Summer 8-15-2015

Genetic Factors that Contribute to the Pathogenesis of Amyotrophic Lateral Sclerosis

Janet Elizabeth Cady
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Division of Biology and Biomedical Sciences
Human and Statistical Genetics

Dissertation Examination Committee:
Matthew Harms, Chair
Alison Goate, Co-Chair
Joseph Dougherty
Christina Gurnett
Timothy Miller
John Rice

Genetic Factors That Contribute to the Pathogenesis of Amyotrophic Lateral Sclerosis

by

Janet Elizabeth Cady

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2015
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would first like to thank my advisor, Dr. Matthew Harms for his support and guidance over the past several years. Matt gave me great opportunities to work independently and try new things which allowed me to learn a great deal in my time in lab. I would like to thank my co-mentor Dr. Alison Goate. This was a process that was new to both Matt and me and Alison was always there to keep us on track.

I would also like to thank the rest my thesis committee, Dr. Timothy Miller, Dr. Christina Gurnett, Dr. Joseph Dougherty, and Dr. John Rice for all of their valuable input. I would especially like to thank Dr. Miller for chairing my committee and helping me navigate the whole dissertation process.

Several current and former members of the Harms lab and the Baloh lab contributed to this work. Jarod Shelton has been a great lab-mate and was always willing to help me out on top of his own work. Jarod, along with Paul Cooper and Shaughn Bell before him, prepared DNA from all of the patient samples that came into lab and screened them all for C9ORF72 repeat expansions. Carolyn Hilliard was responsible for the majority of the ATXN2 repeat sizing.

This project wouldn't have been possible without the help of many other labs as well. The Goate and Cruchaga labs helped in many aspects including providing scripts for analyzing pooled sequencing data, genotyping variants, allowing us to borrow equipment, and sharing ideas. The Miller lab kindly allowed me to use their space for radioactive experiments; Amy Wegener and Kathleen Schoch in particular were a great help. Erica Koval from the Miller lab also performed the TREM2 expression studies. The Southern blot project would not have been possible without Renate Lewis who taught me the protocol and gave me plentiful

troubleshooting advice. Thank you to Gabe Haller and Dave Alvarado from the Gurnett lab for their help with exome sequencing analysis.

This work would also not have been possible without my funding from the NIH through the Genetics Epidemiology T32 training grant (5-T32-HL-83822-5).

I would like to thank my friends for being around to make sure I had some fun. The only people that understand what it's like to be in graduate school are you fellow students. I am so lucky I had all of you around to share our experiences.

Lastly I want to thank my family for being so supportive. My parents and my brothers always found the time to come visit me and listen to me talk about science even though they only understand half of what I say. Thank you for everything!

Janet Cady

Washington University in St. Louis

August 2015

ABSTRACT OF THE DISSERTATION

Genetic Factors That Contribute to the Pathogenesis of Amyotrophic Lateral Sclerosis

By

Janet Elizabeth Cady

Doctor of Philosophy in Biology and Biomedical Sciences

Human and Statistical Genetics

Washington University in St. Louis, 2015

Assistant Professor Matthew Harms, Chair

Professor Alison Goate, Co-Chair

Amyotrophic lateral sclerosis (ALS) is fatal neurodegenerative disease for which there is no cure. The only treatment available extends survival by only a matter of months. There are over 20 genes that are known to cause ALS. Over half of the ALS cases with a family history of disease (FALS) can be explained by mutations in known ALS genes with hexanucleotide repeat expansions in *C9ORF72* accounting for 40% of families. However roughly 90% of cases have no family history of disease (sporadic ALS or SALS) and a much smaller proportion (10%) of these cases can be explained by mutations in known ALS genes. Understanding the genetic factors that cause ALS or influence its progression will help us understand the cellular pathways involved in disease and identify potential therapeutic targets.

We used a pooled-sample sequencing approach to identify mutations in 17 ALS genes in a cohort of FALS and SALS patients to investigate the contribution of these genes to SALS, including the role of rare variants and the effect of mutations in multiple ALS genes in an individual. We identified potentially pathogenic mutations in 64.3% of familial and 27.8% of

sporadic subjects. 3.8% of subjects had mutations in more than one ALS gene and these individuals on average had onset 10 years earlier than those with mutations in only one ALS gene (p=0.0046). There were no individual rare variants that were significantly associated with sporadic ALS, but rare variants in *SOD1* were cumulatively more common in SALS subjects.

In addition we investigated the genetic background and stability of *C9ROF72* repeat expansions in ALS. The presence of a risk haplotype shared between all expansion-carriers led to the prevailing idea of a founder expansion event, however this shared haplotype also supports the hypothesis of a genetic background that is more prone to expansion. We identified a rare variant rs147599399 on this genetic background that is present in some expansion carriers and some non-expansion carriers, indicating that the expansion arose on at least two separate occasions. This raises the possibility that *C9ORF72* repeat expansions in sporadic ALS could be the result of *de novo* expansions on the risk haplotype. Furthermore we showed that expansion carriers with rs147599399 minor allele had longer survival than expansion carriers without the SNP (p=0.00047), indicating that the genetic background surrounding the *C9ORF72* influences the effects of the expansion.

We performed Southern blotting to explore the size and stability of *C9ORF72* repeat expansions. There was a high degree of somatic instability and instability in transmissions between families. There was no difference between expansion sizes in symptomatic and asymptomatic expansion carriers in families an there was no correlation between expansion size in any patient tissues and any clinical characteristics. These results need to be confirmed in a larger sample cohort, but suggest that expansion size alone doesn't determine pathogenicity of *C9ORF72* repeat expansions.

Lastly we examined the candidate gene *TREM2* as a risk factor for ALS. This gene is involved in regulation of microglial activity, which is a known component of ALS pathogenesis, and the rare variant p.R47H was recently associated with risk Alzheimer's disease. We found that the same p.R47H variant was significantly associated with ALS in our cohort and that expression of TREM2 was increased in ALS patients and SOD1 mutant mice compared to controls. A variant in the related gene *TREML4* was marginally associated with ALS, but the effect of this variant is unknown. Mutations in the TREM genes provide a genetic link between to the neuro-inflammatory component of ALS and suggest other genes involved in microglial activation are good candidates for novel variant identification.

Chapter 1

Background and significance

1.1 Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease is the most common form of motor neuron disease. Motor neurons in the brain and spinal cord are affected resulting in loss of motor function, paralysis, muscle wasting, and difficulty breathing and swallowing. Death typically occurs after 3-5 years due to respiratory failure.[1–5] The disease can have either limb onset, affecting the arms or legs first or bulbar onset, affecting speech, swallowing and breathing first. There is no cure for ALS. Riluzole is the only available treatment and only slows the progression of symptoms.[6]

Average age of symptom onset ranges from 40-70, however this can vary greatly with cases in their 20's and even cases of juvenile onset.[7–12] ALS is diagnosed according to the El Escorial criteria[13] which include the presence of signs of upper motor neuron degeneration (weakness, wasting, and fasciculations), signs of lower motor neuron degeneration (increased reflexes, spasticity), progressive spread of symptoms, and the absence of other disease that could explain the clinical findings. The diagnosis can be definite, probable, possible, or suspected depending on the strength of clinical evidence.

Disease progression is typically measured using the ALS functional rating scale (ALSFRS)[14]. This scale ranks the patient's functional impairment using a five-point scale on 12 different tasks with 4 being no loss of function and 0 being complete loss of function. These

score are summed across all 12 tasks to obtain the ALSFRS score. The highest possible score is 48 and the lowest is 0.

Cognitive changes are common in ALS ranging from mild cognitive impairment to frontotemporal dementia (FTD). ALS and FTD, also called frontotemporal lobar degeneration (FTLD), are considered to fall on a spectrum. FTD is present is roughly 25% of cases of ALS while about 50% of FTD cases also have motor neuron degerentation.[15] Even within families, some members can develop either pure ALS, ALS with FTD or pure FTD.

Estimates of the prevalence vary by population with a systematic literature review reporting a range of 1-11.3/100,000 (median =4.48).[16] Prevalence in the United States was 2.9-3.9/100,000[17–19]. Incidence ranged from 0.3-3.6 (median=1.9) worldwide with two studies in the U.S reporting 1.7 and 1.8.[16] When comparing retrospective and prospective studies, prospective studies tended to result in higher estimates of both prevalence and incidence.[16]

There is a family history of disease (typically with an affected first or second degree relative) in 5-10% of cases (FALS)[20–22], although this can difficult to ascertain. The late onset of disease increases the likelihood of death from unrelated causes before developing symptoms, obscuring possible inheritance patterns and there are no clinical distinctions between patients with and without a family history. Other factors such as non-paternity, small family size, incomplete penetrance, and incomplete family history can make familial cases appear to be sporadic (SALS). The heritability of SALS is estimated to be 61% in a twin study[23], but estimates were much lower (12-21%) in genome-wide association studies (GWAS). [24,25]

1.2 Genetics of familial ALS

1.2.1 *SOD1*

The first genetic cause of ALS to be discovered was the Cu/Zn superoxide dismutase *SOD1*. Linkage to the chr21q region was shown in ALS families[26,27] where the causative gene was later discovered to be *SOD1* with 11 different mutations identified in 13 families[28]. There are now 175 distinct mutations in *SOD1* that have been submitted to the ALS mutation database ALSoD[29] (http://alsod.iop.kcl.ac.uk/). These mutations are distributed across the entire gene, including some codons that have many different variants. For instance, six different mutation of the glycine at amino acid position 94 have been reported (p.G94A, p.G94S, p.G94V, p.G94R, p.G94D, p.G94C).

All reported mutations in SOD1 are dominant or dominant with reduced penetrance with the exception p.D91A which is recessive[30] or compound heterozygous with p.D96N.[31] Multiple studies have identified heterozygous p.D91A mutations as well[32,33] suggesting that the mutation can act in a dominant fashion as well. A proposed mechanism was that all recessive p.D91A mutations share a cis-acting protective factor that is absent from dominant p.D91A mutations.[34,35]

Mutations in SOD1 account for roughly 20% of FALS cases and 2% of SALS. Estimate of *SOD1* contribution to FALS in different populations ranged from 1.8% in the Netherlands to 50% in Scotland (only including studies with 10 or more subjects).[36] One study in an Irish population found no mutations in *SOD1*.[37] In SALS populations, *SOD1* accounted for 0-11.3% of cases.[36] In the United States, 23.4% of FALS cases had mutations in *SOD1*[38]. There were no studies in the US looking specifically at the contribution of *SOD1* in SALS, but analyses of mixed SALS/FALS cohorts showed *SOD1* mutations in around 7.5% overall.[39,40]

The protein product of the *SOD1* gene is responsible for converting toxic superoxide anions, a byproduct of oxidative phosphorylation, to hydrogen peroxide. SOD1 protein is

ubiquitously expressed and is predominantly located in the cytoplasm. Mutations in *SOD1* result in protein mis-folding and aggregations.[41,42] The cellular effects of mutant SOD1 are widespread including dysregulation of autophagy[43], ER stress[44], impaired axonal transport[45], and disruption of mitochondrial function.[46]

1.2.2 *TARDBP* and *FUS*

TAR DNA-binding protein (*TARDBP*) and fused-in sarcoma/translated in liposarcoma (*FUS/TLS* or *FUS*) each account for about 4% of FALS and 1% of SALS.[47] Both proteins are structurally similar; they each possess a nuclear localization signal (NLS), glycine-rich domains, RNA-repeat binding motifs (RRMs) and prion-like domains. Both TDP-43 (the protein product of *TARDBP*) and FUS can bind to DNA and RNA and are involved in RNA transcription, splicing, and transport.[48] TDP-43 and FUS are normally located in the nucleus but can also translocate to the cytoplasm. Ubiquitinated TDP-43 and FUS are found in cytoplasmic inclusions in ALS patients.[48]

TARDBP was first selected as a candidate due to the presence of TDP-43 in cytoplasmic inclusions in ALS patients.[49] Sequencing of the gene in patients identified p.A315T[50], and six additional variants[51] as a pathogenic mutations. A total of 52 mutations in *TARDBP* have been reported in ALSoD, most of which lie in the C-terminal glycine-rich domain. The C-terminal domain interacts with heterogeneous nuclear ribonucleoproteins (hnRNP) to modulate alternative-splicing of mRNA transcripts.[52]

Mutations in *FUS* were first discovered in families with linkage[53] or loss of heterozygosity[54] on chromosome 16. There are currently 79 *FUS* mutations listed in ALSoD. These mutations are predominantly located in either the C-terminal NLS or in the glycine-rich

domain.[55,56] Mutations in the C-terminal domain were shown to prevent FUS from entering the nucleus resulting in the formation of cytoplasmic aggregates.[57]

1.2.3 *C9ORF72*

Linkage to the chromosome 9p.21 region was first shown in 2006 in a large Scandinavian family[58]. Additional linkage studies in Dutch[59], Australian[60], French[61], Belgian[62], Welsh[63], North American[64], and French Canadian[65] families identified the same locus. Genome-wide association studies in sporadic ALS subjects showed association with the same region[66,67], however none of these groups were able to identify causative gene. The underlying cause of chromosome 9p.21 ALS wasn't discovered until 2011 when DeJesus Hernazdez *et al.*[68] and Renton *et al.*[69] both identified a hexanucleotide (GGGGCC) repeat expansion in a non-coding region of *C9ORF72* (Figure 1.1).

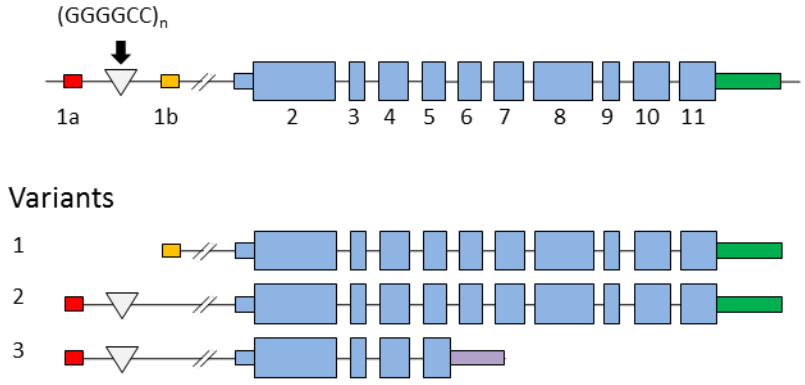

Figure 1.1 The *C9ORF72* gene and its three transcript variants. The gene has two alternate 5'UTR exons (1a and 1b). Transcript 1 uses exon 1b and the expansion lies in the upstream promoter region. Transcript variants 2 and 3 use exon 1a and the repeat is in the intron between exon 1b and exon 2.

Non-pathogenic alleles of the *C9ORF72* repeat are most commonly 2 repeats with a maximum around 25 repeats.[68,70,71] The pathogenic size of repeat expansions is still being

determined, but there is a clear distinction between expanded and unexpanded alleles as determined by repeat-primed PCR (RP-PCR). Southern blots must be used to size expansions larger than 30 repeats and show expansions of roughly 400-8000 repeats.[68,72,73] Expansions in the intermediate range of larger than 20 repeats but smaller than full expansions are seen as well, but their role in pathogenesis is unclear.[72,74–76]

Full repeat expansions are found in 40% of FALS and 7% of SALS cases world-wide.[77] Prevalence of the repeat expansion varies greatly by population with the highest rates in the Finnish population (21% of SALS)[69,77] and lowest rates in Asian populations (0-1.5% of SALS in Japan, China, Korea, and Taiwan).[77–83] In the United States, the *C9ORF72* repeat expansion accounts for about 36% of FALS and 6% of SALS.[77] Full expansions are rare in control populations. Most studies do not find any full expansions in neurologically normal controls however a few have found expansions in a very small number of samples (0.15%-0.4%).[69,72,77]

The pathogenic mechanism of *C9ORF72* repeat expansion is still not fully understood, but is possibly a combination of RNA toxicity[68,84–86], loss of *C9ORF72* function[68], formation of G-quadruplex structures[87–89] and aggregation of di-peptide repeats produced through repeat-associated non-ATG initiated (RAN) translation.[90–92] It is unclear if the toxic effects of RNA foci and di-peptide aggregates are due to the aggregates themselves or the sequestrations of RNA-binding proteins[93,94] and proteins of the ubiquitin-proteasome pathway[95,96], preventing them from carrying out their normal functions. The normal function of C9ORF72 protein is not fully known, but it shares structural similarities with the DENN family of proteins involved in endosomal trafficking.[97,98]

1.2.4 Less common ALS genes

Mutations in over 15 other genes have been identified in ALS, each of which is the cause of less than 1% of ALS cases. The evidence regarding the role of these genes in summarized in table 1.1. The number of mutations reported in ALSoD is included in this table, however it is important to note that this database contains user submitted mutations, some of which are not functional (including intronic and synonymous variants and variants found in controls) and not all mutations that have been identified are submitted. With the exceptions of *UBQLN2, SETX* and *ALS2*, mutations in all of these genes were identified in typical adult onset ALS. Mutations in *SETX* were originally found in juvenile-onset ALS[99] but are also found in adult-onset cases.[37,100,101] *ALS2* was also identified in juvenile-onset ALS[102,103], however no association has been found between *ALS2* and typical ALS.[104]

Table 1.1 Additional genes that cause ALS.

Gene	Discovery Method	Protein function	Evidence	# ALSoD mutations
OPTN	Homozygosity mapping[105]	Mediation of apoptosis through TNFα/NFκB pathway[106], maintenance of Golgi complex/ vesicle trafficking[107], ubiquitin binding	Missense mutations results in neuronal cell death through NFκB pathway[108], truncation mutations unable to bind ubiquitin[105], autophagy-mediated degredation of misfolded proteins impaired[109]	37
ANG	Association, candidate gene[110–112]	Ribonuclease, ribosomal RNA synthesis[113]	Mutations result in decreased ribonuclease activity and decreased nuclear translocation[114]	29
DAO	Linkage analysis[115]	D-amino-acid oxidase, degredation of NMDA receptor agonist D-serine which accumulates with age	Mutation results in reduced enzymatic activity and increased apoptosis[115]	2
DCTN1	Candidate gene[116]	Forms complex which acts as motor for transport along microtubules	Disruption of DCNT1/DNCHC1 complex inhibits axonal transport in motor neurons[117]	7

FIG4	Candidate gene [118]	Phosphatase, regulation of membrane bound signaling lipid $PI(3,5)P_2$ [119]	Mutations reduce FIG4 activity resulting in enlarge vacuoles when transfected in yeast[118], mutations cause neurodegeneration in mouse[120]	10
EWSR1	Candidate gene [121]	RNA binding protein, same family as *FUS*	Mutations result in mis-localization from nucleus to cytoplasm in cultured cells.[121]	1
TAF15	Candidate gene[122,123]	RNA binding protein, same family as *FUS*	Mutations result in cytoplasmic foci in culture and patient spinal cord, cause neurodegeneration in fruit flies[123]	7
SETX	Linkage analysis[99]	DNA/RNA helicase, involved in pre-mRNA processing[124] and DNA damage response[125]	Mutation results in altered gene expression[126]	8
SQSTM1	Candidate gene[127]	Ubiquitin binding[128], mediation of NFκB signaling[129]	Aggregates of protein product p62 co-localized with TARDBP and FUS in ubiquitinated inclusions is ALS patients[130], Zebrafish knockdown has motor neuron phenotype[131]	16
VAPB	Linkage analysis[132]	Vesicle-binding protein, involved in ER functions including unfolded protein response (UPR)[133]	Mutations result in aggregate formation and neurodegenation in fruit flies[134] and induce ubiquitinated inclusions in mice[135], mutations disrupts UPR[133]	2
VCP	Linkage analysis, exome sequencing[136]	ATP-binding protein involved in vesicle transport, required for transport of proteins from ER to cytosol[137]	Transgenic mice have progressive weakness and ubiquitinated inclusions.[138]	7
ALS2	Linkage analysis[102,103]	Modulation of endosome[139]	Interacts with SOD1[140], axonal degeneration in mice[141]	27
UBQLN2	Linkage analysis[142]	Binds to proteasome and ubiquitinated proteins, delivers proteins marked for degradation to proteasome[142,143]	Present in ubiquitinated inclusions in spinal cord from ALS patients[142], mutations disrupt hnRNPA1 binding[144]	6

CHMP2B	Candidate gene[145]	Component of Endosomal Sorting Complex Required for Transport III (ESCRT-III), involved in recycling of cell surface receptors[146]	Mutations result in downregulation of genes involved in axonal transport and translation initiation factors, inhibition of autophagy[147]	6
PFN1	Exome sequening[148]	Regulation of actin dynamics[149]	Mutations result in insoluble ubiquitinated aggregates and inhibit axon growth[148]	4
TUBA4A	Rare variant burden analysis[150]	Microtubule subunit	Mutations result in aggregates and de-stabilization of microtubule networks[150]	12
MATR3	Exome sequencing[151]	Nuclear matrix protein involved in mRNA processing[152]	Mutations result in altered cellular distribution, shown to interact with TDP-43[151]	4
hnRNPA1, hnRNPA2B1	Linkage analysis, exome sequencing[153]	RNA-binding protein involved in RNA processing and shuttling between nucleus and cytoplasm[154]	Interact with TDP-43[52], mutations promote formation of fibrils, enhance recruitment of hnRNPA1, hnRNPA2 to stress granules[153]	2
ERBB4	Linkage analysis, exome sequencing[155]	Epidermal growth factor family of receptor tyrosine kinases, stimulated by neuregulin and involved in synaptic plasticity[156]	Expressed in motor neurons in rats[157], knock-out mice have defects in motor neuron development[158]	2

1.3 Genetics of sporadic ALS

1.3.1 Known FALS genes in SALS

As previously stated, the heritability of SALS is as high as 61%[23], meaning there is still a significant genetic component to sporadic disease. Mutations in many of the genes described above have been identified in SALS as well as in familial forms.[37,159–161] The majority of these

studies have focused only on the most commonly involved genes (*C9ORF72*, *SOD1*, *TARDBP*, and *FUS)* and the extent to which additional FALS genes cause SALS is relatively unexplored.

It is possible that mutations that have been found in SALS in Mendelian ALS genes are not fully penetrant and require additional genetic and environmental factors to cause disease. The idea of oligogenic inheritance, where mutations in more than one ALS gene are needed for disease to develop, has recently been investigated as a mechanism for sporadic ALS.[37,159,161] It has also been proposed that ALS is a multistep process requiring an average of 6 different steps or processes involved in disease development.[162]

Another possibility is that mutations may arise *de novo* in sporadic ALS patients. De-novo mutations have been reported in in both adult and juvenile onset SALS cases *FUS*,[163–165] *SOD1*[166], and *ERBB4.*[155] Exome sequencing of SALS trios with unaffected parents identified *de novo* mutations in the *CREST(SS18L1)* gene.[167] Identification of *de novo* mutations can be difficult because DNA from parents is often unavailable due to the late onset of disease.

1.3.2 Risk loci

In addition to the known FALS genes, intermediate length CAG repeat expansions in the *ATXN2* gene are associated with increased risk for sporadic ALS.[168] Expansions larger than 34 repeats are the cause of Spinocerebellar Ataxia type 2 (OMIM 183090), but CAG tracts in the intermediate range of 27-33 repeats were associated with SALS. A meta-analysis showed that the optimal cutoff to distinguish between ALS and controls was actually 29.[169]

Several genome-wide association studies (GWAS) have been performed in attempts to identify common genetic variants that increase the risk of sporadic ALS. The only consistently replicated genome-wide associations are with *C9ORF72* and *UNC13A*.[66,67,170,171] Associations

were identified in *DDP6* in an Irish population,[172] *FGGY* (FLJ10986) in a North American population of European descent[173], *ITPR2* in the Netherlands[174], and *ELP3* in the UK.[175] These signals have all failed to replicate in additional studies.[115–120] It is likely that this is due to genetic heterogeneity and that the effects of these variants is population specific.

With the failure of GWAS to identify variants that explain a significant portion of SALS, researchers have started investigating the role of rare variants in disease. Rare variants are more likely to have a larger effect size than common variants and while each individual variant may not contribute a significant amount to the heritability of disease, the cumulative effects of rare variants across entire genes can be significant. Associations of rare variants in SALS have been reported in the nicotinic acetylcholine receptors[182,183] and in protein disulfide isomerase genes.[184] A recent whole exome sequencing study identified an excess of variants in *TBK1*, which is involved in phosphorylation of OPTN and SQSTM1.[185]

1.3.3 Genes that modify phenotype

Three genes have been identified that modify ALS phenotype. A genome-wide association study showed that rs1541160 in *KIFAP3* was significantly associated with survival and reduced expression of *KIFAP3*.[186] Other studies have failed to replicate this association[187,188], so additional study is required to determine the significance of this gene in survival. The genome-wide susceptibility locus *UNC13A* was also found to be significantly associated with survival in two independent studies.[171,189] rs12608932 was the associated SNP in both studies which was the same SNP that was significantly associated with disease susceptibility. A morpholino screen of genes that rescued the motor neuron degeneration in

zebrafish model of *SOD1* identified *EPHA4.* Rare mutations in this gene were subsequently identified in a handful of individuals with unusually long survival from ALS.[190]

1.3.4 Environmental risk factors

With around 60% of ALS being inherited genetically, up to 40% could be accounted for by environmental causes. Many studies have investigated possible environmental exposures and the risk of ALS, however these studies often have methodological issues. There are hundreds of possible environmental factors to investigate and the importance of the timing during development of such exposures is unknown Some commonly studied factors include smoking, physical fitness, heavy metals, pesticides, service in the armed forces, and occupations.[191,192] These studies have produced conflicting results with some smoking studies, for instance, showing increased risk, some showing increased risk only in women, and some showing no increased risk.[191] High levels of physical fitness have also been reported as a risk factor for ALS with varying results depending on the measurement of physical fitness (i.e. prevalence of ALS in former varsity athletes[193] or population studies of occupational vs. leisure activity[194]). It was proposed that the increase in ALS risk with leisure activity but not with occupational activity means that an active lifestyle rather than fitness level is a risk factor.[194] Meta-analyses will is difficult or impossible due to differences in measurements of risk factors and outcomes between studies. It is also likely environmental risk factors are dependent upon genetic background, diluting our ability to identify relevant exposures. Increasing our understanding of the underlying genetics of the disease will allow us to determine biologically plausible environmental risk factors to perform a more directed analysis.

1.4 Neuroinflammation

The identification of genes that cause ALS has shed light on many of the pathogenic mechanisms of neurodegeneration, but pathological features of disease can provide us with candidate genes for novel variant detection. The inflammatory response in the nervous system is as important component of neurodegeneration[195] yet no mutations in genes that regulate this response have been identified in ALS patients. Microglia are the main immune cells of the CNS and have a vital role in ALS. Mice that express mutant SOD1 only in the motor neurons have delayed or no neurodegeneration compared to mice with ubiquitous mutant SOD1 expression.[196–198] Activated microglia are detected in sites of motor neuron injury in the motor cortex and spinal cord of ALS patients.[195,199,200]

In addition to responding to pathogens, microglia are required for clearance of cellular debris including injured neurons. Microglia have two different activation states: M1 which is cytotoxic and M2 which can be protective. It is unclear whether dysregulated microglial activation in ALS is detrimental due to a loss of the neuroprotective effects of removal of cellular debris or an over-activation of the cytotoxic effects resulting in damage to healthy neurons. There is evidence that microglia start in the neuroprotective state and transition to a neurotoxic state as disease progresses.[201] Genes that are involved in the mediation of microglial activity are excellent candidates for identification of variants that increase the risk of ALS.

1.5 Summary

Previous genetic studies in ALS have been vital to our understanding of disease pathogenesis. The genes that have been identified in families with ALS implicate several important pathways including oxidative stress, RNA processing, protein misfolding and aggregation, and inability to remove proteins targeted for degradation. Identification of novel

genetic causes of disease can help us find new pathways and can also refine our understanding of the pathways we already know.

Of particular interest is mechanism of the *C9ORF72* repeat expansion which is the most common cause of familial and sporadic ALS. There are several pathologic features associated with the repeat expansion, but it is unknown which ones are the driving force behind pathogenesis and which ones are byproducts of defects in other pathways. Understanding what characteristics of the repeat effect disease presentation (i.e. expansion size, somatic instability, and cis-acting modifiers) could give us clues about how they are causing ALS.

A higher than expected portion of SALS is caused by the hexanucleotide expansion in *C9ORF72* despite the fact that this mutation is also the most common cause of familial ALS. It is presently unclear if these sporadic *C9ORF72* cases are due to incomplete penetrance of the expansion and small family sizes, or if they are *de novo* expansions. The expansions are known to be unstable, therefore an asymptomatic “pre-expansion” carrier could pass on a pathogenic expansion. Distinguishing between these mechanisms of inheritance will be important to our understanding of how expansions cause disease.

Based on heritability studies in sporadic ALS, there is still a large genetic component of SALS that has yet to be discovered. The higher heritability estimates from twin studies compared to GWAS suggests that rare variants, which are not captured by GWAS play a significant role. Mutations in genes that cause FALS have also been found in sporadic ALS, but they do not account for a large portion of disease. The possibility of oligogenic inheritance in sporadic ALS has been supported by a few studies. Identification of additional cases with mutations in multiple ALS genes might reveal patterns in genes that are commonly mutated together which could isolate pathways that work together in neurodegeneration.

1.6 References

1. Chiò A, Logroscino G, Hardiman O, et al. Prognostic factors in ALS: A critical review. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2009;10(5-6):310–323.

2. Del Aguila MA, Longstreth WT, McGuire V, et al. Prognosis in amyotrophic lateral sclerosis: a population-based study. Neurology 2003;60(5):813–819.

3. Jordan H, Fagliano J, Rechtman L, et al. Effects of demographic factors on survival time after a diagnosis of amyotrophic lateral sclerosis. Neuroepidemiology 2015;44(2):114–120.

4. Jordan H, Fagliano J, Rechtman L, et al. Population-based surveillance of amyotrophic lateral sclerosis in New Jersey, 2009-2011. Neuroepidemiology 2014;43(1):49–56.

5. Mitsumoto H, Przedborski S, Gordon PH, editors. Amyotrophic lateral sclerosis. New York: Taylor & Francis; 2006.

6. Miller RG, Mitchell JD, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). Cochrane Database Syst. Rev. 2012;3:CD001447.

7. Testa D, Lovati R, Ferrarini M, et al. Survival of 793 patients with amyotrophic lateral sclerosis diagnosed over a 28-year period. Amyotroph. Lateral Scler. Mot. Neuron Disord. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2004;5(4):208–212.

8. Logroscino G, Traynor BJ, Hardiman O, et al. Incidence of amyotrophic lateral sclerosis in Europe. J. Neurol. Neurosurg. Psychiatry 2010;81(4):385–390.

9. Huisman MHB, de Jong SW, van Doormaal PTC, et al. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J. Neurol. Neurosurg. Psychiatry 2011;82(10):1165–1170.

10. Uenal H, Rosenbohm A, Kufeldt J, et al. Incidence and geographical variation of amyotrophic lateral sclerosis (ALS) in Southern Germany--completeness of the ALS registry Swabia. PloS One 2014;9(4):e93932.

11. Piemonte and Valle d'Aosta Register for Amyotrophic Lateral Sclerosis (PARALS). Incidence of ALS in Italy: evidence for a uniform frequency in Western countries. Neurology 2001;56(2):239–244.

12. Fang F, Valdimarsdóttir U, Bellocco R, et al. Amyotrophic lateral sclerosis in Sweden, 1991-2005. Arch. Neurol. 2009;66(4):515–519.

13. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular

Diseases and the El Escorial “Clinical limits of amyotrophic lateral sclerosis” workshop contributors. J. Neurol. Sci. 1994;124 Suppl:96–107.

14. Cedarbaum JM, Stambler N, Malta E, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J. Neurol. Sci. 1999;169(1-2):13–21.

15. Lillo P, Hodges JR. Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. J. Clin. Neurosci. Off. J. Neurosurg. Soc. Australas. 2009;16(9):1131–1135.

16. Chiò A, Logroscino G, Traynor BJ, et al. Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. Neuroepidemiology 2013;41(2):118–130.

17. Stickler DE, Royer JA, Hardin JW. Validity of hospital discharge data for identifying cases of amyotrophic lateral sclerosis. Muscle Nerve 2011;44(5):814–816.

18. Turabelidze G, Zhu B-P, Schootman M, et al. An epidemiologic investigation of amyotrophic lateral sclerosis in Jefferson County, Missouri, 1998-2002. Neurotoxicology 2008;29(1):81–86.

19. Mehta P, Antao V, Kaye W, et al. Prevalence of amyotrophic lateral sclerosis - United States, 2010-2011. Morb. Mortal. Wkly. Rep. Surveill. Summ. Wash. DC 2002 2014;63 Suppl 7:1–14.

20. Gros-Louis F, Gaspar C, Rouleau GA. Genetics of familial and sporadic amyotrophic lateral sclerosis. Biochim. Biophys. Acta 2006;1762(11-12):956–972.

21. KURLAND LT, MULDER DW. Epidemiologic investigations of amyotrophic lateral sclerosis. 2. Familial aggregations indicative of dominant inheritance. I. Neurology 1955;5(3):182–196.

22. KURLAND LT, MULDER DW. Epidemiologic investigations of amyotrophic lateral sclerosis. 2. Familial aggregations indicative of dominant inheritance. II. Neurology 1955;5(4):249–268.

23. Al-Chalabi A, Fang F, Hanby MF, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J. Neurol. Neurosurg. Psychiatry 2010;81(12):1324–1326.

24. Fogh I, Ratti A, Gellera C, et al. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. Hum. Mol. Genet. 2014;23(8):2220–2231.

25. Keller MF, Ferrucci L, Singleton AB, et al. Genome-wide analysis of the heritability of amyotrophic lateral sclerosis. JAMA Neurol. 2014;71(9):1123–1134.

26. Siddique T, Pericak-Vance MA, Brooks BR, et al. Linkage analysis in familial amyotrophic lateral sclerosis. Neurology 1989;39(7):919–925.

27. Siddique T, Figlewicz DA, Pericak-Vance MA, et al. Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. N. Engl. J. Med. 1991;324(20):1381–1384.

28. Rosen DR, Siddique T, Patterson D, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 1993;362(6415):59–62.

29. Wroe R, Wai-Ling Butler A, Andersen PM, et al. ALSOD: the Amyotrophic Lateral Sclerosis Online Database. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2008;9(4):249–250.

30. Andersen PM, Nilsson P, Ala-Hurula V, et al. Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. Nat. Genet. 1995;10(1):61–66.

31. Hand CK, Mayeux-Portas V, Khoris J, et al. Compound heterozygous D90A and D96N SOD1 mutations in a recessive amyotrophic lateral sclerosis family. Ann. Neurol. 2001;49(2):267–271.

32. Aguirre T, Matthijs G, Robberecht W, et al. Mutational analysis of the Cu/Zn superoxide dismutase gene in 23 familial and 69 sporadic cases of amyotrophic lateral sclerosis in Belgium. Eur. J. Hum. Genet. EJHG 1999;7(5):599–602.

33. Robberecht W, Aguirre T, Van den Bosch L, et al. D90A heterozygosity in the SOD1 gene is associated with familial and apparently sporadic amyotrophic lateral sclerosis. Neurology 1996;47(5):1336–1339.

34. Al-Chalabi A, Andersen PM, Chioza B, et al. Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor. Hum. Mol. Genet. 1998;7(13):2045–2050.

35. Parton MJ, Broom W, Andersen PM, et al. D90A-SOD1 mediated amyotrophic lateral sclerosis: a single founder for all cases with evidence for a Cis-acting disease modifier in the recessive haplotype. Hum. Mutat. 2002;20(6):473.

36. Van Es MA, Dahlberg C, Birve A, et al. Large-scale SOD1 mutation screening provides evidence for genetic heterogeneity in amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatry 2010;81(5):562–566.

37. Kenna KP, McLaughlin RL, Byrne S, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J. Med. Genet. 2013;50(11):776–783.

38. Cudkowicz ME, McKenna-Yasek D, Sapp PE, et al. Epidemiology of mutations in superoxide dismutase in amyotrophic lateral sclerosis. Ann. Neurol. 1997;41(2):210–221.

39. Brown JA, Min J, Staropoli JF, et al. SOD1, ANG, TARDBP and FUS mutations in amyotrophic lateral sclerosis: a United States clinical testing lab experience. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2012;13(2):217–222.

40. Andersen PM, Sims KB, Xin WW, et al. Sixteen novel mutations in the Cu/Zn superoxide dismutase gene in amyotrophic lateral sclerosis: a decade of discoveries, defects and disputes. Amyotroph. Lateral Scler. Mot. Neuron Disord. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2003;4(2):62–73.

41. Lindberg MJ, Tibell L, Oliveberg M. Common denominator of Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis: decreased stability of the apo state. Proc. Natl. Acad. Sci. U. S. A. 2002;99(26):16607–16612.

42. Stathopulos PB, Rumfeldt J a. O, Scholz GA, et al. Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. Proc. Natl. Acad. Sci. U. S. A. 2003;100(12):7021–7026.

43. Chen S, Zhang X, Song L, Le W. Autophagy dysregulation in amyotrophic lateral sclerosis. Brain Pathol. Zurich Switz. 2012;22(1):110–116.

44. Kikuchi H, Almer G, Yamashita S, et al. Spinal cord endoplasmic reticulum stress associated with a microsomal accumulation of mutant superoxide dismutase-1 in an ALS model. Proc. Natl. Acad. Sci. U. S. A. 2006;103(15):6025–6030.

45. Morfini GA, Bosco DA, Brown H, et al. Inhibition of fast axonal transport by pathogenic SOD1 involves activation of p38 MAP kinase. PloS One 2013;8(6):e65235.

46. Vande Velde C, McDonald KK, Boukhedimi Y, et al. Misfolded SOD1 associated with motor neuron mitochondria alters mitochondrial shape and distribution prior to clinical onset. PloS One 2011;6(7):e22031.

47. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. 2014;17(1):17–23.

48. Lagier-Tourenne C, Polymenidou M, Cleveland DW. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Hum. Mol. Genet. 2010;19(R1):R46–64.

49. Neumann M, Sampathu DM, Kwong LK, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 2006;314(5796):130–133.

50. Gitcho MA, Baloh RH, Chakraverty S, et al. TDP-43 A315T mutation in familial motor neuron disease. Ann. Neurol. 2008;63(4):535–538.

51. Sreedharan J, Blair IP, Tripathi VB, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science 2008;319(5870):1668–1672.

52. Buratti E, Brindisi A, Giombi M, et al. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. J. Biol. Chem. 2005;280(45):37572–37584.

53. Sapp PC, Hosler BA, McKenna-Yasek D, et al. Identification of two novel loci for dominantly inherited familial amyotrophic lateral sclerosis. Am. J. Hum. Genet. 2003;73(2):397–403.

54. Kwiatkowski TJ Jr, Bosco DA, Leclerc AL, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 2009;323(5918):1205–1208.

55. Peters OM, Ghasemi M, Brown RH. Emerging mechanisms of molecular pathology in ALS. J. Clin. Invest. 2015;125(5):1767–1779.

56. Hewitt C, Kirby J, Highley JR, et al. Novel FUS/TLS mutations and pathology in familial and sporadic amyotrophic lateral sclerosis. Arch. Neurol. 2010;67(4):455–461.

57. Vance C, Scotter EL, Nishimura AL, et al. ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules. Hum. Mol. Genet. 2013;22(13):2676–2688.

58. Morita M, Al-Chalabi A, Andersen PM, et al. A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology 2006;66(6):839–844.

59. Vance C, Al-Chalabi A, Ruddy D, et al. Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. Brain J. Neurol. 2006;129(Pt 4):868–876.

60. Luty AA, Kwok JBJ, Thompson EM, et al. Pedigree with frontotemporal lobar degeneration--motor neuron disease and Tar DNA binding protein-43 positive neuropathology: genetic linkage to chromosome 9. BMC Neurol. 2008;8:32.

61. Le Ber I, Camuzat A, Berger E, et al. Chromosome 9p-linked families with frontotemporal dementia associated with motor neuron disease. Neurology 2009;72(19):1669–1676.

62. Gijselinck I, Engelborghs S, Maes G, et al. Identification of 2 Loci at chromosomes 9 and 14 in a multiplex family with frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Arch. Neurol. 2010;67(5):606–616.

63. Pearson JP, Williams NM, Majounie E, et al. Familial frontotemporal dementia with amyotrophic lateral sclerosis and a shared haplotype on chromosome 9p. J. Neurol. 2011;258(4):647–655.

64. Boxer AL, Mackenzie IR, Boeve BF, et al. Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. J. Neurol. Neurosurg. Psychiatry 2011;82(2):196–203.

65. Valdmanis PN, Dupre N, Bouchard J-P, et al. Three families with amyotrophic lateral sclerosis and frontotemporal dementia with evidence of linkage to chromosome 9p. Arch. Neurol. 2007;64(2):240–245.

66. Shatunov A, Mok K, Newhouse S, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet Neurol. 2010;9(10):986–994.

67. Laaksovirta H, Peuralinna T, Schymick JC, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol. 2010;9(10):978–985.

68. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 2011;72(2):245–256.

69. Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 2011;72(2):257–268.

70. Ratti A, Corrado L, Castellotti B, et al. C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. Neurobiol. Aging 2012;33(10):2528.e7–14.

71. Rutherford NJ, Heckman MG, Dejesus-Hernandez M, et al. Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype. Neurobiol. Aging 2012;33(12):2950.e5–7.

72. Beck J, Poulter M, Hensman D, et al. 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. 2013;92(3):345–353.

73. Van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E, et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol. 2013;12(10):978–988.

74. Byrne S, Heverin M, Elamin M, et al. Intermediate repeat expansion length in C9orf72 may be pathological in amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Front. Degener. 2014;15(1-2):148–150.

75. Beer AM, Cooper-Knock J, Higginbottom A, et al. Intermediate length C9orf72 expansion in an ALS patient without classical C9orf72 neuropathology. Amyotroph. Lateral Scler. Front. Degener. 2015;16(3-4):249–251.

76. Simón-Sánchez J, Dopper EGP, Cohn-Hokke PE, et al. The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. Brain J. Neurol. 2012;135(Pt 3):723–735.

77. Majounie E, Renton AE, Mok K, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. 2012;11(4):323–330.

78. Jang J-H, Kwon M-J, Choi WJ, et al. Analysis of the C9orf72 hexanucleotide repeat expansion in Korean patients with familial and sporadic amyotrophic lateral sclerosis. Neurobiol. Aging 2013;34(4):1311.e7–9.

79. Zou Z-Y, Li X-G, Liu M-S, Cui L-Y. Screening for C9orf72 repeat expansions in Chinese amyotrophic lateral sclerosis patients. Neurobiol. Aging 2013;34(6):1710.e5–6.

80. Konno T, Shiga A, Tsujino A, et al. Japanese amyotrophic lateral sclerosis patients with GGGGCC hexanucleotide repeat expansion in C9ORF72. J. Neurol. Neurosurg. Psychiatry 2013;84(4):398–401.

81. Ogaki K, Li Y, Atsuta N, et al. Analysis of C9orf72 repeat expansion in 563 Japanese patients with amyotrophic lateral sclerosis. Neurobiol. Aging 2012;33(10):2527.e11–16.

82. Jiao B, Tang B, Liu X, et al. Identification of C9orf72 repeat expansions in patients with amyotrophic lateral sclerosis and frontotemporal dementia in mainland China. Neurobiol. Aging 2014;35(4):936.e19–22.

83. Soong B-W, Lin K-P, Guo Y-C, et al. Extensive molecular genetic survey of Taiwanese patients with amyotrophic lateral sclerosis. Neurobiol. Aging 2014;35(10):2423.e1–6.

84. Lagier-Tourenne C, Baughn M, Rigo F, et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc. Natl. Acad. Sci. U. S. A. 2013;110(47):E4530–4539.

85. Sareen D, O'Rourke JG, Meera P, et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci. Transl. Med. 2013;5(208):208ra149.

86. Lee Y-B, Chen H-J, Peres JN, et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep. 2013;5(5):1178–1186.

87. Fratta P, Mizielinska S, Nicoll AJ, et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci. Rep. 2012;2:1016.

88. Šket P, Pohleven J, Kovanda A, et al. Characterization of DNA G-quadruplex species forming from C9ORF72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Neurobiol. Aging 2015;36(2):1091–1096.

89. Haeusler AR, Donnelly CJ, Periz G, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 2014;507(7491):195–200.

90. Mori K, Weng S-M, Arzberger T, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science 2013;339(6125):1335–1338.

91. Mizielinska S, Grönke S, Niccoli T, et al. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science 2014;345(6201):1192–1194.

92. Ash PEA, Bieniek KF, Gendron TF, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron 2013;77(4):639–646.

93. Cooper-Knock J, Walsh MJ, Higginbottom A, et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain J. Neurol. 2014;137(Pt 7):2040–2051.

94. Mori K, Lammich S, Mackenzie IRA, et al. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. Acta Neuropathol. (Berl.) 2013;125(3):413–423.

95. May S, Hornburg D, Schludi MH, et al. C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. Acta Neuropathol. (Berl.) 2014;128(4):485–503.

96. Kwon I, Xiang S, Kato M, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science 2014;345(6201):1139–1145.

97. Levine TP, Daniels RD, Gatta AT, et al. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinforma. Oxf. Engl. 2013;29(4):499–503.

98. Farg MA, Sundaramoorthy V, Sultana JM, et al. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum. Mol. Genet. 2014;23(13):3579–3595.

99. Chen Y-Z, Bennett CL, Huynh HM, et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am. J. Hum. Genet. 2004;74(6):1128–1135.

100. Hirano M, Quinzii CM, Mitsumoto H, et al. Senataxin mutations and amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2011;12(3):223–227.

101. Zhao Z, Chen W-Z, Wu Z, et al. A novel mutation in the senataxin gene identified in a Chinese patient with sporadic amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2009;10(2):118–122.

102. Hadano S, Hand CK, Osuga H, et al. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. Nat. Genet. 2001;29(2):166–173.

103. Yang Y, Hentati A, Deng HX, et al. The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. Nat. Genet. 2001;29(2):160–165.

104. Al-Chalabi A, Hansen VK, Simpson CL, et al. Variants in the ALS2 gene are not associated with sporadic amyotrophic lateral sclerosis. Neurogenetics 2003;4(4):221–222.

105. Maruyama H, Morino H, Ito H, et al. Mutations of optineurin in amyotrophic lateral sclerosis. Nature 2010;465(7295):223–226.

106. Zhu G, Wu C-J, Zhao Y, Ashwell JD. Optineurin negatively regulates TNFalpha-induced NF-kappaB activation by competing with NEMO for ubiquitinated RIP. Curr. Biol. CB 2007;17(16):1438–1443.

107. Sahlender DA, Roberts RC, Arden SD, et al. Optineurin links myosin VI to the Golgi complex and is involved in Golgi organization and exocytosis. J. Cell Biol. 2005;169(2):285–295.

108. Akizuki M, Yamashita H, Uemura K, et al. Optineurin suppression causes neuronal cell death via NF-κB pathway. J. Neurochem. 2013;126(6):699–704.

109. Shen W-C, Li H-Y, Chen G-C, et al. Mutations in the ubiquitin-binding domain of OPTN/optineurin interfere with autophagy-mediated degradation of misfolded proteins by a dominant-negative mechanism. Autophagy 2015;11(4):685–700.

110. Hayward C, Colville S, Swingler RJ, Brock DJ. Molecular genetic analysis of the APEX nuclease gene in amyotrophic lateral sclerosis. Neurology 1999;52(9):1899–1901.

111. Greenway MJ, Alexander MD, Ennis S, et al. A novel candidate region for ALS on chromosome 14q11.2. Neurology 2004;63(10):1936–1938.

112. Greenway MJ, Andersen PM, Russ C, et al. ANG mutations segregate with familial and “sporadic” amyotrophic lateral sclerosis. Nat. Genet. 2006;38(4):411–413.

113. Tsuji T, Sun Y, Kishimoto K, et al. Angiogenin is translocated to the nucleus of HeLa cells and is involved in ribosomal RNA transcription and cell proliferation. Cancer Res. 2005;65(4):1352–1360.

114. Wu D, Yu W, Kishikawa H, et al. Angiogenin loss-of-function mutations in amyotrophic lateral sclerosis. Ann. Neurol. 2007;62(6):609–617.

115. Mitchell J, Paul P, Chen H-J, et al. Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. Proc. Natl. Acad. Sci. U. S. A. 2010;107(16):7556–7561.

116. Münch C, Sedlmeier R, Meyer T, et al. Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. Neurology 2004;63(4):724–726.

117. LaMonte BH, Wallace KE, Holloway BA, et al. Disruption of dynein/dynactin inhibits axonal transport in motor neurons causing late-onset progressive degeneration. Neuron 2002;34(5):715–727.

118. Chow CY, Landers JE, Bergren SK, et al. Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. Am. J. Hum. Genet. 2009;84(1):85–88.

119. Volpicelli-Daley L, De Camilli P. Phosphoinositides' link to neurodegeneration. Nat. Med. 2007;13(7):784–786.

120. Chow CY, Zhang Y, Dowling JJ, et al. Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. Nature 2007;448(7149):68–72.

121. Couthouis J, Hart MP, Erion R, et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Hum. Mol. Genet. 2012;21(13):2899–2911.

122. Ticozzi N, Vance C, Leclerc AL, et al. Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. Am. J. Med. Genet. Part B Neuropsychiatr. Genet. Off. Publ. Int. Soc. Psychiatr. Genet. 2011;156B(3):285–290.

123. Couthouis J, Hart MP, Shorter J, et al. A yeast functional screen predicts new candidate ALS disease genes. Proc. Natl. Acad. Sci. U. S. A. 2011;108(52):20881–20890.

124. Suraweera A, Lim Y, Woods R, et al. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. Hum. Mol. Genet. 2009;18(18):3384–3396.

125. Yüce Ö, West SC. Senataxin, defective in the neurodegenerative disorder ataxia with oculomotor apraxia 2, lies at the interface of transcription and the DNA damage response. Mol. Cell. Biol. 2013;33(2):406–417.

126. Fogel BL, Cho E, Wahnich A, et al. Mutation of senataxin alters disease-specific transcriptional networks in patients with ataxia with oculomotor apraxia type 2. Hum. Mol. Genet. 2014;23(18):4758–4769.

127. Fecto F, Yan J, Vemula SP, et al. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. Arch. Neurol. 2011;68(11):1440–1446.

128. Kirkin V, McEwan DG, Novak I, Dikic I. A role for ubiquitin in selective autophagy. Mol. Cell 2009;34(3):259–269.

129. Sanz L, Diaz-Meco MT, Nakano H, Moscat J. The atypical PKC-interacting protein p62 channels NF-kappaB activation by the IL-1-TRAF6 pathway. EMBO J. 2000;19(7):1576–1586.

130. Hiji M, Takahashi T, Fukuba H, et al. White matter lesions in the brain with frontotemporal lobar degeneration with motor neuron disease: TDP-43-immunopositive

inclusions co-localize with p62, but not ubiquitin. Acta Neuropathol. (Berl.) 2008;116(2):183–191.

131. Lattante S, de Calbiac H, Le Ber I, et al. Sqstm1 knock-down causes a locomotor phenotype ameliorated by rapamycin in a zebrafish model of ALS/FTLD. Hum. Mol. Genet. 2015;24(6):1682–1690.

132. Nishimura AL, Mitne-Neto M, Silva HCA, et al. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. Am. J. Hum. Genet. 2004;75(5):822–831.

133. Kanekura K, Nishimoto I, Aiso S, Matsuoka M. Characterization of amyotrophic lateral sclerosis-linked P56S mutation of vesicle-associated membrane protein-associated protein B (VAPB/ALS8). J. Biol. Chem. 2006;281(40):30223–30233.

134. Sanhueza M, Zechini L, Gillespie T, Pennetta G. Gain-of-function mutations in the ALS8 causative gene VAPB have detrimental effects on neurons and muscles. Biol. Open 2014;3(1):59–71.

135. Chattopadhyay D, Sengupta S. First evidence of pathogenicity of V234I mutation of hVAPB found in Amyotrophic Lateral Sclerosis. Biochem. Biophys. Res. Commun. 2014;448(1):108–113.

136. Johnson JO, Mandrioli J, Benatar M, et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron 2010;68(5):857–864.

137. Ye Y, Meyer HH, Rapoport TA. The AAA ATPase Cdc48/p97 and its partners transport proteins from the ER into the cytosol. Nature 2001;414(6864):652–656.

138. Weihl CC, Miller SE, Hanson PI, Pestronk A. Transgenic expression of inclusion body myopathy associated mutant p97/VCP causes weakness and ubiquitinated protein inclusions in mice. Hum. Mol. Genet. 2007;16(8):919–928.

139. Otomo A, Hadano S, Okada T, et al. ALS2, a novel guanine nucleotide exchange factor for the small GTPase Rab5, is implicated in endosomal dynamics. Hum. Mol. Genet. 2003;12(14):1671–1687.

140. Kanekura K, Hashimoto Y, Niikura T, et al. Alsin, the product of ALS2 gene, suppresses SOD1 mutant neurotoxicity through RhoGEF domain by interacting with SOD1 mutants. J. Biol. Chem. 2004;279(18):19247–19256.

141. Yamanaka K, Miller TM, McAlonis-Downes M, et al. Progressive spinal axonal degeneration and slowness in ALS2-deficient mice. Ann. Neurol. 2006;60(1):95–104.

142. Deng H-X, Chen W, Hong S-T, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature 2011;477(7363):211–215.

143. Ko HS, Uehara T, Tsuruma K, Nomura Y. Ubiquilin interacts with ubiquitylated proteins and proteasome through its ubiquitin-associated and ubiquitin-like domains. FEBS Lett. 2004;566(1-3):110–114.

144. Gilpin KM, Chang L, Monteiro MJ. ALS-linked mutations in ubiquilin-2 or hnRNPA1 reduce interaction between ubiquilin-2 and hnRNPA1. Hum. Mol. Genet. 2015;24(9):2565–2577.

145. Parkinson N, Ince PG, Smith MO, et al. ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). Neurology 2006;67(6):1074–1077.

146. Martin-Serrano J, Yarovoy A, Perez-Caballero D, et al. Divergent retroviral late-budding domains recruit vacuolar protein sorting factors by using alternative adaptor proteins. Proc. Natl. Acad. Sci. U. S. A. 2003;100(21):12414–12419.

147. Cox LE, Ferraiuolo L, Goodall EF, et al. Mutations in CHMP2B in lower motor neuron predominant amyotrophic lateral sclerosis (ALS). PloS One 2010;5(3):e9872.

148. Wu C-H, Fallini C, Ticozzi N, et al. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature 2012;488(7412):499–503.

149. Mockrin SC, Korn ED. Acanthamoeba profilin interacts with G-actin to increase the rate of exchange of actin-bound adenosine 5’-triphosphate. Biochemistry (Mosc.) 1980;19(23):5359–5362.

150. Smith BN, Ticozzi N, Fallini C, et al. Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. Neuron 2014;84(2):324–331.

151. Johnson JO, Pioro EP, Boehringer A, et al. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. Nat. Neurosci. 2014;17(5):664–666.

152. Salton M, Elkon R, Borodina T, et al. Matrin 3 binds and stabilizes mRNA. PloS One 2011;6(8):e23882.

153. Kim HJ, Kim NC, Wang Y-D, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature 2013;495(7442):467–473.

154. Michael WM, Choi M, Dreyfuss G. A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. Cell 1995;83(3):415–422.

155. Takahashi Y, Fukuda Y, Yoshimura J, et al. ERBB4 mutations that disrupt the neuregulin-ErbB4 pathway cause amyotrophic lateral sclerosis type 19. Am. J. Hum. Genet. 2013;93(5):900–905.

156. Huang YZ, Won S, Ali DW, et al. Regulation of neuregulin signaling by PSD-95 interacting with ErbB4 at CNS synapses. Neuron 2000;26(2):443–455.

157. Pearson RJ, Carroll SL. ErbB transmembrane tyrosine kinase receptors are expressed by sensory and motor neurons projecting into sciatic nerve. J. Histochem. Cytochem. Off. J. Histochem. Soc. 2004;52(10):1299–1311.

158. Gassmann M, Casagranda F, Orioli D, et al. Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. Nature 1995;378(6555):390–394.

159. Lattante S, Conte A, Zollino M, et al. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology 2012;79(1):66–72.

160. Conte A, Lattante S, Luigetti M, et al. Classification of familial amyotrophic lateral sclerosis by family history: effects on frequency of genes mutation. J. Neurol. Neurosurg. Psychiatry 2012;83(12):1201–1203.

161. Van Blitterswijk M, van Es MA, Hennekam EAM, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum. Mol. Genet. 2012;21(17):3776–3784.

162. Al-Chalabi A, Calvo A, Chio A, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. Lancet Neurol. 2014;13(11):1108–1113.

163. Chiò A, Calvo A, Moglia C, et al. A de novo missense mutation of the FUS gene in a "true" sporadic ALS case. Neurobiol. Aging 2011;32(3):553.e23–26.

164. DeJesus-Hernandez M, Kocerha J, Finch N, et al. De novo truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis. Hum. Mutat. 2010;31(5):E1377–1389.

165. Zou Z-Y, Cui L-Y, Sun Q, et al. De novo FUS gene mutations are associated with juvenile-onset sporadic amyotrophic lateral sclerosis in China. Neurobiol. Aging 2013;34(4):1312.e1–8.

166. Alexander MD, Traynor BJ, Miller N, et al. "True" sporadic ALS associated with a novel SOD-1 mutation. Ann. Neurol. 2002;52(5):680–683.

167. Chesi A, Staahl BT, Jovičić A, et al. Exome sequencing to identify de novo mutations in sporadic ALS trios. Nat. Neurosci. 2013;16(7):851–855.

168. Elden AC, Kim H-J, Hart MP, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 2010;466(7310):1069–1075.

169. Van Damme P, Veldink JH, van Blitterswijk M, et al. Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. Neurology 2011;76(24):2066–2072.

170. Van Es MA, Veldink JH, Saris CGJ, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat. Genet. 2009;41(10):1083–1087.

171. Diekstra FP, Van Deerlin VM, van Swieten JC, et al. C9orf72 and UNC13A are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: a genome-wide meta-analysis. Ann. Neurol. 2014;76(1):120–133.

172. Cronin S, Berger S, Ding J, et al. A genome-wide association study of sporadic ALS in a homogenous Irish population. Hum. Mol. Genet. 2008;17(5):768–774.

173. Dunckley T, Huentelman MJ, Craig DW, et al. Whole-genome analysis of sporadic amyotrophic lateral sclerosis. N. Engl. J. Med. 2007;357(8):775–788.

174. Van Es MA, Van Vught PW, Blauw HM, et al. ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. Lancet Neurol. 2007;6(10):869–877.

175. Simpson CL, Lemmens R, Miskiewicz K, et al. Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. Hum. Mol. Genet. 2009;18(3):472–481.

176. Cronin S, Tomik B, Bradley DG, et al. Screening for replication of genome-wide SNP associations in sporadic ALS. Eur. J. Hum. Genet. EJHG 2009;17(2):213–218.

177. Chiò A, Schymick JC, Restagno G, et al. A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. Hum. Mol. Genet. 2009;18(8):1524–1532.

178. Fernández-Santiago R, Sharma M, Berg D, et al. No evidence of association of FLJ10986 and ITPR2 with ALS in a large German cohort. Neurobiol. Aging 2011;32(3):551.e1–4.

179. Fogh I, D'Alfonso S, Gellera C, et al. No association of DPP6 with amyotrophic lateral sclerosis in an Italian population. Neurobiol. Aging 2011;32(5):966–967.

180. Kwee LC, Liu Y, Haynes C, et al. A high-density genome-wide association screen of sporadic ALS in US veterans. PloS One 2012;7(3):e32768.

181. Van Es MA, Van Vught PWJ, Veldink JH, et al. Analysis of FGGY as a risk factor for sporadic amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2009;10(5-6):441–447.

182. Sabatelli M, Eusebi F, Al-Chalabi A, et al. Rare missense variants of neuronal nicotinic acetylcholine receptor altering receptor function are associated with sporadic amyotrophic lateral sclerosis. Hum. Mol. Genet. 2009;18(20):3997–4006.

183. Sabatelli M, Lattante S, Conte A, et al. Replication of association of CHRNA4 rare variants with sporadic amyotrophic lateral sclerosis: the Italian multicentre study. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 2012;13(6):580–584.

184. Gonzalez-Perez P, Woehlbier U, Chian R-J, et al. Identification of rare protein disulfide isomerase gene variants in amyotrophic lateral sclerosis patients. Gene 2015;

185. Cirulli ET, Lasseigne BN, Petrovski S, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 2015;347(6229):1436–1441.

186. Landers JE, Melki J, Meininger V, et al. Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. Proc. Natl. Acad. Sci. U. S. A. 2009;106(22):9004–9009.

187. Van Doormaal PTC, Ticozzi N, Gellera C, et al. Analysis of the KIFAP3 gene in amyotrophic lateral sclerosis: a multicenter survival study. Neurobiol. Aging 2014;35(10):2420.e13–14.

188. Traynor BJ, Nalls M, Lai S-L, et al. Kinesin-associated protein 3 (KIFAP3) has no effect on survival in a population-based cohort of ALS patients. Proc. Natl. Acad. Sci. U. S. A. 2010;107(27):12335–12338.

189. Chiò A, Mora G, Restagno G, et al. UNC13A influences survival in Italian amyotrophic lateral sclerosis patients: a population-based study. Neurobiol. Aging 2013;34(1):357.e1–5.

190. Van Hoecke A, Schoonaert L, Lemmens R, et al. EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models and in humans. Nat. Med. 2012;18(9):1418–1422.

191. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. Nat. Rev. Neurol. 2013;9(11):617–628.

192. Fang F, Ingre C, Roos P, et al. Risk factors for amyotrophic lateral sclerosis. Clin. Epidemiol. 2015;181.[cited 2015 Aug 6]

193. Scarmeas N, Shih T, Stern Y, et al. Premorbid weight, body mass, and varsity athletics in ALS. Neurology 2002;59(5):773–775.

194. Huisman MHB, Seelen M, de Jong SW, et al. Lifetime physical activity and the risk of amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatry 2013;84(9):976–981.

195. Engelhardt JI, Appel SH. IgG reactivity in the spinal cord and motor cortex in amyotrophic lateral sclerosis. Arch. Neurol. 1990;47(11):1210–1216.

196. Pramatarova A, Laganière J, Roussel J, et al. Neuron-specific expression of mutant superoxide dismutase 1 in transgenic mice does not lead to motor impairment. J. Neurosci. Off. J. Soc. Neurosci. 2001;21(10):3369–3374.

197. Lino MM, Schneider C, Caroni P. Accumulation of SOD1 mutants in postnatal motoneurons does not cause motoneuron pathology or motoneuron disease. J. Neurosci. Off. J. Soc. Neurosci. 2002;22(12):4825–4832.

198. Clement AM, Nguyen MD, Roberts EA, et al. Wild-type nonneuronal cells extend survival of SOD1 mutant motor neurons in ALS mice. Science 2003;302(5642):113–117.

199. Engelhardt JI, Tajti J, Appel SH. Lymphocytic infiltrates in the spinal cord in amyotrophic lateral sclerosis. Arch. Neurol. 1993;50(1):30–36.

200. Turner MR, Cagnin A, Turkheimer FE, et al. Evidence of widespread cerebral microglial activation in amyotrophic lateral sclerosis: an [11C](R)-PK11195 positron emission tomography study. Neurobiol. Dis. 2004;15(3):601–609.

201. Liao B, Zhao W, Beers DR, et al. Transformation from a neuroprotective to a neurotoxic microglial phenotype in a mouse model of ALS. Exp. Neurol. 2012;237(1):147–152.

Chapter 2

Clinical characterization of a North American ALS cohort

2.1 Abstract

ALS is a clinically heterogeneous disease with variant in age of onset, site of symptom onset, survival, and rate of progression. Understanding the factors that lead to variation in the disease phenotype, for instance factors that lead to onset at a younger age or decrease survival time, is important for understanding disease pathogenesis. It is also valuable information for designing clinical trials in determining which demographic variables should be taken into account for patient stratification and which variables to control for in determining treatment outcomes. We analyzed the clinical characteristics of our ALS cohort and found that gender was a significant predictor of site of onset with women more likely to have bulbar onset. Site of onset was the most significant predictor of age of onset and subjects with bulbar onset had later age of onset. Age of onset was the most significant predictor of both rate of progression and survival. A younger age of onset was associated with slower progression and longer survival. Site of onset was also significant to progression, as those with bulbar onset had more rapid progression than those with limb onset. It will be important to take factors in consideration when analyzing the impact of genetic and environmental factors in disease and in evaluation of treatment options.

2.2 Introduction

Amyotrophic lateral sclerosis is a universally fatal neurodegenerative disease. Onset typically occurs in mid-to late life with age of onset typically ranging from 40-70 with an average in the mid-60s.[1–6] Death occurs within 3-5 years generally due to respiratory failure.[7–11]

There is no cure for the disease and the only treatment available, Riluzole, only extends survival by 2-3 months.[12]

Understanding the clinical characteristics and what influences them will not only aid in patient counseling, but can provide insight into pathogenesis. Furthermore, knowing which factors influence clinical aspects such as age of onset, survival, and progression will be crucial in evaluation of novel therapies as they move to clinical trials. To delineate the clinical characteristics of our ALS cohort, we looked at the relationships between family history, gender, ethnicity, and site of onset, age of onset, survival, and progression.

2.3 Methods

2.3.1 Subjects

The work presented here includes a total of 873 subjects collected by the Washington University Neuromuscular Genetics Project. Contributing sites included the Neuromuscular Disease Center in St. Louis, Missouri (WUSM), the Virginia Mason Medical Center (VMMC), Houston Methodist Hospital, University of Utah, and Cedars Sinai Medical Center. Subjects from the Neuromuscular Genetics Projects were utilized for different studies included in this work, including poooled-sample sequencing (n=391), exome sequencing (n=724), TREM2 (n=522), C9ORF72 discovery (n= 389), and C9ORF72 haplotype analysis (n=101). Most of the subjects were included in multiple studies. Only subjects that were diagnosed with definite or probable ALS according to El Escorial criteria were included in the analysis of clinical characteristics (n=778).

2.3.2 Data analysis

All analysis was performed in R version 3.1.1. Fisher's exact tests were used to compare categorical variables (gender, family history, site of onset, *C9ORF72* status, ethnicity and progression). Wilcoxon rank-sum tests were used to compare age of onset. Survival analysis was performed using a multivariate Cox propotional hazards model with family history, gender, site of onset, ethnicity, *C9ORF72* status, and age of onset as covariates. Kaplan Meier curves were generated for variables that were significantly associated with survival in the Cox proportional hazard model. Age of onset was divided into quartiles for graphing (<50 years, 50-60 years, 60-68 years, and >68 years). Spearman's correlation was used to assess the relationship between progression and age of onset.

2.4 Results

2.4.1 General characteristics

Clinical characteristics the 778 subjects diagnoses with definite or probable ALS are summarized in table 2.1. Age of onset and survival were within previously reported ranges. Here, survival was calculated for subjects who were deceased only. The male to female ratio was 1.3, which was also typical.[13,14] Our cohort was predominantly of white, non-Hispanic origin. The rate of family history and *C9ORF72* repeat expansions in this cohort are not representative of the ALS population as a whole since some sub-cohorts were ascertained based on either the presence or absence of these characteristics. *C9ORF72* was the only genetic factor we were able to include since mutations in other genes are not common enough to perform analysis.

Table 2.1 Clinical characteristics of 778 ALS patients.

% FALS (n=497)	39.2%
% *C9ORF72* positive (n=773)	8.92%
% Male (n=778)	56.56%
Ethnicity (n=761)	
White (non-Hispanic)	691 (91%)
African American	21 (2.8%)
Hispanic	32 (4.2%)
Other	17 (2%)
Age of onset, yrs. (n=766, median; range)	60; 14-85.9
Site of onset (n=769)	
Spinal	532 (69%)
Bulbar	213 (28%)
Respiratory	11 (1.4%)
FTD	6 (0.8%)
Diffuse	2 (0.3%)
Survival, mo. (n=295, median; range)*	30; 6-220

2.4.2 *C9ORF72*

There were a total of 69 expansion carriers. The *C9ORF72* expansion is significantly more common in subjects with a family history of disease. 44/65 expansion carriers had a family history of disease compared to 151/429 non-expansion carriers (only including subjects for which *C9ORF72* status and family history were known). The expansion was also slightly more common in females compared to males (38/69 expansion carriers and 407/704 non expansion carriers were men (Table2.2)).

Table 2.2 Differences in *C9ORF72* expansion rates by demographic categories.

	Expanded	**Not expanded**	**p-value**
Gender (%Male)	44.9%	57.8%	0.04222
Ethnicity (%White, non-Hispanic)	95.6%	90.3%	0.1889
Family History (%+)	67.7%	35%	1.077×10^{-6}

2.4.3 Site of onset

To analyze site of onset, all subjects with non-spinal onset (bulbar, diffuse, FTD, and respiratory) were combined into a single category (Table 2.3). Non-spinal onset was significantly more common in females compared to males. Non-spinal onset was also more common in FALS compared to SALS, but was not significant when the same analysis was performed in only non-C9 expansion carriers. There was a trend toward non-spinal onset in C9 expansion carriers, but this was not statistically significant on its own.

Table 2.3 Differences in site of onset by demographic categories.

	Spinal	**Not Spinal**	**P-value**
Family History	35.6%	45.6%	0.04282
Family History (C9-)	32.3%	39.5%	0.1729
C9ORF72	8.0%	11.5%	0.1338
Male	60%	48.5%	0.003557
White, non-Hispanic	93.1%	89.8%	0.1095

Table 2.4 shows the site of first limb affected for subjects with spinal onset. There were 3 individuals whose onset was listed as “spinal” with no further information, and 2 with initial symptoms in the trunk that are not listed in this table. The most common site of onset was in the right upper limb and overall, upper limb onset was more common than lower limb. It was not uncommon for subjects to have symptoms begin in both right and left limbs simultaneously, but only 2 subjects had symptoms begin in both upper and lower limbs simultaneously. Women were significantly more likely to have lower limb onset with 60% of women having lower limb onset compared to 38% of men ($p=1.319x10^{-6}$).

Table 2.4 Site of initial symptoms in cases with spinal onset.

	Left	Right	Both	Unspecified	Total
Upper	80	100	31	64	275
Lower	76	69	38	64	247
Both	1	0	1	0	2
Unspecified	0	0	0	3	3
Total	157	169	70	131	527

Within subjects with spinal onset, upper vs lower limb and right vs left limb had no effect on age of onset, progression, or survival (data not shown). Progression rate was however significantly associated with onset in left and right limbs simultaneously compared to onset in either a left or right limb (p=0.007399). There was a higher prevalence of rapid progression, but also a higher prevalence of very slow progression in those with both right and left onset (Figure 2.1).

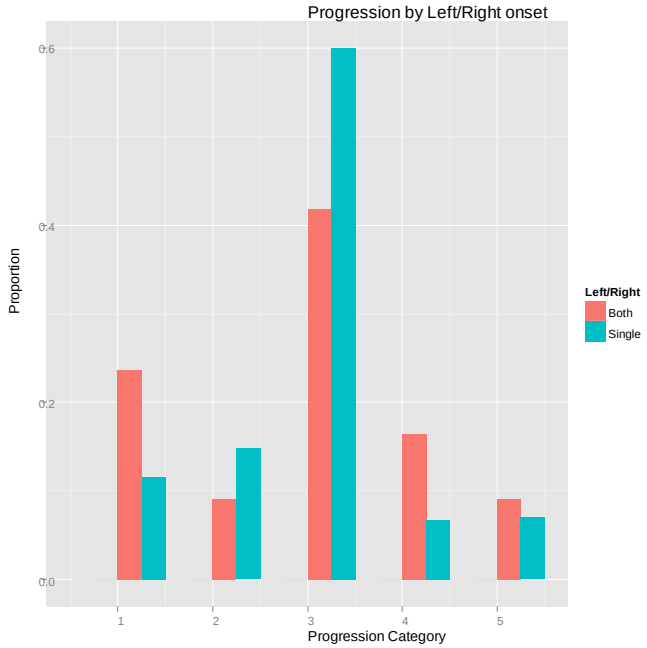

Figure 2.1 Progression rate in patients with either left or right or left and right limb onset.

2.4.4 Age of Onset

Age of onset significantly differed for all categories tested (gender, family history, *C9ORF72* expansion status, ethnicity, and site of onset) except for *C9ORF72* expansion status (Figure 2.2). Interestingly, there was a significantly earlier onset in FALS subjects compared to SALS; however the difference was due mostly to the higher proportion of *C9ORF72* expansion carriers in the FALS cohort, despite the fact that there was no significant association with *C9ORF72* status alone (Figure 2.2 A, B, C). The later onset in females compared to males was due to the later onset in non-spinal onset, which is more common in females (Figure 2.2 D, E, F). We also observed later onset in white, non-Hispanic subjects compared to other ethnicities (Figure2.2 G).

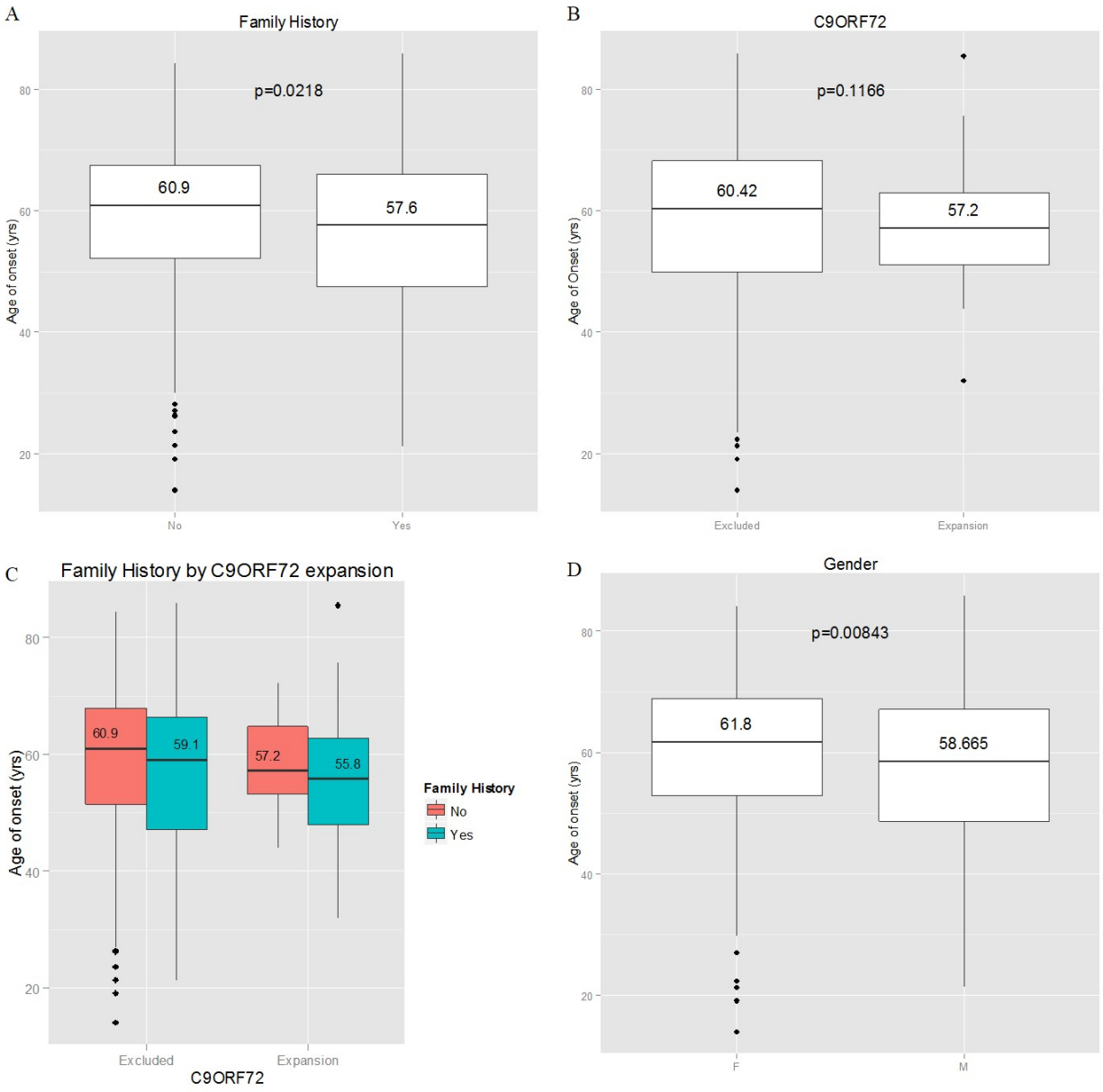

Figure 2.2 Age of onset by A. Family history of ALS, B. C9ORF72 expansion status, C. Family history stratified by C9ORF72 status, D. Gender, E. Site of onset, F. Gender stratified by site of onset, and G. Ethnicity. The median is age of onset in years is shown on the boxplots.

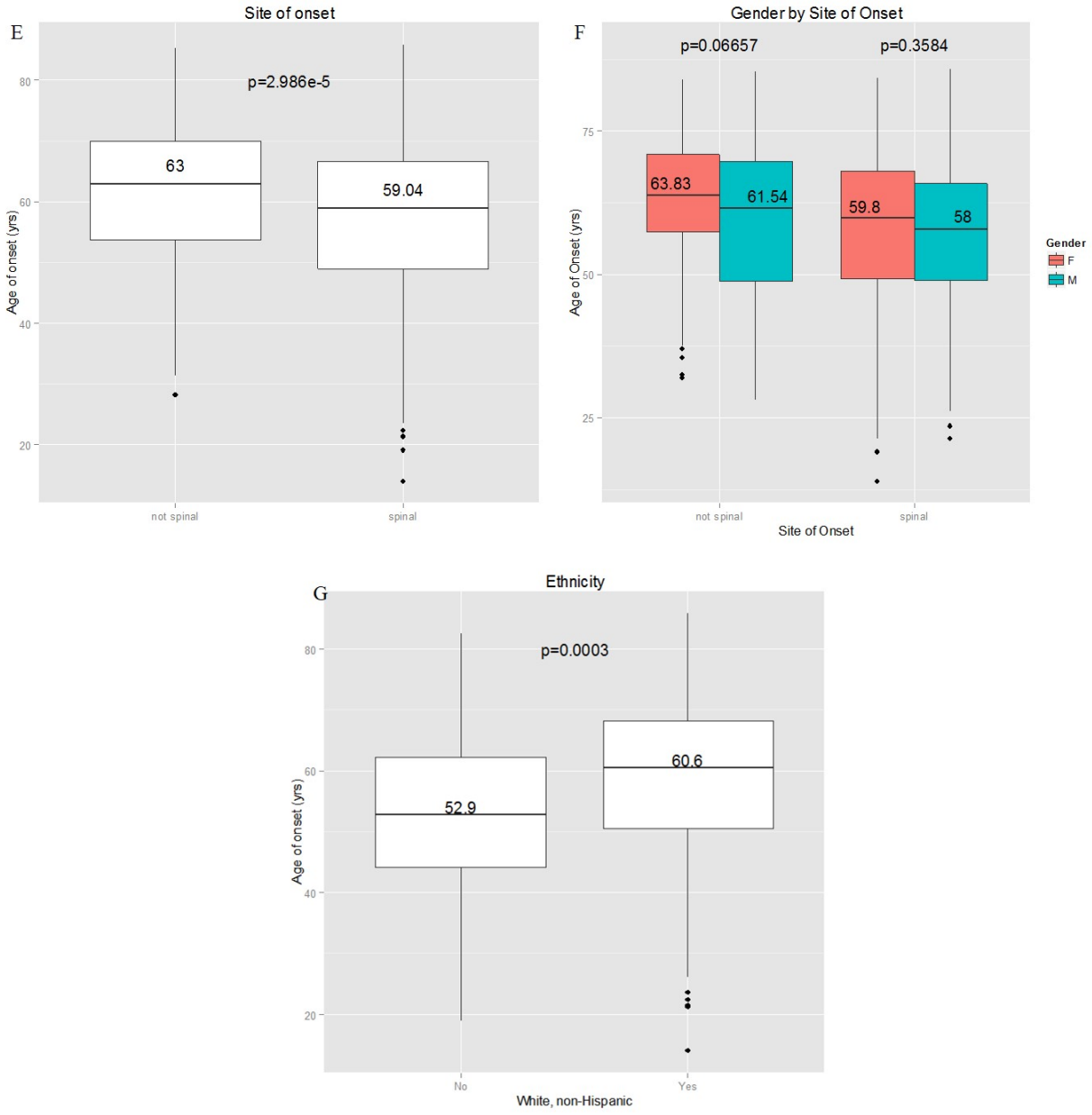

Figure 2.2 continued.

2.4.5 Progression

Progression rates were available for 637 subjects. However, rates were derived from three different metrics. Subjects that had ALSFRS measured at multiple time-points were calculated as change in ALSFRS/time between visits (in months). When this was unavailable, progression was calculated for subjects who were deceased as 48/duration in months. Subjects collected at VMCC were assigned to categories slow, typical-slow, typical, typical-fast, and fast based on expert evaluation. In order to combine these data, ALSFRS/time and 48/duration were assigned to categories 1 (slow), 2(typical-slow), 3 (typical), 4 (typical-fast), and 5 (fast) based on percentile of their respective distributions such that the up to 10% = 1, 10-20% = 2, 20-80% = 3, 80-90% = 4, and 90-100% = category 5. The distribution of the proportion of subjects falling into each category was roughly the same for all methods of calculation (Figure 2.3).

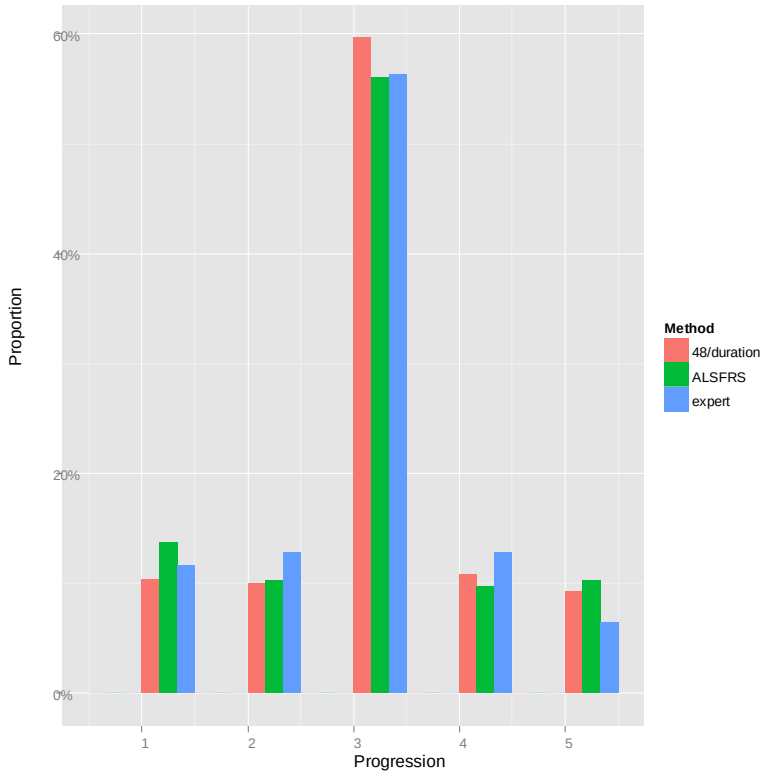

Figure 2.3 Proportion of subjects falling into each progression category stratified by method of calculating progression.

The distribution of progression categories differed between *C9ORF72* expansion carriers and non- expansion carriers (p=0.03057). *C9ORf72* carriers had a higher proportion with typical progression (category 3), while non-expansion carriers had more with slower progression (category 1 and 2). Expansion carriers also had a higher proportion of rapid progression (category 5) (Figure 2.4 A). There was also a significant difference between males and females (p=0.008689). Females appeared to have a higher proportion of category 3 compared to males, however males had a higher proportion in both categories 2 and 5 (Figure 2.4 B). There was a clear association between site of onset and progression (p=0.003356) with a higher proportion of cases with non-spinal onset in categories 4 and 5 (more rapid progression) and more cases with spinal onset in categories 1 and 2 (slower progression) (Figure 2.4 C). We observed no associations with family history or ethnicity (Figure 2.4 D, E).

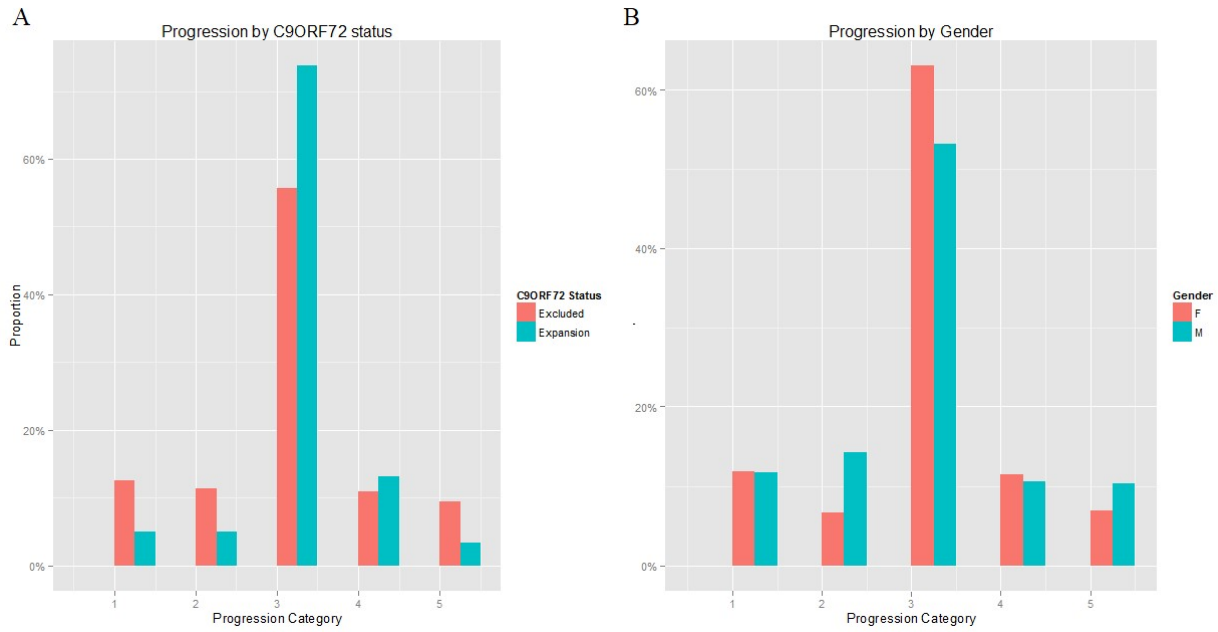

Figure 2.4 Proportion of subjects falling into each progression category stratified by A. Gender, B. Family history of ALS, C. *C9ORF72* expansion status, D. Ethnicity, and E. Site of onset.

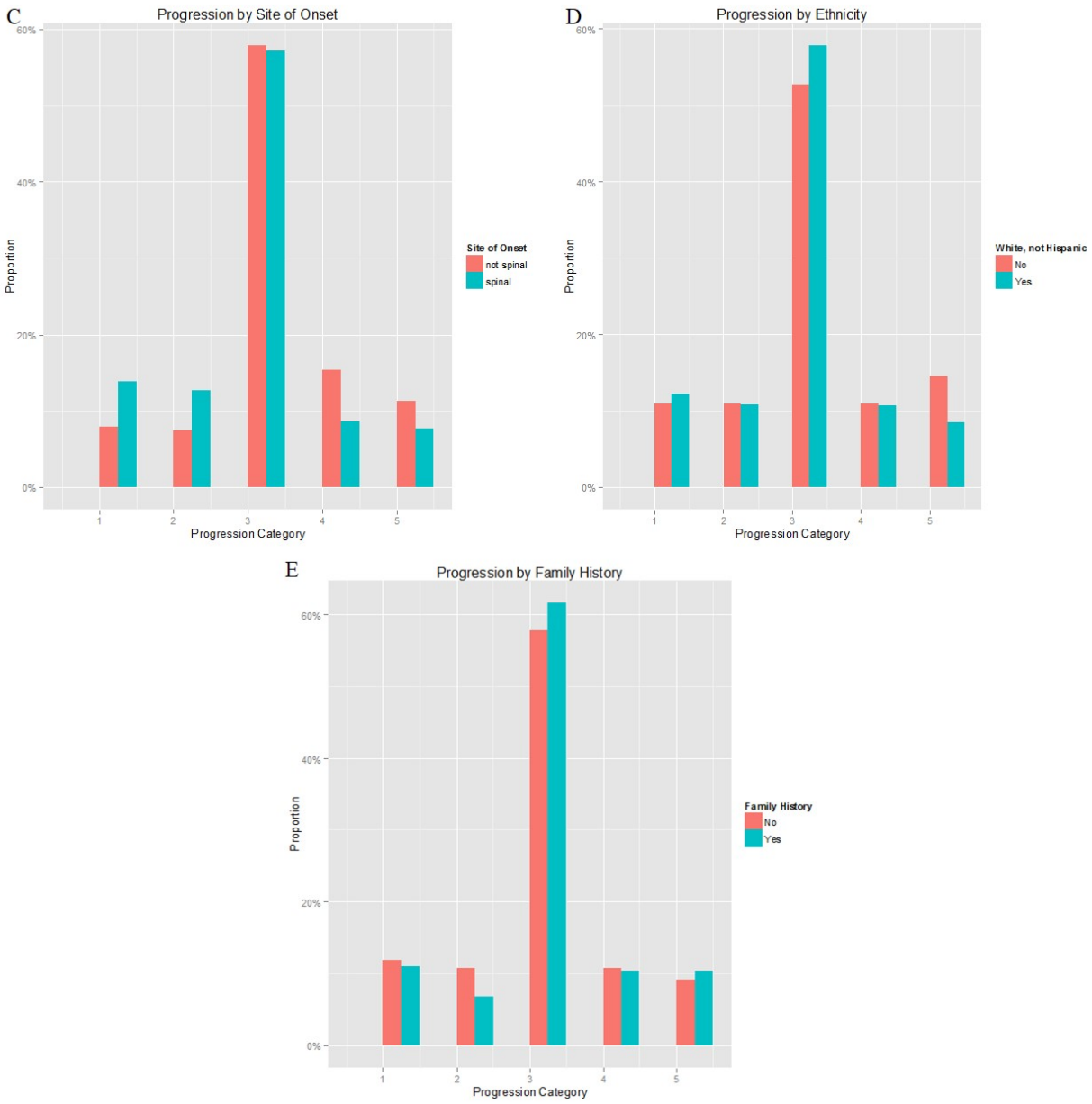

Figure 2.4 continued.

There was a significant correlation between age of onset and progression ($p=3.92x10^{-4}$, spearman's rho=0.14) with more rapid progression in subjects with later onset (median age of onset for category1: 54yrs, category2: 60yrs, category3: 63.65yrs, category4: 61yrs).

The relationship between progression and survival was not assessed since survival time was used to calculate progression for many of the cases. Furthermore, there was no survival data for those whose progression was determined by expert analysis.

2.4.6 Survival

The median survival time for subjects with known disease duration (due to documented death) was 30 months. Using a Kaplan Meier survival analysis including all subjects (Figure 2.5), the median survival was estimated to be 42 (95% CI: 39-46).

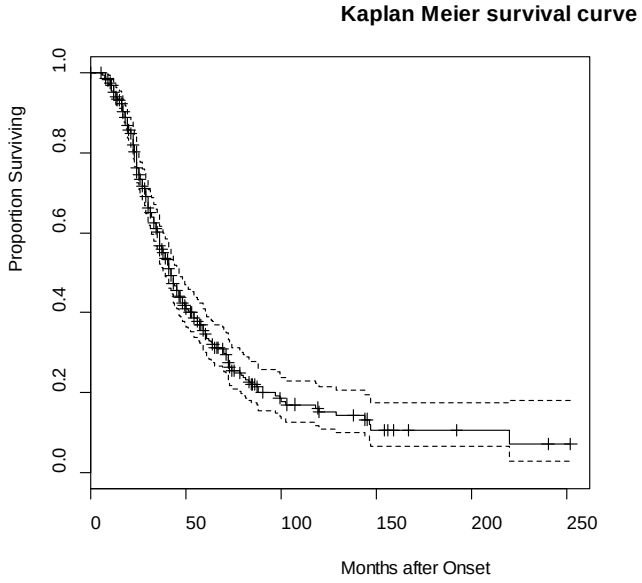

Figure 2.5 Kaplan Meier Survival curve of whole ALS cohort. Dotted lines represent 95% confidence interval.

We fit a multivariate Cox proportional hazards model to determine which variables significantly impacted survival times (Table 2.5). Age of onset was the most statistically significant predictor of survival time with earlier age of onset resulting in longer survival (Figure 2.6 A). *C9ORf72* expansions also significantly impacted the probability of death with a hazard

ratio over 2 meaning death was greater than twice as likely for expansion carriers compared to non-expansion carriers (Figure 2.6 B). Spinal onset was significantly associated with longer survival (Figure 2.6 C); however we have also shown that subjects with non-spinal onset have a later age of onset. Therefore, when age of onset is included in the multivariate model, the p-value is barely significant. Ethnicity was also significant in this model with longer survival in non-white subjects, (Figure 2.6 D).

Table 2.5 Cox-proportional hazards model. Median survival is calculated from Kaplan Meier curves

	Median Survival			
Variable	**Yes**	**No**	**Hazard Ratio**	**P-value**
Family History	41	48	1.062	0.69
Male	48	39	0.785	0.09
C9ORF72 (+)	32	43	2.159	$1.6x10^{-5}$
Spinal Onset	48	30	0.742	0.05
White, not Hispanic	41	NA	2.252	0.037
Age of onset	--	--	1.039	$4x10^{-10}$

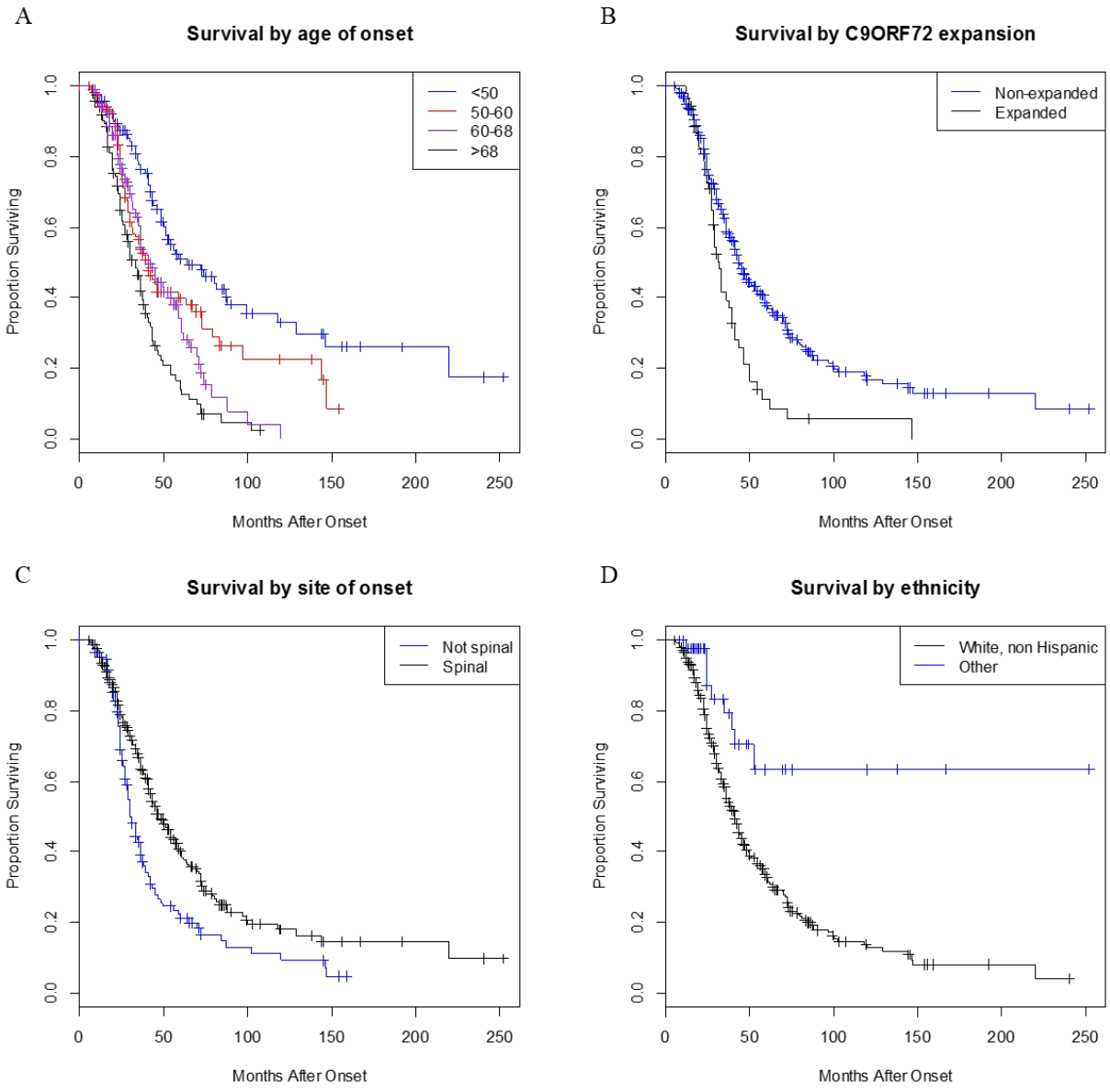

Figure 2.6 Kaplan Meier survival curves stratified by A. Age of onset, B. *C9ORF72* expansion status, C. Site of onset, and D. Ethnicity.

2.5 Discussion

Here we have provided an analysis of the some of the clinical characteristics of amyotrophic lateral sclerosis in a North American cohort. The patients included in this analysis

were derived from multiple centers and were included in a variety of genetic studies described in subsequent chapters. There is some amount of ascertainment bias for certain characteristics. For instance, family history of ALS is more likely to be recorded when it is positive, but is left blank rather than recorded as "sporadic" otherwise. This artificially inflates the rate of FALS in our cohort, but should not affect the relationship between family history and clinical presentation.

We initially found bulbar onset to be more common in cases with a family history of ALS, however there was no difference in site of onset when we stratified for *C9ORF72* status despite the fact that there was not a significant difference in site of onset between expansion carriers and non-expansion carriers. This is consistent with our previously reported finding that bulbar onset is more common in expansion carriers with FALS but not in expansion carriers with SALS.[15] Gender was the only characteristic that had a significant association with site of onset with women more likely to have bulbar onset than men, an association that is well known.[13,14,16–19] Further stratifying spinal onset by limb showed that women with spinal onset were more likely to have lower limb onset than men. To our knowledge, this association has not been previously reported.

Site of onset was the factor most significantly associated with age of onset with subjects with spinal onset having earlier onset than those with bulbar and other onset. While it appeared that onset was later in women, this effect was due to the increase in bulbar onset in women compared to men. Similarly, the later onset we observed in familial cases was due to the increase in bulbar onset in FALS expansion carriers. We also noted that the age of onset was earlier in non-White subjects.

We found age of onset to be the most important indicator of survival as other studies have also shown (reviewed by Chiò et al.[7]).We also noted that *C9ORF72* expansions were significantly associated with shorter survival time. Ethnicity and site of onset were associated with survival, although less so, even when controlling for age of onset. Studies have consistently shown shorter survival times in bulbar onset.[20–26] The studies investigating the role of ethnicity in survival times have conflicting results however most have shown no difference in survival times between ethnicities.[8,10,27] It is possible that the association we see is due to the low numbers of non-white subjects, but differences in follow-up play a role as well. Obituaries are often used to determine survival since many patients don't return to the same clinic and non-white patients are less likely to have obituaries published, resulting in more censored data.

Progression rate was significantly associated with both age and site of onset which was consistent with the association with survival. In addition to more rapid progression in patients with bulbar onset, we also showed that within subjects with spinal onset, there progression was more rapid in when symptoms began in right and left limbs simultaneously. We also saw an association with *C9ORF72* with non-expansion carriers having higher numbers of slower progression, however there were also more non-expansion carriers in the most rapid category. This is possibly caused by specific mutations in other genes, such as *SOD1,* that are associated with rapid progression, however we did not include mutations other than the *C9ORF72* expansion in this analysis. We did not examine the relationship between progression and survival since survival time was used to calculate progression rate in many cases.

There were other clinical characteristics, such as frontotemporal dementia, and used of Riluzole we did not include in analysis due to differences in data collection. Patients in this cohort were collected from four different centers, all of which collected different information.

We noted that cognitive changes were not always recorded and assessment methods were not consistent. While this analysis is not completely comprehensive, we did find several significant associations that need to be taken into account in further analysis.

2.6 References

1. Testa, D., Lovati, R., Ferrarini, M., Salmoiraghi, F. & Filippini, G. Survival of 793 patients with amyotrophic lateral sclerosis diagnosed over a 28-year period. Amyotroph. Lateral Scler. Other Motor Neuron Disord. 5, 208–212 (2004).

2. Logroscino, G. et al. Incidence of amyotrophic lateral sclerosis in Europe. J. Neurol. Neurosurg. Psychiatr. 81, 385–390 (2010).

3. Huisman, M. H. B. et al. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J. Neurol. Neurosurg. Psychiatr. 82, 1165–1170 (2011).

4. Uenal, H. et al. Incidence and geographical variation of amyotrophic lateral sclerosis (ALS) in Southern Germany--completeness of the ALS registry Swabia. PLoS ONE 9, e93932 (2014).

5. Piemonte and Valle d'Aosta Register for Amyotrophic Lateral Sclerosis (PARALS). Incidence of ALS in Italy: evidence for a uniform frequency in Western countries. Neurology 56, 239–244 (2001).

6. Fang, F. et al. Amyotrophic lateral sclerosis in Sweden, 1991-2005. Arch. Neurol. 66, 515–519 (2009).

7. Chiò, A. et al. Prognostic factors in ALS: A critical review. Amyotrophic Lateral Sclerosis 10, 310–323 (2009).

8. Del Aguila, M. A., Longstreth, W. T., McGuire, V., Koepsell, T. D. & van Belle, G. Prognosis in amyotrophic lateral sclerosis: a population-based study. Neurology 60, 813–819 (2003).

9. Jordan, H., Fagliano, J., Rechtman, L., Lefkowitz, D. & Kaye, W. Population-based surveillance of amyotrophic lateral sclerosis in New Jersey, 2009-2011. Neuroepidemiology 43, 49–56 (2014).

10. Jordan, H., Fagliano, J., Rechtman, L., Lefkowitz, D. & Kaye, W. Effects of demographic factors on survival time after a diagnosis of amyotrophic lateral sclerosis. Neuroepidemiology 44, 114–120 (2015).

11. Amyotrophic lateral sclerosis. (Taylor & Francis, 2006).

12. Miller, R. G., Mitchell, J. D. & Moore, D. H. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). Cochrane Database Syst Rev 3, CD001447 (2012).

13. Manjaly, Z. R. et al. The sex ratio in amyotrophic lateral sclerosis: A population based study. Amyotroph Lateral Scler 11, 439–442 (2010).

14. McCombe, P. A. & Henderson, R. D. Effects of gender in amyotrophic lateral sclerosis. Gend Med 7, 557–570 (2010).

15. Harms, M. B. et al. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Neurobiol. Aging 34, 2234.e13–19 (2013).

16. Atsuta, N. et al. Age at onset influences on wide-ranged clinical features of sporadic amyotrophic lateral sclerosis. J. Neurol. Sci. 276, 163–169 (2009).

17. Traynor, B. J. et al. Incidence and prevalence of ALS in Ireland, 1995–1997 A population-based study. Neurology 52, 504–504 (1999).

18. Nalini, A., Thennarasu, K., Gourie-Devi, M., Shenoy, S. & Kulshreshtha, D. Clinical characteristics and survival pattern of 1153 patients with amyotrophic lateral sclerosis: Experience over 30 years from India. Journal of the Neurological Sciences 272, 60–70 (2008).

19. The Scottish Motor Neuron Disease Register: a prospective study of adult onset motor neuron disease in Scotland. Methodology, demography and clinical features of incident cases in 1989. J. Neurol. Neurosurg. Psychiatr. 55, 536–541 (1992).

20. Traynor, B. J. et al. Clinical features of amyotrophic lateral sclerosis according to the El Escorial and Airlie House diagnostic criteria: A population-based study. Arch. Neurol. 57, 1171–1176 (2000).

21. Scotton, W. J. et al. Prognostic categories for amyotrophic lateral sclerosis. Amyotroph Lateral Scler 13, 502–508 (2012).

22. Chancellor, A. M. et al. The prognosis of adult-onset motor neuron disease: a prospective study based on the Scottish Motor Neuron Disease Register. J. Neurol. 240, 339–346 (1993).

23. Forbes, R. B., Colville, S., Cran, G. W., Swingler, R. J. & Scottish Motor Neurone Disease Register. Unexpected decline in survival from amyotrophic lateral sclerosis/motor neurone disease. J. Neurol. Neurosurg. Psychiatr. 75, 1753–1755 (2004).

24. Lee, J. R., Annegers, J. F. & Appel, S. H. Prognosis of amyotrophic lateral sclerosis and the effect of referral selection. J. Neurol. Sci. 132, 207–215 (1995).

25. Chiò, A. et al. Early symptom progression rate is related to ALS outcome: a prospective population-based study. Neurology 59, 99–103 (2002).

26. Tysnes, O. B., Vollset, S. E. & Aarli, J. A. Epidemiology of amyotrophic lateral sclerosis in Hordaland county, western Norway. Acta Neurol. Scand. 83, 280–285 (1991).

27. Gundogdu, B., Al-Lahham, T., Kadlubar, F., Spencer, H. & Rudnicki, S. A. Racial differences in motor neuron disease. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration 15, 114–118 (2013).

Chapter 3

Genetic Characterization of a North American ALS cohort

The work presented in this chapter resulted in the following publications

1. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. Cady J, Allred P, Bali T, Pestronk A, Goate A, Miller TM, Mitra RD, Ravits J, Harms MB, Baloh RH.Ann Neurol. 2015 Jan;77(1):100-13

2. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Harms MB, Cady J, Zaidman C, Cooper P, Bali T, Allred P, Cruchaga C, Baughn M, Libby RT, Pestronk A, Goate A, Ravits J, Baloh RH. Neurobiol Aging. 2013 Sep;34(9):2234.e13-9

3.1 Abstract

To define the genetic landscape of amyotrophic lateral sclerosis (ALS) and assess the contribution of possible oligogenic inheritance, we comprehensively analyzed 17 known ALS genes in 391 ALS patients from the United States. Targeted pooled-sample sequencing was used to identify variants in 16 ALS genes and repeat-primed PCR was used to detect expansions in *C9ORF72* and *ATXN2*. 64.3% of familial and 27.8% of sporadic subjects carried potentially pathogenic novel or rare coding variants in the genes under investigation. 3.8% of subjects had variants in more than one ALS gene, and these individuals had disease onset ten years earlier (p=0.0046) than subjects with variants in a single gene. The number of potentially pathogenic

coding variants did not influence disease duration or site of onset. The rates of subjects carrying variants were significantly higher than previous reports using less comprehensive sequencing approaches. A significant number of subjects carried variants in more than one gene, which influenced the age of symptom onset and supports oligogenic inheritance as relevant to disease pathogenesis.

3.2 Introduction

Amyotrophic lateral sclerosis (ALS) is caused by degeneration of upper and lower motor neurons which results in progressive paralysis and ultimately death. As the most common motor neuron disease, the incidence of ALS is 0.44-3.2/100,000 person years[1] and data from the National ALS Registry demonstrates a prevalence of 3.9/100,000 cases in the US.[2] 5-10% of ALS patients have a family history of the disease (FALS)[3–5] and the genetic analysis of these FALS pedigrees has fueled the discovery of more than 20 ALS genes, some with high-penetrance and others with lower penetrance or tentative associations to disease (reviewed in Harms and Baloh, Andersen and Al-Chalabi).[6,7] Mutations in many of these genes are also found in patients without a family history of ALS (sporadic ALS [SALS]), with high-penetrance mutations found in ~10%.[8–14] Recently, the heritability of SALS has been estimated to be 12-21% from genome-wide association studies[15,16] and as high as 61% in twin studies[17] suggesting additional genetic influences on ALS risk remain to be identified.

The emergence of next-generation sequencing techniques has driven down sequencing costs and made it feasible for studies to abandon sequential candidate gene sequencing in favor of analyzing larger numbers of genes simultaneously. One of the more powerful and cost-effective sequencing techniques for screening moderate number of genes in medium sized

cohorts is termed pooled-sample or pooled-DNA sequencing (Figure 3.1).[18] In this method, DNA samples from multiple patients are pooled prior to PCR amplification of target regions. PCR products are then combined and sequenced *en masse* using short-read/next-generation sequencing platforms.[18] Analysis programs such as SPLINTER utilize statistical algorithms to identify potential variants with high sensitivity, and are capable of detecting single alleles in pools of up to 500 individuals.[19] Pooled-sample sequencing therefore overcomes the resource and time-intensive draw-backs of traditional Sanger sequencing approaches at a fraction of the cost.[18,19]

As a result of next-generation sequencing advances, studies have begun addressing the relative contributions of individual genes in ALS subjects with and without family histories, revealing significant heterogeneity between populations.[8–12,20] Furthermore, screening multiple ALS genes in parallel has also uncovered a number of patients that carry potentially pathogenic variants in more than one known ALS gene.[12] The unexpected frequency of this phenomenon has raised the hypothesis that some fraction of apparently sporadic ALS[8,12] could be caused by the co-occurrence of two or more genetic variants with additive or synergistic deleterious effects. Each variant alone could be tolerated but when combined with a second variant would exceed the threshold required for neurodegeneration. Although several papers have reported cases with multiple variants in ALS genes, no effect on phenotype or disease manifestations has been noted.[9,12]

We have used pooled-sample sequencing as the major technique to analyze 17 ALS-associated genes in 391 ALS subjects from a United States clinic-based cohort. In creating the most comprehensively-sequenced North American ALS cohort to date, this study measures the burden of rare and novel variants in known ALS genes and defines the frequency of potential

oligogenic cases. Importantly, we demonstrate that subjects with rare or unique variants in more than one ALS gene have earlier ages of disease onset.

3.3 Methods

3.3.1 Subjects

Between 2005 and 2011, patients diagnosed with ALS at the Washington University Neuromuscular Disease Center in St. Louis, Missouri (WUSM) or at the Virginia Mason Medical Center (VMMC) were systematically asked to participate in genetic studies. All subjects provided informed and written consent for clinical-genetic correlation studies of ALS that had been approved by institutional ethics review boards. At WUSM, subjects with or without a family history of ALS were included, while only sporadic cases were enrolled at VMMC. All subjects had been evaluated by neuromuscular specialists and diagnosed with probable or definite ALS according to El Escorial criteria.[21] A subset of included subjects (mostly with FALS) also underwent sequencing for one or more ALS genes at commercial reference laboratories, which identified 6 subjects with *SOD1* or *TARDBP* mutations.

3.3.2 Genetic investigations

Sequencing of ALS-associated genes: All coding exons and 20 flanking bases of *SOD1*, *FUS*, *TARDBP*, *ANG*, *OPTN*, *VCP*, *VAPB*, *DAO*, *DCTN1*, *FIG4*,*SETX*, *TAF15*, *EWSR1*, *UBQLN2* , *SQSTM1*, and *C9ORF72* were sequenced in our cohort using the pooled-sample method as previously described in detail and schematized in Figure 1.[18,22] Genomic DNA was extracted from whole blood or saliva of individual subjects according to standard protocols. Double-

stranded DNA was carefully quantified by fluorimetry based on SYBR gold fluorescence. Pooled-sample gDNA pools were then created by combining equimolar amounts of DNA from multiple individuals: two pools containing 21 samples each were used to validate the method, while the remaining samples were divided into 8 pools of 30-50 samples each.

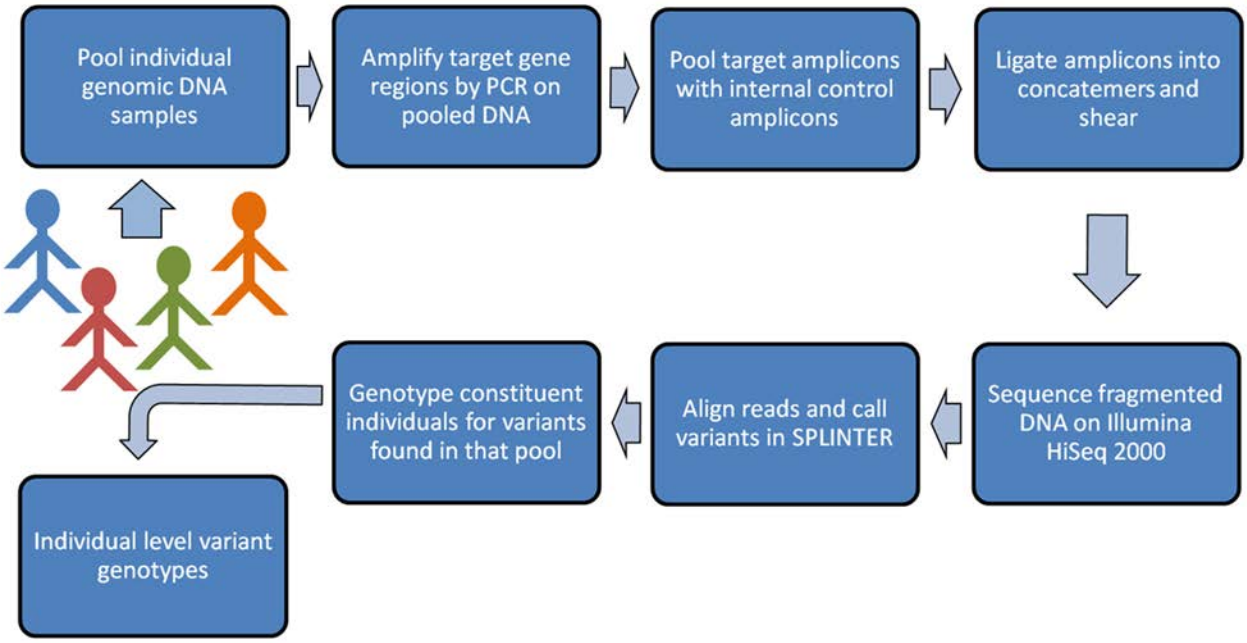

Figure 3.1 Schema of pooled-sample sequencing workflow

Primer pairs for all coding exons and at least 20bp of flanking sequence were designed using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and the RefSeq gene annotations found in GRCh37/hg19 (accession numbers NM_000454.4, NM_004960.3, NM_007375.3, NM_001145.4, NM_001008211.1, NM_007126.3, NM_004738.4, NM_001917.4, NM_004082.4, NM_014845.5, NM_015046.5, NM_139215.1, NM_013986.3, NM_013444.3, NM_003900.4, and NM_001256054). Primer sequences are available upon request. Amplicons from each pool were sequenced on one lane of HiSeq2000 (Illumina), with single-end 42bp reads. *UBQLN2*, *SQSTM1*, and *C9ORF72* were reported after initial sequencing

was underway and all subjects were sequenced as part of 6 pools across two lanes of Illumina HiSeq2000. Exon 1 of *SQSTM1* was not sufficiently covered using pooled-sample methods and required Sanger sequencing of each individual subject. Mutations in *PFN1* were reported after analysis was already underway so this gene was not assessed.[23]

Bioinformatic analysis: Sequence alignment and variant calling were performed using "short indel prediction by large deviation inference and nonlinear true frequency estimation by recursion" (SPLINTER).[19] The SPLINTER program generates an error model based on the negative control for each run. The error model is used to calculate a p-value for each SNP that is detected. SPLINTER calculates the p-value cutoff that has the highest sensitivity and specificity to distinguish true variants in the positive control vector and uses the ratio of sequencing reads with and without variant nucleotides to estimate the frequency of a given variant within a pool.

All variants called by SPLINTER were filtered for variants within exons or the 10 flanking bases and then visually inspected using Integrated Genomics Viewer(IGV)[24,25] after realignment to Hg19 using Novoalign (http://www.novocraft.com) and SAMtools.[26] Variants were annotated using SeattleSeq (http://sngs.washington.edu/SeattleSeqAnnotation131/), SIFT (http://sift.jcvi.org/), MutationTaster (http://www.mutationtaster.org/), and PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/). The effect of splice-site mutations was predicted by Human Splicing Finder (http://www.umd.be/HSF). Population frequencies for each variant were determined in dbSNP, the 1000 Genomes Project, and the NHLBI Exome Sequencing Project version ESP6500 exome variant server (ESP6500) (http://ESP6500.gs.washington.edu/ESP6500/ [1 Dec 2013]).

Variant validation and classification: Novel and rare (<1% MAF in ESP6500) non-synonymous and splicing variants that passed visual inspection were genotyped in individual

DNA samples by either Sequenom or Sanger sequencing to both validate the variant and determine which subject(s) carry them. Validated variants were assigned to four categories based on their presence in the ALS literature and frequencies in population databases. Category 1 variants have been previously reported in ALS patients but are absent from population databases. Category 2 variants have been reported in ALS patients but are present in population databases. Category 3 variants are novel (i.e. they have not been reported in ALS patients or population databases). Category 4 variants have not been reported in ALS patients but are present in population databases. Pathogenicity prediction algorithms were not utilized for category assignments because of their poor track-record in predicting disease-causing mutations.[27,28] All four categories of variants were considered to be potentially pathogenic mutations.

C9ORF72 repeat expansion detection: All subjects were also screened for *C9ORF72* repeat expansions using standard repeat-primed PCR.[29] Identification of a decrementing saw-tooth pattern with 6-bp periodicity and more than 30 peaks was considered positive for an expanded repeat, as in previous applications of this assay. In practice, however, all samples considered positive in this study showed more than 60 peaks.

ATXN2 repeat size: The CAG repeat region was amplified using primers 5' FAM-CCC CGC CCG GCG TGC GAG CCG GTG TAT G 3' and 5' CGG GCT TGC GGA CAT TGG 3'. PCR was performed with *Phusion*High-Fidelity *PCR* Master Mix with HF Buffer (New England BioLabs) with cycles as follows: 30 seconds 98°C, 35 cycles (10 seconds 98°C, 30 seconds 72°C), and 2 minutes 72°C. Repeat lengths were determined by fluorescent capillary gel electrophoresis. While intermediate-length alleles were originally considered to be 27-33 repeats[30], subsequent meta-analysis has shown 29 repeats to be the optimal cutoff to distinguish

ALS subjects from controls.[31] Therefore we considered repeat sizes of 29-33 to be of intermediate length.

3.3.3 Statistical Analysis

Statistical tests comparing patient characteristics were performed in R v3.0.1. Wilcoxon rank-sum tests were used to assess age of onset and survival, and Fisher's exact tests were used to analyze site of symptom onset, family history, gender, and ethnicity.

To identify rare and novel SNPs that might be over-represented in sporadic ALS subjects, we used Fisher's exact tests to compare each candidate SNV's allele frequency in sporadic ALS versus controls. By genotyping variants across a range of frequencies, we found that SPLINTER predicted frequencies and genotyped frequencies were highly correlated (r^2=0.9596) as in prior studies[19] (data not shown). Therefore SPLINTER-predicted frequencies were used for ALS SNPs that were not genotyped. We included only SALS samples with self-reported non-Hispanic white backgrounds (n=309) and used subjects with European ancestry in ESP6500 and the 1000 Genomes Project (1000genomes.org)[32] as our control population(n=4679). Variants were selected for replication based on p-values and potential functional significance. Selected variants were genotyped in 552 ALS cases and 464 neurologically normal controls from the Coriell DNA repository. Tests were performed in R v3.0.1.

Gene-based tests comparing the burden of rare variants in cases compared to controls were performed using SKAT-O.[33] We included the same individuals as were used for single-variant testing. Only missense and nonsense variants with MAF<0.1 in the control cohort were included in analysis. We used a Bonferonni correction to account for multiple testing

(α=0.0036). *UBQLN2* was excluded from analysis since SKAT does not handle data from the X-chromosome.

3.4 Results

3.4.1 Subject Characteristics

Demographic and disease characteristics for the 391 sequenced subjects with ALS are shown in Table 3.1. Age at onset, site of first symptom, and overall disease survival were similar to other population-based and referral center-based cohorts (reviewed in Harms and Baloh).[6] 42 subjects (10.7%) had a family history of ALS, which is comparable to other ALS referral center-based cohorts.

Table 3.1 Subject demographics and disease characteristics. Data for [a]ethnicity, [b]site of onset, and [c]age at onset was missing for 14, 5, and 4 subjects respectively. [d]Survival data was available for 172 subjects.

Total ALS cases	391
Subjects with family history of ALS	10.7%
Self-reported Caucasian (%)[a]	93.1%
Male sex (%)	57%
Limb Onset (%)[b]	69.2%
Age at Onset (mean, stdev)[c]	59.7±12.8
Age at Onset (median, range)	61,14-85
Survival in months (mean, stdev)[d]	41.3±27.7
Survival in months (median, range)	34, 7-147

3.4.2 Sequencing Results

In total, 152 PCR amplicons were required to amplify 203 exons of the 16 genes analyzed by pooled-sample sequencing. Twelve lanes of next-generation sequencing yielded 1.2 billion

total reads (~3 million per subject) to produce a coverage depth exceeding 67x per allele for all amplicons across all pools. Most amplicons showed considerably higher coverage (range: 67-987, median=458.68, IQR=355.98-560.51).

Sensitivity for single alleles (i.e. heterozygous variants present in a single individual within a given pool) was 98% (100% in 12 of 16 pools and 92% in the remainder), as determined by the detection rate of positive control singleton variants. We also compared pooled-sample results to whole-exome data for 35 subjects and found no missed variants in targeted genes. Finally, we detected all six previously-found mutations with the correctly assigned singleton frequencies.

To assess the false positive rate at the low allele frequencies in which we were interested, we performed validation genotyping of 100 allele calls for 67 non-synonymous or splice-site variants that were either rare (<1% minor allele frequency) or absent in population databases. 13 of 100 calls (8 SNPs, including 4 SNPs that were identified and validated in other pools) were not validated by subsequent genotyping, resulting in a false-positive rate of 13%. The false-positives included 5 calls that were true in other pools.

After filtering and validation, 67 rare or novel coding variants were identified (65 by pooled-sample sequencing and two by direct sequencing of exon 1 of *SQSTM1*). Variants were identified in all sequenced genes except for *UBQLN2*.

3.4.3 Variant Identification and Classification

Coverage of targeted bases was ≥130 fold for each subject. Sensitivity for detecting a variant present as a single allele within the pool of normal alleles averaged 98% (100% in 12 of

16 pools and 92% in the remainder). Based on validation genotyping, 13% of variants were determined to be false-positives.

In total, we found 66 rare or novel coding or splice-site variants (Table 3.2). One-third of these (n=23) were previously reported in ALS patients. Ten of these ALS-associated variants were not found in population databases of genetic variation (Category 1) and review of the literature showed that all of them are well-established causal mutations. The remainder of variants previously reported in ALS (n=13) were found to be present in population databases (Category 2). With the exception of *SOD1* D91A (where causality is clear), these variants lacked conclusive evidence of causation in the literature. Two-thirds (n=43) of variants we identified have not been previously reported in ALS, including 17 that are absent from population databases (Category 3) and 26 that are rare in population databases (Category 4).

Table 3.2 Novel and rare coding variants identified in ALS genes. Rare was considered a global minor allele frequency <1%. [a]GRCh37/hg19 [b]dbSNP138 [c]cDNA location and predicted protein changes refer to isoforms listed in Methods. [d]Allele counts are listed as alternate alleles found/total alleles assayed. [e]For all but the *ATXN2* and *C9ORF72* repeats, global allele counts were calculated from all subjects in the 1000 Genomes and NHLBI Exome Sequencing Projects. Global allele counts for *C9ORF72* repeat expansions were derived from[34] while intermediate CAG repeats in *ATXN2* were derived from [30,31]. [f]Population allele count refers to the population most closely matching that of the ALS subjects(s) carrying the variant. Unless indicated by a symbol, this is European ancestry (EA subjects from ESP6500 and 1000genomesEUR). Symbols used to denote other populations: †African American (AA subjects from ESP6500+1000genomesAFR); ‡Hispanic (1000genomesAMR) §Asian (1000genomesASN). *indicates at least one subject carrying that specific variant also carried another variant(s) in an analyzed ALS gene

	Gene Name	Genomic location[a]	dbSNP ID[b]	Predicted cDNA change[c]	Predicted protein change[c]	Allele Counts[d]			
						FALS	SALS	Global[e]	Population[f]
Expansions	*ATXN2*	12:112036785	rs193922927	c.532_534CAG	Q188(29-33)	1/84	11/698*	19/2982	19/2982
	C9ORF72	9:27573539	-			14/84*	21/698	11/7598	11/7598
Category 1: -Reported in ALS -Not in databases	*FUS*	16:31202739	rs121909668	c.1561C>G	R521G	1/84	0/698	0/15190	0/9358
	FUS	16:31202752	-	c.1574C>T	P525L	0/84*	1/698	0/15190	0/9358
	SOD1	21:33032096	rs121912442	c.14C>T	A5V	1/84*	1/698	0/15190	0/9358
	SOD1	21:33036142	rs121912431	c.112G>A	G38R	1/84*	0/698	0/15190	0/9358
	SOD1	21:33038821	-	c.229G>T	D77Y	1/84	0/698	0/15190	0/9358
	SOD1	21:33039600	-	c.269C>T	A90V	0/84	1/698	0/15190	0/9358
	SOD1	21:33039672	rs121912441	c.341T>C	I114T	1/84	1/698	0/15190	0/9358
	TARDBP	1:11082325	rs80356719	c.859G>A	G287S	0/84	1/698*	0/15190	0/9358
	TARDBP	1:11082409	rs80356726	c.943G>A	A315T	1/84	0/698	0/15190	0/9358
	VCP	9:35065360	rs121909329	c.464G>A	R155H	1/84	0/698	0/15190	0/9358
Category 2:	*ANG*	14:21161845	rs121909536	c.122A>T	K41I	0/84	1/698*	27/15190	23/9358

-Reported in ALS -Rare in databases	*ANG*	14:21162130	rs121909543	c.407C>T	P136L	1/84*	0/698	1/15190	1/9358
	DCTN1	2:74588717	rs72466496	c.3746C>T	T1249I	2/84*	1/698*	44/15190	39/9358
	DCTN1	2:74592252	rs72659383	c.3146G>A	R1049Q[35]	0/84	1/698*	22/15190	21/9358
	FIG4	6:110036336	rs121908287	c.122T>C	I41T[36]	0/84	1/698	17/15190	16/9358
	OPTN	10:13166053	rs142812715	c.941A>T	Q314L	0/84	1/698	3/15190	3/9358
	SETX	9:135140020	rs151117904	c.7640T>C	I2547T	2/84	8(1hom)/698*	76/15190	71/9358
	SETX	9:135202325	rs112089123	c.4660T>G	C1554G	0/84	6/698*	47/15190	40/9358
	SOD1	21:33039603	rs80265967	c.272A>C	D91A	1/84	2(1hom)/698	9/15190	9/9358
	SQSTM1	5:179248034	rs200396166	c.98T>T	A33V	0/84	2/698	6/15190	6/9358
	SQSTM1	5:179251013	rs145056421	c.457G>A	V153I	1/84*	0/698	9/15190	9/9358
	SQSTM1	5:179252184	rs11548633	c.712A>G	K238E	0/84	5/698*	41/15190	32/9358
	TAF15	17:34171525	rs200175347	c.1222C>T	R408C	0/84*	1/698	2/15190	2/9358
<u>Category 3:</u> -Not reported in ALS -Not in databases	*DAO*	12:109278977	-	c.194+1G>A	Splice donor	0/84	1/698	0/15190	0/9358
	DCTN1	2:74588653	-	c.3810C>A	H1270Q	0/84	1/698*	0/15190	0/9358
	DCTN1	2:74590527	-	c.3239C>T	S1080F	0/84	1/698	0/15190	0/4898†
	EWSR1	22:29682932	-	c.620C>G	T207S	0/84	1/698	0/15190	0/9358
	FIG4	6:110087935	-	c.1588_1589delTT	F530Ter	0/84	1/698	0/15190	0/9358
	FUS	16:31202282	-	c.1394-2delA	Splice site	1/84	0/698	0/15190	0/9358
	OPTN	10:13160964	-	c.703C>T	Q235Ter	0/84	1/698	0/15190	0/9358
	SETX	9:135202223	-	c.4762G>A	A1588T	0/84	1/698	0/15190	0/572§
	SETX	9:135203632	-	c.3353C>A	T1118K	0/84	1/698	0/15190	0/9358
	SETX	9:135206694	-	c.980A>T	E327V	0/84	1/698	0/15190	0/9358

	SETX	9:135210013	-	c.820A>G	M274V	0/84	1/698*	0/15190	0/9358
	SETX	9:135211743	-	c.658A>C	K220Q	0/84	1/698	0/15190	0/9358
	SETX	9:135211898	-	c.503G>A	R168Q	0/84	1/698*	0/15190	0/9358
	SETX	9:135224775	-	c.41C>T	T14I	0/84	1/698	0/15190	0/9358
	SOD1	21:33038791	-	c.199C>G	P67A	1/84*	0/698	0/15190	0/9358
	SQSTM1	5:179248079	-	c.143T>T	L48P	0/84	1/698	0/15190	0/9358
	TARDBP	1:11082589	-	c.1123A>G	S375G	0/84	1/698	0/15190	0/9358
Category 4: -Not reported in ALS -Rare in databases	*ANG*	14:21161973	rs17560	c.250A>G	K84E	0/84	1/698	70/15190	69/4898†
	DAO	12:109294259	rs4262766	c.992G>A	G331E	0/84	1/698	4/15190	0/9358
	DAO	12:109294301	rs143732132	c.1034C>T	S345F	1/84*	0/698	3/15190	3/9358
	DCTN1	2:74593101	rs145130328	c.2805C>G	I935M	0/84	1/698	4/15190	0/362‡
	DCTN1	2:74598723	rs55862001	c.586A>G	I196V	1/84	4/698*	77/15190	70/9358
	DCTN1	2:74604801	rs374419252	c.332C>G	S111C	0/84	1/698	1/15190	0/9358
	EWSR1	22:29682919	rs144503053	c.607T>A	S203T	0/84	1/698	1/15190	1/9358
	FIG4	6:110081543	rs142463699	c.1228A>C	T410P	0/84	1/698	1/15190	0/362‡
	FIG4	6:110107636	rs143531641	c.2080A>G	M694V	0/84	1/698*	3/15190	3/9358
	FIG4	6:110113852	rs375414729	c.2444T>C	F815S	0/84	1/698	1/15190	1/9358
	FUS	16:31201719	rs186547381	c.1292C>T	P431L	0/84	1/698	3/15190	2/9358
	FUS	16:31202343	rs201772423	c.1453C>T	R485W	0/84	1/698*	1/15190	1/9358
	SETX	9:135140063	rs202121071	c.7597C>T	H2533Y	1/84	0/698	1/15190	1/9358
	SETX	9:135147182	rs150673589	c.7114G>A	D2372N	0/84	1/698	19/15190	6/362‡
	SETX	9:135202120	rs140781535	c.4865C>T	P1622L	0/84	1/698	1/15190	0/9358
	SETX	9:135204004	rs149546633	c.2981A>G	D994G	0/84	1/698	31/15190	0/9358
	SETX	9:135204235	rs376022544	c.2750T>C	M917T	0/84	1/698	1/15190	1/9358

	SETX	9:135205116	rs139200312	c.1869A>C	E623D	0/84	1/698	3/15190	3/9358
	SETX	9:135205594	rs200614765	c.1391C>T	S464L	0/84	1/698	11/15190	6/9358
	SETX	9:135206706	rs372193033	c.968G>A	S323N	0/84	2/698*	1/15190	1/9358
	SETX	9:135218103	rs145438764	c.472T>G	L158V	0/84	1/698	53/15190	48/9358
	SQSTM1	5:179263547	rs201239306	c.1277C>T	A426V	0/84	1/698	1/15190	0/9358
	TAF15	17:34171358	rs140268553	c.1163G>A	R388H	0/84	1/698	7/15190	7/9358
	VAPB	20:57014075	rs146459055	c.390T>G	D130E	0/84	1/698	11/15190	11/9358
	VAPB	20:57016076	rs143144050	c.510G>A	M170I	0/84	5/698*	19/15190	18/9358

3.4.4 Variant Data by Gene

ANG

Three *ANG* variants were identified in ALS patients, all of which were also present in controls. Two variants have previously been associated with ALS. p.P136L (P112L) was identified in an FALS subject with an established SOD1 variant and was in 0.006% of controls. The variant was previously shown to have a deleterious effect on ANG function in cells.[37] pK41I (K17I) was found in ALS subjects in several studies [37–40] however it was also identified in a healthy control and a subject with a FUS variant. The variant was shown to have a deleterious effect on protein function[37], but is also present in 0.1% of public database controls. The p.K84E was previously unidentified in ALS, but is in 0.046% of controls. Only p.P136L was predicted to be damaging by at least two algorithms.

ATXN2

Twelve patients (11 SALS) had one intermediate length *ATXN2* alleles of 30-33 CAG repeats. One of these patients had alleles of 27 and 30 repeats, while the rest had 22 or 23 repeats on their second allele. A patient with a de novo *FUS* p.P525L variant had an ATXN2 allele of 31 repeats which was inherited from the unaffected father. An additional 11 patients had one allele of 27 repeats. Interestingly, two patients with 27 repeats also tested positive for the *C9ORF72* repeat expansion. Clinical characteristics of the group carrying intermediate expansions did not differ from the sporadic ALS cohort as a whole (Table 3.3).

Table 3.3 Comparison of patient characteristics of ATXN2 intermediate-length expansion carriers
[a]n=375, [b]n=365, [c]n=7, [d]n=165

	Intermediate (n=12)	Normal (n=379)	p-value
Age of onset (median,range)	63.5, 14-79	61, 21-85[a]	0.764
%FALS	8.3	10.8	1
%Self-reported Caucasian	91.7	93.2[b]	0.61
%Male	50	57.3	0.77
Survival in months (median, range)	34, 18-61[c]	34, 7-147[d]	0.95

C9ORF72

There were 35 subjects that tested positive for the full hexanucleotide repeat expansion in *C9ORF72* (14 FALS, 21 SALS). In addition, we identified one missense variant by pooled-sample sequencing, p.T49R in an FALS subject who also carried the known *FUS* p.R521G variant (Table 3.4). We were unable to test for segregation of p.T49R; however an affected family member was later confirmed by commercial sequencing to carry the R521G mutation. This variant is present in EVS and 1000genomes at a frequency of 0.01% combined. The p.T49R variant is predicted to be pathogenic by MutationTaster, but not by PolyPhen2. SIFT was unable to give a prediction. We did not consider this variant to be pathogenic since there is no indication that point mutations in *C9ORF72* are causative.

DAO

A novel splice-site variant in *DAO* was discovered in an SALS subject. The c.194+1G>A variant is predicted to destroy the splice donor site the first coding exon, but tissue from this individual was not available to determine DAO levels or effects on splicing. We identified two additional rare variants in the gene: p.G331E in one SALS and p.S345F in one FALS subject. Both were damaging in at least two predictions.

DCTN1

We detected two novel variants in *DCTN1*, both in SALS subjects. The p.S1080F variant was predicted to be pathogenic by two of three algorithms, while p.H1270Q variant was predicted to be benign by all three. The p.T1249I variant, which was found in 2 FALS and 1 SALS subjects, was originally considered pathogenic. It has since been shown that the variant does not segregate with disease and is present in multiple control cohorts (0.29% in ESP6500 and 1000genomes combined) [35]. The p.R1049Q variant was identified in patients with Parkinson's disease and patients with frontotemportal lobar degeneration[35] and is predicted to be damaging by MutationTaster and PolyPhen2. There was one SALS subject in our cohort with this variant and 0.14% of controls. We identified three more rare variants were found in our cohort; p.I935M was found in 1 SALS subject and 4 controls (0.026%) and has two damaging predictions, p.S111C was found in 1 SALS subject and 1 control (0.006%) and has three damaging predictions, and p.I196V was found in 1 FALS and 4 SALS subjects and 77 controls (0.51%). p.I196V is only predicted to be damaging by MutationTaster.

EWSR1

We identified only two variants in *EWSR1*: a novel variant (p.T207S) in an SALS patient and a rare variant (p.S203T) also in an SALS patient. Previous studies of *EWSR1* have focused on the C-terminal domain, where pathogenic variants in the related genes *TARDBP*, *FUS*, and *TAF15* are found.[41] Both of these variants fall within the prion domain, a regions that has not previously been screened for variants. The novel variant is not predicted as damaging, but the rare variant (present in 1 control out of 7595) is predicted damaging by MutationTaster and PolyPhen2.

FIG4

Five variants were identified in *FIG4*. A novel 2bp deletion, c.1588_1589delTT, in an SALS subject results in the formation of a stop codon (ttaTTTgag->ttaTGAg) and is predicted to cause nonsense-mediated decay. The p.I41T variant identified in an SALS patient was reported in a compound heterozygous CMT4J patient [36], and is found at a MAF of 0.11% in controls. p.I41T was predicted to be damaging by all three predictions. Three additional variants have not been reported in SALS and were very rare in controls. p.T410P and p.F815S were each only found in one control (0.006%) but had less than two damaging predictions and p.M694V was found in three (0.02%) and had two damaging predictions.

FUS

Five patients had variants of interest in *FUS*. The known FUS p.R521G[42] variant was found in an FALS subject. This variant was confirmed in an additional affected family member by commercial testing. A patient with disease onset at age 14 had the p.P525L variant. Both parents were negative for the variant. This variant has been previously identified in juvenile onset cases as a *de novo* variant.[42] We identified a novel deletion c.1394-2delA in the splice acceptor site of exon 13. An A>G substitution has been reported in the same position and was shown to cause skipping of exon 14.[43] DNA from the affected father was unavailable to test for segregation. Two variants in FUS p.R485W and p.P431L were present in controls but at very low frequencies (only 1 sample and 3 samples respectively out of 7595). pR485W was predicted to be damaging by SIFT and MutationTaster, but not PolyPhen2 while p.P431L was predicted to be damaging by all three.

OPTN

We found a novel *OPTN* nonsense variant (p.Q235X) in an apparently sporadic subject. Nonsense variants both upstream and downstream of this codon have been reported in other subjects with ALS. p.Q23X and p.Q165X were both dominantly inherited and predicted by SIFT to undergo nonsense-mediated decay (NMD).[44,45] The downstream variant p.Q398X was recessively inherited and NMD was experimentally observed.[46,47] We also identified the variant p.Q314L in one SALS subject. This variant was reported as a disease variant[44], but is now present in controls at a very low frequency (0.02%). It is predicted to be damaging by all three algorithms.

SETX

A notable finding was the abundance of novel *SETX* variants. We found 18 *SETX* variants in 30 subjects. Seven of the variant were novel. Two variants were novel, but in the same residue as a previously reported variant. We identified a p.T1118K variant (p.T1118I was identified in a Chinese cohort[48]) and a p.M274V variant (p.M274I was reported in an autosomal recessive ataxia with peripheral neuropathy[49]) both in SALS patients. The remaining 5 novel variants (p.E327V, p. T14I, p.R168Q, p.A1588T, and p.K220Q) were identified in SALS patients as well. Only p.A1588T and p.M274V were not damaging in at least two predictions.

There were 8 SALS and 2 FALS subjects that carried the p.I2547T variant which has previously been reported in ALS.[50] One SALS subject was homozygous for the variant. Recessive mutations in *SETX* are the cause of ataxia-ocular apraxia 2 (AOA2)[51] and to our knowledge, this is the first report of a possibly recessive mutations in *SETX* causing ALS. Another patient was compound heterozygous for the p.I2547T variant along with the novel p.T14I variant, however they also carried a pathogenic p.R408C variant in *TAF15.* The variant is present at a MAF of 0.5% in controls with one homozygous subject out of 7595. It is predicted to

be damaging by MutationTaster, but not SIFT or PolyPhen2. The variant p.C1554G was also reported as pathogenic[50] and is present in controls (MAF=0.3%). One subject was compound heterozygous for this variant and a novel p.R168Q variant. It was found in 5 additional SALS subjects and is predicted to be damaging by all three predictions.

Nine of the *SETX* variants were never reported in ALS and are rare in controls. Five of these (p.H2533Y (0.006%), p.D2372N (0.1%), p.P1622L (0.006%), p.D994G (0.2%), and p.M917T (0.006%)) were predicted to be damaging in fewer than two predictions. p.E623D (0.02%), p.S464L (0.07%) and p.S323N (0.006%) were predicted to be damaging by two algorithms and p.L158V (0.3%) was predicted to be damaging by all three. Despite the high number of variants in *SETX*, we did not find a significant association with individual SNPs or combined across the gene.

SOD1

We identified 7 different *SOD1* variants in 10 subjects: p.A5V(A4V)[52], p.G38R(G37R)[52], p.D77Y(D76Y)[53], p.A90V(A89V)[54], p.D91A(D90A)[55], p.I114T(I113T)[52], and the novel variant p.67S. The 6 known variants were all predicted to be damaging by at least two predictions, except for p.D91A, which was not predicted to be damaging in any. The p.D91A variant has been observed in both recessive and dominant inheritance.[55] We identified one SALS subject that was homozygous for the variant. One individual with definite FALS was heterozygous for the p.D91A variant, but two affected family members in their pedigree that were also included in the study did not possess this variant. Instead, those two individuals were positive for the *C9ORF72* repeat expansion. Previously known p.A5V and p.I114T variants were detected in SALS subjects. DNA from family members was unavailable for segregation testing in both cases. We also discovered a novel p.P67A variant in a subject with probable FALS that is likely

causative. This variant segregated with disease in additional affected family members and is predicted to be damaging by PolyPhen2, SIFT, and MutationTaster. Furthermore, two variants in the same amino-acid (p.P67S(P66S)[56] and p.P67R(P66R)[40]) have previously been reported to cause disease.

SQSTM1

Four variants in *SQSTM1* were identified in SALS patients and one in an FALS patient. p.L48P and was novel and predicted pathogenic in two out of three predictions. The p.A33V (2 SALS), p.V153I (1 FALS), and p.K283E (5 SALS) variants were previously identified as pathogenic[57,58], but were seen in 0.04%, 0.06%, and 0.27% of controls respectively. p.A33V and p.K283E were both predicted damaging at least twice, but p.V153I was only predicted to be damaging once. One more variant with a MAF of 0.006% in controls, p.A46V, was identified in an SALS subject. This variant was predicted to be damaging twice.

TAF15

Two *TAF15* variants, p.R408C and p.R388H, were present in SALS subjects. Both of these variants were previously found in ALS patients; however p.R388H was identified in controls at the same time and was not considered pathogenic. The p.R408C variant has now also been found in controls, but was previously shown to result in cytoplasmic foci in transfected cells.[59] Both variants were predicted to be damaging in at least two predictions.

TARDBP

We identified three *TARDBP* variants, one of which is novel. A subject from a previously reported family was found to carry p.A315T variant.[60] A sporadic subject was found to carry a known pathogenic variant, p.G287S.[61] Finally, a novel variant, p.S375G, was found in

another SALS subject. The variant was inherited from a parent who was unaffected at age 80 while the subject with ALS had age of onset of 41. It is unclear whether this variant is benign or pathogenic and incompletely penetrant. The p.S375G was not predicted to be damaging by any of the three algorithms used, however the two known pathogenic variants are both only predicted to be damaging by MutationTaster.

VAPB

There were no novel or disease-associated variants in *VAPB*. Two variants, p.D130E and p.M170I, were rare in controls (0.072% and 0.13% respectively). p.D130E was in 1 SALS subject and was not predicted to be damaging. p.M170I was found in 5 SALS subjects and was predicted to be damaging by MutationTaster and PolyPhen2.

VCP

One known *VCP* variant p.R155H[62] was detected in a subject with FALS and segregated with disease among siblings, one of whom has Paget's disease. Their affected parent died of ALS complicated by non-Alzheimer dementia.

3.4.5 Prevalence of variants in ALS genes

We considered 65 of the rare and novel variants identified by sequencing to be potentially pathogenic. 83 subjects (21.2% overall, 35.7% in FALS and 19.5% in SALS) carried at least one of these variants. The *C9ORF72* repeat expansion (found in 8.7% of subjects, n=34) and *ATXN2* intermediate-length CAG repeats (found in 3.1% of subjects, n=12) were also considered to be potentially pathogenic. In total, 124 subjects (31.7% overall; 64.3% across all categories of FALS, 27.8% in SALS) carried one or more of these potentially pathogenic variants (Supplementary Table S1), a higher number than reported in many prior studies.[8–12,14]

The proportion of subjects carrying a potentially pathogenic variant was heavily influenced by the strength of evidence for familial transmission of ALS (Table 3.4), with the highest rate of variant discovery in definite FALS (81.6%). The frequency of variants declined with less evidence for transmission, but even 27.8% of sporadic/simplex subjects were carriers.

Table 3.4 Prevalence of variants in ALS genes by family categorization. Familial ALS categories were assigned according to recently proposed criteria.[13,63] Definite FALS (38% of families): at least 2 first- or second-degree relatives with ALS. Probable A FALS (7% of families): 1 first-degree relative with ALS. Probable B FALS (43% of families): 1 second-degree relative with ALS. Possible FALS (12% of families): 1 distant relative with ALS. Sporadic ALS (89% of entire cohort): all subjects not meeting criteria for any FALS category. ALS = amyotrophic lateral sclerosis; FALS = familial ALS.

	Sporadic	All FALS	Definite	Probable A	Probable B	Possible
Total Subjects	349	42	16	3	18	5
ANG	2	1	1	-	-	-
ATXN2	11	1	-	-	1	-
C9ORF72	21	15	8	1	6	-
DAO	2	1	1	-	-	-
DCTN1	10	3	2	-	-	1
EWSR1	2	-	-	-	-	-
FIG4	5	-	-	-	-	-
FUS	3	2	1	-	-	1
OPTN	2	-	-	-	-	-
SETX	29	3	1	-	2	-
SOD1	4	6	4	-	2	-
SQSTM1	9	1	-	-	1	-
TAF15	2	-	-	-	-	-
TARDBP	2	1	1	-	-	-
UBQLN2	-	-	-	-	-	-
VAPB	6	-	-	-	-	-
VCP	-	1	-	1	-	-
Total Variants	110	35	19	2	12	2
Subjects with any variant	97	27	13	2	10	2
% Subjects with variants	27.8%	64.3%	81.3%	66.7%	55.6%	40.0%

We identified 4 sporadic subjects with potentially recessive causes of their ALS (Table 3.5). One subject was homozygous for *SOD1* p.D91A (D90A), while three others carried two mutations in *SETX*. One subject tested homozygous for *SETX* p.I2547T, but we did not exclude

the possibility of a deletion on one allele. The two additional subjects could each be compound heterozygotes comprised of a rare variant (p.C1554G or p.I2547T with 0.3% and 0.5% MAF in population database respectively) and a novel variant (p.R168Q or p.T14I respectively). The subject carrying p.I2547T and p.T14I was also heterozygous for *TAF15* p.R408C which has previously been reported in a subject with SALS.[59] Due to the absence of additional family members for segregation or tissue for cDNA sequencing, we were unable to determine if these *SETX* variants are in *cis* or *trans*. Because recessive mutations in *SETX* are associated with ataxia-ocular apraxia type 2 (OMIM 606002) and *SETX*-associated ALS is dominantly inherited, we reviewed the medical records of these 3 individuals. All three showed typical ALS disease course without clinically apparent eye movement abnormalities or ataxia.

We also identified a pedigree with FALS with independently-segregating causative mutations (Figure 2). The proband, three affected siblings and a first cousin once-removed all tested positive for the *C9ORF72* repeat expansion. Another first cousin once-removed was diagnosed with ALS at another center test but tested negative for the expansion, including by Southern blot (See figure X in chapterX). Instead, this individual was found to carry a heterozygous *SOD1* p.D91A mutation detected by pooled-sample sequencing.

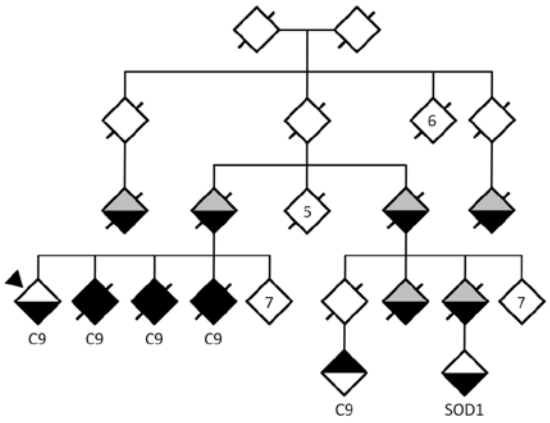

Figure 3.2 Segregation of independent mutations in an ALS family

3.4.6 Prevalence of potential oligogenic subjects

We assessed the number of genes with potentially pathogenic rare variants in each individual. Fifteen subjects (3.8% overall, 14% in FALS, 2.6% in SALS) harbored potentially pathogenic variants in at least two ALS genes: 11 with variants in two ALS genes, while 4 had variants in three genes each (Table 3.5). Although one subject had a *FUS* p.R521G variant and a *C9ORF72* p.T49R variant, they were not included as potentially oligogenic since we did not consider point mutations in *C9ORF72* as pathogenic.

Six potentially oligogenic subjects had a family history of ALS subjects and in all cases one of their variants was either the *C9ORF72* repeat expansion or a missense variant in *SOD1* in combination with additional rare or novel variant(s), several of which have also been previously reported in ALS subjects. Interestingly, one FALS proband carried 3 variants, each of which has previously been reported as pathogenic: *SOD1* p.G38R, *ANG* p.P136L, and *DCTN1* p.T1249I.

Nine apparently sporadic subjects had variants in multiple genes (Table 3.5), but only two were well-established ALS mutations: *TARDBP* p.G287S was found in combination with *VAPB* p.M170I while a subject with juvenile-onset ALS carried a *de novo FUS* p.P525L mutation with a paternally-inherited intermediate-sized CAG expansion in *ATXN2*. Two SALS patients carried multiple ALS-associated variants that are rare in population databases (*ANG* p.K41I with *VAPB* p.M170I and *TAF15* p.R408C with *SETX* p.I2547T and *SETX* p.T14I).

To compare the frequency of “oligogenic” cases in ALS to other populations, we used exomes from 114 subjects diagnosed with CMT1 (n=4), dHMN (n=1), EA (n=5), EVO (n=2), GSM1 (n=7), IBM (n=40), LGMD (n=5), MDC (n=11), MELAS (n=1), MFM (n=6), MPD

(n=8) , NVS (n=1), SFN (n=20), SMA (n=3). We compared these exomes to 614 ALS subjects (some of which were also included in pooled-sample sequencing). Since the non-ALS subjects were not screened for repeat expansions in *C9ORF72* and *ATXN2*, we performed our comparison using only the 15 other genes under investigation. The proportion of subjects carrying variants in more than one gene was not higher in ALS subjects (6.02%) compared to non ALS subjects (6.1%).

Table 3.5 Subjects with multiple variants in ALS genes

	Variant 1	**Variant 2**	**Variant 3**	**Possible model**
Sporadic	*SOD1*(p.D91A)	*SOD1*(p.D91A)	-	Homozygous recessive
Sporadic	*SETX*(p.I2547T)	*SETX*(p.I2547T)	-	Homozygous recessive
Sporadic	*SETX*(p.C1554G)	*SETX*(p.R168Q)	-	Potential compound het
Familial	*SOD1*(p.A5V)	*DAO*(p.S345F)	-	Oligogenic
Familial	*SOD1*(p.P67A)	*SETX*(p.I2547T)	-	Oligogenic
Familial	*C9ORF72*	*DCTN1*(p.I196V)	-	Oligogenic
Familial	*C9ORF72*	*SQSTM1*(p.V153I)	-	Oligogenic
Familial	*C9ORF72*	*SETX*(p.I2547T)	-	Oligogenic
Familial	*FUS*(p.R521G)	*C9ORF72*(p.T49R)	-	-
Sporadic	*ANG*(p.K41I)	*VAPB*(p.M170I)	-	Oligogenic
Sporadic	*ATXN2*(22/31)	*SQSTM1*(p.K238E)	-	Oligogenic
Sporadic	*FUS*(p.R485W)	*SETX*(p.I2547T)	-	Oligogenic
Sporadic	*DCTN1*(p.R1049Q)	*SETX*(p.S323N)	-	Oligogenic
Sporadic	*FUS*(p.P525L)	*ATXN2*(23/31)	-	Oligogenic
Sporadic	*TARDBP*(p.G287S)	*VAPB*(p.M170I)	-	Oligogenic
Familial	*SOD1*(p.G38R)	*ANG*(p.P136L)	*DCTN1*(p.T1249I)	Oligogenic
Sporadic	*ATXN2*(22/32)	*DCTN1*(p.T1249I)	*SETX*(p.M274V)	Oligogenic
Sporadic	*TAF15*(p.R408C)	*SETX*(p.I2547T)	*SETX*(p.T14I)	Oligogenic, potential compound het
Sporadic	*SETX*(p.C1554G)	*DCTN1*(p.H1270Q)	*FIG4*(p.M694V)	Oligogenic

3.4.7 Correlation of variant genes with disease characteristics

In an oligogenic model of disease, the additive or synergistic effects of multiple variants can influence not only the risk of developing disease, but also phenotypic manifestations of the

disease. Age at symptom onset was significantly earlier in cases carrying variants in multiple genes (median=46, IQR=39-61) compared to all other subjects (median=61, IQR=51-70, p=0.0046) and when compared to cases with mutations in just one gene (median=60, IQR=48-60, p=0.017). Even when the subject with juvenile onset was removed a difference of 10 years earlier remained (median=50.5, IQR=40.25-61.5, p=0.013 against all others and p=0.041 against other single-gene variant carriers). Furthermore, there was a weak, but statistically significant negative correlation between age of onset and the number of genes with variants (spearman's rho=-0.11, p=0.024). The number of ALS genes with variants did not influence disease duration or site of onset in our cohort.

3.4.8 Rare variants as modifiers of ALS risk

We also used our sequencing results to search for single variants in known ALS genes that increased or decreased ALS risk. To do so, we analyzed all coding variants found in our ALS cohort and also present in population databases (n=61, with 47 having a population MAF <1%). Three SNPs in *SETX* (rs1183768, rs543573, and rs2296871) were in perfect linkage disequilibrium and were considered to be one signal represented by rs2296871. We included only ALS subjects of European ancestry and compared to controls of European ancestry from ESP6500 and the 1000 Genomes Project. SPLINTER-predicted allele frequencies were used for common variants that were not confirmed by genotyping in ALS subjects. Using a Bonferonni-corrected significance level of 8.2×10^{-4}, 3 variants were significantly more common in our ALS discovery cohort (rs3739927 and rs882709 in *SETX*, and rs41311143 in *EWSR1*). To follow up, we genotyped these 3 SNPs and 28 additional candidate variants in a validation cohort of 552 sporadic ALS cases and 464 controls from Coriell reference panels. However, none of the 31

tested variants showed a significant association with ALS in either direction (Supplementary Table S2).

We also asked whether the burden of rare coding variants in any of the tested ALS genes was higher in sporadic subjects compared to controls using SKAT (Table 3.6).[33] After correcting for multiple tests ($\alpha=3.57\times10^{-3}$), *SOD1* was the only gene that showed a significant association ($p=1.59\times10^{-5}$) while *TARDBP* and *VAPB* approached statistical significance ($p=5.57\times10^{-3}$ and $p=5.99\times10^{-3}$ respectively).

Table 3.6 Gene-based rare variant association tests. Association tests were performed with SKAT using the optimal.adj method and the default linear, weighted kernel, with significance level=3.57×10^{-3}. Only coding variants with minor allele frequencies <1% were included in the analysis. Only subjects of European Ancestry were used from our cohort and controls (ESP6500+1000genomesEUR).

Gene	P-value	# Markers
SOD1	1.59×10^{-3}	4
TARDBP	5.57×10^{-3}	10
VAPB	5.99×10^{-3}	8
SQSTM1	0.126	39
SETX	0.165	125
FUS	0.323	25
DAO	0.425	26
DCTN1	0.443	58
EWSR1	0.450	21
ANG	0.487	9
TAF15	0.573	34
VCP	0.693	5
OPTN	0.765	20
FIG4	0.863	32

3.5 Discussion

Rapid progress toward defining the genetic landscape of ALS has been fueled by the emergence of next-generation sequencing. In this study, we used the efficiency and power of pooled-sample sequencing to investigate the frequency of pathogenic and potentially-pathogenic variants in known ALS genes in a large cohort of US patients. Our approach produced highly accurate sequence data for 16 known genes in a time, sample, and cost-efficient manner. We estimated that this study required 83% less DNA per subject and cost 10% of performing the equivalent study by traditional Sanger sequencing. In doing so, we have generated the most comprehensively sequenced North American cohort to date.

In this group of subjects we identified 27 novel variants (i.e. not found in databases of variation) and an additional 39 that are very rare in control populations. Not surprisingly, the highest rate of variant detection occurred in families with the strongest ALS histories: we found explanatory mutations in 80% of these pedigrees. This rate is higher than many previous reports of all FALS[8–14] and partially stems from our use of a strict definition of familial ALS favoring pedigrees with clearly dominant transmission patterns that undoubtedly enrich for Mendelian genes. Our elevated variant detection rate is also influenced by the large number genes analyzed in each family. Because our cohort was a clinic-based, we cannot address whether differences in populations are also involved.

Although the frequency of variant detection in our sporadic ALS subjects was lower than in familial ALS, it was still 28%. This is considerably higher than other studies [8–12,14], likely due in part to the large number of genes we sequenced. In support of this, we note that the frequency of variants in commonly sequenced genes (e.g. *C9ORF72, SOD1, TARDBP*) was within

previously reported ranges. To directly compare our findings with a similar study of an Irish population[9], we limited both data sets to genes shared between the two studies and only included novel variants (i.e. not seen in any population database). The total number of subjects found with at least one potentially pathogenic mutation was 16.4% in this study compared to 12.8% in the Irish population. This difference is not statistically significant (p=0.12) and was driven by the absence of SOD1 mutations in the Irish cohort. These broad differences in populations need to be given appropriate consideration when genetic testing or counseling is being provided to patients.

Based on previous reports of oligogenic inheritance in ALS, we looked for subjects with potentially pathogenic variants in more than one ALS gene. We found mutations in at least two ALS genes in 3.8% of our subjects (14% in FALS, 2.6% in SALS). In most cases, 1 of the identified variants is a known mutation with clearly established pathogenicity; however, many of the additional variants are of unknown significance. It is possible that these additional variants co-occur with pathogenic mutations by chance. However, the finding that subjects with potentially pathogenic variants in >1 gene had disease onset 10 years earlier than other subjects supports a model of ALS where the additive or synergistic effects of multiple defective genes increase risk and influence disease phenotype.

The rate of potentially oligogenic cases in our cohort is higher than in prior reports, but direct comparisons are prevented by differences in i) which genes were sequenced, ii) how complete variant ascertainment was, iii) relative numbers of familial and sporadic cases, and iv) which variants were considered to be potentially pathogenic. We were able to use exome sequencing from subjects with ALS to compare the frequency of subjects with mutations in more than one ALS gene to other non-ALS disorders. Both cohorts had a rate of ~6% overall.

Although this is higher than the rate we discovered in our pooled-sample sequencing cohort, this could be explained by the different sequencing techniques. In contrast to the ALS cohort the majority of non-ALS cases did not carry a well-established ALS mutation. An equal rate of oligogenic cases in ALS and non-ALS cases supports the hypothesis that carrying mutations in multiple genes is not a risk factor for disease, but influences disease progression.

This study evaluated known ALS genes only. With many efforts underway to generate exome and genome-wide variant data on large numbers of ALS patients, these types of interactions should become easier to detect and validate. These large datasets should also allow unbiased searches for new ALS genes using rare variant burden testing. As a test of this principle, we asked whether rare variant burden testing would identify any of the known ALS genes we sequenced. In our modestly-sized cohort we demonstrated a significant association for *SOD1* and suggestive associations for *TARDBP* and *VAPB*. We also noticed an abundance of variants in the *SETX* gene, an intriguing finding that was also evident in a prior study.[9] These findings predict that well-powered genome-wide studies will identify new ALS genes.

Our study also highlights important lessons regarding mutation screening in ALS. First, a significant number of individuals will harbor more than one potentially pathogenic mutation. This fact dramatically influences estimates of transmission risk and even prognosis. Therefore, comprehensive screening of known genes is preferable to single-gene testing and made more cost-effective by next-generation approaches to sequencing. Second, as our pedigree with independently segregating *SOD1* and *C9ORF72* mutations demonstrates, even once a causative mutation has been identified in a pedigree, each affected individual should be sequenced for confirmation. Third, despite the frequency with which our study found variants in ALS subjects,

36% of FALS and 74% of SALS subjects had no variants in any of 17 ALS genes we analyzed. Efforts are therefore needed to identify additional genes influencing ALS risk.

Finally, we note that many of the novel and rare variants identified by this study and others are of unknown significance and will require further study to validate a possible contribution to ALS pathogenesis. The complexity of determining pathogenicity of variants is highlighted by the 13 variants we identified that had been previously associated with ALS but have since been found in control databases at rates higher than expected for moderate or high penetrance mutations. Although these variants could represent mutations with reduced penetrance, or the presence of pre-symptomatic individuals in control populations, they most likely result from including limited controls in the original studies. In fact, many variants previously reported as pathogenic for ALS and other diseases are now found in the 1000 Genomes Project or the Exome Sequencing Project at frequencies exceeding those expected for moderately or highly penetrant mutations.[64] To prevent the literature from becoming confused with disease-associated variants that are not pathogenic, we support increased attention to variants are reported in disease populations, including the creation of levels of genetic evidence for pathogenicity as recently proposed.[65]

3.6 Supplementary tables

Table S1. Clinical characteristics of variant carriers.

Patient Variant(s)	Family Type	Ethnicity	Sex	Age of Onset (years)	Site of onset	Survival (months)
ANG(p.K41I), *VAPB*(p.M170I)	Simplex	Caucasian	Female	45	Limb	unknown
ANG(p.K84E)	Simplex	African American	Female	43	Limb	>78
ATXN2(22/30)	Simplex	Caucasian	Male	65	Limb	unknown
ATXN2(22/30)	Simplex	Caucasian	Female	57	Unknown	unknown
ATXN2(22/30)	Simplex	Caucasian	Male	66	Limb	61
ATXN2(22/31)	Simplex	Caucasian	Female	79	Limb	unknown
ATXN2(22/31)	Probable B	Caucasian	Male	70	Bulbar	33
ATXN2(22/31)	Simplex	Caucasian	Female	47	Limb	42
ATXN2(22/32)	Simplex	Caucasian	Male	71	Limb	30
ATXN2(23/30)	Simplex	African American	Male	41	Limb	>25
ATXN2(27/30)	Simplex	Caucasian	Female	67	Bulbar	22
C9ORF72	Simplex	Unknown	Female	58	Limb	unknown
C9ORF72	Simplex	Caucasian	Female	60	Limb	unknown
C9ORF72	Simplex	Caucasian	Female	59	Limb	unknown
C9ORF72	Simplex	Caucasian	Male	59	Limb	unknown
C9ORF72	Simplex	Caucasian	Male	72	Limb	unknown
C9ORF72	Simplex	Caucasian	Female	44	Bulbar	29
C9ORF72	Simplex	Caucasian	Male	58	Bulbar	25
C9ORF72	Definite	Caucasian	Male	57	Bulbar	37
C9ORF72	Simplex	Caucasian	Male	50	Limb	29
C9ORF72	Probable B	Caucasian	Female	63	Limb	12
C9ORF72	Definite	Caucasian	Male	45	Limb	>71
C9ORF72	Definite	Caucasian	Male	46	Limb	57
C9ORF72	Definite	Caucasian	Male	64	Limb	62
C9ORF72	Simplex	Caucasian	Male	53	Limb	31
C9ORF72	Probable B	Caucasian	Male	46	Limb	>40

C9ORF72	Simplex	Caucasian	Male	69	Limb	19
C9ORF72	Simplex	Caucasian	Male	48	Bulbar	28
C9ORF72	Definite	Caucasian	Female	51	Bulbar	29
C9ORF72	Simplex	Caucasian	Female	57	Limb	21
C9ORF72	Simplex	African American	Female	52	Limb	24
C9ORF72	Simplex	Caucasian	Male	65	Limb	54
C9ORF72	Probable B	Caucasian	Female	48	Bulbar	41
C9ORF72	Simplex	Caucasian	Female	57	Limb	42
C9ORF72	Simplex	Caucasian	Female	51	Limb	27
C9ORF72	Simplex	Caucasian	Female	56	Bulbar	30
C9ORF72	Probable B	Caucasian	Male	44	FTD	146
C9ORF72	Probable B	Caucasian	Male	63	Limb	46
C9ORF72	Probable A	Caucasian	Female	76	Limb	17
C9ORF72	Simplex	Caucasian	Female	63	Limb	36
C9ORF72	Simplex	Caucasian	Male	56	Limb	46
C9ORF72	Simplex	Caucasian	Female	70	Limb	27
C9ORF72	Simplex	Caucasian	Female	63	Limb	50
C9ORF72, *DCTN1*(p.I196V)	Definite	Caucasian	Male	60	Bulbar	20
C9ORF72, *SETX*(p.I2547T)	Definite	Caucasian	Male	46	Bulbar	33
C9ORF72, *SQSTM1*(p.V153I)	Probable B	Caucasian	Female	63	Limb	32
DAO(p.G331E)	Simplex	Caucasian	Male	57	Limb	unknown
DAO(splice donor)	Simplex	Caucasian	Female	71	Bulbar	29
DCTN1(p.H1270Q), *FIG4*(p.M694V), *SETX*(p.C1554G)	Simplex	Caucasian	Female	41	Limb	>50
DCTN1(p.I196V)	Simplex	Caucasian	Male	57	Bulbar	unknown
DCTN1(p.I196V)	Simplex	Caucasian	Female	30	Limb	unknown
DCTN1(p.I196V)	Simplex	Caucasian	Female	38	Limb	unknown
DCTN1(p.I196V)	Simplex	Caucasian	Female	69	Bulbar	22

DCTN1(p.I935M)	Simplex	Hispanic	Male	81	Limb	unknown
DCTN1(p.R1049Q), *SETX*(p.S323N)	Simplex	Caucasian	Male	69	Limb	38
DCTN1(p.S1080F)	Simplex	African American	Male	64	Limb	>135
DCTN1(p.S111C)	Simplex	Caucasian	Female	71	Limb	>54
DCTN1(p.T1249I)	Possible	Caucasian	Male	69	Bulbar	65
DCTN1(p.T1249I), *SETX*(p.M274V), *ATXN2*(22/32)	Simplex	Caucasian	Female	62	Limb	unknown
EWSR1(p.S203T)	Simplex	Caucasian	Male	61	Limb	78
EWSR1(p.T207S)	Simplex	Caucasian	Female	60	Limb	unknown
FIG4(F530*)	Simplex	Caucasian	Male	82	Limb	16
FIG4(p.F815S)	Simplex	Caucasian	Male	38	Bulbar	unknown
FIG4(p.I41T)	Simplex	Caucasian	Female	77	Bulbar	7
FIG4(p.T410P)	Simplex	Hispanic	Male	78	Bulbar	unknown
FUS(p.c.1394-2delA)	Possible	Caucasian	Male	32	Limb	8
FUS(p.P431L)	Simplex	Caucasian	Male	60	Limb	unknown
FUS(p.P525L), *ATXN2*(23/31)	Simplex	Caucasian	Female	14	Limb	44
FUS(p.R485W), *SETX*(p.I2547T)	Simplex	Caucasian	Male	55	Limb	>62
FUS(p.R521G)	Definite	Caucasian	Female	46	Limb	20
OPTN(p.Q235*)	Simplex	Caucasian	Male	72	Limb	14
OPTN(p.Q314L)	Simplex	Caucasian	Female	63	Bulbar	unknown
SETX(p.A1588T)	Simplex	Asian	Male	48	Limb	unknown
SETX(p.C1554G)	Simplex	Caucasian	Male	78	Bulbar	unknown
SETX(p.C1554G)	Simplex	Caucasian	Male	69	Limb	unknown
SETX(p.C1554G)	Simplex	Caucasian	Male	74	Limb	unknown
SETX(p.C1554G)	Simplex	Caucasian	Female	69	Limb	unknown
SETX(p.C1554G), *SETX*(p.R168Q)	Simplex	Caucasian	Female	67	Bulbar	unknown
SETX(p.D2372N)	Simplex	Hispanic	Male	53	Bulbar	unknown

SETX(p.D994G)	Simplex	Caucasian	Male	70	Unknown	unknown
SETX(p.E327V)	Simplex	Caucasian	Male	60	Limb	>205
SETX(p.E623D)	Simplex	Caucasian	Male	67	Limb	>28
SETX(p.H2533Y)	Probable B	Caucasian	Female	59	Limb	41
SETX(p.I2547T)	Simplex	Caucasian	Male	61	Limb	unknown
SETX(p.I2547T)	Simplex	Caucasian	Male	35	Bulbar	unknown
SETX(p.I2547T)	Simplex	Caucasian	Female	77	Respiratory	unknown
SETX(p.I2547T)	Simplex	Unknown	Male	68	Limb	unknown
SETX(p.I2547T) homozygous	Simplex	Caucasian	Male	51	Limb	32
SETX(p.I2547T), *SETX*(p.T14I), *TAF15*(p.R408C)	Simplex	Caucasian	Female	40	Limb	36
SETX(p.K220Q)	Simplex	Caucasian	Male	72	Bulbar	19
SETX(p.L158V)	Simplex	Caucasian	Male	43	Limb	>55
SETX(p.M917T)	Simplex	Caucasian	Male	40	Limb	16
SETX(p.P1622L)	Simplex	Caucasian	Male	41	Limb	59
SETX(p.S323N)	Simplex	Caucasian	Female	79	Limb	unknown
SETX(p.S464L)	Simplex	Caucasian	Female	39	Limb	129
SETX(p.T1118K)	Simplex	Caucasian	Male	62	Limb	unknown
SOD1(p.A5V)	Simplex	Caucasian	Female	46	Limb	unknown
SOD1(p.A5V), *DAO*(p.S345F)	Definite	Caucasian	Female	57	Limb	13
SOD1(p.A90V)	Simplex	Caucasian	Male	unknown	Limb	unknown
SOD1(p.D77Y)	Probable B	Caucasian	Male	75	Limb	41
SOD1(p.D91A)	Simplex	Caucasian	Male	68	Limb	unknown
SOD1(p.D91A)	Definite	Caucasian	Female	unknown	Limb	unknown
SOD1(p.G38R), *ANG*(p.P136L), *DCTN1*(p.T1249I)	Definite	Caucasian	Female	21	Limb	unknown
SOD1(p.I114T)	Simplex	Caucasian	Male	66	Limb	unknown
SOD1(p.I114T)	Definite	Caucasian	Female	39	Limb	unknown

SOD1(p.P67A), *SETX*(p.I2547T)	Probable B	Caucasian	Female	36	Limb	>95
SQSTM1(p.A33V)	Simplex	Caucasian	Male	63	Bulbar	unknown
SQSTM1(p.A33V)	Simplex	Caucasian	Male	82	Bulbar	25
SQSTM1(p.A426V)	Simplex	Caucasian	Male	76	Bulbar	84
SQSTM1(p.K238E)	Simplex	Caucasian	Male	75	Limb	unknown
SQSTM1(p.K238E)	Simplex	Caucasian	Male	34	Limb	unknown
SQSTM1(p.K238E)	Simplex	Caucasian	Male	61	Limb	unknown
SQSTM1(p.K238E)	Simplex	Caucasian	Male	39	Limb	>108
SQSTM1(p.K238E), *ATXN2*(22/31)	Simplex	Caucasian	Male	38	Bulbar	18
SQSTM1(p.L48P)	Simplex	Caucasian	Male	64	Limb	100
TAF15(p.R388H)	Simplex	Caucasian	Female	62	Limb	unknown
TARDBP(p.A315T)	Definite	Caucasian	Male	48	Limb	118
TARDBP(p.G287S), *VAPB*(p.M170I)	Simplex	Caucasian	Female	72	Bulbar	102
TARDBP(p.S375G)	Simplex	Caucasian	Female	41	Limb	10
VAPB(p.D130E)	Simplex	Caucasian	Male	54	Limb	73
VAPB(p.M170I)	Simplex	Caucasian	Male	74	Limb	unknown
VAPB(p.M170I)	Simplex	Caucasian	Male	67	Bulbar	unknown
VAPB(p.M170I)	Simplex	Caucasian	Male	38	Limb	>71
VCP(p.R155H)	Probable A	Caucasian	Male	42	Limb	>137

Table S2 Individual variant association testing results. Only SNPs found ALS samples with European ancestry. Discovery Cohort= ALS: genotyped frequencies. SPLINTER predictions used when SNPs not genotyped. Only samples of European ancestry used. Controls: Frequency in ESP6500-EA+1000genomesEUR. Significance level adjusted for multiple comparisons $p=8.2x10^{-4}$. Validation Cohort: 552 ALS cases and 464 neurologically normal controls from Coriell, all of European Ancestry. Yellow highlights three variants with significant association in the Discovery Cohort, but that did not show an association with the Validation Cohort.

SNP ID	Hg19 Location	Gene	Variant	Alleles	Discovery Cohort				Validation Cohort			
					Minor Alleles		pval	OR (95%CI)	Minor Alleles		pval	OR (95%CI)
					ALS	Controls			ALS	Controls		
rs80356719	1:11082325	*TARDBP*	G287S	A/G	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
Novel	1:11082589	*TARDBP*	S375G	G/A	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
Novel	2:74588653	*DCTN1*	H1270Q	A/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs72466496	2:74588717	*DCTN1*	T1249I	T/C	1/618	39/9358	0.515	0.39 (0.01,2.3)	6/1080	1/910	0.13	5.08 (0.61,233.59)
rs72659383	2:74592252	*DCTN1*	R1049Q	A/G	1/618	21/9358	1	0.72 (0.02,4.5)	not tested			
rs17721059	2:74596527	*DCTN1*	R495Q	A/G	19/618	175/9358	0.048	1.66 (0.97,2.7)	24/1076	19/908	0.88	1.07 (0.56,2.08)
rs13420401	2:74597937	*DCTN1*	L287M	A/C	2/618	6/9358	0.084	5.06 (0.5,28.37)	0/1084	0/738	1	0 (0,Inf)
rs55862001	2:74598723	*DCTN1*	I196V	G/A	4/618	70/9358	1	0.86 (0.23,2.32)	6/1070	6/904	0.78	0.84 (0.23,3.17)
rs374419252	2:74604801	*DCTN1*	S111C	G/C	1/618	0/9358	0.062	Inf (0.39,Inf)	0/1082	0/738	1	0 (0,Inf)
rs200396166	5:179248034	*SQSTM1*	A33V	T/C	2/618	6/9358	0.105	4.4 (0.43,24.67)	0/1062	2/904	0.21	0 (0,4.53)
Novel	5:179248079	*SQSTM1*	L48P	C/T	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs11548633	5:179252184	*SQSTM1*	K238E	G/A	5/618	32/9358	0.076	2.38 (0.72,6.18)	5/1082	1/738	0.41	3.42 (0.38,161.93)
rs55793208	5:179260099	*SQSTM1*	E274D	T/G	5/618	235/9358	4.11E-03	0.32 (0.1,0.75)	27/1090	22/920	1	1.04 (0.56,1.92)
rs201239306	5:179263547	*SQSTM1*	A426V	T/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			

rs121908287	6:110036336	*FIG4*	I41T	C/T	1/618	16/9358	1	0.95 (0.02,6.11)	0/1044	3/868	0.09	0 (0,2.01)
rs2295837	6:110064928	*FIG4*	M364L	T/A	14/618	345/9358	0.073	0.61 (0.33,1.04)	41/1090	45/918	0.22	0.76 (0.48,1.2)
rs9885672	6:110107517	*FIG4*	V654A	C/T	80/618	1432/9358	0.118	0.82 (0.64,1.05)	not tested			
rs143531641	6:110107636	*FIG4*	M694V	G/A	1/618	3/9358	0.226	5.05 (0.1,63.11)	0/1082	0/738	1	0 (0,Inf)
rs375414729	6:110113852	*FIG4*	F815S	C/T	1/618	1/9358	0.12	15.15 (0.19,1177.85)	not tested			
rs3739927	9:135139826	*SETX*	S2612G	G/A	34/618	249/9358	2.21E-04	2.13 (1.43,3.09)	25/1044	19/866	0.88	1.09 (0.57,2.12)
rs1056899	9:135139901	*SETX*	I2587V	G/A	191/618	2752/9358	0.439	1.07 (0.9,1.28)	300/1046	276/866	0.13	0.86 (0.7,1.05)
rs151117904	9:135140020	*SETX*	I2547T	C/T	7/618	71/9358	0.338	1.5 (0.58,3.27)	8/1084	2/738	0.22	2.74 (0.54,26.51)
rs2296871	9:135173685	*SETX*	T1855A	G/A	126/618	1439/9358	1.35E-03	1.41 (1.14,1.73)	169/1090	146/918	0.81	0.97 (0.76,1.25)
rs140781535	9:135202120	*SETX*	P1622L	T/C	1/618	0/9358	0.062	Inf (0.39,Inf)	0/1074	0/738	1	0 (0,Inf)
rs112089123	9:135202325	*SETX*	C1554G	G/T	6/618	40/9358	0.063	2.28 (0.79,5.45)	3/1078	2/738	1	1.03 (0.12,12.33)
rs1185193	9:135203409	*SETX*	D1192E	T/G	102/618	1282/9358	0.054	1.25 (0.99,1.56)	144/1074	132/906	0.47	0.91 (0.7,1.18)
rs3739922	9:135203530	*SETX*	F1152C	G/T	27/618	317/9358	0.209	1.3 (0.84,1.95)	32/1042	32/866	0.52	0.83 (0.49,1.41)
Novel	9:135203632	*SETX*	T1118K	A/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs149546633	9:135204004	*SETX*	D994G	G/A	1/618	0/9358	0.062	Inf (0.39,Inf)	0/1046	0/868	1	0 (0,Inf)
rs61742937	9:135204010	*SETX*	K992R	G/A	19/618	156/9358	0.016	1.87 (1.09,3.05)	21/1090	11/916	0.22	1.62 (0.74,3.73)
rs376022544	9:135204235	*SETX*	M917T	C/T	1/618	1/9358	0.12	15.15 (0.19,1177.85)	0/1044	0/868	1	0 (0,Inf)
rs882709	9:135205006	*SETX*	A660G	G/C	67/618	548/9358	3.97E-06	1.95 (1.47,2.56)	59/1082	51/912	0.92	0.97 (0.65,1.46)
rs139200312	9:135205116	*SETX*	E623D	C/A	1/618	3/9358	0.226	5.05 (0.1,63.11)	not tested			

rs200614765	9:135205594	*SETX*	S464L	T/C	1/618	6/9358	0.361	2.53 (0.05,20.86)	0/1000	1/824	0.45	0 (0,32.14)
Novel	9:135206694	*SETX*	E327V	T/A	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs372193033	9:135206706	*SETX*	S323N	A/G	2/618	1/9358	0.011	30.36 (1.58,1768.42)	not tested			
Novel	9:135210013	*SETX*	M274V	G/A	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
Novel	9:135211743	*SETX*	K220Q	C/A	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
Novel	9:135211898	*SETX*	R168Q	A/G	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs145438764	9:135218103	*SETX*	L158V	G/T	1/618	48/9358	0.369	0.31 (0.01,1.84)	9/1080	4/738	0.58	1.54 (0.43,6.88)
rs79740039	9:135224757	*SETX*	R20H	A/G	4/618	101/9358	0.415	0.6 (0.16,1.58)	9/1090	10/918	0.65	0.76 (0.27,2.08)
Novel	9:135224775	*SETX*	T14I	T/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs11258194	10:13152400	*OPTN*	M98K	A/T	15/618	297/9358	0.341	0.76 (0.42,1.28)	not tested			
Novel	10:13160964	*OPTN*	Q235*	T/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs142812715	10:13166053	*OPTN*	Q314L	T/A	1/618	3/9358	0.226	5.05 (0.1,63.11)	not tested			
rs4262766	12:109294259	*DAO*	G331E	A/G	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs121909536	14:21161845	*ANG*	K41I	T/A	1/618	23/9358	1	0.66 (0.02,4.07)	4/1046	2/868	0.7	1.66 (0.24,18.4)
rs186547381	16:31201719	*FUS*	P431L	T/C	1/618	2/9358	0.175	7.58 (0.13,145.89)	not tested			
rs201772423	16:31202343	*FUS*	R485W	T/C	1/618	1/9358	0.12	15.15 (0.19,1177.35)	not tested			
Novel	16:31202752	*FUS*	P525L	T/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs140268553	17:34171358	*TAF15*	R388H	A/G	1/618	7/9358	0.401	2.16 (0.05,16.89)	not tested			
rs200175347	17:34171525	*TAF15*	R408C	T/C	1/618	2/9358	0.175	7.58 (0.13,145.89)	0/1048	0/866	1	0 (0,Inf)

rs146459055	20:57014075	*VAPB*	D130E	G/T	1/618	11/9358	0.536	1.38 (0.03,9.5)	not tested			
rs143144050	20:57016076	*VAPB*	M170I	A/G	5/618	18/9358	0.012	4.23 (1.22,11.88)	3/1088	2/916	1	1.26 (0.14,15.15)
rs121912442	21:33032096	*SOD1*	A5V	T/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
Novel	21:33039600	*SOD1*	A90V	T/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs80265967	21:33039603	*SOD1*	D91A	C/A	2/618	9/9358	0.146	3.37 (0.35,16.35)	not tested			
rs121912441	21:33039672	*SOD1*	I114T	C/T	1/618	1/9358	0.12	15.15 (0.19,1177.85)	not tested			
rs144503053	22:29682919	*EWSR1*	S203T	A/T	1/618	1/9358	0.12	15.15 (0.19,1177.35)	1/1076	0/906	1	Inf (0.02,Inf)
Novel	22:29682932	*EWSR1*	T207S	G/C	1/618	0/9358	0.062	Inf (0.39,Inf)	not tested – singleton in ALS			
rs41311143	22:29693915	*EWSR1*	G470S	A/G	24/618	119/9358	6.70E-06	3.14 (1.92,4.94)	12/1086	13/916	0.55	0.78 (0.32,1.86)

3.7 References

1. Abhinav, K. et al. Amyotrophic lateral sclerosis in South-East England: a population-based study. The South-East England register for amyotrophic lateral sclerosis (SEALS Registry). Neuroepidemiology 29, 44–48 (2007).

2. Mehta, P. et al. Prevalence of amyotrophic lateral sclerosis - United States, 2010-2011. Morb. Mortal. Wkly. Rep. Surveill. Summ. Wash. DC 2002 63 Suppl 7, 1–14 (2014).

3. Gros-Louis, F., Gaspar, C. & Rouleau, G. A. Genetics of familial and sporadic amyotrophic lateral sclerosis. Biochim. Biophys. Acta 1762, 956–972 (2006).

4. KURLAND, L. T. & MULDER, D. W. Epidemiologic investigations of amyotrophic lateral sclerosis. 2. Familial aggregations indicative of dominant inheritance. I. Neurology 5, 182–196 (1955).

5. KURLAND, L. T. & MULDER, D. W. Epidemiologic investigations of amyotrophic lateral sclerosis. 2. Familial aggregations indicative of dominant inheritance. II. Neurology 5, 249–268 (1955).

6. Harms, M. B. & Baloh, R. H. Clinical neurogenetics: amyotrophic lateral sclerosis. Neurol. Clin. 31, 929–950 (2013).

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

8. Lattante, S. et al. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology 79, 66–72 (2012).

9. Kenna, K. P. et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J. Med. Genet. 50, 776–783 (2013).

10. Chiò, A. et al. Extensive genetics of ALS: a population-based study in Italy. Neurology 79, 1983–1989 (2012).

11. Kwon, M.-J. et al. Screening of the SOD1, FUS, TARDBP, ANG, and OPTN mutations in Korean patients with familial and sporadic ALS. Neurobiol. Aging 33, 1017.e17–23 (2012).

12. Van Blitterswijk, M. et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum. Mol. Genet. 21, 3776–3784 (2012).

13. Conte, A. et al. Classification of familial amyotrophic lateral sclerosis by family history: effects on frequency of genes mutation. J. Neurol. Neurosurg. Psychiatry 83, 1201–1203 (2012).

14. Renton, A. E., Chiò, A. & Traynor, B. J. State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. 17, 17–23 (2014).

15. Fogh, I. et al. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. Hum. Mol. Genet. 23, 2220–2231 (2014).

16. Keller, M. F. et al. Genome-Wide Analysis of the Heritability of Amyotrophic Lateral Sclerosis. JAMA Neurol. 71, 1123 (2014).

17. Al-Chalabi, A. et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J. Neurol. Neurosurg. Psychiatry 81, 1324–1326 (2010).

18. Druley, T. E. et al. Quantification of rare allelic variants from pooled genomic DNA. Nat. Methods 6, 263–265 (2009).

19. Vallania, F. L. M. et al. High-throughput discovery of rare insertions and deletions in large cohorts. Genome Res. 20, 1711–1718 (2010).

20. Brown, J. A. et al. SOD1, ANG, TARDBP and FUS mutations in amyotrophic lateral sclerosis: a United States clinical testing lab experience. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 13, 217–222 (2012).

21. Brooks, B. R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial ‘Clinical limits of amyotrophic lateral sclerosis’ workshop contributors. J. Neurol. Sci. 124 Suppl, 96–107 (1994).

22. Harms, M. B. et al. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Neurobiol. Aging 34, 2234.e13–19 (2013).

23. Wu, C.-H. et al. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature 488, 499–503 (2012).

24. Robinson, J. T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).

25. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2012).

26. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 25, 2078–2079 (2009).

27. Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. Hum. Mutat. 32, 358–368 (2011).

28. Flanagan, S. E., Patch, A.-M. & Ellard, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. Genet. Test. Mol. Biomark. 14, 533–537 (2010).

29. DeJesus-Hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72, 245–256 (2011).

30. Elden, A. C. et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. Nature 466, 1069–1075 (2010).

31. Van Damme, P. et al. Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. Neurology 76, 2066–2072 (2011).

32. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012).

33. Wu, M. C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. 89, 82–93 (2011).

34. Beck, J. et al. 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, 345–353 (2013).

35. Vilariño-Güell, C. et al. Characterization of DCTN1 genetic variability in neurodegeneration. Neurology 72, 2024–2028 (2009).

36. Chow, C. Y. et al. Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. Nature 448, 68–72 (2007).

37. Wu, D. et al. Angiogenin loss-of-function mutations in amyotrophic lateral sclerosis. Ann. Neurol. 62, 609–617 (2007).

38. Greenway, M. J. et al. ANG mutations segregate with familial and ‘sporadic’ amyotrophic lateral sclerosis. Nat. Genet. 38, 411–413 (2006).

39. Van Es, M. A. et al. A case of ALS-FTD in a large FALS pedigree with a K17I ANG mutation. Neurology 72, 287–288 (2009).

40. Millecamps, S. et al. SOD1, ANG, VAPB, TARDBP, and FUS mutations in familial amyotrophic lateral sclerosis: genotype-phenotype correlations. J. Med. Genet. 47, 554–560 (2010).

41. Couthouis, J. et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. Hum. Mol. Genet. 21, 2899–2911 (2012).

42. Kwiatkowski, T. J., Jr et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science 323, 1205–1208 (2009).

43. DeJesus-Hernandez, M. et al. De novo truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis. Hum. Mutat. 31, E1377–1389 (2010).

44. Del Bo, R. et al. Novel optineurin mutations in patients with familial and sporadic amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatry 82, 1239–1243 (2011).

45. Van Blitterswijk, M. et al. Novel optineurin mutations in sporadic amyotrophic lateral sclerosis patients. Neurobiol. Aging 33, 1016.e1–7 (2012).

46. Maruyama, H. et al. Mutations of optineurin in amyotrophic lateral sclerosis. Nature 465, 223–226 (2010).

47. Kamada, M. et al. Clinicopathologic features of autosomal recessive amyotrophic lateral sclerosis associated with optineurin mutation. Neuropathol. Off. J. Jpn. Soc. Neuropathol. (2013). doi:10.1111/neup.12051

48. Zhao, Z. et al. A novel mutation in the senataxin gene identified in a Chinese patient with sporadic amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 10, 118–122 (2009).

49. Asaka, T., Yokoji, H., Ito, J., Yamaguchi, K. & Matsushima, A. Autosomal recessive ataxia with peripheral neuropathy and elevated AFP: novel mutations in SETX. Neurology 66, 1580–1581 (2006).

50. Hirano, M. et al. Senataxin mutations and amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. 12, 223–227 (2011).

51. Moreira, M.-C. et al. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. Nat. Genet. 36, 225–227 (2004).

52. Rosen, D. R. et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 362, 59–62 (1993).

53. Andersen, P. M. et al. Phenotypic heterogeneity in motor neuron disease patients with CuZn-superoxide dismutase mutations in Scandinavia. Brain J. Neurol. 120 (Pt 10), 1723–1737 (1997).

54. Rezania, K. et al. A rare Cu/Zn superoxide dismutase mutation causing familial amyotrophic lateral sclerosis with variable age of onset, incomplete penetrance and a sensory neuropathy. Amyotroph. Lateral Scler. Mot. Neuron Disord. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 4, 162–166 (2003).

55. Andersen, P. M. et al. Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. Nat. Genet. 10, 61–66 (1995).

56. Keckarević, D., Stević, Z., Keckarević-Marković, M., Kecmanović, M. & Romac, S. A novel P66S mutation in exon 3 of the SOD1 gene with early onset and rapid progression. Amyotroph. Lateral Scler. Off. Publ. World Fed. Neurol. Res. Group Mot. Neuron Dis. 13, 237–240 (2012).

57. Fecto, F. et al. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. Arch. Neurol. 68, 1440–1446 (2011).

58. Rubino, E. et al. SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Neurology 79, 1556–1562 (2012).

59. Couthouis, J. et al. A yeast functional screen predicts new candidate ALS disease genes. Proc. Natl. Acad. Sci. U. S. A. 108, 20881–20890 (2011).

60. Gitcho, M. A. et al. TDP-43 A315T mutation in familial motor neuron disease. Ann. Neurol. 63, 535–538 (2008).

61. Kabashi, E. et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat. Genet. 40, 572–574 (2008).

62. Johnson, J. O. et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron 68, 857–864 (2010).

63. Byrne, S. et al. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. J. Neurol. Neurosurg. Psychiatry 82, 623–627 (2011).

64. Xue, Y. et al. Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. Am. J. Hum. Genet. 91, 1022–1032 (2012).

65. MacArthur, D. G. et al. Guidelines for investigating causality of sequence variants in human disease. Nature 508, 469–476 (2014).

Chapter 4

Risk Haplotype of the *C9ORF72* Locus

4.1 Abstract

The most common genetic cause of ALS to date is the hexanucleotide repeat expansion in the gene *C9ORF72*. The presence of a common haplotype, tagged by rs3849942, shared between all expansion carriers led to the hypothesis of an expansion founder event, however the shared haplotype could alternatively be a genetic background that is permissive to expansion. We found that all subjects in our cohort of 153 expansion carriers shared at least part of the common haplotype. Sequencing a BAC containing *C9ORF72* locus identified the rare SNP rs147599399 within the risk haplotype that was not shared between all expansion carriers and was also present in the general population. Additionally, survival in expansion carriers with the minor allele at rs147599399 was roughly 20 months longer than expansion carriers that were homozygous for the reference allele (p=0.00047). These results show that both rs3849942 and rs147599399 are potential cis-modifiers of the expansion with rs3849942 promoting the formation of *C9ORF72* repeat expansions and rs147599399 mediating their effects. A genetic background that is permissive to expansions raises that possibility that expansions can arise *de novo* and would account for the prevalence of the expansion in sporadic cases of ALS. Understanding how the genetic background promotes expansion could allow us to create strategies to intervene and prevent repeats from expanding.

4.2 Introduction

Hexanucleotide (GGGGCC) repeat expansions in the *C9ORF72* gene are the most common cause of both familial and sporadic ALS to date.[1,2] They account for approximately

34% of familial cases and 6% of sporadic cases overall with frequencies varying by population. The expansion is most prevalent in ALS subjects in Sweden and Finland and lowest in Asian populations (reviewed in Van Blitterswijk et al.)[3] *C9ORF72* repeat expansions also account for a significant portion of frontotemporal dementia (FTD) and have been found in other neurological disorders. While the pathogenic range of repeat expansions has yet to be clearly delineated, expansions larger than 26 are very rare in normal controls.[1,2,4,5]

The repeat expansion is thought to have arisen from a single founder event in the Scandanavian population.[1,2,5–9] A 42-SNP haplotype spanning 232 kb was identified in a GWAS of Finnish SALS subjects with the A allele of rs3849942 tagging the haplotype.[10,11] With the discovery of the *C9ORF72* repeat expansion, it was shown that all expansion carriers shared the rs3849942 risk allele[1]. Analysis of the entire risk haplotype in wider populations revealed that all expansion carriers possessed at least part of the risk haplotype.[8] The prevalence of the *C9ORF72* expansion is lowest in Asian populations with have the least amount of Scandanavian admixture.[3]

This observation has been widely interpreted to mean that the expansion arose once approximately 1500 years ago. The alternative hypothesis to this single founder event is that the risk haplotype is a genetic background that is permissive to expansion. Multiple studies have shown that the rs3849942 A allele is associated with longer repeat lengths in those without the full expansion.[1,5,12] It is possible that these are pre-mutations that can expand further into a pathogenic range. The risk haplotype is at lower frequencies in Asian populations[5] and therefore *C9ORF72* repeat expansions would be less likely to occur in these populations. Two other studies have provided evidence of multiple origins of the expansions. Beck *et al.* used microsattelite markers to show that expansions carriers in the UK did not share recent common

ancestry[12] and Fratta *et al* used Southern blotting to show an individual with a small expansion in blood but a large expansion in brain.[13]

Evidence of multiple origins of the expansion raises the possibility that expansions can arise *de novo* on the risk haplotype and would explain the prevalence of *C9ORF72* expansions in sporadic disease. This would also have implications for estimating the risk of ALS for genetic counseling. Furthermore evidence of multiple expansions would indicate that the risk haplotype is permissive to expansion and allow us to identify the cis-acting modification that promotes repeat expansions in *C9ORF72.*

In this study we first investigated the known risk haplotype in our cohort of expansion carriers to determine if there were alternate genetic backgrounds that would support multiple origins of expansion. In addition we reasoned that rare variation would be more informative with regards to the recent history of the expansion, therefore we use a BAC containing the *C9ORF72* locus derived from a patient carrying the full expansion to identify rare variants within the risk haplotype.

4.3 Methods

4.3.1 Subjects

This study used a total of 153 subjects that carried full *C9ORF72* repeat expansion as diagnosed by repeat-primed PCR. The clinical diagnoses of the subjects were 67 SALS, 67 FALS from 48 families (including asymptomatic relatives), 14 Alzheimer's disease (AD), and 5 unaffected controls. We also included 57 intermediate-length (19-30 repeats) expansion carriers (SALS=7, AD=40, Parkinson's disease=7, Controls=3). Of the 141 ALS subjects, 42 were from Coriell panels NDPT026, NDPT100, NDPT103, NDPT106, and NDPT116. The remaining ALS

subjects were collected at the Washington University Neuromuscular Disease Center in St. Louis, Missouri (WUSM), the Virginia Mason Medical Center (VMMC), or Houston Methodist Hospital.

4.3.2 Genotyping

Eight SNPs from the known risk haplotype rs1110264, rs2225389, rs2589054, rs10511816, rs4879515, rs868856, rs3849942, and rs2453556) were genotyped in 181 subjects from 168 families using Sequenom. An additional 31 subjects were genotyped for rs3849942 using a custom assay (KASPar; KBioscience). Genotyping of rs147599399 was performed by PCR amplification with primers 5' TCT TTA GCC TAG GTG GGG AGA 3' and 5' TGA CAT TTG TAG AGC ACA GCA 3' and subsequent enzymatic digestion with BstN1. This enzyme only cuts sequences that contain the alternate allele. rs147599399 was genotyped in all ALS subjects and asymptomatic relatives that carried full expansions, 5 AD subjects with full expansions, and Coriell panels NDPT026, NDPT100, NDPT103, NDPT106, and NDPT116, a total of 567 subjects.

4.3.3 MiSeq

We obtained DNA from a BAC containing an insert spanning chr9:27527137-27683106 derived from a subject carrying a full repeat expansion confirmed by Southern blotting (Figure 4.1). We sequenced this BAC on the Illumina MiSeq (GTAC). The sequencing reads were aligned to the Hg19 human genome using Novoalign and variants were called using Samtools. We annotated variants using SeattleSeq and obtained variant frequencies from the 1000 genomes phase 3 data.

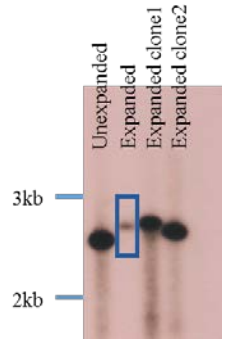

Figure 4.1 Southern blot confirmation of *C9ORF72* repeat expansion in a BAC. The blue box indicates the clone used for sequencing

4.3.4 Statistical analysis

All statistical tests were performed in R version 3.1.1. Comparisons of allelic frequencies were done using Fisher's exact test. Association with age of onset was tested using Wilcoxon rank-sum test. Association with gender and site of onset were tested using Fisher's exact tests. Survival analysis was performed using the Survival package to create Kaplan-Meier curves and Cox-proportional hazards models.

4.4 Results

4.4.1 Known risk haplotype

The 8 SNPs from the known risk haplotype were genotyped in 167 subjects in total. We included subjects that carried the full expansion (n=110) as well as intermediate length expansions (n=57). Confirming the findings in other populations, all subjects carried at least part of the risk haplotype (Figure 4.2 and Figure 4.3). There were 3 subjects that were homozygous

for the non-risk allele at the tag-SNP rs3849942, two of which were intermediate-length expansion carriers. The full expansion carrier was of African American ancestry. In all three cases with the non-risk allele at the tag SNP, both the upstream and downstream SNPs were risk alleles.

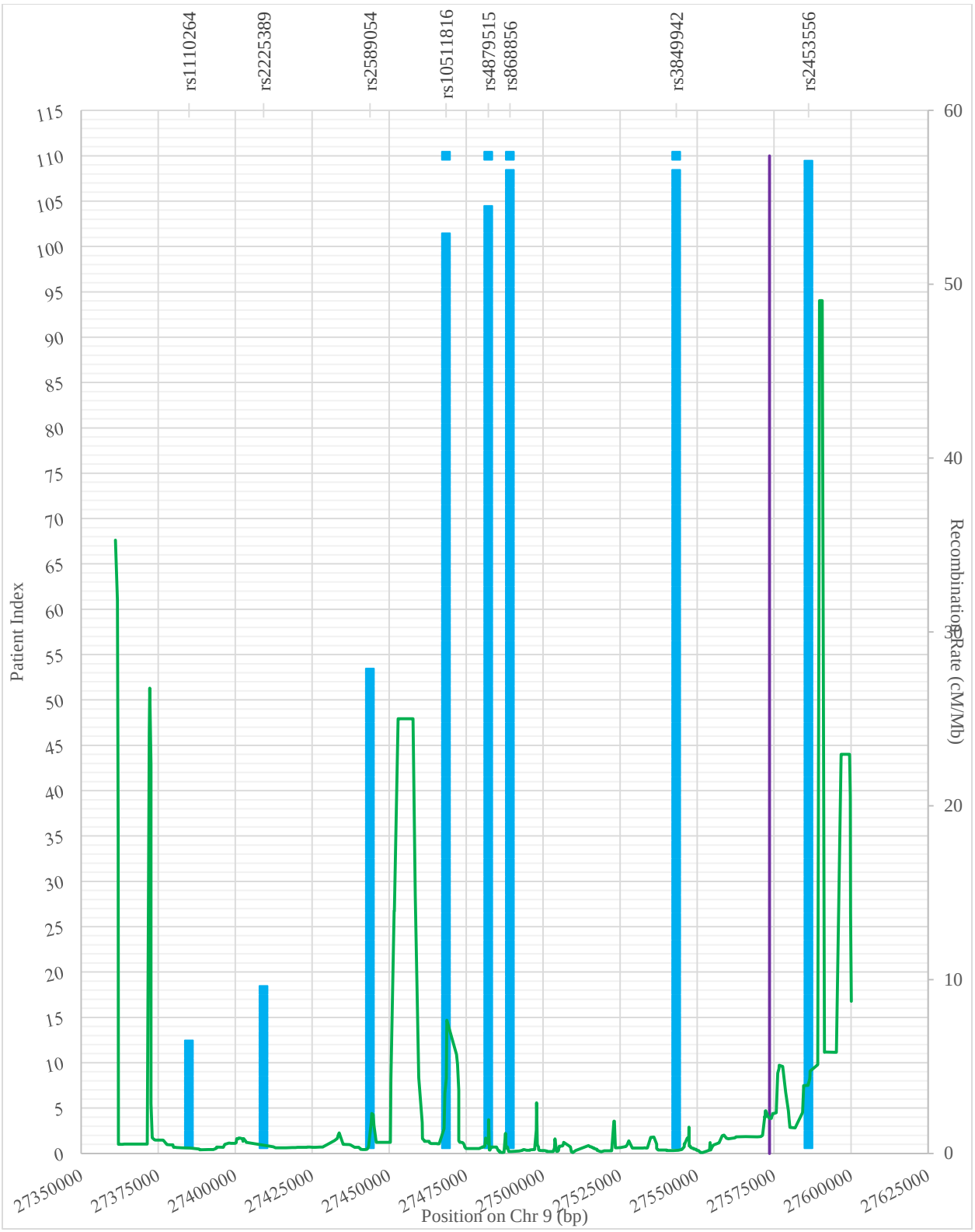

Figure 4.2 Largest continuous risk haplotype in patients with full *C9ORF72* repeat expansions. The green line represents the recombination rate. The vertical purple line represents the location of the repeat expansion. Each row represents a single individual and a blue square indicates the subject carried the risk allele at that SNP.

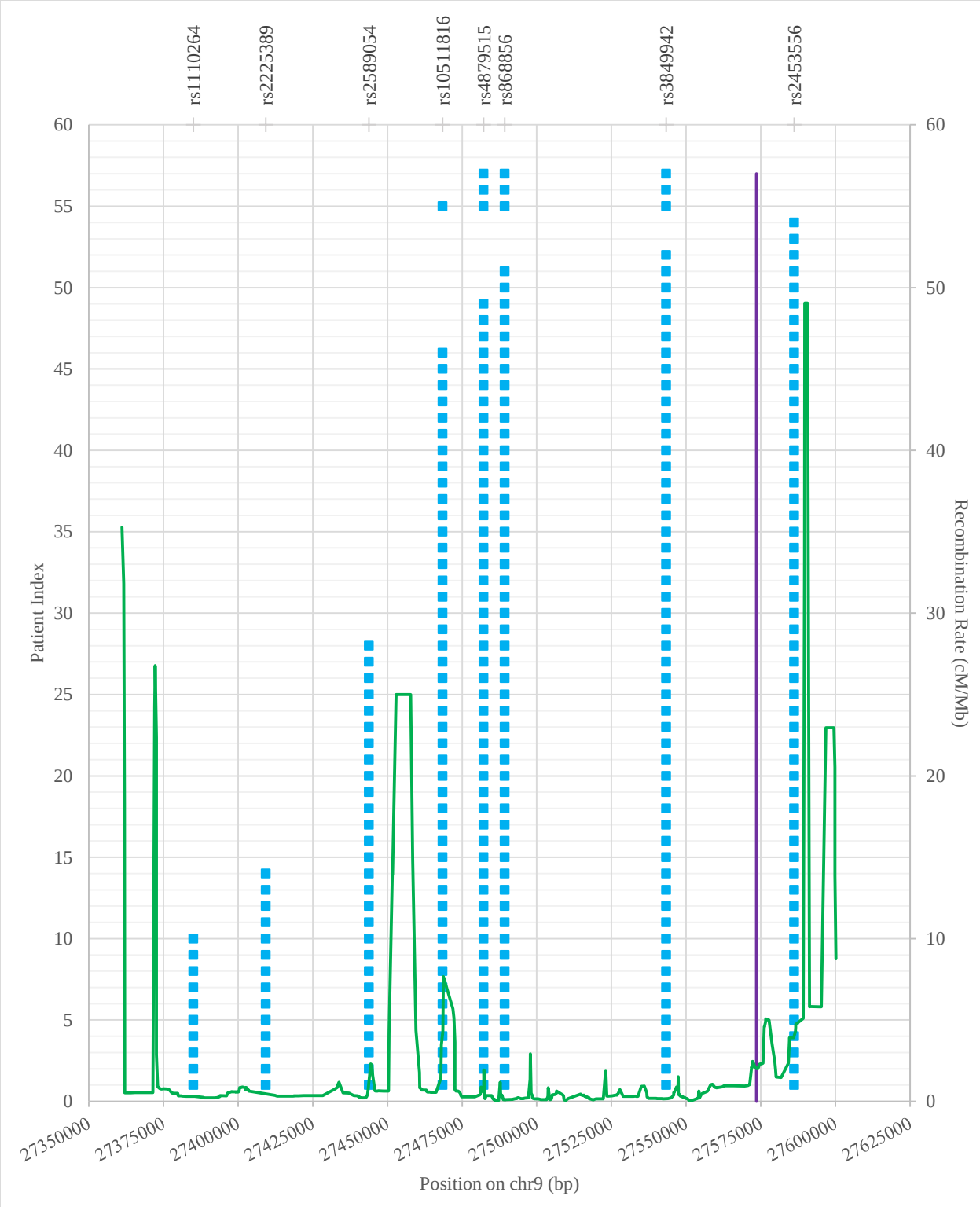

Figure 4.3 Largest continuous risk haplotype with intermediate-length *C9ORF72* repeat expansions (19-30 repeats). The green line represents the recombination rate. The vertical purple line represents the location of the repeat expansion. Each row represents a single individual and a blue square indicates the subject carried the risk allele at that SNP.

4.4.2 Rare SNPs identified by MiSeq

In order to determine if there were rare variants within the risk haplotype that would give us additional insight into the origins of the repeat expansion, we sequenced a BAC containing the section of the hapltyope nearest to the repeat and spanning rs3849942. There were 4 variants within the range of the risk haplotype (Table 4.1). All variants were confirmed by sequencing and all are on the same haplotype as the expansion since they were discovered in the BAC. Interestingly, none of these was novel to this individual. We focused on rs147599399 because of its proximity to the tag SNP rs3849942 (600bp away), because this SNP is very rare in the subjects included in phase 3 of the 1000genomes project, and because it is not present in the Finnish population where the expansion is thought to have arisen (Table 4.2).

Table 4.1 Rare variants on the *C9ORF72* risk haplotype discovered by MiSeq. *Denotes subject used to create BAC.

Position	dbSNP ID	Ref allele	Alt allele	MAF(%) 1000genomes phase3	BAC genotype	Affected Carrier*	Affected Carrier	Unaffected Carrier
chr9:27543629	rs147599399	A	G	0.339	G/G	G/A	G/A	G/A
chr9:27601617	rs75747155	G	A	0.778	A/A	G/A	G/A	G/A
chr9:27617365	rs79652154	C	A	0.572	A/A	C/A	C/A	C/A
chr9:27673058	rs138129614	A	G	0.179	G/G	A/G	A/G	A/G

4.4.3 Prevalence of rs147599399 in additional expansion carriers

We genotyped rs147599399 in a total of 106 full expansion carriers from 86 families, including the family in which the variant was originally identified (WUNM0026). There was one family in addition to WUNM0026 that carried the variant; all expansion carriers were also heterozygous for rs147599399 in both families confirming the SNP is in phase with the C9ORF72 repeat expansion. A total of 19 out of the 86 families carried the variant resulting in a

minor allele frequency of 11%. This is dramatically increased compared to 0.339% in all subjects included in 1000genomes phase3 ($p<2.2X10^{-16}$, OR=35.9, 95%CI=17.5-76.2). The vast majority of expansion carriers in our cohort are of European ancestry so we also compared only those of white, non-Hispanic ethnicity to 1000 genomes subjects with European ancestry ($p=1.625x10^{-10}$, OR=12.8, 95%CI=5.5-31.5). Because our study included expansion carriers of mixed phenotypes, we also looked at rs147599399 in AD. Five of the 87 families had Alzheimer's disease, one of which had the rs147599399 G allele.

To validate this finding, we genotyped rs147599399 in 460 ALS samples from Coriell. The G allele was found in 6/35 expansion carriers compared to 16/425 non-expanded ALS subjects (p=0.004, OR=4.87, 95%CI=1.5-13.7). The frequency of rs147599399 in non-expanded ALS was higher than in the total 1000 genomes cohort ($p=3.46x10^{-6)}$, but not higher than in the 1000 genomes subjects of European ancestry (p=0.116)

Table 4.2 Frequency of rs147599399 in all 1000 genomes populations

Population	Allele count	MAF (%)
Tosaini in Italy (TSI)	209 (A) / 5 (G)	2.336
Colombian in Medellin, Colombia (CLM)	185 (A) / 3 (G)	1.596
Iberian in Spain (IBS)	211 (A) / 3 (G)	1.402
British in England and Scotland (GBR)	180 (A) / 2 (G)	1.099
Americans of African Ancestry in SW USA (ASW)	121 (A) / 1 (G)	0.82
African Caribbeans in Barbados (ACB)	191 (A) / 1 (G)	0.521
Punjabi from Lahore, Pakistan (PJL)	191 (A) / 1 (G)	0.521
Puerto Ricans from Puerto Rico (PUR)	207 (A) / 1 (G)	0.481
Bengali from Bangladesh (BEB)	172 (A)	0
Chinese Dai in Xishuangbanna, China (CDX)	186 (A)	0
Utah Residents (CEPH) with Northern and Western European Ancestry (CEU)	198 (A)	0
Han Chinese in Beijing, China (CHB)	206 (A)	0
Southern Han Chinese in China (CHS)	210 (A)	0
Esan in Nigeria (ESN)	198 (A)	0
Finnish in Finland (FIN)	198 (A)	0
Gujarati Indian from Houston, Texas (GIH)	206 (A)	0
Gambian in Western Divisions in the Gambia (GWD)	226 (A)	0
Indian Telugu from the UK (ITU)	204 (A)	0
Japanese in Tokyo, Japan (JPT)	208 (A)	0
Kinh in Ho Chi Minh City, Vietnam (KHV)	198 (A)	0
Luhya in Webuye, Kenya (LWK)	198 (A)	0
Mende in Sierra Leone (MSL)	170 (A)	0
Mexican ancestry in Los Angeles (MXL)	128 (A)	0
Peruvian in Lima, Peru (PEL)	170 (A)	0
Sri Lankan Tamil from the UK (STU)	204 (A)	0
Yoruba in Ibadan, Nigeria (YRI)	216 (A)	0
ALL	4991 (A) / 17 (G)	0.339

4.4.4 Association of rs147599399 with clinical characteristics in ALS

Within our ALS expansion carriers, there were no differences between gender, site of onset, or age of onset with A/A versus A/G genotypes at rs147599399. However, there was a

difference in survival with the A/G individuals having longer survival (median=50 months) than A/A individuals (median=32) (figure 4.3). We fit a cox-proportional hazards model controlling for age of onset and found that rs147599399 was significantly associated with survival (hazard ratio (A/G vs A/A =0.235, p=0.00047) with the G allele associated with longer survival. There was no difference in survival by rs147599399 genotype in non-expansion carriers.

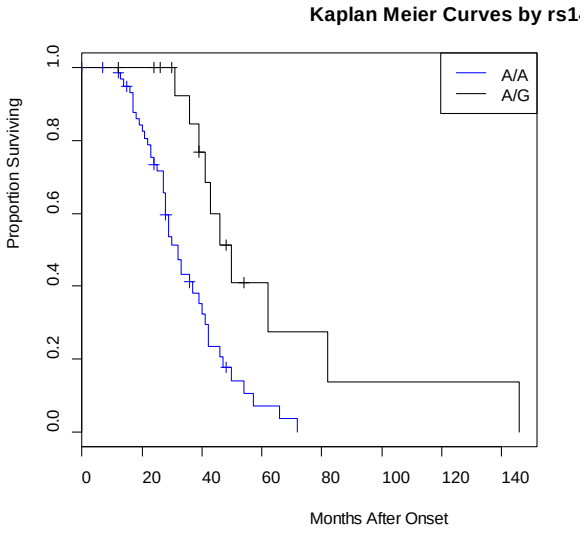

Figure 4.4 Survival curves by rs147599399 genotypes.

4.5 Discussion

Analysis of the *C9ORF72* locus has shown that there is a common haplotype shared between expansion carriers. The simplest explanation for this is observation is a single founder expansion event. Here we provide direct evidence of multiple expansion events based on the SNP rs147599399. This SNP is present in both non-expansion carriers and expansion carriers, but is not present in all expansion carriers. It is less likely that the SNP arose in both groups than the expansion arising on both backgrounds, especially given the known instability of the repeat.[12–14] Since the rs147599399-G carriers in the phased 1000 genomes data have the risk

allele both upstream and downstream of the repeat on the same chromosome, it is also unlikely that a recombination event between rs147599399 and the repeat in an expansion carrier led to the rs147599399-G allele in non-expansion carriers.

An alternate possibility is that there was a founder pre-mutation on the originally identified risk haplotype that is more likely to continue to expand into a pathogenic range as is the case with Myotonic Dystrophy type 1[15,16] and Friedreich Ataxia.[17] In this scenario, the rs147599399 SNP would have occurred on the pre-mutation background and a subset of these would pass on full expansions to their offspring and a subset would remain in the pre-mutation range and not develop disease (Figure 4.5). Further analysis of transmission of intermediate-length alleles will be required to determine if this is a possibility.

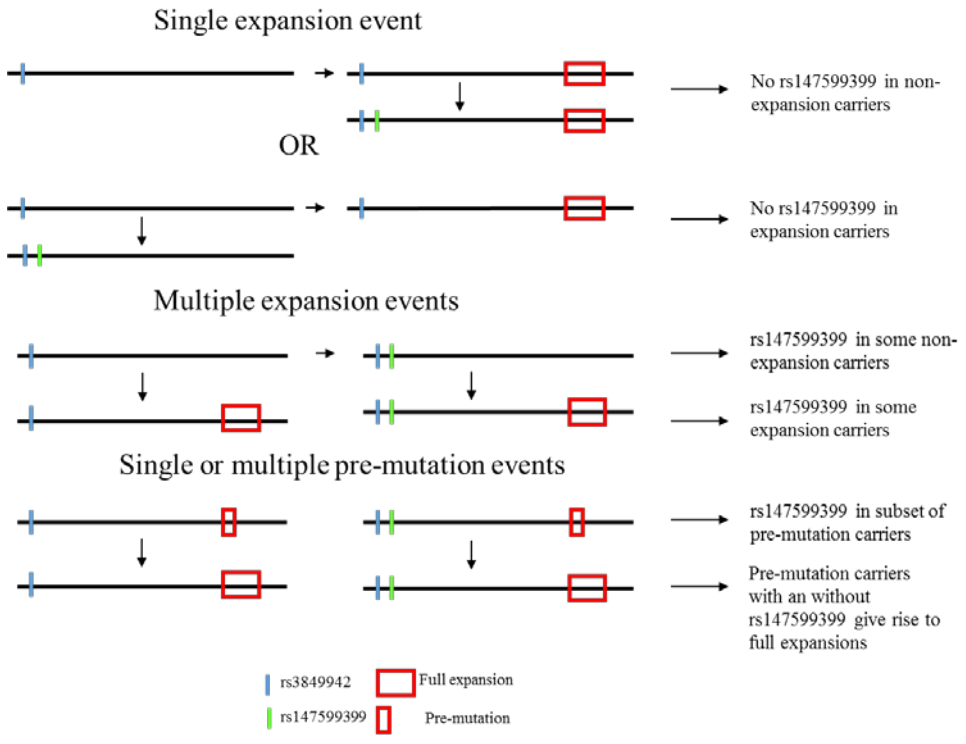

Figure 4.5 Models of expansion events

In addition to showing multiple origins of the expansion, we have discovered that rs147599399 modifies survival in expansion carriers with the G allele resulting in an almost 20 month increase in disease duration. rs147599399 does not influence survival in ALS patients without the repeat expansion, suggesting that the SNP exerts an effect specifically on the repeat expansion. It could be acting by stabilizing the expansion, altering transcription of *C9ORF72*, or altering translation of dipeptide-repeats.[18] This is another in a growing list of genetic factors that have been shown to modify the disease phenotype in *C9ORF72*-associated disease along with *TMEM106B*[19,20], *ATXN2*[21], and several others.[22] However, to our knowledge, this is the first modifier in cis with the expansion.

This is not the first evidence of cis-acting genetic factors influencing repeat expansions. Haplotype is an important factor in Huntington Disease (HD) with multiple origins of the expansion arising on population-specific haplotypes, but the exact mechanism of this is not known.[23–27] Possibilities include changes in replication origins which can alter stability of repeats during replication,[28,29] and formation of secondary structures, which can in turn alter the effects of DNA mismatch repair machinery.[30] Methylation of flanking sequences can also alter stability of repeats as in SCA7[31] and fragile-X syndrome.[32]

We show here that genetic context is an important factor in the pathogenicity of *C9ORF72* through both the risk background associated with rs3849942 and the increased survival associated with rs147599399. This finding suggests that *de novo* expansions on the permissive haplotype are possible, which might explain why the expansion is relatively common in sporadic ALS patients. Understanding the mechanisms behind these phenomena will allow us to develop strategies to intervene.

4.6 References

1. DeJesus-Hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72, 245–256 (2011).

2. Renton, A. E. et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72, 257–268 (2011).

3. Van Blitterswijk, M., DeJesus-Hernandez, M. & Rademakers, R. How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? Curr. Opin. Neurol. 25, 689–700 (2012).

4. Harms, M. B. et al. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Neurobiol. Aging 34, 2234.e13–19 (2013).

5. Smith, B. N. et al. The C9ORF72 expansion mutation is a common cause of ALS+/−FTD in Europe and has a single founder. Eur. J. Hum. Genet. 21, 102–108 (2013).

6. Mok, K. et al. The chromosome 9 ALS and FTD locus is probably derived from a single founder. Neurobiol. Aging 33, 209.e3–209.e8 (2012).

7. Ratti, A. et al. C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. Neurobiol. Aging 33, 2528.e7–2528.e14 (2012).

8. Majounie, E. et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. 11, 323–330 (2012).

9. Van der Zee, J. et al. A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats. Hum. Mutat. 34, 363–373 (2013).

10. Shatunov, A. et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet Neurol. 9, 986–994 (2010).

11. Laaksovirta, H. et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol. 9, 978–985 (2010).

12. Beck, J. et al. 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, 345–353 (2013).

13. Fratta, P. et al. Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion. Neurobiol. Aging 36, 546.e1–546.e7 (2015).

14. Van Blitterswijk, M. et al. 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 (2013).

15. Tishkoff, S. A. et al. A Global Haplotype Analysis of the Myotonic Dystrophy Locus: Implications for the Evolution of Modern Humans and for the Origin of Myotonic Dystrophy Mutations. Am. J. Hum. Genet. 62, 1389–1402 (1998).

16. Imbert, G., Kretz, C., Johnson, K. & Mandel, J.-L. Origin of the expansion mutation in myotonic dystrophy. Nat. Genet. 4, 72–76 (1993).

17. Cossée, M. et al. Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proc. Natl. Acad. Sci. U. S. A. 94, 7452–7457 (1997).

18. Mori, K. et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science 339, 1335–1338 (2013).

19. Van Blitterswijk, M. et al. TMEM106B protects C9ORF72 expansion carriers against frontotemporal dementia. Acta Neuropathol. (Berl.) 127, 397–406 (2014).

20. Gallagher, M. D. et al. TMEM106B is a genetic modifier of frontotemporal lobar degeneration with C9orf72 hexanucleotide repeat expansions. Acta Neuropathol. (Berl.) 127, 407–418 (2014).

21. Van Blitterswijk, M. et al. Ataxin-2 as potential disease modifier in C9ORF72 expansion carriers. Neurobiol. Aging (2014). doi:10.1016/j.neurobiolaging.2014.04.016

22. Van Blitterswijk, M. et al. Genetic modifiers in carriers of repeat expansions in the C9ORF72 gene. Mol. Neurodegener. 9, 38 (2014).

23. Warby, S. C. et al. HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. Eur. J. Hum. Genet. 19, 561–566 (2011).

24. Warby, S. C. et al. CAG Expansion in the Huntington Disease Gene Is Associated with a Specific and Targetable Predisposing Haplogroup. Am. J. Hum. Genet. 84, 351–366 (2009).

25. Baine, F. K. et al. Huntington disease in the South African population occurs on diverse and ethnically distinct genetic haplotypes. Eur. J. Hum. Genet. 21, 1120–1127 (2013).

26. Andrew, S. et al. DNA analysis of distinct populations suggests multiple origins for the mutation causing Huntington disease. Clin. Genet. 43, 286–294 (1993).

27. Squitieri, F. et al. DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. Hum. Mol. Genet. 3, 2103–2114 (1994).

28. Cleary, J. D., Nichol, K., Wang, Y.-H. & Pearson, C. E. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nat. Genet. 31, 37–46 (2002).

29. Gerhardt, J. et al. The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. Mol. Cell 53, 19–31 (2014).

30. Moore, H., Greenwell, P. W., Liu, C. P., Arnheim, N. & Petes, T. D. Triplet repeats form secondary structures that escape DNA repair in yeast. Proc. Natl. Acad. Sci. U. S. A. 96, 1504–1509 (1999).

31. Libby, R. T. et al. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet. 4, e1000257 (2008).

32. Nichol Edamura, K., Leonard, M. R. & Pearson, C. E. Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. Am. J. Hum. Genet. 76, 302–311 (2005).

Chapter 5

Characterization of size and stability of *C9ORF72* repeat expansions

5.1 Abstract

Despite the prevalence of *C9ORF72* repeat expansions in ALS, little is known about the importance of repeat sizes and stability. Understanding these characteristics can have a great impact in our knowledge of the pathogenic mechanisms of the repeat expansion and of how it is transmitted from generation to generation. We used Southern blot analysis to determine size of the repeat expansion in a variety of patient tissues. We observed a high degree of somatic instability between tissues with larger expansions in most regions of the brain than in blood. Expansions size did not correlate with clinical characteristics in any tissue and there was no difference in expansion size in blood between affected and unaffected expansion carriers. There was instability in transmission of expansions within families, but unlike the transmissions in repeat disorders with anticipation, the expansions showed increases and decreases between generations. It will be important to continue this work with additional samples in order to perform more concrete analysis.

5.2 Introduction

Large expansions in *C9ORF72* are the most common cause of ALS, however the effects of expansion sizes on pathogenesis are not well understood.[1,2] Many pathogenic mechanisms of *C9ORF72* repeats have been proposed which could be influenced by expansion sizes. GGGGCC

repeats form RNA-DNA heteroduplexes called G-quadruplexes[3,4] and increases in repeat number altered G-quadruplex conformation in small repeats.[5] Expanded *C9ORF72* also forms nuclear RNA foci which sequester RNA-binding proteins.[1,6–8] Expression of constructs with 8,38, or 72 repeats in cell lines and zebrafish showed length dependent effects on the formation of RNA foci, especially in neuronal cells.[9] The role of repeat length in repeat-associated non-ATG initiated translation into aggregating dipeptide repeats[10–13] has not been investigated, but it is possible that larger expansions would promote increased production of dipeptides and increased aggregation in a similar fashion to RNA foci.

Variability in expansion size and stability could explain disease characteristics such as tissue specificity, progressive degeneration, and incomplete penetrance. For example, in the related repeat expansion disorder Myotonic Dystrophy (DM), somatic instability has been observed in the CTG repeat tract of *DMPK* with larger repeats in muscle (the affected tissue) compared to blood.[14] Additionally the repeat is unstable and continues to expand with age accounting for the progressive nature of the disease.[15] Like many other repeat expansion disorders DM1 also shows anticipation with larger repeat expansions and earlier onset in progressive generations.[16,17]

C9ORF72 expansions are detected by repeat-primed PCR (RP-PCR) which does not size expansions larger than 30 repeats, therefore Southern blots must be used to detect the number of repeats in full expansions. Due to the difficulty of performing Southern blots, very few studies have been done observing the impact of expansion sizes, most of which have focused on expansions in blood-derived DNA.[1,18–24] All studies have shown expansion sizes of hundreds to thousands of repeats in all tissues and variability of expansion sizes between tissues. Although

some studies have reported clinical associations with expansion sizes in certain tissues, these results are mostly inconsistent between studies.[18,19,23]

In this study, we use Southern blot analysis to determine expansion lengths in multiple tissue types in a cohort of ALS subjects. In addition, we analyze the stability of repeat expansions both in cell culture and in transmission in families.

5.3 Methods

5.3.1 Samples

All patients included in this study were diagnosed with definite or probable ALS at either Washington University Neuromuscular Disease Center in St. Louis, Missouri (WUSM) or the Virginia Mason Medical Center (VMMC). Patients were screened for *C9ORF72* repeat expansion by repeat-primed PCR as previously described (Figure 5.1).[1] DNA from patient samples was prepared by the Hope Center DNA/RNA purification core at Washington University in St. Louis. All participants had provided signed informed consent for studies approved by local institutional review boards.

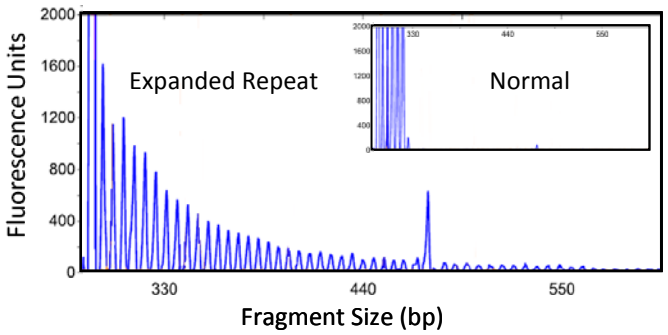

Figure 5.1 Example of repeat-primed PCR to detect C9ORF72 repeat expansions

5.3.2 Southern blots

Southern blots were performed as previously described.[25] Briefly, genomic DNA was digested using either XbaI or XbaI and Taq$^{\alpha}$I. 3.5-5ug of digested DNA was run out on a gel and transferred to the membrane. Hybridization was performed using a 590bp probe labeled with ^{32}P that hybridizes adjacent to the repeat. Blots were exposed on film for 1-7 overnights. Band sizes were determined using GelAnalyzer (ww.gelanalyzer.com). Either the 1kb plus DNA ladder (Invitrogen) or the 1kb DNA ladder (NEB) was used as a standard to measure expansion size. The maximum, minimum and peak for each band were measured. The number of repeats was calculated by subtracting the digest size of 0 repeats (2359 for XbaI digest or 1941 for XbaI+Taq$^{\alpha}$I) from the measured band size and dividing by 6.

5.3.3 Statistical Analysis

All statistical analyses were performed in R v3.1.3. For samples that had multiple expansion size, the largest was used in analysis. When samples were blotted more than once, the average was used. Differences in number of repeats between affected and unaffected expansion carriers and between different tissues were assessed by Wilcoxon-rank sum tests. Differences in number of repeats in different tissues within individuals were assessed using Wilcoxon paired signed -rank tests. Correlations between repeat sizes and age of onset and between repeat sizes from different tissues were performed using Spearman correlations Survival was assessed using cox-proportional hazards models controlling for age of onset.

5.4 Results

5.4.1 Expansions in blood-derived DNA

We were able to perform Southern blots on blood-derived DNA from 39 different expansion carriers. Expansions in blood-derived DNA generally showed long smears on Southern blots (Figure 5.2). The median size of the smear peaks was 1702 and ranged from 705 to 3520 repeats. Smear minima were between 477 and 2297 repeats with a median of 967. Smear maxima ranged from 840 to 3947 repeats.

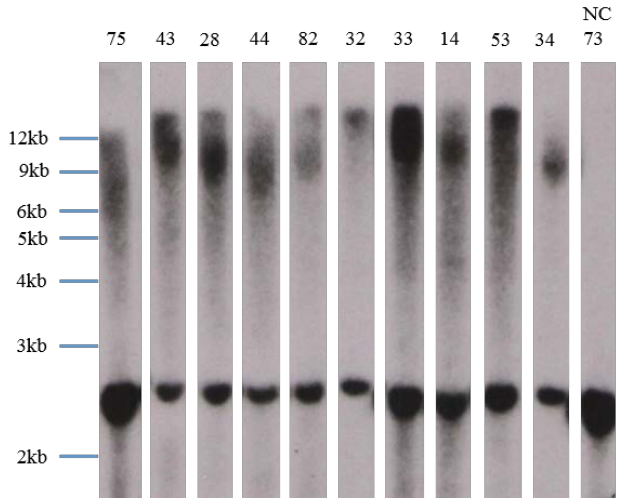

Figure 5.2 Representative Southern blots of blood-derived DNA. Sample numbers are listed above. The far right sample is an ALS patient that was negative for the expansion by RP-PCR and is shown as a negative control. Samples were digested with XbaI resulting in wild-type alleles of ~2359bp.

There was no difference in the minima (p=0.56), maxima (p=0.41) or peaks (p=0.33) between the 26 symptomatic and 13 asymptomatic expansion carriers (peaks shown in figure 5.3 A). There was also no correlation between repeat size and age of onset in symptomatic repeat carriers (p=0.60 for minima, p=0.56 for maxima, p=0.94 for peaks [shown in Figure 5.3B]). Number of repeats also had no effect on survival (p=0.43, p=0.39, p=0.95 for peaks, maxima, and minima respectively).

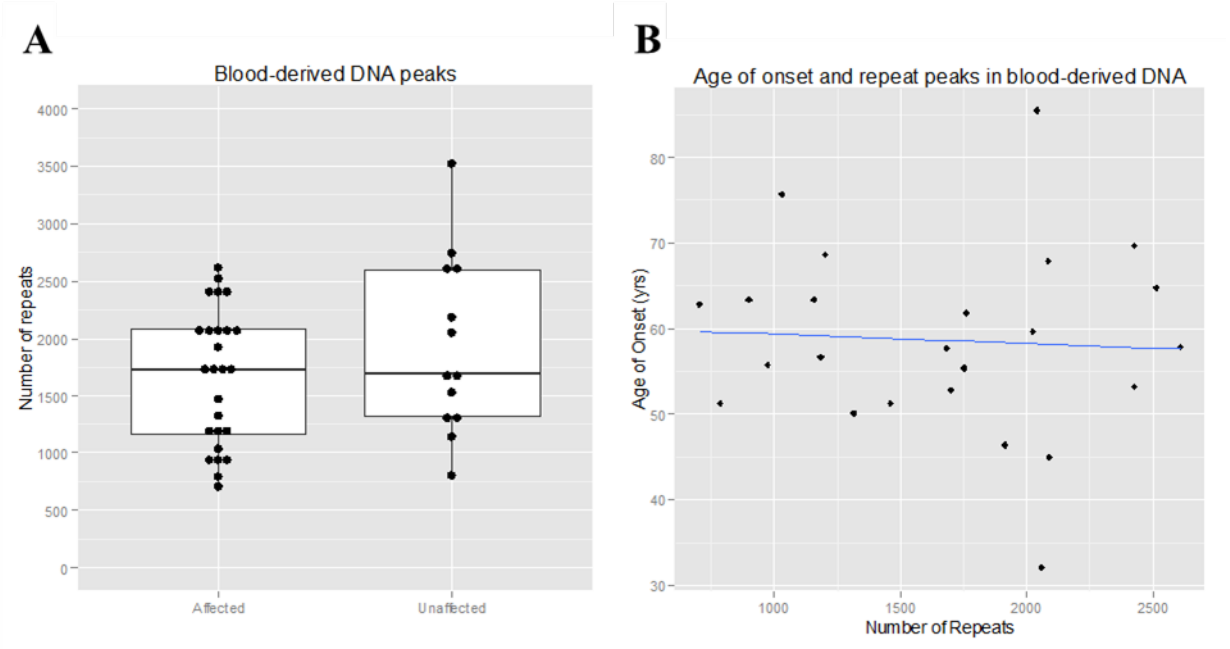

Figure 5.3 Relationship between *C9ORF72* repeat sizes and phenotype. A. Boxplot of peak number of repeats in blood-derived DNA in affected and unaffected expansion carriers. The data are transposed on top of the boxplot. B. Scatterplot of peak number of repeats in bood-derived DNA versus age of onset. The blue line represents the least-squares regression line.

5.4.2 Expansion sizes across families

Blood derived DNA was available from 7 families with multiple expansion carriers. Three families had multiple causes of disease. Family WUNM0026 (Figure 5.4 A) included one distant affected relative that tested negative for the expansion, but was found to be a heterozygous carrier of the *SOD1* p. D91A variant. This individual had another affected cousin who was positive for the expansion. Southern blotting of this family confirmed the absence of the expansion in the p.D91A carrier and presence of the expansion in the other affected members. In family WUNM0085 the proband was positive for the expansion and their distant relative who also had ALS tested negative and was confirmed by Southern blot (Figure 5.4 D). The cause of disease is still unknown in this individual. In family WUNM0194 (Figure 5.4 G)

the proband and two unaffected sisters carried the expansion. They had one affected brother who was negative for the expansion and whose cause of disease is unknown.

We observed both stable and unstable transmission patterns in families. The expansion was roughly the same size in all expansion carriers in families WUNM0026, WUNM0160, WUNM0226, WUNM0521, and WUNM0194 (Figure 5.4 A, B, C, E, G). Family WUNM0176 (Figure 5.4 F) has a proband whose daughter, mother, and brother are all asymptomatic and all have different sized repeats. In fact the proband has the smallest repeat out of the four family members.

There was no relationship between expansion size and whether or not the carrier was symptomaticwhich is consistent with the statistical analysis performed above (which includes these families). Four families (WUNM0026, WUNM0226, WUNM0512, WUNM0194) all had affected and unaffected individuals with expansions of roughly the same size. In family WUNM0226, both expansion carriers were identical twins but only one had ALS. In family WUNM0176 the proband had a smaller repeat than three asymptomatic carriers.

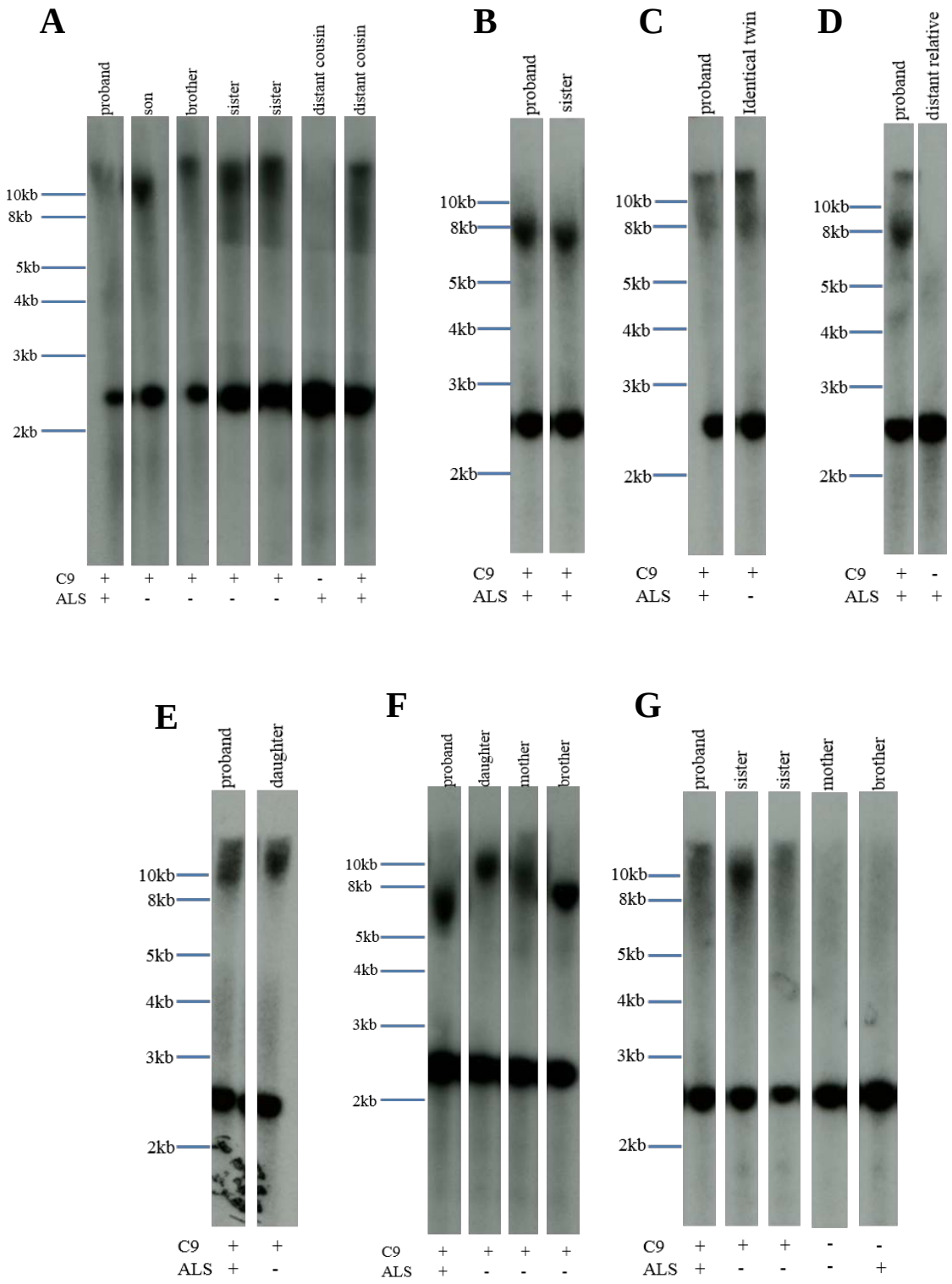

Figure 5.4 Southern blots of blood-derived DNA from 7 ALS families. Samples were digested with XbaI resulting in wild-type alleles of ~2359b.p A.WUNM0026, B. WUNM0160, C. WUNM0226, D. WUNM0085, E. WUNM0521, F. WUNM0176, G. WUNM0194

5.4.3 Expansions in cultured cells

To observe how *C9ORF72* repeat expansions behave in cultured cells, we performed Southern blots on fibroblast-derived DNA from 19 expansion carriers, all of whom were symptomatic. Expansions in fibroblasts were uniformly smaller than blood-derived DNA and peak sizes ranged from 340-960 repeats (Figures 5.5 and 5.6A) with the exception of one sample with a small expansion of ~35 repeats. The median peak number of repeats was 703. Some fibroblast lines showed multiple sizes of expansions (Figure 5.6A). Unsurprisingly, due to the narrow range of expansion sizes in fibroblasts, there was no correlation between number of repeats in fibroblast-derived DNA and age of onset. There were also no associations between number of repeats and survival (p=0.16, p=0.17, p=0.09 for minima, peaks, and maxima respectively).

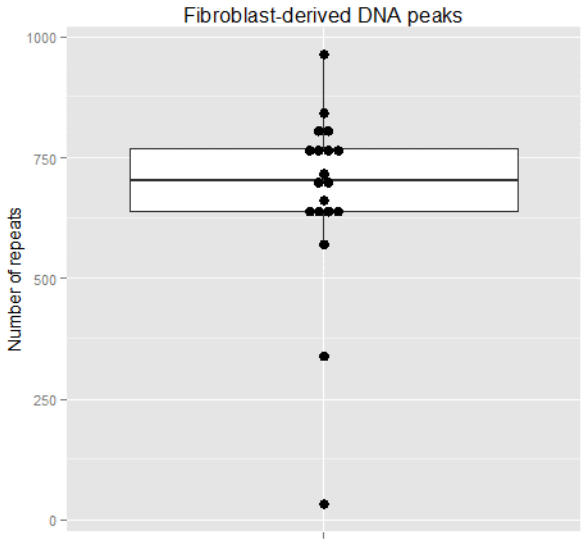

Figure 5.5 Peak expansion sizes in fibroblast-derived DNA

Subject 32 showed multiple expansion sizes in fibroblast-derived DNA (Figure 5.6 A) that were both smaller than the single band observed in the blood-derived DNA from the same patient (Figure 5.6 A). There was no correlation between peak expansion size in fibroblast-derived DNA and peak expansion size in blood-derived DNA (data not shown). There were no significant correlations between peak, maximum, or minimum number of repeats in fibroblast-derived and blood-derived DNA.

We performed Southern blots of fibroblasts at different passages to observe changes in expansion size through cell divisions. The expansion was larger in later passages in two different cell lines (Figure 5.6 B). We also performed Southern blots of a lymphoblast cell line from patient ND08980 from the Coriell repository (Figure 5.6 C). This cell line showed multiple expansion sizes and was also unstable with the latest passage (p17) only showing one expansion size.

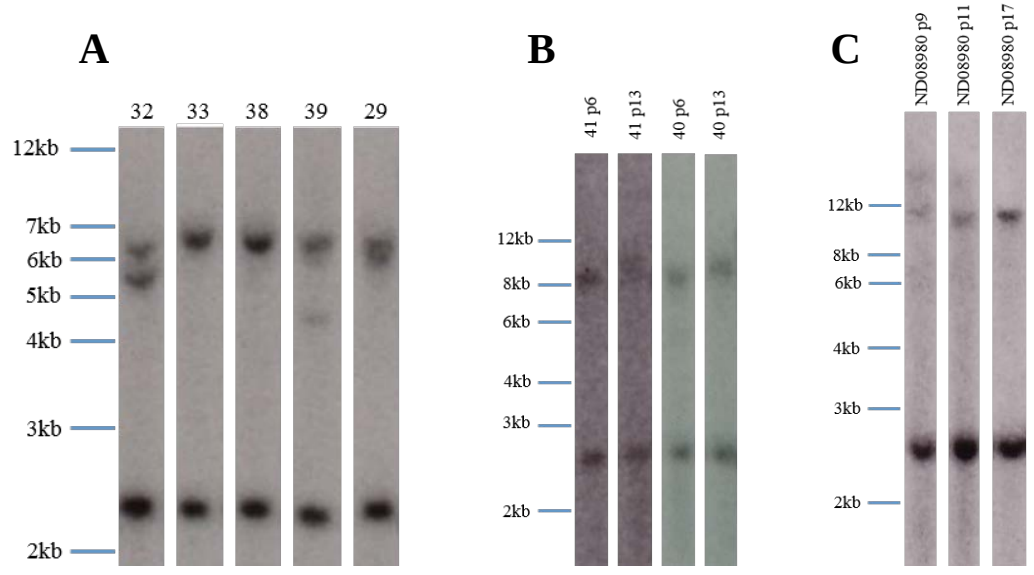

Figure 5.6 Southern blots of cultured cells. Samples were digested with XbaI resulting in wild-type alleles of ~2359bp. A. Representative blot of cultured patient fibroblast-derived DNA. B. Fibroblast-derived DNA from two patients at passages 6 and 13. C. Lymphoblast-derived DNA from one patient at passages 9, 11, and 17

5.4.4 Expansions in different tissues

Next we used tissue from patient autopsies to observe somatic instability of the repeat expansion. In initial studies we noted that expansions in brain derived DNA were larger than in blood and fibroblasts from the same individual (Figure 5.7), so we expanded our analysis to a range of tissues across the entire body and multiple regions of the CNS.

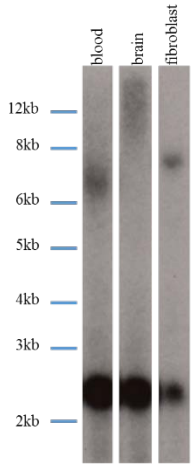

Figure 5.7 Expansions in blood, brain and fibroblast derived DNA from a single individual (patient 18). Sample were digested with XbaI resulting in wild-type alleles of ~2359bp.

We observed variability in expansion size across tissue samples from the same patients (Figure 5.8). The most obvious pattern was that expansions in DNA from the cerebellum are smaller than other brain regions (seen in Figure 5.8A). This was consistent across all individuals. Another interesting finding was that expansions in non-CNS tissues are within the same size range as CNS tissues. In fact it appears that expansions in the kidney are in fact larger than in multiple brain regions (Figure 5.9). Expansions in DNA derived from the testes was also very large, although they also showed a significant amount of DNA that did not migrate out of the

wells of the gel. It is unclear if this is DNA that has very large expansions or if it is simply less soluble.

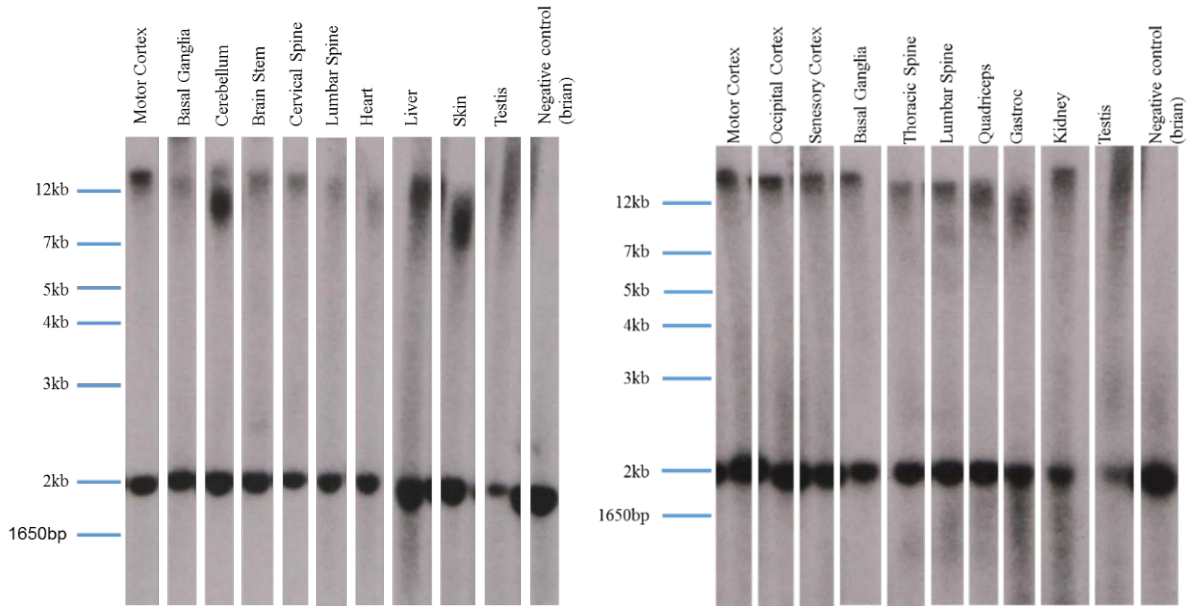

Figure 5.8 Southern blots across multiple tissues. Samples were digested with XbaI+Taq$^{\alpha}$I resulting in wild-type alleles of ~1941bp. A. Patient 47, B. Patient 68

We compared number of repeats across different tissue types for which there were at least three subjects with usable measurements. This included cerebellum (n=6, median =1570, range=996-1761), frontal cortex (n=4, median=2317, range=1588-2413), occipital cortex (n=5, median=2088, range=1290-3375), sensory cortex (n=3, median=2391, range= 2145-2524), motor cortex (n=6, median=2334, range=1447-2716), cervical spine (n=4, median=2171, range=1290-3677), lumbar spine (n=4, median=2326, range=1930-3615), kidney (n=3, median=2951, range=2187-4296), and blood (n=39, median=1702, range=705-3520). Each tissue was compared to all of the others and we found that expansions in the cerebellum were on

average smaller than in frontal cortex, sensory cortex, lumbar spine, and kidney. No other comparisons were statistically significant (Table 5.1). We also performed paired tests comparing the differences in expansion sizes between tissue types within individuals. None of these comparisons were significant (Table 5.2). There were also no significant correlations between expansion size in a given tissue and expansion size in any other tissue (data not shown).

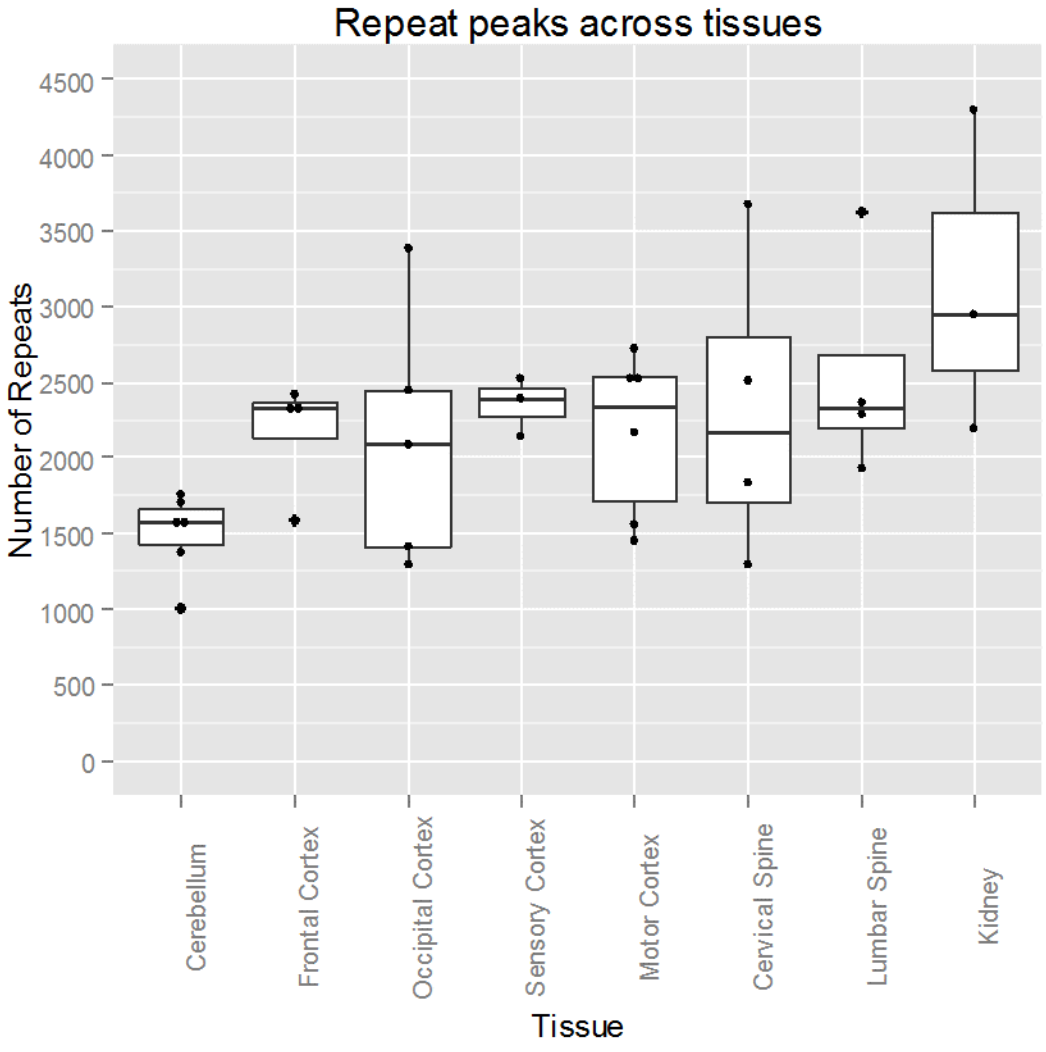

Figure 5.9 Repeat sizes across multiple tissues. Only tissues with at least three subjects blotted were included

Table 5.1 P-values from Wilcoxon rank-sum test comparing expansion sizes in different tissue types. Only peak repeat numbers were used.

	Cerebellum	Cervical Spine	Motor Cortex	Occipital Cortex	Frontal Cortex	Kidney	Lumbar Spine	Sensory Cortex	Blood
Cerebellum	-	-	-	-	-	-	-	-	-
Cervical Spine	0.1714	-	-	-	-	-	-	-	-
Motor Cortex	0.1320	1.0000	-	-	-	-	-	-	-
Occipital Cortex	0.3290	0.8057	0.5368	-	-	-	-	-	-
Frontal Cortex	**0.0381**	0.8857	0.9143	0.9048	-	-	-	-	-
Kidney	**0.0238**	0.4000	0.1667	0.2500	0.4000	-	-	-	-
Lumbar Spine	**0.0095**	0.8857	0.7619	0.5556	0.8857	0.6286	-	-	-
Sensory Cortex	**0.0238**	0.8571	1.0000	0.5714	0.6286	0.4000	0.8571	-	-
Blood	0.4133	0.2502	0.1287	0.3180	0.2333	0.0191	0.0689	0.0873	-

Table 5.2 P-values from Wilcoxon paired signed rank tests comparing expansion sizes in different tissue types within individuals. Only peak number of repeats were used

	Cerebellum	Cervical Spine	Motor Cortex	Occipital Cortex	Frontal Cortex	Kidney	Lumbar Spine	Sensory Cortex	Blood
Cerebellum	-	-	-	-	-	-	-	-	-
Cervical Spine	0.375	-	-	-	-	-	-	-	-
Motor Cortex	0.0625	0.875	-	-	-	-	-	-	-
Occipital Cortex	0.25	1	0.625	-	-	-	-	-	-
Frontal Cortex	0.125	0.5	0.25	0.25	-	-	-	-	-
Kidney	0.25	0.5	0.25	0.5	0.5	-	-	-	-
Lumbar Spine	0.125	0.75	0.875	1	0.75	0.25	-	-	-
Sensory Cortex	0.25	1	0.25	1	0.5	0.5	0.25	-	-
Blood	1	0.25	0.125	0.125	0.5	0.25	0.25	0.5	-

We analyzed the relationship between expansion size in different tissue regions and clinical course (Table 5.3). There was a moderate association that was not statistically significant

between age of onset and expansion size in the sensory cortex, however there were only three observations for this tissue.

Table 5.3 Associations between expansion sizes in different tissues and clinical presentation. P-values from spearman correlation and cox-proportional hazards models are given for associations with age of onset and survival respectively.

	Age of onset	Survival
Cerebellum	0.8569	0.27
Cervical Spine	0.6428	0.5
Motor Cortex	0.4441	0.5
Occipital Cortex	0.9779	0.27
Frontal Cortex	0.5549	1
Kidney	0.7727	1
Lumbar Spine	0.8416	0.55
Sensory Cortex	0.0611	1

5.5 Discussion

Based on characteristics of other repeat expansion disorders and the relative lack of knowledge regarding characteristics of repeat expansions, we investigated these characteristics of the *C9ORF72* repeat expansion in our ALS cohort. We found all repeat expansions to be very large, ranging from ~700 repeats to ~4500 repeats, and showing a high degree of instability. Within patients, expansion size varied with the smallest expansions in the cerebellum (not including cultured fibroblasts). There were no significant differences between different CNS regions or between CNS and non-CNS tissues (blood and kidney). The same pattern of smaller repeats in the cerebellum compared to other brain regions has been observed in all studies comparing *C9ORF72* expansions across brain regions.[18,22,23]

We did observe instability of the repeat during transmission within families. Unlike repeat expansion disorders with anticipation, in which expansions become larger with each subsequent generation, we saw both families with increases or decreases in expansion size and families with no change in expansion size from parents to offspring. One family in fact showed a decrease in size from an unaffected parent to the affected proband and then an increase in size from the proband to their daughter.

We saw no difference in expansion size in blood-derived DNA between symptomatic and asymptomatic expansion carriers from ALS families. There was in fact a set of monozygotic twins with identical repeat sizes that were discordant for disease. Another set of monozygotic twins with *C9ORF72* expansions that are discordant for disease has been reported as well.[26] It is possible that symptomatic expansion carriers have larger expansions in CNS tissues, however no tissue from asymptomatic carriers was available for comparison.

Our analysis did not identify any associations between repeat length in any tissue and either age of onset or survival. While other studies have reported associations, they have not been consistent between studies. Beck *et al.*[19] found smaller expansions in blood to be associated with earlier age of onset, while others did not.[18,23] van Blitterswijk *et al*[18] found smaller repeats in the frontal cortex were associated with earlier age of onset and larger repeats in cerebellum were associated with shorter survival in patients with FTD. There were no associations in ALS. Nordin *et al.* [23]found smaller expansions in the cerebellum and parietal lobe to be associated with earlier age at onset and larger expansions in the parietal lobe to be association with shorter survival.

The lack of correlations with clinical characteristics in our study may be due to the small sample-size. The technical difficulty and amount of sample required to perform Southern blots

limited the number of usable observations for analysis. Additionally, the imprecise nature of obtaining the number of repeats from Southern blot images decreases the accuracy of our analysis. Size estimation of expansions larger than 20kb becomes even more inaccurate as there small errors in measurement translate into large differences in estimated size. For these reasons, the development of novel methods, such as single-molecule real-time sequencing of long reads used to sequence expansions in *FMR1*[27], would be useful for continued study of this repeat expansion.

One interpretation of these results is that the number of repeats does not determine toxicity. It is possible that there is a threshold effect where expansions become toxic once they reach a certain size and additional repeats do not have any added effect. The fact that we see no difference in expansion sizes between affected and unaffected tissues points to a tissue specific ability to cope with toxic repeat expansions. Furthermore, each tissue tested contains a heterogeneous mix of cell types (neurons, microglia, astrocytes, etc.). It is possible that differences will become more apparent if we look specifically at different types of neurons. Additionally, the lack of difference in expansion size between unaffected carriers indicates that additional genetic and environmental factors are involved in pathogenesis. Further studies of repeat sizes in larger cohorts are needed to confirm these findings and to improve the analysis of clinical characteristics.

5.6 References

1. DeJesus-Hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72, 245–256 (2011).

2. Renton, A. E. et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72, 257–268 (2011).

3. Haeusler, A. R. et al. C9orf72 Nucleotide Repeat Structures Initiate Molecular Cascades of Disease. Nature 507, 195–200 (2014).

4. Fratta, P. et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. Sci Rep 2, 1016 (2012).

5. Šket, P. et al. Characterization of DNA G-quadruplex species forming from C9ORF72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Neurobiology of Aging 36, 1091–1096 (2015).

6. Lagier-Tourenne, C. et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc. Natl. Acad. Sci. U.S.A. 110, E4530–4539 (2013).

7. Cooper-Knock, J. et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain 137, 2040–2051 (2014).

8. Sareen, D. et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. Sci Transl Med 5, 208ra149 (2013).

9. Lee, Y.-B. et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. Cell Rep 5, 1178–1186 (2013).

10. Ash, P. E. A. et al. Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. Neuron 77, 639–646 (2013).

11. Mori, K. et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science 339, 1335–1338 (2013).

12. Kwon, I. et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science 345, 1139–1145 (2014).

13. Mizielinska, S. et al. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. Science 345, 1192–1194 (2014).

14. Anvret, M. et al. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. Hum. Mol. Genet. 2, 1397–1400 (1993).

15. Martorell, L. et al. Progression of somatic CTG repeat length heterogeneity in the blood cells of myotonic dystrophy patients. Hum. Mol. Genet. 7, 307–312 (1998).

16. Mahadevan, M. et al. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. Science 255, 1253–1255 (1992).

17. Buxton, J. et al. Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. Nature 355, 547–548 (1992).

18. Van Blitterswijk, M. et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. The Lancet Neurology 12, 978–988 (2013).

19. Beck, J. et al. Large C9orf72 Hexanucleotide Repeat Expansions Are Seen in Multiple Neurodegenerative Syndromes and Are More Frequent Than Expected in the UK Population. The American Journal of Human Genetics 92, 345–353 (2013).

20. Dobson-Stone, C. et al. C9ORF72 repeat expansion in Australian and Spanish frontotemporal dementia patients. PLoS ONE 8, e56899 (2013).

21. Dols-Icardo, O. et al. Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. Hum. Mol. Genet. 23, 749–754 (2014).

22. Buchman, V. L. et al. Simultaneous and independent detection of C9ORF72 alleles with low and high number of GGGGCC repeats using an optimised protocol of Southern blot hybridisation. Mol Neurodegener 8, 12 (2013).

23. Nordin, A. et al. Extensive size variability of the GGGGCC-expansion in C9orf72 in both neuronal and non-neuronal tissue in 18 patients with ALS or FTD. Human Molecular Genetics (2015). doi:10.1093/hmg/ddv064

24. Fratta, P. et al. Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion. Neurobiology of Aging 36, 546.e1–546.e7 (2015).

25. Harms, M. B. et al. Lack of C9ORF72 coding mutations supports a gain of function for repeat expansions in amyotrophic lateral sclerosis. Neurobiol. Aging 34, 2234.e13–19 (2013).

26. Xi, Z. et al. Identical twins with the C9orf72 repeat expansion are discordant for ALS. Neurology 83, 1476–1478 (2014).

27. Loomis, E. W. et al. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Res. 23, 121–128 (2013).

Chapter 6

Identification of mutations in the TREM family genes

The work presented in this chapter resulted in the following publication:

1. The *TREM2* variant p.R47H is a risk factor for sporadic amyotrophic lateral sclerosis. Cady J, Koval ED, Benitez BA, Zaidman C, Jockel-Balsarotti J, Allred P, Baloh RH, Ravits J, Simpson E, Appel SH, Pestronk A, Goate AM, Miller TM, Cruchaga C, Harms MB. JAMA Neurol. 2014 Apr;71(4):449-53. doi: 10.1001/jamaneurol.2013.6237

6.1 Abstract

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease in which microglia play a significant and active role. Recently, a rare missense variant (p.R47H) in the microglial activating gene *TREM2* was found to increase the risk of several neurodegenerative diseases, including Alzheimer's disease. Here we show that the *TREM2* variant p. R47H was more common in subject with ALS than in controls and is therefore a significant risk factor for ALS (OR=2.40; 95%CI=1.29-4.15; $p=4.1x10^{-3}$). Furthermore, *TREM2* expression was increased in spinal cords from ALS patients and $SOD1^{G93A}$ mice ($p=2.8x10^{-4}$, $p=2.8x10^{-9}$ respectively), confirming dysregulated TREM2 in disease. *TREM2* expression in human spinal cord was negatively correlated with survival (p=0.04), but not other phenotypic aspects of disease. This study demonstrates that the *TREM2* p.R47H variant is a potent risk factor for sporadic amyotrophic lateral sclerosis. These findings identify the first genetic influence on neuro-inflammation in ALS and highlight the TREM2 signaling pathway as a therapeutic target in ALS and other neurodegenerative diseases.

6.2 Introduction

Activated microglia in the vicinity of degenerating neurons are a long-recognized pathological feature of ALS[1], but whether such activation is a beneficial response or injurious contributor to the disease process remains unclear. In fact the answer may be both- mouse model data show that microglia express both neuroprotective and neurotoxic factors simultaneously[2] and may transition from a neuroprotective phenotype at symptom onset to become more neurotoxic later in the disease course.[3]

There are many signaling pathways governing microglial phenotype, including a complex formed by TREM2 (MIM 605086) and TYROBP (also known as DAP12, MIM 604142).[4] Activation of TREM2/TYROBP results in a potentially neuroprotective microglial state, with improved phagocytosis of apoptotic cellular debris and down-regulation of inflammatory cytokines.[5] The importance of signaling through TREM2/TYROBP is made clear by the fact that recessive mutations in either gene cause early onset frontotemporal-like dementia, either in isolation[6] or as part of the recessive human disease polycystic lipomembranous osteodysplasia with sclerosing leukoencephalopathy (PLOSL or Nasu-Hakola Disease, MIM 221770).[7,8] Furthermore, recent studies have demonstrated that a rare non-synonymous variant in *TREM2*, rs75932628 (encoding p.R47H) is a strong risk-factor for Alzheimer's disease (AD), another neurodegenerative disease characterized by microglial activation.[9–13] Other studies have implicated the same variant in frontotemporal dementia and Parkinson's disease.[14,15] It has been hypothesized that this variant impairs TREM2/TRYOBP signaling, thereby blunting neuroprotective microglial activation and exacerbating the disease process.[9,11]

In this study we demonstrate that p.R47H is a risk factor for sporadic ALS, and demonstrate upregulation of TREM2 in human ALS spinal cord and in spinal cord from the G93A mouse model of *SOD1* ALS.[16]

6.3 Methods

6.3.1 Study subjects and *TREM2* p.R47H genotyping

923 sporadic ALS (SALS) subjects and 1854 normal controls, all of self-reported non-Hispanic white background, were included and provided written informed consent at their contributing institution. Diagnoses of probable or definite ALS were made by neuromuscular specialists according to El Escorial criteria (Washington University in St. Louis, n= 273; Virginia Mason Medical Center, n=143; Methodist Neurological Institute, n= 47; Coriell plates NDPT025, NDPT026, NDPT100, NDPT103, NDPT106, n=460). Control subjects were without ALS, Parkinson's disease, or dementia and were collected from ongoing studies (Washington University, n=1390) or Coriell panels (NDPT020, NDPT079, NDPT082, NDPT095, NDPT096, n=464). 55% of SALS cases were male and age at DNA collection was 61.0±11.6 years old (mean ± stdev), while the control cohort was 44% male and 68±13.6 years old (mean ± stdev). An additional control group of 25,023 individuals of European or European American descent was collated from published studies (Table 1) and from the unrelated European Americans (EA) genotyped by whole-exome sequencing as part of the NHLBI's Exome Sequencing Project or ESP (http://ESP6500.gs.washington.edu/ESP6500/ [1 Dec 2013]).) Icelandic controls were not included given the isolation of the population and significantly higher MAF at rs75932628.[9]

DNA was extracted from blood or saliva using standard methods and genotyped for rs75932628 (*TREM2* p.R47H) using a custom KASPar (KBioscience) assay.[12] Genotype call rate was 99.7% in both cases and controls. p.R47H carriers were validated by sequencing.

6.3.2 TREM2 expression analysis

Expression in human lumbar spinal cord: Total RNA was extracted from snap-frozen transverse sections of lumbar spinal cord of 18 autopsied subjects with ALS (Table 2) and 12 controls without neurological disease using the miRNeasy kit (Qiagen). Extracted RNA was quantified and 40ng was used as input for the Express One-Step Superscript qRT-PCR Universal (Invitrogen: 11781-200) with validated Taqman assays for human TREM2 and three endogenous controls, GAPDH, PPIA, and RPLPO (Applied Biosystems 4331182, 4333764F, 4333763F, 4333761F respectively). Reactions were run in duplicate on an ABI 7500 fast thermocycler. *TREM2* expression was normalized to the geometric mean of the three endogenous controls. 10 subjects had provided separate informed consent for genetic analysis, but none were found to carry the p.R47H variant.

Expression in mouse spinal cord: Total RNA was extracted from saline-perfused and snap-frozen spinal cords of 8 end-stage $SOD1^{G93A}$ transgenic mice (Jackson Lab B6.Cg-Tg(SOD1*G93A)1Gur/J) and 6 negative littermate controls. *TREM2* expression was quantified using a mouse-specific *TREM2* Taqman assay (Applied Biosystems 4331182) and normalized to the endogenous control SMRT using primers and probe from IDT DNA (Probe: AGACGTCTCACACAAGGAAGGACTCGCC, Forward primer: GGGTATATTTTTGATACCTTCAATGAGTTA, Reverse primer

TCTGAAACAGTAGGTAGAGACCAAAGC). Reactions were run in duplicate on an ABI 7500 fast thermocycler.

6.3.3 Exome Sequencing

Whole exome sequencing was generated from 735 patients either at Washington University or at Duke University as part of a large ALS exome sequencing initiative.[17] Sequences were aligned to the human genome reference Hg19 using Novoalign (http://www.novocraft.com) and variants were called by Samtools.[18] PCA analysis was used to identify 608 subjects of European ancestry that were used for analysis. Variants in the TREM locus (chr6: 41116999-41254457) and *TYROBP* (chr19: 36395303-36399211) were queried using vcftools.[19] Variants in these regions were then annotated using SeattleSeq (http://sngs.washington.edu/SeattleSeqAnnotation131/), and filtered them to included variants that resulted in coding changes in any transcript of a particular gene. Variants in the same genomic regions were extracted from the non-Finnish European population in the ExAC database (Exome Aggregation Consortium, Cambridge, MA, http://exac.broadinstitute.org [May,2015]) for comparison of variant frequencies.

6.3.4 Statistical analysis

All statistics were computed using R version 3.0.1 except as noted. Fisher's exact test was used to compare proportions of p.R47H carriers in cases and controls. Comparisons of TREM2 expression utilized student t-tests, while correlations between TREM2 expression and subject characteristics utilized Spearman correlations (continuous variables) or Mann-Whitney U (dichotomous variables). Logistic regression was performed in PLINK with age and gender as covariates using cases and controls for whom this data was available (913/920 of ALS subjects and 1803/1848 of controls). All tests were two-tailed, with the significance level set at p=0.01 to

correct for multiple comparisons. Single-variant association tests in the TREM family genes were performed using fisher's exact tests. Gene-based tests were performed using SKAT as described in chapter 3.

6.4 Results

6.4.1 *TREM2* p.R47H in sporadic ALS

1.09% (10/920) of sporadic ALS subjects and 0.162% (3/1848) of normal controls were heterozygous carriers of the p.R47H variant, showing a significant enrichment in ALS (OR=6.77; 95% CI 1.86-24.65; p=0.0016). No cases or controls were homozygous for this allele. Because the proportion of p.R47H carriers in the population declines with age[7] and our cases were younger than controls, we also analyzed our data by logistic regression with age and gender as covariates. This produced a similar risk estimate (OR=7.38; 95%CI = 1.95-27.9, p= 0.0032). To provide a more conservative estimate of effect size, we also compared our sporadic ALS cohort to an aggregate control population of European ancestry gleaned from published studies and databases (n=25,023, Table 6.1). We again observed an enrichment in sporadic ALS, albeit with a lower effect size (OR=2.81; CI 1.31-5.41; $p=4.8x10^{-3}$). A prior study of a smaller cohort of North American ALS patients found a non-significant but increased frequency in cases versus controls (0.7% vs. 0.45%).[14] A combined analysis of this study with ours compared to all available controls also showed a significant association (OR=2.40; 95%CI=1.29-4.15; $p=4.1x10^{-3}$), confirming that *TREM2* p.R47H is a risk factor for ALS. *TREM2* p.R47H carriers in our cohort showed no difference in age of symptom onset, site of first symptom, or in survival compared to those without. However, the rarity of the variant limited our power to detect such a

difference. A larger cohort of p.R47H carriers with ALS will be required to definitively determine effects on disease parameters.

Table 6.1 *TREM2* p.R47H carriers in published cohorts and this study. Abbreviations: MAF= minor allele frequency; ESP-EA= Exome Sequencing Project European American; N. America=North American

Cohort		No. Subjects	p.R47H carriers	MAF (%)
Controls	ESP-EA	4,300	22	0.26
	Spain[12]	550	0	0.00
	Georgia, USA[9]	402	1	0.12
	Germany[9]	1,891	7	0.19
	Netherlands[9]	4,950	15	0.15
	Norway[9]	2,484	8	0.16
	N. America/UK[11]	5,166	20	0.19
	Utah, USA[13]	2,540	12	0.24
	France[10]	783	4	0.26
	N. America/Ireland/Poland[14]	1,957	8	0.20
	This Study (N. America)	1,848	3	0.08
	Total Controls	26,871	100	0.19
ALS	This Study (N. America)	920	10	0.54
	N. America[14]	765	5	0.33
	Total ALS Subjects	1685	15	0.45

6.4.2 TREM2 expression in spinal cords from humans with ALS and SOD1^{G93A} mice

In collaboration with Tim Miller's lab, we examined spinal cord expression of TREM2 in lumbar spinal cord sections from 18 subjects with ALS and found a 2.8-fold upregulation compared to controls ($p=2.8x10^{-4}$; Figure 6.1 panel A). Expression levels did not correlate with age of onset, site of symptom onset, or presence of a known disease-causing mutation (Table 6.2). However, the degree of upregulation showed a modest inverse correlation with disease survival that was not statistically significant after correction for multiple comparisons. Because markers of microglial activation are also upregulated in models of SOD1 ALS[20], we evaluated

TREM2 expression in SOD1^{G93A} transgenic mice and found a 13-fold increase compared to non-transgenic littermates ($p=2.8x10^{-9}$; Figure 6.1 panel B).

Table 6.2 ALS autopsy subjects studied for TREM2 expression in lumbar spinal cord.
[a] Survival defined as symptom onset to death or full-time ventilation.
[b] 3 subjects had *SOD1* mutations (A4V, G85R, I133T) and 3 had *C9ORF72* repeat expansions.
[c] Spearman correlation, unadjusted for multiple comparisons.
[d] Mann-Whitney U rank test, unadjusted for multiple comparisons.

Demographic Category	Metric		Correlation	P-value
Age at onset (years, n=17)	Mean ± stdev (range)	61±13 (29-75)	r=-0.03	0.9[c]
Survival[a] (months, n=18)	Mean ± stdev (range)	31.3±28.0 (4-108)	r=-0.49	0.04[c]
Postmortem interval (hours, n=11)	Mean ± stdev (range)	12.4±7.5 (2-28)	r=-0.45	0.17[c]
Site of onset (n=16)	% Bulbar (n)	31 (5)		0.21[d]
Genetic cause[b] (n=18)	% With known gene (n)	33 (6)		0.21[d]

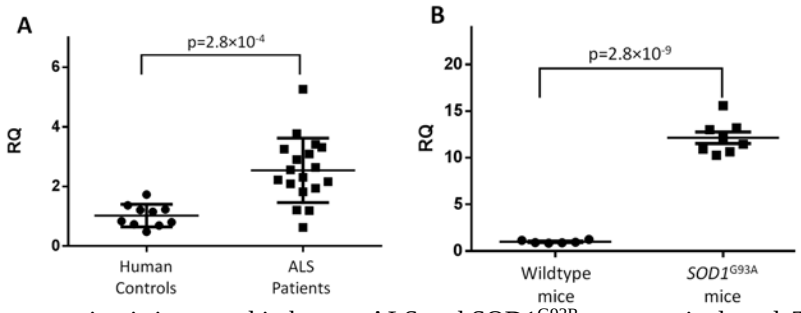

Figure 6.1 TREM2 expression is increased in human ALS and SOD1^{G93R} mouse spinal cord. TREM2 expression was measured by qPCR in A) lumbar spinal cord sections from 18 ALS subjects and 12 controls and normalized to the geometric mean of three endogenous control genes. B) In mice, expression was measured in spinal cords from 8 SOD^{G93R} mice and 6 wild-type littermates with normalization to an endogenous control. p-values were calculated using two-tailed student's t-test.

6.4.3 Additional mutations in *TREM2* and TREM family genes

We used whole-exome sequences of ALS patients to determine if there were additional *TREM2* mutations that conferred a risk of ALS. There were 7 coding variants in *TREM2* in the exomes of 608 ALS patients with European ancestry, none of which were novel (table 6.3). When compared to the population of non-Finnish Europeans (NFE) from the ExAC browser, no variants were significantly associated with disease status. Although there is overlap between the

samples included in whole exome sequencing and the cohort we originally genotyped, the p.R47H signal did not replicate in this set. The majority of the association came from the Coriell samples which were not included in these exomes. The frequency of rare coding variants in *TREM2* was 0.020 in ALS patients and 0.021 in the NFE population in the ExAC browser. The distribution was not different by SKAT analysis (p=0.99).

Table 6.3 Variants in *TREM2* detected by exome sequencing.

					Allele Counts				
Chr	**Hg19 Position**	**Amino Acids**	**SNP ID**	**Alleles (Ref/Alt)**	**Alt ALS**	**Ref ALS**	**Alt NFE**	**Ref NFE**	**p-value**
6	41129252	R47H	rs75932628	C/T	3	1215	172	66102	1
6	41129207	R62H	rs143332484	C/T	14	1226	762	65900	1
6	41129133	D87N	rs142232675	C/T	4	1238	119	66609	0.292
6	41129105	T96K	rs2234253	G/T	1	1241	57	66671	1
6	41127543	H157Y	rs2234255	G/A	1	1241	20	66546	0.322
6	41126655	L211P	rs2234256	A/G	1	1239	64	66670	1
6	41126429	W191X	rs2234258	C/T	1	1241	19	62339	0.326

There are 4 additional genes *TREM1, TREML1, TREML2* and *TREML4* in the TREM family of genes, all located in the same genomic region as *TREM2*. The SNP rs3747742 in the related gene *TREML2* has been shown to be protective in Alzheimer's disease.[21] We looked for variants in these four genes in addition to *TYROBP*.

We saw no difference in the frequency of rs3747742 (MAF=29.6% in ALS and MAF= 29.9% in NFE). Aside from this SNP, we only analyzed rare variants that were less than 1% frequency in controls (table 6.4). There were a total of 8 novel variants in these genes, all of which were found in a single individual. None of the variants passed the corrected significance threshold (0.002). The variant that came closest to significance was rs112680060 in *TREML4* (OR=1.9, 95% CI=1.15-3.08).

Table 6.4 Variants in *TYROBP*, *TREM1*, *TREML4*, *TREML2*, and *TREML1*.

						Allele Counts				
Gene	**Chr**	**Hg19 position**	**Amino Acids**	**SNP ID**	**Alleles (Ref/Alt)**	**Alt ALS**	**Ref ALS**	**Alt NFE**	**Ref NFE**	**p-value**
TYROBP	19	36398357	A74T	.	C/T	1	1236	0	65890	0.0184
TYROBP	19	36398414	V55L	rs77782321	C/A	13	1216	637	65217	0.6602
TREM1	6	41250175	P122A	.	G/C	1	1241	0	66466	0.0183
TREM1	6	41248810	C163F	rs201082267	C/A	1	1231	14	66722	0.2400
TREM1	6	41243926	F214L	rs2234245	G/C	1	1241	29	66695	0.4250
TREML4	6	41196169	A2S	rs112680060	G/T	19	1223	490	61320	0.0091
TREML4	6	41197301	T146K	rs9471515	C/A	1	1241	43	66687	0.5559
TREML4	6	41204306	G197S	.	G/A	1	1241	0	66662	0.0183
TREML2	6	41166063	V54F	rs147506354	C/A	4	1238	83	66653	0.0755
TREML2	6	41162556	R131K	.	C/T	1	1237	0	64452	0.0188
TREML2	6	41162395	A185T	.	C/T	1	1241	13	66493	0.2282
TREML2	6	41162371	T193A	rs145455750	T/C	3	1239	167	66287	1.0000
TREML2	6	41162204	S248R	rs115991880	G/T	4	1238	212	26352	0.0668
TREML2	6	41162173	P259T	.	G/T	1	1237	0	12334	0.0912
TREML2	6	41160572	V285I	rs35521209	C/T	1	1241	11	56969	0.2280
TREML1	6	41121702	E57A	.	T/G	1	1239	0	66302	0.0184
TREML1	6	41118614	V171I	rs41273760	C/T	2	1236	7	66627	0.0110
TREML1	6	41118000	M207T	rs35929443	A/G	3	1239	114	65144	0.4838
TREML1	6	41117594	D228E	rs138237630	A/T	4	1226	282	66444	0.8232

We compared total variant burden in each of these genes using SKAT. *TREML4* was the only gene that was significant (table 6.5). This appears to be mainly driven by the higher frequency of rs112680060 in ALS.

Table 6.5 SKAT variant burden testing.

Gene	**Total variant frequency (%) ALS**	**Total variant frequency (%) NFE**	**SKAT p-value**
TREM2	2.01	2.13	0.99
TYROBP	1.12	1.17	0.86
TREML2	1.21	0.90	0.26
TREM1	0.24	0.28	0.77
TREML1	0.80	0.78	0.51
TREML4	1.69	1.05	0.003

6.5 Discussion

Our study demonstrates that a rare variant in *TREM2* (p.R47H), more than doubles the risk of ALS. In addition to identifying a novel risk factor for ALS, this finding provides the first link between genetic variation and microglial activation in ALS pathogenesis. This is important in light of a recent study demonstrating that higher degrees of microglial activation on pathological examination were correlated with both the degree of upper motor neuron symptoms and a more rapid disease progression.[22] Interestingly, our evaluation of TREM2 expression in ALS spinal cord showed a similar trend, with higher levels of TREM2 correlating with shorter survival. Furthermore, our finding that TREM2 expression is also increased in spinal cords from $SOD1^{G93A}$ mice is congruent with recent studies of isolated microglia from this same model[2] and suggests that studies of microglial activation in this model may provide insights relevant to human ALS.

We identified no novel variants in *TREM2* by exome sequencing and none of the known variants were associated with disease state. Interestingly, we did not identify any p.R47H carriers in the additional subjects we analyzed; all three p.R47H carriers were previously discovered in our original cohort. We did not have ages of the controls used for this analysis and were therefore not able to control for possible changes in frequency with age as we did for p.R47H. It is possible that we were unable to detect associations because of this.

p.R47H was first shown to increase risk for Alzheimer's disease with subsequent associations with frontotemporal dementia and Parkinson's disease.[9–14] How the p.R47H variant affects TREM2 function and predisposes to neurodegeneration is currently unknown. Because TREM2 signaling mediates potentially neuroprotective microglial activities (including phagocytosis of apoptotic cells and secretion of anti-inflammatory cytokines), one model

hypothesizes that p.R47H is a loss-of-function allele. Inadequate clean-up of cellular debris and counter-productive inflammation would predispose to symptomatic disease. The p.R47H variant is located in the extracellular domain of TREM2 where it could interfere with binding to unidentified ligand(s) or disrupt signaling through its receptor complex partner TYROBP.

We also looked for variants in the four other genes in the TREM family of membrane receptors, all of which are located in the same genomic location with *TREM2*.[23] There was an increase in variant burden in ALS in *TREML4* with the variant rs112680060 contributing most of the signal. *TREML4* is expressed abundantly in the spleen and recognizes apoptotic and necrotic cells.[24] Like TREM2, TREML4 associates with TYROBP.[24] Although *TREML4* is not normally expressed in microglia, it was shown to be significantly upregulated in microglia in response to damage to dopaminergic neurons.[25] There were no associations with any of the other TREM genes or with *TYROBP*.

As dysregulated TREM2 signaling confers risk for several neurodegenerative disorders, insights gleaned from the study of TREM2 in ALS are likely to be applicable to other diseases and vice versa. This includes the important possibility that manipulation of TREM2 signaling or microglial activation would be a worthwhile therapeutic strategy.

6.6 References

1. Engelhardt, J. I. & Appel, S. H. IgG reactivity in the spinal cord and motor cortex in amyotrophic lateral sclerosis. Arch. Neurol. 47, 1210–1216 (1990).

2. Chiu, I. M. et al. A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. Cell Rep 4, 385–401 (2013).

3. Liao, B., Zhao, W., Beers, D. R., Henkel, J. S. & Appel, S. H. Transformation from a neuroprotective to a neurotoxic microglial phenotype in a mouse model of ALS. Exp. Neurol. 237, 147–152 (2012).

4. Kierdorf, K. & Prinz, M. Factors regulating microglia activation. Front Cell Neurosci 7, 44 (2013).

5. Takahashi, K., Rochford, C. D. P. & Neumann, H. Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. J. Exp. Med. 201, 647–657 (2005).

6. Guerreiro, R. J. et al. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. JAMA Neurol 70, 78–84 (2013).

7. Paloneva, J. et al. Loss-of-function mutations in TYROBP (DAP12) result in a presenile dementia with bone cysts. Nat. Genet. 25, 357–361 (2000).

8. Paloneva, J. et al. Mutations in two genes encoding different subunits of a receptor signaling complex result in an identical disease phenotype. Am. J. Hum. Genet. 71, 656–662 (2002).

9. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N. Engl. J. Med. 368, 107–116 (2013).

10. Pottier, C. et al. TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. J. Alzheimers Dis. 35, 45–49 (2013).

11. Guerreiro, R. et al. TREM2 variants in Alzheimer's disease. N. Engl. J. Med. 368, 117–127 (2013).

12. Benitez, B. A. et al. TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiol. Aging 34, 1711.e15–17 (2013).

13. Gonzalez Murcia, J. D. et al. Assessment of TREM2 rs75932628 association with Alzheimer's disease in a population-based sample: the Cache County Study. Neurobiol. Aging (2013). doi:10.1016/j.neurobiolaging.2013.06.004

14. Rayaprolu, S. et al. TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. Mol Neurodegener 8, 19 (2013).

15. Benitez, B. A., Cruchaga, C. & United States–Spain Parkinson's Disease Research Group. TREM2 and neurodegenerative disease. N. Engl. J. Med. 369, 1567–1568 (2013).

16. Gurney, M. E. et al. Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. Science 264, 1772–1775 (1994).

17. Cirulli, E. T. et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 347, 1436–1441 (2015).

18. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).

19. Danecek, P. et al. The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011).

20. Hall, E. D., Oostveen, J. A. & Gurney, M. E. Relationship of microglial and astrocytic activation to disease onset and progression in a transgenic model of familial ALS. Glia 23, 249–256 (1998).

21. Benitez, B. A. et al. Missense variant in TREML2 protects against Alzheimer's disease. Neurobiol. Aging 35, 1510.e19–26 (2014).

22. Brettschneider, J. et al. Microglial activation correlates with disease progression and upper motor neuron clinical symptoms in amyotrophic lateral sclerosis. PLoS ONE 7, e39216 (2012).

23. Allcock, R. J. N., Barrow, A. D., Forbes, S., Beck, S. & Trowsdale, J. The human TREM gene cluster at 6p21.1 encodes both activating and inhibitory single IgV domain receptors and includes NKp44. Eur. J. Immunol. 33, 567–577 (2003).

24. Hemmi, H. et al. A New Triggering Receptor Expressed on Myeloid Cells (Trem) Family Member, Trem-Like 4, Binds to Dead Cells and Is a DNAX Activation Protein 12-Linked Marker for Subsets of Mouse Macrophages and Dendritic Cells. J Immunol 182, 1278–1286 (2009).

25. Thomas, D. M., Francescutti-Verbeem, D. M. & Kuhn, D. M. Gene expression profile of activated microglia under conditions associated with dopamine neuronal damage. FASEB J. 20, 515–517 (2006).

Chapter 7

Conclusions and future directions

7.1 Contribution of known ALS genes

Amyotrophic lateral sclerosis (ALS) is a degenerative disease of the motor neurons which is universally fatal. Over 20 genes have been identified that cause ALS. Mutations in these genes explain up to 70% of ALS cases with a family history of disease (FALS), with hexanucleotide repeat expansions accounting for 40% of cases alone. The vast majority of ALS cases have no family history and are considered sporadic (SALS). Some of these cases also have mutations in FALS genes, with about 6% caused by *C9ORF72*. Current hypotheses for genetic contributions to SALS include oligogenic inheritance, where mutations in multiple ALS genes are required for disease, and rare variants in ALS genes that increase risk of ALS.

Using pooled-sample sequencing we investigated variations in 17 known ALS genes and found that potentially pathogenic mutations in 64.3% of familial ALS cases and 27.8% of sporadic ALS cases. This was higher than any previous study and while this is possibly due to genetic heterogeneity it is more likely due to differences in which genes were included. In the past, only the most commonly mutated genes in ALS were screened in large cohorts (*SOD1, TARDBP, FUS, ANG, OPTN, VCP,* and recently *C9ORF72).* The decreasing cost of next-generation sequencing enabled our pooled sequencing strategy and other targeted sequencing techniques to screen larger numbers of genes.[1,2]

Furthermore, the proportion of potentially pathogenic mutations was highly influenced by the definition of which mutations are considered pathogenic. The use of common guidelines in

reporting novel mutations will improve consistency and allow us to make accurate comparisons between studies.[3] Several criteria should be considered when evaluating potentially pathogenic variants including the segregation of the variant with affected members in a family, absence of the variant in control populations, and functional evidence. Due to the late onset of disease, we are less likely to have DNA from parents and other family members to test for segregation. Therefore variant frequency in controls and functional evidence will be more important in the assessment of pathogenicity.

We considered all mutations that had a frequency of less than 1% in control databases to be potentially pathogenic. Other studies typically either exclude all variants detected in public databases or internally sequenced controls. We found that excluding all control database variants removed several mutations that have previously been reported to cause ALS while all of these remained when using a threshold of 1% minor allele frequency. The presence of these mutations in controls is likely a reflection of the small number of controls used when they were identified, prior to our knowledge of the abundance of rare variation in all individuals.[4,5] Alternatively, the presence of these mutations in controls could represent lower penetrance alleles or contamination with individuals at-risk for ALS but not yet manifesting symptoms.

Statistical significance is important to consider when comparing novel variants to control populations; however most studies lack the power to detect statistically significant associations of rare variants. We compared frequencies of all rare variants identified in our sporadic ALS cohort and did not find any that were significantly more common in ALS. This includes mutations that are known to be pathogenic, such as *SOD1* mutations p.A5V and p.I114T. Additionally, some genes are more tolerant to variation than others and therefore mutations would be less likely to have an effect.[6,7] Gene-based test such as SKAT take this into account by

comparing the mutation frequencies across a region in cases and controls and can show whether a gene is more commonly mutated in affected individuals compared to the mutation frequency in the general population. In a gene-based test looking at total variation in *SOD1,* we did achieve statistical significance, however this was the only gene that did. This type of collapsing method can be further refined to identify specific domains of a gene or expanded to identify entire pathways and networks that are more frequently mutated in disease.

There are several *in silico* pathogenicity prediction algorithms that can be used to prioritize variants for follow-up, however we have not used them here to exclude potentially pathogenic variants. It is important to consider the specific gene and disease under investigation when using these *in silico* methods. These algorithms evaluate pathogenicity of variants using evolutionary conservation and changes in structural and biochemical properties. Drastic changes in amino acid properties are, however, more likely to result in early onset disease[8], therefore a mutation involved in a late onset disease such as ALS might be more likely to be predicted as benign. Furthermore, mutations that result in late onset disease are not under purifying selection and the positions will not be as highly conserved, again making them more likely to be predicted to be benign. There is also evidence that different *in silico* predictions and combinations of multiple predictions are more accurate for different genes.[9]

Experimental evidence will be necessary to validate the predicted functional effects of potentially pathogenic variants. The use of induced pluripotent stem cells (iPSCs) and direct conversion of adult tissues such as fibroblasts into other tissue types such as neurons or microglia will be particularly valuable to show the effects of mutations in the context of the patient's entire genome. Advances in genome editing technologies such as CRISPR/Cas9 will allow us to determine if a particular mutation is causative by converting it to wild-type and

observing if this corrects the pathogenic phenotype.[10] This technology can also be used to delete large segments of DNA and could potentially remove the *C9ORF72* repeat expansion.[11] This will be particularly interesting in patients with mutations in multiple ALS genes. Based on our hypothesis that mutations in multiple genes are not required for disease but act synergistically, we would expect that correcting only one mutation would not fully correct the phenotype.

Despite finding mutations in a higher than expected proportion of subjects, nearly 70% of our cases remain unexplained. The falling cost of next-generation sequencing has made it possible to perform whole exome sequencing (WES) on larger numbers of patients to identify novel genes involved in disease. Recently, our patients were included in an ALS consortium that sequenced 2869 ALS subjects and identified the gene *TBK1* as risk factor.[12] As additional WES data continues to be generated, novel approaches to analysis and refinement of *in silico* pathogenicity predictions will be required to tease out meaningful associations from massive amounts of information. Current analysis focus on the effects of coding variants and ignores the hundreds of mutations in regulatory regions and microRNAs. Data generated from the ENCODE project, which aims to annotate all functional elements of the human genome[13], will help us interpret variants that may be causing disease through changes in gene regulation.

7.2 *C9ORF72*

One of the intriguing aspects of the *C9ORF72* repeat expansion is its prevalence in sporadic ALS. We pursued two different avenues that would give us insight into this observation and also help us understand their pathogenicity. The first was to determine whether all expansion carriers were descended from a single expansion event. Evidence of a single founder event would indicate that the sporadic ALS carriers were cases of incomplete penetrance in the parents.

We identified a rare variant rs147599399 on the risk haplotype that was present in both some expansion carriers and some non-expansion carriers, showing that the expansion arose at least twice in history. This SNP also influenced disease presentation by extending survival in expansion carriers. This finding shows that the risk haplotype, marked by rs3849942, is one that is permissive to expansion and that both rs147599399 and rs3849942 could be cis-acting modifiers of the expansion.

Further sequencing of the risk haplotype in *C9ORF72* expansion carriers may identify more rare variants that would provide further evidence of multiple expansions. One group used targeted capture to sequence the risk haplotype and did not identify rs147599399 or other variants that would provide evidence against a single founder, however all subjects were of Northern European ancestry.[14] It is possible that expansions in subjects of Northern European ancestry are indeed descended from the same founder and that more heterogeneous populations such as ours are needed to reveal rare haplotype variants. Furthermore, it may be that rs147599399 is not itself cis-acting modifier of the expansion. Sequencing the full risk haplotype may reveal that rs147599399 is in linkage disequilibrium (LD) with the true modifier and may identify additional cis-acting modifiers.

Determining the effects of these SNPs is an important step in comprehending the origin of repeat expansions in *C9ORF72*. Studies have shown that the distance between the repetitive sequence and origin of replication is important factor in instability of trinucleotide repeats[15] and one study showed that a SNP on the predisposing haplotype of the *FMR1* expansion fell within a mapped replication origin altering the timing of DNA replication in that region.[16,17] This could be the mechanism through which rs3849942 (or a different SNP on the risk haplotype) predisposes the *C9ORF72* repeat toward further expansion.

There are a few approaches that use next-generation sequencing to identify replication origins genome-wide. Repli-seq involves labeling newly synthesizes DNA in dividing cells with BrdU, sorting cells by cell-cycle progression and sequencing the BrdU labeled DNA. This provides a map of the replication timing through the cell cycle.[18] However, due to the differences in replication timing between different cell lineages[18] and the fact that we observe contraction of the repeat in the BAC, it is unclear whether observations in cell culture models will be an accurate representation of how the repeat locus behaves in disease. Another technique of observing replication origins uses ChIP-Seq targeting a component of the origin replication complex (ORC1) to identify replication origins.[19] This method was performed using cell lines, however could potentially be used on frozen patient tissues.[20]

Cis-acting modifiers in other repeat expansion disorders include altered methylation status and formation of secondary structures.[21–24] Methylation profiles and G-quadruplex structures, which are present in expanded GGGGCC repeats[25], can both be evaluated using frozen patient tissues.[26] The downstream effects of these changes, such as increased or decreased transcription, alternative splicing, and changes in RAN translation levels of *C9ORF72*, can also be compared by rs147599399 genotype. RNAseq was recently used to show that *C9ORF72* expansion carriers had altered transcription levels and splicing patterns of certain genes compared to non-C9 ALS subjects.[27] It would be interesting to see if transcriptome profiles also differed by r147599399 genotype within expansion carriers.

Another way we explored the repeat expansion was by using Southern blot analysis to ascertain expansion sizes. Instability of the repeat expansion could lead to sporadic disease through anticipation or through somatic instability. Progressive expansion of the repeat through generations, as is characteristic of anticipation, could produce SALS patients with parents whose

expansions do not meet a threshold for pathogenicity. Somatic instability could result in SALS patients with larger expansions in CNS tissues leading to neurodegeneration whereas their unaffected parents would not have larger expansions in their CNS tissues.

Instability of the repeat in control families from the CEPH panel was shown with a change from 21 repeats in paternal grandparent to 22 repeats in the father to 20 repeats in his son. All families with changes in repeat size between generations had more than 10 repeat units and all unstable transmissions were paternal. [28] To our knowledge, we have shown the first evidence of instability in families with full-length expansions. Unstable transmissions came from paternal and maternal lines. This instability was not the typical pattern that is seen in anticipation since the repeat was not always expanding between generations, however there are other repeat expansion disorders with anticipation that have both expansions and contractions as well.[29,30]

Somatic instability of the repeat has been shown in multiple reports. All of these studies consistently show that expansions in the blood are smaller than most brain regions except for the cerebellum.[28,31–33] We observed the same somatic instability, but did not identify any correlations between expansion size and clinical features in any of the observed tissues. Other studies have produced mixed results regarding the effect of expansion size in specific tissues on age of onset and survival. Beck et al. found expansion sizes in blood to be positively correlated with age of onset in a variety of neurodegenerative diseases including ALS.[28] Van Blitterswijk et al. found no correlations between age of onset or survival in ALS patients, but did find a two associations in patients with frontotemporal dementia: a positive correlation between expansion size in the frontal cortex and age of onset in and an association between larger expansions in the cerebellum with reduced survival.[31] The Nordin et al. study included the largest number of tissues and only found correlations with expansion sizes in the cerebellum and parietal lobe.

Smaller expansions in the cerebellum and parietal lobe were associated with earlier age of onset in ALS while larger expansions in the parietal lobe were associated with shorter survival in ALS patients and the overall cohort.[32] Additional studies will be required to clarify these findings.

We observed no difference in expansions in blood-derived DNA of affected and unaffected carriers, although it is possible that affected carriers have larger expansions in CNS tissues or more specifically in motor neurons compared with unaffected carriers. These data suggest that expansion size is not the only factor in pathogenicity. There is evidence that environmental stress can induce trinucleotide repeat instability in cells[34] so it is possible the cellular stress in motor neurons could induce further expansion and result in degeneration of those neurons. There could also be other cis- or trans-acting genetic factors that interact with expansions and reduce their penetrance. For example a common variant in *TMEM106B* protects against fronto-temporal dementia in expansion carriers[35] and intermediate expansions in *ATXN2* predispose towards motor neuron degeneration.[36]

One aspect of the expansions that we were unable to study was whether the repeat sequence is pure or contains interruptions of non GGGGCC units. Interruptions in the repeat sequence of other expansion disorders reportedly stabilize expansions and prevent them from continuing to expand by preventing slippage of DNA replication machinery during cell division.[37–39] Due to the length and GC content of the repeats, they cannot be sequenced by conventional methods. Real-time long-read sequencing was recently used to sequence through expanded CGG repeats in the *FMR1* gene showing the presence of AGG interruptions in the sequence and simultaneously determining the size of the expansions.[40] Preliminary studies sequencing the BAC on the MinIon long-read sequencing platform were able to generate a size distribution for expansion length (Figure 7.1), but base-calling is not yet accurate enough to

detect sequence interruptions. This method requires the preparation of sequencing libraries containing the repeat, so while it can be used to sequence the BAC, we are working on developing methods to isolate expanded *C9ORF72* alleles from patients genomic DNA for sequencing.

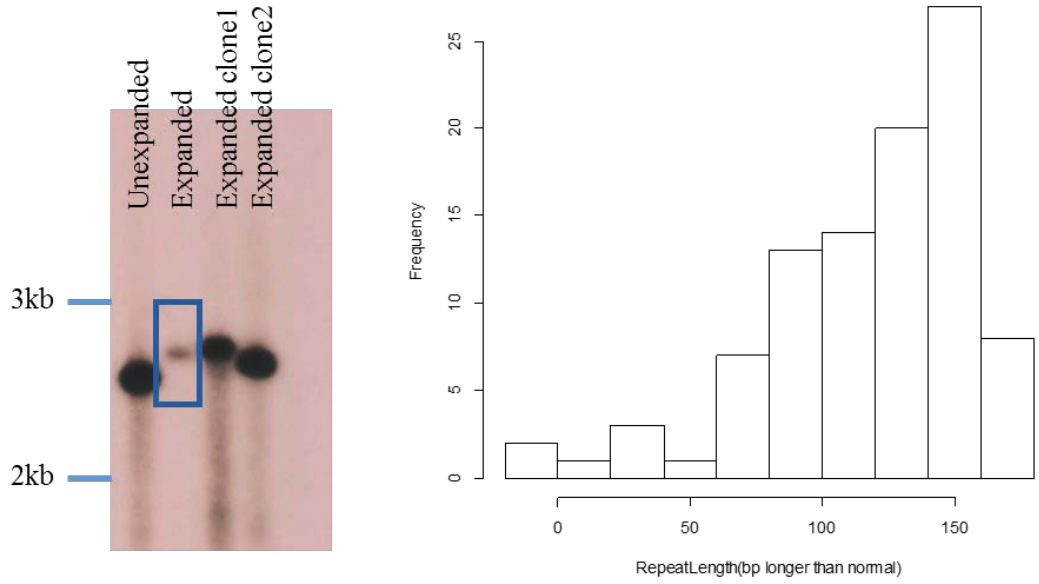

Figure 7.1 Preliminary sequnecing of a *C9ORF72* BAC on the MinIon sequencing platform. A. Southern blot showing the size of the expansion in the BAC compared to a BAC with an unexpanded allele (~30 repeats in the expanded BAC). B. Histogram of repeat lengths from a MinIon sequencing run. The peak expansion size was 150bp longer than wild-type, corresponding to ~28 repeats.

7.3 Neuroinflammation

In order to find risk factors that are associated with sporadic ALS, we selected *TREM2* as a candidate gene. *TREM2* encodes the triggering receptor expressed on myeloid cells 2 and is involved in activation of microglia.[41] The rare p.R47H SNP was shown to be associated with

Alzheimer's disease which also presents with microglial activation,[42–46] therefore we genotyped this SNP in our cohort.

We showed that this same SNP was significantly associated with disease in our SALS cohort. TREM2 expression was increased in spinal cords from ALS patients and $SOD1^{G93A}$ mice. It is unclear whether this is a specific effect of TREM2 or a reflection of the recruitment of microglia to the site of motor neuron injury. We were unable to determine if p.R47H had any effect on expression levels. We hypothesized that p.R47H was a loss of function mutation that resulted in an inability of microglia to remove cellular debris. A recent study of *TREM2* in AD showed that the SNP did not alter the expression or signaling ability of TREM2, but reduced its ability to bind ligands including phospholipids.[47]

In order to determine if there were other mutations in the TREM family that were associated with ALS, we looked for variants in *TREM2, TREM1, TREML1, TREML2, TREML4*, and the binding partner of TREM2, *TYROBP*. There were no additional variants in *TREM2* or any of the other genes that were significantly associated with ALS. Variants in *TREML4* approached significance, mainly due to the SNP rs112680060. TREML4 is also involved in recognition of apoptotic cells and associates with TYROBP[48] and while it is not normally expressed in microglia, it was upregulated in response to neuronal damage.[49] A role for *TREML4* in ALS has not been reported, so it will be important to see if TREML4 is expressed in spinal cords of ALS patients and if the rs112680060 SNP has any effect on its expression. If *TREML4* acts similarly to *TREM2* this SNP could alter the ability of *TREML4* to recognize degenerating neurons for clearance.

Current studies in our lab are attempting to clarify the role of *TREM2* in ALS. We are measuring the expression levels of TREM2 mRNA levels in additional spinal cords to confirm

our initial observations. In addition we are measuring TREM2 in blood and CSF from patients to determine its viability as an ALS biomarkers. Mouse models exploring the effects of *TREM2* genetic knock-out in the $SOD1^{G93A}$ are underway to explore the role of *TREM2* in motor neuron degeneration. Additionally anti-sense oligos are being generated to knock-down TREM2 expression in symptomatic $SOD1^{G93A}$ mice to study the specific effects of *TREM2* in disease progression.

7.4 References

1. Kenna, K. P. et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J. Med. Genet. 50, 776–783 (2013).

2. Liu, Z.-J. et al. Identify mutation in amyotrophic lateral sclerosis cases using HaloPlex target enrichment system. Neurobiol. Aging 35, 2881.e11–15 (2014).

3. MacArthur, D. G. et al. Guidelines for investigating causality of sequence variants in human disease. Nature 508, 469–476 (2014).

4. 1000 Genomes Project Consortium et al. An integrated map of genetic variation from 1,092 human genomes. Nature 491, 56–65 (2012).

5. Tennessen, J. A. et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337, 64–69 (2012).

6. Bustamante, C. D. et al. Natural selection on protein-coding genes in the human genome. Nature 437, 1153–1157 (2005).

7. Petrovski, S., Wang, Q., Heinzen, E. L., Allen, A. S. & Goldstein, D. B. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. PLoS Genetics 9, e1003709 (2013).

8. Subramanian, S. & Kumar, S. Evolutionary anatomies of positions and types of disease-associated and neutral amino acid mutations in the human genome. BMC Genomics 7, 306 (2006).

9. Leong, I. U. S., Stuckey, A., Lai, D., Skinner, J. R. & Love, D. R. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. BMC Med. Genet. 16, 34 (2015).

10. Xie, F. et al. Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac. Genome Research 24, 1526–1533 (2014).

11. Canver, M. C. et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. 289, 21312–21324 (2014).

12. Cirulli, E. T. et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science 347, 1436–1441 (2015).

13. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).

14. Smith, B. N. et al. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. Eur. J. Hum. Genet. 21, 102–108 (2013).

15. Cleary, J. D., Nichol, K., Wang, Y.-H. & Pearson, C. E. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nature Genetics 31, 37–46 (2002).

16. Gerhardt, J. et al. The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. Mol. Cell 53, 19–31 (2014).

17. Gerhardt, J. et al. Cis-acting DNA sequence at a replication origin promotes repeat expansion to fragile X full mutation. J. Cell Biol. 206, 599–607 (2014).

18. Hansen, R. S. et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proceedings of the National Academy of Sciences 107, 139–144 (2010).

19. Dellino, G. I. et al. Genome-wide mapping of human DNA-replication origins: Levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Research 23, 1–11 (2013).

20. Savic, D., Gertz, J., Jain, P., Cooper, G. M. & Myers, R. M. Mapping genome-wide transcription factor binding sites in frozen tissues. Epigenetics & Chromatin 6, 30 (2013).

21. Moore, H., Greenwell, P. W., Liu, C. P., Arnheim, N. & Petes, T. D. Triplet repeats form secondary structures that escape DNA repair in yeast. Proc. Natl. Acad. Sci. U.S.A. 96, 1504–1509 (1999).

22. Libby, R. T. et al. Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. Hum. Mol. Genet. 12, 41–50 (2003).

23. Libby, R. T. et al. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. PLoS Genet. 4, e1000257 (2008).

24. Nichol Edamura, K., Leonard, M. R. & Pearson, C. E. Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. Am. J. Hum. Genet. 76, 302–311 (2005).

25. Šket, P. et al. Characterization of DNA G-quadruplex species forming from C9ORF72 G4C2-expanded repeats associated with amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Neurobiol. Aging 36, 1091–1096 (2015).

26. Lam, E. Y. N., Beraldi, D., Tannahill, D. & Balasubramanian, S. G-quadruplex structures are stable and detectable in human genomic DNA. Nat Commun 4, 1796 (2013).

27. Prudencio, M. et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nature Neuroscience 18, 1175–1182 (2015).

28. Beck, J. et al. Large C9orf72 Hexanucleotide Repeat Expansions Are Seen in Multiple Neurodegenerative Syndromes and Are More Frequent Than Expected in the UK Population. The American Journal of Human Genetics 92, 345–353 (2013).

29. Martorell, L., Monckton, D. G., Sanchez, A., Lopez De Munain, A. & Baiget, M. Frequency and stability of the myotonic dystrophy type 1 premutation. Neurology 56, 328–335 (2001).

30. Ashizawa, T. et al. Characteristics of intergenerational contractions of the CTG repeat in myotonic dystrophy. Am. J. Hum. Genet. 54, 414–423 (1994).

31. Van Blitterswijk, M. et al. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. The Lancet Neurology 12, 978–988 (2013).

32. Nordin, A. et al. Extensive size variability of the GGGGCC-expansion in C9orf72 in both neuronal and non-neuronal tissue in 18 patients with ALS or FTD. Human Molecular Genetics (2015). doi:10.1093/hmg/ddv064

33. Fratta, P. et al. Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion. Neurobiology of Aging 36, 546.e1–546.e7 (2015).

34. Chatterjee, N., Lin, Y., Santillan, B. A., Yotnda, P. & Wilson, J. H. Environmental stress induces trinucleotide repeat mutagenesis in human cells. Proc. Natl. Acad. Sci. U.S.A. 112, 3764–3769 (2015).

35. Van Blitterswijk, M. et al. TMEM106B protects C9ORF72 expansion carriers against frontotemporal dementia. Acta Neuropathol. 127, 397–406 (2014).

36. Van Blitterswijk, M. et al. Ataxin-2 as potential disease modifier in C9ORF72 expansion carriers. Neurobiology of Aging (2014). doi:10.1016/j.neurobiolaging.2014.04.016

37. Pearson, C. E. et al. Interruptions in the triplet repeats of SCA1 and FRAXA reduce the propensity and complexity of slipped strand DNA (S-DNA) formation. Biochemistry 37, 2701–2708 (1998).

38. Gao, R. et al. Instability of expanded CAG/CAA repeats in spinocerebellar ataxia type 17. Eur. J. Hum. Genet. 16, 215–222 (2008).

39. Choudhry, S., Mukerji, M., Srivastava, A. K., Jain, S. & Brahmachari, S. K. CAG repeat instability at SCA2 locus: anchoring CAA interruptions and linked single nucleotide polymorphisms. Hum. Mol. Genet. 10, 2437–2446 (2001).

40. Loomis, E. W. et al. Sequencing the unsequenceable: expanded CGG-repeat alleles of the fragile X gene. Genome Res. 23, 121–128 (2013).

41. Kierdorf, K. & Prinz, M. Factors regulating microglia activation. Front Cell Neurosci 7, 44 (2013).

42. Benitez, B. A. et al. TREM2 is associated with the risk of Alzheimer's disease in Spanish population. Neurobiol. Aging 34, 1711.e15–17 (2013).

43. Guerreiro, R. et al. TREM2 variants in Alzheimer's disease. N. Engl. J. Med. 368, 117–127 (2013).

44. Gonzalez Murcia, J. D. et al. Assessment of TREM2 rs75932628 association with Alzheimer's disease in a population-based sample: the Cache County Study. Neurobiol. Aging (2013). doi:10.1016/j.neurobiolaging.2013.06.004

45. Jonsson, T. et al. Variant of TREM2 associated with the risk of Alzheimer's disease. N. Engl. J. Med. 368, 107–116 (2013).

46. Pottier, C. et al. TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. J. Alzheimers Dis. 35, 45–49 (2013).

47. Wang, Y. et al. TREM2 Lipid Sensing Sustains the Microglial Response in an Alzheimer's Disease Model. Cell 160, 1061–1071 (2015).

48. Hemmi, H. et al. A New Triggering Receptor Expressed on Myeloid Cells (Trem) Family Member, Trem-Like 4, Binds to Dead Cells and Is a DNAX Activation Protein 12-Linked Marker for Subsets of Mouse Macrophages and Dendritic Cells. J Immunol 182, 1278–1286 (2009).

49. Thomas, D. M., Francescutti-Verbeem, D. M. & Kuhn, D. M. Gene expression profile of activated microglia under conditions associated with dopamine neuronal damage. FASEB J. 20, 515–517 (2006).