The genetic landscape of amyotrophic lateral sclerosis

Sarah Louise Morgan

July 2016

UCL, Institute of Neurology, PhD in Neurogenetics

Supervisors Professor John Hardy and Dr Alan Pittman

DECLARATION

I, Sarah Morgan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ACKNOWLEDGEMENTS

I wish to dedicate this thesis to my mother, an inspirational woman, who has ascended high enough in her field to be worth tweeting about from space. I hope to be able to follow in her footsteps. I would like to thank both my parents for their constant motivation and deft guidance.

I would like to express my sincere gratitude to my supervisors, Professor John Hardy and Dr Alan Pittman, for guiding me through a most rewarding project. John, also for being the backbone of the department and always expressing his moral character without losing his humour. Alan, for answering all my questions and supporting me constantly, day-to-day.

I would like to thank my fellow co-workers for the stimulating discussions and friendship especially Dr Phillip Smethurst, Michael Chou, Dr Maryam Shaoi, Claire Hall, Aarti Singh, Michael Thor, Deborah Hughes, Dr Niccolo Mencacci, Sara Bandrés Ciga, Demis Kia, Heather Ging and Léa R'Bibo.

I am grateful to the three clinicians at the National who were crucial to the assembly of the clinical material: Drs Richard Orrell, Pietro Fratta and Katie Sidle, and to Professor Ammar Al-Chalabi and his team, our collaborators at King's College.

This project would not have been possible without the generous donations from patients and without the work of the MND association to whom I am greatly indebted.

Lastly but not least, I must thank my partner, Dr Parashkev Nachev, who has stimulated my development into an ever evolving person. Thank you for all the times you have leapt to assist me and for inspiring me in a way I never knew was possible.

ABSTRACT

Next-generation sequencing (NGS) technologies have a vast number of advantages that have caused a growth in their application for uncovering the genetics of complex diseases. Amyotrophic lateral sclerosis (ALS) is one such disease that could benefit from this technique. As a rapid-onset disease, the time to diagnosis must match this speed if we want to increase our chances of finding a treatment drug that works. In a number of ALS cases, the diagnosis can be aided by genetics. However, we currently do not understand the full genetic background of ALS and so to address this issue, I have designed a screening panel to sequence 25 ALS-associated genes in 1,235 patients. This data was compared against 613 controls to perform a case-control analysis. Alongside mutation burden tests and tests for an oligogenic basis, I have additionally created a novel method, a pipeline assisted by machine learning, for uncovering high-dimensional genetic patterns that predispose an individual to ALS.

The results indicate that there is an increase burden of rare variants in the UTRs of the genes *SOD1*, *TARDBP*, *FUS*, *VCP*, *OPTN* and *UBQLN2* collectively. Additionally, we discovered an increased number of patients with two mutations in different ALS genes than would be expected by chance alone. Encompassed in these results is the finding of a novel ALS gene, *MATR3*, which we aided the first publication of. We have also screened *CHCHD10* in ALS and frontotemporal dementia (FTD) finding confirmations of previously published mutations plus additional novel variants. A selection of 26 Argentinian ALS samples were included in the study which reveal 27 known and novel mutations across 17 patients. Lastly, machine learning methods are able to perform better than chance at predicting patients on the basis of their genetics.

In conclusion, many cases of ALS, sporadic included, show a complex genetic interplay which, combined with the overall mutation burden, determine the risk and course of ALS.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

1000g – 1000 genomes project

ACMG – American college of medical genetics and genomics

AD – Alzheimer's disease

CDCV – Common disease-common variant

CDRV – Common disease-rare variant

CG69 – Complete genomics

CMD – Congenital muscular dystrophy

CMT – Charcot-Marie-tooth disease

dbSNP – Single nucleotide polymorphism database

ddNTP – Dideoxynucleotides

DNA – Deoxyribonucleic acid

dNTP – Deoxynucleotides

DPR – Dipeptide repeat

DSMO – Dimethyl sulfoxide

ESCRT – Endosomal sorting complex required for transport

EVS6500 – Exome variant server

ExAC – Exome aggregation consortium

fALS – Familial ALS

FTD – Frontotemporal dementia

GATK – Genome analysis toolkit

GoF – Gain of function

GWAS – Genome-wide association study

HGMD – Human Gene Mutation Database

HSP – Hereditary spastic paraplegia

IBM – Inclusion body myositis

IBMPFD – Inclusion body myopathy with Paget disease and frontotemporal dementia

LMN – Lower motor neuron

LoF – Loss of function

MND – Motor neuron disease

mRNA – Messenger RNA

NF-κB – Nuclear factor kappa-light-chain-enhancer of activated B cells

NGS – Next generation sequencing

NIH – National Institutes of Health

OMIM – Online Mendelian inheritance in man

PBP – progressive bulbar palsy

PCA – Principle components analysis

PCR – Polymerase chain reaction

PD – Parkinson's disease

PDB – Paget disease of bone

PLS – Primary lateral sclerosis

PMA – Progressive muscular atrophy

RNA – Ribonucleic acid

sALS – Sporadic ALS

SBS – Sequencing by synthesis

SCA – Spinocerebellar ataxia

SKAT – SNP-set (sequence) kernel association test

SMA – Spinal muscular atrophy

SNP – Single nucleotide polymorphism

SVM – Support vector machine

TBE – Tris-borate-ethylenediaminetetraacetic acid

TSCA – TruSeq custom amplicon

t-SNE – t-distributed stochastic neighbour embedding

UCSC – University of California, Santa Cruz

UMN – Upper motor neuron

UPS – Ubiquitin proteasome system

VCPDM – Vocal cord pharyngeal distal myopathy

VQSR – Variant quality score recalibration

WES – Whole-exome sequencing

GENE ABBREVIATIONS

ANG – Angiogenin

APEX – Apyrimidinic endodeoxyribonuclease 1

APOE4 – Apolipoprotein E

ATXN – Ataxin

C9orf72 – Chromosome 9 open reading frame 72

CHCHD10 – Coiled-coil-helix-coiled-coil-helix domain containing 10

CHMP2B – Charged multivesicular body protein 2B

DAO – D-amino-acid oxidase

DCTN1 – Dynactin 1

FIG4 – Factor-induced 4

FUS – Fused in sarcoma

LMNB1 – Lamin B1

LRRK2 – Leucine-Rich Repeat Kinase 2

MATR3 – Matrin 3

OPTN – Optineurin

PFN1 – Profilin 1

PON – Paraoxonase

PRPH – Peripherin

SETX – Senataxin

SOD1 – Superoxide dismutase 1

SPG11 – Spatacsin 11

SQSTM1 – Sequestosome 1

TARDBP/ TDP43 – Transactive response DNA binding protein 43

TLS – Translocated in sarcoma

TREM2 – Triggering receptor expressed on myeloid cells 2

UBQLN2 – Ubiquilin 2

VAPB – Vesicle-associated membrane protein-associated protein B/C

VEGF – Vascular endothelial growth factor

CHAPTER 1. INTRODUCTION

1.1. BACKGROUND

1.1.1. MOTOR NEURON DISEASE

The complexity of the molecular mechanisms implicated in neurodegenerative diseases like amyotrophic lateral sclerosis (ALS) is equally paralleled by multifaceted genetics which still, in today's era, are not fully understood. ALS is a severe, debilitating disease with an average life expectancy of three years from diagnosis. Patients nominally present with progressive voluntary muscle weakness affecting their ability to walk, then to talk, until eventually reducing the patient's ability to breathe. Approximately 10% of cases are familial (fALS) which generally present at a younger age compared to sporadic (sALS) cases. Frontotemporal dementia (FTD) often accompanies ALS with 15% of patients meeting the FTD diagnostic criteria and as many as 50% displaying mild cognitive dysfunction (Ferrari *et al.*, 2011; Abrahams *et al.*, 2014). The incidence of ALS in the general European population is estimated at 2 per 100,000 person-years with an overrepresentation of males (Logroscino *et al.*, 2010). The most common ages of onset lie between 40 and 70, averaging at 55 years old. After 80 years of age, new cases become extremely rare.

Of course, ALS is just one form of motor neuron disease (MND) and indeed the most common, but there are several other diseases within this spectrum which have a great similarity to ALS either in clinical features or pathological findings, or even both. The main motor neuron diseases include: ALS, primary lateral sclerosis (PLS), progressive muscular atrophy (PMA) and progressive bulbar palsy (PBP). ALS commonly begins with a limb site of onset but can also present with speech problems, termed bulbar onset, in a quarter of patients. The underlying aetiology involves degeneration of the upper and lower motor neurons (UMN and LMN). PLS has an isolated UMN phenotype, while PMA displays pure LMN symptoms and PBP is strictly bulbar. The motor cortex, brain stem and spinal cord are all affected in ALS and as the motor neurons degenerate in these areas, muscle atrophy and spasticity occurs. Pathologically brisk reflexes are a clear sign of UMN involvement while fasciculations (spontaneous muscle contractions) are due to the LMNs (Kiernan *et al.*, 2011).

There is no single diagnostic test for ALS as a whole, and consequently clinicians use the El Escorial criteria to categorise patients and require evidence of UMN and LMN manifestations

in the limbs as well as the exclusion of other disorders known to produce similar symptoms (Wijesekera and Leigh, 2009). Post-mortem examination reveals the presentation of ubiquitinated neuronal inclusion bodies which mostly reside in the motor neurons recognised as vulnerable in ALS (Ince *et al.*, 1998). The role of these aggregates in the disease is still unsolved, we do not know if they are toxic, protective or a by-product of other disease mechanisms. This observation is not specific to ALS and protein misfolding is observed in many neurodegenerative disorders like Alzheimer's and Parkinson's disease (Soto, 2003).

1.1.2. GENETICS

1.1.2.1. OVERVIEW

Genetics is the study of heredity and deals with the variation observed in organisms as a consequence of their DNA. Many human disorders have had genetics implicated in their cause ranging from the classic cystic fibrosis to conditions traditionally considered as solely environmental e.g. bulimia and obesity (Müller *et al.*, 2012). With the invention of Sanger sequencing in 1977, it became feasible to explore the DNA and the traits hidden within it.

Variation of the DNA within genes is a normal process with the average individual containing ten million variants compared to our standard reference genome which is three billion base pairs long. However, this reference is composed of only 13 individuals which results in some mutations having a frequency of 90% in the population. Additionally, the word *mutation* bears with it a connotation that the change is damaging. *Polymorphism* on the other hand implies benign variation, *poly* describing multiple occurrences of it across the population. Nevertheless, common variation is involved in human disease, for example in schizophrenia and *APOE* in Alzheimer's (Purcell *et al.*, 2009; Guerreiro *et al.*, 2012). In this thesis, I use the term *mutation* to describe all variation, be it causal or not. I will also put all gene abbreviations in italics when referring to the gene itself, and not when speaking of the protein.

The classic type of inheritance, the passing of genetic traits from one generation to the next, is based on Mendel's laws which are a set of rules founded on studies in peas during the 1800s. In humans, Mendelian diseases are caused by dominant or recessive, highly penetrant mutations which are present in multiple patients within the same family.

Mutations which follow a Mendelian pattern of inheritance are the *low-hanging fruit* in terms of finding genetic causes, for the affected family members will all have the causal variant while the unaffected will not. In such cases, linkage analysis can be performed on families with dominant or recessive inheritance to uncover the causal mutation (Figure 1A). However there are a number of methods which can be employed in the search for causal genes and the mutations within them (Figure 1B-F).

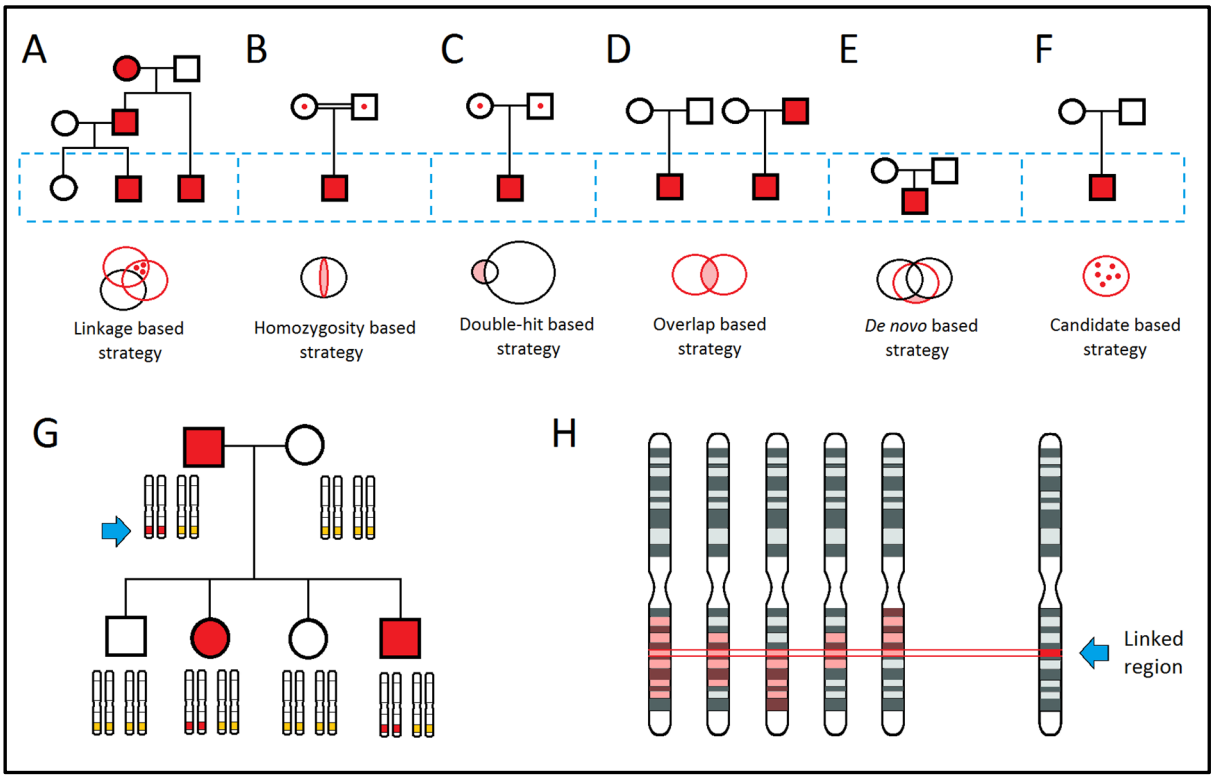

Figure 1. (A-F) Different strategies used for finding disease-causing genes; subjects within the dashed blue box would be sequenced; red subjects = affected. Redrawn from Gilissen *et al*., (2012). (A) Linkage analysis can be performed on a family with both affected and unaffected individuals. (B) When the parents are consanguineous, the causal mutant is often a rare homozygous variant which narrows down the search field for researchers. (C) If a recessive disorder is suspected, one technique is to look for two different mutations within the same gene producing a compound heterozygous effect. (D) Examining patients from different families can reveal the same gene being mutated in both. (E) Sequencing trios (mother, father and child) can reveal *de novo* variants which may be the cause of the disease. (F) Lastly, it is possible to only examine candidate genes which are suspected to be involved in the disease. (G) Family tree with affected (red) and unaffected individuals exhibiting dominant inheritance. Arrow points to the haplotype in red containing the causative mutation. Linkage analysis can be performed to identify this area. (H) The five chromosomes one the left are each from a different family and show the linked region (shaded red). These can be combined to narrow the search window (blue arrow) to find the causative mutation. Drawn in Paint v1511.

Linkage analysis can be performed by assaying a selection of SNPs from across the whole genome to find regions of shared DNA (haplotypes) which segregate with the disease (Figure 1G-H). This method exploits the fact that genomic regions in close proximity to each other are more likely to be inherited together. Therefore, non-pathogenic SNPs located near a disease-causing variant are able to help in pinpointing the mutation we are looking for. However, *linked* areas can contain hundreds of genes, the sequencing of which would be extremely laborious with traditional Sanger sequencing methods.

In a similar vein, homozygosity mapping can be used to look for recessive disorders across different families (Figure 1B). This method exploits the fact that inbred populations with an arising recessive disease often have genomic regions shared amongst the affected individuals which will display homozygous markers located near the disease-causing gene. This enables an economical screening of the genome in a small number of patients to reveal a region associated with a disease which can then be sequenced fully.

For the strategies in Figure 1C-E, sequencing larger areas of the genome may be required to achieve a high confidence in the results. This can be ascertained via exome sequencing (described subsequently). Lastly, the least profitable of all the tactics is to only examine candidate genes which are suspected to be involved in the disease (Figure 1F). This relises heavily on our current knowledge of which genes are related in their function and all the functions each gene performs.

Unfortunately, in rapidly fatal diseases like ALS it can be very problematic to find large families displaying Mendelian inheritance with DNA available from multiple affected and unaffected members. To complicate matters, many cases are not as simple as this and factors such as reduced penetrance, gene interaction, epigenetics and environmental effects may augment the puzzle. Genetically complex diseases have a range of mutation types defined by effect size and frequency in the population (Figure 2A).

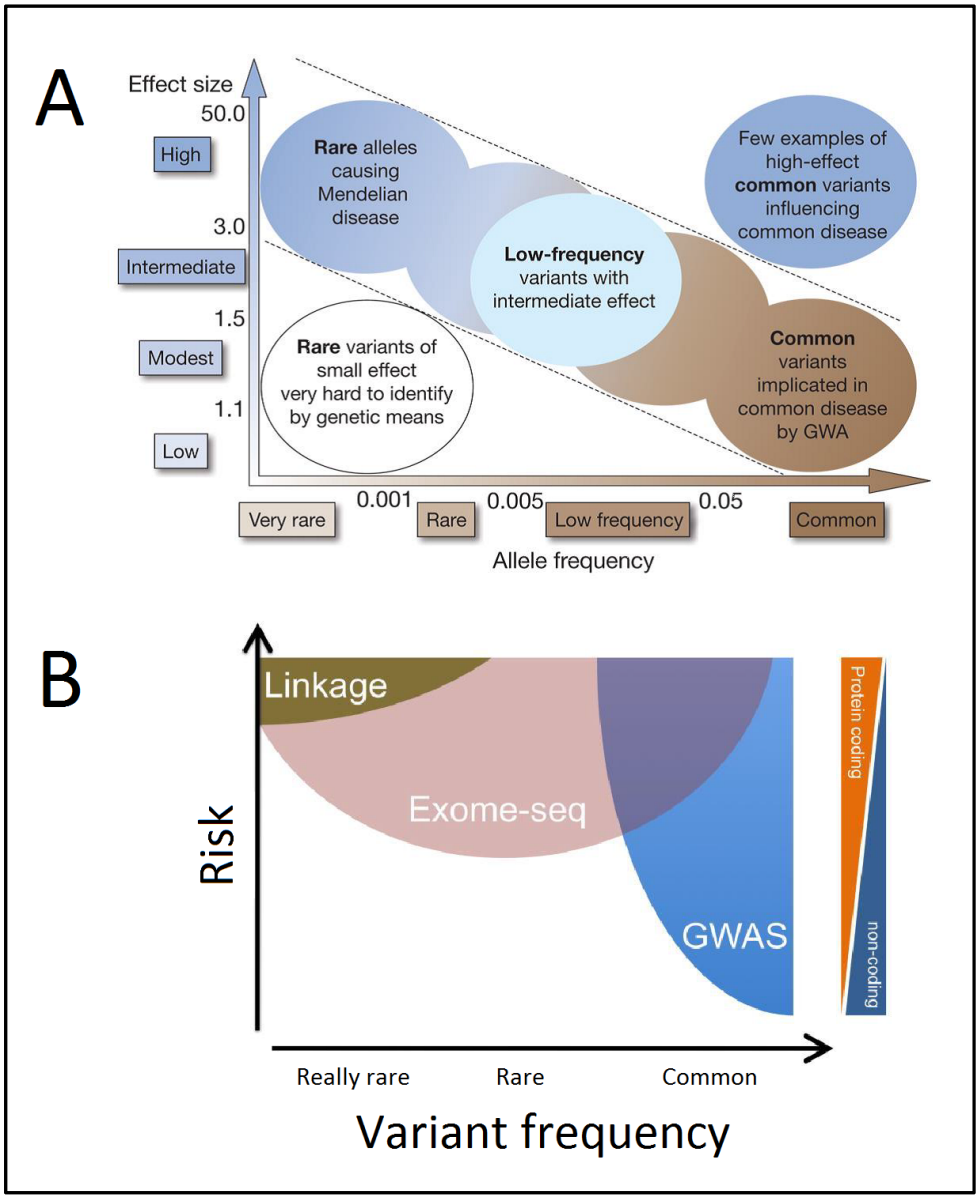

Figure 2. (A) Graph showing the predicted different types of mutations associated with disease based on the variant effect size and frequency in the general population, defined as the genetic architecture of disease (Manolio et al., 2009). (B) Different genetic techniques and predictions on the types of mutations each captures (Singleton *et al.*, 2010).

Different gene discovery strategies have been developed to target these variant types (Figure 2B). This includes genome-wide association studies (GWAS) where subjects are genotyped by microarrays for common variation in the genome. The power of this study is accomplished by using thousands of individuals, numbers that are achievable due to the relatively cheap cost: a test presently in May 2016 costs £35 per patient. This unbiased technique looks for association of common variation with disease or an observable trait, even when variants have a relatively minor effect, first proposed by Risch and Merikangas (1996). This falls within the overlap-based strategy (Figure 1D).

Lastly, exome sequencing, also known as targeted exome capture, is a relatively recent method that utilises next-generation sequencing, first described by Hodges *et al.*, in 2007. Whole-exome sequencing (WES) is intended to cover all coding regions of the genome which accounts for about 1-2% of all DNA and contains 30 million bases (~£200 per patient). Whole-genome sequencing (WGS) is a similar technique which attempts to read all 3 billion bases making it substantially more expensive (~£1000 per patient). Since predictions calculate that disease-causing mutations are 85% coding, most studies opt for WES over WGS (Singleton *et al.*, 2010). Although the price of sequencing has dropped significantly in the past few years (Figure 3), it may occasionally be preferable to design a targeted capture to only include the genes of interest for a particular study and in a clinical setting this would be desirable in order to save money. This technology is discussed further in Chapter 2.

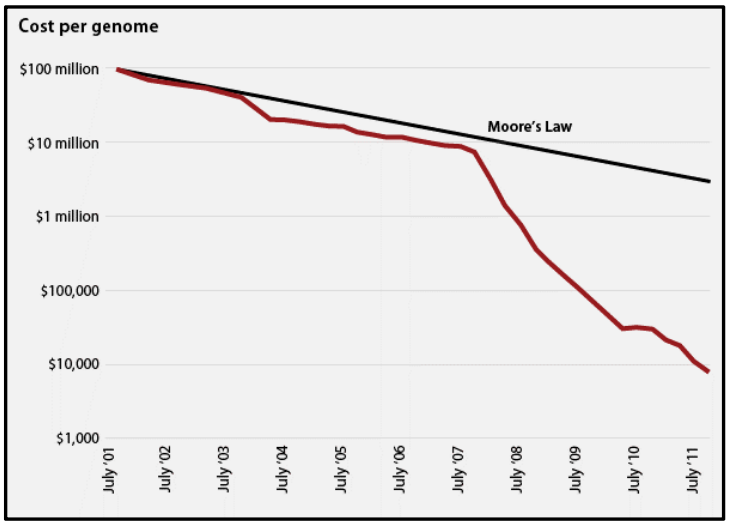

Figure 3. The cost of sequencing a genome since 2001. With the introduction of exome sequencing in 2007, this cost has decreased rapidly (Alterovitz, 2014).

1.2. GENETICS OF ALS

At present, over 100 genes have been associated with ALS to varying degrees of significance. The major players are *SOD1*, *TARDBP*, *C9orf72* and *FUS* (Figure 4). Here, the involvement of these genes in ALS is common, certain and undisputed. However, it is the dubious genes which would be of interest to fully explore with the aim of gaining a full understanding of the disease.

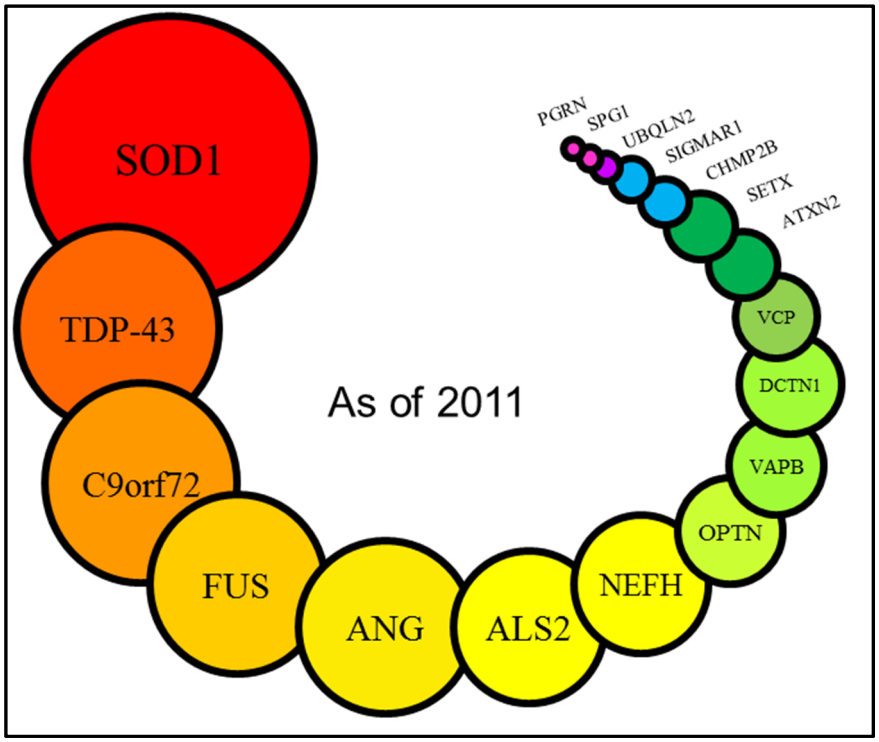

Figure 4. List of genes implicated in ALS with circle size representing relative contribution to disease. This was based on our understanding as of 2011. By the end of this year, however, *C9orf72* had already surpassed *SOD1* as the top ALS gene. I aim to update this diagram by the end of this thesis (redrawn from Jones, 2011).

Understanding ALS genetics helps inform us about the underlying processes that are disrupted in the disease in the hope of designing treatments which alleviate symptoms or prolong life. Examining these genes and mechanisms creates a highly complex picture which needs to be collectively examined if we are to solve the puzzle of ALS (Figure 5).

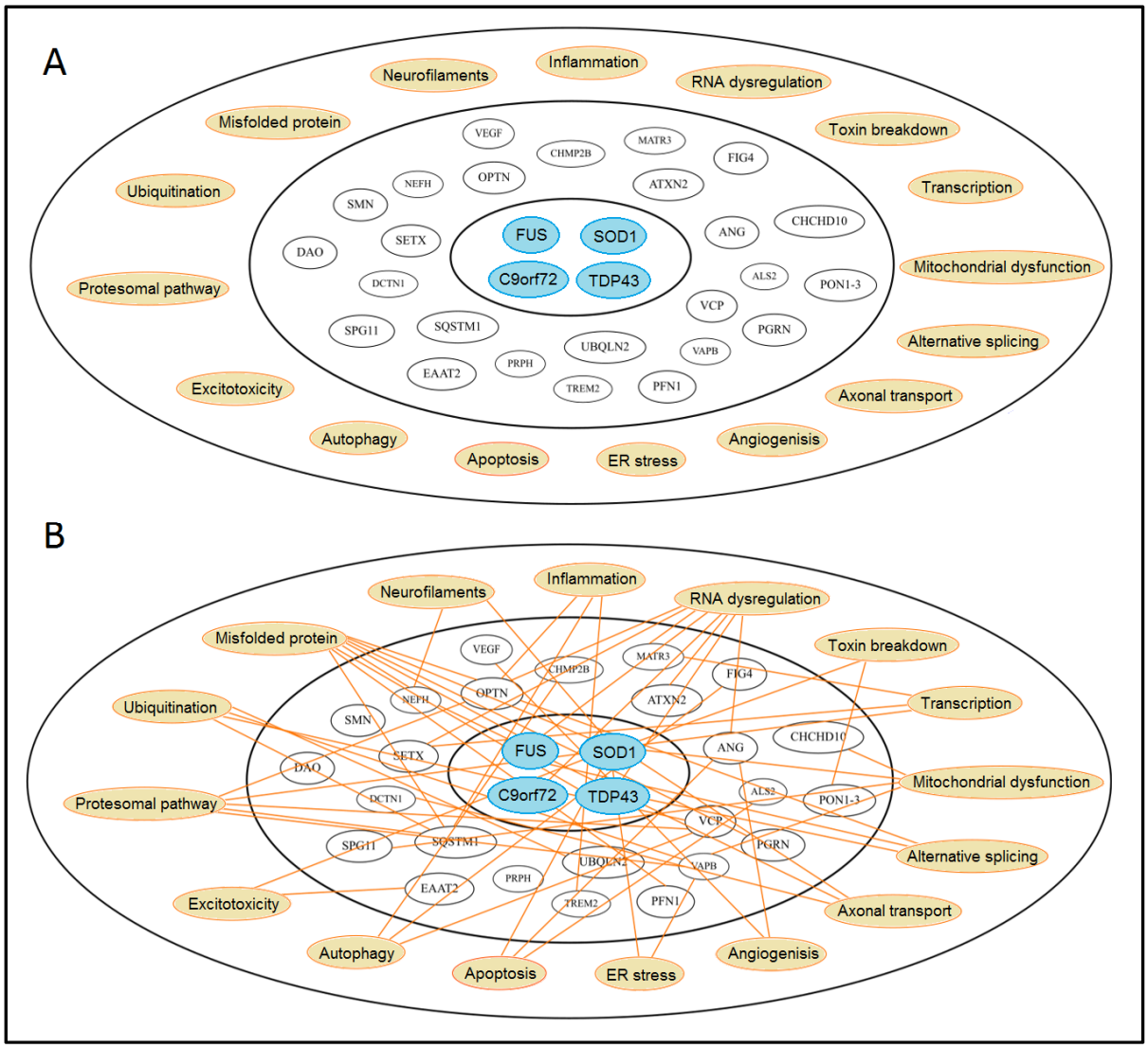

Figure 5. (A) Diagram to show the main ALS genes in blue (inner circle), other associated genes in black (middle circle) and the molecular mechanisms these genes incriminate in yellow (outer circle). (B) The same diagram with lines connecting the genes to the implicated underlying mechanisms. Drawn in Inkscape v0.91.

Previous work looking to divide familial patients up by disease classification (definite, probable and possible fALS) resulted in significantly different rates of mutation identification in each group: 62%, 42% and 11% respectively (Conte *et al.*, 2012). *Definite* fALS was determined if two or more family members had ALS or FTD while *probable* described a single first- or second-degree affected relative and lastly *possible* was defined as having a distant relative with the disease.

The next chapters are ordered by the knowledge as of 2011 (Figure 4). One aim of this thesis is to update this order and accurately describe each gene's relative contribution.

1.2.1. *SOD1*

The first ALS gene to be discovered was superoxide dismutase 1 (*SOD1*) and for a long time, this gene was considered to be the most common genetic cause of familial (12-23%) and sporadic (1-7%) cases. At a time when some clinicians still believed there was no genetic element to ALS, studies in families with multiple affected individuals displayed strong linkage to chromosome 21 (Siddique *et al.*, 1991) which eventually revealed mutations in *SOD1* (Rosen *et al.*, 1993). This gene is implicated in the destruction of free superoxide radicals giving rise to an oxidative stress hypothesis in ALS. Nonetheless, *SOD1* is suspected to be involved in a variety of processes and the mutant SOD1 protein is observed to alter gene expression, cause mitochondrial malfunction, produce cytoskeletal abnormalities and induce dysfunctional axonal transport (Tu *et al.*, 1996; Vos *et al.*, 2007; Maximino *et al.*, 2014; Tafuri *et al.*, 2015).

More than 170 different *SOD1* mutations have been uncovered with the majority of these almost certainly causing the disease in the patients harbouring them (Figure 6). Mutations occur in all five *SOD1* exons with most clustering in exon 4 and 5. Nearly all of these are dominantly inherited except for a few like D90A and D96N which have been shown to cause ALS in a homozygous or compound heterozygous state (Hand *et al.*, 2001). A single amino acid substitution represents 96% of the causal *SOD1* alterations with the rest being due to a small number of insertions and deletions.

D90A most common variant in Europe while A4V is the most frequent in North America (Andersen *et al.*, 2003). The former is largely associated with a slow disease course while the latter often involves an aggressive phenotype. However, Andersen *et al.* (2003) suggested that the atypical clinical outcomes in the American D90A cases were potentially due to modifiers, like other mutations or environmental factors. They also proposed that the mutant SOD1 protein is required for its toxicity given that null mutants have not been identified. Whether this gene is indeed a gain-of-function as this evidence implies or a loss-of-function is still contested and may, in fact, be both (Sau *et al.*, 2007).

The phenotype produced by *SOD1* mutations varies greatly, for example, N86S causes a juvenile onset (Hayward *et al.*, 1998), then L126S and F20C exhibit a low penetrance which is restricted to females (Murakami *et al.*, 2001; Kim *et al.*, 2007), while L117V and D101N have

a slow and rapid progression respectively (Synofzik *et al.*, 2012; Ayers *et al.*, 2014) and lastly I113T has extreme phenotypic variability amongst those who carry it (Lopate *et al.*, 2010). While it is certain that *SOD1* causes disease in a number of familial ALS patients, these appear to be distinct from most other familial and sporadic cases of ALS that have TARDBP aggregation (Farrawell *et al.*, 2015).

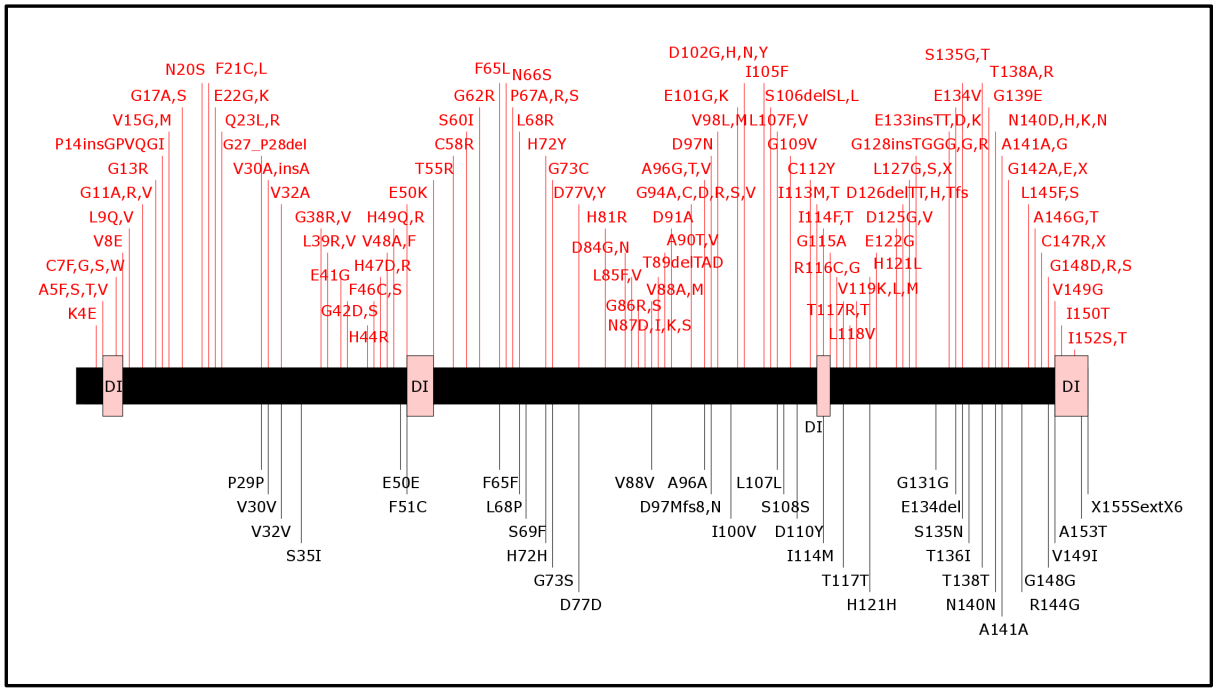

Figure 6. Mutations in *SOD1* in patients (above gene) and controls (below gene). Red =likely pathogenic. DI = dimer interface. Drawn in Inkscape v0.91.

1.2.2. *TARDBP*

The finding of transactive response DNA binding protein 43 (*TDP-43* or *TARDBP*) in ALS was a seminal step forward in our understanding of this disease, for not only did the TARDBP protein present in the aggregates of dying neurons but mutations in its gene was found to cause ALS (Sreedharan *et al.*, 2008). This publication unconventionally found the variants first and the linkage afterwards. TARDBP-positive aggregates are also found in FTD cases (Neumann *et al.*, 2006) but not in ALS cases caused by *SOD1* mutations (Mackenzie *et al.*, 2007). As *TARDBP* was further studied, it became certain that this gene was important in ALS with *TARDBP's* genetic contribution presenting in 3-5% of familial and 1-2% for sporadic patients with a number of implicated mutations (Figure 7).

TARDBP brought forward the idea of altered RNA and DNA processing as an underlying faulty mechanism in ALS, for its protein is involved in transcriptional activation, RNA stability and mRNA alternative splicing. There are two RNA-recognition motifs in *TARDBP* and a glycine-rich C-terminal. Nominally found in the nucleus, mutated TARDBP forms aggregates in the cytoplasm leaving very little protein behind to perform its job. Therefore the disease may arise from this loss of protein function. Of course, mutations in this gene might actually confer a toxic gain-of-function which is suggested by the presence of TARDBP in ALS-related aggregates, however, that engages the contentious and unsolved debate over the aggregate's role in the disease. Studies *in vitro* and *in vivo* within zebrafish cultures and transgenic models suggest both loss and gain mechanisms exist in ALS (Kabashi *et al.*, 2010).

Most causative mutations are single amino acid alterations clustering in exon 6 in the glycine-rich domain. The importance of this exon is emphasised by the fact that it constitutes 60% of the protein and 70% of the mRNA transcript (Pesiridis *et al.*, 2009). Typical ALS, ALS-FTD and pure FTD have all been observed with *TARDBP* mutations which are inherited in a highly penetrant, dominant manner (Borroni *et al.*, 2009).

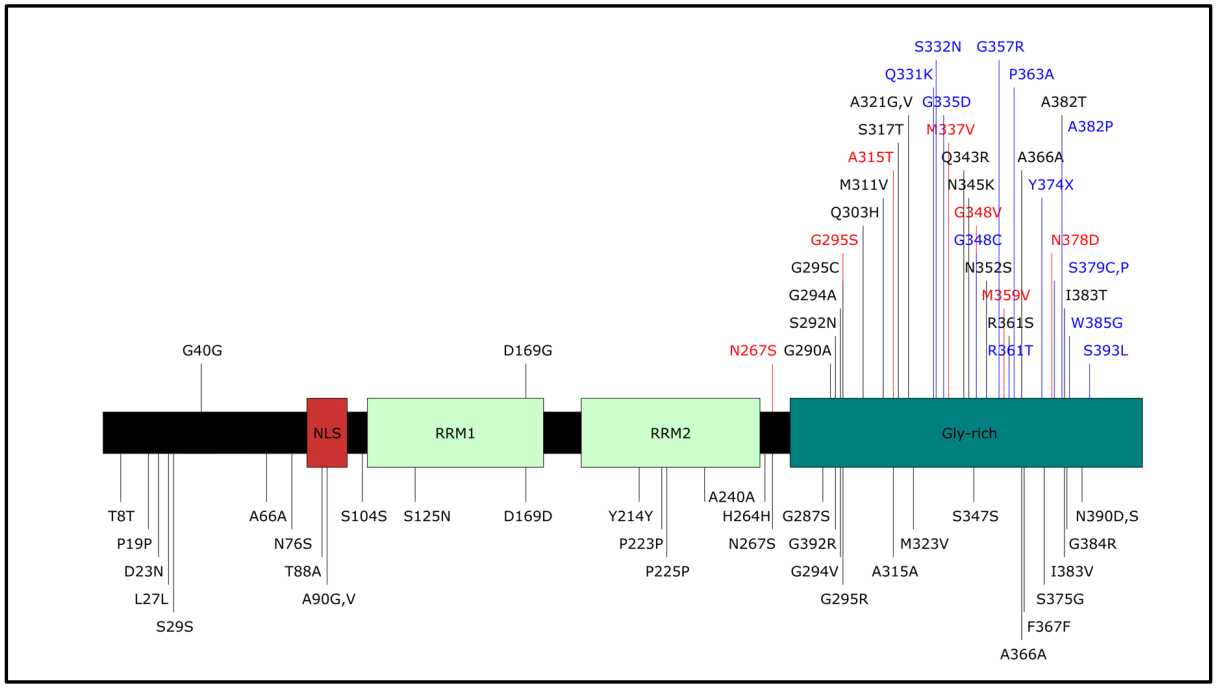

Figure 7. Mutations in *TARDBP* in patients (above gene) and controls (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; NLS = nuclear location signal; RRM = RNA recognition motif. Drawn in Inkscape v0.91.

1.2.3. *C9ORF72*

In contrast to most of the genes that will be described, the largest genetic contributor to ALS is not with SNPs but instead with a repeat expansion in the intron or promoter of *C9orf72* (Figure 8). Despite the results from linkage studies and hits in GWAS pointing to a locus on chromosome 9p21, the causal gene remained unknown for a lengthy period of time due to the nature of large repetitive regions being extremely problematic to sequence. Furthermore, the repeats were composed of GGGGCC making the area highly GC rich so adding to the difficulty in sequencing. Two independent groups discovered this mutation at the same time in ALS-FTD pedigrees (DeJesus-Hernandez *et al.*, 2011b; Renton *et al.*, 2011).

In the general European population, 90% of people will have between two and ten repeats in this gene (Renton *et al.*, 2011). Patients seen with this polyQ expansion have hundreds to thousands of repeats but the exact cut off between pathogenic and benign remains unclear. Another neurodegenerative repeat expansion disorder is Huntington's disease where CAG repeats of 36+ in the huntingtin gene have the potential to be causal. An observed phenomenon in Huntington's is *anticipation* where the severity of the disease increases in the offspring of patients. This has been attributed to the instability of the expansion which gets larger with each generation. However, variable evidence supports the same effect occurring in *C9orf72*, with some reports linking expansion size and severity of onset (Rohrer *et al.*, 2015; Gijselinck *et al.*, 2015).

Inclusions are often TARDBP-positive but more commonly, dipeptide repeat (DPR) proteins are the core pathological finding in *C9orf72* cases. Both gain- and loss-of-function mechanisms have been hypothesised with the key three theories focusing on haploinsufficiency, RNA toxicity and DPR toxicity (Gitler and Tsuiji, 2016). The expansion is unstable in some tissues resulting in differences between the sizes of the expansion in blood compared to those in the brain (van Blitterswijk *et al.*, 2012a).

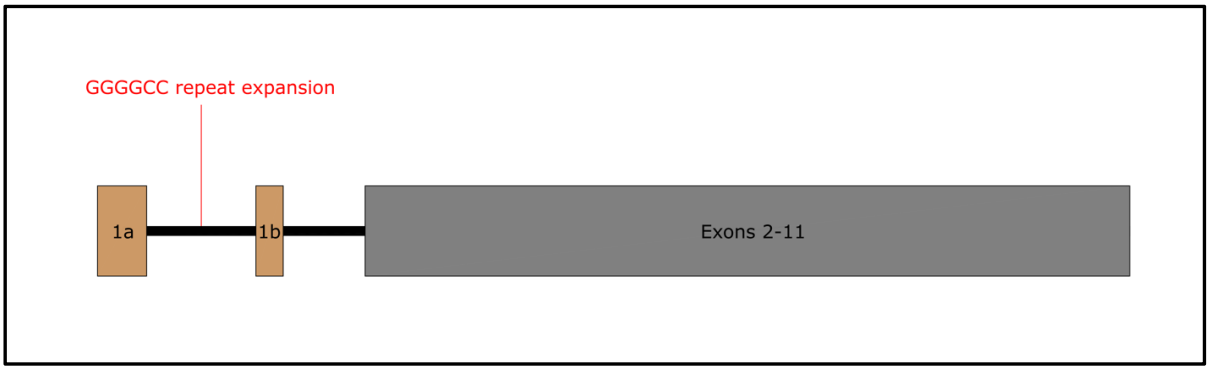

Figure 8. The repeat expansion in C9orf72 is located either in the intron or within the promoter region of the gene. Drawn in Inkscape v0.91.

1.2.4. *FUS*

Another big player in the genetics of ALS is fused in sarcoma (*FUS*), also known as translocated in sarcoma (*TLS*; Figure 9). This gene was uncovered by two groups simultaneously in families with dominantly inherited ALS (Vance *et al.*, 2009) and recessive ALS (Kwiatkowski *et al.*, 2009). In a similar vein to *TARDBP*, *FUS* regulates mRNA alternative splicing and transcription while additionally being concerned with DNA repair to reduce genetic damage. ALS mutations in this gene induce a malfunctioning RNA metabolism and, accordingly, often disturb the region for RNA binding: the C-terminal domain (Lagier-Tourenne and Cleveland, 2009). Herein lies the link to *TARDBP* again for the same C-terminal region exhibits a clustering of disease-causing mutations (Sleegers and Van Broeckhoven, 2009). However, exons 3, 5 and 6 also contain mutation sites which, in contrast to the C-terminal, appear more frequently in sporadic patients rather than familial (Lattante *et al.*, 2013b).

The FUS protein is located in the nucleus (like *TARDBP*) yet mutant forms are also observed to accumulate in the cytoplasm as misfolded protein (Bosco *et al.*, 2010). However, TARDBP-positive inclusions which are found in most cases of ALS, are not present in patients with *FUS* mutations suggesting that FUS toxicity is independent of TARDBP (Lagier-Tourenne and Cleveland, 2009). In short, *FUS* is likely to be part of the same pathogenic cascade as *TARDBP* involving abnormal RNA processing and protein aggregation. Predictions concerning *FUS*'s contribution to fALS ranges between 0.6 and 20% whilst for sALS it is approximately 0.4-2% of cases. More than 80% of reported variants are substitutions and roughly 10% are deletions. *De novo* variants have been found (DeJesus-Hernandez *et al.*, 2010) as well as a range of phenotypes, including a few of juvenile onset (Bäumer *et al.*, 2010).

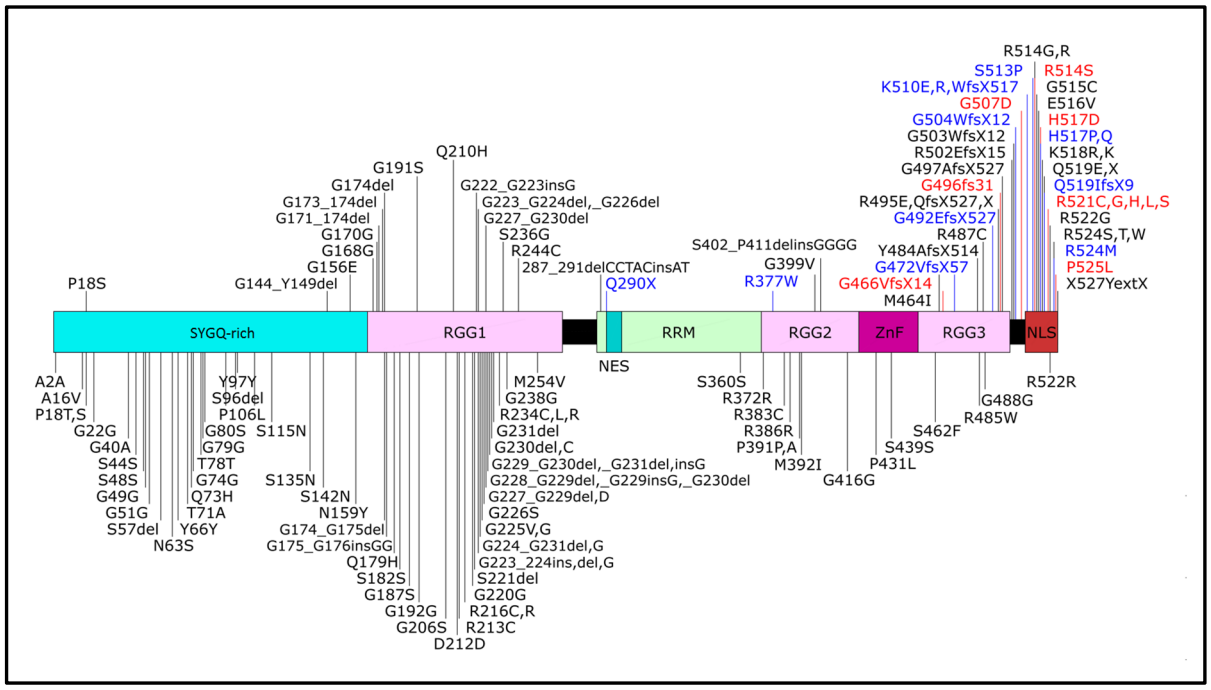

Figure 9. Mutations in *FUS* in patients (above gene) and controls (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; RGG = arginine-glycine-glycine-rich; RRM = RNA recognition motif; ZnF = zinc finger; NLS = nuclear localisation signal. Drawn in Inkscape v0.91.

1.2.5. *ANG*

Thus far, the genes mentioned above can produce high-risk, causal variants whereas not all genes implicated in ALS are such. One example is angiogenin (*ANG*) whose mutations have been seen to increase the risk of not only ALS, but Parkinson's disease and ALS-FTD as well (van Es *et al.*, 2009; van Es *et al.*, 2011). Greenway and colleagues (2004) selected angiogenin as a potential ALS gene after evaluation of a previous paper on the *APEX* gene which contained a synonymous polymorphism associated with ALS (Hayward *et al.*, 1999). *ANG* is in close proximity to *APEX* and so they are likely to be inherited together. Additionally, *ANG's* function is similar to that of another gene implicated in ALS: *VEGF*, as discussed later. For these reasons, *ANG* was examined to find a synonymous polymorphism (rs11701) associated with ALS and eventually missense variants in 15 patients (Greenway *et al.*, 2004; Greenway *et al.*, 2006).

Incidence of potentially damaging variants accounts for 1-2% of fALS and 1% of sALS cases. Most mutations are single amino acid alterations and are suspected to induce a loss-of-function disease mechanism. Both polymorphisms and rare variants have been selected for

analysis against ALS and each have displayed some promising results which are marginally more often replicated than not amongst different populations (Greenway *et al.*, 2006; Wu *et al.*, 2007; Gellera *et al.*, 2008; Paubel *et al.*, 2008; Fernández-Santiago *et al.*, 2009; Millecamps *et al.*, 2010b). Additionally, angiogenin concentrations in the serum of ALS patients are documented as abnormally high (Cronin *et al.*, 2006).

ANG is obviously involved in angiogenesis, however, characterisation of known ALS-causing mutations implicates a reduced ribonucleolytic activity in the disease i.e. modifies rRNA (Crabtree *et al.*, 2007) or also in the supposed 'neuroprotective' ability of angiogenin (Subramanian *et al.*, 2008). Since variants identified in patients also occur in controls, it is likely that *ANG* confers a very small risk for ALS, if any (Figure 10).

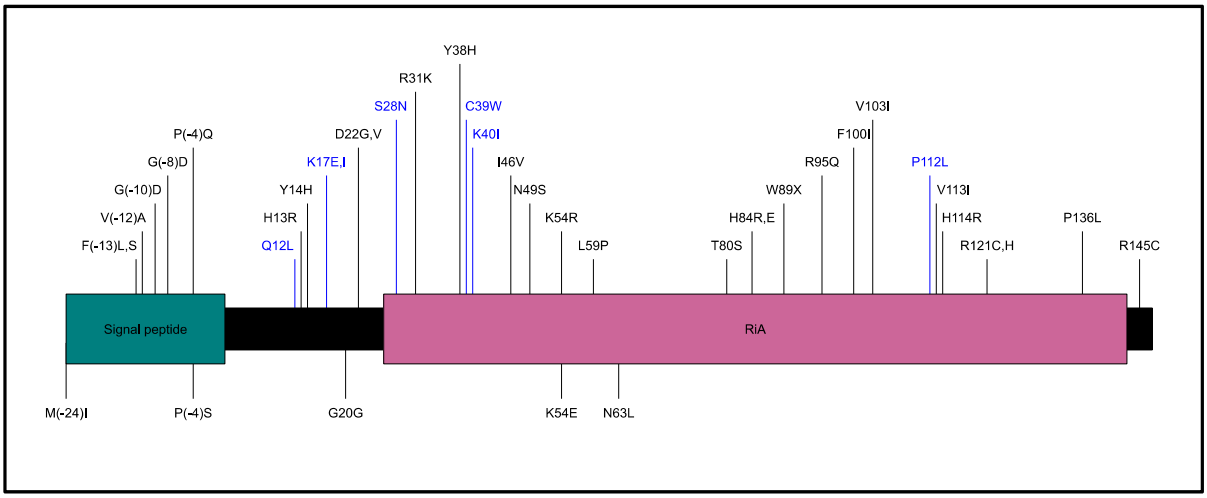

Figure 10. Mutations in *ANG* found in patients and control (above gene) and solely controls (below gene). Blue mutation = likely pathogenic; black = unknown; RiA = Ribonuclease A-domain. Drawn in Inkscape v0.91.

1.2.6. *ALS2*

Alsin (*ALS2*) is a gene initially found in ALS but is now more commonly associated with other MNDs, particularly hereditary spastic paraplegia (HSP). Hentati *et al.* (1994) first provided the linkage location for this gene at 2q33 where a homozygous *ALS2* deletion was later found in two separate families (Hadano *et al.*, 2001). Both families had multiple affected individuals with juvenile-onset ALS. The first was a large Tunisian pedigree where DNA was available from 14 affected and 10 unaffected to reveal the ALS2 gene. This loci was then sequenced in a smaller Kuwaiti family which also showed segregation in 4 affected and 3 unaffected individuals. Both families were inbred. Alsin is abundant in motor neurons and has been

implicated in GTPase activation, cell division, differentiation, apoptosis and endocytosis (Yamanaka *et al.*, 2003).

There are a number of publications on families with mutated *ALS2* causing infantile ascending HSP (Eymard-Pierre *et al.*, 2002; Lesca *et al.*, 2003; Devon *et al.*, 2003; Verschuuren-Bemelmans *et al.*, 2008; Herzfeld *et al.*, 2009; Eker *et al.*, 2014) or juvenile PLS (Yang *et al.*, 2001; Panzeri *et al.*, 2006; Shirakawa *et al.*, 2009; Mintchev *et al.*, 2009). As documented in these two diseases, in ALS this gene's role is strictly limited to a recessively-inherited, juvenile-onset disease (Yang *et al.*, 2001; Kress *et al.*, 2005; Luigetti *et al.*, 2013; Siddiqi *et al.*, 2014). Simple substitutions occur in homozygous or compound heterozygous form to induce a disease of normally slow progression (Figure 11). The association in sporadic cases is still currently doubtful (Beleza-Meireles and Al-Chalabi, 2009) with negative results in any study examining the adult-onset form of the disease (Hand *et al.*, 2003; Al-Chalabi *et al.*, 2003; Brugman *et al.*, 2007).

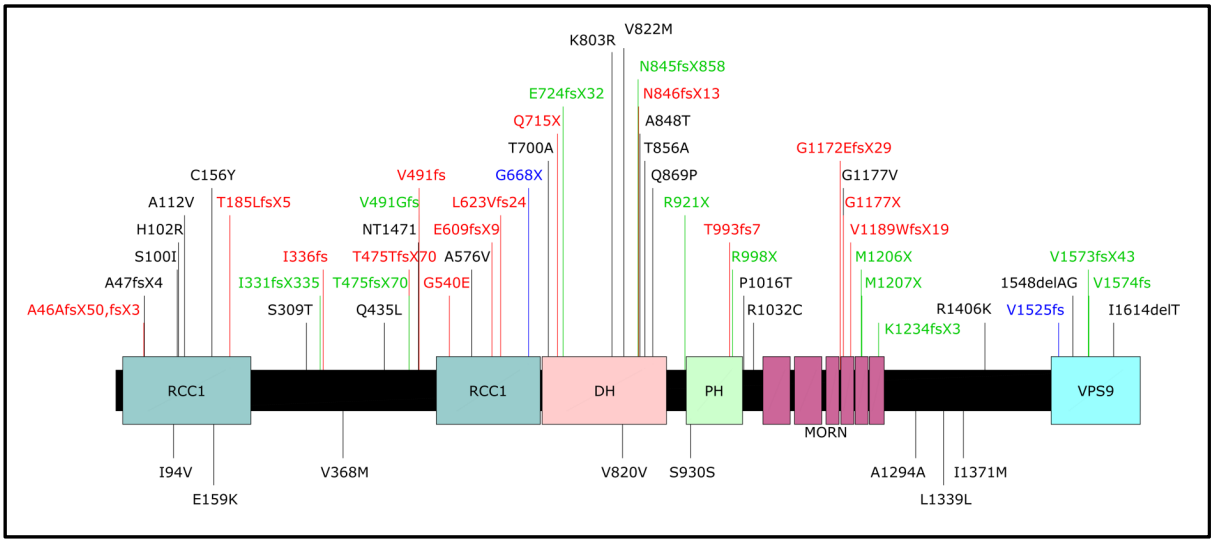

Figure 11. Mutations in *ALS2* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; green = IAHSP; RCC = Regulator of chromosome condensation; DH = Dbl homology; PH = PH domain-like; MORN = Membrane Occupation and Recognition Nexus; VPS9 = vacuolar protein sorting-associated. Drawn in Inkscape v0.91.

1.2.7. *NEFH*

The neurofilament, heavy polypeptide (*NEFH*) gene encodes a protein which forms a structural network with other neurofilaments to support neurons and give the cell its shape. These proteins are especially abundant in motor neurons where they are required for the

adequate functioning of the extensive and lengthy projections. A frequent observation in the motor neurons of sporadic and familial patients is an accumulation of abnormal neurofilaments, some of which form aggregates in the cell body (Hirano *et al.*, 1984; Xu *et al.*, 1993). Additionally, mice models with mutated or overexpressed neurofilaments develop motor neuron abnormalities and axon transport deficiencies (Collard *et al.*, 1995). However, it was later found that crossing *SOD1* transgenic mice with overexpressed neurofilament mice resulted in a survival increase (Couillard-Després *et al.*, 1998; Kong and Xu, 2000). There is evidence to suggest that this protective effect is achieved through neurofilament aggregation which sequesters harmful proteins (Nguyen *et al.*, 2001) therefore neurofilament mutations may be a modifier of *SOD1* toxicity. Conversely, removal of the neurofilament, light polypeptide (*NEFL*) also increases survival in *SOD1* mice models (Williamson *et al.*, 1998).

The 'KSP motif' in *NEFH* is named such to describe the high occurrence of these three consecutive amino acids within this region (Figure 12). This repetitive segment is a largely conserved area important for the protein's role in interacting with other molecules, all of which is vital for intracellular transport to axons and dendrites (Millecamps and Julien, 2013). Neurofilaments are currently being examined as potential biomarkers for ALS to aid diagnosis (Mendonça *et al.*, 2011).

There have been a few negative studies regarding *NEFH* in ALS (Rooke *et al.*, 1996; Vechio *et al.*, 1996; Garcia *et al.*, 2006) whilst positive reports find that most variants occur in sporadic cases rather than familial, in the form of deletions and insertions in the KSP domain (Figlewicz *et al.*, 1994; Tomkins *et al.*, 1998; Al-Chalabi *et al.*, 1999; Skvortsova *et al.*, 2004). In part, this explains the theory that *NEFH* alterations behave as a risk factor for ALS rather than being capable of causing disease alone (Cleveland and Rothstein, 2001). Moreover, excessive glutamate stimulation has been shown to affect neurofilaments and cause axon transport deficits in cell cultures (Ackerley *et al.*, 2000) leading to the idea that excitotoxicity in ALS may precede neurofilament malfunction which then exacerbates or alleviates the problem.

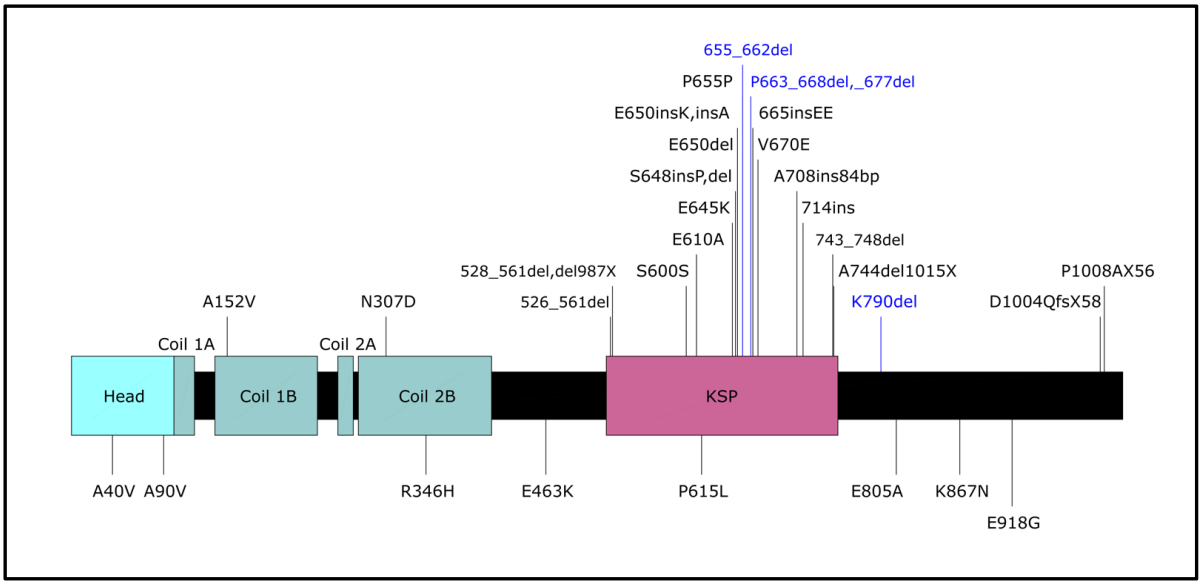

Figure 12. Mutations in *NEFH* found in patients (above gene) and controls (below gene). Blue mutation = likely pathogenic; black = unknown; KSP = lysine-serine-proline repeats. Drawn in Inkscape v0.91.

1.2.8. *OPTN*

Optineurin (*OPTN*) is unusual in that it is one of the few genes to be discovered by homozygosity mapping and can cause both dominant and recessive forms of ALS (Maruyama *et al.*, 2010; Figure 13). Using this technique on 6 individuals from 5 different families with consanguineous marriages revealed 17 genes linked to the disease and *OPTN* as the most likely candidate within these. A further 683 ALS patients were sequenced which confirmed mutations in *OPTN*, none were present in 2509 Japanese controls. The disease course in these patients was generally less severe than typical ALS. All three identified mutations appear to affect regions in *OPTN* important for binding to ubiquitin. Prior to this and more often than ALS, this gene causes primary open-angle glaucoma (POAG). Here, mutations exhibit an inhibition in the NF-κB activity of *OPTN* whereas ALS-associated variants do not, alternatively, they increase this activity (Maruyama *et al.*, 2010). Following this, investigations into these mutations provided a potential link between ALS and the immune system (Sakaguchi *et al.*, 2011). The OPTN protein is also implicated in Golgi maintenance, membrane trafficking, exocytosis and autophagy.

Mutations in this gene are predominantly found in Japanese and Italian populations, and are seemingly rarer outside these areas (Solski *et al.*, 2012). It is hypothesised that dominant *OPTN* cases vary in the mechanism of disease compared to recessive *OPTN* cases and that

of POAG too (Andersen and Al-Chalabi, 2011). Most variations are single amino acid changes which occur in <1% of fALS and <1% of sALS. Optineurin-positive inclusions have been detected in a small number of ALS patients and more commonly in those with causal mutations (Hortobágyi *et al.*, 2011; Ito *et al.*, 2011).

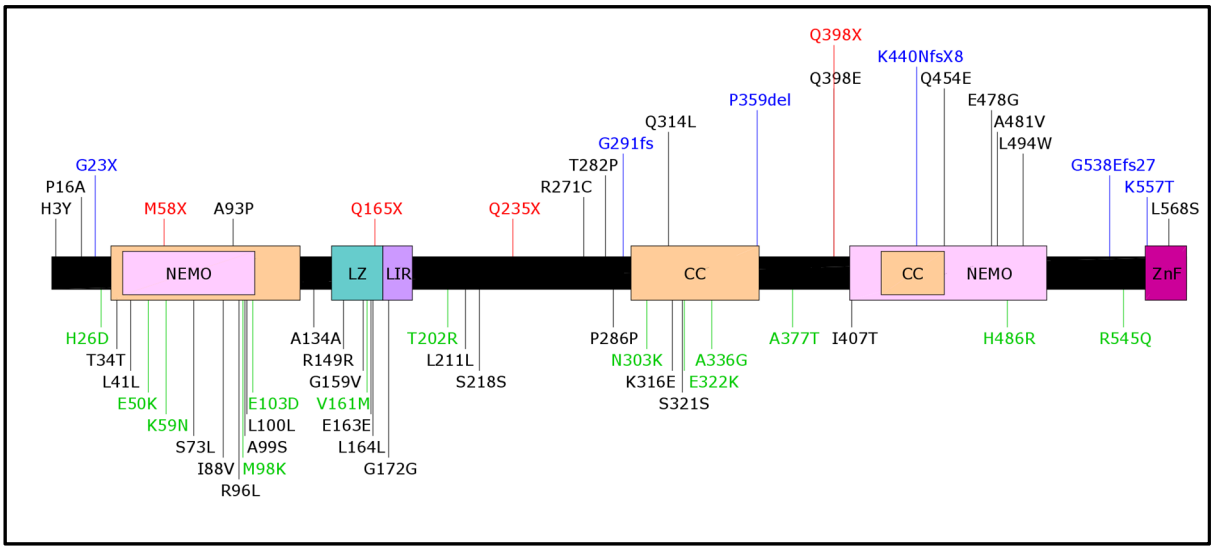

Figure 13. Mutations in *OPTN* found in patients (above gene) and controls or unrelated diseases (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; green = glaucoma; NEMO = NF-kappa-B essential modulator; LZ = Leucine zipper; LIR = LC3-interacting region; CC = coiled coil; ZnF = zinc finger. Drawn in Inkscape v0.91.

1.2.9. *VAPB*

VAPB (vesicle-associated membrane protein-associated protein B/C) is a minor but replicated ALS-causing gene (Figure 14). Linkage of eight families with typical and atypical phenotypes from Brazil found the location 20q13 and sequencing in one of these families established the *VAPB* variant P56S. This segregated with ALS in 12 affected and 4 unaffected (Nishimura *et al.*, 2004). The authors then demonstrated that this variant was absent from 400 matched controls and was present in 22 individuals from the other 6 families, providing strong genetic evidence for this gene. The same mutation in *VAPB* has been seen to cause ALS in some while causing spinal muscular atrophy (SMA) in others (Nishimura *et al.*, 2004). The most common mutations are heterozygous substitutions with P56S being the most frequent variant, especially in Brazilian populations due to a common founder in this population (Nishimura *et al.*, 2005; Landers *et al.*, 2008a; Vinay Kumar *et al.*, 2014).

Associated with the endoplasmic reticulum (ER), *VABP's* functions vary from lipid metabolism to activating the unfolded protein response (UPR) and vesicular trafficking. ER-stress and the UPR are both renowned as faulty in MND (Kanekura *et al.*, 2009). The UPR is involved in the removal of unwanted waste within cells and the dysfunction of which causes aggregates to assemble – a hallmark of ALS. Functional studies have joined the P56S mutation with causing aberrant localisation of the VAPB protein so that it no longer associates with the ER and instead forms aggregates (Nishimura *et al.*, 2004).

Other variants in *VAPB* include S160del which shows segregation in a family (excluding those of younger age) of two affected and five unaffected. However, it is also exists in the general population albeit infrequently (0.45%), and doesn't disrupt VAPB protein in neurons suggesting this variant is not pathogenic (Landers *et al.*, 2008a). Additionally, the variant T46I was only observed in a single patient but did affect *VAPB's* cellular functions by inhibiting the UPR making the variant a more likely candidate (Chen *et al.*, 2010).

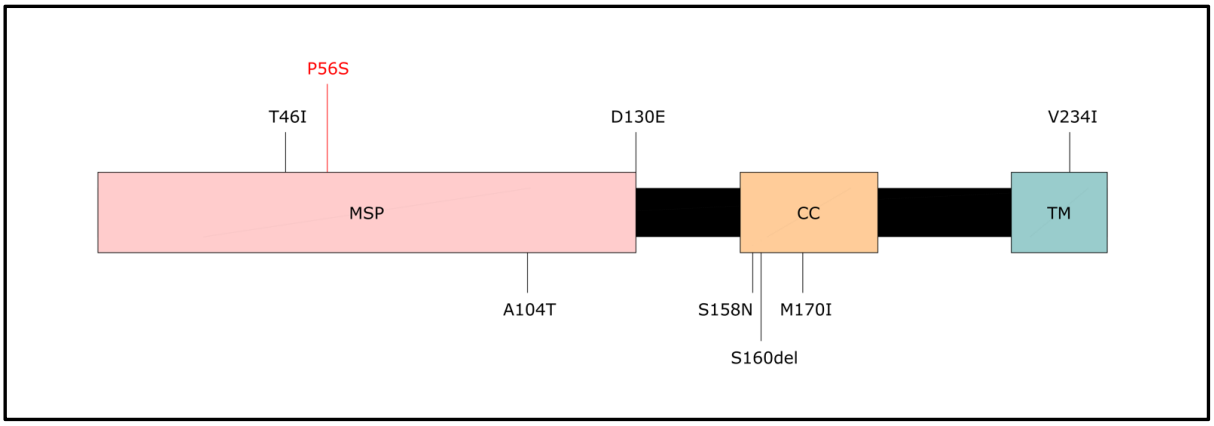

Figure 14. Mutations in *VAPB* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; black = unknown; MSP = major sperm protein domain; CC = coiled coil; TM = transmembrane domain. Drawn in Inkscape v0.91.

1.2.10. *DCTN1*

Dynactin 1 (*DCTN1*) is attributed to roles in retrograde transport via microtubules and vesicle trafficking. In mice, when the function of the dynactin protein and its associate protein dynein is harmed, a phenotype reminiscent of motor neuron disease is observed (Teuling *et al.*, 2008). These two proteins form a complex within cells and are considered crucial to neuronal maintenance.

Multiple neurodegenerative disorders have been connected to *DCTN1* including ALS, Perry syndrome, PD and FTD (Münch *et al.*, 2005; Newsway *et al.*, 2010; Uribe, 2010; Araki *et al.*, 2014). In ALS, point mutations are seen to be dominantly causal (Münch *et al.*, 2004; Figure 15). Research into published alterations demonstrate abnormal folding of the protein and an inability to bind appropriately with other associated compounds (Puls *et al.*, 2003). Malfunctioning retrograde transport is implicated in the disease process of ALS with *DCTN1* adding to this theory.

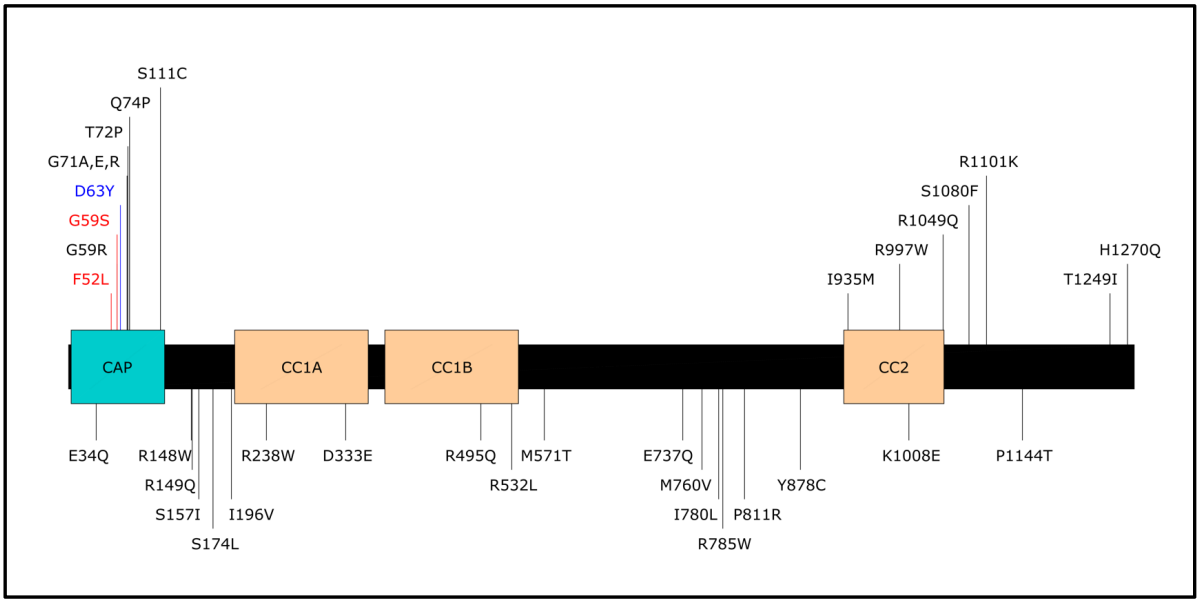

Figure 15. Mutations in *DCTN1* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; CAP = Cytoskeleton-associated protein; CC = coiled coil. Drawn in Inkscape v0.91.

1.2.11. *VCP*

In 2010, valosin containing protein (*VCP; P97*) was identified by WES as a causal gene for ALS (Johnson *et al.*, 2010). This gene had previously been connected to inclusion body myopathy with Paget disease and frontotemporal dementia (IBMPFD), Charcot-Marie-Tooth (CMT) and pure FTD cases which both have an overlap in their genetic aetiology with ALS (Bersano *et al.*, 2009). In the initial study, linkage and exome sequencing revealed four variants in different genes which segregated with the disease (Figure 16). VCP was picked from these four as the most likely candidate given its history with causing IBMPFD (Johnson *et al.*, 2010). The authors then went on to sequence an additional 288 patients finding four more VCP mutants and so making a case for the pathogenic involvement of this gene in ALS.

VCP is implicated in the ubiquitin-proteasome system which, considering the pathological hallmark of ALS is ubiquitinated aggregates, links in agreeably with the current literature (Blokhuis *et al.*, 2013). This gene is additionally connected to another related mechanism of protein degradation, namely, autophagy (Ju and Weihl, 2010; Tresse *et al.*, 2010). As will be highlighted in the rest of this report, autophagy continually reappears in the story of ALS. Furthermore, VCP regulates mitochondria-associated proteins which is an organelle recognised to be faulty in ALS (Xu *et al.*, 2011).

Current estimates for *VCP*'s involvement in familial and sporadic ALS stand at 1-2% and 1% respectively. Not all studies have replicated this finding suggesting it is either rarer than predicted or restricted to certain populations (Miller *et al.*, 2012; Tiloca *et al.*, 2012; Williams *et al.*, 2012a; Zou *et al.*, 2013a). Most mutations reported in *VCP* are either substitutions or intronic variants and cause dominantly-inherited ALS (DeJesus-Hernandez *et al.*, 2011a; Abramzon *et al.*, 2012; Hirano *et al.*, 2015).

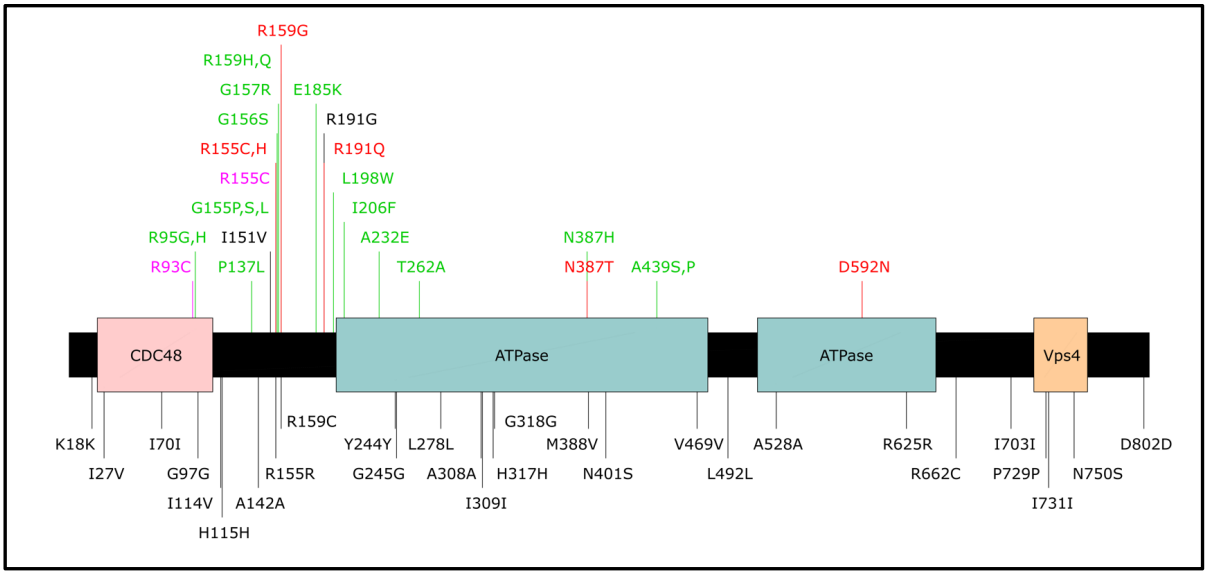

Figure 16. Mutations in *VCP* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; pink = in IBMPFD; green = in CMT; CDC48 = N-terminal domain; Vps4 = oligomerisation C-terminal. Drawn in Inkscape v0.91.

1.2.12. *ATXN1-2*

Other genes united to ALS by repeat expansions include the *ATXN1-2* genes. These code for the proteins ataxin 1 and ataxin 2 which are both ubiquitously expressed and located in the cytoplasm (*ATXN1*) and nucleus (*ATXN2*). Spinocerebellar ataxia (SCA) is a degenerative

movement disorder which can be caused by 40+ and 32+ CAG repeat expansions in the ataxin 1 and 2 genes respectively (Banfi *et al.*, 1994; Lorenzetti *et al.*, 1997). Both *ATXN* genes appear to be involved in processing RNA, a common pathway in ALS pathology.

The *ATXN2* gene was largely introduced into ALS genetics when an American group used an unbiased screen in yeast to look for genes involved in the toxicity of *TARDBP* (Elden *et al.*, 2010). They found 27 genes in which overexpression increased *TARDBP*-induced cell death. One of these genes was *ATXN2* and given its past in causing SCA, the authors explored this gene in fly models and in humans to find that intermediate repeats of 27-33 glutamines generated a significant risk for ALS in a cohort of 915 patients compared to 980 controls. Previous to this study, there have been reports of SCA and MND within the same patient (Infante *et al.*, 2004; Nanetti *et al.*, 2009). Consequently, these two diseases have been suggested to overlap in a similar fashion to FTD and ALS, in that they extend along the same spectrum of a syndrome (Andersen and Al-Chalabi, 2011). The finding of intermediate repeats in ALS is a replicated finding (van Damme *et al.*, 2011; Yu *et al.*, 2011; Liu *et al.*, 2013; Lattante *et al.*, 2014; Chiò *et al.*, 2015a; Figure 17) with reports also linking ATXN1 expansion repeats to ALS as well (Conforti *et al.*, 2012; Spataro and La Bella, 2014).

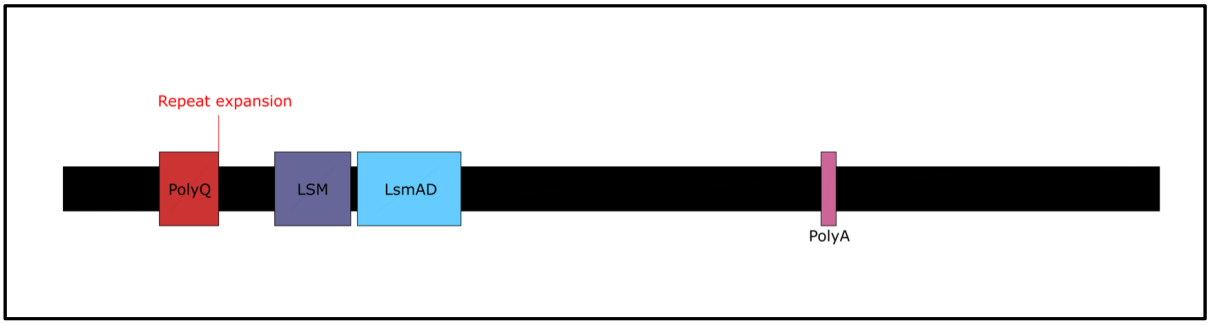

Figure 17. Repeat expansion in *ATXN2*. PolyQ = polyglutamine region; LSM = like-sm domain; LsmAD = lsm-associated domain; PolyA = poly(A)-binding. Drawn in Inkscape v0.91.

1.2.13. *SETX*

Mutations in senataxin (*SETX*) were discovered to cause ataxia with oculomotor apraxia type 2 (AOA2) in the homozygous form in a number of families (Moreira *et al.*, 2004). AOA2 is a movement disorder where the eyes are also affected. Not long after this publication, linkage and sequencing also tied *SETX* variants to ALS4 in a heterozygous capacity (Chen *et al.*, 2004). ALS4 describes a dominant, juvenile-onset form of the disease with a slow progression. Segregation was observed in three families confirming its likely role in the disease.

The function of this gene is unknown but theories generally centre on DNA/RNA processing and repair for this gene contains a helicase domain (De Amicis *et al.*, 2011). Like its yeast homolog, the SETX protein responds to oxidative stress and SETX-deficient cell models show a reduced binding of RNA polymerase II to a number of genes and hence lowering their expression (Suraweera *et al.*, 2009).

There have been a number of positive studies in *SETX* with many patients presenting as an early-onset disease with slow progression (Chance *et al.*, 1998; Chen *et al.*, 2004; Zhao *et al.*, 2009; Avemaria *et al.*, 2011; Saracchi *et al.*, 2014; Figure 18). Although it has been claimed that this gene is associated with a high risk of ALS (Leblond *et al.*, 2014) some studies report relatives of a *SETX* patient also harbouring the alteration with a proposition that other modifiers affect this mutation (Hirano *et al.*, 2011). It is of course also feasible that the reported variant was not actually pathogenic in the initial ALS cases. Rudnik-Schöneborn *et al.* (2012) reported a family with proximal spinal muscular atrophy harbouring *SETX* mutations where the affected father had a previously reported pathogenic mutation while both affected children had this and another *SETX* variant of unknown consequence from their unaffected mother. Both children had an earlier age of onset by 20 years suggesting these variants combined to produce a more severe phenotype.

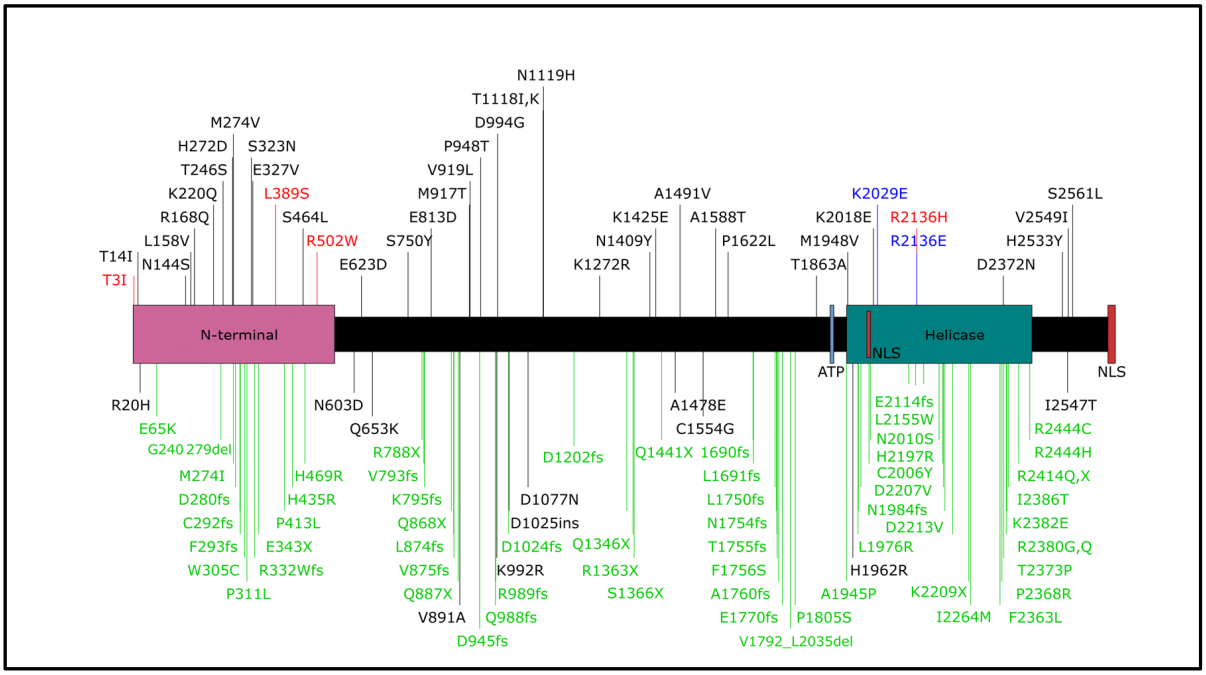

Figure 18. Mutations in *SETX* found in patients (above gene) and controls or another disease (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; green = AOA2; ATP = ATP-binding; NLS = nuclear localisation signal. Drawn in Inkscape v0.91.

1.2.14. *CHMP2B*

Genetic overlap between FTD and ALS arose again with the discovery of the chromatin modifying protein 2B (*CHMP2B*) gene. Linkage analysis and eventual sequencing identified mutations in a six-generation Danish family with FTD (Skibinski *et al.*, 2005). This mutation lay in the splice site for exon 6 and segregated in the 11 affected and 14 unaffected individuals screened. Expressing this mutant in cell cultures exhibited abnormal CHMP2B protein location, which is usually found dispersed in the cytosol, and an atypical endosome morphology (Skibinski *et al.*, 2005). Following this strong evidence for *CHMP2B* as a disease-causing gene, publications in both FTD (Eskildsen *et al.*, 2009; Lindquist *et al.*, 2008; Holm *et al.*, 2009) and ALS (Parkinson *et al.*, 2006; Cox *et al.*, 2010; van Blitterswijk *et al.*, 2012d; Figure 19). Indeed there have been some negative results published in this gene indicating it is a rare cause of disease (Cannon *et al.*, 2006; Rizzu *et al.*, 2006; Schumacher *et al.*, 2007; Blair *et al.*, 2008).

CHMP2B provides another link to dysfunctional autophagy regulation in ALS. As part of the endosomal sorting complex required for transport (ESCRT) machinery, *CHMP2B* aids the formation of endosomes and hence the removal of unwanted proteins. Lee *et al.* (2007) assessed that mutations in *CHMP2B* or loss of ESCRT function caused dendrites to retract leading to eventual neuronal cell death. Lastly, part of the ESCRT complex also interacts with spastin, a protein which causes hereditary spastic paraplegia displaying a clear link between these proteins and neurodegeneration (Reid *et al.*, 2005). However, CHMP2B is ubiquitously expressed so it is curious how specific neuronal cells would be vulnerable to its effects.

Missense variants have been reported in FTD however it is hypothesised these are unlikely to be damaging and only those which delete the C-terminal are considered pathogenic (Ferrari *et al.*, 2010). Finally, inclusions in mutant *CHMP2B* cases were positive for *SQSTM1* which, as discussed later, is a common feature in neurodegenerative diseases (Parkinson *et al.*, 2006).

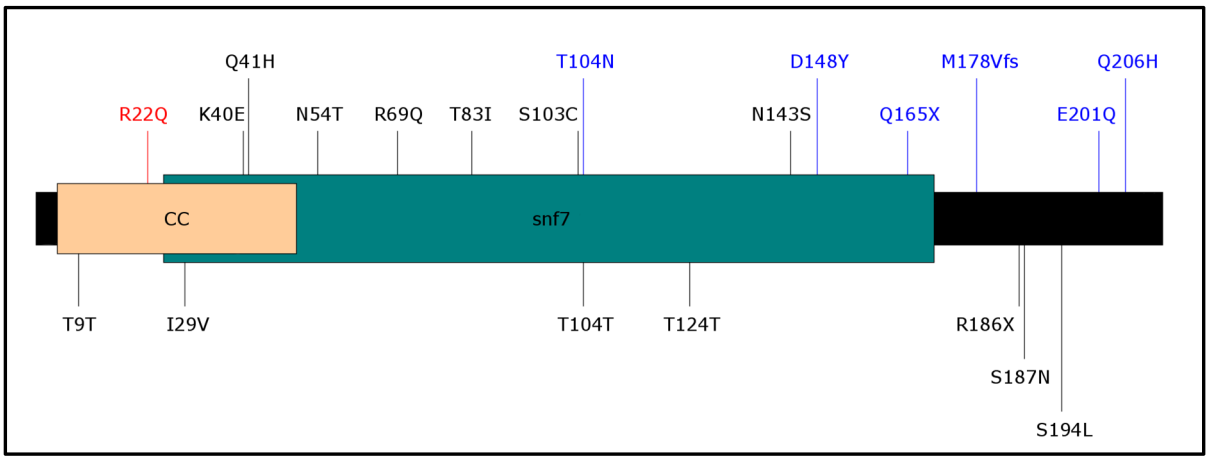

Figure 19. Mutations in *CHMP2B* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; CC = coiled coil; snf7 = vacuolar sorting protein snf7. Drawn in Inkscape v0.91.

1.2.15. *UBQLN2*

The chief X-chromosome gene of importance is ubiquilin 2 (*UBQLN2*) which causes dominantly inherited ALS. Mutations are observed as both juvenile- and adult-onset and, as implied by its name, *UBQLN2* is highly involved in the ubiquitin-mediated protein degradation pathway. The first report of this gene came from a five-generation family with ALS and ALS-FTD. It was observed that there was no male-to-male transmission was present and linkage analysis pointed towards an area on the X-chromosome (Deng *et al.*, 2011). This pedigree displayed a reduced penetrance and lower age of onset in females. The causal mutation was identified as P497H which lies in the PXX-repeat domain and was not found in 928 controls. To examine this gene further, 188 more fALS or ALS-FTD patients were sequenced to find four more mutants all affecting proline residues in the same PXX domain. Functional work revealed that inclusions in these patients and in 47 other ALS cases were all positive for UBQLN2. Additionally, the ubiquitin proteasome system (UPS) is disrupted by mutations in *UBQLN2* (Deng *et al.*, 2011).

Predicted to occur in just under 1% of both familial and sporadic ALS, most reported publications document proline amino acid substitutions as the causal origin (Figure 20). Three group found the P506S mutation in different populations (Gellera *et al.*, 2013; Vengoechea *et al.*, 2013; Özoğuz *et al.*, 2015), one of which was located in a large family where X-linked inheritance was suspected. This family had a range of diagnoses including ALS, FTD, HSP, bulbar palsy and multiple sclerosis.

Four amino acids upstream of the PXX region lies the T487I variant found to segregate in two different families with a likely common ancestor (Williams *et al.*, 2012b). Outside of these areas, variants have been identified however the evidence of their pathogenicity is not as strong. Daoud *et al.* (2012a) found two rare mutations in 590 patients but only sequenced 190 controls while Synofzik *et al.* (2012) found three variants in 206 patients, none of which were in 1450 controls. Despite such a high number of controls, this paper does not report the variants uncovered in them which would have been beneficial for the research community. There have also been many negative studies in this gene (Millecamps *et al.*, 2012; van Doormaal *et al.*, 2012; Hernández *et al.*, 2012; McLaughlin *et al.*, 2014; Kim *et al.*, 2014).

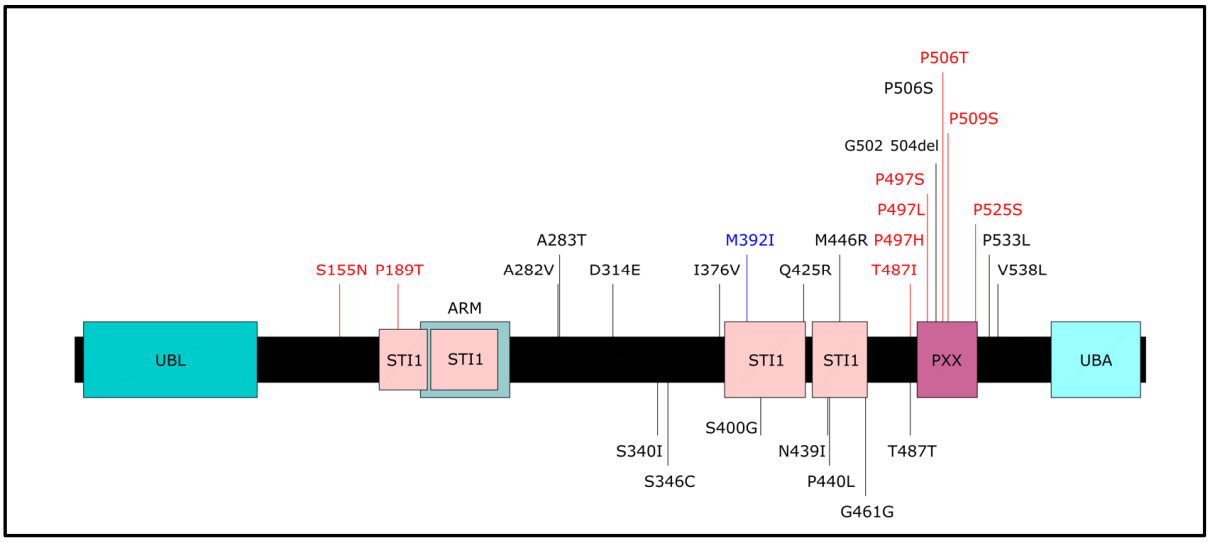

Figure 20. Mutations in *UBQLN2* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; blue = likely pathogenic; black = unknown; UBL = ubiquitin-like domain; STI= stress-inducible phosphoprotein; PXX = proline XX repeat region; UBA = ubiquitin-associated domain. Drawn in Inkscape v0.91.

1.2.16. *SQSTM1*

Paget disease of bone (PDB) is a disorder producing bone deformities where the osteoclast cells become overactive, a process involving the NF-κB protein complex which, as described previously (Chapter 1.2.8. *OPTN*), has a wide variety of functions. Linkage analysis and eventual sequencing of 24 PDB families revealed sequestosome 1 (SQSTM1/p62) as the causative agent (Laurin *et al.*, 2002). As seen in *VCP*, this disease appears to have genetic overlap with ALS.

The SQSTM1 protein has several functions within cells and is located in both the nucleus and cytoplasm. It has been observed to perform as a scaffold protein and associates with ubiquitin, linking in the malfunctioning protein degradation pathway in ALS anew. The inclusion bodies from several neurodegenerative disorders immunostain for SQSTM1 including Parkinson's and Alzheimer's disease (Ciani *et al.*, 2003). *SQSTM1* responds to cellular stress, for example oxidative stress – an implicated molecular pathway in ALS. The aggregates that form in transgenic *SOD1* mice have a composition that includes the SQSTM1 protein, the overexpression of which increases the number of these inclusions (Gal *et al.*, 2007). This is dependent on the ubiquitin-associated (UBA) domain in *SQSTM1*. Conversely, overexpression of this gene also results in a reduction of TARDBP aggregates *in vitro* which is dependent on autophagy (Brady *et al.*, 2011)

Variants giving rise to ALS occur in about 1% of familial and <1% of sporadic cases (Rubino *et al.*, 2012; Hirano *et al.*, 2013; Figure 21) although one study reported mutations in 2-3% of ALS patients (Fecto *et al.*, 2011). Moreover, FTD cases with *SQSTM1* mutations have also been published (Le Ber *et al.*, 2013) as well as a recent discovery of this gene causing atypical apraxia (Boutoleau-Bretonnière *et al.*, 2015).

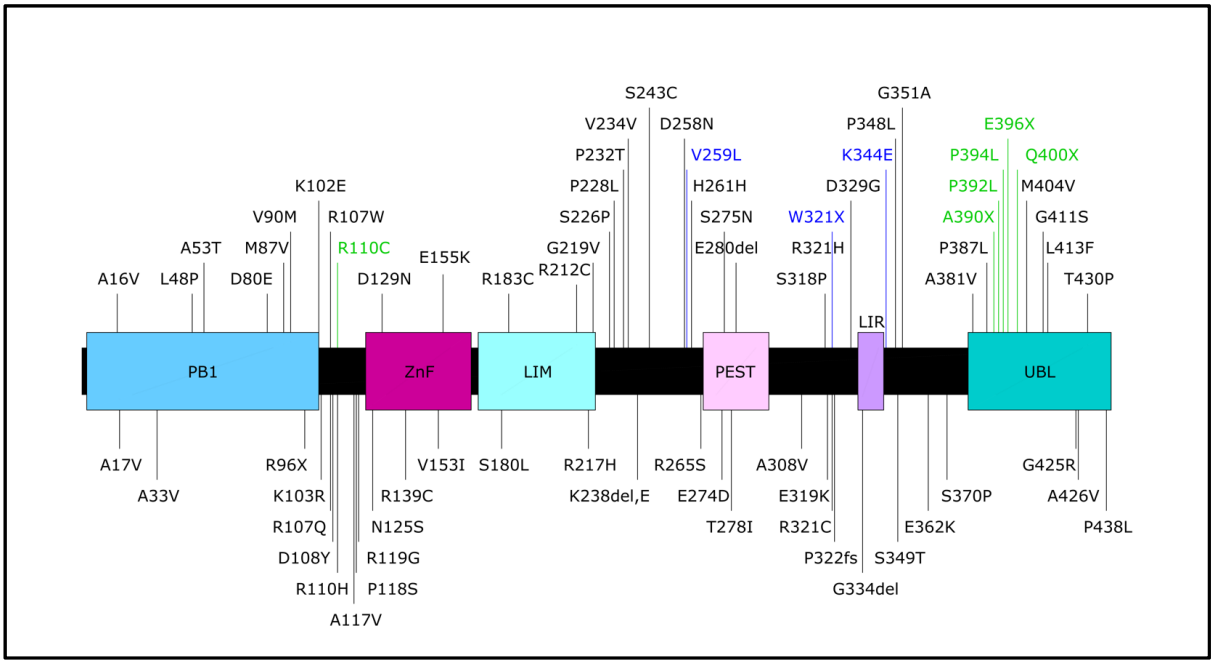

Figure 21. Mutations in *SQSTM1* found in patients (above gene) and controls (below gene). Blue mutation = likely pathogenic; black = unknown; green = PBD; PB1 = Phox and Bem1 domain; ZnF = zinc finger; LIM = LIM-binding; PEST = proline-glutamic acid-serine-threonine rich; LIR = LC3-interacting region; UBL = ubiquitin-like domain. Drawn in Inkscape v0.91.

1.2.17. *SPG11*

With 40 exons to its name, spastic paraplegia 11 (*SPG11*) is a colossal gene which causes autosomal recessive spastic paraplegia and more rarely ALS. The protein, spatacsin, is of unknown function though it is speculated to be involved in the maintenance of mitochondria and in the transport of proteins and vesicles (Murmu *et al.*, 2011).

The main publications of this gene in ALS include an exome sequencing study on a family with recessive, juvenile ALS and HSP which revealed compound heterozygous variants in SPG11 (Daoud *et al.*, 2012b). Then sequencing in a large cohort of sporadic and familial ALS patients from Turkey where a homozygous stopgain mutation was observed in a single familial subject (Özoğuz *et al.*, 2015). Beyond this, almost all reports in SPG11 are specific to spastic paraplegia.

1.2.18. *PRPH*

In a similar vein as *NEFH*, peripherin (*PRPH*) forms filaments which provide structural support to cells and, on occasion, they also bond with neurofilaments. Additionally, the PRPH protein is synthesised following nerve injury to promote neuronal repair (Belecky-Adams *et al.*, 1993). Accordingly, it was believed that *PRPH* has neural regenerative abilities but this is contested by the observation that transgenic mice overexpressing *PRPH* develop selective dysfunction and eventual death of the motor neurons (Beaulieu *et al.*, 1999). Like *NEFH*, peripherin is recognised as forming inclusions in ALS patient motor neurons (Corbo and Hays, 1992) and is also found to be abnormal in *SOD1* transgenic mice (Tu *et al.*, 1996). A similar gene, *PRPH2*, is published as causing retinal neuronal degeneration (Jordan *et al.*, 1992).

Mutations in *PRPH* are seen to cause autosomal recessive ALS however very few studies have been published in this gene (Figure 22). Examining 190 ALS and 380 control subjects revealed two mutations, one intronic and the other causing a premature protein truncation (Gros-Louis *et al.*, 2004). This latter variant was expressed in cell cultures which exhibited abnormal neurofilament structures. Leung *et al.* (2004) detected the homozygous D141Y variant which was then later confirmed in an Italian cohort of 122 ALS and 245 controls (Corrado *et al.*, 2011). This mutant caused the PRPH protein to be predisposed to aggregation

and surviving motor neurons in the affected patient's spinal cord contained peripherin-positive inclusions (Leung *et al.*, 2004). However, these two variants have been called into question for causing divergent forms of inheritance (Schymick *et al.*, 2007).

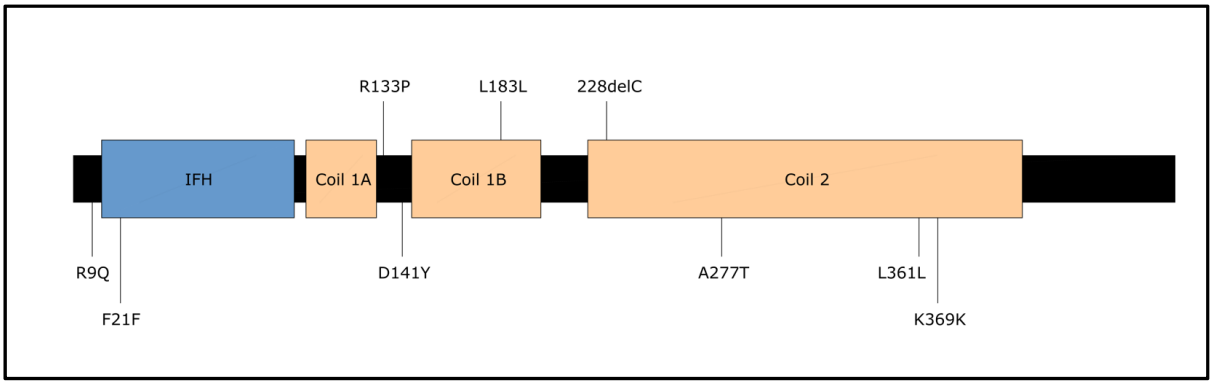

Figure 22. Mutations in *PRPH* found in patients (above gene) and controls (below gene). All mutations are of uncertain pathogenicity. IFH = intermediate filament head. Drawn in Inkscape v0.91.

1.2.19. *FIG4*

Known by a few names, *FIG4* (factor induced gene; ALS11; *SAC3*) functions as a polyphosphoinositide phosphatase and causes a disease identified as CMT, mentioned in section 1.2.11, which affects both the sensory and motor neurons. In rarer circumstances, *FIG4* is also associated with ALS and PLS where it is hypothesised to be a risk factor for these disorders (Figure 23). The function of *FIG4* is not entirely known but research suggests it is involved in autophagy (Ferguson *et al.*, 2009), as previously mentioned is a commonly observed faulty system in neurodegenerative disease (Martinez-Vicente and Cuervo, 2007). The FIG4 protein is present on endosome membranes where it aids protein trafficking (Di Paolo and De Camilli, 2006).

Mice models with a loss of *FIG4* function exhibit neurodegeneration which is thought to represent CMT rather than ALS (Chow *et al.*, 2007). The neurons of ALS and FTD patients do not stain for FIG4 but interestingly, the inclusions observed in Pick's disease and Parkinson's both present as positive for FIG4 (Kon *et al.*, 2014). Despite containing seven conserved motifs, *FIG4*'s role in ALS is not always replicated and is likely to be present in only a handful of cases worldwide (Verdiani *et al.*, 2013).

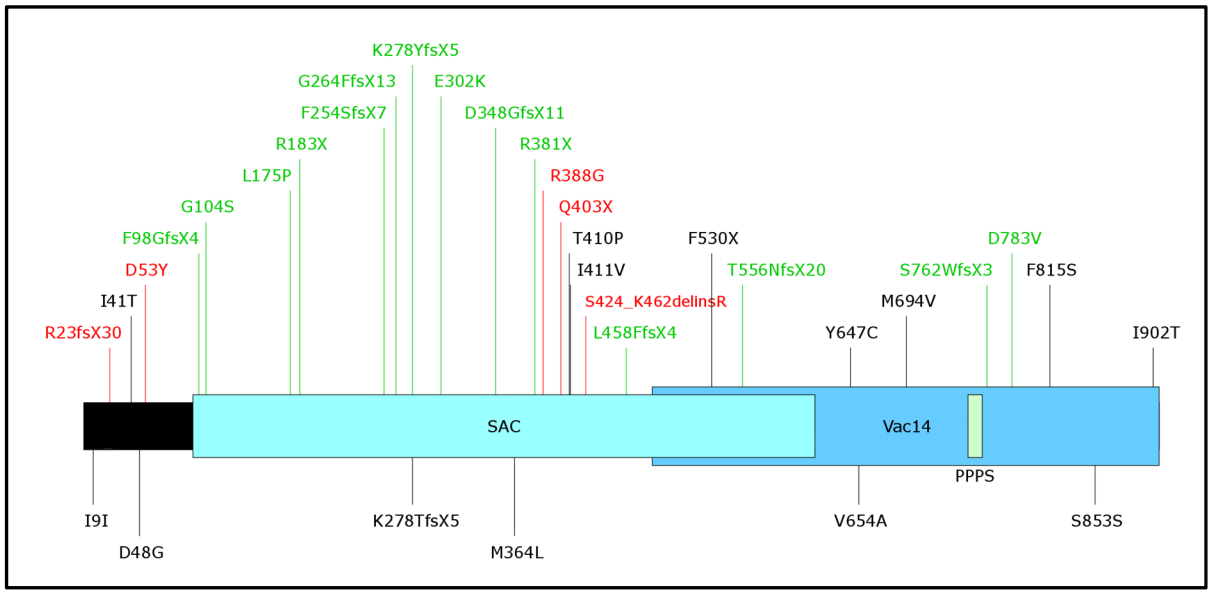

Figure 23. Mutations in *FIG4* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; black = unknown; green = CMT4J; SAC = phosphatase; Vac14 = Vac14-binding; PPPS = poly-proline-poly-serine. Drawn in Inkscape v0.91.

1.2.20. *TREM2*

A recent discovery in the accumulating number of ALS genes is triggering receptor expressed on myeloid cells 2 (*TREM2*; Figure 24). Unusually for proteins affecting motor neurons, this gene is expressed in bone marrow and has several roles in the inflammatory response which follows injury or disease (Ford and McVicar, 2009). Aside from bone-related disorders, homozygous mutations are also seen to cause early-onset dementia and FTD (Chouery *et al.*, 2008; Guerreiro *et al.*, 2013). Interestingly, heterozygous mutants have been repeatedly connected to the risk of developing various neurodegenerative diseases including ALS and Alzheimer's (Jonsson *et al.*, 2013; Rayaprolu *et al.*, 2013; Cady *et al.*, 2014). Additionally, TREM2 expression is altered in sporadic ALS (Figueroa-Romero *et al.*, 2012). It is for these reasons that variants in *TREM2* are hypothesised to contribute to mutation burden in neurodegenerative disease (Giraldo *et al.*, 2013).

In the nervous system, *TREM2* is mostly expressed in the microglia where they support neurons and respond to injury by prompting repair and regeneration (Painter *et al.*, 2015). The R47H variant is the most studied in neurodegenerative diseases, with Alzheimer's often displaying an association (Benitez *et al.*, 2013; Jin *et al.*, 2014; Abduljaleel *et al.*, 2014). Investigations into the different reported mutations support the idea of a loss-of-function

mechanism (Kleinberger *et al.*, 2014). Regarding ALS, R47H has been studied producing both positive and negative results (Cady *et al.*, 2014; Lill *et al.*, 2015; Chen *et al.*, 2015).

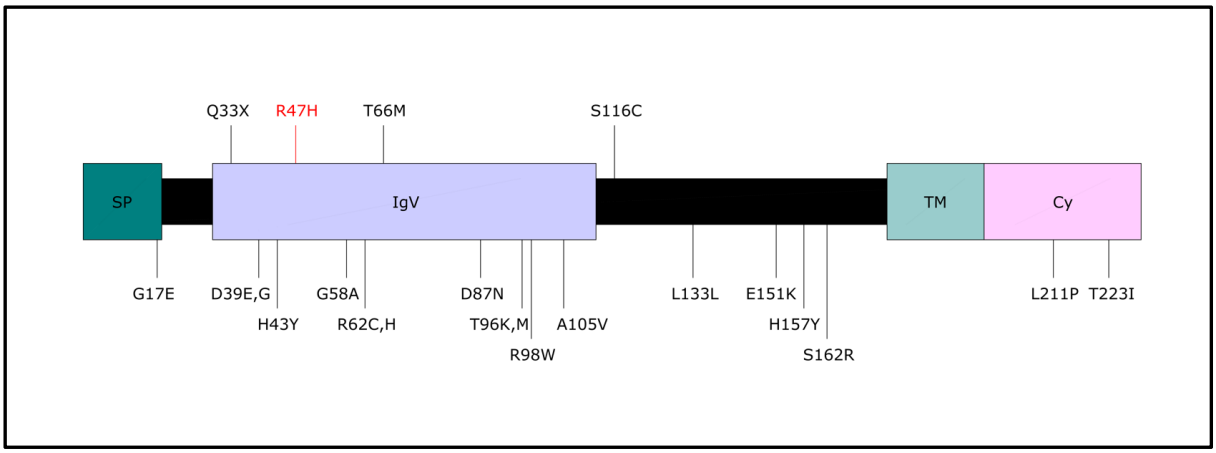

Figure 24. Mutations in *TREM2* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; black = unknown; SP = Signal peptide; IgV = IgV-set domain; TM = Transmembrane domain; Cy = Cytoplasmic domain. Drawn in Inkscape v0.91.

1.2.21. *PFN1*

In 2012, cytoskeletal deficits in ALS were once again confirmed by the function of a novel disease-causing gene: exome sequencing in two ALS pedigrees revealed mutations in profilin 1 (*PFN1*) as a cause of ALS (Wu *et al.*, 2012). Transfected cell cultures of these mutants displayed inclusions, many of which were TARDBP-positive and highly ubiquitinated (Tanaka *et al.*, 2016). The PFN1 protein is involved in controlling actin a microfilament, within the cell. This could place *PFN1* alongside *NEFH* and *DCTN1* as affecting the cytoskeleton architecture and axonal transport in motor neurons.

Although the initial data on *PFN1* is extremely promising, studies since have had mixed results with positive (Ingre *et al.*, 2013) and negative publications (Zou *et al.*, 2013b; Daoud *et al.*, 2013; Lattante *et al.*, 2013a). The most attention has been focused on rare variants that arise in an abnormally high number of patients. It has since been shown that these variants increase a patient's probability of developing ALS (Dillen *et al.*, 2013; Tiloca *et al.*, 2013; van Blitterswijk *et al.*, 2013a; Smith *et al.*, 2015; Fratta *et al.*, 2014; Figure 25).

A study in 550 Chinese ALS patients and 545 matched controls detected the synonymous variant L112L as significantly associated with the disease (Chen *et al.*, 2013). Traditionally,

synonymous variants are disregarded when looking for a disease's cause but recent evidence has been challenging this. Studies in schizophrenia and autism show an enrichment of synonymous mutations in regions that likely affect splicing regulation (Takata *et al.*, 2016). The L112L variant is located at three amino acid residues from the start of an exon however it is not an evolutionary conserved position.

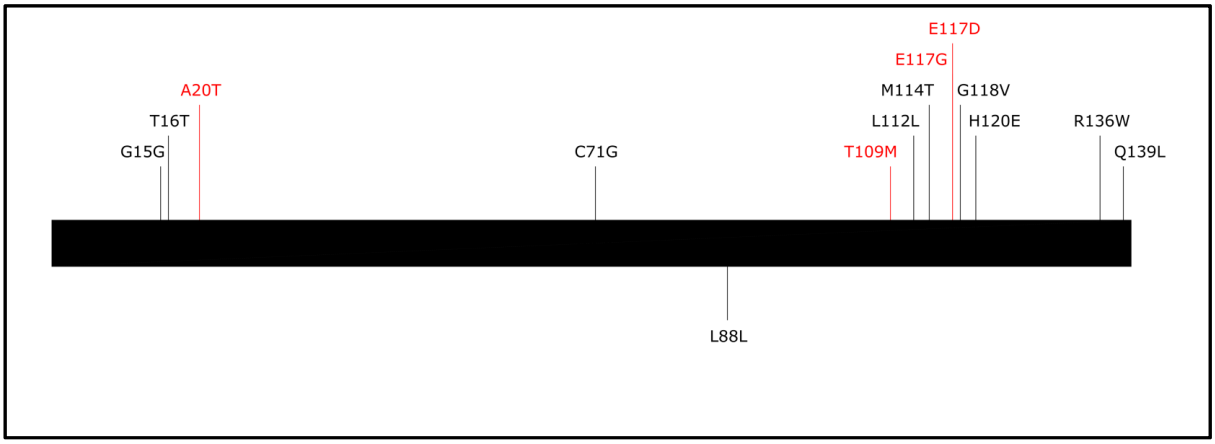

Figure 25. Mutations in *PFN1* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; black = unknown. Drawn in Inkscape v0.91.

1.2.22. *MATR3*

Another addition to the long list of ALS genes found recently is matrin 3 (*MATR3*). This gene was discovered by Bryan Traynor, and his team at the National Institutes of Health (NIH), to dominantly segregate in a family of ALS patients (Johnson *et al.*, 2014b). In light of this (at the time) unpublished information, we included this gene in our study. The full history and analysis of this gene is explored in Chapter 4.

1.2.23. *PON1-3*

The three paraoxonase genes (*PON1-3*) are enzymes involved in the hydrolysis of organophosphates, in other words, the breakdown of potentially damaging chemicals like pesticides and nerve agents. Several common polymorphisms in these genes display an altered protein ability at tackling these compounds. For example, Q192R affects protein function with each allele being more efficient at breaking down a different set of chemicals while L55M alters levels of PON1 expression as do promoter SNPs. The potential reason for

the relationship of this gene with ALS is illuminated by the fact that environmental toxins may conceivably increase one's chances of MND as observed in agricultural workers and hypothesised in Gulf War veterans (McGuire *et al.*, 1997; Haley, 2003; Kamel *et al.*, 2012). One study found that pesticide exposure was a risk factor for ALS but this was completely reliant on self-reported exposure (Morahan and Pamphlett, 2006)

These variants have been examined with respect to ALS producing varied results. C311S and Q192R were associated with ALS in a Polish cohort but the authors did not seem to correct for multiple testing, in which case the results were not significant (Slowik *et al.*, 2006). Next the intronic variants rs10487132 and rs11981433 in North Americans were correlated with ALS in trios (parents and an affected offspring) but not in unrelated case-control analyses (Saeed *et al.*, 2006). While in Irish populations the L55M variant combined with an intronic SNP was significant (Cronin *et al.*, 2007). Morahan *et al.* (2007) found that promoter SNPs which reduce expression were more common in ALS in Australia. In contrast, C311S and nearby SNPs were affiliated with ALS in France, Quebec and Sweden (Valdmanis *et al.*, 2008), and Landers *et al.* (2008b) found significance with the two intronic SNPs rs2074351 and rs705382 only. Negative studies have been published (Kasperaviciute *et al.*, 2007; Ricci *et al.*, 2011) and large meta-analyses have revealed negative results when including GWAS results (Wills *et al.*, 2009) or restricting the SNPs examined to only Q192R and L55M (Lee *et al.*, 2015).

Most candidate gene studies report positive results which are all slightly different from each other despite including the same regions, whereas meta-analyses are negative. This could either reflect the different populations each was tested in or could simply mean the initial studies are false positives owing to small samples sizes. The *PON* genes remain contentious on their involvement in ALS.

1.2.24. *VEGF*

Being the archetype angiogenic factor, vascular endothelial growth factor A (*VEGFA;* commonly abbreviated to *VEGF*) was one of the first proteins to be examined after perfusion deficits were recorded in neurodegenerative diseases. Whether this is a consequence of these disorders or whether it contributes to them is not entirely known (Rosenstein *et al.*, 2010). With

respect to ALS, interest in this gene first peaked when investigators attempted to study the effect of VEGF depletion. Since knockout models cause embryonic lethality, researchers created transgenic rodents with an altered *VEGF* gene (*VEGF*$^{a/a}$) to reduce protein expression while still allowing survival into adulthood. This lead to a pathological phenotype reminiscent of human ALS (Oosthuyse *et al.*, 2001). Furthermore, when VEGF is administered to *SOD1* mice models, their survival rate and motor function improves (Storkebaum *et al.*, 2004; Azzouz *et al.*, 2004). Deletion of the hypoxia response element of *VEGF* in mice is damaging (Oosthuyse *et al.*, 2001) but evidence of this effect in humans is lacking (Gros-Louis *et al.*, 2003; Van Vught *et al.*, 2005).

Positive outcomes in *VEGFA* are achieved only when common polymorphisms are examined but this is rarely replicated (Lambrechts *et al.*, 2003; Fernández-Santiago *et al.*, 2006).

1.2.25. *DAO*

D-amino-acid oxidase (*DAO*) was documented as dominantly segregating in a large ALS kindred with the pathogenic variant of R199W (Mitchell *et al.*, 2010; Figure 26). *DAO* is clearly involved in regulating D-serine which is the proposed cause of degeneration in mice models where the DAO activity has been inhibited (Sasabe *et al.*, 2012). Expression of this mutant by use of lentiviruses instigated the aggregation of proteins and cellular dysfunction (Paul and de Belleroche, 2012). Since this initial report, very few groups have published in *DAO* except to report negative results (Millecamps *et al.*, 2010a).

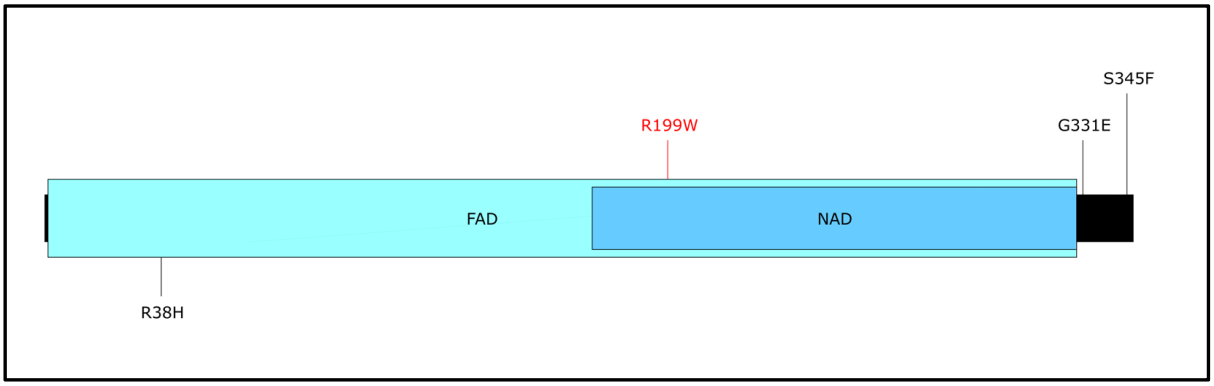

Figure 26. Mutations in *DAO* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; black = unknown; FAD = FAD dependent oxidoreductase; NAD = NAD-binding domain. Drawn in Inkscape v0.91.

As highlighted by studying the genetics of ALS in patients, cell and animal models, there are many molecular pathways implicated in this disease (Table 1). These mechanisms include: oxidative stress, protein aggregation, mitochondrial dysfunction, aberrant transcription and RNA processing, excitotoxicity, defective axonal transport, abnormal ubiquitin-proteasome system and endosomal trafficking, neuroinflammation, endoplasmic reticulum stress, defective glial cells and autophagy, some of which are overlapping pathways (Ferraiuolo *et al.*, 2011).

Mechanism	Associated genes	Details
Faulty protein processing	C9orf72, SOD1, TARDBP, FUS, FIG4, OPTN, SQSTM1, UBQLN1, VCP, PRPH, CHMP2B, VAPB	• Proteins misfold, ubiquitinate and aggregate within cells • Increased ER stress • Faulty endosome-lysosome and autophagosome • Increased apoptosis
Disturbed RNA metabolism and RNA-binding proteins	C9orf72, TARDBP, FUS, ANG, ATXN2, SETX	• Aberrant splicing and transport of RNA • Sequestration of RNA-binding proteins • Expression patterns altered
Cytoskeletal and axon/dendrite abnormalities	SOD1, DCTN1, NEFH, PRPH, SPG11, PFN1, TARDBP	• Impaired actin assembly • Build-up of neurofilaments • Axon retraction • Axon-transport disruption
Neuroinflammation and astrogliosis	SOD1, TREM2	• Hyperactivation of microglia • Excess inflammation
Dysfunctional vesicles	C9orf72, ALS2, CHMP2B, VAPB	• Reduced release of neurotransmitters • Faulty endosome-lysosome
Excitotoxicity	C9orf72, SOD1, FUS, DAO	• Excess glutamate • Reduced glutamate uptake
Defective mitochondria	SOD1, TARDBP	• Reduced energy supply • Increased apoptosis
Impaired DNA repair	FUS, SETX	• Reduced nuclear or mitochondria DNA repair • Increased oxidative stress may damage DNA

Table 1. Description of the main molecular mechanism known to be aberrant in ALS and the genes which can be linked which each, as covered in Chapters 1.2.1-1.2.25.

With each of these processes, it is challenging to determine if it is a cause of disease, an exacerbating downstream effect, a neuroprotective mechanism or the outcome of other processes. It is rare to begin studying a patient with ALS until the disease has already progressed to the point of presenting with symptoms. This then makes it harder to dissect cause from effect. Motor neurons are unique in their size, or more specifically, the length of their axons. Maintaining these long projections requires numerous mitochondria and a

functioning axonal transport system, which has been suggested to be the explanation of their vulnerability seen in ALS. Under the microscope, differences have been observed depending on the genetic status of the patient (Taylor *et al*., 2016; Table 2).

Subtype of ALS	Pathology
C9orf72-familial	Intranuclear RNA foci, SQSTM1-positive inclusions and different cells with TARDBP-positive inclusions
SOD1-famililal	SOD1-positive inclusions
FUS-familial	FUS-positive inclusions
Sporadic and other familial cases	TARDBP-positive inclusions

Table 2. Descriptions of the common pathologies underpinning the subtypes of ALS.

Impaired proteostasis is the most debated and arguably most important abnormal mechanism in ALS and a number of neurodegenerative diseases. Despite this, we still lack conclusive evidence to implicate it as an upstream causal mechanism. I believe that it is not a single molecular mechanism that is solely responsible for ALS but in fact many at the same time. Targeting a single mechanism with treatment will never be effective as a cure due to other pathological pathways still proceeding. It is of high importance to understand all of these convoluted and overlapping pathways as they advance throughout the course of the disease. As with the divergent pathologies, it is also likely that the different genetic subtypes of ALS also have an overlapping but distinct group of molecular mechanisms that underpin each.

1.4. JUSTIFICATION FOR CURRENT STUDY

As has been illuminated above, the genetics of ALS is highly complex and multifaceted. Most studies so far limit themselves on both the number of patients they include and the number of genes sequenced. Genetics is important for understanding diseases like ALS because, as of yet, treatments are lacking and clinical trials are failing. Understanding the full genetics of a disease supports our knowledge on the underlying mechanisms involved. The key here is the plurality of 'mechanism' which results in multiple pathways a treatment must potentially target in order to be efficacious. This complexity is likely to be at least part of the reason for the low success rate of clinical trials. Secondly, a genetic diagnosis of ALS could occur in the presymptomatic phase of the disease and enable potential treatments to be applied at a much earlier stage. This creates a larger window to identify therapies with a positive effect instead of only utilising the drugs once a significant amount of cell death has already occurred. Lastly, ALS is an incredibly short disease meaning rapid diagnosis could make all the difference to getting a patient the right treatment.

Previously, an oligogenic hypothesis of ALS has been proposed and preliminarily confirmed by statistics looking at the effect of rare mutation burden on disease (Cady *et al.*, 2015; van Blitterswijk *et al.*, 2012b). If a significant number of ALS cases are indeed caused by more than one risk variant, then genetic counselling will need to be corrected.

When a heritable element is suspected in any neurodegenerative disease, genetic diagnosis lies with sequential Sanger sequencing. However, with an array of multiple genes causing each disease and, additionally, numerous alterations within each gene being potentially harmful, it can be time consuming and costly to diagnose a patient suspected of harbouring a detrimental genetic variation. Furthermore, the range of genetic tests at each institution can be limited. It is now plausible that next-generation sequencing (NGS) technologies will eliminate many of the issues with traditional Sanger sequencing. To test this possibility, we have developed a single comprehensive assay containing a number of genes which have, to varying degrees, been implicated in ALS.

The aims for my project include:

1. To investigate more than a thousand cases of ALS (we have sequenced 1,235)
2. To gather a comprehensive genetic dataset on these cases including all genes of major and minor importance
3. To determine the important pathogenic and risk variants
4. To determine the benign variation within ALS genes
5. To investigate a potential oligogenic basis of ALS
6. To explore novel ALS genes which are published during my project
7. To develop a list of mutations which require additional functional studies to confirm their pathogenicity
8. To explore the possibility of a genetic pattern which is associated with ALS by use of a new method involving machine learning techniques
9. To make judgements on the contribution of each gene studied to ALS
10. To contribute to the field of the genetic landscape of ALS

I have been given a huge cohort of more than a thousand ALS cases which was made possible by the MND Association. This charity is involved in the funding of many projects like the current one with the sole objective of discovering a cure for MND. In order to analyse these samples adequately, the bulk of this project will implement next-generation sequencing to extrapolate information on the genetic landscape of ALS. This fast, comprehensive technique will allow us to incorporate many ALS genes into the study. Using this data I will be able to address many of the aims laid out for this project. I will also complete additional projects of rapid sequencing in novel ALS genes and *C9orf72*.

I hypothesise that the genetics of ALS may be more complex than we have previously thought, due to the techniques used and the many single-gene studies completed. With the numbers of patients we have gathered, we should be able to confirm any potential oligogenic basis of ALS. We additionally theorise that we will be able to discredit some genes (and specific mutations) as not being involved in ALS outside the initial families they were discovered in.

CHAPTER 2 METHODS AND MATERIALS

Methods described here are relevant to more than one chapter while unique methods are described in the individual chapters.

2.1. PATIENTS

A total of 1,417 DNA samples were kindly provided by the MNDA DNA bank. This comprised of 51 controls, 111 fALS and 1,220 sALS. A further 547 samples were obtained from University College London and Partners (UCLP) MND clinics which were composed of 452 ALS and 95 FTD. Additionally, I was given 510 in-house controls which have whole-exome sequencing data completed on them as part of the IPDGC consortium (Simon-Sanchez *et al.*, 2009), and 85 in-house controls to run in my own study, 33 of which overlapped between the two groups. Controls were of Caucasian ethnicity, were over 60 years old and free from any neurological disorder. Additionally, subjects were excluded from the study if they had a first-degree relative with a neurological disorder including Alzheimer's (AD), ALS, ataxia, autism, bipolar disorder, cerebrovascular disease, dementia, dystonia, Parkinson's (PD) and schizophrenia. Lastly, I was also given 26 Argentinian samples with ALS and FTD. All patients provided written consent prior to the study.

2.2. DNA QUANTIFICATION

To ensure all samples were of adequate concentration for sequencing, all 1,502 were run on the Qubit 2.0 fluorometer (Life Technologies, UK). This machine detects the amount of fluorescent dye bound to DNA (or RNA), therefore providing a direct measurement of DNA quantity. The Qubit reagent does not bind to degraded DNA and is considered to be more accurate than the competing popular DNA measuring device: the NanoDrop, which can also measure contamination, for example, from protein. All samples were diluted to a concentration of 50ng/ml.

2.3. SANGER SEQUENCING

2.3.1. PRIMER DESIGN

The Primer3 (v0.4.0) online software was used to design all oligonucleotide primers for Sanger sequencing ensuring they cover the exons and flanking UTRs of the gene (Untergasser *et al.*, 2012; http://bioinfo.ut.ee/primer3-0.4.0). Input regions were obtained from the reference sequence provided on Ensembl website. The resulting primer sequences were checked using the In-Silico PCR tool on the UCSC website to ensure they were specific for the desired location (https://genome.ucsc.edu/cgi-bin/hgPcr).

2.3.2. PCR

The Nobel Prize winning invention of polymerase chain reaction (PCR) is a core technique in every genetics laboratory. This *in vitro* method amplifies chosen regions of denatured DNA using short DNA sequences (primers) that are complementary to the target binding areas. Thermal cycling enables the denaturing of double-stranded DNA, annealing of the primers, followed by exponential replication using the enzyme DNA polymerase as each new product adds to the collection of replicable templates. The four deoxynucleotides (dNTPs; aka DNA precursors) are required for this experiment: dGTP, dCTP, dATP and dTTP as well as GoTaq DNA Polymerase and GoTaq Reaction Buffer, all obtained from Promega (UK). Some reactions required dimethyl sulfoxide (DSMO) or Betaine solution (both Sigma, UK). Samples were run on an automatic Eppendorf Mastercycler Pro S. All primers were optimised on control DNA to ensure the best experimental conditions, these are described separately in Chapters 4 and 5.

2.3.3. GEL ELECTROPHORESIS

To determine if the PCR was successful, samples were run on a 3% agarose gel. This technique provides a porous gel in which DNA fragments can move. When an electrical current is applied across the gel, the negatively charged DNA will travel at a rate relative to its size. Given that the size of the PCR products is known, this test enables a fast detection of their presence in the sample. To create a gel, agarose powder (Sigma, UK) is heated and dissolved in TBE buffer (see Table 3 for components) with GelRed (Biotium, US). Once the gel has set,

it is placed into the electrophoresis tank in TBE buffer. After DNA is loaded into the gel alongside a 100bp ladder (Life Technologies, UK) a 120V current was applied for 30 minutes using PowerPac HC v1.07 (Bio-Rad, UK). The samples were then visualised using a UV transilluminator GelDoc-It Imaging System (Ultra-Violet Products Ltd, UK).

Reagent	Quantity	Company
Trizma base	12.11g	Sigma, UK
Boric acid	6.18g	Sigma, UK
Ethylenediaminetetraacetic acid (EDTA)	0.74g	Sigma, UK
Nuclease-free water	Make up to final volume of 1L	Qiagen, UK

Table 3. Recipe for 1x TBE (tris/borate/EDTA) buffer, reagents are dissolved in the order presented in this table.

2.3.4. PCR PURIFICATION

Enzymatic clean-up of PCR products was achieved by use of Fast-AP which deals with any abundance of dNTPs and Exonuclease I which targets single-stranded DNA for degradation (both Life Technologies, UK; Figure 27).

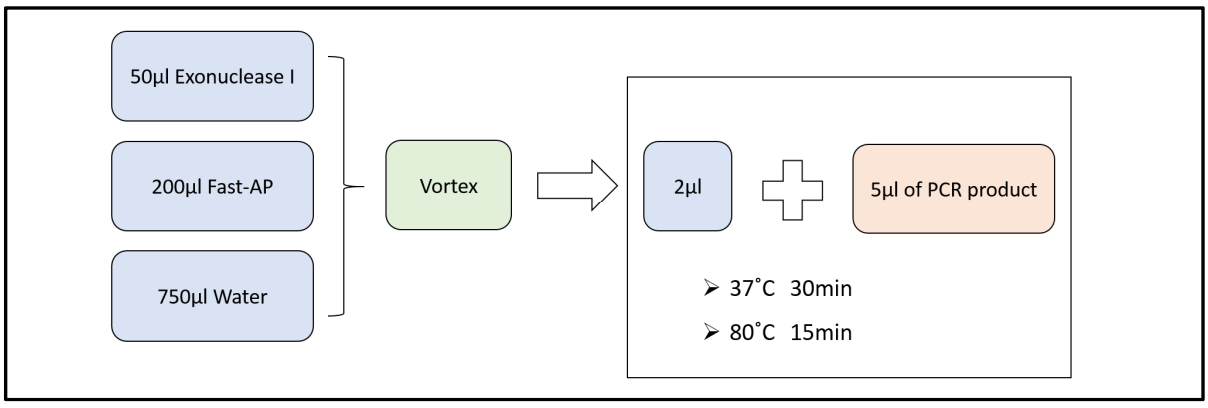

Figure 27. PCR purification protocol.

2.3.5. BIGDYE

The sequencing reaction was achieved by use of BigDye Terminator v1.3 Cycle sequencing kit and 5X sequencing buffer (Applied Biosystems, USA) following the method in Figure 28.

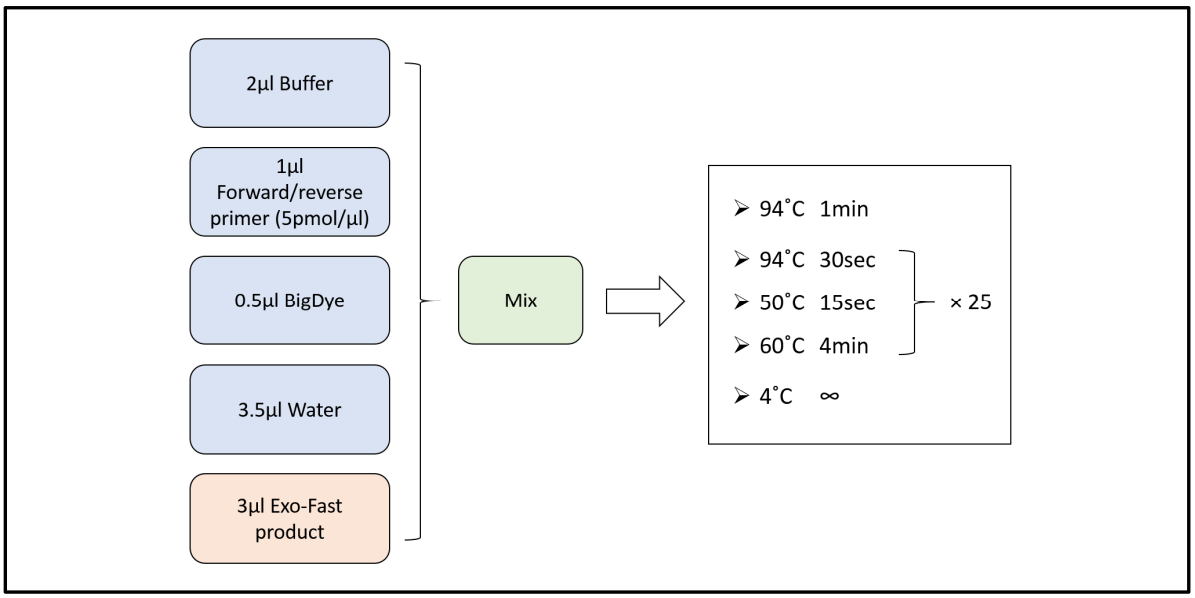

Figure 28. BigDye protocol.

2.3.6. CLEAN-UP

The clean-up step for the sequencing reaction is completed with Sephadex G-50 Bioreagent (Sigma, UK) and a FiltrEX 96 well filter plate (Corning, US) as per Figure 29.

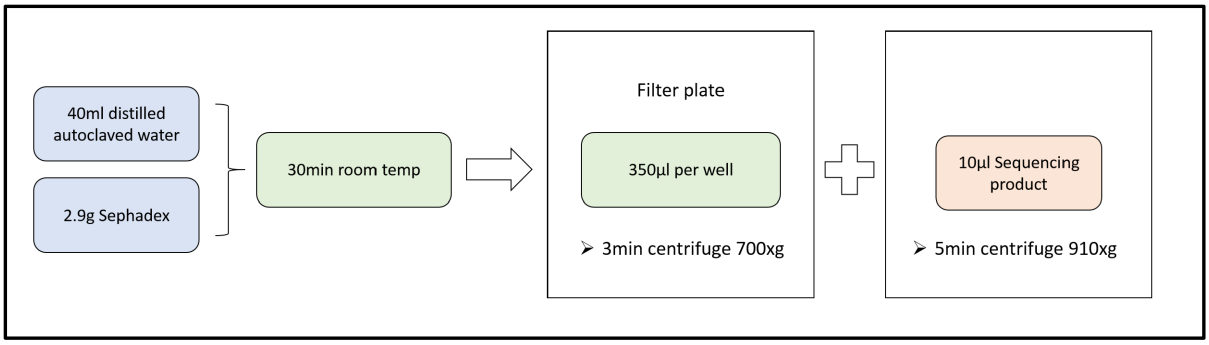

Figure 29. PCR clean-up protocol.

2.3.7. SEQUENCING

The final product is run on an ABI 3730xl DNA analyser (Applied Biosystems, USA). Traditionally (pre-1986), this system performed PCR amplification of the DNA molecule, dividing the sample into four reactions (Figure 30). Each of these have all four dNTPs while additionally containing one extra base in the form of a dideoxynucleotide (ddNTP). This modified base prevents any other dNTPs from being added. This results in many fragments of the DNA at different lengths with a tagged nucleotide on the end. Capillary electrophoresis then separates out all the fragments allowing a one base pair resolution of the product which was manually read to obtain the DNA sequence. In 1986, the ddNTPs were fluorescently labelled which meant the mixture could stay as a single sequencing reaction rather than four (Karger and Guttman, 2009). Sequencing reading could then be automated and generates the classic chromatogram which is then examined for mutations by means of the software Sequencher (v4.1.4.).

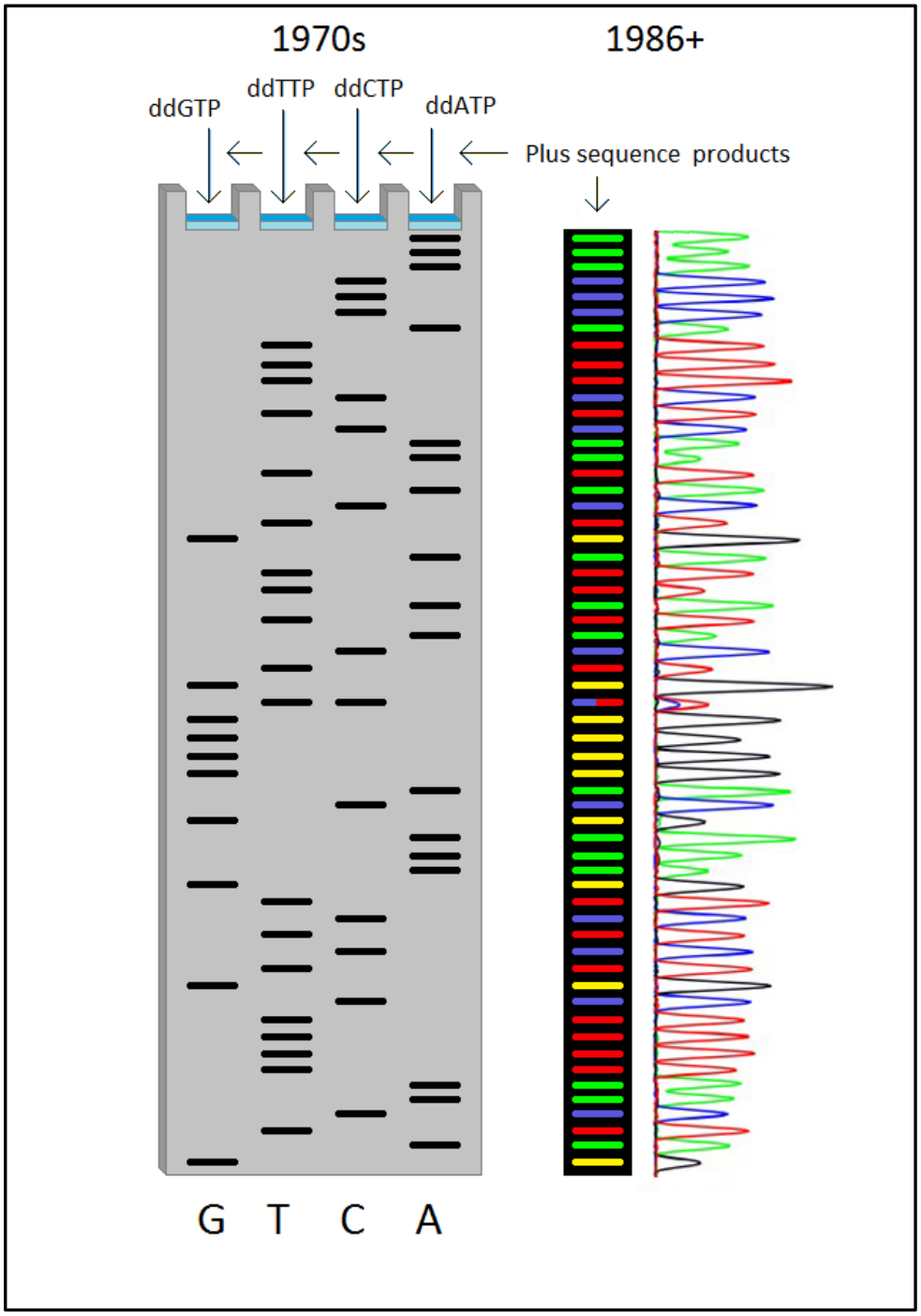

Figure 30. Capillary electrophoresis used on fragments of product which are labelled with ddNTPs and run in four separate columns (left; 1970s) or in a single column (right; 1986+). The latter of these is automatically detected forming the chromatogram on the right. Drawn in Paint v1511.

2.4. NEXT-GENERATION SEQUENCING

2.4.1. OVERVIEW

The creation of NGS has exponentially increased the amount of DNA and number of patients we can study. To compare, running a single 96-well plate using the traditional Sanger sequencing method determines on average 300-800 base pairs in the 96 patients whereas NGS can sequence thousands, millions or billions of base pairs in a single run. This unprecedented high-throughput has enabled us to study the genetics of human traits and disease in a high-dimensional fashion never before possible. We can now ask and answer more complex genetic questions to understand the underlying biology. When we want to assess the amount of DNA adequately read by NGS, we measure the *coverage* or *read depth*, which both describe how well each nucleotide was evaluated i.e. because the technology can produce errors, we check each nucleotide multiple times so these terms simply refer to the number of times the DNA was read. This is often described as, for example, 10x which means the average depth across the desired area was 10 reads.

The mechanics underlying NGS developed by Solexa/Illumina are based on sequencing by synthesis (SBS) technology where fluorescently-labelled nucleotides bind to the DNA emitting a light which can be detected at each base along the chain. Each nucleotide will discharge a different colour which is imaged and then the dye is removed (Figure 31). This process is completed in a massively parallel way with thousands of DNA amplicons being read simultaneously. The collection of DNA molecules created for sequencing is called a *library*.

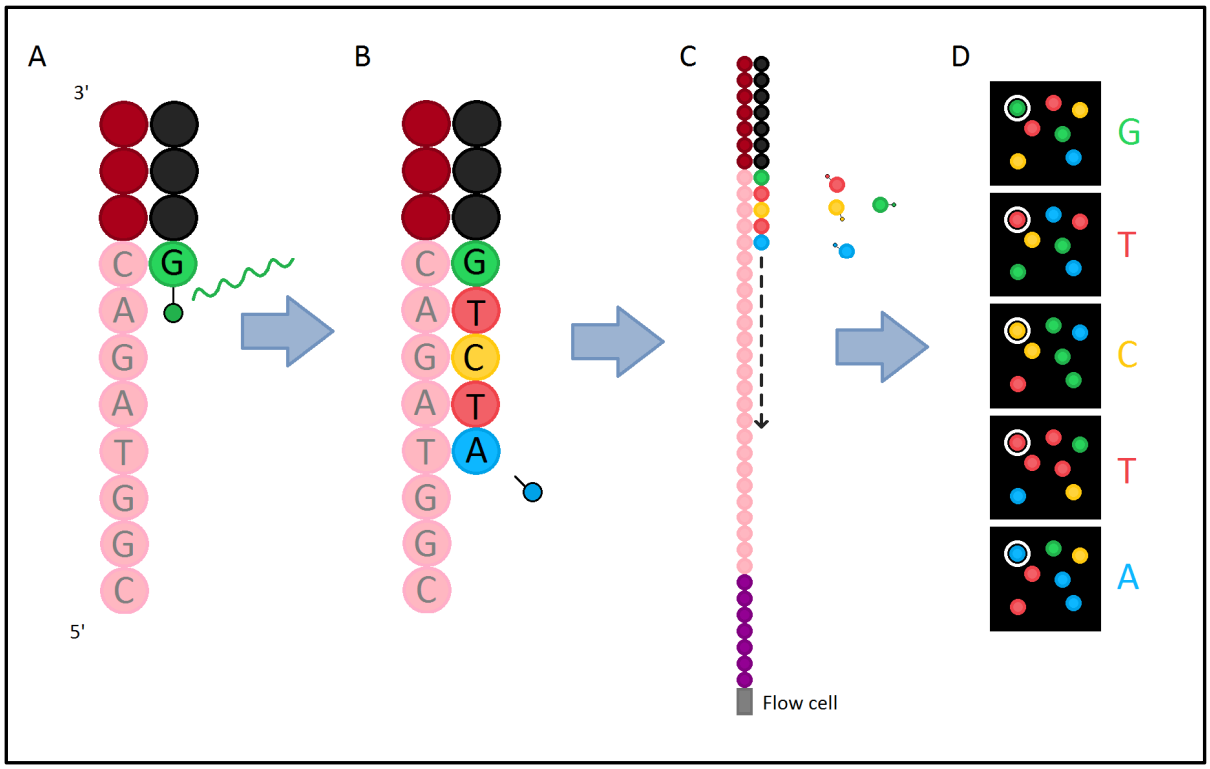

Figure 31. Diagram of Illumina's sequencing by synthesis method. (A) A newly attached nucleotide both stops the addition of new bases and fluoresces which is detected by an optical sensor. (B) The chain is able to grow as the fluorescence is cleaved and washed away. (C) The four fluorescently labelled bases compete for the next spot in the chain. (D) Imaging from above reveals the locations of each amplicon on the flow cell and each nucleotide is read sequentially to reveal the nucleotide sequence. Each coloured dot actually represents thousands of the same amplicon, termed a cluster, described in more detail in the next chapter. Drawn in Paint v1511.

2.4.2. TRUSEQ CUSTOM AMPLICON PROTOCOL

The NGS method I utilised in this study is the TruSeq Custom Amplicon (TSCA) assay v1.5 (Illumina, UK). This technique utilises PCR amplicon-based target enrichment and enables rapid design and sequencing with huge flexibility in selecting the areas of the genome to be covered. The content of the custom design can include up to 652,800 base pairs. Oligo probes were designed using DesignStudio v1.6 online which provides metrics on quality of the designed oligonucleotides (http://www.illumina.com/applications/designstudio.ilmn). Figure 32 displays an example of the DNA amplicons produced by custom oligo pairs in the *FUS* gene.

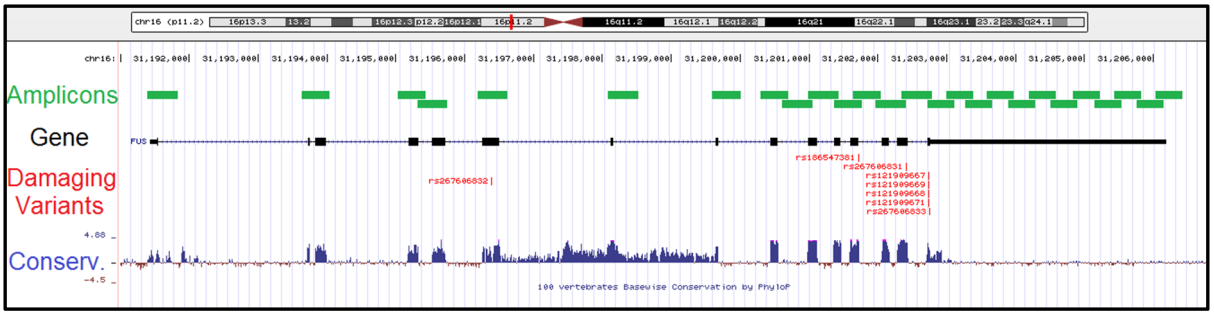

Figure 32. Amplicons (green) are designed to adequately cover important areas of the gene *FUS* i.e. reported pathogenic variants (red) and all exons (black boxes) and UTRs (thinner black boxes at the ends of the gene). Conservation across different species (blue) show that exons are often highly conserved as gauged by PhyloP. Modified from the UCSC website and adapted in Paint v1511.

To begin the experiment, the designed probes are incubated with each DNA sample to enable hybridisation of the custom oligos. These two probes bind to the same strand of the DNA to enable greater specificity than other PCR-based amplicon techniques (Figure 33). Unlike Sanger sequencing, hundreds of primers can be added to the same well containing one DNA sample.

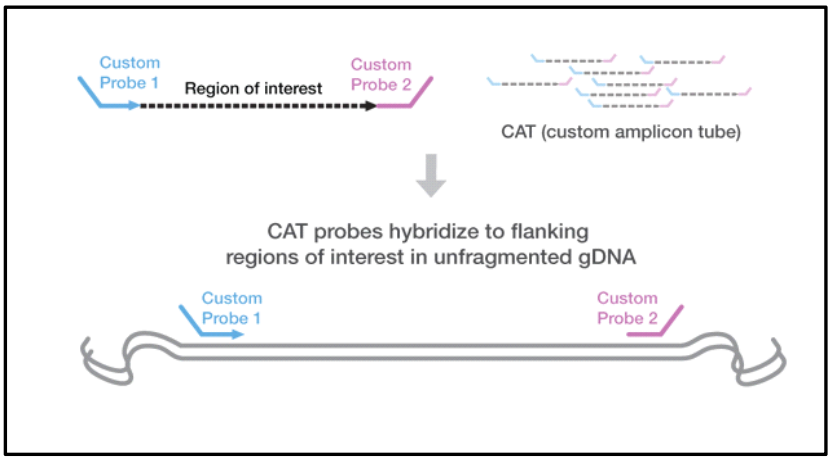

Figure 33. Custom probes are designed to cover the regions of interest. The probes bind to genomic DNA. (Illumina 2015).

Next the samples are washed using a filter plate to remove any oligo primers which have not bound to the DNA. Next the extension-ligation step bridges the two probes using DNA polymerase to create copies of the target region. This is flanked by sequences which aid PCR amplification in the next step. Here, each patient is barcoded with a unique section of genetic code enabling massively parallel sequencing as multiple patients can be pooled in the same mixture (Figure 34).

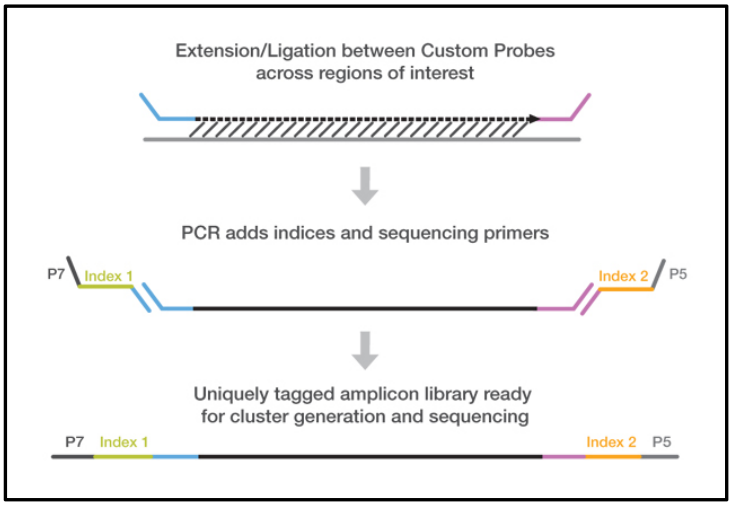

Figure 34. The extension and ligation step copies the DNA while then the PCR step allows for the addition of unique DNA to identify each patient and allow for binding to the flow cell. (Illumina 2015).

After PCR clean-up, library normalisation is achieved by essentially the same method: using magnetic beads at a set concentration which bind to a limited amount of DNA while allowing the wash-off of all other products. Lastly, 96 samples are then pooled together and denatured for use on the MiSeq platform (Illumina, UK).

On the MiSeq instrument, a flow cell is used – a glass slide with DNA oligos protruding from the surface. These oligos bind specifically to the adapters on the amplicons which undergo bridge amplification and cluster generation (Figure 35). This describes the process of an amplicon bending over and attaching at both ends to the flow cell. A copy of this amplicon is then created and both amplicons with make new bridges for further amplification. The resulting mass of amplicons is called a *cluster*. From a single DNA molecule, 100-200 million clusters of identical molecules are created.

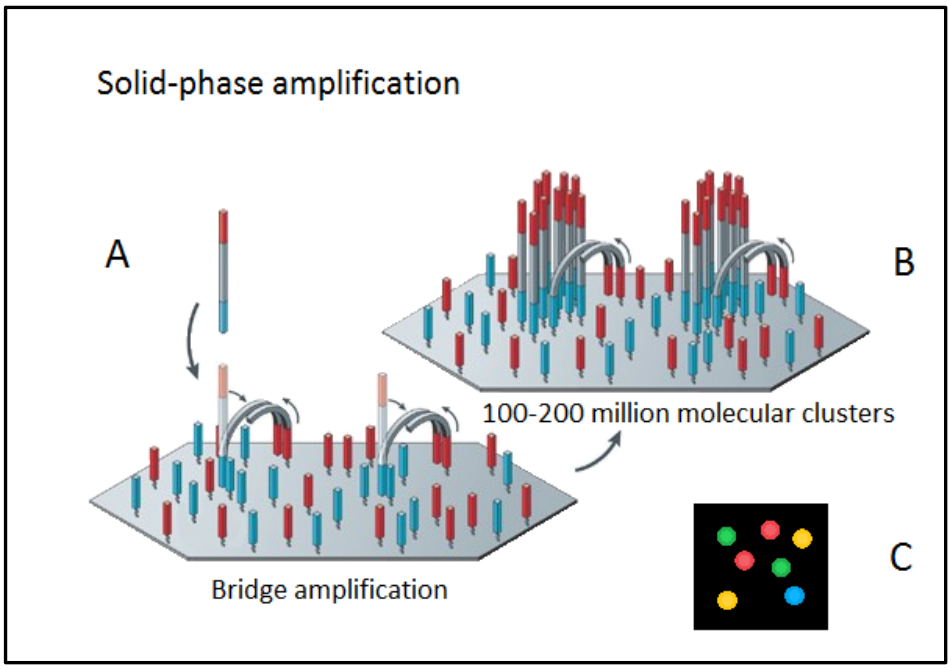

Figure 35. (A) One DNA molecule (amplicon) binds to the flow cell and undergoes bridge amplification. (B) Millions of copies are produced (cluster growth). (C) Clusters are vital because they increase the fluorescent signal making it easier to detect. The actual images in the machine have millions of these clusters. (Modified from Metzker, 2010).

Illumina machines also employ paired-end sequencing (Figure 36) meaning they sequence every amplicon from both ends. This improves the data quality and makes it easier to align to the reference genome.

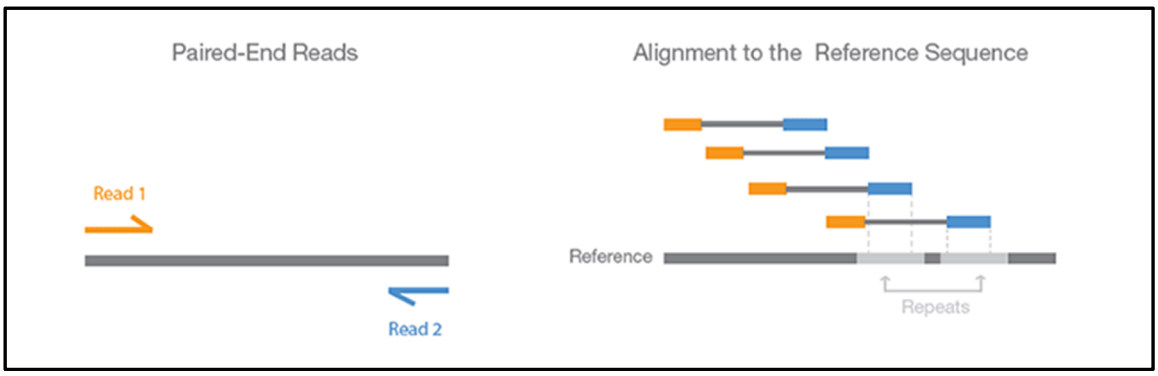

Figure 36. Sequencing from both ends of the DNA can help combat repetitive regions which are notoriously difficult to sequence and align to the genome. (Illumina, 2015).

Although NGS has rapidly increased the genetic information we can now gather, this doesn't come without drawbacks. We can now produce up to 1.8TB in a single run on the latest technology (HiSeq X) which presents as a huge task to store, analyse and interpret efficiently. The required infrastructure for this task is expensive due to the sequencing machines and

supercomputing as well as the personnel expertise. Also, some might argue we produce more data then we know what to do with, as most rare variants end up in the box of *unknown significance*. Additionally, false positive calls are present in the data and although we have computational techniques for removing them, confirmation with Sanger sequencing is still often performed as the gold standard. Lastly, NGS remains inexpensive only with large quantities of patients and is usually run in batches of 95 samples. Our study cost approximately £0.0004 per base pair. If NGS gene panels are to be clinically useful, patients with rarer diseases might need to wait longer for enough other individuals to also be present.

2.5. BIOINFORMATICS

2.5.1. OVERVIEW

Generating huge quantities of genetic data requires a sophisticated chain of data processing elements, also known as a *pipeline*. That's not to say each step occurs one after the other, instead, sections of a pipeline will often be computed in parallel to speed up an otherwise lengthy procedure. Pipelines are constantly updated and adapted to each project they are required for but genetic pipelines generally follow the same chain of events: process the reads to form a single genome (usually by aligning to a reference), report any variation detected in the reads (termed variant *calling*), remove low quality data (QC) and annotate the remaining variants with a range of information. For my data, I designed an array of scripts to manipulate the data and analyse it appropriately.

2.5.2. METHOD

The MiSeq has its own reporting software (MiSeq Reporter v2.5) which produces the raw genetic data in the form of FASTQ files. It then aligns the reads to the designed input genomic areas and calls variants using GATK. However, we additionally realign these FASTQ files to the whole reference genome as there may be off-target reads which would more appropriately map to other regions of the DNA (Figure 37). We use Novoalign v3 to perform this task using the reference human genome (UCSC hg19).

Variants were extracted using the Maq model in SAMtools v0.1.18 and filtered by the following criteria:

- Consensus quality > 30
- SNP quality > 30
- Root mean square mapping quality > 30

SAMtools calculated quality scores for each variant. I recalled variants using UnifiedGenotyper from GATK v3.5 (genome analysis toolkit) which is able to jointly call variants across multiple samples to combine them into a single file. Calls from the different methods were examined in parallel to ensure all variation was captured. These variants were curated by use of ANNOVAR (Nov2014; Wang *et al.*, 2010), one of the leading programmes in the annotation of genetic data. This software provides details about each detected variant in terms of position, associated gene, exonic function, amino acid change, frequency in control databases (ExAC, 1000 genomes, ESP6500 and cg69), presence in ClinVar and predictions from the algorithms: PhyloP, SIFT, PolyPhen2, LRT, MutationTaster, GERP and CADD. These are discussed further below (Tables 4 and 5). I then filtered variants in each sample, removing:

- Artefacts common to multiple wells that are either not predicted to be so frequent or appear to be a batch effect i.e. only present in one plate and commonly so (likely false positives)
- Artefacts with a low read balance (likely false positive, Figure 40)
- Artefacts with a low read depth (low quality)

For the most part, we are not interested in synonymous or common variants (MAF >1% of the Caucasian population). Unless stated otherwise, these were removed from the data. I will use the term *rare* to describe any variant(s) below this frequency. The software PLINK (v1.9) is a publically available resource for analysing genetic data (Purcell *et al.*, 2007). It can perform tests to ensure that variants conform to the Hardy-Weinberg principle, that samples are not related or duplicated and lastly it can perform a chi-square test to examine phenotype correlations and frequencies of specific SNPs in cases verses controls.

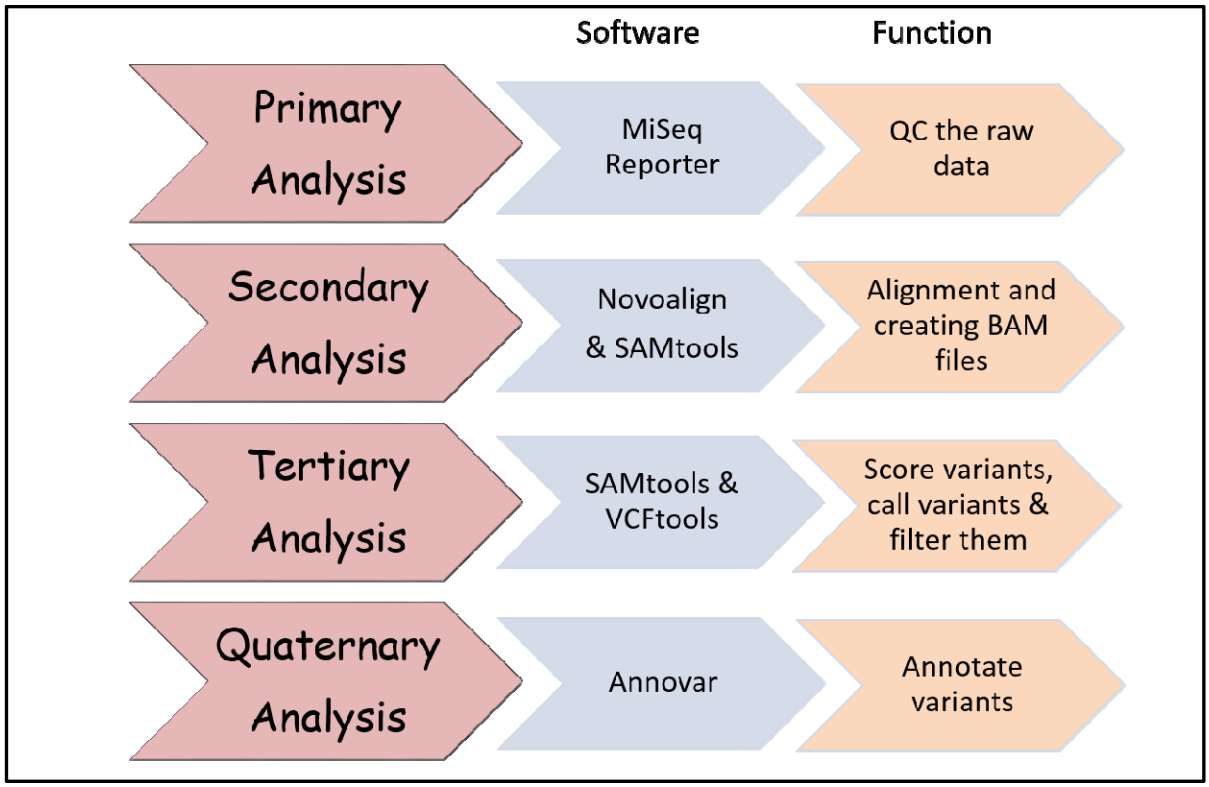

Figure 37. A general overview of the pipeline implemented in this study. Drawn in Microsoft PowerPoint 2013.

The coverage of the panel was estimated using CovCheck, an in-house analysis software which calculates the number of reads each region of interest is covered by using BedTools (v2.10.1).

As an extra quality check, I also build an automated program to visually analyse variants within GenomeBrowse (v2.1.0. Golden Helix). I wanted to get a feel for the data and perform visual checks on variants with medium-low quality scores. Additionally a number of the controls were sequenced on a different platform so it was vital to remove all false positives as they might introduce false differences between the two groups. An example of two good quality variants are seen in Figure 38 and poor quality variants in Figure 39 and 40. I also took at least one picture of every nucleotide with a reported variant. This resulted in 40,000+ pictures being examined by eye to guarantee high-quality data.

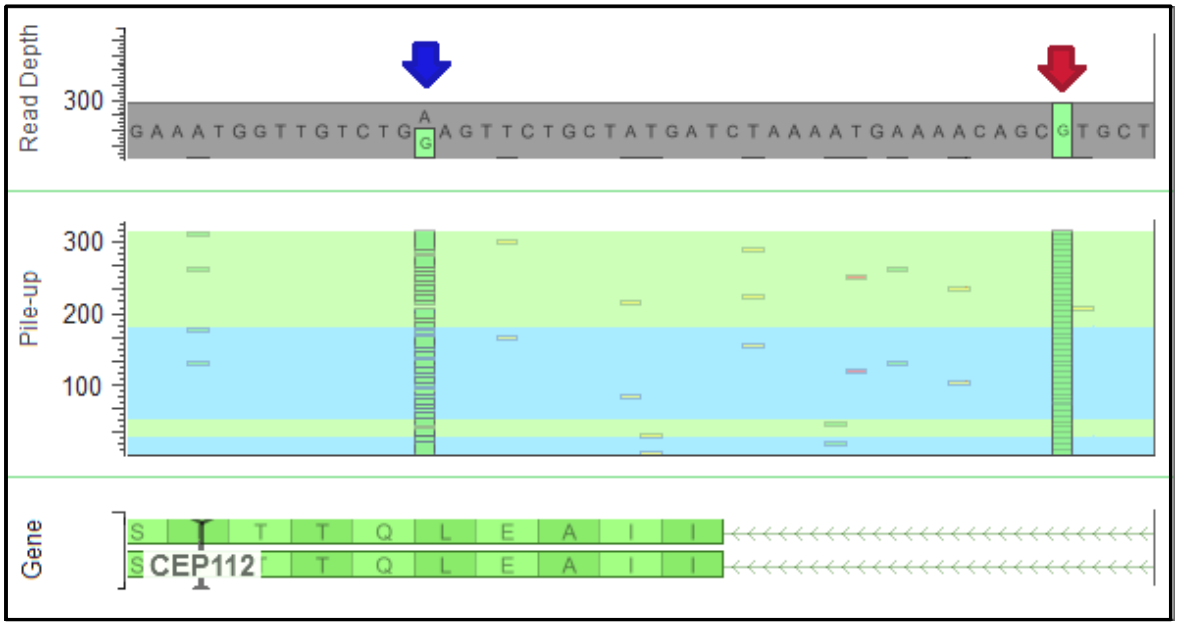

Figure 38. Example of a good quality heterozygous (blue arrow) and homozygous (red arrow) variant in the CEP112 gene. The former is within the exon while the latter is intronic. Both variants are captured in the forward and reverse sequencing (green and blue bars in the pile-up section) and are covered by 300 reads which gives them high quality scores. Grey nucleotides in the read depth section represent the read call being identical to the human reference (h19).

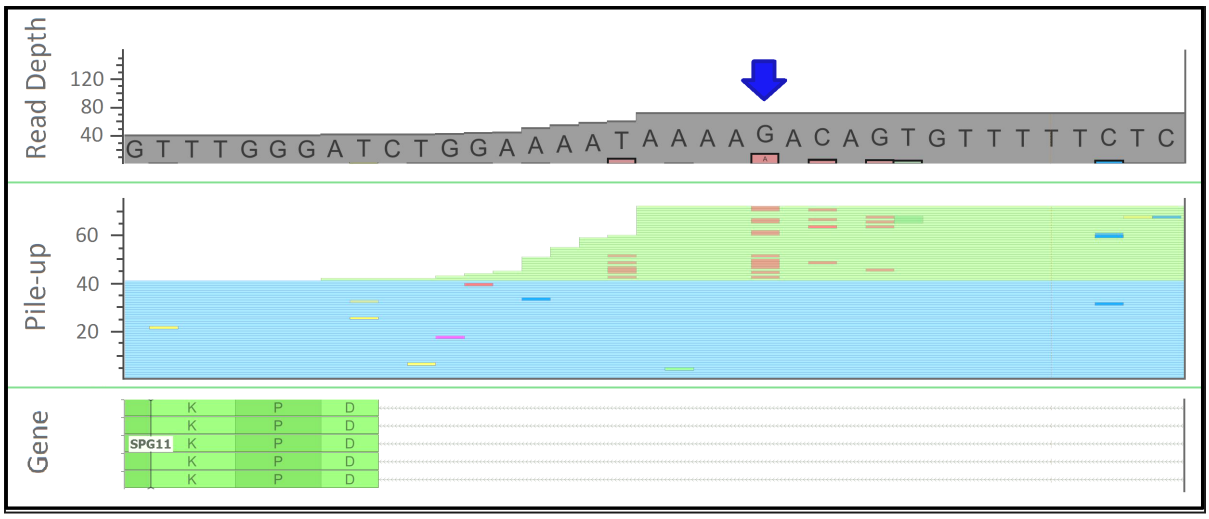

Figure 39. Example of a poor quality call which is almost certainly a false positive (blue arrow). The call is only present in the reverse strand (green bars in the pile-up section). There are also several other false positives in close proximity which is another clue to a variant being false.

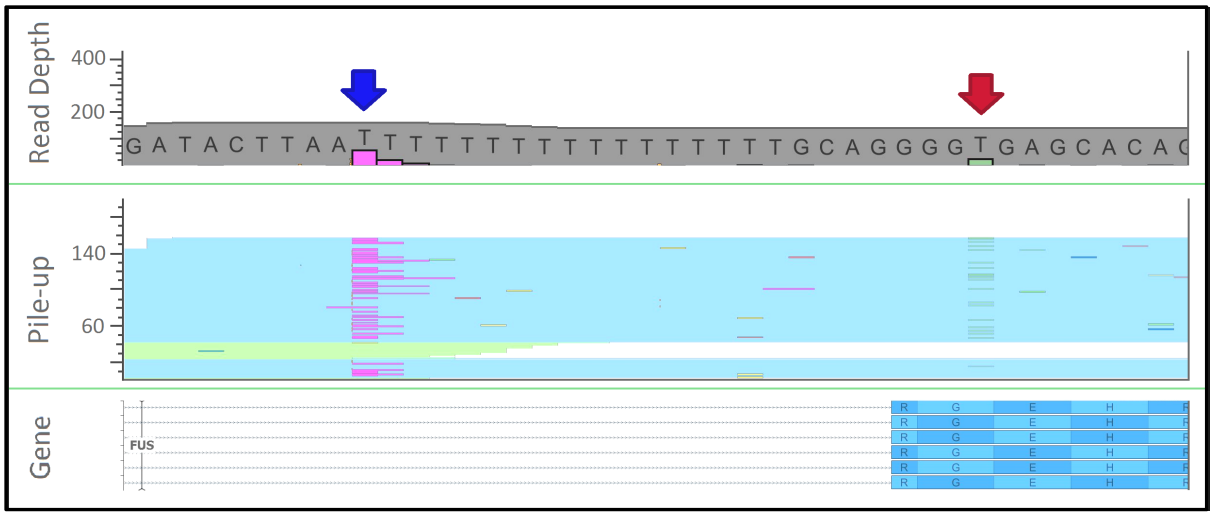

Figure 40. Example of two poor quality variants. The first is a deletion which is likely due to the mononucleotide repeat present in the DNA which is known to cause sequencing problems (blue arrow). The second is a variant with low read balance, only 14% of the reads are called as the alternative allele which means it is unlikely to be real (red arrow).

2.5.3. RESOURCES

A huge number of resources and tools are required for adequate analysis of genetic data. I have listed those that I have been heavily using during my PhD in Table 4 and 5.

Database	Website	Description
Exome Aggregation Consortium (ExAC)	http://exac.broadinstitute.org/	Exome data on 60,706 unrelated individuals sequenced as part of research projects. Includes disease cohorts.
Exome Variant Server (EVS6500)	http://evs.gs.washington.edu/EVS	Contains exome data on 6,503 unrelated subjects which is composed of 2,203 African-Americans and 4,300 European-Americans.
1000 Genomes Project (1000g)	http://browser.1000genomes.org	Has wider diversity with 2,504 individuals from 26 populations, however some of the data is WGS with only 2-4X coverage and hence contains some lower-quality data and some related individuals.
UK10K	http://www.uk10k.org/	WGS from control populations, families and disease cohorts (neurodevelopmental, obesity and rare disease). Holds data on 10,000 subjects.
Complete Genomics (CG69)	http://www.completegenomics.com/public-data/69-genomes/	Contains data from 69 individuals who are genetically diverse, some are related.
The Single Nucleotide Polymorphism Database (dbSNP)	http://www.ncbi.nlm.nih.gov/snp	Collection from many sources of data on short genetic variations (typically ≤50 bp). May contain pathogenic variants. Release 129 is pre-NGS and so is considered to contain no NGS-specific false positives compared to 137.
Ensembl	http://www.ensembl.org/	Genome database with list of transcripts for each gene and their sequence.
Online Mendelian Inheritance in Man (OMIM)	http://www.omim.org	Provides a list of genes that have an association a Mendelian disorder. Also covers a small number of mutations in these genes known to be causal.
UCSC	http://genome.ucsc.edu/	An online tool for visualising the genome and a huge number of annotations from many sources. Also provides BLAT for searching for sequences in the genome and In-Silico PCR for examining primer pairs.
GeneCards	http://www.genecards.org/	Database on genes and a number of annotations from multiple sources.
Human Gene Mutation Database (HGMD)	http://www.hgmd.org	A more comprehensive list of variants associated with disease but is not free to access.
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar	Database linking variants to any clinical significance discovered. However the archive is considerably smaller than the published literature.
ALSoD	alsod.iop.kcl.ac.uk/	Database of mutations reported in ALS.
ALSgene	http://alsgene.org/	Database of mutations reported in ALS.
Pubmed	www.ncbi.nlm.nih.gov/pubmed	Database of scientific publications
Google Scholar	scholar.google.co.uk/	Interface for searching for publications

Table 4. List of websites and databases used in this project. Control databases can contain both pathogenic variants and related individuals and so are not considered to be perfect. This and the presence of low quality calls might skew variant frequencies.

Name	Website	Description
SIFT	http://sift.jcvi.org	The effect of amino acid change on protein function. Uses closely related sequences and measures variation within these areas to judge novel variants.
MutationTaster	http://www.mutationtaster.org	Examines the effect on protein structure and function adding in evolutionary conservation.
PolyPhen2	http://genetics.bwh.harvard.edu/pph2	Bases predictions on a missense amino acid change and how common this change has occurred in nature. Also determines if the site is hypermutable.
PhyloP	http://compgen.bscb.cornell.edu/phast/help-pages/phyloP.txt	Conservation or acceleration p-values stemming from the expected rate of evolution.
CADD	http://cadd.gs.washington.edu	Integrates multiple annotations including regulatory effects. Works on SNPs, indels and non-coding variants
GERP	http://mendel.stanford.edu/sidowlab/downloads/gerp/index.html	Uses evolutionally rate profiling i.e. looks for substitution deficits which imply constraint due to functional importance.
LRT	http://annovar.openbioinformatics.org/	Likelihood ratio test based on functional disruption at highly conserved, coding residues.

Table 5. List of *in silico* tools used in this thesis.

To aid variant interpretation with respect to disease, an array of *in silico* tools have been designed to make predictions about their pathogenicity (Table 5). The underlying mechanics of these algorithms often overlap and so the American College of Medical Genetics and Genomics (ACMG) Standards and Guidelines state that they should be used collectively as a single piece of evidence (Richards *et al.*, 2015). This group is composed of experts in the field of genetics and attended multiple workshops to achieve a common consensus on how the scientific community should report and interpret genetic variants. It was decided that *in silico* tools should only be used when they all come to an agreement on the prediction of either benign or pathogenic. If there is a split between the predictions, then they should not be used to infer any information about the mutation. Assessing the ability of these tools show that known pathogenic variants are predicted as such 65-80% of the time (Tavtigian *et al.*, 2008; Hicks *et al.*, 2011; Thusberg *et al.*, 2011; Thompson *et al.*, 2013). However the specificity was much lower and non-pathogenic mutations are also predicted to be pathogenic. Splice site predictions perform better on both sensitivity and specificity: 90-100% and 60-80% respectively (Vreeswijk *et al.*, 2009; Houdayer *et al.*, 2012). For these reasons, it was decided by the ACMG workshop that these tools should be strictly applied to genes of known relevance

to a disease. Deleterious (for the protein) variants will exist in other genes but this may have no relevance for the disease in question.

Additionally, Richards *et al*. (2015) say that the classification terminology of variants should be standardised so that we describe alterations as being in one of five categories (Table 6). The authors do note that this will greatly increase the number of variants graded as *uncertain significance* however there will be fewer false positives in the literature. The only point I would contest in this list is 'observed in trans/cis with a dominant variant' because there have been incidences of patients harbouring two known pathogenic mutations like the *C9orf72* repeat expansion and a known *TARDBP* mutations (van Blitterswijk *et al.*, 2012b).

Category	Certainty of claim	Lines of evidence
Pathogenic	99%	• Null variant in a gene where LoF is a known cause of disease in this gene. • Statistically more in patients than controls. • Same alteration as previously known to cause disease • Relevant functional studies confirming a damaging effect • Segregation in multiple family members, some of which are distantly related • *De novo* confirmed
Likely pathogenic	90-98% sure variant is pathogenic	• Absent in control databases • *In silico* pathogenic predictions • Affects same codon as one previously published as causal • Truncation mutation • Missense in gene with low mutation rate • Located in a known mutation hotspot or studied functional domain which causes disease when mutated • *De novo* not confirmed • If recessive disease/gene variant found in trans with known pathogenic variant • Co-segregation in a few family members
Uncertain significance	Everything in-between	• Mixed evidence or missing evidence
Likely benign	90-98% sure variant is not pathogenic	• Computational evidence suggest benign • Missense variant in a gene where only published causal variants are truncating • Synonymous variant predicted not to affect splicing • In-frame indels in repetitive regions that aren't known to have any function • Observed in trans/cis with a dominant known variant • Found in an unrelated disease
Benign	99%	• Doesn't segregate with disease • Established functional studies related to disease mechanism show no effect of the variant • MAF in controls is too high for disease in question or observed in controls in a manner not consistent with disease penetrance

Table 6. List of the five-tiers of variant classification and the evidence to aid variant interpretation. To be placed in each category, a variant must fulfil multiple criteria within the box (Richards *et al.*, 2015).

The other decisions of note that this group made include those on transcripts and assays. The reference gene transcript used should be the most relevant i.e. most expressed, which is normally the longest however others should be kept in mind as they may compensate for a deleterious variant that only presents in a single transcript. Functional assays should be biologically relevant and not be restricted to one component of a gene's multifaceted functions, unless we know this is specific for the disease.

2.6. REPEAT SIZING

C9orf72 is the primary gene implicated in ALS but unfortunately the causal repeat expansion requires a unique method to detect its presence: repeat-primed PCR (Table 7, Figure 41). Additionally, a sizing PCR is used to determine the smaller allele. GeneScan Liz 500 size standard (Life technologies, UK) is used to calibrate the repeat sizes against each other. I used GeneMapper v4 (Life Technologies, UK) to analyse the repeat sizes.

Reagent	Volume (µl)	Sequence	Supplier
Nuclease-free water	395.35	-	Qiagen, UK
Forward	16.5	6-FAMCAAGGAGGGAAACAACCGCAGCC	Sigma, UK
Anchor	16.5	CAGGAAACAGCTATGACC	Sigma, UK
Reverse	1.65	GCAGGCACCGCAACCGCAG	Sigma, UK

Table 7. Primer sequences and the volumes required to create the primer mix used in the repeat-primed PCR for *C9orf72* repeat expansion size. Primer sequences from DeJesus-Hernandez *et al.* (2011b).

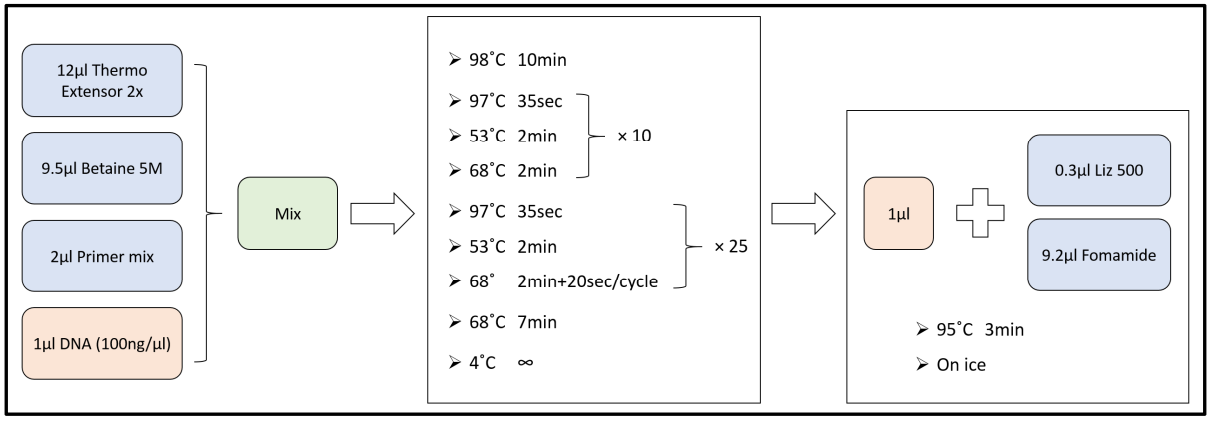

Figure 41. Protocol for repeat-primed PCR.

2.7. SUPPLIERS

Applied Biosystems Inc – 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

Bio-Rad Laboratories Ltd – Maxted Road, Hemel Hempstead, Hertfordshire, HP2 7DX, UK.

Corning – One Riverfront Plaza, Corning, NY, 14831, USA.

Illumina – Chesterford Research Park, Little Chesterford, Essex, CB10 1XL, UK.

Life Technologies – 3 Fountain Drive, Inchinnan Business Park, Paisley PA4 9RF, UK.

Promega – Delta House, Chilworth Research Centre, Southampton, SO16 7NS, UK.

Qiagen – Skelton House Lloyd Street North Manchester M15 6SH UK

Sigma-Aldrich Company Ltd – Fancy Road, Poole, Dorset, BH12 4QH, UK.

Ultra-Violet Products Ltd – Trinity Hall Farm Estate, Nuffield Road, Cambridge CB4 1TG, UK.

CHAPTER 3 PILOT TEST PLATE

3.1. INTRODUCTION

NGS is a powerful tool for assessing genetic variation in multigene disorders like ALS. However, one of the issues with NGS is the coverage i.e. small regions of the DNA will have low or missing reads, with most studies reporting between 85-98% coverage of the desired area. Using predesigned kits often results in a higher coverage but currently there is no commercially available gene panel for ALS. Therefore, we decided it was prudent to complete a proof-of-principle pilot study and optimise the technology to ensure high-quality data.

3.2. MATERIALS AND METHODS

3.2.1. STUDY DESIGN

I performed a literature search on the genetics of ALS to understand which genes were of importance in the disease. This information was then combined with the opinions of experts in the field to obtain a list of ALS genes we wanted to test. We decided that there were five categories of ALS genes we wanted to include in the study. Firstly the most important genes: *C9orf72*, *FUS*, *OPTN*, *SOD1*, *TARDBP*, *UBQLN2* and *VCP*. Then the genes of minor importance included: *ALS2*, *ANG*, *DCTN1*, *NEFH*, *PRPH*, *SETX*, *SQSTM1* and *VAPB*. Genes that causes similar diseases were included: *CHMP2B*, *FIG4* and *SPG11*. Genes where our knowledge was contradictory or lacking: *PON1*, *PON2*, *PON3*, *VEGFA.* And lastly genes where no knowledge was currently published in ALS: *MATR3*, *BSCL2* and *CEP112.* These final three genes were requests from our collaborators based on their own preliminary results. Experiment procedure and analytics were completed as described previously in Chapter 2.4 and 2.5. The samples have been described in Chapter 2.1.

3.2.2. PROBE DESIGN

The ALS gene panel was designed for TSCA to cover the 25 genes of interest. These were split into two groups depending on the desired coverage. The first group in which full exon

sequencing was desired included mostly major genes and unknown genes: *BSCL2, CEP112, FUS, MATR3, OPTN, SOD1, SPG11, TARDBP, UBQLN2* and *VCP*. The second group contained genes whose involvement in ALS is minor or questionable, and was restricted to specific areas where disease-causing mutations cluster: *ALS2, ANG, CHMP2B, DAO, DCTN1, FIG4, NEFH, PON1, PON2, PON3, PRPH, SETX, SQSTM1, VAPB* and *VEGFA*.

I ran 43 designs through the DesignStudio software, tweaking it every time to get the best metrics for coverage and low probe interaction.

3.3. RESULTS

The sequencing panel was tested on 95 samples, one of which was a SOD1-positive control, from both familial (n=22) and sporadic (n=73) ALS patients obtained from UCLP. In this cohort, NGS technology identified 43 rare variants (30 unique variants) in 33 patients that affect an ALS-associated gene either by an amino acid change or an alteration to a predicted splicing site (fALS = 6/22 patients; sALS = 27/73 patients). Of these variants, 13 have been previously reported with respect to ALS: 5 were determined as causal (*ALS2* I94V x3; *FUS* R521C; *SOD1* I114T), 4 were hypothesised as likely causal (*OPTN* 1401+4A>G; *PRPH* R9Q x3) and 4 had unknown significance (*SPG11* V270I x2; *TARDBP* L5P x2; Table 8). A small number of the remainder presented in more than one patient. The positive control *SOD1* mutation was detected in this trial. 28 variants remained open for further examination.

Gene	Variant	Amino acid change	No. of patients	PhyloP	SIFT	PolyPhen	MT	dbSNP	Quality Score	Genomic Position	Disease-causing? (no. of patients)	Patient Status (no. of patients)	Reference if previously reported
ALS2	A280G	I94V	3	B	B	B	B	rs3219154	391-897	202626437	Yes (1) No (2)	sALS (3)	Hand *et al.*, (2003) NS; Herzfeld *et al.*, (2009) in Spastic paraplegia
FUS	C1561T	R521C	1	B	D	D	D	rs121909670	2690	31202739	Yes (1)	fALS (1)	Vance *et al.*, (2009); Suzuki *et al.*, (2010) Drepper *et al.*, (2011)
OPTN	1401+2T >G	N/A	1	-	-	-	-	-	3818	13169905	Likely (1)	sALS (1)	Del Bo *et al.*, (2011) similar mutation
PRPH	G26A	R9Q	3	C	D	P	D	rs57451017	2524-5762	49689009	Likely (3)	fALS (1) sALS (2)	Gros-Louis *et al.*, (2004)
SOD1	T341C	I114T	2	C	-	D	D	rs121912441	727-1100	33039672	Yes (2)	fALS (2)	Gellera *et al.*, (2001; Stewart *et al.*, (2006)
SPG11	G808A	V270I	2	C	B	P	B	rs80338868	6261-6704	44949354	Unknown (2)	sALS (2)	Stevanin *et al.*, (2008) NS
TARDBP	T14C	L5P	2	-	-	-	-	rs61730366	193-259	11073982	Likely (2)	fALS (1) sALS (1)	Guerreiro *et al.*, (2008); Gijselinck *et al.*, (2009); Kirby *et al.*, (2010)

Table 8. Known variants identified in this study using the test ALS panel. NS = not significant; MT = mutation taster; B = benign; C = conserved (red); D = damaging (red); P = possibly damaging (blue).

The coverage of the panel was estimated using CovCheck which calculated that 92% of the desired area was covered by at least 10 reads (Figure 42).

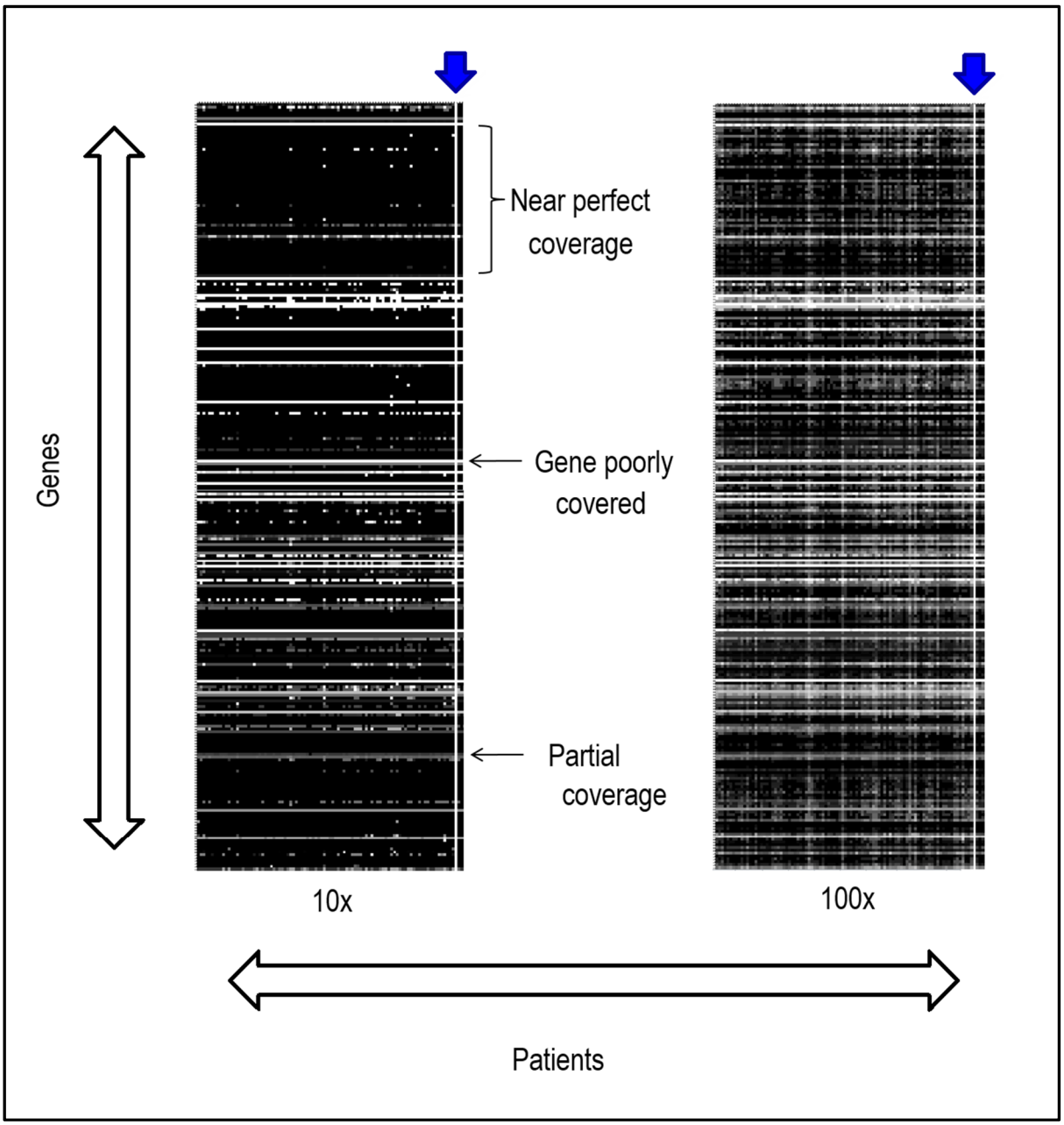

Figure 42. Coverage represented on a scale from black to white representing high and low read depth respectively, for all patients across the desired areas at 10x and 100x. Blue arrow denotes negative control.

The small regions of the genome that failed to sequence (Table 9) mostly lay in the genes *BSCL2*, *FUS*, *OPTN*, *SPG11*, *SETX* and *VEGF*. This was caused by a significantly high GC content of approximately 70% or, in the case of *FUS*, due to a high number of simple repeats and segmental duplication.

Gene	Length (bp)	Coverage			
		2x	10x	20x	50x
ALS2	1535	100%	100%	99%	96%
ANG	470	100%	100%	94%	87%
BSCL2	2139	86%	85%	74%	62%
CCDC46	3894	99%	99%	97%	92%
CHMP2B	6	100%	100%	99%	96%
DAO	3	100%	100%	100%	100%
DCTN1	16	92%	91%	87%	81%
FIG4	170	100%	100%	100%	100%
FUS	5104	83%	75%	70%	66%
MART3	5747	97%	94%	91%	86%
NEFH	383	100%	100%	100%	100%
OPTN	3582	90%	86%	79%	71%
PON1	492	99%	97%	94%	90%
PON2	138	99%	99%	98%	96%
PON3	14	100%	100%	100%	94%
PRPH	1012	94%	92%	84%	80%
SETX	15	89%	87%	86%	84%
SOD1	960	100%	100%	97%	88%
SPG11	7734	91%	87%	82%	76%
SQSTM1	50	93%	93%	87%	81%
TARDBP	4211	98%	94%	90%	87%
UBQLN2	3417	99%	99%	91%	82%
VAPB	7930	99%	97%	96%	94%
VCP	3846	96%	91%	84%	82%
VEGFA	219	85%	83%	78%	69%
		2x	10x	20x	50x
All genes	53087	95%	92%	87%	82%

Table 9. List of genes sequenced and coverage of the desired genomic areas. *BSCL2*, *FUS* and *VEGFA* had the lowest coverage and the overall coverage at 10x was 92%.

3.4. DISCUSSION

In this pilot study, we demonstrate the feasibility of NGS as a research and potential diagnostic tool for patients with ALS. 43 rare variants were identified amongst 33 of 95 patients (35%) with varying degrees of significance. 28 of these are completely novel. As predicted by genetic modelling, only 9 of 43 variants were in patients classed as familial and some mutations known to be disease causing were found in apparently sporadic patients (Al-Chalabi and Lewis, 2011).

A novel variant in *MATR3* (P154S) was detected and subsequently published along with data from a collaborator as a novel gene connected to causing ALS (Johnson *et al.*, 2014b). This is discussed further in Chapter 4. Other interesting findings included five patients who presented with different sets of two heterozygous mutations in *SPG11* which is known to cause spastic paraplegia via compound heterozygosity (Paisan-Ruiz *et al.*, 2008; Samaranch *et al.*, 2008). The c.T14C polymorphism found in *TARDBP* was also formerly reported in three different ALS studies but was disregarded because the variant is synonymous in the primary *TARDBP* transcript (Guerreiro *et al.*, 2008; Gijselinck *et al.*, 2009; Kirby *et al.*, 2010). On examination of another transcript (uc010oap.2), which has been found to be expressed in the brain (Ramasamy *et al.*, 2014), we have identified this SNP as non-synonymous and therefore hypothesise, with caution, that it may potentially be causal. One of the most common mutations in *FUS* (R521C) was detected in a familial patient from this study. 40 patients had both the SNPs rs854560 and rs10487132 in the *PON* genes which, in combination with each other, have previously been labelled as a risk factor (Cronin *et al.*, 2007). The amino acid change R9Q in *PRPH* has been previously reported by Gros-Louis *et al.*, (2004) who predicted it to be damaging due to its conserved nature, however, failed to detect it at a significant difference from controls. *OPTN*, which has presented as both a dominant and recessive cause of ALS, contained a heterozygous insertion at a splice site in one of our patients (c.1401+2T>G). A similar insertion at the same position has been flagged before in an ALS case (Del Bo *et al.*, 2011). The remaining 28 variants require further scrutiny for a possible causal role in ALS.

Nothing of interest was uncovered in the genes BSCL2 and CEP112 so these genes were dropped from further research.

The advantages of NGS over Sanger are primarily in the speed and in the expense. For example, this study costs approximately £150 per patient and takes 72 hours from start to finish to process 96 samples whereas to Sanger an equal quantity of DNA would cost nearly 33x more (approximately £5,000) per patient and take at least 300,000 hours of hands-on time, if not significantly more. Additionally, it is relatively easy to include new genes on the panel which may be of interest or even remove genes to increase the coverage of the remaining targets. This technology does, however, require 95 patient samples to be run simultaneously, in a single lane, in order to achieve the pricing stated above. This and similar NGS technologies are already finding successful application in other genetically-complex diseases such as hereditary ataxia and dementia (Sailer *et al.*, 2012; Beck *et al.*, 2014). The miss rate on this platform is approximately 8% but given the huge advantages of NGS over Sanger and the ease with which to identify these missing areas, we believe that these positives outweigh the marginally increased miss rate. To conclude, NGS technology shows promise for the diagnosis of both familial and sporadic ALS; our rapid high-throughput method is suitable for large scale genetic studies.

This work has been published:

Morgan S, Shoai M, Fratta P, Sidle K, Orrell R, Sweeney MG, et al. Investigation of next-generation sequencing technologies as a diagnostic tool for amyotrophic lateral sclerosis. Neurobiol. Aging 2015; 36: 1600.e5-8.

CHAPTER 4. MATR3

4.1. INTRODUCTION

As discussed briefly in Section 1.2.22 we were alerted to the possibility of a new ALS gene from the team at the NIH, namely *MATR3*. Including this gene in our pilot study revealed the P154S variant in a single patient (Figure 43; Chapter 3).

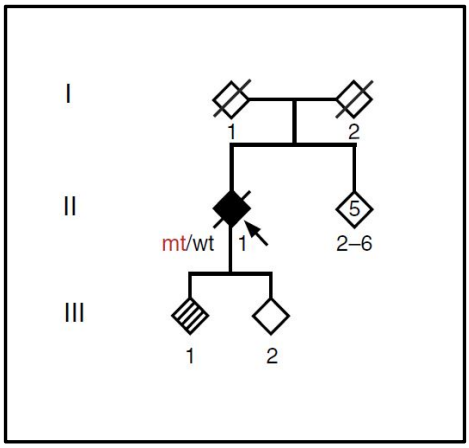

Figure 43. Family tree of P154S *MATR3* patient indicated by the arrow. Solid diamond indicates patient affected with ALS while multiple diagonal lines represents autism and a single diagonal line signifies deceased.

Our collaborators at the NIH had performed exome sequencing in an ALS-FTD family which singled out mutations in two different genes: *LMNB1* and *MATR3,* both of which segregated with the disease (Figure 44). The former is implicated in mitosis while the latter is an RNA- and DNA-binding protein and appears to be involved in editing RNA and transcription (Zhang and Carmichael, 2001; Salton *et al.*, 2011). Additionally, *MATR3* had been published with respect to a disease called vocal cord pharyngeal distal myopathy (VCPDM; Senderek *et al.*, 2009).

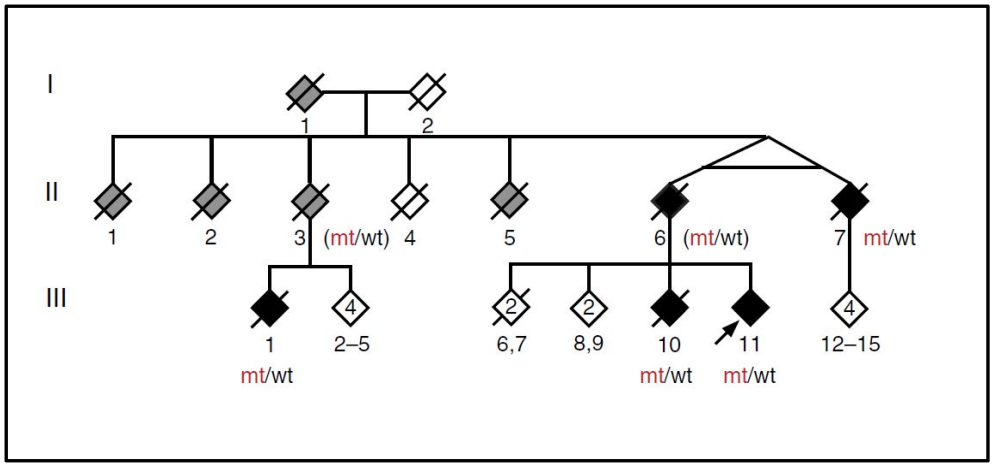

Figure 44. Pedigree with the F115C variant. Arrow indicates proband. Solid diamond represents affected individual while grey is possibly affected. A single diagonal line signifies deceased. mt = mutant alleles; wt = wild-type alleles; brackets indicates presumed obligate carriers.

With this new knowledge of *MATR3* being a candidate gene in ALS, Bryan Traynor and Howard Feit (the neurologist who initially characterised the VCPDM family) re-examined the clinical data and decided that a slowly progressive ALS was a more accurate diagnosis (Figure 45). This was mainly due to the observation of brisk reflexes in four of six patients and all patients showing the classic 'split-hand' characteristic of ALS. Myopathy patients should not have brisk reflexes, these are considered to be a symptom which would rule out this diagnosis. The causal variant was published as S85C.

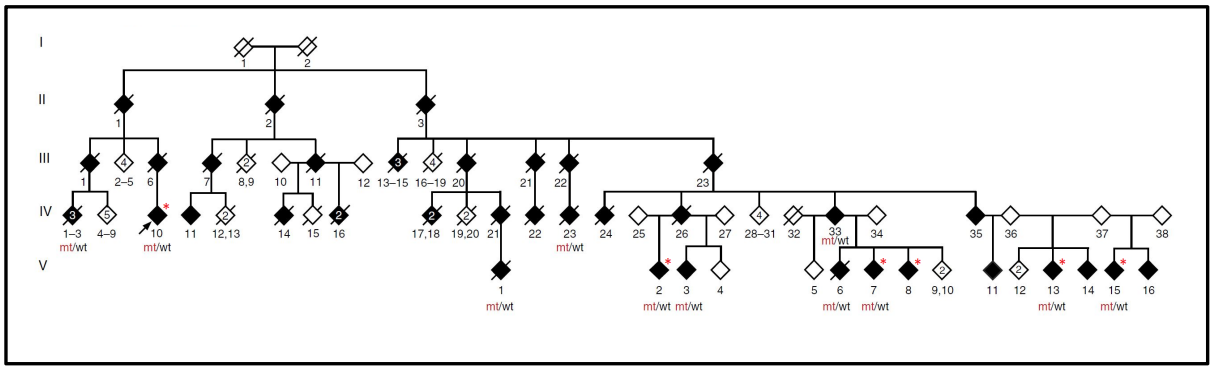

Figure 45. Family originally diagnosed with VCPDM, red asterisks = individuals who were clinically re-examined; solid diamond = affected; arrows points to proband.

Furthermore, they then examined 108 familial cases to discover another mutation: T622A which was in the proband and their affected cousin (Figure 46).

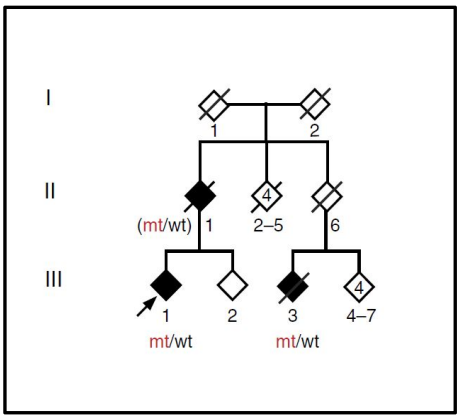

Figure 46. The family trees of the four patients with *MATR3* mutations. A single diagonal line signifies deceased; mt = mutant alleles; wt = wild-type alleles; brackets = presumed obligate carriers; arrow symbolises proband.

All of these results have been published (Johnson *et al.*, 2014b). The *LMNB1* gene was investigated further to find nothing of interest.

The many functions of *MATR3* are not entirely known however it is located within the nucleus and interacts with the TARDBP protein in an RNA-dependent fashion. It also has been linked to neuronal death following glutamate stimulation (Giordano *et al.*, 2005). However, if *MATR3* is pathogenic alone, it would seem more likely that this effect is through its interaction with RNA and its regulation of the expression of 77 other genes (Salton *et al.*, 2011). In yeast, MATR3 interacts with many proteins involved in RNA metabolism (Zeitz *et al.*, 2009).

Only S85C had an observable effect on the TARDBP protein while mutant F115C increased the nuclear staining of MATR3. Although the different effects of the mutations are disconcerting, there is evidence of mutations within known causal genes producing different pathogenic mechanisms for example, *LRRK2* in Parkinson's disease (Cookson, 2012) and *FUS* in ALS (van Blitterswijk *et al.*, 2013b).

Given this initial work, we decided to further characterise *MATR3* in ALS.

I performed Sanger sequencing (Figure 47) using the primer pairs shown in Table 10.

Exon	Forward primer	Reverse primer	Amplicon size (bp)	Optimum annealing temperature (˚C)
2	CTGCACGCCTTGCTAGTTTA	TTTCTTCCCAATCATCCCTA	396bp	58
2	CAACAAGGAGCTCATAGTGCA	GGGTTGAGACTAGGACCACG	398bp	56
12	TGGTGTGTCCTTTTGATTTCAG	GGTTCCTGCTCTGTCTGGTC	187bp	59

Table 10. Primers used to sequence *MATR3* and the annealing temperatures used for each.

Variants were examined against the transcript ENST00000394800 as it produces the longest protein length however within the exons of interest, all variants are the same between this and ENST00000618441, ENST00000361059 and ENST00000509990 which are potentially other transcripts of importance. Control data was gathered from whole-exome sequencing.

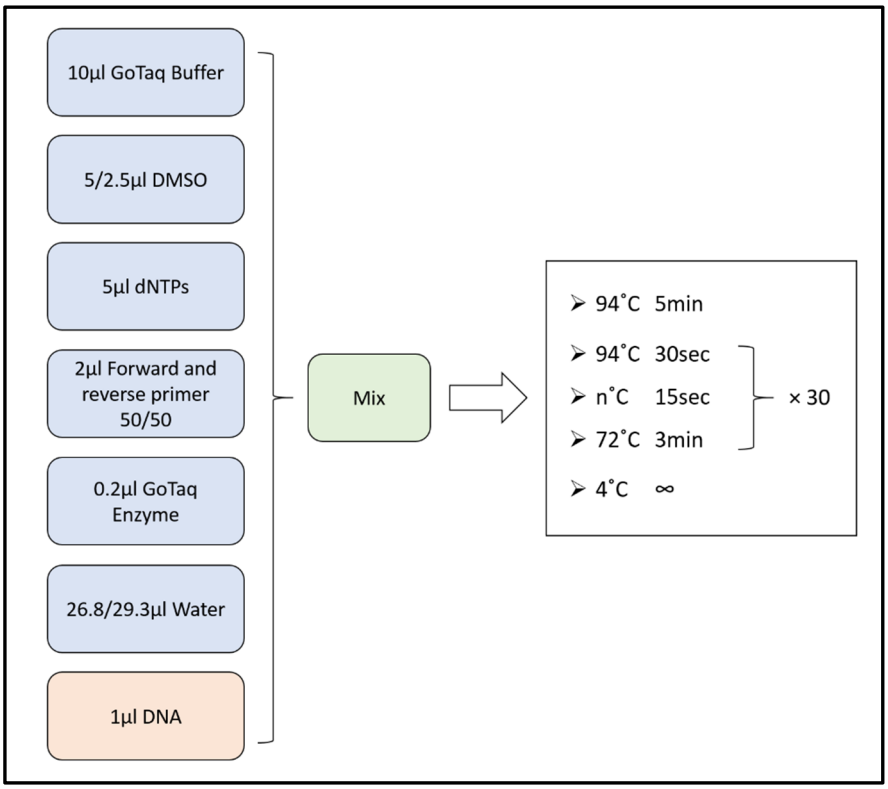

Figure 47. Conditions for PCR; n = optimised temperature seen in Table 10.

Sequencing was performed on 582 patients (88% were sALS) which identified the nucleotide alteration c.1867G>A causing E623K in the protein (Figure 48). This variant was not found in ExAC or 1000 genomes. The female patient presented with an apparently sporadic disease at the age of 45 and classified as definite ALS as gauged by the El Escorial criteria.

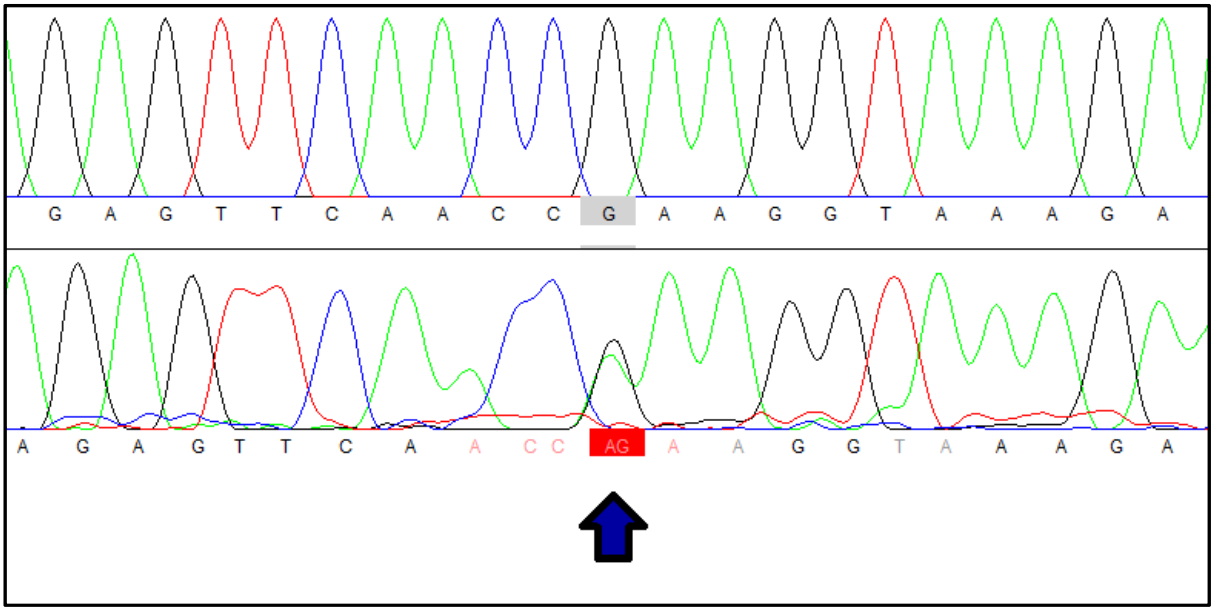

Figure 48. Sanger trace of the heterozygous E623K variant in *MATR3* (blue arrow).

This alteration is located on chromosome five at position 138658375 (h19) or 139322686 (hg38). It is predicted to be disease causing by MutationTaster and is conserved across many species (Figure 49). It is also predicted to affect splicing.

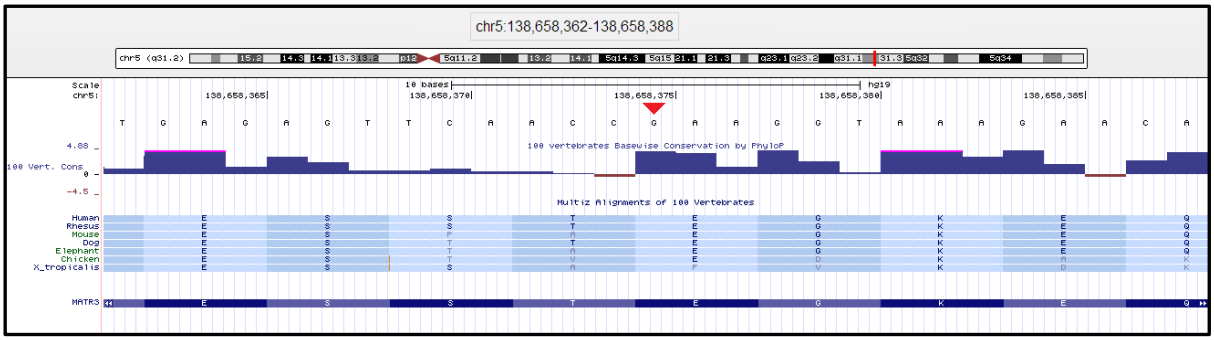

Figure 49. Location of the E623K (red arrow) and its conservation across different species. Modified from UCSC website in Paint v1511.

The synonymous variant E633E was also discovered in a single patient and was not predicted to affect splicing. This was not found in ExAC or 1000 genomes.

Exome data on 510 controls revealed no variants in exon 12 and four variants in exon 2 (Table 11). Additionally five variants were found amongst the rest of the gene but these areas were not examined in patients.

Variant	Nucleotide	Exon	Result	ExAC	ESP6500	Base pair
S98F	C293T	2	Non-synonymous	.	.	138643397
E168E	G504A	2	Synonymous	.	.	138643608
N277S	A830G	2	Non-synonymous	.	.	138643934
C293C	C879T	2	Synonymous	.	.	138643983
A378T (two subjects)	G1132A	6	Non-synonymous	0.019	0.0003	138652744
G463G	C1389A	8	Synonymous	0.039	.	138654677
L731L	T2191C	13	Synonymous	0.15	0.0002	138661171
V815V	T2445G	14	Synonymous	.	.	138661925

Table 11. List of variants uncovered in *MATR3* in control subjects. All are heterozygous.

4.4. DISCUSSION

Sequencing *MATR3* in a follow-up cohort identified the variant E623K in a 45-year-old sporadic patient. This alteration sits directly next to the previously published T622A and exists at a highly conserved nucleotide. However this latter mutant was present in a familial patient with an age of onset of 66.

There have been a number of publications since the original *MATR3* finding, the results are mixed with negative publications, variants of uncertain significance and the replication of mutations segregating in VCPDM. The S85C variant initially found in the family with VCPDM who were then clinically reclassified, was published again in this disease in 16 patients from 6 families in Germany (Müller *et al.*, 2014b). None of the subjects had any evidence of LMN involvement and all families shared the same haplotype around this mutation suggesting a common ancestor. Three of 150 German controls also had this same haplotype. Two other VCPDM pedigrees from Japan and America also confirmed S85C involvement (Yamashita *et al.*, 2015; Palmio *et al.*, 2016) yet this Japanese family displayed phenotype variability within mutation carriers and had LMN symptoms but no UMN signs. Muscle biopsies in these

patients revealed aggregates positive for SQSTM1, TARDBP and MATR3. Potentially, the S85C variant might be modified by other effects to produce different phenotypes.

Millecamps *et al.* (2014) examined patients with ALS and FTD which did not find any mutations in *MATR3*. Neither did Fifita *et al.* (2015) in an Australian cohort of familial ALS however both these studies examined only 153 and 106 patients respectively which might be lower than the requirement to capture potential *MATR3* variants.

Functional work *in vitro* on the initial reported mutations found no changes in the localisation of MATR3 or any presence of inclusions with overexpression of the mutant protein (Gallego-Iradi *et al.*, 2015). However, taking muscle biopsies from patients with *MATR3* mutations rather than inducing mutations produced different results: MATR3 was depleted in the nuclei of patients with the S85C variant (Palmio *et al.*, 2016). The authors did not perform any functional work specific to the disease so the relevance of MATR3 mislocalisation is questionable. Coelho *et al.* (2015) examined the function of *MATR3* to find that it affects the alternative splicing of many genes, implying the effects of mutations might be multifaceted. Johnson *et al*., (2014b) examined the ability of wild-type and mutant MATR3 protein to bind to TARDBP in HEK cells. Of the three mutations found in ALS, only S85C increased the binding of these two proteins. Nevertheless, given that no functional work has conclusively associated *MATR3* with an ALS-like cellular phenotype, the link is currently unconvincing.

R147W was published in an Italian cohort examining 200 ALS samples (Origone *et al.*, 2015). This was not present in 500 controls and is predicted to be pathogenic. A single Chinese patient was found with a S610F alteration but the *MATR3* mutation rate was higher in controls (Xu *et al.*, 2016). However the authors claim this without completing statistics and examination of the reported numbers reveals that this trend is not significant. 207 Taiwanese ALS subjects revealed A72T which was not in 500 controls (Lin *et al.*, 2015). Then Leblond *et al.* (2016) reported three variants V394M, c.-399+2T>A (5'UTR) and c.48+1G>T in ALS. The last of these affects splicing but only in an isoform that is 559 amino acids long rather than the full 847. It is unclear if this transcript is expressed as it is not reported in Ensembl. However, the variants reported in all the four papers all remain as uncertain significance with none providing strong evidence of pathogenicity.

It is clear that *MATR3* is associated in VCPDM but its involvement in ALS is still contentious. It seems likely to be involved in the families first published and it's possible that the reclassified VCPDM pedigree have both diseases since most, but not all, patients had a more likely diagnosis of ALS. The variant we have identified in this follow-up adds to the variant collection of unknown consequence.

Part of this work has been published:

Johnson JO, Pioro EP, Boehringer A, Chia R, Feit H, Renton AE, et al. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. Nat. Neurosci. 2014; 17: 664–666.

CHAPTER 5 CHCHD10

5.1. INTRODUCTION

In 2014, whole-exome sequencing on two distantly related patients provided a list of candidate disease genes which was narrowed down to *CHCHD10* by confirmation with Sanger sequencing in a further 6 affected and 2 unaffected (Figure 50; Bannwarth *et al.*, 2014). This large family presented with a range of phenotypes, including ALS and FTD, who all harboured a missense variant in exon 2 (S59L). Muscle biopsies from the patients revealed a mitochondrial dysfunction as did overexpression of mutant *CHCHD10* in cell cultures. The authors screened a further 21 families and established the same mutation in another pedigree providing good preliminary evidence for this gene being causal in ALS.

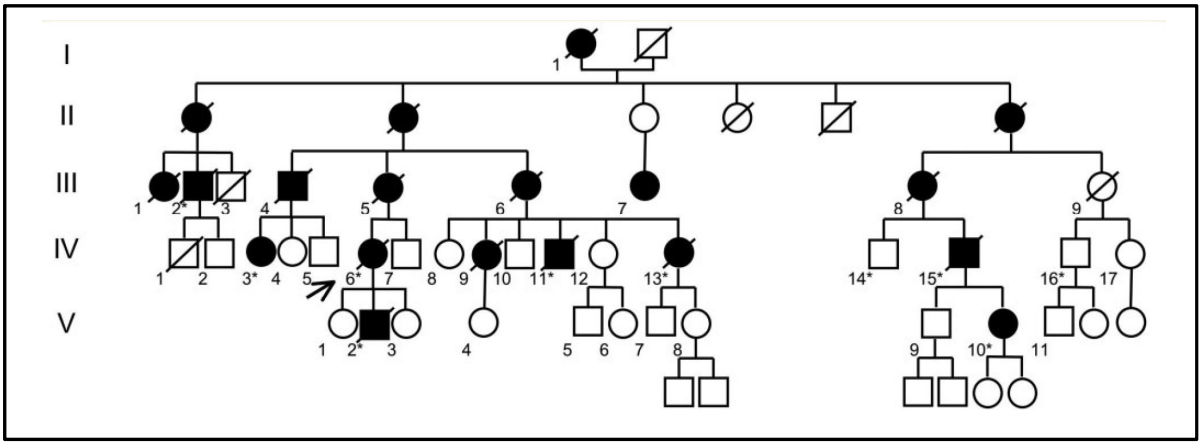

Figure 50. Pedigree of the family with the S59L mutant in *CHCHD10*. Solid symbols denote those affected and an asterisk is given to those who underwent genetic testing. Arrow points to proband.

Named Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10 (*CHCHD10*), this gene is located in the mitochondria and is hypothesised to hinder protein-protein interactions (Bannwarth *et al.*, 2014). Since this initial study, limited functional work on ALS/FTD mutations has been completed, however, they are theorised to act via a gain-of-function mechanism (Ronchi *et al.*, 2015). The authors have not disclosed why they came to this conclusion but it may be due to the mutation they uncovered (P80L) not being situated in any of the domains or binding sites thought to be important. This patient had a complex IV deficiency, an enzyme involved in the production of ATP. The CHCHD10 protein is ubiquitously expressed and a partial knockdown in HELA cell cultures reveal a mitochondrial dysfunction (Martherus *et al.*, 2010). This suggests a loss-of-function mechanism and is due to *CHCHD10*'s role in ATP

synthesis. Mitochondrial dysfunction is documented in ALS but more work is needed to confirm if this is the cause of disease in these patients or an unrelated comorbidity.

A number of other groups subsequently sequenced ALS and FTD patients for *CHCHD10* mutations with a number of these reporting the P34S alteration. Chiò *et al.* (2015b) found this variant in three of 224 sporadic patients (1.3%) while Chaussenot *et al.* (2014) established it in two of 80 sporadic individuals (2.5%). These groups also reported this variant to be absent in 165 and 200 controls respectively. If this variant is pathogenic, it would place it as the second biggest known cause of sporadic ALS for a single mutation after the hexanucleotide expansion in *C9orf72*.

The aim of the present study is to determine the prevalence of *CHCHD10* mutations in British ALS and FTD patients.

5.2. MATERIALS AND METHODS

We screened a cohort of 547 UK patients (452 ALS and 95 FTD) for mutations in *CHCHD10* by use of Sanger sequencing. Patient samples were obtained from the London UCLP MND Network and were all clinically diagnosed. Patients were previously determined to be negative for mutations in *C9orf72*, *TARDBP*, *FUS*, *SOD1 VCP*, *PGRN*, and *MAPT*.

CHCHD10 primers are described in Table 12 and the optimised experiment conditions are in Figure 51. Chi-square tests were performed in PLINK providing an odds ratio for any variants uncovered.

Exon	Forward primer	Reverse primer	Amplicon size (bp)	Optimum annealing temperature (˚C)
1	CGTAAAGGCCGTTAGTGTCG	GGGAGGAAGCAGGGTTAATC	1053	56
2	CTCCTCACTGGACACTTGGG	GGTCGTTTCCAGGAGCTG	355	58
3	AGCCTGGCCAACATAGTGAA	GAGTCTGCACCGACCTCTT	660	59
4	ACCTCATCAGCCAGGGAG	CCAACCCTCCTCTTGCAC	293	58

Table 12. Details of the primers used in this study to capture each exon of *CHCHD10*.

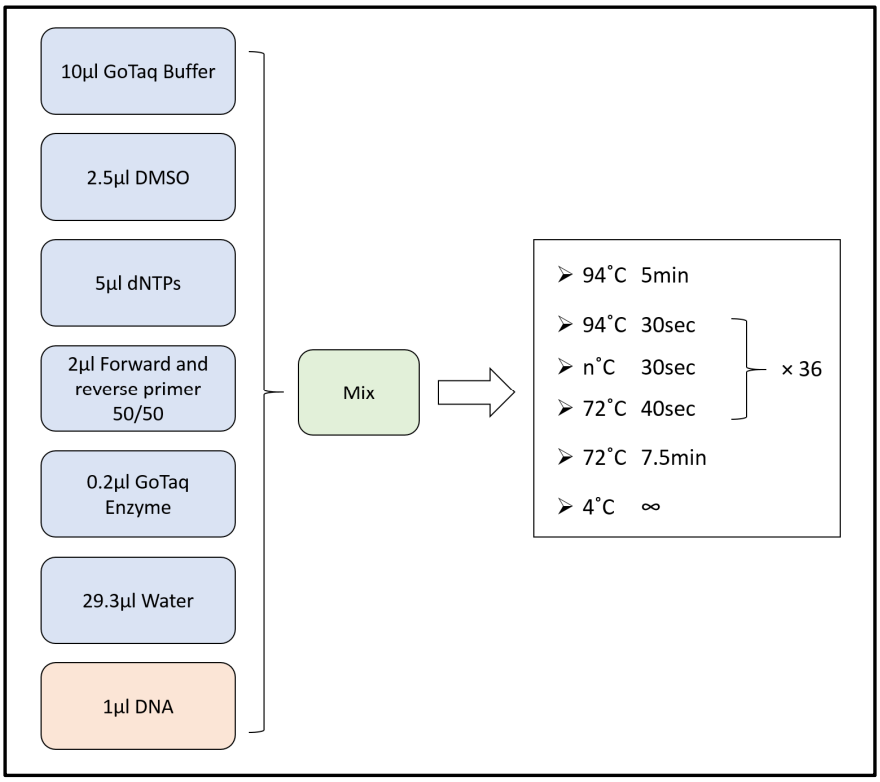

Figure 51. Conditions for PCR. n = optimised temperature seen in Table 12.

5.3. RESULTS

Sequencing of *CHCHD10* revealed four unique variants in eight patients (Table 13 and 14) which included the P34S variant in five patients with either ALS or FTD (Figure 52). P96T is present in up to 6% of control databases and so is likely a benign polymorphism (Table 14). The three other variants all have a frequency of less than 1% which is our threshold for being considered too common to cause disease. Predicative algorithms only provided a consensus for Y135H which is concluded to be benign.

Examining the UK10K MAFs of the four *CHCHD10* variants found that P34S was present in 29 of 4777 individuals (0.61%) as opposed to 0.91% of our cohort (odds ratio of 1.51, 95% confidence interval: 0.58, 3.9; P = 0.4). The Y135H variant (0.18% in our cohort) was present in 4 of 5232 individuals (0.076%) while S77G was absent in this database. This latter variant was present in a patient with Cuban ancestry.

Disease	Age of onset	Site of onset	Sex	Amino acid change
ALS	67	Upper limb	F	P34S
ALS	68	Limb and bulbar	M	P34S
ALS	60	Lower limb	M	P34S
ALS	66	Upper and lower	M	P34S
FTD	65	Dementia	M	P34S
ALS*	-	-	-	S77G
ALS	-	-	-	P96T
ALS-FTD	70	Dementia	M	Y135H

Table 13. Summary of the variants in *CHCHD10* in our cohort. *Cuban origin.

Nuc. change	Amino acid change	Exon	Base pair	1000G	ExAC	ESP6500	cg46	dbSNP	SIFT	PolyPhen 2	LRT	MT
100C>T	P34S	2	24109722	0.0004	0.001	.	.	rs551521196	B	B	D	D
229A>G	S77G	2	24109593	0.002	0.0003	0.0003	.	rs370872556	B	P	D	D
286C>A	P96T	3	24108438	0.06	0.03	0.06	0.04	rs111677724	B	P	B	P
403T>C	Y135H	3	24108321	.	0.0003	0.0005	.	rs145649831	B	B	B	B

Table 14. Details on the four variants identified in *CHCHD10*; Nuc = Nucleotide; MT = MutationTaster; B = benign; D = damaging (red); P = possibly damaging (blue).

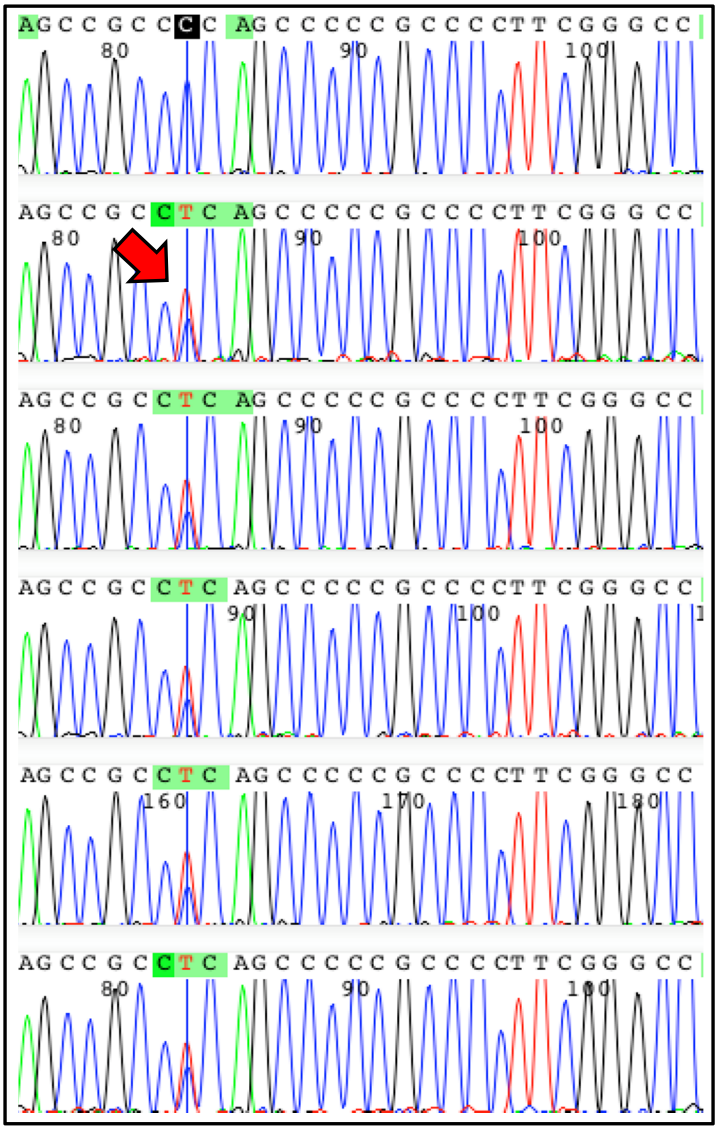

Figure 52. Sanger traces of control subject (top panel) and the P34S variant (red arrow) in five patients.

5.4. DISCUSSION

Sequencing in 547 subjects with ALS and FTD revealed four distinct variants including the published P34S variant in five patients. This variant is located in exon 2 where most *CHCHD10* variants are reported and lies in the non-structured N-terminal domain. All patients harbouring this alteration presented in their seventh decade of life with mostly a limb-onset disease in contrast to previous studies linking this variant to bulbar-onset (Chiò *et al.*, 2015b; Chaussenot *et al.*, 2014). Previous work considered this variant to be pathogenic however it is present in ExAC at a frequency of 0.2%, which increases to 0.6% when only Europeans are examined.

This is just below the threshold for rarity which raises questions of its pathogenicity. Using control data from the UK10K cohort, we performed statistical analysis of the P34S alteration to find no significant difference. Of note is that exon 2 of *CHCHD10* is not adequately covered in WES including in ExAC with only 5,179 individuals (9%) having results for the P34 location (Figure 53). This low quality may contribute to the absence of common variants in this region within control databases. Both Y135H and S77G remain as uncertain significance while P96T is likely benign.

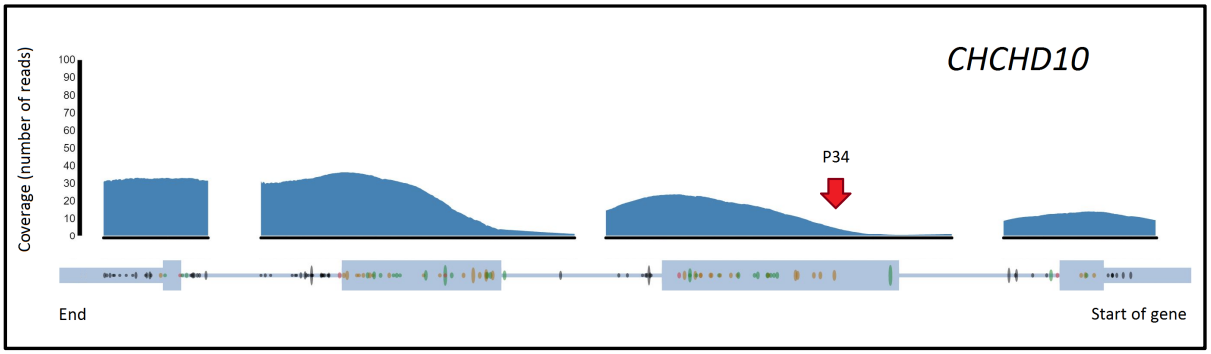

Figure 53. Average coverage of *CHCHD10* in ExAC. The whole gene has a lower than average read depth with the worst coverage at the beginning of exon 2 where the P34 codon is located (red arrow). Adapted from the ExAC website.

Other studies into *CHCHD10* are very promising. The R15L transition has been seen to display segregation in multiple publications albeit one was with an expected incomplete penetrance owing to a single obligate carrier (Johnson *et al.*, 2014a; Müller *et al.*, 2014a; Kurzwelly *et al.*, 2015). Although none of these groups performed functional work, this variant is not present in ExAC which is promising. Zhang *et al.*, (2015) completed a comprehensive study in a number of neurodegenerative disorders: ALS, PD, AD and FTD with 497 control subjects. They also identified R15L in a sporadic patient as well as three other variants in ALS or FTD including P80L which was confirmed by Ronchi *et al*., (2015) in two patients. In the PD and AD subjects, only P34S was present but this did not segregate with the disease providing further evidence that this variant is benign. Chinese cohorts of FTD but not ALS are particularly enriched with *CHCHD10* mutations which were present in 7.7% of sporadic FTD (Jiao *et al.*, 2015; Li *et al.*, 2016).

Outside of ALS and FTD, variants have also been observed to segregate in a family with mitochondrial myopathy (Ajroud-Driss *et al.*, 2015) and in 17 pedigrees with spinal motor neuronopathy (Penttilä *et al.*, 2015). The former group performed functional studies which suggest that G58R is causal.

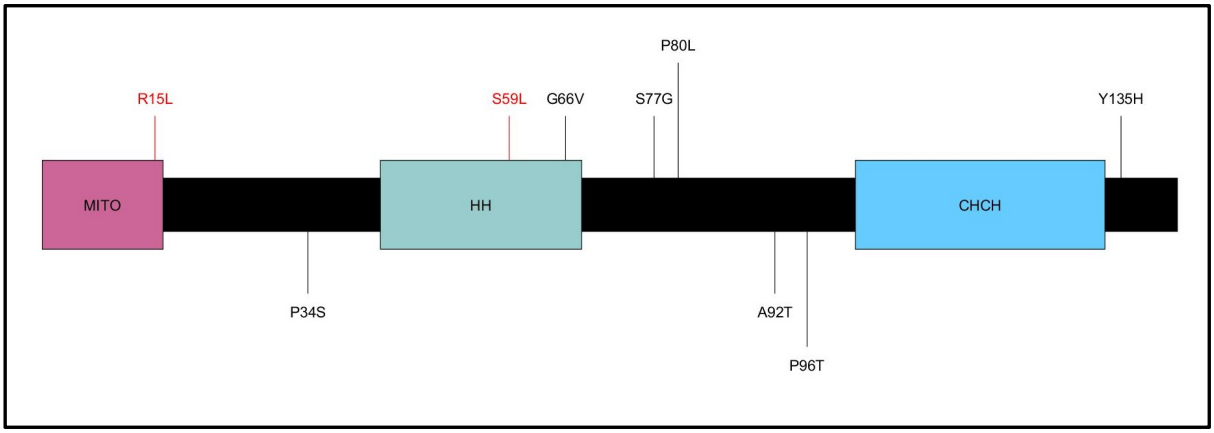

Figure 54. Mutations in *CHCHD10* found in patients (above gene) and controls (below gene). Red mutation = pathogenic; black = unknown; MITO = mitochondria targeting; HH = hydrophobic helix; CHCH = coiled coil 1-helix 1-coiled coil 2-helix 2. Drawn in Inkscape v0.91.

In conclusion, our results do not support P34S as a pathogenic variant but the literature does advocate *CHCHD10* as a causal gene in both ALS and FTD, providing another genetic link between these two diseases and reinforcing the hypothesis of mitochondrial dysfunction as a mechanism for ALS pathogenesis. I belive that there is stong evidence to support both R15L and the initial mutation (S59L) in being pathogenic for ALS (Figure 54).

This work has been published:

Abdelkarim S, Morgan S*, Plagnol V, Lu C-H, Adamson G, Howard R, et al. CHCHD10 Pro34Ser is not a highly penetrant pathogenic variant for amyotrophic lateral sclerosis and frontotemporal dementia. Brain 2015: awv223.*

CHAPTER 6 CORE STUDY: USING THE GENE PANEL TO ANALYSE A LARGE COHORT OF ALS AND CONTROL SAMPLES

6.1. OVERVIEW

We performed a pilot study using NGS as a tool for examining the genetics of ALS in 95 patients (Chapter 3). This method revealed a number of interesting results and provided a basis for the principle investigation where we aimed to implement this technology on more than one thousand patients.

In terms of the loci included in the gene panel, there were a few containing regions of DNA which failed to sequence. With this in mind, the genes *BSCL2*, *CEP112* and *VEGF* were removed from the project as their connection to ALS still remains poor and we cannot achieve high quality data on them. Additionally, the amplicons covering *FUS*, *OPTN* and *SETX* were redesigned to obtain optimum coverage. Lastly, although the gene *SPG11* has a potential involvement in ALS, the sheer size of it presented as a challenge to both sequence and analyse especially considering the fact that compound heterozygous cases have been observed and the chances of two rare variants occurring in this gene are much higher. It is for this reason that we decided not take this gene through to the next part of our experiment. However, since I have WES data on the controls, I also examined the removed genes within this control cohort to compare to the test plate data. Since the completion of the test plate, two other genes were gaining popularity as risk factors for ALS, namely *TREM2* and *PFN1*. Therefore these genes were added to the ALS panel.

6.1.1. METHODS

A requirement for the final design of the project was to balance the available funds with both the number of patients and amount of DNA targeted. The resulting design extensively covered all exons and both untranslated regions (UTRs) of *SOD1, TARDBP, FUS*, *VCP, OPTN*, and *UBQLN2*. Then the rest of the genes were covered at mutation hotspots: *ALS2, ANG, CHMP2B, DAO, DCTN1, FIG4, NEFH, PFN1, PON1, PON2, PON3, PRPH, SETX, SQSTM1, TREM2* and *VAPB*. These hotspots were ascertained by use of an ALS variant database I created discussed in Chapter 9. The TSCA protocol was run using this design and analysed

as per the descriptions in Chapter 2. I performed identical analysis on whole-exome sequencing data from 510 control patients pulling out the same genomic regions which were covered by the MiSeq panel. I included four common sex markers to ensure that subjects matched their stated gender.

As mentioned NGS is unable to reliably assay long repeats such as those in *C9orf72* and *ATXN2*. Therefore, repeat-primed PCR was used to detect the expansion mutation in *C9orf72.* Our collaborators at Kings completed standard fragment length analysis for the microsatellite repeat in *ATXN2*. Although, because our control samples did not have *ATXN2* data, it was not included in our case-control analysis.

6.1.2. RESULTS

A total of 1,131 subjects were run on the MiSeq which included 100 controls, 124 fALS and 917 sALS. The majority of these (n = 1,074) were from the MNDA DNA bank while 33 of the controls were from IPDGC and the remainder were Argentinian samples (18 sALS and 6 fALS) which are discussed separately in Chapter 7 since their ethnicity varied from the rest of the cohort. WES data was provided on 510 controls. Chapter 2.1 contains a more detailed description of all samples. 1.6% of samples completely failed to sequence which included five controls (all from the WES), eleven sporadic and two familial subjects. A further three sporadic patients and five controls failed to sequence adequately for some of the desired genomic area but this varies slightly depending on loci or gene region under examination. Therefore some genes, when examined independently, had slightly higher subject numbers.

Following initial standard quality checks to remove false positive calls, 52,804 variants remained. As per the method described in Section 2.5.2 a total of 29,930 images were taken of flagged variants which were then examined by eye. Of these, 8,654 were kept which would normally be discarded by a computer, while 4,621 which passed quality checks were clearly false positives (8.8% of all calls). The final number of mutations averages at 41 alterations per person (range 24-72) which is mostly due to common polymorphisms and includes intronic and synonymous SNPs. There were some minor regions which were covered adequately by only one of the technologies used (either WES or MiSeq) and variation within these regions were not included in most analyses except when examining an individual variant against the

published literature. 317 patients did not have complete *C9orf72* data because of insufficient DNA. Of those typed for repeat expansions in this gene, 45 of 654 (6.9%) sporadic patients had the mutation as did 11 of 72 (15.2%) familial cases.

Comparing the 33 controls post-filtering which had been examined using whole-exome and MiSeq-targeted sequencing revealed identical calls in all subjects except for a few intronic variants (which WES does not capture) and for indels of two or more nucleotides. The differences in indel calling lay mostly with mononucleotide repeats which are known to cause problems in NGS. All indels of two or more nucleotides across patients and controls were removed from the analysis to ensure no technological biases were driving the differences between the two groups. For SNPs, WES and targeted sequencing both produced the same results and therefore the former can be reliably used as controls for my dataset.

One of the major difficulties in NGS data is how to interpret variants, especially those which are novel or extremely rare. We found 906 alterations which are defined as such, of which 225 are exonic and 138 are previously published with respect to ALS or another disease, however, some of these also occur in our control cohort. Variants were deemed likely to be causal if they were published previously and not found in control cohorts. Under this interpretation of pathogenicity, 103 patients in 1,007 can be explained (10.2%; Table 15) by mostly *C9orf72* repeat expansions (4.9%) but also *SOD1* (2%), *TARDBP* and *FUS* (both 1%). However, as mentioned *C9orf72* is potentially higher than this frequency due to missing data in a number of patients.

SPG11, although not examined in patients, was examined independently in the control dataset. Subjects had an average of 1.7 mutations in this recessively causal gene, with 24% of individuals harbouring a homozygous variant and 42% with a potential compound heterozygous variant. A total of 1% of controls had 5 mutations in *SPG11* showing it is a highly mutated gene. Additionally, one control had a stopgain mutation in this gene.

Gene	Familial		Sporadic		Controls	
	All	Likely pathogenic	All	Likely pathogenic	All	Likely pathogenic
ALS2	3.7%	0.9%	6.6%	0%	8%	0%
ANG	0%	0%	0.4%	0%	1.3%	0.3%
CHMP2B	2.8%	0%	2.9%	0%	11%	0%
DAO	0.9%	0.9%	0.2%	0.1%	0%	0%
DCTN1	0%	0%	0.9%	0%	1.5%	0%
FIG4	0.9%	0%	0.6%	0.1%	7.2%	0%
FUS	2.8%	1.9%	2.4%	0.6%	12%	0.2%
NEFH	27%	0%	18.6%	0%	38%	0%
PFN1	4.6%	0%	7%	0.2%	6.3%	1.2%
PRPH	2.8%	0%	1.4%	0%	2%	0%
SETX	3.7%	0%	4.2%	0%	6%	0%
SOD1	8.3%	7.3%	1.1%	0.8%	0.3%	0%
SQSTM1	1.9%	0%	1.6%	0%	1.7%	0%
TARDBP	6.5%	4.6%	1.8%	0.8%	1%	0%
TREM2	0.9%	0.9%	0.8%	0.7%	0.7%	0.7%
UBQLN2	1.9%	0%	2.6%	0.1%	1.7%	0%
VAPB	0%	0%	1.6%	0%	1.2%	0%
VCP	1.9%	0%	1.6%	0.1%	7.7%	0%

Table 15. Percentage of patients with coding mutations in each gene. Variants were considered to be likely pathogenic if they fulfilled several criteria from Table 6. These results do not take into account missing data and so actual numbers may be slightly higher.

6.2. BURDEN OF RARE VARIANTS

6.2.1. OVERVIEW

One of the techniques for identifying the presence of causal mutations is to examine the collective mutation burden in selected genomic regions in patients set against a control cohort. If there are locations in patient DNA with increased rare variation, then it can be assumed that some or most of these mutations are detrimental and are associated with the disease in question. The test plate results were included in these assessments to increase power.

6.2.2. METHOD

A region-based test comparing the rare-variant burden in cases versus controls was completed by use of the SNP-set (sequence) kernel association test (SKAT v1.1.2; Wu *et al.*,

2011). This test is a collapsing method which combines results from multiple variants into a single score for the selected region. It is computationally efficient and increases the power to detect the effect of rare variation in a case-control study while correcting for covariates.

The rational for the loci selected for this test were based on an *a priori* understanding of areas likely to be involved in the disease. I based this on the results discussed in Chapter 9 which collected together all published variation within these genes and their determined association with ALS. Given that this test examines rare SNPs only, the genes *ATXN2* and *C9orf72* were excluded as the nature of their association lies within repeat expansions. *PON1-3* and *VEGFA* were also removed as the interest in these genes remains with common variation. The remaining genes were all included but filtered for an MAF of less than 0.01 and for high quality ensuring there were no differences in missingness between cases and controls. SNPs could be intronic, exonic, synonymous, coding, novel or known to be pathogenic. Sex was used as a covariate and a dichotomous test was implemented for case-control analysis regardless of patient status (familial or sporadic). Although previous work on mutation burden had only found significant results in familial patients, most of our power comes from the number of sporadic individuals we have obtained for this study so we decided to include them.

A total of five tests were carried out, firstly on the entire region sequenced and then on specific areas we were interested in. Reported P-values are all corrected using the formula:

$$B = 1 - (1 - P)^n$$

Where *P* is the critical P-value obtained in the test, *n* is the number of tests completed and *B* is the Bonferroni corrected P-value. All stated values in the results section are adjusted for this multiple testing and are considered to be significant if less than the 0.05 threshold.

6.2.3. RESULTS

The first imputation saw 26 individuals removed for missingness and 33 WES controls removed as they were duplicates of the MiSeq controls. SKAT analysis revealed an increased number of rare variants in cases compared to controls when all genomic regions are included (P = 0.003). This was based on the 393 variants which passed all quality checks and did not have more than 10% missingness. Since this may be solely due to known pathogenic variants,

previously published SNPs in ALS were excluded (regardless of whether they were hypothesised to be pathogenic or not) and the test was repeated. The result was still significant (P = 0.01). This significant difference in the burden of rare variants lay in the UTR and intronic areas of the genes rather than the exons. We therefore tested coding and non-coding regions independently: UTR and intronic loci P = 0.04 whereas exonic variation did not show significant association P = 0.1 (synonymous) and P = 0.1 (non-synonymous). We did not have sufficient power to further analyse the UTR and intronic regions independently of each other, however, there were more rare variants in the UTRs than within introns of patients (Table 16).

Source	Description	Unique SNPs	Total number of mutations found	Corrected P-value
Coding and non-coding	All	393	888	0.003
Coding	Known causal	49	112	NA
Coding and non-coding	All minus known causal	334	724	0.01
Coding	All	195	469	NA
Coding	Nonsynonymous minus known	87	193	0.1
Coding	Synonymous	59	164	0.1
Non-coding	All	198	419	0.04
Non-coding	UTR	102	205	NA
Non-coding	Intronic	96	214	NA

Table 16. Description of the burden test results and corrected P-values. This includes the number of individual unique SNPs and then the total number of mutations found across all subjects for these unique SNPs; NA = region not tested based on *a priori* decision of which areas we wanted to examine.

6.2.4. DISCUSSION

Using SKAT to examine any increase in rare variation revealed a significant burden as one would expect given that known pathogenic mutations were included (n=393). However the test remained significant even after these previously reported mutants were removed (n=49), almost entirely due to the UTR and intronic SNPs included. We believe that the UTRs contain most of this burden over the introns based on their function, the trend of increased rare variants in the current study and previous ALS work on these areas as discussed below. As for the exonic regions, without the reported pathogenic variants, there was no significant burden. This indicates that most (or all) of the causal variants have been discovered already within coding regions. This is likely due to these areas being exhaustively examined over the years while very few studies have concentrated on non-coding regions.

The UTRs of genes are often ignored in genetic studies of disease, partly owing to the difficulty in the interpretation of any discovered variation. We had decided at the beginning of the project to include the UTRs to address this lack of knowledge, especially within *SOD1*, *TARDBP*, *FUS*, *OPTN*, *VCP* and *UBQLN2*. Additionally, smaller areas in the remaining genes were also covered with high-quality reads in these regions owing to the primer designs prioritising efficiency of exome coverage rather than strictly sequencing only the exons and therefore included some of the UTRs in select loci.

There have been a handful of publications also suggesting this effect to varying degrees. Sabatelli *et al.* (2013) examined a large cohort of 420 ALS and 480 controls of Italian origin specifically for variants in the 3'UTR of *FUS*. They found an increased number of these non-coding mutations within patients: four unique rare variants amongst five individuals while no rare variants were present in controls. Three of these were studied further in primary fibroblast cultures (c.*59G>A, c.*108C>T and c.*110G>A) in comparison to one patient with the known R521C *FUS* alteration, two patients with no causal mutations and four control subjects. The UTR modifications and the known pathogenic *FUS* variant all caused a mislocalisation of the FUS protein whereas this effect was not present in the other ALS subjects or controls. This provides evidence for the theory that these mutants contribute to ALS. (Dini Modigliani *et al.*, 2014) reported a c.*48G>A variant in two ALS patients with a severe phenotype. Functional work revealed that this alteration increased FUS expression dramatically. Correspondingly, overexpression of wild-type FUS causes an ALS-like syndrome in mice (Mitchell *et al.*, 2013). The 3'UTR of *FUS* is known to be involved in a feedback loop for its own expression via the alternative splicing of exon 7 (Zhou *et al.*, 2013). Inclusion of exon 7 increased protein expression and vice versa. The authors also examined disease-causing mutations to discover that they altered this autoregulation method and exon 7 was unable to be repressed, therefore increasing FUS protein concentration. FUS-knockout mice exhibit abnormalities but none that relate specifically to ALS (Kino *et al.*, 2015) while ExAC reports no LoF *FUS* variants in all 58,787 individuals sequenced despite 28.6 being expected to have occurred. These studies combined suggest that a tight control of FUS expression is necessary in humans (Dini Modigliani *et al.*, 2014).

Within the *TARDBP* gene, Gitcho *et al.* (2009) published c.*2076G>A as segregating in a family with ALS and FTD. However, they specifically only mention two affected family members and do not divulge if they sequenced unaffected family members. The 3'UTR variant was not present in 982 control subjects and caused TARDBP RNA expression to increase

twofold compared to 40 controls and FTD patients with other mutations. Like *FUS*, *TARDBP* also regulates its own expression through the 3'UTR (Ayala *et al.*, 2011). However this is achieved by inducing RNA instability rather than through splicing. Again, as with *FUS*, there are no LoF variants in *TARDBP* in ExAC when 11.8 are predicted to occur.

Additionally, there are a number of studies which report UTR variants in ALS with uncertain pathogenicity. Rutherford *et al.* (2008) report six UTR alterations in *TARDBP* but do not reveal how many cases each variant was in or if these were present in controls. In 2009, a French cohort of 285 sporadic ALS cases were sequenced to reveal one patient harbouring c.*1462T>C in *TARDBP* which was not in 360 controls (Daoud *et al.*, 2009). Then four UTR alterations were reported in this gene in 410 ALS/FTD cases in Belgium (Gijselinck *et al.*, 2009). Two FUS 3'UTR changes were present in an Italian ALS cohort but the authors did not reveal if they were present in their 376 controls which seems likely given that c.*41G>A is a common polymorphism (Ticozzi *et al.*, 2009). The *FUS* c.*24G<C change was detected in a fALS proband which was absent from 970 controls and two affected relatives indicating it is likely to be a rare polymorphism (Groen *et al.*, 2010). Another study revealed UTR variants in *FUS* and *SOD1* in ALS but it is not explicitly stated if they also sequenced their 700 controls for these regions (DeJesus-Hernandez *et al.*, 2010). Drepper *et al.* (2011) and Zou *et al.* (2012a) both report these non-coding *FUS* changes in ALS that are absent from controls while in *VCP*, c.*12C<T is present in cases and not 1,205 controls (Abramzon *et al.*, 2012). Lastly, *ANG* and *FUS* were reported to have UTR mutations not present in controls, one of which we also found solely in cases, namely c.*132C<A (Brown *et al.*, 2012). While this collection of papers seem to indicate a burden of UTR variation in ALS, not a single report mentions if any rare variation existed solely in controls. This piece of information is lacking in many publications and not just for UTR data.

While the variants of uncertain significance require more investigation, our data and the published work on functional studies, segregation and burden of rare variants in ALS all point towards the UTRs having a likely involvement in the disease especially within *FUS* and *TARDBP*.

6.3. OLIGOGENIC BASIS

6.3.1. OVERVIEW

As discussed in Chapter 1, the hypothesis of an oligogenic basis in ALS is starting to be explored by a few different groups.

6.3.2. METHOD

To examine this further, a selection of variants were run through a binomial test in R v3.2.3 as described previously (van Blitterswijk *et al.*, 2012b). However, heterozygous and homozygous hits were both treated equally as a single mutation unlike the aforementioned study. Variants were selected based on their likelihood of being pathogenic, so they had to fulfil several of the criteria mentioned in Table 6 and had to be either nonsynonymous coding variants or directly within a known splicing region.

To calculate the binomial distribution of the data, we used the formula:

$$f(x) = \binom{n}{x} p^x (1-p)^{(n-x)} \quad where\ x = 0, 1, 2, \dots, x$$

Where f(x) is the probability of getting the result achieved which is based upon the number of independent trials completed (n), the probability of obtaining a single successful trial (p), and the number of successful trials (x). Each trial has a binary outcome: either success or failure.

This is used to compute the probability of multiple mutations occurring in cases given the probability distributions of mutations in both cases and controls. We performed this test twice, firstly on reported ALS-variants only and then on variation found in *C9orf72*, *SOD1*, *TARDBP*, *FUS*, *ANG*, *ALS2*, *VCP*, *OPTN*, *NEFH* and *UBQLN2* where we included all rare, coding variation in these genes (excluding *C9orf72* where only repeat expansions were included). Reported P-values are corrected for these two investigations using the formula in Section 6.2.2.

6.3.3. RESULTS

A binomial test performed on variants which were previously published in ALS revealed that there is not an increased number of patients with two mutations than expected by chance alone (P = 0.4). Since the data collected in the current cohort was much richer than simply selecting the list of known mutations found, the published literature on specific variants was disregarded and instead genes were tested as a whole for only: *C9orf72, SOD1, TARDBP, FUS, ANG, ALS2, VCP, OPTN, NEFH* and *UBQLN2*. There were 11 of 1,112 patients with more than one mutation in an ALS gene which is significantly higher than expected by chance based on the mutation rates in both cases and control (P = 0.001).

6.3.4. DISCUSSION

Implementing a binominal test to look for an increased number of cases with multiple mutations exposed such an effect in rare variants deemed likely to cause disease. This test is based on the probability distribution of mutations in both cases and controls. Restricting the analysis to known variants did not replicate previous findings. A potential explanation for this may be due to the published literature containing inaccurate information on the pathogenic status of variants (and genes) whereas previous work only included variants in a smaller number of genes, that is, those more likely to be associated with the disease: *C9orf72*, *SOD1*, *FUS*, *TARDBP* and *ANG* (van Blitterswijk *et al.*, 2012b). With this idea in mind, we performed a hypothesis-driven binomial test selecting only coding variants in the genes we believed to be most associated with ALS. In our cohort, 1% of patients had two or more of these mutations which was significantly higher than expected by chance based on known mutation rates. Most of these patients had the repeat expansion in *C9orf72* with another mutation for example with *VCP* R155H, *TARDBP* A321V or *FUS* R521C (all known variants). A single control subject also had two mutations, namely P372R in *ALS2* and A90V in *TARDBP*. The former is a heterozygous variant in a gene known to be recessively damaging and so is unlikely to be pathogenic alone while the latter is hypothesised as a risk variant rather than fully causal as it has been found in some control cohorts yet displays abnormal localisation and aggregation of TARDBP (Guerreiro *et al.*, 2008; Winton *et al.*, 2008). This control patient with two potential risk variants demonstrates we should be careful in our interpretation of such mutations.

It is reassuring that dual mutations occur even without *ANG* and *NEFH*, as these are genes which some may question over their involvement in ALS. A recent paper by Nakamura *et al.* (2016) reported pathogenic mutations alongside "potentially pathogenic" variants in the same patients. However, this latter cluster of mutations all lie in the *uncertain significance* box according to guidelines and I would consider many of them not to be pathogenic, for example, missense variants in *SPG11* which, as discussed later in Section 6.5.18, is a highly mutable gene. Likewise, Kenna *et al.* (2013) published some convincing and some not convincing combinations of mutations with 1.6% of patients harbouring two alterations in ALS genes (4% of fALS and 1.3% sALS). However, there have been reports of two known pathogenic variants converging for example in two families where the proband has both a known *TARDBP* variant and the *C9orf72* expansion (Chiò *et al.*, 2012). The authors link this combination to a more severe phenotype and an earlier age of onset, as does Cady *et al.* (2015) who examined 391 cases for 17 ALS genes to discover that 3.8% of patients had more than one mutation and an earlier age of onset by ten years. As discussed, van Blitterswijk *et al.* (2012b) found five families with multiple mutations (5% of their familial cohort) which was statistically more than that expected by chance. They found *C9orf72* repeat expansions jointly with mutations in *FUS* or *TARDBP*. Another study by van Blitterswijk *et al.* (2012c) found a novel *VAPB* variant alongside the *C9orf72* mutation. However the pathogenicity of the *VAPB* mutation was not confirmed and so remains as a variant of uncertain significance. Lastly, Bury *et al.* (2016) published a patient with mutations in both *OPTN* and *C9orf72*. The aggregates within motor and non-motor neurons were studied in this patient to reveal OPTN staining even in cells absent for TARDBP-positive inclusions.

In short, the present study recapitulates previous findings of an increased number of patients with multiple mutations.

6.4. *PON1-3* AND *VEGFA*

6.4.1. OVERVIEW

For the genes *PON1-3* and *VEGFA*, their association with ALS remains with common variation, therefore these areas were selected out of the data and filtered for variants present in dbSNP. Chi-squared SNP-based association tests in cases versus controls were performed using PLINK for these 20 loci. P-values were corrected for the 20 iterations.

6.4.3. RESULTS

The genes *PON1-3* and *VEGFA* have both been reported as potential risk factors for ALS. We selected 20 loci of common variation within these genes to analyse but we did not find any significant differences in SNP frequencies between controls and cases and, in fact, some reported “important” SNPs were present at a higher rate in controls than in cases. It should be noted that the frequencies in our cohort were higher than that observed in the ExAC database.

6.4.4. DISCUSSION

Previously there have been several studies looking at common polymorphisms in the genes *VEGFA* and *PON1-3* which present mixed results in their association with ALS (Lambrechts *et al.*, 2003; Wills *et al.*, 2009). We characterised 20 loci in patients and in controls to find no relationship with ALS but concede that this may be due to low call rates in patients for these particular genes. The variants rs7493 and rs12026 were associated with controls however only 95 patients could be adequately typed for these locations. It was observed that 14 of the 20 loci were found in higher frequencies in our controls compared to public databases (ExAC and 1000 genomes) while 6 of these were considerably higher. This highlights the importance of collecting adequate controls for each study rather than solely relying on these available databases.

The *PON* genes, if truly associated with ALS, have an environmental element to their effect. This may cloud any analysis as a subject’s exposure to different chemicals is difficult to measure accurately. It relies on the patient’s knowledge and is a much more confounded than many other environmental factors like smoking or exercise.

6.5. DISCUSSION

6.5.1. OVERVIEW

We performed NGS, microsatellite repeat allele sizing, and repeat-primed PCR on 1736 subjects across 24 genes associated with ALS. 138 variants were detected which have been previously published in ALS, other diseases or controls while 845 rare SNPs of uncertain significance were also found. In this discussion I will focus on the coding variation uncovered since non-coding has been discussed in Section 6.2.

6.5.2. *SOD1*

Since the publication of many *SOD1* mutations, the recognised functional transcript has been altered by one amino acid resulting in discrepancies in the numbering of mutations between recent publications and those published more than a year ago.

Within *SOD1*, I found 11 unique exonic variants in 21 individual. All but two were strictly observed in patients; these were the synonymous variant A141A (previously A140A) which has been published in both patients and one control (Blumen *et al.*, 2010; Weber *et al.*, 2012; Di Vito *et al.*, 2013; Gamez *et al.*, 2006) and D91A (D90A) which was heterozygous in 2 patients and homozygous in 1 control. This D91A finding does not match the published literature in that homozygous variants are assigned with causing disease while, in a heterozygous state, it is hypothesised to be either benign, a risk factor, have reduced penetrance or be influenced by genetic modifiers to alter inheritance patterns (Robberecht *et al.*, 1996; Al-Chalabi *et al.*, 1998; Winter *et al.*, 2000; Luigetti *et al.*, 2009; Luisa Conforti *et al.*, 2009; Andersen *et al.*, 1995). Al-Chalabi *et al.* (1998) explored the haplotypes of reported dominant and recessive pedigrees to discover that all families associated with a causal homozygous D91A had a common founder. The authors suggest that this haplotype contains a protective variant which inhibits disease in the heterozygous form. However, this theory needs to be tested in a sufficiently large control cohort. It has also been hypothesised that this mutant co-segregates with another deleterious variant outside of the Scandinavian populations where this mutation is most common (van Blitterswijk *et al.*, 2012d). Felbecker *et al.* (2010) reported two large families with the D91A variant however this did not segregate with the disease. In both pedigrees only three of five affected individuals were homozygous

for this mutation whereas the others were homozygous for the wild-type allele. This collection of evidence points towards D91A and A141A not being involved in the disease and, correspondingly, both are the only two *SOD1* variants identified in this cohort which are also found in public databases, albeit rarely.

The only novel variant found within *SOD1* is T40A which was present in a sporadic individual. This is located between the known L39R and G42S mutations (previously L38R and G41S; Brown *et al.*, 2012; Boukaftane *et al.*, 1998; Andersen *et al.*, 2003; Millecamps *et al.*, 2010b). D77Y was discovered in three patients while the most common *SOD1* variant in our cohort was I114T which was ascertained in five patients (six including the test plate results).

The total number of *SOD1* mutations in familial patients (7.3%) is significantly lower than previously published however this may be due to *SOD1* being one of the primary candidate genes for initial sequencing when a familial patient arises in the clinic and, if found, might result in their exclusion from further studies.

6.5.3. *TARDBP*

In *TARDBP*, 16 coding variants were found in a total of 23 patients and 6 controls. One of these is located on a splice site and ten are non-synonymous, with A90V being the only one in this group found in controls. This variant is also present in ExAC and ESP (0.02%) while we located it in 0.2% of patients and 0.3% of controls. Previously, it was published in an ALS-FTLD case and a single control in 1,385 individuals (Winton *et al.*, 2008). This paper also presented *in vivo* functional work on this variant showing that 22% of A90V-transfected cells have abnormal localisation of TARDBP compared to almost no mislocalisation with the wild-type protein. This led the authors to postulate that A90V is a risk variant for ALS-FTLD. However, other groups reporting this mutation all exclusively observe it in controls with 1/185 in France and Quebec, 1/872 in British and Australian and lastly 2/806 in a Caucasian cohort (Kabashi *et al.*, 2008; Sreedharan *et al.*, 2008; Guerreiro *et al.*, 2008). It is possible that some mislocalisation of TARDBP is tolerated and does not cause disease.

The only other non-synonymous *TARDBP* variant found in public databases is G287S. This variant is in a single subject in ExAC amongst 60,704 individuals. We ascertained G287S

solely in sporadic individuals as did two previous publications (Kabashi *et al.*, 2008; Corrado *et al.*, 2009). Therefore it is still possible that this variant is pathogenic.

Stopgain mutations obviously have a severe consequence on the protein which is eventually synthesised and are likely causal in genes of high importance like *TARDBP*. One such variant is Y374X which we detected in a familial subject and has been verified before and creates a damaging truncated protein (Daoud *et al.*, 2009; Del Bo *et al.*, 2009). Unusually, one patient had two novel mutations in this gene: the E14K alteration in exon 2 and N179D which is in exon 4.

Finally, four known alterations were uncovered in familial (M337V, G348V and N378D) and sporadic (A321V) patients which have all been published in at least two studies and so are likely to be pathogenic (M337V: Sreedharan *et al.*, 2008; Tamaoka *et al.*, 2010; Corrado *et al.*, 2009; G348V: Kirby *et al.*, 2010; Zou *et al.*, 2012b; N378D: Tsai *et al.*, 2011; Ticozzi *et al.*, 2011; A321V: Kirby *et al.*, 2010; Cooper-Knock *et al.*, 2012).

6.5.4. *FUS*

The two main mutation hotspots in *FUS* are within the C-terminal domain and in the glycine-rich region. The most commonly reported codon in ALS is R521 which in our patient cohort has been mutated into R521C, R521H and R521L. These are all known variants and have been published in multiple studies (R521C: Vance *et al.*, 2009; Tateishi *et al.*, 2010; Suzuki *et al.*, 2010; Morgan *et al.*, 2015; R521H: Vance *et al.*, 2009; Blair *et al.*, 2010 R521L: Deng *et al.*, 2010; Zou *et al.*, 2012a).

Within the glycine-rich domain is a S221 deletion found in a single case. Lattante *et al.* (2012) also reported this variant in cases and not 793 controls however it is located in close proximity to the G223 deletion reported by Belzil *et al.* (2011b) in three controls potentially reducing the likelihood that S221del is damaging as some deletions in this area are clearly tolerated. Accordingly, 15 indels are presented in ExAC among 254 individuals between codons R216 and G231. Furthermore, at the beginning of this glycine-rich domain is G167-168del which we identified in a control subject and so is also likely to be benign. Another deletion found only in

controls is Y55-56del which is near the reported S57delTCT which is hypothesised to be causal (Belzil *et al.*, 2009).

A single sporadic individual harboured the P431L variant which has been published in essential tremor (ET) in a single case (Merner *et al.*, 2012). This group performed exome sequencing on a family with ET to uncover Q290X in *FUS* as segregating with the disorder. No LoF mutations have ever been reported in controls and only two stopgain alterations (R495X and Q519X) and 11 frameshift indels which cause a premature stop codon (all between G466 and Q519) have been detected in ALS (Deng *et al.*, 2014). The Q290X modification is located in exon 9 whereas all the ALS associated variants mentioned are in exons 14 or 15. Functional work revealed that FUS expression was significantly lower with Q290X than with ALS-associated mutations and could be restored by blocking nonsense-mediated decay which was not the case for the ALS mutants. This indicates different mechanisms underlying the two diseases and therefore it seems improbable that mutations could cause both. The P431L variant in the single familial case with ET could not be verified for segregation amongst other family members but given that we have exposed this alteration in ALS, I think it is unlikely to cause disease alone.

We identified S135N in a single control, which has been reported by Rademakers *et al.* (2010) in a sporadic case of ALS but the authors did concede that it was present in dbSNP and so unlikely to be pathogenic. Later, Huey *et al.* (2012) found S135N in one of 659 controls confirming this theory.

G507D is a variant which ticks many boxes for being considered pathogenic, it is absent from all control databases, it is predicted to be damaging *in silico* and is found in multiple studies, including our own, in cases and not controls (Corrado *et al.*, 2010; Lai *et al.*, 2011). While R487H is present one of our controls and in two individuals from ExAC and the 1000 genomes. This codon has been reported before in a single sporadic case in 1,192 individuals as R487C which was not in 970 controls (van Blitterswijk *et al.*, 2012d).

6.5.5. *VCP*

The full exonic regions of *VCP* were sequenced in our cohort revealing 17 coding SNPs. All of these are rare and 9 of which are non-synonymous. Examining the data in ExAC for this gene reveals a significantly low number of polymorphisms amongst the general public for both missense and loss-of-function variants (z-score = 6.47 and pLI = 1). This indicates that these types of mutations are more likely to be damaging.

Two of the coding variants were both non-synonymous and found only in patients, namely, G523V which is novel and the published I114V which I have identified in sporadic individuals (Koppers *et al.*, 2012; González-Pérez *et al.*, 2012). G523V is predicted to be disease-causing by all algorithms while I114V is not. This latter variant was previously found in a familial patient and in the unaffected side of the family suggesting it is benign (González-Pérez *et al.*, 2012).

The I27V alteration was initially reported in two patients, one with FTD and the other with isolated progressive dysarthria (Rohrer *et al.*, 2011). This variant was not present in their 451 healthy controls. Later, Beck *et al.* (2014) also identified I27V in a single dementia case who also had a mutation in *PSEN1*. Although this variant is rare, predicted as damaging and lies close to a cluster of known pathogenic variants, we demonstrate it within a control subject as does Majounie *et al.* (2012) who also found it in a PD patient. Following this, Weihl *et al.* (2015) published I27V as causing inclusion body myositis (IBM). In this study they identify R95C and I27V and compare these to the known R155H mutation for *in vitro* analysis. They report that only R115H and R95C increase ATP hydrolysis but that all three mutations increase the expression of P62 and LC3II which have been linked to disease causation. However, examining the western blots presented, it is clear that I27V produces less of these proteins particularly LC3II. Therefore I27V may not be sufficient alone to be pathogenic. Lastly, H404P and A698P were both only located in controls and reside within exon 11 and 15 respectively. They represent very rare benign polymorphisms.

6.5.6. *OPTN*

Rare variants detected in optineurin totalled 24, with 19 of these being non-synonymous. Two are stopgain variants, two are on splicing sites and two are frameshift deletions. One

frameshift and one splicing change was found in a control subject and the rest were constrained to patients. The Q441X in a sporadic individual lies next to the reported K440fs in a family with an aggressive phenotype (Weishaupt *et al.*, 2013). This frameshift variant introduces a stop site 8 amino acids later and potentially disrupts the UBA domain. The other stopgain variant we detected was Q146X which was observed in the homozygous state in a single sporadic case. Amongst the 19 found mutations, the only variant to be previously reported as causing ALS is Q398E which we found in a control patient (Kenna *et al.*, 2013).

M98K, N303K and R545Q were uncovered in both cases and controls and have been previously identified as causing glaucoma (Rezaie *et al.*, 2002; Melki *et al.*, 2003; Sripriya *et al.*, 2006; Caixeta-Umbelino *et al.*, 2009; Weishaupt *et al.*, 2013; Buentello-Volante *et al.*, 2013). Since mutations causing this disease do not overlap with ALS, it seems implausible that these are disease causing in this instance. The M98K variant was in 5.5% of 605 controls, 5.5% of 992 patients and 10% of patients sequenced on the test plate (n = 94) clearly presenting the ease with which to find differences in variant frequencies when using low numbers of subjects. Notably, studies finding a difference between glaucoma patients and controls all used 100-200 patients. Rezaie *et al.* (2002) observed this variant in 13.6% of glaucoma patients (n = 169) compared to 2% of controls which was significantly different. However the latest update in EVS puts this variant at 11.8% and 1000 genomes marks it as 7%. Sripriya *et al.* (2006) compared a frequency in glaucoma of 4-6% in different cohorts to their 100 controls who were all absent for this variant. Given these new findings, it seems reasonable that these papers were reporting false positives due to being under powered. Rezaie *et al.* (2002) reported R545Q in 2.2% of POAG subjects while Weishaupt *et al.* (2013) found this variant in ALS and hypothesised a link between variants causing glaucoma also being risk factors for ALS, however, given there is only a single paper which reports these two conditions occurring together, it would seem unlikely that they are affiliated by genetics. Additionally, functional work on known pathogenic variants display a different underlying mechanism for glaucoma and ALS (Maruyama *et al.*, 2010).

A single control subject harboured the K557R alteration which is located at the same codon as the published K557T (Del Bo *et al.*, 2011). K557 is located within the C-terminal zinc-finger domain which is highly conserved and is hypothesised to indirectly affect apoptosis. Both variants are rare and predicted to be damaging. K557T was in a single familial patient with no further evidence to support its pathogenicity. Finding this codon mutated in a control subject

lowers the probability of the K557T variant being pathogenic, but only slightly. Functional studies are needed to understand this mutation further.

A single sporadic subject harboured R271H which is in the same codon as the published R271C (Iida *et al.*, 2012) although both these variants have only been found in patients, they are not evolutionarily conserved and are predicted as benign so Iida and colleagues suggest their variants is a rare polymorphism.

6.5.7. *ANG*

Two controls presented with the known K41I variant (previously K17I) which has been found across ALS patients from different populations (van Es *et al.*, 2009; Greenway *et al.*, 2006; Cady *et al.*, 2015). Assays examining this variant have exhibited a complete loss of ANG function (Wu *et al.*, 2007) yet we found no patients with this variant. *ANG* remains controversial as an ALS gene with some groups questioning its role in the disease. K41I is present in all control databases and ExAC places it in 0.1% of its individuals (172 of 60,705). This is the same as the initial *ANG* variant first reported as segregating with disease in a family linked to the region 14q11.2 (Greenway *et al.*, 2006). The authors did not sequence this entire area and instead picked *ANG* as a candidate gene. It is therefore possible that they missed the real causal agent. However the authors do note that this association was only present in Irish and Scottish populations while absent from US, English and Swedish ALS cases.

K78E was verified in two subjects and previously reported as a homozygous variant in a sporadic individual while not present in 616 controls (Fernández-Santiago *et al.*, 2009). K84E is observed in one sporadic and six controls in our cohort while previously only being reported in ALS (Brown *et al.*, 2012; Cady *et al.*, 2015). Although, the former study sequenced 55 controls for this region and the latter did not include a healthy cohort while both sequenced more than 1000 patients.

6.5.8. *ALS2*

As a recessive cause of ALS, the list of potentially causal variants are fewer in number for *ALS2*. As a result, only a single patient has a potentially causal mutation: the homozygous frameshift variant W1179fs (c.3536delG) in exon 22. The age of onset of this subject was 55. Since no reports in this gene have been associated with an adult-onset ALS, either something has protected this patient or this variant is not the cause of their disease.

All other rare variants of interest were heterozygous. However, firstly, *ALS2* was not covered in its entirety since the sequencing was restricted to previously reported mutation hotspots. Therefore it is conceivable that coding variants were missed and in fact some of these individuals were expressing another mutation. Secondly, heterozygous variants in ALS2 have not been examined as potential risk factors for ALS in humans but studies in mice and zebrafish indicate that variation in *ALS2* may be a risk factor for ALS (Cai *et al.*, 2005; Gros-Louis *et al.*, 2008).

6.5.9. *NEFH*

Of the 132 variants documented in *NEFH*, 84 had a read count of below 20 and only 24 of these were awarded a quality score of above 50. Obviously these variants must be examined with caution and highlights the drop in ability with NGS to tackle long repetitive sequences as recognised in this gene.

A number of missense and synonymous variants were uncovered in our cohort however almost all of the published literature focus on indels in the KSP-region of *NEFH.* The exception to this is the study by Daoud *et al.* (2011) who report missense variants in ALS of unknown consequence with no functional work and no comparison to control databases. Examining ExAC reveals that some of these variants appear in low frequencies and lots of natural variation occurs in this gene.

One of the previously published indels in *NEFH* which is relatively accepted by some as causal, K790del (c.2368-2370del; Figlewicz *et al.*, 1994), was found in a single sporadic

individual but also in two controls throwing some doubt onto this variant's pathogenicity. Two other indels were identified, Q465-469del in one familial (c.1394-1405del) and E500delinsETK in a sporadic patient (c.1498-1499insAAACAA). Both are absent from EVS and 1000 genomes but present in ExAC (0.004% and 0.2% respectively). Although both of these frequencies are below our arbitrary cut-off, the latter is close.

6.5.10. *VAPB*

Only three rare variants were uncovered in *VAPB* and none of these appear to have any involvement in the disease. A deletion of three nucleotides across codons 158-159 (c. 474-476del) was observed in eight sporadic individuals and four controls. This was located next to the known S160del (Landers *et al.*, 2008a). However this change is also in public databases, albeit rarely (0.45%), and does not disrupt the localisation of VAPB in functional studies (Landers *et al.*, 2008a). The missense transition M170I is found in both ALS (van Blitterswijk *et al*., 2012b; Cady *et al*., 2015) and a small number of controls (van Blitterswijk *et al*., 2012b) so it was pondered to be a risk factor. We detected it in 5 sporadic (0.56%) and 3 control subjects (0.5%) which indicates a more benign nature of this mutation. Finally, the recognised R184Q alteration was located in a single sporadic individual, previously found in only PD (0.04%; Kun-Rodrigues *et al.*, 2015) rather than controls, however it is present in ESP at a similar rate (0.03%) and in our patient cohort (0.1%).

6.5.11. *DCTN1*

Within the *DCTN1* gene, a total of eight non-synonymous, rare alternative alleles were detected. The R785W substitution has been declared by Münch *et al.* (2004) in two familial patients and two unaffected relatives. The authors suggest the possibility of a reduced penetrance however in the current cohort, one patient and two healthy controls also had this change indicating it is benign.

The T1249I variant is an example of a rare modification which was connected to ALS in studies that sequenced none or few controls whilst we uncovered it in 4 control subjects and no patients (Münch *et al.*, 2004; Cady *et al.*, 2015). Stockmann *et al.* (2013) expressed this mutant

in cells and found no complications of interest compared to causal variants providing further evidence towards a benign character. The three novel substitutions T12A, V73I and V1081M all occurred in a single patient and were absent from both our controls and all public databases making them good candidates to perform functional assays on. The latter of these is close to the published V1081M found by Cady *et al.* (2015) however this study did not sequence controls so the variant has very minor evidence to support its role in ALS.

Previously, *DCTN1* missense variants have been detected in 3% of fALS patients yet none segregated with the disease so alterations should be examined with caution (Vilariño-Güell *et al.*, 2009). In contrast, Stockmann *et al*. (2013) found that mutations which occurred mostly in patients plus a small number of controls displayed altered protein function and sometimes no function at all. The authors propose that *DCTN1* is a modifier of ALS rather than a disease causer.

6.5.12. *SETX*

The results for *SETX* included a number of published mutations of varying consequence. K1425E was previously found in two affected members of a family with inherited peripheral neuropathy, however, the authors note that this variant is predicted as benign and correspondingly we located it to two control subjects (Drew *et al.*, 2015). I2547T was reported in the same paper as not segregating with the disease and in MND patients in other studies (our cohort; Hirano *et al.*, 2011; Cady *et al.*, 2015) and in five of 340 controls (our cohort; Arning *et al.*, 2012).

Two variants next to each other presented in patients (D1553G) and both patients and controls (C1554G). The latter of these is described by Hirano et al (2011) in a single patient and not in 100 healthy controls in addition to two other studies which did not use controls (Cady *et al.*, 2015; Ghani *et al.*, 2015). While Arning *et al*. (2012) did examine controls finding 1/240 harbouring this variant. This highlights the importance of using an adequate number of controls to reduce false positive results.

One sporadic individual possessed V2549A which is located on the same codon as V2549I found in two patients and was absent from 305 controls (Kenna *et al.*, 2013). We also found

the similar A1478V and A1478E in patients only, with the second substitution previously segregating in two affected and two unaffected subjects within an ALS pedigree (Arning *et al.*, 2012). However they also report it in a German control cohort (1/1,090). The D1077N modification was established in a number of patients but failed to segregate in a family with ALS (Arning *et al.*, 2012) and accordingly we located it in both patients and a control subject. Cady *et al*. (2015) sequenced a huge number of patients finding D994G which was absent from public databases however we detected it in two healthy controls. Another two controls had the S2G alteration which is adjacent to the known T3I which was presented in the first *SETX* publication segregating in a large family of 17 members (Chen *et al.*, 2004). This indicates that S2G has not disturbed the same function as T3I. Another six rare, novel variants found only in patients remain open for further investigation, four of which are absent from control databases.

It's hard to say that any of these mutations in our cohort have strong evidence for their pathogenicity.

6.5.13. *CHMP2B*

The most commonly reported pathogenic variants in *CHMP2B* are deletions of the C-terminal and from the six alterations reported in this study, none accomplishes this effect.

The I29V alteration was first published in FTD and 1/100 controls with the authors concluding that this was a rare benign polymorphism (Cannon *et al.*, 2006). However, following that, I29V was published in ALS and reported to be absent from 640 controls and public databases (Parkinson *et al.*, 2006). Cox *et al.* (2010) reported another two MND subjects with this variant and an absence from 500 controls. HEK-293 cells were used in this study to examine the mutation. They determined large cytoplasmic vacuoles in cells containing *CHMP2B*^{I29V}. However, this group also examined other *CHMP2B* variants including T104N which in addition to generating vacuoles, produced inclusions which were positive for CHMP2B, which was not the case for I29V (Cox *et al.*, 2010). Han and colleagues (2012) could not recapitulate this in neuronal cultures with I29V but did observe a detrimental effect with T104N. We have now discovered this variant in two cases and two controls possibly indicating that CHMP2B-positive

inclusions are required for a pathogenic consequence or that other factors play an involvement in I29V patients.

The two variations R69Q and T83I were both reported by van Blitterswijk (2012d) in PMA and ALS respectively, and not in 750 controls. We located the former in a sporadic individual while the latter was in 7% of our cohort. Although the calls have adequate quality scores, this number is suspiciously high for a very rare mutation and needs to be confirmed with Sanger sequencing. The UTR of *CHMP2B* was also partially covered in this study, finding c.-151C>A in 25 patients and 21 controls replicating previous studies (Cox *et al*., 2010).

6.5.14. *UBQLN2*

The first causal mutations described in *UBQLN2* were in a large pedigree with P497H which segregated in the affected family members. We have also found this variant in a single sporadic patient which potentially may be due to a *de novo* mutation or from other P497H carriers in the family dying before the typical age of onset for ALS. Transgenic mice expressing this mutant develop UBQLN2-positive inclusions in the brain and likewise, in HeLa cells, there is increased protein aggregation (Xia *et al.*, 2014; Gorrie *et al.*, 2014). Aside from this substitution which has clear pathogenicity, five novel, rare variants of uncertain significance were uncovered in sporadic individuals: L87F, Q460R, 496-499del, A603D and T334M, the last of these was present in both the heterozygous (n = 2) and homozygous state (n = 1).

6.5.15. *TREM2*

Associated with a number of different neurodegenerative disorders, the *TREM2* R47H variant was genotyped in our cohort finding near identical frequencies between cases (0.11%) and controls (0.1%). Previous work has shown significant differences in the frequency of R47H in PD and FTD patients (1.3-2.1%) compared to controls (0.45%), with a non-significant increase in ALS (0.7%) patients (Rayaprolu *et al.*, 2013). (Pottier *et al.*, 2013) confirmed this in AD as well (2% versus 0.5% in controls). Both these control frequencies are higher than our two cohort groups. The largest study in ALS was completed by Cady *et al*. (2015) who examined 26,871 control subjects for the R47H mutation and 1,685 ALS patients. There was a

significantly higher number of mutations in cases (0.45%) versus controls (0.19%). This group examined individuals from 11 different countries with R47H varying in prevalence between them (0-0.26% for controls and 0.33-0.54% for ALS). We will require many more subjects in our cohort to detect any significant differences in ALS. S16P was also detected in our cohort however, in a heterozygous state, the relevance of this variant is unknown.

6.5.16. *PFN1*

Examining *PFN1* in our cohort reveals the E117G variant in both cases (0.11%) and controls (0.99%). Formerly, this alteration has been established by different groups as appearing in cases more than controls (Table 17), with a large meta-analysis finding a significant association with ALS (Fratta *et al.*, 2014). Our cohort results do not replicate these findings but we are only a sixth of the total patient number and a twentieth of the controls that Fratta and colleagues examined.

Freq. in cases (no.)	Freq. in controls (no.)	Reference
0.4% (1,090)	0.09% (1,089)	Wu *et al.*, 2012
0.29% (342)	0.17% (1,167)	van Blitterswijk *et al.*, 2013a
0.09% (1,168)	0% (1,512)	Tiloca *et al.*, 2013
0.2% (715)	N/A	Yang *et al.*, 2013
0.9% (328)	0.3% (864)	Dillen *et al.*, 2013
0.25% (5,118)*	0.11% (13,089)*	Fratta *et al.*, 2014
0.11% (880)	0.99% (599)	This study
N/A	0.066% (33,352)	ExAC (European only)
N/A	0.02% (6,503)	ESP
N/A	0.05% (2,504)	1000 genomes

Table 17. Frequencies of the E117G variant in different studies. * This publication also contains results from the other reports.

At the equivalent codon, E117D has been captured by Yang *et al.* (2013) in 0.2% sporadic ALS, however, no controls were sequenced in this study. We identified this change in a single sporadic patient and single control. Like with E117G, it is highly probable we need many more subjects to detect any differences that might be present.

The synonymous polymorphism L112L is mostly ignored in publications except for Chen *et al.* (2013) who found a significant difference in this variant, with more controls possessing the mutation (16% of 550 ALS and 21% in 545 controls). A few groups have reported this variant in ALS (2-26%) but did not sequence controls (Daoud *et al.*, 2013; Lattante *et al.*, 2013a; Yang

et al., 2013). A larger study also reported a minor trend of 3.8% (n = 1,168) in cases and 4.3% (n = 1,512) in controls (Tiloca *et al.*, 2013). However this is not confirmed in our study (6.4% versus 5.1%) or by Zou et al (2013b) 31% versus 22% in cases and controls respectively.

Lastly, we identified a heterozygous W4X variant in a single sporadic case. This is a conserved location and the variant is absent from all control databases. Since the literature specifically focuses on missense variants producing a small risk effect, this stopgain variant would need to be examined further to understand if LoF variants contribute to ALS. There are no LoF variants in ExAC for this gene, however, because *PFN1* is relatively small, only 3.6 are expected and so the constraint metric is not significant. There are also a low number of missense variants (18 in total compared to 68.1 expected).

6.5.17. *SQSTM1*

Eleven patients (1% of this cohort) presented with an alteration at the highly conserved K238E position in *SQSTM1* as did three controls. This has been previously published in ALS (Fecto *et al.*, 2011; Le Ber *et al.*, 2013; van der Zee *et al.*, 2014; Cady *et al.*, 2015) and sporadic ALS-FTD (Rubino *et al.*, 2012) and controls (van der Zee *et al.*, 2014; Le Ber *et al.*, 2013). It is situated within a tumour necrosis factor receptor-associated factor 6 (TRAF6) binding site which is a vital element for the protein's ability to interact with NFκB.

P392L, the most frequent *SQSTM1* variant in ALS, was revealed in 0.3% of patients and 0.7% of controls. Despite appearing in healthy subjects, P392L is still described as being associated with disease appearing in up to 2.3% of fALS, up to 46% of familial PDB and 0.3% of 5,999 controls (Hocking *et al.*, 2002; Laurin *et al.*, 2002; Chung *et al.*, 2008; Fecto *et al.*, 2011; Teyssou *et al.*, 2013; Kwok *et al.*, 2014).

Lastly V153I, which has been reported by Fecto *et al*. (2011) and Cady *et al*. (2015) as only appearing in cases, was found in a single control subject in our study and 0.07% of controls by van der Zee *et al*. (2014). Larger numbers of patients and controls will be required to elucidate whether these variants are risk factors for ALS or not.

6.5.18. SPG11

Although we removed *SPG11* from the gene panel in this study, we had whole-exome sequencing data on most of the controls which included coverage of this gene. ExAC reveals that this gene has an increased number of missense variants compared to the expected mutation rate. In our control cohort, 24% of individuals had a homozygous variant in this recessively causal gene, while 1% had five mutations and a single control subject harboured the stopgain variant W683X. These results highlight the difficulty in interpreting variation within *SPG11* and caution should be taken when segregation analysis is unavailable.

6.5.19. *PRPH*

Despite 12 variants unearthed in *PRPH*, no homozygous alterations were discovered in patients within this recessively-inherited gene. The only PRPH mutation to be consistently reported in ALS is D141Y (Leung *et al.*, 2004; Gros-Louis *et al.*, 2004; Corrado *et al.*, 2011). However despite conservation of this amino acid down to zebrafish, only one of these studies claimed definitive causation with a homozygous alteration (Leung *et al*., 2004). In publically available genetic datasets, this variant is reported to occur in between 0.05-0.1% of the general population whereas our study revealed an incidence of 2.9% of fALS and 1.1% of sALS and controls. Similarly, in sporadic subjects, it has previously been established in 1.5% of cases (Gros-Louis et al., 2004).

A study investigating the effect of PRPH expression on a *SOD1* mouse model found that neither diminishing nor augmenting PRPH protein levels affected any common outcome measure, for example, severity or onset, which implies little if any involvement in ALS (Larivière *et al.*, 2003).

6.5.21. *FIG4*

Five missense variants were detected in *FIG4* with only one of these (I41T) previously being described in CMD (Lenk *et al.*, 2011; Chow *et al.*, 2007) but in combination with a null allele.

FIG4 exhibits recessive inheritance in CMT disease yet heterozygosity had been associated with ALS (Chow *et al.*, 2009). In this study, they described two stopgain, two splice site and two missense mutations which they deemed as deleterious while four other missense variants displayed little or no damaging effect in the functional studies the researchers performed. It seems less likely that missense mutants are damaging in this case. However, we uncovered the novel homozygous D206G aberration in a sporadic patient which is absent from all control databases and predicted to be damaging by *in silico* tools. The patient is classed as probable ALS and his age of onset was 51. It would be interesting to perform a functional assay on this variant.

The only other report of significance published five rare variants in a cohort of 698 sALS which were all either absent from public databases or present at very low frequencies (Cady *et al*., 2015). There are very few other publications exploring this gene with respect to ALS, however, it has been associated with causing Yunic-Varón syndrome, a recessive, infantile disorder which mainly affects the bones, indicating diverse roles of *FIG4* within the body (Campeau *et al.*, 2013).

6.5.21. *DAO*

As far as the authors are aware, the only known pathogenic *DAO* variant R199W has not been replicated until the present study (Mitchell *et al.*, 2010). This may be due to the fact that few studies have sequenced such a large cohort of patients for this gene. The familial patient we identified harbouring the R199W variant was female with a disease onset at age 44. Additionally, another female patient with sporadic ALS had a similar mutation at the same codon: R199Q. However the disease onset was much later at 71. Lastly, another alteration two amino acids away from these was detected: Q201R, in a sporadic ALS patient with an onset at 59. All three of these mutants are predicted as damaging by PhyloP, SIFT, PolyPhen and LRT, and did not occur in our control samples.

6.5.22 *C9ORF72*

As the most common cause of ALS, it was necessary to include *C9orf72* in the present study. As a result, 15.2% of familial cases sequenced had the expansion as well as 6.9% of sporadic subjects and none of the controls. This is slightly lower than expected, for inherited ALS, which could be due to the small number of fALS individuals included, only 72 were adequately typed for the hexanucleotide repeat. This may have introduced a sampling error meaning the frequency of mutations is not representative of a larger cohort. The initial studies on *C9orf72* reported expansions in 46% of fALS and 21% of sALS for Finnish patients (Renton et al., 2011), and 23.5% and 4.1% respectively which were in patients from Canada and USA (DeJesus-Hernandez *et al.*, 2011b). The two highest reports of expansions in familial cases comes from Greek (50%) and Sardinian patients (57.1%) however the former study only included a total of ten subjects (Chiò *et al.*, 2012; Mok *et al.*, 2012). Japanese populations have very few expansions with none reported in familial cases and only 2 in 1,021 sporadic individuals (Ogaki *et al.*, 2012; Nakamura *et al.*, 2016). German and Italian cohorts have been reported with 22-24% of fALS harbouring a pathogenic repeat while the UK is double that at 43% (Ratti *et al.*, 2012; Chiò *et al.*, 2012; Cooper-Knock *et al.*, 2012). This latter result was gauged on only 63 patients so it is likely that we need more familial cases to get an accurate estimate of pathogenic expansions in the UK.

6.6. CONCLUSION

In summary, we have completed a large-scale sequencing study in ALS uncovering an increased number of rare mutations in the UTRs of selected genes, as well as an increased number of patients with multiple mutations potentially causing their disease. This further endorses the oligogenic hypothesis of ALS. Lastly, we believe that caution should be taken when novel variants are revealed in a disease cohort, especially when an inadequate amount of controls have been sequenced.

This work has been accepted in Brain:

Morgan S, Shatunov A, Sproviero W, Shoai M, Orrell R, Fratta P, Hardy J, Pittman A, Al-Chalabi A. A comprehensive analysis of rare genetic variation in ALS in the UK. 2016. Brain.

CHAPTER 7 ARGENTINIAN COHORT

7.1. INTRODUCTION

Next-generation sequencing studies have disproportionally examined the genetics Western and Asian populations over African and South American individuals. This is especially the case for the genetics of rare diseases like ALS which affects people worldwide. I was given the opportunity to sequence a small cohort of Argentinian ALS patients to search for potential disease-causing variation. Previous work into the genetic makeup of Argentina revealed a heterogeneous population with a composition of 65% European, 31% Indigenous American and 4% African (Avena *et al.*, 2012). Another paper has analysed *C9orf72* in a small Argentinian ALS cohort to discover expansions in a single (1/47) sporadic subject and a single (1/3) familial case (Itzcovich *et al.*, 2016). The aim of the present study was to explore the genetics of 26 ALS patients from Argentina.

7.2. MATERIALS AND METHODS

A total of 26 ALS patients from Buenos Aires, one of which had ALS-FTD, were analysed. Firstly for *C9orf72* repeat expansions and then for 24 relevant genes using the panel described in Chapter 6. Subjects included in the study consisted of 8 familial ALS, 19 sporadic ALS and 1 sporadic ALS-FTD. The average age of onset was 61 (range 45-83) and the ethnicity was self-reported as Latin except for two of Caucasian heritage. However, the clinician who examined these patients believes all but one patient to be of European heritage. Additionally *ATXN2* screening was performed on all subjects to detect repeat expansions. My colleague, Lucia Schottlaender, performed the analysis of *C9orf72* and *ATXN1* using methods previously described (Koutsis *et al.*, 2012; Schottlaender *et al.*, 2015).

7.3. RESULTS

Two patients harboured the *C9orf72* repeat expansion (8%). In the remaining 24 subjects, 122 variants were uncovered by NGS, 75 of which were common SNPs (present in >1% of the population), 14 were synonymous, 34 were intronic and 9 were rare, coding mutations. Three

patients had an *ATXN2* intermediate repeat. Of the 24 patients fully sequenced, 8 had potentially damaging modifications in *SOD1*, *TARDBP*, *FUS*, *UBQLN2*, *VCP*, *CHMP2B*, *SETX* and *ATXN2* which may have caused their disease (Table 18-19).

Subject	Gene	Variant	Disease	El Escorial	Site of onset	Age of Onset	Evolution (months)	*ATXN2* screen	Gender	Ethnicity
1	*SOD1*	G86S	fALS (probable)	Possible	Limbs	50	17	Normal range	F	Latin
2	*FUS*	S135N	sALS	Possible	Limbs	74	24	Normal range	M	Latin
3	*SETX*	D1077N	sALS-FTD	Probable	Bulbar + FTD	72	30	Normal range	F	Caucasian
4	*VCP*	R155H	fALS (probable)	Probable	Limbs	45	48	22/27 repeats	M	Latin
5	*TARDBP*	N378D	fALS (probable)	Definite	Limbs	52	96	22/27 repeats	F	Latin
6	*FUS*	D490N	fALS (probable)	Definite	Bulbar	48	7	22/27 repeats	F	Latin
	UBQLN2	P497S								
7	*SOD1*	D84G	fALS (probable)	Possible	Limbs	47	24	Normal range	M	Latin
8	*CHMP2B*	R32Q	fALS (probable)	Probable	Limbs	83	24	Normal range	M	Caucasian
9	*SOD1*	E22G	sALS	Probable	Limbs	48	36	Normal range	M	Latin
10	*SOD1*	E22G	sALS	Possible	Limbs	67	15	Normal range	M	Latin
11	*SOD1*	E22G	sALS	Definite	Bulbar/ upper limbs	56	48	Normal range	F	Latin
12	*C9orf72*	Expansion	sALS	Probable	Limbs	65	20	Normal range	M	Latin
13	*C9orf72*	Expansion	sALS	Definite	Bulbar	67	24	Normal range	F	Latin

Table 18. List of Argentinian patients with rare mutations.

Subject	Gene	Variant	Nuc. change	Exon	Chr	Position	Genotype	dbSNP	*in silico* prediction	Our verdict
1	*SOD1*	G86S	G256A	4	21	33039587	het	-	D	D
2	*SOD1*	E22G	A65G	1	21	33032147	het		U	D
3	*FUS*	S135N	G404A	5	16	31195598	het	rs61732970	T	T
4	*SETX*	D1077N	G3229A	10	9	135203756	het	rs145097270	U	T
5	*VCP*	R155H	G464A	5	9	35065360	het	rs121909329	U	D
6	*TARDBP*	N378D	A1132G	6	1	11082598	het	-	U	D
7	*FUS*	D490N	G1468A	14	16	31202358	het	-	U	D
	UBQLN2	P497S	C1489T	1	X	56591795	hom	-	U	U
8	*SOD1*	D84G	A251G	4	21	33039582	het	-	D	D
9	*CHMP2B*	R32Q	G95A	2	3	87289909	het	-	U	U

Table 19. Details on rare mutations uncovered; D = damaging; U = unknown; T = tolerated; Nuc = nucleotide.

7.4. DISCUSSION

A selection of 26 ALS patients from Argentina were scrutinised in 26 genes of interest for any disease-causing mutations. 13 of these patients have one or more likely damaging variants which may have contributed to their disease. Two of these included the *SOD1* mutations D84G and G86S which are both in exon 4. The latter of these has been observed previously as an aggressive ALS phenotype in a Japanese kindred (Takazawa *et al.*, 2010). This matches the clinical progression of our patient who was diagnosed within a month of onset and lived for only 17 months.

Another patient presented with the known alteration N378D in *TARDBP* (Tsai *et al.*, 2011; Ticozzi *et al.*, 2011). This patient also had an intermediate CAG repeat expansion in *ATXN2* which is known to increase a patient's risk of ALS (Elden *et al.*, 2010). However, at 27 repeats, this is at the cusp of what is considered a risk with many papers declaring a higher number in order to get significant results (van Damme *et al.*, 2011; Conforti *et al.*, 2012). For this reason it has been suggested that the lower limit may diverge in different populations (Lee *et al.*, 2011). A *FUS* S135N variant in a sporadic sample was established which has both been observed in ALS (Rademakers *et al.*, 2010) and in one control subject (Huey *et al.*, 2012) indicating it is likely to be benign. Additionally, another patient presented with a novel *FUS* D490N variant which lies in the RGG-rich domain between the disease-causing variants R487C and R495X (van Blitterswijk *et al.*, 2012b). This patient additionally has an *ATXN2* intermediate expansion and the homozygous P497S variant in *UBQLN2*. This variant is located on the same codon as P497H which was shown to impair the protein degradation pathway in cell cultures (Deng *et al.*, 2011).

VCP revealed only one known mutation which was found in an individual with a family history of ALS: R155H which has been reported previously in both ALS and IBMPFD as one of the most common *VCP* variants (Johnson *et al.*, 2010; González-Pérez *et al.*, 2012; Cady *et al.*, 2015; IBMPFD: Watts *et al.*, 2004; Viassolo *et al.*, 2008; Jacquin *et al.*, 2013). Examination of *CHMP2B* revealed the novel variant R32Q which coincided with an intermediate *ATXN2* repeat expansion. The subject with FTD and Caucasian ethnicity harboured a D1077N variant in *SETX* however this variant did not segregate in a family with ALS (Arning *et al.*, 2012). Lastly there were 34 intronic variants of unknown significance.

Excluding the two SNPs which were questioned in previous publications, the incidence of potentially damaging variants in our ALS cohort for each gene stands at 8% in C9orf72, 19% in *SOD1*, 12% in *ATXN2* and 4% in *TARDBP*, *FUS*, *UBQLN2* and *VCP*. This amounts to 34% of patients able to be genetically explained, which includes 62% of familial cases and 28% of sporadic subjects. This is higher than might be expected which could be due to the low subject count.

Additionally, the frequency of multiple mutations occurring in the same patient stands at 12%. This was entirely from the familial subjects. Most publications examining the prevalence of *ATXN2* repeat expansions excluded patients where a mutation was already discovered in another gene. In the present report, all three of the subjects with an intermediate expansion had a mutation in a known ALS gene, therefore, these would have all been excluded in other studies.

As the number of patients in this cohort are relatively few, a larger replication study needs to be completed. NGS has shown that the genetics of ALS within South American countries are potentially similar to that of European decent with *SOD1* being the most common cause of ALS. Given that there is a chance most of these patients are in fact of European decent, this finding is not surprising. We will be performing further analysis on these patients to determine their genetic makeup.

This work is currently in preparation for submission:

Morgan S, Schottlaender L, Hardy J, Holden H, Pittman A. Comprehensive investigation of causal genes in an Argentinian cohort with amyotrophic lateral sclerosis. 2016.

CHAPTER 8 MACHINE LEARNING

8.1. INTRODUCTION

With the discovery of *C9orf72*, the majority of patients with a familial classification can now be explained genetically. The latest estimations of the number of patients with known genetic causes puts definite fALS at 61-81% (two relatives also affected), probable fALS at 27-66% (one affected relative), possible fALS at 11-40% (one distantly affected relative or relative with similar disorder) and lastly sALS at 11-28% (Sabatelli *et al.*, 2016). However, there are still many unsolved cases and with recent work highlighting the involvement of multiple genes within some patients, unsolved cases may be due to the interaction of multiple variants. One method to investigate this is to use machine learning which is able to uncover high-dimensional genetic patterns that could predispose an individual to ALS. I hypothesise that in the complex genetic interplay of this condition, genetic interactions combine with the overall mutation burden to determine the risk and course of ALS, in a way that high-dimensional analyses might reveal.

Machine learning describes high-dimensional pattern recognition software that is often termed as *intelligent* based on its ability to learn information and patterns, and make informative predictions from highly complex data without explicitly being programmed to do so. Big companies like Google and Amazon have invested heavily in machine learning techniques and now depend greatly upon them. The basic principle of this method is to be presented with some data (a training set), build a model (learn about the data) and make decisions or predictions that are driven by the data (Figure 55). The computer must come up with its own program for solving this problem. There are a vast array of underlying algorithms to achieve this, each with various advantages and disadvantages. However the main rule for machine learning is that you need big data. This data can be split into training data (for the algorithm to learn from) and test data (to implement the learned model on and test its performance). If the technique performs well then it can be applied to a new cohort of data for validation of the model.

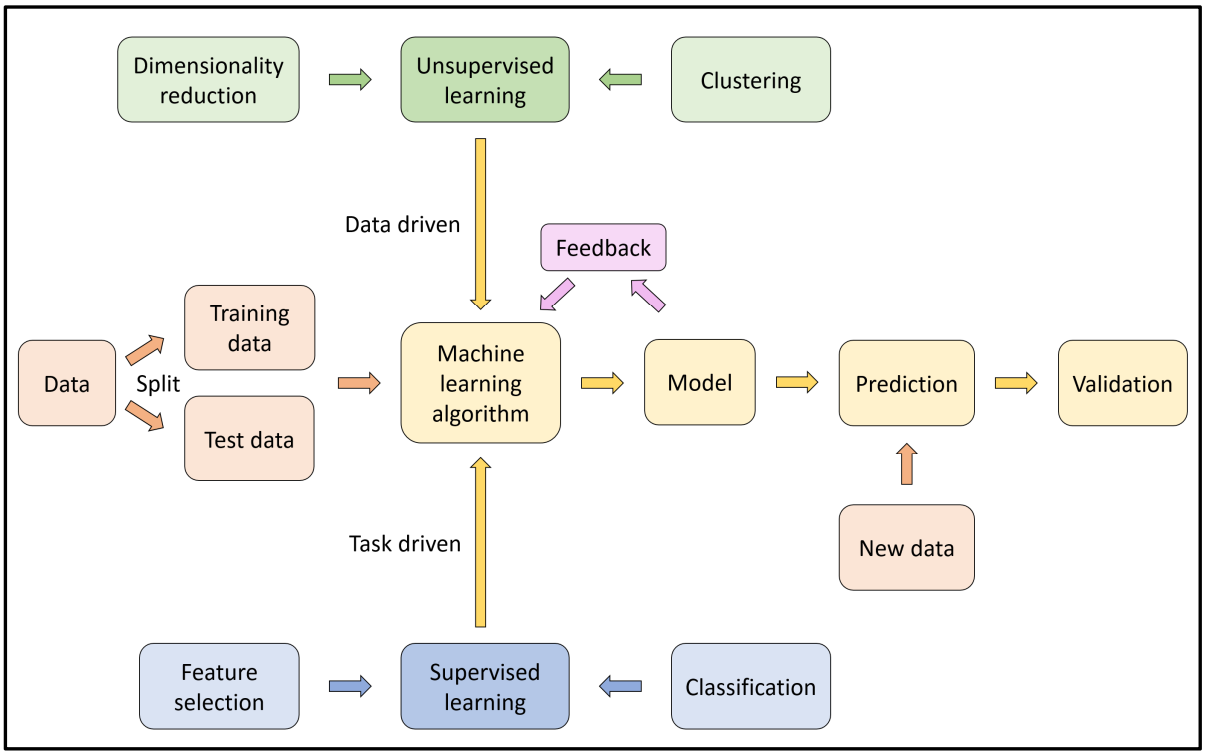

Figure 55. Flowchart on the process involved with machine learning techniques. Drawn in Powerpoint 2013.

The types of machine learning I will be focusing on include unsupervised and supervised learning methods. *Unsupervised* describes data exploration, where the machine is given the whole dataset and is tasked with finding underlying relationships within the data. For example, Amazon take all the information about their customers to cluster people into groups of individuals that are similar. Similar people are more likely to buy the same products and so this information is used to help recommend new items to each customer. *Supervised* learning is where the computer is given the data, but this time it comes with labels. For example, you might present an algorithm with spam emails and non-spam emails (two labels) and have the machine learn the *patterns* in these types of email. This learning can then be applied to novel emails where the classification of the label is unknown to predict which emails are spam, therefore creating a spam filter which can be constantly learning and improving from new emails. The more data these algorithms receive, the easier it is for them to make intelligent decisions.

As the first step in examining my data, I wanted to visualise any underlying relationships that might be present. A traditional method which can achieve this is principal component analysis (PCA). This is often used in genetics for examining population stratification and accomplishes this by dimensionality reduction i.e. the thousands of parameters (mutations in my case) are expressed on two dimensions allowing for each subject to be represented by a single point and can display a clustering of individuals who are genetically similar (Figure 56). Humans find it easier to think and observe in up to three dimensions, however, computers are perfectly able to deal with data in millions of dimensions. In order to generalise across these multiple dimensions, you need representations from all of the features present. PCA completes this task by plotting all of the traits against each other and creating multiple axes to fit this data. The first two axes will capture most of the variation in the data and so these form the basis of a two-dimensional representation of the initial data. I therefore used PCA to examine my data.

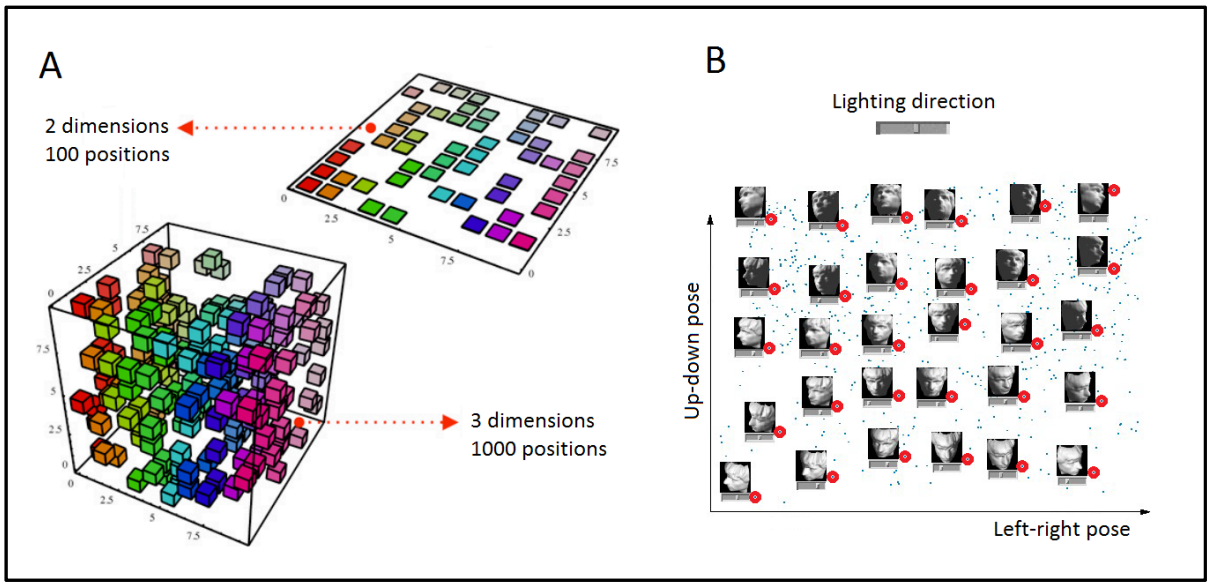

Figure 56. Examples of dimensionality reduction. (A) Schematic representing how you might scale from three dimensions to two (Hunt, 2015). (B) Plotting points which represent images of a face on three dimensions, namely, orientation of the face up-down, left-right pose and the lighting direction. Red circles display location of the associated picture nearby (Tenenbaum *et al.*, 2000).

An award-winning algorithm for visualising high-dimensional data is t-Distributed Stochastic Neighbour Embedding (t-SNE; (Maaten and Hinton, 2008). This dimensionality reduction method outperforms the popular PCA by placing importance on local structure and correcting for large disparities in the sizes of features. I implemented this algorithm in MATLAB (R2015b).

The next step in my work was to perform supervised classification in an attempt to make binary predictions about data of unknown outcome, that is, to classify subjects of unknown disease status. Here the cases and controls are split into a training (80%) and testing group (20%) and if the classifier is able to make correct predictions, it is possible to uncover which of the features (mutations) are driving this. Using a holdout of the dataset, in the form of a testing group, prevents the overfitting of an algorithm to the training data and allows for the accuracy of the method to be determined.

A popular technique to achieve this classification is to use a linear support vector machine (SVM). This algorithm maps the data onto a feature space which optimises the separation of the two known outcomes. This mapping is known as a *transformation* and uses known mathematical functions called kernels to succeed in this. The *support vectors* are those which are most important for classifying the data and will be the closest points to the separation line, also known as a *hyperplane* or *decision surface* (Figure 57). Once the hyperplane has been created, new samples can be introduced to discern which group they map to.

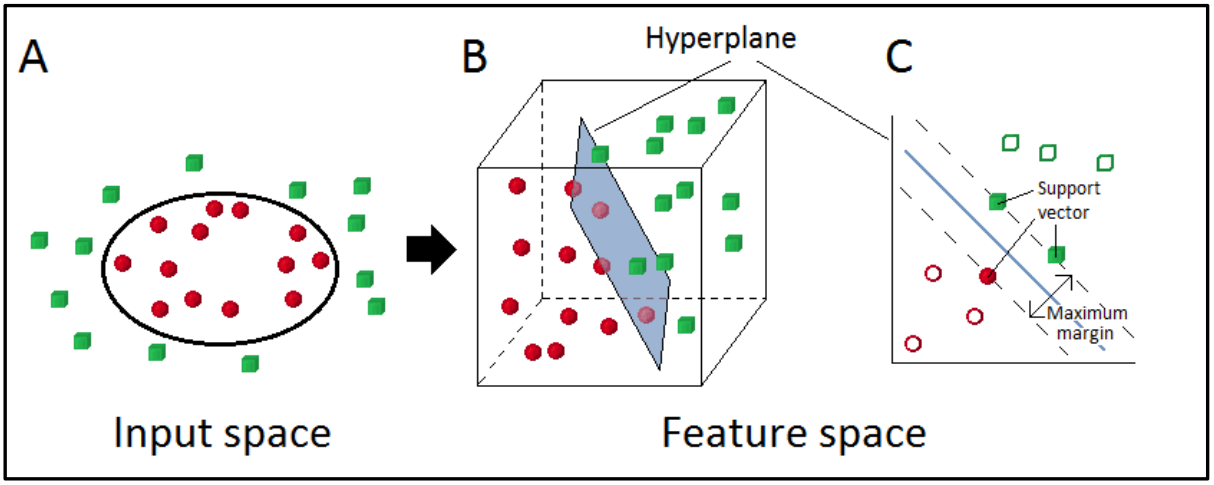

Figure 57. Schematics describing the process of creating a hyperplane using the SVM method. (A) Example of data that is complex in low dimensions. (B) Mapping this data to the feature space which reveals how to separate the red and green groups using a hyperplane. (C) A slice through the hyperplane where the solid points closest to this decision line (within the maximum margin) are used as the support vectors. Drawn in Paint v1511.

I also implemented the Boosted Trees algorithm which employs decision trees in order to make outcome predictions (Figure 58). This technique builds up many weak prediction models into a single stronger model. This method can be more suitable for sparse data like that seen in complex-disease genetics. It is also competent at learning to rank features and so a variation of Boosted Trees is employed by the likes of Yahoo for their search engine.

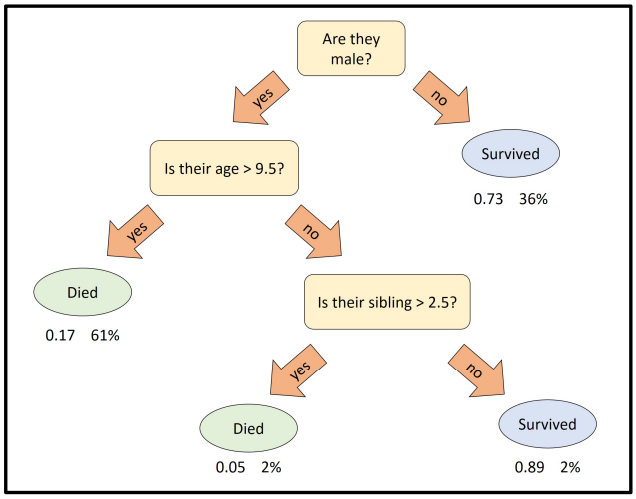

Figure 58. Decision tree documenting the chances of surviving the final titanic voyage. Orange = branch; blue and green = leaves; yellow = decisions, with the first decision being described as the root; leaf numbers indicate the probability of surviving (left) and the percentages of people in each leaf (right).

A confusion matrix is one way of displaying the ability of a classifier. It includes the numbers of true positives (TP), true negatives (TN), false positives and false negatives. From this we can infer an accuracy of the method using the formula:

$$Acc = \frac{(TP + TN)}{Total\ number\ of\ subjects}$$

Other simple measurements include specificity (rate of correctly classifying controls) and precision (how often does it correctly assign a patient with ALS).

One of the first problems I encountered with this method was due to the unequal numbers of patients (1031) and controls (610). Therefore I was generously given 599 extra "controls" to solve this drawback. This data is from a cohort of patients with other diseases not related to ALS. The age range of this cohort included subjects lower than the normal onset age for ALS, however, given that this disease is quite rare it is unlikely that many, if any, subjects in this cohort will later develop ALS. The samples were called using HaplotypeCaller and filtered using Variant Quality Score Recalibration (VQSR) with GATK.

We also decided to include *C9orf72* in the final test since it has been published alongside other mutations as highlighted in Chapter 6 and is part of the oligogenic basis of ALS.

Firstly, I completed a standard PCA on my data (Figure 59). There were no obvious clusters to be identified by this test.

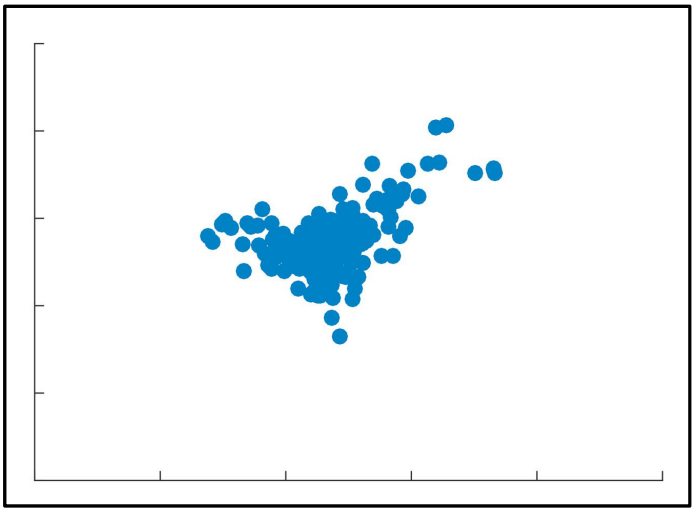

Figure 59. PCA results for the original dataset.

Executing the t-SNE algorithm on my data, including common variation but not *C9orf72*, revealed potentially four groups of interest (Figure 60A). One minor group was a clear outlier which gave grounds for the removal of these subjects from the study. For the patterns in the remaining samples, we speculated if sex or age were contributing factors to these divides however, neither influenced the structure of the data (Figure 60B and C). When homozygous variants are examined exclusively, it can be observed that these four groups partition clearly as an expression of the common variation within several haplotypes present in the subjects. Controls and patients are distributed evenly between these groups and so all of these subjects are likely to be of similar ethnicity.

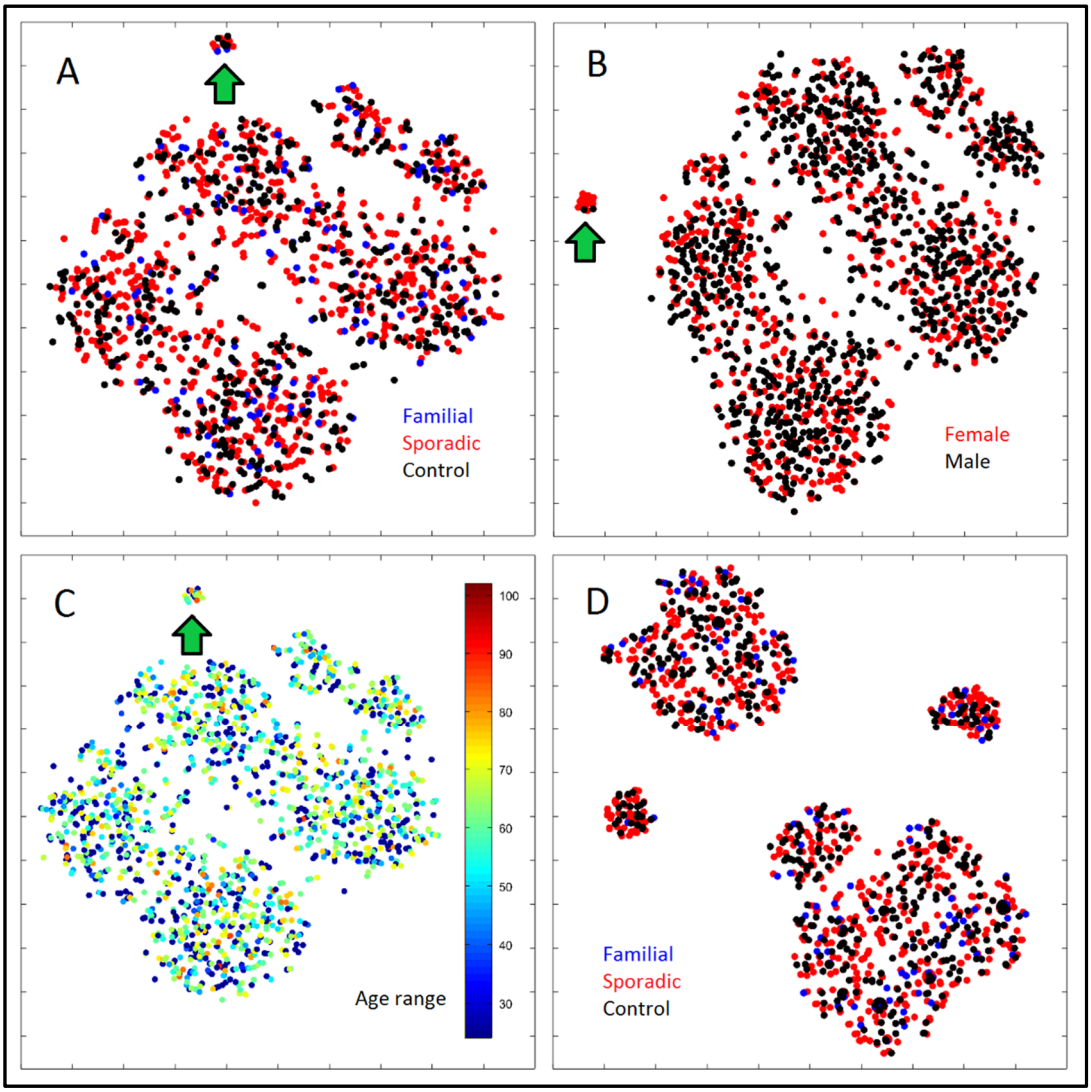

Figure 60. t-SNE results with roughly four clusters in each image. (A) All data has been included in the analysis, green arrow indicates outlier. (B) Same data with gender labelled. (C) Same data with age labelled. (D) Only homozygous variants included in the analysis.

On the original dataset, both the linear SVM and Boosted Trees methods perform badly by predicting most subjects as a patient (Figure 61).

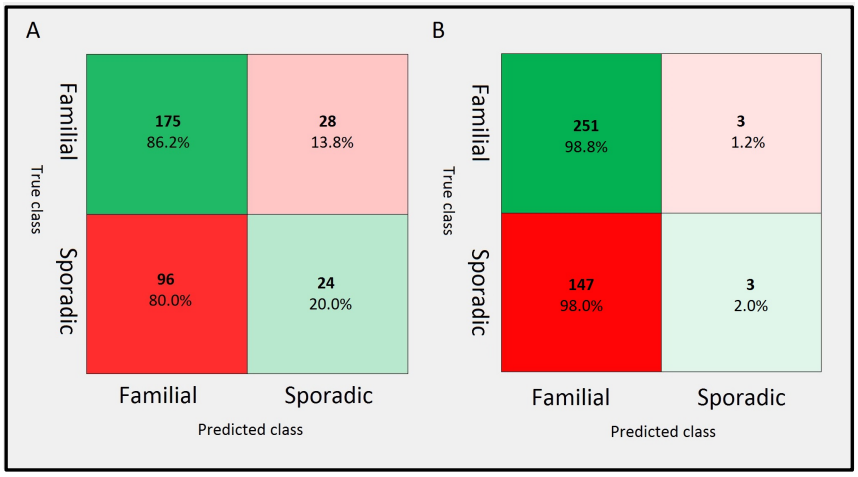

Figure 61. Confusion matrix of the classifiers (A) linear SVM and (B) Boosted Trees which achieved an accuracy of 62 and 63% respectively on the original dataset collected for this study. 459 mutations were included in these tests.

To combat this issue, I added 599 extra individuals to obtain more equal numbers of cases and controls. From the 459 variants of high-quality, 198 variants were filtered out due to inadequate coverage in the new control cohort along with 26 subjects with more than 30% missingness. I performed t-SNE on this new dataset (Figure 62).

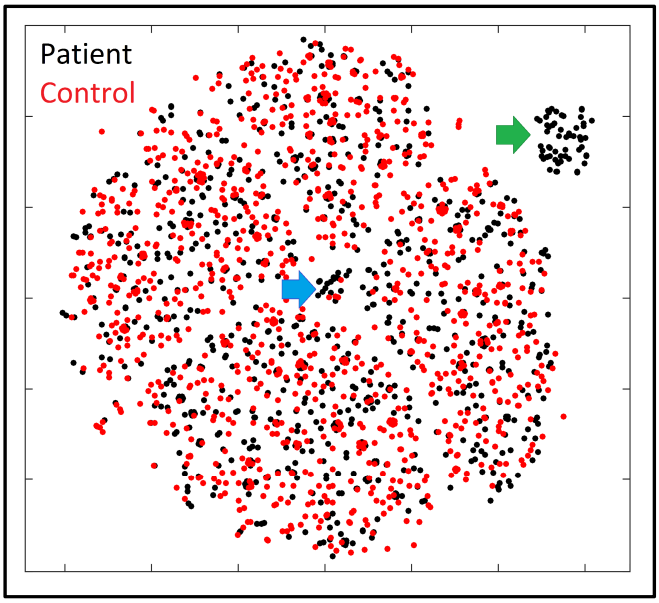

Figure 62. t-SNE results, including new controls. Green arrow indicates C9orf72 positive patients while the blue arrow identifies a small group of subjects who do not harbour the common FUS -54A>G variant.

Running these tests on the larger cohort increased the ability of the classifiers (Figure 63).

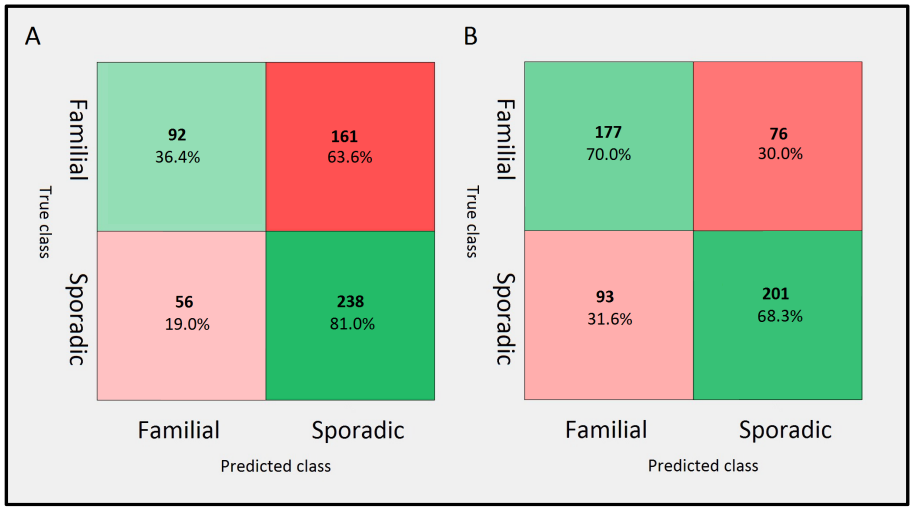

Figure 63. Confusion matrix on the whole dataset including new controls (A) linear SVM 60% accuracy and (B) Boosted Tress with 69% accuracy.

Exporting the variant classifiers from the learnt model reveals six main and five minor variants which push a subject in favour of being classed as a patient, with only one strong variant which helps classify controls (Figure 64).

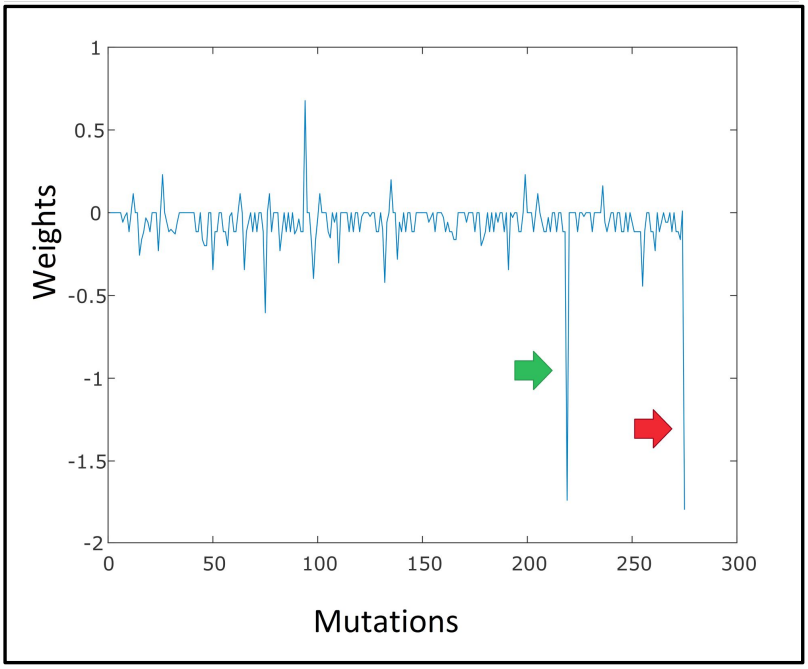

Figure 64. The classifiers used to determine disease status in the final cohort with their respective weights. Positive numbers push the classification towards a control and negative numbers towards a patient. Red arrow = *C9orf72* expansion; Green arrow = *FUS* c.-54A>G.

The biggest predictor of ALS is *C9orf72* and the second biggest is -54A>G in FUS, however, although the A allele of this latter variant is more common in cases, many of these cases are from the Argentinian cohort. This variant remains positive even without these patients. The MAF of this variant is >0.9 indicating that the rarer allele (A) has been misrepresented as the common allele in the reference genome. The G "mutation" may therefore be considered protective.

Examining familial cases against sporadic reveals 89% accuracy, however, again this is due to the uneven sized groups (Figure 65). I will require many more familial cases in order to complete these techniques appropriately.

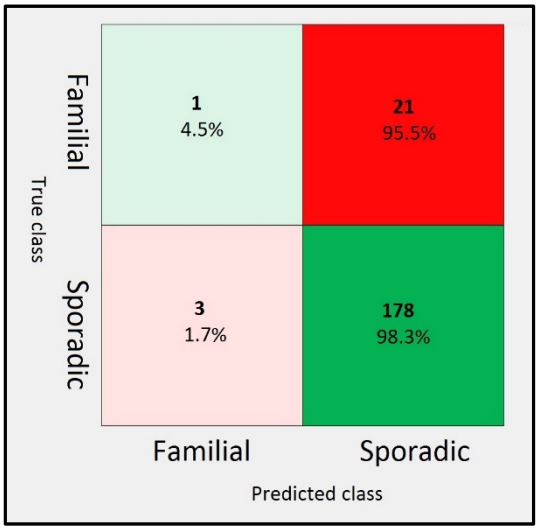

Figure 65. Confusion matrix of sporadic versus familial patients for the linear SVM method.

8.4. DISCUSSION

To explore the possibility of explaining more cases of ALS, I implemented several machine learning techniques to probe for genetic patterns associated with the disease. Completing the t-SNE algorithm on my dataset revealed different sub groups within the data which were mostly driven by common variation and the subtly different haplotypes present in my cohort. There was one outlier group which had significantly different common variation and so were likely to be of an alternate ethnicity than stated on their clinical results. These subjects were removed from the analysis. Otherwise, patients and controls were mixed relatively evenly in

each cluster presenting that the control cohort is from the correct population for comparison against the patients. PCA did not identify these groups and was far less informative on the data.

One of the issues with classification learning is its inability to adequately handle the comparison of two datasets of different sizes. This was especially highlighted when I executed SVM and Boosted Trees on double the number of patients as controls; the algorithms could score highly by predicting everyone as a patient. Additionally this was the case for examining sporadic patients separate to familial. To address this issue, we collected 599 extra controls to match the numbers of patients used in the study. While these controls were not perfect, they significantly reduced the bias due to sample size. They were also obtained using a different technique, therefore the covered regions was fewer than that just tested. The t-SNE results pull out two interesting observations. First, as a proof-of-principle, subjects with *C9orf72* expansions were easily separated from the rest of the group. Secondly, a minor group containing more patients than controls was due to not harbouring a UTR mutation in *FUS* c.-54A>G. This variant has a frequency of 94-98% in 1000 genomes and ExAC but it is interesting that fewer patients have this common mutation. With such a high MAF, it is clear that the reference genome is misleading in this case and the rare allele (A) might be associated with ALS. As discussed in Chapter 6.2, the UTRs of *FUS* play a role in its regulation. A mutation in this region could potentially cause RNA instability and affect the downregulation of the FUS protein directly. Further work is required to determine the effect of the reference allele of this SNP.

When the classification techniques are employed on this new dataset, they display an ability to predict more patients correctly, however, the accuracy is not as high as we had hoped for. Yet the ability to assign patients correctly still remains even when *C9orf72* is removed. There is clearly a number of features present in the data that help us classify cases of ALS but this is nowhere near the predictive ability required in order to be implemented as a useful tool. However, it is interesting that common variation is included in the classification of this cohort suggesting we shouldn't disregard common variation entirely. Considering that only selected genes were included in this analysis, it is entirely possible that using WES or WGS data might increase this accuracy of these classifiers.

CHAPTER 9 ALS VARIANT DATABASE

One of the difficulties with the current genetic literature on ALS is that there is no standard procedure for reporting variants associated with disease. A number of authors will report the amino acid alteration and codon number whereas others only the nucleotide change, dbSNP ID or genomic location. On occasion, merely the genetic sequence shown by Sanger sequencing is published without identifying the position. For some genes, the recognised transcript has altered over the course of a gene's publications. Collectively, these cause the task of searching for one's own results in the published literature to be all the more challenging. Currently, there are a few public databases documenting the genetics of ALS however they are all lacking either in numbers or in information. Most do not even link the reported mutation to the original publication. They are often difficult to search and none report which variants were found in controls or how many controls were sequenced in the study. Lastly, it is also of interest if a variant has previously been discovered within a close proximity of the novel one uncovered in one's own project. For this reason, it was necessary to create my own database of genetic reports in ALS-associated genes. This currently contains over 3000 variants from 500 papers which have focused on any of the genes I have mentioned in this report plus a few others of interest. These numbers are constantly increasing with every new genetic report in ALS. An example from this database is presented in Table 20.

This database documents:

- Gene of interest
- Associated disease
- Paper's guess on pathogenicity
- My prediction on pathogenicity
- Nucleotide alteration and location
- Amino acid alteration and position
- Any previous names of the variant
- The dbSNP reference
- Zygosity
- The exon/intron number
- Domain affected
- Disease phenotype
- The predicted result of the mutation
- The makeup and n-number of the cohort
- The number of controls used
- Nationality of the patients
- Onset of disease
- Whether functional work was carried out
- The original paper reference
- Any extra notable comment

When all this information is available together, we are able to observe the genomic areas which are highly mutated in ALS. This database has aided me hugely in variant interpretation and could be a great resource for other scientists. The largest free ALS database currently available is ALSoD and while this website is extremely useful, it focuses more on the patients harbouring the alterations rather than variant interpretation. Also, there are only a total of 658 mutations listed. With my database, I have tried to use the literature to make judgements on the pathogenicity of each mutation according to Table 6. In order to help others attempting to achieve the same goal, I will be transforming this data into an easily searchable website with a ranking system for pathogenicity.

Disease	Nucleotide change	Disease causing?	Amino Acid change	dbSNP	Dominant or Recessive	Zygosity	Result of mutation	Exon	Domain	Disease cohort	Nationality	Controls	Reference
Control	605G>T	no	A99S					5		N/A	African	2 in 371	Ayala-Lugo 2007
Glaucoma	1274G>A	no	E322K	rs523747				10	Coil-Coiled	1 in 314	Caucasian	371	Ayala-Lugo 2007
ALS	123G>A	unlikely	L41L	rs11591687			Silent	4		42 fALS & 47 sALS	British	none	Bury 2016
ALS	293T>A	unlikely	M98K	rs11258194				5	Coil-coiled	42 fALS & 47 sALS	British	4.15% of 375	Bury 2016
ALS	102G>A	unlikely	T34T	rs2234968			Silent	4		42 fALS & 47 sALS	British	none	Bury 2016
POAG	603T>A	no	M98K	rs11258194				5		5.6% of 785	Australian	17 in 218	Craig 2006
fALS	67G>T	likely	G23X		D	het	Nonsense	4		161 fALS & 113 sALS	Italian	7080	del bo 2013
PLS	844A>C	likely	T282P			het	Missense	9	Coil-Coiled	161 fALS & 113 sALS	Italian	7080	del bo 2013
ALS	1743A>G	yes	E478G	rs267606929		het		14		Case study	Japanese	none	Ito 2011
ALS	402A>C	no	A134A	rs113955718			Silent	6		75 FALS & 420 SALS	American	none	Johnson 2012
ALS	858G>A	no	P286P	rs151065414			Silent	9	Coil-Coiled	75 FALS & 420 SALS	American	none	Johnson 2012
fALS	1743A>G	yes	E478G	rs267606929				14		Six inbred families	Japanese	781	Maruyama 2010
sALS	1502C>T	yes	Q398X	rs267606928	R	hom	Stopgain	12		Six inbred families	Japanese	781	Maruyama 2010
ALS	287G>T	unlikely	R96L			het		5		126 FALS	French	509	Millecamps 2011
ALS & control	964AG	no	E322K	rs523747				10	Coil-Coiled	218 SALS & 18 FALS	Japanese	found in 271	Naruse 2012
ALS & control	293T>A	no	M98K	rs11258194				5		218 SALS & 18 FALS	Japanese	found in 271	Naruse 2012
FTLD	703C>T	yes	Q235X		R	hom	Stopgain	8	Coil-Coiled	107 patients FTLD	American	155	Pottier 2015
POAG	603T>A	sig	M98K	rs11258194				5		6.5% of patients	Russian	1% of controls	Rakhmanov 2005
POAG	433g>A	unsure	L41L	rs11591687			Silent	4		2.9% of patients	Russian	1% of controls	Rakhmanov 2005
POAG	458G.A	yes	E50K					4		7/52 POAG	-	270	Rezaie 2002
ALS	218C>T	no	S73L			het		5		96 ALS families	Australian	480	Solski 2012
ALS	799A>G	no	E163E	rs113811959			Silent	6		563 sALS & 124 FALS	Caucasian	none	Sugihara 2011
ALS	1274G>A	no	E322K	rs523747				10	Coil-Coiled	563 sALS & 124 FALS	Caucasian	none	Sugihara 2011
ALS	433G>A	no	L41L	rs11591687			Silent	4		563 sALS & 124 FALS	Caucasian	none	Sugihara 2011
fALS	493C>T	yes	Q165X		D	het	Stopgain	6	Rab8 binding	64 SALS & a family	Danish	2070	Tumer 2011
fALS	493C>T	yes	Q165X		D	het	Stopgain	6	Rab8 binding	64 SALS & a family	Danish	2070	Tumer 2011

Table 20. An extract from the ALS gene mutation database I created using OPTN as the gene of interest. Some columns were removed for aesthetics. The control column presents a number to represent numbers sequenced that were absent for the mutation in that paper.

For the website, I want to create a ranking system so that variant interpretation can be streamlined. Table 21 shows an example of how this might work for *OPTN*. The score system is designed to reward rare variants which have turned up in multiple studies, only in patients and match the mechanism of disease while penalising novel genes and mutations which only exist in a single person with no functional work attached to it. I picked an arbitrary threshold of 15 points to be considered likely pathogenic and those in the 10-15 range as variants we need to perform functional analysis on.

Gene	Location	Variant	Present in public Databases	Prediction in silico	Present in ALS	Present in other disease	Present in controls	No. affected individuals	Mutation type	Known hotspot/ correct domain	Segregation	Functional work	Gene ranking	De novo	Nonsynon./pred. to affect splicing	Gene mutation rate	Case-control analysis	Score
OPTN	10:13109129C>T	p.H3Y	2	-1	1	1	-1	1	0	1	0	0	2	0	1	0	0	7
OPTN	10:13109168C>G	p.P16A	1	-1	1	1	-1	1	0	1	0	0	2	0	1	0	0	6
OPTN	10:13109189G>T	p.G23X	2	1	1	1	1	1	1	1	0	0	2	0	1	1	0	13
OPTN	10:13109224G>A	p.T34T	-2	0	1	-3	-5	3	-2	1	0	0	2	0	0	0	0	-5
OPTN	10:13109245G>A	p.L41L	0	0	1	-3	-5	3	-2	1	0	0	2	0	0	0	0	-3
OPTN	10:13109270G>A	p.E50K	2	1	-10	-3	1	0	0	1	0	-1	2	0	1	0	0	-6
OPTN	10:13109354A>G	c.166+66A>G	-2	0	1	1	-1	1	0	0	0	0	2	0	0	0	0	2
OPTN	10:13110284G>C	p.K59N	2	1	1	-3	-5	1	0	1	0	0	2	0	1	0	0	1
OPTN	10:13110325C>T	p.S73L	1	1	1	1	1	1	0	1	-5	0	2	0	1	0	0	5
OPTN	10:13110369A>G	p.I88V	2	-1	-10	1	-5	0	0	1	0	0	2	0	1	0	0	-9
OPTN	10:13110384G>C	p.A93P	2	1	1	1	1	1	0	1	0	0	2	0	1	0	0	11
OPTN	10:13110394G>T	p.R96L	1	-1	1	1	1	1	0	1	0	1	2	0	1	0	0	9
OPTN	10:13110400T>A	p.M98K	-2	-1	1	-3	-5	3	0	1	0	-1	2	0	1	0	0	-1
OPTN	10:13110402G>T	p.A99S	1	-1	-10	1	-5	0	0	1	0	0	2	0	1	0	0	-10
OPTN	10:13112465->AG	c.382_383insAG	2	0	1	-3	1	1	1	0	0	1	2	0	1	1	0	8
OPTN	10:13112530G>A	p.R149R	0	0	1	1	-1	1	-2	1	0	0	2	0	0	0	0	3
OPTN	10:13112559G>T	p.G159V	1	1	1	1	-5	1	0	1	0	0	2	0	1	0	0	4
OPTN	10:13112564G>A	p.V161M	1	0	1	-3	-5	1	0	1	0	0	2	0	1	0	0	-1
OPTN	10:13112572A>G	p.E163E	0	0	1	-3	-5	2	-2	1	0	0	2	0	0	0	0	-4
OPTN	10:13112576C>T	p.Q165X	2	1	1	1	1	3	1	1	2	0	2	0	1	1	0	17
OPTN	10:13112599C>T	p.G172G	2	0	1	1	-1	1	-2	1	0	0	2	0	0	0	0	5
OPTN	10:13112635G>-	c.552+1delG	2	0	1	1	1	1	1	0	0	0	2	0	1	0	0	10
OPTN	10:13116262C>T	c.553-5C>T	-2	0	1	1	-1	1	1	0	0	0	2	0	1	0	0	4
OPTN	10:13116364G>A	c.626+24G>A	-2	0	1	1	-1	1	0	0	0	0	2	0	0	0	0	2
OPTN	10:13118964C>T	p.Q235X	2	1	1	1	1	2	1	0	0	0	2	0	1	1	0	13
OPTN	10:13122332T>C	c.780-53T>C	-2	0	1	1	-1	1	0	0	0	0	2	0	0	0	0	2
OPTN	10:13122416C>T	p.R271C	1	-1	1	1	1	1	0	0	0	0	2	0	1	0	0	7
OPTN	10:13122449A>C	p.T282P	1	-1	1	1	1	1	0	0	0	0	2	0	1	0	0	7
OPTN	10:13124021C>A	p.N303K	0	-1	-10	-3	1	0	0	0	0	0	2	0	1	0	0	-10
OPTN	10:13124053A>T	p.Q314L	0	1	1	1	1	2	0	0	0	0	2	0	1	0	0	9
OPTN	10:13124058G>A	p.A316T	2	0	1	1	-1	1	0	0	0	0	2	0	1	0	0	7
OPTN	10:13124076G>A	p.E322K	2	0	1	-3	-5	4	0	0	0	0	2	0	1	0	0	2
OPTN	10:13125413C>T	c.999-5C>T	2	0	1	1	-1	1	0	0	0	0	2	0	1	0	0	7
OPTN	10:13125989C>G	p.Q398E	1	-1	1	1	1	1	0	0	0	0	2	0	1	0	0	7
OPTN	10:13125989C>T	p.Q398X	2	1	1	1	1	2	1	0	0	2	2	0	1	1	0	15
OPTN	10:13126017T>C	p.I407T	0	0	-10	-3	1	0	0	0	0	0	2	0	1	0	0	-9
OPTN	10:13127821A>-	c.1320delA	2	0	1	1	1	2	1	0	0	0	2	0	1	1	0	12
OPTN	10:13127862C>G	p.Q454E	2	1	1	1	1	1	0	1	0	0	2	0	1	0	0	11
OPTN	10:13127907A>G	c.1401+4A>G	1	0	1	1	1	1	0	0	0	0	2	0	1	0	0	8
OPTN	10:13132098A>G	p.E478G	2	1	1	1	1	4	0	1	0	2	2	0	1	0	0	16
OPTN	10:13132107C>T	p.A481V	0	1	1	1	1	2	0	1	0	0	2	0	1	0	0	10
OPTN	10:13132146T>G	p.L494W	0	1	1	1	1	1	0	1	0	0	2	0	1	0	0	9
OPTN	10:13132269G>A	c.1532+72G>A	-2	0	1	1	-1	1	0	0	0	0	2	0	0	0	0	2
OPTN	10:13136697C>A	c.1613-48C>A	-2	0	1	1	-1	1	0	0	0	0	2	0	0	0	0	2
OPTN	10:13136766G>A	p.R545Q	1	-1	1	-3	-5	2	0	0	0	0	2	0	1	0	0	-2
OPTN	10:13136802A>C	p.K557T	2	1	1	1	-1	1	0	0	0	0	2	0	1	0	0	8
OPTN	10:13136835T>C	p.L568S	2	1	-10	-3	-5	0	0	0	0	0	2	0	1	0	0	-12

Table 21. Example of how the variant interpretation table might be presented for a website. The blue column represents the aggregated score with darker colours indicating a higher score.

The scoring system may need to be adjusted as I based the numbers on my own observations and judgement on which information was most important for variant interpretation (Table 22).

Measure	Outcome (score)				
Present in public databases	Absent (2)	Rare <1% (1)	Between 1 and 5% (0)	Common >5% (-2)	
Prediction in silico	Damaging (1)	Predictions disagree (0)	Benign (-1)		
Present in ALS	Present (1)	Not present (-10)			
Present in other unrelated disease	Not present (1)	Present (-3)			
Present in controls	Not present (1)	None sequenced (-1)	Present (-5)		
No. affected individuals	More than three (3)	Two (2)	One (1)	None (0)	
Mutation type	Matches known cause (1)	Could be correct (0)	Wrong type (-2)		
Known hotspot/ correct domain	Yes (1)	No (0)			
Segregation	Yes (2)	Never found (0)	Didn't segregate (-5)		
Functional work	Confirms pathogenic effect (4)	Probably confirms (2)	Loosely confirms (1)	None completed (0)	No pathogenic effect (-5)
Gene ranking	Known common gene (5)	Known rare gene (2)	Known dubious gene (0)	Unknown gene (-3)	
De novo	Found and tested (3)	N/A (0)			
Nonsynonymous/ pred. to affect splicing	Yes (1)	No (0)			
Gene mutation rate	ExAC shows significant loss of this type of mutation (3)	ExAC shows some loss of this type of mutation (1)	ExAC shows no loss of this type of mutation (0)		
Case-control analysis	Significant in meta-analysis (3)	Significant in single study (1)	N/A (0)		

Table 22. Ranking system for variant interpretation.

CHAPTER 10 CONCLUSIONS AND FUTURE DIRECTIONS

10.1. OVERVIEW

In this study, we demonstrate the feasibility of NGS as a research and potential diagnostic tool for patients with ALS. Examining the data presented in this thesis to infer each gene's contribution to ALS, it is clear that *C9orf72*, *SOD1*, *TARDBP* and *FUS* are all undisputed in their involvement. The additional genes that are the supported by my data include *ATXN2*, *OPTN*, *VCP* and *UBQLN2*. Next the genes that are questionably supported are *ALS2*, *DAO*, *FIG4* and *MATR3* (Figure 66). Lastly the genes which have no evidence for their pathogenicity in this study are comprised of *ANG*, *CHCHD10*, *CHMP2B*, *DCTN1*, *NEFH*, *PFN1*, *PON1-3*, *PRPH*, *SETX*, *SPG11*, *SQSTM1*, *TREM2*, *VAPB* and *VEGFA*. However within this latter group, I believe that the literature provides adequate evidence for the association of *CHCHD10*, *CHMP2B*, *SPG11*, *SQSTM1*, *VAPB* and a very minor risk with *PFN1* and *TREM2*.

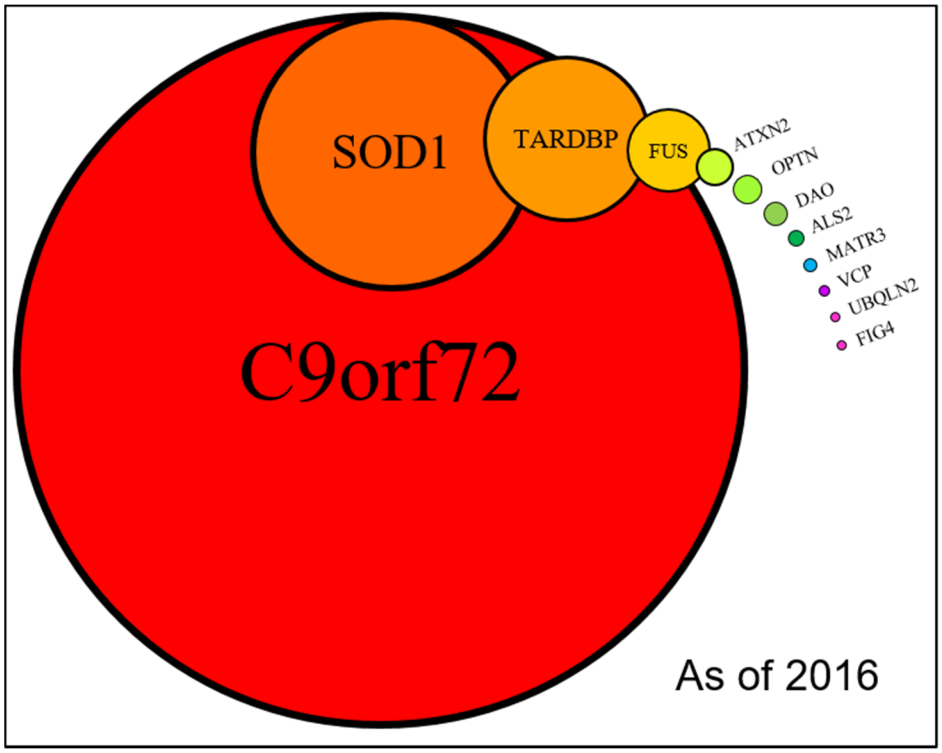

Figure 66. Updated version of Figure 4 based on the results uncovered in this thesis. List of genes implicated in ALS with circle size representing relative contribution to disease.

Examining the test plate, core study and Argentinian cohort together, variants which are likely pathogenic results in a total of 50 (37%) fALS cases able to be explained genetically as well as 104 (11%) sALS subjects. A number of other variants within this cohort may also be

pathogenic and functional work is the next step to determining this and potentially increasing the numbers of explained patients.

10.2. ISSUES WE NEED TO ADDRESS

The biggest problem in disease genetics is the interpretation of rare variants, especially novel mutations never recorded before. We need to follow the recommended guidelines and not exaggerate claims of causality for publications. It is hoped that my database will address some of these issues by providing a source of background knowledge for each variant and ranking them. I have recorded a number of publications which claim to have found a novel variant when in fact the mutation has been published before in ALS.

One of the other principal problems in disease genetics is acquiring adequate control data. Firstly, sequencing patients with a disorder is more profitable in terms of producing an interesting result. Therefore there is a tendency to reuse the same controls for different studies so as not to waste extra money on them (as performed in the current study) or not to sequence any controls and simply rely upon public databases as a control group. Secondly, many publications do not report if their controls harboured any rare mutations. Some of these cohorts are of high enough numbers that it would be statistically next to impossible for there not to be rare mutations given the known background mutation rates within each gene. The drive for not reporting these variants is that they might make the reported disease-implicated variants appear less convincing or the authors may have believed they were not important enough to mention. However these factors only result in the literature producing an incomplete picture of a disease gene's variation. Additionally, a reliance on public databases could produce false negative results as they may contain pathogenic mutations for they are not perfect as a control group. They contain a mixture of individuals of all ages with minimal health checks and some related individuals. Additionally, they may also contribute to false positive outcomes since they may not be similar enough in ethnicity to a study's patient cohort (or have other ethnicities mixed in) which could make common variation in the patients look significant higher than controls. In our control cohort, 185 unique variants were uncovered which were absent from all public databases and 140 of these were also absent from our patients. It is always essential to remember that rare variation is actually quite common and just further stresses the importance of sequencing control cohorts. Of course, as time advances, we will have access to more data than ever which hopefully should reduce these issues.

Another issue in the literature is the use of the phrase “segregated with disease”. Many genes that have been associated with a disease, have had a major part of their causal evidence lie in the fact that the mutation segregates with the disorder in a number of members in a large family. However, I have noticed a number of publications using this expression when only two individuals in a family were examined or worse, when they do not disclose how many relatives were sequenced in the study. Obviously it makes a better story if this information is in the supplemental notes of a paper but this knowledge is vital for those trying to understand a variant’s (or gene’s) involvement in a disease.

10.3. FUTURE OF GENETICS

Genetics is a numbers game; as we get more samples, patients and controls, we will be able to more accurately analyse the data and gain a fuller understanding on all the pathways involved in diseases like ALS. Year by year we are acquiring larger datasets but this does not come without its challenges. To handle this expansion requires efficient computation and a large infrastructure of supercomputers and knowledgeable personnel to interpret the data. One of the slowest sections of the analysis pipeline is aligning to the reference genome. The most challenging alignments are reads which contain indels with the larger of these causing the most issues. One tactic which addresses this problem is the use of a reference genome composed using graph theory (Figure 66).

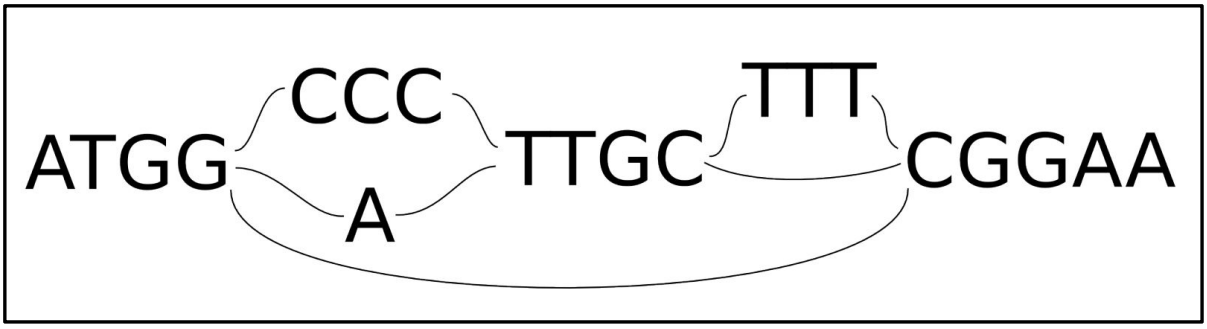

Figure 67. A segment of DNA with different possible mutations as represented by graph theory. An individual being mapped to this reference could have any combination following the lines from start to finish. It is also possible to include connections enabling the sequence to go back on itself and take another path.

In contrast to a linear reference, a whole genome variation graph allows for multiple matches along its length. Computers prefer to solve this type of problem and so using this new

reference speeds up alignment and reduces reference bias making the results more accurate particularly for sizable indels. Richard Durbin, the name behind the 1000 genomes project, is one such individual aiming to create these reference maps. These sequences have the ability to contain information from thousands of individuals without requiring masses of disk storage space. It is also anonymous for all individuals within the reference for no single trajectory across the genome can be identified without intersecting others.

A revolutionary project currently underway is Genomics England. The aim of this project is to sequence 100,000 entire genomes from individuals with a range of disorders. The results from these subjects will hugely aid the discovery of novel genes and pathways related to disease. It will also provide a move towards personalised healthcare where a patient's genetic results will determine the diagnosis and treatments prescribed.

Not every causal variant (or gene) will produce a perfect Mendelian inheritance of ALS which can be uncovered by analysing enough members of the same family. Therefore we need more work on the functional consequences of mutations. Many of these experiments are expensive and time consuming, potentially taking a couple of years to complete for some cell cultures and so I believe we need to create some high throughput, low cost functional studies to analyse variants of uncertain significance, even if it's only to select those to carry on into more robust experiments.

10.4. FINAL CONCLUSION

To conclude this thesis, the impact of this work can be summarised in three major findings. Firstly, I have updated the genetic landscape of ALS, altered the order of the genes to reflect their relative contribution and established that the genetics are more complex than previously assumed. Secondly, I have highlighted the importance of considering non-coding variation when examining ALS genetics, given the increase in rare variant burden within these regions. Lastly, I have explored a novel machine learning approach in ALS which uncovered a potentially protective mutation within FUS. A method that may be beneficial for many complex genetic disorders. All three of these outcomes provide new and exciting research questions to be answered as the field develops in the aim of one day finding a cure for ALS.

REFERENCES

Abduljaleel, Z., Al-Allaf, F.A., Khan, W., Athar, M., Shahzad, N., Taher, M.M., Elrobh, M., Alanazi, M.S. and El-Huneidi, W. 2014. Evidence of trem2 variant associated with triple risk of Alzheimer's disease. *PloS One*. **9**(3),p.e92648.

Abrahams, S., Newton, J., Niven, E., Foley, J. and Bak, T.H. 2014. Screening for cognition and behaviour changes in ALS. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*. **15**(1–2),pp.9–14.

Abramzon, Y., Johnson, J.O., Scholz, S.W., Taylor, J.P., Brunetti, M., Calvo, A., Mandrioli, J., Benatar, M., Mora, G., Restagno, G., Chiò, A. and Traynor, B.J. 2012. Valosin-containing protein (VCP) mutations in sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(9),p.2231.e1-2231.e6.

Ackerley, S., Grierson, A.J., Brownlees, J., Thornhill, P., Anderton, B.H., Leigh, P.N., Shaw, C.E. and Miller, C.C. 2000. Glutamate slows axonal transport of neurofilaments in transfected neurons. *The Journal of Cell Biology*. **150**(1),pp.165–176.

Ajroud-Driss, S., Fecto, F., Ajroud, K., Lalani, I., Calvo, S.E., Mootha, V.K., Deng, H.-X., Siddique, N., Tahmoush, A.J., Heiman-Patterson, T.D. and Siddique, T. 2015. Mutation in the novel nuclear-encoded mitochondrial protein CHCHD10 in a family with autosomal dominant mitochondrial myopathy. *Neurogenetics*. **16**(1),pp.1–9.

Al-Chalabi, A., Andersen, P.M., Chioza, B., Shaw, C., Sham, P.C., Robberecht, W., Matthijs, G., Camu, W., Marklund, S.L., Forsgren, L., Rouleau, G., Laing, N.G., Hurse, P.V., Siddique, T., Leigh, P.N. and Powell, J.F. 1998. Recessive Amyotrophic Lateral Sclerosis Families with the D90A SOD1 Mutation Share a Common Founder: Evidence for a Linked Protective Factor. *Human Molecular Genetics*. **7**(13),pp.2045–2050.

Al-Chalabi, A., Andersen, P.M., Nilsson, P., Chioza, B., Andersson, J.L., Russ, C., Shaw, C.E., Powell, J.F. and Leigh, P.N. 1999. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. *Human Molecular Genetics*. **8**(2),pp.157–164.

Al-Chalabi, A., Hansen, V.K., Simpson, C.L., Xi, J., Hosler, B.A., Powell, J.F., McKenna-Yasek, D., Shaw, C.E., Leigh, P.N. and Brown, R.H. 2003. Variants in the ALS2 gene are not associated with sporadic amyotrophic lateral sclerosis. *Neurogenetics*. **4**(4),pp.221–222.

Al-Chalabi, A. and Lewis, C.M. 2011. Modelling the effects of penetrance and family size on rates of sporadic and familial disease. *Human Heredity*. **71**(4),pp.281–288.

Alterovitz, G. 2014. Medical Informatics: Introduction. [Accessed 1 February 2016]. Available from: https://math.mit.edu/research/highschool/primes/materials/2014/conf/10-Alterovitz.pdf.

Andersen, P.M. and Al-Chalabi, A. 2011. Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nature Reviews. Neurology*. **7**(11),pp.603–615.

Andersen, P.M., Nilsson, P., Ala-Hurula, V., Keränen, M.L., Tarvainen, I., Haltia, T., Nilsson, L., Binzer, M., Forsgren, L. and Marklund, S.L. 1995. Amyotrophic lateral sclerosis

associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. *Nature Genetics*. **10**(1),pp.61–66.

Andersen, P.M., Sims, K.B., Xin, W.W., Kiely, R., O'Neill, G., Ravits, J., Pioro, E., Harati, Y., Brower, R.D., Levine, J.S., Heinicke, H.U., Seltzer, W., Boss, M. and Robert H Brown, J. 2003. Sixteen novel mutations in the Cu/Zn superoxide dismutase gene in amyotrophic lateral sclerosis: a decade of discoveries, defects and disputes. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*. **4**(2),pp.62–73.

Araki, E., Tsuboi, Y., Daechsel, J., Milnerwood, A., Vilarino-Guell, C., Fujii, N., Mishima, T., Oka, T., Hara, H., Fukae, J. and Farrer, M.J. 2014. A novel DCTN1 mutation with late-onset parkinsonism and frontotemporal atrophy. *Movement Disorders: Official Journal of the Movement Disorder Society*. **29**(9),pp.1201–1204.

Arning, L., Epplen, J.T., Rahikkala, E., Hendrich, C., Ludolph, A.C. and Sperfeld, A.-D. 2012. The SETX missense variation spectrum as evaluated in patients with ALS4-like motor neuron diseases. *neurogenetics*. **14**(1),pp.53–61.

Avemaria, F., Lunetta, C., Tarlarini, C., Mosca, L., Maestri, E., Marocchi, A., Melazzini, M., Penco, S. and Corbo, M. 2011. Mutation in the senataxin gene found in a patient affected by familial ALS with juvenile onset and slow progression. *Amyotrophic Lateral Sclerosis*. **12**(3),pp.228–230.

Avena, S., Via, M., Ziv, E., Pérez-Stable, E.J., Gignoux, C.R., Dejean, C., Huntsman, S., Torres-Mejía, G., Dutil, J., Matta, J.L., Beckman, K., Burchard, E.G., Parolin, M.L., Goicoechea, A., Acreche, N., Boquet, M., Ríos Part, M.D.C., Fernández, V., Rey, J., Stern, M.C., Carnese, R.F. and Fejerman, L. 2012. Heterogeneity in genetic admixture across different regions of Argentina. *PloS One*. **7**(4),p.e34695.

Ayala, Y.M., De Conti, L., Avendaño-Vázquez, S.E., Dhir, A., Romano, M., D'Ambrogio, A., Tollervey, J., Ule, J., Baralle, M., Buratti, E. and Baralle, F.E. 2011. TDP-43 regulates its mRNA levels through a negative feedback loop. *The EMBO journal*. **30**(2),pp.277–288.

Ayers, J., Lelie, H., Workman, A., Prudencio, M., Brown, H., Fromholt, S., Valentine, J., Whitelegge, J. and Borchelt, D. 2014. Distinctive features of the D101N and D101G variants of superoxide dismutase 1; two mutations that produce rapidly progressing motor neuron disease. *Journal of Neurochemistry*. **128**(2),pp.305–314.

Azzouz, M., Ralph, G.S., Storkebaum, E., Walmsley, L.E., Mitrophanous, K.A., Kingsman, S.M., Carmeliet, P. and Mazarakis, N.D. 2004. VEGF delivery with retrogradely transported lentivector prolongs survival in a mouse ALS model. *Nature*. **429**(6990),pp.413–417.

Banfi, S., Servadio, A., Chung, M., Kwiatkowski, T.J., McCall, A.E., Duvick, L.A., Shen, Y., Roth, E.J., Orr, H.T. and Zoghbi, H.Y. 1994. Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nature Genetics*. **7**(4),pp.513–520.

Bannwarth, S., Ait-El-Mkadem, S., Chaussenot, A., Genin, E.C., Lacas-Gervais, S., Fragaki, K., Berg-Alonso, L., Kageyama, Y., Serre, V., Moore, D.G., Verschueren, A., Rouzier, C., Le Ber, I., Augé, G., Cochaud, C., Lespinasse, F., N'Guyen, K., de Septenville, A., Brice, A., Yu-Wai-Man, P., Sesaki, H., Pouget, J. and Paquis-Flucklinger, V. 2014. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain: A Journal of Neurology*. **137**(Pt 8),pp.2329–2345.

Bäumer, D., Hilton, D., Paine, S.M.L., Turner, M.R., Lowe, J., Talbot, K. and Ansorge, O. 2010. Juvenile ALS with basophilic inclusions is a FUS proteinopathy with FUS mutations. *Neurology*. **75**(7),pp.611–618.

Beaulieu, J.M., Nguyen, M.D. and Julien, J.P. 1999. Late onset of motor neurons in mice overexpressing wild-type peripherin. *The Journal of Cell Biology*. **147**(3),pp.531–544.

Beck, J., Pittman, A., Adamson, G., Campbell, T., Kenny, J., Houlden, H., Rohrer, J.D., de Silva, R., Shoai, M., Uphill, J., Poulter, M., Hardy, J., Mummery, C.J., Warren, J.D., Schott, J.M., Fox, N.C., Rossor, M.N., Collinge, J. and Mead, S. 2014. Validation of next-generation sequencing technologies in genetic diagnosis of dementia. *Neurobiology of Aging*. **35**(1),pp.261–265.

Belecky-Adams, T., Wight, D.C., Kopchick, J.J. and Parysek, L.M. 1993. Intragenic sequences are required for cell type-specific and injury-induced expression of the rat peripherin gene. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. **13**(12),pp.5056–5065.

Beleza-Meireles, A. and Al-Chalabi, A. 2009. Genetic studies of amyotrophic lateral sclerosis: controversies and perspectives. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **10**(1),pp.1–14.

Belzil, V.V., Daoud, H., St-Onge, J., Desjarlais, A., Bouchard, J.-P., Dupre, N., Lacomblez, L., Salachas, F., Pradat, P.-F., Meininger, V., Camu, W., Dion, P.A. and Rouleau, G.A. 2011. Identification of novel FUS mutations in sporadic cases of amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **12**(2),pp.113–117.

Belzil, V.V., Valdmanis, P.N., Dion, P.A., Daoud, H., Kabashi, E., Noreau, A., Gauthier, J., Team, for the S., Hince, P., Desjarlais, A., Bouchard, J.-P., Lacomblez, L., Salachas, F., Pradat, P.-F., Camu, W., Meininger, V., Dupré, N. and Rouleau, G.A. 2009. Mutations in FUS cause FALS and SALS in French and French Canadian populations. *Neurology*. **73**(15),pp.1176–1179.

Benitez, B.A., Cooper, B., Pastor, P., Jin, S.-C., Lorenzo, E., Cervantes, S. and Cruchaga, C. 2013. TREM2 is associated with the risk of Alzheimer's disease in Spanish population. *Neurobiology of Aging*. **34**(6),p.1711.e15-17.

Bersano, A., Del Bo, R., Lamperti, C., Ghezzi, S., Fagiolari, G., Fortunato, F., Ballabio, E., Moggio, M., Candelise, L., Galimberti, D., Virgilio, R., Lanfranconi, S., Torrente, Y., Carpo, M., Bresolin, N., Comi, G.P. and Corti, S. 2009. Inclusion body myopathy and frontotemporal dementia caused by a novel VCP mutation. *Neurobiology of Aging*. **30**(5),pp.752–758.

Blair, I.P., Vance, C., Durnall, J.C., Williams, K.L., Thoeng, A., Shaw, C.E. and Nicholson, G.A. 2008. CHMP2B mutations are not a common cause of familial or sporadic amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*. **79**(7),pp.849–850.

Blair, I.P., Williams, K.L., Warraich, S.T., Durnall, J.C., Thoeng, A.D., Manavis, J., Blumbergs, P.C., Vucic, S., Kiernan, M.C. and Nicholson, G.A. 2010. FUS mutations in amyotrophic lateral sclerosis: clinical, pathological, neurophysiological and genetic analysis. *Journal of Neurology, Neurosurgery, and Psychiatry*. **81**(6),pp.639–645.

van Blitterswijk, M., Baker, M.C., Bieniek, K.F., Knopman, D.S., Josephs, K.A., Boeve, B., Caselli, R., Wszolek, Z.K., Petersen, R., Graff-Radford, N.R., Boylan, K.B., Dickson, D.W. and Rademakers, R. 2013a. Profilin-1 mutations are rare in patients with amyotrophic lateral sclerosis and frontotemporal dementia. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*. **14**(5–6),pp.463–469.

van Blitterswijk, M., DeJesus-Hernandez, M. and Rademakers, R. 2012a. How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? *Current Opinion in Neurology*. **25**(6),pp.689–700.

van Blitterswijk, M., van Es, M.A., Hennekam, E.A.M., Dooijes, D., van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., de Bakker, P.I.W., Veldink, J.H. and van den Berg, L.H. 2012b. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Human Molecular Genetics*. **21**(17),pp.3776–3784.

van Blitterswijk, M., van Es, M.A., Koppers, M., van Rheenen, W., Medic, J., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., Veldink, J.H. and van den Berg, L.H. 2012c. VAPB and C9orf72 mutations in 1 familial amyotrophic lateral sclerosis patient. *Neurobiology of Aging*. **33**(12),p.2950.e1-4.

van Blitterswijk, M., Vlam, L., van Es, M.A., van der Pol, W.-L., Hennekam, E.A.M., Dooijes, D., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., Veldink, J.H. and van den Berg, L.H. 2012d. Genetic overlap between apparently sporadic motor neuron diseases. *PloS One*. **7**(11),p.e48983.

van Blitterswijk, M., Wang, E.T., Friedman, B.A., Keagle, P.J., Lowe, P., Leclerc, A.L., van den Berg, L.H., Housman, D.E., Veldink, J.H. and Landers, J.E. 2013b. Characterization of FUS mutations in amyotrophic lateral sclerosis using RNA-Seq. *PloS One*. **8**(4),p.e60788.

Blokhuis, A.M., Groen, E.J.N., Koppers, M., van den Berg, L.H. and Pasterkamp, R.J. 2013. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathologica*. **125**(6),pp.777–794.

Blumen, S.C., Inzelberg, R., Nisipeanu, P., Carasso, R.L., Oved, D., Aizenstein, O., Drory, V.E., Bergstrom, C. and Andersen, P.M. 2010. Aggressive familial ALS with unusual brain MRI and a SOD1 gene mutation. *Amyotrophic Lateral Sclerosis*. **11**(1–2),pp.228–231.

Borroni, B., Bonvicini, C., Alberici, A., Buratti, E., Agosti, C., Archetti, S., Papetti, A., Stuani, C., Di Luca, M., Gennarelli, M. and Padovani, A. 2009. Mutation within TARDBP leads to frontotemporal dementia without motor neuron disease. *Human Mutation*. **30**(11),pp.E974-983.

Bosco, D.A., Lemay, N., Ko, H.K., Zhou, H., Burke, C., Kwiatkowski, T.J., Sapp, P., McKenna-Yasek, D., Brown, R.H. and Hayward, L.J. 2010. Mutant FUS proteins that cause amyotrophic lateral sclerosis incorporate into stress granules. *Human Molecular Genetics*. **19**(21),pp.4160–4175.

Boukaftane, Y., Khoris, J., Moulard, B., Salachas, F., Meininger, V., Malafosse, A., Camu, W. and Rouleau, G.A. 1998. Identification of six novel SOD1 gene mutations in familial amyotrophic lateral sclerosis. *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*. **25**(3),pp.192–196.

Boutoleau-Bretonnière, C., Camuzat, A., Le Ber, I., Bouya-Ahmed, K., Guerreiro, R., Deruet, A.-L., Evrard, C., Bras, J., Lamy, E., Auffray-Calvier, E., Pallardy, A., Hardy, J., Brice, A., Derkinderen, P. and Vercelletto, M. 2015. A phenotype of atypical apraxia of speech in a family carrying SQSTM1 mutation. *Journal of Alzheimer's disease: JAD*. **43**(2),pp.625–630.

Brady, O.A., Meng, P., Zheng, Y., Mao, Y. and Hu, F. 2011. Regulation of TDP-43 aggregation by phosphorylation and p62/SQSTM1. *Journal of Neurochemistry*. **116**(2),pp.248–259.

Brown, J.A., Min, J., Staropoli, J.F., Collin, E., Bi, S., Feng, X., Barone, R., Cao, Y., O'Malley, L., Xin, W., Mullen, T.E. and Sims, K.B. 2012. SOD1, ANG, TARDBP and FUS mutations in amyotrophic lateral sclerosis: a United States clinical testing lab experience. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **13**(2),pp.217–222.

Brugman, F., Eymard-Pierre, E., van den Berg, L.H., Wokke, J.H.J., Gauthier-Barichard, F. and Boespflug-Tanguy, O. 2007. Adult-onset primary lateral sclerosis is not associated with mutations in the ALS2 gene. *Neurology*. **69**(7),pp.702–704.

Buentello-Volante, B., Elizondo-Olascoaga, C., Miranda-Duarte, A., Guadarrama-Vallejo, D., Cabral-Macias, J. and Zenteno, J.C. 2013. Association study of multiple gene polymorphisms with the risk of adult-onset primary open-angle glaucoma in a Mexican population. *Experimental Eye Research*. **107**,pp.59–64.

Bury, J.J., Highley, J.R., Cooper-Knock, J., Goodall, E.F., Higginbottom, A., McDermott, C.J., Ince, P.G., Shaw, P.J. and Kirby, J. 2016. Oligogenic inheritance of optineurin (OPTN) and C9ORF72 mutations in ALS highlights localisation of OPTN in the TDP-43-negative inclusions of C9ORF72-ALS. *Neuropathology: Official Journal of the Japanese Society of Neuropathology*. **36**(2),pp.125–134.

Cady, J., Allred, P., Bali, T., Pestronk, A., Goate, A., Miller, T.M., Mitra, R.D., Ravits, J., Harms, M.B. and Baloh, R.H. 2015. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. *Annals of Neurology*. **77**(1),pp.100–113.

Cady, J., Koval, E.D., Benitez, B.A., Zaidman, C., Jockel-Balsarotti, J., Allred, P., Baloh, R.H., Ravits, J., Simpson, E., Appel, S.H., Pestronk, A., Goate, A.M., Miller, T.M., Cruchaga, C. and Harms, M.B. 2014. TREM2 variant p.R47H as a risk factor for sporadic amyotrophic lateral sclerosis. *JAMA neurology*. **71**(4),pp.449–453.

Cai, H., Lin, X., Xie, C., Laird, F.M., Lai, C., Wen, H., Chiang, H.-C., Shim, H., Farah, M.H., Hoke, A., Price, D.L. and Wong, P.C. 2005. Loss of ALS2 function is insufficient to trigger motor neuron degeneration in knock-out mice but predisposes neurons to oxidative stress. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*. **25**(33),pp.7567–7574.

Caixeta-Umbelino, C., Vasconcellos, J.P.C. de, Costa, V.P., Kasahara, N., Paolera, M.D., Almeida, G.V. de, Cohen, R., Jr, C.M., Rocha, M.N., Richeti, F., Longui, C.A. and Melo, M.B. de 2009. Lack of Association Between Optineurin Gene Variants T34T, E50K, M98K, 691_692insAG and R545Q and Primary Open Angle Glaucoma in Brazilian Patients. *Ophthalmic Genetics*. **30**(1),pp.13–18.

Campeau, P.M., Lenk, G.M., Lu, J.T., Bae, Y., Burrage, L., Turnpenny, P., Román Corona-Rivera, J., Morandi, L., Mora, M., Reutter, H., Vulto-van Silfhout, A.T., Faivre, L., Haan, E., Gibbs, R.A., Meisler, M.H. and Lee, B.H. 2013. Yunis-Varón syndrome is caused by mutations in FIG4, encoding a phosphoinositide phosphatase. *American Journal of Human Genetics*. **92**(5),pp.781–791.

Cannon, A., Baker, M., Boeve, B., Josephs, K., Knopman, D., Petersen, R., Parisi, J., Dickison, D., Adamson, J., Snowden, J., Neary, D., Mann, D., Hutton, M. and Pickering-Brown, S.M. 2006. CHMP2B mutations are not a common cause of frontotemporal lobar degeneration. *Neuroscience Letters*. **398**(1–2),pp.83–84.

Chance, P.F., Rabin, B.A., Ryan, S.G., Ding, Y., Scavina, M., Crain, B., Griffin, J.W. and Cornblath, D.R. 1998. Linkage of the gene for an autosomal dominant form of juvenile amyotrophic lateral sclerosis to chromosome 9q34. *American Journal of Human Genetics*. **62**(3),pp.633–640.

Chaussenot, A., Le Ber, I., Ait-El-Mkadem, S., Camuzat, A., de Septenville, A., Bannwarth, S., Genin, E.C., Serre, V., Augé, G., French research network on FTD and FTD-ALS, Brice, A., Pouget, J. and Paquis-Flucklinger, V. 2014. Screening of CHCHD10 in a French cohort confirms the involvement of this gene in frontotemporal dementia with amyotrophic lateral sclerosis patients. *Neurobiology of Aging*. **35**(12),p.2884.e1-4.

Chen, H.-J., Anagnostou, G., Chai, A., Withers, J., Morris, A., Adhikaree, J., Pennetta, G. and de Belleroche, J.S. 2010. Characterization of the properties of a novel mutation in VAPB in familial amyotrophic lateral sclerosis. *The Journal of Biological Chemistry*. **285**(51),pp.40266–40281.

Chen, X., Chen, Y., Wei, Q., Guo, X., Cao, B., Ou, R., Zhao, B. and Shang, H.-F. 2015. Assessment of TREM2 rs75932628 association with amyotrophic lateral sclerosis in a Chinese population. *Journal of the Neurological Sciences*. **355**(1–2),pp.193–195.

Chen, Y., Zheng, Z.-Z., Huang, R., Chen, K., Song, W., Zhao, B., Chen, X., Yang, Y., Yuan, L. and Shang, H.-F. 2013. PFN1 mutations are rare in Han Chinese populations with amyotrophic lateral sclerosis. *Neurobiology of Aging*. **34**(7),p.1922.e1-5.

Chen, Y.-Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., Nicholson, G.A., Auer-Grumbach, M., Wagner, K., De Jonghe, P., Griffin, J.W., Fischbeck, K.H., Timmerman, V., Cornblath, D.R. and Chance, P.F. 2004. DNA/RNA Helicase Gene Mutations in a Form of Juvenile Amyotrophic Lateral Sclerosis (ALS4). *The American Journal of Human Genetics*. **74**(6),pp.1128–1135.

Chiò, A., Calvo, A., Moglia, C., Canosa, A., Brunetti, M., Barberis, M., Restagno, G., Conte, A., Bisogni, G., Marangi, G., Moncada, A., Lattante, S., Zollino, M., Sabatelli, M., Bagarotti, A., Corrado, L., Mora, G., Bersano, E., Mazzini, L., D'Alfonso, S. and PARALS 2015. ATXN2 polyQ intermediate repeats are a modifier of ALS survival. *Neurology*. **84**(3),pp.251–258.

Chiò, A., Mora, G., Sabatelli, M., Caponnetto, C., Traynor, B.J., Johnson, J.O., Nalls, M.A., Calvo, A., Moglia, C., Borghero, G., Monsurrò, M.R., La Bella, V., Volanti, P., Simone, I., Salvi, F., Logullo, F.O., Nilo, R., Battistini, S., Mandrioli, J., Tanel, R., Murru, M.R., Mandich, P., Zollino, M., Conforti, F.L., ITALSGEN Consortium, Brunetti, M., Barberis, M., Restagno, G., Penco, S. and Lunetta, C. 2015. CHCH10 mutations in an Italian cohort of familial and sporadic amyotrophic lateral sclerosis patients. *Neurobiology of Aging*. **36**(4),p.1767.e3-6.

Chiò, A., Restagno, G., Brunetti, M., Ossola, I., Calvo, A., Canosa, A., Moglia, C., Floris, G., Tacconi, P., Marrosu, F., Marrosu, M.G., Murru, M.R., Majounie, E., Renton, A.E., Abramzon, Y., Pugliatti, M., Sotgiu, M.A., Traynor, B.J. and Borghero, G. 2012. ALS/FTD phenotype in two Sardinian families carrying both C9ORF72 and TARDBP mutations. *Journal of neurology, neurosurgery, and psychiatry*. **83**(7),pp.730–733.

Chouery, E., Delague, V., Bergougnoux, A., Koussa, S., Serre, J.-L. and Mégarbané, A. 2008. Mutations in TREM2 lead to pure early-onset dementia without bone cysts. *Human Mutation*. **29**(9),pp.E194-204.

Chow, C.Y., Landers, J.E., Bergren, S.K., Sapp, P.C., Grant, A.E., Jones, J.M., Everett, L., Lenk, G.M., McKenna-Yasek, D.M., Weisman, L.S., Figlewicz, D., Brown, R.H. and Meisler, M.H. 2009. Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. *American Journal of Human Genetics*. **84**(1),pp.85–88.

Chow, C.Y., Zhang, Y., Dowling, J.J., Jin, N., Adamska, M., Shiga, K., Szigeti, K., Shy, M.E., Li, J., Zhang, X., Lupski, J.R., Weisman, L.S. and Meisler, M.H. 2007. Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. *Nature*. **448**(7149),pp.68–72.

Chung, P.Y.J., Beyens, G., Guañabens, N., Boonen, S., Papapoulos, S., Karperien, M., Eekhoff, M., Van Wesenbeeck, L., Jennes, K., Geusens, P., Offeciers, E., Van Offel, J., Westhovens, R., Zmierczak, H., Devogelaer, J.-P. and Van Hul, W. 2008. Founder effect in different European countries for the recurrent P392L SQSTM1 mutation in Paget's Disease of Bone. *Calcified Tissue International*. **83**(1),pp.34–42.

Ciani, B., Layfield, R., Cavey, J.R., Sheppard, P.W. and Searle, M.S. 2003. Structure of the ubiquitin-associated domain of p62 (SQSTM1) and implications for mutations that cause Paget's disease of bone. *The Journal of Biological Chemistry*. **278**(39),pp.37409–37412.

Cleveland, D.W. and Rothstein, J.D. 2001. From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. *Nature Reviews. Neuroscience*. **2**(11),pp.806–819.

Coelho, M.B., Attig, J., Bellora, N., König, J., Hallegger, M., Kayikci, M., Eyras, E., Ule, J. and Smith, C.W.J. 2015. Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. *The EMBO journal*. **34**(5),pp.653–668.

Collard, J.-F., Côté, F. and Julien, J.-P. 1995. Defective axonal transport in a transgenic mouse model of amyotrophic lateral sclerosis. *Nature*. **375**(6526),pp.61–64.

Conforti, F.L., Spataro, R., Sproviero, W., Mazzei, R., Cavalcanti, F., Condino, F., Simone, I.L., Logroscino, G., Patitucci, A., Magariello, A., Muglia, M., Rodolico, C., Valentino, P., Bono, F., Colletti, T., Monsurrò, M.R., Gambardella, A. and La Bella, V. 2012. Ataxin-1 and ataxin-2 intermediate-length PolyQ expansions in amyotrophic lateral sclerosis. *Neurology*. **79**(24),pp.2315–2320.

Conte, A., Lattante, S., Luigetti, M., Grande, A.D., Romano, A., Marcaccio, A., Marangi, G., Rossini, P.M., Neri, G., Zollino, M. and Sabatelli, M. 2012. Classification of familial amyotrophic lateral sclerosis by family history: effects on frequency of genes mutation. *Journal of Neurology, Neurosurgery & Psychiatry*.,p.jnnp-2012-302897.

Cookson, M.R. 2012. Cellular effects of LRRK2 mutations. *Biochemical Society Transactions*. **40**(5),pp.1070–1073.

Cooper-Knock, J., Hewitt, C., Highley, J.R., Brockington, A., Milano, A., Man, S., Martindale, J., Hartley, J., Walsh, T., Gelsthorpe, C., Baxter, L., Forster, G., Fox, M., Bury, J., Mok, K., McDermott, C.J., Traynor, B.J., Kirby, J., Wharton, S.B., Ince, P.G., Hardy, J. and Shaw, P.J. 2012. Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. *Brain: A Journal of Neurology*. **135**(Pt 3),pp.751–764.

Corbo, M. and Hays, A.P. 1992. Peripherin and neurofilament protein coexist in spinal spheroids of motor neuron disease. *Journal of Neuropathology and Experimental Neurology*. **51**(5),pp.531–537.

Corrado, L., Bo, R.D., Castellotti, B., Ratti, A., Cereda, C., Penco, S., Sorarù, G., Carlomagno, Y., Ghezzi, S., Pensato, V., Colombrita, C., Gagliardi, S., Cozzi, L., Orsetti, V., Mancuso, M., Siciliano, G., Mazzini, L., Comi, G.P., Gellera, C., Ceroni, M., D'Alfonso, S. and Silani, V. 2010. Mutations of FUS gene in sporadic amyotrophic lateral sclerosis. *Journal of Medical Genetics*. **47**(3),pp.190–194.

Corrado, L., Carlomagno, Y., Falasco, L., Mellone, S., Godi, M., Cova, E., Cereda, C., Testa, L., Mazzini, L. and D'Alfonso, S. 2011. A novel peripherin gene (PRPH) mutation identified in one sporadic amyotrophic lateral sclerosis patient. *Neurobiology of Aging*. **32**(3),p.552.e1-552.e6.

Corrado, L., Ratti, A., Gellera, C., Buratti, E., Castellotti, B., Carlomagno, Y., Ticozzi, N., Mazzini, L., Testa, L., Taroni, F., Baralle, F.E., Silani, V. and D'Alfonso, S. 2009. High frequency of TARDBP gene mutations in Italian patients with amyotrophic lateral sclerosis. *Human Mutation*. **30**(4),pp.688–694.

Couillard-Després, S., Zhu, Q., Wong, P.C., Price, D.L., Cleveland, D.W. and Julien, J.P. 1998. Protective effect of neurofilament heavy gene overexpression in motor neuron disease induced by mutant superoxide dismutase. *Proceedings of the National Academy of Sciences of the United States of America*. **95**(16),pp.9626–9630.

Cox, L.E., Ferraiuolo, L., Goodall, E.F., Heath, P.R., Higginbottom, A., Mortiboys, H., Hollinger, H.C., Hartley, J.A., Brockington, A., Burness, C.E., Morrison, K.E., Wharton, S.B., Grierson, A.J., Ince, P.G., Kirby, J. and Shaw, P.J. 2010. Mutations in CHMP2B in Lower Motor Neuron Predominant Amyotrophic Lateral Sclerosis (ALS). *PLOS ONE*. **5**(3),p.e9872.

Crabtree, B., Thiyagarajan, N., Prior, S.H., Wilson, P., Iyer, S., Ferns, T., Shapiro, R., Brew, K., Subramanian, V. and Acharya, K.R. 2007. Characterization of human angiogenin variants implicated in amyotrophic lateral sclerosis. *Biochemistry*. **46**(42),pp.11810–11818.

Cronin, S., Greenway, M.J., Ennis, S., Kieran, D., Green, A., Prehn, J.H.M. and Hardiman, O. 2006. Elevated serum angiogenin levels in ALS. *Neurology*. **67**(10),pp.1833–1836.

Cronin, S., Greenway, M.J., Prehn, J.H.M. and Hardiman, O. 2007. Paraoxonase promoter and intronic variants modify risk of sporadic amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*. **78**(9),pp.984–986.

van Damme, P., Veldink, J.H., van Blitterswijk, M., Corveleyn, A., van Vught, P.W.J., Thijs, V., Dubois, B., Matthijs, G., van den Berg, L.H. and Robberecht, W. 2011. Expanded

ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. *Neurology*. **76**(24),pp.2066–2072.

Daoud, H., Dobrzeniecka, S., Camu, W., Meininger, V., Dupré, N., Dion, P.A. and Rouleau, G.A. 2013. Mutation analysis of PFN1 in familial amyotrophic lateral sclerosis patients. *Neurobiology of Aging*. **34**(4),p.1311.e1-2.

Daoud, H., Suhail, H., Szuto, A., Camu, W., Salachas, F., Meininger, V., Bouchard, J.-P., Dupré, N., Dion, P.A. and Rouleau, G.A. 2012a. UBQLN2 mutations are rare in French and French-Canadian amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(9),p.2230.e1-2230.e5.

Daoud, H., Valdmanis, P.N., Gros-Louis, F., Belzil, V., Spiegelman, D., Henrion, E., Diallo, O., Desjarlais, A., Gauthier, J., Camu, W., Dion, P.A. and Rouleau, G.A. 2011. Resequencing of 29 candidate genes in patients with familial and sporadic amyotrophic lateral sclerosis. *Archives of Neurology*. **68**(5),pp.587–593.

Daoud, H., Valdmanis, P.N., Kabashi, E., Dion, P., Dupré, N., Camu, W., Meininger, V. and Rouleau, G.A. 2009. Contribution of TARDBP mutations to sporadic amyotrophic lateral sclerosis. *Journal of Medical Genetics*. **46**(2),pp.112–114.

Daoud, H., Zhou, S., Noreau, A., Sabbagh, M., Belzil, V., Dionne-Laporte, A., Tranchant, C., Dion, P. and Rouleau, G.A. 2012b. Exome sequencing reveals SPG11 mutations causing juvenile ALS. *Neurobiology of Aging*. **33**(4),p.839.e5-9.

De Amicis, A., Piane, M., Ferrari, F., Fanciulli, M., Delia, D. and Chessa, L. 2011. Role of senataxin in DNA damage and telomeric stability. *DNA repair*. **10**(2),pp.199–209.

DeJesus-Hernandez, Desaro, P., Johnston, A., Ross, O.A., Wszolek, Z.K., Ertekin-Taner, N., Graff-Radford, N.R., Rademakers, R. and Boylan, K. 2011a. Novel p.Ile151Val mutation in VCP in a patient of African American descent with sporadic ALS. *Neurology*. **77**(11),pp.1102–1103.

DeJesus-Hernandez, Kocerha, J., Finch, N., Crook, R., Baker, M., Desaro, P., Johnston, A., Rutherford, N., Wojtas, A., Kennelly, K., Wszolek, Z.K., Graff-Radford, N., Boylan, K. and Rademakers, R. 2010. De novo truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis. *Human Mutation*. **31**(5),pp.E1377–E1389.

DeJesus-Hernandez, Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y.R., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R. and Rademakers, R. 2011b. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*. **72**(2),pp.245–256.

Del Bo, R., Ghezzi, S., Corti, S., Pandolfo, M., Ranieri, M., Santoro, D., Ghione, I., Prelle, A., Orsetti, V., Mancuso, M., Sorarù, G., Briani, C., Angelini, C., Siciliano, G., Bresolin, N. and Comi, G.P. 2009. TARDBP (TDP-43) sequence analysis in patients with familial and sporadic ALS: identification of two novel mutations. *European Journal of Neurology*. **16**(6),pp.727–732.

Del Bo, R., Tiloca, C., Pensato, V., Corrado, L., Ratti, A., Ticozzi, N., Corti, S., Castellotti, B., Mazzini, L., Sorarù, G., Cereda, C., D'Alfonso, S., Gellera, C., Comi, G.P., Silani, V. and SLAGEN Consortium 2011. Novel optineurin mutations in patients with familial

and sporadic amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*. **82**(11),pp.1239–1243.

Deng, H., Gao, K. and Jankovic, J. 2014. The role of FUS gene variants in neurodegenerative diseases. *Nature Reviews Neurology*. **10**(6),pp.337–348.

Deng, H.-X., Chen, W., Hong, S.-T., Boycott, K.M., Gorrie, G.H., Siddique, N., Yang, Y., Fecto, F., Shi, Y., Zhai, H., Jiang, H., Hirano, M., Rampersaud, E., Jansen, G.H., Donkervoort, S., Bigio, E.H., Brooks, B.R., Ajroud, K., Sufit, R.L., Haines, J.L., Mugnaini, E., Pericak-Vance, M.A. and Siddique, T. 2011. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature*. **477**(7363),pp.211–215.

Deng, H.-X., Zhai, H., Bigio, E.H., Yan, J., Fecto, F., Ajroud, K., Mishra, M., Ajroud-Driss, S., Heller, S., Sufit, R., Siddique, N., Mugnaini, E. and Siddique, T. 2010. FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. *Annals of Neurology*. **67**(6),pp.739–748.

Devon, R.S., Helm, J.R., Rouleau, G.A., Leitner, Y., Lerman-Sagie, T., Lev, D. and Hayden, M.R. 2003. The first nonsense mutation in alsin results in a homogeneous phenotype of infantile-onset ascending spastic paralysis with bulbar involvement in two siblings. *Clinical Genetics*. **64**(3),pp.210–215.

Di Paolo, G. and De Camilli, P. 2006. Phosphoinositides in cell regulation and membrane dynamics. *Nature*. **443**(7112),pp.651–657.

Di Vito, L., de Biase, D., Pession, A., Visani, M., Liguori, R., Zambito Marsala, S., Leta, V., De Carolis, P. and Donadio, V. 2013. Brachial amyotrophic diplegia associated with the a140a superoxide dismutase 1 mutation. *Neurogenetics*. **14**(3–4),pp.255–256.

Dillen, L., Van Langenhove, T., Engelborghs, S., Vandenbulcke, M., Sarafov, S., Tournev, I., Merlin, C., Cras, P., Vandenberghe, R., De Deyn, P.P., Jordanova, A., Cruts, M., Van Broeckhoven, C., van der Zee, J. and BELNEU consortium 2013. Explorative genetic study of UBQLN2 and PFN1 in an extended Flanders-Belgian cohort of frontotemporal lobar degeneration patients. *Neurobiology of Aging*. **34**(6),p.1711.e1-5.

Dini Modigliani, S., Morlando, M., Errichelli, L., Sabatelli, M. and Bozzoni, I. 2014. An ALS-associated mutation in the FUS 3′-UTR disrupts a microRNA–FUS regulatory circuitry. *Nature Communications*. **5**,p.4335.

van Doormaal, P.T.C., van Rheenen, W., van Blitterswijk, M., Schellevis, R.D., Schelhaas, H.J., de Visser, M., van der Kooi, A.J., Veldink, J.H. and van den Berg, L.H. 2012. UBQLN2 in familial amyotrophic lateral sclerosis in The Netherlands. *Neurobiology of Aging*. **33**(9),p.2233.e7-2233.e8.

Drepper, C., Herrmann, T., Wessig, C., Beck, M. and Sendtner, M. 2011. C-terminal FUS/TLS mutations in familial and sporadic ALS in Germany. *Neurobiology of Aging*. **32**(3),p.548.e1-4.

Drew, A.P., Zhu, D., Kidambi, A., Ly, C., Tey, S., Brewer, M.H., Ahmad-Annuar, A., Nicholson, G.A. and Kennerson, M.L. 2015. Improved inherited peripheral neuropathy genetic diagnosis by whole-exome sequencing. *Molecular Genetics & Genomic Medicine*. **3**(2),pp.143–154.

Eker, H.K., Unlü, S.E., Al-Salmi, F. and Crosby, A.H. 2014. A novel homozygous mutation in ALS2 gene in four siblings with infantile-onset ascending hereditary spastic paralysis. *European Journal of Medical Genetics*. **57**(6),pp.275–278.

Elden, A.C., Kim, H.-J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M., Padmanabhan, A., Clay-Falcone, D., McCluskey, L., Elman, L., Juhr, D., Gruber, P.J., Rüb, U., Auburger, G., Trojanowski, J.Q., Lee, V.M.-Y., Van Deerlin, V.M., Bonini, N.M. and Gitler, A.D. 2010. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*. **466**(7310),pp.1069–1075.

van Es, M.A., Diekstra, F.P., Veldink, J.H., Baas, F., Bourque, P.R., Schelhaas, H.J., Strengman, E., Hennekam, E. a. M., Lindhout, D., Ophoff, R.A. and van den Berg, L.H. 2009. A case of ALS-FTD in a large FALS pedigree with a K17I ANG mutation. *Neurology*. **72**(3),pp.287–288.

van Es, M.A., Schelhaas, H.J., van Vught, P.W.J., Ticozzi, N., Andersen, P.M., Groen, E.J.N., Schulte, C., Blauw, H.M., Koppers, M., Diekstra, F.P., Fumoto, K., LeClerc, A.L., Keagle, P., Bloem, B.R., Scheffer, H., van Nuenen, B.F.L., van Blitterswijk, M., van Rheenen, W., Wills, A.-M., Lowe, P.P., Hu, G., Yu, W., Kishikawa, H., Wu, D., Folkerth, R.D., Mariani, C., Goldwurm, S., Pezzoli, G., Van Damme, P., Lemmens, R., Dahlberg, C., Birve, A., Fernández-Santiago, R., Waibel, S., Klein, C., Weber, M., van der Kooi, A.J., de Visser, M., Verbaan, D., van Hilten, J.J., Heutink, P., Hennekam, E.A.M., Cuppen, E., Berg, D., Brown, R.H., Silani, V., Gasser, T., Ludolph, A.C., Robberecht, W., Ophoff, R.A., Veldink, J.H., Pasterkamp, R.J., de Bakker, P.I.W., Landers, J.E., van de Warrenburg, B.P. and van den Berg, L.H. 2011. Angiogenin variants in Parkinson disease and amyotrophic lateral sclerosis. *Annals of Neurology*. **70**(6),pp.964–973.

Eskildsen, S.F., Østergaard, L.R., Rodell, A.B., Østergaard, L., Nielsen, J.E., Isaacs, A.M. and Johannsen, P. 2009. Cortical volumes and atrophy rates in FTD-3 CHMP2B mutation carriers and related non-carriers. *NeuroImage*. **45**(3),pp.713–721.

Eymard-Pierre, E., Lesca, G., Dollet, S., Santorelli, F.M., di Capua, M., Bertini, E. and Boespflug-Tanguy, O. 2002. Infantile-onset ascending hereditary spastic paralysis is associated with mutations in the alsin gene. *American Journal of Human Genetics*. **71**(3),pp.518–527.

Farrawell, N.E., Lambert-Smith, I.A., Warraich, S.T., Blair, I.P., Saunders, D.N., Hatters, D.M. and Yerbury, J.J. 2015. Distinct partitioning of ALS associated TDP-43, FUS and SOD1 mutants into cellular inclusions. *Scientific Reports*. **5**,p.13416.

Fecto, F., Yan, J., Vemula, S.P., Liu, E., Yang, Y., Chen, W., Zheng, J.G., Shi, Y., Siddique, N., Arrat, H., Donkervoort, S., Ajroud-Driss, S., Sufit, R.L., Heller, S.L., Deng, H.-X. and Siddique, T. 2011. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Archives of Neurology*. **68**(11),pp.1440–1446.

Felbecker, A., Camu, W., Valdmanis, P.N., Sperfeld, A.-D., Waibel, S., Steinbach, P., Rouleau, G.A., Ludolph, A.C. and Andersen, P.M. 2010. Four familial ALS pedigrees discordant for two SOD1 mutations: are all SOD1 mutations pathogenic? *Journal of Neurology, Neurosurgery & Psychiatry*. **81**(5),pp.572–577.

Ferguson, C.J., Lenk, G.M. and Meisler, M.H. 2009. Defective autophagy in neurons and astrocytes from mice deficient in PI(3,5)P2. *Human Molecular Genetics*. **18**(24),pp.4868–4878.

Fernández-Santiago, R., Hoenig, S., Lichtner, P., Sperfeld, A.-D., Sharma, M., Berg, D., Weichenrieder, O., Illig, T., Eger, K., Meyer, T., Anneser, J., Münch, C., Zierz, S., Gasser, T. and Ludolph, A. 2009. Identification of novel Angiogenin (ANG) gene missense variants in German patients with amyotrophic lateral sclerosis. *Journal of Neurology*. **256**(8),pp.1337–1342.

Fernández-Santiago, R., Sharma, M., Mueller, J.C., Gohlke, H., Illig, T., Anneser, J., Münch, C., Ludolph, A., Kamm, C. and Gasser, T. 2006. Possible gender-dependent association of vascular endothelial growth factor (VEGF) gene and ALS. *Neurology*. **66**(12),pp.1929–1931.

Ferraiuolo, L., Kirby, J., Grierson, A.J., Sendtner, M. and Shaw, P.J. 2011. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nature Reviews. Neurology*. **7**(11),pp.616–630.

Ferrari, R., Kapogiannis, D., Huey, E.D., Grafman, J., Hardy, J. and Momeni, P. 2010. Novel missense mutation in charged multivesicular body protein 2B in a patient with frontotemporal dementia. *Alzheimer Disease and Associated Disorders*. **24**(4),pp.397–401.

Ferrari, R., Kapogiannis, D., Huey, E.D. and Momeni, P. 2011. FTD and ALS: a tale of two diseases. *Current Alzheimer research*. **8**(3),pp.273–294.

Fifita, J.A., Williams, K.L., McCann, E.P., O'Brien, A., Bauer, D.C., Nicholson, G.A. and Blair, I.P. 2015. Mutation analysis of MATR3 in Australian familial amyotrophic lateral sclerosis. *Neurobiology of Aging*. **36**(3),p.1602.e1-1602.e2.

Figlewicz, D.A., Krizus, A., Martinoli, M.G., Meininger, V., Dib, M., Rouleau, G.A. and Julien, J.-P. 1994. Variants of the heavy neurofilament subunit are associated with the development of amyotrophic lateral sclerosis. *Human Molecular Genetics*. **3**(10),pp.1757–1761.

Figueroa-Romero, C., Hur, J., Bender, D.E., Delaney, C.E., Cataldo, M.D., Smith, A.L., Yung, R., Ruden, D.M., Callaghan, B.C. and Feldman, E.L. 2012. Identification of epigenetically altered genes in sporadic amyotrophic lateral sclerosis. *PloS One*. **7**(12),p.e52672.

Ford, J.W. and McVicar, D.W. 2009. TREM and TREM-like receptors in inflammation and disease. *Current Opinion in Immunology*. **21**(1),pp.38–46.

Fratta, P., Charnock, J., Collins, T., Devoy, A., Howard, R., Malaspina, A., Orrell, R., Sidle, K., Clarke, J., Shoai, M., Lu, C.-H., Hardy, J., Plagnol, V. and Fisher, E.M.C. 2014. Profilin1 E117G is a moderate risk factor for amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*. **85**(5),pp.506–508.

Gal, J., Ström, A.-L., Kilty, R., Zhang, F. and Zhu, H. 2007. p62 accumulates and enhances aggregate formation in model systems of familial amyotrophic lateral sclerosis. *The Journal of Biological Chemistry*. **282**(15),pp.11068–11077.

Gallego-Iradi, M.C., Clare, A.M., Brown, H.H., Janus, C., Lewis, J. and Borchelt, D.R. 2015. Subcellular Localization of Matrin 3 Containing Mutations Associated with ALS and Distal Myopathy. *PloS One*. **10**(11),p.e0142144.

Gamez, J., Corbera-Bellalta, M., Nogales, G., Raguer, N., García-Arumí, E., Badia-Canto, M., Lladó-Carbó, E. and Alvarez-Sabín, J. 2006. Mutational analysis of the Cu/Zn

superoxide dismutase gene in a Catalan ALS population: should all sporadic ALS cases also be screened for SOD1? *Journal of the Neurological Sciences*. **247**(1),pp.21–28.

Garcia, M.L., Singleton, A.B., Hernandez, D., Ward, C.M., Evey, C., Sapp, P.A., Hardy, J., Brown, R.H. and Cleveland, D.W. 2006. Mutations in neurofilament genes are not a significant primary cause of non-SOD1-mediated amyotrophic lateral sclerosis. *Neurobiology of Disease*. **21**(1),pp.102–109.

Gellera, C., Castellotti, B., Riggio, M.C., Silani, V., Morandi, L., Testa, D., Casali, C., Taroni, F., Di Donato, S., Zeviani, M. and Mariotti, C. 2001. Superoxide dismutase gene mutations in Italian patients with familial and sporadic amyotrophic lateral sclerosis: identification of three novel missense mutations. *Neuromuscular disorders: NMD*. **11**(4),pp.404–410.

Gellera, C., Colombrita, C., Ticozzi, N., Castellotti, B., Bragato, C., Ratti, A., Taroni, F. and Silani, V. 2008. Identification of new ANG gene mutations in a large cohort of Italian patients with amyotrophic lateral sclerosis. *Neurogenetics*. **9**(1),pp.33–40.

Gellera, C., Tiloca, C., Del Bo, R., Corrado, L., Pensato, V., Agostini, J., Cereda, C., Ratti, A., Castellotti, B., Corti, S., Bagarotti, A., Cagnin, A., Milani, P., Gabelli, C., Riboldi, G., Mazzini, L., Sorarù, G., D'Alfonso, S., Taroni, F., Comi, G.P., Ticozzi, N., Silani, V. and SLAGEN Consortium 2013. Ubiquilin 2 mutations in Italian patients with amyotrophic lateral sclerosis and frontotemporal dementia. *Journal of Neurology, Neurosurgery, and Psychiatry*. **84**(2),pp.183–187.

Ghani, M., Lang, A.E., Zinman, L., Nacmias, B., Sorbi, S., Bessi, V., Tedde, A., Tartaglia, M.C., Surace, E.I., Sato, C., Moreno, D., Xi, Z., Hung, R., Nalls, M.A., Singleton, A., St George-Hyslop, P. and Rogaeva, E. 2015. Mutation analysis of patients with neurodegenerative disorders using NeuroX array. *Neurobiology of Aging*. **36**(1),p.545.e9-14.

Gijselinck, I., Sleegers, K., Engelborghs, S., Robberecht, W., Martin, J.-J., Vandenberghe, R., Sciot, R., Dermaut, B., Goossens, D., van der Zee, J., De Pooter, T., Del-Favero, J., Santens, P., De Jonghe, P., De Deyn, P.P., Van Broeckhoven, C. and Cruts, M. 2009. Neuronal inclusion protein TDP-43 has no primary genetic role in FTD and ALS. *Neurobiology of Aging*. **30**(8),pp.1329–1331.

Gijselinck, I., Van Mossevelde, S., van der Zee, J., Sieben, A., Engelborghs, S., De Bleecker, J., Ivanoiu, A., Deryck, O., Edbauer, D., Zhang, M., Heeman, B., Bäumer, V., Van den Broeck, M., Mattheijssens, M., Peeters, K., Rogaeva, E., De Jonghe, P., Cras, P., Martin, J.-J., de Deyn, P.P., Cruts, M. and Van Broeckhoven, C. 2015. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Molecular Psychiatry*.

Gilissen, C., Hoischen, A., Brunner, H.G. and Veltman, J.A. 2012. Disease gene identification strategies for exome sequencing. *European journal of human genetics: EJHG*. **20**(5),pp.490–497.

Giordano, G., Sánchez-Pérez, A.M., Montoliu, C., Berezney, R., Malyavantham, K., Costa, L.G., Calvete, J.J. and Felipo, V. 2005. Activation of NMDA receptors induces protein kinase A-mediated phosphorylation and degradation of matrin 3. Blocking these effects prevents NMDA-induced neuronal death. *Journal of Neurochemistry*. **94**(3),pp.808–818.

Giraldo, M., Lopera, F., Siniard, A.L., Corneveaux, J.J., Schrauwen, I., Carvajal, J., Muñoz, C., Ramirez-Restrepo, M., Gaiteri, C., Myers, A.J., Caselli, R.J., Kosik, K.S., Reiman, E.M. and Huentelman, M.J. 2013. Variants in triggering receptor expressed on myeloid cells 2 are associated with both behavioral variant frontotemporal lobar degeneration and Alzheimer's disease. *Neurobiology of Aging*. **34**(8),p.2077.e11-18.

Gitcho, M.A., Bigio, E.H., Mishra, M., Johnson, N., Weintraub, S., Mesulam, M., Rademakers, R., Chakraverty, S., Cruchaga, C., Morris, J.C., Goate, A.M. and Cairns, N.J. 2009. TARDBP 3'-UTR variant in autopsy-confirmed frontotemporal lobar degeneration with TDP-43 proteinopathy. *Acta Neuropathologica*. **118**(5),pp.633–645.

Gitler, A.D. and Tsuiji, H. 2016. There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS. *Brain Research*.

González-Pérez, P., Cirulli, E.T., Drory, V.E., Dabby, R., Nisipeanu, P., Carasso, R.L., Sadeh, M., Fox, A., Festoff, B.W., Sapp, P.C., McKenna-Yasek, D., Goldstein, D.B., Brown, R.H. and Blumen, S.C. 2012. Novel mutation in VCP gene causes atypical amyotrophic lateral sclerosis. *Neurology*. **79**(22),pp.2201–2208.

Gorrie, G.H., Fecto, F., Radzicki, D., Weiss, C., Shi, Y., Dong, H., Zhai, H., Fu, R., Liu, E., Li, S., Arrat, H., Bigio, E.H., Disterhoft, J.F., Martina, M., Mugnaini, E., Siddique, T. and Deng, H.-X. 2014. Dendritic spinopathy in transgenic mice expressing ALS/dementia-linked mutant UBQLN2. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(40),pp.14524–14529.

Greenway, M.J., Alexander, M.D., Ennis, S., Traynor, B.J., Corr, B., Frost, E., Green, A. and Hardiman, O. 2004. A novel candidate region for ALS on chromosome 14q11.2. *Neurology*. **63**(10),pp.1936–1938.

Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., Morrison, K.E., Green, A., Acharya, K.R., Brown, R.H. and Hardiman, O. 2006. ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nature Genetics*. **38**(4),pp.411–413.

Groen, E.J.N., van Es, M.A., van Vught, P.W.J., Spliet, W.G.M., van Engelen-Lee, J., de Visser, M., Wokke, J.H.J., Schelhaas, H.J., Ophoff, R.A., Fumoto, K., Pasterkamp, R.J., Dooijes, D., Cuppen, E., Veldink, J.H. and van den Berg, L.H. 2010. FUS mutations in familial amyotrophic lateral sclerosis in the Netherlands. *Archives of Neurology*. **67**(2),pp.224–230.

Gros-Louis, F., Kriz, J., Kabashi, E., McDearmid, J., Millecamps, S., Urushitani, M., Lin, L., Dion, P., Zhu, Q., Drapeau, P., Julien, J.-P. and Rouleau, G.A. 2008. Als2 mRNA splicing variants detected in KO mice rescue severe motor dysfunction phenotype in Als2 knock-down zebrafish. *Human Molecular Genetics*. **17**(17),pp.2691–2702.

Gros-Louis, F., Larivière, R., Gowing, G., Laurent, S., Camu, W., Bouchard, J.-P., Meininger, V., Rouleau, G.A. and Julien, J.-P. 2004. A Frameshift Deletion in Peripherin Gene Associated with Amyotrophic Lateral Sclerosis. *Journal of Biological Chemistry*. **279**(44),pp.45951–45956.

Gros-Louis, F., Laurent, S., Lopes, A.A.S., Khoris, J., Meininger, V., Camu, W. and Rouleau, G.A. 2003. Absence of mutations in the hypoxia response element of VEGF in ALS. *Muscle & Nerve*. **28**(6),pp.774–775.

Guerreiro, R.J., Gustafson, D.R. and Hardy, J. 2012. The genetic architecture of Alzheimer's disease: beyond APP, PSENs and APOE. *Neurobiology of Aging*. **33**(3),pp.437–456.

Guerreiro, R.J., Lohmann, E., Brás, J.M., Gibbs, J.R., Rohrer, J.D., Gurunlian, N., Dursun, B., Bilgic, B., Hanagasi, H., Gurvit, H., Emre, M., Singleton, A. and Hardy, J. 2013. Using exome sequencing to reveal mutations in TREM2 presenting as a frontotemporal dementia-like syndrome without bone involvement. *JAMA neurology*. **70**(1),pp.78–84.

Guerreiro, R.J., Schymick, J.C., Crews, C., Singleton, A., Hardy, J. and Traynor, B.J. 2008. TDP-43 Is Not a Common Cause of Sporadic Amyotrophic Lateral Sclerosis. *PLOS ONE*. **3**(6),p.e2450.

Hadano, S., Hand, C.K., Osuga, H., Yanagisawa, Y., Otomo, A., Devon, R.S., Miyamoto, N., Showguchi-Miyata, J., Okada, Y., Singaraja, R., Figlewicz, D.A., Kwiatkowski, T., Hosler, B.A., Sagie, T., Skaug, J., Nasir, J., Brown, R.H., Scherer, S.W., Rouleau, G.A., Hayden, M.R. and Ikeda, J.E. 2001. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nature Genetics*. **29**(2),pp.166–173.

Haley, R.W. 2003. Excess incidence of ALS in young Gulf War veterans. *Neurology*. **61**(6),pp.750–756.

Han, J.-H., Ryu, H.-H., Jun, M.-H., Jang, D.-J. and Lee, J.-A. 2012. The functional analysis of the CHMP2B missense mutation associated with neurodegenerative diseases in the endo-lysosomal pathway. *Biochemical and Biophysical Research Communications*. **421**(3),pp.544–549.

Hand, C.K., Devon, R.S., Gros-Louis, F., Rochefort, D., Khoris, J., Meininger, V., Bouchard, J.-P., Camu, W., Hayden, M.R. and Rouleau, G.A. 2003. Mutation screening of the ALS2 gene in sporadic and familial amyotrophic lateral sclerosis. *Archives of Neurology*. **60**(12),pp.1768–1771.

Hand, C.K., Mayeux-Portas, V., Khoris, J., Briolotti, V., Clavelou, P., Camu, W. and Rouleau, G.A. 2001. Compound heterozygous D90A and D96N SOD1 mutations in a recessive amyotrophic lateral sclerosis family. *Annals of Neurology*. **49**(2),pp.267–271.

Hayward, C., Brock, D.J., Minns, R.A. and Swingler, R.J. 1998. Homozygosity for Asn86Ser mutation in the CuZn-superoxide dismutase gene produces a severe clinical phenotype in a juvenile onset case of familial amyotrophic lateral sclerosis. *Journal of Medical Genetics*. **35**(2),p.174.

Hayward, C., Colville, S., Swingler, R.J. and Brock, D.J. 1999. Molecular genetic analysis of the APEX nuclease gene in amyotrophic lateral sclerosis. *Neurology*. **52**(9),pp.1899–1901.

Hentati, A., Bejaoui, K., Pericak-Vance, M.A., Hentati, F., Speer, M.C., Hung, W.Y., Figlewicz, D.A., Haines, J., Rimmler, J. and Ben Hamida, C. 1994. Linkage of recessive familial amyotrophic lateral sclerosis to chromosome 2q33-q35. *Nature Genetics*. **7**(3),pp.425–428.

Hernández, I., Espinosa, A., Real, L.M., Galán, J.J., Mauleón, A., Roca, M.R., Tárraga, L., Ruiz, A. and Boada, M. 2012. Molecular evaluation of human Ubiquilin 2 gene PXX

domain in familial frontotemporal dementia patients. *Journal of Neurology*. **259**(11),pp.2488–2490.

Herzfeld, T., Wolf, N., Winter, P., Hackstein, H., Vater, D. and Müller, U. 2009. Maternal uniparental heterodisomy with partial isodisomy of a chromosome 2 carrying a splice acceptor site mutation (IVS9-2A>T) in ALS2 causes infantile-onset ascending spastic paralysis (IAHSP). *Neurogenetics*. **10**(1),pp.59–64.

Hicks, S., Wheeler, D.A., Plon, S.E. and Kimmel, M. 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Human Mutation*. **32**(6),pp.661–668.

Hirano, A., Donnenfeld, H., Sasaki, S. and Nakano, I. 1984. Fine structural observations of neurofilamentous changes in amyotrophic lateral sclerosis. *Journal of Neuropathology and Experimental Neurology*. **43**(5),pp.461–470.

Hirano, M., Nakamura, Y., Saigoh, K., Sakamoto, H., Ueno, S., Isono, C., Mitsui, Y. and Kusunoki, S. 2015. VCP gene analyses in Japanese patients with sporadic amyotrophic lateral sclerosis identify a new mutation. *Neurobiology of Aging*. **36**(3),p.1604.e1-6.

Hirano, M., Nakamura, Y., Saigoh, K., Sakamoto, H., Ueno, S., Isono, C., Miyamoto, K., Akamatsu, M., Mitsui, Y. and Kusunoki, S. 2013. Mutations in the gene encoding p62 in Japanese patients with amyotrophic lateral sclerosis. *Neurology*. **80**(5),pp.458–463.

Hirano, M., Quinzii, C.M., Mitsumoto, H., Hays, A.P., Roberts, J.K., Richard, P. and Rowland, L.P. 2011. Senataxin mutations and amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **12**(3),pp.223–227.

Hocking, L.J., Lucas, G.J.A., Daroszewska, A., Mangion, J., Olavesen, M., Cundy, T., Nicholson, G.C., Ward, L., Bennett, S.T., Wuyts, W., Hul, W.V. and Ralston, S.H. 2002. Domain-specific mutations in sequestosome 1 (SQSTM1) cause familial and sporadic Paget's disease. *Human Molecular Genetics*. **11**(22),pp.2735–2739.

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M.N., Smith, S.W., Middle, C.M., Rodesch, M.J., Albert, T.J., Hannon, G.J. and McCombie, W.R. 2007. Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*. **39**(12),pp.1522–1527.

Holm, I.E., Isaacs, A.M. and Mackenzie, I.R.A. 2009. Absence of FUS-immunoreactive pathology in frontotemporal dementia linked to chromosome 3 (FTD-3) caused by mutation in the CHMP2B gene. *Acta Neuropathologica*. **118**(5),pp.719–720.

Hortobágyi, T., Troakes, C., Nishimura, A.L., Vance, C., Swieten, J.C. van, Seelaar, H., King, A., Al-Sarraj, S., Rogelj, B. and Shaw, C.E. 2011. Optineurin inclusions occur in a minority of TDP-43 positive ALS and FTLD-TDP cases and are rarely observed in other neurodegenerative disorders. *Acta Neuropathologica*. **121**(4),pp.519–527.

Houdayer, C., Caux-Moncoutier, V., Krieger, S., Barrois, M., Bonnet, F., Bourdon, V., Bronner, M., Buisson, M., Coulet, F., Gaildrat, P., Lefol, C., Léone, M., Mazoyer, S., Muller, D., Remenieras, A., Révillion, F., Rouleau, E., Sokolowska, J., Vert, J.-P., Lidereau, R., Soubrier, F., Sobol, H., Sevenet, N., Bressac-de Paillerets, B., Hardouin, A., Tosi, M., Sinilnikova, O.M. and Stoppa-Lyonnet, D. 2012. Guidelines

for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. *Human Mutation*. **33**(8),pp.1228–1238.

Huey, E.D., Ferrari, R., Moreno, J.H., Jensen, C., Morris, C.M., Potocnik, F., Kalaria, R.N., Tierney, M., Wassermann, E.M., Hardy, J., Grafman, J. and Momeni, P. 2012. FUS and TDP43 genetic variability in FTD and CBS. *Neurobiology of aging*. **33**(5),p.1016.e9-1016.17.

Iida, A., Hosono, N., Sano, M., Kamei, T., Oshima, S., Tokuda, T., Kubo, M., Nakamura, Y. and Ikegawa, S. 2012. Optineurin mutations in Japanese amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry*. **83**(2),pp.233–235.

Ince, P.G., Lowe, J. and Shaw, P.J. 1998. Amyotrophic lateral sclerosis: current issues in classification, pathogenesis and molecular pathology. *Neuropathology and Applied Neurobiology*. **24**(2),pp.104–117.

Infante, J., Berciano, J., Volpini, V., Corral, J., Polo, J.M., Pascual, J. and Combarros, O. 2004. Spinocerebellar ataxia type 2 with Levodopa-responsive parkinsonism culminating in motor neuron disease. *Movement Disorders: Official Journal of the Movement Disorder Society*. **19**(7),pp.848–852.

Ingre, C., Landers, J.E., Rizik, N., Volk, A.E., Akimoto, C., Birve, A., Hübers, A., Keagle, P.J., Piotrowska, K., Press, R., Andersen, P.M., Ludolph, A.C. and Weishaupt, J.H. 2013. A novel phosphorylation site mutation in profilin 1 revealed in a large screen of US, Nordic, and German amyotrophic lateral sclerosis/frontotemporal dementia cohorts. *Neurobiology of Aging*. **34**(6),p.1708.e1-6.

Ito, H., Nakamura, M., Komure, O., Ayaki, T., Wate, R., Maruyama, H., Nakamura, Y., Fujita, K., Kaneko, S., Okamoto, Y., Ihara, M., Konishi, T., Ogasawara, K., Hirano, A., Kusaka, H., Kaji, R., Takahashi, R. and Kawakami, H. 2011. Clinicopathologic study on an ALS family with a heterozygous E478G optineurin mutation. *Acta Neuropathologica*. **122**(2),pp.223–229.

Itzcovich, T., Xi, Z., Martinetto, H., Chrem-Méndez, P., Russo, M.J., de Ambrosi, B., Uchitel, O.D., Nogués, M., Silva, E., Rojas, G., Bagnatti, P., Amengual, A., Campos, J., Rogaeva, E., St George-Hyslop, P., Allegri, R., Sevlever, G. and Surace, E.I. 2016. Analysis of C9orf72 in patients with frontotemporal dementia and amyotrophic lateral sclerosis from Argentina. *Neurobiology of Aging*. **40**,p.192.e13-15.

Jacquin, A., Rouaud, O., Soichot, P., Bejot, Y., Dygai-Cochet, I., Sarazin, M., Stojkovic, T., Lemesle-Martin, M., Giroud, M. and Moreau, T. 2013. Psychiatric Presentation of Frontotemporal Dementia Associated with Inclusion Body Myopathy due to the VCP Mutation (R155H) in a French Family. *Case Reports in Neurology*. **5**(3),pp.187–194.

Jiao, B., Xiao, T., Hou, L., Gu, X., Zhou, Y., Zhou, L., Tang, B., Xu, J. and Shen, L. 2015. High prevalence of CHCHD10 mutation in patients with frontotemporal dementia from China. *Brain*.,p.awv367.

Jin, S.C., Benitez, B.A., Karch, C.M., Cooper, B., Skorupa, T., Carrell, D., Norton, J.B., Hsu, S., Harari, O., Cai, Y., Bertelsen, S., Goate, A.M. and Cruchaga, C. 2014. Coding variants in TREM2 increase risk for Alzheimer's disease. *Human Molecular Genetics*. **23**(21),pp.5838–5846.

Johnson, J.O., Glynn, S.M., Gibbs, J.R., Nalls, M.A., Sabatelli, M., Restagno, G., Drory, V.E., Chiò, A., Rogaeva, E. and Traynor, B.J. 2014a. Mutations in the CHCHD10 gene are a common cause of familial amyotrophic lateral sclerosis. *Brain: A Journal of Neurology*. **137**(Pt 12),p.e311.

Johnson, J.O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V.M., Trojanowski, J.Q., Gibbs, J.R., Brunetti, M., Gronka, S., Wuu, J., Ding, J., McCluskey, L., Martinez-Lage, M., Falcone, D., Hernandez, D.G., Arepalli, S., Chong, S., Schymick, J.C., Rothstein, J., Landi, F., Wang, Y.-D., Calvo, A., Mora, G., Sabatelli, M., Monsurrò, M.R., Battistini, S., Salvi, F., Spataro, R., Sola, P., Borghero, G., ITALSGEN Consortium, Galassi, G., Scholz, S.W., Taylor, J.P., Restagno, G., Chiò, A. and Traynor, B.J. 2010. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron*. **68**(5),pp.857–864.

Johnson, J.O., Pioro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J., Gibbs, J.R., Nalls, M.A., Morgan, S., Shoai, M., Hardy, J., Pittman, A., Orrell, R.W., Malaspina, A., Sidle, K.C., Fratta, P., Harms, M.B., Baloh, R.H., Pestronk, A., Weihl, C.C., Rogaeva, E., Zinman, L., Drory, V.E., Borghero, G., Mora, G., Calvo, A., Rothstein, J.D., ITALSGEN Consortium, Drepper, C., Sendtner, M., Singleton, A.B., Taylor, J.P., Cookson, M.R., Restagno, G., Sabatelli, M., Bowser, R., Chiò, A. and Traynor, B.J. 2014b. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nature Neuroscience*. **17**(5),pp.664–666.

Jones, A. 2011. Relationship between ALS and FTD genes. [Accessed 16 May 2015]. Available from: http://alsod.iop.kcl.ac.uk/als_ftd.aspx.

Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P.V., Snaedal, J., Bjornsson, S., Huttenlocher, J., Levey, A.I., Lah, J.J., Rujescu, D., Hampel, H., Giegling, I., Andreassen, O.A., Engedal, K., Ulstein, I., Djurovic, S., Ibrahim-Verbaas, C., Hofman, A., Ikram, M.A., van Duijn, C.M., Thorsteinsdottir, U., Kong, A. and Stefansson, K. 2013. Variant of TREM2 associated with the risk of Alzheimer's disease. *The New England Journal of Medicine*. **368**(2),pp.107–116.

Jordan, S.A., Farrar, G.J., Kumar-Singh, R., Kenna, P., Humphries, M.M., Allamand, V., Sharp, E.M. and Humphries, P. 1992. Autosomal dominant retinitis pigmentosa (adRP; RP6): cosegregation of RP6 and the peripherin-RDS locus in a late-onset family of Irish origin. *American Journal of Human Genetics*. **50**(3),pp.634–639.

Ju, J.-S. and Weihl, C.C. 2010. p97/VCP at the intersection of the autophagy and the ubiquitin proteasome system. *Autophagy*. **6**(2),pp.283–285.

Kabashi, E., Lin, L., Tradewell, M.L., Dion, P.A., Bercier, V., Bourgouin, P., Rochefort, D., Bel Hadj, S., Durham, H.D., Vande Velde, C., Rouleau, G.A. and Drapeau, P. 2010. Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Human Molecular Genetics*. **19**(4),pp.671–683.

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.-P., Lacomblez, L., Pochigaeva, K., Salachas, F., Pradat, P.-F., Camu, W., Meininger, V., Dupre, N. and Rouleau, G.A. 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature Genetics*. **40**(5),pp.572–574.

Kamel, F., Umbach, D.M., Bedlack, R.S., Richards, M., Watson, M., Alavanja, M.C., Blair, A., Hoppin, J.A., Schmidt, S. and Sandler, D.P. 2012. Pesticide exposure and amyotrophic lateral sclerosis. *Neurotoxicology*. **33**(3),pp.457–462.

Kanekura, K., Suzuki, H., Aiso, S. and Matsuoka, M. 2009. ER stress and unfolded protein response in amyotrophic lateral sclerosis. *Molecular Neurobiology*. **39**(2),pp.81–89.

Karger, B.L. and Guttman, A. 2009. DNA Sequencing by Capillary Electrophoresis. *Electrophoresis*. **30**(Suppl 1),pp.S196–S202.

Kasperaviciute, D., Weale, M.E., Shianna, K.V., Banks, G.T., Simpson, C.L., Hansen, V.K., Turner, M.R., Shaw, C.E., Al-Chalabi, A., Pall, H.S., Goodall, E.F., Morrison, K.E., Orrell, R.W., Beck, M., Jablonka, S., Sendtner, M., Brockington, A., Ince, P.G., Hartley, J., Nixon, H., Shaw, P.J., Schiavo, G., Wood, N.W., Goldstein, D.B. and Fisher, E.M.C. 2007. Large-scale pathways-based association study in amyotrophic lateral sclerosis. *Brain: A Journal of Neurology*. **130**(Pt 9),pp.2292–2301.

Kenna, K.P., McLaughlin, R.L., Byrne, S., Elamin, M., Heverin, M., Kenny, E.M., Cormican, P., Morris, D.W., Donaghy, C.G., Bradley, D.G. and Hardiman, O. 2013. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. *Journal of Medical Genetics*. **50**(11),pp.776–783.

Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R. and Zoing, M.C. 2011. Amyotrophic lateral sclerosis. *Lancet (London, England)*. **377**(9769),pp.942–955.

Kim, H.-J., Kwon, M.-J., Choi, W.-J., Oh, K.-W., Oh, S.-I., Ki, C.-S. and Kim, S.H. 2014. Mutations in UBQLN2 and SIGMAR1 genes are rare in Korean patients with amyotrophic lateral sclerosis. *Neurobiology of Aging*. **35**(8),p.1957.e7-8.

Kim, H.Y., Ki, C.-S., Koh, S.-H., Park, K.-H., Sunwoo, I.-N. and Kim, S.H. 2007. Clinical characteristics of familial amyotrophic lateral sclerosis with a Phe20Cys mutation in the SOD1 gene in a Korean family. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **8**(2),pp.73–78.

Kino, Y., Washizu, C., Kurosawa, M., Yamada, M., Miyazaki, H., Akagi, T., Hashikawa, T., Doi, H., Takumi, T., Hicks, G.G., Hattori, N., Shimogori, T. and Nukina, N. 2015. FUS/TLS deficiency causes behavioral and pathological abnormalities distinct from amyotrophic lateral sclerosis. *Acta Neuropathologica Communications*. [Online]. **3**. [Accessed 18 June 2016]. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4408580/.

Kirby, J., Goodall, E.F., Smith, W., Highley, J.R., Masanzu, R., Hartley, J.A., Hibberd, R., Hollinger, H.C., Wharton, S.B., Morrison, K.E., Ince, P.G., McDermott, C.J. and Shaw, P.J. 2010. Broad clinical phenotypes associated with TAR-DNA binding protein (TARDBP) mutations in amyotrophic lateral sclerosis. *Neurogenetics*. **11**(2),pp.217–225.

Kleinberger, G., Yamanishi, Y., Suárez-Calvet, M., Czirr, E., Lohmann, E., Cuyvers, E., Struyfs, H., Pettkus, N., Wenninger-Weinzierl, A., Mazaheri, F., Tahirovic, S., Lleó, A., Alcolea, D., Fortea, J., Willem, M., Lammich, S., Molinuevo, J.L., Sánchez-Valle, R., Antonell, A., Ramirez, A., Heneka, M.T., Sleegers, K., Zee, J. van der, Martin, J.-J., Engelborghs, S., Demirtas-Tatlidede, A., Zetterberg, H., Broeckhoven, C.V., Gurvit, H., Wyss-Coray, T., Hardy, J., Colonna, M. and Haass, C. 2014. TREM2

mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Science Translational Medicine*. **6**(243),p.243ra86-243ra86.

Kon, T., Mori, F., Tanji, K., Miki, Y., Toyoshima, Y., Yoshida, M., Sasaki, H., Kakita, A., Takahashi, H. and Wakabayashi, K. 2014. ALS-associated protein FIG4 is localized in Pick and Lewy bodies, and also neuronal nuclear inclusions, in polyglutamine and intranuclear inclusion body diseases. *Neuropathology: Official Journal of the Japanese Society of Neuropathology*. **34**(1),pp.19–26.

Kong, J. and Xu, Z. 2000. Overexpression of neurofilament subunit NF-L and NF-H extends survival of a mouse model for amyotrophic lateral sclerosis. *Neuroscience Letters*. **281**(1),pp.72–74.

Koppers, M., van Blitterswijk, M.M., Vlam, L., Rowicka, P.A., van Vught, P.W.J., Groen, E.J.N., Spliet, W.G.M., Engelen-Lee, J., Schelhaas, H.J., de Visser, M., van der Kooi, A.J., van der Pol, W.-L., Pasterkamp, R.J., Veldink, J.H. and van den Berg, L.H. 2012. VCP mutations in familial and sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(4),p.837.e7-13.

Koutsis, G., Pemble, S., Sweeney, M.G., Paudel, R., Wood, N.W., Panas, M., Kladi, A. and Houlden, H. 2012. Analysis of spinocerebellar ataxias due to expanded triplet repeats in Greek patients with cerebellar ataxia. *Journal of the Neurological Sciences*. **318**(1–2),pp.178–180.

Kress, J.A., Kühnlein, P., Winter, P., Ludolph, A.C., Kassubek, J., Müller, U. and Sperfeld, A.-D. 2005. Novel mutation in the ALS2 gene in juvenile amyotrophic lateral sclerosis. *Annals of Neurology*. **58**(5),pp.800–803.

Kun-Rodrigues, C., Ganos, C., Guerreiro, R., Schneider, S.A., Schulte, C., Lesage, S., Darwent, L., Holmans, P., Singleton, A., (ipdgc), I.P.D.G.C., Bhatia, K. and Bras, J. 2015. A systematic screening to identify de novo mutations causing sporadic early-onset Parkinson's disease. *Human Molecular Genetics*. **24**(23),pp.6711–6720.

Kurzwelly, D., Krüger, S., Biskup, S. and Heneka, M.T. 2015. A distinct clinical phenotype in a German kindred with motor neuron disease carrying a CHCHD10 mutation. *Brain: A Journal of Neurology*. **138**(Pt 9),p.e376.

Kwiatkowski, T.J., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., Valdmanis, P., Rouleau, G.A., Hosler, B.A., Cortelli, P., de Jong, P.J., Yoshinaga, Y., Haines, J.L., Pericak-Vance, M.A., Yan, J., Ticozzi, N., Siddique, T., McKenna-Yasek, D., Sapp, P.C., Horvitz, H.R., Landers, J.E. and Brown, R.H. 2009. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science (New York, N.Y.)*. **323**(5918),pp.1205–1208.

Kwok, C.T., Morris, A. and de Belleroche, J.S. 2014. Sequestosome-1 (SQSTM1) sequence variants in ALS cases in the UK: prevalence and coexistence of SQSTM1 mutations in ALS kindred with PDB. *European journal of human genetics: EJHG*. **22**(4),pp.492–496.

Lagier-Tourenne, C. and Cleveland, D.W. 2009. Rethinking ALS: the FUS about TDP-43. *Cell*. **136**(6),pp.1001–1004.

Lai, S.-L., Abramzon, Y., Schymick, J.C., Stephan, D.A., Dunckley, T., Dillman, A., Cookson, M., Calvo, A., Battistini, S., Giannini, F., Caponnetto, C., Mancardi, G.L., Spataro, R.,

Monsurro, M.R., Tedeschi, G., Marinou, K., Sabatelli, M., Conte, A., Mandrioli, J., Sola, P., Salvi, F., Bartolomei, I., Lombardo, F., Mora, G., Restagno, G., Chiò, A. and Traynor, B.J. 2011. FUS mutations in sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging*. **32**(3),p.550.e1-550.e4.

Lambrechts, D., Storkebaum, E., Morimoto, M., Del-Favero, J., Desmet, F., Marklund, S.L., Wyns, S., Thijs, V., Andersson, J., van Marion, I., Al-Chalabi, A., Bornes, S., Musson, R., Hansen, V., Beckman, L., Adolfsson, R., Pall, H.S., Prats, H., Vermeire, S., Rutgeerts, P., Katayama, S., Awata, T., Leigh, N., Lang-Lazdunski, L., Dewerchin, M., Shaw, C., Moons, L., Vlietinck, R., Morrison, K.E., Robberecht, W., Van Broeckhoven, C., Collen, D., Andersen, P.M. and Carmeliet, P. 2003. VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nature Genetics*. **34**(4),pp.383–394.

Landers, J.E., Leclerc, A.L., Shi, L., Virkud, A., Cho, T., Maxwell, M.M., Henry, A.F., Polak, M., Glass, J.D., Kwiatkowski, T.J., Al-Chalabi, A., Shaw, C.E., Leigh, P.N., Rodriguez-Leyza, I., McKenna-Yasek, D., Sapp, P.C. and Brown, R.H. 2008a. New VAPB deletion variant and exclusion of VAPB mutations in familial ALS. *Neurology*. **70**(14),pp.1179–1185.

Landers, J.E., Shi, L., Cho, T.-J., Glass, J.D., Shaw, C.E., Leigh, P.N., Diekstra, F., Polak, M., Rodriguez-Leyva, I., Niemann, S., Traynor, B.J., McKenna-Yasek, D., Sapp, P.C., Al-Chalabi, A., Wills, A.-M.A. and Brown, R.H. 2008b. A common haplotype within the PON1 promoter region is associated with sporadic ALS. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **9**(5),pp.306–314.

Larivière, R.C., Beaulieu, J.-M., Nguyen, M.D. and Julien, J.-P. 2003. Peripherin is not a contributing factor to motor neuron disease in a mouse model of amyotrophic lateral sclerosis caused by mutant superoxide dismutase. *Neurobiology of Disease*. **13**(2),pp.158–166.

Lattante, S., Conte, A., Zollino, M., Luigetti, M., Del Grande, A., Marangi, G., Romano, A., Marcaccio, A., Meleo, E., Bisogni, G., Rossini, P.M. and Sabatelli, M. 2012. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. *Neurology*. **79**(1),pp.66–72.

Lattante, S., Le Ber, I., Camuzat, A., Brice, A. and Kabashi, E. 2013a. Mutations in the PFN1 gene are not a common cause in patients with amyotrophic lateral sclerosis and frontotemporal lobar degeneration in France. *Neurobiology of Aging*. **34**(6),p.1709.e1-2.

Lattante, S., Millecamps, S., Stevanin, G., Rivaud-Péchoux, S., Moigneu, C., Camuzat, A., Da Barroca, S., Mundwiller, E., Couarch, P., Salachas, F., Hannequin, D., Meininger, V., Pasquier, F., Seilhean, D., Couratier, P., Danel-Brunaud, V., Bonnet, A.-M., Tranchant, C., LeGuern, E., Brice, A., Le Ber, I., Kabashi, E. and French Research Network on FTD and FTD-ALS 2014. Contribution of ATXN2 intermediary polyQ expansions in a spectrum of neurodegenerative disorders. *Neurology*. **83**(11),pp.990–995.

Lattante, S., Rouleau, G.A. and Kabashi, E. 2013b. TARDBP and FUS mutations associated with amyotrophic lateral sclerosis: summary and update. *Human Mutation*. **34**(6),pp.812–826.

Laurin, N., Brown, J.P., Morissette, J. and Raymond, V. 2002. Recurrent mutation of the gene encoding sequestosome 1 (SQSTM1/p62) in Paget disease of bone. *American Journal of Human Genetics*. **70**(6),pp.1582–1588.

Le Ber, I., Camuzat, A., Guerreiro, R., Bouya-Ahmed, K., Bras, J., Nicolas, G., Gabelle, A., Didic, M., De Septenville, A., Millecamps, S., Lenglet, T., Latouche, M., Kabashi, E., Campion, D., Hannequin, D., Hardy, J., Brice, A. and French Clinical and Genetic Research Network on FTD/FTD-ALS 2013. SQSTM1 mutations in French patients with frontotemporal dementia or frontotemporal dementia with amyotrophic lateral sclerosis. *JAMA neurology*. **70**(11),pp.1403–1410.

Leblond, C.S., Gan-Or, Z., Spiegelman, D., Laurent, S.B., Szuto, A., Hodgkinson, A., Dionne-Laporte, A., Provencher, P., de Carvalho, M., Orrù, S., Brunet, D., Bouchard, J.-P., Awadalla, P., Dupré, N., Dion, P.A. and Rouleau, G.A. 2016. Replication study of MATR3 in familial and sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging*. **37**,p.209.e17-21.

Leblond, C.S., Kaneb, H.M., Dion, P.A. and Rouleau, G.A. 2014. Dissection of genetic factors associated with amyotrophic lateral sclerosis. *Experimental Neurology*. **262 Pt B**,pp.91–101.

Lee, J.-A., Beigneux, A., Ahmad, S.T., Young, S.G. and Gao, F.-B. 2007. ESCRT-III dysfunction causes autophagosome accumulation and neurodegeneration. *Current biology: CB*. **17**(18),pp.1561–1567.

Lee, T., Li, Y.R., Ingre, C., Weber, M., Grehl, T., Gredal, O., de Carvalho, M., Meyer, T., Tysnes, O.-B., Auburger, G., Gispert, S., Bonini, N.M., Andersen, P.M. and Gitler, A.D. 2011. Ataxin-2 intermediate-length polyglutamine expansions in European ALS patients. *Human Molecular Genetics*. **20**(9),pp.1697–1700.

Lee, Y.H., Kim, J.-H., Seo, Y.H., Choi, S.J., Ji, J.D. and Song, G.G. 2015. Paraoxonase 1 Q192R and L55M polymorphisms and susceptibility to amyotrophic lateral sclerosis: a meta-analysis. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*. **36**(1),pp.11–20.

Lenk, G.M., Ferguson, C.J., Chow, C.Y., Jin, N., Jones, J.M., Grant, A.E., Zolov, S.N., Winters, J.J., Giger, R.J., Dowling, J.J., Weisman, L.S. and Meisler, M.H. 2011. Pathogenic mechanism of the FIG4 mutation responsible for Charcot-Marie-Tooth disease CMT4J. *PLoS genetics*. **7**(6),p.e1002104.

Lesca, G., Eymard-Pierre, E., Santorelli, F.M., Cusmai, R., Di Capua, M., Valente, E.M., Attia-Sobol, J., Plauchu, H., Leuzzi, V., Ponzone, A., Boespflug-Tanguy, O. and Bertini, E. 2003. Infantile ascending hereditary spastic paralysis (IAHSP): clinical features in 11 families. *Neurology*. **60**(4),pp.674–682.

Leung, C.L., He, C.Z., Kaufmann, P., Chin, S.S., Naini, A., Liem, R.K.H., Mitsumoto, H. and Hays, A.P. 2004. A Pathogenic Peripherin Gene Mutation in a Patient with Amyotrophic Lateral Sclerosis. *Brain Pathology*. **14**(3),pp.290–296.

Li, X.L., Shu, S., Li, X.G., Liu, Q., Liu, F., Cui, B., Liu, M.S., Peng, B., Cui, L.Y. and Zhang, X. 2016. CHCHD10 is not a frequent causative gene in Chinese ALS patients. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*.,pp.1–3.

Lill, C.M., Rengmark, A., Pihlstrøm, L., Fogh, I., Shatunov, A., Sleiman, P.M., Wang, L.-S., Liu, T., Lassen, C.F., Meissner, E., Alexopoulos, P., Calvo, A., Chio, A., Dizdar, N.,

Faltraco, F., Forsgren, L., Kirchheiner, J., Kurz, A., Larsen, J.P., Liebsch, M., Linder, J., Morrison, K.E., Nissbrandt, H., Otto, M., Pahnke, J., Partch, A., Restagno, G., Rujescu, D., Schnack, C., Shaw, C.E., Shaw, P.J., Tumani, H., Tysnes, O.-B., Valladares, O., Silani, V., van den Berg, L.H., van Rheenen, W., Veldink, J.H., Lindenberger, U., Steinhagen-Thiessen, E., SLAGEN Consortium, Teipel, S., Perneczky, R., Hakonarson, H., Hampel, H., von Arnim, C.A.F., Olsen, J.H., Van Deerlin, V.M., Al-Chalabi, A., Toft, M., Ritz, B. and Bertram, L. 2015. The role of TREM2 R47H as a risk factor for Alzheimer's disease, frontotemporal lobar degeneration, amyotrophic lateral sclerosis, and Parkinson's disease. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*. **11**(12),pp.1407–1416.

Lin, K.-P., Tsai, P.-C., Liao, Y.-C., Chen, W.-T., Tsai, C.-P., Soong, B.-W. and Lee, Y.-C. 2015. Mutational analysis of MATR3 in Taiwanese patients with amyotrophic lateral sclerosis. *Neurobiology of Aging*. **36**(5),p.2005.e1-4.

Lindquist, S.G., Braedgaard, H., Svenstrup, K., Isaacs, A.M., Nielsen, J.E. and FReJA Consortium 2008. Frontotemporal dementia linked to chromosome 3 (FTD-3)--current concepts and the detection of a previously unknown branch of the Danish FTD-3 family. *European Journal of Neurology*. **15**(7),pp.667–670.

Liu, X., Lu, M., Tang, L., Zhang, N., Chui, D. and Fan, D. 2013. ATXN2 CAG repeat expansions increase the risk for Chinese patients with amyotrophic lateral sclerosis. *Neurobiology of Aging*. **34**(9),p.2236.e5-2236.e8.

Logroscino, G., Traynor, B.J., Hardiman, O., Chiò, A., Mitchell, D., Swingler, R.J., Millul, A., Benn, E., Beghi, E. and EURALS 2010. Incidence of amyotrophic lateral sclerosis in Europe. *Journal of Neurology, Neurosurgery, and Psychiatry*. **81**(4),pp.385–390.

Lopate, G., Baloh, R.H., Al-Lozi, M.T., Miller, T.M., Fernandes Filho, J.A., Ni, O., Leston, A., Florence, J., Schierbecker, J. and Allred, P. 2010. Familial ALS with extreme phenotypic variability due to the I113T SOD1 mutation. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **11**(1–2),pp.232–236.

Lorenzetti, D., Bohlega, S. and Zoghbi, H.Y. 1997. The expansion of the CAG repeat in ataxin-2 is a frequent cause of autosomal dominant spinocerebellar ataxia. *Neurology*. **49**(4),pp.1009–1013.

Luigetti, M., Conte, A., Madia, F., Marangi, G., Zollino, M., Mancuso, I., Dileone, M., Grande, A.D., Lazzaro, V.D., Tonali, P.A. and Sabatelli, M. 2009. Heterozygous SOD1 D90A mutation presenting as slowly progressive predominant upper motor neuron amyotrophic lateral sclerosis. *Neurological Sciences*. **30**(6),pp.517–520.

Luigetti, M., Lattante, S., Conte, A., Romano, A., Zollino, M., Marangi, G. and Sabatelli, M. 2013. A novel compound heterozygous ALS2 mutation in two Italian siblings with juvenile amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*. **14**(5–6),pp.470–472.

Luisa Conforti, F., Sprovieri, T., Mazzei, R., Patitucci, A., Ungaro, C., Zoccolella, S., Magariello, A., Bella, V.L., Tessitore, A., Tedeschi, G., Simone, I.L., Majorana, G., Valentino, P., Citrigno, L., Gabriele, A., Bono, F., Monsurrò, M.R., Muglia, M. and Quattrone, A. 2009. Further evidence that D90A-SOD1 mutation is recessively inherited in ALS patients in Italy. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **10**(1),pp.58–60.

Maaten, L. van der and Hinton, G. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*. **9**(Nov),pp.2579–2605.

Mackenzie, I.R.A., Bigio, E.H., Ince, P.G., Geser, F., Neumann, M., Cairns, N.J., Kwong, L.K., Forman, M.S., Ravits, J., Stewart, H., Eisen, A., McClusky, L., Kretzschmar, H.A., Monoranu, C.M., Highley, J.R., Kirby, J., Siddique, T., Shaw, P.J., Lee, V.M.-Y. and Trojanowski, J.Q. 2007. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Annals of Neurology*. **61**(5),pp.427–434.

Majounie, E., Traynor, B.J., Chiò, A., Restagno, G., Mandrioli, J., Benatar, M., Taylor, J.P. and Singleton, A.B. 2012. Mutational analysis of the VCP gene in Parkinson's disease. *Neurobiology of Aging*. **33**(1),p.209.e1-2.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A. and Visscher, P.M. 2009. Finding the missing heritability of complex diseases. *Nature*. **461**(7265),pp.747–753.

Martherus, R.S.R.M., Sluiter, W., Timmer, E.D.J., VanHerle, S.J.V., Smeets, H.J.M. and Ayoubi, T.A.Y. 2010. Functional annotation of heart enriched mitochondrial genes GBAS and CHCHD10 through guilt by association. *Biochemical and Biophysical Research Communications*. **402**(2),pp.203–208.

Martinez-Vicente, M. and Cuervo, A.M. 2007. Autophagy and neurodegeneration: when the cleaning crew goes on strike. *The Lancet. Neurology*. **6**(4),pp.352–361.

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H., Komure, O., Matsuura, S., Kobatake, K., Morimoto, N., Abe, K., Suzuki, N., Aoki, M., Kawata, A., Hirai, T., Kato, T., Ogasawara, K., Hirano, A., Takumi, T., Kusaka, H., Hagiwara, K., Kaji, R. and Kawakami, H. 2010. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature*. **465**(7295),pp.223–226.

Maximino, J.R., de Oliveira, G.P., Alves, C.J. and Chadi, G. 2014. Deregulated expression of cytoskeleton related genes in the spinal cord and sciatic nerve of presymptomatic SOD1(G93A) Amyotrophic Lateral Sclerosis mouse model. *Frontiers in Cellular Neuroscience*. **8**,p.148.

McGuire, V., Longstreth, W.T., Nelson, L.M., Koepsell, T.D., Checkoway, H., Morgan, M.S. and van Belle, G. 1997. Occupational exposures and amyotrophic lateral sclerosis. A population-based case-control study. *American Journal of Epidemiology*. **145**(12),pp.1076–1088.

McLaughlin, R.L., Kenna, K.P., Vajda, A., Byrne, S., Bradley, D.G. and Hardiman, O. 2014. UBQLN2 mutations are not a frequent cause of amyotrophic lateral sclerosis in Ireland. *Neurobiology of Aging*. **35**(1),p.267.e9-11.

Melki, R., Belmouden, A., Akhayat, O., Brézin, A. and Garchon, H.-J. 2003. The M98K variant of the OPTINEURIN (OPTN) gene modifies initial intraocular pressure in patients with primary open angle glaucoma. *Journal of Medical Genetics*. **40**(11),pp.842–844.

Mendonça, D.M.F., Martins, S.C.S., Higashi, R., Muscara, M.N., Neto, V.M., Chimelli, L. and Martinez, A.M.B. 2011. Neurofilament heavy subunit in cerebrospinal fluid: a biomarker of amyotrophic lateral sclerosis? *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **12**(2),pp.144–147.

Merner, N.D., Girard, S.L., Catoire, H., Bourassa, C.V., Belzil, V.V., Rivière, J.-B., Hince, P., Levert, A., Dionne-Laporte, A., Spiegelman, D., Noreau, A., Diab, S., Szuto, A., Fournier, H., Raelson, J., Belouchi, M., Panisset, M., Cossette, P., Dupré, N., Bernard, G., Chouinard, S., Dion, P.A. and Rouleau, G.A. 2012. Exome sequencing identifies FUS mutations as a cause of essential tremor. *American Journal of Human Genetics*. **91**(2),pp.313–319.

Metzker, M.L. 2010. Sequencing technologies - the next generation. *Nature Reviews. Genetics*. **11**(1),pp.31–46.

Millecamps, S., Corcia, P., Cazeneuve, C., Boillée, S., Seilhean, D., Danel-Brunaud, V., Vandenberghe, N., Pradat, P.-F., Le Forestier, N., Lacomblez, L., Bruneteau, G., Camu, W., Brice, A., Meininger, V., LeGuern, E. and Salachas, F. 2012. Mutations in UBQLN2 are rare in French amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(4),p.839.e1-3.

Millecamps, S., Da Barroca, S., Cazeneuve, C., Salachas, F., Pradat, P.-F., Danel-Brunaud, V., Vandenberghe, N., Lacomblez, L., Le Forestier, N., Bruneteau, G., Camu, W., Brice, A., Meininger, V. and LeGuern, E. 2010a. Questioning on the role of D amino acid oxidase in familial amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(26),p.E107; author reply E108.

Millecamps, S., De Septenville, A., Teyssou, E., Daniau, M., Camuzat, A., Albert, M., LeGuern, E., Galimberti, D., French research network on FTD and FTD-ALS, Brice, A., Marie, Y. and Le Ber, I. 2014. Genetic analysis of matrin 3 gene in French amyotrophic lateral sclerosis patients and frontotemporal lobar degeneration with amyotrophic lateral sclerosis patients. *Neurobiology of Aging*. **35**(12),p.2882.e13-15.

Millecamps, S. and Julien, J.-P. 2013. Axonal transport deficits and neurodegenerative diseases. *Nature Reviews. Neuroscience*. **14**(3),pp.161–176.

Millecamps, S., Salachas, F., Cazeneuve, C., Gordon, P., Bricka, B., Camuzat, A., Guillot-Noël, L., Russaouen, O., Bruneteau, G., Pradat, P.-F., Le Forestier, N., Vandenberghe, N., Danel-Brunaud, V., Guy, N., Thauvin-Robinet, C., Lacomblez, L., Couratier, P., Hannequin, D., Seilhean, D., Le Ber, I., Corcia, P., Camu, W., Brice, A., Rouleau, G., LeGuern, E. and Meininger, V. 2010b. SOD1, ANG, VAPB, TARDBP, and FUS mutations in familial amyotrophic lateral sclerosis: genotype-phenotype correlations. *Journal of Medical Genetics*. **47**(8),pp.554–560.

Miller, J.W., Smith, B.N., Topp, S.D., Al-Chalabi, A., Shaw, C.E. and Vance, C. 2012. Mutation analysis of VCP in British familial and sporadic amyotrophic lateral sclerosis patients. *Neurobiology of Aging*. **33**(11),p.2721.e1-2.

Mintchev, N., Zamba-Papanicolaou, E., Kleopa, K.A. and Christodoulou, K. 2009. A novel ALS2 splice-site mutation in a Cypriot juvenile-onset primary lateral sclerosis family. *Neurology*. **72**(1),pp.28–32.

Mitchell, J., Paul, P., Chen, H.-J., Morris, A., Payling, M., Falchi, M., Habgood, J., Panoutsou, S., Winkler, S., Tisato, V., Hajitou, A., Smith, B., Vance, C., Shaw, C., Mazarakis, N.D. and de Belleroche, J. 2010. Familial amyotrophic lateral sclerosis is associated with a mutation in D-amino acid oxidase. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(16),pp.7556–7561.

Mitchell, J.C., McGoldrick, P., Vance, C., Hortobagyi, T., Sreedharan, J., Rogelj, B., Tudor, E.L., Smith, B.N., Klasen, C., Miller, C.C.J., Cooper, J.D., Greensmith, L. and Shaw, C.E. 2013. Overexpression of human wild-type FUS causes progressive motor neuron degeneration in an age- and dose-dependent fashion. *Acta Neuropathologica*. **125**(2),pp.273–288.

Mok, K.Y., Koutsis, G., Schottlaender, L.V., Polke, J., Panas, M. and Houlden, H. 2012. High frequency of the expanded C9ORF72 hexanucleotide repeat in familial and sporadic Greek ALS patients. *Neurobiology of Aging*. **33**(8),p.1851.e1-5.

Morahan, J.M. and Pamphlett, R. 2006. Amyotrophic lateral sclerosis and exposure to environmental toxins: an Australian case-control study. *Neuroepidemiology*. **27**(3),pp.130–135.

Morahan, J.M., Yu, B., Trent, R.J. and Pamphlett, R. 2007. A gene-environment study of the paraoxonase 1 gene and pesticides in amyotrophic lateral sclerosis. *Neurotoxicology*. **28**(3),pp.532–540.

Moreira, M.-C., Klur, S., Watanabe, M., Németh, A.H., Ber, I.L., Moniz, J.-C., Tranchant, C., Aubourg, P., Tazir, M., Schöls, L., Pandolfo, M., Schulz, J.B., Pouget, J., Calvas, P., Shizuka-Ikeda, M., Shoji, M., Tanaka, M., Izatt, L., Shaw, C.E., M'Zahem, A., Dunne, E., Bomont, P., Benhassine, T., Bouslam, N., Stevanin, G., Brice, A., Guimarães, J., Mendonça, P., Barbot, C., Coutinho, P., Sequeiros, J., Dürr, A., Warter, J.-M. and Koenig, M. 2004. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature Genetics*. **36**(3),pp.225–227.

Morgan, S., Shoai, M., Fratta, P., Sidle, K., Orrell, R., Sweeney, M.G., Shatunov, A., Sproviero, W., Jones, A., Al-Chalabi, A., Malaspina, A., Houlden, H., Hardy, J. and Pittman, A. 2015. Investigation of next-generation sequencing technologies as a diagnostic tool for amyotrophic lateral sclerosis. *Neurobiology of Aging*. **36**(3),p.1600.e5-8.

Müller, K., Andersen, P.M., Hübers, A., Marroquin, N., Volk, A.E., Danzer, K.M., Meitinger, T., Ludolph, A.C., Strom, T.M. and Weishaupt, J.H. 2014a. Two novel mutations in conserved codons indicate that CHCHD10 is a gene associated with motor neuron disease. *Brain: A Journal of Neurology*. **137**(Pt 12),p.e309.

Müller, T.D., Greene, B.H., Bellodi, L., Cavallini, M.C., Cellini, E., Di Bella, D., Ehrlich, S., Erzegovesi, S., Estivill, X., Fernández-Aranda, F., Fichter, M., Fleischhaker, C., Scherag, S., Gratacòs, M., Grallert, H., Herpertz-Dahlmann, B., Herzog, W., Illig, T., Lehmkuhl, U., Nacmias, B., Ribasés, M., Ricca, V., Schäfer, H., Scherag, A., Sorbi, S., Wichmann, H.-E., Hebebrand, J. and Hinney, A. 2012. Fat mass and obesity-associated gene (FTO) in eating disorders: evidence for association of the rs9939609 obesity risk allele with bulimia nervosa and anorexia nervosa. *Obesity Facts*. **5**(3),pp.408–419.

Müller, T.J., Kraya, T., Stoltenburg-Didinger, G., Hanisch, F., Kornhuber, M., Stoevesandt, D., Senderek, J., Weis, J., Baum, P., Deschauer, M. and Zierz, S. 2014b. Phenotype

of matrin-3-related distal myopathy in 16 German patients. *Annals of Neurology*. **76**(5),pp.669–680.

Münch, C., Rosenbohm, A., Sperfeld, A.-D., Uttner, I., Reske, S., Krause, B.J., Sedlmeier, R., Meyer, T., Hanemann, C.O., Stumm, G. and Ludolph, A.C. 2005. Heterozygous R1101K mutation of the DCTN1 gene in a family with ALS and FTD. *Annals of Neurology*. **58**(5),pp.777–780.

Münch, C., Sedlmeier, R., Meyer, T., Homberg, V., Sperfeld, A.D., Kurt, A., Prudlo, J., Peraus, G., Hanemann, C.O., Stumm, G. and Ludolph, A.C. 2004. Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. *Neurology*. **63**(4),pp.724–726.

Murakami, T., Warita, H., Hayashi, T., Sato, K., Manabe, Y., Mizuno, S., Yamane, K. and Abe, K. 2001. A novel SOD1 gene mutation in familial ALS with low penetrance in females. *Journal of the Neurological Sciences*. **189**(1–2),pp.45–47.

Murmu, R.P., Martin, E., Rastetter, A., Esteves, T., Muriel, M.-P., El Hachimi, K.H., Denora, P.S., Dauphin, A., Fernandez, J.C., Duyckaerts, C., Brice, A., Darios, F. and Stevanin, G. 2011. Cellular distribution and subcellular localization of spatacsin and spastizin, two proteins involved in hereditary spastic paraplegia. *Molecular and Cellular Neurosciences*. **47**(3),pp.191–202.

Nakamura, R., Sone, J., Atsuta, N., Tohnai, G., Watanabe, H., Yokoi, D., Nakatochi, M., Watanabe, H., Ito, M., Senda, J., Katsuno, M., Tanaka, F., Li, Y., Izumi, Y., Morita, M., Taniguchi, A., Kano, O., Oda, M., Kuwabara, S., Abe, K., Aiba, I., Okamoto, K., Mizoguchi, K., Hasegawa, K., Aoki, M., Hattori, N., Tsuji, S., Nakashima, K., Kaji, R. and Sobue, G. 2016. Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. *Neurobiology of Aging*. **39**,p.219.e1-219.e8.

Nanetti, L., Fancellu, R., Tomasello, C., Gellera, C., Pareyson, D. and Mariotti, C. 2009. Rare association of motor neuron disease and spinocerebellar ataxia type 2 (SCA2): a new case and review of the literature. *Journal of Neurology*. **256**(11),pp.1926–1928.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., McCluskey, L.F., Miller, B.L., Masliah, E., Mackenzie, I.R., Feldman, H., Feiden, W., Kretzschmar, H.A., Trojanowski, J.Q. and Lee, V.M.-Y. 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (New York, N.Y.)*. **314**(5796),pp.130–133.

Newsway, V., Fish, M., Rohrer, J.D., Majounie, E., Williams, N., Hack, M., Warren, J.D. and Morris, H.R. 2010. Perry syndrome due to the DCTN1 G71R mutation: a distinctive levodopa responsive disorder with behavioral syndrome, vertical gaze palsy, and respiratory failure. *Movement Disorders: Official Journal of the Movement Disorder Society*. **25**(6),pp.767–770.

Nguyen, M.D., Larivière, R.C. and Julien, J.P. 2001. Deregulation of Cdk5 in a mouse model of ALS: toxicity alleviated by perikaryal neurofilament inclusions. *Neuron*. **30**(1),pp.135–147.

Nishimura, A.L., Al-Chalabi, A. and Zatz, M. 2005. A common founder for amyotrophic lateral sclerosis type 8 (ALS8) in the Brazilian population. *Human Genetics*. **118**(3–4),pp.499–500.

Nishimura, A.L., Mitne-Neto, M., Silva, H.C.A., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J.R.M., Gillingwater, T., Webb, J., Skehel, P. and Zatz, M. 2004. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *American Journal of Human Genetics*. **75**(5),pp.822–831.

Ogaki, K., Li, Y., Atsuta, N., Tomiyama, H., Funayama, M., Watanabe, H., Nakamura, R., Yoshino, H., Yato, S., Tamura, A., Naito, Y., Taniguchi, A., Fujita, K., Izumi, Y., Kaji, R., Hattori, N., Sobue, G. and Japanese Consortium for Amyotrophic Lateral Sclerosis research (JaCALS) 2012. Analysis of C9orf72 repeat expansion in 563 Japanese patients with amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(10),p.2527.e11-16.

Oosthuyse, B., Moons, L., Storkebaum, E., Beck, H., Nuyens, D., Brusselmans, K., Van Dorpe, J., Hellings, P., Gorselink, M., Heymans, S., Theilmeier, G., Dewerchin, M., Laudenbach, V., Vermylen, P., Raat, H., Acker, T., Vleminckx, V., Van Den Bosch, L., Cashman, N., Fujisawa, H., Drost, M.R., Sciot, R., Bruyninckx, F., Hicklin, D.J., Ince, C., Gressens, P., Lupu, F., Plate, K.H., Robberecht, W., Herbert, J.M., Collen, D. and Carmeliet, P. 2001. Deletion of the hypoxia-response element in the vascular endothelial growth factor promoter causes motor neuron degeneration. *Nature Genetics*. **28**(2),pp.131–138.

Origone, P., Verdiani, S., Bandettini Di Poggio, M., Zuccarino, R., Vignolo, M., Caponnetto, C. and Mandich, P. 2015. A novel Arg147Trp MATR3 missense mutation in a slowly progressive ALS Italian patient. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*. **16**(7–8),pp.530–531.

Özoğuz, A., Uyan, Ö., Birdal, G., Iskender, C., Kartal, E., Lahut, S., Ömür, Ö., Agim, Z.S., Eken, A.G., Sen, N.E., Kavak, P., Saygı, C., Sapp, P.C., Keagle, P., Parman, Y., Tan, E., Koç, F., Deymeer, F., Oflazer, P., Hanağası, H., Gürvit, H., Bilgiç, B., Durmuş, H., Ertaş, M., Kotan, D., Akalın, M.A., Güllüoğlu, H., Zarifoğlu, M., Aysal, F., Döşoğlu, N., Bilguvar, K., Günel, M., Keskin, Ö., Akgün, T., Özçelik, H., Landers, J.E., Brown, R.H. and Başak, A.N. 2015. The distinct genetic pattern of ALS in Turkey and novel mutations. *Neurobiology of Aging*. **36**(4),p.1764.e9-18.

Painter, M.M., Atagi, Y., Liu, C.-C., Rademakers, R., Xu, H., Fryer, J.D. and Bu, G. 2015. TREM2 in CNS homeostasis and neurodegenerative disease. *Molecular Neurodegeneration*. **10**,p.43.

Paisan-Ruiz, C., Nath, P., Wood, N.W., Singleton, A. and Houlden, H. 2008. Clinical heterogeneity and genotype-phenotype correlations in hereditary spastic paraplegia because of Spatacsin mutations (SPG11). *European Journal of Neurology*. **15**(10),pp.1065–1070.

Palmio, J., Evilä, A., Bashir, A., Norwood, F., Viitaniemi, K., Vihola, A., Huovinen, S., Straub, V., Hackman, P., Hirano, M., Bushby, K. and Udd, B. 2016. Re-evaluation of the phenotype caused by the common MATR3 p.Ser85Cys mutation in a new family. *Journal of Neurology, Neurosurgery, and Psychiatry*. **87**(4),pp.448–450.

Panzeri, C., Palma, C.D., Martinuzzi, A., Daga, A., Polo, G.D., Bresolin, N., Miller, C.C., Tudor, E.L., Clementi, E. and Bassi, M.T. 2006. The first ALS2 missense mutation associated with JPLS reveals new aspects of alsin biological function. *Brain*. **129**(7),pp.1710–1719.

Parkinson, N., Ince, P.G., Smith, M.O., Highley, R., Skibinski, G., Andersen, P.M., Morrison, K.E., Pall, H.S., Hardiman, O., Collinge, J., Shaw, P.J., Fisher, E.M.C., MRC Proteomics in ALS Study and FReJA Consortium 2006. ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). *Neurology*. **67**(6),pp.1074–1077.

Paubel, A., Violette, J., Amy, M., Praline, J., Meininger, V., Camu, W., Corcia, P., Andres, C.R., Vourc'h, P. and French Amyotrophic Lateral Sclerosis (ALS) Study Group 2008. Mutations of the ANG gene in French patients with sporadic amyotrophic lateral sclerosis. *Archives of Neurology*. **65**(10),pp.1333–1336.

Paul, P. and de Belleroche, J. 2012. The role of D-amino acids in amyotrophic lateral sclerosis pathogenesis: a review. *Amino Acids*. **43**(5),pp.1823–1831.

Penttilä, S., Jokela, M., Bouquin, H., Saukkonen, A.M., Toivanen, J. and Udd, B. 2015. Late onset spinal motor neuronopathy is caused by mutation in CHCHD10. *Annals of Neurology*. **77**(1),pp.163–172.

Pesiridis, G.S., Lee, V.M.-Y. and Trojanowski, J.Q. 2009. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Human Molecular Genetics*. **18**(R2),pp.R156–R162.

Pottier, C., Wallon, D., Rousseau, S., Rovelet-Lecrux, A., Richard, A.-C., Rollin-Sillaire, A., Frebourg, T., Campion, D. and Hannequin, D. 2013. TREM2 R47H variant as a risk factor for early-onset Alzheimer's disease. *Journal of Alzheimer's disease: JAD*. **35**(1),pp.45–49.

Puls, I., Jonnakuty, C., LaMonte, B.H., Holzbaur, E.L.F., Tokito, M., Mann, E., Floeter, M.K., Bidus, K., Drayna, D., Oh, S.J., Brown, R.H., Ludlow, C.L. and Fischbeck, K.H. 2003. Mutant dynactin in motor neuron disease. *Nature Genetics*. **33**(4),pp.455–456.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. and Sham, P.C. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. **81**(3),pp.559–575.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., (Leader), S.M.P., Ruderfer, D.M., McQuillin, A., Morris, D.W., O'Dushlaine, C.T., Corvin, A., Holmans, P.A., O'Donovan, M.C., Macgregor, S., Gurling, H., Blackwood, D.H.R., Craddock, N.J., Gill, M., Hultman, C.M., Kirov, G.K., Lichtenstein, P., Muir, W.J., Owen, M.J., Pato, C.N., Scolnick, E.M., Clair, D.S., (Leader), P.S., Williams, N.M., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Toncheva, D., Milanova, V., Thelander, E.F., Sullivan, P., O'Dushlaine, C.T., Kenny, E., Quinn, E.M., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Crombie, C., Fraser, G., Kuan, S.L., Walker, N., McGhee, K.A., Pickard, B., Malloy, P., Maclean, A.W., Beck, M.V., Pato, M.T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J.A., Ferreira, C.P., Macedo, A., Azevedo, M.H., Kirby, A.N., Ferreira, M.A.R., Daly, M.J., Chambert, K., Kuruvilla, F., Gabriel, S.B., Ardlie, K. and Moran, J.L. 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. **460**(7256),pp.748–752.

Rademakers, R., Stewart, H., Dejesus-Hernandez, M., Krieger, C., Graff-Radford, N., Fabros, M., Briemberg, H., Cashman, N., Eisen, A. and Mackenzie, I.R.A. 2010. Fus

gene mutations in familial and sporadic amyotrophic lateral sclerosis. *Muscle & Nerve*. **42**(2),pp.170–176.

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T., UK Brain Expression Consortium, North American Brain Expression Consortium, Coin, L., de Silva, R., Cookson, M.R., Singleton, A.B., Hardy, J., Ryten, M. and Weale, M.E. 2014. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neuroscience*. **17**(10),pp.1418–1428.

Ratti, A., Corrado, L., Castellotti, B., Del Bo, R., Fogh, I., Cereda, C., Tiloca, C., D'Ascenzo, C., Bagarotti, A., Pensato, V., Ranieri, M., Gagliardi, S., Calini, D., Mazzini, L., Taroni, F., Corti, S., Ceroni, M., Oggioni, G.D., Lin, K., Powell, J.F., Sorarù, G., Ticozzi, N., Comi, G.P., D'Alfonso, S., Gellera, C., Silani, V. and SLAGEN Consortium 2012. C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. *Neurobiology of Aging*. **33**(10),p.2528.e7-14.

Rayaprolu, S., Mullen, B., Baker, M., Lynch, T., Finger, E., Seeley, W.W., Hatanpaa, K.J., Lomen-Hoerth, C., Kertesz, A., Bigio, E.H., Lippa, C., Josephs, K.A., Knopman, D.S., White, C.L., Caselli, R., Mackenzie, I.R., Miller, B.L., Boczarska-Jedynak, M., Opala, G., Krygowska-Wajs, A., Barcikowska, M., Younkin, S.G., Petersen, R.C., Ertekin-Taner, N., Uitti, R.J., Meschia, J.F., Boylan, K.B., Boeve, B.F., Graff-Radford, N.R., Wszolek, Z.K., Dickson, D.W., Rademakers, R. and Ross, O.A. 2013. TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease. *Molecular Neurodegeneration*. **8**,p.19.

Reid, E., Connell, J., Edwards, T.L., Duley, S., Brown, S.E. and Sanderson, C.M. 2005. The hereditary spastic paraplegia protein spastin interacts with the ESCRT-III complex-associated endosomal protein CHMP1B. *Human Molecular Genetics*. **14**(1),pp.19–38.

Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A.M., Kaganovich, A., Scholz, S.W., Duckworth, J., Ding, J., Harmer, D.W., Hernandez, D.G., Johnson, J.O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R.J., Orrell, R.W., Neal, J., Murray, A., Pearson, J., Jansen, I.E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J.B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M.A., Peuralinna, T., Jansson, L., Isoviita, V.-M., Kaivorinne, A.-L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., ITALSGEN Consortium, Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J.D., Sendtner, M., Drepper, C., Eichler, E.E., Alkan, C., Abdullaev, Z., Pack, S.D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N.M., Heutink, P., Pickering-Brown, S., Morris, H.R., Tienari, P.J. and Traynor, B.J. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. **72**(2),pp.257–268.

Rezaie, T., Child, A., Hitchings, R., Brice, G., Miller, L., Coca-Prados, M., Héon, E., Krupin, T., Ritch, R., Kreutzer, D., Crick, R.P. and Sarfarazi, M. 2002. Adult-Onset Primary Open-Angle Glaucoma Caused by Mutations in Optineurin. *Science*. **295**(5557),pp.1077–1079.

Ricci, C., Battistini, S., Cozzi, L., Benigni, M., Origone, P., Verriello, L., Lunetta, C., Cereda, C., Milani, P., Greco, G., Patrosso, M.C., Causarano, R., Caponnetto, C., Giannini,

F., Corbo, M. and Penco, S. 2011. Lack of association of PON polymorphisms with sporadic ALS in an Italian population. *Neurobiology of Aging*. **32**(3),p.552.e7-13.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L. and ACMG Laboratory Quality Assurance Committee 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*. **17**(5),pp.405–424.

Risch, N. and Merikangas, K. 1996. The future of genetic studies of complex human diseases. *Science (New York, N.Y.)*. **273**(5281),pp.1516–1517.

Rizzu, P., van Mil, S.E., Anar, B., Rosso, S.M., Donker Kaat, L., Heutink, P. and van Swieten, J.C. 2006. CHMP2B mutations are not a cause of dementia in Dutch patients with familial and sporadic frontotemporal dementia. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*. **141B**(8),pp.944–946.

Robberecht, W., Aguirre, T., Bosch, L.V.D., Tilkin, P., Cassiman, J.J. and Matthijs, G. 1996. D90A heterozygosity in the SOD1 gene is associated with familial and apparently sporadic amyotrophic lateral sclerosis. *Neurology*. **47**(5),pp.1336–1339.

Rohrer, J.D., Isaacs, A.M., Mizielinska, S., Mead, S., Lashley, T., Wray, S., Sidle, K., Fratta, P., Orrell, R.W., Hardy, J., Holton, J., Revesz, T., Rossor, M.N. and Warren, J.D. 2015. C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. *The Lancet. Neurology*. **14**(3),pp.291–301.

Rohrer, J.D., Warren, J.D., Reiman, D., Uphill, J., Beck, J., Collinge, J., Rossor, M.N., Isaacs, A.M. and Mead, S. 2011. A novel exon 2 I27V VCP variant is associated with dissimilar clinical syndromes. *Journal of Neurology*. **258**(8),pp.1494–1496.

Ronchi, D., Riboldi, G., Del Bo, R., Ticozzi, N., Scarlato, M., Galimberti, D., Corti, S., Silani, V., Bresolin, N. and Comi, G.P. 2015. CHCHD10 mutations in Italian patients with sporadic amyotrophic lateral sclerosis. *Brain: A Journal of Neurology*. **138**(Pt 8),p.e372.

Rooke, K., Figlewicz, D.A., Han, F.Y. and Rouleau, G.A. 1996. Analysis of the KSP repeat of the neurofilament heavy subunit in familiar amyotrophic lateral sclerosis. *Neurology*. **46**(3),pp.789–790.

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P. and Deng, H.X. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature*. **362**(6415),pp.59–62.

Rosenstein, J.M., Krum, J.M. and Ruhrberg, C. 2010. VEGF in the nervous system. *Organogenesis*. **6**(2),pp.107–114.

Rubino, E., Rainero, I., Chiò, A., Rogaeva, E., Galimberti, D., Fenoglio, P., Grinberg, Y., Isaia, G., Calvo, A., Gentile, S., Bruni, A.C., St George-Hyslop, P.H., Scarpini, E., Gallone, S., Pinessi, L. and TODEM Study Group 2012. SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Neurology*. **79**(15),pp.1556–1562.

Rudnik-Schöneborn, S., Arning, L., Epplen, J.T. and Zerres, K. 2012. SETX gene mutation in a family diagnosed autosomal dominant proximal spinal muscular atrophy. *Neuromuscular Disorders*. **22**(3),pp.258–262.

Rutherford, N.J., Zhang, Y.-J., Baker, M., Gass, J.M., Finch, N.A., Xu, Y.-F., Stewart, H., Kelley, B.J., Kuntz, K., Crook, R.J.P., Sreedharan, J., Vance, C., Sorenson, E., Lippa, C., Bigio, E.H., Geschwind, D.H., Knopman, D.S., Mitsumoto, H., Petersen, R.C., Cashman, N.R., Hutton, M., Shaw, C.E., Boylan, K.B., Boeve, B., Graff-Radford, N.R., Wszolek, Z.K., Caselli, R.J., Dickson, D.W., Mackenzie, I.R., Petrucelli, L. and Rademakers, R. 2008. Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. *PLoS genetics*. **4**(9),p.e1000193.

Sabatelli, M., Marangi, G., Conte, A., Tasca, G., Zollino, M. and Lattante, S. 2016. New ALS-Related Genes Expand the Spectrum Paradigm of Amyotrophic Lateral Sclerosis. *Brain Pathology (Zurich, Switzerland)*. **26**(2),pp.266–275.

Sabatelli, M., Moncada, A., Conte, A., Lattante, S., Marangi, G., Luigetti, M., Lucchini, M., Mirabella, M., Romano, A., Del Grande, A., Bisogni, G., Doronzio, P.N., Rossini, P.M. and Zollino, M. 2013. Mutations in the 3' untranslated region of FUS causing FUS overexpression are associated with amyotrophic lateral sclerosis. *Human Molecular Genetics*. **22**(23),pp.4748–4755.

Saeed, M., Siddique, N., Hung, W.Y., Usacheva, E., Liu, E., Sufit, R.L., Heller, S.L., Haines, J.L., Pericak-Vance, M. and Siddique, T. 2006. Paraoxonase cluster polymorphisms are associated with sporadic ALS. *Neurology*. **67**(5),pp.771–776.

Sailer, A., Scholz, S.W., Gibbs, J.R., Tucci, A., Johnson, J.O., Wood, N.W., Plagnol, V., Hummerich, H., Ding, J., Hernandez, D., Hardy, J., Federoff, H.J., Traynor, B.J., Singleton, A.B. and Houlden, H. 2012. Exome sequencing in an SCA14 family demonstrates its utility in diagnosing heterogeneous diseases. *Neurology*. **79**(2),pp.127–131.

Sakaguchi, T., Irie, T., Kawabata, R., Yoshida, A., Maruyama, H. and Kawakami, H. 2011. Optineurin with amyotrophic lateral sclerosis-related mutations abrogates inhibition of interferon regulatory factor-3 activation. *Neuroscience Letters*. **505**(3),pp.279–281.

Salton, M., Elkon, R., Borodina, T., Davydov, A., Yaspo, M.-L., Halperin, E. and Shiloh, Y. 2011. Matrin 3 binds and stabilizes mRNA. *PloS One*. **6**(8),p.e23882.

Samaranch, L., Riverol, M., Masdeu, J.C., Lorenzo, E., Vidal-Taboada, J.M., Irigoyen, J., Pastor, M.A., de Castro, P. and Pastor, P. 2008. SPG11 compound mutations in spastic paraparesis with thin corpus callosum. *Neurology*. **71**(5),pp.332–336.

Saracchi, E., Castelli, M., Bassi, M.T., Brighina, E., Cereda, D., Marzorati, L., Patassini, M., Appollonio, I., Ferrarese, C. and Brighina, L. 2014. A novel heterozygous SETX mutation in a patient presenting with chorea and motor neuron disease. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*. **15**(1–2),pp.138–140.

Sasabe, J., Miyoshi, Y., Suzuki, M., Mita, M., Konno, R., Matsuoka, M., Hamase, K. and Aiso, S. 2012. D-amino acid oxidase controls motoneuron degeneration through D-serine. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(2),pp.627–632.

Sau, D., Biasi, S.D., Vitellaro-Zuccarello, L., Riso, P., Guarnieri, S., Porrini, M., Simeoni, S., Crippa, V., Onesto, E., Palazzolo, I., Rusmini, P., Bolzoni, E., Bendotti, C. and

Poletti, A. 2007. Mutation of SOD1 in ALS: a gain of a loss of function. *Human Molecular Genetics*. **16**(13),pp.1604–1618.

Schottlaender, L.V., Polke, J.M., Ling, H., MacDoanld, N.D., Tucci, A., Nanji, T., Pittman, A., de Silva, R., Holton, J.L., Revesz, T., Sweeney, M.G., Singleton, A.B., Lees, A.J., Bhatia, K.P. and Houlden, H. 2015. Analysis of C9orf72 repeat expansions in a large series of clinically and pathologically diagnosed cases with atypical parkinsonism. *Neurobiology of Aging*. **36**(2),p.1221.e1-6.

Schumacher, A., Friedrich, P., Diehl-Schmid, J., Ibach, B., Eisele, T., Laws, S.M., Förstl, H., Kurz, A. and Riemenschneider, M. 2007. No association of chromatin-modifying protein 2B with sporadic frontotemporal dementia. *Neurobiology of Aging*. **28**(11),pp.1789–1790.

Schymick, J.C., Talbot, K. and Traynor, B.J. 2007. Genetics of sporadic amyotrophic lateral sclerosis. *Human Molecular Genetics*. **16**(R2),pp.R233–R242.

Senderek, J., Garvey, S.M., Krieger, M., Guergueltcheva, V., Urtizberea, A., Roos, A., Elbracht, M., Stendel, C., Tournev, I., Mihailova, V., Feit, H., Tramonte, J., Hedera, P., Crooks, K., Bergmann, C., Rudnik-Schöneborn, S., Zerres, K., Lochmüller, H., Seboun, E., Weis, J., Beckmann, J.S., Hauser, M.A. and Jackson, C.E. 2009. Autosomal-dominant distal myopathy associated with a recurrent missense mutation in the gene encoding the nuclear matrix protein, matrin 3. *American Journal of Human Genetics*. **84**(4),pp.511–518.

Shirakawa, K., Suzuki, H., Ito, M., Kono, S., Uchiyama, T., Ohashi, T. and Miyajima, H. 2009. Novel compound heterozygous ALS2 mutations cause juvenile amyotrophic lateral sclerosis in Japan. *Neurology*. **73**(24),pp.2124–2126.

Siddiqi, S., Foo, J.N., Vu, A., Azim, S., Silver, D.L., Mansoor, A., Tay, S.K.H., Abbasi, S., Hashmi, A.H., Janjua, J., Khalid, S., Tai, E.S., Yeo, G.W. and Khor, C.C. 2014. A novel splice-site mutation in ALS2 establishes the diagnosis of juvenile amyotrophic lateral sclerosis in a family with early onset anarthria and generalized dystonias. *PloS One*. **9**(12),p.e113258.

Siddique, T., Figlewicz, D.A., Pericak-Vance, M.A., Haines, J.L., Rouleau, G., Jeffers, A.J., Sapp, P., Hung, W.Y., Bebout, J. and McKenna-Yasek, D. 1991. Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. *The New England Journal of Medicine*. **324**(20),pp.1381–1384.

Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G., Kruger, R., Federoff, M., Klein, C., Goate, A., Perlmutter, J., Bonin, M., Nalls, M.A., Illig, T., Gieger, C., Houlden, H., Steffens, M., Okun, M.S., Cookson, M., Foote, K.D., Fernandez, H.H., Traynor, B.J., Schreiber, S., Arepalli, S., Zonozi, R., Gwinn, K., van der Brug, M., Lopez, G., Chanock, S.J., Schatzkin, A., Park, Y., Hollenbeck, A., Gao, J., Huang, X., Wood, N.W., Lorenz, D., Deuschl, G., Chen, H., Riess, O., Hardy, J.A., Singleton, A.B. and Gasser, T. 2009. Genome-Wide Association Study reveals genetic risk underlying Parkinson's disease. *Nature genetics*. **41**(12),pp.1308–1312.

Singleton, A.B., Hardy, J., Traynor, B.J. and Houlden, H. 2010. Towards a complete resolution of the genetic architecture of disease. *Trends in genetics: TIG*. **26**(10),pp.438–442.

Skibinski, G., Parkinson, N.J., Brown, J.M., Chakrabarti, L., Lloyd, S.L., Hummerich, H., Nielsen, J.E., Hodges, J.R., Spillantini, M.G., Thusgaard, T., Brandner, S., Brun, A., Rossor, M.N., Gade, A., Johannsen, P., Sørensen, S.A., Gydesen, S., Fisher, E.M.C. and Collinge, J. 2005. Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. *Nature Genetics*. **37**(8),pp.806–808.

Skvortsova, V., Shadrina, M., Slominsky, P., Levitsky, G., Kondratieva, E., Zherebtsova, A., Levitskaya, N., Alekhin, A., Serdyuk, A. and Limborska, S. 2004. Analysis of heavy neurofilament subunit gene polymorphism in Russian patients with sporadic motor neuron disease (MND). *European journal of human genetics: EJHG*. **12**(3),pp.241–244.

Sleegers, K. and Van Broeckhoven, C. 2009. Motor-neuron disease: Rogue gene in the family. *Nature*. **458**(7237),pp.415–417.

Slowik, A., Tomik, B., Wolkow, P.P., Partyka, D., Turaj, W., Malecki, M.T., Pera, J., Dziedzic, T., Szczudlik, A. and Figlewicz, D.A. 2006. Paraoxonase gene polymorphisms and sporadic ALS. *Neurology*. **67**(5),pp.766–770.

Smith, B.N., Vance, C., Scotter, E.L., Troakes, C., Wong, C.H., Topp, S., Maekawa, S., King, A., Mitchell, J.C., Lund, K., Al-Chalabi, A., Ticozzi, N., Silani, V., Sapp, P., Brown, R.H., Landers, J.E., Al-Sarraj, S. and Shaw, C.E. 2015. Novel mutations support a role for Profilin 1 in the pathogenesis of ALS. *Neurobiology of Aging*. **36**(3),p.1602.e17-27.

Solski, J.A., Williams, K.L., Yang, S., Nicholson, G.A. and Blair, I.P. 2012. Mutation analysis of the optineurin gene in familial amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(1),p.210.e9-10.

Soto, C. 2003. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews Neuroscience*. **4**(1),pp.49–60.

Spataro, R. and La Bella, V. 2014. A case of amyotrophic lateral sclerosis with intermediate ATXN-1 CAG repeat expansion in a large family with spinocerebellar ataxia type 1. *Journal of Neurology*. **261**(7),pp.1442–1443.

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., Baralle, F., de Belleroche, J., Mitchell, J.D., Leigh, P.N., Al-Chalabi, A., Miller, C.C., Nicholson, G. and Shaw, C.E. 2008. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science (New York, N.Y.)*. **319**(5870),pp.1668–1672.

Sripriya, S., Nirmaladevi, J., George, R., Hemamalini, A., Baskaran, M., Prema, R., Ve Ramesh, S., Karthiyayini, T., Amali, J., Job, S., Vijaya, L. and Kumaramanickavel, G. 2006. OPTN gene: profile of patients with glaucoma from India. *Molecular Vision*. **12**,pp.816–820.

Stevanin, G., Azzedine, H., Denora, P., Boukhris, A., Tazir, M., Lossos, A., Rosa, A.L., Lerer, I., Hamri, A., Alegria, P., Loureiro, J., Tada, M., Hannequin, D., Anheim, M., Goizet, C., Gonzalez-Martinez, V., Le Ber, I., Forlani, S., Iwabuchi, K., Meiner, V., Uyanik, G., Erichsen, A.K., Feki, I., Pasquier, F., Belarbi, S., Cruz, V.T., Depienne, C., Truchetto, J., Garrigues, G., Tallaksen, C., Tranchant, C., Nishizawa, M., Vale, J., Coutinho, P., Santorelli, F.M., Mhiri, C., Brice, A., Durr, A. and SPATAX consortium 2008. Mutations in SPG11 are frequent in autosomal recessive spastic paraplegia

with thin corpus callosum, cognitive decline and lower motor neuron degeneration. *Brain: A Journal of Neurology*. **131**(Pt 3),pp.772–784.

Stewart, H.G., Andersen, P.M., Eisen, A. and Weber, M. 2006. Corticomotoneuronal dysfunction in ALS patients with different SOD1 mutations. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*. **117**(8),pp.1850–1861.

Stockmann, M., Meyer-Ohlendorf, M., Achberger, K., Putz, S., Demestre, M., Yin, H., Hendrich, C., Linta, L., Heinrich, J., Brunner, C., Proepper, C., Kuh, G.F., Baumann, B., Langer, T., Schwalenstöcker, B., Braunstein, K.E., von Arnim, C., Schneuwly, S., Meyer, T., Wong, P.C., Boeckers, T.M., Ludolph, A.C. and Liebau, S. 2013. The dynactin p150 subunit: cell biology studies of sequence changes found in ALS/MND and Parkinsonian syndromes. *Journal of Neural Transmission (Vienna, Austria: 1996)*. **120**(5),pp.785–798.

Storkebaum, E., Lambrechts, D. and Carmeliet, P. 2004. VEGF: once regarded as a specific angiogenic factor, now implicated in neuroprotection. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*. **26**(9),pp.943–954.

Subramanian, V., Crabtree, B. and Acharya, K.R. 2008. Human angiogenin is a neuroprotective factor and amyotrophic lateral sclerosis associated angiogenin variants affect neurite extension/pathfinding and survival of motor neurons. *Human Molecular Genetics*. **17**(1),pp.130–149.

Suraweera, A., Lim, Y., Woods, R., Birrell, G.W., Nasim, T., Becherel, O.J. and Lavin, M.F. 2009. Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Human Molecular Genetics*. **18**(18),pp.3384–3396.

Suzuki, N., Aoki, M., Warita, H., Kato, M., Mizuno, H., Shimakura, N., Akiyama, T., Furuya, H., Hokonohara, T., Iwaki, A., Togashi, S., Konno, H. and Itoyama, Y. 2010. FALS with FUS mutation in Japan, with early onset, rapid progress and basophilic inclusion. *Journal of Human Genetics*. **55**(4),pp.252–254.

Synofzik, M., Maetzler, W., Grehl, T., Prudlo, J., Vom Hagen, J.M., Haack, T., Rebassoo, P., Munz, M., Schöls, L. and Biskup, S. 2012. Screening in ALS and FTD patients reveals 3 novel UBQLN2 mutations outside the PXX domain and a pure FTD phenotype. *Neurobiology of Aging*. **33**(12),p.2949.e13-17.

Tafuri, F., Ronchi, D., Magri, F., Comi, G.P. and Corti, S. 2015. SOD1 misplacing and mitochondrial dysfunction in amyotrophic lateral sclerosis pathogenesis. *Frontiers in Cellular Neuroscience*. [Online]. **9**. [Accessed 7 May 2016]. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4548205/.

Takata, A., Ionita-Laza, I., Gogos, J.A., Xu, B. and Karayiorgou, M. 2016. De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia. *Neuron*. **89**(5),pp.940–947.

Takazawa, T., Ikeda, K., Hirayama, T., Kawabe, K., Nakamura, Y., Ito, H., Kano, O., Yoshii, Y., Tanaka, F., Sobue, G. and Iwasaki, Y. 2010. Familial amyotrophic lateral sclerosis with a novel G85S mutation of superoxide dismutase 1 gene: clinical features of lower motor neuron disease. *Internal Medicine (Tokyo, Japan)*. **49**(2),pp.183–186.

Tamaoka, A., Arai, M., Itokawa, M., Arai, T., Hasegawa, M., Tsuchiya, K., Takuma, H., Tsuji, H., Ishii, A., Watanabe, M., Takahashi, Y., Goto, J., Tsuji, S. and Akiyama, H. 2010. TDP-43 M337V Mutation in Familial Amyotrophic Lateral Sclerosis in Japan. *Internal Medicine*. **49**(4),pp.331–334.

Tanaka, Y., Nonaka, T., Suzuki, G., Kametani, F. and Hasegawa, M. 2016. Gain-of-function profilin 1 mutations linked to familial amyotrophic lateral sclerosis cause seed-dependent intracellular TDP-43 aggregation. *Human Molecular Genetics*. **25**(7),pp.1420–1433.

Tateishi, T., Hokonohara, T., Yamasaki, R., Miura, S., Kikuchi, H., Iwaki, A., Tashiro, H., Furuya, H., Nagara, Y., Ohyagi, Y., Nukina, N., Iwaki, T., Fukumaki, Y. and Kira, J. 2010. Multiple system degeneration with basophilic inclusions in Japanese ALS patients with FUS mutation. *Acta Neuropathologica*. **119**(3),pp.355–364.

Tavtigian, S.V., Greenblatt, M.S., Lesueur, F., Byrnes, G.B. and IARC Unclassified Genetic Variants Working Group 2008. In silico analysis of missense substitutions using sequence-alignment based methods. *Human Mutation*. **29**(11),pp.1327–1336.

Taylor JP, Brown Jr RH, Cleveland DW. Decoding ALS: from genes to mechanism. *Nature*. 2016; 539: 197–206.

Tenenbaum, J.B., de Silva, V. and Langford, J.C. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science (New York, N.Y.)*. **290**(5500),pp.2319–2323.

Teuling, E., van Dis, V., Wulf, P.S., Haasdijk, E.D., Akhmanova, A., Hoogenraad, C.C. and Jaarsma, D. 2008. A novel mouse model with impaired dynein/dynactin function develops amyotrophic lateral sclerosis (ALS)-like features in motor neurons and improves lifespan in SOD1-ALS mice. *Human Molecular Genetics*. **17**(18),pp.2849–2862.

Teyssou, E., Takeda, T., Lebon, V., Boillée, S., Doukouré, B., Bataillon, G., Sazdovitch, V., Cazeneuve, C., Meininger, V., LeGuern, E., Salachas, F., Seilhean, D. and Millecamps, S. 2013. Mutations in SQSTM1 encoding p62 in amyotrophic lateral sclerosis: genetics and neuropathology. *Acta Neuropathologica*. **125**(4),pp.511–522.

Thompson, B.A., Greenblatt, M.S., Vallee, M.P., Herkert, J.C., Tessereau, C., Young, E.L., Adzhubey, I.A., Li, B., Bell, R., Feng, B., Mooney, S.D., Radivojac, P., Sunyaev, S.R., Frebourg, T., Hofstra, R.M.W., Sijmons, R.H., Boucher, K., Thomas, A., Goldgar, D.E., Spurdle, A.B. and Tavtigian, S.V. 2013. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. *Human Mutation*. **34**(1),pp.255–265.

Thusberg, J., Olatubosun, A. and Vihinen, M. 2011. Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*. **32**(4),pp.358–368.

Ticozzi, N., LeClerc, A.L., van Blitterswijk, M., Keagle, P., McKenna-Yasek, D.M., Sapp, P.C., Silani, V., Wills, A.-M., Brown Jr., R.H. and Landers, J.E. 2011. Mutational analysis of TARDBP in neurodegenerative diseases. *Neurobiology of Aging*. **32**(11),pp.2096–2099.

Ticozzi, N., Silani, V., LeClerc, A.L., Keagle, P., Gellera, C., Ratti, A., Taroni, F., Kwiatkowski, T.J., McKenna-Yasek, D.M., Sapp, P.C., Brown, R.H. and Landers, J.E.

2009. Analysis of FUS gene mutation in familial amyotrophic lateral sclerosis within an Italian cohort. *Neurology*. **73**(15),pp.1180–1185.

Tiloca, C., Ratti, A., Pensato, V., Castucci, A., Sorarù, G., Del Bo, R., Corrado, L., Cereda, C., D'Ascenzo, C., Comi, G.P., Mazzini, L., Castellotti, B., Ticozzi, N., Gellera, C., Silani, V. and SLAGEN Consortium 2012. Mutational analysis of VCP gene in familial amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(3),p.630.e1-2.

Tiloca, C., Ticozzi, N., Pensato, V., Corrado, L., Del Bo, R., Bertolin, C., Fenoglio, C., Gagliardi, S., Calini, D., Lauria, G., Castellotti, B., Bagarotti, A., Corti, S., Galimberti, D., Cagnin, A., Gabelli, C., Ranieri, M., Ceroni, M., Siciliano, G., Mazzini, L., Cereda, C., Scarpini, E., Sorarù, G., Comi, G.P., D'Alfonso, S., Gellera, C., Ratti, A., Landers, J.E., Silani, V. and SLAGEN Consortium 2013. Screening of the PFN1 gene in sporadic amyotrophic lateral sclerosis and in frontotemporal dementia. *Neurobiology of Aging*. **34**(5),p.1517.e9-10.

Tomkins, J., Usher, P., Slade, J.Y., Ince, P.G., Curtis, A., Bushby, K. and Shaw, P.J. 1998. Novel insertion in the KSP region of the neurofilament heavy gene in amyotrophic lateral sclerosis (ALS). *Neuroreport*. **9**(17),pp.3967–3970.

Tresse, E., Salomons, F.A., Vesa, J., Bott, L.C., Kimonis, V., Yao, T.-P., Dantuma, N.P. and Taylor, J.P. 2010. VCP/p97 is essential for maturation of ubiquitin-containing autophagosomes and this function is impaired by mutations that cause IBMPFD. *Autophagy*. **6**(2),pp.217–227.

Tsai, C.-P., Soong, B.-W., Lin, K.-P., Tu, P.-H., Lin, J.-L. and Lee, Y.-C. 2011. FUS, TARDBP, and SOD1 mutations in a Taiwanese cohort with familial ALS. *Neurobiology of Aging*. **32**(3),p.553.e13-553.e21.

Tu, P.H., Raju, P., Robinson, K.A., Gurney, M.E., Trojanowski, J.Q. and Lee, V.M. 1996. Transgenic mice carrying a human mutant superoxide dismutase transgene develop neuronal cytoskeletal pathology resembling human amyotrophic lateral sclerosis lesions. *Proceedings of the National Academy of Sciences*. **93**(7),pp.3155–3160.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Research*. **40**(15),p.e115.

Uribe, V. 2010. DCTN1 mutations are implicated in multiple neurodegenerative disorders. *Clinical Genetics*. **77**(1),pp.32–34.

Valdmanis, P.N., Kabashi, E., Dyck, A., Hince, P., Lee, J., Dion, P., D'Amour, M., Souchon, F., Bouchard, J.-P., Salachas, F., Meininger, V., Andersen, P.M., Camu, W., Dupré, N. and Rouleau, G.A. 2008. Association of paraoxonase gene cluster polymorphisms with ALS in France, Quebec, and Sweden. *Neurology*. **71**(7),pp.514–520.

Van Vught, P.W.J., Sutedja, N.A., Veldink, J.H., Koeleman, B.P.C., Groeneveld, G.J., Wijmenga, C., Uitdehaag, B.M.J., de Jong, J.M.B.V., Baas, F., Wokke, J.H.J. and Van den Berg, L.H. 2005. Lack of association between VEGF polymorphisms and ALS in a Dutch population. *Neurology*. **65**(10),pp.1643–1645.

Vance, C., Rogelj, B., Hortobágyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., Ganesalingam, J., Williams, K.L., Tripathi, V., Al-Saraj, S., Al-Chalabi, A., Leigh, P.N., Blair, I.P., Nicholson, G., de Belleroche, J., Gallo, J.-M., Miller, C.C. and Shaw, C.E. 2009. Mutations in FUS, an RNA processing

protein, cause familial amyotrophic lateral sclerosis type 6. *Science (New York, N.Y.)*. **323**(5918),pp.1208–1211.

Vechio, J.D., Bruijn, L.I., Xu, Z., Brown, R.H. and Cleveland, D.W. 1996. Sequence variants in human neurofilament proteins: absence of linkage to familial amyotrophic lateral sclerosis. *Annals of Neurology*. **40**(4),pp.603–610.

Vengoechea, J., David, M.P., Yaghi, S.R., Carpenter, L. and Rudnicki, S.A. 2013. Clinical variability and female penetrance in X-linked familial FTD/ALS caused by a P506S mutation in UBQLN2. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*. **14**(7–8),pp.615–619.

Verdiani, S., Origone, P., Geroldi, A., Poggio, M.B.D., Mantero, V., Bellone, E., Mancardi, G., Caponnetto, C. and Mandich, P. 2013. The FIG4 gene does not play a major role in causing ALS in Italian patients. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*. **14**(3),pp.228–229.

Verschuuren-Bemelmans, C.C., Winter, P., Sival, D.A., Elting, J.-W., Brouwer, O.F. and Müller, U. 2008. Novel homozygous ALS2 nonsense mutation (p.Gln715X) in sibs with infantile-onset ascending spastic paralysis: the first cases from northwestern Europe. *European journal of human genetics: EJHG*. **16**(11),pp.1407–1411.

Viassolo, V., Previtali, S., Schiatti, E., Magnani, G., Minetti, C., Zara, F., Grasso, M., Dagna-Bricarelli, F. and Di Maria, E. 2008. Inclusion body myopathy, Paget's disease of the bone and frontotemporal dementia: recurrence of the VCP R155H mutation in an Italian family and implications for genetic counselling. *Clinical Genetics*. **74**(1),pp.54–60.

Vilariño-Güell, C., Wider, C., Soto-Ortolaza, A.I., Cobb, S.A., Kachergus, J.M., Keeling, B.H., Dachsel, J.C., Hulihan, M.M., Dickson, D.W., Wszolek, Z.K., Uitti, R.J., Graff-Radford, N.R., Boeve, B.F., Josephs, K.A., Miller, B., Boylan, K.B., Gwinn, K., Adler, C.H., Aasly, J.O., Hentati, F., Destée, A., Krygowska-Wajs, A., Chartier-Harlin, M.-C., Ross, O.A., Rademakers, R. and Farrer, M.J. 2009. Characterization of DCTN1 genetic variability in neurodegeneration. *Neurology*. **72**(23),pp.2024–2028.

Vinay Kumar, C., Kumar, K.M., Swetha, R., Ramaiah, S. and Anbarasu, A. 2014. Protein aggregation due to nsSNP resulting in P56S VABP protein is associated with amyotrophic lateral sclerosis. *Journal of Theoretical Biology*. **354**,pp.72–80.

Vos, K.J.D., Chapman, A.L., Tennant, M.E., Manser, C., Tudor, E.L., Lau, K.-F., Brownlees, J., Ackerley, S., Shaw, P.J., McLoughlin, D.M., Shaw, C.E., Leigh, P.N., Miller, C.C.J. and Grierson, A.J. 2007. Familial amyotrophic lateral sclerosis-linked SOD1 mutants perturb fast axonal transport to reduce axonal mitochondria content. *Human Molecular Genetics*. **16**(22),pp.2720–2728.

Vreeswijk, M.P.G., Kraan, J.N., van der Klift, H.M., Vink, G.R., Cornelisse, C.J., Wijnen, J.T., Bakker, E., van Asperen, C.J. and Devilee, P. 2009. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. *Human Mutation*. **30**(1),pp.107–114.

Wang, K., Li, M. and Hakonarson, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. **38**(16),p.e164.

Watts, G.D.J., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P. and Kimonis, V.E. 2004. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nature Genetics*. **36**(4),pp.377–381.

Weber, M., Neuwirth, C., Thierbach, J., Schweikert, K., Czaplinski, A., Petersen, J., Jung, H.H., Birve, A., Marklund, S.L. and Andersen, P.M. 2012. ALS patients with SOD1 mutations in Switzerland show very diverse phenotypes and extremely long survival. *Journal of Neurology, Neurosurgery & Psychiatry*. **83**(3),pp.351–353.

Weihl, C.C., Baloh, R.H., Lee, Y., Chou, T.-F., Pittman, S.K., Lopate, G., Allred, P., Jockel-Balsarotti, J., Pestronk, A. and Harms, M.B. 2015. Targeted sequencing and identification of genetic variants in sporadic inclusion body myositis. *Neuromuscular Disorders*. **25**(4),pp.289–296.

Weishaupt, J.H., Waibel, S., Birve, A., Volk, A.E., Mayer, B., Meyer, T., Ludolph, A.C. and Andersen, P.M. 2013. A novel optineurin truncating mutation and three glaucoma-associated missense variants in patients with familial amyotrophic lateral sclerosis in Germany. *Neurobiology of Aging*. **34**(5),p.1516.e9-15.

Wijesekera, L.C. and Leigh, P.N. 2009. Amyotrophic lateral sclerosis. *Orphanet Journal of Rare Diseases*. **4**,p.3.

Williams, K.L., Solski, J.A., Nicholson, G.A. and Blair, I.P. 2012a. Mutation analysis of VCP in familial and sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(7),p.1488.e15-16.

Williams, K.L., Warraich, S.T., Yang, S., Solski, J.A., Fernando, R., Rouleau, G.A., Nicholson, G.A. and Blair, I.P. 2012b. UBQLN2/ubiquilin 2 mutation and pathology in familial amyotrophic lateral sclerosis. *Neurobiology of Aging*. **33**(10),p.2527.e3-2527.e10.

Williamson, T.L., Bruijn, L.I., Zhu, Q., Anderson, K.L., Anderson, S.D., Julien, J.P. and Cleveland, D.W. 1998. Absence of neurofilaments reduces the selective vulnerability of motor neurons and slows disease caused by a familial amyotrophic lateral sclerosis-linked superoxide dismutase 1 mutant. *Proceedings of the National Academy of Sciences of the United States of America*. **95**(16),pp.9631–9636.

Wills, A.-M., Cronin, S., Slowik, A., Kasperaviciute, D., Es, M.A.V., Morahan, J.M., Valdmanis, P.N., Meininger, V., Melki, J., Shaw, C.E., Rouleau, G.A., Fisher, E.M.C., Shaw, P.J., Morrison, K.E., Pamphlett, R., Berg, L.H.V. den, Figlewicz, D.A., Andersen, P.M., Al-Chalabi, A., Hardiman, O., Purcell, S., Landers, J.E. and Brown, R.H. 2009. A large-scale international meta-analysis of paraoxonase gene polymorphisms in sporadic ALS. *Neurology*. **73**(1),pp.16–24.

Winter, S.M., Claus, A., Oberwittler, C., Völkel, H., Wenzler, S. and Ludolph, A.C. 2000. Recessively inherited amyotrophic lateral sclerosis: a Germany family with the D90A CuZn-SOD mutation. *Journal of Neurology*. **247**(10),pp.783–786.

Winton, M.J., Van Deerlin, V.M., Kwong, L.K., Yuan, W., Wood, E.M., Yu, C.-E., Schellenberg, G.D., Rademakers, R., Caselli, R., Karydas, A., Trojanowski, J.Q., Miller, B.L. and Lee, V.M.-Y. 2008. A90V TDP-43 variant results in the aberrant localization of TDP-43 in vitro. *FEBS letters*. **582**(15),pp.2252–2256.

Wu, C.-H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D.M., Kost, J.E., Gonzalez-Perez, P., Fox, A.D., Adams, J., Taroni, F., Tiloca, C., Leclerc, A.L., Chafe, S.C., Mangroo, D., Moore, M.J., Zitzewitz, J.A., Xu, Z.-S., van den Berg, L.H., Glass, J.D., Siciliano, G., Cirulli, E.T., Goldstein, D.B., Salachas, F., Meininger, V., Rossoll, W., Ratti, A., Gellera, C., Bosco, D.A., Bassell, G.J., Silani, V., Drory, V.E., Brown, R.H. and Landers, J.E. 2012. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature*. **488**(7412),pp.499–503.

Wu, D., Yu, W., Kishikawa, H., Folkerth, R.D., Iafrate, A.J., Shen, Y., Xin, W., Sims, K. and Hu, G.-F. 2007. Angiogenin loss-of-function mutations in amyotrophic lateral sclerosis. *Annals of Neurology*. **62**(6),pp.609–617.

Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *American Journal of Human Genetics*. **89**(1),pp.82–93.

Xia, Y., Yan, L.H., Huang, B., Liu, M., Liu, X. and Huang, C. 2014. Pathogenic mutation of UBQLN2 impairs its interaction with UBXD8 and disrupts endoplasmic reticulum-associated protein degradation. *Journal of Neurochemistry*. **129**(1),pp.99–106.

Xu, L., Li, J., Tang, L., Zhang, N. and Fan, D. 2016. MATR3 mutation analysis in a Chinese cohort with sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging*. **38**,p.218.e3-4.

Xu, S., Peng, G., Wang, Y., Fang, S. and Karbowski, M. 2011. The AAA-ATPase p97 is essential for outer mitochondrial membrane protein turnover. *Molecular Biology of the Cell*. **22**(3),pp.291–300.

Xu, Z., Cork, L.C., Griffin, J.W. and Cleveland, D.W. 1993. Involvement of neurofilaments in motor neuron disease. *Journal of Cell Science. Supplement*. **17**,pp.101–108.

Yamanaka, K., Vande Velde, C., Eymard-Pierre, E., Bertini, E., Boespflug-Tanguy, O. and Cleveland, D.W. 2003. Unstable mutants in the peripheral endosomal membrane component ALS2 cause early-onset motor neuron disease. *Proceedings of the National Academy of Sciences of the United States of America*. **100**(26),pp.16041–16046.

Yamashita, S., Mori, A., Nishida, Y., Kurisaki, R., Tawara, N., Nishikami, T., Misumi, Y., Ueyama, H., Imamura, S., Higuchi, Y., Hashiguchi, A., Higuchi, I., Morishita, S., Yoshimura, J., Uchino, M., Takashima, H., Tsuji, S. and Ando, Y. 2015. Clinicopathological features of the first Asian family having vocal cord and pharyngeal weakness with distal myopathy due to a MATR3 mutation. *Neuropathology and Applied Neurobiology*. **41**(3),pp.391–398.

Yang, S., Fifita, J.A., Williams, K.L., Warraich, S.T., Pamphlett, R., Nicholson, G.A. and Blair, I.P. 2013. Mutation analysis and immunopathological studies of PFN1 in familial and sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging*. **34**(9),p.2235.e7-10.

Yang, Y., Hentati, A., Deng, H.X., Dabbagh, O., Sasaki, T., Hirano, M., Hung, W.Y., Ouahchi, K., Yan, J., Azim, A.C., Cole, N., Gascon, G., Yagmour, A., Ben-Hamida, M., Pericak-Vance, M., Hentati, F. and Siddique, T. 2001. The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nature Genetics*. **29**(2),pp.160–165.

Yu, Z., Zhu, Y., Chen-Plotkin, A.S., Clay-Falcone, D., McCluskey, L., Elman, L., Kalb, R.G., Trojanowski, J.Q., Lee, V.M.-Y., Van Deerlin, V.M., Gitler, A.D. and Bonini, N.M. 2011. PolyQ repeat expansions in ATXN2 associated with ALS are CAA interrupted repeats. *PloS One*. **6**(3),p.e17951.

van der Zee, J., Van Langenhove, T., Kovacs, G.G., Dillen, L., Deschamps, W., Engelborghs, S., Matěj, R., Vandenbulcke, M., Sieben, A., Dermaut, B., Smets, K., Van Damme, P., Merlin, C., Laureys, A., Van Den Broeck, M., Mattheijssens, M., Peeters, K., Benussi, L., Binetti, G., Ghidoni, R., Borroni, B., Padovani, A., Archetti, S., Pastor, P., Razquin, C., Ortega-Cubero, S., Hernández, I., Boada, M., Ruiz, A., de Mendonça, A., Miltenberger-Miltényi, G., do Couto, F.S., Sorbi, S., Nacmias, B., Bagnoli, S., Graff, C., Chiang, H.-H., Thonberg, H., Perneczky, R., Diehl-Schmid, J., Alexopoulos, P., Frisoni, G.B., Bonvicini, C., Synofzik, M., Maetzler, W., vom Hagen, J.M., Schöls, L., Haack, T.B., Strom, T.M., Prokisch, H., Dols-Icardo, O., Clarimón, J., Lleó, A., Santana, I., Almeida, M.R., Santiago, B., Heneka, M.T., Jessen, F., Ramirez, A., Sanchez-Valle, R., Llado, A., Gelpi, E., Sarafov, S., Tournev, I., Jordanova, A., Parobkova, E., Fabrizi, G.M., Testi, S., Salmon, E., Ströbel, T., Santens, P., Robberecht, W., De Jonghe, P., Martin, J.-J., Cras, P., Vandenberghe, R., De Deyn, P.P., Cruts, M., Sleegers, K. and Van Broeckhoven, C. 2014. Rare mutations in SQSTM1 modify susceptibility to frontotemporal lobar degeneration. *Acta Neuropathologica*. **128**(3),pp.397–410.

Zeitz, M.J., Malyavantham, K.S., Seifert, B. and Berezney, R. 2009. Matrin 3: chromosomal distribution and protein interactions. *Journal of Cellular Biochemistry*. **108**(1),pp.125–133.

Zhang, M., Xi, Z., Zinman, L., Bruni, A.C., Maletta, R.G., Curcio, S.A.M., Rainero, I., Rubino, E., Pinessi, L., Nacmias, B., Sorbi, S., Galimberti, D., Lang, A.E., Fox, S., Surace, E.I., Ghani, M., Guo, J., Sato, C., Moreno, D., Liang, Y., Keith, J., Traynor, B.J., St George-Hyslop, P. and Rogaeva, E. 2015. Mutation analysis of CHCHD10 in different neurodegenerative diseases. *Brain: A Journal of Neurology*. **138**(Pt 9),p.e380.

Zhang, Z. and Carmichael, G.G. 2001. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell*. **106**(4),pp.465–475.

Zhao, Z., Chen, W.-Z., Wu, Z., Wang, N., Zhao, G., Chen, W. and Murong, S. 2009. A novel mutation in the senataxin gene identified in a Chinese patient with sporadic amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*. **10**(2),pp.118–122.

Zhou, Y., Liu, S., Liu, G., Oztürk, A. and Hicks, G.G. 2013. ALS-associated FUS mutations result in compromised FUS alternative splicing and autoregulation. *PLoS genetics*. **9**(10),p.e1003895.

Zou, Z.-Y., Liu, M.-S., Li, X.-G. and Cui, L.-Y. 2013a. Screening of VCP mutations in Chinese amyotrophic lateral sclerosis patients. *Neurobiology of Aging*. **34**(5),p.1519.e3-1519.e4.

Zou, Z.-Y., Peng, Y., Feng, X.-H., Wang, X.-N., Sun, Q., Liu, M.-S., Li, X.-G. and Cui, L.-Y. 2012a. Screening of the FUS gene in familial and sporadic amyotrophic lateral sclerosis patients of Chinese origin. *European Journal of Neurology*. **19**(7),pp.977–983.

Zou, Z.-Y., Peng, Y., Wang, X.-N., Liu, M.-S., Li, X.-G. and Cui, L.-Y. 2012b. Screening of the TARDBP gene in familial and sporadic amyotrophic lateral sclerosis patients of Chinese origin. *Neurobiology of Aging*. **33**(9),p.2229.e11-2229.e18.

Zou, Z.-Y., Sun, Q., Liu, M.-S., Li, X.-G. and Cui, L.-Y. 2013b. Mutations in the profilin 1 gene are not common in amyotrophic lateral sclerosis of Chinese origin. *Neurobiology of Aging*. **34**(6),p.1713.e5-6.