasized in this study.

Firstly, we sequenced known candidate genes in our FALS cases excluded for known mutations. *VCP* and *SQSTM1* genes were sequenced. We did not identify any coding changes in *VCP* but report a 5' hexanucleotide expansion exclusively found in ALS. Known and novel *SQSTM1* mutations, P392L and E155K, were identified in FALS kindred presenting with a history of Paget's disease of bone.

Secondly, we carried out association studies for two candidate genes on Chromosome 17, *P4HB* and *NPLOC4*, and showed that they were risk factors for FALS and SALS respectively. The association of *P4HB* SNPs with FALS survival time indicates that it is a modifier gene.

Thirdly, to explore novel genes in ALS, we investigated Variable number tandem repeats (VNTR) from top candidate genes selected based on association signals from previous Genome wide association (GWA) studies and protein functions. VNTRs in *NIPA1* and *HSPB8* gene were associated with FALS and SALS respectively. Finally, we characterized the size of the reported hexanucleotide GGGGCC expansion in the *C9orf72* gene using Southern blot analysis in our FALS cohort and interim results are presented.

DECLARATION

I declare that this thesis is my own composition and data present herein is my own work unless otherwise stated. This thesis has not been submitted in any form for another degree or diploma at any university or institute.

Kwok, Chun Tak

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor Professor Jackie de Belleroche for giving me the precious opportunity to study Neurogenetics, an area that always interests me. During my study, Jackie experiencedly guided me through all PhD milestones and encouraged me writing for publications.

I would like to thank Dr. Alex Morris for his daily guidance. Alex patiently took me through all practical steps in my research projects from pipetting to statistics with his broad knowledge in genetics and basic sciences.

I would also like to acknowledge people who carried out the experiments for some results that are presented in this thesis. Ms Jennifer Frampton carried out an initial association study of the *P4HB* gene in her MSc Project (2008-2009). Ms Hsiang-Ya Wang and Dr. Alex Morris carried out the sequencing of 5' and 3'UTR, Exon 1, 13 and 17 of the *VCP* gene and the screening of the hexanucleotide repeat in ALS cases and Controls in Wang's MSc Project (2010-2011). Dr. Alex Morris also carried out fragment length analysis of the *NPLOC4* TG repeat.

Finally I would like to thank my family for unconditionally supporting my study.

DEDICATION

This thesis is dedicated to my parents, Kin Lun Kwok and Siu Yin Kwok.

CONTENTS

Chapter 1 Introduction to ALS

Chapter 2 Methodology and Materials

Chapter 3 Sequence Analysis of *VCP* Gene in the Imperial College ALS Cohort

Chapter 4 Sequence Analysis of *SQSTM1* Gene in the Imperial College ALS Cohort

Chapter 5 Association Studies of Candidate Genes in ALS

Chapter 6 VNTR analysis in ALS

Chapter 7 Characterization of *C9orf72* GGGGCC expansion

Chapter 8 Discussion

DETAILED CONTENTS

ABSTRACT 2

DECLARATION 3

ACKNOWLEDGEMENTS 4

DEDICATION 5

CONTENTS 6

DETAILED CONTENTS 7

LIST OF FIGURES 12

LIST OF TABLES 14

ABBREVIATIONS 16

CHAPTER 1 INTRODUCTION TO ALS 19

1.1 Clinical features of ALS: Symptoms begin in various sites and progress throughout the body 20

1.2 The diagnosis of ALS is based on clinical and electrophysiological findings 21

1.3 Current treatments for ALS are limited 26

1.4 Pathological findings in ALS 26

1.5 Genetic studies have led to the understanding of underlying pathogenesis of ALS 27

1.6 Mutant SOD1 causes oxidative stress, a common feature in ALS 35

1.7 Mutant SOD1 proteins are prone to aggregate 38

1.8 ER Stress, Unfolded Protein Response (UPR) and *VAPB* Mutations (ALS8) 39

1.9 Mitochondrial dysfunction 42

1.10 Neurofilaments, impaired axonal transport and *DCTN1* mutations 43

1.11 Excitotoxicity and *DAO* mutations 44

1.12 Apotosis: a possible route to neuronal death 45

1.13 Other forms of protein aggregates, *FUS* (ALS6) and *TARDBP* (ALS10) mutations 46

1.14 The Ubiquitin-Proteasome System (UPS), autophagy and *VCP* mutations (ALS14) 49

1.15 The NF-kB pathway: TDP-43 aggregates, *SQSTM1* and *OPTN* mutations 52

1.16 ALS-FTD and *C9orf72* mutations 53

1.17 Other ALS Loci 55

1.17.1 Alsin (ALS2) mutations 55

1.17.2 18q21 (ALS3) 56
1.17.3 *SETX* (Senataxin, ALS4) 56
1.17.4 *SPG11*, Spatacsin (ALS5) 57
1.17.5 20p13 (ALS7) 58
1.17.6 ALS9: Angiogenin (*ANG*) 58
1.17.7 *FIG4* (ALS11) 60
1.17.8 *OPTN* (ALS12) 61
1.17.9 *ATXN2* (Ataxin-2, ALS13) 62
1.17.10 *UBQLN2* (ALS15) 63
1.17.11 *SIGMAR1* (ALS16) 64
1.17.12 *CHMP2B* (ALS17) 65
1.17.13 *PFN-1* (ALS18) 65

1.18 Sequence analysis and the importance of studying known candidate genes in extended cohorts in ALS 66

1.19 Aims of study 69
1.19.1 Characterizing known genes in the IC-FALS cohort: Sequence analysis and Southern's Blot 69
1.19.2 Association studies 70
1.19.3 Characterization of VNTR length in candidate genes for ALS 70

CHAPTER 2 METHODOLOGY AND MATERIALS 72

2.1 Subjects 72

2.2 DNA Extraction 74
2.2.1 DNA extraction from whole blood or the buffy coat layer 74
2.2.2 DNA purification from agarose gels 74
2.2.3 Total RNA extraction from blood 75

2.3 Polymerase chained reaction (PCR) 75

2.4 cDNA Synthesis and Endpoint RT-PCR 77

2.5 DNA Sequencing 77
2.5.1 DNA Purification 77
2.5.2 DNA Sequencing 78

2.6 Restriction digest 78

2.7 Electrophoresis 79
2.7.1 Agarose gel electrophoresis 79
2.7.2 Denaturing Polyacrylamide Gel Electrophoresis (PAGE) 80

2.8 SNP Genotyping using Kompetitive Allele Specific PCR (KASP^{TM}) assay 84

2.9 Fragment analysis 84

2.10 Southern Blot 84
2.10.1 Sample preparation and electrophoresis 85
2.10.2 Preparation for DNA transfer 86
2.10.3 Assembling the blot 86
2.10.4 Dissembling the blot and preparation for hybridization 88
2.10.5 Hybridization using Digoxigenin (DIG) labelled probes 88

2.10.6 Chemiluminesence detection 90
2.10.7 Dot Blot 91
2.11 Statistics 92
2.12 Stock solutions 94

CHAPTER 3 SEQUENCE ANALYSIS OF *VCP* GENE IN THE IC ALS COHORT 96

3.1 Introduction 96
3.2 Method 97
3.2.1 Sample collection and DNA extraction 97
3.2.2 DNA Sequencing 97
3.2.3 Gel electrophoresis 98
3.2.4 Total RNA extraction and RT-PCR 98
3.2.5 Statistics 98
3.3 Results 100
3.3.1 Sequence analysis of the *VCP* gene 100
3.3.2 Screening for the hexanucleotide expansion in additional ALS cohorts 107
3.3.3 SNPs identified in the *VCP* gene 107
3.4 Discussion 111

CHAPTER 4 SEQUENCE ANALYSIS OF *SQSTM1* GENE IN THE IC ALS COHORT 116

4.1 Introduction 116
4.2 Methods 117
4.2.1 Sample collection 117
4.2.2 Genotyping 117
4.2.3 Data analysis 118
4.3 Results 120
4.3.1 Identification of *SQSTM1* sequence variants in a UK-FALS cohort 120
4.3.2 Founder haplotype of P392L *SQSTM1* kindred 130
4.4 Discussion 134

CHAPTER 5 ASSOCIATION STUDIES OF CANDIDATE GENES IN ALS 136

5.1 General introduction: Association study and its use for identifying susceptibility genetic variants in ALS 136
5.2 Association studies indicate that *P4HB* is a risk factor in Amyotrophic lateral sclerosis 140
5.2.1 Introduction 140
5.2.2 Methods for *P4HB* association study 142
5.2.2.1 Sample collection 142
5.2.2.2 Identification and genotyping of SNPs 142
5.2.2.3 DNA Sequencing 143
5.2.2.4 Statistics 143
5.2.3 Results 144
5.2.3.1 Genotype Analysis of P4HB yielded Significant Genotypic Associations with ALS 147

5.2.3.2 Rare Haplotypes are associated with FALS and SALS 149
5.2.3.3 Genotypes and Haplotypes of P4HB affect the Survival of ALS 155
5.2.4 Discussion: *P4HB* 157

5.3 Association study of the *NPLOC4* Gene: 161
5.3.1 Introduction 161
5.3.2 Methods for *NPLOC4* association study 162
5.3.2.1 Subjects 162
5.3.2.2 Fragment analysis 164
5.3.2.3 Statistics 164
5.3.3 Results 165
5.3.3.1 GWAS SNPs tagged for independent haplotype blocks in the NPLOC4 gene 165
5.3.3.2 SNPs in NPLOC4 gene are significantly associated with Sporadic ALS 166
5.3.3.3 Linkage disequilibrium and Haplotype Analysis 170
5.3.3.4 Frequencies of [TG]n repeats 172
5.3.3.5 Gender and age of onset effects of NPLOC4 SNPs *173*
5.3.4 Discussion: *NPLOC4* 174

CHAPTER 6 VNTR ANALYSIS IN ALS **178**

6.1 Introduction 178

6.2 Backrgounds for candidate genes 180
6.2.1 *SSBP3* 180
6.2.2 *HSPB8* 180
6.2.3 *EIF2AK2* 181
6.2.4 *YWHAQ* 181
6.2.5 *GPX1* 182
6.2.6 *SLC12A2* 183
6.2.7 *RBM23* 183
6.2.8 *NIPA1* 183
6.2.9 *CAPNS1* 183
6.2.10 *UBQLN3* 184
6.2.11 *TMEM158* 184
6.2.12 *IRX2, IRX3* and *IRX4* 185
6.2.13 *CTNND2* 185
6.2.14 *ID4* 185
6.2.15 *TXNDC5* 186
6.2.16 *RXRB* 186
6.2.17 *FAM120C* 187

6.3 Methodology 187
6.3.1 Subjects 187
6.3.2 Genotyping and DNA sequencing 188
6.3.3 Data analysis and Statistics. 190

6.4 Results 191
6.4.1 Identification of candidate Variable number tandem repeats for ALS 191
6.4.2 Identification of polymorphic VNTRs 195

6.4.3 VNTR association with ALS: allelic and genotypic associations and loss of heterozygosity (LOH) tests 198
6.4.4 Genotypic correlation with phenotypes and expression 204
6.5 Discussions 204

CHAPTER 7 CHARACTERIZATION OF *C9orf72* GGGGCC EXPANSION 208

7.1 Introduction 208
7.2 Methodology 209
7.2.1 Subjects and sample preparation 209
7.2.2 Southern Blot analysis 210
7.2.3 Statistics 210
7.3 Results 211
7.3.1 Optimization of Southern blot protocol using different probes. 211
7.3.2 Characterisation of *C9orf72* expansion size using Digoxigenin-based Southern Blots. 212
7.3.3 Correlation between *C9orf72* expansion size and clinical phenotypes. 212
7.4 Discussion 217

CHAPTER 8 DISCUSSION 220

8.1 Overview: the current and changing view of ALS genetics and proteostasis as a pathogenic mechanism 220
8.2 Sequence analysis 223
8.3 Association studies 225
8.4 VNTR analysis 227
8.5 Characterization of *C9orf72* expansion 228
8.6 Future work 230

PUBLICATIONS LED BY THIS STUDY 233

REFERENCES 234

APPENDIX 276

Appendix I: R Codes used for a quick assessment of genotypes/allelic counts and association tests 276

Appendix II: Linux codes used for retrieving raw sequencing data from the 1000 genome project and calling for indels using SAMTools, Dindel and BEDTools programs 283

Appendix III: R Codes used for re-locating a list of VNTR (Kozlowski 2010) to the reference genome (hg19) using BSgenome package and the cross-referencing with GWAS data 286

LIST OF FIGURES

Figure 1-1. *SOD1* functions, structures and mutations found in ALS 36

Figure 1-2. Signal transduction in Unfolded protein response (UPR) 41

Figure 2-1. Summary of subjects in this study 73

Figure 2-2. Setup of Polyacrylamide gel electrophoresis (PAGE) 83

Figure 2-3. Setup of Southern transfer 87

Figure 3-1. Published coding mutations in *VCP* in ALS and IBMPFD 104

Figure 3-2. The c.-360G>C Pedigree 106

Figure 3-3. Electrohporesis for the *VCP* tandem repeats 108

Figure 4-1. *SQSTM1* sequence variants in ALS and PDB 121

Figure 4-2. Pedigrees of the FALS index cases carrying P392L and E155K *SQSTM1* mutations 125

Figure 4-3. Mutation frequency of P392L *SQSTM1* in different populations 126

Figure 4-4. Linkage disequilibrium and Survival analysis of *SQSTM1* SNPs 132

Figure 5-1. Details of *P4HB* SNPs investigated in this study 146

Figure 5-2. SNP LD, haplotype and diplotype associations 151

Figure 5-3. Effects of *P4HB* genotype on site at onset of disease 156

Figure 5-4. Kaplan-Meyer curves showing the survival times of FALS patients with different *P4HB* genotypes 156

Figure 5-5. SNPs, haplotype and diplotypes in the *NPLOC4* gene 167

Figure 5-6. Genotype frequency of TG repeat and Survival analysis of the *NPLOC4* SNPs 172

Figure 6-1. SNPs that flank ± 100kb of known tri-nucleotide repeats throughout the genome 193

Figure 6-2. A summary of VNTRs in the current study 194

Figure 6-3. Allele frequencies of VNTR with ≥3 alleles in this study 200

Figure 6-4. Grouped Allelic and genotype frequencies of VNTRs with ≥ 3 alleles 203

Figure 7-1. Representative Southern blot results and the sizing of expanded alleles 213

Figure 7-2. *C9orf72* Southern blot results using Buffy coat DNA 214

Figure 7-3. *C9orf72* Southern blot results using LBC DNA 215

Figure 7-4. Correlation and regression analyzes of clinical phenotypes against *C9orf72* expansion sizes 216

Figure 8-1. ALS mutations involved in Proteostasis pathways 222

LIST OF TABLES

Table 1-1. The revised El- Escorial Criteria 24

Table 1-2. Mendelian genes in ALS 30

Table 2-1. Preparation of PCR reactions 76

Table 2-2. Concentration of agarose gels used for the separation of DNA fragments of different sizes as summarized by (Barril, 2012) 80

Table 2-3. Formulations for commonly used gel percentages and the amount of solutions used for each 50ml gels (adopted from national diagnostics) 81

Table 2-4. Preparation of PCR reactions for the synthesis of DIG-labelled Probes 89

Table 2-5. A combined power heatmap with table of effect size w-Values, for different Case/Control MAFs at a=0.05, and a total sample size of 270 subjects 92

Table 2-6. Penetrance functions used for determining model of associations (Modified from Clarke et al 2011) 93

Table 2-7. Preparation of stock solutions used in this study 94

Table 3-1. Primers used for Polymerase Chained Reactions for the VCP Study 99

Table 3-2. Frequencies of *VCP* mutations 102

Table 3-3. Frequencies of hexanucleotide expansions in cases and controls 109

Table 3-4. Prevalence of *VCP* SNPs detected in this study 110

Table 3-5. Summary of phenotypes exhibited for known *VCP* mutations 112

Table 4-1. Primers used for the Sequencing of *SQSTM1* Gene 119

Table 4-2. Summary of coding variants in *SQSTM1* in FALS, SALS and FTD from different studies 127

Table 4-3. Combined allelic frequencies of *SQSTM1* variants that are predicated as pathogenic in cases and controls 129

Table 4-4. Founder haplotypes for *SQSTM1* mutations 131

Table 4-5. Common SNPs (MAF > 10%) detected in this study 133

Table 5-1. Primers used for PCR of *P4HB* SNPs 143

Table 5-2. Description of ALS and Control cohorts used in the *P4HB* study 145

Table 5-3. Power calculation for the *P4HB* association study 145

Table 5-4. Genotypic and allelic associations of *P4HB* SNPs 1 and 4 147

Table 5-5. Genotypic and allelic associations of *P4HB* SNPs 148

Table 5-6. Phased haplotypes for *P4HB* in ALS ranked according to their frequencies in controls 152

Table 5-7. Diplotype frequencies of *P4HB* SNPs 154

Table 5-8. Clinical details available for phenotype analysis in this study 155

Table 5-9. Cox-regression models used to analyse the combined effects of *C9orf72* and *P4HB* on the survival of FALS 157

Table 5-10. Power calculations 163

Table 5-11. Primers used for polymerase chained reaction of the *NPLOC4* SNPs 164

Table 5-12. SNP associations of *NPLOC4* SNPs 168

Table 5-13. Non-associated *NPLOC4* SNPs (SNPs 4, 5 and SNP 1, 2 for FALS) 169

Table 5-14. *NPLOC4* Haplotypic associations 171

Table 5-15. Mantel-Haenszel analysis with respect to gender 173

Table 5-16. Haplotype association in Female SALS 174

Table 5-17. *NPLOC4* SNPs tagged by those analysed in this study 177

Table 6-1. Summary of the cases and controls employed in this study 188

Table 6-2. Primers and PCR Conditions used in this study 189

Table 6-3. Criteria used for the selection of candidate genes 192

Table 6-4. Non-polymorphic VNTR repeats in FALS and IC-Controls 195

Table 6-5. VNTR Allele frequencies in FALS and Controls 196

Table 6-6. VNTR Genotype frequencies 198

Table 6-7. Frequencies of *HSPB8* repeat length (*rs112223147*) in SALS 202

ABBREVIATIONS

^{o}C	Degrees Centigrade
+ve	Positive
-ve	Negative
µg	Microgram
µl	Microliter
µM	Micromolar
ALS	Amyotrophic lateral sclerosis
AOO	Age of onset
ATF6	bZiP-containing activating transcription factor 6
BH	Benjamini-Hochberg
bp	Base pair
C9orf72	Chromosome 9 open reading frame 72
cDNA	Complementary DNA
Chr	Chromosome
cM	Centimorgan
ddNTP	Dideoxynucleotide
DNA	Doxyribonucleic acid
dNTP	Deoxynucleotide
EA	European American
EAAT2	Excitatory amino acid transporter 2
EDTA	Ethylene diamine tetra acetate
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation
Ero1	ER oxidoreductin 1
EUR	European
FALS	Familial Amyotrophic lateral sclerosis
FDR	False discovery rate
FOR	Forward
FTD	Frontotemporal dementia
FTLD	Frontotemporal lobar degeneration
FVC	Forced vital capacity

g	Gram
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
GWAS	Genome wide association study
HCI	Hyaline conglomerate inclusion
HE Staining	Hematoxylin and eosin staining
HSP	Heat shock protein
HSPB8	Heat shock protein beta-8
IBMPFD	Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia
IC	Imperial College
IF	Intermediate filaments
IRE1	Transmembrane receptors Inositol requiring kinase 1
LBHI	Lewy body like Hyaline inclusion
LC-3	Autophagic marker light chain 3 (LC3)
LD	Linkage disequilibrium
LMN	Lower motor neuron
LOD	Logarithm of odds
MAF	Minor allele frequency
Mb	Megabase
mRNA	Messenger RNA
NF-kB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NGS	Next generation sequencing
NHLBI ESP EVS	NIH Heart, Lung and Blood Institute GO Exome Sequencing Project Exome Variant Server
NIPA1	Non-imprinted in Prader-Willi/Angelman syndrome region protein 1
NMDA	N-methyl-D aspartate
NPLOC4/Npl4	Nuclear protein localization protein 4 homolog
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chained reaction
PDB	Paget's disease of bone
PDI	Protein disulphide isomerase
PERK	Protein kinase like ER kinase

PTq	Taq Polymerase
qrtPCR	Quantitative reverse transcription Polymerase chained reaction
REV	Reverse
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
rtPCR	Reverse transcription Polymerase chained reaction
SALS	Sporadic Amyotrophic lateral sclerosis
SMA	Spinal muscular atrophy
SNP	Single Nucleotide polymorphism
SOD1	Superoxide dismutase 1
SQSTM1/p62	Sequestosome 1
SSC	Saline sodium citrate
SSCP	Single strand conformation polymorphism
TARDBP	Transactive response DNA binding protein 43 kDa
TBE	Tris borate EDTA
Ub	Ubiquitin
UBI	Ubiquitinated inclusion
Ufd1	Ubiquitin Fusion Protein 1
UMN	Upper motor neuron
UPR	Unfolded protein response
UPS	Ubiquitin proteasome system
UTR	Untranslated region
UV	Ultraviolet light
VAPB	Vesicle-associated membrane protein-associated protein B
VCP/p97	Valosin containing protein/
VNTR	Variable number tandem repeat
WT	Wild type
YWHAQ	Tyrosine 3-monooxygenase-tryptophan 5-monooxygenase activation protein.

Chapter 1

Introduction to ALS

Amyotrophic lateral sclerosis (ALS) is a lethal neurodegenerative disease caused by progressive loss of motor neurons in the spinal cord, brain stem and cerebral cortex. This results in characteristic phenotypes comprising both upper motor neuron (UMN) and lower motor neuron (LMN) symptoms. UMN symptoms include spasticity and pathological reflexes, whereas LMN symptoms include muscle atrophy and paralysis. Sensory functions are preserved in ALS. Treatments for the disease are currently limited and death usually occurs within 3-5 years as a result of respiratory failure.

The disease has incidence and prevalence of 1.89/100,000 per year and 5.2/ 100,000 respectively in western countries, with a Male to Female ratio of 1.5:1 (Wijesekera 2009). The distribution of these parameters seems to be uniform across different countries (Worms 2001), except for several endemic areas in the West pacific region, such as Kii Peninsula, Islands of Guam and West New Guinea. Increased incidence has also been reported in US veterans and Italian soccer players (Mitsumoto 2010). The clustering has been attributed to environmental factors.

About 5-10% of ALS cases are familial (FALS) whilst the remaining are Sporadic (SALS). In general, there is no clinical distinction between familial and sporadic cases, and the current views of disease mechanisms are largely inferred from the discoveries of causal genes in ALS families. In recent years, there has been substantial progress in identifying ALS mutations, allowing a classification for the disease. This also initiated our attempt to allocate genetic causes for each kindred in our FALS cohort. In the meantime, although risk factors, such as gender, smoking, and heavy metal exposure, have been recognized, there is a need for further investigation of genetic risk factors or modifiers for ALS. In this study, we began with the screening and characterization of a group of known genes in our cohort, which was followed by association studies using single nucleotide polymorphisms (SNP) and screening for VNTR in novel candidate genes. The genes were selected using

a candidate-gene-approach, which emphasis on functional evidence obtained from earlier research.

In this thesis, I have started by describing the background to ALS from a genetic point of view, in which each pathway is illustrated with the corresponding candidate genes. General experimental protocols and statistics are described in the Methodology (**Chapter 2**). Each of the following Chapters contain the specific background and methodologies used for the genes being analyzed, followed by separate discussions. The overall impact of this study is discussed in the final Chapter.

1.1 Clinical features of ALS: Symptoms begin in various sites and progress throughout the body

The natural history of ALS begins with a pre-clinical phase, which lacks any detectable clinical symptoms. In the presence of predisposing factors, subtle pathological changes such as size of motor neurons may occur at this stage (Brooks, 1996). In the Clinical phase, neurophysiological changes, such as abnormalities in single-fiber density and motor unit count, may precede the onset of symptoms, which typically occurs in patients in their fifties (Haverkamp et al., 1995). After onset, different groups of muscles may be affected depending on the sites of lesions. Asymmetric muscle weakness and atrophy are common lower motor neuron symptoms presenting in the early phase of disease. Upper motor neuron symptoms, such as muscle stiffness and fasciculation, are due to corticospinal involvement and hyperactivity of tendon reflexes may be elicited in physical examination. In 75% of patients these symptoms begin at their limbs, i.e. limb onset, whereas in 25% symptoms start in the bulbar region, resulting in bulbar palsy. Due to the damage of motor components in cranial nerves IX, X and XII, patients with bulbar onset may have difficulties in speaking, swallowing and loss of tongue mobility. Pseudobulbar symptoms including dysarthria and exaggeration of facial expression may be seen as a result of the loss of corticobulbar innervation (Brown, 2001).

As disease progresses, focal symptoms may spread to other parts of body based on a topographical pattern, resulting in an asymmetric distribution of weakened muscle groups. It has been proposed that initial symptoms may be caused by discrete UMN and LMN

lesions which are responsible for the same region, and the lesions subsequently spread independently during disease progression, leading to a more complex phenotype (Ravits and La Spada, 2009). In general, the progression of symptoms, which are correlated with neuronal loss, is fastest among anatomical contiguous areas and it favours a rostral-caudal direction. For example, in arm-onset patients, LMN symptoms first spread to contralateral arm and then ipsilateral foot, whereas UMN symptoms spread to ipsilateral foot prior to contralateral arm (Ravits et al., 2007). Arm symptoms occur sooner in bulbar-onset patients than the occurrence of bulbar symptoms in arm-onset patients (Brooks, 1991). However, it has also been reported that caudal-rostral spread occurs faster than rostral-caudal spread within the spinal column (Brooks, 1996), which is supported by the observation that there is a higher percentage of LMN loss in cervical than lumbar anterior horn in patients with trunk onset (Ravits et al., 2007). The extent of muscle loss can be measured by isometric muscle strength, which deteriorates as a function of time. The maximum rate of loss is seen in the early phase of disease and distal muscles are more severely affected. Respiratory function is compromised and deteriorates with disease progression. Forced vital capacity (FVC), a measurement of pulmonary function, has been reported as prognostic factor for ALS, in which patient with baseline FVC<75% have shortened survival time (Czaplinski et al., 2006). Complications such as inability to feed, aspiration and pneumonia may occur in late stage of disease, and death usually occurs due to respiratory failure within 3-5 years of onset. Nevertheless, different rates of progression are observed in FALS patients caused by different mutations (**Section 1.5**). Rapid progression is observed in patients with bulbar onset, whereas slow progression has been reported in SALS patients <40 years old.

Intellectual reasoning, vision, hearing and sense function are not affected in ALS and sexual, bowel and bladder functions are mostly preserved.

1.2 The diagnosis of ALS is based on clinical and electrophysiological findings

In 1998, the World Federation of Neurology proposed a revised version of ALS diagnosis criteria, known as revised El Escorial criteria (R-EEC) (Brooks et al., 2000). The criteria, which stratify the confidence of diagnosis into four categories, are mainly based on clinical grounds and have been the gold standard for diagnosis (**Table 1-1**). For example, a

definite diagnosis requires the presence of upper and lower motor neuron symptoms in bulbar and two different spinal regions or greater than three spinal regions. Four anatomical regions are appointed for examination. Bulbar region refers to muscles of the jaw, face, palate, tongue and laryngeal muscles, whereas spinal region is categorized into cervical, thoracic and lumbosacral regions. In ALS, loss of nerve innervation results in membrane instability and reinnervation and these can be demonstrated in electrophysiological examinations. EMG findings compatible with active and chronic denervation, such as positive sharp wave and fibrillation potentials, are required to confirm LMN degeneration.

On the other hand, as an effort to increase the sensitivity and facilitate early diagnosis, the Awaji criteria (AC) recommend that neurophysiological evidence should be taken as equivalent to clinical information when confirming regional involvements. The criteria also appreciate the sufficiency of fasciculation potentials, a characteristic electrophysiological feature of ALS, in the evaluation of active degeneration, allowing the feasibility of diagnosis without the presence of fibrillation potentials and positive sharp waves (de Carvalho et al., 2008). It has been reported that the time interval between onset and diagnosis using AC is 6.2 month earlier than that using the R-EEC (Okita et al., 2011).

The diagnosis of ALS must be also accompanied by pathological, neuroimaging and other laboratory investigations to exclude other conditions that mimic ALS. These include (1) other motor neuron diseases of restricted involvement such as progressive muscular atrophy (PMA, lower motor neurons only), primary lateral sclerosis (PLS, upper motor neurons only) and progressive bulbar palsy (PBP, bulbar symptoms only) (Belsh, 1999; Norris et al., 1993). These disorders, however, may eventually progress to ALS when LMN or UMN symptoms subsequently develop and they are often considered as the same category of disease. Also in this category, flail arm syndrome (Hu et al., 1998) and flail leg syndrome are two distinctive phenotypes in which weakness and wasting are confined to upper and lower limbs respectively with improved survival compared to classical and bulbar onset ALS (Wijesekera et al., 2009). (2) ALS-like symptoms of other causes such as benign fasciculation, Parkinson's disease, Kennedy's disease, brainstem stroke, lumbosacral stenosis, cervical myelopathy, carpal tunnel syndrome, brachial plexopathy,

neuropathy, adult-onset spinal muscular atrophy (SMA) and heavy metal intoxication should be excluded before diagnosis (Belsh, 1999; Mitsumoto, 1997).

Noteworthy, the term ALS variant describes the situation where certain clinical, genetic or epidemiological features develop in parallel with ALS. Familial ALS is most commonly classified based on the mutations that segregate within affected families or locus linked to the phenotype. Although phenotypic heterogeneity has been well recognized in patients with different point mutations in *SOD1* gene, the most common causal gene caused by point mutations, patients with *SOD1* mutations are collectively referred as ALS-1. At least 18 FALS subtypes have been reported according to the genetic classification to date (**Section 1.5**). Types of ALS can also be distinguished by the presence of other coexisting conditions that extend beyond the pyramidal tract, such as Parkinsonism and Frontotemporal dementia (FTD) (**Section 1.16**), or geographic clustering (Hudson, 1981). There is a high prevalence of Parkinsonism-Dementia variant of ALS (ALS-PDC) in the indigenous Chamorro population from the island of Guam, where clusters of ALS cases have been reported (Sundar et al., 2007).

Muscle strength and Forced vital capacity (FVC) are two important factors predicting the survival in ALS (Voustianiouk et al., 2008). The Revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) is a questionnaire- based scoring system for monitoring the course of ALS (Cedarbaum et al., 1999), whereas the Appel ALS Score (AALS) is an examination based counterpart. Both scales correlate with disease progression and survival and ALSFRS-R, which involves assessment of gross motor tasks, fine motor tasks, bulbar functions and respiratory functions, is also correlated with quality of life and FVC.

Table 1-1. The revised El- Escorial Criteria.

Requirement of Diagnosis
Presence of 1 Evidence of LMN Degeneration by Clinical, electrophysiological or neuropathology exam. 2 Evidence of UMN degeneration by clinical exam 3 Progressive spread of signs within a region or to other regions, with the Absence of a. Electrophysiological evidence of other disease process that might explain the signs of LMN or UMN degenerations, and b. Neuroimaging evidence of other disease process that might explain the observed clinical and electrophysiological signs.
Categories of Diagnosis
1 Clinically Definite ALS: UMN+ LMN signs together in bulbar and ≥ 2 spinal regions; or UMN+ LMN signs together in ≥ 3 spinal regions. 2 Clinically Probable ALS: UMN+ LMN signs together in ≥ 2 regions with UMN rostral to LMN signs. 3 Clinically Probable ALS- Laboratory supported: UMN+ LMN signs together in 1 region; or UMN in 1 region + LMN defined by EMG in ≥ 2 regions. 4 Clinically Possible ALS: UMN+ LMN signs together in 1 region, or UMN sign alone in ≥ 2 regions, or LMN signs are rostral to UMN signs.
Electrophysiological Features
Electrophysiological studies should be performed to: 1 Confirm LMN dysfunction in clinically affected regions. 2 Detect electrophysiological evidence of LMN dysfunction in clinically uninvolved regions 3 Exclude other pathophysiological processes. Criteria: 1 The diagnosis requires the combination of features of: a. Active denervation: Fibrillation potentials (FP), Positive sharp waves (Psw). b. Chronic denervation: Large MUAPs, reduced interference pattern with firing rates higher than 10Hz, unstable motor unit potentials. 2 Fasciculation potentials are characteristic clinical feature of ALS. 3 These signs should be found in at least 2 regions of: brain stem, thoracic spinal cord, cervical and lumbosacral spinal cords. 4 Nerve conduction studies are required for diagnosis and to exclude other disorders.

Other Supporting Features

1. Absence of neuroimaging features supporting other diagnosis.
2. Laboratory features excluding other conditions or those found in ALS-related conditions:
    - Monoclonal gammopathy: elevation in monoclonal anti-neural antigen antibody.
    - Dysimmune motor system degeneration: elevation in polyclonal anti-neural antigen antibody
    - Nonmalignant endocrine abnormalities: elevation in parathyroid hormone, thyroid hormone and other endocrine abnormalities
    - Lymphoma: abnormalities consistent with lymphoma
    - Infection: HIV-1, HTLV-1, encephalitis lethargica, VZV etc.
    - Acquired enzyme defects: detoxification enzymes etc.
    - Exogenous toxins: evidence of intoxication
    - Physical injury: antecedent electrical or radiation injury or severe trauma.
    - Vasculitis: elevated ESR and CSF abnormalities consistent with spinal vasculitis or ischemic injury.
    - Spondylotic myelopathy: spinal compression.
3. Pathological evidences from muscle biopsy: disseminated single angulated muscle fibers.
4. Postmortem spinal cord biopsies.

The above Criteria are summarized according to Brooks et al (2000).

1.3 Current treatments for ALS are limited

Treatments for ALS are currently limited. The only available prescription for ALS, Riluzole, a presynaptic modulator of glutamate transmission, was shown to protect neurons against excitotoxicity and prolong survival time of patients by 3- 5 months (Wokke, 1996). The β-lactam antibiotic ceftriaxone is another potential anti-glutamate drug that decreases synaptic glutamate level by facilitating astrocyte glutamate uptake through the EAAT2 transporter and showed neuroprotective effects (Rothstein et al., 2005). Compounds or substances with beneficial effects in Protein folding (Arimoclomol), Autophagy (Lithium), Muscle function (Myostatin Inhibitor), Neurotrophins (VEGF, IGF-1) and Stem cell therapy are potential alternatives for ALS [Reviewed by (Zinman and Cudkowicz, 2011)].

1.4 Pathological findings in ALS

The pathological features of ALS have been well characterized. There is loss of motor neurons in both spinal cord (Lower motor neurons in the ventral horn) and the cerebral cortex (Betz's cells), gliosis, and the presence of Neuronal cytoplasmic inclusions (NCIs) in lower motor neurons (Brownell et al., 1970). Different types of inclusions exist, including Bunina bodies, Ubiquitinated inclusions (UBIs) and Hyaline conglomerate inclusions (HCIs) (Wijesekera and Leigh, 2009).

Bunina bodies, which can be visualized in HE staining, are small, eosinophilic and round inclusions seen in the cytoplasm and dendrites (Okamoto et al., 2008). Generally immuno-negative for ubiquitin, Bunina bodies are found in 70% of SALS and are considered as a marker for the disease. Cystatin C and Transferrin have been reported as being components of Bunina bodies. However, the most common inclusions, which were found in 90-100% of SALS cases, are the UBIs. UBIs are immune-positive for ubiquitin and can be morphologically classified into Skeine-like (filamentous) inclusions and Lewy-body like (Round shape) inclusions. TDP-43 is the major component of UBIs, which are often also positive for p62 and are found in SALS and FALS cases that are unrelated to *SOD1* mutations (Neumann et al., 2006; Tan et al., 2007). On the other hand, HCIs are inclusions with glassy appearance under HE staining. These inclusions are mainly composed of accumulated intermediate filament (IF) proteins, i.e. hyperphosphorylated IF

subunits and peripherins, and have been found in other neurodegenerative conditions. They are therefore not a specific marker for ALS (Wood et al., 2003).

FALS associated with *SOD1* mutations encompass a different histo-pathologic entity. In contrast to the classic form of ALS, in which degeneration is restricted to the motor neurons, *SOD1* FALS generally demonstrate multisystem degeneration, in which regions such as posterior thoracic nucleus and spinocerebellar tract are also involved (Tagawa et al., 2007). In addition, *SOD1* FALS is pathologically characterised by the presence of Lewy body like Hyaline inclusions (LBHIs). These inclusions are positive for ubiquitin and SOD1, but not TDP-43 (Mackenzie et al., 2007). Not surprisingly, these *SOD1* positive inclusions are rare in SALS and *SOD1* unrelated FALS, and have been considered as the distinction between *SOD1*- positive and negative cases. Notably, the pathological features of *SOD1*- negative FALS are indistinguishable from SALS, suggesting a common pathological mechanism of FALS and SALS (Tagawa et al., 2007; Tan et al., 2007).

1.5 Genetic studies have led to the understanding of underlying pathogenesis of ALS

Genetic studies are important for understanding ALS as they ① have direct impacts on clinical diagnosis and subclassification of disease; ② may reveal general pathogenic mechanisms, risk factors, and possibly therapeutic targets; ③ lead to the development of models that can be used for further investigation; and ④ understanding of gene and protein functions (Hardy and Orr, 2006).

As mentioned above, patterns of inheritance are found in 5-10% of ALS patients (FALS), and the clinical appearances FALS are similar to the majority of Sporadic cases (SALS). It is therefore important to investigate the genetic causes of FALS, which provide invaluable information about the general mechanisms (Pasinelli and Brown, 2006). Localization of disease loci was made possible by linkage studies. In brief, because the possibility of recombination between two loci decreases with the proximity between them, adjacent genes are more likely to segregate together during meiosis. Using suitable pedigrees, the linkage distance, measured by a LOD score, which is the logarithm of the odds of the

likelihood that two loci are linked against the likelihood that the loci are unlinked, between markers and the disease locus can be calculated.

By performing linkage studies in 23 multi-generation autosomal dominant FALS families, Siddiuqe et al (1991) linked the disease locus to markers representing an approximately 10cM region (~10 Mb) in chromosome 21. Dinucleotide repeats linked to the *SOD1* gene were subsequently shown to exhibit a high linkage signal with these markers (Rosen et al., 1992). Sequencing of the gene revealed 11 *SOD1* mutations in 13 FALS families (Rosen et al., 1993). *SOD1* mutations were subsequently found to account for 20% of FALS cases and more than 160 different mutations throughout gene have been reported. Most of the common mutations, such as A4V, I113T and L144F, caused autosomal dominant inheritance but recessive forms, such as D90A (Al-Chalabi et al., 1998) and N86S, were occasionally seen (**Figure 1-1D**). The mean survival time of SOD1-FALS is 4.6 years (Orrell et al., 1999) and clinical features may vary. Rapid progression is seen in A4V, L84V mutation carriers whereas slow progression has been reported in H46R carriers. Incomplete penetrance was obvious for some mutations which were found in obligate carriers and SALS (de Belleroche et al., 1995). The discovery of *SOD1* mutations was considered as a milestone and the hypotheses proposed to explain the effect *SOD1* mutations also recapitulated the general features of ALS, including *elevated oxidative stress*, *protein aggregation*, *mitochondrial dysfunction*, *cytoskeleton abnormality* and *excitotoxicity* (Bendotti and Carrì, 2004).

In addition to *SOD1*, mutations in more than 18 genes have been identified in different FALS pedigrees using linkage studies (Until DEC 2013) (**Table 1-2**). Currently, it seems to be unfeasible to predict clinical phenotypes based on the functions of candidate genes alone, as variants at different position may cause different consequences on protein structure and activities. However, there are instances where characteristic phenotypes are associated with mutations in certain genes. Juvenile onset cases have been reported in carriers of Alsin (*ALS2*, **Section 1.17.1**), Senataxin (*SETX*, **Section 1.17.3**), Spatacsin (*SPG11*, **Section, 1.17.4**) and Sigma-1 Receptor (*SIGMAR1*, **Section 1.17.11**) mutations, whereas VAMP associated membrane protein B (*VAPB*, **Section 1.8**) and Dynactin (*DCTN*, **Section 1.10**) mutations may cause adult-onset, slow progressive ALS. These known causal genes provide important clues about disease mechanisms and can be

generally categorized into *four* groups according to their functions: ① *Redox reactions,* ② *Membrane trafficking,* ③ *mRNA processing* and ④ *Protein synthesis* and *degradation*. Noteworthy, the fact that mutations in several recently found genes, such as *TARDBP*, *VCP* and *C9orf72*, are present in both FALS and Familial FTD supports the hypothesis that these conditions are extremities of the same spectrum of disorders caused by common mechanisms (**Section 1.16**).

The genetics of the majority of cases, SALS, is still unclear. Although mutations in FALS-genes have been identified in SALS cases, they were not verified by linkage signals or segregation analysis. SALS is often considered as a complex disease and genetic risk factors are located through association studies. Using a *candidate gene approach*, which determinates genetic risk on a trait based on *a prior* hypothesis that the genes might play roles in the aetiology of disease (Tabor, 2002), association signals defined by Single nucleotide polymorphisms (SNPs), Variable number tandem repeats (VNTRs) or haplotypes have been reported in *APEX1*, *ATXN2*, *HFE*, *NEFH*, *SMN1*, *SMN2*, *PON1*, *PON2*, *PON3* and *VEGF* genes. This approach has also applied to the identification of genetic modifiers in familial cases. Genome-wide association studies (GWAS), in contrast, analyse hundreds of thousands of SNPs throughout the genome regardless of their functional properties. GWA studies have led to the identification of a number of risk factors in ALS which will be further discussed in **Chapter 5**. The association signals represent the susceptibility of disease conferred by the corresponding genes, which may have a partial contribution to the pathological mechanisms. In some cases, some of these association signals have led to the discovery of confirmed causal mutations.

Table 1-2. Mendelian genes in ALS.

Subtypes	Genes	Inheritance	Functions	Pathology	Remarks
ALS					
ALS1	*SOD1* (21q)[6]	AD	-Reduce oxidative stress by facilitating disulphide bond exchange.	-HCI which contains SOD1 aggregations [2, 3].	-Most common mutation. ~20% of FALS. -Generally Typical ALS. Some mutations are associated with rapid progression or bulbar onset.
ALS2	*ALSIN* (2q33)	AR	-GTPase regulator[1] ; -Related to Vesicle transport & Trafficking.	-Mild reduction of axons.	-Juvenile ALS (Onset <10 years old). Slow progression.
ALS3	18q21 (18q21)	AD	-Unknown.	-Unknown.	-French family (20 affected). -Typical ALS, Leg Onset at 45, duration 5 years[4].
ALS4	*SETX* (9q34, Senataxin)	AD	-RNA processing.	-Unknown	-Juvenile ALS (Onset <25 years), initially affect distal Limbs, slow progression[5].
ALS5	*SPG11* (15q21, Spatacsin)	AR	-Gene expression -Protein trafficking.	-Betz cell loss. -Neuronal loss and strocytosis in IX and XII nuclei. Anterior horn atrophy. -No bunina bodies/ skein like inclusions.	-Juvenile ALS (mean onset= 16 years), long term survival. -Variety of symptoms, principally distal weakness, *Hoffmann always +ve*. No sensory or cognitive impairment [7].

Subtypes	Genes	Inheritance	Functions	Pathology	Remarks
ALS6	*FUS* (16p11)	AD/ AR	-TDP-43 homolog. -Regulation of transcription, -RNA Splicing.	-Severe LMN loss in spinal cord. -FUS protein is mis- localized and form inclusions[8]. -No skein like inclu- sions were observed.	-~4% FALS. Typical ALS. Slightly rariler onset (44.5), normal progression (33m). No cognitive impairments[8]. -Also found in two Juvenile sporadic cases, one with mental retardation. -FTD is also described in 3 families[10].
ALS7	? (20p13) [9]	AD	-Unknown	-Unknown	-Onset at 57 years old. Survival 3 years.
ALS8	*VAPB* (20q13)	AD	-Formation of presynaptic vesicles. -Vesicle trafficking. -Unfolded protein response.	-Transgenic mices expressing mutant VAPB have +ve TDP-43 aggregates[13].	-Brazilian Family. Onset between 25 to 55 years old [11]. Also in one English FALS case [12]. -Phenotypic heterogeneity: late onset SMA, atypical ALS, typical ALS.
ALS9	*ANG* (14q11, Angiogenin)	AD/ Sporadic	-Ribosomes synthesis and protein translation.	-Unknown	-Onset under 50 years old. Survival from 6m to 5years. Typical ALS with incomplete penetrance. One patient demonstrated Parkinsonism and FTD symptoms [14,15].
ALS10	*TARDBP* (1p36, TDP-43)	AD/ AR/ Sporadic	-mRNA maturation. -Neurotoxicity.	-TDP-43 inclusions[17].	-Mean onset at 47, survive for 5.5 years[16]. Typical ALS. Most with slow progression.
ALS11	*FIG4* (6q21)	AD/ Sporadic	-Endosomal membrane fusion and Autophagy[19].	-Unknown	-Also found in CMT4J. Mean onset 56 years old. Either ALS or PLS. Mean 9 years duration[18].

Subtypes	Genes	Inheritance	Functions	Pathology	Remarks
ALS12	*OPTN* (10p15)	AD/ AR	-Regulates NF-κB activation by TNF-α.	-OPTN +ve, cytoplasmic inclusions.	-Also in POAG. Onset at 30-60 years old. Slow progression[20].
ALS13	*ATXN2* (CAG repeat with CAA interruptions[22])	AD	-modifier of TDP-43 toxicity.	-Abnormal localization of ATXN2 in ALS.	-~4.7% (24-34 repeats) of ALS[21]. Earlier age of onset in repeat +ve patients.
ALS14	*VCP* (9p13)	AD	-ERAD, Autophagy.	-TDP-43 aggregates and Bunina bodies.	-Found in IBMPFD and ~1% ALS. +ve Family history for FTD, Age of onset 37-53, survival 29m- 12 years [23].
ALS15	*UBQLN2* (Xp11)	XD	-Ubiquitin ligation in Proteasome degradation pathway.	-Skeine like inclusions, which contain ubiquilin 2.	-5 Families, 90% penetrance at 70 yro. Adult onset, survival 2-7 years. Dementia was prominent in several cases [24].
ALS16	*SIGMAR1* (9p13)	AR	-ER Chaperone, Ion channel modulation.	-Abnormal subcellular distribution.	-Saudi Arabian consanguious family of Juvenile ALS[25]. Lower limb spasicity at onset (1-2 year old), slow progression. Normal Cognitive functions.
ALS18	*PFN1* (17p13.2)	AD	-Converts G-actin to filamentous F-actin.	-Mutant PFN1 may form aggregates.	-Typical ALS, Mean age of onset in FALS 44.8 years [26]. -PFN1 is associated with Miller Dieker Syndrome.
ALS-FTD					

Subtypes	Genes	Inheritance	Functions	Pathology	Remarks
ALS-FTD1	? (9q21-22)	AD/ Sporadic	- Unknown.	- Unknown.	-ALS and FTD. Mean onset 54 years old, mean survival 3.8 years. Develop together[27].
ALS-FTD2	*C9orf72* (9p21)	AD/ Sporadic	-Unknown.	-Various TDP-43 pathology	-Due to intronic GGGGCC expansion. Most common mutation in FALS and FALS/FTD[33]. -Associated with a founder haplotype.
ALS-FTD3 (ALS17)	*CHMP2B* (3p11)	Sporadic	-Endosomal pathway for the degradation of transmembrane proteins.	-p62 +ve, Ub +ve inclusions.	-ALS and FTD. Onset at 6th to 8th decade. Account for ~1% of ALS. With bulbar dysfunction [28].
OTHERS					
	DAO (12q24.11)	AD	-D-serine metabolism and Excitotoxicity.	-Reduced DAO activity in spinal cords of obligate carrier.	-Rapid progression with mean age of death of 44 years [29].
	DCTN1 (2p13.1)	AR	-Axonal transport.	-Binding to microtubule was decreased in mutant protein .	-Early onset, slow progressive ALS [30].
	NEFH (22q12)	AD/ Sporadic	-Neurofilament heavy chain.	-Not available.	-Variable clinical manifestations [31].
	SPG11 (15q21.1)	AR	-Axonal transport and vesicle trafficking.	-Anterior horn cell loss.	-ARJALS (Type1). Mean onset 16 years, slow progressive. Also found in ARHSP-TCC [32].
	PRPH (12q12)	Sporadic	-Type III intermediate filament.	-Cytoplasmic inclusions.	-Typical ALS.

The table summarizes the key features of ALS subtypes and the corresponding causing genes. ALS-related genes that have not been allocated to any subtypes are shown in *OTHERS*. AD: Autosomal Dominant; AR: Autosomal Recessive; ARJALS: Autosomal Recessive Juvenile Onset ALS. Information in this table was obtained from Washington University Neuromuscular Disease Center (*http://neuromuscular.wustl.edu/synmot.html#Hereditaryals*) and the following References:

1 (Kanekura, 2006)
2 (Orrell et al., 1995)
3 (Hays et al., 2006)
4 (Hand et al., 2002)
5 (Chen et al., 2004)
6 (Rosen et al., 1993)
7 (Orlacchio et al., 2010)
8 (Vance et al., 2009)
9 (Sapp et al., 2003)
10 (Yan et al., 2010)
11 (Nishimura et al., 2004a)
12 (Chen et al., 2010)
13 (Tudor, 2010)
14 (van Es et al., 2009a)
15 (Greenway et al., 2006)
16 (Sreedharan et al., 2008)
17 (Kabashi et al., 2008)
18 (Chow et al., 2009)
19 (Ferguson et al., 2009)
20 (Maruyama et al., 2010)
21 (Elden et al., 2010)
22 (Yu et al., 2011)
23 (Johnson et al., 2010)
24 (Deng et al., 2011)
25 (Al-Saif et al., 2011)
26 (Wu et al., 2012)
27 (Hosler et al., 2000)
28 (Cox et al., 2010)
29 (Mitchell et al., 2010)
30 (Puls et al., 2003)
31 (Figlewicz et al., 1994)
32 (Hentati et al., 1998).
33 (Renton, 2011).

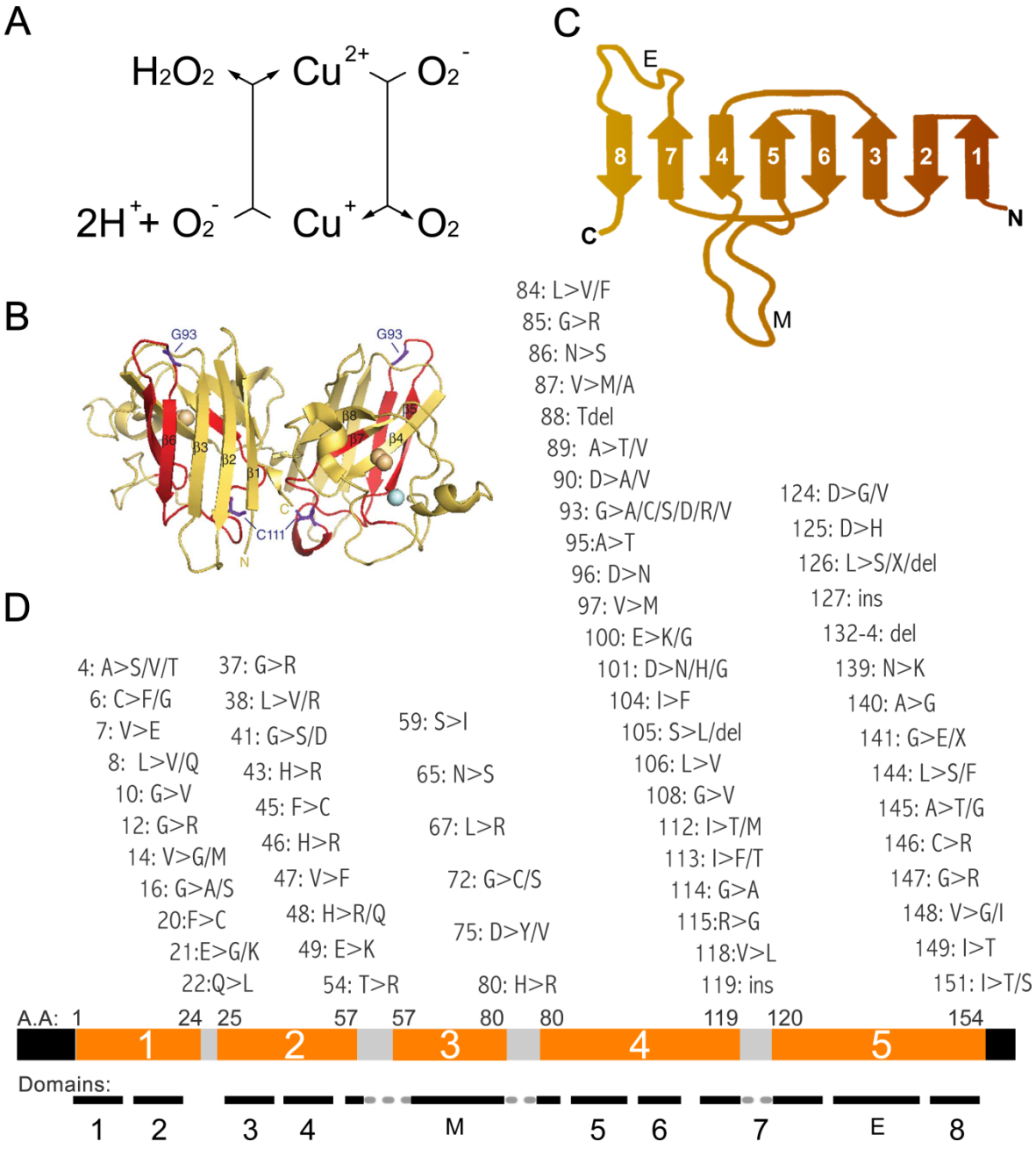
1.6 Mutant SOD1 causes oxidative stress, a common feature in ALS

Superoxide dismutase is a ubiquitously expressed protein that has a protective function against oxidative stress by clearing free radicals generated in normal cell processes. The enzyme is a homodimer (**Figure 1-1B**), and each monomer binds Zn and Cu ions, the latter being an active site for thescavenging activity in which hydroxyl and superoxide ions are catalysed to form hydrogen peroxide and oxygen (**Figure 1-1A**).

The structure of the SOD1 protein has been well characterized. Each monomer consists of an eight-stranded beta barrel, a metal binding loop (residue 49-84) and an electrostatic loop (residue 122-143). These strands are arranged in antiparallel beta-sheet structure, forming a “Greek-key” motif (Tainer et al., 1982) (**Figure 1-1C**). The catalytic copper ion is held by His46, His48, His63 and His120, whereas the Zinc ion, which is important for structural stability, is held by His63, His71, His80 and Asp83 (Rakhit and Chakrabartty, 2006). The dimer is held together by a hydrophobic interface, which is normally buried inside the molecule, and each monomer is stabilized by intramolecular disulphide bonds, which will be further discussed in **Chapter 5 and 8**.

More than 160 *SOD1* mutations have been reported in ALS to date (*http://alsod.iop.kcl.ac.uk*). These mutations distribute throughout the gene but are more frequent in exon 4 and 5 (**Figure 1-1D**). Their effects on protein structure vary depending on the location. For example, A4V, the most common mutation, abolishes the hydrophobic interface (Cardoso et al., 2002), whereas the H43R mutation disrupts hydrophobic packing at one end of the beta-barrel (DiDonato et al., 2003). Rather than isolated effects on enzyme activities, the hazardous consequences of *SOD1* mutations are attributed to destabilizing effects on protein structure which may initiate subsequent pathological processes. It has been concluded that mutant SOD1 toxicity can be mediated through two major gain-of-function mechanisms – oxidative stress and protein aggregation (Bendotti and Carrì, 2004).

Figure 1-1. *SOD1* functions, structures and mutations found in ALS.



A. In the catalytic cycle of SOD1, superoxide ions originating from various sources are catalyzed to form oxygen and hydrogen peroxide through a “Ping-Pong mechanism” in the Cupper-containing active site of SOD1 protein (Rakhit and Chakrabartty, 2006). **B**. SOD1 protein exists as a homo-dimer (Bosco et al., 2010). **C**. Within each monomer there are 8 antiparallel beta sheets, arranged in a 3-dimentional barrel shape (Khare et al., 2005), a metal binding loop (M), and an electrostatic loop (E). **D** shows *SOD1* mutations identified in ALS as summarized by (Bendotti and Carrì, 2004).

Mutant SOD1 is a source of oxidative stress as a result of altered enzymatic activity. The mutant protein accepts nonconventional substrates such as hydrogen peroxide (H_2O_2) and peroxynitrite ($ONOO^-$) and catalyzes the formation of hydroxyradicals ($OH^\cdot$) and the nitration of tyrosine residues of SOD1 itself in the presence of the substrates, respectively. In addition, mutant SOD1 proteins lacking the zinc ion may, in a reverse direction, catalyze the formation of superoxide anion ($O_2^{\cdot-}$) from oxygen molecules (Pasinelli and Brown, 2006). A possible explanation for these aberrant activities was that the mutations mediate conformational changes allowing more access of nonconventional substrates to the Cu ion, the active site for catalyzation (Beckman et al., 1993).

Although it appears that the reactive oxygen species (ROS) and reactive nitrogen species (RNS) generated from the aberrant chemistry are not prerequisite for the development of ALS (Luty et al., 2010), there were several evidence supporting the involvement of oxidative stress in the disease. It has been shown that the concentrations of oxidized nitric oxide products are increased in Cerebrospinal fluid (CSF) of SALS patients (Zou et al., 2013), and ROS levels are elevated in the CSF of G93A transgenic mice in vivo (DeJesus-Hernandez et al., 2011a). The hazardous consequences caused by free radicals have been validated by findings of elevated markers of oxidative damage to protein (Muller et al., 1999), lipid (Shibata et al., 2001) and DNA (Zhang et al., 2007a) in postmortem tissues of SALS. Furthermore, the exposure of cultured neurons to ROS causes toxicity and reduced viability (Weihl et al., 2008). 4-hydroxynoneal (HNE), a toxic product resulting from lipid peroxidation, has been reported to have an elevated conjugation rate with other proteins in the spinal cord of ALS patients, possibly by compromising the functions of the conjugated proteins (Mizuno et al., 2003).

In addition, oxidative stress has been shown to interact with other pathogenic mechanisms in ALS. Firstly, ROS/ RNS exacerbate excitotoxicity by inhibiting glutamate transporters. This results in the accumulation of synaptic glutamate, which activates glutamate receptors, and allows the influx of Calcium ions and causing neurotoxicity (Barber and Shaw, 2010). Secondly, prolonged oxidative stress may cause mitochondrial damages (Barber and Shaw, 2010). Mitochondrial dysfunction (**Section 1.9**) is one of the pathogenic mechanisms in ALS, in which morphological changes of mitochondria and dysfunctions of the respiratory chain have been reported. In addition, mutation rate of

mitochondrial DNA was found to be elevated in the brain of SALS patients (Dhaliwal and Grewal, 2000). In SOD-FALS, direct damage to the mitochondria may be caused by the ROS generated by the mutant SOD1 proteins that are localized to the mitochondrial intermembraneous space (Blauw et al., 2012). Thirdly, oxidative stress can be incurred during oxidative folding, when disulphide bonds are introduced between cysteine residues. The formation of inappropriate disulphide bonds has been implicated in ALS and will be discussed further in **Chapter 5**.

1.7 Mutant SOD1 proteins are prone to aggregate

An alternative pathogenic hypothesis was that the harmful effects of mutant are due to formation of SOD1 aggregates and this is potentially intriguing. Mutant SOD1 can aggregate to form amyloid like fibrils through interaction between the loops and beta-barrels. High molecular weight, detergent insoluble SOD1 aggregates have been detected in inclusions in neurons and astroglia from SOD1- FALS patients and transgenic mice (Johnston et al., 2000). The aggregates are formed before the onset of symptoms and their levels are correlated with stages of disease. The aggregates may contribute to disease by affecting axonal transport, proteasomal degradation and the sequestering properties of molecular chaperones (Elam et al., 2003).

It has been shown that *SOD1* mutations not only impair metal binding ability, but also give rise to a non-native protein- protein interaction interface that promotes aggregation (Elam et al., 2003). Indeed, SOD1 proteins lacking metal binding (Apo-SOD1), are unstable and can form fibrillar structures under destabilizing conditions. The aggregates, however, mainly comprise WT, full length SOD1 protein and it was therefore deduced that the aggregation takes place in a two-step process. In the nucleation process, a nucleus composed of abnormal SOD1 is formed and this is initiated by the cross-linking of disulphide-reduced apo-SOD1 through intermolecular disulphide bonds. The nucleus then recruits WT-SOD1 for an elongation process in which partially metallized SOD1 are mainly involved (Chattopadhyay et al., 2008). Indeed, oxidation of wild type SOD1 results in conformational changes that are similar to that caused by *SOD1* mutations, suggesting that SOD1 may also play a role in ALS without *SOD1* mutations (Bosco et al., 2010).

1.8 ER Stress, Unfolded Protein Response (UPR) and *VAPB* Mutations (ALS8)

In addition to the cytoplasm, mutant SOD1 proteins are also localized to different organelles including the endoplasmic reticulum (ER), where proteins targeting the secretory pathway are folded (Urushitani et al., 2008). These proteins are translocated to the ER and assembled into functional structures under an oxidative environment, and those that fail to fold properly are degraded through the ubiquitin-proteasome system (UPS) or autophagy (Malhotra and Kaufman, 2007). As a mean of protein quality control, accumulation of unfolded proteins in ER activates the Unfolded protein response (UPR), which mediates a decrease in global transcription rates, up-regulation of ER chaperones and apoptosis (Ron and Walter, 2007). As summarized in **Figure 1-2**, this is accomplished by the detection of ER stress by the molecular chaperone BiP/Grp78 and the transduction of signals through the IRE1, ATF6 and PERK pathways.

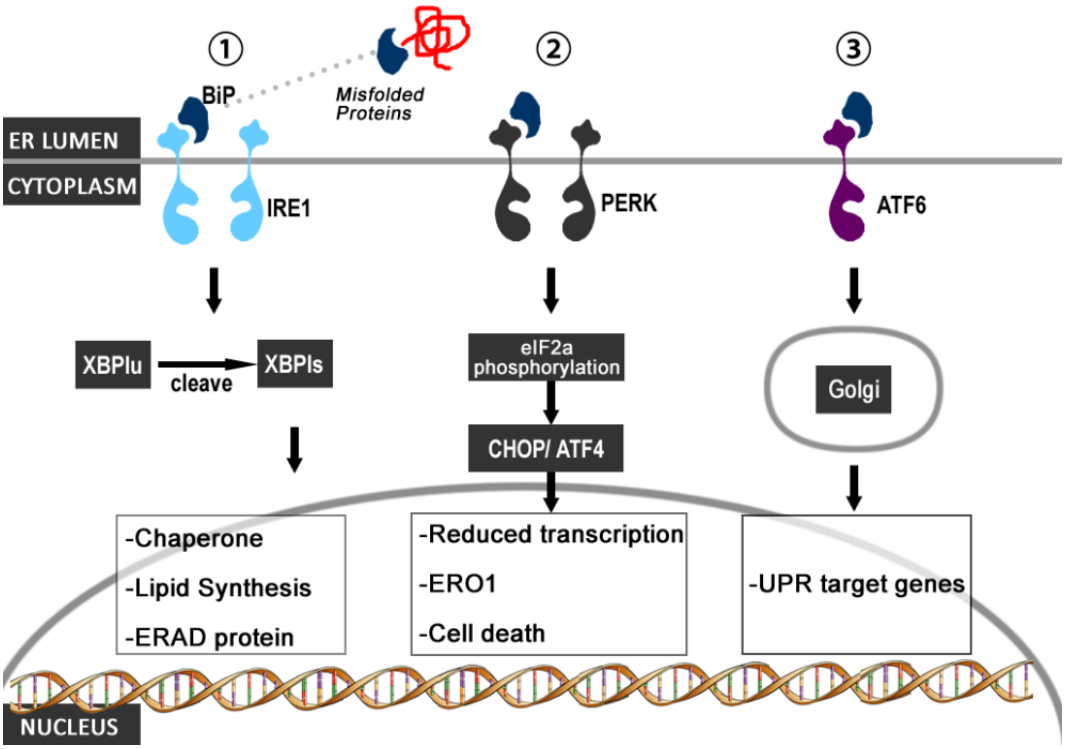
Importantly, prolonged ER stress may induce apoptosis via pathways coupled to the UPR [reviewed by (Kadowaki et al., 2004)]. First, the transcription factor CHOP can be activated through the PERK pathway, represses the expression of BCL-2 and promotes apoptosis. Second, ER stress also promotes the efflux of Ca^{2+} to the cytoplasm mediated by the BCL-2 family members, Bax and Bak, which are located on the ER membrane. Increased cytoplasmic Ca^{2+} activates Calpain, which in turn activates Caspase 4 (Caspase 12 in rodents), and causes cell death (Orrenius et al., 2003). Third, Caspase-12 can be activated as a result of the interaction between activated IRE1 and TNF receptor associated factor 2 (TRAF2). IRE1-TRAF2 also activates the ASK1-JNK pathway and lead to apoptosis (Kadowaki et al., 2004).

The involvement of ER stress in ALS is supported by evidence that the UPR markers IRE1, PERK and ATF6 are overexpressed in the spinal cord of SALS (Atkin et al., 2008). Expression of another marker, BiP/Grp78, is increased in cultured cells and transgenic mice models prior to the onset of symptoms (Tobisawa et al., 2003). It has been demonstrated that Caspase-12 is activated in mutant SOD1 transgenic mice (Wootz et al., 2004), and CHOP is up-regulated in both transgenic mice and SALS patients (Ito et al., 2009), suggesting the involvement of UPR-coupled apoptosis. Indeed, instead of merely being a source of stress, mutant SOD1 also represses the clearance of misfolded proteins by interacting with essential components, such as Derlin 1, in the retrotranslocation

machinery in Endoplasmic reticulum associated protein degradation (ERAD) (Nishitoh et al., 2008).

An ALS candidate gene related to ER stress is Vesicle-associated membrane protein-associated protein B (*VAPB*). *VAPB* mutations were first identified in a multigeneration Brazilian FALS pedigree linked to chromosome 20q13 (Nishimura et al., 2004b). The phenotypes of *VAPB* mutation carriers varied and were classified into three categories: ① late-onset Spinal muscular atrophy (SMA); ② atypical ALS; and ③ typical ALS (Nishimura et al., 2004a). Both typical (survival <5 years) and Slow progressive cases (up to 18years) have been observed (Landers et al., 2008). VAPB, an ER protein, is a member of the VAP family interacting with Vesicle-associated membrane protein (VAMP) during exocytosis of neurotransmitters. VAPB also plays a regulatory role in the UPR by activating the IRE1 pathway and both P56S and T46I mutations impair this property. In addition, mutant VAPB also inhibits the function of wild type VAPB, which dimerizes with the mutant form and becomes inactivated (Chen et al., 2010; Kanekura, 2006). Therefore, a proposed pathogenic mechanism for *VAPB* was that the cellular functions necessary for counteracting the build-up of protein aggregates are abolished by the mutations.

Figure 1-2. Signal transduction in Unfolded protein response (UPR).



As reviewed by Malhotra et al (2007), the signal transduction of UPR relies on transmembrane receptors Inositol requiring kinase 1 (IRE1), Protein kinase like ER kinase (PERK) and bZiP-containing activating transcription factor 6 (ATF6). Without ER stress, activation of these receptors is inhibited by the binding of the molecular chaperone BiP, which has affinity for hydrophobic residues. Upon the accumulation of misfolded proteins, BiP is released and the receptors are activated. ① Firstly, IRE1 undergoes homo-dimerization and trans-autophosphorylation, acquiring endoribonuclease activity. Acitvated IRE1 cleaves the mRNA of XBPI to form alternatively spliced XBPIs, which act as transcription factors binding ERSE elements in UPR target genes such as ERAD proteins. ②Similarly, the homo-dimerization and auto-phosphorylation of PERK activates its kinase activity, which catalyzes the phosphorylation of eIF2α. This pathway is responsible for the reduction in global transcription rate and up-regulation of genes related to oxidative stress and apoptosis. ③ Upon release from BiP, the activated ATF6s are trafficked to the Glogi complex and cleaved. The cleaved forms become transcription factors binding ATF/cAMP and ERSE-1 elements in the UPR target genes such as *P4HB* (PDI), *CRT* and *PDIA4* (ERP72) (Okada et al., 2002).

1.9 Mitochondrial dysfunction

In addition to being an energy-generating organelle, mitochondria also play roles in different cellular functions and pathways such as calcium storage, apoptosis, steroid synthesis and mitochondrial ROS signaling, and there is accumulating evidence suggesting that mitochondrial functions are disrupted in ALS. Morphological abnormalities such as vacuole formation and conglomerates of dark mitochondria have been observed in proximal axons and cell bodies of motor neurons in ALS postmortem tissues, and impaired respiratory chain functions have been demonstrated (Manfredi and Xu, 2005). These include dysfunctions in the redox carriers Complex I (NADH:CoQ oxidoreductase) and IV (Cytochrome c oxidase, COX), which are located at the inner mitochondrial membrane (Vielhaber et al., 2000). Besides, mitochondira is an important source of ROS, which may in turn induce mitochondrial DNA mutations (Murphy, 2009). Human mitochondrial DNA encodes 37 genes that are crucial for the respiratory functions (Swerdlow et al., 1998). Comi et al(1998) identified a 5' microdeletion that causes premature termination of the mitochondrial *COX1* gene in a patient with motor neuron degeneration . The inhibition of COX1 may induce ROS by enhancing electron leak through Complex I and III and damage other cellular functions (Chen et al., 2003).

Another body of evidence for mitochondrial dysfunction comes from mutant SOD1 transgenic models. It has been observed in SOD1 G93A transgenic mice that the formation of vacuoles are due to detachment between inner and outer mitochondrial membranes and their size is associated with disease progression (Sasaki et al., 2004). Mitochondria are the second largest Ca^{2+} reservoir in cells, however, the Ca^{2+} loading function is impaired in SOD1 transgenic mice (Damiano et al., 2006), and this may result in an increase of cytoplasmic Ca^{2+} level causing neuronal death (Carrì et al., 1997). Similarly to what was found in ALS patients, oxidative phosphorylation in mitochondria is also impaired in mutant SOD1 transgenics, giving rise to a decrease in ATP production and increase in oxidative damage (Mattiazzi et al., 2002). The reason for these changes can be ascribed to the localization of mutant SOD1 in mitochondria. The mutant protein was found to aggregate in mitochondria and co-localize with cytochrome c and peroxisomal membranes (Takeuchi et al., 2002). These mitochondrial mutant SOD1 proteins were hypothesized to cause neuronal damage by facilitating the release of pro-apoptotic molecules, such as cytochrome c, to the cytosol by either forming pores in the

outer membrane or impairing the import of functional proteins to the mitochondria through the TOM/TIM complex (Pasinelli and Brown, 2006). Lastly, mutant SOD1 may also inhibit the function of the anti-apoptotic protein Bcl-2 promoting apoptosis (Pasinelli et al., 2004).

1.10 Neurofilaments, impaired axonal transport and *DCTN1* mutations

Aberrant accumulation of neurofilaments in the anterior horn neurons is characteristic of both sporadic and SOD1- linked ALS, indicating dysfunctions in axonal transport (Hirano et al., 1984; Munoz et al., 1988). Neurofilaments (NFs) are intermediate filaments forming cytoskeleton in neurons and contain light (NF-L), medium (NF-M) and heavy (NF-H) subunits. Overexpression of both NF light (*NEFL*) (Xu et al., 1993) and heavy subunits (*NEFH*) (Côté et al., 1993) in transgenic mice causes motor neuron degeneration and variants in these two genes are associated with ALS patients (Figlewicz et al., 1994). Accumulation of neurofilaments has been observed in SOD1- FALS patients (Rouleau et al., 1996) and it has been shown that NF-L may contribute to the selective vulnerability of motor neurons, since its depletion prolongs the survival of SOD1 transgenic mice (Williamson et al., 1998). However, overexpressing NF-L and NF-H also prolongs the survival of the SOD1 mice (Kong and Xu, 2000) and this was accredited to the protective effects of perikaryal, but not axonal, accumulation of NFs, which may ameliorate damage mediated by Ca^{2+}, glutamate, CDK5 and SOD1 (Cleveland and Rothstein, 2001).

One of the hypotheses for the mechanisms underlying the accumulation of neurofilaments is impaired axonal transport. There are two types of axonal transport: fast and slow axonal transports. Vesicles and mitochondria are transported down the axon via fast axonal transport, which employs proteins of the kinesin family, whereas enzymes and soluble proteins are transported via slow axonal transport (Williamson and Cleveland, 1999). After being translated in the cell body, NFs are transported to the axon through slow axonal transport and provide mechanical support for neuronal structures. Impairments in fast and slow axonal transports have been documented in ALS patients and SOD1 transgenic mice respectively (Sasaki and Iwata, 1996; Williamson and Cleveland, 1999) and these may result in the accumulation of NFs in the cell body. Nevertheless, the observation that axonal transport is reduced in transgenic mice overexpressing NF-H and peripherin as a result of disorganized NFs indicates that impaired axonal transport could also be a consequence of neurofilament accumulation itself. The lack of axonal transport eventually

leads to the depletion of essential components for the axon and neurodegeneration (Collard et al., 1995; Millecamps et al., 2006).

Mutations in Dynactin (*DCTN1*), a multi-subunit complex that attaches cargos to dynein and microtubules during retrograde axonal transport, have been identified in a dominantly inherited FALS kindred presenting with early onset, slowly progressive ALS (Puls et al., 2003) and SALS. The mutation (G59S), located in the p150Glued subunit, causes aggregation of the subunit and induces cell death (Levy et al., 2006).

1.11 Excitotoxicity and *DAO* mutations

Being one of the earliest pathogenic mechanisms proposed in ALS, excitotoxicity was reported before the identification of *SOD1* mutations. Glutamate is a common neurotransmitter in the CNS, but continuous accumulation of glutamate in the synapse mediates neuronal damages by activating NMDA or AMPA receptors, which allow influx of calcium ions (NMDA receptors). Calcium overload triggers apoptosis and activates proteases, phospholipases and nucleases, resulting in tissue damages (Foran and Trotti, 2009). There is evidence indicating altered glutamate metabolism in ALS, including that ① glutamate levels were increased in the CSF of ALS patients (Rothstein et al., 1990); ② CSF from SALS patients induces excitotoxicity in cultured neurons; and ③ the clinically proven effects of the routinely used anti-glutamate drug Riluzole in ALS.

The clearance of glutamate from the synaptic space greatly relies on glutamate transporters residing on postsynaptic membranes and astrocytes. EAAT2, a subtype of glutamate transporters expressed on the surface of astrocytes and other glial cells, mediates the transportation of most glutamate in the CNS (Foran and Trotti, 2009). The expression of *EAAT2* is decreased and aberrant transcripts are present in the spinal cord of ALS patients (Lin et al., 1998). Mutations in the *EAAT2* were identified in FALS and SALS patients but these variants were not proved to contribute to the alternative splicing (Aoki et al., 1998). The N206S mutation, which was found in SALS, reduced glycoslysation of the transporter and impaired glutamate reuptake (Trotti et al., 2001). However, despite lacking EAAT2 accelerated neurodegeneration in SOD1 transgenic mice (Pardo et al., 2006), it was argued that excitotoxicity is a collateral damage rather

than a primary one, since EAAT2 expression was reduced in the presence of other pathogenic sources such as ROS (Rao et al., 2003) and Caspase-3 activation (Boston-Howes et al., 2006).

A recently proposed pathway related to excitotoxicity is D-amino acid metabolism. By sequencing a region in Chromosome 12q22-23 linked to a FALS kindred presenting with early onset and short survival time, Mitchell et al (2010) reported a R199W mutation in D-amino acid oxidase (*DAO*) gene. Gene expression was unaffected but enzyme activity was compromised in both an obligate carrier of the mutation and cultured cells expressing the mutant protein, suggesting a loss of function effect (Mitchell et al., 2010). DAO is a peroxisomal flavin adenine dinucleotide (FAD) dependent oxidase catalyzing the deamination, i.e. degradation, of D-amino acids including D-serine, whose formation is catalyzed by serine racemase in both neurons and glial cells. (Paul and de Belleroche, 2012; Wolosker et al., 2008). Like glutamate, D-serine is a NMDA agonist and enhancer of NMDA- dependent toxicity (Sasabe et al., 2007). Elevated D-serine level was observed in ALS spinal cords and SOD1 transgenic mice, and there is a reverse correlation between DAO activity and D-serine level in spinal cord. Suppression of DAO activity also enhances the activation of NMDA receptors, proposing a direct link towards excitotoxicity (Sasabe et al., 2012). However, although this evidence is consistent with the loss of function hypothesis that DAO mutations cause pathogenicity by mediating the build-up of synaptic D-serine, how the mutants promote ubiquitinated inclusions and cell death in cultured cells remains to be elucidated (Mitchell et al., 2010).

1.12 Apotosis: a possible route to neuronal death

Although whether apoptosis plays a role in ALS pathogenicity is controversial, there are many links between them. As opposed to necrosis (Type III cell death), a term describing premature cell death characterized by karyolysis, Apoptosis, or Type I cell death, is a genetically controlled non-inflammatory process of cell death executed through sophisticated signaling pathways. Apoptosis is characterized by a series of morphological changes, including cell shrinkage, pyknosis, blebbing of cytoplasm and karyorrhexis (Elmore, 2007). These changes are due to protein cleavage, crosslinking and DNA fragmentation mediated by proteases, transglutaminase and endonucleases respectively.

Apoptosis can be activated via intrinsic or extrinsic pathways and finally converges on the execution pathway, which employs effector caspases, i.e. caspase-3, 6 and 7, to activate enzymes such as endonuclease CAD, decomposing DNA at regular interval. Stimuli targeting the intrinsic pathway initiate apoptosis by affecting the permeability of mitochondrial membranes, after which cytochrome c is released to the cytosol and activates caspase-9, which in turns activates caspase-3.

A link between ALS and apoptosis is Calcium overload and this can be induced by the activation of NMDA receptors during excitotoxicity. Elevation of cytoplasmic calcium level contributes to apoptosis by activating different components of the apoptotic cascade (Orrenius et al., 2003). Oxidative stress may also increase cytoplasmic Ca^{2+} concentration by damaging Ca^{2+} transporting systems for the ER and mitochondria , the latter being the main reservoir of Ca^{2+} (McConkey and Orrenius, 1997). As described in **Section 1.8**, apoptosis takes place when different mechanisms are triggered as during prolonged ER stress caused by either protein misfolding or disrupted ER calcium homeostasis (Nakagawa et al., 2000). Furthermore, there is evidence showing that Caspase-1 and Caspase-3 are activated in SOD1 transgenic mice (Li et al., 2000; Pasinelli et al., 1998) and that overexpression of BCL-2 delayed the onset and prolonged the survival of SOD1 transgenic mice (Kostic et al., 1997). However, it was also argued that morphological changes in the motor neurons of these models did not match those seen in apoptosis (Migheli et al., 1999).

1.13 Other forms of protein aggregates, *FUS* (ALS6) and *TARDBP* (ALS10) mutations

TDP-43 is known as the major component of Ubiquitinated Inclusions (UBI), a pathological hallmark of both ALS and Ubiquitin-positive frontal-temporal lobe dementia (FTLD-U, **Section 1.16**), suggesting a unifying pathology between these conditions (Neumann et al., 2006). Congruously, mutations in the gene *TARDBP*, which encodes TDP-43, were found in autosomal-dominant inherited FALS (ALS10) (Sreedharan et al., 2008) and FTD families with or without motor neuron degeneration (Benajiba et al., 2009; Borroni et al., 2009). FALS patients with *TARDBP* mutations developed typical ALS with mean age of onset and survival time of 47 years and 5.5 years respectively (Sreedharan et al., 2008).

In addition, *TARDBP* mutations also cause TDP-43 positive pathology reminiscent of those seen in SALS and FTLD-U (Van Deerlin et al., 2008). TDP-43 is a ubiquitously expressed hnRNP protein that regulates DNA transcription, alternative splicing and RNA stability and is normally localized in the nucleus (Tollervey et al., 2011). However, in ALS/FTLD-U, TDP-43 is depleted from the nucleus forming hyperphosphorylated, ubiquinated C-terminal aggregates in the cytoplasm (Neumann et al., 2006). Therefore, pathogenic effects may be mediated by either loss of normal functions or gain of toxic functions.

TDP-43 is a 414 amino acid protein containing two RNA Recognition Motifs (RRMs) and a glycine-rich C-terminal region implicated in the binding of single stranded DNA and RNA molecules (Mackenzie et al., 2010). TDP-43 preferentially binds UG rich regions in long introns and may regulate splicing or expression of more than 100,000 targets genes (Tollervey et al., 2011). It is known that, by interacting with the UG repeats in Exon 9 of *CFTR* pre-mRNA, TDP-43 induces the skipping of Exon 9 (Buratti et al., 2001). Other recognized splicing mechanisms regulated by TDP-43 include *APOAII* Exon 3, *SMN1/2* Exon7, *SC35*, *HDAC6*, and *S6K1* (Buratti and Baralle, 2010; Fiesel et al., 2011). TDP-43 is also known to be a component of transport granules, processing bodies and stress granules, which are important structures for the transportation, processing and degradation of mRNA (Gendron et al., 2010).

A toxic gain of function effect may be mediated by the cytoplasmic TDP-43 aggregates. The aggregates may directly contribute to axonal damages since overexpression of TDP-43 (wild type) recapitulates ALS- phenotype and TDP-43 pathologies in transgenic mice (Wils et al., 2010). TDP-43 aggregates induce cellular toxicity and apoptosis, impair neurite growth (Zhang et al., 2009), and deplete normal nuclear TDP-43 (Yang et al., 2010). Although mechanisms underlying aggregate formation remain elusive, it was shown that a 25kDa TDP-43 C-terminal fragment, which can be generated by caspases, was prone to form ubiquitin- positive inclusions (Zhang et al., 2009). The RRM2 motif was necessary, but not sufficient, for aggregate formation (Yang et al., 2010). Caspases have been implicated in different neurodegenerative conditions, for example, Caspase-3 is activated in ALS transgenic mice (Li et al., 2000) and the knockdown of Progranulin (*PRGN*) gene, a causal gene for FTLD, causes caspase-3 dependent cleavage and

redistribution of TDP-43 (Zhang et al., 2007b). Another ALS candidate gene, *VCP*, also regulates the distribution of TDP-43 (Gitcho et al., 2009).

The identification of *TARDBP* mutations led to the search of genes containing similar domains within a linked region in Chromosome 16, and mutations in *FUS* gene were identified in FALS kindred. Patients with *FUS* mutations developed typical ALS without cognitive impairments. The mean age of onset was 44.5 years and mean survival time was 33 months. In these patients, ubiquitin- positive and p62- positive inclusions were rare. TDP-43- positive inclusions were absent and FUS- positive cytoplasmic inclusions were found (Vance et al., 2009). FUS was subsequently shown to be a common component of inclusions that are also immune-reactive for TDP-43 and ubiquitin in non-SOD1 FALS and SALS (Deng et al., 2010). *FUS* mutations were also found in FTLD and the FUS protein was identified in ubiquitin- positive, tau-negative, TDP-43 negative inclusions found in FTLD patients, known as FTLD-FUS.

Fused in Sacroma (FUS) is a ubiquitously expressed nuclear protein sharing certain functional and structural features with TDP-43. FUS was known to regulate gene transcription, RNA splicing and nucleo-cytosolic mRNA transportation (Vance et al., 2009). Most *FUS* mutations were found in the C-terminal region, which contains a R/G rich region, a RRM domain and a Zinc finger domain (Lagier-Tourenne et al., 2010). It has been demonstrated that cytoplasmic mislocalization of FUS can be caused by disrupting a 32 amino-acid nuclear localization signal at the C-terminal. This can be affected by some truncating mutations. In addition, it has been shown that FUS mislocalisation is related to methylation of Arginine residues which modulates Transportin-1 mediated nuclear import of mutant FUS protein (Dormann et al 2012). Interestingly, mutant FUS protein interacts with stress granule markers and, probably, processing bodies (Gal et al., 2011) and overexpression of mutant FUS protein causes progressive paralysis reminiscent of ALS accompanied by ubiquitinated inclusions in rats (Huang et al., 2011). Indeed, mutant FUS proteins with defective RNA binding sites failed to induce neurodegenerative phenotypes as the RNA-binding competent mutant FUS did in fly brains, indicating that RNA binding capacity is required for mediating the pathogenic effects (Daigle et al., 2013). Taken together, *FUS* and *TARDBP* mutations may cause deficits in mRNA quality control

systems and these are important pathogenic mechanisms for ALS/FTLD-U (Ito and Suzuki, 2011).

1.14 The Ubiquitin-Proteasome System (UPS), autophagy and *VCP* mutations (ALS14)

As summarized above, like many other neurodegenerative conditions, such as AD, PD and HD, protein aggregation is an important feature of ALS. These aggregates exist in the form of cytosolic inclusions and may mediate various toxic effects, causing neuronal damages. Their formation is probably facilitated by defects in the protein-degrading pathways. There are two systems through which misfolded proteins are eliminated in eukaryotes—The Ubiquitin-Proteasome system (UPS) and Autophagy-Lysosome system. One of the candidate genes characterized in this study, *VCP*, is involved in both systems.

In the secretory pathway, newly synthesized peptides are folded into their nascent state in aid of molecular chaperones, such as heat shock protein 70, in the ER. This process involves the formation of correct disulphide bonds which is allowed by the highly oxidative environment in the ER. However, a considerable proportion of peptides that are not properly folded or subjected to post-synthetically damage are degraded through the UPS. In fact, the binding of misfolded proteins to HSP70 not only helps their refolding to the proper structure, but also mediates their degradation (Goldberg, 2003). The UPS also degrades normal, short-lived proteins and this was known to be important for cellular processes such as the NF-kB pathway. UPS is a multi-staged system, in which the proteins to be degraded are labeled with ubiquitin and have their peptide bonds hydrolyzed in the 26S proteasome. Indeed, the fact that ubiquitin is present in most inclusions, regardless of their component, found in ALS suggests that UPS is actively involved in the elimination of these aggregates. This is supported by evidence showing direct involvement of UPS in degrading mutant SOD1 proteins, but not wild type, in ALS (Urushitani et al., 2002) as well as inclusions found in other neurodegenerative conditions (Taylor et al., 2002).

The mechanisms of the UPS have been extensively reviewed (Ciechanover, 2006; Glickman and Ciechanover, 2002). Brief speaking, the first step being the recognition of

substrates, which are conjugated with ubiquitin in a three- step mechanism. Ubiquitin molecules(Ub) are firstly activated by the ubiquitin-activating enzyme, E1, in an ATP-requiring reaction generating E1-S-ubiquitin intermediate. The activated Ub then form intermediates with Ubiquitin-conjugating enzymes, E2, which transfer the Ub to substrates that are bound to the substrate-specific Ubiquitin ligases, E3. Catalyzed by E3, the Ub molecules are covalently conjugated to the substrates and can be successively added to the internal lysine residue of the previous Ub molecule, forming poly-Ub chains. The second step in UPS is proteasomal degradation carried out in the 26S proteasome, which consists a 20S core particle and a 19S regulatory particle. The 20S core particle is enzymatically active, whereas the 19S regulatory particle is responsible for the recognition of substrates. After degradation, Ub are released from the short peptides by deubiquitinating enzymes and recycled.

Since ubiquitination and degradation take place in the cytoplasm, the substrates need to be extracted from the ER prior to degradation through a process known as retrotranslocation, and the whole machinery comprising recognition, translocation, ubiquitination and degradation is known as ER-associated degradation (ERAD) (Yoshida, 2007). In the very first step of ERAD, unfolded proteins are recognized by ERAD components such as EDEM, OS9, and XTP3B. After that, the misfolded proteins are unfolded with the aid of enzymes that cleave disulphide bonds such as PDI and BiP (Yoshida, 2007). The unfolded substrates then pass through a retrotranslocation channel formed by components such as Derlin-1,VIMP and UBX2 on the ER membrane and this process requires the protein VCP (p97), encoded by the gene *VCP* in which mutations have been identified in FALS. VCP has been shown to connect Derlin-1 through VIMP, recognise the substrates and present them to ubiquitin ligases E3 (Ye et al., 2004). VCP can either promote or inhibit ubiquitination depending on the existence of different cofactors (Halawani and Latterich, 2006). In fact, VCP is able to recognise both non-ubiquitinated and ubiquitinated substrates through interacting with hydrophobic peptide segments and poly-Ub chains, respectively. Once substrates are bound, the ATPase activity of VCP is activated (Ye et al., 2003). Finally, VCP also participates in the delivery of ubiquitinated substrates to the 26S proteasome for degradation (Weihl et al., 2009).

There are further indications for the involvement of UPS in the pathogenic mechanism of ALS. First, in addition to VCP, mutations in another UPS- related gene, *UBQLN2*, have been identified in ALS and this will be discussed in **Section 1.17.10**. Second, UPS has been shown to be a target of the toxic SOD1 aggregates (Urushitani et al., 2002). In the presence of certain cytokines during infection or inflammation, the constitutive subunits of the 20S catalytic core of the proteasome can be replaced by other homologs to form the inducible proteasome (Bendotti et al., 2012). It has been shown in SOD1 mice that the inducible subunits were overexpressed in the spinal cord, whereas the constitutive subunits were decreased along with the activities of the UPS (Cheroni et al., 2009).

Autophagy, on the other hand, is a highly conserved pathway for the degradation of long-lived proteins or cell components. There are three types of Autophagy, i.e. Macroautophagy, Microautophagy, and Chaperone-mediated autophagy. Marcoautophagy, the one implicated in neurodegeneration and a variety of pathogenic conditions, refers to a process in which the cellular components to be degraded are sequestered in vesicles and delivered to auto-lysosomes for breakdown (Yorimitsu and Klionsky, 2005). This involves four distinct stages, which are vesicle nucleation, vesicle elongation, formation of autophagosome and fusion with lysosome (Levine and Kroemer, 2008).

In mamillian cells, Autophagy is regulated in a complicated manner and can be induced by a variety of stimuli (Mizushima, 2007). A well-known condition for induction is nutrient starvation, such as the depletion of amino acids. Autophagy is suppressed by insulin and induced by glucagon in liver cells. Different stimuli converge on autophagy regulators such as mTOR, a Rapamycin target that inhibits autophagy by interacting with Autophagy-related proteins 1 (Atg1) (Jung et al., 2010), and Beclin1 (Atg6)/PI3KIII, which, in contrast, interacts with Atg9 and induces autophagy. After the induction stage, nucleation and elongation stages take place in the participation of different members from homologs of the *ATG* gene family. Membranes are recruited to from an elongating structure known as phagophore, which sequesters the components to be degraded, and a double-membrane vesicle known as autophagosome is then formed. The inclusion of substrates relies on the cargo receptor LC-3, a mammalian homolog of Atg8, found in the inner membrane of the autophagosome. The substrates targeted for autophagy can be recognized by

Sequestosome-1 (SQSTM1, **Chapter 4**), which interacts with LC-3, and shuttled into the autophagosome (Bjørkøy et al., 2005). In the final stage of Autophagy, autophagosomes are fused with lysosomes to form autolysosomes. The contents are eventually degraded by lysosomal hydrolases and released to the cytosol.

There is considerable evidence suggesting that Autophagy may contribute to the pathogenesis of ALS, and this has been recently reviewed (Chen et al., 2012). The evidence can be summarized into **four** aspects. *First*, there is increased autophagic activity in the spinal cord of both ALS patients and SOD1 transgenic mice. (Hetz et al., 2009; Zhang et al., 2011) and suppression of autophagy induced the formation of neuronal inclusions and motor neuron degeneration in mice (Hara et al., 2006). *Second*, both SOD1 and TDP-43 aggregates can be degraded through autophagy and enhancing autophagy using rapamycin reduces both mutant SOD1 and TDP-43 aggregates in cultured cells (Caccamo et al., 2009; Kabuta et al., 2006). However, it should be noted that rapamycin also induces apoptosis in SOD1 transgenic mice (Zhang et al., 2011). *Third*, Lithium, a modulator for autophagy, mediates neuroprotective effects and delays disease progression in ALS patients (Fornai et al., 2008), although further confirmation is needed (Aggarwal et al., 2010). *Fourth*, most importantly, several ALS related genes play roles in pathways associated with autophagy. They are *UBQLN2* (**Section 1.17.10**), *CHMP2B* (**Section 1.17.12**), *SQSTM1* (Chapter 3), *DCTN1* and *VCP*. Mutations in *DCTN1,* which encodes the p150 unit of dynactin, has been found in both FALS and SALS (Münch et al., 2004). Dynactin is an activator for the motor protein dynein and mediates the binding of cellular components to microtubules during vesicle transport, including autophagosomes, and the fusion of autolysosomes may be interrupted when this is impaired (Chen et al., 2012). In addition, VCP also has been shown to participate in autophagy as loss of VCP causes accumulation of autophagosomes and failure of the formation of autolysosomes (Ju et al., 2009). A detailed background of VCP will be discussed in **Chapter 3**.

1.15 The NF-kB pathway: TDP-43 aggregates, *SQSTM1* and *OPTN* mutations

ALS candidate genes *SQSTM1* and *OPTN* both participate in the NF-kB pathway. NF-kB is a widely expressed protein complex that regulates gene expression. Targets of NF-kB are known to be involved in the regulation of immune responses, cellular stress, cell

differentiation and apoptosis (Oeckinghaus et al., 2011). In the presence of various stimuli, NF-kB can be activated through either canonical (common) or non-canonical pathways, after which the activated NF-kB translocates to the nucleus and binds DNA sequences known as kB sites (Reviewed by (Oeckinghaus et al., 2011)). In the canonical pathway, prior to activation, the NF-kB molecule, composed of two subunits, p65(RelA) and p50, is bound to the inhibitor IkB, and this is known as the latent state. The pathway can be activited by cytokine receptors such as IL-1R, TNFR and TLR4. Upon activitation, IkB is phosphorylated by IKK, an IkB kinase consisting NEMO, IKKα and IKKβ, and degraded by the proteasome. This process releases NF-kB, which then enters the nucleus and regulates transcription. In the non-canonical pathway, in contrast, the two subunits of NF-kB complex are p100 and RelB. After induction of this pathway, activated IKKα phosphorylates the p100 subunit, which is then partially degraded. Then, the active form of NF-kB, consisting of p52 and RelB, enters the nucleus and regulates transcription.

Interestingly, the NF-kB pathway has been recently implicated as a possible route for TDP-43 mediated toxicity (Swarup et al., 2011).It was shown in the study that TDP-43 activates the p65 subunit of NF-kB and that the expression levels of both TDP-43 and p65 are elevated in the spinal cord of SALS patients. The activation of NF-kB also contributes to neuronal vulnerability caused by excitotoxicity and inflammation. Furthermore, ALS candidate genes *OPTN* (**Section1.17.8**) and *SQSTM1* (p62, **Chapter 4**) are both involved in the NF-kB pathway. OPTN acts as a NF-kB inhibitor, whereas p62 mediates its activation. When the Nk-kB pathway is activated through membrane receptors, an adaptor protein TRAF-6 is recruited, and the enzyme aPKC is employed to activate the IKKs. p62 functions as a scaffold protein linking these two components. This pathway mediates the NF-kB activation through the RANK receptor, which regulates the activation and differentiation of osteoclasts when activated by RANK-L, and, therefore, Paget's disease of bone (PDB) is caused when the pathway is disrupted with *SQSTM1* mutations (McManus and Roux, 2012).

1.16 ALS-FTD and *C9orf72* mutations

Despite being characterized by degeneration of the motor system, cognitive impairment reminiscent of Frontotemporal dementia (FTD) is frequently noticed in ALS. FTD is a

heterogeneous group of conditions caused by severe frontotemporal lobar degeneration (FTLD) and it encompasses three groups of symptoms (Lillo and Hodges, 2009). ① The behavioural variant of FTD (bvFTD) is characterized by profound changes in personality and social behavior, apathy and alterations in appetite; ② Semantic dementia (SD) is marked by loss in semantic memory causing anomic aphasia, i.e. difficulties in naming and comprehensions; ③ In Progressive non-fluent aphasia (PNFA), there are progressive losses in phonological and grammatical abilities with preserved word comprehensions. Cognitive impairments had been reported in 51% of SALS patients, 15% of which met the diagnostic criteria for FTD (Ringholz et al., 2005). Conversely, motor neuron degeneration had also been described in patients with initial diagnosis of FTD (Giordana et al., 2011). ALS-FTD was reported to have shortened survival times compared to classical ALS (Olney et al., 2005).

As previously mentioned, the clinical overlap between ALS and FTD is supported by both pathological and genetic findings. Histological subtypes of FTLD are distinguished by immunoreactive profiles (Giordana et al., 2011). A type of FTLD presents with tau-positive inclusions which are found in other conditions such as Pick's disease, corticobasal degeneration and progressive supranuclear palsy, whereas TDP-43 is the major component of tau-negative, ubiquitin-positive inclusions that are present in the most common histological variety, FTLD-U (FTLD-TDP-43) (Weder et al., 2007). In addition, FTLD-FUS refers to conditions where FUS-positive inclusions are found (Mackenzie et al., 2010). TDP-43 and FUS were both found in ALS inclusions, suggesting a common disease mechanism.

ALS and FTD co-exist in familial form of autosomal dominant inheritance and has been linked to chromosome 9p21 (Vance et al., 2006). Analysis of Genome-wide association signals narrowed the susceptible region to a 232kb LD block spanning three genes: *MOBKL2B*, *C9orf72*, and *IFNK* (Mok et al., 2012) and it was recently confirmed that a hexanucleotide GGGGCC expansion in the intron 1 of *C9orf72* gene is responsible for the linkage signal and segregates with disease (Renton et al., 2011). This will be further discussed in **Chapter 7**.

1.17 Other ALS Loci

1.17.1 Alsin (ALS2) mutations

Autosomal recessive juvenile amyotrophic lateral sclerosis (ARJALS) refers to a group of rare forms of ALS presenting with onset before 25 years and long progression (Hamida et al., 1990). There are three types of ARJALS and the loci for type 3 (ALS2) and type 1 (ALS5) were mapped to chromosome 2q33 and 15q15-22 respectively.

Sequence analysis of transcripts from Chromosome 2q33 resulted in the identification of mutations in a gene with 34 Exons, *ALS2* (Hadano et al., 2001). ALS 2 (ARJALS Type 3) is characterized by juvenile onset, long survival time and prominent upper motor neuron symptoms such as spasticity of limb, facial and tongue muscles resembling primary lateral sclerosis (Hentati et al., 1998). *ALS2* mutations are also responsible for Infantile onset ascending hereditary spastic paralysis (IAHSP) and Juvenile primary lateral sclerosis (PLSJ), which causes pure upper motor neuron symptoms.

Multiple mutations have been identified in this gene, most of which are frameshift and nonsense mutations causing truncation of the protein. ALS2 is an 184kDA protein containing motifs that are homologous to members of Guanine nucleotide exchange factors (GEFs), which promote the dissociation of GDP from inactive GTPases. There are three GEF domains in the Alsin protein. The RCC-1 like domain locating near the N-terminal is a regulator for Ran GTPase, which was implicated in nuclear transfer and chromatin condensation. The DH/PH domain, which locates in the middle part of protein, and VPS9 domain, which located at the C-terminal, regulate the GTPases Rho and Rab5 respectively. Rho plays a role in the regulation of cytoskeleton and neuronal morphogenesis, whereas Rab5 regulates vesicle trafficking (Otomo et al., 2003). There are two alternative spliced transcripts of *ALS2* producing a long, 1,659 amino acids, and a short, 396 amino acids, proteins respectively. As reviewed by Chandran et al (2007), phenotypes were initially reported to be associated with the length of mutants. Lower motor neurons tend to be unaffected when the short form of *ALS2* remains intact, leading to PLSJ or IAHSP, but this has been complicated by a later observation that lower motor neurons were not affected in some mutations affecting the short form of *ALS2*.

More pathogenic consequences of *ALS2* mutations have been revealed by the catalytic activity of the VPS9 domain for Rab5. ALS2 may play a role in vesicle trafficking as it promotes endosome formation though the activation of Rab5 located in endosomal compartments (Otomo et al., 2003). The RLD domain, however, appeared to be a negative regulator for Rab5 activation and it is able to bind the PDZ domains of Glutamate receptor interacting protein (GRIP1), which interacts with a calcium-impermeable GluR2 subunit of the AMPA receptor. It has been shown that dysfunction of ALS2 reduces the presentation of GluR2 to the plasma membrane and enhances glutamate- induced excitotoxicity (Lai et al., 2006). Although the lack of *ALS2* was insufficient to induce a motor neuron disease phenotype in knock-out mice, it did increase the vulnerability to oxidative stress in cultured neurons (Cai et al., 2005).

1.17.2 18q21 (ALS3)

Chromosome 18q21 has been linked to a European kindred presenting autosomal dominant typical ALS. The mean age of onset was 45 years and survival time was 5 years. The region defined by D18S846 and D18s1109 spans 7.5cM (8Mb) and contains 50 genes. No mutation has been identified to date (Hand et al., 2002).

1.17.3 SETX (Senataxin, ALS4)

ALS4 is the only form of dominantly inherited juvenile onset ALS. Onset is often seen before the second decade witha slow progression and a normal life span. Affected individuals usually present with distal weakness and wasting, pyramidal signs and normal sensation (Chen et al., 2004). Linkage analysis of an 11-generation autosomal dominant juvenile onset ALS pedigree from Maryland mapped the locus to a 5cM interval in chromosome 9p34 (Chance et al., 1998). A mutation was subsequently identified in the *SETX* gene in this kindred and two other mutations in the same gene were confirmed in additional kindreds (Chen et al., 2004). *SETX* mutations were also found to be a cause of Ataxia-oculomotor apraxia type 2, a syndrome comprising early onset cerebrellar atrophy, axonal sensorimotor neuropthy, oculomotor apraxia and elevated serum AFP (AOA2, OMIM 606002)

Senataxin shows homology to the fungal Sen1p protein, which possesses helicase activity and is required for splicing and termination of tRNA, small nuclear RNA and small

nucleolar RNA (Moreira et al., 2004). Senataxin is a 302.8 kD protein with a DNA/RNA helicase domain at the C-terminal (OMIM 608465) and Chen et al (2004) described that there are two main transcripts of 11.5kb and 9kb respectively. A nuclear localization signal and an ATP/GTP binding site which is essential for the activity of helicases were also identified . In addition, the C-terminal of Senataxin also shows homology to two other members of the DExxQ-box family of helicases: RENT1/Upf1 and IGHMBP2. RENT1 plays a role in non-sense mediated RNA decay (NMD), whereas IGHMBP2 has been known to bind a specific DNA sequence in the immunoglobin mu chain switch region (Moreira et al., 2004). IGHMBP2 mutations have also been implicated in Spinal muscular atrophy with respiratory distress type 1 (Tachi et al., 2005).

Recent research has revealed more about the roles that Senataxin plays in neurodegeneration. Expression of full-length Senataxin alleviates oxidative DNA damage in lymphoblastoid cell lines derived form AOA2 patients, indicating that the lack of helicase activity may abolish the cell's ability to repair double strand breakage (DSB) damage when subjected to oxidative stress (Suraweera et al., 2007). However, this can also be a consequence of alteration in expression of other genes required for DNA damage repair, which can be regulated by the Senataxin homolog Sen1 through regulating RNA polymerase II (Steinmetz et al., 2006). In agreement with this, it has been recently shown that Senataxin regulates transcription termination by allowing the access of 5'-3' exonuclease Xrn2 to RNA/DNA hybrid structures (Skourti-Stathaki et al., 2011). Senataxin also regulates mRNA splicing (Suraweera et al., 2009) and neuronal differentiation by promoting the expression of fibroblast growth factor 8 (FGF8) (Vantaggiato et al., 2011). In conclusion, although the exact way in which *SETX* mutations cause neurodegeneration has not been clarified, it is probably achieved by altering the expression of other genes that are important for the survival of neurons.

1.17.4 SPG11, Spatacsin (ALS5)

Mutations in the *SPG11* (Spatacsin) gene located in the Chromosome 15q15-22 locus were recently found to be responsible for another form of ARJALS (type 1) , also known as ALS5 (Orlacchio et al., 2010). This gene is also causal for Autosomal recessive hereditary spastic paraplegia with thin corpus callosum (ARHSP-TCC) (Stevanin et al., 2007). Onset of ARHSP-TCC is usually before the second decade. In ARHSP-TCC, there

is progressive spasticity and stiffness of lower limb and mental retardation. Cognitive impairments are occasionally observed (OMIM 604360). Age of onset of ALS5 ranges from 8 -18 years old and progression is usually longer than ten years. In contrast to ALS2, lower motor neuron symptoms being in limbs and upper motor neuron symptoms are moderate in ALS5 (Hentati et al., 1998).

Very little is known about the functions and pathogenic roles of Spatascin. Spatascin is a 2,443 amino acid protein that belongs to the Aromatic compound dioxygenase superfamily. There are four transmembrane domains indicating functions as receptor or transporter. Spatascin mutations are likely to cause disease in a loss of function manner, since most of them are truncating mutations (Orlacchio et al., 2010). Spatascin is expressed in neuronal cell bodies and co-localizes with markers for mictrotubules, endoplasmic reticulum, mitochondria and vesicles involved in protein trafficking, suggesting roles in axonal transport and vesicle trafficking (Murmu et al., 2011). Indeed, the major pathological feature in Hereditary spastic paraplegia is the retrograde degeneration of long nerve fibres in corticospinal tract and dorsal columns as a result of disrupted axonal transport, cytoskeleton organization, membrane trafficking and mitochondrial metabolism (Salinas et al., 2008). Given the fact that upper motor neuron symptoms also predominates in ALS5, it is likely that Spatascin has a crucial function in maintaining adequate axonal nutrient supplies for neurons in the corticospinal tract.

1.17.5 20p13 (ALS7)

A 5Mb locus on 20p13 was reported to be shared by two out of 15 siblings in a FALS kindred from Boston with a LOD score of >3.0. This was designated as ALS7. The mean age of onset was 56.5 years and mean survival time was 2.9 years. However, this locus has not been identified in any other kindred and additional information is needed for confirmation (Sapp et al., 2003).

1.17.6 ALS9: Angiogenin (ANG)

The gene *ANG* encodes a protein essential for angiogenesis, Angiogenin, and is located on chromosome 14q11. Angiogenin (ANG) facilitates the formation of new vessels by mediating the degradation of basal membranes degradation and allowing migration of endothelial cells. ANG not only binds to receptors on the endothelial cells, such as alpha-

actin and ANG receptor, but also undergoes nuclear translocation in the target cells. Target genes harbouring ANG-binding elements, i.e. [CT]n repeats, can be regulated in this way. RNase activity, another property of ANG, is also necessary for its functions [reviewed by (Gao and Xu, 2008)]. It is noteworthy that ANG is required for cell proliferation induced by other angiogenic proteins, such as Fibroblast growth factors (FGF), Epidermal growth factor (EGF) and *Vascular endothelial growth factor* (VEGF) (Kishimoto et al., 2005).

During hypoxia, Hypoxia inducible factors (HIF) upregulates VEGF and promotes angiogenesis. Mice lacking HIF binding region in the promoter of *VEGF* gene developed symptoms reminiscent of ALS, probably due to impairment in neural vascular perfusion (Oosthuyse et al., 2001). Meta-analysis of subjects from Sweden, Belgium and England showed that a risk haplotype in the promoter region of *VEGF* both conferrs risks of ALS and correlates with the expression of the gene (Lambrechts et al., 2003).

Association studies indicated that a common coding SNP in *APEX* gene, located on chromosome 14q11.2-12, is associated with ALS (Hayward et al., 1999). However, this association was then later allocated to *ANG,* which locates 237Kb downstream to *APEX* (Greenway et al., 2004). In the same study it was demonstrated that the SNP rs11701 in the *ANG* gene is significantly associated with ALS and coding mutations were subsequently identified in both FALS and SALS patients from the Irish/Scottish population (Greenway et al., 2006). *ANG* mutations have been identified in ALS patients from America, Italy, France, Netherlands and Germany (Fernandez-Santiago et al., 2009). ANG- ALS cases are typical ALS. Age of onset ranges from 27 to 76 years and survival time ranges from 0.8 to 10 years (Greenway et al., 2006). K40I and C39W *ANG* mutations may affect the active sites or folding of the protein, whereas S28N mutation was located adjacent to the nuclear localisaziton sequence (NLS) and may impair nuclear import. Wu et al (2007) showed that K17I, S28N and P112L mutants failed to induce angiogenesis and the RNase activity essential for ANG function was abolished. Q12L and K40I mutants impaired neurite extension and reduced the survival of cultured motor neurons (Subramanian et al., 2008). In addition, the cytotoxic effects of ANG mutants could be explained by the motor neurons' tendency to form stress granules (SGs), a structure composed of sequestered ribonucleoproteins during cellular stress, when mutant ANG

failed to induce RNA cleavage (Thiyagarajan et al., 2012). This provides RNA for the assembly of SGs. RNA may be necessary for the assembly of SGs. For example, it has been recently shown that TDP-43 inclusions are colocalized with SG markers, causing a variety of pathogenic consequences (Li et al., 2013). RNase treatment abolished the binding between endogenous TDP-43 and SG protein TIA-1, suggesting that the interaction was dependent on RNA, which can be used as a therapeutic target (Liu-Yesucevitz et al., 2010).

1.17.7 FIG4 (ALS11)

There is a genetic overlap between ALS11 and Charcot-marie-tooth disease (CMT) type 4J (CMT4J), which are both caused by *FIG4* mutations. *FIG4* encodes for phosphastidylinositol 3,5- bisphosphate (abbrev: PI(3,5)P2) 5-phosphatase, also known as SAC domain containing protein 3 (Sac3). The phosphastase FIG4 links with molecules ArPIKfyve (VAC14) and PIKfyve, which has kinase activity, to form a complex (PAS complex) that regulates the metabolism of PI(3,5)P2. The PAS complex therefore possesses both phosphostase activity that removes a phosphate group of PI(3,5)P2 to form PI3P and kinase activity that catalyzes the opposite reaction. PI(3,5)P2 is a less abundant type of inositol phospholipids, which are membrane molecules playing important cellular functions such as controlling membrane-cytosol interface, defining organelle identity, membrane trafficking and cytoskeleton organization (Di Paolo and De Camilli, 2006). PI(3,5)P2 was known for its function in regulating membrane homeostasis in late endosomes and a possible role in the retrograde transport from late endosome to trans-glogi network via various effectors (Michell et al., 2006).

FIG4 mutations have been identified in both FALS and SALS. Clinical presentations were typical and the FALS kindred were of dominant inheritance. Average age of onset of these patients was 56 years and average duration time was 9 years (Chow et al., 2009a). Variants found in *FIG4* included two truncating mutations located upstream of the SAC active site and two splice site mutations that were predicted to induce exon skipping. Two missense mutations, D53Y and R388G, were shown to have deleterious effects on protein function. I41T, the commonest *FIG4* mutation in CMT4J (Chow et al., 2007), was not reported in ALS.

A pathogenic model of *FIG4* mutation has been proposed using I41T transgenic mice. It was shown that the stability of FIG4 was maintained by VAC14, and the I41T mutation, which abolished the interaction with VAC14, greatly reduced FIG4 activity and caused neurodegeneration (Lenk et al., 2011). Loss of FIG4 in turn reduced PI(3,5)P2 and resulted in vacuolization and accumulation of autophagy intermediates causing neuronal damages (Ferguson et al., 2009a; Lenk et al., 2011).

1.17.8 OPTN (ALS12)

Homozygosity mapping and direct sequencing in an autosomal recessive consanguineous Japanese ALS family revealed mutations in the Optineurin (*OPTN*) gene, including a homozygous deletion of exon 5 caused by AluJ-mediated recombination and a nonsense Q398X mutation. Q398X was further detected in SALS cases and an additional E478G mutation was detected in an autosomal dominant FALS kindred (Maruyama et al., 2010). Skein-like inclusions positive for OPTN, Ubiquitin and TDP-43 antibodies were observed in postmortem spinal cord tissue of the E478G patient, and OPTN was further shown to colocalize with SOD1 and FUS inclusions in patients with corresponding mutations (Ito et al., 2011b; Maruyama et al., 2010). Subsequent screening identified *OPTN* mutations in cases from Germany (Weishaupt et al., 2013), Italy (Del Bo et al., 2011) and Holland (van Blitterswijk et al., 2012b), giving mutation frequencies of 1-4% in FALS and <1% in SALS. *OPTN* patients present with typical ALS with variable age of onset ranging from 24 to 83 years, and their survival time ranges from 0.75 to 25 years. In addition, *OPTN* is a causal gene for primary open angle glaucoma (POAG) and a risk factor for Paget's disease of bone (PDB) (Albagha et al., 2010), and the latter can also be caused by *SQSTM1* mutations. The fact that both *SQSTM1* and *OPTN* are implicated PDB and ALS highlights their functions in NF-kB pathway (**Section 1.15**).

OPTN is known as an inhibitor of NF-kβ by competing for the binding sites of NF-kβ essential modulator (NEMO), which activates the TNF-α dependent NF-kβ pathway (Zhu et al., 2007). In POAG mouse models, overexpression of an *OPTN* E50K mutation led to neuronal death probably by abolishing the interaction with Rab8 and propagating oxidative stress (Chi et al., 2010). In contrast, ALS- associated *OPTN* mutations mediate loss of function effects. Mutant OPTN proteins lack the ability to inhibit NF-kβ (Maruyama et al., 2010) and knockdown of OPTN in cultured cells results in inappropriate activation of

NF-kβ and promotes cell detath (Akizuki et al., 2013). OPTN also interacts with different proteins involved in vesicle trafficking. One of such is Myosin VI, which is involved in both endocytic and exocytic pathways. OPTN anchores Myosin VI to the Glogi complex and this is required for exocytosis. Moreover, the glogi complex is fragmented in the absence of OPTN (Sahlender et al., 2005). OPTN interacts with Rab8 and Huntingtin (HTT), both of which are implicated in vesicle trafficking, coordinating motor functions and maintaining Glogi structures (Sahlender et al., 2005).

Indeed, that fact that OPTN co-localizes with cytoplasmic inclusions could be explained by a possible role of OPTN in autophagy. It has been recently demonstrated in SOD1 cells that OPTN recognizes and mediates the clearance of protein aggregates through the autophagy- lysosome pathway in an ubiquitin- independent manner and this necessitates the phosphorylation by a kinase TBK1 (Korac et al., 2013).

1.17.9 ATXN2 (Ataxin-2, ALS13)

Spinocerebellar ataxia 2 (SCA2) is an autosomal dominant form of progressive cerebellar ataxia characterized by incoordination of gait and limb movements, ophthalmoplegia, pyramidal signs, mild dementia and peripheral neuropathy (OMIM183090). SCA2 is caused by trinucleotide CAG expansions in the Ataxin-2 (*ATXN2*) gene (Imbert et al., 1996). The expansion codes for a polyglutamine tract in the 1,313-amino-acid protein localized to the Glogi complex and ER. CAG alleles of less than 22 repeats are commonly found in the healthy population. Alleles up to 31 repeats remain non-pathogenic, whereas those of more than 33 repeats are exclusively observed in SCA2 patients (Geschwind et al., 1997).

PBP1, a yeast ortholog of human *ATXN2*, was identified as a modifier that enhances TDP-43 toxicity and it was shown that intermediate *ATXN2* expansions (27-33 repeats) are over-represented in ALS (Elden et al., 2010). In subsequent screening intermediate (31-33 repeats) and longe (32-39 repeats) *ATXN2* expansions were identified in FALS and SALS patients with frequencies of 1.1% and 0.5% respectively. All these patients were typical ALS. The mean age of onset of these patients was 57 years and mean survival time was 35 months (Van Damme et al., 2011). Most expansions with ≥ 34 repeats are interrupted with CAA units, and these imperfect units were associated with

early age of onset in ALS (Ross et al., 2011; Yu et al., 2011). Of note, phenotypic overlap between ALS and SCA2 has been reported in patients with *ATXN2* expansions of ≥33 repeats (Nanetti et al., 2009).

ATXN2 co-localizes and interacts with TDP-43 in a RNA- dependent manner and has a distinct cellular distribution in ALS (Elden et al., 2010). In addition to TDP-43, ATXN2 has been shown to interact with other proteins possessing RNA recognition motifs (RRM) as well as components of P-bodies and Stress granules, suggesting a role in RNA processing (Nonhoff et al., 2007). However, Farg et al (2013) demonstrated that ATXN2 interacts with FUS in a RNA- independent manner. The intermediate poly-glutamine expansions promoted the translocation of mutant FUS to the cytoplasm and enhanced ER stress, Glogi fragmentation and apoptosis induced by mutant FUS protein . These effects may be caused by perturbed vesicle trafficking, another known physiological function of ATXN2 (Farg et al., 2013). ATXN2 has been shown to interact with components involved in endocytosis (Nonis et al., 2008). Expression of expanded ATXN2 abolishes its normal localization to the Glogi complex, triggers Glogi fragmentation and apoptosis (Huynh et al., 2003). Neuronal viability may be compromised when the proper trafficking of essential molecules is disrupted.

1.17.10 UBQLN2 (ALS15)

UBQLN2 is the only known causal gene for dominant, X-linked ALS. The *UBQLN2* P497H mutation was first identified in a five-generation FALS kindred with no male-to-male transmissions and reduced penetrance in females (Deng et al., 2011). Four mutations, P497S, P506T, P509S and P525S, were then identified in a different kindred and it was estimated that *UBQLN2* mutations account for <1% of FALS. Most mutations are located in codons for Proline residues in the characteristic PXX tandem repeat region in the intron-less gene. Age of onset of these patients ranges from 16 to 71 years and the survival time of female appears to be longer than male (Deng et al., 2011).

UBQLN2 positive skein-like neuronal inclusions were identified in UBQLN2-FALS cases. These inclusions are also positive for TDP-43, p62, FUS and OPTN. UBQLN2 positive, p62 positive inclusions were also identified in ALS patients without *UBQLB2* mutations and FTD patients, suggesting ubiquitinopathy is a common feature of ALS/FTD (Fecto

and Siddique, 2012). Intriguingly, Brettschneider et al (2012) described a dinstinctive UBQLN2 distribution that coincided with the *C9orf72* expansion.

UBQLN2 belongs to the Ubiquilin family and has a direct functional link to protein clearance. The protein has three domains, including a N-terminal ubiquitin-like domain, which binds to proteasome, a variable middle part, and a C-terminal ubiquitin-associated domain, which binds to polyubiquitin chains (Fecto and Siddique, 2012). Therefore, the widespread localization of UBQLN2 in neuronal inclusions may be explained by the hypothesis that UBQLN2 plays a role in targeting ubiquitinated substrates for proteasomal degradation. It has also been shown that the C-terminal of TDP-43 interacts with UBQLN2 in cultured cells and mutant UBQLN2 protein impairs the functions of UPS (Deng et al., 2011).

1.17.11 SIGMAR1 (ALS16)

By analyzing shared haplotypes in an autosomal recessive Saudi Arabian ALS family, a homozygous missense mutation, causing p.E102Q substitution, in the *SIGMAR1* gene was found to segregate with disease (Al-Saif et al., 2011). The mutation caused juvenile onset (1 to 2 years), slow progressive ALS without cognitive impairments. In addition, mutations in the 3'UTR of the gene have been identified in ALS/FTD families in which patients have distinct TDP-43 pathology (Luty et al., 2010). *SIGMAR1* encodes for a sigma-receptor 1, which is primarily expressed in spinal cord motor neurons (Mavlyutov et al., 2010). The trans-membranous receptor binds a variety of ligands and functions as a modulator for ion channels and ER chaperones. Moreover, SIGMAR1 has been implicated in neuroprotection and may suppress accumulation of misfolded proteins in the ER (Hayashi and Su, 2007). The p.E102Q mutation changed subcellular localization of the protein and enhanced apoptosis (Al-Saif et al., 2011) and, in addition, a 3'UTR variant was shown to increase SIGMAR1 expression and relocate TDP-43 to the cytoplasm (Luty et al., 2010). Therefore, SIGMAR1 may be related to multiple functional aspects of neurodegeneration and the deleterious effects could be mediated through either gain of function or loss of function mechanisms.

1.17.12 CHMP2B (ALS17)

Charged multivesicular body protein 2b (*CHMP2B*) is a gene that has been linked to FTD (Skibinski et al., 2005). Parkinson et al (2006) reported *CHMP2B* mutations in two unrelated ALS patients. One patient harbouring a Q206H mutation developed ALS at the age of 75 and died at 15 months after onset. There was self-reported family history of ALS and no cognitive impairment was noted. Remarkably, autopsy showed LMN inclusions that were positive for both ubiquitin and p62/SQSTM1. The other patient, who harboured the p.I29V mutation, was a SALS case with a confirmed diagnosis of FTD. Age of onset of this patient was 65 years and survival time was 6 years. Subsequent screening identified one further mutation, p.T104N, and it was found that most of the carriers presented prominent lower motor neuron symptoms. Surprisingly, dementia was not a feature in these cases. *CHMP2B* mutations account for ~1% of ALS (Cox et al., 2010).

CHMP2B is a 213-amino-acid protein containing a Coiled-coil, a Snf-7 and an acidic C-terminal domains. The yeast orthologue of human CHMP2B, VPS2, is a component of Endosomal secretory complex required for transport complex III (ESCRTIII) (Cox et al., 2010). Transmembrane proteins are degraded through the endosomal pathway, in which ubiquitinated substrates to be degraded are firstly contained within vesicles and then matured into late endosomes/multivesicular bodies (MVBs). MVB finally fuses with lysosomes and this process is facilitated by the ESCRT complexes (Hurley and Emr, 2006). MVBs also fuse with autophagosomes (Fader and Colombo, 2009). Therefore, malfunction of the complexes may result in the build-up of ubiquitinated inclusions. It has been reported that CHMP2B is associated with autophagy and the loss of this protein causes neurodegeneration (Lee et al., 2007b). In cell models transfected with ALS associated *CHMP2B* mutations, large vacuoles were found and the turnover of LC3, a marker of autophagic activity, was inhibited, indicating that the vesicle fusion with lysosome and autophagic activity are compromised (Cox et al., 2010).

1.17.13 PFN-1 (ALS18)

Exome sequencing of FALS cases from two unrelated Caucasian and Sephardic Jewish kindred revealed mutations in Profilin 1, *PFN1* (Wu et al., 2012). All families with *PFN1* mutations demonstrated dominant inheritance. The patients presented typical ALS and the mean of age of onset was 44.8 years. Dementia was not reported. Mutant PFN1

gained propensity to form ubiquitinated aggregates, which also contained TDP-43, in cell models, suggesting that the mutants may induce TDP-43 aggregation (Wu et al., 2012). PFN1 regulates actin polymerization, which is important for the rearrangement of cytoskeleton (Theriot and Mitchison, 1993). It has been shown that *PFN1* mutations impair the actin-binding ability of the protein and reduce neurite growth, which requires normal actin dynamics (Wu et al., 2012). It is also noteworthy that PFN1 also interacts with VCP (Witke et al., 1998).

Also using Exome sequencing, mutations in the Matrin 3 (*MATR3*) gene were recently identified in ALS families presenting variable phenotypes (Johnson et al 2014). Matrin 3 is a RNA binding protein that interacts with TDP-43 and Matrin 3 pathology was identified in ALS patients with or without *MATR3* mutations.

1.18 Sequence analysis and the importance of studying known candidate genes in extended cohorts in ALS

Genotypes of a DNA polymorphism/ variant can be determined by the nature of the variant, such as Restriction fragment length polymorphism (RFLP), Amplified fragment length polymorphism (AFLP), Single strand conformation polymorphism (SSCP) or DNA sequencing. The latter method allows a direct determination of DNA sequences and has been widely used for the screening of deleterious genetic variants in human diseases. The Sanger sequencing method, which was developed in the 1970s, is an enzyme based method relying on the random inhibition of chain elongation in the presence of dideoxynucleotides (ddNTPs) [Reviewed by (Strachan, 2011)]. In brief, a DNA template to be analyzed is amplified in four parallel reactions; each contains DNA polymerase, a single primer, dNTP (all four types of bases) and ddNTP (a single type of base). During the reaction, strand elongation is terminated when the ddNTP is incorporated, generating a mixture containing fragments of different sizes corresponding to the positions of the bases. The ddNTPs are labelled with isotopes and the sequence is finally determined by electrophoresis, in which four different lanes are used for the four reactions. In automated Sanger sequencing, the ddNTPs are labelled with fluorescent dyes with different emission lengths, and, therefore, all reactions can be separated in a single electrophoresis lane, allowing better efficiency.

Massively parallel DNA sequencing, also known as the Next generation sequencing (NGS), allows rapidly generating massive amount of sequencing data. NGS relies on detecting signals generated during the incorporation of bases, i.e. sequencing-by-synthesis, instead of the separation of DNA fragments. In most NGS platforms, the DNA templates are fragmentized, attached to an adapter sequence which allows non-specific amplification, and clonally amplified via emlusion PCR or Solid-phase amplification [Reviewed by (Metzker, 2010)]. In commonly used platforms such as Illumina, the sequencing chemistry is based on Cyclic reversible termination (CRT). The DNA amplification process is terminated after the incorporation of a fluorescently labelled dNTP that is chemically modified to stop the reaction. Unlike ddNTP, the modification can be reversibly removed. After that, unincorporated dNTPs are removed by washes and the fluorescent signal can be detected. At the end of each cycle, the chemical modification of dNTP is removed by chemical cleavage allowing the start of the next cycle.

As reviewed above, the identification of DNA mutations has extended the understanding of ALS pathogenesis. The Sanger sequencing method was traditionally employed for mutation screening in a disease-linked locus inferred by linkage studies using multi-generation kindred. Such a locus may, however, still contain a large number of genes that may need to be prioritized for screening using a candidate gene approach. The difficulty has been overcome by NGS-based approaches, such as Whole exome sequencing and Deep resequencing, which, indeed, have recently led to the identification of *VCP*, *C9orf72*, *PFN1*, *ERBB4* and *hnRNPA2B1* mutations in ALS. However, in such methods, massive amount of genetic variants are generated and additional genetic or bio-informatic approaches are often required to identify the candidate variants (Cooper and Shendure, 2011). For example, linkage studies may be carried out in parallel with NGS in a kindred to exclude the variants that are not located within linked regions. Secondly, the deleterious effect of the variant may be predicted according to the location and type of the variant as well as gene functions. This can be achieved by a number of algorithms *in silico* (Kircher et al., 2014). Thirdly, it is important to show co-segregation of a variant with disease and to establish a mutation frequency in different cohorts of cases and controls. Therefore, screening of known mutation in larger cohorts is almost always necessitated.

Sanger sequencing remains the method of choice for the screening of known candidate genes in extended cohorts or novel genes that are strongly implicated in disease. In an effort to define genetic causes of all families in our FALS cohort (**Section 1.19**), we carried out Sanger sequencing on two known ALS candidate genes, *VCP* and *SQSTM1*, in our FALS cohort. Both genes are of particular functional interests as they are implicated in the maintenance of ER proteostasis which is disturbed in ALS (Hetz and Mollereau, 2014). VCP/p97 plays a role in extracting misfolded proteins from the ER to cytosol during ERAD, autophagy and nuclear transport and *VCP* mutations were identified in ALS with or without Inclusion body myopathy, Paget's disease and Frontotemporal dementia (IBMPFD). Some or all of these symptoms were found to be manifest in kindred with *VCP* or *hnRNPA2B1* mutations and, therefore, a term Multisystem Proteinopathy (MSP) has been recently adopted to describe the spectrum of disease (Benatar et al., 2013; Kim et al., 2013). On the other hand, SQSTM1/p62 is involved in targeting the misfolded proteins for autophagic degradation and *SQSTM1* mutations were independently identified in ALS and Paget's disease of bone as well as kindred presenting with both disorders, supporting the notion that ALS and PDB may share a common mechanism.

1.19 Aims of study

(1) In view of the importance of defining causal genes in the classification and understanding of disease mechanism of ALS, we aim to define genetic causes of all families from the Imperial College FALS (IC-FALS) cohort. **(2)** We also sought to identify genetic risk factors and modifiers for FALS and SALS using association studies and **(3)** characterize VNTRs in novel candidate genes for ALS.

1.19.1 Characterizing known genes in the IC-FALS cohort: Sequence analysis and Southern Blot

A total of 208 unrelated kindred have been recruited to this study, predominantly of British descent from the UK. Most of these families have confirmed diagnosis of ALS with DNA available for screening but others have a family history but lack DNA samples. Extensive linkage and FALS mutation screening has been completed with 134 well documented kindred. Investigation of the IC-FALS cohort had previously led to the identification of mutations in *TARDBP* (Sreedharan et al., 2008), *DAO* (Mitchell et al., 2010), *VAPB* (Chen et al., 2010) and *FUS* (Vance et al., 2006). Defining the mutations present in the cohort is important for the understanding of genetic causes of ALS and their prevalence in the UK. Screening for mutations in identified FALS genes in the IC-FALS cohort allowed the mutation frequencies of the following genes to be established: *SOD1* (19.0 %), *TARDBP* (4.5%), *FUS* (2.2%), *VAPB* (0.8%) and *DAO* (0.8%) in **134** kindred. This cohort has more recently also been screened for the *C9orf72* GGGGCC expansion using Repeat-primed PCR method and it was shown that 31.3% of kindred in this cohort are positive for the expansion. Thus, 41.4% of this cohort was of unknown status. In this project, 98 IC FALS kindred were investigated which lacked mutations in *SOD1*, *FUS*, *TARDBP*, *VAPB* or *DAO* and these consisted of 42 *C9orf72* positive kindred identified after the start of this project together with 56 kindred lacking known mutations. Additional kindred were included during the course of the study.

To further clarify mutation frequencies and genotype-phenotype relationships in the IC-FALS cohort lacking mutations in the *SOD1*, *TARDBP*, *FUS*, *VAPB* and *DAO* genes, we carried out sequence analysis of coding regions in two known ALS-causing genes, *VCP* and *SQSTM1.* Mutations in these genes are found in ALS and Paget's disease of bone (PDB) and play roles in the clearance of protein aggregates. Exons harboring most known

mutations, i.e. Exons 1, 2, 3, 5, 6, 7, 10, 13, 14, 17, and the 5' and 3' UTRs, of the *VCP* gene and all 8 coding exons of the *SQSTM1* gene were screened using Sanger sequencing. A novel small hexanucleotide expansion in the 5' UTR of the *VCP* gene and a novel coding mutation in the *SQSTM1* gene were identified. Our results also show the coexistence of ALS and PDB in *SQSTM1* kindred. (**Chapter 3**)

Although the IC-FALS has been screened for the *C9orf72* expansion using the Repeat-primed PCR method, the sizes of expansion need to be characterized. We aimed to optimize a non-radioactive Southern's blot protocol for the detection of the sizes of expansion in FALS patients that are known to be positive, and investigate their relationship with phenotypes of disease. Interim results of this project are presented in **Chapter 7**.

1.19.2 Association studies

Our next aim was to carry out Single nucleotide polymorphism (SNP) association studies to investigate genetic risk factors and disease modifiers that may be present in the IC-FALS and IC-SALS cohorts. The IC-SALS cohort consists of UK ALS patients lacking a family history of ALS collected within the Imperial College Healthcare Trust. Using a candidate-gene-approach, we prioritized SNPs in two genes functioning in protein quality control, *P4HB* and *NPLOC4*, located at the telomeric region of Chromosome 17. The SNPs were genotyped using restriction digest, which were confirmed by sequencing, or Competitive Allele Specific PCR (KASP™) Service carried out by LGC genomics, UK. The SNPs were tested for allele, genotype, haplotype associations and disease phenotype correlation, i.e. age of onset, disease duration and site of onset. Our results indicate that *P4HB* gene is a risk factor and modifier of FALS, whereas the *NPLOC4* gene confers risk of SALS (**Chapter 5**).

1.19.3 Characterization of VNTR length in candidate genes for ALS

The last aim of this study was to investigate novel disease causing genes in ALS. Variable number tandem repeats (VNTR) has been associated with a variety of neurodegenerative conditions including ALS and we investigated whether VNTR length from a group of candidate genes was associated with ALS. 20 VNTRs from 19 genes, known to be expressed in the human spinal cord, located close to SNPs that are associated with ALS

or implicated in the pathogenic mechanism of disease were characterized in this study. Using PCR and electrophoresis, we investigated whether any products indicative of abnormal expansions existed in the IC-FALS cohort. One of the candidates, *HSPB8*, which previously showed altered expression in SALS, was further investigated in the IC-SALS cohort. The effects of the *HSPB8* repeat on gene expression, which has been previously characterized, and disease phenotype was investigated. Our results show no abnormal expansion indicative of repeat instability in FALS. However, it was shown that long alleles in the *HSPB8* repeat are risk factors for SALS. In addition, the allele frequencies of the *NIPA1* repeat, which has been previously shown to be associated with FALS, was further characterized and confirmed in this study (**Chapter 6**).

-END OF CHAPTER 1-

Chapter 2

Methodology and Materials

2.1 Subjects

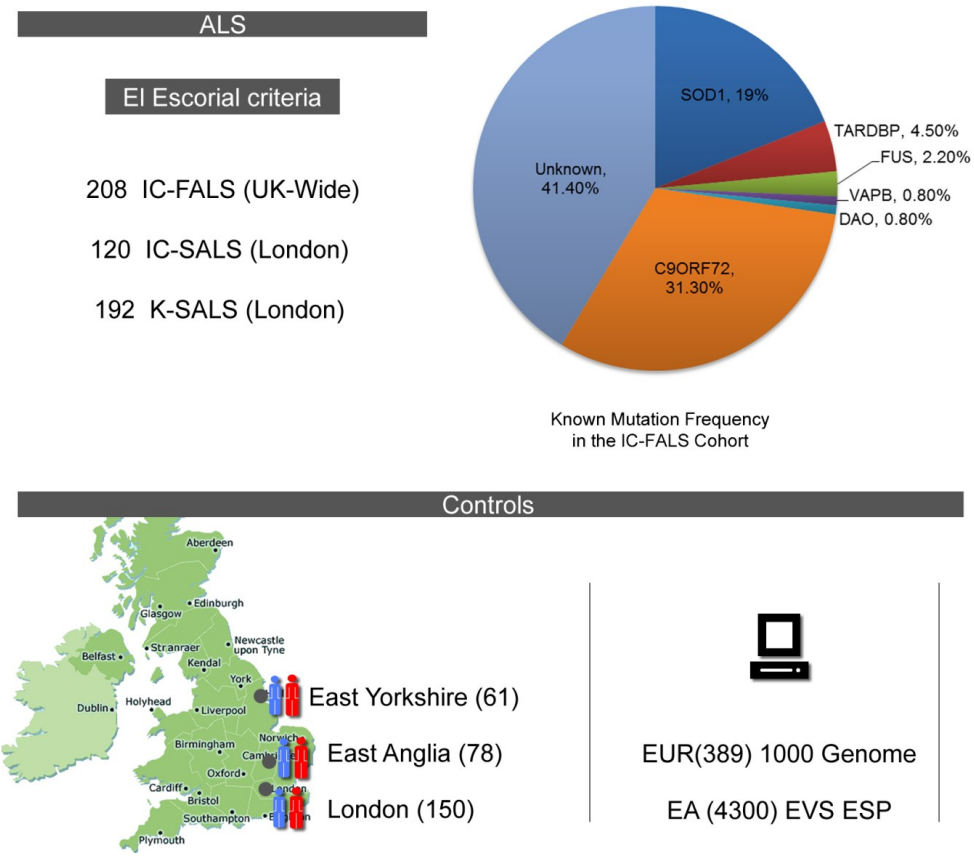
The study population comprise ALS cases and controls recruited from Imperial College NHS Healthcare Trust clinics as summarized in **Figure 2-1**. All ALS patients were diagnosed according to the El Escorial criteria (**Table 1-1**) which requires presentation with both upper motor neuron and lower motor neuron symptoms (Brooks et al., 2000).

As summarised in **Section 1.19**, The IC-FALS cohort contained a total of 208 FALS kindred and full screening for known FALS mutations was complete for 134 FALS kindred. Using DNA from index cases, each from an unrelated kindred, 27.3% of these kindred were positive for mutations in *SOD1*, *FUS*, *TARDBP* and *DAO* genes (**Figure 2-1**). The remaining 72.7% comprised the main body of subjects used for initial mutation analysis (*VCP* gene) performed prior to the discovery of the *C9orf72* expansions. The screening of the *SQSTM1* gene and VNTRs was carried out using *C9orf72*-negative cases (41.4% of characterised cohort of 134 FALS kindred). The *C9orf72* positive cases were used for measuring expansion sizes (**Chapter 7**).

The IC-SALS cohort consisted of 120 Sporadic ALS cases obtained from the Imperial College NHS Healthcare Trust clinics. In addition, a second cohort was used for some studies, the K-SALS cohort, and this contained 192 Sporadic ALS cases obtained from the King's college NHS Healthcare Trust clinics through collaboration with Professor Christopher Shaw and colleagues. These cases were investigated in the *P4HB* and *NPLOC4* association studies (**Chapter 5**). The IC-Control cohort contained neurologically healthy individuals collected from East Anglian (78 individuals), East Yorkshire (61 individuals) and London populations (150 individuals). All controls were of UK ethnicity. Genotype data from the European subgroup (EUR) of the 1000 genome project (*http://www.1000genomes.org/*) included 98 Toscani from Italy (TSI), 93 Finnish from Finland (FIN), 89 British form England and Scotland (GBR), 14 Iberian from Spain (IBR)

and 85 Utah residents with Northern and western European ancestry (CEU); and the European American subgroup from the NHLBI Exome sequencing project (*http://evs.gs.washington.edu/EVS/*) were also included as controls. Stratification tests between controls were carried out as described for each study.

Figure 2-1. Summary of subjects in this study.



Frequencies of known mutations in the IC-FALS cohort are shown in the upper panel. Controls were obtained from the UK and public databases as indicated.

2.2 DNA Extraction

2.2.1 DNA extraction from whole blood or the buffy coat layer

DNA used for genotyping and sequencing was extracted from whole blood or the buffy coat layer, which contains leukocytes, using a QIAamp® DNA Mini Kit (Qiagen, UK) according to protocols provided by the manufacturer. In summary, cells were first lysed to release DNA, 200µl of whole blood/ buffy coat was incubated with 20µl QIAGEN® Protease and buffer AL for 10 minutes at 56^{o}C. The mixture was briefly centrifuged to bring down evaporated drops, after which 200µl 96-100% ethanol was added to the mixture and vortexed for 15 seconds. To bind DNA, the mixture was applied to a QIAmp Mini spin column and centrifuged for 1 minute at 6,000 g (8000 rpm). The filtrate was discarded. A wash was carried out using 500µl of buffer AW1 which was added to the column and centrifuged for 1 minute at 6,000 g. The filtrate was discarded and the column was placed in a new collection tube. 500µl of buffer AW2 was then added to the column and the column was centrifuged for 3 minutes at 20,000 g (14,000 rpm). The filtrate was discarded and the column was placed in a new collection tube. To elute DNA from the filter, 200µl of ddH_2O was added to the column and incubated for 2 minutes, after which the column was centrifuged for 1 minute at 6,000 g.

2.2.2 DNA purification from agarose gels

To isolate DNA from Agarose gels, desired bands were excised using a scalpel or excising tips under UV light and purified using a QIAGEN gel extraction kit (QIAGEN, UK) according to the Manufacurer's instructions. To do this, the excised gels were incubated with 3 times gel volumes (1g = 100µl) of solubilisation buffer QG for 10 minutes at 50oC. After the gels are completely dissolved, 1 gel volume of isopropanol was added to the mixture. The mixture was vortexed and applied to the QIAquick column and centrifuged for 1 minute at 17,900 g. The filtrate was discarded and, for washing, 750µl of washing buffer PE was added to the column, which was then centrifuged for 1 minute at 17,900 g. The filtrate was discarded and the column was centrifuged for 1 minute at 17,900 g again to remove residual washing buffers. To elute DNA, the column was placed in a new collection tube. 50 µl of ddH_2O was added to the column, incubated for 2 minutes and centrifuged for 1 minute at 17,900 g.

2.2.3 Total RNA extraction from blood

To investigate whether DNA variants affect expression or cause alternative splicing, we performed RT-PCR using RNA extracted from patient's whole blood or buffy coat samples, which were stored at -80^{o}C. RNA was extracted using TRI Reagent® (SIGMA, UK) and Direct-zol ™ RNA MiniPrep Kits (ZYMO Research, USA).

For each extraction, 100µl of whole blood/ buffy coat was mixed with 300µl of TRI Reagent®, vortexed and incubated for 5 minutes at room temperature. The mixture (400µl) was then mixed and vortexed with 400µl of absolute alcohol and loaded onto the Zymo-Spin IIC Column, which was placed in a collection tube. The tube was centrifuged for 1 minute at 16,000 g, allowing RNA binding to the column. Filtrate was discarded. To pre-wash, 400µl Direct-zol RNA PreWash was added to the column and centrifuged for 1 minute at 16,000g. Filtrate was discarded and the pre-wash was repeated once more. To wash, 700µl of RNA Wash Buffer was added to the column and centrifuged for 1 minute at 16,000 g. Filtrate was discarded. After that, the tube was centrifuged for an additional 2 minutes to ensure removal of washing buffers. To elute RNA, the column was placed into a new tube and 25µl of RNase-free water was added, incubated for 2 minutes and centrifuged for 1 minute at 16,000 g.

The eluted RNA was immediately used for measuring of concentration and reverse transcription. Unused RNA was stored at -80^{o}C.

2.3 Polymerase chained reaction (PCR)

Desired DNA fragments were amplified from extracted genomic DNA for genotyping using PCR. PCR Primers flanking ~100 base pairs at either end of the target sequence were designed using the PRIMER 3 (*http://frodo.wi.mit.edu/primer3*) program. The optimum length of primers and PCR products used were 18 to 22 bps and less than 700 bps respectively. No more than 4 identical consecutive nucleotides were allowed in the primers and the maximum difference in melting temperatures (Tm) between forward and reverse primers was 5^{o}C. Purchased Primers (Invitrogen, UK; SIGMA, UK) were diluted to a stock concentration of 10µM and stored at -20^{o}C. The details of primers are specified in the methodology of each project.

Table 2-1. Preparation of PCR reactions.

Reagents	Stock Concentration	Final Concentration	Volume Needed for Single Reaction.
Buffer	5X	1.00 X	6.00µl
Mg^{2+}	25 mM	1.50 mM	1.80µl
dNTPs	10 mM	0.10 mM	0.30µl
Forward Primer	10 µM	0.50 µM	1.50µl
Reverse Primer	10 µM	0.50 µM	1.50µl
PTq	5 U/µl	0.02 U/µl	0.12µl
Template	10ng/µl	0.5 ng/µl	1.5µl
Water			17.28µl
Total Volume			*30.00µl*

Components of reagents used for a standard single PCR reaction. The stock dNTP mix contained 10mM of each nucleotide: dATP, dCTP, dGTP, dTTP.

PCR reactions were carried out in a standard 30µl solution containing 1X buffer, 1.5mM Mg^{2+}, 0.1mM dNTP, 0.5µM forward primers, 0.5µM reverse primers, 0.05U/µl GoTaq® DNA Polymerase (Promega, UK) and 0.5ng/µl templates (**Table 2-1**).

A standard 35 cycles was used for most reactions. Annealing temperature (T_{ann}) was taken as being 5^{o}C below then melting temperature (T_m). PCR results were visualized by separating an aliquot (8µl) of reaction mix in 2% agarose gels. If further optimisation was required, temperature gradient PCRs were carried out, in which 12 reactions were cycled under different T_{ann} ranging from 50 to 65°C.

PCRx Enhancers System (Invitrogen, UK) was used in optimizing unsuccessful PCRs. 7-deaza-dGTP (THERMOPOL®, New England Biolabs, UK), which lowers melting temperatures of the G-C duplexes, was used for the amplification of GC rich regions.

PCRs were carried out in thermal cyclers (Techne, UK) using the following cycles:
Initial denaturation: 94°C for 5 minutes. For each cycle : Denaturation step (94°C) for 30 seconds, Annealing step (T_{ann}°C) for 30 seconds, Extension step (72°C) for 45 seconds
Final elongation: 72°C for 5 minutes.

2.4 cDNA Synthesis and Endpoint RT-PCR

To perform RT-PCR, cDNA was synthesized from RNA using Cloned Avian Myeloblastosis Virus (AMV) Reverse Transcriptase (AMV-RT) immediately after RNA extraction, after which the cDNA was used as template for an ordinary PCR reaction as described in **Section 2.3**. Random hexamers were used as primers for the reverse transcription. The Cloned AMV First-Strand cDNA Synthesis Kit (Invitrogen, UK) was used.

For each reaction, 8µl (~24ng, after measurement) of RNA was mixed with 1µl of random hexamers (50ng/µl), 2µl of dNTP (10mM) and 1µl of RNase-free water, giving a final volume of 12µl. The mixture was then incubated for 5 minutes at 65^{o}C for denaturation. After that, the tube was vortexed, placed on ice and mixed with a separate 8µl mixture containing 4µl of cDNA synthesis buffer (5X), 1µl of DTT (0.1M), 1µl of RNaseOUT^{TM} (40U/µl), 1µl of RNase-free water and 1µl of Cloned AMV-RT (1µl). Next, the mixture was placed in a preheated thermal cycler and incubated for 10 minutes at 25^{o}C, followed by 50^{o}C for 50 minutes. Reaction was then stopped by heating for 5 minutes at 80^{o}C and proceeded for ordinary PCR reaction or stored at -20^{o}C.

2.5 DNA Sequencing

2.5.1 DNA Purification

Prior to sequencing, DNA was purified from the PCR mixture using SureClean (Bioline, UK) or an Exonuclease I- Shrimp Alkaline Phosphatase (Exo-SAP) method to remove proteins, primers and dNTPs from the mix.

SureClean is a column-free protocol allowing the precipitation of nucleic acids ≥ 75 bp. For each clean-up, 20µl of SureClean was added to 20µl of PCR products and incubated for 10 minutes. The mixture was centrifuged for 40 minutes at 4000 rpm and supernatant was removed. Then the sample was resuspended and participated with 40µl of ethanol, centrifuged for 40 minutes at 4000 rpm again, after which alcohol was removed by centrifuging the tube/ plate upside down for 1 minute at 400 rpm. Finally, the platelet was resuspended in 10µl of H_2O and ready for sequencing.

In the Exo-SAP method, excess primers were removed by exonuclease I and dNTPs were removed by shrimp alkaline phosphatase. For each clean-up, 10µl of Exo-SAP mix containing 0.025µl exonuclease I, 0.25µl of shrimp alkaline phosphatase and 9.725µl of distilled water was added into 30 µl of PCR reaction mix. The 40 µl mixture was incubated at 37°C for 30 minutes and then 95°C for 5 minutes in a PCR machine. A 10 µl aliquot was then withdrawn for sequencing and the remainder was stored at -20°C.

2.5.2 DNA Sequencing

DNA sequences were obtained using traditional Sanger sequencing method, which has been reviewed in **Section 1.18**. The reactions were carried out using the automated 3730xl DNA Analyzer system (Applied Biosystems, UK) at the Imperial College genomics laboratory (*http://genomics.csc.mrc.ac.uk/*). The sizes of PCR products used for sequencing were optimized to 200-600 bps. For each sequencing reaction, 10µl of cleaned PCR products, as described above, was mixed with 1µl (6.4 pmol) of forward primer or reverse primers for sequencing. The chromatograms were visualized using Codon Code Aligner (*http://www.codoncode.com/aligner/*) and Seqdoc program (*http://research.imb.uq. edu.au/seqdoc/*) (Crowe, 2005).

2.6 Restriction digest

In addition to DNA sequencing, genotypes of DNA variants can be determined using restriction fragment length polymorphism (RFLP) analysis. Restriction endonucleases, or restriction enzymes, are enzymes that cleave DNA at or near a specific sequence recognized by the enzyme, also known as restriction sites. When an allele of a DNA variant creates a restriction site with flanking sequences, it is possible to tell whether this allele is present using enzymes recognizing that site. The cleaved products can be visualized in agarose gel electrophoresis.

Restriction enzymes, identified using NEBCUTTER (*http://tools.neb.com/NEBcutte r2*), discriminating different alleles of DNA variants were used for Single nucleotide variant (SNV) genotyping in this study. The desired fragments was first amplified in PCR (**Section 2.3**), after which a 10 µl mixture containing 1X Buffer, 1 to 3 U restriction enzyme and distilled water was added to digest each 30 µl (<1 µg) of PCR products. Compatible

buffers and digesting temperatures were chosen according to instructions by the manufacturer (New England Biolabs, UK).The addition of 100 µg/ml of Bovine Serum Albumin (BSA) was required for some enzymes. Conditions used for each digest will be specified in the methodology sections of the project. The digestion of genomic DNA used for Southern's blot will be mentioned in **Section 2.10**.

2.7 Electrophoresis

2.7.1 Agarose gel electrophoresis

The casting of agarose gels and performing electrophoresis were carried out using routine laboratory practices for visualizing PCR and restriction digest results. Agarose is a polysaccharide polymer material that crosslinks to form matrices containing channels that allow the passage of DNA. When electric field is applied, the negatively charged DNA molecules in the agarose gel migrate to the positively charged anode with a rate proportional to their sizes. The optimum percentage [w/v] of agarose used depends on the sizes of fragments to be analysed (**Table 2-2**). Ethidium bromide, a DNA dye that becomes fluorescent under UV lights, was used for the staining of DNA, and, after electrophoresis, the DNA fragments can be visualized in a UV transilluminator.

To prepare a 2% agarose gel, which was routinely used in this study, 11g of agarose powder (ELECTRAN® , VWR, UK) was dissolved in 500ml of 0.5X TBE buffer and microwaved for 7 minutes. The mix was cooled down in water and 10µl of Ethidium Bromide was added when the temperature decreased to 60^{o}C and mixed. The mix was then poured into the gel casting tank and remained in room temperature for 1 hour.

Samples (100-1000 bps) were then loaded and electrophoresied at 210V (7V/cm) for 1 hour. 10µl of ethidium bromide was added to the anode buffers prior to electrophoresis. A gel documentation system (Geldoc) was used to visualize the result. PhiX-174-HaeIII (Thermo scientific, UK) was used as the DNA marker in the standard gel.

Table 2-2. Concentration of agarose gels used for the separation of DNA fragments of different sizes as summarized by (Barril, 2012).

% Agarose [w/v]	Fragment Size (bp)
0.2	5000-40000
0.4	5000-30000
0.6	3000-10000
0.8	1000-7000
1	500-5000
1.5	300-3000
2	200-1500
3	100-1000

To discriminate between small differences in fragment sizes, a long electrophoresis was carried out using 2.8% agarose gels. This method was able to detect down to 3 bps difference in fragment sizes and has been applied to analysis of short tandem repeats or indels in this study. To prepare the gel, 14g of agarose powder was microwaved at 800W for 10 minutes in 500 ml TBE. 10µl of ethidium bromide was added when the gel was cooled down to 60^{o}C and 10µl of ethidium bromide was added in the anode buffer prior to electrophoresis. Electrophoresis was carried out at 140V (4.5V/cm) for 16 hours. A 10 bp DNA ladder (Invitrogen, UK) was used as a marker for long-electrophoreses.

2.7.2 Denaturing Polyacrylamide Gel Electrophoresis (PAGE)

While larger DNA fragments can be resolved in agarose gels, polyacrylamide gels enable better resolutions down to single base pair differences in size. Acrylamide gel is a polymerized matrix containing acrylamide and bisacrylamide molecules. The polymerization of acrylamide takes place in the presence of bisacrylamide, a cross-linking agent that binds two acrylamide molecules when initiators, such as ammonium persulphate and TMEMD, are added. The ratio of acrylamide to bisacrylamide is typically 19:1 and different concentrations of acrylamide are selected depending on fragment sizes (**Table 2-3**). A denaturing polyacrylamide gel, in addition, also contains a DNA denaturant that keeps the DNA single-stranded during electrophoresis and avoids the formation of heteroduplexes. Therefore, compared to agarose gel electrophoresis, a denaturing PAGE

was preferably used for the accurate sizing of short tandem repeats and small indels and urea was used as the denaturant in this study.

Table 2-3. Formulations for commonly used gel percentages and the amount of solutions used for each 50ml gels (adopted from national diagnostics).

% Monomer [w/v]	Fragment Size (bp)
4	>200
5	80-200
6	60-150
8	40-100
12	10-50
20	<20

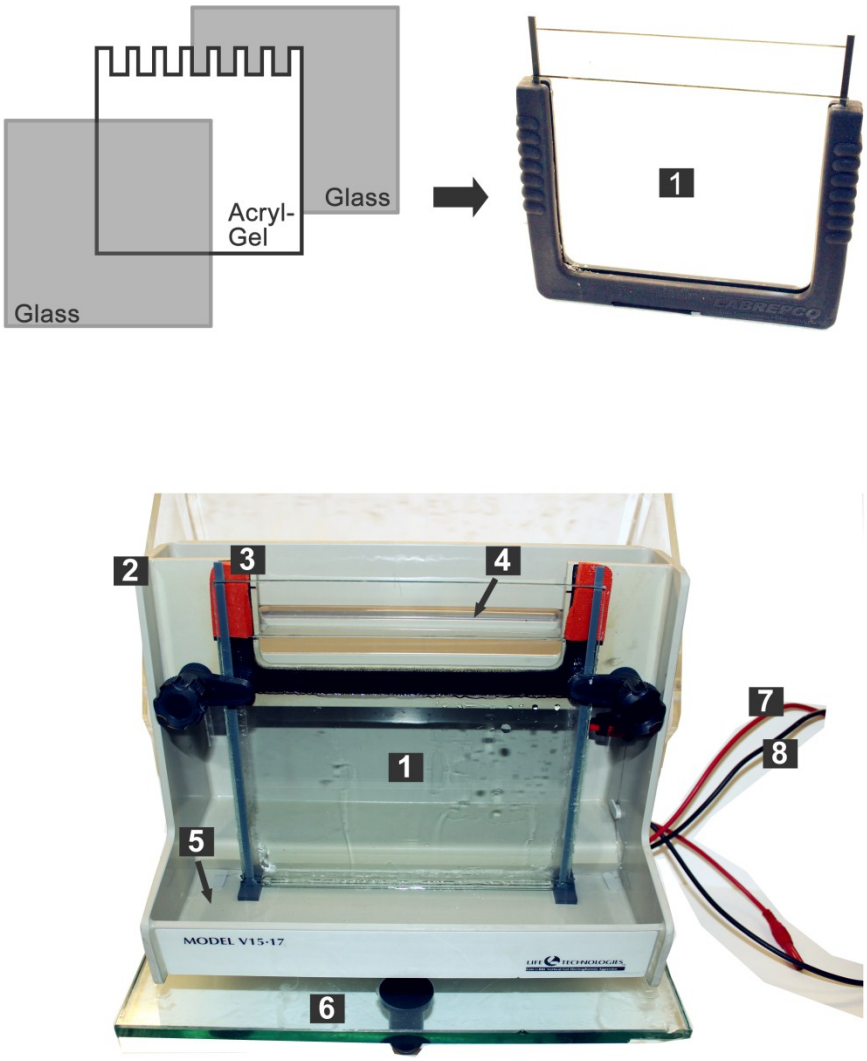
In this study, gels were cast using a 50ml 4% Urea polyacrylamide gel (SequaGel UreaGel System, National Diagnostics, UK). For each gel, 8ml of UreaGel Concentrate (237.5g/L acrylamide, 12.5g/L methylenebisacrylamide, 7.5M urea), 37ml of UreaGel Diluent (7.5M urea) and 5ml of UreaGel Buffer (10X TBE and 7.5M urea) were mixed in a measuring cylinder. After that, 20µl of TMEMD and 400µl of freshly prepared 10% ammonium persulphate was added into the mix. The cylinder was then sealed with parafilm and mixed by turning upside down. The mix was then poured into the 1mm space between the glass plates of the casting setup as shown in **Figure2-2, 1**. The gel was left overnight for polymerization.

Prior to loading, the samples were denatured using a formamide loading buffer containing 95% formamide (Applied Biosystems, UK), 0.9µg/ml Xylene Cyanol and 0.005M EDTA. Equal volume of samples and loading buffer were mixed and incubated at 99^{o}C for 10 minutes, and then chilled on ice. The PAGE apparatus was placed on a leveller. Both anode and cathode chambers of the PAGE apparatus were filled with 1XTBE. The gel sandwich was placed into the PAGE apparatus (**Figure 2-2, 2-8**) and pre-electrophoresed at 15W for 30 minutes, which allowed the gel to be warmed up to ~50^{o}C. Blow out the wells, which were filled with urea, of the pre-heated gel with syringe and needle.

After that, depending on the DNA concentration, 1-10µl (100-1000ng) of the loading mix was loaded into each well. Samples were kept on ice when loading. The gel was electrophoresed at 15W for 3 hours and 45 minutes to give the maximum resolution of 200-300bp products. The time needed can be estimated by the migration of Xylene cyanol, whose migration rate is approximately equivalent to that of a 155bp fragment in a 4% gel.

When finished, the setup was dismantled and the gel was detached from the glass plates by soaking in distilled water. The gel was placed in a UV-transparent container. 50ml of 1X TBE and 5µl of 10,000X SYBR Gold Nucleic Acid Gel Stain (Invitrogen, UK), making a 1X solution, were added into the container. The gel was incubated in the dark for 30 minutes and visualized in a UV trans-illuminator.

Figure 2-2. Setup of Polyacrylamide gel electrophoresis (PAGE).



1: Assembly of the acrylamide gel sandwich; **2**: the PAGE apparatus; **3**: A water-resistant sponge isolating the cathode chamber; **4**: Cathode buffer (1X TBE); **5**: Anode buffer (1X TBE); **6**: leveller; **7**: Anode cable; **8**: Cathode cable.

2.8 SNP Genotyping using Kompetitive Allele Specific PCR (KASP™) assay

KASP is a commercially available service allowing SNP genotyping based on competitive allelic specific PCR. The modified PCR reaction utilizes a reverse primer and two forward primers that are labelled with different fluorescent dyes (FAM, HEX or ROX). Each forward primer is complementary to one allele of a SNP and is initially bound to a quencher that inactivates fluorescence signals. Upon denaturation, the labelled primers are released from the quenchers and take part in PCR reactions that exponentially generate fluorescent signals representing different genotypes. Genomic DNA was prepared following the specifications of sample quantity and packaging, and genotyping was carried out by LGC genomics, Middlesex, UK, (*http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/*).

2.9 Fragment analysis

Fragment analysis is a well-established method for analysing microsatellites and has been widely applied in linkage analysis and DNA fingerprinting. In principle, DNA fragments are amplified using a fluorescently labelled primer and separated in capillary electrophoresis. Genotypes of the microsatellite are then discriminated by the relative positions of fluorescent signals in the electropherogram. In this study, fragment analysis was performed on a dinucleotide TG repeat in the *NPLOC4* gene (**Section 5.3**). The electrophoresis was carried out using an ABI 3730x1 DNA Analyzer at the Imperial College Genomic Core Facility.

To prepare samples, we first optimized PCR conditions, as described in **Section 2.3**, using a 6-FAM labelled forward primer. 1µl of PCR products and 10µl of 95% Hi-Di Formamide (Applied Biosystems,UK) were aliquot and mixed in a well of 96-well PCR plate. The plate was wrapped in aluminium foil and sent for analysis. Electropherograms were visualized in GeneMapper® software v4.1.

2.10 Southern Blot

Southern blotting is a technique used to detect DNA of a specific sequence in a sample. The technique can be used to analyse DNA from different sources such as PCR products,

cloned fragments and genomic DNA. In principle, the samples are first fragmented into desired fragments using restriction digest and separated by size in agarose gel electrophoresis. DNA molecules in the gel are then denatured and transferred onto a nylon membrane in the presence of capillary force, which can be created by a stack of paper towels. After that, the membrane, on which the DNA fragments attach, is hybridized with an oligonucleotide DNA probe complementary to the desired sequence and labelled with materials that can be detected in autoradiography (32P) or chemiluminescent assays (Digoxigenin). Southern blot protocols using both detection methods are summarised below.

2.10.1 Sample preparation and electrophoresis

DNAs used for southern blot were extracted from whole blood or lymphoblastoid cell line as described in **Section 2.2.1**. 5-30µg of DNA was concentrated to a volume of ~30µl using a centrifuging concentrator (Eppendorf 5301) or Ethanol precipitation. To perform ethanol precipitation, 1/10 volume of 3M Sodium acetate, pH 5.2, was added to the sample. The tube was vortexed and 2 to 3 volumes of 100% ethanol was added, followed by vortexing. The mixture was placed on ice for 20 minutes, followed by centrifugation for 15 minutes at maximum speed. After that, supernatant was removed and 1 volume of 70% ethanol was added into the tube. The tube was centrifuged briefly and supernatant was removed. The tube was airdried for 3 to 5 hours at room temperature and pellet was reconstituted to desired volume using distilled water.

50U of XbaI, or with 75U of EcoRI (New England Biolabs, UK), buffer and water were added, making a final volume of 50µl and incubated at 37^{o}C overnight. After adding 5µl of loading dye samples were loaded into every other lane in an 8mm-thick 0.8% agarose gel. The gel was 14.5cm x 12cm and we used a 7mm wide comb (12 samples per gel) so the 50µl samples can be fully contained.

Different markers were used depending on the detection method. For chemiluminenescent assay, 8µl of DIG labelled DNA Molecular Marker III (10ng/µl) and VII (10ng/µl) (Roche, UK) were mixed with 3µl of 6X loading dye and loaded into the first two lanes respectively. For radioactive detection, 1µl of 1kb plus DNA ladder (Stock: 250ng/µl) (Invitrogen, UK) was used. The gel was then electrophoresed in 1X TBE buffer

at 30V (1.5V/cm) for 16 hours. Ethidium bromide was added to the anode before the run. This is suitable for an appropriate migration of bands >1kb in a 14.5 cm gel. For a better separation, a voltage of 3-4V/cm can be used but bands <2kb would not be detectable. An image was kept after the run.

2.10.2 Preparation for DNA transfer

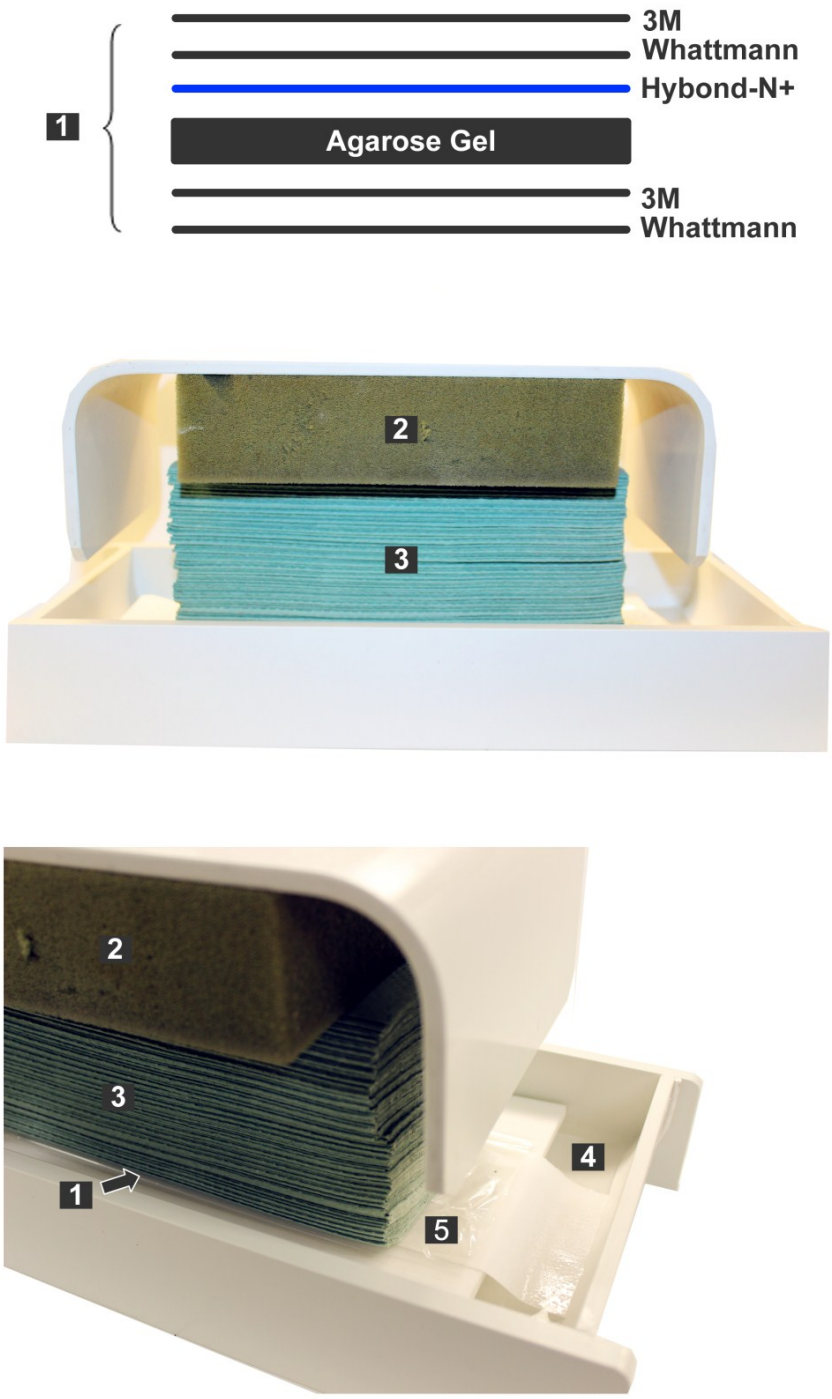
Depurination was an *optional* step allowing random DNA fragmentation and was used for the transfer of larger (>3kb) DNA fragments. The gel was placed in large volumes of depurination solution (0.25M HCl) and rocked for 10 minutes. The gel was rinsed 3 times in distilled water afterwards. Bromophenol blue (BPB), which turns yellow in acidic solutions, can be used as a pH indicator for this process.

The gel was then placed in large volumes (~500ml) of denaturation solution, which contained 1.5M NaCl and 0.4M NaOH, for 30 minutes on a rocker mixer. Reagents were changed every 15 minutes. After this, the gel was neutralized in neutralization solution, which contained 1.5M NaCl, 1M Tis-HCl (pH7.4) and placed in ~500ml of neutralization solution for 15 minutes on the rocker mixer.

2.10.3 Assembling the blot

In this stage the denatured DNA was transferred onto a positively charged nylon membrane (Hybond-XL). First, the reservoir of the blotting apparatus was filled with 20XSSC. To prepare the wick, trim a piece of 3M Whattmann filter paper, place it on the plastic platform and soak both ends in the buffer. After that, to assemble the gel sandwich, trim 4 sheets of 3M Whattman into the size of the gel, soak in 20XSSC and place 2 of them on the wick. Place the gel on the Whattman papers with the flat surface (sample wells) facing up. Trim a piece of Hybond-XL and and carefully place it on the gel without moving after making contact and place the other 2 sheets of pre-wet Whattman papers on top of the Hybond-XL. In order to prevent direct contact between the wick and the paper towels, surfaces surrounding the gel were covered using saran wrap. A pile, approximately 10 cm height, of paper towels was placed on the gel sandwich. Finally, a ~500g weight was placed on the pile of paper towels. The transfer was left undisturbed overnight. Details of the assembly are shown in **Figure2-3**.

Figure 2-3. Setup of Southern transfer.



1: The gel sandwich; **2**: A 500g weight; **3**: Stack of paper towels; **4**: Transfering buffer (20X SSC); **5**: Saran wrap preventing direct contact between the paper towl and the wick.

2.10.4 Dissembling the blot and preparation for hybridization

The blot was dismantled after overnight transfer. The paper towels and saran wrap were removed and the gel sandwich was placed on clean filter paper with the membrane side facing down. The gel was kept in place, the well positions on the membrane were marked with a pencil and the gel was removed. The successfulness of the transfer was confirmed by visualizing the gel in a UV transilluminator confirming no DNA was left in the gel and the dye could be seen on the membrane indicating a successful transfer.

Next, the membrane was washed in 2X SSC at room temperature for 5 minutes on the rocker mixer. The membrane was placed with the DNA side facing up, on a clean filter paper to air dry for 15 minutes. The membrane was placed in a UV cross-linker and cross-linked at *120,000 microjoules*/cm^2. At this stage, the membrane could be stored at 4^oC in a plastic folder or used for hybridization.

2.10.5 Hybridization using Digoxigenin (DIG) labelled probes

During hybridization, digoxigenin (DIG) labelled DNA probes were denatured and incubated with the membrane. The DIG-probes were synthesized using a PCR DIG Probe synthesis kit (Roche, UK), in which the probes were synthesized in a PCR reaction using DIG-labelled dUTP and a cloned plasmid DNA as template. The reactions were optimized using a dNTP mix with DIG-dUTP:dTTP ratio of 1:6, which can be simply prepared by mixing equal volumes of vial 2 (ratio of 1:3) and vial 4 (unlabelled dNTP) in the Kit. The labelling reactions were carried out as following described in **Table 2-4** and the yield was checked by electrophoresis. Generally, migration of labelled probes is slower than the unlabelled products and the yield is decreased when a high DIG-dUTP:dTTP ratio is used. To prepare the hybridization buffer, 64ml of distilled water was added into a bottle of DIG Easyhyb Granules (Roche, UK) and dissolved by incubating at 37^oC for 5 minutes. The Easyhyb was then preheated to the hybridization temperature (T_{opt}), which can be calculated using the following equations, where Tm is the melting temperature and l is the length of the oligonucleotide:

$$T_m = 40.82 + 0.41(\%G+C) – (600/l)$$

$$T_{opt} = T_m - (20\sim25^oC)$$

The hybridization temperature used was 48°C. The membrane was placed in a glass bottle, 8ml of Easyhyb and 2ml of Salmon Sperm DNA (Roche, UK) was added and the membrane was incubated at 48°C for 3 hours in a rolling hybridization oven. This pre-hybridization step blocks the membrane for non-specific bindings. Alternatively, the incubation can be carried out in a sealed plastic bag, which increases buffer contact and prevents drying of the membrane.

Table 2-4. Preparation of PCR reactions for the synthesis of DIG-labelled Probes.

Reagents	Stock Concentr-ation	Final Concentr-ation	Volume needed for Single Reaction
Water	-	-	29.25µl
PCR Buffer with Mg^{2+} [vial3]	10X	1X	5µl
1:6 DIG- dNTP mix	2mM	200µM	5µl
Forward Primer	10µM	0.5µM	2.5µl
Reverse Primer	10µM	0.5µM	2.5µl
Polymerase [vial 1]	-	-	0.75µl
Template	10ng/µl	0.5 ng/µl	5µl

The concentrations of the stock 1:6 DIG-dNTP mix were 2mM dATP, 2mM dCTP, 2mM dGTP, 1.65mM dTTP and 0.35mM dUTP. Purified plasmid DNA containing a fragment of the probe was used as template. The PCR reactions were carried out at an annealing temperature of 55°C.

At the end of the pre-hybridization step, the DIG-probe was denatured by incubating at 95°C for 10 minutes and rapidly cooled on ice. The hybridization buffer was poured into a clean falcon tube and mixed with the DIG-probe. The volumes of probe needed depend on the size of probe and the yield of labelling. In general, 2 to 10µl of probe (~200ng) is needed. The mix was then hybridized with the membrane at 48°C for 16 hours.

Following the hybridization step, the hybridization buffer was discarded, or stored at -20°C and stringency washes were carried out to remove non-specific binding. The membrane was treated, with constant agitation, as follow:

- Low stringency wash: 2XSSC, 0.1%SDS at room temperature for 15 minutes.

- Low stringency wash: 2XSSC, 0.1%SDS at 68^{o}C for 15 minutes.
- High stringency wash: 0.1XSSC, 0.1%SDS at 68^{o}C for 15 minutes.
- High stringency wash: 0.1XSSC, 0.1%SDS at 68^{o}C for 15 minutes.

The high temperature steps were carried out in the rolling hybridization oven and temperature was monitored at 15 minute intervals. The chemiluminescent detection washes are carried out immediately after the washes.

2.10.6 Chemiluminesence detection

The DIG-labelled probes were detected by anti-DIG monoclonal antibodies conjugated with alkaline phosphatase (AP). AP catalyzes its substrates for the emission of visible light, which can be captured by X ray films. The membrane was washed and blocked using the DIG Wash and Block Buffer Set (Roche, UK) and incubated with Anti-Digoxigenin-AP Fab Fragments (Roche, UK) as follows:

- Rinse membrane in 1X Washing Buffer (DIG wash set) at room temperature for 5 minutes.
- Blocking: This step blocks non-specific binding sites for the antibody. Incubate the membrane in 100 ml 1X blocking solution for 30 minutes. To prepare the solution, dilute the maleic acid buffer (10X stock) to 1X with distilled water and dilute the blocking solution (10X stock) to 1X using the 1X Maleic acid buffer.
- Prepare Antibody solution: Centrifuge the antibody solution, which is ready to use, for 5 minutes at full speed. Use 1µl of antibody per ml (37.5mU/ml). Mix 40ml of the 1X blocking solution with 4µl of antibody taken from the surface.
- Wash the membrane in the antibody solution at room temperature for 30 minutes on the rocker mixer.
- Discard the antibody solution and wash the membrane in 1X DIG washing buffer for 2 x 15 minutes at room temperature on the rocker mixer.
- Take out the membrane and equilibrate in in 20 ml 1X detection buffer. Place the membrane in a plastic folder with DNA side up.
- Add 0.5-1ml (20 drops) of CDP-STAR (Roche, UK). Close the folder and incubate at room temperature for 5 minutes.
- Finally, squeeze out excess liquid and expose to X ray film for 15 minutes to 6 hours depending on the strength of signal.

The membrane should be kept wet during all procedures to control backgrounds. The membrane can be stripped and re-probed after washing in Stripping solution (0.2M HCl, 0.1% SDS) at 37^{o}C for 2 x 20 minutes under constant agitation. The preparations of all stock solutions are listed in **Table 2-7**.

2.10.7 Dot Blot

Dot blot was carried out to test and optimize for the hybridization conditions of Southern's blot. To do this, DNA samples were denatured in NaOH manually applied onto positively charged nylon membrane:

- Prepare a serial dilution of DNA templates to be blotted.
- Mix 1 µl of template, 1.45 µl of 1M NaOH and 0.5 µl of 200mM EDTA pH 8.2 in a microcentrifuge tube, giving a final volume of 2.4 µl mixture containing 0.4M NaOH,
- Heat denature the mixture at 95^{o}C for 10 minutes, and immediately place it on ice.
- Cut a piece of Hybond N+ membrane (GE) and slowly apply the samples (2.4 µl) onto the membrane. A dye can be used to aid visualizing the samples.
- UV crosslink the membrane.
- The membrane can then be processed for pre-hybridization and hybridization (**Section 2.10.5**).

2.11 Statistics

Statistical methods used for each study will be specified in each of the following chapters. In general, the power of each study was calculated using G*Power, which calculates power based on w (effect sizes):

$$w = \sqrt{\sum \frac{(p_0 - p_1)^2}{p_0}}$$

,where w is the effect size. p_0 and p_1 are allele frequencies in unaffected and affected cases respectively. Power of study can then be determined as following (**Table 2-5**).

Table 2-5. A combined power heatmap with table of effect size w-Values, for different Case/Control MAFs at a=0.05, and a total sample size of 270 subjects.

	Increase of MAF in Cases		
Control MAF	7%	8%	10%
50%	0.14	0.16	0.2
40%	0.143	0.163	0.20
30%	0.153	0.175	0.22
20%	0.175	0.2	0.25
10%	0.233	0.267	0.33

Power <70%
Power >73%
Power >85%

For SNP analysis, departure from Hardy-Weinberg Equilibrium (HWE) was tested by comparing observed and expected genotype frequencies using 2 x 3 Fisher's test with 1 degree of freedom. 2 x 2 Fisher's test, which was used for testing allelic association, was implemented using PLINK Program. 2 x 3 Fisher's test, which was used for testing genotypic associations, was implemented using the *fisher.test()* function in R (**Appendix I**). Models of associations were determined by comparing disease penetrance between individuals of different genotypes (**Table 2-6**) (Clarke et al., 2011).

Table 2-6. Penetrance functions used for determining model of associations (Modified from Clarke et al 2011).

Model	Penetrance			Relative Risk	
	d/d	*d/D*	*D/D*	*d/D*	*D/D*
Addtive	f_0	$f_0\gamma$	$2f_0\gamma$	γ	2γ
Recessive	f_0	f_0	$f_0\gamma$	1	γ
Dominant	f_0	$f_0\gamma$	$f_0\gamma$	γ	γ
Multiplicative	f_0	$f_0\gamma$	$f_0\gamma^2$	γ	γ^2

Disease penetrance functions for different genotypes *d/d, d/D and D/D* are shown. *D* is the associated allele and the non-associated genotype ***d/d*** has a baseline disease penetrance f_0 = 0. The right panel shows the relative risk of disease for genotypes *d/D* and *D/D* compared to wild type genotype *d/d,* where the genetic penetrance parameter is γ>1.

Linkage disequilibrium (LD) and haplotypes were visualized using Haploview (Barrett et al., 2005), which adopts the Partition-ligation-expectation-maximization (PL-EM) algorithm for the reconstruction of Haplotypes. r^2 value was used as the measurement of LD (Lewontin, 1964). Haplotype phasing was performed using PLINK and haplotypes with the best posterior probability were chosen. Each of the haplotypes was tested for associations with disease using 2 x 2 Fisher's tests. A mode-free permutation omnibus test for haplotype association was carried out using CLUMP Program.

All tests were corrected for multiple testing using Benjamini- Hochberg correction, which was implemented using the *p.adjust()* function in R.

For genotype- phenotype correlations, allele frequencies were compared in patients with different site of onsets, i.e. bulbar onset or limb onset. Survival analysiswas performed using Kaplan-Meyer curves which were compared using a log-rank test, implemented using SPSS and Graphpad.

2.12 Stock solutions

Table 2-7. Preparation of stock solutions used in this study.

Solutions	Volume	Components
Stock solutions for electrophoresis		
1X TBE	20L	Tris Base, 216g Boric Acid, 110g 0.5M EDTA pH 8, 40ml Distilled Water, to 20L
Denaturing Solution for PAGE	10ml	95% Formamide, 9.9ml Xylene Cyanol, 9µg 0.5M EDTA pH 8, 100µl
0.5M EDTA pH 8	500ml	84.05g EDTA 400ml H_2O Titrate with 5M NaOH until pH=8.0, and bring up volume to 500ml.
Stock solutions for Southern's Blot		
20X SSC	2L	Sodium Chloride, 350.6g Sodium Citrate, 176.4g Titrate with HCl until pH=7.
2X SSC	1L	20X SSC, 100ml Distilled water, 900ml
Depurination Solution	2L	37% HCl, 41.39ml Distilled water, 1958ml
Denaturation Solution	2L	NaOH, 32g NaCl, 175.32g Distilled water, bring up to 2L

Neutralization Soluction	2L	Tris, 242.28g NaCl, 175.32g Distilled water, bring up to 1.5 L Titrate with HCl to pH 7.4 and bring up to 2L
Pre-Hybridization Solution	10ml	DIG-Easyhyb, 8ml Salmon Sperm DNA, 2ml
Low Stringency Wash Solution	1L	2x SSC, 1L SDS, 1g
High Stringency Wash Solution	1L	2x SSC, 50ml Water, 950ml SDS, 1g
Stripping Solution	1L	NaOH, 8g SDS, 1g Distilled water, 1L

-END OF CHAPTER 2-

Chapter 3

Sequence Analysis of *VCP* Gene in the Imperial College ALS Cohort

3.1 Introduction

VCP mutations were first identified in ALS pedigrees by Johnson et al (Johnson et al., 2010) using exome sequencing but were already known to be a cause of an autosomal dominant form of Inclusion body myopathy with Paget's disease of bone and fronto-temporal dementia (IBMPFD), also characterized by TDP-43 inclusions (Weihl et al., 2008) as seen in ALS cases with *VCP* mutations (Johnson et al., 2010). In fact, a family history of cognitive impairment, dementia, myopathy or Paget's disease of bone (PDB) is not uncommon in ALS and coexistence of FALS and PDB has been reported in pedigrees with p62/*SQSTM1* mutations (Fecto and Siddique, 2011). Importantly, $VCP^{R155H/+}$ knock-in mice develop slow-progressive motor neuron degeneration and extensive TDP-43 pathology (Yin et al., 2012).

Valonsin containing protein (VCP) is a ubiquitously expressed protein involved in a variety of cellular activities but its most well established roles lie in the translocation of misfolded proteins from the endoplasmic reticulum (ER) during ER-associated protein degradation (ERAD) and their subsequent degradation by the proteasome and through autophagy. *VCP* mutations induce ER stress and the unfolded protein response (UPR) (Gitcho et al., 2009), impair ERAD function causing accumulation of misfolded proteins in the ER (Weihl et al., 2006), decrease proteasome activity and impair autophagy (Ju and Weihl, 2010). *VCP* mutations have also been shown to cause mitochondrial dysfunction and impaired ATP production (Bartolome et al., 2013). Most recently, the essential role of VCP in autophagy has been linked to the removal of stress granules (SGs), which are formed in cellular stress and accumulate RNA binding proteins such as TDP-43 and FUS. Indeed the accumulation of cytoplasmic inclusions containing stress granule constituents is widespread in ALS, FTLD and other neurodegenerative conditions. Furthermore, overexpression of ALS-associated mutations in *VCP*, A232E and R155H, leads to the

formation of SG-like structures containing the SG markers, eIF3, TDP-43 and VCP, suggesting a common effect of SG promoting mutations in *TARDBP, FUS* and *VCP* that may contribute to ALS pathogenesis (Buchan et al., 2013).

In order to further investigate the prevalence of *VCP* mutations and potentially identify ALS multigenerational pedigrees in which *VCP* mutations segregate with disease, we screened a UK cohort of FALS cases for mutations/ DNA variants in *VCP.* Although no mutations were identified in coding regions, we report an unusual hexanucleotide expansion located in the 5'UTR. In addition, a novel single nucleotide substitution, c.-360G>C, was also identified in the 5'UTR. Both of these DNA variants were predicted to be pathogenic.

3.2 Method

3.2.1 Sample collection and DNA extraction

Cohorts of 102 FALS index cases and 90 SALS cases were obtained with consent (Imperial College NHS Healthcare Trust). The diagnosis of ALS was made according to revised El Escorial criteria, which requires the presence of both upper motor neuron and lower motor neuron symptoms. An additional cohort of 96 FALS cases obtained from King's College London (Kings College NHS Healthcare Trust) was screened for the hexanucleotide repeat. DNA samples of 184 healthy individuals were used as controls. DNA was extracted from whole blood or buffy coats using QIAmp DNA Mini Kit (QIAGEN, UK) according to the Manufacturer's instructions.

3.2.2 DNA Sequencing

Exons 1, 2, 3, 5, 6, 7, 10, 13, 14, 17, including ~400bps of the 5' and 3' UTRs and ~100bps flanking regions of *VCP* were amplified using Platinum Taq DNA Polymerase (Invitrogen, UK) and sequenced. PCR Primers (Invitrogen, UK) were designed using Primer 3 program *(http://frodo.wi.mit.edu/)* (**Table 3-1).** PCR products were purified using SureClean (BIOLINE, UK) and sequenced using an ABI 3730XL sequencer (Imperial College Genomics laboratory). Sequences were aligned and analyzed using Condon Code Aligner (*http://www.codoncode.com/aligner/*) and Seqdoc programs (*http://research.imb.uq.edu.au/seqdoc/*).

3.2.3 Gel electrophoresis

Long electrophoresis was used to discriminate between DNA fragments of similar size using 2.8% agarose (Electran®, VWR BDH Prolabo, UK) gels. Electrophoresis was carried out for 16 hours at 140V, 55mA and data analysed using a Gel Doc 2000 (BIO-RAD, UK) system.

3.2.4 Total RNA extraction and RT-PCR

Expression of VCP was studied using RT-PCR. RNA was extracted from whole blood or buffy coats using Direct-zolTM RNA Mini-Prep Kit (Zymo Research) and cDNA was generated using SuperScript® III First-Strand Synthesis System (Invitrogen, UK) according to the Manufacturer's instructions. PCR reactions were carried out using cDNA primers shown in **Table 3-1**.

3.2.5 Statistics

Genotypic and allelic associations of SNPs were tested by 2 x 3 and 2 x 2 Fisher's tests respectively. The genotypic frequencies of hexanucleotide repeats between cases and controls were compared using 2 x 4 Fisher's tests. To obtain a frequency of the hexanucleotide repeat in the general population, raw DNA/RNA sequencing data from the 1000 genome project and the Geuvadis project (*http://www.geuvadis.org/*) were retrieved using SAMTools (*http://samtools.sourceforge.net/*) and called for indels using Dindel (Albers et al., 2011) programs. Depth and Breadth of coverage were evaluated using BEDTools (*https://code.google.com/p/bedtools/*) program (**Appendix II**). Haploview (*http://www.broadinstitute.org/haploview*) was used to reconstruct haplotypes and calculate r^2 values. Survival data was analyzed using Kaplan-Meyer curves and compared using a log-rank test. The Bioinformatic programs, SIFT (*http://sift.jcvi.org*), Polyphen 2 (*http://genetics.bwh.harvard.edu/pph2/*) and Mutation taster (*http://www.mutationtaster.org*) were used to predict the effect of DNA variants. Phylogibbs (*http://www.phylogibbs.unibas.ch/cgi-bin/phylogibbs.pl*), a Gibbs sampler incorporating phylogenetic information, was used to analyse the 5' region.

Table 3-1. Primers used for Polymerase Chained Reactions for the VCP Study.

Exon	Primers	Product Size
5'UTR and Exon 1	1a F: TGTGTGTTCTGTGGTTGCCC 1a R: CCGCAGATCACAGCCAATCA	583
	1b F: GCGTGTCGCATCACTGAG 1b R: CTGCATGACACAGCACGAT	719
	Hex F: GATTCGGCTCTTCTCGGCTC Hex R: TAACGGCTACGAGCGGTGG	250
2-3	F: GCTTTCTGGTCTAGGGACAGC R: CAAGAACTTGGTCCTGCCTG	685
5	F: GAGCTTGGCATTTTGACCC R: CCCAGTCCTGACAGTTACCAC	301
6	F: ACCATGCCGGGTTGAGAATC R: CCCTCTAATCCAAGGCAATAATGA	382
7	F: CCCTCTCTGGAGCGCTAGTC R: AAAAGGATGTGTTCATAAGTGCTC	269
10	F: AGAGTGACCAACCACCCTGG R: TGCCAACTCCCATTTCCTGG	449
13	F: AGGTTTGAGGCACTAAGGAGTC R: CAGTTGAGCAGCCAGCACTA	600
14	F: GTGTGAGCCACCACGTTTG R: CCCAGTGGAATCTTGTCCAG	471
17	F: TGGGAGCATTAGACAGTGCTT R: TGCAGATGCTTTACTGTGGCA	597
VCP-cDNA Primers (Exon 3 and 5)	F: CGAGGTGACACAGTGTTGCT R: TTGAACTCCACAGCACGCAT	335
GAPDH-cDNA Primers (Exon 7 and 8)	F: CCTGCACCACCAACTGCTTA R: GAACATCATCCCTGCCTCTAC	181

Exon1 and the 5'UTR were sequenced using two pairs of primers. The Hex primers were used for amplifying the hexanucleotide repeat for electrophoresis. Successful PCR of exon 10 requires 2.5mM Mg^{2+} concentration. The VCP cDNA PCR was carried out for 55 cycles at annealing temperature of 65^{o}C, whereas the GAPDH cDNA PCR was carried out for 38 cycles at annealing temperature of 55^{o}C.

3.3 Results

3.3.1 Sequence analysis of the VCP gene

Twelve non-synonymous DNA variants have been detected in the *VCP* gene in ALS to-date (FALS and SALS) cohorts (**Table 3-2**). In this study, we screened *VCP* in a further UK cohort of FALS and SALS cases. As most pathogenic mutations in *VCP* that have been detected in IBMPFD and ALS are found in the N-terminal domain, the L1 linker and the D1/ ATPase domain (**Figure 3-1 A**), we screened ten exons (1, 2, 3, 5, 6, 7, 10, 13, 14, 17) harbouring most known mutations found in ALS and IBMPFD together with 5'UTR and 3'UTR regions of the gene that have not been extensively studied before. The FALS cohort consisted of 102 index cases, from which known mutations in *SOD1*, *VAPB*, *TARDBP*, *FUS* and *DAO* genes had been previously excluded. No coding mutations in *VCP* were detected. However, three novel variants were found in regulatory regions in three separate index cases, a single base pair substitution c.-360G>C located in the 5'UTR, a c.2421+94C>T substitution located in the 3'UTR and an 18bp insertion, c.-221_-220insCTGCCACTGCCACTGCCG present in the 5'UTR. The latter sequence was inserted in the middle of a repeat sequence of imperfect CTGCCR hexanucleotide repeats, in which the last nucleotide could either be A or G, and this is equivalent to an expansion of 3 additional hexanucleotide repeats (i.e. an 8-repeat variant) within the control sequence that contains 5 repeats (**Figure 3-1 B**). These variants are not present in any public databases. All three index cases harbouring these variants were heterozygous and also positive for the *C9orf72* hexanucleotide expansion.

The c.-360G>C substitution and the 18 bps hexunucleotide expansion were predicted to be pathogenic using the Mutation taster program as they may interrupt the 5' regulatory region (**Figure 3-1 C**). The c.2421+94C>T was predicted to be polymorphic. The proband carrying the hexanucleotide expansion was a female patient with limb onset at 41 years who died at the age of 47 years. There was no record of PDB, myopathy or dementia in the pedigree (**Figure 3-1 D**).

The carrier of c.-360G>C (**Figure 3-2**) was a male patient with bulbar onset at the age of 64 who died at 68 years of age, 48 months after disease onset. He developed ALS with dementia and his mother and brother were also diagnosed with ALS (refer **Figure 3-2** for details). The carrier of c.2421+94C>T was a female patient with bulbar onset at the age of

73 years who died at 75 years, 24 months after disease onset. The only FALS case possessing the 18bp insertion is described in main text.

Interestingly, female FALS patients who possessed the minor allele of SNP2 had significantly decreased survival time (median 15.5months) compared to those possessing the major allele (median 25 months) (**P=0.0174**) or compared to male patients possessing the minor allele (**P=0.0242**). No significant changes in onset and site of onset were found for any of the polymorphisms detected.

Table 3-2. Frequencies of *VCP* mutations.

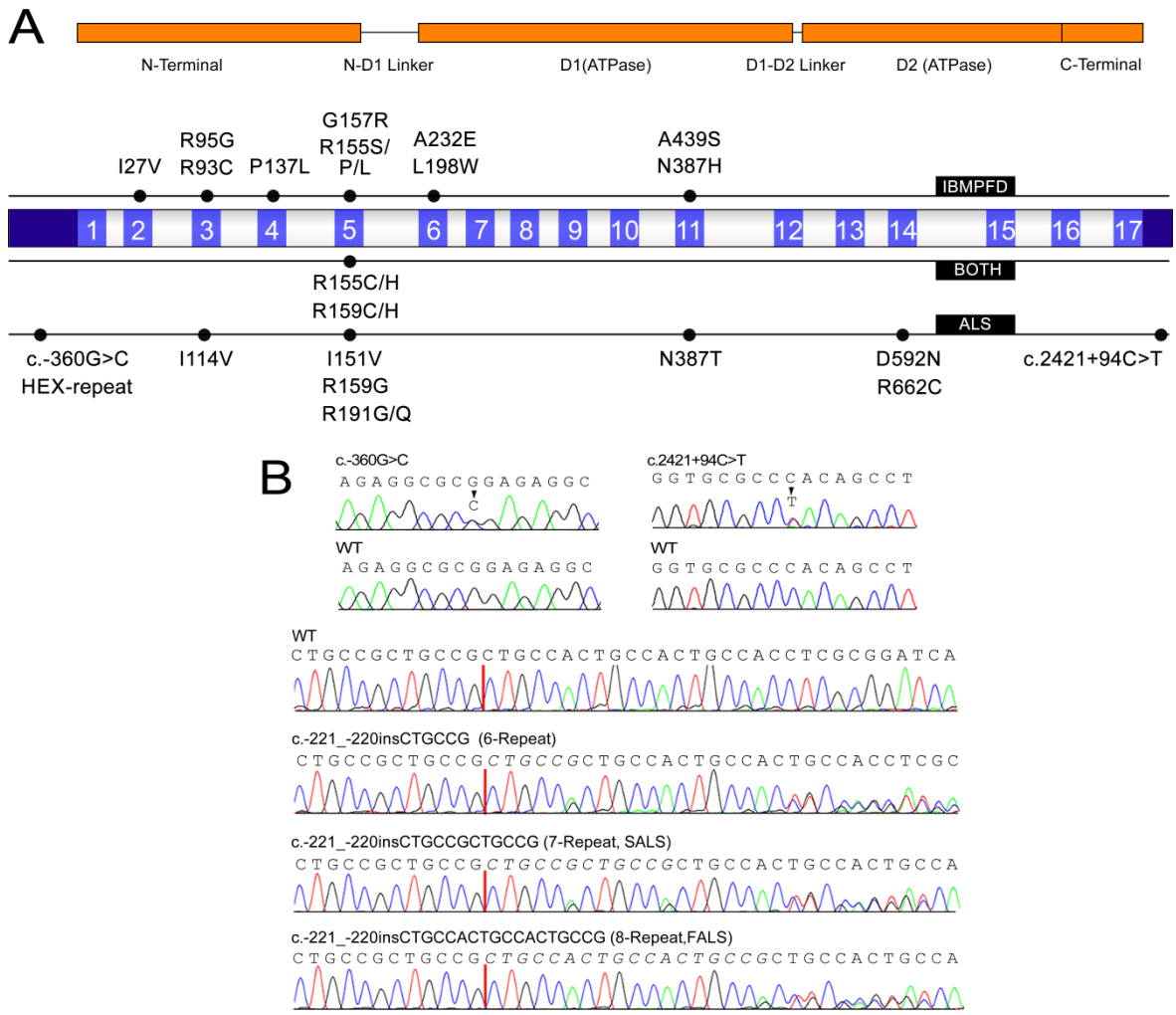
Protein/ Region	Position	rs	Changes	Population Prevalence			Prediction		
				FALS	SALS	Control	SIFT	PP	MT
5' Variant	**35072710**		**c.-360G>C**	**1/102[a]**		**0/379[k]**	**na**	**na**	**+**
5' Variant	**35072570**		**8-repeat***	**1/198[a]**	**0[a]/88[a]**	**0/219[a]**	**na**	**na**	**+**
5' Variant	**35072570**		**7-repeat****	**0/198[a]**	**1[a]/88[a]**	**0/219[a]**	**na**	**na**	**+**
I114V	35066777		c.340A>G	1[i]/876[b,d,e-g,i,j]	1[d]/2243[b-d,f-j]	0/5558[d,i,k]	-	-	+
I151V	35065373	rs367881889	c.451A>G	0/1074 [a,b,d,e-g,i,j] [b,d,e-	1[c]/2505 [b-d,f-j]	0/5781[c,d,k]	-	-	+
R155C[§]	35065361	rs121909330	c.463C>T	1[i]/1074 [a,b,d,e-g,i,j]	0/2505 [b-d,f-j]	0/5558[d,i,k]	+	+	+
R155H[§]	35065360	rs121909333	c.464G>A	2 [b,i]/1074 [a,b,d,e-g,i,j,l]	0/2505 [b-d,f-j]	0/7127[b,d,i,k,]	-	-	+
R159G[§]	35065349	rs387906789	c.475C>G	1[b]/1074 [a,b,d,e-g,i,j,l]	0/2505 [b-d,f-j]	0/6943[b,d,k]	-	+	+
R159C[§]	35065349		c.475C>T	1[i]/1074 [a,b,d,e-g,i,j]	1[h]/2505 [b-d,f-j]	0/7127[d,h,l,k]	+	+	+
R159H[§]	35065348	rs121909335	c.476G>A	1[d]/1074 [a,b,d,e-g,i,j]	0/2505 [b-d,f-j]	0/5392[d,k]	-	-	+
R191G[§]	35065253		c.571C>G	1[i]/1074 [a,b,d,e-g,i,j]	0/2505 [b-d,f-j]	0/5558[d,i,k]	-	+	+
R191Q[§]	35065252	rs121909334	c.572G>A	3 [b,i]/1074 [a,b,d,e-g,i,j]	0/2505 [b-d,f-j]	0/7127[b,d,i,k]	+	+	+
I300I	35062259	rs372839296	c.900C>T	1[g]/876 [b,d,e-g,i,j]	0/1642 [b-d,f-j]	1/5374[d,k]	-		+
T330T				0/876 [b,d,e-g,i,j]	1[d]/1642 [b-d,f-j]				
N387T	35061608		c.1160A>C	0/978 [b,d,e-g,i,j]	1[h]/1642 [b-d,f-j]	0/6943[d,h,k]	+	+	+
L414L				0/876 [b,d,e-g,i,j]	1[d]/1642 [b-d,f-j]				
I479I				1[d]/876 [b,d,e-g,i,j]	0/1642 [b-d,f-j]				
A528A				0/978 [b,d,e-g,i,j]	2[d]/1642 [b-d,f-j]				
Q568Q	35059790	rs142577424	c.1704A>G	3[e]/978 [b,d,e-g,i,j]	0/2016 [b-d,f-j]	46/5659[d,e,k]			+
D592N	35059720	rs387906790	c.1774G>A	1[b]/978 [b,d,e-g,i,j]	0/2016 [b-d,f-j]	0/7188[b,d,j,k]	+	+	+
R625R	35059619	rs201410035	c.1875G>T	0/978 [b,d,e-g,i,j]	1/2016 [b-d,f-j]	0/5619[j]			+
L661L				0/978 [b,d,e-g,i,j]	1[d]/2016 [b-d,f-j]				
R662C	35059510		c.1984C>T	0/978 [b,d,e-g,i,j]	1[h]/2016 [b-d,f-j]	0/7188[d,h,j,k]	-	+	+
3' Variant	**35057020**		**c.2421+94C>T**	**1[a]/102[a]**		**0/379[k]**	**na**	**na**	**+**

Frequencies of *VCP* variants found in ALS to date are summarized. Variants identified in this study are in bold. Predictions on the consequences of the variants using SIFT, Polyphen2 (PP) and Mutation Taster(MT) are shown: "+" : predicted as pathogenic; "-" : predicted as tolerated; "§" denotes the mutations that have been found in ALS families with histories of Paget's disease or inclusion body myopathy. *8-repeat hexanucleotide expansion: c.-221_-220insCTGCCACTGCCACTGCCG **7-repeat hexanucleotide expansion: c.-221_-220insCTGCCGCTGCCG.

References for Table 3-2.

a Present study (102 FALS for Exon1, 2, 3, 5, 6, 7, 10, 13, 14, 17, 5' and 3' UTRs; 96 FALS for HEX-Repeat, 88 SALS for HEX-Repeat, 219 Controls for Hex repeat). **b** (Johnson et al., 2010) (211 FALS for all Exons, 78 SALS for all Exons, 1569 Controls for R155H, R159G, R191Q and D592N). **c** (DeJesus-Hernandez et al., 2011a) (1 SALS for All Exons + 112 SALS for Exon 5; 407 Controls for I151V). **d** (Koppers et al., 2011) (93 FALS from 80 Families for all Exons; 377 SALS for all Exons+ 377 SALS for Exon 4 and 5+ 58 SALS-FTD for All Exons; 713 Controls for R159H, 685 for T330T, 674 for L414L and I479I, 662 for A528A, 594 for L661L and 695 for the rest of Exons). **e** (Tiloca et al., 2011) (166 FALS+ 14 FALS/FTD for all Exons, 285 Controls for Q568Q). **f** (Miller et al., 2012) (75 FALS for all Exons; 101 SALS for all Exons + 150 SALS for Exons 5 and 14). **g** (Williams et al., 2012) (131 FALS for all Exons; 48 SALS for all Exons). **h** (Abramzon et al., 2012) (701 SALS for all Exons; 1569 Controls (same as **b**) for R159C, N387T and R662C). **i** (Gonzalez-Perez et al., 2012) (179 FALS for all Exons + 96 FALS for Exon 5; 178 SALS for all Exons; 184 Controls for I114V, R155C, R155H, R159C, R191G and R191Q). **j** (Zou et al., 2013) (20 FALS for all Exons; 100 SALS for all Exons + 224 SALS for Exons 4, 5, 10, 14; 245 Controls for Exon 14). **k** 1000 genome project EUR subgroup (379 Controls) or with NHLBI EVS (4300 Controls).**l** Data from (Benatar et al., 2013) was not included in this Table as frequency data was not available but is included in **Table 3-5**. na indicates not appropriate to use for non-coding changes.

Figure 3-1. Published coding mutations in *VCP* in ALS and IBMPFD.



A.VCP is an AAA^+ protein comprising N, N-D1 linker, D1, D1-D2 Linker, D2 and C domains. Two conserved motifs, Walker1 and Walker2, serve as ATP binding sites and catalyze hydrolysis, whereas N-domain and C-domain allow binding of substrates and cofactors. FALS-related *VCP* mutations have been identified throughout the gene and are more frequent in exon 5 and 14, which correspond to N and D2 domains respectively. Conditions associated with the variants are shown separately.

B shows the chromatograms of the variants identified in this study. The positions where the hexanucleotide repeats are expanded are marked by red lines and the mutated sequences are shown in the expansion carriers.

C shows the 5'UTR sequence of the *VCP* gene. The hexanucleotide repeat region is underlined and the transcription start site is denoted by asterisk (*). Predicted transcription factor binding sites are shown in orange boxes and the c.-360G>C variant (bold) is located in a potential MAZ binding site.

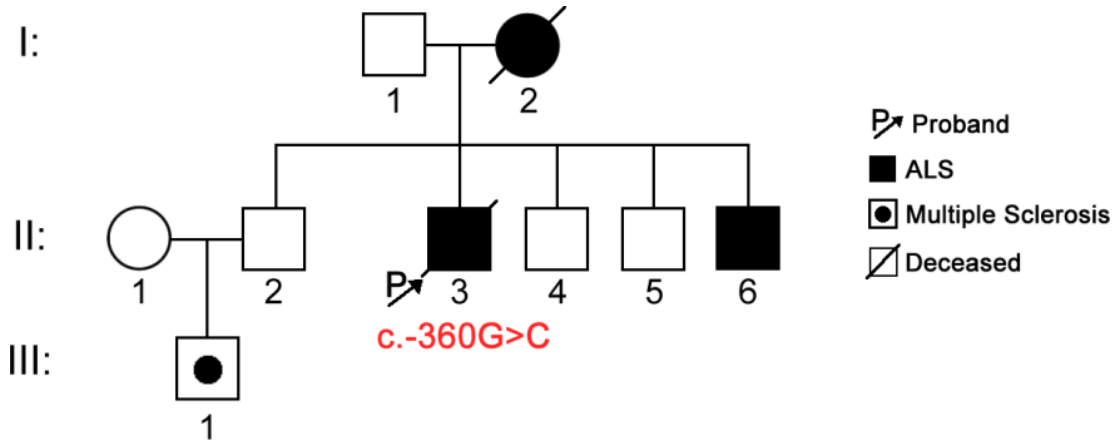
D shows the pedigree of the FALS case carrying the 8-repeat hexanucleotide expansion.

E and F show the frequencies of phenotypes existing in published ALS kindred caused by different *VCP* mutations. The mutations are grouped according to the number of symptoms seen in the kindred.

(E) For mutations causing ALS alone or with 1 additional phenotype, i.e. I114V, D592N, R159H, R191G and R159C, Dementia, PDB and Myopathy co-occurred with ALS at the same frequency.

(F) Dementia and PDB are more frequently observed than Myopathy in kindred in which mutations causing ALS with 2 or 3 phenotypes are present , i.e. R155H, R191G and R191Q. I114V and D592N have only been described in pure ALS kindred, whereas R155H and R191Q were found in kindred presenting all four phenotypes.

Figure 3-2. The c.-360G>C Pedigree.



Individual I:2 had severe bulbar onset and died at 63 years. The proband, II:3, onset at 63 years and died at 68 years. He developed hyperreflexia, fasciculations, muscular atrophy, bulbar signs and dementia. III:1 was a multiple sclerosis patient and developed ALS at 42 years. Hyperreflexia, fasciculations, muscular atrophy and bulbar signs were present in this patient.

3.3.2 Screening for the hexanucleotide expansion in additional ALS cohorts

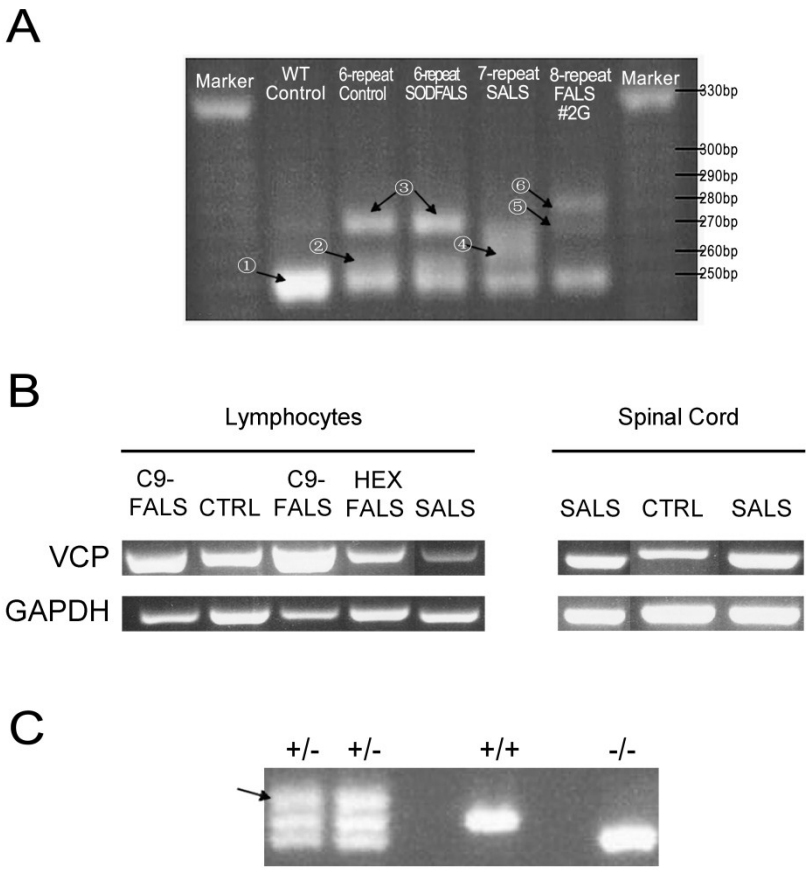
In order to further investigate the prevalence of the hexanucleotide expansion, the region containing the repeat was amplified in four extra cohorts including FALS **(n=96)**, FALS with *SOD1* mutations **(n=20)**, SALS **(n=88)** and Controls **(n=219).** The products were then separated by long electrophoresis to discriminate 3 base pair differences in PCR product sizes (**Figure 3-3A**). Representative samples were confirmed by sequencing. No variation was found in the additional FALS cohort. A 7-repeat variant, c.-221_-220insCTGCCGCTGCCG, was found in one SALS case, and a 6-repeat variant, c.-221_-220insCTGCCG, was found in six controls and one SOD-FALS case. Unlike the 8-repeat variant, the 7-repeat and the 6-repeat variants both contain homogenous inserted sequences, $[CTGCCG]_n$, at the same site as the 8-repeat. All these variants were heterozygous. The 5-repeat variant is the major allele in all cohorts. A significant genotypic association of the 8-repeat with ALS was found using a 2 x 4 Fisher's test **(P=0.0064, Table 3-3).**

RT-PCR was carried out to investigate the effect of the hexanucleotide repeat on VCP expression in lymphocytes. VCP expression was readily detected in spinal cord and lymphocytes from control, SALS and FALS cases **(Figure 3-3B)**. Although a similar level of expression was found in lymphocytes from the 8-repeat carrier, the lack of further samples does not allow any conclusions to be made about effects on expression.

3.3.3 SNPs identified in the VCP gene

An 11 bp indel close to the intron 9/Exon 10 junction (rs11272867) detected in FALS cases was further genotyped in a group of controls (**n=184**) using long electrophoresis (**Figure 3-3C**). No significant difference in SNP frequency was observed between FALS cases and controls (MAF = 30.2%). A 3' UTR SNP, rs62544156, was detected in 1 FALS case. The SNP was absent from EUR controls from the 1000 genome project but was found in 0.1% of chromosomes in the European-American samples from the NHLBI Exome Variant Server. We further tested SNP associations with FALS cases with or without the *C9orf72* mutations in the common SNPs identified in this study (**Table 3-4)** but found no significant allelic and genotypic association (after FDR correction) with FALS.

Figure 3-3. Electrohporesis for the *VCP* tandem repeats.



A. Agarose gel electrophoresis: ① Represents the wild type 5-repeat variant seen in most samples; ② is a faint band representing the 6-repeat variant; bands ③ and ⑥ are heteroduplexes generated as a result of mixed templates. The seven-repeat and eight-repeat variants are represented by ④ and ⑤ respectively. These were confirmed by sequencing.

B. RT-PCR of the *VCP* and *GAPDH* genes were carried out in the lymphocytes and spinal cords of ALS and Controls. The 8-repeat FALS case is indicated by HEX. Each lane is from a different individual.

C. The PCR products of the 11-bp indel were also separated on agarose electrophoresis. The first 2 lanes show heterozygous products accompanied by heteroduplexs (arrow). Lane 4 shows the large PCR product which represents homozygous insertions, and wild type product is seen in Lane 6.

Table 3-3. Frequencies of hexanucleotide expansions in cases and controls.

Population	$[CTGCCR]_8$	$[CTGCCR]_7$	$[CTGCCR]_6$	$[CTGCCR]_5$	*n*
FALS	1	0	0	197	198
SALS	0	1	0	87	88
SODFALS	0	0	1	19	20
CONTROL	0	0	6	213	219

The wild type 5-repeat is the most frequent allele in all cohorts. The 6-repeat variant was found in Control and SOD-FALS individuals, whereas the 7 and 8 repeat variants were only found in SALS and FALS cohorts.

A Fisher's 2 x4 Table was used to compare the frequency of the VNTRs between cases and controls, and significant results were obtained between FALS cases and controls **(P=0.0208)** and pooled ALS cases and controls **(P=0.0064)**. The analysis showed that rare higher number repeats were only found in FALS and SALS and not in controls, suggesting pathogenic effects are caused by these variants.

Table 3-4. Prevalence of *VCP* SNPs detected in this study.

Known Genetic variations of the VCP gene in FALS cases and Controls

Population	Chrs	MAF %	Genotype Frequency%			Genotype Fisher's P (FDR)	Allelic Fisher's P (FDR)
***SNP1**, rs10972300, Int2, G/A*			GG	GA	AA		
EUR	762	21	63	32	5		
FALS	188	27	55	36	8	0.1803 (0.3335)	0.0767 (0.2018)
C9+	74	19	68	27	5	0.7868 (0.9659)	0.8800 (0.8800)
C9-	114	32	47	42	11	**0.0299** (0.1495)	**0.0108** (0.0900)
***SNP2**, rs514492, Int7, A/G*			AA	AG	GG		
EUR	762	23	60	33	7		
FALS	204	20	67	27	6	0.4695 (0.5859)	0.2594 (0.3243)
C9+	72	22	55	28	17	0.9555 (0.9659)	0.8782 (0.8800)
C9-	98	26	55	33	12	0.8284 (0.8481)	0.6999 (0.8749)
***SNP3**, rs11272867, Int9, -/TTGTGTACTGT*			WT/WT	WT/INS	INS/INS		
IC Contros	368	70	7	47	46		
FALS	194	75	7	35	58	0.1312 (0.3335)	0.2005 (0.3243)
C9+	58	72	7	41	52	0.7726 (0.9659)	0.7589 (0.8800)
C9-	90	70	11	38	51	0.3675 (0.6185)	1 (1)
***SNP4**, rs684562, Int13, G/A*			GG	GA	AA		
EUR	762	31	50	39	11		
FALS	186	28	53	39	9	0.6266 (0.6266)	0.3771 (0.3771)
C9+	64	28	53	38	9	0.9659 (0.9659)	0.6741 (0.8800)
C9-	86	36	39	49	12	0.3711 (0.6185)	0.3937 (0.8749)
***SNP5**, rs62544156, 3'UTR ,C/T*			CC	CT	TT		
EUR	-	-	-	-	-	-	-
FALS	196	1	99	1	0	-	-
***SNP6**, rs10553318, 3'UTR, G/T*			GG	GT	TT		
EUR	762	21	63	32	5		
FALS	186	15	73	24	3	0.2001 (0.3335)	0.0807 (0.2018)
C9+	64	13	78	19	3	0.2325 (0.9659)	0.1434 (0.7170)
C9-	86	19	67	28	5	0.8481 (0.8481)	0.6758 (0.8749)

Six known SNPs were detected in this study. Four of these are intronic, which were identified in the flanking regions of exon sequences, and two were found in the 3'UTR (Exon17). SNPs 1, 2, 4, 6 were polymorphic with minor allele frequencies > 10%, whereas SNP 5 was found in only one FALS case. The frequencies of these SNPs in FALS were compared with the EUR subgroup from the 1000Genome project using Fisher's tests. All comparisons were corrected for multiple testing using Benjamini-Hochbergh correction (FDR). No significant association was detected. There was no available frequency data for SNP3 and SNP5 in any accessible database.

3.4 Discussion

In this study, we investigated whether any potentially pathogenic DNA variants were present in the *VCP* gene in a cohort of 102 index FALS cases. No known or novel mutations were found in the ten exons which harbour most known mutations. Taking into account the results from previous studies, it can be concluded that *VCP* mutations are rare in ALS (**Table 3-2**).

To date nine coding mutations have been detected in 14 kindred. In general, most mutation carriers are positive for a family history of multiple symptoms **(Table 3-5)**. Amongst the previously reported families with at least one confirmed case of ALS, dementia was present in seven families, myopathy in five families and PDB in seven families **(Table 3-5).** Age at onset of ALS varied between 28 and 63 years and disease duration between 1 and 26 years with considerable heterogeneity within some kindred (**Table 3-5**). For some mutations (R159G, R159C, R159H, 191G, R191Q or D592N), ALS was the predominant phenotype where 90% of carriers possessing these mutations developed ALS, whereas the clinical manifestations of other mutations such as R155H were more variable (only 17.6% of carriers developing ALS). Notably, only the R191G mutation has been shown to segregate with ALS (Gonzalez-Perez et al., 2012). Two mutations, I114V and D592N, have only been found in kindred presenting with a pure ALS phenotype (**Figure 3-1 E-F, Table 3-5**). Two coding changes (I114V and R159C) found in FALS have also been reported in SALS together with 3 further novel coding changes (I151V, N387T and R662C). In addition, *VCP* mutations have been reported both in families with ALS, and also families that present with PDB with or without dementia and IBMPDB (e.g. R155H) (**Table 3-2**).

Table 3-5. Summary of phenotypes exhibited for known *VCP* mutations.

Mutat-ions	Symptoms seen in *VCP* mutation carriers from FALS families (No. of families positive for the symptoms)				Age of onset (Years)	Survival Time (Years)	References
	ALS	Dementia	Paget's disease	Myopathy	Median (Range), n	Median (Range), n	
I114V	1/2 (1/1)	0/2 (0/1)	0/2 (0/1)	0/2 (0/1)	45, n=1	2.25 n=1	(Gonzalez-Perez et al., 2012)
R155C	?/? (1/1)	?/? (1/1)	?/? (1/1)	?/?(1/1)	na	na	(Benatar et al., 2013)§
R155H	*3/17 (4/4)	*2/17 (3/4)	*3/17 (4/4)	8/17 (3/4)	41(28-63), n=19	15.45 (1-26) n=10	(Benatar et al., 2013; Gonzalez-Perez et al., 2012; Johnson et al., 2010)§
R159G	3/3 (2/2)	2/3 (2/2)	1/3 (1/2)	0/3 (0/2)	52 (46-53), n=3	5 (2-13), n=3	(Benatar et al., 2013)§
R159C	2/3 (1/1)	0/3 (0/1)	1/3 (1/1)	0/3 (0/1)	55 (53-57), n=2	-	Gonzalez-Perez et al., 2012)
R159H	1/1 (1/1)	*0/1 (1/1)	0/1 (0/1)	0/1 (0/1)	59 n=1	1.9 n=1	(Koppers et al., 2011)
R191G	6/7 (1/1)	0/7 (0/1)	0/7 (0/1)	4/7 (1/1)	45 (42-45), n=3	9 n=1	(Gonzalez-Perez et al., 2012)
R191Q	6/6 (3/3)	1/6 (1/3)	*0/6 (1/3)	*0/6 (1/3)	50 (37-53), n=7	2.8 (2.4-14.9), n=3	(Johnson et al., 2010) (Gonzalez-Perez et al., 2012)
D592N	1/1 (1/1)	0/1 (0/1)	0/1 (0/1)	0/1 (0/1)	53 n=1	<1 n=1	(Johnson et al., 2010)

Phenotypes of *VCP* mutation carriers from Familial ALS cases published to date. Pedigrees with multiple affected and at least one confirmed ALS case are classified as FALS pedigrees. Values indicate the number of cases with a confirmed mutation exhibiting a specific phenotype divided by the number of individuals with a confirmed mutation. * indicates that not all individuals within the family with this specified phenotype had been tested for the mutation. The values in brackets indicate the number of positive families with a defined phenotype divided by the total number of familes with the specified mutation. In summary, most mutation carriers were positive for a familial history of multiple symptoms. The clinical manifestations of R155H were most variable (only 17.6% of carriers developed ALS). In contrast, no additional symptoms were seen in carriers of I114V and D592N mutations and this may be explained by the rapidly progressive phenotypes caused by these mutations. Using the data where ethnicity is specified, 2 pedigrees, carrying R159H and R191Q mutations, were from Italian and Dutch cohorts, whereas 9 pedigrees, carrying the R155C, R155H, R159G, R191Q and D592N mutations, were from US cohorts. § Data from Benatar et al (2013) was not included in **Table 3-2** as frequency data was not given. na =not available.

In view of the important regulatory function of 5' and 3'UTRs, these regions were also included in the study. An imperfect hexanucleotide expansion and a single base pair substitution were identified in the 5'UTR region, which were both predicted to be pathogenic, were found in two index cases. Although both of these cases were also positive for a *C9orf72* expansion mutation, the co-existence of multiple mutations, including *C9orf72* and *TARDBP,* has been reported previously in ALS (van Blitterswijk et al., 2012a). The expression of the *VCP* gene is regulated by the transcription factors, PBX-1 and ELF2, which bind to the 5' upstream region of the *VCP* gene (Qiu, 2007). The transcription start site and a predicted transcription factor binding site for E-box, are specifically located in the region of the hexanucleotide repeat expansions, which potentially may interrupt the formation of the initiation complex. The Phylogibbs program also predicted the presence of a possible motif for transcription factor binding within the hexanucleotide repeat. In addition, the c.-360G>C variant is located within a predicted transcription factor binding site for MAZ (Qiu, 2007) **(Figure 3-1C)**. Clearance of existing ubiquitinated aggregates may be impaired as a consequence of decreased levels of VCP expression (Kobayashi et al., 2007).

Whilst the 5 and 6-repeat expansions were common in controls, the 7- and 8-repeat variants were only found in SALS and FALS respectively. It is possible that pathogenicity increases with the length of the repeat. In fact, after analysing the raw data from the Geuvadis RNA sequencing project of 1000 Genomes samples (*http://www.geuvadis.org*/) using the Dindel program (Albers et al., 2011) (**Appendix II**), we found that a 6-repeat, but not 7 or 8-repeats, was present in 324 well-covered samples (all with >20 depth of coverage and >0.99 breadth of coverage in a 200bp window flanking the repeat). Whilst most pedigrees with *VCP* mutations also have family histories of PDB, myopathy and dementia (**Table 3-5**), the 8-repeat family only presented with ALS. It is possible that the early age of onset in the proband (47 years) is the result of a combined effect of both *VCP* and *C9orf72* expansions. However, no genomic DNA was available from other family members to study the patterns of segregation and anticipation, and no other FALS cases possessed this expansion.

Tandem repeats are frequently associated with neurodegenerative diseases, such as Huntington's disease and spinocerebellar ataxias and have recently been shown be the

most common mutation found in ALS, FTD and ALS/FTD, with the identification of hexanucleotide repeat expansions in the first intron of *C9orf72* gene, located on chromosome 9p21 accounting for 18-55% of FALS cases according to geographic location, being most prevalent in Western Europe (DeJesus-Hernandez et al., 2011b; Moreira et al., 2004; Renton, 2011). Tri-nucleotide expansions in *NIPA1* (Blauw et al., 2012) and *ATXN2* (Elden et al., 2010) genes also confer risk of ALS (**Chapter 6**).

In conclusion, the overall frequency of *VCP* mutations in Europe and US is low (<1%). However, three novel variants were found in the 5' UTR of FALS cases, including a hexanucleotide expansion which may be pathogenic and merit further investigation.

-END OF CHAPTER 3-

Chapter 4

Sequence Analysis of *SQSTM1* Gene in the Imperial College ALS Cohort

4.1 Introduction

TDP-43/FUS -positive poly-ubiquitinated inclusions are the pathological hallmark of most sporadic ALS cases. These inclusions not only contain TDP-43 or FUS, but also p62/sequestosome 1 (p62/SQSTM1), optineurin (OPTN) and ubiquilin-2 (UBQLN2). Indeed, mutations in several components of protein homeostasis, such as *UBQLN2*, Charged multivesicular body protein 2B (*CHMP2B*), *OPTN* and valosin containing protein (*VCP*), are also present in less prevalent forms of ALS and ALS/FTD (Fecto and Siddique, 2011). Interestingly, mutations in *OPTN* and *VCP* are also seen in other allelic forms characterised by TDP-43-positive poly-ubiquitinated inclusions, primary open angle glaucoma (POAG) and inclusion body myopathy with Paget's disease of bone and frontotemporal dementia (IBMPFD), respectively.

Screening of four large ALS cohorts from US, Europe and Japan for mutations in *SQSTM1* has revealed the presence of a number of novel or rare coding mutations in FALS, FTLD and SALS, some of which are predicted to be pathogenic (Fecto et al., 2011; Hirano et al., 2013; Rubino et al., 2012; Teyssou et al., 2013). However, due to lack of DNA in multiple members of FALS kindred, to date it has not been proved that any of these mutations are transmitted with disease in ALS. Nevertheless, amongst the coding mutations is P392L, which is common in Paget's disease of bone (PDB) (Morissette et al., 2006; Rea et al., 2009), where there is evidence of transmission of the mutation with disease in multiple kindred. In order to further investigate and assess whether they are independently causal for ALS or whether p62/*SQSTM1* DNA variants or modified forms of p62/SQSTM1 contribute to a significant predisposition to disease, we have examined a further FALS cohort from the UK.

p62/SQSTM1 is abundantly expressed in spinal cord, especially in motor neurons and is present in distinctive p62/SQSTM1-positive fibrillar or compact TDP-43-positive inclusions found in ALS and FTD-ALS (Keller et al., 2012). In addition, a significant portion of *C9orf72* positive FTD/ALS patients have p62 positive, TDP-43 negative inclusions, indicating that multiple substrates are recognised by p62 in these conditions (Al-Sarraj et al., 2011). SQSTM1 plays multiple roles in protein homeostasis (Geetha and Wooten, 2002), in particular, the interaction of p62/SQSTM1 with the autophagic marker light chain 3 (LC3) is essential for the generation of autophagosomes necessary for the autophagic degradation of ubiquitinated protein aggregates (Lamark and Johansen, 2010). Depletion of p62 protein levels inhibits LC3 recruitment to autophagosomes and has been shown to increase cell death induced by mutant huntingtin (Bjørkøy et al., 2005).

4.2 Methods

4.2.1 Sample collection

The FALS cohort used in the initial screen of *SQSTM1* mutations consisted of 61 FALS kindred that are known to lack mutations in *SOD1*, *TARDBP*, *FUS*, *VAPB*, *DAO*, *VCP* and *C9orf72*. This is a subset of a larger cohort as described in **Section 1.19.1**. This study employed ALS cases, presenting with motor neuron symptoms and diagnosed as ALS according to EL-Escorial criteria from the United Kingdom (Imperial College Healthcare NHS trust). All ALS patients were positive for a familial history (FALS) and each patient was an index case from a separate kindred. Research governance regulations of the college were satisfied with appropriate informed consent form subjects. Controls included consented UK cases obtained within the Trust and genotype data from samples from the EUR subgroup of 1000 genome project (*http://www.1000genomes.org*) and the European American subgroup of NHLBI EVS (*http://evs.gs.washington.edu/EVS/*). DNA was extracted from whole blood or the buffy coat layer using a DNA extraction kit (Qiagen). In addition a further group of 26 *C9orf72* positive FALS cases were also screened.

4.2.2 Genotyping

All eight exons of the *SQSTM1* gene that encode the major 440 amino acid isoform (SQSTM1-001, NM_003900.4) were amplified using Polymerase chain reaction (PCR). Primers flanking at least 50bp of these exons were designed using Primer 3 program (*http://frodo.wi.mit.edu/primer3*) (**Table 4-1**) and purchased from Invitrogen (UK). PCR

products were purified using Sureclean (Bioline, UK), and sequenced using an ABI Prism BigDye terminator kit (Applied Biosystems, Warrington, Cheshire, UK). The presence of the P392L mutation was confirmed by RFLP analysis using the restriction enzyme BstU1 (New England Biolabs, UK). DNA fragments were separated on a 2% agarose gel (Electran, VWR) and stained with 0.5 µg/ml ethidium bromide for 1 h at 11 V/cm. φX-174-HaeIII (New England Biolabs, UK) was used as DNA marker. Gel profiles were visualized and analysed using a GelDoc system (Bio-Rad UK, Hertfordshire, UK).

4.2.3 Data analysis

Mutation pathogenicity was evaluated using SIFT (*http://sift.jcvi.org*), Polyphen 2 (*http://genetics.bwh.harvard.edu/pph2/*) and Mutation taster (*http://www.mutationtaster.org*). The effects of intronic variants on splicing were analysed using Human splicing finder program (*http://www.umd.be/HSF/*). Genotype and allelic distributions of variants were compared with controls using 2x2 and 2x3 Fisher's tests and subjected to Benjamini-Hochberg correction. To examine genetic background, haplotypes were reconstructed using the Expectation-Maximization algorithm, implemented in PLINK. Linkage disequilibrium was measured by r^2 values calculated using Haploview (*http://www.broadinstitute.org/haploview*). We also analysed the age of onset and survival time of FALS using Kaplan-Meyer curve and log-rank test. The combined effect of pathogenic variants was calculated using the Mantel Haenszel Test.

Table 4-1. Primers used for the Sequencing of *SQSTM1* Gene.

Exon	Primers	Product Size
Exon1	E1F: GCCTCCGCGTTCGCTACA E1R: GTCACCACTCCAGTCACCAG	478
Exon2	E2F: GTCTTGCCTCTCACTCCTGC E2R: CCACACCTGGCCTATGTCTC	396
Exon3	E3F: GGATTCCATGCTGGAGAGCAG E3R: TTCACCTTCCGGAGCCAG	479
Exon4	E4F: ACTTGTGTAGCGTCTGCGAG E4R: TTGTAGGGCACCAGGAAGGT	452
Exon 5	E5F: CACAGGGACCTTGGCAAGAA E5R: TGAGGCAACAAATCCTCACCA	296
Exon 6	E6F: TCTGTAGTCTCCACAGGCCA E6R CTGCAGAGGTGCTGAGGATG	391
Exon 7	E7F: CCCTGCAGCCTTAACTGCAC E7R: TGTCGCTGAAATCAGAGGAGG	397
Exon 8	E8F: CCAAGGCAGCAGGGTATGTG E8R: TGGCTTCTTGCACCCTAACC	372
rs10277	SNP1F:CTGCTGAGGCCTTCTCTTGA SNP1R: GGCCTGACATGGAAGGTGAA	287
rs1065154	SNP2F: GGACTCCATAGCTCCTTCCCA SNP2R: TGTCCCAGCCTGACAGCTT	249

The reaction mixture contained 1× buffer, 1.5 mM Mg^{2+}, 0.1 mM dNTP, 0.5 μM primers, 0.05 U/μl GOtaq Taq DNA polymerase (Promega, UK), and 0.5 ng/μl templates, and thermal cycle was carried out as follow: 94^{o}C for 210s [94^{o}C for 30s, 55^{o}C for 30s, 72^{o}C for 45s] X38 cycles, 72^{o}C for 360s. Successful amplification of Exon 1 requires the addition of 1x PCR Enhancer (Invitrogen, UK) and 45 cycles at annealing temperature of 65 ^{o}C.

4.3 Results

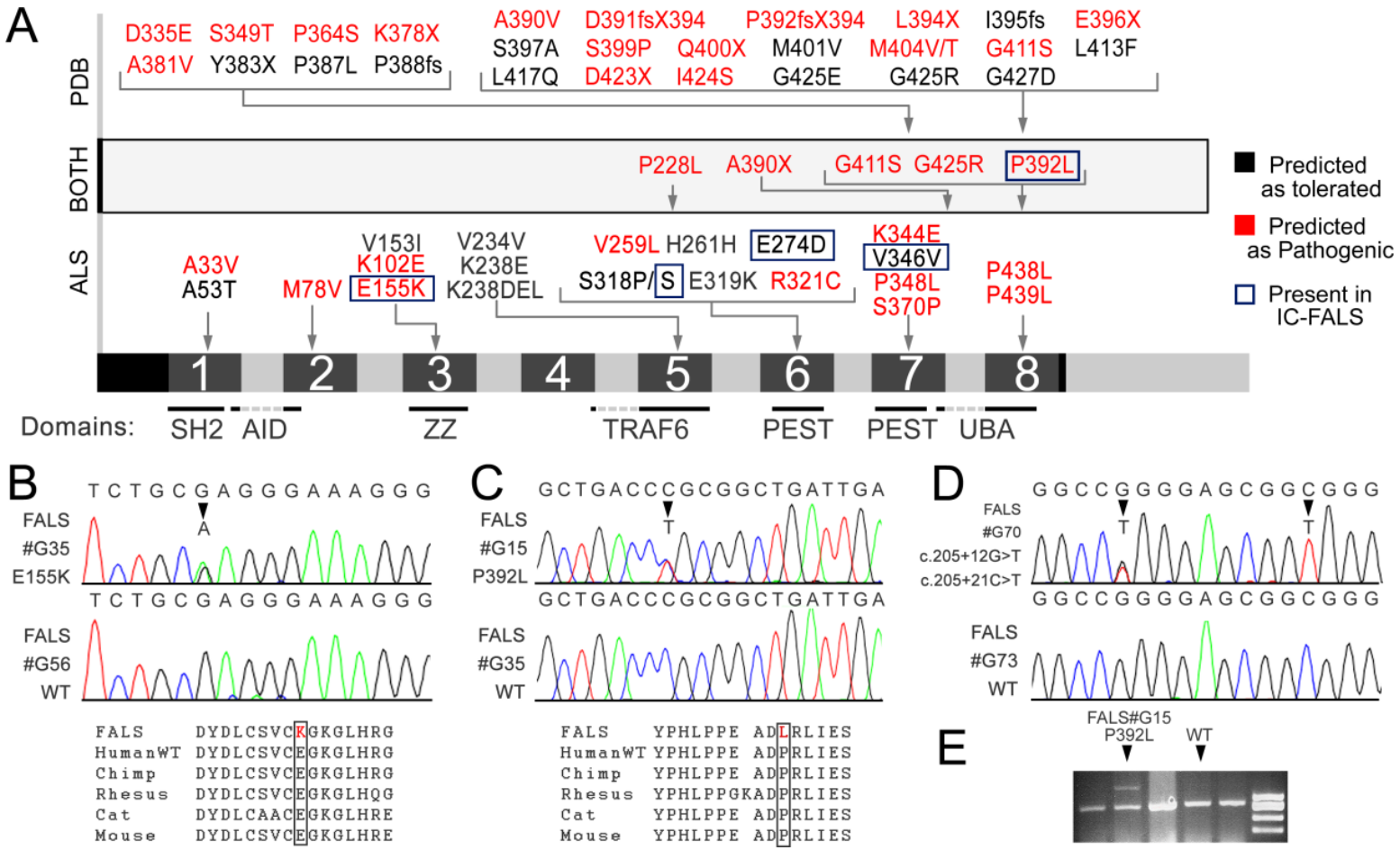
4.3.1 Identification of SQSTM1 sequence variants in a UK-FALS cohort

We screened all eight coding exons of the *SQSTM1* gene (**Figure 4-1 A**), in a cohort of 61 Familial ALS patients. All subjects lacked mutations in *SOD1*, *TARDBP*, *FUS* (exons 14 and 15, which harbour most known mutations found in ALS), *VAPB*, *DAO*, *VCP* and *C9orf72* genes, and each individual was an index case from unrelated families. Six exonic variants c.463G>A (E155K) (**Figure 4-1 B**), c.822G>C (E274D), c.888G>T (P296P), c.954C>T (S318S), c.1038G>A (V346V), and c.1175C>T (P392L) (**Figure 4-1 C**) were identified in 5 FALS index cases, three of which were non-synonymous and three were synonymous. One index case harboured 3 variants (E274D, S318S and P296P) and a second index case harboured 2 variants (E274D and S318S).

The c.822G>C (rs55793208, E274D) substitution has been previously reported in ALS by Rubino et al (Rubino et al., 2012) and results in a glutamate to aspartate change in the PEST domain of SQSTM1. All three synonymous variants have been previously reported, but not in ALS. The E274D and other synonymous variants were predicted to be tolerated, using SIFT, Polyphen2 and Mutation taster programs. However, whilst such programs can predict pathogenicity for recessive mutations, their predictions may not always apply to dominant mutations which may have subtle effects on protein structure or splicing (Flanagan et al., 2010; Valdmanis et al., 2009). There were no significant differences in allele frequency between cases and controls for the E274D and other synonymous variants using a 2x2 Fisher's test.

However, the substitution c.1175C>T (P392L) disrupts a highly conserved sequence in the UBA domain(Geetha and Wooten, 2002) and may abolish the ubiquitin-binding ability of the protein. Using SIFT, Polyphen2 and Mutation taster programs, this mutation was predicted to be pathogenic by all three programmes. To confirm the presence of this mutation, we digested the PCR product with BstU1, which selectively cuts the major allele leaving the mutated allele uncut (**Figure 4-1 E**).

Figure 4-1. *SQSTM1* sequence variants in ALS and PDB.



A Summarizes *SQSTM1* mutations previously reported in ALS and PDB. Probable pathogenic variants are in red and variants in the present study are outlined in blue. Domains are annotated as summarized by Geetha and Wooten (2002): SH2 (src homology 2 binding domain), AID (acidic interaction domain), ZZ (ZZ finger), TRAF6 (binding site for ring-finger protein tumor necrosis factor), PEST (PEST sequence) and UBA (Ubiquitin-associated domain). **B**-**D.** Representative chromatograms showing the heterozygous c.463G>A (E155K), c.1175C>T (P392L), c.205+12G>T and c.205+21C>T variants and control sequences indicated by arrows. The protein sequences are conserved in mammals and the mutated residues are highlighted in red. **E**. The P392L mutation was confirmed by RFLP. Upper band showing the mutation only appeared in lane 2 (FALS#G15), whereas samples with normal sequences (Lane 3, FALS#G35) were completely digested.

The proband bearing the P392L mutation **(P392L Family I, Figure 4-2 A)** was male and diagnosed with ALS at the age of 53 years. Upper and lower motor symptoms were present and there was significant bulbar involvement but no cognitive impairment. The patient had previously been diagnosed with Paget's disease of bone from bone scans and analysis of serum phosphate levels. The patient died of respiratory failure 6 years after diagnosis at the age of 59 years. There was a family history in which the father (IV:11) and paternal aunt (IV:9) of the proband were also diagnosed with ALS. Individual IV:11, presented with prominent lower motor neuron symptoms and bulbar involvement, onset occurred at 60 years and the patient died at 67 years of age. Individual IV:9 demonstrated significant muscle wasting and died at 52 years. However, no DNA was available from either of these cases. We were able to obtain DNA from the brother (V:5) and son (VI:5) of the proband. Individual V:5 was not a mutation carrier and had no symptoms of ALS when 64 years old. VI:5, in contrast, was heterozygous for the mutation but no conclusion on the phenotype can be made yet, as the individual was 21 years at the time of sampling.

We also screened for the P392L mutation in sporadic ALS cases (n=86), controls (n=78) and FALS cases with hexanucleotide expansions in *C9orf72* (n=26). Cases with *C9orf72* expansions were included at this stage as they are known to exist together with other FALS mutations (van Blitterswijk et al., 2012a). One further P392L mutation was detected in these FALS cases and confirmed by sequencing but the mutation was absent from all controls and SALS cases that were screened. This FALS case (**P392L Family B**, **Figure 4-2 B**) showed typical ALS features with prominent bulbar features including marked tongue spasticity. Disease onset occurred at 63 years of age and disease duration was 27 months. Her sibling showed similar disease duration and died at the age of 62 years. Both brothers and parents were unaffected by ALS, the mother died at 86 years and the father in his mid-fifties with nephritis. The brothers both died in their 70s from cancer. There was no evidence of PDB in this kindred.

The overall prevalence of the P392L substitution is 2.3% in our UK-FALS cohort, which is significantly higher than the general population (**P=0.0455**, F**igure 4-3**).

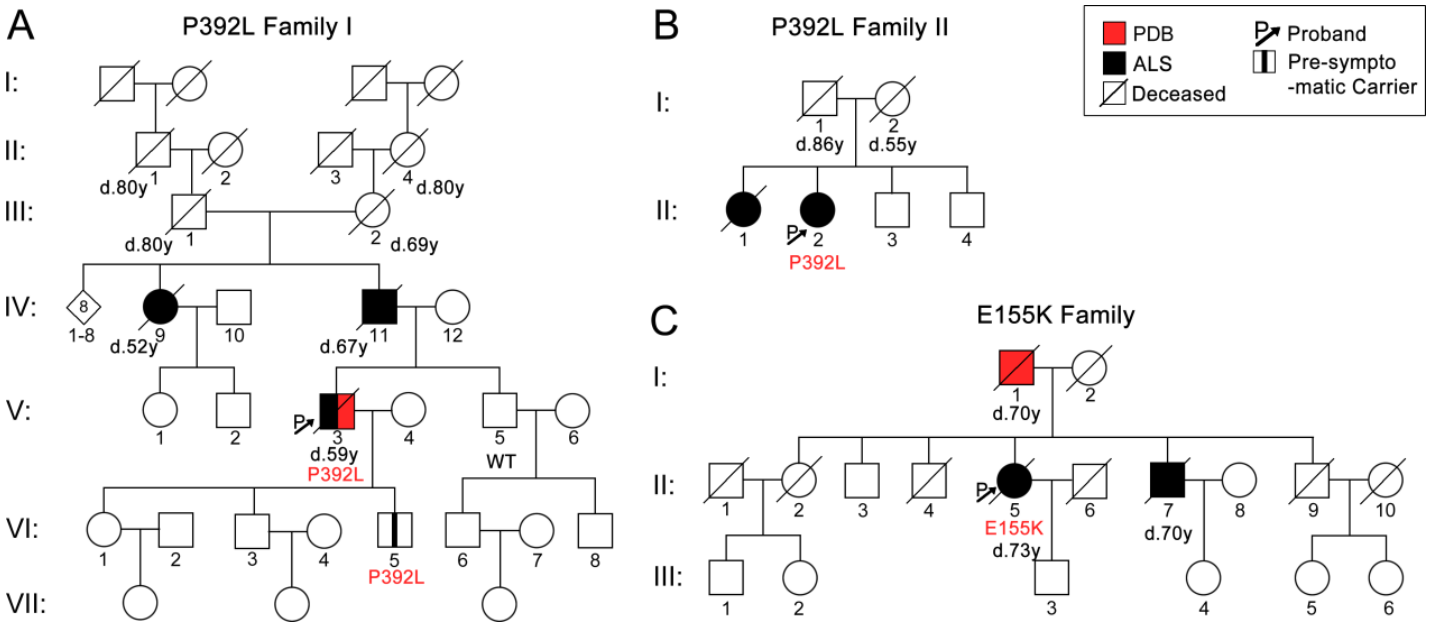
c.463G>A (E155K) is a novel non-synonymous variant found in the ZZ-type zinc finger domain, which is responsible for protein-protein interactions and interacts with RIP in

TNFα/NFkB pathway (Geetha and Wooten, 2002). The mutation is absent from all available databases and was predicted to be pathogenic by all three programmes. The proband was diagnosed with ALS at 71 years and survived for 21 months and was heterozygous for the variant. There was no record of PDB in this case. The father of the proband developed PDB but not ALS, while the brother of the proband was diagnosed with ALS at 69 years. He survived for 12 months and there was no record of PDB in this case (**Figure 4-2 C**). No DNA was available for these samples.

All coding mutations in *SQSTM1* detected in ALS (SALS and FALS) in this and previous studies are presented in **Table 4-2** together with their prevalence in ALS and control populations and pathogenicity predictions. Analysis of all *SQSTM1* variants that are predicted as being pathogenic in published cohorts (Fecto et al., 2011; Hirano et al., 2013; Rubino et al., 2012; Teyssou et al., 2013) and the current study demonstrated that pathogenic *SQSTM1* variants are significantly associated with FALS with a combined odds ratio of 3.85 ($\mathbf{P_{MH}=0.000}3$, **Table 4-3**). No significant association was found for SALS cases. In combination with previous studies, it can be concluded that P392L, which has been identified in 5 out of 524 FALS families, is the most common *SQSTM1* mutation in FALS and PDB to date with a mutation frequency of 2.3% in our UK-FALS cohort and 0.95% in pooled FALS populations published to date (**Table 4-2; Figure 4-3**).

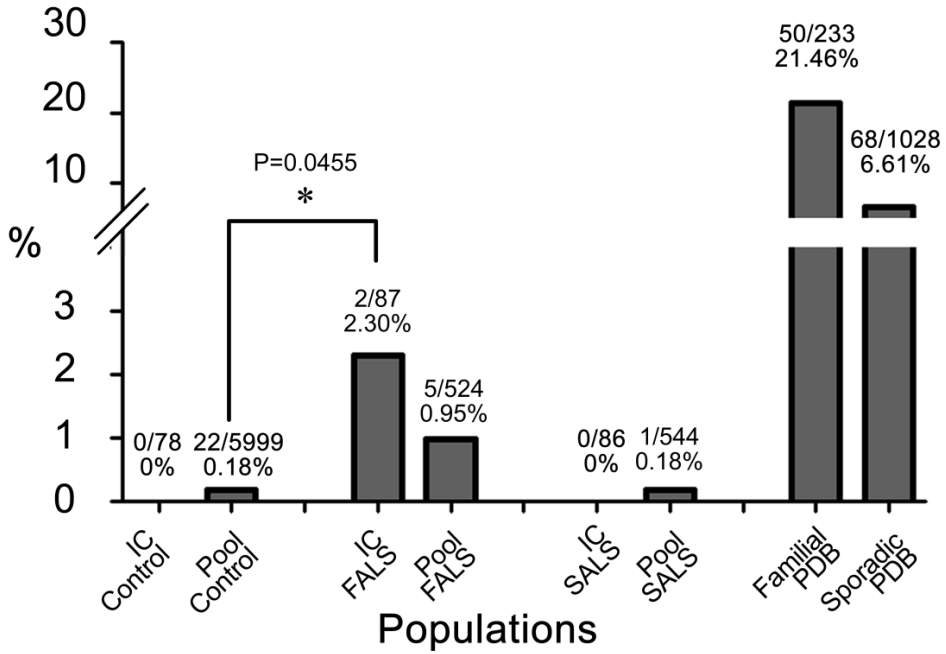
We also identified two novel intronic variants, c.205+12G>T and c.205+21C>T, which are located at 12 bps and 21 bps from the exon 1/intron 1 junction respectively. These variants were present in the same FALS individual, who was heterozygous for the first variant and homozygous for the second (**Figure 4-1 D**). Their effects on splicing were analysed using the Human splicing finder program, which showed that the mutant allele of c.205+21C>T may give rise to a novel donor site with a consensus value of 82.06 (+ 48.59% variation compared to wildtype) and elongate exon 1 by 16 bps.

Figure 4-2. Pedigrees of the FALS index cases carrying P392L and E155K *SQSTM1* mutations.



Age of death and genotypes are indicated where available. **A**. in P392L Family I, individual II:1 and II:4 died of senility, and III:1 and III:2 died of arteriosclerosis and heart disease respectively. DNA was available for three individuals. The index case developed both PDB and ALS, and the phenotypes of generations VI and VII were unknown at the time of sampling. **B**. In Family II, individual I:2 died of nephritis and no DNA was available for other family members. **C**. In the E155K Family, the father of the proband developed Paget's disease and died at 70 years.

Figure 4-3. Mutation frequency of P392L *SQSTM1* in different populations.



The frequency of the P392L substitution is shown as a percentage of individuals carrying this mutation in different populations. IC-groups show the data obtained in the current study. There is a significant increase in prevalence in IC-FALS compared to the combined control group using a 2 x 2 Fisher's test. The frequency in PDB was summarized as reported by Morissette et al (2006) and Rea et al (2009).

Table 4-2. Summary of coding variants in *SQSTM1* in FALS, SALS and FTD from different studies.

Protein	Position	rs	Changes	Population Prevalence				Prediction		
				FALS	SALS	FTD	Control	SIFT	PP	MT
A33V	179248034	rs200396166	c.98C>T	1/498[a]	2/458[a]		6/4418[a,f]	+	-	+
A53T	179248093		c.157G>A		1/458[c]		0/4550[c,f]	-	-	-
M87V	179250011		c.259A>G		1/458[d]		0/4660[d,f]	+	-	+
K102E	179250056		c.304A>G		1/458[d]		0/4660[d,f]	+	-	+
V153I	179251013	rs145056421	c.457G>A		2/458[a]		9/5024[a,f]	-	-	-
E155K	**179251019**		**c.463G>A**	**1/498[e]**			0/4300[e,f]	+	+	+
P228L	179252155	rs151191977	c.683C>T		1/458[a]		4/5024[a,f]	+	-	+
V234V	179252174		c.702G>A	1/498[a]			0/5024[a,f]	-		-
K238E	179252184	rs11548633	c.712A>G		1/458[b]		32/4824[b,f,g]	+	+	+
K238del	179252186-8		c.714-716delGAA		1/458[a]		0/724[a]			+
V259L	179260052		c.775G>C			1/170[b]	0/4445[b,f]	+	-	+
H261H	179260060	rs145001811	c.783C>T		1/458[a]		1/5038[a,f]	-		-
E274D	179260099	rs55793208	**c.822G>C**	**2/498[b,e]**	11/458[b]	5/170[b]	236/4824[b,f,g]	-	-	-
P296P	**179260165**	**rs148984239**	**c.888G>T**	**1/498[e]**			1/4679[f,g]			
S318P	179260229		c.952T>C	1/498[a]			0/5038[a,f]	-	-	-
S318S	179260231	rs56092424	**c.954C>T**	**2/498[e]**			199/4674[f,g]	-		-
E319K	179260232	rs61748794	c955G>A			1/170[b]	2/4817[b,f,g]	-	-	-
R321C	179260238	rs140226523	c.961C>T		1/458[a]		5/5407[a,f,g]	+	-	-
K344E	179260647		c.1032A>G			1/170[b]	0/4805[b,d,f]	+	+	+
V346V	**179260655**	**rs150470670**	**c.1038G>A**	**1/498[e]**			4/4300[f]	-		+
P348L	179260660		c.1044C>T		1/458[b]		0/4805[b,d,f]	+	+	+
S370P	179260725	rs143956614	c.1108T>C	1/498[a]			0/5393[a,b,f]	-	-	+
A390X	179260783		c.1165+1G>A		1/458		-			+
P392L	179263445	rs104893941	**c.1175C>T**	**5/524[a,d,e]**	1/544[a,e]		22/5999[a,b,d,e,f,g]	+	+	+
G411S	179263501	rs143511494	c.1231G>A	1/498[a]			0/5397[a,d,f]	+	+	+
G425R	179263543	CM041449	c.1273G>A	1/498[a]			0/5397[a,d,f]	+	+	+
P438L	179263586		c.1313C>T		1/458[b]		1/4805[b,d,f]	+	+	+
P439L	179263676	rs199854262	c.1316C>T		1/458[c]		0/4910[c,d,f]	+	+	+

Mutations identified in this study are in bold and the frequencies comprise heterozygous and homozygous minor genotypes. All changes shown are exonic except that A390X, also known as IVS7+1 G>A, occurred at the splice site of intron 7. Data from the EUR subgroup of 1000 genome project and EA subgroup of NHLBI Exome sequencing project and 'in house' controls published in previous studies were combined as reference controls. The effects of mutations were predicted using SIFT, Polyphen 2(PP) and Mutation taster (MT): "+" = probable damaging; "–" =tolerated.

[a] Fecto et al (2011) (FALS, n=340; SALS, n=206; control, n=724);

[b] Rubino et al (2012) (SALS, n=124; control=145);

[c] Hirano et al (2013) (FALS, n=7, SALS, n=54; control n=250);

[d] Teyssou et al (2013) (FALS, n=90; SALS, n=74; control, n=360);

[e] Current study (FALS, n=61+26, SALS,n=86, control, n=78);

[f] Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: *http://evs.gs.washington.edu/EVS/*) [Apr, 2013];

[g] 1000 genomes (EUR, n=379).

Table 4-3. Combined allelic frequencies of *SQSTM1* variants that are predicated as pathogenic in cases and controls.

Protein	CTRL	FALS (POOL)			SALS (POOL)		
Changes	MAF%	MAF%	OR	CI_{95}	MAF%	OR	CI_{95}
A33V	0.07	0.10	1.48	0.18-12.3			
M87V	0				0.11	30.54	1.24-750.32
K102E	0				0.11	31.54	1.24-750.32
E155K	0	0.10	25.92	1.06-636.7			
P228L	0.04				0.11	2.7443	0.31-24.58
K238E	0.33				0.11	0.33	0.05-2.41
K238del	0				0.11	4.75	0.19-116.65
S318P	0						
R321C	0.05				0.11	2.36	0.28-20.24
K344E	0						
P348L	0				0.11	31.49	1.28-773.66
S370P	0	0.10	32.51	1.32- 798.5			
A390X	-				0.11		
P392L	0.18	0.48	2.60	0.98-6.88	0.09	0.50	0.07-3.72
G411S	0	0.10	32.53	1.32-799.1			
G425R	0	0.10	33.00	1.34-810.5			
P438L	0.10				0.11	10.50	0.66-168.04
P439L	0				0.11	32.18	1.31-790.57
	χ^2_{MH}		13.23			1.46	
	P_{MH}		**0.0002**			0.2276	
	P_{Exact}		**0.0010**			0.1504	
	Combined OR		3.85	1.86-7.94		1.63	0.83-3.22
	P_{Woolf}		0.1993			**0.0480**	

We analysed the combined risk carried by these variants using Mantel-Haenszel analysis. Odds ratios were calculated for 2 x 2 contingency tables for each variant. It can be summarised that all FALS-associated variants are more frequent in FALS than control with odds ratios >1 and Woolf's test indicated no significant heterogeneity between the odds ratios in FALS. Significant Mantel-Haenszel Test and Exact conditional test indicates that pathogenic *SQSTM1* variants are associated with FALS with a combined odds ratio of 3.85. However, we did not observe this association in SALS. Significant values are bolded. Refer **Table 4-2** for the counts of genotype.

4.3.2 Founder haplotype of P392L SQSTM1 kindred

A four-SNP founder haplotype, H_2, has been previously described to carry most P392L mutations in PDB patients of British descent (Lucas et al., 2005). We carried out haplotype analysis to investigate whether this was also the case in FALS. We first reconstructed the haplotypes in the control population (1000 genome EUR group) where all 4 SNPs are in high LD (r^2>0.8) (**Figure 4-4 A**) showing that H_1 (T-A-C-T) and H_2 (C-G-T-G) accounted for 95% of haplotypes, as previously reported (Lucas et al., 2005) (**Table 4-4**). All P392L mutations carriers in this study were heterozygous for H_2 and $H_{2\alpha}$ (C-G-C-G) haplotypes and the E155K carrier was heterozygous for H_1 and H_2 haplotypes. In the FALS cohort as a whole, there were no significant differences in the frequency of a two-SNP haplotype between cases and controls (**Table 4-4**). Survival data was available for 16 samples and there were no significant differences between haplotypes (**Figure 4-4 B**).

The frequencies of common non-coding variants detected in this study are summarised in **Table 4-5,** the allele frequencies between cases and controls were compared using 2 x 2 Fisher's test and corrected for multiple testing. No marked changes were detected.

All index cases in P392L FALS families and the pre-symptomatic son in P392L S*QSTM1* family I were heterozygous for H_2 and C-G-C-G ($H_2\alpha$) haplotypes. The unaffected brother in P392L family I was homozygous for H_1 haplotype.The index case in the E155K family was heterozygous for the H_1 and H_2 haplotypes.

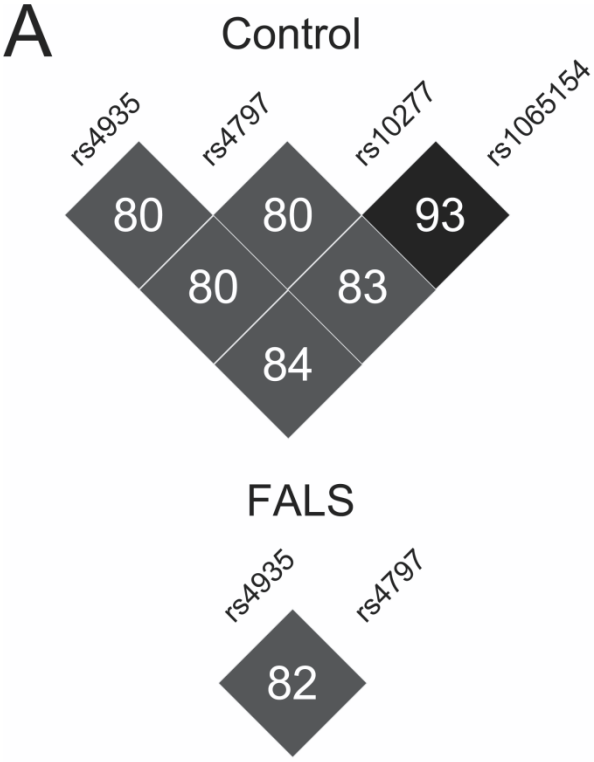
A two-SNP haplotype comprising positions 1 and 2 was reconstructed in both FALS and Control (first column) but there were no significant changes in haplotype frequencies between cases and controls using a 2 x 2 Fisher's test (data not shown).

Table 4-4. Founder haplotypes for *SQSTM1* mutations.

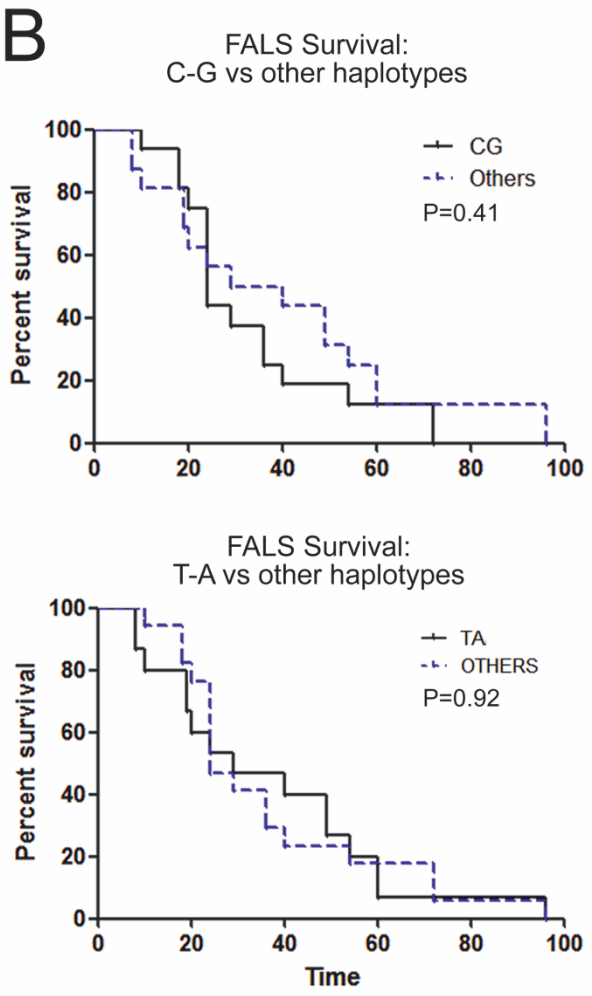
Haplotypes		Frequency (%)	
1, 2	1, 2, 3, 4	Control	FALS
C-G		**48.3**	**40.2**
	C-G-T-G (H_2)	46.2	
	C-G-C-G ($H_2\alpha$)	1.2	
	C-G-C-T	0.9	
T-A		**46.7**	**54.9**
	T-A-C-T (H_1)	45.9	
	T-A-C-G	0.3	
	T-A-T-G	0.5	
T-G		**2.8**	**3.3**
	T-G-T-T	0.1	
	T-G-T-G	1.5	
	T-G-C-T	1.2	
C-A		**2.2**	**1.6**
	C-A-T-G	1.3	
	C-A-C-T	0.9	

Haplotype frequencies of the p62/*SQSTM1*gene are shown. The reported four-SNP haplotype comprising position 1 (rs4935), 2 (rs4797), 3 (rs10277) and 4 (rs1065154) was reconstructed in control samples derived from 1000 genome (column 2), and 4 individuals were heterozygous for the P392L mutation, 2 for K238E, 16 for E274D, 16 for S319S and 2 for R321C. All P392L and K238E were contained within H_2 haplotype, whereas the others were contained within H_1 haplotype. None of these variants were contained within the $H_2\alpha$ haplotype.

Figure 4-4. Linkage disequilibrium and Survival analysis of *SQSTM1* SNPs



A. Linkage disequilibrium (LD) between the SNPs that form the founder haplotype. r^2 values, a measurement of LD, are shown in the boxes.



B. Kaplan-Meyer plots showing the percentage of surviving individuals of different p62/SQSTM1 haplotype at different time intervals. A subgroup of FALS (**n=16**) was used in this analysis. Curves were compared using Log-rank test and P values are shown. No significant differences were observed.

Table 4-5. Common SNPs (MAF > 10%) detected in this study.

	Chr	MAF	Genotype			HWE	Allelic				Genotype	
	(2n)	%	%				P	FDR	OR	CI_{95}	P	FDR
rs4700700 G/A												
EUR	758	6.3	88	11	1	0.05						
FALS	114	13	77	19	4	0.23	0.0177	0.0568	2.24	1.21-4.15	**0.0384**	0.0640
rs2241349 G/A												
EUR	758	32	45	46	9	0.41						
FALS	66	47	27	52	21	1	0.1985	0.1985	1.88	1.13-3.12	**0.0354**	0.0640
rs4935 C/T												
EUR	758	49	26	50	25	1						
FALS	122	58	13	57	30	0.2	0.0794	0.1323	1.42	0.97-2.09	0.0897	0.1121
rs4797 G/A												
EUR	758	49	25	53	23	0.36						
FALS	122	57	11	64	25	0.04	0.1432	0.1790	1.36	0.93-2.00	0.5673	0.5673
rs155787 A/G												
EUR	758	51	23	51	26	0.84						
FALS	40	33	35	65	0	0.06	**0.0227**	0.0568	0.45	0.23-0.89	**0.0097**	**0.0485**
rs10277 C/T												
EUR	758	50	25	50	25	1						
FALS												
rs1065154 G/T												
EUR	758	49	26	50	24	1						
FALS												

Allelic and genotypic frequencies of common SNPs captured in this study are shown. All SNPs are in Hardy-Weinberg equilibrium (HWE) in control. There was significant difference in genotypic frequency of rs155787 using a 2 x 3 Fisher's test. All P values were subjected to Benjamini-Hochberg correction as shown in the FDR column. rs7711505 (intronic), rs753636646 (intronic) and rs155790 (3' UTR) were also detected in this study at similar frequencies in control and ALS populations.

4.4 Discussion

In this study, we screened all exons of the *SQSTM1* gene and confirmed the presence of *SQSTM1* mutations in FALS cases of UK descent. We identified six candidate exonic variants, including a novel E155K mutation, and two novel intronic variants in FALS index cases. There were positive familial histories of Paget's disease of bone in two families carrying either the P392L or the E155K mutation, which were predicted to be pathogenic.

The E274D and other synonymous exonic variants detected in this study were not predicted to be pathogenic. In contrast to the findings of Rubino et al (2012), we found no significant differences in allelic and genotypic frequencies of the E274D or other synonymous variants between our ALS cohorts and controls derived from a UK control cohort and publically available databases. In contrast, the properties of the P392L and E155K mutations are more likely to contribute pathogenic effects. Firstly, these mutations are located in highly conserved domains that are required for the basic functions of p62/SQSTM1. The P392L mutation occurs in the UBA domain which is required for the binding of ubiquitin (Geetha and Wooten, 2002) and modifies the structure of this domain by extending the N-terminus of helix1 (Ciani et al., 2003). This results in a reduced affinity for mono-ubiquitin and shortened polyubiquitin chains (Cavey et al., 2006). The ZZ domain, which accommodates the E155K mutation, is a conserved structural component in different ZIP/p62 homologs and plays an important role in the formation of complexes that participate in the NF-kB signalling pathway (Moscat et al., 2007). Secondly, the coexistence of both PDB and ALS in the two families reported here and in other studies (Teyssou et al., 2013) is indicative that common pathogenic mechanisms underlie both these diseases. p62/SQSTM1 has an established causal role in PDB and the P392L mutation is, in fact, the most common cause of PDB (Morissette et al., 2006; Rea et al., 2009). The P392L mutation segregates with PDB and, although the binding of polyubiquitin remains intact, it causes a phenotype that is indistinguishable from those caused by mutations that truncate the entire domain (Hocking et al., 2004). Thirdly, multiple studies have confirmed the existence of the P392L mutation in ALS from different populations (Fecto et al., 2011; Teyssou et al., 2013) and, in combination with the data presented here, it can be concluded that it is the most common mutation in the gene in ALS to date (**Table 4-2**). The coexistence of the P392L mutation in a family with a *C9orf72* expansion is not unexpected as *C9orf72* has also been found in FALS cases harbouring

SOD1, *FUS* or *TARDBP* mutations (van Blitterswijk et al., 2012a). The *SQSTM1* P392L mutation index case with the *C9orf72* mutation (Family II) showed classical features of ALS with no evidence of cognitive impairment and shorter survival (age at death: 62 to 65 years) compared to Family I that lacked a *C9orf72* expansion (age at death: 52 to 67 years). It is possible that the severity of the disease may have been affected by the co-existence of both mutations but whether the effect of the *SQSTM1* mutation is potentiated by the expansion remains to be established.

The fundamental role of p62 in autophagy is well established and relevant to multiple tissues. However, more recently considerable evidence has accumulated for the specific involvement of p62 in ALS and ALS with fronto-temporal lobar degeneration (ALS/FTLD), where p62 is associated with both TDP-43-positive and TDP-43-negative inclusions in spinal cord and cerebral cortex. Furthermore, similar inclusions have been demonstrated in ALS cases harbouring *SQSTM1* mutations (Teyssou et al., 2013).

In spite of the phenotypic heterogeneity found in ALS, ALS/FTLD and PDB and related disorders such as IBMPFD and POAG, the discovery of mutations in overlapping candidate genes, such as *VCP*, *UBQLN2*, *CHMP2B* and *OPTN*, points to a common pathogenic mechanism affecting the formation and clearance of misfolded proteins (Fecto and Siddique, 2011). In the current study, we also report that the genotypes of Protein disulphide isomerase (PDI), a redox enzyme that reduces formation of aberrant disulphide bonds, are associated with survival time of FALS (**Chapter 5**), supporting the hypothesis that factors interfering with proteostasis may modify disease progression and can be considered as therapeutic targets for ALS. The investigation of p62/*SQSTM1* mutations in multiple cohorts provides further support for their contribution to a significant predisposition to disease in ALS.

-END OF CHAPTER 4-

Chapter 5

Association Studies of ALS Candidate Genes

5.1 General introduction: Association study and its use for identifying susceptibility genetic variants in ALS.

Linkage studies have been successful in allocating oligogenetic causes to a number of Mendelian traits, and such causes may have major contribution to the traits. However, the majority of human diseases cannot be explained by a single genetic cause and appear to be underlined by multiple genetic factors, each exerting partial effects on the traits (Risch, 2000). Although inheritance is sometimes identified in rare forms of common diseases, linkage studies aiming to determinate disease locus in common conditions that are not inherited in a simple Mendelian manner, also known as *complex diseases*, have been unsuccessful (Altmuller et al., 2001). *Association studies*, on the other hand, have taken advantage of a greater statistical power in detecting genetic risk factors for complex diseases and have been widely applied for this purpose in the past few decades. *Genetic association* is a statistical observation of "a tendency of two characters (disease, marker alleles etc.) to occur together at non-random frequencies which can sometimes be caused by *linkage disequilibrium*" (Strachan, 2011). In contrast to *linkage*, which describes the relationship between a marker and the disease locus segregating in affected kindred, the association between a genetic variant and disease is a result of shared ancestral chromosomal segments or *haplotypes* that cause or confer susceptibility of the disease in a population. An association study determines whether an individual carrying specific genetic variants, mostly common *Single nucleotide polymorphisms* (SNPs), is at increased risk of disease (Lewis, 2002).

Different designs for association studies have been described, including family based designs and population based designs (Cordell and Clayton, 2005). A family based *case-parent triad* design compares the genotype transmitted from heterozygous parents to affected individuals to those that are non-transmitted using a transmission/disequilibrium (TDT) test. This design eliminates population stratification, however, the disadvantage is

that it has limited power and requires a large trio cohort that is relatively difficult to obtain compared to a *case-control* design (Cordell and Clayton, 2005). In a population-based *case-control* design, which is the most commonly used design for association studies, allele and genotype frequencies of a SNP are compared between cases and controls using Chi-square or Fisher's tests (Hirschhorn et al., 2002). Other statistical methods, such as Logistic regression, Cochran-Armitage trend test and Permutation tests are also employed with the disease status as the outcome.

The advance in microarray technology has given rise to *Genome-wide association studies* (GWAS), which assesses SNPs throughout the genome at even intervals [reviewed by (Hirschhorn and Daly, 2005)]. Thus, GWAS enables an unbiased assessment of genetic associations with diseases without making assumptions about gene functions and generates massive amounts of genotype data in the populations being studied. However, it has not always been possible to replicate the top-hit association signals in follow-up studies. In addition, association tests in GWAS studies are subjected to a stringent genome-wide threshold for *multiple corrections* and cohorts of large numbers are usually required to gain sufficient *power*. Having said this, it remains possible that SNPs having genuine effects that do not exceed the genome-wide threshold are overlooked in such analyses.

On the other hand, in a *candidate gene approach*, the SNPs being studied are selected based on an *a priori* assumption that they may directly or indirectly contribute to the trait being investigated, such as mediating changes in gene expression or protein function or being in linkage disequilibrium with these changes (Tabor et al., 2002). Although, in the end, the identification of significant associations is not always accompanied by functional changes in mRNA or protein levels, it has been suggested that such results should be considered as informative clues to potential disease pathways rather than evidence of causality (Tabor et al., 2002). In a way, these two approaches are complementary to each other as, with publicly available GWAS databases, such as *dbGAP* (*http://www.ncbi.nlm.nih.gov/gap*) and *GWAS Central* (*http://www.gwascentral.org*), independent *candidate gene* studies can be initiated by taking into account both functional properties of the candidate gene and previous GWAS results, including the below - threshold signals. Indeed, some traditional challenges to the candidate gene approach,

such as population stratification and Type1 error and the lack of SNP functional impacts, can nowadays be overcome *in silico* using sophisticated databases and prediction algorithms (Patnala et al., 2013), which are also taken account in this study.

Despite the identification of more than 18 causal genes in FALS, no major gene has been shown to be causal for SALS. Using the candidate gene approach, a number of genetic risk factors have been reported for SALS, such as Apurinic endonuclease (*APEX*), Angiogenin (*ANG*), Chromatic modifying protein 2b (*CHMP2B*), Dynactin (*DCTN1*), Haemochromatosis (*HFE*), Neurofilaments (*NEFL*, *NEFM*, *NEFH*), Paraoxonase (*PON*), Peripherin (*PRPH*), Progranulin (*PGRN*), SOD1, Survival motor neuron 1 and 2 (*SMN1* and *SMN2*) and Vascular endothelial cell growth factor (*VEGF*) genes [reviewed by (Schymick et al., 2007a)]. Subsequent screening of these genes has indeed led to the identification of mutations in FALS or SALS patients as summarized in **Chapter 1**. In the GWA studies carried out in ALS, SNPs in *FGGY* (Schymick et al., 2007b), *ITPR2* (van Es et al., 2007), *DPP6* (van Es et al., 2008) and *UNC13A* (van Es et al., 2009b) genes have been shown to exceed the genome wide threshold. Noteworthy, in a GWAS study using a Finish ALS cohort comprising both FALS and SALS patients (Laaksovirta et al., 2010), a significant association signal was observed in the 9p21 region, which overlapped with the linked region in ALS/FTD families. The association was mainly driven by FALS and a 42-SNP haplotype was found in both FALS and SALS patients that accounted for most of the association (Laaksovirta et al., 2010; Mok et al., 2012). The *C9orf72* expansion, which was covered by this haplotype, was subsequently identified as the most common cause of FALS/FTD (Renton, 2011). Indeed, SNPs in the 9p21 region were also evident for SALS populations in independent studies, suggesting common variants in this region may be risk factors for both FALS and SALS (Shatunov et al., 2010; van Es et al., 2009b). However, the possibility of the existence of a common founder in SALS with reduced penetrance cannot be excluded (Shatunov et al., 2010), and disease loci other than *C9orf72* may exist in this region as residual association was retained after removal of *C9orf72* positive cases (Jones et al., 2013).

In this study, we carried out two independent association studies on two candidate genes in the telomeric region of the chromosome 17, *P4HB* and *NPLOC4*. Protein disulphide isomerase A1 (PDIA1), which is encoded by the *P4HB* gene, is a redox enzyme that

facilitates the exchange of disulphide bonds and prevents the formation of aberrant protein aggregates, however, this gene was not well-represented in GWA studies. *NPLOC4* is a gene located ~200 kb upstream of *P4HB* and functionally being a cofactor for VCP. Association signals in this gene were shown in previous GWAS studies and this was further investigated in our study. Our results show that SNPs and haplotypes for *P4HB* (**Section 5.2**) and *NPLOC4* (**Section 5.3**) genes are associated in our FALS and SALS cohorts respectively.

5.2 Association studies indicate that *P4HB* is a risk factor in Amyotrophic lateral sclerosis

5.2.1 Introduction

A prominent pathological feature of ALS as well as many other neurodegenerative diseases is the build-up of misfolded or aberrant proteins that might, for example, arise due to oxidative damage, impaired calcium homeostasis, endoplasmic reticulum (ER) stress and/or impaired protein quality control. There is strong evidence demonstrating that ER stress occurs in sporadic cases of ALS and also in animal models of ALS (Atkin et al., 2008; Ilieva et al., 2007) resulting in the recruitment of the unfolded protein response (UPR), through IRE1, ATF6 and PERK pathways inducing the production of ER chaperones and protein disulphide isomerase (PDI, also known as prolyl 4-hydroxylase subunit beta [P4HB]) which facilitate protein folding and transport. In addition, the removal of misfolded proteins is facilitated through ER-associated degradation (ERAD), the ubiquitinated proteasomal system (UPS) and autophagy which are all evident in ALS tissue. However, if ER stress is not resolved, activation of apoptosis is initiated in the ER (Yoshida, 2007). Interestingly, several mutations known to cause familial ALS occur in genes mediating these processes such as VAPB which regulates the IRE1 component of the UPR (Chen et al., 2010; Nishimura et al., 2004a), SOD1 which interacts with components of ERAD, and VCP (Johnson et al., 2010) and UBQLN2 (Deng et al., 2011) that regulate UPS and autophagy. Furthermore, levels VAPB are known to decrease significantly in sporadic cases of ALS (Anagnostou et al., 2010).

In view of the importance of protein quality control in ALS pathogenesis, we have focused on one particular ER component, PDI (PDIA1) that not only prevents the formation of misfolded proteins by facilitating the exchange of disulphide bonds but also has oxidoreductase and protein isomerase activities. PDI together with other ER stress markers are significantly elevated in ALS spinal cord compared to controls and also in the G93A SOD1 mouse model of ALS (Atkin, 2006; Atkin et al., 2008; Massignan et al., 2007). Elevated levels of 5-nitrosylated PDI, an inactivated form, are also found in ALS which suggests that despite the elevated levels, PDI is functionally inactive (Walker et al., 2010). Overexpression of PDI decreases both the accumulation of SOD1 aggregates and also neuronal death in the SOD1 mouse model of ALS, whereas inhibition of PDI using bacitracin increases the formation of aggregates (Atkin, 2006; Atkin et al., 2008). In ALS,

the major inclusions associated with most forms of both sporadic and FALS are the TDP-43-positive ubiquitinated protein aggregates that accumulate in the cytoplasm and recent studies show co-localisation with PDI (Honjo et al., 2011). FUS-positive ubiquitinated inclusions found in a smaller subset of ALS cases also co-localise with PDI (Farg et al., 2012).

Although the application of genome wide association studies (GWAS) has raised expectations of detecting genetic risk factors in sporadic ALS, this has not been realised, despite the use of large cohorts. A limited number of associations have been reported but most have not been replicated. The reasons are multiple, GWAS methodology uses SNPs covering the entire genome and consequently a large statistical correction for the number of SNPs tested (usually 300K to 550K) must be applied. Hence large combined cohorts are necessary which may introduce additional variation due to population stratification of SNP allele frequencies. In practice, GWAS can rarely detect multiple variants or a single rare variant at a single locus (Andersen and Al-Chalabi, 2011) and SNPs within a gene of interest may be poorly represented which is true for *P4HB* where only one tag SNP (rs2070871) is present in the Illumina HumanHap SNP array (550K SNPs). We have used a candidate gene approach which is hypothesis driven and therefore dependent on rigorous standards of biological plausibility. Indeed, the results of a recent simulation study showed that a candidate gene approach has greater power than GWAS (Amos et al., 2011). The evidence outlined above, for a role of PDI in ALS pathogenesis, provides a compelling case for testing the hypothesis that the corresponding gene (*P4HB*) is a risk factor for ALS.

In this study, we have investigated whether *P4HB* is a genetic risk factor for familial or sporadic ALS by analysis of associations between SNP markers and disease and disease severity, assessed by age at onset and duration. Our results show significant genotypic association of SNPs in P4HB, with FALS and that specific haplotypes confer risk for FALS or SALS. We also identified two genotypes and one diplotype that modify disease survival.

5.2.2 Methods for P4HB association study

5.2.2.1 Sample collection.

ALS patients were recruited from the UK (Imperial College Healthcare NHS Trust (ICHT) and Kings College Healthcare NHS Trust (KCHT)), diagnosed according to the El Escorial criteria (Brooks et al., 2000) and satisfying Research Governance regulations for both Institutions. Each FALS patient was an index case from a separate kindred. The control cohort consists of UK subjects of European origin and the EUR subgroup from the 1000 Genome Project Database. Informed consent was given for all samples. DNA was extracted from whole blood or the buffy coat layer using a DNA Extraction kit (QIAGEN), according to protocols provided by the manufacturers.

5.2.2.2 Identification and genotyping of SNPs.

SNPs in *P4HB* were identified in ENSEMBL (*http://www.ensembl.org*) and dbSNP (*http://www.ncbi.nlm.nih.-gov/projects/SNP/*). Only SNPs with Minor allele frequency (MAF) of > 15% in the European population and having natural restriction sites were selected. Primers flanking a 100-bp region around the SNPs were designed using PRIMER 3 program (*http://frodo.wi.mit.edu/ primer3*) and purchased from Invitrogen UK. Details of the primers are shown in **Table 5-1**. PCR was carried out in a standard 30µl volume containing 1X buffer, 1.5mM Mg^{2+}, 0.1mM dNTP, 0.5µM Primers, 0.05U/µl GOtaq® Taq DNA polymerase (Promega, UK) and 0.5ng/µl templates.

SNPs were genotyped using restriction digestion. SNP1 and SNP3 were digested with the restriction enzyme MspI, and SNP 2 and SNP 4 were digested with BstNI and AvaII respectively. PCR products were incubated with 1U of enzyme at appropriate temperatures overnight, after which the DNA fragments were separated on a 2% agarose gel (VWR) stained with 0.5 µg/ml ethidium bromide for 1 hour at 11V/cm. PhiX-174-HaeIII (New England Biolabs, UK) was used as DNA marker. Results were visualized using the GELDOC (Bio-Rad UK, Hemel Hempstead Hertfordshire, UK) system and profiles quantified where necessary. In order to rule out the possibility of genotyping errors, all rare haplotypes not present in controls and ambiguous results were sequenced for confirmation.

Table 5-1. Primers used for PCR of *P4HB* SNPs.

SNPs	Primers	Product Size
SNP 1	F: TCTAGTGGACTCCAGAGAAT R: CGAACCCTGTGTTGTCAC	211
SNP 2	F: GTGAGCTCTGACTTCCAG R: GTTCTGTCTCCATTCTCTG	210
SNP 3	F: CTACATCCAGGCTGGTCCT R: GTGAAATCAAGACTCACATC	242
SNP 4	F: GACTGATCATGGCTCTTG R: TGTAGAGAGGCCAGTGGT	190

5.2.2.3 DNA Sequencing

PCR products were purified with Sureclean (Bioline, UK) according to the manufacturer's instructions, and centrifuged for 40 mins at 4000 rpm. The supernatant was removed and the pellet was resuspended in 70% ethanol, centrifuged again at 4000 rpm for 40 mins and the supernatant was removed. Finally, 10µl H_2O containing 6.4 pmol of primer was added and sequencing was carried out using an ABI Prism BigDye terminator kit (Applied Biosystems, Warrington, Cheshire, UK) according to the manufacturer's instructions. The sequences were analyzed using Codon Code Aligner and Seqdoc programs (*http://research.imb.uq.edu.au/seqdoc*).

5.2.2.4 Statistics

Allelic and genotypic associations, as well as departures from Hardy-Weinberg equilibrium, were tested with 2 x 2 and 2 x 3 Fisher's tests using the PLINK program. Linkage disequilibrium (LD) was estimated using r^2 values which were calculated and represented using Haploview (Barrett et al., 2005). Haplotypes were reconstructed using the Expectation-maximization (EM) algorithm implemented in PLINK. Haplotypes and diplotypes were inferred in a four-SNP window and those with posterior probabilities >0.9 were accepted. Fisher's test was also used to test the associations for each haplotype in a 2 x 2 contingency table comparing the counts of haplotypes in ALS against controls. The P values for both genotypic and haplotypic tests were corrected for multiple comparisons using the Benjamini-Hochberg method. We compared the overall haplotypic distributions

in a 2 x N contingency table using the CLUMP program, a model-free algorithm which assesses significance using Monte-Carlo simulations (Sham and Curtis, 1995). For genotype-phenotype relationships, the frequencies of site of onset were compared for each genotype using the Fisher's test. Data for age at onset and survival were presented in Kaplan-Meyer curves using SPSS and Graphpad and Hazard ratios were estimated using the Cox-regression test. Differences between curves were tested by the log-rank test. The power of the study was evaluated using the G*Power program (Version 3).

5.2.3 Results

We investigated whether SNPs in *P4HB*, which encodes PDI, are associated with disease and/or disease progression in ALS. The ALS cohort consisted of index cases from 200 families, lacking mutations in *TARDBP*, *FUS*, *SOD1*, *VABP* and *DAO* and 282 cases of sporadic ALS, totalling 482 ALS cases (Imperial College Healthcare Trust and Kings College Hospital MND Care and Research Clinic) and a control group of 169 UK individuals (Imperial College Healthcare NHS Trust), which was further expanded with genotyping data from 367 EUR cases originating from the UK, Italy, Finland and Utah in the 1,000 genomes project, giving a total of 536 control individuals (Samples are summarized in **Table 5-2**).

There was no significant population stratification between the control groups (2 x 2 Fisher's test). The power of these cohorts to detect a significant association with FALS was greater than 95% to detect a difference in MAF frequency between cases and controls of >6% (G* power) and >69% for a 4% change in MAF frequency (**Table 5-3**). The familial cases were screened for *C9orf72* expansions using repeat primed PCR (Renton et al., 2011) and this was taken into account in a sub-analysis.

Four SNPs were selected with MAF frequencies of 18 – 19%, which were located in introns 4 and 6 and in the 3'UTR, rs876017, rs876016, rs2070872 and rs8324, referred to as SNP1, 2, 3 and 4 respectively (**Figure 5-1**). None of these SNPs deviated from Hardy-Weinberg (HW) equilibrium in controls. A preliminary analysis of a small FALS cohort (104 index cases and the control cohort) was carried out using all 4 SNPs, which showed an increase in minor allele frequency for all SNPs but significant genotypic association was

only evident for SNPs 2 and 3 and therefore warranted examination in the main study. Preliminary data for SNPs 1 and 4 is shown in **Table 5-4**.

Table 5-2. Description of ALS and Control cohorts used in the *P4HB* study.

Cohorts	No. of individuals	Gender Ratio (Male: Female)
Control	536	1 : 1.13
UK (ICHT)	169	
EUR(1000 Genome Project)	367	
FALS	200	1 : 0.88
UK (ICHT)	104	
UK (KCHT)	96 (SNP2 and 3)	
SALS	282	1 : 0.49
UK (ICHT)	90	
UK (KCHT)	192 (SNP2 only)	
Total	***1018***	

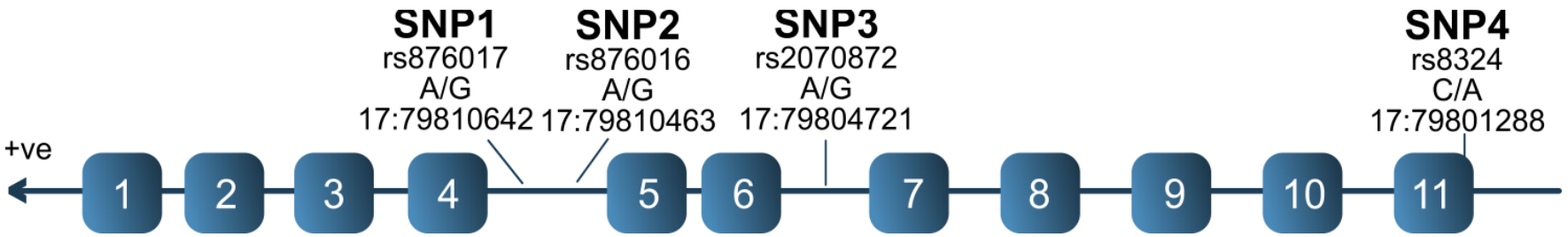
DNA samples of cases and controls were obtained from existing cohorts. To evaluate the effect of gender on genetic associations, we compared allelic and genotypic distributions of the SNPs between Male and Female controls and no significant differences were observed.

Table 5-3. Power calculation for the *P4HB* association study.

Control MAF	Increase in MAF			
	4%	5%	6%	7%
19% (SNP3)	69.13%	87.78%	96.67%	99.39%
18% (SNP2)	71.06%	89.15%	97.26%	99.54%

Power of FALS study for different effect sizes, as measured by w-values and control MAFs, where α=0.05 and total sample size=736 (excluding SALS), are shown.

Figure 5-1. Details of *P4HB* SNPs investigated in this study.



Alleles on the coding strand are shown.

Table 5-4. Genotypic and allelic associations of *P4HB* SNPs 1 and 4.

	Chr (2n)	MAF %	Genotype %			Allelic P	Genotype P
SNP 1 (rs876017)			*AA*	*AG*	*GG*		
Control	1066	18	67	30	3		
FALS	208	21	60	38	2	0.2843	0.2329
└C9 -ve	140	21	59	40	1	0.3531	0.2233
SALS	180	19	64	33	2	0.7546	0.8288
Pool ALS	388	20	62	36	2	0.3613	0.2800
SNP 4 (rs8324)			*CC*	*CA*	*AA*		
Control	1062	19	66	31	3		
FALS	202	19	64	33	3	0.8442	0.9336
└C9 -ve	136	22	60	35	4	0.3535	0.5328
SALS	180	24	60	32	8	0.1035	0.1248
Pool ALS	382	21	62	32	5	0.2285	0.3939

Genotype frequencies are summarized in the left panel. The FALS group contains patients with and without the hexanucleotide expansion in the *C9orf72* gene, whereas the C9–ve group contains only samples lacking the expansion. All P values were calculated using the Fisher's test and are uncorrected for multiple testing.

5.2.3.1 Genotype Analysis of P4HB yielded Significant Genotypic Associations with ALS.

A full analysis of familial cases and controls was carried out using two SNPs, rs876016 and rs2070872, referred to as SNPs and 2 and 3 respectively. Significant genotypic associations with FALS were evident for both SNPs using 2 x 3 tests, following FDR correction using the Benjamini-Hochberg method (**Table 5-5**). There was an increase in heterozygote frequency for SNPs 2 and 3 of 11% and 13% respectively which gives rise to the differences in genotype distribution. In addition to genotypic association, SNP 3 also showed deviation from HW equilibrium in FALS cases (P = 0.0008). Using a 2 x 2 test to compare heterozygote frequency, the A/G genotype of SNP 3 (**$OR_{AG/AA}$=1.67, P=0.0014**) was shown to confer a significant risk of FALS.

There were no significant associations with SALS cases but when all ALS cases were pooled (FALS and SALS), significant (FDR corrected) allelic and genotypic associations were found for SNP2 and a significant genotypic association for SNP 2 (**Table 5-5**). This result suggests that it is reasonable to conclude that the minor alleles of SNPs 2 and 3 are risk factors for ALS.

Table 5-5. Genotypic and allelic associations of *P4HB* SNPs.

	Chr	MAF	Genotype			Allelic				Genotype	
	(2n)	%	%			P	FDR	OR	CI_{95}	P	FDR
SNP 2 (rs876016)			*AA*	*AG*	*GG*						
Control	1064	18	68	29	3						
FALS	394	25	55	40	5	**0.0039**	**0.0155**	1.51	1.15-2.00	**0.0100**	**0.0198**
└C9 –ve	268	25	54	42	4	**0.0153**	0.0612	1.49	1.09-2.05	**0.0153**	**0.0307**
SALS	530	22	60	36	4	0.0511	0.2045	1.29	1.00-1.68	0.1134	0.2495
Pool ALS	924	23	58	37	5	**0.0037**	**0.0148**	1.39	1.11-1.72	**0.0082**	**0.0280**
SNP 3 (rs2070872)			*AA*	*AG*	*GG*						
Control	1056	19	66	30	4						
FALS*	384	23	56	43	1	0.1816	0.3632	1.22	0.92-1.62	**0.0011**	**0.0046**
└C9 -ve	256	24	53	46	1	0.1189	0.2379	1.30	0.94-1.80	**0.0012**	**0.0047**
SALS	178	24	60	34	6	0.1868	0.2491	1.30	0.89-1.89	0.3385	0.4513
Pool ALS*	562	23	57	40	3	0.0929	0.1858	1.25	0.97-1.60	**0.0153**	**0.0306**

Genotypic and allelic associations of *P4HB* SNPs. Genotype frequencies are summarized in the left panel. The FALS group contains patients with and without the hexanucleotide expansion in the *C9orf72* gene, whereas the C9–ve group contains only samples excluded for the expansion. All p values were calculated using Fisher's test and significant associations are in bold. The FDR columns show p values corrected for 4 SNPs using the Benjamini-Hochberg FDR correction. *There were no deviations from HW equilibrium except for SNP 3 FALS and Pool ALS groups, $P < 0.0008$ and 0.001, respectively.

5.2.3.2 Rare Haplotypes are associated with FALS and SALS.

Four SNP haplotypes, which included two additional flanking SNPs, rs876017 (SNP 1) and rs8324 (SNP 4), were examined in ALS cases compared to controls. SNPs 1-4 form an LD block in controls extending across the *P4HB* gene from exon 5 to the 3'UTR as indicated from r^2 values between these SNPs which were > 0.95 in the control samples obtained from the 1000 genome database. High LD was also observed in the matched controls (Imperial College Healthcare NHS Trust). However, LD was decreased in both FALS and SALS (**Figure 5-2A**). Haplotypes and diplotypes were reconstructed using PLINK and the frequencies are summarized in **Table 5-6** and **Table 5-7** respectively. Individuals possessing rare haplotypes were sequenced to ensure correct genotyping.

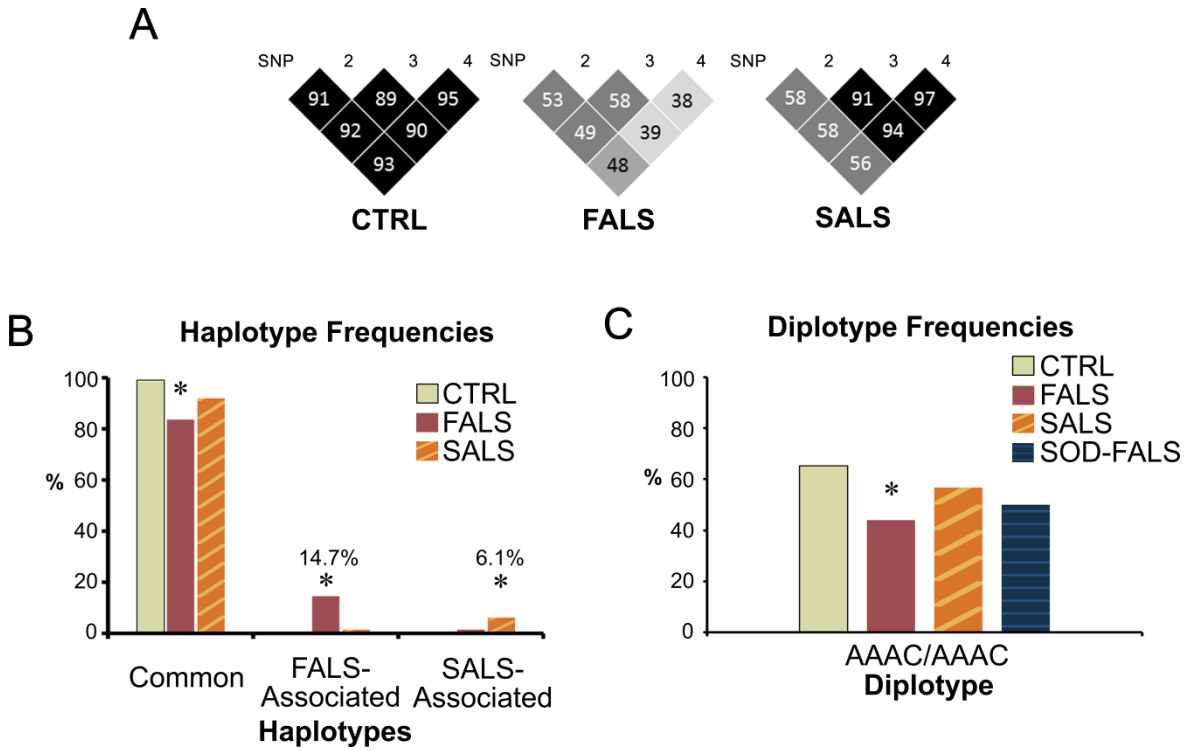
In total, 15 haplotypes were identified in cases and controls. The haplotype AAAC, which is composed of major alleles, was the most common in all groups, but its frequency was significantly decreased by 11.9% in FALS (**P_{FDR}=0.0013**) compared to controls. Consistent with the changes in LD, rare haplotypes were more prevalent in cases than in controls. Amongst these, seven haplotypes were significantly overrepresented in FALS and one haplotype was significantly over-represented in SALS (**Table 5-6**). The combined frequencies of the corresponding risk haplotypes were 14.7% in FALS, 6.1% in SALS and 0.5% in controls (**Figure 5-2B**).

Since the functions of PDI that are affected in ALS have been primarily focused on the folding of mutant SOD1, we also analysed the *P4HB* haplotype in 20 additional FALS cases possessing SOD1 mutations. One haplotype, AAGC, was more prevalent in SOD1-FALS than in controls. Remarkably, AAGC, which was not found in SALS, was found in 4 individuals with SOD-FALS with a frequency of 10% (**P=0.0025**) but only in 1% of non-SOD1-FALS cases, indicating a specific association with *SOD1* mutations. We then used the CLUMP program (Sham and Curtis, 1995) which contains an omnibus permutation test to examine differences in haplotype distributions between ALS groups and control (**Table 5-6**). This analysis showed that the specific haplotype distributions for FALS and SALS that did not occur in controls under 10^8 simulations (**P<0.0001**). We did not find any haplotypes exclusively associated with the *C9orf72* expansion, and no significant difference in distribution was observed between expansion positive and negative groups.

AAAC/AAAC was the most common diplotype in all groups, and most of the remaining individuals were heterozygous. The frequency of AAAC/AAAC was significantly decreased

in FALS (**P=0.0015**) (**Figure 5-2C, Table 5-7**), suggesting that individuals lacking the homozygous haplotype are at higher risk of FALS in the UK population. This association can be largely attributed to the risk haplotypes as most of the over-represented diplotypes contained one of the risk-haplotypes. We identified 3 FALS cases, which were homozygous for rare FALS-risk haplotypes. Moreover, the only individual who was homozygous for a rare haplotype in SALS also carried the SALS-risk haplotype, AGGA. Neverthe less the most significant association with FALS contained one common haplotype and one FALS-risk haplotype.

Figure 5-2. SNP LD, haplotype and diplotype associations.



A. r^2 values for controls were >89% which indicates a high LD, whereas values in FALS were consistently reduced**.** In SALS, the LD between SNP1 and other SNPs were decreased but the other three SNPs were still in LD, suggesting a recombination between SNP1 and SNP2. Haplotype analysis showed that more rare haplotypes were present in FALS and SALS than in controls.

B. Haplotypes were grouped according to the presence of significant association with ALS shown in **Table 5-6**, and the combined frequency of these haplotypes is represented for each phenotype (control, FALS, SALS and FALS with SOD1 mutations. The FALS group consisted of the seven haplotypes significantly associated with FALS and indicated in red in **Table 5-6**. The SALS associated haplotype consisted of one haplotype significantly associated with SALS and the FALS-SOD1 associated haplotype consisted of the one haplotype associated with FALS-SOD1, as indicated in **Table 5-6**. The corrected P values were for FALS and SALS were **1.8E-17** and **3.6E-7** respectively.

C. The diplotype AAAC/AAAC was less prevalent in FALS than controls, with corrected P values of **0.0018**.

Table 5-6. Phased haplotypes for *P4HB* in ALS ranked according to their frequencies in controls.

HAP	CTRL	FALS						SALS		SOD-FALS	
	%	All %	FDR	C9+ %	FDR	C9- %	FDR	%	FDR	%	FDR
AAAC	80.5	68.6	**0.0013**	71.2		67.4	**0.0039**	73.9		65.0	
GGGA	17.2	12.7		10.6		13.8		17.2		12.5	
GAGA	0.7										
AAGC	0.5	1.0		1.5		0.7				10.0	**0.0025**
AGGA	0.4	1.5				2.2		6.1	**5.42E-06**	5.0	
AAGA	0.3										
AGAC	0.2	3.9	**0.0001**	6.1	**0.0022**	2.9	**0.0063**	0.6		2.5	
GGAC	0.1	1.0		1.5		0.7		0.6			
GGGC	0.1	3.4	**0.0001**	4.5	**0.0055**	2.9	**0.0039**				
AAAA	0.1	2.0	**0.0086**			2.9	**0.0039**			2.5	
GAAA		1.0	**0.0485**	1.5		0.7					
GGAA		2.0	**0.0025**	1.5		2.2	**0.0056**				
AGGC		1.0	**0.0485**			1.4	**0.0282**			2.5	
GAAC		1.5	**0.0103**	1.5		1.4	**0.0282**	1.1			
AGAA		0.5				0.7		0.6			
Omnibus		**P< 0.0001**						**P< 0.0001**		**P=2.3E-4**	

Each haplotype was tested with a 2 x 2 Fisher's test with FDR correction for 15 tests. Significant P values are shown in bold. The data shows that the most common haplotype, highlighted in black has a frequency of >65% in all populations and was significantly decreased in FALS, in which rare haplotypes were more prevalent. Seven haplotypes were significantly associated with FALS and these are highlighted in red. One haplotype was significantly associated with SALS and is highlighted in orange. One haplotype was found in 10% of SOD-FALS individuals and is highlighted in brown. Significant P values were obtained in Omnibus tests for haplotypic distributions between cases and controls. Non-significant FDR values are not shown.

Table 5-7. Diplotype frequencies of *P4HB* SNPs.

	Diplotype Frequencies %		P for AAAC/AAAC	FDR	OR	CI_{95}
	AAAC/AAAC	Heterozygous Diplotypes				
Control	65.2	31.8				
FALS	44.0	54.0*	**1.18E-4**	**0.0015**	2.38	1.54-3.67
SALS	56.7	41.1	0.1244	0.5383	1.43	0.91-2.25
SOD-FALS	50.0	50.0	0.2322	0.8625	1.87	0.76-4.58
Pool ALS	50.0	47.4	**0.0002**	**0.0070**	1.87	1.34-2.62

Diplotypes were grouped according to whether they were homozygous for AAAC, tested for associations using the 2 by 2 Fisher's test and corrected for 30 tests. The percentage of samples lacking the homozygous diplotype was significantly higher in FALS and Pooled ALS and the odds ratio was highest in FALS. * Heterozygous diplotypes are significantly more prevalent in FALS (P=3.5E-5), and all FALS-associated diplotypes were heterozygous for the AAAC haplotype (data not shown).

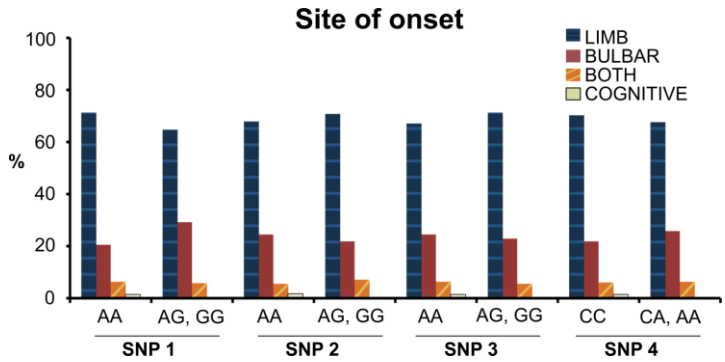
5.2.3.3 Genotypes and Haplotypes of P4HB affect the Survival of ALS.

We also examined the genotype-phenotype relationships between the *P4HB* gene and ALS. Clinical details, including gender, age of onset, age of death and site of onset, were available for a subset of FALS and SALS patients as summarized in **Table 5-8**. None of the SNPs were associated with site of onset (**Figure 5-3**) or gender. We then investigated whether the genotypes of each SNP affected survival. Affected cases were grouped into two groups according to the presence of risk alleles and the survival curves were compared using the Log-rank test. In FALS, the minor alleles of SNPs 3 and 4 were associated with shortened survival times (median=21 months for SNP 3; 24 months for SNP 4) compared to the major allele (median=29 months for SNP 3; 29 months for SNP 4, **Figure 5-4 A and B**). Since the lack of the AAAC/AAAC diplotype was associated with a higher risk of FALS, we tested whether this also modifies survival. We found that FALS patients with the AAAC/AAAC diplotype have significantly longer survival time (median=34.5 months) compared with those lacking this diplotype (median=22 months, **Figure 5-4 C**). This correlation was not affected by gender. However, the risk haplotypes did not have any significant effects on survival time, suggesting a protective role of the common genotypes and haplotype in FALS. Furthermore, in order to investigate whether the *C9orf72* expansion had an effect on survival, we performed a Cox-regression in which both *P4HB* genotypes and the presence of the *C9orf72* expansion were treated as covariates, but no significant interaction was observed (**Table 5-9**).

Table 5-8. Clinical details available for phenotype analysis in this study.

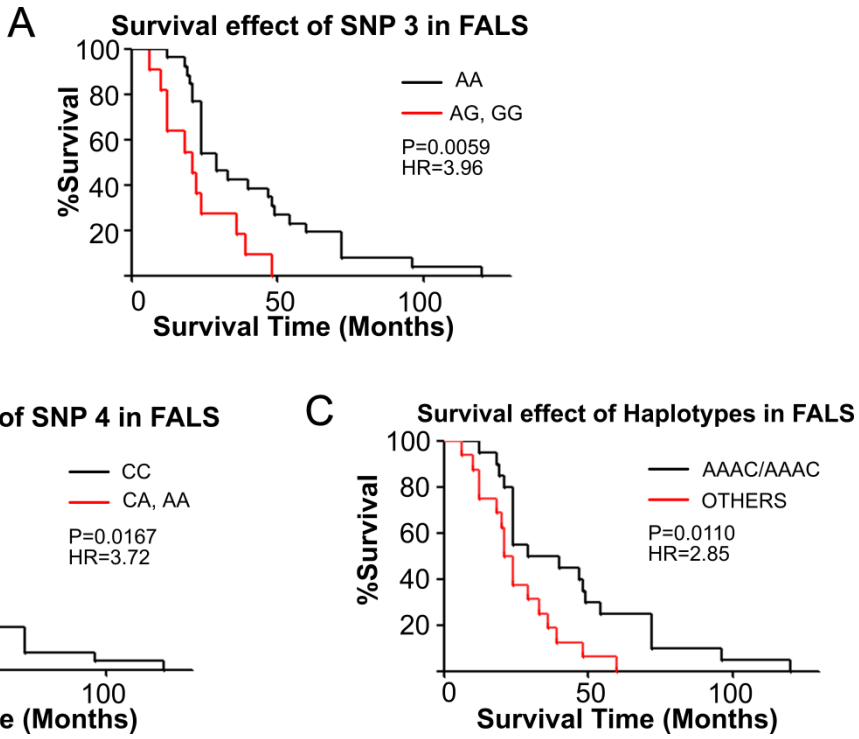
Clinical details	No. of patients with clinical details	
	FALS	SALS
Age of onset	142	24
Survival Time	37	23
Site of onset		
Limb	49	18
Bulbar	18	5
Both	6	
Cognitive	1	

Figure 5-3. Effects of *P4HB* genotype on site at onset of disease.



Frequencies of each site at onset are shown for each SNP genotype. No significant differences were found.

Figure 5-4. Kaplan-Meyer curves showing the survival times of FALS patients with different *P4HB* genotypes.



A. Effect of SNP 3 genotypes AA compared to AG and GG combined on survival, with the following number of individuals, n_{AA}=26, $n_{AG,GG}$=11, **B**. Effect of SNP 4 genotypes CC compared to CA and AA combined on survival, with the following number of individuals, n_{CC}=27, $n_{CT,TT}$=9. **C**.Effect of diplotype on survival for the diplotype containing only major alleles AAAC/AAAC compared to the remainder with the following number of individuals, $n_{AAAC/AAAC}$=20, n_{others}=16. Nominal p values. P vales and Hazard ratios values are given for each comparison.

Table 5-9. Cox-regression models used to analyse the combined effects of *C9orf72* and *P4HB* on the survival of FALS.

Models	Covariants	Exp (B)	P
*SNP3 + C9+ SNP3*C9*	C9	1.247	0.730
	SNP3*C9orf72	0.648	0.559
*SNP4 + C9+ SNP4*C9*	C9	0.620	0.555
	SNP4*C9orf72	1.359	0.734
*AAAC + C9+ AAAC*C9*	C9	1.833	0.245
	AAAC*C9	0.431	0.239

Cox-regression was used to analyse the combined effects of *C9orf72* and *P4HB* on the survival of FALS. In the first two rows, genotype of SNP3, SNP 4, the presence of the expansion and their interaction were included in the model, and the significances for the coefficients are shown. In the last row, the presence of the homozygous major haplotype, AAAC, was used instead of a single SNP genotype. Exp (B) represents the changes in hazard ratio explained by a unit change of the dichotomous covariates.

5.2.4 Discussion: P4HB

In this study, we found significant genotype associations with FALS for two SNPs in the *P4HB* gene (rs876016 and rs2070872). This is a novel finding. No significant association was detected in SALS cases which is consistent with GWAS studies, although this gene is poorly represented in the widely used SNP arrays. All subjects are of European ancestry and cases and controls were matched to eliminate population stratification with no departure from Hardy-Weinberg equilibrium in controls. In fact, the deviation from Hardy-Weinberg equilibrium found for one associated SNP (rs2070872) in FALS cases provides extra support for association with disease (Deng et al., 2001; Gyorffy et al., 2004). The FALS cohort consisted of index cases lacking mutations in *SOD1*, *TARDBP*, *FUS*, *VAPB* and *VCP* but included cases with *C9orf72* expansions. A significant genotypic association for these two SNPs was retained after removal of *C9orf72* cases despite the reduced sample size indicating that the *P4HB* may have a stronger genotypic association in patients without the expansion.

Haplotype analysis was used to define the haplotypes that conferred risk of susceptibility for FALS and SALS. The haplotypes were tested using the 2 x 2 Fisher's test individually

and subjected to correction for multiple comparisons. In all groups, the most frequent haplotypes were composed of either all major alleles or all minor alleles, but greater variation in heterozygous haplotypes was seen in the FALS and SALS groups. Most importantly, there was a highly significant decrease in the frequency of the common haplotype, AAAC, in both the combined FALS cohort and FALS cases lacking the *C9orf72* expansion. Interestingly, among the haplotypes that were significantly elevated in cases, two were exclusive to ALS. One haplotype, AAGC, was selectively over-represented in *SOD1*-FALS, and, considering the known functional link between PDI and *SOD1* mutations, it is possible that this haplotype may influence penetrance in these individuals. Based on the haplotype data, the seven FALS haplotypes significantly associated with FALS were grouped together as a FALS-risk group and the SALS associated haplotype was referred to as the SALS-risk group for subsequent analysis. The frequencies of these two groups were each significantly different from controls (**Figure 5-2B**). Diplotype analysis showed that most of the risk-haplotypes existed in the form of heterozygotes, and most diplotypes that are significantly increased in ALS also contain at least one of the risk-haplotypes.

The survival data, which was available for a subset of patients, was presented in Kaplan-Meyer plots and the Hazard ratios were estimated using the Cox-regression test. The finding that the presence of the minor allele of SNP 3 (rs2070872) and the lack of the homozygous AAAC diplotype are associated with reduced survival is suggestive of a protective effect for the AAAC haplotype. The Hazard ratios also indicate that there is a **2.85** fold risk of reduced survival after onset in FALS patients possessing the minor allele of SNP3, compared to the major allele (**Figure 5-4 C**). Again, in order to account for the effect of the *C9orf72* expansion, it was included as a covariate in the regression but no significant effect was seen.

Based on these findings, we propose that *P4HB* is a genetic modifier for ALS. PDI is a multifunctional protein that catalyses the formation, cleavage and rearrangement of disulphide bonds, through the oxido-reductase properties of its two thioredoxin domains. PDI also functions as a chaperone that inhibits aggregation of misfolded proteins. PDI has been implicated in ALS, as together with other UPR markers it is up-regulated in the spinal cord of both SALS cases and ^{G93A}SOD1 transgenic mice, in which PDI preferentially

binds to SOD1 aggregates. Several studies have shown the co-localisation of PDI with TDP-43, FUS or SOD1-positive ubiquitinated protein inclusions in ALS (Farg et al., 2012; Honjo et al., 2011) .

The mechanisms involved in the formation of SOD1 aggregates have been extensively investigated and show that reduction of disulphide bonds plays a crucial role in the nucleation process, which is followed by the formation of mature fibrils via non-covalent interactions between intact SOD1 molecules (Chattopadhyay et al., 2008). The nucleation process limits the aggregation rate of SOD1 (Chia et al., 2010), which is related to the survival of FALS patients (Wang et al., 2008). Hence an inhibitory effect of PDI on protein aggregation may extend this process and prolong the survival of ALS. The neuroprotective action of PDI is further supported by the effect of deletion of a PDI regulator, reticulon-4A, which accelerates motor neuron degeneration in the SOD1 transgenic mouse (Yang et al., 2009).

The role of PDI in redox signalling pathways may also contribute to ALS pathogenesis. Many studies have shown that oxidative damage to proteins, lipids and DNA, caused by ROS occur in ALS and may affect disease progression (Barber et al., 2006). In addition to catalysing the generation of ROS via a variety of aberrant reactions, mutant SOD1 also prolongs the activation of NADPH oxidase, one of the main sources from which ROS is generated, increasing the generation of ROS (Ferraiuolo et al., 2011). Remarkably, PDI is involved in both generation and consumption of ROS. In the presence of ER stress, H_2O_2 can be generated through the PDI/Ero1 system as PDI undergoes oxidation, however, H_2O_2 is also consumed through the Prdx4/PDI and Gpx7/8/PDI systems which also oxidize the protein (Laurindo et al., 2012). Therefore, pathogenicity may be caused when the homeostasis maintained by PDI is disrupted. Indeed, it is possible that the normal function of PDI is disrupted in ALS. Walker et al (2010) demonstrated that, as also happens in Parkinson's disease and Alzheimer's disease, the active cysteine residues of PDI are S-nitrosylated in SALS and G93A transgenic mouse spinal cords, which inactivates the enzyme. Although none of the SNPs used in this study are located in coding regions, it is possible that the haplotype block associated with ALS may contain other genetic variations associated with gene function.

Several modifier genes related to the redox signalling pathway in ALS have been proposed. Knockdown of the *Nox1 or Nox2* genes, encoding NADPH oxidase, prolongs the survival of the G93A transgenic mouse, where homozygous deletion of Nox2 led to a four-fold increase in survival index (Marden et al., 2007). In addition, Mitchell et al (2009) reported the association of two SNPs located in the thioredoxin reductase 1 gene (*TXNRD1*) with FALS. This gene, encoding (ThxR1) catalyses the NADPH-dependent reduction of thioredoxin (Powis et al., 1995) and was also proposed as a modifier gene as a minor allele in one SNP was associated with the age of onset of FALS (Mitchell et al., 2009). PDI and thioredoxin share similar active site sequences and they both can be reduced by thioredoxin reductase (Lundström and Holmgren, 1990). Taken together, our findings are consistent with the known cytoprotective effects of PDI demonstrated in previous studies and suggest that DNA variants in this gene linked to the SNP markers used in this study may impair its physiological function. Further investigations are needed to further characterize the modifier effect of this gene and therapeutic potential of this protein.

5.3 Association study of the *NPLOC4* Gene:

5.3.1 Introduction

Using a candidate gene approach that focused on genes involved in oxidative protein folding, we reported the associations of *TXRDN1* and *P4HB* with FALS. We then annotated and examined the telomeric region of chromosome 17, where *P4HB* locates, for association signals with SALS in a publicly available GWAS dataset and found that the signals were enriched in a gene locates ~200 kb upstream of *P4HB*-- the *NPLOC4* gene.

NPLOC4 encodes the protein nuclear protein localization protein 4 homolog, Npl4, which is ubiquitously expressed in the central nervous system and was originally known to have a regulatory role in nuclear import. The Npl4 protein contains an ubiquitin associated domain, a zinc finger domain, a Npl4-homology domain and a C- terminal zinc-finger domain (NZF) (Lass et al., 2008). Npl4 forms a heterodimer with Ufd1 and serves as a co-factor of Valosin containing protein (*VCP*), a known causal gene for FALS. VCP is involved in multiple cellular processes and its function depends on the binding of different co-factors. The most well-characterized function of the VCP complex with Npl4 and Ufd1 lies in the transfer of ubiquitinated misfolded proteins from the ER for proteasomal degradation, a process known as endoplasmic reticulum associated protein degradation (ERAD). This process requires the binding of the cofactor Npl4-Ufd1, which contain ubiquitin binding domains recognising lys-48 linked polyubiquitin chains (Meyer et al., 2002). Ubiquitin binding by Npl4 is mediated by the NZF domain. Accumulation of misfolded proteins in the ER induces ER stress, a well-established feature of ALS. Furthermore, ER stress can also be induced by the over-expression of mutant Npl4 proteins (Lass et al., 2008).

TDP-43 is a nuclear protein, but in the pathological state it is depleted from the nucleus and forms toxic cytoplasmic aggregates, which are hallmarks of most SALS and FTD cases. The typical hallmark of both ALS and FTLD are cytoplasmic ubiquitinated protein inclusions containing the nuclear protein, TDP-43, suggesting a defect in nuclear import or protein clearance (Neumann et al., 2006). Similar TDP-43- positive cytoplasmic inclusions are also associated with *VCP* mutations present in multiple phenotypes of multi-systematic proteinopathy including ALS and FTD, Paget's disease and Inclusion body myositis (IBMPFD). TDP-43 possesses a nuclear localization signal (NLS), which causes

cytoplasmic aggregations when disrupted (Winton et al., 2008), and is imported to the nucleus through by interacting with karyopherin-β1, a transport receptor in the Nuclear pore complex (NPC) (Nishimura et al., 2010). Npl4 is another structural component of the NPC and regulates nuclear protein import (DeHoratius and Silver, 1996). Moreover, the importance of nuclear import in the pathogenesis of ALS can be illustrated by the observation that mutations impairing the activity of a component in the nuclear import machinery, Ran-GTPase, redistributes FUS, a TDP-43 homolog that also forms cytoplasmic aggregates in ALS, to the cytoplasm (Ito et al., 2011a).

An important function of TDP-43 is regulating the maturation of pre-mRNA. TDP-43 possesses two RNA recognition motifs, RRM 1 and RRM 2, found in a variety of RNA binding proteins. These motifs preferentially bind UG rich intronic regions and regulate the splicing and expression of its targets (Tollervey et al., 2011). Npl4 is one of these targets. The expression of an alternatively spliced Npl4 isoform lacking the last exon, which encodes the NZF domain, was increased in SALS and, as a consequence, the ubiquitin-binding ability of the VCP-Npl4-Ufd1 complex may be impaired, exacerbating ER stress (Xiao et al., 2011). Furthermore, amplification of a CNV containing a fragment of *NPLOC4* has been found in a single SALS patient (Pamphlett et al., 2011).

In order to investigate whether common variants of the *NPLOC4* gene are risk factors of ALS, we genotyped six SNPs in the Imperial College ALS and Control cohorts and showed that 3 of them are risks factors for SALS. Distinct gender effects were observed. We also hypothesized that a TG dinucleotide repeat adjacent to the alternatively spliced exon, c.1669+1387TG(11_30), may be responsible for the TDP-43 biding and confer risks of ALS. We showed that this repeat is polymorphic but failed to identify any alleles associated with SALS.

5.3.2 Methods for NPLOC4 association study

5.3.2.1 Subjects

The study included 89 FALS and **97** SALS UK cases recruited from Imperial College Healthcare NHS Trust) and **174** additional SALS cases recruited from and Kings College Healthcare NHS Trust. Research governance regulations for both institutions were

fulfilled. The control populations contained **276** neurologically normal European individuals from the U.K. (IC-Controls) and **379** healthy samples from the EUR subgroup of 1000 Genomes Project (*http://www.1000genomes.org/*). These panels provide **>90%** power to detect a **5%** change in minor allele frequency for SNPs 1-3 in SALS (**Table 5-10**).DNA was extracted from whole blood or the buffy coat layer using a DNA extraction kit (QIAGEN, UK), according to protocols provided by the manufacturer.

Table 5-10. Power calculations.

	Control MAF %	% Increase				
		4	5	6	7	8
(SNP2)	19	0.8734	0.9724	0.9965	0.9997	1
(SNP3)	24	0.8133	0.9455	0.9897	0.9988	0.9999
(SNP1)	30	0.7569	0.9131	0.9785	0.9964	0.9996

The power of detecting alterations in MAFs in our cases (excluding FALS) and controls (*Total 926 Samples*) when α=0.05.

The Selection and Genotyping of Single nucleotide polymorphisms (SNPs)

Six SNPs were selected (MAF>15%) showing LD (r^2>0.9) with other SNPs and present in across five separate LD blocks (only SNPs 4 and 5 were present in the same LD block) in the European population. PCR primers are shown in **Table 5-11**. Genotyping was carried out using restriction digests or Kompetitive Allele Specific PCR (KASP™), a PCR based method using fluorescently labelled allele-specific primers. For the restriction digests, 30µl PCR products were incubated with 1U of enzyme for 16 hours as follows: DdeI for SNP1, PflMI for SNP2, NlaIV for SNP4, HinfI for SNP5. KASP™ was used for SNPs 3 and 5 that lacked restriction sites and was carried out by LGC Genomics (UK) following the instructions provided for sample preparation. A subset of samples was genotyped by both methods and no mismatch was observed confirming genotyping accuracy.

Table 5-11. Primers used for polymerase chained reaction of the *NPLOC4* SNPs.

SNP	Primers	Product Size
rs6565612	FOR: GCTTCCCAAAGTGCTGGGAT REV: ACCAGTCCCTCGTTAAGTTGTG	265
rs8075102	FOR: CAGGGCCTACAAAGAGGCAG REV: TCGTTGTACTCAGGTATTCTCTGG	281
rs7405450	FOR: ATTGGAGCTCTCAGACCCGA REV: GGAGTTGGATCTGTAACTGACTGT	299
rs3934711	FOR: TGTAAGTTGTGGCAGTTTGCAT REV: CCACTGCTCAGACCTCTTCC	241
rs9912074	FOR: CCAGGTCCTGAACGATTCCC REV: GAGGTTGCAGTGAGGCAAGA	302
Primers for TG repeat	FOR: 6-FAM-5'-TTGCCTGATGCTGAGGTTGG REV: 6-FAM-5'-CAGGAGGCAGAGTTTGCAGT	203

Primers used for polymerase chained reaction. The PCR thermal cycles were carried out as following: 94^{o}C for 3m30s, [94^{o}C for 30s, 55^{o}C for 30s, 72^{o}C for 45s] x35 cycles, 72^{o}C for 5m. 6-FAM: 6-carboxyfluorescein.

5.3.2.2 Fragment analysis

The genotypes of the dinucleotide repeat, c.1669+1387TG(11_30), were analysed using fragment analysis. A DNA Fragment containing the repeat was amplified using a 6-carboxyfluorescein (6-FAM) labelled forward primer and unlabelled reverse primer. Capillary electrophoresis was carried out using ABI 3730x1 DNA Analyzer by the Imperial College Genomic Core Facility. Electrophoregrams were visualized using GeneMapper v4.1 program (Life Technologies).

5.3.2.3 Statistics

The power of the study was evaluated using G*Power program (v3). Allelic, genotypic associations and departures from Hardy-Weinberg equilibrium (HWE) were tested using 2 x 2 and 2 x 3 Fisher's tests using in PLINK program v0.99 (*http://pngu.mgh.harvard.edu/~purcell/plink/*) and R. Mantel-Haenszel analysis and Woolf's test were employed for association tests in stratified genders. Linkage disequilibrium (LD) was estimated using r^2 values calculated by Haploview program

(Barrett et al., 2005). Haplotypes were phased in 3-SNP and 5-SNP windows using PLINK, which adopts the Expectation–Maximization algorithm, and those with the highest posterior probability were accepted. The haplotypes were also tested for deviations from HWE using a program based on Likelihood ratio tests (LRTs) under models causing the deviations (Mao et al., 2013). Each haplotype was tested for association using 2 x 2 Fisher's test. All association tests were corrected for multiple comparisons using the Benjamini-Hochberg method. The program CLUMP, a model-free algorithm using Monte Carlo simulations, was used for omnibus tests for Haplotypic distributions. Distribution of TG the repeat between cases and controls were tested using Mann-Whitney-U test. The genotype and haplotype's effects on age of onset and survival times were investigated using Kaplan-Meyer curve and Hazard ratio computed by GraphPad. Differences between curves were tested by the log-rank tests.

5.3.3 Results

5.3.3.1 GWAS SNPs tagged for independent haplotype blocks in the *NPLOC4* gene

Tagged SNPs within the region of the *NPLOC4* gene were examined for enriched association signals using a publically available GWAS dataset of ALS (dbGAP study accession no: phs000101.v3.p1) (Schymick et al., 2007b). Significant associations (uncorrected) were found for SNPs *rs6565612* (**SNP1**) and *rs7405450* (**SNP3**), the latter being the most significant SNP in the region (**P=0.0019**). In the EUR population (*http://www.1000genomes.org/*), these tagged SNPs ($\mathbf{r^2 \geq 0.9}$) arise from different non-overlapping haplotype blocks (Blocks 1 and 3, **Figure 5-5 A**). SNP *rs8075102* (**SNP2**), which showed significance in a follow-up GWAS study (Chio et al., 2009), tagged an independent block (Block 2) between Blocks 1 and 3. SNPs *rs3934711* (**SNP4**) and *rs9912074* (**SNP5**), although not highlighted by the GWAS studies, were located in the same block (Block 4) that covers a region implicated in alternative splicing and locates immediately adjacent to a putative TDP-43 binding site (**Figure 5-5 A**). To characterize the haplotype in this region, a downstream exonic SNP *rs4073997* (**SNP6**), was also investigated. Subjects analysed in this study were comprising **89** FALS index cases (Imperial College), **271** SALS cases (97 from Imperial College Health Care Trust and 174

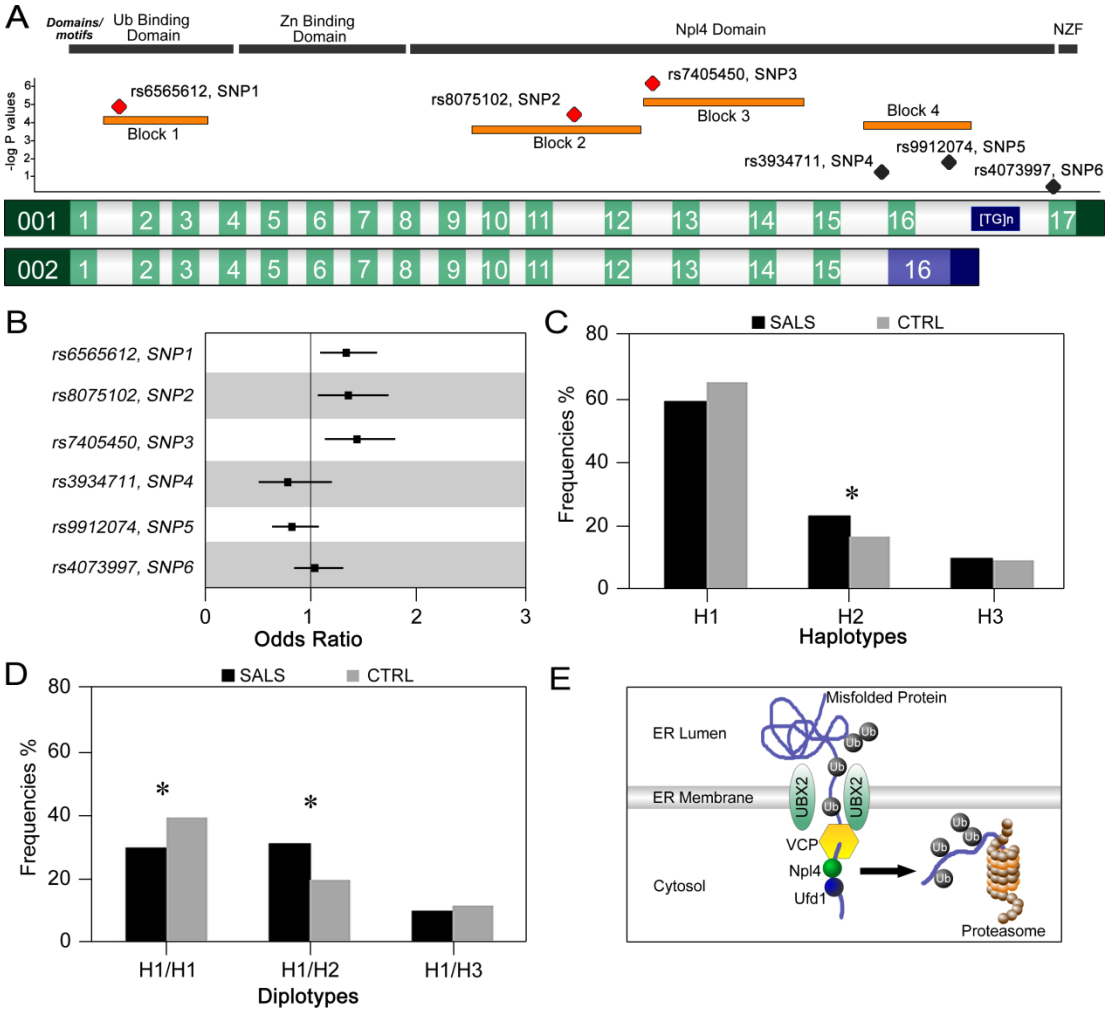
from King's College London) and **276** normal controls (Imperial College London). Controls from the EUR population (**n=379**) were also included in the analysis.

5.3.3.2 SNPs in *NPLOC4* gene are significantly associated with Sporadic ALS

SNPs 1, 2 and 3 were genotyped in SALS and control cohorts, All SNPs were in HWE in controls and no significant population stratification was seen between the controls (Imperial College London) and the EUR population (using 2 x 2 Fisher's tests). Minor allele frequencies of SNP 1, SNP 2 and SNP 3 were increased in SALS, yielding significant allelic and genotypic associations after Benjamini-Hochberg correction (**P_{FDR}=0.0307** for SNP1**, P_{FDR}=0.0307** for SNP 2**, P_{FDR}=0.0131** for SNP3) (**Table 5-12**). SNPs 1 and 2 were further studied in the FALS cohort (**Table 5-13**) but no significant association was evident although the association of SNP1 and 2 remained significant in pooled ALS (**P_{FDR}=0.0277** for SNP1**, P_{FDR}=0.0367** for SNP 2). Of note, SNP 2 deviated from HWE in FALS (**P=0.0360**) and SALS (**P=0.0013**). Odds ratios of these associations in SALS ranged from 1.32 to 1.97 (**Figure 5-5 B**).

As SNP 4 is in complete LD with SNP 5 in both SALS and Controls (**r^2=1,** Details not shown), only the latter SNP was genotyped in the full set of samples. No significant allelic or genotypic associations were found for these two SNPs (**Table 5-13**).

Figure 5-5. SNPs, haplotype and diplotypes in the *NPLOC4* gene.



A shows the structure and domains of the *NPLOC4* gene (NM_017921.2). The constitutively spliced transcript 001 (ENST00000331134) encodes a 608 amino acid protein, whereas the alternatively spliced transcript 002 (ENST00000374747), in which exon 17 is replaced with an incomplete intron 16, encodes an elongated protein of 617 amino acids. Haplotypes tagged by the *NPLOC4* SNPs span different functional domains and are indicated in orange. Block 1 resides the Ub-binding domain, which interacts with VCP, whereas Block 2, 3 and 4 are in the conserved Npl4 domain. The NZF domain, which allows ubiquitin binding, is lost in transcript 002. The TG repeat located between SNPs 5 and 6 is indicated in blue and the –log P values of the SNPs in the current study are indicated in red. **B**. Odds ratios and CI_{95} between SALS and Controls from the current study are indicated by squares and lines respectively. **C** shows the frequencies of the three most common haplotypes and (**D**) diplotypes (3-SNP window) in SALS and controls. * indicates P<0.05 in 2 x 2 Fisher's tests after multiple corrections. **E** shows the role of NPLOC4 as a co-factor for VCP during ERAD.

Table 5-12. SNP associations of *NPLOC4* SNPs.

	Chr (2n)	MAF %	Genotype %			Allelic P	FDR	OR	CI_{95}	Genotype P	FDR
rs6565612, SNP 1			AA	AG	GG						
Control	1262	30	49	42	9						
♀	644	29	51	40	9						
♂	594	30	48	43	9						
ALS	718	35	41	48	11	**0.0139**	**0.0277**	1.28	1.06-1.56	**0.0250**	**0.0250**
SALS	540	36	38	52	11	**0.0110**	**0.0307**	1.32	1.07-1.64	**0.0112**	**0.0223**
♀	176	40	33	55	13	**0.0076**	**0.0147**	1.63	1.15-2.30	**0.0088**	**0.0176**
♂	344	34	42	49	9	0.3087	0.7933	1.17	0.88-1.56	0.4027	0.7719
rs8075102, SNP 2			CC	CA	AA						
Control	1282	19	66	31	3						
♀	658	17	59	28	3						
♂	600	20	53	34	3						
ALS*	720	23	56	42	2	**0.0367**	**0.0367**	1.27	1.02-1.59	**0.0012**	**0.0024**
SALS**	542	24	55	43	2	**0.0153**	**0.0307**	1.36	1.06-1.73	**0.0017**	**0.0102**
♀	176	26	50	48	2	**0.0098**	**0.0147**	1.69	1.14-2.50	**0.0022**	**0.0066**
♂	344	23	57	41	2	0.4086	0.7933	1.15	0.83-1.58	0.3234	0.7719
rs7405450, SNP 3			AA	AG	GG						
Control	1280	24	59	35	6						
♀	654	20	65	30	6						
♂	602	27	53	40	7						
SALS	508	31	48	43	9	**0.0022**	**0.0131**	1.44	1.14-1.80	**0.0079**	**0.0223**
♀	172	34	44	44	12	**0.0004**	**0.0023**	1.97	1.37-2.85	**0.0018**	**0.0066**
♂	318	29	51	41	8	0.5864	0.7933	1.10	0.81-1.49	0.8136	0.9674
rs4073997, SNP 6			GG	GC	CC						
Control	1272	31	50	42	10						
♀	652	27	54	39	7						
♂	596	35	43	45	12						
SALS	490	32	46	44	9	0.6877	0.6877	1.05	0.84-1.31	0.7365	0.7365
♀	158	38	39	46	15	**0.0061**	**0.0147**	1.68	1.17-2.42	**0.0207**	**0.0311**
♂	318	29	49	44	7	0.0763	0.4577	0.77	0.57-1.03	0.1462	0.7719

The associations of SNPs 1-3 and 6 are summarized. Genotype and allelic frequencies are shown in the left side. Significant values are in bold. All SNPs had >10% MAF. No deviation from HWE was noticed in controls and SNP2 deviated in cases: * P=7.72E-5; ** P=0.0013. The allelic and genotypic associations were calculated using 2 x 2 and 2 x 3 Fisher's tests, respectively, and subjected to Benjamini- Hochberg corrections (FDR). Significant P values are in boldface. Female and males are denoted by "♀" and "♂" and the associations stratified by gender are shown separately. The gender split was taken account in multiple corrections.

Table 5-13. Non-associated *NPLOC4* SNPs (SNPs 4, 5 and SNP 1, 2 for FALS).

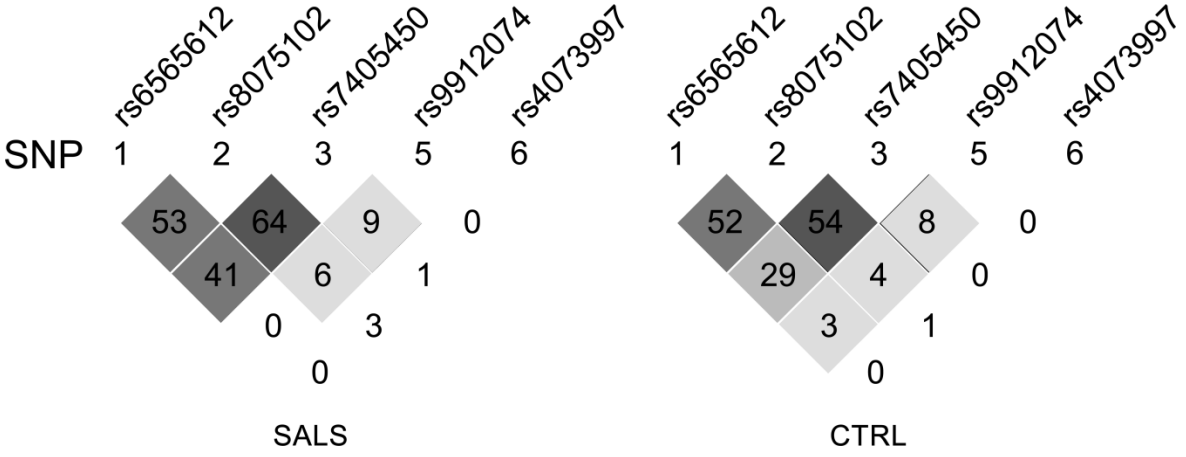
	Chr	MAF	Genotype			Allelic				Genotype	
	(2n)	%	%			P	FDR	OR	CI_{95}	P	FDR
rs6565612, SNP 1			AA	AG	GG						
FALS	172	33	47	42	12	0.4788	0.9576	1.14	0.81-1.60	0.6529	0.6529
♀	86	38	37	49	14	0.0796	0.1592	1.53	0.96-2.45	0.1700	0.1700
♂	80	24	60	33	8	0.2422	0.2972	0.71	0.41-1.22	0.4207	0.5761
rs8075102, SNP 2			CC	CA	AA						
FALS	172	19	63	37	0	1	1	0.99	0.66-1.49	0.1365	0.2729
♀	86	21	58	42	0	0.4524	0.4524	1.26	0.72-2.21	0.1261	0.1700
♂	80	15	70	30	0	0.2972	0.2972	0.69	0.36-1.32	0.5761	0.5761
rs3934711, SNP 4			CC	CT	TT						
EUR	758	22	61	34	5						
♀	402	23	60	35	5						
♂	356	21	63	33	4						
SALS	172	18	65	34	1	0.3014	0.3617	0.78	0.51-1.20	0.3042	0.4263
♀	44	9	82	18	0	**0.0338**	**0.0406**	0.34	0.12 - 0.97	0.1506	0.1807
♂	110	22	58	40	2	0.7907	0.7933	1.06	0.63 - 1.79	0.5146	0.7719
rs9912074, SNP 5			TT	TC	CC						
Control	1272	21	63	33	4						
♀	652	22	60	35	5						
♂	596	19	66	30	4						
SALS	518	18	68	29	3	0.1702	0.2553	0.82	0.63-1.07	0.3552	0.4263
♀	166	17	70	27	4	0.1370	0.1370	0.70	0.45-1.10	0.2860	0.2860
♂	334	18	67	29	4	0.7933	0.7933	0.94	0.67-1.33	0.9674	0.9674

SNP frequencies and of non-associated SNPs are summarized, in complementary to **Table 5-12**. Significant values are in bold.

5.3.3.3 Linkage disequilibrium and Haplotype Analysis

No apparent LD amongst SNPs 1, 2, 3, 5 and 6 was observed in cases or controls (**Figure 5-6**). In view of the significant associations that were found, we investigated whether a specific haplotype was associated with SALS. Five SNPs were phased in 3-SNP and 5-SNP windows and haplotype frequencies were compared using 2 x 2 fisher's tests with Benjamini-Hochberg corrections. The omnibus tests for haplotype distributions were significant for both windows (**Table 5-14**). In the 3-SNP window containing SNPs1, 2 and 3, three haplotypes accounted for **>90%** of chromosomes analysed, A-C-A (H_1), G-A-G (H_2) and G-C-A (H_3), and these were the most common haplotypes found in both cases and controls (**Table 5-14**). There was no deviation from HWE based on likelihood ratio tests for different causal models (Data not shown) (Mao et al., 2013). The frequency of H_2, comprising all minor alleles, was increased by **6.5% in SALS** (**P_{FDR}=0.0106**, **Figure 5-5 C**). Therefore, we propose that H_2 is a risk haplotype. In the 5-SNP window, only the haplotype G-A-G-T-G showed significant association (**P_{FDR}=0.0283**). Furthermore, the diplotype analysis showed that homozygosity of H_1 (H_1/H_1 haplotype) was decreased in SALS (**P_{BH}=0.0222**). The H_2 risk haplotype mainly existed in a heterozygous form of H_1/H_2, which was also increased in cases (**P_{BH}=0.0072**) (**Figure 5-5 D**).

Figure 5-6. Linkage disequilibrium between *NPLOC4* SNPs.



r^2 values between the NPLOC4 SNPs are indicated by grayscale. SNP4 was in complete LD (r^2=1) with SNP5 in both SALS and Controls and and therefore not shown.

Table 5-14. *NPLOC4* Haplotypic associations.

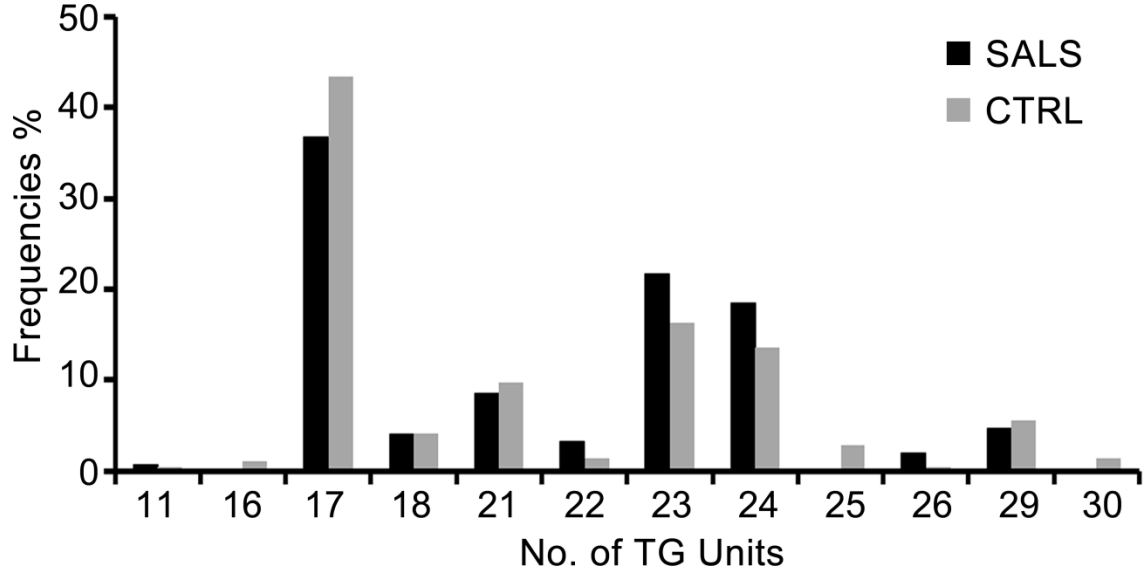
Windows	Haplotype		Frequencies%		P	FDR
			CTRL	SALS		
3	ACA (H_1)	+++	65.1	59.2	**0.0194**	0.0516
5	ACATG	+++++	37.1	32.4	0.0592	0.2393
5	ACATC	++++-	17.0	17.6	0.7325	0.9418
5	ACACG	+++-+	9.8	8.9	0.5984	0.9179
5	ACACC	+++--	0.7	0.0	0.0665	0.2393
3	GAG (H_2)	---	16.9	23.4	**0.0013**	**0.0106**
5	GAGTG	---++	14.4	20.6	**0.0016**	**0.0283**
5	GAGTC	---+-	2.2	2.7	0.6119	0.9179
3	GCA (H_3)	-++	9.3	10.0	0.6630	0.8278
5	GCACG	-++-+	3.1	1.9	0.2053	0.4619
5	GCACC	-++--	6.0	6.8	0.5211	0.9179
5	GCATG	-++++	0.2	0.9	0.0505	0.2393
5	GCATC	-+++-	0.7	0.6	1.0000	1.0000
			3-SNP Window Omnibus:		**0.0043**	
			5-SNP Window Omnibus:		**0.0065**	

The common haplotypes, phased in 3-SNP (SNP 1, 2, 3) and 5-SNP (SNP 1, 2, 3, 5, 6) windows, in SALS (n=271 for 3 SNPs, n=264 for 5 SNPs) and Controls (n=645) are shown**.** These account for >90% of haplotypes in cases and controls. "+" denotes major alleles and "-" denotes minor alleles. Each haplotype was tested using 2 x 2 Fisher's test and the 3-SNP and 5-SNP haplotypes were corrected for 8 and 18 comparisons respectively using Benjamini-Hochberg procedure (FDR). Significant P values, which are in boldface, were retained in two 3-SNP haplotypes, H_1 and H_2, and one 5-SNP haplotype, CAGTG. Model-free omnibus tests were calculated for the both windows.

5.3.3.4 Frequencies of [TG]n repeats

Intronic TG repeats are probable binding sites for TDP-43, which regulates alternative splicing (Polymenidou et al., 2011). To investigate whether a TG dinucleotide repeat located in intron 16 (**Figure 5-5 A**), c.1669+1387TG(11_30), is associated with SALS, fragment analysis was carried out in a subset of **76** SALS and **144** controls. The repeat was polymorphic. We observed the presence of 12 alleles ranging from 11 to 30 repeats, in which the reference allele, $[TG]_{17}$, was the commonest in both cases and controls (**Figure 5-6**). No alleles or genotypes of this repeat confer additional risk of SALS in our cohorts.

Figure 5-6. Genotype frequency of the TG repeat in the *NPLOC4* gene.



Fragment analysis of the TG dinucleotide repeat, c.1669+1387TG(11_30), was performed in a subset of Imperial College controls (n=144) and SALS (n=76) and the allelic frequencies are shown. 9 alleles were found in SALS (median= 21.5 repeats) and 10 alleles were found in Controls (median= 21 repeats).

5.3.3.5 Gender and age of onset effects of *NPLOC4* SNPs

When cohorts were analysed by gender, very marked allelic and genotypic differences were seen in females for SNPs 1-3 (P_{FDR}= **0.0147**, OR= 1.63 for SNP 1; P_{FDR}=**0.0147**, OR=1.69 for SNP 2; P_{FDR}=**0.0023**, OR=1.97 for SNP 3), whilst in contrast SNP associations in males were diminished. In addition, a significant allelic association of SNP 6 was found in females SALS cases alone (P_{FDR}=**0.0147**). Associations of SNPs 1, 2, 3 remained significant when gender was accounted using a Mantel-Haenszel analysis and heterogeneity of odds ratio was evident for SNP 3 and 6 following a Woolf's test (**Table 5-15**).

When haplotype analysis was stratified for gender, a **12.7%** decrease in H_1 and a **10.4%** increase in H_2 were found in female SALS cases (**Table 5-16**). No significant effects of site of onset, age at onset or survival on SNP frequency were detected.

Table 5-15. Mantel-Haenszel analysis with respect to gender.

	Odds ratios				CMH	Woolf
	Female	Male	Combined	CI_{95}	P	P
RS6565612,SNP1	1.63	1.17	1.33	1.07-1.67	**0.0106**	0.1498
RS8075102,SNP2	1.69	1.15	1.33	1.04- 1.71	**0.0228**	0.1333
RS7405450, SNP3	1.97	1.10	1.38	1.09-1.74	**0.0069**	**0.0157**
RS3934711,SNP4	0.33	1.06	0.79	0.50-1.24	0.3116	0.0657
RS9912074, SNP5	0.70	0.94	0.84	0.64-1.11	0.2164	0.3164
RS4073997, SNP6	1.68	0.77	1.03	0.82-1.30	0.7837	**0.0010**

Odds ratios for SALS association within males female,sand when combined are shown. All associations (SNP1-3) remained significant when gender stratification was accounted for, as shown by the Cochran–Mantel–Haenszel (CMH) statistics. However, the size effects for SNP3 and 6 were significantly heterogeneous between genders as indicated by the Woolf test.

Table 5-16. Haplotype association in Female SALS

Windows	Haplotype		Frequencies %		P	FDR
			CTRL♀	SALS♀		
3	ACA (H_1)	+++	66.7	54.0	**0.0026**	**0.0103**
5	ACATG	+++++	41.5	25.6	**0.0001**	**0.0025**
5	ACATC	++++-	13.9	20.9	**0.0317**	0.1904
5	ACACG	+++-+	10.7	7.6	0.2559	0.6293
5	ACACC	+++--	0.9	0.0	0.3540	0.6293
3	GAG (H_2)	---	15.2	25.6	**0.0017**	**0.0103**
5	GAGTG	---++	13.0	22.1	**0.0038**	**0.0340**
5	GAGTC	---+-	1.8	2.9	0.3689	0.6293
3	GCA (H_3)	-++	10.5	10.2	1	1
5	GCACG	-++-+	3.5	1.2	0.1353	0.4871
5	GCACC	-++--	6.1	8.1	0.3846	0.6293
5	GCATG	-++++	0.2	0.6	0.3725	0.6293
5	GCATC	-+++-	0.8	0.6	1.0000	1.0000
			3-SNP Window Omnibus:		**0.0118**	
			5-SNP Window Omnibus:		**0.0044**	

Haplotype association in Female SALS (**n=88 for 3 SNPs, n=86 for 5 SNPs**) and gender matched controls (**n=302 for 3 SNPs; n=300 for 5 SNPs**). "+" denotes major alleles and "-" denotes minor alleles. Compared to the complete cohorts, as shown in **Table 5-14**, the differences in H_1 and H_2 frequencies were augmented.

5.3.4 Discussion: NPLOC4

We carried out an independent study in a UK cohort on *NPLOC4* to investigate a potential association with SALS, based on the major role played by Npl4 in proteostasis pathways strongly implicated in ALS pathogenesis and previous GWAS data,. In addition to the SNPs (SNP 1, 3, 4) genotyped in the GWAS dataset, we analysed 3 additional SNPs and showed significant associations of 3 out of 6 SNPs with SALS. These include two known associations from the GWAS Dataset and we propose that *NPLOC4* is a genetic risk factor for SALS. Although no association was evident in FALS cases, two SNPs remained significant in the pooled ALS group. In addition to the initial GWAS study (Schymick et al., 2007b), SNPs 1-3 have been shown to be significant (uncorrected) in a replication stage in which their major alleles were reported as risk alleles (Chio et al., 2009). We showed that minor alleles were over-represented in ALS cases, suggesting heterogeneity in

effects which may be caused by ethnicity. The replication stage of the GWAS study contained ALS cases from the US, Italy and Germany, whereas, in this study, population stratification was controlled by using ethnically homogeneous samples from the UK. Using 2 x 2 Fisher's tests, there were no significant MAF differences of all SNPs between the IC-Controls and the EUR from 1000 genome project. However, the IC-Controls were significantly different from the controls used in the GWAS study for SNP1 and 2 (**P=0.0384** for SNP1; **P=0.0035** for SNP2; **P=0.0850** for SNP3) (Chio et al., 2009).

We examined the linkage disequilibrium (LD) of the SNPs with MAF>0.1 in the *NPLOC4* gene in the EUR population prior to haplotype analysis and showed that, instead of a single LD block, discrete blocks were formed throughout the gene. Using tagged SNPs we showed that SNP 1, 2 and 3 were each in LD (r^2>0.9) with >2 tagged SNPs spanning from 6 to 22 kb (**Table 5-17**), suggesting that adequate amount of variation were captured by these SNPs. In addition, we were interested in a separate LD block, covering intron 16, which is predicted to interact with the TG repeat. Three SNPs in (SNP 4 and 5) or adjacent (SNP 6) to the block were therefore analysed in the SALS cohort, in which no changes in LD was observed (Data not shown). Only SNPs 1, 2, 3, 5 and 6 were used for haplotype analysis, which was carried out in both 3-SNP and 5-SNP windows. In the 3-SNP window, the risk alleles of SNP 1-3 constituted a risk haplotype for SALS, H_2 (G-A-G), covering the first half of the gene (Exon 2 -12), and it was shown in the 5-SNP window that H_2 was further defined by the risk haplotype, G-A-G-T-G.

TDP-43 is known to bind TG repeats through its RRM 1 motif and mediate exon skipping of the *CFTR* gene (Buratti and Baralle, 2008). Splicing events are regulated through different mechanisms and it is likely that in the case of *NPLOC4*, TDP-43 promotes the removal of intron 16. Xiao et al (2011) further validated the binding of TDP-43 to a site of *NPLOC4* pre-mRNA adjacent to the alternative spliced exon, suggesting a functional role of the intron 16 TG repeat. Nevertheless, no significant differences in genotypic or allelic distributions of the repeat were observed between cases and controls. There are two possible explanations, firstly, the variation in the numbers of repeat detected in our cohorts may not be sufficient to interfere with TDP-43 binding as RNA containing $(UG)_{11}$ repeats efficiently bind to TDP-43 (Buratti et al., 2001),which is equivalent to the lowest number detected in the current study. Alternatively, if there is a quantitative relationship

between repeat length and TDP-43 binding efficiency, the resulting alteration in Npl4 isoform levels may be tolerated. It has been shown that, in the Npl4-Ufd1 complex, loss of the Npl4 NZF domain, which is the case in the 002 isoform (**Figure 5-5A**), does not completely abolish the ubiquitin–binding ability as this can be compensated by the UT3 domain of Ufd1 (Ye et al., 2003a). However, we do not exclude the possibility that a total loss of the TG repeat may result in profound pathogenic effects.

NPLOC4 interacts with several known pathogenic pathways occurring in ALS and its probable functional links with TDP-43 may underlie the specific association with SALS. We compared cases and controls within the same gender category and demonstrated gender effects in which both single SNP and haplotype associations were strengthened in female SALS. As can be seen in **Table 5-6**, the MAFs of all associated SNPs were lower in female in controls and this trend was inverted in cases. The size effects, which were in consistent in direction for both genders, were appreciably increased in female SALS (**Table 5-15**). The incidence of ALS has been known to be higher in male than female, with a ratio of ~1.3, and female patients tend to have a later age of onset with bulbar involvement (McCombe and Henderson, 2010). Gender effects have been also reported in *SOD1* transgenic mices in which survival times were prolonged in females (Heiman-Patterson et al., 2005) and a SNP, rs1570360, in the *VEGF* gene showed female-dependent association with SALS patients (Fernandez-Santiago et al., 2006). Furthermore, there is considerable evidence suggesting that oestrogen is a neuroprotective agent (Garcia-Segura et al., 2001) and it is possible that Npl4 may play a role in the regulation of oestrogen-mediated signalling pathways (Nawaz et al., 1999; Wojcikiewicz, 2004).

Using an approach complementary to GWA studies, this report provides an initial evidence for *NPLOC4* as a susceptibility gene of SALS that is strongest in females. Combined with our previous report (**Section 5.1**, Kwok et al., 2013), two genes playing roles in the maintenance of ER proteostasis at the telomeric region of chromosome 17 , *P4HB* and *NPLOC4*, have been associated with our FALS and SALS cohorts respectively and replications are merited to further validate these findings in different cohorts.

Table 5-17. *NPLOC4* SNPs tagged by those analysed in this study.

SNP	Position	Exon/inron
SNPs Tagged by rs6565612		
rs74002480	79611355	5’
rs35570626	79611084	5’
rs62075721	79610246	5’
rs7406219	79609410	5’
rs9747119	79607017	5’
rs7220310	79606056	5’
*		Intron1
rs8079963	79603285	
rs9890852	79600110	Intron1
rs12946906	79599573	Intron1
rs7405937	79597811	Intron1
rs7406382	79597786	Intron1
rs6565609	79592267	Intron2
rs7502337	79588470	Intron3
SNPs Tagged by rs8075102		
rs7405966	79566727	Intron9
rs7405469	79566596	Intron9
rs7405646	79551064	Intron12
rs7406991	79548614	Intron12
SNPs tagged by rs7405450		
rs9893365	79541555	Intron12
rs8081883	79540598	Intron12
rs7406408	79535532	Intron14
SNPs tagged by rs3934711		
rs7406704	79535171	Intron14
rs3936237	79533264	Intron15
rs9911739	79532109	Intron16
rs9912074	79531888	Intron16
rs9905026	79531572	Intron16
rs72855627	79531183	Intron16

SNPs that are tagged by those analysed in this study. Regions of SNPs tagged by rs6565612= 22,885 bps; Region tagged by rs8075102= 18,113 bps; Region tagged by rs7405450= 6,023 bps.

-END OF CHAPTER 5-

Chapter 6

Investigation of Variable Number Tandem Repeat (VNTR) in Novel ALS Candidate Genes

6.1 Introduction

Dynamic mutations in variable number tandem repeats (VNTR) have been reported to be associated with progressive neurodegenerative conditions and in some cases may cause pathogenic consequences depending on their location (Paulson and Fischbeck, 1996). The most well documented expansions in coding regions are present in polyglutamine and polyalanine tracts in disorders such as Huntington's disease, Spinalbulbar Muscular Atrophy (SBMA), Dentatorubral-Pallidoluysian Atrophy (DRPLA), Spinocerebellar Ataxia (SCA) types 1-3 and 17 and Oculopharyngeal Muscular Dystrophy (OPMD). Triplet expansions in introns and 5' and 3' Untranslated regions (UTR) are seen in Myotonic Dystrophy (DM), Friedrich's Ataxia and SCA types 8 and 12, causing pathogenicity by affecting gene expression or mediating RNA dependent toxicity. Furthermore, age at onset and severity of disease are often associated with the length of expansion, which has the propensity to increase in successive generations depending on perfectness and length of the repeat. Although the majority of known dynamic mutations are trinucleotide expansions, tetra- and penta-nucleotide expansions have been reported in conditions such as DM2 (Liquori et al., 2001) (CCTG), SCA10 (Matsuura et al., 2000) (AATCT).

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease affecting both upper and lower motor neurons and is pathologically characterized by the presence of cytoplasmic inclusions. Ubiquitinated inclusions containing TDP-43 are found in most Sporadic ALS (SALS) patients and Frontotemporal dementia (FTD) patients, indicating a common pathogenic mechanism underlying these conditions (Neumann et al., 2006). A hexanucleotide GGGGCC expansion in intron 1 of *C9orf72* gene was recently identified as the most common cause of Familial ALS (FALS) with or without FTD (Renton, 2011). The effect of the expansion is as yet not established but may cause different pathogenic consequences (**Chapter 7**), indicating further functional consequences of noncoding

repeats (Mori et al., 2013b). We also reported that a small hexanucleotide expansion in 5'UTR of the *VCP* gene was exclusively found in FALS and SALS in **Chapter 3**.

Genome-wide assessment of VNTR length in ALS is difficult as genotyping of repetitive sequences is limited by the read length obtained using current Next-generation sequencing (NGS) platforms (Figley et al., 2014; Treangen and Salzberg, 2012). Using candidate-gene approaches, a number of VNTR in genes causing relevant neurological conditions have been investigated in ALS, and VNTRs present in the *ATXN2* (Elden et al., 2010) and *NIPA1* (Blauw et al., 2012) genes have been shown to be associated with susceptibility to SALS. In view of the importance of VNTRs in ALS, we explored a list of twenty candidate VNTR in genes based either on evidence of altered expression in SALS, location adjacent to associated SNPs in GWAS studies (Chio et al., 2009) or involvement in known pathogenic mechanisms present in ALS. Trinucleotide repeats (TNR) in coding polyalanine tracts and those located in the 5' and 3' regulatory regions were prioritised for initial study.

Four candidate genes, *EIF2AK2*, *CAPNS1*, *YWHAQ* and *FAM120C*, were amongst 48 genes revealed in our previous reports using Random Rapid Rapid Assay of cDNA Level (RRACE) (Kaushik et al., 1998) and High Density Array Membrane (HDAM) (Kaushik, 2000) based methods for detecting tandem repeat-containing genes that are expressed in the human spinal cord. YWHAQ (14-3-3 protein θ) was subsequently shown to be over-expressed in the spinal cord of SALS patients (Malaspina et al., 2000). Lewy-body like hyaline inclusions (LBHI) that are immunoreactive for YWHAQ and other forms of 14-3-3 proteins have been identified in the spinal cords of SALS patients (Kawamoto et al., 2004) and a FALS patient with a 2 base pair *SOD1* deletion (Kawamoto et al., 2005). Fourteen additional TNRs, previously reported by Kozlowski (Kozlowski et al., 2010), were selected based on the presence of SNPs that are significantly associated ($P<0.005$) with SALS within ±100 kb of the VNTR or being implicated in ALS pathogenesis. They are involved in DNA metabolism (*SSBP3*), Excitotoxity (*TMEM158*), Oxidative stress (*GPX1*), Neurodevelopment (*SLC12A2*, *IRX2*, *IRX3*, *IRX4*), Neuronal Growth (*ID4*), Wnt/TCF and NfkB pathway (*CTNND2*), Retinoid Acid signalling (*RXRB*), Protein folding and degradation through the Ubiquitin-Proteasome System (UPS) (*HSPB8*, *UBQLN3*) and RNA binding (*RBM23*). Of these, the activity of Glutathione peroxidase 1 (*GPX1*), a pivotal

free-radical scavenging enzyme, was reported to be decreased in the motor cortex of ALS patients (Przedborski et al., 1996). Iroquois homeobox protein 3 (*IRX3*) is overexpressed in the spinal cord of SOD1 transgenic mice at 90 days (Ferraiuolo et al., 2007). Heat shock protein 22 (*HSPB8*), a small heat shock protein expressed in the spinal cord, is involved in facilitating protein folding and the degradation of misfolded proteins via autophagy and the UPS (Hu et al., 2007). Expression of HSPB8 is upregulated both in SOD1 transgenic mice (Crippa et al., 2010) and spinal cord of SALS cases (Anagnostou et al., 2010), suggesting a neuroprotective effect. Finally, based on our findings that SNPs in the *P4HB* gene, which encodes Protein disulphide isomerase A1 (PDIA1), is a potential risk factor and modifier of FALS (**Chapter 5**), we also included *TXNDC5,* which encodes PDIA15, another member of the PDI family. The reported association of the *NIPA1* repeat in SALS (Blauw et al., 2012) was also further investigated in our FALS cohort.

VNTRs were screened in ALS and control cohorts for repeat expansions and association studies carried out for polymorphic candidates. Our results show that the *NIPA1* and *HSPB8* repeats are significantly associated with FALS and SALS respectively.

6.2 Backrgounds for candidate genes

6.2.1 SSBP3

The *SSBP3* gene encodes Single strand DNA binding (SSB) protein 3. The SSB proteins, which are characterized by a N-terminal domain and a glycine- and proline-rich domain, bind and stabilize single stranded DNA generated in different DNA metabolism processes such as replication, repair and recombination (Shamoo, 2001), where SSB proteins remove secondary structures, regulate annealing and prevent nuclease-mediated DNA degradation. The human SSB has been shown to regulate the arrest of DNA replication induced by UV radiation (Carty et al., 1994). Although the precise functions of the SSBP3 protein are unknown, it is a human ortholog of the chicken *CSDP* gene and may play a role in DNA repair following oxidative damages in neurodegeneration (Castro et al., 2002; Mantha et al., 2013).

6.2.2 HSPB8

Please refer to discussions.

6.2.3 EIF2AK2

Eukaryotic translation initiation factor 2- alpha kinase 2 (EIF2AK2), also known as Protein kinase R (PKR), is a serine/ threonine kinase that mediates responses to different cellular stresses. EIF2AK2 is capable of binding double-stranded RNA in circumstances such as viral infection, after which it is activated by dimerization and auto-phosphorylation. The activated EIF2AK2 phosporylates the translation initiation factor EIF2A and inhibits mRNA translations, which can be followed by apoptosis (Garcia et al., 2006). This pathway is critical for antiviral effects mediated by interferons and relies on the N-terminal dsRBD and C-terminal kinase domains. Indeed, the phosphorylation of the EIF2A by the ER sensor PERK is a crucial step for activating the Unfolded protein response (UPR), which can also be regulated by PKR in a PERK- independent manner (Lee et al., 2007a). Alternatively, either by exerting kinase activity or directly interacting with other molecules, EIF2AK2 is involved in different signalling pathways such as the NF-kB pathway, which has been implicated in ALS. Moreover, EIF2AK2 is also essential for the stabilization of the mRNA of HSP70, which is important for protein folding (Zhao et al., 2002).

6.2.4 YWHAQ

The *YWHAQ* gene encodes for theta-isotype of the14-3-3 protein, a group of signalling molecules that are abundantly expressed in the brain, comprising 1% of its total soluble protein (Berg et al., 2003). 14-3-3 proteins were known for their roles in the regulation of enzymatic activity and subcellular localization of it substrates, and, in neurons, they are involved in multiple processes including intracellular signalling, division and differentiation, migration, neurite outgrowth, ion channel regulation and apoptosis (Berg et al., 2003; Shimada et al., 2013). 14-3-3 proteins interplay with known pathogenic mechanisms of neurodegeneration, such as oxidative stress, ER stress, protein aggregates formation and degradation, and the isotypes have been implicated in various neurological disorders (Steinacker et al., 2011). Regarding ALS, the zeta-isotype of 14-3-3 has been shown to interact with HSP70 and BAG3, indicating a role in the prevention and elimination of misfolded proteins (Ge et al., 2010). Further evidence of this hypothesis are the direct interaction of the beta, zeta, tau, gamma and eta isotypes with the mRNA of low-molecular weight neurofilament (NF-L) (Ge et al., 2007), which can be also bound and modulated by mutant SOD1 and TDP-43 (Volkening et al., 2009). Furthermore, the

observation that the zeta- and theta- isotypes were upregulated in hypoglossoal nerves during mechanical injury suggests that 14-3-3 proteins may be required for peripheral nerve regeneration (Namikawa et al., 1998). We previously identified the theta-isotype of 14-3-3 from the human spinal cord cDNA library (Kaushik et al., 1998) and, using Nothern's blot, it was demonstrated that the expression of both transcripts (1.8kb and 2.2kb) were consistently up-regulated in the spinal cord of ALS patients. The 5'UTR hexa-nucleotide repeat was characterized in a small group of FALS, SALS, PD patients as well as controls, but no significant differences were noticed. In the current study, we sought to characterize this repeat in an extended cohort of FALS, SALS and Controls.

6.2.5 GPX1

There are substantial evidences indicating that oxidative stress is a mechanism causing neuronal injuries in ALS. 5-10% of Familial ALS cases are caused by mutations in SOD1 gene, a redox enzyme that catalyzes the detoxification of superoxide radicals to form oxygen and hydrogen peroxide. The removal of hydrogen peroxide, however, requires reduced glutathione (GSH), which donates H^+ and e^- to free radicals and undergoes oxidation, forming the oxidized GSSG (Schulz et al., 2000). This process, also involved in the detoxification of peroxynitrite and hydroxyl radicals, is facilitated by glutathione peroxidise 1 (GPX1) (Mills, 1957). Any deficiency in this pathway may cause the build-up of Reactive oxygen species and impair cellular tolerance to oxidative stress (de Haan et al., 1998). GPX1 is a well-characterized, cytosolic, selenium containing enzyme accommodating a poly-Alanine tract, coded by a CCG repeat in Exon 1, at the N-terminal. Loss of homozygosity (LOH) in this repeat, which is polymorphic with 5-, 6- and 7- repeat alleles, has been associated with cancer in lung (Moscow et al., 1994), head and neck (Hu et al., 2004) and breast (Hu and Diamond, 2003) and this was ascribed to the process of tumor evolution. A functional explanation of the association was the accompanying loss of the Pro genotype at codon 198 which sustains the activity of GPX1 stimulated by selenium (Hu and Diamond, 2003). It has been shown that the activity of GPX1 is significantly reduced in the motor cortex (Przedborski et al., 1996), but not the spinal cord (Fujita et al., 1996) of SALS patients.

6.2.6 SLC12A2

The Solute carrier family 12 gene (*SLC12A2*) encodes for Na^+-K^+-$2Cl^-$ co-transporter 1 (NKCC1) expressed in both epithelial and non-epithelial cells throughout the body. Solutes are imported into the cells through NKCC1, allowing the secretion of fluids to lumens of exocrine glands and the regulation of cell volumes (Hebert et al., 2004). There are two isoforms of NKCC1, the first being the constitutive spliced isoform containing a C-terminal domain, 12 transmembrane domains and a N-terminal domain, whereas the second isoform has a shortened C-terminal domain due to the deletion of Exon 21 (Hebert et al., 2004). NKCC1, which is expressed in the developing brain and spinal cord, also regulates the neuronal responses to GABA by modulating chloride equilibrium potential (E_{Cl}) and contributes to the development of neonatal seizures (Delpy et al., 2008; Dzhala et al., 2005).

6.2.7 RBM23

RNA binding motif protein 23, also known as RNPC4 or CAPERbeta, belongs to the Splicing factor U2 auxiliary factor 65 ($U2AF^{65}$) family of RNA binding proteins and regulates steroid hormone receptor mediated transcription and splicing (Dowhan et al., 2005). The protein contains a serine-arginine-rich (RS) domain and two RRM motifs that are typically found in the U2AF proteins, which participates in the formation of spliceosome through binding to the poly-pyrimidine tract at the 3' splice site (Lunde et al., 2007). Defects in pre-RNA maturation have been implicated in ALS as TAR-DNA binding protein 43 (TDP-43), another nuclear RNA-binding protein, has been found in the neuronal inclusions in ALS (Neumann et al., 2006). TDP-43 is depleted from the nucleus, suggesting loss-of-function consequences may contribute to ALS.

6.2.8 NIPA1

Please refer to discussions.

6.2.9 CAPNS1

Calpain small subunit 1 (CAPNS1) is a regulatory unit of Calpains, which are calcium-activated cysteine proteases that have been implicated in different neurodegenerative conditions. The cytoplasmic enzymes, which are widely expressed in different tissues, can be activated by stimuli such as altered Ca^{2+}, phosphorylation by PKA and decrease in

Ca^{2+} binding proteins (Vosler et al., 2008). Upon activation, m-Calpain cleaves pro-Caspase-12, which in turns cleave other effector Caspases and induces apoptosis (Orrenius et al., 2003). This link is interesting as it may connect different pathological mechanisms to cell death in ALS. First, a cause of motor neurons damages in ALS is excitotoxicity, which occurs when NMDA receptors are over-stimulated by extracellular glutamate. The activated receptors in turn allow Ca^{2+} influx through Ca^{2+} channels (Pasinelli and Brown, 2006), creating alterations in cytoplasmic Ca^{2+} levels. Second, it has been shown that activated Calpain cleaves TDP-43, a pathological hallmark of ALS, in transgenic neuronal models of ALS (Yamashita et al., 2012). In fact, C-terminal of TDP-43 formed after the cleavage of Caspase-3, which activates Calpain, forms insoluble TDP-43 aggregates that mimic ALS (Zhang et al., 2009), suggesting these processes may be associated.

6.2.10 UBQLN3

The *UBQLN3* gene encodes one of the four human Ubiquilin proteins, Ubiquilin-3. Ubiquilins have been implicated in Alzheimer's disease and Parkinson's disease. Like other Ubiquilins, Ubiquilin-3 is testis-specific isoform characterized by the N-terminal Ubiquitin-like domain (UBL) and C-terminal Ubiquitin-associated domain (UBA). The UBL domain shares homology with Ubiquitin and may be able to conjugate substrates that are targeted for proteasomal degradation, whereas the UBA domain may interact with other components in Ubiquitin-Proteasome pathway (Conklin et al., 2000). Moreover, Ubiquilins also regulate autophagy by interacting with LC-3 (Rothenberg et al., 2010).

6.2.11 TMEM158

TMEM158 gene, which encodes Transmembrane protein 158, as known as Ras-induced senescence 1 (RIS1), is a candidate tumor suppressor gene that is upregulated specifically in response to the activation of the oncogene *Ras*, causing proliferation senescence similar to that seen in mitosis (Barradas et al., 2002). Interestingly, RIS1 has also been identified as a cell-surface binding protein for brain injury-derived neurotrophic peptide (BINP), which protects neurons against glutamate damages, and it may therefore play a role in the regulation of neuronal death (Hama et al., 2001).

6.2.12 IRX2, IRX3 and IRX4

IRX2 and *IRX3* genes encode for two members from the Iroquois homeobox protein family (Irx), both of which are from the IrxA subgroup. Irx proteins share a homeodomain domain, TALE, and a Iro box domain and are involved in a variety of developmental processes including transcriptional regulation of neural development (Gomez-Skarmeta and Modolell, 2002). The genomic organisation of the Iroquois genes is highly conserved among species. In human, both IRX2 and IRX4 are located in chromosome 5, whereas IRX3 is located in chromosome 16. IRX3 has been shown to be up-regulated at 90 days in motor neurons from G93A SOD1 transgenic mice, suggesting a role in adaptive functions following SOD1 injury (Ferraiuolo et al., 2007).

6.2.13 CTNND2

Delta-Catenin2 (catenin δ2, CTNND2) was known as an Armadillo-repeat containing protein that plays a role in cell adhesion and interacts with essential components for the functioning of neuromuscular junctions (Rodova et al., 2004). The protein has been shown to regulate the Wnt/TCF pathway and the Nf-kB pathways. In disease conditions, CTNND2 was located in a deleted region in chromosome 5p that was responsible for Cri-du-Chat syndrome and the knock down of this gene resulted in severe cognitive impairments (Israely et al., 2004). In fact, another member of this family, beta-Catenin, has been shown to interact with Presenilin-1, a candidate gene in Familial Alzheimer's disease and enhance neuronal vulnerability to beta-amyloids (Zhang et al., 1998).

6.2.14 ID4

ID4 encodes Inhibitor of DNA-binding protein 4, a protein that belongs to the helix-loop-helix (HLH) transcription factor family. However, without any DNA binding motifs, ID proteins function as suppressors of other transcription factors (bHLH) in the same family and represses gene expression induced by cis-elements such as E-box (Perk et al., 2005). ID4 was thought as a tumor suppression gene based on the observation that its expression was altered in various types of tumors (Perk et al., 2005). ID proteins have also been implicated in regulating axonal growth and elongation in neurons, which share molecular similarities with the initiation process of tumor metastasis (Iavarone and Lasorella, 2006).

6.2.15 TXNDC5

We reported in Chapter 5 that the *P4HB* gene, which encodes the ER protein Protein disulphide isomerase (PDI) A1 (PDIA1), is a risk factor and modifier for Familial ALS. Proteins from the PDI family are characterized by the presence of TRX-like domain, which contains catalytic active sites facilitating the exchange of disulphide bonds and prevent the formation of misfolded proteins. *TXNDC5* encodes another member from the PDI family, PDIA15, also known as Erp46 or EndoPDI (Galligan and Petersen, 2012). PDIA15 possesses three TRX-like domains containing three functionally active motifs, CGHC, which are identical to PDIA1, and an ER-localization signal, KDEL (Galligan and Petersen, 2012). It has been shown that PDIA15 was expressed in the brain and it was able to complement the functions of PDIA1 in yeast, suggesting an important role in ER protein folding (Knoblach et al., 2003).

6.2.16 RXRB

There were considerable studies showing protective effects of Retinoic acid (RA) in various neurodegenerative conditions. RA is a metabolic product of retinol (Vitamin A), which binds RBP1 in bloodstream and enters the cells through membrane receptor STRA6. RA may function as a signalling molecule by entering the nucleus and forming a complex with RA receptor (RAR) and Retinoic X receptor (RXR), after which the complex binds target DNA sequences (Maden, 2007). RA signalling has been implicated in the differentiation and development of the nervous system and, in particular, motor neurons, and may therefore have a protective role against neurodegeneration (Maden, 2007). In fact, a link between RA signalling and ALS has been established by the observation that rats depleted of the receptor RAR-alpha, which was also reduced in SALS spinal cords, developed motor neuron degeneration (Corcoran et al., 2002). We previously showed that the expression of Cellular retinoic acid binding protein 1 (CRABP1), an intracellular carrier of RA, was significantly increased in the spinal cord of SALS, suggesting a secondary response to compensate for the loss of RA signalling (Maden, 2007; Malaspina et al., 2001). In addition, RA was shown to regulate the level of Superoxide dismutases (SOD) and protect the neurons from apoptosis and oxidative stress (Ahlemeyer et al., 2001). The *RXRB* gene encodes Retinoic X receptor beta and may therefore important in mediating the plausible protective effects of RA.

6.2.17 FAM120C

FAM120C is a 16-exon gene encoding a putative transmembrane protein spanning 111kb in chromosome Xp11.2 (Holden and Raymond, 2003). A microdeletion involving this gene has been associated with autistic disorder (Qiao et al., 2008).

6.3 Methodology

6.3.1 Subjects

The study population consisted of **200** FALS index cases, **99** SALS cases and **299** Controls from the UK recruited from the Imperial College Healthcare NHS Trust and Kings College Healthcare NHS Trust (**Table 6-1**). Diagnosis of ALS was made using the El Escorial Criteria. All FALS cases presented with a positive familial history in which at least two members were affected in the family and each FALS was an index case from unrelated families. All FALS cases lacked mutations in *SOD1*, *TARDBP*, *FUS*, *DAO* and *VAPB* genes. The *C9orf72* expansion had been screened in the FALS cohort and this was taken into account in the analysis. Informed consent was given by all subjects and the study was carried out according to the Research governance regulations of Imperial College London and Kings College London. DNA was extracted from whole blood or the buffy coat layer using a DNA extraction kit according to instructions of the manufacturer (Qiagen, Manchester, UK).

Table 6-1. Summary of the cases and controls employed in this study.

Cohorts	No. of samples	Gender Ratio (Male: Female)
Control	299	1:1.05
Group 1	129	
Group 2	82	
Group 3	88	
FALS	200	
Imperial College (All)	104	1 : 1.04
King's College (YWHAQ)	96	
SALS		
Imperial College (HSP)	99	1:2.22
Total	***598***	

A total of 598 cases and controls were included in this study and genotyping was performed in different subsets of samples. The Imperial FALS cohort was used for the initial screening of all VNTRs, whereas the King's FALS cohort and the Imperial SALS cohort were only employed in the screening for *YWHAQ* and *HSP* repeats respectively. Control group 1, which was collected from East Anglia and East Yorkshire, was screened for the *EIF2AK2*, *CAPNS1* and *YWHAQ* repeats. Control group 2, which was collected from London, was screened for the *HSPB8* repeat. Control group 3 was collected from around the U.K. and was used in the screening for all other polymorphic VNTRs.

6.3.2 Genotyping and DNA sequencing

DNA fragments were amplified in PCR reactions using primers flanking ~100 bps of the VNTRs. The primers were designed using Primer 3 Program (*http://frodo.wi.mit.edu/primer3*) and purchased from Invitrogen, UK. A standard 30μl PCR reaction mixture was prepared, as shown in **Table 6-2**, and a nested PCR was used for the *GPX1* repeat. The VTNRs are then genotyped using electrophoresis. For an agarose gel electrophoresis which discriminates down to 3 bps size differences, for DNA fragments sized 200- 300 bps, 15 to 30 μl of PCR products were separated in a 20 cm x 20 cm x 0.8 cm, 2.8% agarose gel at 110V for 16 hours. Time may be extended for larger fragments.

For denaturing acrylamide gel electrophoresis (Urea-PAGE), we prepared 50 ml of 4% to 6% urea polyacrylamide gel (1mm thick) using the SequaGel Urea Gel System (National

Diagnostics, UK) according to instructions. Prior to loading, 1 to 10 µl of PCR products were mixed with equal amount of denaturing buffer containing 95% Formamide (Applied Biosystems, UK), 0.9µg/ml Xylene Cyanol and 0.005M EDTA, incubated at 99^{o}C for 10 minutes and chilled on ice. The gel was pre-electrophoresed to 45^{o}C. After loading, electrophoresis was carried out at 15W for 3 to 4 hours depending on fragment sizes. The gel was then stained in 50 ml of 1X SYBR Gold Nucleic Acid Gel Stain (Invitrogen, UK) for 30 minutes. The results of both agarose and acrylamide gel electrophoreses were visualized using GelDoc (Bio-Rad UK, Hemel Hempstead, Hertfordshire, UK) system. The GelAnalyzer Program v 2010a (*http://www.gelanalyzer.com/*) was used for molecular weight calculation. For DNA sequencing, the PCR products were purified using Sureclean (Bioline, UK) according to instructions and sequenced using ABI Prism BigDye terminator kit (Applied Biosystems, Warrington, Cheshire, UK) carried out by the College's core genomic laboratory. Sequences were visualized using Codon Code Aligner and Seqdoc programs (*http://research.imb.uq.edu.au/seqdoc*).

Table 6-2. Primers and PCR Conditions used in this study.

VNTR	Primers	Product Size
UBL3	FOR: CTGGGGCATTTTTCTCCTG REV: GAGGTTCTGGTTCGAAGAGG	277
HSPB8	FOR: GCTCCATCAGGAACCAAGCA REV: CTCCCTCCTTTAGGCATCGC	293
ID4	FOR: GATGAAGGCGGTGAGCCC REV: GCGGCTATAGCAGTCGTTCA	223
SSBP3(A)	FOR: CGTTCGGTTGAGCTCCAAGT REV: CATCGCCCTGGAACTCCTTC	188
SSBP3(B)	FOR: CTGACGCTTTGACAGCTGGA REV: CGCCACTTGCAAAATAGGGC	283
TMEM158	FOR: TAAGCCCATTGAGTCCACGC REV: ACACAGGACGGACACAGGGA	286
*GPX1**	FOR: GAAAACTGCCTGTGCCACG REV: CAGAGGGACGCCACATTCTC NestedFOR:CCGCTGGCTTCTTGGACAAT Nested REV: AGCCCAGGCTCACAGGCT	260, 126
IRX4	FOR: GAGGCCAAGCTGGGGTTTGT REV: CCGTCCACCCAGTTTCTGAG	281
IRX2	FOR: CTCGGCGTTCAGCCCCTA REV: GACTCCTGAGTCGCCAGC	252

VNTR	Primers	Product Size
IRX3	FOR: GAAGCCCAAGATCTGGTCCC REV: AGAGCCCAGCAGGGAGAG	256
SLC12A2	FOR: TCGGTGCCGGAGGATGCTG REV: GCTTCCTCGCTGCCTTTAGC	284
TXNDC5	FOR: CCCAGGACGCCTCCTCCC REV: GCGCGAAGAACATGACGAAG	248
CTNND2	FOR: TAAGTGCGCGTGTTCTCTCC REV: CTGCAGCTTGGTGGGCGAA	265
NIPA1	FOR: CTCTTCCTGCTCCTCCCCC REV: GCACGATGCCCTTCTTCTGT	286
RXRB	FOR: CCAGGGATCATGTCTTGGGC REV: ATAAAGCGGTCACTGGCTCG	289
RBM23	FOR: CAGCTCATGGCAAAACTGGC REV: GCGCACTCCTTCTTTCCAGA	291
YWHAQ	FOR: GCATTGTCTGACGGCGCTC REV: GCTCGGCCTGCTCGGCCAGCTT	198
EIF2AK2	FOR: CCCGTAGCAGACGAGGGCTT REV: GGGACGCAGGATTGGCGAGT	253
FAM120C	FOR: AGATCGTCGCCTATCCTGCT REV: GAGTAGGCACCCAAGGCAG	283
CAPNS1	FOR: AGTGAGTCGCAGCCATGTTC REV: CCCTCTGATAGTCTCCGCCT	246

PCR reactions were carried out in a standard 30µl volume containing 1× buffer, 1.5 mM Mg^{2+}, 0.1 mM dNTP, 0.5 µM primers, 0.05 U/µl GOtaq Taq DNA polymerase (Promega, UK), and 0.5 ng/µl templates. PCRx Enhancers System (Invitrogen, UK) and 7-deaza-dGTP (THERMOPOL®, New England Biolabs, UK) were used when necessary. The thermal cycles were carried out as following: 5 minutes at 94^{o}C, [45s at 94^{o}C, 30s at Annealing Temperature, 30s at 72^{o}C] x 35 Cycles, 5 mintues at 72^{o}C.

*To perform Nested PCR for the *GPX1* repeat, the gDNA templates were, in the first round, amplified using the FOR and REV primers to generate 260-bp products, after which 1µl of an 1:100 fold dilution of the product was taken as the template for the second round of PCR using the Nested primers, which generated the final 126-bp products.

6.3.3 Data analysis and Statistics.

To obtain a list of known Trinucleotide repeats (TNRs) throughout the genome, all TNRs reported by Kozlowski et al (2010) were matched to the reference sequence

(BSgenome.Hsapiens.UCSC.hg19) using the BSgenome package v 1.30.0 (*http://www.bioconductor.org/packages/release/bioc/html/BSgenome.html*) in R (**Appendix II**). The successful matches together with SNPs genotyped in a previous GWA study (dbGAP Study: phs000101, Accession: pha002846.1) (Schymick et al., 2007b) were annotated for gene names and positions using a prediction track obtained from the UCSC Table Browser, assembly GRCh37/hg19. After that, the most significant P value, which had been computed using Cochran-Armitage trend test, within a ± 100 kb window of each matched TNR was obtained and analyzed in a Manhattan plot (Data not shown).

All VNTRs genotyped in this study were tested for allelic and genotypic associations with ALS. For the omnibus tests for allelic distributions, we employed the Mann-Whitney U test, computed using R, and a Monte-Carol test for association, computed using the CLUMP Program (Sham and Curtis, 1995). Alleles were then categorized according to repeat length and a 2 x 2 Fisher test, computed using R, was employed to test the association of each allele or each group of alleles. The 2 x 2 Fisher test was also used to analyse genotype distributions. All tests were corrected for multiple comparisons using the Benjamini-Hochberg method. Student's t test was used to compare HSPB8 expression in samples with different genotypes. To investigate whether the VNTRs are associated with clinical parameters, we compared allele frequencies in cases with different site of onset using the Fisher's test. Survival and age of onset analysis using Kaplan-Meyer curves was carried out for all VNTRs using Graphpad and SPSS Programs.

6.4 Results

6.4.1 Identification of candidate Variable number tandem repeats for ALS.

In order to investigate whether VNTR expansions are involved in ALS pathogenesis, we selected 20 candidate VNTRs for study, using a comprehensive list of VNTRs present throughout the genome (Kozlowski et al., 2010), expression data and relevance to ALS based on existing publications. Publically available GWAS dataset using the BSgenome package in R (*http://cran.r-project.org/*) were also taken into account (**Table 6-3**). Although no significant SNP associations have been reported in candidate genes harbouring the VNTRs selected for this study, the closest GWAS SNPs were some distance from the

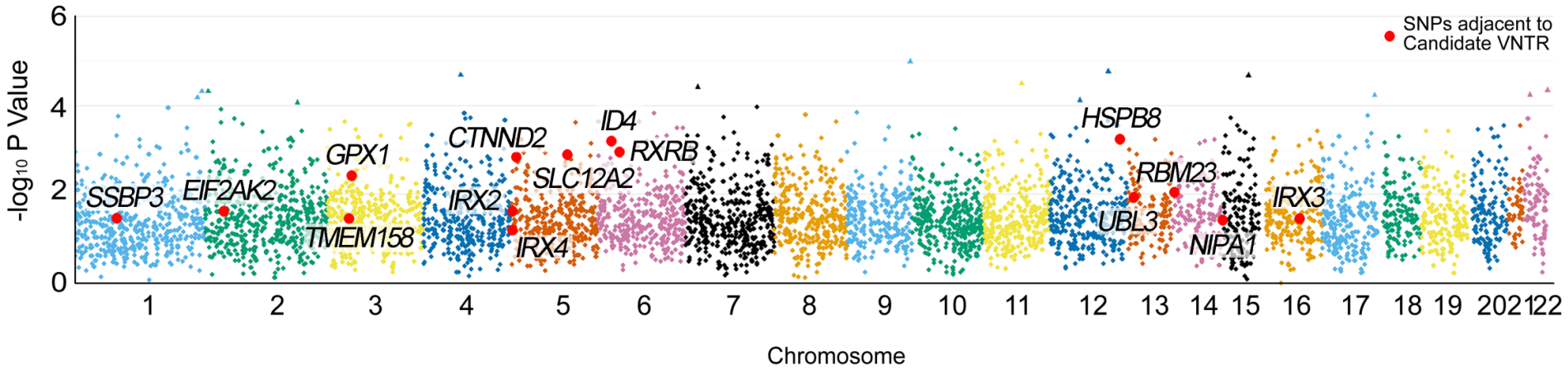
VNTR (up to 100kb) in most cases (**Figure 6-1**). Interestingly, the highest uncorrected SNP association (-log p value ~3) was for one of the selected candidates (*HSPB8*). The candidate VNTRs included 11 VNTRs coding for polyalanine tracts, 1 coding for a polyglycine tract, 1 coding for a polyglutamine tract and 6 located in regulatory regions. With the exception of the *YWHAQ* repeat, which is a hexanuceotide GGGGCC repeat of 2 units, all VNTRs are trinucleotide repeats of ≥ 5 units in the reference sequence (**Figure 6-2**). The *NIPA1* repeat which has previously been been shown to be associated with FALS (Blauw et al., 2012) was also analysed in our cohorts in parallel.

Table 6-3. Criteria used for the selection of candidate genes.

Genes	Gene Functions		GWAS SNP ± 100kb			Expression Profile		
	Neuro-logical Dis-orders	ALS path-ways (DNA/ RNA binding, cell survival and Redox)	P< 0.0 007	P< 0.00 15	P< 0.005	Altered Expression in ALS/ mouse	Spinal Cord TNR Kaushik et al.	Expressed in Brain & Spinal Cord
EIF2AK2		✓					✓	✓
CAPNS1		✓					✓	✓
FAM120C	✓						✓	✓
YWHAQ		✓				✓	✓	✓
HSPB8	✓	✓	✓			✓		✓
GPX1		✓			✓	✓		✓
IRX3		✓				✓		✓
ID4		✓	✓					✓
CTNND2	✓	✓		✓				✓
RXRB		✓		✓				✓
RBM23		✓		✓				✓
SLC12A2				✓				✓
NIPA1	✓	✓						✓
TXNDC5		✓						✓
UBQLN3		✓						✓
SSBP3		✓						✓
TMEM158		✓						✓
IRX4		✓						✓
IRX2		✓						✓

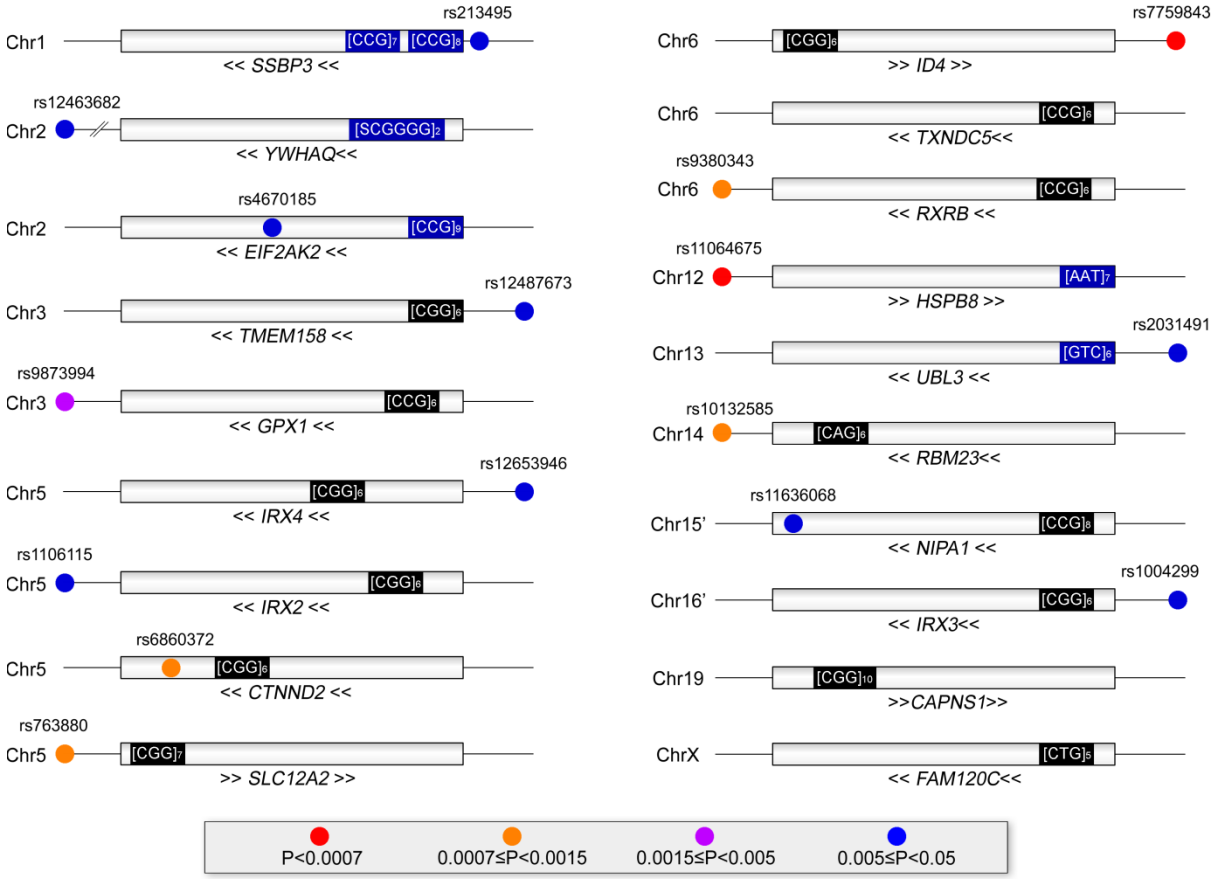
All candidate genes being investigated in study are expressed in the brain and spinal cord, most of which were known to play roles in DNA/RNA binding, cell survival and Redox reactions. Expressions of four candidates were shown to be altered in ALS. *HSPB8* was of particular interest as multiple criteria are fulfilled.

Figure 6-1. SNPs that flank ± 100kb of known tri-nucleotide repeats throughout the genome.



Manhattan-Plot showing SNPs that flank ± 100kb of known tri-nucleotide repeats throughout the genome (Kozlowski et al., 2010). These are the most significant associations with SALS according to a previous GWAS study (dbGAP study accession: phs000101.v3.p1; analysis: pha002846.1) within the region. SNPs adjacent to the candidate VNTRs in this study are indicated by red dots.

Figure 6-2. A summary of VNTRs in the current study.



Relative location between the candidate VNTRs (Kozlowski et al., 2010) and the associated ± 100kb SNPs are shown. In addition, a SNP which is located 133.6kb upstream from the *YWHAQ* hexanucleotide repeat is indicated by a broken line (-//-). Protein coding and non-coding VNTRs are shown in black and blue respectively. The SNPs are indicated by dots of different colours representing their significances.

6.4.2 Identification of polymorphic VNTRs.

All VNTR candidates were genotyped in an IC-FALS cohort consisting of **104** index cases from which known mutations in the *SOD1*, *VCP*, *TARDBP*, *FUS* (Exon 14 and 15), *VAPB* and *DAO* genes had been excluded. FALS cases harboring *C9orf72* expansions were included in this screen in view of the coexistence of *C9orf72* expansions with other FALS genes (van Blitterswijk et al., 2012a). *C9orf72* status was taken into account in all analyses and effects are noted in the text. VNTR containing sequences were amplified by PCR and sized using agarose gel or polyacrylamide gel electrophoresis. The c.*-25CCCCGS(3_4) repeat in the *YWHAQ* gene was further studied in an additional FALS cohort of **96** cases from King's College London and the results are included in the following analysis. Nine of the VNTRs were polymorphic and used for further analysis. Details of the non-polymorphic VNTRs are summarized in **Table 6-4**.

Table 6-4. Non-polymorphic VNTR repeats in FALS and IC-Controls.

Gene	Chr	VNTR	Exon		Alleles	No. Chr. FALS
SSBP3(A)	1	g.*54872037CCG(8)/ c.*-379CGG(8)	5'	rs3033693	Ref: 8 (100)	82
UBL3	13	g.*30423923GTC(6) c.*-265GAC(6)	5'	NA	Ref: 6 (100)	98
TMEM158	3	g.*45266664CGG(5_6) c.*839CCG(5_6)	Ex1	TMP_ESP_3_45266662	Ref: 6 (100)	80
IRX4	5	g.*1878499CGG(5_7) c.*1127CCG(5_7)	Ex5	TMP_ESP_5_1878496	6(100)	78
IRX2	5	g.*2751296CGG(6) c.*215CCG(6)	Ex1	NA	6(100)	58
CTNND2	5	g.*11385276CGG(6) c.*658CCG(6)	Ex7	NA	Ref: 6 (100)	78
ID4	6	g.*19838107CGG(5_6) c.*122CGG(5_6)	Ex1	TMP_ESP_6_19838105	Ref: 6 (100)	80
TXNDC5	6	g.*7910871CCG(5_7) c.*122CGG(5_7)	Ex1	TMP_ESP_6_7910870	Ref: 6 (100)	84
RXRB	6	g.*33168097CCG(5_6) c.*140CGG(5_6)	Ex1	TMP_ESP_6_33168096	Ref: 6 (100)	86
IRX3	16	g.*54318613CGG(5_6) c.*1163CCG(5_6)	Ex2	TMP_ESP_16_54318611	Ref: 6 (100)	38
FAM120C	X	g.*54209521CTG(5_6) c.*97CAG(5_6)	Ex1	TMP_ESP_X_54209518	NP	128

Details of the non-polymorphic VNTRs genotyped in this study.

Table 6-5. VNTR Allele frequencies in FALS and Controls.

Gene	Chr	VNTR	Exon	RS	Alleles	Allele Counts (Frequencies, %)				Association tests			
						Chr (2N)	CTRL	Chr (2N)	FALS	Allelic 2 x 2	FDR	Omnibus CLUMP	FDR
SSBP3	1	g.*54871712**CCG**(5_13) c.*-51CGG(5_13)	5'	TMP_ESP_1_54871709	Short:5-6 Ref: 7 Long: 8-13	150	11 (7.3%) 61 (40.7%) 78 (52.0%)	132	7 (5.3%) 68 (51.5%) 57 (43.2%)	0.6270 0.0734 0.1527	0.8282 0.2760 0.3245	0.3840	0.4934
HSPB8	12	g.*119632512**AAT**(5_9) c.*591+849AAT(5_9)	3'	rs112223147	Short:5-6 Ref: 7 Long: 8-9	290	109 (37.6%) 83 (28.6%) 98 (33.8%)	120	40 (33.3%) 35 (29.2%) 45 (37.5%)	0.4319 0.9052 0.4955	0.7342 1 0.7658	0.4934	0.4934
EIF2 AK2	2	g.*37384056**CCG**(6_9) c.*-215CGG(6_9)	5’	rs72114633	6 Ref:9	240	31 (12.9%) 209 (87.1%)	130	11 (8.5%) 119 (91.5%)	0.2315	0.4373		
YWHAQ	2	g.*9970589S**CGGGG**(3_4) c.*-25CCCCGS(3_4)	Ex2	rs200302461	Ref: 3 4	214	187 (87.4%) 27 (12.6%)	242	212 (87.6%) 30 (12.4%)	1	1		
GPX1	3	g.*49395675**CCG**(4_7) c.*20CGG(4_7)	Ex1 (Ala)	rs56041243 rs17838762	Short: 4-5 Ref: 6 Long: 7	156	110 (70.5%) 45 (28.8%) 1 (0.6%)	122	97 (79.5%) 25 (20.5%) 0 (0%)	0.09744 0.1265 1	0.2760 0.3072 1	0.3296	0.4934
SLC12A2	5	g.*127419933**CGG**(5_7) c.*287CGG(5_7)	Ex1 (Ala)	TMP_ESP_5_127419931	5 Ref: 7	145	5 (3.4%) 140 (96.6%)	85	2 (2.4%) 83 (97.6%)	1	1		
RBM23	14	g.*23371267**CAG**(4_7) c.*1151CTG(4_7)	Ex12 (Ala)	rs61680332	Ref: 6 7	154	116 (75.3%) 38 (24.7%)	84	66 (78.6%) 18 (21.4%)	0.6333	0.8282		
NIPA1 [§]	15	g.*23086366**CCG**(5_9) c.*23CGG(5_9)	Ex1 (Ala)	NA	Short: 5-7 Ref: 8 Long:>8	146	32 (**21.9%**) 109 (**74.7%**) 5 (3.4%)	98	41 (**41.8%**) 48 (**49.0%**) 9 (9.2%)	**0.0010** **6.77E-05** 0.0891	**0.0085** **0.0012** 0.2760	**2.8E-05**	**0.0001**
CAPNS1	19	g.*36632027**CGG**(10_11) c.*114CGG(10_11)	Ex2 (Gly)	TMP_ESP_19_36632024	Ref: 10 11	250	246 (98.4%) 4 (1.6%)	130	128 (98.5%) 2 (1.5%)	1	1		

Four candidates encode for poly-alanine tracts and the *CAPNS1*-repeat encodes for a poly-glycine tract. Counts and frequencies (brackets) of alleles that are shorter than (short), equal to (ref) or longer than (long) the reference allele are shown for both FALS and Control cohorts. The controls used for *HSPB8*-repeat contain the additional **82** UK Samples that were included in the later stage (see main text). Each group was tested for association using 2 x 2 Fisher's tests and model-free omnibus tests were performed for VNTRs with ≥ 3 alleles. All tests were subjected to Benjamini- Hochberg correction (FDR). Significant P values are in bold.

§ The allele frequencies of the *NIPA1* repeat in *C9orf72* negative FALS cases (N=25) are: Short (44%), (**P=0.0035**), Reference (46%) (**P=0.0004**), Long (10%) (**P=0.1273**).

6.4.3 VNTR association with ALS: allelic and genotypic associations and loss of heterozygosity (LOH) tests

Polymorphic VNTRs were genotyped in FALS index cases and matched UK controls and representative bands were sequenced to confirm genotyping. Allele counts and frequencies are summarized in **Figure 6-3**. Association tests were performed and corrected for multiple comparisons (**Table 6-5**). Association tests on all genotypes present in the current study were performed using 2 x 2 Fisher's tests (**Table 6-6**). Loss of heterozygosity (LOH) was also tested for all VNTRs as LOH is known to occur in the *GPX1* repeat in lung cancer (Moscow et al., 1994). Significant associations were detected for two candidates after correction for multiple testing.

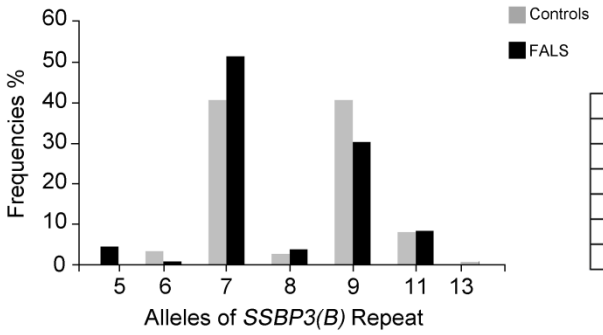
Table 6-6. VNTR Genotype frequencies.

		CTRL		FALS		SALS		
VNTR	Genotype	%	%	P	FDR	%	P	FDR
SSBP3(B)	5/13	0.0	1.5	0.468	1			
	5/7	5.3	6.1	1	1			
	5/9	2.7	1.5	1	1			
	6/7	4.0	1.5	0.623	1			
	6/8	1.3	0.0	1	1			
	6/9	1.3	0.0	1	1			
	7/11	5.3	10.6	0.347	1			
	7/7	9.3	24.2	**0.022**	0.308			
	7/8	1.3	0.0	1	1			
	7/9	46.7	36.4	0.235	1			
	8/9	2.7	7.6	0.252	1			
	9/11	10.7	6.1	0.379	1			
	9/13	1.3	0.0	1	1			
	9/9	8.0	4.5	0.502	1			
	N=	***75***	***66***					
HSPB8	5/5	6.2	5.0	1	1	4.1	0.5698	1
	5/6	12.4	3.3	0.0672	0.8069	5.1	0.0733	0.9135
	5/7	18.6	25.0	0.3429	1	15.3	0.6046	1
	5/8	18.6	10.0	0.1474	1	25.5	0.2064	1
	5/9	0.7	5.0	0.0763	0.8389	0.0	1	1
	6/8	0.0	8.3	**0.0019**	**0.0247**	0.0	1	1
	6/9		1.7	0.2927	1		1	1
	7/7	10.3	5.0	0.2844	1	6.1	0.3525	1
	7/8	14.5	21.7	0.2203	1	22.4	0.1247	1
	7/9	3.4	1.7	0.6734	1	4.1	1	1
	8/8	6.2	1.7	0.2867	1	13.3	0.0703	0.9135
	8/9	7.6	6.7	1	1	4.1	0.4157	1
	9/9	1.4	5.0	0.1509	1	0.0	0.5167	1

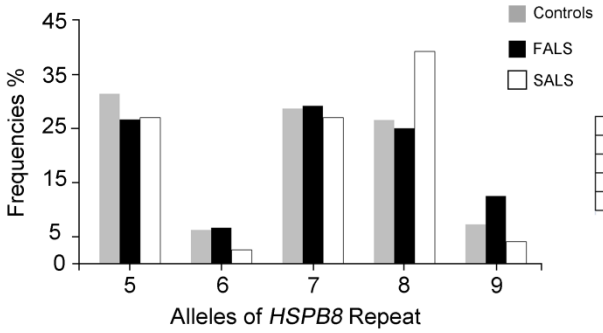
		CTRL		FALS			SALS	
VNTR	Genotype	%	%	P	FDR	%	P	FDR
	N=	***145***	***60***			***98***		
EIF2AK2	6/6	4.2	3.1	1	1			
	6/9	17.5	10.8	0.2845	0.7198			
	9/9	78.3	86.2	0.2399	0.7198			
	N=	***120***	***65***					
	N=	***73***	***49***					
CAPNS1	10/10	96.8	96.9	1	1			
	10/11	3.2	3.1	1	1			
	N=	***125***	***65***					
YWHAQ	3/3	79.4	77.7	0.8718	1			
	3/4	15.9	19.8	0.4918	1			
	4/4	4.7	2.5	0.4793	1			
	N=	***107***	***121***					
GPX1	4/4	20.5	31.1	0.1716	1			
	4/5	21.8	19.7	0.8351	1			
	4/6	30.8	24.6	0.4524	1			
	5/5	5.1	9.8	0.3341	1			
	5/6	15.4	13.1	0.8097	1			
	6/6	5.1	1.6	0.3849	1			
	6/7	1.3	0.0	1	1			
	N=	***78***	***61***					
SLC12A2	A/A	86.7	88.6	1	1			
	A/B	6.7	4.5	1	1			
	A/C	6.7	6.8	1	1			
	N=	***75***	***44***					
RBM23	6/6	52.6	57.1	0.7016	1			
	6/7	47.4	42.9	0.7016	1			
	N=	***76***	***42***					
NIPA1	5/8	4.1	0.0	0.2731	0.5461			
	7/7	2.7	14.3	**0.0293**	0.1307			
	7/8	34.2	55.1	**0.0261**	0.1307			
	8/8	53.4	14.3	**9.8E-6**	**0.0001**			
	8/≥9	4.1	14.3	0.0875	0.2626			
	≥9/≥9	1.4	2.0	1	1			
	N=	***73***	***49***					
CAPNS1	10/10	96.8	96.9	1	1			
	10/11	3.2	3.1	1	1			
	N=	***125***	***65***					

Genotype frequencies of the VNTR in Cases and Controls. Each genotype was tested for association using a 2 x 2 Fisher's test. All tests were corrected for multiple comparisons using the Benjamimi-Hochberg correction (FDR). Significant values are in bold.

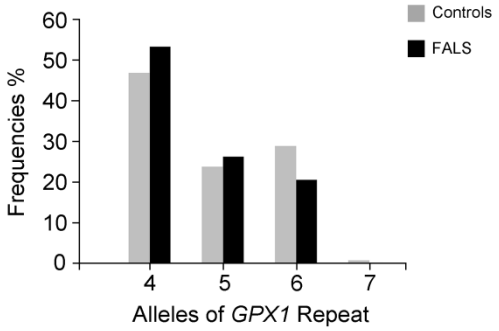
Figure 6-3. Allele frequencies of VNTR with ≥3 alleles in this study.



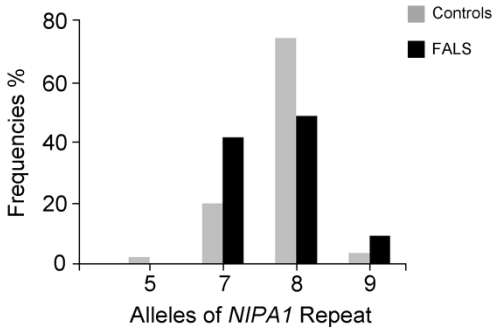
	CTRL		FALS	
	Count	%	Count	%
5	6	4.0	6	4.5
6	5	3.3	1	0.8
7	61	40.7	68	51.5
8	4	2.7	5	3.8
9	61	40.7	40	30.3
11	12	8.0	11	8.3
13	1	0.7	1	0.8



	CTRL		FALS		SALS	
	Count	%	Count	%	Count	%
5	91	31.4	32	26.7	53	27.0
6	18	6.2	8	6.7	5	2.6
7	83	28.6	35	29.2	53	27.0
8	77	26.6	30	25.0	77	39.3
9	21	7.2	15	12.5	8	4.1



	CTRL		FALS	
	Count	%	Count	%
4	73	46.8	65	53.3
5	37	23.7	32	26.2
6	45	28.8	25	20.5
7	1	0.6	0	0.0



	CTRL		FALS	
	Count	%	Count	%
5	3	2.1	0	0.0
7	29	19.9	41	41.8
8	109	74.7	48	49.0
9	5	3.4	9	9.2

From the analysis of allelic associations in FALS in the nine polymorphic VNTRs, significant effects were found for *NIPA1*. Performing the omnibus test for allelic distributions using the CLUMP Program, the alanine repeat c.*23CGG(5_9) in the *NIPA1* gene showed a significant difference between FALS cases and controls (**P_{BH}=0.0001**). Alleles were subsequently grouped into *Short* (short alleles), *Reference* (=reference sequence) and *Long* (expanded) categories which were tested for associations with ALS using 2 x 2 Fisher's tests. The frequency of the *Reference* allele (8-repeat) of the *NIPA1* repeat was significantly decreased from 74.7% in controls to 49.0% in FALS (**P_{BH}=0.0012, OR=0.33**), whereas the *Short* alleles (≤7-repeat) were increased from 21.9% in controls to 41.8% in FALS (**P_{BH}=0.0085, OR=2.56**). The increase in *Short* allele frequency was further enhanced in *C9orf72* negative cases (**P=0.0035, OR=2.80**) (**Figure 6-4 A**). The *Short* allele category consisted of two alleles in controls, the 5-repeat (2.1%) and the 7-repeat (19.9%) but only the 7-repeat was present in FALS. Previously the *Long* allele (≥ 9-repeat), was reported to be associated with SALS (Blauw et al., 2012) with allele frequency increasing from 1.9 to 2.7% in a large SALS cohort (2292 ALS and 2777 control cases). Whilst the control frequency of the long allele in our study was similar (3.4%), no increase could be detected in our FALS cohort reflecting the smaller cohort size. However, it is interesting to note that the increase in short allele frequency in FALS was clearly not evident in SALS indicating a marked difference in VNTR length between FALS and SALS cases.

Consistent with the changes in allelic frequencies, there was a significant decrease in the frequency of the *Reference* 8/8 genotype of the *NIPA1* repeat (**P_{BH}=0.0001, OR=0.15**) together with an increase in genotypes that only contained the *Short* alleles (≤7-repeat), i.e.7/7, in FALS (**P_{BH}=0.0293, OR=5.92**) cases. Similarly, in *C9orf72* negative FALS cases, the frequency of the *Short* genotype was further increased (**P=0.0111, OR=8.88**) (**Figure 6-4 B**). The LOH test showed that heterozygous genotypes of the *NIPA1* repeat were significantly over-represented in FALS (**P=0.0053, OR=3.07, CI_{95}=1.43-6.60**) (**Figure 6-4 C)**.

A significant genotype association was also found for the 6/8 genotype of the 3' UTR repeat c.*591+849AAT (5_9) in the *HSPB8* gene which was absent from controls but present in **8.3%** of FALS cases, (**P_{BH}=0.0247, OR=28.84**), most of which (4/5) were

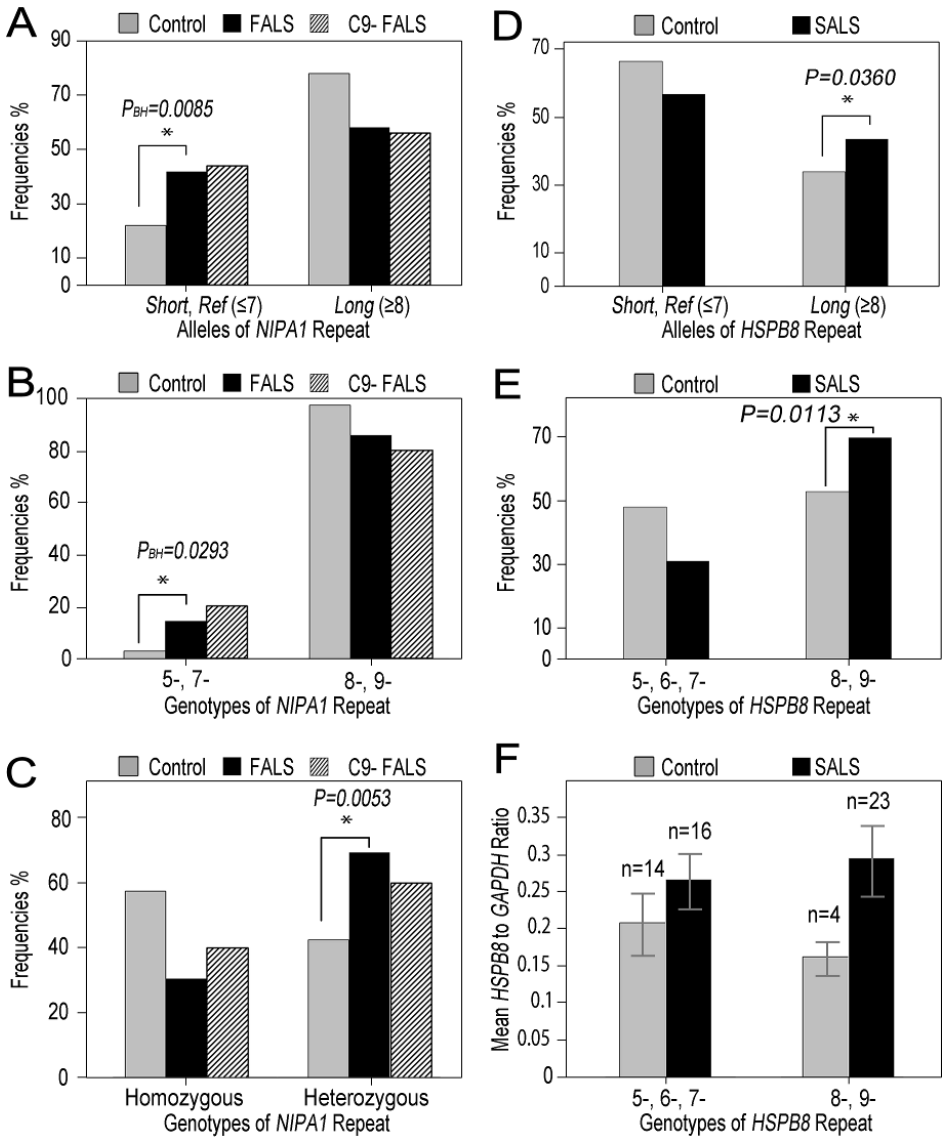
C9orf72 negative (**Table 6-6**). No significant allelic association for the *HSPB8* gene was found in FALS. In view of the significant genotypic association in FALS, this repeat in the *HSPB8* gene was further investigated in SALS cases (n= **99**) and additional UK controls (n= **82**) (total number of control individuals used for *HSPB8* = **170**) (**Figure 6-4 D**). A significant omnibus association for the repeat with SALS was obtained (**P=0.0161**) with an over-representation of the *Long* alleles (≥8 repeats) in SALS (**P=0.0360, OR=1.50**) (**Table 6-7**). Similarly, when *HSPB8* genotypes for SALS cases and controls were pooled according to the length of the expanded allele, we found that those containing at least one copy of a *Long* allele (i.e. 8/-, 9/-) were significantly associated with increased risk of SALS (**P=0.0113, OR=2.06, Figure 6-4 E**). Although smaller effects were seen in terms of an over-representation of long alleles in FALS cases, pooling genotypes for FALS and SALS indicated that they conferred a significant increased risk of ALS applied to both FALS and SALS (**P=0.0141, OR=1.80**).

Table 6-7. Frequencies of *HSPB8* repeat length (*rs112223147*) in SALS

Genotypes	CTRL 2N (%)	SALS 2N (%)	P	OR	CI_{95}	Omnibus CLUMP
Short (<7)	109 (37.6%)	58 (29.6%)	0.0797	0.70	0.47-1.03	
Normal (=7)	83 (28.6%)	53 (27.0%)	0.7576	0.92	0.62-1.39	**P= 0.0161**
Long (≥8)	98 (**33.8**%)	85 (43.4%)	**0.0360**	**1.50**	1.03-2.18	

The table shows allele counts and frequencies (bracket) for the *HSPB8* repeat in SALS and Controls. There is a significant increase in frequency of long repeats in SALS and a significant omnibus test for association. Significant values are in bold.

Figure 6-4. Grouped Allelic and genotype frequencies of VNTRs with ⩾ 3 alleles.



A shows the increase in frequencies of the *Short* alleles (≤7), *Reference and Long* (≥8) alleles of the *NIPA1* repeat in FALS compared to Controls. Association was tested using a 2 x 2 Fisher's test and significances (P<0.05) are indicated by an asterisk (*). **B**. Genotype frequencies that only contained the *Short* alleles (5-, 7-) were significantly increased in FALS. **C** shows the change in heterozygosity of the *NIPA1* repeat in FALS compared to Controls. **D**. *Long* (≥8 repeat) alleles of the *HSPB8* repeat are more frequent in SALS. **E**. Genotypes of the *HSPB8* repeat are grouped according to the size of the longer allele. **F**. *HSPB8* mRNA expression in spinal cord in SALS (n=39) and controls (n=18) is compared for different genotypes using an Unpaired Student's t test with Welch's correction. Data for levels of expression is obtained from Anagnostou et al (2010).

6.4.4 Genotypic correlation with phenotypes and expression

To investigate the effect of genotype on clinical parameters, we first compared the genotype frequencies in cases with different sites of onset (limb or bulbar) and gender using 2 x 2 Fisher's tests. No association with site of onset was observed. Notably, frequency of the 7-repeat genotype in RBM23 was significantly increased in males (29%) compared to females (16%) FALS cases (**P=0.0414**). Age-of-onset and survival analyses were performed (Kaplan-Meyer's curve) but no repeats showed significant effects on survival times.

We have previously shown that the expression level of HSPB8 is increased in spinal cord of SALS cases (Anagnostou et al., 2010) and in view of the association of the *HSPB8* repeat with SALS, we investigated the effect of genotype on HSPB8 expression in spinal cord. *HSPB8* repeat genotypes were determined for 18 Controls and 39 SALS cases with known levels of HSPB8 expression. There were no significant differences in expression between genotypes containing long alleles and those containing short alleles in controls or ALS cases (**Figure 6-4F**).

6.5 Discussions

The finding that hexanucleotide repeat expansions in *C9orf72* are the major cause of familial ALS together with the evidence that trinucleotide repeat polymorphisms in the ATXN2 gene are a risk factor for ALS highlight the importance of these repetitive variants in ALS as also occurs on other neurodegenerative disorders. To further assess whether other VNTRs are involved in the aetiology of ALS, we characterized VTNRs present in 19 further genes in ALS. VNTRs were genotyped using agarose gel and polyacrylamide gel electrophoresis which can detect up to 53 CAG repeats in the *HTT* gene (Goldberg et al., 1993), No alleles exclusively present in FALS were identified. Although it is possible that exceptionally large expansions may have failed to amplify, this is unlikely as no VNTR showed increased homozygosity in FALS (Figley et al., 2014). Nine of the VNTRs were found to be polymorphic in both FALS and controls cases and four of these i.e. *SSBP3*, *GPX1*, *HSPB8* and *NIPA1* repeats, each showed more than two alleles in our cohorts. The alleles were grouped according to size using unifying criteria, in which those identical to the reference sequences (GRCh 37) were categorised as *Reference* alleles, whereas

deleted and expanded alleles were grouped as *Short* and *Long* alleles respectively. We confirmed that the *Reference* alleles were the commonest alleles present in both FALS and Control cohorts for all the VNTRs tested. Allelic and genotype associations test were carried out using two different tests. A model free permutation test for omnibus allelic and genotypic distributions was carried out using the CLUMP program and each individual or allele and genotype group was tested for association using a 2 x 2 Fisher's test.

A VNTR in *NIPA1*, a causal gene for Hereditary spastic paraplegia (HSP) type SPG6, previously found to be associated with sporadic ALS (Blauw et al., 2012) showed a significant omnibus association with FALS. Using a 2 x 2 Fisher's test, we found that *Short* alleles (≤ 7-repeat) were significantly associated with FALS. Although most of the *Short* alleles were present in a heterozygous genotype containing the *Reference* allele, 7/8 (**Table 6-6**), which was not associated, a significant risk of FALS was conferred by the homozygous *Short* genotype, 7/7. The effect sizes of both allelic and genotypic associations were increased in *C9orf72*-negative FALS cases and in the LOH test, we found that heterozygous genotypes were more common in FALS (**P=0.0053**), which has not been reported for SALS, suggesting a FALS-specific shift in allelic and genotypic distributions. Our results, however, do not oppose the previous findings that the rare long alleles containing ≥ 9 repeats are associated with SALS (Blauw et al., 2012) and it is possible that alteration in repeat length of the *NIPA1* repeat is associated with disease penetrance. In our controls, the 7- and 8-repeats are the most common alleles with frequencies of **0.20** and **0.75** respectively (**Figure 6-3**), which is similar to previous reports (**0.20** and **0.78**) (Chai et al., 2003), indicating consistency in genotyping. Indeed, there was a **5.8%** increase in the frequency of the 9-repeat in our FALS cohort compared to controls with an odds ratio of **2.85**, but no significance was observed due to lack of power.

Our results support the conclusion that the *NIPA1* repeat is a risk factor for FALS. *NIPA1* encodes a transmembranous transporter that allows the uptake of Mg^{2+} ions (Goytain et al., 2007), which is also associated with endosome formation and may inhibit Bone morphogenetic protein (BMP) signalling, a crucial pathway in regulating synaptic function and axonal transport (Blackstone, 2012). Mutations in the *NIPA1* gene are the cause of an autosomal dominant form of hereditary spastic paraplegia, SPG6, characterized by axonal degeneration of the cotricospinal and spinocerebellar tracts. Phenotypic and genetic

overlaps between HSP and ALS have been documented (Eymard-Pierre et al., 2002; Stevanin et al., 2008) and, most importantly, TDP-43 pathology has been reported in a familial hereditary spastic paraplegia case with a *NIPA1* mutation who developed ALS-like symptoms and frontal lobe degeneration in later life (Martinez-Lage et al., 2012).

Although the *HSPB8* repeat was not associated with FALS at the allelic level, a significant genotype association was found for the 6/8 genotype which was absent from controls. As we have previously shown evidence of differential expression of HSPB8 in SALS we therefore studied VNTR frequency of *HSPB8* in SALS patients. The length of the repeat ranged from **5** to **9** in both cases and controls which is consistent with the control alignment data from the 1000 genomes project (Data was obtained from ftp://ftp.1000genomes.ebi.ac.uk and visualized using NCBI Genome Workbench v.2.7.15, details not shown). Unlike FALS cases, we found both an allelic and genotypic association of the *HSPB8* VNTR repeat with SALS, which has not been documented before. There was a significant omnibus allelic association together with an over-representation of long alleles (≥8) in SALS compared to controls. The increase is largely contributed by the increased frequency of the heterozygous 7/8 genotype (**Table 6-6**). When genotypes were grouped according to size, a significant association with SALS cases was observed with long alleles (8-, 9-) (**P=0.0113**). There was no significant increase in levels of expression of *HSPB8* in SALS cases Anagnostou et al., 2010) harbouring the associated long alleles compared to controls (**Figure 6-4E**).

The risk of ALS conferred by the VNTR may be associated with the biological functions of HSPB8, which are likely to play a central role in counteracting the formation of different protein aggregates. HSPB8, a 22-kDA small Heat shock protein abundantly expressed in the brain and spinal cord, has been implicated in protein folding, degradation of misfolded proteins and apoptosis. Mutations in *HSPB8* are found in Charcot-Marie-tooth disease (CMT) type 2L and distal hereditary motor neuropathy (dHMN) type II (Irobi et al., 2010). HSPB8 functions as a molecular chaperone assisting correct protein folding mediated by a conserved alpha-crystallin domain (Hu et al., 2007) and it has been shown to prevent the aggregation of the polyglutamate protein Htt43Q (Carra et al., 2005). The observations that HSPB8 is upregulated in both G93A-SOD1 transgenic mice and SALS cases and facilitates the removal of both SOD1 and TDP-43 aggregates via autophagy indicate that

the protein may play an important role in modulating the formation of aberrant protein aggregates and disease progress in ALS (Crippa et al., 2010).

In summary, we show the association of VNTRs in two candidate genes, *NIPA1* and *HSPB8*, with FALS and SALS and respectively. The allele distribution of the NIPA1 repeat found in FALS in this study differed from that found in SALS in a previous study (Blaww et al 2013) but was of greater magnitude. We also demonstrate a novel association of the *HSPB8* repeat with SALS, which highlights the potential relevance of small heat shock proteins to ALS pathogenesis as they are known to participate in the elimination of misfolded proteins. In conclusion, VNTRs may play important roles in the pathogenicity of ALS and the further characterization of VNTRs as causal mutations or risk factors in the disease is merited.

-END OF CHAPTER 6-

Chapter 7

Characterization of *C9orf72* GGGGCC Expansion

7.1 Introduction

A locus for Amyotrophic lateral sclerosis on Chromosome 9p21 (ALS) associated with both ALS and Fromtotemporal dementia (FTD) has been previously reported (Morita et al., 2006; Vance et al., 2006). In 2010, a Genome-wide association study (GWAS) conducted in a Finnish population demonstrated that a single nucleotide polymorphism (SNP), rs3849942, in this 9p21 region was significantly associated with FALS, a significant portion of which shared a 42-SNP haplotype spanning a 232kb LD block (Laaksovirta et al., 2010). The finding indicated the presence of founder mutations in one of the genes contained in this block: *MOBKL2B*, *IFNK* and *C9orf72*. Following next-generation deep re-sequencing, a hexanucleotide GGGGCC (G4C2) expansion in intron 1 of the *C9orf72* gene was found as the cause for the association, being confirmed using a repeat-primed PCR (rpPCR) method which detects up to 60 repeat units (DeJesus-Hernandez et al., 2011b; Renton et al, 2011). The expansion was subsequently shown to be present in 73.2% of FALS (Cruts et al., 2013), 100% of FALS/FTD, (Cruts et al., 2013), 20% of SALS and 30% of FTD (Ling et al., 2013) from different cohorts, being the most common cause of both ALS and FTD. In addition, the *C9orf72* expansion has also been reported in Alzheimer's disease (Harms et al., 2013), and Parkinsonism (Lesage et al., 2013).

The *C9orf72* expansion may mediate multiple pathological consequences through both Gain-of-function and Loss-of-function mechanisms [reviewed by (Ling et al., 2013)]. Haploinsufficiency was implicated as mRNA expression levels of transcript variants 1 and 3, both encoding C9orf72 isoform a, were decreased in lymphoblast cell lines and frontal coretex of *C9orf72* positive cases (DeJesus-Hernandez et al., 2011b). The C9orf72 protein has been reported to play roles in membrane trafficking and autophagy, indicating pathological consequences when normal functions are affected (Ling et al., 2013). Alternatively, whilst most patients with the *C9orf72* expansion have TDP-43 pathology,

TDP-43 negative, p62 and Ubiquilin positive inclusions were also found in *C9orf72* cases. These inclusions were found to contain insoluble dipeptide-repeat protein (DRR) aggregates encoded by the expansion through non-ATG initiated translation, a mechanism also observed in Spinocerebellar Ataxia type 8 (SCA8) (Mori et al., 2013). In addition to DRP aggregates which may contribute to neurotoxicity, RNA foci, which may sequester RNA binding proteins, were found in both cytoplasm and nucleus in the spinal cord and frontal cortex of affected patients (DeJesus-Hernandez et al., 2011b). It was also found that the expansion is capable of forming secondary DNA or RNA structures such as G-quadraplexes and DNA/RNA hybrids, which induce the generation of abnormal transcripts sequestering RNA binding proteins (Haeusler et al., 2014).

A wide range of clinical diversity has been described in *C9orf72* ALS cases. In *C9orf72* positive cases, as detected using rpPCR, there were reports of increased frequencies of severe phenotypes, such as bulbar onset, early onset, reduced survival time and increased FTD comorbidity (Byrne et al., 2012; Cooper-Knock et al., 2014) and bvFTD was the subtype enriched for the *C9orf72* epxansion (Cruts et al., 2013). However, as the rpPCR method is unable to discriminate the size of larger expansions (>60 Units), the relationship between expansion size and clinical phenotype has not been thoroughly investigated. Characterising expansion sizes is also important for understanding of the threshold size for pathogenicity, stability of the expansion and anticipation. Different Southern blot protocols, both radioactive and non-radioactive, have been used for sizing *C9orf72* expansions (Beck et al., 2013; Buchman et al., 2013; DeJesus-Hernandez et al., 2011b). In this study, we performed Southern blots in our FALS cohort which have been previously shown to possess the expansion using rpPCR using a protocol originally proposed by Dejesus et al (2011b).

7.2 Methodology

7.2.1 Subjects and sample preparation

The study population comprises 65 FALS cases known to harbour the *C9orf72* expansion, which account for 31.3% of the IC-FALS cohort (**Section 1.19.1**). Genomic DNA was used from whole blood or buffy coats from 20 cases, from LBCs of 4 cases. Post-mortem spinal cord was used from 4 cases and cerebral cortex extracts were used from a single case.

DNA from blood or LBC sources was extracted using a QIAmp DNA Blood Midi Kit (QIAGEN, UK), whereas those from tissues were extracted using a QIAmp DNA Mini Kit for tissues (QIAGEN, UK). DNA was concentrated to a final volume of 30µl using a centrifuging concentrator (Eppendorf 5301) or ethanol precipitation (**Section 2.10.1**).

7.2.2 Southern Blot analysis

The Southern blot protocol is described in detail in **Section 2.10**. In brief, DNA was digested using 50-100 Units of XbaI (New England Biolabs, UK) in a 50µl reaction mix for 16 hours at 37^{o}C and separated in electrophoresis using 0.8%, 8mm agarose gel at 3-4V/cm for 16 hours. The DNA was then transferred to a nylon membrane using 20X SSC overnight. The membrane was UV-crosslinked and pre-hybridized in 10-20 ml of DIG Easyhyb for 3 hours at 48^{o}C. Oligonucleotide probes labelled with Digoxigenin (DIG) were synthesized using a PCR DIG Probe synthesis kit (Roche, UK) with the following primers: Probe 1 (241 bp): Forward: AGAACAGGACAAGTTGCC; Reverse: AACACACACCTCCTAAACCC. Probe 2 (367bp): Forward: AAGTGCCATCTCACACTTGC; Reverse: CCCTGGTAGGGGACAGCTC. Hybridization was carried out using 0.5-3µl/ml of probe for 16 hours at 48^{o}C and the membrane was washed in 2XSSC, 0.1%SDS for 15 minutes at room temperature; 0.1XSSC, 0.1% SDS for 15 minutes at 68^{o}C twice. Chemiluminesence signals were detected using DIG Wash and Block Buffer Set (Roche, UK) according to the Manufacturer's instructions. In the detection stage, 250µl of CDP-star (Roche, UK) was mixed with 20ml of detection buffer and incubated with the membrane in a sealed hybridization bag for 5 minutes prior to exposure. The membrane was then exposed to X-Ray Film from 30 minutes to 12 hours.

7.2.3 Statistics

Expansion sizes were calculated using GelAnalyzer v2010 (*http://www.gelanalyzer.com/*). To assess the relationship between expansion size and clinical phentypes, correlation and regression analyses were carried out using Graphpad. Survival analyses were carried out using Kaplan-Meyer curves (Graphpad) by splitting the data at 25%, 50% and 75% percentiles of sizes. Frequencies of site of onset were compared in cases harbouring expansions of different sizes using 2 x 2 Fisher's tests.

7.3 Results

7.3.1 Optimization of Southern blot protocol using different probes.

Considering that variable expansion sizes have been described, two commonly used markers were calibrated (**Figure 7-1 A**). The two markers yielded different calibration curves and using the DIG labelled DNA molecular marker III (Roche, UK) alone may result in an error of up to 6kb in the measurement of molecular weight (MW) in <2kb or >5kb regions. The DIG marker VII yielded more reliable size predictions in the 1-5kb region, suggesting that DIG marker III should only be used for the calculation of MW beyond the range of marker VII (**Figure 7-1 B**).

The signal generated by a probe is directly proportional to the amount of DIG-labelled-dUTP which is randomly incorporated during PCR reaction. A longer probe gives rise to a stronger signal and requires a higher annealing temperature. *It was hypothesized that the concomitant binding of two similar sized probes targeting different sequences may enhance the signal at a similar hybridization temperature*. Therefore, we generated and tested two DIG-labelled probes, including a 241bp probe (Probe 1), which was originally used by Dejesus et al (2011), and a 367bp probe (Probe 2), which is a novel "in hosue" probe located 5' upstream from Probe 1 (**Figure 7-1C**). No additional human genomic sequences that might lead to non-specific binding were identified for either probes using BLAST.

To test the specificity of both probes, three independent filters, each blotted with two FALS samples known to carry the *C9orf72* expansion, were hybridized using Probe 1, Probe 2, and Probe 1+2 respectively. Hybridization was carried out at 48^{o}C for 16 hours and the filter was washed in 2x SSC, 0.1%SDS at room temperature for 15 minutes, followed by 0.1XSSC, 0.1%SDS at 68^{o}C for 15 minutes twice. For both probes, the wild type bands were present in both samples and the "large" band for sample (i) was only present when Probe 1 alone was used. For sample (ii), two "large" bands were observed for Probe 1 and and three for probe 2, respectively, only one of which was common for both probes, suggesting a genuine band (**Figure 7-1D**). As expected, the signal was enhanced when both probes were used together but non-specific binding was also increased. These results suggest that the binding of Probe 1 is more specific than Probe 2, and, therefore, only the former was chosen for subsequent experiments.

7.3.2 Characterisation of C9orf72 expansion size using Digoxigenin-based Southern Blots.

Our cohort contains **65** FALS patients possessing the *C9orf72* expansion as previously confirmed using rpPCR (repeat-primed PCR) methodology. Southern blots were carried out using whole blood, spinal cord tissues or LBCs derived from the FALS patients using Probe 1. Expansion sizes were estimated using the GelAnalyzer program (**Figure 7-1D**) and genotypes of **8** samples from **7** unrelated kindred were obtained, all of which were from Buffy coat extracts (**Figure 7-2**) or LBCs (**Figure 7-3**). The wild type alleles were detectable in all of these samples and showed moderate variations. Only discrete bands were taken into account as expansions. The expansion sizes range from **~3.9kb (~240 Units)** to **~20.4kb (~3011 Units)** with a median size of **8.76kb (1049Units) (Figure 7-4A**). The FALS sample G48 was replicated in two independent experiments yielding expansion sizes of **11.6kb** and **12.8kb**, respectively, which were within the sizing error found using the DIG marker III (**Figure 7-2**).

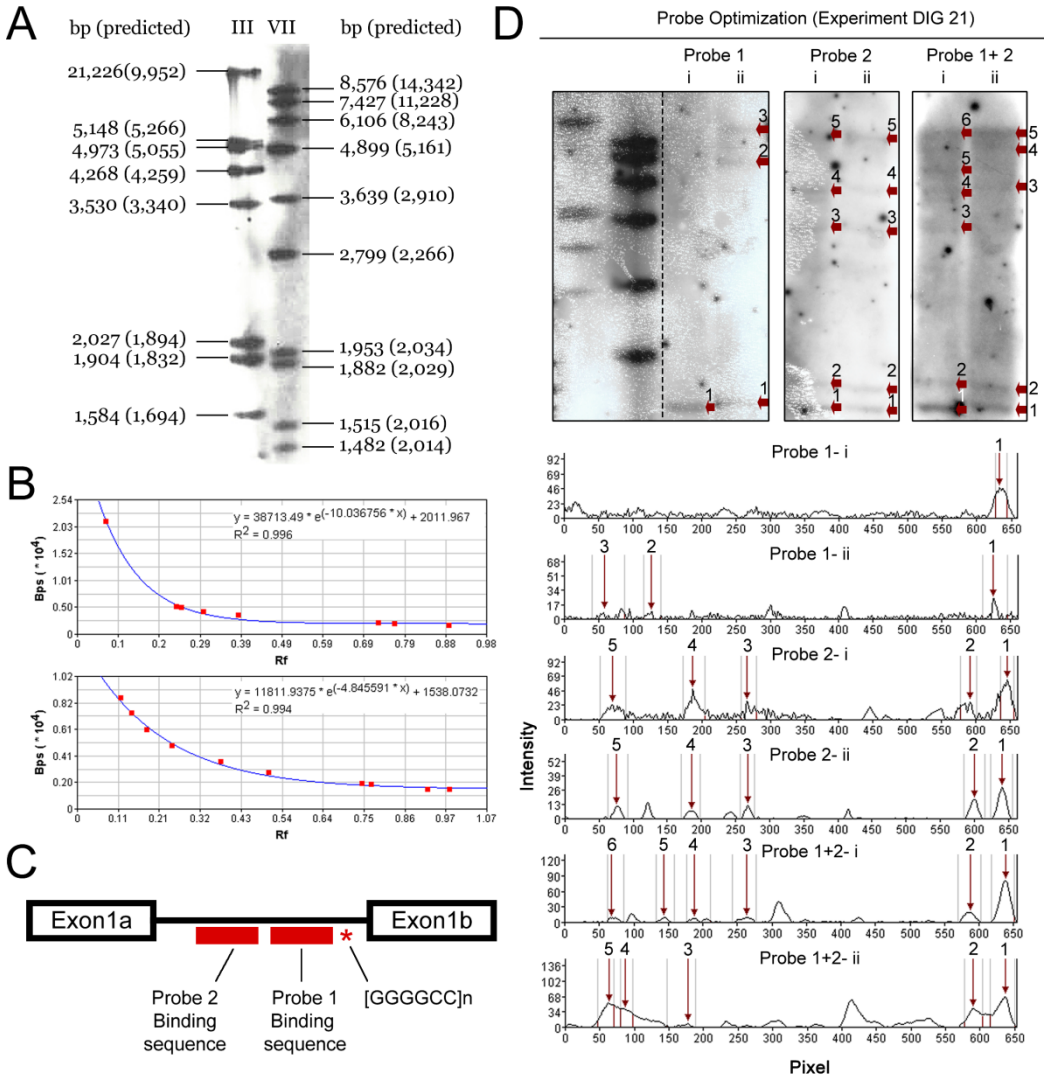
It should be noted that two of the genotyped samples, AO620 **(~20.4kb, 3011 Units)** and AO621 **(~17.5kb, 2534 Units)** were from the same kindred. Their expansion sizes were in the same range but *decreased* in latter sample, which was from a younger generation and had an onset 17 year earlier than the former. In addition, an expanded allele sized **~2.5kb (25 Units)** was detected in a single control sample (buffy coat, data not shown).

7.3.3 Correlation between C9orf72 expansion size and clinical phenotypes.

Details of age of onset were available for all **8** samples and disease duration was known for **7** samples and these were analysed for correlation with expansion sizes. No significant correlation between expansion size and age of onset (**Pearson's r=0.1148, CI95=-0.64-0.76 P=0.7866**) (**Figure 7-4B**) or survival time was observed (**Spearman r=-0.0522,CI95=-0.77-0.73 P=0.9115**) (**Figure 7-4C**). No significant difference was observed in survival analysis or site of onset was observed. Gender information was available for all 8 samples (2 male and 6 female) but no differences could be concluded.

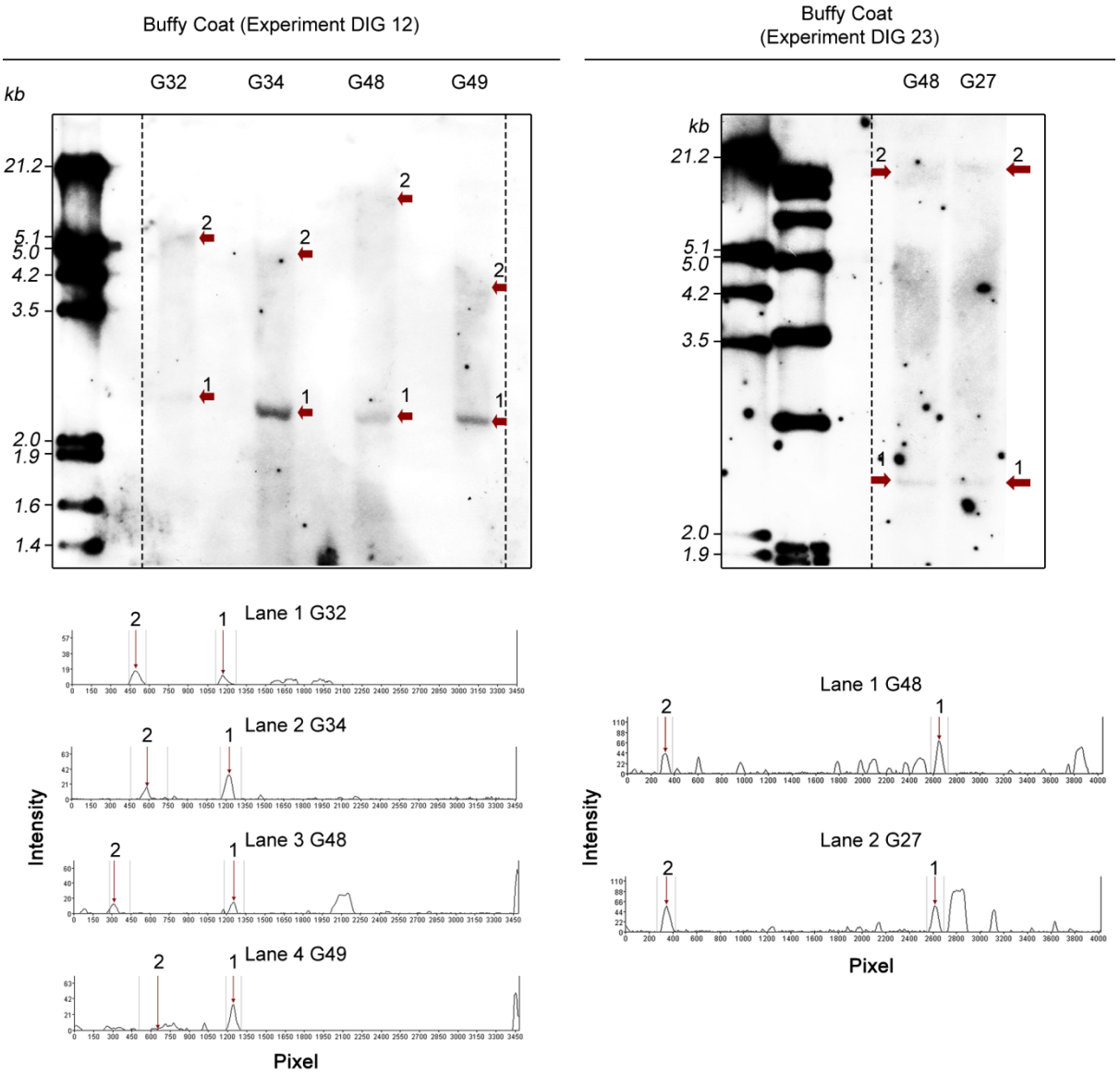
.

Figure 7-1. Representative Southern blot results and the sizing of expanded alleles.



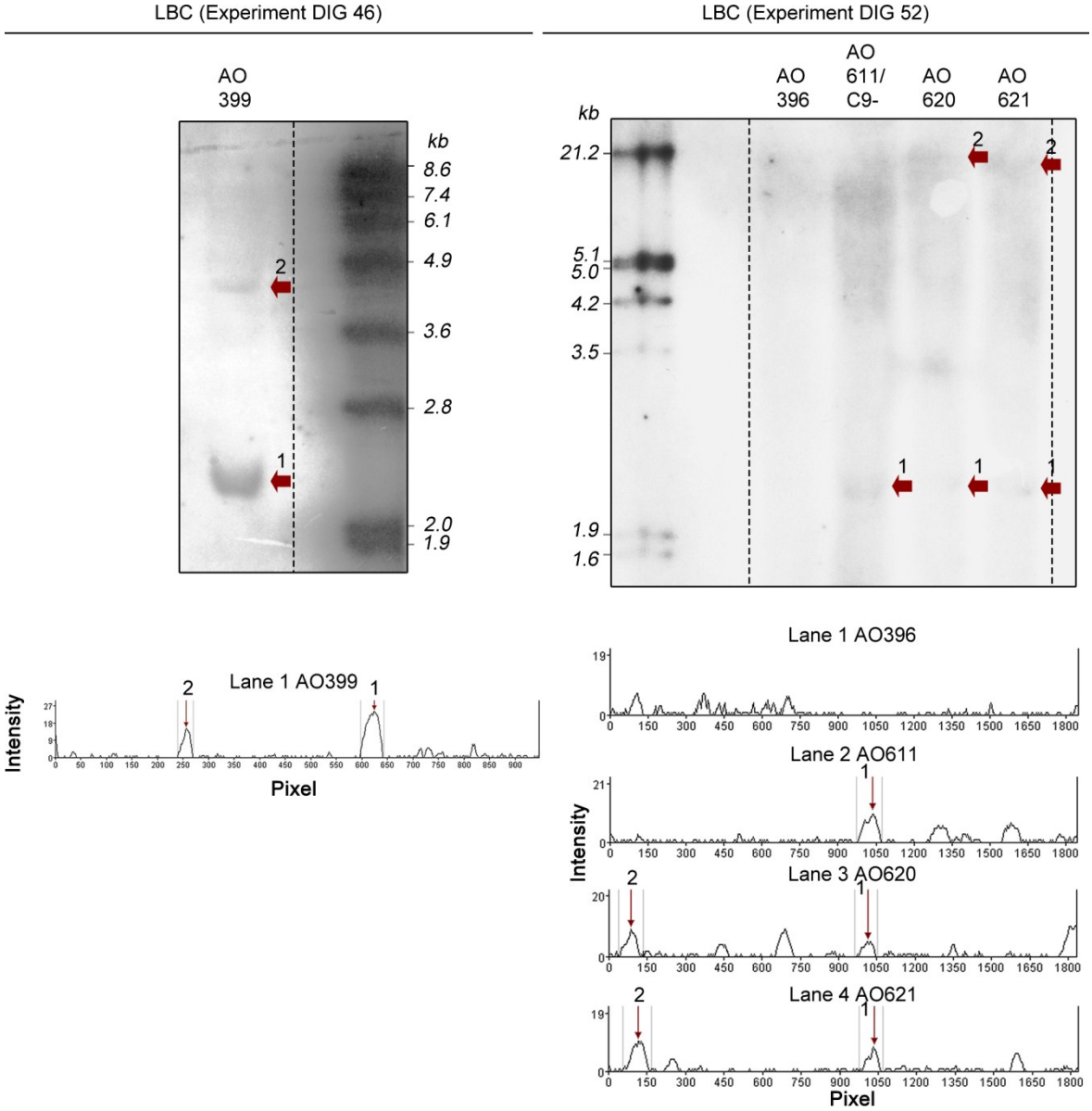
DIG labelled DNA molecular marker III (Roche, UK) and VII in a typical Southern's blot are shown in **A**. To assess the relative migration of bands of different sizes, the actual molecular weights of each marker bands are compared with the predicted weights (brackets) when using the other set of markers as standard. **B** shows the calibration curves for each set of markers, in which the molecular weight (**Bps**) versus relative migration distance (**Rf**) was fit using a logarithmic model (R^2 >0.99 for both markers). Using marker III, the curve becomes flat when **Rf** is >0.39, whereas using marker VII, there is better discrimination in small molecular weight (<8kb) in spite of a smaller detection range. **C**. Probes 1 (241bp) and 2 (357bp) both bind *C9orf72* genomic sequence 5' upstream to the hexanucleotide expansion. **D**. Two *C9orf72* positive FALS cases (i and ii) were hybridized using either probes in the following conditions: 3µg of genomic DNA (buffy coat), digested with 25U of XbaI and hybridized using 3µl/ml of DIG labelled probe (each) in 10ml of EasyHyb for 16 hours at 48°C. Intensity profiles for each lane are shown in the lower panel. Background signals were eliminated using a "rolling ball" method and the peaks corresponding to known bands are indicted by red arrows. The peaks labelled as "1" correspond to the wild type alleles.

Figure 7-2. *C9orf72* Southern blot results using Buffy coat DNA.



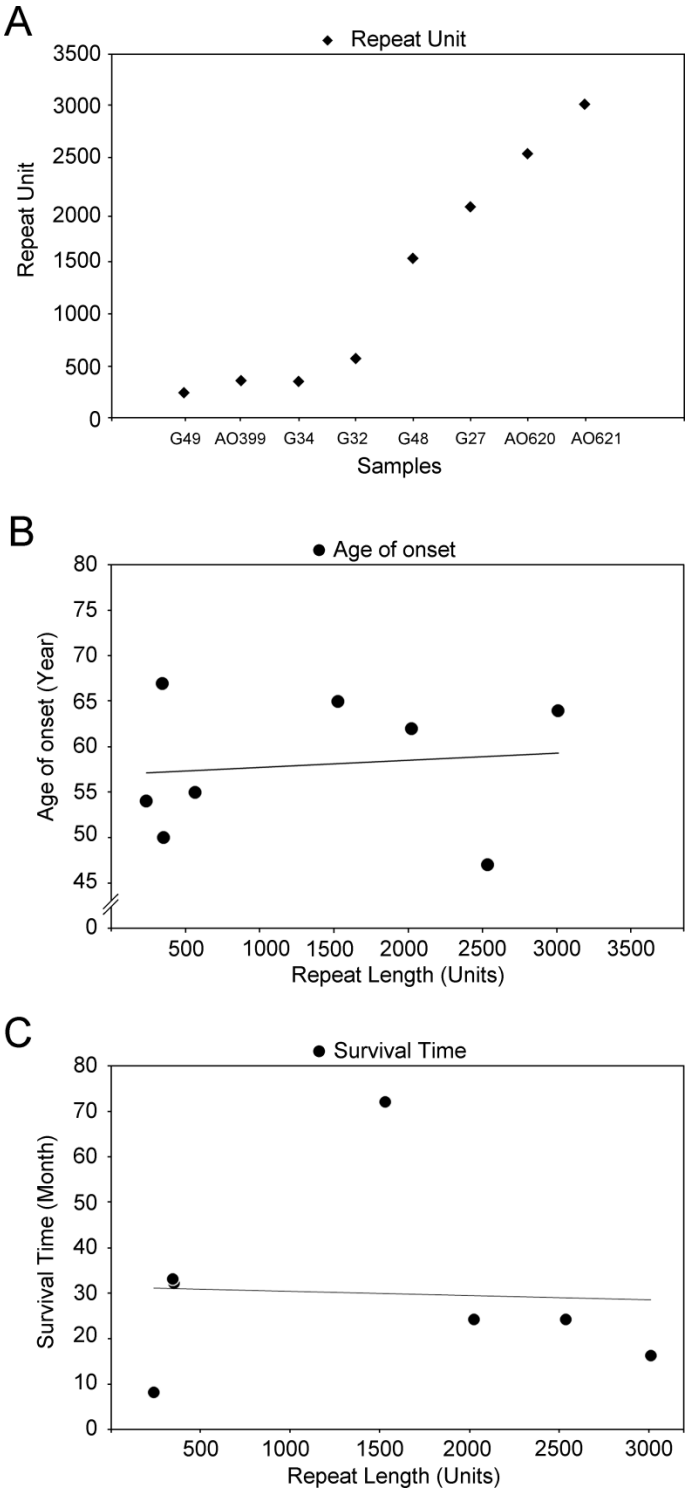
The filters were blotted using 3-10µg of genomic DNA extracted from buffy coat and hybridized using Probe 1. The estimated expansion sizes of G32, G34, G48 and G49 in the left blot are: 5864bp (567 Units), 4519bp (346 Units), 11616bp (1530 Units) and 3878bp (240 Units) bp respectively. For the right blot, the estimated size of G48, which was replicated, and G27 are 12797bp (1741 Units) and 14511bp (2025 Units) respectively. Intensity profiles are shown in lower panels and the bands are indicated using red arrows.

Figure 7-3. *C9orf72* Southern blot results using LBC DNA.



Both filters were blotted using 20-30µg of DNA extracted from LBC. The estimated expansion size of AO399 (left) is 4341 bp (354 Units). No visible bands were observed in the first lane (AO396) in the filter on the right hand side. The second lane, AO611, is a *C9orf72* negative control and a smear was present indicating non-specific binding. Although smears were also present in AO620 (20401bp, 3011Units) and AO621 (17542bp, 2534Units), relatively discrete bands were found as indicated in the intensity plots (arrow "2").

Figure 7-4. Correlation and regression analyzes of clinical phenotypes against *C9orf72* expansion sizes.



A. summarizes expansion sizes of all genotyped samples. Correlation and regression analysis between expansion size and age of onset are shown in **B** and **C** respectively. Pearson's r, r^2 and P values are mentioned in the main text.

7.4 Discussion

The *C9orf72* GGGGCC expansion has been shown to be the most common cause of familial ALS and FTD but the relationship between expansion size and phenotype is not yet confirmed. In this study we optimized a non-radioactive Southern blot protocol, originally proposed by (DeJesus-Hernandez et al., 2011b), using two digoxigenin-labelled probes and demonstrated that the 241bp Probe 1, which was originally used by Dejesus et al, leads to more specific binding than the in-house 371bp, Probe 2. In addition, the co-binding of both probes did not enhance the signal-to-background ratio, which can only be improved by increasing sample quantity (5-30ng) and prolonging exposure time. However, as a result, a marked portion of samples in our cohort could not be genotyped due to insufficient template concentrations. Considering the limited separation of large sized bands, we suggest that the DIG marker III (Roche) should only be used for expansions beyond the range of smaller markers such as DIG marker VII (**Figure 7-1 A**). We characterized the expansion size in 8 FALS cases, all showing discrete bands with consistent sizes, when replicated. The range of expansion sizes coincides with previous reports. A control individual possessing a small expansion with an estimated size of ~25 repeats was observed, which was within the previously reported normal range (< 60Units) (Renton, 2011), but this was not found in any of the FALS cases.

C9orf72 expansions are not specifically found in any particular ALS phenotypes and there were only few studies of the relationship between *C9orf72* repeat size and phenotype. Although variation in the non-expanded alleles (< 32 units) is not associated with any clinical parameters (Rutherford et al., 2012), expansion sizes in the frontal cortex are correlated with age of onset in FTD and those of >1467 Units in the cerebellum are associated with short survival time in a combined cohort of FTD and MND (van Blitterswijk et al., 2013). In addition, a positive correlation between expansion size (LBC) and age of onset has been recently reported in a pure MND cohort (Hubers et al., 2014). In the current study, we were able to investigate phenotypic associations using Southern blots but no significant correlation was found. Our findings are in agreement with other Southern blot-based studies using pure ALS cohorts (Beck et al., 2013; Dols-Icardo et al., 2014). Despite not being significant, it is noteworthy that a positive correlation coefficient between expansion size and age of onset was found in our cohort.

In addition to the relationship between repeat size and phenotype severity, it is important to answer the question of whether *C9orf72* disorders share the two important features of expansion diseases, clinical anticipation and repeat instability (Renton et al., 2014). Anticipation has been previously reported in *C9orf72* ALS kindred, in which children had an earlier onset (average 7 years) than their parents (Chiò et al., 2012), but whether this is related to expansion size was unclear (Cooper-Knock et al., 2014). In the current study, there are two kindred where multiple samples were available. In both kindred, age of onset was younger in children compared to their parents, indicating anticipation. In Kindred I, only the affected offspring AO399 (onset at **50** years) was genotyped and the size of the affected parent AO396 (onset at **75** years) was inconclusive. In kindred II, the offspring (AO621, onset at **47** years) had a smaller expansion than the parent (AO620, onset at **64** years), indicating that the expansion size is unrelated to, if any, anticipation. It should be noted that genotyping in both of these kindred was carried out using DNA from LBCs, which may show considerable size variation compared to genomic DNA form blood (Dols-Icardo et al., 2014). Indeed, it appears that *C9orf72* expansion size may change during life time and may be affected by the age at collection (Cooper-Knock et al., 2014). Therefore, no conclusion can still be made on whether there is tendency in altering size in successive generations. It was demonstrated that the non-expanded repeat varied amongst healthy kindred but no evidence of increased instability or perfectness have been reported to date (Beck et al., 2013).

There are several challenges to characterizing the genotype- phenotype relationships of the *C9orf72* repeat in ALS. First, as discussed above, the separation of high molecular weight alleles is not satisfactory and there can be up to 6kb errors in estimating expansion size using Southern blots. The error can be minimized by using different markers for different size ranges. Secondly, the repeat demonstrates marked somatic mosacism and the sizes obtained from different tissues may show different correlation with clinical phenotypes. It was proposed that the sizes in cerebellum are more stable reflecting the actual size causing diseases (van Blitterswijk et al., 2013), but expansion sizes in the spinal cord, the major tissue being affected in ALS, have not been thoroughly investigated using Southern blots. Although it may not turn out to be a useful disease marker in clinical practice, spinal cord repeat size may provide more insight in the expansion. However, studies using such tissues are likely to have limited power and require extended cohorts.

Thirdly, if expansion size is not simply a reflection of age at collection, the positive correlation between *C9orf72* expansion size with age of onset (Hubers et al., 2014) is inconsistent with known pathogenic mechanisms seen with expansions. Although haploinsufficeny has been proposed as a pathogenic mechanism underlying the expansion, ALS/FTD Patients harbouring homozygous *C9orf72* expansions, as confirmed by Southern blot, did not develop extreme phenotypes beyond the spectrum of heterozygous patients, in agreement with a gain-of-function hypothesis (Fratta et al., 2013). In VNTR disorders caused by such a mechanism, for example, DM1, which is caused by an intronic expansion generating toxic mRNA, large expansion sizes are often associated with early age of onset (Woollacott and Mead, 2014). Having said that, none of the *C9orf72* gain-of-function models, such as forming RNA foci and DPR inclusions have been correlated with neurodegeneration to date [reviewed by (Mackenzie et al., 2014)].

In conclusion, we optimized a Southern blot protocol and carried out initial assessment of the *C9orf72* expansion size in our FALS cohort. No significant association with any clinical parameter was found. Our results do not support the idea that *C9orf72* expansion size underlies clinical anticipation. The power of the current study power is limited by the small sample size used and further investigations in additional samples are required.

-END OF CHAPTER 7-

Chapter 8

Discussion

8.1 Overview: the current and changing view of ALS genetics and proteostasis as a pathogenic mechanism

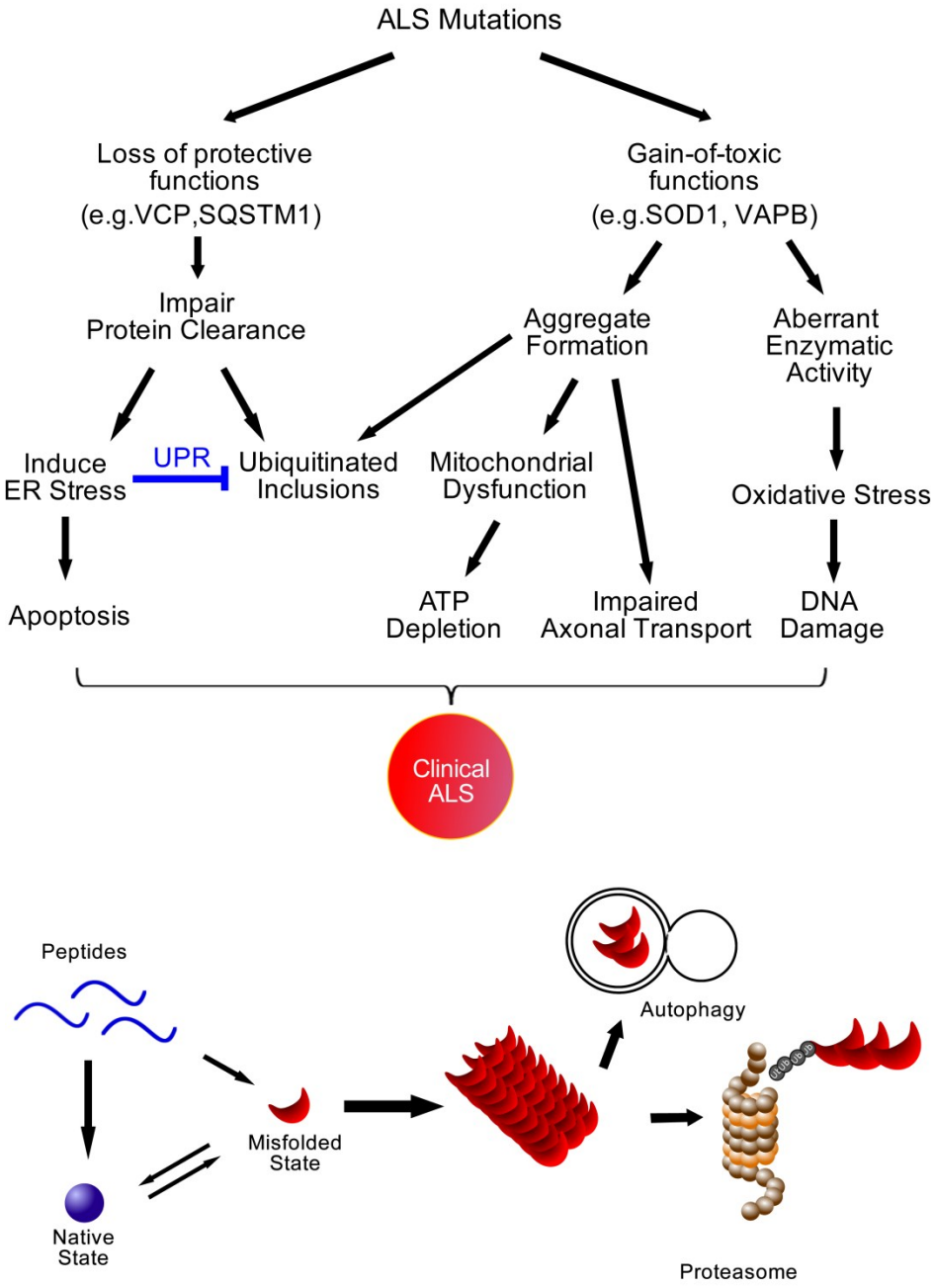
Amyotrophic lateral sclerosis (ALS) is a lethal disorder characterized by the degeneration both upper and lower motor neurons, including Betz cells in the motor cortex, descending axons which form the corticospinal tract and lower motor neurons in the anterior horn of the spinal cord, accompanied by the proliferation of glial cells (Leblond et al., 2014). Despite being clinically indistinguishable, ALS has been empirically classified into Familial ALS (FALS) and Sporadic ALS (SALS) according to family histories. A general consensus for defining a positive familial history has been concluded as the presence of ALS in the first or second degree relatives of the index case (Byrne et al., 2011), yet this can be complicated by the fact that ALS can co-occur with other conditions such as Frontotemproal Dementia (FTD), Paget's disease of bone (PDB) and Inclusion body myopathy (IBM). Considerable advances hsve been made on the genetics of FALS cases satisfying Mendelian inheritance and this has promoted the understanding of disease mechanisms. To date at least 18 genes have been found to cause for FALS since the identification of *SOD1* mutations.

The presence of protein aggregates is an important feature of ALS and the mechanisms for their formation and pathogenic effects have been understood by studying SOD1-FALS models. SOD1 aggregates represent a distinctive pathological entity compared to non-SOD1 ALS cases (Mackenzie et al., 2007) and can mediate multiple downstream effects that are part of the general mechanisms of ALS, such as oxidative stress, impaired axonal transport and mitochondrial dysfunction (**Figure 8-1**). Although it was postulated that the formation of aggregates was triggered by conformational abnormalities of the mutant SOD1 protein, different types of aggregates containing such as TDP-43, FUS and C9orf72 DRP were subsequently found as components of neuronal inclusions, indicating the involvement of a common proteostasis pathway. This is further supported by the

identification of mutations in genes, *VCP* and *SQSTM1*, that are directly involved in the major protein clearance pathways: Ubiquitin-proteasome system (UPS) and Autophagy (**Figure 8-1**).TDP-43 aggregates are characteristic of both ALS and FTD, suggesting that the conditions are, indeed, two extremes of the same spectrum of disease, and it was further demonstrated in *VCP* and *SQSTM1* kindred that both conditions may manifest in the presence of the same genetic causes.

Recent research has led to the re-examination of rare, functional genetic variants as risk factors for ALS (Renton et al., 2014). It has been generally assumed in FALS cases that a single locus is responsible for each kindred which follows Mendelian inhieritance. However, an oligogenic model of inheritance has been recently proposed to explain the facts that penetrance is incomplete for some mutations and, in some kindred, more than one known ALS mutation is present (van Blitterswijk et al., 2012a). It was shown, in such kindred, that a second mutation was required for the development of ALS, and, instead of being “directly pathogenic”, or being fully penetrant, each mutation confers a different risk for disease. This model is consistent with the notion that ALS pathology is consequential to the interaction between different genetic variants and, hence, can be modified by multiple components in relevant pathways. In contrast, whilst SALS was considered as a complex disease and its risk factors mostly being common polymorphisms, germline *de novo* mutations occurring in unaffected parents have been reported to be enriched in ALS cases (Chesi et al., 2013). Known ALS mutations such as *SOD1*, *SQSTM1* and *C9orf72*, have been detected in a substantial portion of SALS giving rise to an *allelic* model for the sporadic cases (reviewed by(Renton et al., 2014). Importantly, both *oligogenic* and *allelic* models may point to the same deduction that low penetrance of a variant is likely to be due to the existence of genetic modifiers which “neutralize” its pathogenic effects. It would also not be surprising that if these modifiers existed, they would have universal effects on both FALS and SALS. Our studies may promote insight into these concepts. First, we performed sequence analysis in both *C9orf72* positive and negative FALS cases and presented candidate variants that co-exist with *C9orf72* expansions. Second, including both FALS and SALS cohorts in the association studies, surprisingly, we found associations that were specific for either FALS or SALS.

Figure 8-1. ALS mutations involved in Proteostasis pathways.



Current views of pathogenic mechanisms of ALS mutations that disrupt proteostasis are summarized in the upper panel, whereas the lower panel shows the pathways of formation and clearance of misfolded proteins. PDI facilitates the exchange between native and misfolded states in the ER (bi-directional arrows). Misfolded proteins, which may oligomerize to form aggregates, can be eliminated through autophagy or proteasomal degradation which take place in the cytoplasm.

In an attempt to clarify the genetic causes in our FALS cohort and to determine genetic risk factors and modifiers for ALS, we carried out sequence analysis and association studies in our cohorts. Genes playing roles in proteostasis pathways comprise the main body of candidate genes sequenced in this study. Sequence analysis was carried out for the *VCP* and *SQSTM1* genes, both take part in the extraction or shuttling of substrates for protein degradation, in our FALS cohort. In association studies we investigated Protein disulphide isomerase (PDI), which plays an important role in protein quality control. The ER redox enzyme prevents inappropriate disulphide bonds formation and therefore is an appropriate candidate as modifier gene. Another candidate in the association study is Nuclear pore localization protein 4 (Npl4), a cofactor for VCP.

Whilst most known variants in ALS had been point mutations, the discovery of the *C9orf72* expansion as the most common cause of ALS/FTD delineated a more complicated landscape of ALS genetics. Two fundamental questions are to be answered: Are there any other expansions that are causal or susceptible for ALS? Does ALS exhibit characteristics that are seen in other neurodegenerative diseases caused by expansions? Addressing these questions, we first attempted to investigate known Variable number tandem repeats (VNTRs) in novel candidate genes as causes or risk factors for ALS, followed by measuring the expansion sizes of the *C9orf72* expansion in our cohorts using Southern blot analysis.

8.2 Sequence analysis

Following the screening for selected exons in the *VCP* gene we did not identify any coding mutations. This study was carried out prior to the identification of the *C9orf72* expansion and C9-positive cases were therefore also included for analysis. The results were not unexpected considering the low frequency (~1%) of *VCP* mutations.

We report an imperfect hexanucleotide repeat, c.-221_-220insCTGCCACTGCCACTGCCG, in the 5' UTR region and demonstrated that the expanded genotypes were exclusively found in ALS. No frequency data was available for this repeat in any public database and, by examining the BAM files from the 1000 genome project, it was found that the repeat was poorly covered for genotyping. The result that

repeat number was increased from Control to SALS and FALS led to the hypothesis that expansion size is associated with susceptibility or penetrance of disease. However, neither segregation with disease nor penetrance could be investigated due to lack of multiple samples in the affected kindred. There was no indication of gaining instability of the repeat, however we do not exclude the possibility that exceptionally large expansions have been lost during PCR. Location of the repeat at the transcription start site of the *VCP* gene is indicative of a functional effect, however, no significant alteration in VCP expression levels can be concluded using a semiquantitative endpoint rtPCR.

VCP mutations are known to be rare and mediate variable phenotypes known as Multisystematic proteinopathy. We conclude that *VCP* mutations are not a major cause of FALS in our FALS cohort. The heaxnucleotide expansion may be pathogenic.

In addition to the functional similarities between *SQSTM1* and *VCP* genes, clinical overlaps are seen in patients harbouring mutations in these genes. Paget's disease of bone (PDB) is common to both entities. All eight exons of the *SQSTM1* gene were screened for mutations in our FALS cohort excluding any known mutations and coding mutations were identified. Six exonic variants in five FALS index cases, three of which were non-synonymous and three were synonymous. One index case harboured three variants (E274D, P296P and S318S), and a second index case harboured two variants (E274D and S318S). Considering coding changes and frequencies, it was concluded that only P392L and E155K mutations were likely to be pathogenic. The P392L mutation has an established pathogenic role and is the commonest mutation in PDB.

We reported two FALS *SQSTM1* kindred in which both ALS and PDB developed. In one P392L kindred, the proband developed both ALS and PDB. In the second E155K kindred, the proband developed pure ALS and his father developed pure PDB. These results indicate connections between ALS and PDB and that *SQSMT1*mutations may result in variable phenotypes resembling multisystematic proteinopathy caused by *VCP* mutations. Although PDB has a higher prevalence than ALS, a primary association between *SQSTM1* mutations and ALS was illustrated in the E155K kindred as the proband developed pure ALS, but not ALS/PDB. Following the screening for the P392L mutation in additional *C9orf72* positive FALS cases, we found one additional P392L kindred

presenting with pure ALS. A shorter survival time was observed in this case compared to the *C9orf72* positive P392L case and we speculated that this may be attributed to the co-existence of both mutations.

It has been shown in British PDB cases that most P392L mutations come from a founder haplotype, and we also showed that this is also the case in ALS. However, the E155K mutation was found to arise from a different haplotype background.

In conclusion, our results confirm the presence of *SQSTM1* mutation in FALS. The frequency of the P392L mutation in this UK FALS cohort was 2.3% and 0.95% overall including three previously screened FALS cohorts. This mutation is the most common *SQSTM1* mutation found in ALS to date, and a likely pathogenicity is supported by having an established causal role in PDB. The occurrence of the same mutation in ALS and PDB is indicative of a common pathogenic pathway that converges on proteostasis.

8.3 Association studies

Common Single nucleotide polymorphisms (SNP) were investigated as risk factors and modifiers for ALS using association studies. The two candidate genes, *P4HB* and *NPLOC4*, are located in telomeric region of the Chromosome 17 and are poorly represented in previous GWA studies. SNPs were investigated in both FALS and SALS cohorts and those showing significances in the initial studies were further studied in expanded cohorts. We selected representative common SNPs, with >15% minor alleles frequencie (MAF), based on linkage disequilibrium (LD) throughout the genes.

The *P4HB* gene is associated with FALS using multiple statistical analyses. Firstly, there were significant genotypic associations for two SNPs in *P4HB* gene with FALS, SNP 2 (rs876016, **P=0.0198**) and SNP 3 (rs2070872**, P=0.0046**), all values being FDR corrected. For allelic associations, significances were obtained for SNP 2 in FALS (**P=0.0155**) and ALS as a combined group of FALS and SALS (**P=0.0148**). Secondly, we examined four SNP haplotypes including two additional flanking SNPs, SNP 1 (rs876017) and SNP 4 (rs8324), and found that, consistent with the decrease in LD, rare haplotypes were more common in ALS cases compared to controls. Seven haplotypes were found to be

significantly associated with FALS and one haplotype was significantly associated with SALS. One rare haplotype, which was present in controls, was overrepresented in a group of *SOD1* positive FALS cases.

P4HB may play a role in the progression of ALS as survival effects were also found. Reduced survival was observed in FALS cases possessing at least one copy of the minor allele of SNP 3 (rs2070872, **P=0.0059**). Although no association with SNP 4 (rs8324) was observed, the SNP also showed survival effects (**P=0.0167**). For a combined effect of all 4 SNPs on survival, we found that individuals lacking the homozygous AAAC/AAAC diplotype had a reduced survival time (**P=0.011**).

The association results are consistent with a protective role of Protein disulphide isomerase (PDI), which reduces the formation of inappropriate bonds. This is best illustrated using the SOD1 model. In wild type SOD1, conserved disulphide bonds are formed between residues Cys57 and Cys146 which maintain protein stability, whereas in mutant SOD1 the bonds are reduced and this is a necessary step for aggregate formation (Ray et al., 2004) (Furukawa et al., 2006). Aberrant disulphide bonds, such as those between Cys6 and Cys111, are then introduced enabling crosslinking between molecules (Banci et al., 2008). Disulphide bonds link a significant portion of mutant SOD1 aggregates (Furukawa et al., 2006) and may also contribute to the crosslinking of TDP-43(Cohen et al., 2012). However, the aberrant structures have decreased stability and this can be prevented through the active exchange of disulphide bonds catalysed by PDI. In conclusion, our results suggest that *P4HB* is a modifier gene for FALS and is a potential therapeutic target.

The *NPLOC4* gene became a candidate for our association study based on several reasons. Firstly, Nuclear protein localization protein 4 homolog (NPLOC4) is an essential co-factor for the function of VCP during ERAD. Secondly, the splicing of NPLOC4 is regulated by TDP-43 and alternative transcripts have been shown to increase in the spinal cord of SALS. Thirdly, the *NPLOC4* gene is located ~200kb upstream of *P4HB* in the chromosome 17 and sub-threshold genome association signals were reported in previous GWAS studies.

Six SNPs were selected based on intragenic LD pattern in the control population. The first 3 SNPs tagged for 3 independent LD blocks from Exon 1 to 15. Interesting, a strong LD block was observed in the 3' end of the gene and is adjacent to a TG dinucleotide repeat in intron 16. With the hypothesis that the repeat is responsible for TDP-43 mediated alternative splicing, we selected two SNPs from this block in order to see whether LD is altered in ALS. Finally, a coding SNP in exon 17 which is independent from any other LD block was included. The TG- rich repeat was characterized in SALS and controls.

After correction for multiple testing, we found significant associations for three *NPLOC4* SNPs, SNP 1 (*rs6565612,***P=0.0243**), SNP 2 (*rs8075102,* **P=0.0260**) and SNP 3 (*rs7405450*, **P=0.0060**) in our SALS cohort. This was accompanied by the association of a 3-SNP haplotype H_2 which consists of the associated SNPs (**P=0.0107**). Remarkably, all three SNPs showed considerable variation in allele frequencies between genders, and the associations were further confirmed by the fact that minor allele frequencies were invariably increased in ALS cases using a Mantel-Haesnzel analysis. Size effects for all associations were enhanced in female patients. Delayed age of onset was observed in pooled ALS cases possessing the minor allele of *rs6565612* but we concluded that this is due to the gender- specific association.

It is interesting to note that, although no association was found for the TG repeat, this neither excluded its possible role as a TDP-43 binding site nor the association between repeat length and binding capacity. However, it can be deduced that, changes in TDP-43, rather than genotypes of the TG repeat, are likely to be the primary cause of the increased alternative splicing. In conclusion, we have investigated the association of *NPLOC4* gene with ALS based on the combined evidences of known gene functions and GWAS signals. Our results indicate that *NPLOC4* is a risk factor for SALS.

8.4 VNTR analysis

Dynamic mutations are frequently associated with neurological disorders and the involvement of repetitive sequences in *C9orf72*, *ATXN2* and *NIPA1* genes indicates that Variable number tandem repeats (VNTRs) may play an important role in ALS. To investigate the role of VNTRs in ALS, we investigated 20 VNTRs in 19 genes in Familial

ALS cases and controls, their selection being based on altered expression in ALS, location adjacent to associated SNPs in GWA studies and/or involvement in pathogenic mechanisms implicated in ALS. The gene *HSPB8* was of particular interest as it satisfies multiple selection criteria. HSPB8 is a molecular chaperone that assists protein folding and is overexpressed in SALS. A SNP located 54.7kb upstream of the gene, rs11064675 (P=0.0006, uncorrected value), was associated with SALS in GWA study (Schymick et al., 2007b). We therefore investigated the *HSPB8* repeat in a sporadic ALS (SALS) cohort.

We screened for abnormal expansions in all VNTRs using PCR and electrophoresis in a FALS cohort and no abnormal expansion was detected. However, as for the *VCP* expansion, it is possible that exceptionally large expansions are missed. Nine VNTRs were polymorphic in the FALS cohort and were further genotyped in Controls.

Significant allelic and genotypic associations of *Short* alleles (≤7-repeat) of the *NIPA1* repeat were detected in FALS, heterozygous genotypes being more common in FALS than controls. The repeat association with FALS was distinct from that in a previous report that the long repeat was associated with SALS (Blauw et al., 2012). We did not find significances for the long repeats, probably due to lack of power, but there was, indeed, a larger effect size compared to the previous report.

We showed that Long alleles (≥8-repeat) of the *HSPB8* VNTR were associated with SALS. Interestingly, the association was mainly contributed by the increased frequency of the 8-repeat. We asked whether the association was due to altered HSPB8 expression level mediated by the repeat but no genotype-expression relationship was observed. Our results suggest that VNTRs in *NIPA1* and *HSPB8* genes are risk factors for FALS and SALS, respectively, and that the association of *NIPA1* Short alleles may be specific for FALS. It is possible that VNTRs in other candidate genes may be risk factors for ALS and further investigations are merited.

8.5 Characterization of *C9orf72* expansion

A GGGGCC expansion in the *C9orf72* gene is the most common cause of Amyotrophic lateral sclerosis and Fromtotemporal dementia. The expansion has been found in a wide

range of clinical phenotypes but no correlation between expansion size and phenotype has been confirmed. In our cohort, **31.3%** of FALS cases had been shown to harbour the expansion using the repeat-primed PCR method. However, large expansions can not be amplified using such method and Southern blot was considered as the method of choice to measure the expansion size (Akimoto et al., 2014).

The use of different probes has been described for Southern blot analysis, the first of those being a 241bp probe labelled with Digoxigenin (DIG) using a PCR reaction (DeJesus-Hernandez et al., 2011b). The 5'-3'-DIG-labelled Probe containing five GGGGCC repeats was suggested to increase the signal but also produced smeared bands in the blot (Beck et al., 2013), whereas Buchmann et al used a ^{32}P labelled probe (Buchman et al., 2013). In this study, we optimized the Sothern blot procedure using two different DIG labelled probes. The first one (Probe 1) being the 241bp probe originally used by Dejesus et al which binds the 5' flanking sequence of the expansion, whereas the second one binds a novel 371bp probe that is located 5' to Probe 1. The probes were used to probe *C9orf72* positive samples both independently and simultaneously and it was concluded that Probe 1 gives more specific binding than Probe 2. However, consistent with other reports, we found that Probe 1 yielded a weak signal and this can only be overcome by increasing template quantitiy.

Unlike other repeat-disorders, no evidence of gaining instability has been reported in successive generations of *C9orf72* cases. There was, however, evidence of anticipation (Chiò et al., 2012). We obtained expansion sizes for 8 FALS cases from 7 unrelated families and demonstrated that the expansion size ranges from **~3.9kb (~240 Units)** to **~20.4kb (~3011 Units)** with a median size of **8.76kb (1049Units)** in our cohort. Of the analysed samples, multiple cases were available from 2 kindred. We found that age of onset was earlier in younger generations in both kindred, indicating possible anticipation. Genotypes were obtained from multiple cases in one of the two kindred and the expansion size was smaller in the offspring. Therefore, if there was anticipation, *no* association with expansion size can be concluded. However, this is to be confirmed in a larger cohort.

No significant correlation with age of onset or survival time was observed. However, this does not necessarily mean that expansion size is not associated with pathogenicity. It has

been reported that the mechanism of translation of dipeptide repeat proteins (DRP) is more efficient with increasing repeat length (Mori et al., 2013). Regarding the oligogenic model of inheritance, which is supported by our findings in *VCP* and *SQSTM1* screenings, it is possible that the effect of *C9orf72* is modified by other genetic variants.

This section presents interim results of a project aiming to characterize expansion size from different tissues, including LBC, blood, spinal cord and temporal cortex, in our FALS cohorts. To date, genotypes have been obtained from a portion of LBC and blood samples. Our results are consistent with most published reports that there is *no* correlation of *C9orf72* expansion size with any clinical parameter.

8.6 Future work

Two sequence variants identified in this study may merit further investigation, including the *VCP* hexanucletodie expansion and the *SQSTM1* E155K mutation. Both mutations are novel rare variants and their assocaition with disease will be supported if they are absent from a larger group of controls or show segregation with disease within affected kindred. To investigate the effect of the *VCP* expansion on expression, we carried out a semiquantitative endpoint rtPCR using whole blood from the carrier but no changes were evident. The effect can be better characterized by measuring expression levels in spinal cord using qrtPCR. A luciferase essay can also be carried out to confirm whether promoter activity is affected (Qiu, 2007) using systems such as the Dual-luciferase system (Promega). To do this, the *VCP* promoter region containing the expansion is amplified using the heterozygous carrier as a template. Wild type and expanded fragments are separated in a long agarose electrophoresis and independently cloned into a Luciferase reporter vector, which lacks eukaryotic promoter/enhancer sequences. After that, cells in which *VCP* are abundantly expressed are transiently co-transfected with the *VCP* construct along with an internal control construct containing a different luciferase gene. A positive control is made by transfecting the cells with a control luciferase vector containing both promoter and enhancer seuqences and lacks the *VCP* promoter. The cells are then lysed and luciferase reagents are added to generate luminescent signals, which represent expression induced by the cloned sequences.

Functional analysis can be carried out to follow up the *SQSTM1* E155K mutation, which is hypothesized to change protein properties. Mutant forms of *SQSTM1* can be generated using a traditional two step mutagenesis PCR using a full length *SQSTM1* cDNA as template and cloned into a mammalian expressing vector containing GFP. The constructs are then used to transfect a neuronal cell line (NSC-34) in which intracellular distribution of WT and mutant SQSTM1 proteins are visualized using immunocytochemistry using anti-GFP antibodies. Western blot is subsequently carried out to quantify apoptotic markers in cells expressing the mutatnt proteins. We have used a similar approach to study pathogenic effects of *VAPB* and *DAO* mutations associated with FALS (Chen et al., 2010; Mitchell et al., 2010; Paul and de Belleroche, 2014).

Results from our association studies also suggest further follow-up studies. Firstly, the *NPLOC4* is a novel candidate gene in ALS and alternative splicing has been shown to be increased in ALS spinal cord upon the depletion of TDP-43. The alternatively spliced transcript lacks exon 17 which is replaced by the original intron 16 encoding an elongated protein. qRT PCR should be carried out to investigate its expression in ALS spinal cord and whether expression of spliced isoforms are associated with any SNPs or haplotypes. In addition, spatial distribution and localization patterns of the NPLOC4 protein can be characterized in autopsy spinal cord tissues using immunohistochemistry. Secondly, although no significant association with expression was concluded for the *HSPB8* repeat, the increased difference in expression between Case/Control possessing the *Long* alleles may indicate its interactions with expression regulatory mechanisms. A luciferase assay, in which the luciferase gene is placed between a constituent promoter and the 3'UTR of interest, can be employed to better understand whether expression is affected.

We optimized a Southern blot protocol and showed *C9orf72* expansion size in a subgroup of FALS patients. Our future plan is to chacterize expansion size in all *C9or72* positive cases in our cohort. This requires further DNA extraction from different available tissues including blood, spinal cord, temporal cortex and lymphoblastoid cell lines. Showing expansion sizes in multiple affected individuals from same kindred would allow further understanding of anticipation and repeat instability. As mentioned, the current method using the 241bp probe gives faint singals for expanded alles unless a large quantity of template is used. Considering the limited amount of each sample available, it would be

worth using a radioactive Southern blot protocol with longer probes labelled with ^{32}P via nick translation, which has been shown to be more sensitive.

-END OF MAIN TEXT-

Publications led by this study

Published Articles:

Kwok, C.T., Morris, A., and de Belleroche, J.S. (2013). Sequestosome-1 (*SQSTM1*) sequence variants in ALS cases in the UK: prevalence and coexistence of *SQSTM1* mutations in ALS kindred with PDB. European Journal of Human Genetics.*22(4)*:492-6.

Kwok, C.T., Morris, A.G., Frampton, J., Smith, B., Shaw, C.E., and de Belleroche, J. (2013). Association studies indicate that Protein disulfide isomerase is a risk factor in amyotrophic lateral sclerosis. Free Radical Biology and Medicine *58*, 81-86.

Scheduled Articles (June 2014):

Kwok, C.T, Wang, H.Y., Morris, A.G., Smith, B., Shaw, C.E. and de Belleroche, J.(2014). Novel hexanucleotide expansion identified in the *VCP* Gene in Familial and Sporadic Amyotrophic Lateral Sclerosis. Neurogenetics. *Article in Review.*

Kwok, C.T., Morris, A.G., de Belleroche, J. (2014). Variable number tandem repeats in the *NIPA1* and *HSPB8* gene are associated with Familial and Sporadic Amyotrophic Lateral Sclerosis. Journal of Neurology, Neurosurgery & Psychiatry. *Article in Review.*

Kwok, C.T., Morris, A.G., de Belleroche, J. (2014). *NPLOC4* is a risk factor of Sporadic Amyotrophic lateral sclerosis. *Ready for submission.*

References

Abramzon, Y., Johnson, J.O., Scholz, S.W., Taylor, J.P., Brunetti, M., Calvo, A., Mandrioli, J., Benatar, M., Mora, G., Restagno, G.*, et al.* (2012). Valosin-containing protein (VCP) mutations in sporadic amyotrophic lateral sclerosis. Neurobiol Aging *33*, 2231 e2231-2231 e2236.

Aggarwal, S., Zinman, L., Simpson, E., McKinley, J., Jackson, K., Pinto, H., Kaufman, P., Conwit, R., Schoenfeld, D., Shefner, J.*, et al.* (2010). Safety and efficacy of lithium in combination with riluzole for treatment of amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. Lancet neurology *9*, 481-488.

Ahlemeyer, B., Bauerbach, E., Plath, M., Steuber, M., Heers, C., Tegtmeier, F., and Krieglstein, J. (2001). Retinoic acid reduces apoptosis and oxidative stress by preservation of SOD protein level. Free Radic Biol Med *30*, 1067-1077.

Akimoto, C., Volk, A.E., van Blitterswijk, M., Van den Broeck, M., Leblond, C.S., Lumbroso, S., Camu, W., Neitzel, B., Onodera, O., van Rheenen, W.*, et al.* (2014). A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories. Journal of medical genetics, *51*:419-24.

Akizuki, M., Yamashita, H., Uemura, K., Maruyama, H., Kawakami, H., Ito, H., and Takahashi, R. (2013). Optineurin suppression causes neuronal cell death via NF-kappaB pathway. J Neurochem *126*, 699-704.

Al-Chalabi, A., Andersen, P., Chioza, B., Shaw, C., Sham, P., Robberecht, W., Matthijs, G., Camu, W., Marklund, S., Forsgren, L.*, et al.* (1998). Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor. Human molecular genetics *7*, 2045-2050.

Al-Saif, A., Al-Mohanna, F., and Bohlega, S. (2011). A mutation in sigma-1 receptor causes juvenile amyotrophic lateral sclerosis. Ann Neurol *70*, 913-919.

Al-Sarraj, S., King, A., Troakes, C., Smith, B., Maekawa, S., Bodi, I., Rogelj, B., Al-Chalabi, A., Hortobágyi, T., and Shaw, C. (2011). p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. Acta neuropathologica *122*, 691-702.

Albagha, O.M.E., Visconti, M.R., Alonso, N., Langston, A.L., Cundy, T., Dargie, R., Dunlop, M.G., Fraser, W.D., Hooper, M.J., Isaia, G.*, et al.* (2010). Genome-wide association study identifies variants at CSF1, OPTN and TNFRSF11A as genetic risk factors for Paget's disease of bone. Nat Genet *42*, 520-524.

Albers, C.A., Lunter, G., MacArthur, D.G., McVean, G., Ouwehand, W.H., and Durbin, R. (2011). Dindel: accurate indel calls from short-read data. Genome research *21*, 961-973.

Altmuller, J., Palmer, L.J., Fischer, G., Scherb, H., and Wjst, M. (2001). Genomewide scans of complex human diseases: true linkage is hard to find. Am J Hum Genet *69*, 936-950.

Amos, W., Driscoll, E., and Hoffman, J.I. (2011). Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? Proceedings of the Royal Society B: Biological Sciences *278*, 1183-1188.

Anagnostou, G., Akbar, M.T., Paul, P., Angelinetta, C., Steiner, T.J., and de Belleroche, J. (2010). Vesicle associated membrane protein B (VAPB) is decreased in ALS spinal cord. Neurobiology of Aging *31*, 969-985.

Andersen, P.M., and Al-Chalabi, A. (2011). Clinical genetics of amyotrophic lateral sclerosis: what do we really know? Nat Rev Neurol *7*, 603-615.

Aoki, M., Lin, C., Rothstein, J., Geller, B., Hosler, B., Munsat, T., Horvitz, H., and Brown, R. (1998). Mutations in the glutamate transporter EAAT2 gene do not cause abnormal EAAT2 transcripts in amyotrophic lateral sclerosis. Annals of Neurology *43*, 645-653.

Atkin, J.D. (2006). Induction of the Unfolded Protein Response in Familial Amyotrophic Lateral Sclerosis and Association of Protein-disulfide Isomerase with Superoxide Dismutase 1. Journal of Biological Chemistry *281*, 30152-30165.

Atkin, J.D., Farg, M.A., Walker, A.K., McLean, C., Tomas, D., and Horne, M.K. (2008). Endoplasmic reticulum stress and induction of the unfolded protein response in human sporadic amyotrophic lateral sclerosis. Neurobiol Dis *30*, 400-407.

Banci, L., Bertini, I., Boca, M., Girotto, S., Martinelli, M., Valentine, J.S., and Vieru, M. (2008). SOD1 and amyotrophic lateral sclerosis: mutations and oligomerization. PLoS One *3*, e1677.

Barber, S., and Shaw, P. (2010). Oxidative stress in ALS: key role in motor neuron injury and therapeutic target. Free radical biology & medicine *48*, 629-641.

Barber, S.C., Mead, R.J., and Shaw, P.J. (2006). Oxidative stress in ALS: A mechanism of neurodegeneration and a therapeutic target. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease *1762*, 1051-1067.

Barradas, M., Gonos, E.S., Zebedee, Z., Kolettas, E., Petropoulou, C., Delgado, M.D., Leon, J., Hara, E., and Serrano, M. (2002). Identification of a candidate tumor-suppressor gene specifically activated during Ras-induced senescence. Exp Cell Res *273*, 127-137.

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics *21*, 263-265.

Barril, P.N., S (2012). Introduction to Agarose and Polyacrylamide Gel Electrophoresis Matrices with Respect to Their Detection Sensitivities. In Gel Electrophoresis – Principles and Basics, S. Magdeldin, ed. INTECH. *ISBN 978-953-51-0458-2.*

Bartolome, F., Wu, H.-C., Burchell, Victoria S., Preza, E., Wray, S., Mahoney, Colin J., Fox, Nick C., Calvo, A., Canosa, A., Moglia, C.*, et al.* (2013). Pathogenic VCP Mutations Induce Mitochondrial Uncoupling and Reduced ATP Levels. Neuron *78*, 57-64.

Beck, J., Poulter, M., Hensman, D., Rohrer, J.D., Mahoney, C.J., Adamson, G., Campbell, T., Uphill, J., Borg, A., Fratta, P.*, et al.* (2013). Large C9orf72 Hexanucleotide Repeat Expansions Are Seen in Multiple Neurodegenerative Syndromes and Are More Frequent Than Expected in the UK Population. Am J Hum Genet *92*, 345-53.

Beckman, J., Carson, M., Smith, C., and Koppenol, W. (1993). ALS, SOD and peroxynitrite. Nature *364*, 584.

Belsh, J.M. (1999). Diagnostic challenges in ALS. Neurology *53*, S26-30; discussion S35-26.

Benajiba, L., Le Ber, I., Camuzat, A., Lacoste, M., Thomas-Anterion, C., Couratier, P., Legallic, S., Salachas, F., Hannequin, D., Decousus, M.*, et al.* (2009). TARDBP mutations in motoneuron disease with frontotemporal lobar degeneration. Ann Neurol *65*, 470-473.

Benatar, M., Wuu, J., Fernandez, C., Weihl, C.C., Katzen, H., Steele, J., Oskarsson, B., and Taylor, J.P. (2013). Motor neuron involvement in multisystem proteinopathy: implications for ALS. Neurology *80*, 1874-1880.

Bendotti, C., and Carrì, M. (2004). Lessons from models of SOD1-linked familial ALS. Trends in molecular medicine *10*, 393-400.

Bendotti, C., Marino, M., Cheroni, C., Fontana, E., Crippa, V., Poletti, A., and De Biasi, S. (2012). Dysfunction of constitutive and inducible ubiquitin-proteasome system in amyotrophic lateral sclerosis: implication for protein aggregation and immune response. Prog Neurobiol *97*, 101-126.

Berg, D., Holzmann, C., and Riess, O. (2003). 14-3-3 proteins in the nervous system. Nature reviews Neuroscience *4*, 752-762.

Bjørkøy, G., Lamark, T., Brech, A., Outzen, H., Perander, M., Øvervatn, A., Stenmark, H., and Johansen, T. (2005). p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. The Journal of Cell Biology *171*, 603-614.

Blackstone, C. (2012). Cellular pathways of hereditary spastic paraplegia. Annual review of neuroscience *35*, 25-47.

Blauw, H.M., van Rheenen, W., Koppers, M., Van Damme, P., Waibel, S., Lemmens, R., van Vught, P.W., Meyer, T., Schulte, C., Gasser, T.*, et al.* (2012). NIPA1 polyalanine repeat expansions are associated with amyotrophic lateral sclerosis. Hum Mol Genet *21*, 2497-2502.

Borroni, B., Bonvicini, C., Alberici, A., Buratti, E., Agosti, C., Archetti, S., Papetti, A., Stuani, C., Di Luca, M., Gennarelli, M.*, et al.* (2009). Mutation within TARDBP leads to frontotemporal dementia without motor neuron disease. Hum Mutat *30*, E974-983.

Bosco, D.A., Morfini, G., Karabacak, N.M., Song, Y., Gros-Louis, F., Pasinelli, P., Goolsby, H., Fontaine, B.A., Lemay, N., McKenna-Yasek, D.*, et al.* (2010). Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. Nat Neurosci *13*, 1396-1403.

Boston-Howes, W., Gibb, S.L., Williams, E.O., Pasinelli, P., Brown, R.H., Jr., and Trotti, D. (2006). Caspase-3 cleaves and inactivates the glutamate transporter EAAT2. J Biol Chem *281*, 14076-14084.

Brettschneider, J., Van Deerlin, V.M., Robinson, J.L., Kwong, L., Lee, E.B., Ali, Y.O., Safren, N., Monteiro, M.J., Toledo, J.B., Elman, L.*, et al.* (2012). Pattern of ubiquilin pathology in ALS and FTLD indicates presence of C9ORF72 hexanucleotide expansion. Acta Neuropathol *123*, 825-839.

Brooks, B.R. (1991). The role of axonal transport in neurodegenerative disease spread: a meta-analysis of experimental and clinical poliomyelitis compares with amyotrophic lateral sclerosis. The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques *18*, 435-438.

Brooks, B.R. (1996). Natural history of ALS: Symptoms, strength, pulmonary function, and disability. Neurology *47*, 71S-82S.

Brooks, B.R., Miller, R.G., Swash, M., and Munsat, T.L. (2000). El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord *1*, 293-299.

Brown, R.H. (2001). Amyotrophic lateral sclerosis and other motor neuron diseases. In Harrison's principles of Internal Medicine, E. Braunwald, ed. (McGraw-Hill), pp. 2412-2415.

Brownell, B., Oppenheimer, D., and Hughes, J. (1970). The central nervous system in motor neurone disease. Journal of neurology, neurosurgery, and psychiatry *33*, 338-357.

Buchan, J.R., Kolaitis, R.M., Taylor, J.P., and Parker, R. (2013). Eukaryotic stress granules are cleared by autophagy and Cdc48/VCP function. Cell *153*, 1461-1474.

Buchman, V., Cooper-Knock, J., Connor-Robson, N., Higginbottom, A., Kirby, J., Razinskaya, O., Ninkina, N., and Shaw, P. (2013). Simultaneous and independent

detection of C9ORF72 alleles with low and high number of GGGGCC repeats using an optimised protocol of Southern blot hybridisation. Molecular neurodegeneration *8*, 12.

Buratti, E., and Baralle, F.E. (2008). Multiple roles of TDP-43 in gene expression, splicing regulation, and human disease. Front Biosci *13*, 867-878.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F.E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. The EMBO journal *20*, 1774-1784.

Byrne, S., Elamin, M., Bede, P., Shatunov, A., Walsh, C., Corr, B., Heverin, M., Jordan, N., Kenna, K., Lynch, C.*, et al.* (2012). Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study. Lancet neurology *11*, 232-240.

Byrne, S., Walsh, C., Lynch, C., Bede, P., Elamin, M., Kenna, K., McLaughlin, R., and Hardiman, O. (2011). Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. Journal of neurology, neurosurgery, and psychiatry *82*, 623-627.

Caccamo, A., Majumder, S., Deng, J., Bai, Y., Thornton, F., and Oddo, S. (2009). Rapamycin rescues TDP-43 mislocalization and the associated low molecular mass neurofilament instability. The Journal of biological chemistry *284*, 27416-27424.

Cai, H., Lin, X., Xie, C., Laird, F., Lai, C., Wen, H., Chiang, H.-C., Shim, H., Farah, M., Hoke, A.*, et al.* (2005). Loss of ALS2 function is insufficient to trigger motor neuron degeneration in knock-out mice but predisposes neurons to oxidative stress. The Journal of neuroscience : the official journal of the Society for Neuroscience *25*, 7567-7574.

Cardoso, R.M.F., Thayer, M.M., DiDonato, M., Lo, T.P., Bruns, C.K., Getzoff, E.D., and Tainer, J.A. (2002). Insights into Lou Gehrig's Disease from the Structure and Instability of the A4V Mutant of Human Cu,Zn Superoxide Dismutase. Journal of Molecular Biology *324*, 247-256.

Carra, S., Sivilotti, M., Chávez Zobel, A., Lambert, H., and Landry, J. (2005). HspB8, a small heat shock protein mutated in human neuromuscular disorders, has in vivo chaperone activity in cultured cells. Human molecular genetics *14*, 1659-1669.

Carrì, M., Ferri, A., Battistoni, A., Famhy, L., Gabbianelli, R., Poccia, F., and Rotilio, G. (1997). Expression of a Cu,Zn superoxide dismutase typical of familial amyotrophic lateral sclerosis induces mitochondrial alteration and increase of cytosolic Ca2+ concentration in transfected neuroblastoma SH-SY5Y cells. FEBS Letters *414*, 365-368.

Carty, M.P., Zernik-Kobak, M., McGrath, S., and Dixon, K. (1994). UV light-induced DNA synthesis arrest in HeLa cells is associated with changes in phosphorylation of human single-stranded DNA-binding protein. EMBO J *13*, 2114-2123.

Castro, P., Liang, H., Liang, J.C., and Nagarajan, L. (2002). A novel, evolutionarily conserved gene family with putative sequence-specific single-stranded DNA-binding activity. Genomics *80*, 78-85.

Cavey, J.R., Ralston, S.H., Sheppard, P.W., Ciani, B., Gallagher, T.R.A., Long, J.E., Searle, M.S., and Layfield, R. (2006). Loss of Ubiquitin Binding Is a Unifying Mechanism by Which Mutations of SQSTM1 Cause Paget's Disease of Bone. Calcified tissue international *78*, 271-277.

Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., and Nakanishi, A. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci *169*, 13-21.

Chai, J.H., Locke, D.P., Greally, J.M., Knoll, J.H.M., Ohta, T., Dunai, J., Yavor, A., Eichler, E.E., and Nicholls, R.D. (2003). Identification of Four Highly Conserved Genes between Breakpoint Hotspots BP1 and BP2 of the Prader-Willi/Angelman Syndromes Deletion Region That Have Undergone Evolutionary Transposition Mediated by Flanking Duplicons. The American Journal of Human Genetics *73*, 898-925.

Chance, P.F., Rabin, B.A., Ryan, S.G., Ding, Y., Scavina, M., Crain, B., Griffin, J.W., and Cornblath, D.R. (1998). Linkage of the gene for an autosomal dominant form of juvenile amyotrophic lateral sclerosis to chromosome 9q34. Am J Hum Genet *62*, 633-640.

Chandran, J., Ding, J., and Cai, H. (2007). Alsin and the molecular pathways of amyotrophic lateral sclerosis. Molecular neurobiology *36*, 224-231.

Chattopadhyay, M., Durazo, A., Sohn, S.H., Strong, C.D., Gralla, E.B., Whitelegge, J.P., and Valentine, J.S. (2008). Initiation and elongation in fibrillation of ALS-linked superoxide dismutase. Proc Natl Acad Sci U S A *105*, 18663-18668.

Chen, H.J., Anagnostou, G., Chai, A., Withers, J., Morris, A., Adhikaree, J., Pennetta, G., and de Belleroche, J.S. (2010). Characterisation of the properties of a novel mutation in VAPB in familial ALS. Journal of Biological Chemistry *285*, 40266-81.

Chen, Q., Vazquez, E.J., Moghaddas, S., Hoppel, C.L., and Lesnefsky, E.J. (2003). Production of reactive oxygen species by mitochondria: central role of complex III. J Biol Chem *278*, 36027-36031.

Chen, S., Zhang, X., Song, L., and Le, W. (2012). Autophagy dysregulation in amyotrophic lateral sclerosis. Brain Pathol *22*, 110-116.

Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A*., et al.* (2004). DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am J Hum Genet *74*, 1128-1135.

Cheroni, C., Marino, M., Tortarolo, M., Veglianese, P., De Biasi, S., Fontana, E., Zuccarello, L.V., Maynard, C.J., Dantuma, N.P., and Bendotti, C. (2009). Functional alterations of the ubiquitin-proteasome system in motor neurons of a mouse model of familial amyotrophic lateral sclerosis. Hum Mol Genet *18*, 82-96.

Chesi, A., Staahl, B.T., Jovicic, A., Couthouis, J., Fasolino, M., Raphael, A.R., Yamazaki, T., Elias, L., Polak, M., Kelly, C.*, et al.* (2013). Exome sequencing to identify de novo mutations in sporadic ALS trios. Nat Neurosci *16*, 851-855.

Chi, Z.L., Akahori, M., Obazawa, M., Minami, M., Noda, T., Nakaya, N., Tomarev, S., Kawase, K., Yamamoto, T., Noda, S.*, et al.* (2010). Overexpression of optineurin E50K disrupts Rab8 interaction and leads to a progressive retinal degeneration in mice. Hum Mol Genet *19*, 2606-2615.

Chia, R., Tattum, M.H., Jones, S., Collinge, J., Fisher, E.M.C., and Jackson, G.S. (2010). Superoxide Dismutase 1 and tgSOD1^{G93A} Mouse Spinal Cord Seed Fibrils, Suggesting a Propagative Cell Death Mechanism in Amyotrophic Lateral Sclerosis. PLoS One *5*, e10627.

Chiò, A., Borghero, G., Restagno, G., Mora, G., Drepper, C., Traynor, B., Sendtner, M., Brunetti, M., Ossola, I., Calvo, A.*, et al.* (2012). Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72. Brain : a journal of neurology *135*, 784-793.

Chio, A., Schymick, J.C., Restagno, G., Scholz, S.W., Lombardo, F., Lai, S.L., Mora, G., Fung, H.C., Britton, A., Arepalli, S.*, et al.* (2009). A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. Hum Mol Genet *18*, 1524-1532.

Chow, C., Landers, J., Bergren, S., Sapp, P., Grant, A., Jones, J., Everett, L., Lenk, G., McKenna-Yasek, D., Weisman, L.*, et al.* (2009). Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. American journal of human genetics *84*, 85-88.

Chow, C., Zhang, Y., Dowling, J., Jin, N., Adamska, M., Shiga, K., Szigeti, K., Shy, M., Li, J., Zhang, X.*, et al.* (2007). Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. Nature *448*, 68-72.

Ciani, B., Layfield, R., Cavey, J.R., Sheppard, P.W., and Searle, M.S. (2003). Structure of the Ubiquitin-associated Domain of p62 (SQSTM1) and Implications for Mutations That Cause Paget's Disease of Bone. Journal of Biological Chemistry *278*, 37409-37412.

Ciechanover, A. (2006). The ubiquitin proteolytic system: from a vague idea, through basic mechanisms, and onto human diseases and drug targeting. Neurology *66*, 19.

Clarke, G.M., Anderson, C.A., Pettersson, F.H., Cardon, L.R., Morris, A.P., and Zondervan, K.T. (2011). Basic statistical analysis in genetic case-control studies. Nat Protoc *6*, 121-133.

Cleveland, D.W., and Rothstein, J.D. (2001). From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. Nat Rev Neurosci *2*, 806-819.

Cohen, T.J., Hwang, A.W., Unger, T., Trojanowski, J.Q., and Lee, V.M. (2012). Redox signalling directly regulates TDP-43 via cysteine oxidation and disulphide cross-linking. The EMBO journal *31*, 1241-1252.

Collard, J.-F., Cote, F., and Julien, J.-P. (1995). Defective axonal transport in a transgenic mouse model of amyotrophic lateral sclerosis. Nature *375*, 61-64.

Comi, G., Bordoni, A., Salani, S., Franceschina, L., Sciacco, M., Prelle, A., Fortunato, F., Zeviani, M., Napoli, L., Bresolin, N.*, et al.* (1998). Cytochrome c oxidase subunit I microdeletion in a patient with motor neuron disease. Annals of Neurology *43*, 110-116.

Conklin, D., Holderman, S., Whitmore, T.E., Maurer, M., and Feldhaus, A.L. (2000). Molecular cloning, chromosome mapping and characterization of UBQLN3 a testis-specific gene that contains an ubiquitin-like domain. Gene *249*, 91-98.

Cooper-Knock, J., Shaw, P., and Kirby, J. (2014). The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype. Acta neuropathologica *127*, 333-345.

Cooper, G.M., and Shendure, J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. Nat Rev Genet *12*, 628-640.

Corcoran, J., So, P., and Maden, M. (2002). Absence of retinoids can induce motoneuron disease in the adult rat and a retinoid defect is present in motoneuron disease patients. Journal of cell science *115*, 4735-4741.

Cordell, H., and Clayton, D. (2005). Genetic association studies. Lancet *366*, 1121-1131.
Côté, F., Collard, J.-F., and Julien, J.-P. (1993). Progressive neuronopathy in transgenic mice expressing the human neurofilament heavy gene: A mouse model of amyotrophic lateral sclerosis. Cell *73*, 35-46.

Cox, L., Ferraiuolo, L., Goodall, E., Heath, P., Higginbottom, A., Mortiboys, H., Hollinger, H., Hartley, J., Brockington, A., Burness, C.*, et al.* (2010). Mutations in CHMP2B in lower motor neuron predominant amyotrophic lateral sclerosis (ALS). PLoS One *5*.

Crippa, V., Sau, D., Rusmini, P., Boncoraglio, A., Onesto, E., Bolzoni, E., Galbiati, M., Fontana, E., Marino, M., Carra, S.*, et al.* (2010). The small heat shock protein B8 (HspB8) promotes autophagic removal of misfolded proteins involved in amyotrophic lateral sclerosis (ALS). Human molecular genetics *19*, 3440-3456.

Crowe, M.L. (2005). SeqDoC: rapid SNP and mutation detection by direct comparison of DNA sequence chromatograms. BMC Bioinformatics *6*, 133.

Cruts, M., Gijselinck, I., Van Langenhove, T., van der Zee, J., and Van Broeckhoven, C. (2013). Current insights into the C9orf72 repeat expansion diseases of the FTLD/ALS spectrum. Trends Neurosci *36*, 450-459.

Czaplinski, A., Yen, A.A., and Appel, S.H. (2006). Forced vital capacity (FVC) as an indicator of survival and disease progression in an ALS clinic population. Journal of Neurology, Neurosurgery & Psychiatry *77*, 390-392.

Daigle, J., Lanson, N., Smith, R., Casci, I., Maltare, A., Monaghan, J., Nichols, C., Kryndushkin, D., Shewmaker, F., and Pandey, U. (2013). RNA-binding ability of FUS regulates neurodegeneration, cytoplasmic mislocalization and incorporation into stress granules associated with FUS carrying ALS-linked mutations. Human molecular genetics *22*, 1193-1205.

Damiano, M., Starkov, A., Petri, S., Kipiani, K., Kiaei, M., Mattiazzi, M., Flint Beal, M., and Manfredi, G. (2006). Neural mitochondrial Ca2+ capacity impairment precedes the onset of motor symptoms in G93A Cu/Zn-superoxide dismutase mutant mice. Journal of Neurochemistry *96*, 1349-1361.

de Belleroche, J., Orrell, R., and King, A. (1995). Familial amyotrophic lateral sclerosis/motor neurone disease (FALS): a review of current developments. Journal of medical genetics *32*, 841-847.

de Carvalho, M., Dengler, R., Eisen, A., England, J.D., Kaji, R., Kimura, J., Mills, K., Mitsumoto, H., Nodera, H., Shefner, J.*, et al.* (2008). Electrodiagnostic criteria for diagnosis of ALS. Clin Neurophysiol *119*, 497-503.

de Haan, J.B., Bladier, C., Griffiths, P., Kelner, M., O'Shea, R.D., Cheung, N.S., Bronson, R.T., Silvestro, M.J., Wild, S., Zheng, S.S.*, et al.* (1998). Mice with a Homozygous Null Mutation for the Most Abundant Glutathione Peroxidase, Gpx1, Show Increased Susceptibility to the Oxidative Stress-inducing Agents Paraquat and Hydrogen Peroxide. Journal of Biological Chemistry *273*, 22528-22536.

DeHoratius, C., and Silver, P. (1996). Nuclear transport defects and nuclear envelope alterations are associated with mutation of the Saccharomyces cerevisiae NPL4 gene. Molecular Biology of the Cell *7*, 1835.

DeJesus-Hernandez, M., Desaro, P., Johnston, A., Ross, O.A., Wszolek, Z.K., Ertekin-Taner, N., Graff-Radford, N.R., Rademakers, R., and Boylan, K. (2011a). Novel p.Ile151Val mutation in VCP in a patient of African American descent with sporadic ALS. Neurology *77*, 1102-1103.

DeJesus-Hernandez, M., Mackenzie, Ian R., Boeve, Bradley F., Boxer, Adam L., Baker, M., Rutherford, Nicola J., Nicholson, Alexandra M., Finch, NiCole A., Flynn, H., Adamson, J.*, et al.* (2011b). Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron *72*, 245-256.

Del Bo, R., Tiloca, C., Pensato, V., Corrado, L., Ratti, A., Ticozzi, N., Corti, S., Castellotti, B., Mazzini, L., Soraru, G.*, et al.* (2011). Novel optineurin mutations in patients with familial and sporadic amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry.

Delpy, A., Allain, A.-E., Meyrand, P., and Branchereau, P. (2008). NKCC1 cotransporter inactivation underlies embryonic development of chloride-mediated inhibition in mouse spinal motoneuron. The Journal of physiology *586*, 1059-1075.

Deng, H.-X., Zhai, H., Bigio, E., Yan, J., Fecto, F., Ajroud, K., Mishra, M., Ajroud-Driss, S., Heller, S., Sufit, R.*, et al.* (2010). FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. Annals of neurology *67*, 739-748.

Deng, H.W., Chen, W.M., and Recker, R.R. (2001). Population admixture: detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits. Genetics *157*, 885-897.

Deng, H.X., Chen, W., Hong, S.T., Boycott, K.M., Gorrie, G.H., Siddique, N., Yang, Y., Fecto, F., Shi, Y., Zhai, H.*, et al.* (2011). Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature *477*, 211-215.

Dhaliwal, G.K., and Grewal, R.P. (2000). Mitochondrial DNA deletion mutation levels are elevated in ALS brains. Neuroreport *11*, 2507-2509.

Di Paolo, G., and De Camilli, P. (2006). Phosphoinositides in cell regulation and membrane dynamics. Nature *443*, 651-657.

DiDonato, M., Craig, L., Huff, M.E., Thayer, M.M., Cardoso, R.M.F., Kassmann, C.J., Lo, T.P., Bruns, C.K., Powers, E.T., Kelly, J.W.*, et al.* (2003). ALS Mutants of Human Superoxide Dismutase Form Fibrous Aggregates Via Framework Destabilization. Journal of Molecular Biology *332*, 601-615.

Dols-Icardo, O., Garcia-Redondo, A., Rojas-Garcia, R., Sanchez-Valle, R., Noguera, A., Gomez-Tortosa, E., Pastor, P., Hernandez, I., Esteban-Perez, J., Suarez-Calvet, M.*, et al.* (2014). Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. Hum Mol Genet *23*, 749-754.

Dormann, D., Madl, T., Valori, C.F., Bentmann, E., Tahirovic, S., Abou-Ajram, C., Kremmer, E., Ansorge, O., Mackenzie, I.R., Neumann, M., et al. (2012). Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. EMBO J *31*, 4258-4275.

Dowhan, D.H., Hong, E.P., Auboeuf, D., Dennis, A.P., Wilson, M.M., Berget, S.M., and O'Malley, B.W. (2005). Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERalpha and CAPERbeta. Mol Cell *17*, 429-439.

Dzhala, V., Talos, D., Sdrulla, D., Brumback, A., Mathews, G., Benke, T., Delpire, E., Jensen, F., and Staley, K. (2005). NKCC1 transporter facilitates seizures in the developing brain. Nature medicine *11*, 1205-1213.

Elam, J., Taylor, A., Strange, R., Antonyuk, S., Doucette, P., Rodriguez, J., Hasnain, S., Hayward, L., Valentine, J., Yeates, T.*, et al.* (2003). Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. Nature structural biology *10*, 461-467.

Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M.*, et al.* (2010). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature *466*, 1069-1075.

Elmore, S. (2007). Apoptosis: a review of programmed cell death. Toxicol Pathol *35*, 495-516.

Eymard-Pierre, E., Lesca, G., Dollet, S., Santorelli, F.M., di Capua, M., Bertini, E., and Boespflug-Tanguy, O. (2002). Infantile-Onset Ascending Hereditary Spastic Paralysis Is Associated with Mutations in the Alsin Gene. The American Journal of Human Genetics *71*, 518-527.

Fader, C.M., and Colombo, M.I. (2009). Autophagy and multivesicular bodies: two closely related partners. Cell Death Differ *16*, 70-78.

Farg, M., Soo, K., Warraich, S., Sundaramoorthy, V., Blair, I., and Atkin, J. (2013). Ataxin-2 interacts with FUS and intermediate-length polyglutamine expansions enhance FUS-related pathology in amyotrophic lateral sclerosis. Human molecular genetics *22*, 717-728.

Farg, M.A., Soo, K.Y., Walker, A.K., Pham, H., Orian, J., Horne, M.K., Warraich, S.T., Williams, K.L., Blair, I.P., and Atkin, J.D. (2012). Mutant FUS induces endoplasmic reticulum stress in amyotrophic lateral sclerosis and interacts with protein disulfide-isomerase. Neurobiol Aging *33*, 2855-68.

Fecto, F., and Siddique, T. (2011). Making Connections: Pathology and Genetics Link Amyotrophic Lateral Sclerosis with Frontotemporal Lobe Dementia. Journal of Molecular Neuroscience *45*, 663-675.

Fecto, F., and Siddique, T. (2012). UBQLN2/P62 cellular recycling pathways in amyotrophic lateral sclerosis and frontotemporal dementia. Muscle & nerve *45*, 157-162.

Fecto, F., Yan, J., and Vemula, S. (2011). SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. Arch Neurol *68*, 1440-1446.

Ferguson, C., Lenk, G., and Meisler, M. (2009). Defective autophagy in neurons and astrocytes from mice deficient in PI(3,5)P2. Human molecular genetics *18*, 4868-4878.

Fernandez-Santiago, R., Hoenig, S., Lichtner, P., Sperfeld, A.D., Sharma, M., Berg, D., Weichenrieder, O., Illig, T., Eger, K., Meyer, T.*, et al.* (2009). Identification of novel Angiogenin (ANG) gene missense variants in German patients with amyotrophic lateral sclerosis. J Neurol *256*, 1337-1342.

Fernandez-Santiago, R., Sharma, M., Mueller, J.C., Gohlke, H., Illig, T., Anneser, J., Munch, C., Ludolph, A., Kamm, C., and Gasser, T. (2006). Possible gender-dependent association of vascular endothelial growth factor (VEGF) gene and ALS. Neurology *66*, 1929-1931.

Ferraiuolo, L., Heath, P.R., Holden, H., Kasher, P., Kirby, J., and Shaw, P.J. (2007). Microarray analysis of the cellular pathways involved in the adaptation to and progression of motor neuron injury in the SOD1 G93A mouse model of familial ALS. J Neurosci *27*, 9201-9219.

Ferraiuolo, L., Kirby, J., Grierson, A.J., Sendtner, M., and Shaw, P.J. (2011). Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. Nat Rev Neurol *7*, 616-630.

Figlewicz, D.A., Krizus, A., Martinoli, M.G., Meininger, V., Dib, M., Rouleau, G.A., and Julien, J.-P. (1994). Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. Human molecular genetics *3*, 1757-1761.

Figley, M.D., Thomas, A., and Gitler, A.D. (2014). Evaluating noncoding nucleotide repeat expansions in amyotrophic lateral sclerosis. Neurobiol Aging *35*, 936 e931-934.

Flanagan, S.E., Patch, A.-M., and Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genetic testing and molecular biomarkers *14*, 533-537.

Foran, E., and Trotti, D. (2009). Glutamate transporters and the excitotoxic path to motor neuron degeneration in amyotrophic lateral sclerosis. Antioxidants & redox signaling *11*, 1587-1602.

Fornai, F., Longone, P., Cafaro, L., Kastsiuchenka, O., Ferrucci, M., Manca, M., Lazzeri, G., Spalloni, A., Bellio, N., Lenzi, P.*, et al.* (2008). Lithium delays progression of amyotrophic lateral sclerosis. Proceedings of the National Academy of Sciences of the United States of America *105*, 2052-2057.

Fratta, P., Poulter, M., Lashley, T., Rohrer, J., Polke, J., Beck, J., Ryan, N., Hensman, D., Mizielinska, S., Waite, A.*, et al.* (2013). Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. Acta neuropathologica *126*, 401-409.

Fujita, K., Yamauchi, M., Shibayama, K., Ando, M., Honda, M., and Nagata, Y. (1996). Decreased cytochrome c oxidase activity but unchanged superoxide dismutase and

glutathione peroxidase activities in the spinal cords of patients with amyotrophic lateral sclerosis. J Neurosci Res *45*, 276-281.

Furukawa, Y., Fu, R., Deng, H.X., Siddique, T., and O'Halloran, T.V. (2006). Disulfide cross-linked protein represents a significant fraction of ALS-associated Cu, Zn-superoxide dismutase aggregates in spinal cords of model mice. Proc Natl Acad Sci U S A *103*, 7148-7153.

Gal, J., Zhang, J., Kwinter, D., Zhai, J., Jia, H., Jia, J., and Zhu, H. (2011). Nuclear localization sequence of FUS and induction of stress granules by ALS mutants. Neurobiology of aging *32*, 40.

Galligan, J., and Petersen, D. (2012). The human protein disulfide isomerase gene family. Human genomics *6*, 6.

Gao, X., and Xu, Z. (2008). Mechanisms of action of angiogenin. Acta biochimica et biophysica Sinica *40*, 619-624.

Garcia-Segura, L.M., Azcoitia, I., and DonCarlos, L.L. (2001). Neuroprotection by estradiol. Prog Neurobiol *63*, 29-60.

Garcia, M.A., Gil, J., Ventoso, I., Guerra, S., Domingo, E., Rivas, C., and Esteban, M. (2006). Impact of protein kinase PKR in cell biology: from antiviral to antiproliferative action. Microbiology and molecular biology reviews : MMBR *70*, 1032-1060.

Ge, F., Li, W.-L., Bi, L.-J., Tao, S.-C., Zhang, Z.-P., and Zhang, X.-E. (2010). Identification of novel 14-3-3ζ interacting proteins by quantitative immunoprecipitation combined with knockdown (QUICK). Journal of proteome research *9*, 5848-5858.

Ge, W.-W., Volkening, K., Leystra-Lantz, C., Jaffe, H., and Strong, M. (2007). 14-3-3 protein binds to the low molecular weight neurofilament (NFL) mRNA 3' UTR. Molecular and cellular neurosciences *34*, 80-87.

Geetha, T., and Wooten, M.W. (2002). Structure and functional properties of the ubiquitin binding protein p62. FEBS Lett *512*, 19-24.

Gendron, T.F., Josephs, K.A., and Petrucelli, L. (2010). Review: transactive response DNA-binding protein 43 (TDP-43): mechanisms of neurodegeneration. Neuropathol Appl Neurobiol *36*, 97-112.

Geschwind, D., Perlman, S., Figueroa, C., Treiman, L., and Pulst, S. (1997). The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients with autosomal dominant cerebellar ataxia. American journal of human genetics *60*, 842-850.

Giordana, M., Ferrero, P., Grifoni, S., Pellerino, A., Naldi, A., and Montuschi, A. (2011). Dementia and cognitive impairment in amyotrophic lateral sclerosis: a review.

Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology *32*, 9-16.

Gitcho, M.A., Strider, J., Carter, D., Taylor-Reinwald, L., Forman, M.S., Goate, A.M., and Cairns, N.J. (2009). VCP mutations causing frontotemporal lobar degeneration disrupt localization of TDP-43 and induce cell death. J Biol Chem *284*, 12384-12398.

Glickman, M., and Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. Physiological reviews *82*, 373-428.

Goldberg, A. (2003). Protein degradation and protection against misfolded or damaged proteins. Nature *426*, 895-899.

Goldberg, Y.P., Andrew, S.E., Clarke, L.A., and Hayden, M.R. (1993). A PCR method for accurate assessment of trinucleotide repeat expansion in Huntington disease. Hum Mol Genet *2*, 635-636.

Gomez-Skarmeta, J.L., and Modolell, J. (2002). Iroquois genes: genomic organization and function in vertebrate neural development. Current opinion in genetics & development *12*, 403-408.

Gonzalez-Perez, P., Cirulli, E.T., Drory, V.E., Dabby, R., Nisipeanu, P., Carasso, R.L., Sadeh, M., Fox, A., Festoff, B.W., Sapp, P.C.*, et al.* (2012). Novel mutation in VCP gene causes atypical amyotrophic lateral sclerosis. Neurology *79*, 2001-2008.

Goytain, A., Hines, R.M., El-Husseini, A., and Quamme, G.A. (2007). NIPA1(SPG6), the basis for autosomal dominant form of hereditary spastic paraplegia, encodes a functional Mg2+ transporter. J Biol Chem *282*, 8060-8068.

Greenway, M., Alexander, M., Ennis, S., Traynor, B., Corr, B., Frost, E., Green, A., and Hardiman, O. (2004). A novel candidate region for ALS on chromosome 14q11.2. Neurology *63*, 1936-1938.

Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J.*, et al.* (2006). ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. Nat Genet *38*, 411-413.

Gyorffy, B., Kocsis, I., and Vasarhelyi, B. (2004). Missed calculations and new conclusions: re-calculation of genotype distribution data published in Journal of Investigative Dermatology, 1998-2003. J Invest Dermatol *122*, 644-646.

Hadano, S., Hand, C., Osuga, H., Yanagisawa, Y., Otomo, A., Devon, R., Miyamoto, N., Showguchi-Miyata, J., Okada, Y., Singaraja, R.*, et al.* (2001). A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. Nature Genetics *29*, 166-173.

Haeusler, A., Donnelly, C., Periz, G., Simko, E., Shaw, P., Kim, M.-S., Maragakis, N., Troncoso, J., Pandey, A., Sattler, R.*, et al.* (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195-200.

Halawani, D., and Latterich, M. (2006). p97: The cell's molecular purgatory? Mol Cell *22*, 713-717.

Hama, T., Maruyama, M., Katoh-Semba, R., Takizawa, M., Iwashima, M., and Nara, K. (2001). Identification and molecular cloning of a novel brain-specific receptor protein that binds to brain injury-derived neurotrophic peptide. Possible role for neuronal survival. J Biol Chem *276*, 31929-31935.

Hamida, M., Hentati, F., and Hamida, C. (1990). HEREDITARY MOTOR SYSTEM DISEASES (CHRONIC JUVENILE AMYOTROPHIC LATERAL SCLEROSIS): CONDITIONS COMBINING A BILATERAL PYRAMIDAL SYNDROME WITH LIMB AND BULBAR AMYOTROPHY. Brain : a journal of neurology.

Hand, C.K., Khoris, J., Salachas, F., Gros-Louis, F., Lopes, A.A., Mayeux-Portas, V., Brewer, C.G., Brown, R.H., Jr., Meininger, V., Camu, W.*, et al.* (2002). A novel locus for familial amyotrophic lateral sclerosis, on chromosome 18q. Am J Hum Genet *70*, 251-256.

Hara, T., Nakamura, K., Matsui, M., Yamamoto, A., Nakahara, Y., Suzuki-Migishima, R., Yokoyama, M., Mishima, K., Saito, I., Okano, H.*, et al.* (2006). Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. Nature *441*, 885-889.

Hardy, J., and Orr, H. (2006). The genetics of neurodegenerative diseases. Journal of neurochemistry *97*, 1690-1699.

Harms, M., Benitez, B.A., Cairns, N., Cooper, B., Cooper, P., Mayo, K., Carrell, D., Faber, K., Williamson, J., Bird, T.*, et al.* (2013). C9orf72 hexanucleotide repeat expansions in clinical Alzheimer disease. JAMA neurology *70*, 736-741.

Haverkamp, L.J., Appel, V., and Appel, S.H. (1995). Natural history of amyotrophic lateral sclerosis in a database population. Validation of a scoring system and a model for survival prediction. Brain : a journal of neurology *118 (Pt 3)*, 707-719.

Hayashi, T., and Su, T.-P. (2007). Sigma-1 receptor chaperones at the ER-mitochondrion interface regulate Ca(2+) signaling and cell survival. Cell *131*, 596-610.

Hays, A.P., Naini, A., He, C.Z., Mitsumoto, H., and Rowland, L.P. (2006). Sporadic amyotrophic lateral sclerosis and breast cancer: Hyaline conglomerate inclusions lead to identification of SOD1 mutation. J Neurol Sci *242*, 67-69.

Hayward, C., Colville, S., Swingler, R.J., and Brock, D.J. (1999). Molecular genetic analysis of the APEX nuclease gene in amyotrophic lateral sclerosis. Neurology *52*, 1899-1901.

Hebert, S., Mount, D., and Gamba, G. (2004). Molecular physiology of cation-coupled Cl-cotransport: the SLC12 family. Pflügers Archiv : European journal of physiology *447*, 580-593.

Heiman-Patterson, T.D., Deitch, J.S., Blankenhorn, E.P., Erwin, K.L., Perreault, M.J., Alexander, B.K., Byers, N., Toman, I., and Alexander, G.M. (2005). Background and gender effects on survival in the TgN(SOD1-G93A)1Gur mouse model of ALS. J Neurol Sci *236*, 1-7.

Hentati, A., Ouahchi, K., Pericak-Vance, M., Nijhawan, D., Ahmad, A., Yang, Y., Rimmler, J., Hung, W., Schlotter, B., Ahmed, A.*, et al.* (1998). Linkage of a commoner form of recessive amyotrophic lateral sclerosis to chromosome 15q15-q22 markers. Neurogenetics *2*, 55-60.

Hetz, C., and Mollereau, B. (2014). Disturbance of endoplasmic reticulum proteostasis in neurodegenerative diseases. Nat Rev Neurosci *15*, 233-49.

Hetz, C., Thielen, P., Matus, S., Nassif, M., Court, F., Kiffin, R., Martinez, G., Cuervo, A., Brown, R., and Glimcher, L. (2009). XBP-1 deficiency in the nervous system protects against amyotrophic lateral sclerosis by increasing autophagy. Genes & development *23*, 2294-2306.

Hirano, A., Nakano, I., Kurland, L.T., Mulder, D.W., Holley, P.W., and Saccomanno, G. (1984). Fine structural study of neurofibrillary changes in a family with amyotrophic lateral sclerosis. J Neuropathol Exp Neurol *43*, 471-480.

Hirano, M., Nakamura, Y., Saigoh, K., Sakamoto, H., Ueno, S., Isono, C., Miyamoto, K., Akamatsu, M., Mitsui, Y., and Kusunoki, S. (2013). Mutations in the gene encoding p62 in Japanese patients with amyotrophic lateral sclerosis. Neurology *80*, 458-463.

Hirschhorn, J., and Daly, M. (2005). Genome-wide association studies for common diseases and complex traits. Nature reviews Genetics *6*, 95-108.

Hirschhorn, J.N., Lohmueller, K., Byrne, E., and Hirschhorn, K. (2002). A comprehensive review of genetic association studies. Genet Med *4*, 45-61.

Hocking, L.J., Lucas, G.J., Daroszewska, A., Cundy, T., Nicholson, G.C., Donath, J., Walsh, J.P., Finlayson, C., Cavey, J.R., Ciani, B.*, et al.* (2004). Novel UBA domain mutations of SQSTM1 in Paget's disease of bone: genotype phenotype correlation, functional analysis, and structural consequences. J Bone Miner Res *19*, 1122-1127.

Holden, S., and Raymond, F.L. (2003). The human gene CXorf17 encodes a member of a novel family of putative transmembrane proteins: cDNA cloning and characterization of CXorf17 and its mouse ortholog orf34. Gene *318*, 149-161.

Honjo, Y., Kaneko, S., Ito, H., Horibe, T., Nagashima, M., Nakamura, M., Fujita, K., Takahashi, R., Kusaka, H., and Kawakami, K. (2011). Protein disulfide isomerase-immunopositive inclusions in patients with amyotrophic lateral sclerosis. Amyotroph Lateral Scler.

Hosler, B.A., Siddique, T., Sapp, P.C., Sailor, W., Huang, M.C., Hossain, A., Daube, J.R., Nance, M., Fan, C., Kaplan, J.*, et al.* (2000). Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22. JAMA *284*, 1664-1669.

Hu, M.T., Ellis, C.M., Al-Chalabi, A., Leigh, P.N., and Shaw, C.E. (1998). Flail arm syndrome: a distinctive variant of amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry *65*, 950-951.

Hu, Y., Dolan, M.E., Bae, R., Yee, H., Roy, M., Glickman, R., Kiremidjian-Schumacher, L., and Diamond, A. (2004). Allelic loss at the GPx-1 locus in cancer of the head and neck. Biol Trace Elem Res *101*, 97-106.

Hu, Y.J., and Diamond, A.M. (2003). Role of glutathione peroxidase 1 in breast cancer: loss of heterozygosity and allelic differences in the response to selenium. Cancer Res *63*, 3347-3351.

Hu, Z., Chen, L., Zhang, J., Li, T., Tang, J., Xu, N., and Wang, X. (2007). Structure, function, property, and role in neurologic diseases and other diseases of the sHsp22. Journal of neuroscience research *85*, 2071-2079.

Huang, C., Zhou, H., Tong, J., Chen, H., Liu, Y.-J., Wang, D., Wei, X., and Xia, X.-G. (2011). FUS transgenic rats develop the phenotypes of amyotrophic lateral sclerosis and frontotemporal lobar degeneration. PLoS genetics *7*.

Hubers, A., Marroquin, N., Schmoll, B., Vielhaber, S., Just, M., Mayer, B., Hogel, J., Dorst, J., Mertens, T., Just, W.*, et al.* (2014). Polymerase chain reaction and Southern blot-based analysis of the C9orf72 hexanucleotide repeat in different motor neuron diseases. Neurobiol Aging *35*, 1214 e1211-1216.

Hudson, A. (1981). Amyotrophic lateral sclerosis and its association with dementia, parkinsonism and other neurological disorders: a review. Brain : a journal of neurology *104*, 217-247.

Hurley, J., and Emr, S. (2006). The ESCRT complexes: structure and mechanism of a membrane-trafficking network. Annual review of biophysics and biomolecular structure *35*, 277-298.

Huynh, D.P., Yang, H.T., Vakharia, H., Nguyen, D., and Pulst, S.M. (2003). Expansion of the polyQ repeat in ataxin-2 alters its Golgi localization, disrupts the Golgi complex and causes cell death. Hum Mol Genet *12*, 1485-1496.

Iavarone, A., and Lasorella, A. (2006). ID proteins as targets in cancer and tools in neurobiology. Trends in Molecular Medicine *12*, 588-594.

Ilieva, E.V., Ayala, V., Jove, M., Dalfo, E., Cacabelos, D., Povedano, M., Bellmunt, M.J., Ferrer, I., Pamplona, R., and Portero-Otin, M. (2007). Oxidative and endoplasmic reticulum stress interplay in sporadic amyotrophic lateral sclerosis. Brain : a journal of neurology *130*, 3111-3123.

Imbert, G., Saudou, F., Yvert, G., Devys, D., Trottier, Y., Garnier, J., Weber, C., Mandel, J., Cancel, G., Abbas, N.*, et al.* (1996). Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. Nature genetics *14*, 285-291.

Irobi, J., Almeida-Souza, L., Asselbergh, B., De Winter, V., Goethals, S., Dierick, I., Krishnan, J., Timmermans, J.-P., Robberecht, W., De Jonghe, P.*, et al.* (2010). Mutant HSPB8 causes motor neuron-specific neurite degeneration. Human molecular genetics *19*, 3254-3265.

Israely, I., Costa, R.M., Xie, C.W., Silva, A.J., Kosik, K.S., and Liu, X. (2004). Deletion of the neuron-specific protein delta-catenin leads to severe cognitive and synaptic dysfunction. Current biology : CB *14*, 1657-1663.

Ito, D., Seki, M., Tsunoda, Y., Uchiyama, H., and Suzuki, N. (2011a). Nuclear transport impairment of amyotrophic lateral sclerosis-linked mutations in FUS/TLS. Annals of neurology *69*, 152-162.

Ito, D., and Suzuki, N. (2011). Conjoint pathologic cascades mediated by ALS/FTLD-U linked RNA-binding proteins TDP-43 and FUS. Neurology *77*, 1636-1643.

Ito, H., Fujita, K., Nakamura, M., Wate, R., Kaneko, S., Sasaki, S., Yamane, K., Suzuki, N., Aoki, M., Shibata, N.*, et al.* (2011b). Optineurin is co-localized with FUS in basophilic inclusions of ALS with FUS mutation and in basophilic inclusion body disease. Acta Neuropathologica *121*, 555-557.

Ito, Y., Yamada, M., Tanaka, H., Aida, K., Tsuruma, K., Shimazawa, M., Hozumi, I., Inuzuka, T., Takahashi, H., and Hara, H. (2009). Involvement of CHOP, an ER-stress apoptotic mediator, in both human sporadic ALS and ALS model mice. Neurobiology of Disease *36*, 470-476.

Johnson, J.O., Pioro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J., et al. (2014). Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. Nat Neurosci *17*, 664-666.

Johnson, J.O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V.M., Trojanowski, J.Q., Gibbs, J.R., Brunetti, M., Gronka, S., Wuu, J.*, et al.* (2010). Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS. Neuron *68*, 857-864.

Johnston, J.A., Dalton, M.J., Gurney, M.E., and Kopito, R.R. (2000). Formation of high molecular weight complexes of mutant Cu, Zn-superoxide dismutase in a mouse model for familial amyotrophic lateral sclerosis. Proc Natl Acad Sci U S A *97*, 12571-12576.

Jones, A., Woollacott, I., Shatunov, A., Cooper-Knock, J., Buchman, V., Sproviero, W., Smith, B., Scott, K., Balendra, R., Abel, O.*, et al.* (2013). Residual association at C9orf72 suggests an alternative amyotrophic lateral sclerosis-causing hexanucleotide repeat. Neurobiology of aging *34*, 22340-22347.

Ju, J.-S., Fuentealba, R., Miller, S., Jackson, E., Piwnica-Worms, D., Baloh, R., and Weihl, C. (2009). Valosin-containing protein (VCP) is required for autophagy and is disrupted in VCP disease. The Journal of Cell Biology *187*, 875-888.

Ju, J.S., and Weihl, C.C. (2010). Inclusion body myopathy, Paget's disease of the bone and fronto-temporal dementia: a disorder of autophagy. Hum Mol Genet *19*, R38-45.

Jung, C.H., Ro, S.H., Cao, J., Otto, N.M., and Kim, D.H. (2010). mTOR regulation of autophagy. FEBS Lett *584*, 1287-1295.

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F.*, et al.* (2008). TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet *40*, 572-574.

Kabuta, T., Suzuki, Y., and Wada, K. (2006). Degradation of amyotrophic lateral sclerosis-linked mutant Cu,Zn-superoxide dismutase proteins by macroautophagy and the proteasome. The Journal of biological chemistry *281*, 30524-30533.

Kadowaki, H., Nishitoh, H., and Ichijo, H. (2004). Survival and apoptosis signals in ER stress: the role of protein kinases. J Chem Neuroanat *28*, 93-100.

Kanekura, K. (2006). Characterization of Amyotrophic Lateral Sclerosis-linked P56S Mutation of Vesicle-associated Membrane Protein-associated Protein B (VAPB/ALS8). Journal of Biological Chemistry *281*, 30223-30233.

Kaushik, N. (2000). Characterization of trinucleotide-and tandem repeat-containing transcripts obtained from human spinal cord cDNA library by high-density filter hybridization. DNA and cell biology *19*, 265.

Kaushik, N., Malaspina, A., Schalling, M., Baas, F., and de Belleroche, J. (1998). Isolation and characterization of trinucleotide repeat containing partial transcripts in human spinal cord. Neurogenetics *1*, 239-247.

Kawamoto, Y., Akiguchi, I., Fujimura, H., Shirakashi, Y., Honjo, Y., and Sakoda, S. (2005). 14-3-3 proteins in Lewy body-like hyaline inclusions in a patient with familial amyotrophic lateral sclerosis with a two-base pair deletion in the Cu/Zn superoxide dismutase (SOD1) gene. Acta Neuropathol *110*, 203-204.

Kawamoto, Y., Akiguchi, I., Nakamura, S., and Budka, H. (2004). 14-3-3 proteins in Lewy body-like hyaline inclusions in patients with sporadic amyotrophic lateral sclerosis. Acta Neuropathol *108*, 531-537.

Keller, B.A., Volkening, K., Droppelmann, C.A., Ang, L.C., Rademakers, R., and Strong, M.J. (2012). Co-aggregation of RNA binding proteins in ALS spinal motor neurons: evidence of a common pathogenic mechanism. Acta Neuropathol *124*, 733-747.

Khare, S., Wilcox, K., Gong, P., and Dokholyan, N. (2005). Sequence and structural determinants of Cu, Zn superoxide dismutase aggregation. Proteins *61*, 617-632.

Kim, H.J., Kim, N.C., Wang, Y.D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A.*, et al.* (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*, 467-473.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet *46*, 310-315.

Kishimoto, K., Liu, S., Tsuji, T., Olson, K., and Hu, G.-F. (2005). Endogenous angiogenin in endothelial cells is a general requirement for cell proliferation and angiogenesis. Oncogene *24*, 445-456.

Knoblach, B., Keller, B.O., Groenendyk, J., Aldred, S., Zheng, J., Lemire, B.D., Li, L., and Michalak, M. (2003). ERp19 and ERp46, new members of the thioredoxin family of endoplasmic reticulum proteins. Molecular & cellular proteomics : MCP *2*, 1104-1119.

Kobayashi, T., Manno, A., and Kakizuka, A. (2007). Involvement of valosin-containing protein (VCP)/p97 in the formation and clearance of abnormal protein aggregates. Genes to Cells *12*, 889-901.

Kong, J., and Xu, Z. (2000). Overexpression of neurofilament subunit NF-L and NF-H extends survival of a mouse model for amyotrophic lateral sclerosis. Neurosci Lett *281*, 72-74.

Koppers, M., van Blitterswijk, M.M., Vlam, L., Rowicka, P.A., van Vught, P.W., Groen, E.J., Spliet, W.G., Engelen-Lee, J., Schelhaas, H.J., de Visser, M.*, et al.* (2011). VCP mutations in familial and sporadic amyotrophic lateral sclerosis. Neurobiol Aging *33*, 837.e7-13.

Korac, J., Schaeffer, V., Kovacevic, I., Clement, A., Jungblut, B., Behl, C., Terzic, J., and Dikic, I. (2013). Ubiquitin-independent function of optineurin in autophagic clearance of protein aggregates. Journal of cell science *126*, 580-592.

Kostic, V., Jackson-Lewis, V., de Bilbao, F., Dubois-Dauphin, M., and Przedborski, S. (1997). Bcl-2: prolonging life in a transgenic mouse model of familial amyotrophic lateral sclerosis. Science *277*, 559-562.

Kozlowski, P., de Mezer, M., and Krzyzosiak, W.J. (2010). Trinucleotide repeats in human genome and exome. Nucleic Acids Res *38*, 4027-4039.

Kwok, C.T., Morris, A.G., Frampton, J., Smith, B., Shaw, C.E., and de Belleroche, J. (2013). Association studies indicate that protein disulfide isomerase is a risk factor in amyotrophic lateral sclerosis. Free Radical Biology and Medicine *58*, 81-86.

Laaksovirta, H., Peuralinna, T., Schymick, J.C., Scholz, S.W., Lai, S.L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D.G., Gibbs, J.R.*, et al.* (2010). Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol *9*, 978-985.

Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum Mol Genet *19*, R46-64.

Lai, C., Xie, C., McCormack, S., Chiang, H.-C., Michalak, M., Lin, X., Chandran, J., Shim, H., Shimoji, M., Cookson, M.*, et al.* (2006). Amyotrophic lateral sclerosis 2-deficiency leads to neuronal degeneration in amyotrophic lateral sclerosis through altered AMPA receptor trafficking. The Journal of neuroscience : the official journal of the Society for Neuroscience *26*, 11798-11806.

Lamark, T., and Johansen, T. (2010). Autophagy: links with the proteasome. Current Opinion in Cell Biology *22*, 192-198.

Lambrechts, D., Storkebaum, E., Morimoto, M., Del-Favero, J., Desmet, F., Marklund, S., Wyns, S., Thijs, V., Andersson, J., van Marion, I.*, et al.* (2003). VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. Nature genetics *34*, 383-394.

Landers, J.E., Leclerc, A.L., Shi, L., Virkud, A., Cho, T., Maxwell, M.M., Henry, A.F., Polak, M., Glass, J.D., Kwiatkowski, T.J.*, et al.* (2008). New VAPB deletion variant and exclusion of VAPB mutations in familial ALS. Neurology *70*, 1179-1185.

Lass, A., McConnell, E., Fleck, K., Palamarchuk, A., and Wójcik, C. (2008). Analysis of Npl4 deletion mutants in mammalian cells unravels new Ufd1-interacting motifs and suggests a regulatory role of Npl4 in ERAD. Experimental cell research *314*, 2715-2723.

Laurindo, F.R., Pescatore, L.A., and de Castro Fernandes, D. (2012). Protein disulfide isomerase in redox cell signaling and homeostasis. Free Radic Biol Med *52*, 1954-1969.

Leblond, C., Kaneb, H., Dion, P., and Rouleau, G. (2014). Dissection of genetic factors associated with amyotrophic lateral sclerosis. Experimental neurology *S0014-4886*, 00115-0.

Lee, E.S., Yoon, C.H., Kim, Y.S., and Bae, Y.S. (2007a). The double-strand RNA-dependent protein kinase PKR plays a significant role in a sustained ER stress-induced apoptosis. FEBS Lett *581*, 4325-4332.

Lee, J.-A., Beigneux, A., Ahmad, S., Young, S., and Gao, F.-B. (2007b). ESCRT-III dysfunction causes autophagosome accumulation and neurodegeneration. Current biology : CB *17*, 1561-1567.

Lenk, G., Ferguson, C., Chow, C., Jin, N., Jones, J., Grant, A., Zolov, S., Winters, J., Giger, R., Dowling, J.*, et al.* (2011). Pathogenic mechanism of the FIG4 mutation responsible for Charcot-Marie-Tooth disease CMT4J. PLoS genetics *7*.

Lesage, S., Le Ber, I., Condroyer, C., Broussolle, E., Gabelle, A., Thobois, S., Pasquier, F., Mondon, K., Dion, P.A., Rochefort, D.*, et al.* (2013). C9orf72 repeat expansions are a rare genetic cause of parkinsonism. Brain : a journal of neurology *136*, 385-391.

Levine, B., and Kroemer, G. (2008). Autophagy in the pathogenesis of disease. Cell *132*, 27-42.

Levy, J.R., Sumner, C.J., Caviston, J.P., Tokito, M.K., Ranganathan, S., Ligon, L.A., Wallace, K.E., LaMonte, B.H., Harmison, G.G., Puls, I.*, et al.* (2006). A motor neuron disease-associated mutation in p150Glued perturbs dynactin function and induces protein aggregation. J Cell Biol *172*, 733-745.

Lewis, C.M. (2002). Genetic association studies: design, analysis and interpretation. Briefings in bioinformatics *3*, 146-153.

Lewontin, R.C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. Genetics *49*, 49-67.

Li, M., Ona, V.O., Guegan, C., Chen, M., Jackson-Lewis, V., Andrews, L.J., Olszewski, A.J., Stieg, P.E., Lee, J.P., Przedborski, S.*, et al.* (2000). Functional role of caspase-1 and caspase-3 in an ALS transgenic mouse model. Science *288*, 335-339.

Li, Y., King, O., Shorter, J., and Gitler, A. (2013). Stress granules as crucibles of ALS pathogenesis. The Journal of cell biology *201*, 361-372.

Lillo, P., and Hodges, J. (2009). Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia *16*, 1131-1135.

Lin, C., Bristol, L., Jin, L., Dykes-Hoberg, M., Crawford, T., Clawson, L., and Rothstein, J. (1998). Aberrant RNA processing in a neurodegenerative disease: the cause for absent EAAT2, a glutamate transporter, in amyotrophic lateral sclerosis. Neuron *20*, 589-602.

Ling, S.-C., Polymenidou, M., and Cleveland, D. (2013). Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron *79*, 416-438.

Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*, 864-867.

Liu-Yesucevitz, L., Bilgutay, A., Zhang, Y.-J., Vanderweyde, T., Vanderwyde, T., Citro, A., Mehta, T., Zaarur, N., McKee, A., Bowser, R.*, et al.* (2010). Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. PloS one *5*.

Lucas, G.J., Hocking, L.J., Daroszewska, A., Cundy, T., Nicholson, G.C., Walsh, J.P., Fraser, W.D., Meier, C., Hooper, M.J., and Ralston, S.H. (2005). Ubiquitin-associated domain mutations of SQSTM1 in Paget's disease of bone: evidence for a founder effect in patients of British descent. J Bone Miner Res *20*, 227-231.

Lunde, B., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. Nature reviews Molecular cell biology *8*, 479-490.

Lundström, J., and Holmgren, A. (1990). Protein disulfide-isomerase is a substrate for thioredoxin reductase and has thioredoxin-like activity. Journal of Biological Chemistry *265*, 9114-9120.

Luty, A.A., Kwok, J.B., Dobson-Stone, C., Loy, C.T., Coupland, K.G., Karlstrom, H., Sobow, T., Tchorzewska, J., Maruszak, A., Barcikowska, M.*, et al.* (2010). Sigma nonopioid intracellular receptor 1 mutations cause frontotemporal lobar degeneration-motor neuron disease. Ann Neurol *68*, 639-649.

Mackenzie, I., Frick, P., and Neumann, M. (2014). The neuropathology associated with repeat expansions in the C9ORF72 gene. Acta neuropathologica *127*, 347-357.

Mackenzie, I., Rademakers, R., and Neumann, M. (2010). TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. Lancet neurology *9*, 995-1007.

Mackenzie, I.R., Bigio, E.H., Ince, P.G., Geser, F., Neumann, M., Cairns, N.J., Kwong, L.K., Forman, M.S., Ravits, J., Stewart, H.*, et al.* (2007). Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann Neurol *61*, 427-434.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. Nat Rev Neurosci *8*, 755-765.

Malaspina, A., Kaushik, N., and Belleroche, J.d. (2000). A 14-3-3 mRNA Is Up-Regulated in Amyotrophic Lateral Sclerosis Spinal Cord. Journal of Neurochemistry *75*, 2511-2520.

Malaspina, A., Kaushik, N., and de Belleroche, J. (2001). Differential expression of 14 genes in amyotrophic lateral sclerosis spinal cord detected using gridded cDNA arrays. Journal of neurochemistry *77*, 132-145.

Malhotra, J., and Kaufman, R. (2007). The endoplasmic reticulum and the unfolded protein response. Seminars in cell & developmental biology *18*, 716-731.

Manfredi, G., and Xu, Z. (2005). Mitochondrial dysfunction and its role in motor neuron degeneration in ALS. Mitochondrion *5*, 77-87.

Mantha, A.K., Sarkar, B., and Tell, G. (2013). A short review on the implications of base excision repair pathway for neurons: Relevance to neurodegenerative diseases. Mitochondrion.

Mao, W.-G., He, H.-Q., Xu, Y., Chen, P.-Y., and Zhou, J.-Y. (2013). Powerful Haplotype-Based Hardy-Weinberg Equilibrium Tests for Tightly Linked Loci. PLoS One *8*, e77399.

Marden, J.J., Harraz, M.M., Williams, A.J., Nelson, K., Luo, M., Paulson, H., and Engelhardt, J.F. (2007). Redox modifier genes in amyotrophic lateral sclerosis in mice. The Journal of clinical investigation *117*, 2913-2919.

Martinez-Lage, M., Molina-Porcel, L., Falcone, D., McCluskey, L., Lee, V.Y., Deerlin, V., and Trojanowski, J. (2012). TDP-43 pathology in a case of hereditary spastic paraplegia with a NIPA1/SPG6 mutation. Acta Neuropathologica *124*, 285-291.

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H.*, et al.* (2010). Mutations of optineurin in amyotrophic lateral sclerosis. Nature *465*, 223-226.

Massignan, T., Casoni, F., Basso, M., Stefanazzi, P., Biasini, E., Tortarolo, M., Salmona, M., Gianazza, E., Bendotti, C., and Bonetto, V. (2007). Proteomic analysis of spinal cord of presymptomatic amyotrophic lateral sclerosis G93A SOD1 mouse. Biochem Biophys Res Commun *353*, 719-725.

Matsuura, T., Yamagata, T., Burgess, D.L., Rasmussen, A., Grewal, R.P., Watase, K., Khajavi, M., McCall, A.E., Davis, C.F., Zu, L.*, et al.* (2000). Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet *26*, 191-194.

Mattiazzi, M., D'Aurelio, M., Gajewski, C.D., Martushova, K., Kiaei, M., Beal, M.F., and Manfredi, G. (2002). Mutated human SOD1 causes dysfunction of oxidative phosphorylation in mitochondria of transgenic mice. J Biol Chem *277*, 29626-29633.

Mavlyutov, T., Epstein, M., Andersen, K., Ziskind-Conhaim, L., and Ruoho, A. (2010). The sigma-1 receptor is enriched in postsynaptic sites of C-terminals in mouse motoneurons. An anatomical and behavioral study. Neuroscience *167*, 247-255.

McCombe, P., and Henderson, R. (2010). Effects of gender in amyotrophic lateral sclerosis. Gender medicine *7*, 557-570.

McConkey, D., and Orrenius, S. (1997). The role of calcium in the regulation of apoptosis. Biochem Biophys Res Commun *239*, 357-366.

McManus, S., and Roux, S. (2012). The adaptor protein p62/SQSTM1 in osteoclast signaling pathways. Journal of molecular signaling *7*, 1.

Metzker, M.L. (2010). Sequencing technologies - the next generation. Nat Rev Genet *11*, 31-46.

Meyer, H., Wang, Y., and Warren, G. (2002). Direct binding of ubiquitin conjugates by the mammalian p97 adaptor complexes, p47 and Ufd1-Npl4. The EMBO journal *21*, 5645-5652.

Michell, R., Heath, V., Lemmon, M., and Dove, S. (2006). Phosphatidylinositol 3,5-bisphosphate: metabolism and cellular functions. Trends in biochemical sciences *31*, 52-63.

Migheli, A., Atzori, C., Piva, R., Tortarolo, M., Girelli, M., Schiffer, D., and Bendotti, C. (1999). Lack of apoptosis in mice with ALS. Nature medicine *5*, 966-967.

Millecamps, S., Robertson, J., Lariviere, R., Mallet, J., and Julien, J.P. (2006). Defective axonal transport of neurofilament proteins in neurons overexpressing peripherin. J Neurochem *98*, 926-938.

Miller, J.W., Smith, B.N., Topp, S.D., Al-Chalabi, A., Shaw, C.E., and Vance, C. (2012). Mutation analysis of VCP in British familial and sporadic amyotrophic lateral sclerosis patients. Neurobiol Aging.

Mills, G.C. (1957). HEMOGLOBIN CATABOLISM: I. GLUTATHIONE PEROXIDASE, AN ERYTHROCYTE ENZYME WHICH PROTECTS HEMOGLOBIN FROM OXIDATIVE BREAKDOWN. Journal of Biological Chemistry *229*, 189-197.

Mitchell, J., Morris, A., and Debelleroche, J. (2009). Thioredoxin reductase 1 haplotypes modify familial amyotrophic lateral sclerosis onset. Free Radical Biology and Medicine *46*, 202-211.

Mitchell, J., Paul, P., Chen, H.J., Morris, A., Payling, M., Falchi, M., Habgood, J., Panoutsou, S., Winkler, S., Tisato, V.*, et al.* (2010). Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. Proceedings of the National Academy of Sciences *107*, 7556-7561.

Mitsumoto, H. (1997). Diagnosis and progression of ALS. Neurology *48*.

Mizuno, Y., Hori, S., Kakizuka, A., and Okamoto, K. (2003). Vacuole-creating protein in neurodegenerative diseases in humans. Neuroscience Letters *343*, 77-80.

Mizushima, N. (2007). Autophagy: process and function. Genes Dev *21*, 2861-2873.

Mok, K., Traynor, B.J., Schymick, J., Tienari, P.J., Laaksovirta, H., Peuralinna, T., Myllykangas, L., Chiò, A., Shatunov, A., Boeve, B.F.*, et al.* (2012). The chromosome 9 ALS and FTD locus is probably derived from a single founder. Neurobiology of Aging *33*, 209.e203-209.e208.

Moreira, M.C., Klur, S., Watanabe, M., Nemeth, A.H., Le Ber, I., Moniz, J.C., Tranchant, C., Aubourg, P., Tazir, M., Schols, L.*, et al.* (2004). Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. Nat Genet *36*, 225-227.

Mori, K., Weng, S.-M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C.*, et al.* (2013a). The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. Science *339*, 1335-1338.

Morissette, J., Laurin, N., and Brown, J.P. (2006). Sequestosome 1: Mutation Frequencies, Haplotypes, and Phenotypes in Familial Paget's Disease of Bone. Journal of bone and mineral research *21*, P38-P44.

Morita, M., Al-Chalabi, A., Andersen, P.M., Hosler, B., Sapp, P., Englund, E., Mitchell, J.E., Habgood, J.J., de Belleroche, J., Xi, J.*, et al.* (2006). A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology *66*, 839-844.

Moscat, J., Diaz-Meco, M.T., and Wooten, M.W. (2007). Signal integration and diversification through the p62 scaffold protein. Trends in biochemical sciences *32*, 95-100.

Moscow, J.A., Schmidt, L., Ingram, D.T., Gnarra, J., Johnson, B., and Cowan, K.H. (1994). Loss of heterozygosity of the human cytosolic glutathione peroxidase I gene in lung cancer. Carcinogenesis *15*, 2769-2773.

Muller, J.M., Meyer, H.H., Ruhrberg, C., Stamp, G.W., Warren, G., and Shima, D.T. (1999). The mouse p97 (CDC48) gene. Genomic structure, definition of transcriptional regulatory sequences, gene expression, and characterization of a pseudogene. J Biol Chem *274*, 10154-10162.

Münch, C., Sedlmeier, R., Meyer, T., Homberg, V., Sperfeld, A., Kurt, A., Prudlo, J., Peraus, G., Hanemann, C., Stumm, G.*, et al.* (2004). Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. Neurology *63*, 724-726.

Munoz, D.G., Greene, C., Perl, D.P., and Selkoe, D.J. (1988). Accumulation of phosphorylated neurofilaments in anterior horn motoneurons of amyotrophic lateral sclerosis patients. J Neuropathol Exp Neurol *47*, 9-18.

Murmu, R., Martin, E., Rastetter, A., Esteves, T., Muriel, M.-P., El Hachimi, K., Denora, P., Dauphin, A., Fernandez, J., Duyckaerts, C*., et al.* (2011). Cellular distribution and subcellular localization of spatacsin and spastizin, two proteins involved in hereditary spastic paraplegia. Molecular and cellular neurosciences *47*, 191-202.

Murphy, M.P. (2009). How mitochondria produce reactive oxygen species. Biochem J *417*, 1-13.

Nakagawa, T., Zhu, H., Morishima, N., Li, E., Xu, J., Yankner, B., and Yuan, J. (2000). Caspase-12 mediates endoplasmic-reticulum-specific apoptosis and cytotoxicity by amyloid-beta. Nature *403*, 98-103.

Namikawa, K., Su, Q., Kiryu-Seo, S., and Kiyama, H. (1998). Enhanced expression of 14-3-3 family members in injured motoneurons. Brain Res Mol Brain Res *55*, 315-320.

Nanetti, L., Fancellu, R., Tomasello, C., Gellera, C., Pareyson, D., and Mariotti, C. (2009). Rare association of motor neuron disease and spinocerebellar ataxia type 2 (SCA2): a new case and review of the literature. J Neurol *256*, 1926-1928.

Nawaz, Z., Lonard, D.M., Dennis, A.P., Smith, C.L., and O'Malley, B.W. (1999). Proteasome-dependent degradation of the human estrogen receptor. Proc Natl Acad Sci U S A *96*, 1858-1862.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M*., et al.* (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science *314*, 130-133.

Nishimura, A., Mitneneto, M., Silva, H., Richiericosta, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J., Gillingwater, T., and Webb, J. (2004a). A Mutation in the Vesicle-Trafficking Protein VAPB Causes Late-Onset Spinal Muscular Atrophy and Amyotrophic Lateral Sclerosis. The American Journal of Human Genetics *75*, 822-831.

Nishimura, A., Zupunski, V., Troakes, C., Kathe, C., Fratta, P., Howell, M., Gallo, J.-M., Hortobágyi, T., Shaw, C., and Rogelj, B. (2010). Nuclear import impairment causes cytoplasmic trans-activation response DNA-binding protein accumulation and is associated with frontotemporal lobar degeneration. Brain : a journal of neurology *133*, 1763-1771.

Nishimura, A.L., Mitne-Neto, M., Silva, H.C.A., Oliveira, J.R.M., Vainzof, M., and Zatz, M. (2004b). A novel locus for late onset amyotrophic lateral sclerosis/motor neurone disease variant at 20q13. Journal of medical genetics *41*, 315-320.

Nishitoh, H., Kadowaki, H., Nagai, A., Maruyama, T., Yokota, T., Fukutomi, H., Noguchi, T., Matsuzawa, A., Takeda, K., and Ichijo, H. (2008). ALS-linked mutant SOD1 induces ER stress- and ASK1-dependent motor neuron death by targeting Derlin-1. Genes Dev *22*, 1451-1464.

Nonhoff, U., Ralser, M., Welzel, F., Piccini, I., Balzereit, D., Yaspo, M.-L., Lehrach, H., and Krobitsch, S. (2007). Ataxin-2 interacts with the DEAD/H-box RNA helicase DDX6 and interferes with P-bodies and stress granules. Molecular biology of the cell *18*, 1385-1396.

Nonis, D., Schmidt, M.H., van de Loo, S., Eich, F., Dikic, I., Nowock, J., and Auburger, G. (2008). Ataxin-2 associates with the endocytosis complex and affects EGF receptor trafficking. Cellular signalling *20*, 1725-1739.

Norris, F., Shepherd, R., Denys, E., U, K., Mukai, E., Elias, L., Holden, D., and Norris, H. (1993). Onset, natural history and outcome in idiopathic adult motor neuron disease. Journal of the Neurological Sciences *118*, 48-55.

Oeckinghaus, A., Hayden, M.S., and Ghosh, S. (2011). Crosstalk in NF-kappaB signaling pathways. Nature immunology *12*, 695-708.

Okada, T., Yoshida, H., Akazawa, R., Negishi, M., and Mori, K. (2002). Distinct roles of activating transcription factor 6 (ATF6) and double-stranded RNA-activated protein kinase-like endoplasmic reticulum kinase (PERK) in transcription during the mammalian unfolded protein response. The Biochemical journal *366*, 585-594.

Okamoto, K., Mizuno, Y., and Fujita, Y. (2008). Bunina bodies in amyotrophic lateral sclerosis. Neuropathology : official journal of the Japanese Society of Neuropathology *28*, 109-115.

Okita, T., Nodera, H., Shibuta, Y., Nodera, A., Asanuma, K., Shimatani, Y., Sato, K., Izumi, Y., and Kaji, R. (2011). Can Awaji ALS criteria provide earlier diagnosis than the revised El Escorial criteria? Journal of the neurological sciences *302*, 29-32.

Olney, R., Murphy, J., Forshew, D., Garwood, E., Miller, B., Langmore, S., Kohn, M., and Lomen-Hoerth, C. (2005). The effects of executive and behavioral dysfunction on the course of ALS. Neurology *65*, 1774-1777.

Oosthuyse, B., Moons, L., Storkebaum, E., Beck, H., Nuyens, D., Brusselmans, K., Van Dorpe, J., Hellings, P., Gorselink, M., Heymans, S*., et al.* (2001). Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration. Nature genetics *28*, 131-138.

Orlacchio, A., Babalini, C., Borreca, A., Patrono, C., Massa, R., Basaran, S., Munhoz, R.P., Rogaeva, E.A., St George-Hyslop, P.H., Bernardi, G*., et al.* (2010). SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. Brain *133*, 591-598.

Orrell, R., Habgood, J., Malaspina, A., Mitchell, J., Greenwood, J., Lane, R., and deBelleroche, J. (1999). Clinical characteristics of SOD1 gene mutations in UK families with ALS. Journal of the neurological sciences *169*, 56-60.

Orrell, R.W., King, A.W., Hilton, D.A., Campbell, M.J., Lane, R.J., and de Belleroche, J.S. (1995). Familial amyotrophic lateral sclerosis with a point mutation of SOD-1: intrafamilial heterogeneity of disease duration associated with neurofibrillary tangles. J Neurol Neurosurg Psychiatry *59*, 266-270.

Orrenius, S., Zhivotovsky, B., and Nicotera, P. (2003). Regulation of cell death: the calcium-apoptosis link. Nat Rev Mol Cell Biol *4*, 552-565.

Otomo, A., Hadano, S., Okada, T., Mizumura, H., Kunita, R., Nishijima, H., Showguchi-Miyata, J., Yanagisawa, Y., Kohiki, E., Suga, E*., et al.* (2003). ALS2, a novel guanine nucleotide exchange factor for the small GTPase Rab5, is implicated in endosomal dynamics. Human molecular genetics *12*, 1671-1687.

Pamphlett, R., Morahan, J.M., and Yu, B. (2011). Using case-parent trios to look for rare de novo genetic variants in adult-onset neurodegenerative diseases. J Neurosci Methods *197*, 297-301.

Pardo, A.C., Wong, V., Benson, L.M., Dykes, M., Tanaka, K., Rothstein, J.D., and Maragakis, N.J. (2006). Loss of the astrocyte glutamate transporter GLT1 modifies disease in SOD1(G93A) mice. Exp Neurol *201*, 120-130.

Parkinson, N., Ince, P.G., Smith, M.O., Highley, R., Skibinski, G., Andersen, P.M., Morrison, K.E., Pall, H.S., Hardiman, O., Collinge, J*., et al.* (2006). ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). Neurology *67*, 1074-1077.

Pasinelli, P., Belford, M.E., Lennon, N., Bacskai, B.J., Hyman, B.T., Trotti, D., and Brown, R.H., Jr. (2004). Amyotrophic lateral sclerosis-associated SOD1 mutant proteins bind and aggregate with Bcl-2 in spinal cord mitochondria. Neuron *43*, 19-30.

Pasinelli, P., Borchelt, D.R., Houseweart, M.K., Cleveland, D.W., and Brown, R.H., Jr. (1998). Caspase-1 is activated in neural cells and tissue with amyotrophic lateral sclerosis-associated mutations in copper-zinc superoxide dismutase. Proc Natl Acad Sci U S A *95*, 15763-15768.

Pasinelli, P., and Brown, R.H. (2006). Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat Rev Neurosci *7*, 710-723.

Patnala, R., Clements, J., and Batra, J. (2013). Candidate gene association studies: a comprehensive guide to useful in silico tools. BMC genetics *14*, 39.

Paul, P., and de Belleroche, J. (2012). The role of D-amino acids in amyotrophic lateral sclerosis pathogenesis: a review. Amino acids *43*, 1823-1831.

Paul, P., and de Belleroche, J. (2014). The role of D-serine and glycine as co-agonists of NMDA receptors in motor neurone degeneration and amyotrophic lateral sclerosis (ALS). Frontiers in Synaptic Neuroscience *16*, 6-10.

Paulson, H.L., and Fischbeck, K.H. (1996). Trinucleotide repeats in neurogenetic disorders. Annu Rev Neurosci *19*, 79-107.

Perk, J., Iavarone, A., and Benezra, R. (2005). Id family of helix-loop-helix proteins in cancer. Nature reviews Cancer *5*, 603-614.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Huelga, S.C., Moran, J., Liang, T.Y., Ling, S.-C., Sun, E., Wancewicz, E., Mazur, C.*, et al.* (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat Neurosci *14*, 459-468.

Powis, G., Briehl, M., and Oblong, J. (1995). Redox signalling and the control of cell growth and death. Pharmacology & therapeutics *68*, 149-173.

Przedborski, S., Donaldson, D., Jakowec, M., Kish, S.J., Guttman, M., Rosoklija, G., and Hays, A.P. (1996). Brain superoxide dismutase, catalase, and glutathione peroxidase activities in amyotrophic lateral sclerosis. Ann Neurol *39*, 158-165.

Puls, I., Jonnakuty, C., LaMonte, B.H., Holzbaur, E.L., Tokito, M., Mann, E., Floeter, M.K., Bidus, K., Drayna, D., Oh, S.J.*, et al.* (2003). Mutant dynactin in motor neuron disease. Nat Genet *33*, 455-456.

Qiao, Y., Liu, X., Harvard, C., Hildebrand, M.J., Rajcan-Separovic, E., Holden, J.J., and Lewis, M.E. (2008). Autism-associated familial microdeletion of Xp11.22. Clin Genet *74*, 134-144.

Qiu, Y. (2007). Pre-B-cell leukemia transcription factor 1 regulates expression of valosin-containing protein, a gene involved in cancer growth. The American journal of pathology *170*, 152.

Rakhit, R., and Chakrabartty, A. (2006). Structure, folding, and misfolding of Cu,Zn superoxide dismutase in amyotrophic lateral sclerosis. Biochim Biophys Acta *1762*, 1025-1037.

Rao, S.D., Yin, H.Z., and Weiss, J.H. (2003). Disruption of glial glutamate transport by reactive oxygen species produced in motor neurons. J Neurosci *23*, 2627-2633.

Ravits, J., Paul, P., and Jorg, C. (2007). Focality of upper and lower motor neuron degeneration at the clinical onset of ALS. Neurology *68*, 1571-1575.

Ravits, J.M., and La Spada, A.R. (2009). ALS motor phenotype heterogeneity, focality, and spread: deconstructing motor neuron degeneration. Neurology *73*, 805-811.

Ray, S.S., Nowak, R.J., Strokovich, K., Brown, R.H., Jr., Walz, T., and Lansbury, P.T., Jr. (2004). An intersubunit disulfide bond prevents in vitro aggregation of a superoxide dismutase-1 mutant linked to familial amytrophic lateral sclerosis. Biochemistry *43*, 4899-4905.

Rea, S.L., Walsh, J.P., Ward, L., Magno, A.L., Ward, B.K., Shaw, B., Layfield, R., Kent, G.N., Xu, J., and Ratajczak, T. (2009). Sequestosome 1 Mutations in Paget's Disease of Bone in Australia: Prevalence, Genotype/Phenotype Correlation, and a Novel Non-UBA Domain Mutation (P364S) Associated With Increased NF-κB Signaling Without Loss of Ubiquitin Binding. Journal of bone and mineral research *24*, 1216-1223.

Renton, A. (2011). A Hexanucleotide Repeat Expansion in *C9ORF72* Is the Cause of Chromosome 9p21-Linked ALS-FTD. Neuron *72*, 257-268.

Renton, A., Chiò, A., and Traynor, B. (2014). State of play in amyotrophic lateral sclerosis genetics. Nature neuroscience *17*, 17-23.

Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L.*, et al.* (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72*, 257-268.

Ringholz, G., Appel, S., Bradshaw, M., Cooke, N., Mosnik, D., and Schulz, P. (2005). Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology *65*, 586-590.

Risch, N. (2000). Searching for genetic determinants in the new millennium. Nature *405*, 847-856.

Rodova, M., Kelly, K.F., VanSaun, M., Daniel, J.M., and Werle, M.J. (2004). Regulation of the rapsyn promoter by kaiso and delta-catenin. Mol Cell Biol *24*, 7188-7196.

Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. Nature Reviews Molecular Cell Biology *8*, 519-529.

Rosen, D., Sapp, P., O'Regan, J., Horvitz, H., Donaldson, D., Nussbaum, C., Gusella, J., Haines, J., Pestka, S., and Jung, V. (1992). Dinucleotide repeat polymorphisms (D21S223 and D21S224) at 21q22.1. Human molecular genetics *1*, 547.

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X.*, et al.* (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature *362*, 59-62.

Ross, O.A., Rutherford, N.J., Baker, M., Soto-Ortolaza, A.I., Carrasquillo, M.M., Dejesus-Hernandez, M., Adamson, J., Li, M., Volkening, K., Finger, E.*, et al.* (2011). Ataxin-2 repeat-length variation and neurodegeneration. Hum Mol Genet *20*, 3207-12.

Rothenberg, C., Srinivasan, D., Mah, L., Kaushik, S., Peterhoff, C., Ugolino, J., Fang, S., Cuervo, A., Nixon, R., and Monteiro, M. (2010). Ubiquilin functions in autophagy and is degraded by chaperone-mediated autophagy. Human molecular genetics *19*, 3219-3232.

Rothstein, J., Tsai, G., Kuncl, R., Clawson, L., Cornblath, D., Drachman, D., Pestronk, A., Stauch, B., and Coyle, J. (1990). Abnormal excitatory amino acid metabolism in amyotrophic lateral sclerosis. Annals of Neurology *28*, 18-25.

Rothstein, J.D., Patel, S., Regan, M.R., Haenggeli, C., Huang, Y.H., Bergles, D.E., Jin, L., Dykes Hoberg, M., Vidensky, S., Chung, D.S.*, et al.* (2005). Beta-lactam antibiotics offer neuroprotection by increasing glutamate transporter expression. Nature *433*, 73-77.

Rouleau, G.A., Clark, A.W., Rooke, K., Pramatarova, A., Krizus, A., Suchowersky, O., Julien, J.-P., and Figlewicz, D. (1996). SOD1 mutation is assosiated with accumulation of neurofilaments in amyotrophic lateral scelaries. Annals of Neurology *39*, 128-131.

Rubino, E., Rainero, I., Chio, A., Rogaeva, E., Galimberti, D., Fenoglio, P., Grinberg, Y., Isaia, G., Calvo, A., Gentile, S.*, et al.* (2012). SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Neurology *79*, 1556-1562.

Rutherford, N., Heckman, M., Dejesus-Hernandez, M., Baker, M., Soto-Ortolaza, A., Rayaprolu, S., Stewart, H., Finger, E., Volkening, K., Seeley, W.*, et al.* (2012). Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype. Neurobiology of Aging *33*, 2950.e5-7.

Sahlender, D.A., Roberts, R.C., Arden, S.D., Spudich, G., Taylor, M.J., Luzio, J.P., Kendrick-Jones, J., and Buss, F. (2005). Optineurin links myosin VI to the Golgi complex and is involved in Golgi organization and exocytosis. The Journal of Cell Biology *169*, 285-295.

Salinas, S., Proukakis, C., Crosby, A., and Warner, T.T. (2008). Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. Lancet Neurol *7*, 1127-1138.

Sapp, P.C., Hosler, B.A., McKenna-Yasek, D., Chin, W., Gann, A., Genise, H., Gorenstein, J., Huang, M., Sailer, W., Scheffler, M.*, et al.* (2003). Identification of two novel loci for dominantly inherited familial amyotrophic lateral sclerosis. Am J Hum Genet *73*, 397-403.

Sasabe, J., Chiba, T., Yamada, M., Okamoto, K., Nishimoto, I., Matsuoka, M., and Aiso, S. (2007). D-Serine is a key determinant of glutamate toxicity in amyotrophic lateral sclerosis. The EMBO journal *26*, 4149-4159.

Sasabe, J., Miyoshi, Y., Suzuki, M., Mita, M., Konno, R., Matsuoka, M., Hamase, K., and Aiso, S. (2012). D-amino acid oxidase controls motoneuron degeneration through D-serine. Proceedings of the National Academy of Sciences of the United States of America *109*, 627-632.

Sasaki, S., and Iwata, M. (1996). Impairment of fast axonal transport in the proximal axons of anterior horn neurons in amyotrophic lateral sclerosis. Neurology *47*, 535-540.

Sasaki, S., Warita, H., Murakami, T., Abe, K., and Iwata, M. (2004). Ultrastructural study of mitochondria in the spinal cord of transgenic mice with a G93A mutant SOD1 gene. Acta neuropathologica *107*, 461-474.

Schulz, J.B., Lindenau, J., Seyfried, J., and Dichgans, J. (2000). Glutathione, oxidative stress and neurodegeneration. Eur J Biochem *267*, 4904-4911.

Schymick, J., Talbot, K., and Traynor, B. (2007a). Genetics of sporadic amyotrophic lateral sclerosis. Human molecular genetics *16 Spec No. 2*, 42.

Schymick, J.C., Scholz, S.W., Fung, H.C., Britton, A., Arepalli, S., Gibbs, J.R., Lombardo, F., Matarin, M., Kasperaviciute, D., Hernandez, D.G.*, et al.* (2007b). Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol *6*, 322-328.

Sham, P.C., and Curtis, D. (1995). Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Annals of Human Genetics *59*, 97-105.

Shamoo, Y. (2001). Single-stranded DNA-binding Proteins. eLS John Wiley & Sons Ltd, Chichester *http://www.els.net*.

Shatunov, A., Mok, K., Newhouse, S., Weale, M., Smith, B., Vance, C., Johnson, L., Veldink, J., van Es, M., van den Berg, L.*, et al.* (2010). Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet neurology *9*, 986-994.

Shibata, N., Nagai, R., Uchida, K., Horiuchi, S., Yamada, S., Hirano, A., Kawaguchi, M., Yamamoto, T., Sasaki, S., and Kobayashi, M. (2001). Morphological evidence for lipid peroxidation and protein glycoxidation in spinal cords from sporadic amyotrophic lateral sclerosis patients. Brain Res *917*, 97-104.

Shimada, T., Fournier, A., and Yamagata, K. (2013). Neuroprotective Function of 14-3-3 Proteins in Neurodegeneration. BioMed research international *2013*, 564534.

Siddique, T., Figlewicz, D., Pericak-Vance, M., Haines, J., Rouleau, G., Jeffers, A., Sapp, P., Hung, W., Bebout, J., and McKenna-Yasek, D. (1991). Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. The New England journal of medicine *324*, 1381-1384.

Skibinski, G., Parkinson, N., Brown, J., Chakrabarti, L., Lloyd, S., Hummerich, H., Nielsen, J., Hodges, J., Spillantini, M., Thusgaard, T.*, et al.* (2005). Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. Nature Genetics *37*, 806-808.

Skourti-Stathaki, K., Proudfoot, N.J., and Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. Mol Cell *42*, 794-805.

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E.*, et al.* (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science *319*, 1668-1672.

Steinacker, P., Aitken, A., and Otto, M. (2011). 14-3-3 proteins in neurodegeneration. Seminars in cell & developmental biology *22*, 696-704.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol Cell *24*, 735-746.

Stevanin, G., Azzedine, H., Denora, P., Boukhris, A., Tazir, M., Lossos, A., Rosa, A.L., Lerer, I., Hamri, A., Alegria, P.*, et al.* (2008). Mutations in SPG11 are frequent in autosomal recessive spastic paraplegia with thin corpus callosum, cognitive decline and lower motor neuron degeneration. Brain *131*, 772-784.

Stevanin, G., Santorelli, F.M., Azzedine, H., Coutinho, P., Chomilier, J., Denora, P.S., Martin, E., Ouvrard-Hernandez, A.M., Tessa, A., Bouslam, N.*, et al.* (2007). Mutations in SPG11, encoding spatacsin, are a major cause of spastic paraplegia with thin corpus callosum. Nat Genet *39*, 366-372.

Strachan, T.R., A. (2011). Human Molecular Genetics, 4 edn. Garland Science. *ISBN 978-0-815-34149-9*.

Subramanian, V., Crabtree, B., and Acharya, K.R. (2008). Human angiogenin is a neuroprotective factor and amyotrophic lateral sclerosis associated angiogenin variants affect neurite extension/pathfinding and survival of motor neurons. Human Molecular Genetics *17*, 130-149.

Sundar, P., Yu, C.-E., Sieh, W., Steinbart, E., Garruto, R., Oyanagi, K., Craig, U.-K., Bird, T., Wijsman, E., Galasko, D.*, et al.* (2007). Two sites in the MAPT region confer genetic risk for Guam ALS/PDC and dementia. Human molecular genetics *16*, 295-306.

Suraweera, A., Becherel, O.J., Chen, P., Rundle, N., Woods, R., Nakamura, J., Gatei, M., Criscuolo, C., Filla, A., Chessa, L.*, et al.* (2007). Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. J Cell Biol *177*, 969-979.

Suraweera, A., Lim, Y., Woods, R., Birrell, G., Nasim, T., Becherel, O., and Lavin, M. (2009). Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. Human molecular genetics *18*, 3384-3396.

Swarup, V., Phaneuf, D., Dupre, N., Petri, S., Strong, M., Kriz, J., and Julien, J.P. (2011). Deregulation of TDP-43 in amyotrophic lateral sclerosis triggers nuclear factor kappaB-mediated pathogenic pathways. J Exp Med *208*, 2429-2447.

Swerdlow, R.H., Parks, J.K., Cassarino, D.S., Trimmer, P.A., Miller, S.W., Maguire, D.J., Sheehan, J.P., Maguire, R.S., Pattee, G., Juel, V.C*., et al.* (1998). Mitochondria in sporadic amyotrophic lateral sclerosis. Exp Neurol *153*, 135-142.

Tabor, H., Risch, N., and Myers, R. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature reviews Genetics *3*, 391-397.

Tabor, H.K. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature reviews Genetics *3*, 391.

Tachi, N., Kikuchi, S., Kozuka, N., and Nogami, A. (2005). A new mutation of IGHMBP2 gene in spinal muscular atrophy with respiratory distress type 1. Pediatr Neurol *32*, 288-290.

Tagawa, A., Tan, C.-F., Kikugawa, K., Fukase, M., Nakano, R., Onodera, O., Nishizawa, M., and Takahashi, H. (2007). Familial amyotrophic lateral sclerosis: a SOD1-unrelated Japanese family of bulbar type with Bunina bodies and ubiquitin-positive skein-like inclusions in lower motor neurons. Acta neuropathologica *113*, 205-211.

Tainer, J.A., Getzoff, E.D., Beem, K.M., Richardson, J.S., and Richardson, D.C. (1982). Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. Journal of Molecular Biology *160*, 181-217.

Takeuchi, H., Kobayashi, Y., Ishigaki, S., Doyu, M., and Sobue, G. (2002). Mitochondrial localization of mutant superoxide dismutase 1 triggers caspase-dependent cell death in a cellular model of familial amyotrophic lateral sclerosis. The Journal of biological chemistry *277*, 50966-50972.

Tan, C.-F., Eguchi, H., Tagawa, A., Onodera, O., Iwasaki, T., Tsujino, A., Nishizawa, M., Kakita, A., and Takahashi, H. (2007). TDP-43 immunoreactivity in neuronal inclusions in familial amyotrophic lateral sclerosis with or without SOD1 gene mutation. Acta neuropathologica *113*, 535-542.

Taylor, J.P., Hardy, J., and Fischbeck, K.H. (2002). Toxic proteins in neurodegenerative disease. Science *296*, 1991-1995.

Teyssou, E., Takeda, T., Lebon, V., Boillée, S., Doukouré, B., Bataillon, G., Sazdovitch, V., Cazeneuve, C., Meininger, V., LeGuern, E*., et al.* (2013). Mutations in SQSTM1 encoding p62 in amyotrophic lateral sclerosis: genetics and neuropathology. Acta neuropathologica, 1-12.

Theriot, J.A., and Mitchison, T.J. (1993). The three faces of profilin. Cell *75*, 835-838.

Thiyagarajan, N., Ferguson, R., Subramanian, V., and Acharya, K. (2012). Structural and molecular insights into the mechanism of action of human angiogenin-ALS variants in neurons. Nature communications *3*, 1121.

Tiloca, C., Ratti, A., Pensato, V., Castucci, A., Soraru, G., Del Bo, R., Corrado, L., Cereda, C., D'Ascenzo, C., Comi, G.P.*, et al.* (2011). Mutational analysis of VCP gene in familial amyotrophic lateral sclerosis. Neurobiol Aging.

Tobisawa, S., Hozumi, Y., Arawaka, S., Koyama, S., Wada, M., Nagai, M., Aoki, M., Itoyama, Y., Goto, K., and Kato, T. (2003). Mutant SOD1 linked to familial amyotrophic lateral sclerosis, but not wild-type SOD1, induces ER stress in COS7 cells and transgenic mice. Biochem Biophys Res Commun *303*, 496-503.

Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., Konig, J., Hortobagyi, T., Nishimura, A.L., Zupunski, V.*, et al.* (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci.

Treangen, T., and Salzberg, S. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature reviews Genetics *13*, 36-46.

Trotti, D., Aoki, M., Pasinelli, P., Berger, U.V., Danbolt, N.C., Brown, R.H., Jr., and Hediger, M.A. (2001). Amyotrophic lateral sclerosis-linked glutamate transporter mutant has impaired glutamate clearance capacity. J Biol Chem *276*, 576-582.

Tudor, E.L. (2010). Amyotrophic lateral sclerosis mutant vesicle-associated membrane protein-associated protein-B transgenic mice develop TAR-DNA-binding protein-43 pathology. Neuroscience *167*, 774.

Urushitani, M., Ezzi, S., Matsuo, A., Tooyama, I., and Julien, J.-P. (2008). The endoplasmic reticulum-Golgi pathway is a target for translocation and aggregation of mutant superoxide dismutase linked to ALS. FASEB journal : official publication of the Federation of American Societies for Experimental Biology *22*, 2476-2487.

Urushitani, M., Kurisu, J., Tsukita, K., and Takahashi, R. (2002). Proteasomal inhibition by misfolded mutant superoxide dismutase 1 induces selective motor neuron death in familial amyotrophic lateral sclerosis. J Neurochem *83*, 1030-1042.

Valdmanis, P.N., Verlaan, D.J., and Rouleau, G.A. (2009). The proportion of mutations predicted to have a deleterious effect differs between gain and loss of function genes in neurodegenerative disease. Human mutation *30*, E481-E489.

van Blitterswijk, M., DeJesus-Hernandez, M., Niemantsverdriet, E., Murray, M.E., Heckman, M.G., Diehl, N.N., Brown, P.H., Baker, M.C., Finch, N.A., Bauer, P.O.*, et al.* (2013). Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol *12*, 978-988.

van Blitterswijk, M., van Es, M.A., Hennekam, E.A., Dooijes, D., van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., van der Kooi, A.J., de Visser, M.*, et al.* (2012a). Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum Mol Genet *21*, 3776-3784.

van Blitterswijk, M., van Vught, P.W., van Es, M.A., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., Veldink, J.H., and van den Berg, L.H. (2012b). Novel optineurin mutations in sporadic amyotrophic lateral sclerosis patients. Neurobiol Aging *33*, 1016 e1011-1017.

Van Damme, P., Veldink, J., van Blitterswijk, M., Corveleyn, A., van Vught, P., Thijs, V., Dubois, B., Matthijs, G., van den Berg, L., and Robberecht, W. (2011). Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. Neurology *76*, 2066-2072.

Van Deerlin, V.M., Leverenz, J.B., Bekris, L.M., Bird, T.D., Yuan, W., Elman, L.B., Clay, D., Wood, E.M., Chen-Plotkin, A.S., Martinez-Lage, M.*, et al.* (2008). TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. Lancet Neurol *7*, 409-416.

van Es, M.A., Diekstra, F.P., Veldink, J.H., Baas, F., Bourque, P.R., Schelhaas, H.J., Strengman, E., Hennekam, E.A., Lindhout, D., Ophoff, R.A.*, et al.* (2009a). A case of ALS-FTD in a large FALS pedigree with a K17I ANG mutation. Neurology *72*, 287-288.

van Es, M.A., Van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Andersen, P.M., Van Den Bosch, L., de Jong, S.W., van 't Slot, R., Birve, A.*, et al.* (2007). ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. Lancet Neurol *6*, 869-877.

van Es, M.A., van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Van den Bosch, L., de Jong, S.W., de Jong, V., Baas, F., van't Slot, R.*, et al.* (2008). Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. Nat Genet *40*, 29-31.

van Es, M.A., Veldink, J.H., Saris, C.G., Blauw, H.M., van Vught, P.W., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J., Huisman, M.H.*, et al.* (2009b). Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet *41*, 1083-1087.

Vance, C., Al-Chalabi, A., Ruddy, D., Smith, B., Hu, X., Sreedharan, J., Siddique, T., Schelhaas, H., Kusters, B., Troost, D.*, et al.* (2006). Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. Brain : a journal of neurology *129*, 868-876.

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P.*, et al.* (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science *323*, 1208-1211.

Vantaggiato, C., Bondioni, S., Airoldi, G., Bozzato, A., Borsani, G., Rugarli, E., Bresolin, N., Clementi, E., and Bassi, M. (2011). Senataxin modulates neurite growth through fibroblast growth factor 8 signalling. Brain : a journal of neurology *134*, 1808-1828.

Vielhaber, S., Kunz, D., Winkler, K., Wiedemann, F.R., Kirches, E., Feistner, H., Heinze, H.J., Elger, C.E., Schubert, W., and Kunz, W.S. (2000). Mitochondrial DNA abnormalities in skeletal muscle of patients with sporadic amyotrophic lateral sclerosis. Brain : a journal of neurology *123 (Pt 7)*, 1339-1348.

Volkening, K., Leystra-Lantz, C., Yang, W., Jaffee, H., and Strong, M. (2009). Tar DNA binding protein of 43 kDa (TDP-43), 14-3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS). Brain research *1305*, 168-182.

Vosler, P., Brennan, C., and Chen, J. (2008). Calpain-mediated signaling mechanisms in neuronal injury and neurodegeneration. Molecular neurobiology *38*, 78-100.

Voustianiouk, A., Seidel, G., Panchal, J., Sivak, M., Czaplinski, A., Yen, A., Appel, S., and Lange, D. (2008). ALSFRS and appel ALS scores: discordance with disease progression. Muscle & nerve *37*, 668-672.

Walker, A.K., Farg, M.A., Bye, C.R., McLean, C.A., Horne, M.K., and Atkin, J.D. (2010). Protein disulphide isomerase protects against protein aggregation and is S-nitrosylated in amyotrophic lateral sclerosis. Brain : a journal of neurology *133*, 105-116.

Wang, Q., Johnson, J.L., Agar, N.Y., and Agar, J.N. (2008). Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. PLoS Biol *6*, e170.

Weder, N., Aziz, R., Wilkins, K., and Tampi, R. (2007). Frontotemporal dementias: a review. Annals of general psychiatry *6*, 15.

Weihl, C.C., Dalal, S., Pestronk, A., and Hanson, P.I. (2006). Inclusion body myopathy-associated mutations in p97/VCP impair endoplasmic reticulum-associated degradation. Hum Mol Genet *15*, 189-199.

Weihl, C.C., Pestronk, A., and Kimonis, V.E. (2009). Valosin-containing protein disease: inclusion body myopathy with Paget's disease of the bone and fronto-temporal dementia. Neuromuscul Disord *19*, 308-315.

Weihl, C.C., Temiz, P., Miller, S.E., Watts, G., Smith, C., Forman, M., Hanson, P.I., Kimonis, V., and Pestronk, A. (2008). TDP-43 accumulation in inclusion body myopathy muscle suggests a common pathogenic mechanism with frontotemporal dementia. J Neurol Neurosurg Psychiatry *79*, 1186-1189.

Weishaupt, J.H., Waibel, S., Birve, A., Volk, A.E., Mayer, B., Meyer, T., Ludolph, A.C., and Andersen, P.M. (2013). A novel optineurin truncating mutation and three glaucoma-

associated missense variants in patients with familial amyotrophic lateral sclerosis in Germany. Neurobiol Aging *34*, 1516 e1519-1515.

Wijesekera, L.C., and Leigh, P.N. (2009). Amyotrophic lateral sclerosis. Orphanet Journal of Rare Diseases *4*, 3.

Wijesekera, L.C., Mathers, S., Talman, P., Galtrey, C., Parkinson, M.H., Ganesalingam, J., Willey, E., Ampong, M.A., Ellis, C.M., Shaw, C.E*., et al.* (2009). Natural history and clinical features of the flail arm and flail leg ALS variants. Neurology *72*, 1087-1094.

Williams, K.L., Solski, J.A., Nicholson, G.A., and Blair, I.P. (2012). Mutation analysis of VCP in familial and sporadic amyotrophic lateral sclerosis. Neurobiol Aging *33*, 1488 e1415-1486.

Williamson, T.L., Bruijn, L.I., Zhu, Q., Anderson, K.L., Anderson, S.D., Julien, J.P., and Cleveland, D.W. (1998). Absence of neurofilaments reduces the selective vulnerability of motor neurons and slows disease caused by a familial amyotrophic lateral sclerosis-linked superoxide dismutase 1 mutant. Proc Natl Acad Sci U S A *95*, 9631-9636.

Williamson, T.L., and Cleveland, D.W. (1999). Slowing of axonal transport is a very early event in the toxicity of ALS-linked SOD1 mutants to motor neurons. Nat Neurosci *2*, 50-56.

Winton, M., Igaz, L., Wong, M., Kwong, L., Trojanowski, J., and Lee, V. (2008). Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. The Journal of biological chemistry *283*, 13302-13309.

Witke, W., Podtelejnikov, A., Di Nardo, A., Sutherland, J., Gurniak, C., Dotti, C., and Mann, M. (1998). In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. The EMBO journal *17*, 967-976.

Wojcikiewicz, R.J. (2004). Regulated ubiquitination of proteins in GPCR-initiated signaling pathways. Trends Pharmacol Sci *25*, 35-41.

Wokke, J. (1996). Riluzole. Lancet *348*, 795-799.

Wolosker, H., Dumin, E., Balan, L., and Foltyn, V.N. (2008). D-amino acids in the brain: D-serine in neurotransmission and neurodegeneration. FEBS J *275*, 3514-3526.

Wood, J., Beaujeux, T., and Shaw, P. (2003). Protein aggregation in motor neurone disorders. Neuropathology and applied neurobiology *29*, 529-545.

Woollacott, I., and Mead, S. (2014). The C9ORF72 expansion mutation: gene structure, phenotypic and diagnostic issues. Acta neuropathologica *127*, 319-332.

Wootz, H., Hansson, I., Korhonen, L., Napankangas, U., and Lindholm, D. (2004). Caspase-12 cleavage and increased oxidative stress during motoneuron degeneration in transgenic mouse model of ALS. Biochem Biophys Res Commun *322*, 281-286.

Wu, C.-H., Fallini, C., Ticozzi, N., Keagle, P., Sapp, P., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D.*, et al.* (2012). Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature *488*, 499-503.

Wu, D., Yu, W., Kishikawa, H., Folkerth, R.D., Iafrate, A.J., Shen, Y., Xin, W., Sims, K., and Hu, G.F. (2007). Angiogenin loss-of-function mutations in amyotrophic lateral sclerosis. Ann Neurol *62*, 609-617.

Xiao, S., Sanelli, T., Dib, S., Sheps, D., Findlater, J., Bilbao, J., Keith, J., Zinman, L., Rogaeva, E., and Robertson, J. (2011). RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. Mol Cell Neurosci *47*, 167-180.

Xu, Z., Cork, L.C., Griffin, J.W., and Cleveland, D.W. (1993). Increased expression of neurofilament subunit NF-L produces morphological alterations that resemble the pathology of human motor neuron disease. Cell *73*, 23-33.

Yamashita, T., Hideyama, T., Hachiga, K., Teramoto, S., Takano, J., Iwata, N., Saido, T.C., and Kwak, S. (2012). A role for calpain-dependent cleavage of TDP-43 in amyotrophic lateral sclerosis pathology. Nat Commun *3*, 1307.

Yan, J., Deng, H.X., Siddique, N., Fecto, F., Chen, W., Yang, Y., Liu, E., Donkervoort, S., Zheng, J.G., Shi, Y.*, et al.* (2010). Frameshift and novel mutations in FUS in familial amyotrophic lateral sclerosis and ALS/dementia. Neurology *75*, 807-814.

Yang, C., Tan, W., Whittle, C., Qiu, L., Cao, L., Akbarian, S., and Xu, Z. (2010). The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism. PLoS One *5*, e15878.

Yang, Y.S., Harel, N.Y., and Strittmatter, S.M. (2009). Reticulon-4A (Nogo-A) redistributes protein disulfide isomerase to protect mice from SOD1-dependent amyotrophic lateral sclerosis. The Journal of neuroscience : the official journal of the Society for Neuroscience *29*, 13850-13859.

Ye, Y., Meyer, H., and Rapoport, T. (2003). Function of the p97-Ufd1-Npl4 complex in retrotranslocation from the ER to the cytosol: dual recognition of nonubiquitinated polypeptide segments and polyubiquitin chains. The Journal of cell biology *162*, 71-84.

Ye, Y., Shibata, Y., Yun, C., Ron, D., and Rapoport, T.A. (2004). A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. Nature *429*, 841-847.

Yin, H.Z., Nalbandian, A., Hsu, C.I., Li, S., Llewellyn, K.J., Mozaffar, T., Kimonis, V.E., and Weiss, J.H. (2012). Slow development of ALS-like spinal cord pathology in mutant valosin-containing protein gene knock-in mice. Cell Death Dis *3*, e374.

Yorimitsu, T., and Klionsky, D.J. (2005). Autophagy: molecular machinery for self-eating. Cell Death Differ *12 Suppl 2*, 1542-1552.

Yoshida, H. (2007). ER stress and diseases. FEBS J *274*, 630-658.

Yu, Z., Zhu, Y., Chen-Plotkin, A., Clay-Falcone, D., McCluskey, L., Elman, L., Kalb, R., Trojanowski, J., Lee, V., Van Deerlin, V.*, et al.* (2011). PolyQ repeat expansions in ATXN2 associated with ALS are CAA interrupted repeats. PloS one *6*.

Zhang, B., Tomita, Y., Qiu, Y., He, J., Morii, E., Noguchi, S., and Aozasa, K. (2007a). E74-like factor 2 regulates valosin-containing protein expression. Biochem Biophys Res Commun *356*, 536-541.

Zhang, X., Li, L., Chen, S., Yang, D., Wang, Y., Zhang, X., Wang, Z., and Le, W. (2011). Rapamycin treatment augments motor neuron degeneration in SOD1(G93A) mouse model of amyotrophic lateral sclerosis. Autophagy *7*, 412-425.

Zhang, Y.-J., Xu, Y.-F., Cook, C., Gendron, T.F., Roettges, P., Link, C.D., Lin, W.-L., Tong, J., Castanedes-Casey, M., Ash, P.*, et al.* (2009). Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. Proceedings of the National Academy of Sciences *106*, 7607-7612.

Zhang, Y.J., Xu, Y.F., Dickey, C.A., Buratti, E., Baralle, F., Bailey, R., Pickering-Brown, S., Dickson, D., and Petrucelli, L. (2007b). Progranulin mediates caspase-dependent cleavage of TAR DNA binding protein-43. J Neurosci *27*, 10530-10534.

Zhang, Z., Hartmann, H., Do, V.M., Abramowski, D., Sturchler-Pierrat, C., Staufenbiel, M., Sommer, B., van de Wetering, M., Clevers, H., Saftig, P.*, et al.* (1998). Destabilization of beta-catenin by mutations in presenilin-1 potentiates neuronal apoptosis. Nature *395*, 698-702.

Zhao, M., Tang, D., Lechpammer, S., Hoffman, A., Asea, A., Stevenson, M.A., and Calderwood, S.K. (2002). Double-stranded RNA-dependent Protein Kinase (pkr) Is Essential for Thermotolerance, Accumulation of HSP70, and Stabilization of ARE-containing HSP70 mRNA during Stress. Journal of Biological Chemistry *277*, 44539-44547.

Zhu, G., Wu, C.J., Zhao, Y., and Ashwell, J.D. (2007). Optineurin negatively regulates TNFalpha- induced NF-kappaB activation by competing with NEMO for ubiquitinated RIP. Curr Biol *17*, 1438-1443.

Zinman, L., and Cudkowicz, M. (2011). Emerging targets and treatments in amyotrophic lateral sclerosis. Lancet Neurol *10*, 481-490.

Zou, Z.Y., Liu, M.S., Li, X.G., and Cui, L.Y. (2013). Screening of VCP mutations in Chinese amyotrophic lateral sclerosis patients. Neurobiol Aging *34*, 1519 e1513-1514.

Appendix

Appendix I: R Codes used for a quick assessment of genotypes/allelic counts and association tests

individual	clusterbx	population	status	snp_3	snp_4	snp_6	snp_9	c9orf72
477	5B	FALS	2	AG	AA	TT	GT	POS
478	5C	FALS	2	AG	AA	TT	GT	UNKNOWN
479	5D	FALS	2	AA	AG	TT	GG	UNKNOWN
480	5E	FALS	2	AA	AG	TT	GG	POS
481	5F	FALS	2	AA	NA	TC	NA	POS
482	5G	FALS	2	AA	AG	TT	GG	UNKNOWN
483	5H	FALS	2	AA	AA	NA	NA	UNKNOWN

```
# Genotype 3.0: Usage (with the above data format for example,
major allele can not be missing)

snp_3allele=c("AA","AG","GG")
casepop=file[which(population=="FALS"),]
controlpop=file[which(population=="CONTROL"),]

> genotypetabfin_all("snp_3",snp_3allele,casepop,controlpop)
> genotypefisher_all("snp_3",snp_3allele,casepop,controlpop)
> alleletabfin_all("snp_3",snp_3allele,casepop,controlpop)
> allelefisher_all("snp_3",snp_3allele,casepop,controlpop)
> allelefisherwithod_all("snp_3",snp_3allele,casepop,controlpop)
> hwepvalue_all("snp_3",snp_3allele,casepop,controlpop)

#=========Atomic Functions============
genotypefisher=function(SNP,ALLELE,POP1,POP2){
firstpopulation=subset(POP1,select=SNP)
firstpopulationmatrix1=as.matrix(firstpopulation)
firstpop=na.omit(firstpopulationmatrix1)
firstpop_stand=c(" ")
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[1])
firstpop_stand[i]="AA"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[2])
firstpop_stand[i]="AB"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[3])
firstpop_stand[i]="BB"}
firstpopulationmatrix=as.matrix(firstpop_stand)
firstgeno<-genotype(firstpopulationmatrix,sep="")
firsttab<-table(firstgeno)
firsttabadj<-c(firsttab,0,0,0)
firsttabfin<-firsttabadj[1:3]
```

```
secondpopulation=subset(POP2,select=SNP)
secondpopulationmatrix1=as.matrix(secondpopulation)
secondpop=na.omit(secondpopulationmatrix1)
secondpop_stand=c(" ")
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[1])
secondpop_stand[i]="AA"}
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[2])
secondpop_stand[i]="AB"}
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[3])
secondpop_stand[i]="BB"}
secondpopulationmatrix=as.matrix(secondpop_stand)
secondgeno<-genotype(secondpopulationmatrix,sep="")
secondtab<-table(secondgeno)
secondtabadj<-c(secondtab,0,0,0)
secondtabfin<-secondtabadj[1:3]
conttable=matrix(c(firsttabfin,secondtabfin),3)
fisher.test(conttable)$p.value
}

genotypetabfin=function(SNP,ALLELE,POP1){
firstpopulation=subset(POP1,select=SNP)
firstpopulationmatrix1=as.matrix(firstpopulation)
firstpop_stand=c(" ")
firstpop=na.omit(firstpopulationmatrix1)
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[1])
firstpop_stand[i]="AA"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[2])
firstpop_stand[i]="AB"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[3])
firstpop_stand[i]="BB"}
firstpopulationmatrix=as.matrix(firstpop_stand)
firstgeno<-genotype(firstpopulationmatrix,sep="")
firsttab<-table(firstgeno)
firsttabadj<-c(firsttab,0,0,0)
firsttabfin<-firsttabadj[1:3]
firsttabfin
}

allelefisher=function(SNP,ALLELE,POP1,POP2){
firstpopulation=subset(POP1,select=SNP)
firstpopulationmatrix1=as.matrix(firstpopulation)
firstpop=na.omit(firstpopulationmatrix1)
```

```
firstpop_stand=c(" ")
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[1])
firstpop_stand[i]="AA"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[2])
firstpop_stand[i]="AB"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[3])
firstpop_stand[i]="BB"}
firstpopulationmatrix=as.matrix(firstpop_stand)
firstgeno<-genotype(firstpopulationmatrix,sep="")
firsttab<-table(firstgeno)
firsttabadj<-c(firsttab,0,0,0)
firsttabfin<-firsttabadj[1:3]
secondpopulation=subset(POP2,select=SNP)
secondpopulationmatrix1=as.matrix(secondpopulation)
secondpop=na.omit(secondpopulationmatrix1)
secondpop_stand=c(" ")
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[1])
secondpop_stand[i]="AA"}
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[2])
secondpop_stand[i]="AB"}
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[3])
secondpop_stand[i]="BB"}
secondpopulationmatrix=as.matrix(secondpop_stand)
secondgeno<-genotype(secondpopulationmatrix,sep="")
secondtab<-table(secondgeno)
secondtabadj<-c(secondtab,0,0,0)
secondtabfin<-secondtabadj[1:3]
alleletab<-
rbind(c((2*firsttabfin[1]+firsttabfin[2]),(2*firsttabfin[3]+firstt
abfin[2])),c((2*secondtabfin[1]+secondtabfin[2]),(2*secondtabfin[3
]+secondtabfin[2])))
fisher.test(alleletab)$p.value
}

alleletabfin=function(SNP,ALLELE,POP1){
firstpopulation=subset(POP1,select=SNP)
firstpopulationmatrix1=as.matrix(firstpopulation)
firstpop=na.omit(firstpopulationmatrix1)
firstpop_stand=c(" ")
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[1])
firstpop_stand[i]="AA"}
```

```
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[2])
firstpop_stand[i]="AB"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[3])
firstpop_stand[i]="BB"}
firstpopulationmatrix=as.matrix(firstpop_stand)
firstgeno<-genotype(firstpopulationmatrix,sep="")
firsttab<-table(firstgeno)
firsttabadj<-c(firsttab,0,0,0)
firsttabfin<-firsttabadj[1:3]
alleletab<-
c((2*firsttabfin[1]+firsttabfin[2]),(2*firsttabfin[3]+firsttabfin[
2]))
alleletab
}

hwepvalue=function(SNP,ALLELE,POP1){
firstpopulation=subset(POP1,select=SNP)
firstpopulationmatrix1=as.matrix(firstpopulation)
firstpop=na.omit(firstpopulationmatrix1)
firstpop_stand=c(" ")
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[1])
firstpop_stand[i]="AA"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[2])
firstpop_stand[i]="AB"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[3])
firstpop_stand[i]="BB"}
firstpopulationmatrix=as.matrix(firstpop_stand)
firstgeno<-genotype(firstpopulationmatrix,sep="")
HWE.exact(firstgeno)$p.value
}
allelefisherod=function(SNP,ALLELE,POP1,POP2){
firstpopulation=subset(POP1,select=SNP)
firstpopulationmatrix1=as.matrix(firstpopulation)
firstpop=na.omit(firstpopulationmatrix1)
firstpop_stand=c(" ")
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[1])
firstpop_stand[i]="AA"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[2])
firstpop_stand[i]="AB"}
for(i in 1:length(firstpop)) {
if(firstpop [i]==ALLELE[3])
```

```
firstpop_stand[i]="BB"}
firstpopulationmatrix=as.matrix(firstpop_stand)
firstgeno<-genotype(firstpopulationmatrix,sep="")
firsttab<-table(firstgeno)
firsttabadj<-c(firsttab,0,0,0)
firsttabfin<-firsttabadj[1:3]
secondpopulation=subset(POP2,select=SNP)
secondpopulationmatrix1=as.matrix(secondpopulation)
secondpop=na.omit(secondpopulationmatrix1)
secondpop_stand=c(" ")
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[1])
secondpop_stand[i]="AA"}
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[2])
secondpop_stand[i]="AB"}
for(i in 1:length(secondpop)) {
if(secondpop [i]==ALLELE[3])
secondpop_stand[i]="BB"}
secondpopulationmatrix=as.matrix(secondpop_stand)
secondgeno<-genotype(secondpopulationmatrix,sep="")
secondtab<-table(secondgeno)
secondtabadj<-c(secondtab,0,0,0)
secondtabfin<-secondtabadj[1:3]
alleletab<-
rbind(c((2*secondtabfin[1]+secondtabfin[2]),(2*secondtabfin[3]+sec
ondtabfin[2])),c((2*firsttabfin[1]+firsttabfin[2]),(2*firsttabfin[
3]+firsttabfin[2])))
as.vector(fisher.test(alleletab)$estimate)
}

#==========GENOTYPE COUNT 3.0========
genotypetabfin_all=function(SNP,SNP_ALLELE,CASEPOP,CONTROLPOP){
obj1=matrix(0,nrow=(length(CONTROLPOP)),ncol=3)
obj2=matrix(0,nrow=(length(CASEPOP)),ncol=3)
for(j in 1:length(CONTROLPOP)) {
obj1[j,]=eval(call("genotypetabfin",SNP,SNP_ALLELE,as.symbol(CONTR
OLPOP[j])))
}
for(i in 1:length(CASEPOP)) {
obj2[i,]=eval(call("genotypetabfin",SNP,SNP_ALLELE,as.symbol(CASEP
OP[i])))
}
result=rbind(obj1,obj2)
colnames(result)=c("AA","Aa","aa")
rownames(result)=c(CONTROLPOP,CASEPOP)
result}
```

```
#usage: genotypetabfin_all("snp_3",snp_3allele,casepop,controlpop)

#========GENOTYPE TEST COUNT ======
genotypefisher_all=function(SNP,SNP_ALLELE,CASEPOP,CONTROLPOP){
obj1=matrix(0,ncol=(length(CASEPOP)),nrow=(length(CONTROLPOP)))
for(i in 1:length(CASEPOP)){
for(j in 1:length(CONTROLPOP)){
obj1[j,i]=eval(call("genotypefisher",SNP,SNP_ALLELE,as.symbol(CASE
POP[i]),as.symbol(CONTROLPOP[j])))
}
}
print("======SUPER FISHERS TEST  OUTPUT  :)=======")
colnames(obj1)=CASEPOP
rownames(obj1)=CONTROLPOP
obj1}
# usage:
genotypefisher_all("snp_3",snp_3allele,casepop,controlpop)

#=========ALLELE COUNT 3.0============
alleletabfin_all=function(SNP,SNP_ALLELE,CASEPOP,CONTROLPOP){
obj1=matrix(0,nrow=(length(CONTROLPOP)),ncol=2)
obj2=matrix(0,nrow=(length(CASEPOP)),ncol=2)
for(j in 1:length(CONTROLPOP)) {
obj1[j,]=eval(call("alleletabfin",SNP,SNP_ALLELE,as.symbol(CONTROL
POP[j])))
}
for(i in 1:length(CASEPOP)) {
obj2[i,]=eval(call("alleletabfin",SNP,SNP_ALLELE,as.symbol(CASEPOP
[i])))
}
result=rbind(obj1,obj2)
colnames(result)=c("A","a")
rownames(result)=c(CONTROLPOP,CASEPOP)
result}

# usage: alleletabfin_all("snp_4",snp_4allele,casepop,controlpop)

#===========ALLELE TEST 3.0==============
allelefisher_all=function(SNP,SNP_ALLELE,CASEPOP,CONTROLPOP){
obj1=matrix(0,ncol=(length(CASEPOP)),nrow=(length(CONTROLPOP)))
for(i in 1:length(CASEPOP)){
for(j in 1:length(CONTROLPOP)){
obj1[j,i]=eval(call("allelefisher",SNP,SNP_ALLELE,as.symbol(CASEPO
P[i]),as.symbol(CONTROLPOP[j])))
}
}
print("========SUPER ALLELE FISHER's TEST  OUTPUT  :)======")
colnames(obj1)=CASEPOP
```

```
rownames(obj1)=CONTROLPOP
obj1}
# usage: allelefisher_all("snp_3",snp_3allele,casepop,controlpop)

#==========ALLELE TEST WITH ODD RATIO ==============
allelefisherwithod_all=function(SNP,SNP_ALLELE,CASEPOP,CONTROLPOP)
{
obj1=matrix(0,ncol=2*(length(CASEPOP)),nrow=length(CONTROLPOP))
column_name=rep(0,times=2*length(CASEPOP))
for(i in 1:length(CASEPOP)){
column_name[(2*i-1)]=CASEPOP[i]
column_name[(2*i)]=c("OR")
for(j in 1:length(CONTROLPOP)){
obj1[j,(2*i-
1)]=eval(call("allelefisher",SNP,SNP_ALLELE,as.symbol(CASEPOP[i]),
as.symbol(CONTROLPOP[j])))
obj1[j,(2*i)]=eval(call("allelefisherod",SNP,SNP_ALLELE,as.symbol(
CASEPOP[i]),as.symbol(CONTROLPOP[j])))
}
}
print("======SUPER ALLELE FISHER's TEST with ODD RATIO OUTPUT
:)===")
colnames(obj1)=column_name
rownames(obj1)=CONTROLPOP
obj1}
# usage:
allelefisherwithod_all("snp_4",snp_4allele,casepop,controlpop)

#==============HWEPVALUE 3.0================
hwepvalue_all=function(SNP,SNP_ALLELE,CASEPOP,CONTROLPOP){
obj1=matrix(0,nrow=(length(CONTROLPOP)),ncol=1)
obj2=matrix(0,nrow=(length(CASEPOP)),ncol=1)
for(j in 1:length(CONTROLPOP)) {
obj1[j,]=eval(call("hwepvalue",SNP,SNP_ALLELE,as.symbol(CONTROLPOP
[j])))
}
for(i in 1:length(CASEPOP)) {
obj2[i,]=eval(call("hwepvalue",SNP,SNP_ALLELE,as.symbol(CASEPOP[i]
)))
}
result=rbind(obj1,obj2)
colnames(result)=c("HWE p values")
rownames(result)=c(CONTROLPOP,CASEPOP)
print("============HWE P VALUE OUTPUT  :)============")
result}
#usage: hwepvalue_all("snp_3",snp_3allele,casepop,controlpop)
```

Appendix II: Linux codes used for retrieving raw sequencing data from the 1000 genome project and calling for indels using SAMTools, Dindel and BEDTools programs

```
#!/bin/sh

BASE="ftp://ftp.ebi.ac.uk/pub/databases/microarray/data/experiment
/GEUV/E-GEUV-1/processed/"
REGION="chr9:35072459-35072659"
REFERENCE_FILE="chr9.fa"
OUTPUTFOLDER="VCPHEXSAM2test"
DINDEL_WINDOW="120"
MAXIMUM_DISTANCE_BETWEEN_WINDOW="0"
FILTER_QUALITY="0"
BEDFILE="vcphex.bed"

INDIVIDUAL="\
HG00096.1.M_111124_6 \
HG00097.7.M_120219_2 "

mkdir ${OUTPUTFOLDER}

for c in $INDIVIDUAL ; do
      samtools/samtools view -h ${BASE}/${c}.bam ${REGION} >
      ${OUTPUTFOLDER}/${c}.sam
      samtools/samtools view -bS ${OUTPUTFOLDER}/${c}.sam >
      ${OUTPUTFOLDER}/${c}.bam
      samtools/samtools sort ${OUTPUTFOLDER}/${c}.bam
      ${OUTPUTFOLDER}/${c}sorted
      samtools/samtools index ${OUTPUTFOLDER}/${c}sorted.bam
      rm -f -r ${OUTPUTFOLDER}/${c}.sam
      rm -f -r ${OUTPUTFOLDER}/${c}.bam
      rm -f -r ${OUTPUTFOLDER}/${c}.bam.bai
done


for c in $INDIVIDUAL ; do
      mkdir ${OUTPUTFOLDER}/${c}windows
      mkdir ${OUTPUTFOLDER}/${c}glf

      dindel/dindel --analysis getCIGARindels --bamFile
      ${OUTPUTFOLDER}/${c}sorted.bam \
      --outputFile ${OUTPUTFOLDER}/${c}.dindel_output --ref
      ${REFERENCE_FILE}

      dindel/makeWindows.py -i
      ${OUTPUTFOLDER}/${c}.dindel_output.variants.txt \
```

```
      -w ${OUTPUTFOLDER}/${c}windows/${c}.sample.realign_windows -m
      ${MAXIMUM_DISTANCE_BETWEEN_WINDOW} -n ${DINDEL_WINDOW}

echo Number of Windows is $(seq 1 $(find
${OUTPUTFOLDER}/${c}windows -type f | wc -l))

            for d in $(seq 1 $(find ${OUTPUTFOLDER}/${c}windows -
            type f | wc -l)); do
                  dindel/dindel --analysis indels \
                  --doDiploid \
                  --bamFile ${OUTPUTFOLDER}/${c}sorted.bam \
                  --ref ${REFERENCE_FILE} \
                  --varFile
                  ${OUTPUTFOLDER}/${c}windows/${c}.sample.realign_wi
                  ndows.${d}.txt \
                  --libFile
                  ${OUTPUTFOLDER}/${c}.dindel_output.libraries.txt \
                  --outputFile
                  ${OUTPUTFOLDER}/${c}glf/${c}.sample.dindel.stage_2
                  _output_windows.${d}
            done ;

find ${OUTPUTFOLDER}/${c}glf -type f >
${OUTPUTFOLDER}/${c}.glffiles.txt

      dindel/mergeOutputDiploid.py \
      --inputFiles ${OUTPUTFOLDER}/${c}.glffiles.txt \
      --outputFile {OUTPUTFOLDER}/${c}.glffiles.txtvariantCalls.txt \
      --refFile ${REFERENCE_FILE} \
      --filterQual ${FILTER_QUALITY}

      rm -f -r ${OUTPUTFOLDER}/${c}.dindel_output.variants.txt
      rm -f -r ${OUTPUTFOLDER}/${c}.dindel_output.libraries.txt
      rm -f -r ${OUTPUTFOLDER}/${c}windows
      rm -f -r ${OUTPUTFOLDER}/${c}glf
      rm -f -r ${OUTPUTFOLDER}/${c}.glffiles.txt
      rm -f -r ${c}.bam.bai
done;


mkdir ${OUTPUTFOLDER}/coverageBed_result
for c in $INDIVIDUAL ; do
      bedtools/bin/bedtools coverage \
      -abam ${OUTPUTFOLDER}/${c}sorted.bam \
      -b ${BEDFILE} >
      ${OUTPUTFOLDER}/coverageBed_result/${c}sorted.bed_coverage.txt;
done;
```

```
find ${OUTPUTFOLDER}/coverageBed_result/*.txt -type f -exec cat {}
>> ${OUTPUTFOLDER}/coverageBed_result/Combined_Coverage.txt \;
```

Appendix III: R Codes used for re-locating a list of VNTR (Kozlowski 2010) to the reference genome (hg19) using BSgenome package and the cross-referencing with GWAS data.

```
#=============LOAD DATA=============
chio_file<-"E:/08ALS/VNTRs/Kozlowski 2010/phs000101.pha002846.txt"
#defining the link of ped file
chio_file<-read.delim(file=chio_file,header=T,sep="\t")

repeat_file<-"E:/08ALS/VNTRs/Kozlowski 2010/nar-02659-n-2009-
File008.txt" #defining the link of ped file
repeat_file<-read.delim(file=repeat_file,header=T,sep="\t")

repeat_intergenic_file<-"E:/08ALS/VNTRs/Kozlowski 2010/nar-02659-
n-2009-File007.txt" #defining intergenic repeat file
repeat_intergenic_file<-
read.delim(file=repeat_intergenic_file,header=T,sep="\t")

track_file<-"E:/08ALS/VNTRs/Kozlowski 2010/prediction_track.txt"
#defining the link of ped file
track_file<-read.delim(file=track_file,header=T,sep="\t")

#=============Load BSgenome=============
source("http://www.bioconductor.org/biocLite.R")
biocLite("BSgenome")
library(BSgenome)
biocLite("BSgenome.Hsapiens.UCSC.hg19") #installs the human genome
(~850 MB download).
library(BSgenome.Hsapiens.UCSC.hg19)

#=============For each Chromosome========
print("Chr1")

chio_file_chr=chio_file[which(chio_file$chr==1),]
repeat_intergenic_file_chr=repeat_intergenic_file[which(repeat_int
ergenic_file$chr==1),]

track_file_chr=track_file[which(track_file$chr==1),]

table1=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#P value
table2=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#Rs number
table3=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#Position
```

```
table4=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#Ref Seq Gene Start
table5=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#Ref Seq Gene end
table6=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#No. of Significant SNPs
table7=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#matched position
table8=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#matched gene_name
table9=matrix(0,nrow=length(repeat_intergenic_file_chr$position),n
col=1)#matched number

for (i in 1:length(repeat_intergenic_file_chr$position)){
      rep_unit=repeat_intergenic_file_chr$rep_unit[i]
      period=repeat_intergenic_file_chr$period[i]
      repeat_sequence=paste(rep(rep_unit,each=period),collapse="")
      repeat_sequence_reverse=as.character(reverseComplement(DNAStr
      ing(repeat_sequence)))
      repeat_start=c(start(matchPattern(repeat_sequence,Hsapiens$ch
      r1)),start(matchPattern(repeat_sequence_reverse,Hsapiens$chr1
      )))
      matched_position_all=repeat_start[repeat_start<repeat_interge
      nic_file_chr$position[i]+2000000 &
      repeat_start>repeat_intergenic_file_chr$position[i]-2000000]
      matched_position=matched_position_all[which.min(abs(matched_p
      osition_all-repeat_intergenic_file_chr$position[i]))]
      gene_name=track_file_chr[which(track_file_chr$start<matched_p
      osition & track_file_chr$end>matched_position),]$gene_id
      gene_start=min(track_file_chr[which(as.character(track_file_c
      hr$gene_id)==as.character(gene_name)),]$start)
      gene_end=max(track_file_chr[which(as.character(track_file_chr
      $gene_id)==as.character(gene_name)),]$end)
      pvalue=chio_file_chr[which(chio_file_chr$position<matched_pos
      ition+100000 & chio_file_chr$position>matched_position-
      100000),]$p_value
      rs=chio_file_chr[which(chio_file_chr$position<matched_positio
      n+100000 & chio_file_chr$position>matched_position-
      100000),]$snp_id
      rsid=rs[which.min(pvalue)]
      position=chio_file_chr[which(chio_file_chr$position<matched_p
      osition+100000 & chio_file_chr$position>matched_position-
      100000),]$position
      positionid=position[which.min(pvalue)]
      significant_pvalue=chio_file_chr[which(chio_file_chr$position
      <matched_position+100000 &
      chio_file_chr$position>matched_position-100000 &
      chio_file_chr$p_value<0.05),]$p_value
```

```
      table1[i,]=min(pvalue)
      table2[i,]=as.character(rsid[1])
      table3[i,]=as.character(positionid[1])
      table4[i,]=as.character(gene_start)
      table5[i,]=as.character(gene_end)
      table6[i,]=length(as.vector(significant_pvalue))
      table7[i,]=as.character(matched_position[1])
      table8[i,]=as.character(gene_name[1])
      table9[i,]=length(matched_position_all)
      }
result1=cbind(table1,table2,table3,table4,table5,table6,table7,tab
le8,table9)
result1
```

-END OF Thesis-