The University Of Sheffield.

Profiling of somatic, population and phenotypic heterogeneity in genetic causes of ALS/MND

Shaila Haque

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Sheffield Institute for Translational Neuroscience (SITraN)
Department of Neuroscience
Faculty of Medicine, Dentistry and Health
University of Sheffield

April 2023

Abstract

Amyotrophic Lateral Sclerosis (ALS) or Motor Neurone Disease (MND) is a rapidly progressive and universally fatal neurological disorder. While some genetic causes have been discovered, much remains unknown, especially in different populations. Brachial monomelic amyotrophy (MMA) is a rare condition that mimics MND, but with a better prognosis. The genetic basis of MMA is currently unknown. In this study, DNA samples from the motor cortex of ALS cases were genotyped using whole genome sequencing (WGS) to compare results with those previously found in peripheral blood to screen for the presence of somatic mutations. Targeted sequencing of ALS genes in 28 Bangladeshi MND patients, including 11 MMA patients, was performed and compared with results from 8 UK MMA patients to establish if there was genetic link between the two diseases. Two *SOD1* and a *FUS* mutation were found in the motor cortex samples but had previously been found in peripheral blood, not supporting the hypothesis of somatic heterogeneity. Screening of Bangladesh samples revealed 99 variants in 32 ALS risk genes, including pathogenic *SOD1* mutations in 3 patients. Forty-three variants of conflicting interpretations (n=17), uncertain significance (n=12), risk factor (n=1) and unknown (n=13) were found in MMA patients, highlighting a shared genetic basis of MMA and MND. This is supported by the presence of TDP-43 pathology in UK MMA patient samples. The study discovered the genetic basis of MND in Bangladesh was overlapping but distinct compared to the UK, with no *C9ORF72* mutations found in the Bangladeshi MND cases. Further work is needed to validate the variants currently designated as variants of uncertain significance in both MMA and MND patients.

Acknowledgement

First of all, I would like to express my deepest gratitude to Allah, the most compassionate and merciful, for enabling me to turn my dream into a reality.

I am deeply grateful to my admired supervisor, Professor Janine Kirby, for her constant support, guidance, and insightful contributions to my research. Her wisdom and expertise have played a significant role in my academic and scientific development during my PhD journey. I would also like to extend my gratitude to my supervisor Dr. Johnathan Cooper-Knock for his strong guidance and support throughout my PhD studies.

I would like to express my appreciation to Dr. Tom Jenkins for his valuable contributions and to Professor Oliver Bandmann for his insightful feedback. I would also like to thank Professor Laura Ferraiuolo and the technical team in SITraN for their support. I would not have been able to complete this thesis without the support and encouragement of my lab members, including Ilaria Giovannelli, Allan Shaw, Raquel Rua Martins, Tomas R Marlow, and Katie Bowden.

I am immensely grateful to Professor Dame Pamela J Shaw for her inspiration and support. I also thank the Bangladesh Government for awarding me the Prime Minister Scholarship and the University of Barishal for granting me a study leave. I am thankful to my colleagues who supported me during the document processing stage.

I have many friends who have supported me throughout my journey, but I would like to give a special shout-out to my first friends in England, Mara and Arwa, who kept me going during the COVID pandemic by making me pancakes, falafel, and planning wild trips. I am also thankful to my friends Alovya, Sadiah, Maleeha, Anushka, Emily, the Magnolia, and the Mafia Group for their constant inspiration and support. I cannot

express enough gratitude to my thesis mentor, Dr. Neda Azarmehr, for her invaluable support and guidance.

Last, but not least, I am extremely grateful to my parents and my brother who have endured my craziness during this work and showered me with their blessings, love and support.

March 2023

Dedicated to

My Parents

For their unconditional love and never-ending support

Table of Contents

List of Figures

List of Supplementary Figure

List of Tables

List of Supplementary Tables

Abbreviations

Abbreviation	*Definition*
A-T	Ataxia-Telangiectasia
ACMG	American College Of Medical Genetics And Genomics
AD	Autosomal Dominant
AlzD	Alzheimer's Disease
AF	Allele Frequencies
ALS	Amyotrophic Lateral Sclerosis
ALS-FTD	Amyotrophic Lateral Sclerosis-Frontotemporal Dementia
ALSDI	ALS Diagnostic Index
AMBRoSIA	A Multicentre Biomarker Resource Strategy In Als
AMPA	A-Amino-3-Hydroxy-5-Methyl-4-Isoxazolepropionic Acid
AOA	Ataxia with Oculomotor Apraxia
APP	Amyloid Precursor Protein
AR	Autosomal Recessive
ASOs	Allele-Specific Antisense Oligonucleotides
Bactin	β-Actin
BAM	Binary Alignment/Map
BBB	Blood-Brain Barrier
BD	Bangladesh
BD-MND	Bangladesh MND Sample
BDNF	Brain-Derived Neurotrophic Factor
BER	Base-Excision Repair
C-MYC	Cellular Myc (Myc Proto-Oncogene, Bhlh Transcription Factor)

CADD	Combined Annotation Dependent Depletion
cDNA	Complementary DNA
CNS	Central Nervous System
CNVs	Copy Number Variants
CS	Cockayne Syndrome
CTFs	C-Terminal Fragments
DALYs	Disability-Adjusted Life-Years
DDR	DNA Damage Response
Del	Deletion
DLB	Dementia With Lbs
DMEM	Dulbecco's Modified Eagles Medium
DMSO	Dimethyl Sulfoxide
DNs	Dystrophic Neurites
DPR	Dipeptide Repeat
DSBs	Double-Strand Breaks
EAAT2	Excitatory Amino Acid Transporter 2
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
ER	Endoplasmic Reticulum
ex	Exome
FALS	Familial Amyotrophic Lateral Sclerosis
FBS	Fetal Bovine Serum
FDA	Food And Drug Administration
FGF	Fibroblast Growth Factors
FTD	Frontotemporal Dementia
FTLD	Frontotemporal Lobar Degeneration

GCIs	Glial Cytoplasmic Inclusions
GDNF	Glial-Derived Neurotrophic Factor
ge	Genome
HM3	The Third Phase Of The International Hapmap Project
Hoechst	2'-[4-Ethoxyphenyl]-5-[4-Methyl-1-Piperazinyl]-2,5'-Bi-1H-Benzimidazole Trihydrochloride Trihydrate
HPO	Human Phenotype Ontology
HR	Homologous Recombination
HRE	Hexanucleotide Repeat Expansion
HREM	Hexanucleotide Repeat Expansion Mutation
HRP	Horseradish Peroxidase
HSP	Hereditary Spastic Paraplegia
HTA	Human Tissue Authority
iA	Induced astrocyte
ICC	Immunocytochemistry
IGF-I	Insulin-Like Growth Factor-I
InDels	Insertion/Deletions
iNPCs	Induced Neural Progenitor Cells
Ins	Insertion
ir	Immunoreactive
IRAS	Integrated Research Application System
KLF4	KLF Transcription Factor 4
LATE	Limbic-predominant age-related TDP-43 encephalopathy
LCT	Lactase
LD	Linkage Disequilibrium
LINE	Long Interspersed Nuclear Elements

LMN	Lower Motor Neuron
LMNs	Lower Motor Neurons
MAF	Minor Allele Frequency
MHC	Histocompatibility Complex
MIXED	Multiple-Nucleotide and an Indel
MMA	Monomelic Amyotrophy
MMR	Mismatch Repair
MND	Motor Neurone Disease
MNDA	Motor Neurone Disease Association
MNP	Multiple-Nucleotide Polymorphism
MRC	Medical Research Council
MSA	Multiple System Atrophy
MT	Mutationtaster
NCIs	Neuronal Cytoplasmic Inclusions
NER	Nucleotide Excision Repair
Nestin	Neuroepithelial Stem Cell Protein
NfL	Neurofilament Light Chain
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End-Joining
NHS	National Health Service
NHSer	Normal Horse Serum
NIIs	Neuronal Intranuclear Inclusions
NINS&H	National Institute Of Neurosciences & Hospital
NPCs	Neural Stem/Progenitor Cells
OCT3/4	Transcription Factor Protein Encoded By Pou5f1
OMIM	Online Mendelian Inheritance In Man

OPCs	Oligodendrocyte Precursor Cells
OXPHOS	Oxidative Phosphorylation
PAX6	Paired Box Protein Pax-6
PBL	Peripheral Blood Leukocytes
PBP	Progressive Bulbar Palsy
PBS	Phosphate-Buffered Saline
PD	Parkinson's Disease
PDD	Parkinson's Disease With Dementia
PEG	Polyethylene Glycol
PLS	Primary Lateral Sclerosis
PMA	Progressive Muscular Atrophy
pNfH	Phosphorylated Neurofilament Heavy Subunit
QC	Quality Control
RDBPs	RNA/DNA-Binding Proteins
REC	Research Ethics Committee
RHH	Royal Hallamshire Hospital
ROS	Reactive Oxygen Species
SALS	Sporadic Amyotrophic Lateral Sclerosis
SAM	Sequence Alignment/Map
SBTB	Sheffield Brain Tissue Bank
SDGS	Sheffield Diagnostic Genetics Service
SDS PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SGs	Stress Granules
ShARC	Sheffield Advanced Research Computer
SITraN	Sheffield Institute For Translational Neuroscience
SNCA	Alpha-Synuclein

SNP	Single Nucleotide Polymorphism
SNVs	Single Nucleotide Variants
SOX2	SRY (Sex Determining Region Y)-Box 2
SSB	Single Strand Breaks
SSBR	Single Strand Breaks Repair
TBST	Tris Buffered Saline, With Tween® 20
TDP-25/CTF25	C-Terminal Fragments 25 kDa
TDP-35/CTF35	C-Terminal Fragments 35 kDa
TUoS	The University Of Sheffield
UCSC	UCSC Genome Browser
UK	The United Kingdom
UKGTN	UK Genetic Testing Network
UMNs	Upper Motor Neurons
VCF	Variant Call Format
VEGF	Vascular Endothelial Growth Factor
VUS	Variants of Uncertain Significance
WB	Western Blot
WES	Whole Exome Sequencing
WGA	Whole Genome Amplification
XD	X-Linked (XD) Inheritance
XP	Xeroderma Pigmentosum

Chapter 1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a rapidly progressive terminal neurodegenerative disease related to the progressive death of motor neurons (Klim et al., 2019, Peters and Brown Jr, 2023). ALS typically develops in adults during midlife. In general, death takes place due to respiratory weakness after a disease course of three to five years (Haverkamp et al., 1995, Corcia et al., 2008, Udine et al., 2023).

1.1. Classification

ALS is the most common motor neurone disease (MND), an umbrella term which is used and also includes primary lateral sclerosis (PLS), progressive muscular atrophy (PMA), progressive bulbar palsy (PBP), pseudobulbar palsy, monomelic amyotrophy (MMA) (WHO, 2018, Harrison et al., 2021). Depending on the affected motor neurons, MND can be divided into three major groups: 70% ALS (van Es et al., 2017) involving upper motor neurons (UMNs) and lower motor neurons (LMNs), 5% primary lateral sclerosis (PLS) involving only UMNs and 25% progressive muscular atrophy (PMA), involving only LMNs (Swinnen and Robberecht, 2014, Grad et al., 2017). Benign brachial monomelic amyotrophy (MMA), also called Hirayama syndrome (Hirayama, 2008), is a rare type of motor neuron disorder. It primarily affects the upper limbs, causing muscle wasting and weakness. At the onset, it can resemble amyotrophic lateral sclerosis, but it has a benign prognosis and is more commonly found in males (Aundhakar et al., 2017).

ALS can be classified in different ways such as progression rate, mode of inheritance, site of onset and age of onset (**Figure 1.1**) (Ravits et al., 2013, Kennedy et al., 1998). ALS can be fast/rapid or slow depending on the rate of the progression of the disease (Müller et al., 2018). Usually, fast progression is defined as a disease duration of <2 years; slow progression is a disease duration of >4 years (van den Berg-Vos et al., 2003). ALS can be familial (FALS) or sporadic (SALS). Sporadic Amyotrophic Lateral Sclerosis (SALS) is the most common form of ALS accounting for 90% of cases that occur randomly, with no history of the disease in the family (Sabatelli et al., 2013). If one or more first- or second-degree relatives are reported to suffer from ALS, it is known as FALS (Valdmanis and Rouleau, 2008). While the majority of cases exhibit autosomal dominant (AD) inheritance there are rare cases of autosomal recessive (AR), or X-linked (XD) inheritance (Rowland and Shneider, 2001, Nguyen et al., 2018a, Cooper-Knock et al., 2017). SALS also has a strong genetic component – estimated heritability from twin studies is 61% (Al-Chalabi et al., 2010). SALS likely results from a complex interaction of genetics and the environment although, to date, there is disagreement as to which environmental risk factors are important.

ALS can also be subdivided based on the site of onset: most commonly spinal-onset ALS/limb-onset ALS (70%), bulbar onset ALS (20-15%), though rare respiratory onset and generalized onset where multiple regions are involved simultaneously have also been reported (Brotman et al., 2021). Progressive bulbar palsy (PBP), a subtype of PLS was originally thought to be a different disease from ALS based on a distinct survival profile associated with dysarthria or dysphagia due to bulbar muscle weakness and

atrophy (Swinnen and Robberecht, 2014, Couratier et al., 2021). Two additional regional variants are described associated with prolonged survival: flail arm syndrome/brachial amyotrophic diplegia and flail leg syndrome/leg amyotrophic diplegia in which proximal muscles are preferentially affected (Swinnen and Robberecht, 2014). Other variants include amyotrophic lateral sclerosis-frontotemporal dementia (ALS-FTD), a form of ALS referring to patients who fulfil the diagnostic criteria for both ALS and frontotemporal dementia (FTD) (van Es et al., 2017). A very rare variant is respiratory onset, with orthopnoea or dyspnoea, and mild or even absent spinal or bulbar signs at presentation (Chiò et al., 2011). Usually, ALS starts in the fifth or sixth decade of life which is described as the peak age of onset. Onset at a younger age has been described as either young onset of ALS before age 45 (Turner et al., 2012) or juvenile-onset of ALS affected before age 25 (Orban et al., 2007).

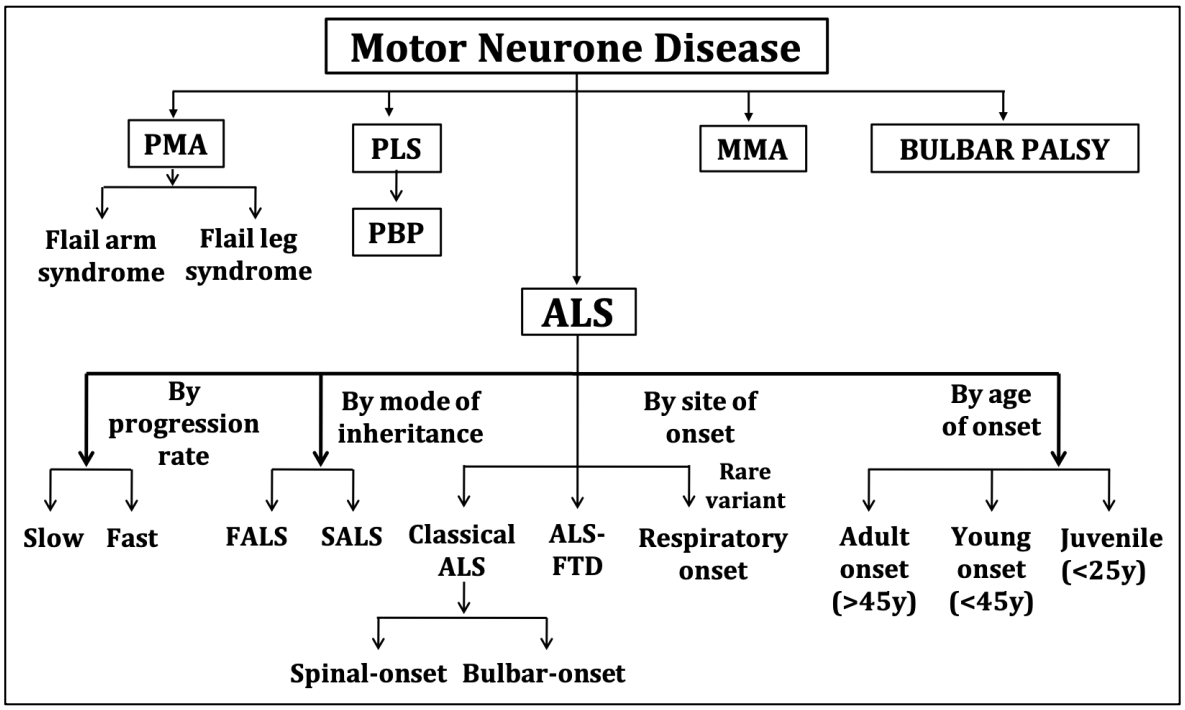


Figure 1.1. Overview of MND and frequent ALS classifications; it should be noted that certain patients can have, for example, bulbar onset associated with PLS.

1.2. Signs and symptoms

The onset of ALS is often subtle such that the symptoms may be barely noticeable at the early stage but the weakness becomes more visible over time. The progression of ALS differs from person to person. Some of the early symptoms and signs include:

1. Trouble in carrying out daily activities such as fastening buttons and using a fork
2. Weakness in the feet, hands, legs, and ankles
3. Muscle cramps and fasciculations (muscle twitches) in the arms, shoulders, and tongue
4. Worsening posture and a hard time holding the head up

5. Cognitive changes (~3 to 30% of ALS patients develop mild cognitive impairment which can progress to FTD (Grossman et al., 2007))
6. Slurred and nasal speech
7. Difficulty in breathing, chewing and swallowing (Saudagar et al., 2019).

1.3. Pathophysiology/Pathogenesis of ALS

1.3.1. ALS Mechanism

The mechanisms involved in the development of ALS neurodegeneration are still not understood although a set of consistent candidates has been described (**Figure 1.2**) (Kiernan et al., 2011). Many mechanisms have been implicated in both SALS and FALS, including protein aggregation, excitotoxicity, RNA toxicity, hypermetabolism, inflammation, prion-like spreading, and decreased neurotrophic support from non-neuronal cells (van Es et al., 2017). Table 1.1 shows the major pathophysiological mechanisms involved in ALS along with a set of risk genes associated with each mechanism.

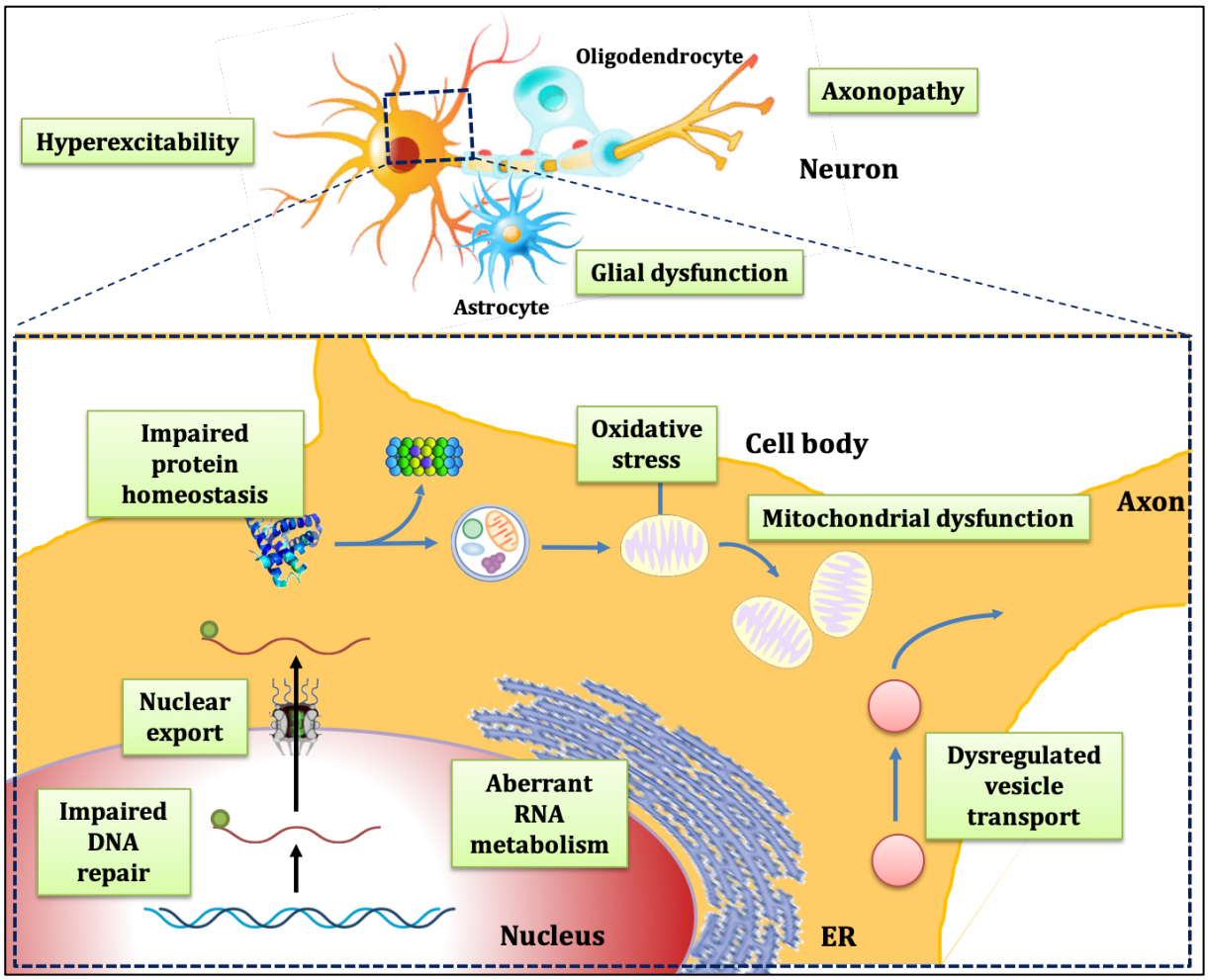


Figure 1.2. The cellular and molecular mechanism mediating ALS neurodegeneration.

The arrows indicate the links between the mechanisms. ER, endoplasmic reticulum.

The figure is adapted from Hardiman et al., 2017 (Hardiman et al., 2017).

Table 1.1. The main pathophysiological mechanisms involved in ALS and the ALS genes are shown to be associated with each of these mechanisms. For abbreviation see Table 1.2.

Implicated disease mechanisms	Gene	Refs
Hyperexcitability	*SOD1*	(Kim et al., 2017)
Impaired DNA repair	*C21ORF2, FUS, NEK1* and *SPG11*	(Sun et al., 2020)
Aberrant RNA metabolism	*ANG, ATXN2, C9ORF72, ELP3, EWSR1, FUS, HNRNPA1, HNRNPA2B1, MATR3, SETX, TAF15* and *TARDBP*	(Donnelly et al., 2014)
Nuclear export	*C9ORF72* and *FUS*	(Jovicic et al., 2016)
Impaired protein homeostasis	*ALS2, C9ORF72, CHMP2B, OPTN, SIGMAR1, SQSTM1, UBQLN2, SOD1, TBK1, VAPB* and *VCP*	(Maruyama et al., 2010a, Johnson et al., 2010b, Luty et al., 2010, Al-Saif et al., 2011, Fecto et al., 2011, Weidberg and Elazar, 2011)
Oxidative stress	*ALS2, SOD1* and *TARDBP*	(Rosen, 1993, Siddique et al., 1991)
Mitochondrial dysfunction	*CHCHD10, SOD1* and *TARDBP*	(Bannwarth et al., 2014)
Dysregulated vesicle transport	*ALS2, CHMP2B, FIG4, OPTN, SOD1, UNC13A* and *VAPB*	(Nishimura et al., 2004, Sreedharan et al., 2008)
Glial dysfunction	*C9ORF72* and *SOD1*	(DeJesus-Hernandez et al., 2011, Renton et al., 2011)
Axonopathy	*DCTN, NEFH, PFN1, PRHPH, SOD1, SPG11* and *TUBA4A*	(Moloney et al., 2014, Wu et al., 2012)

SOD1 has been linked to the largest number of associated pathophysiological mechanisms, probably because it was the cornerstone of disease modelling from the time of the discovery that autosomal dominant mutations within *SOD1* cause ALS (Rosen et al., 1993, Masrori and Van Damme, 2020) until relatively recently. Mechanisms linked to ALS associated with *SOD1* include oxidative stress,

mitochondrial dysfunction, dysregulated vesicle transport, glial cells, axon dysfunction, neuronal hyperexcitability and SOD1 protein aggregates (Hardiman et al., 2017). Intracellular ubiquitinated protein aggregates which are a hallmark of ALS are proposed to be harmful to neurons. Some of the most prominent mechanisms will now be discussed in detail.

1.3.1.1. Protein Aggregation

Protein aggregation is identified as a key feature of ALS (Ince et al., 1998a, Ince et al., 1998b, Piao et al., 2003). The TDP-43 protein, encoded from *TARDBP* was found to be the key protein aggregated in >98% of both familial and sporadic ALS (Sreedharan et al., 2008). Exceptions include cases of FALS with mutations in *SOD1* and *FUS*, misfolded SOD1 and cytoplasmic inclusions containing mutant fused in sarcoma (FUS) protein are detected respectively (Rakhit et al., 2007, Groen et al., 2010, Hewitt et al., 2010b). Mutations in genes responsible for protein processing may contribute directly to ALS. For example, recently mutation within valosin-containing protein (VCP) which regulates protein homeostasis, autophagy, organelle biogenesis and cell signalling, and mutation within ubiquitin 2 (*UBQLN2*) which regulates the protein degradation pathway, have been associated with monogenic ALS (Johnson et al., 2010a, Ritson et al., 2010, Deng et al., 2011). In 2022 Gosset et al proposed the prionoid hypothesis that protein aggregates may be toxic to neurons via a prion-like behaviour in which misfolded protein aggregates induce aggregation in spatially associated wild type protein (Gosset et al., 2022). Previous studies showed a prion-like activity of SOD1 *in vivo* that suggested that the expression of human wildtype SOD1 can influence the

phenotype of mice expressing ALS-linked mutated SOD1 (Deng et al., 2006, Prudencio et al., 2010, Wang et al., 2009). TAR DNA-binding protein of 43 kDa (TDP-43) exhibits prionoid characteristics in addition to Cu/Zn superoxide dismutase 1 (*SOD1*) and fused in sarcoma (*FUS*) (McAlary et al., 2019, Polymenidou and Cleveland, 2011, Bertolotti, 2018).

It is interesting that TDP-43 proteinopathy is observed in various neurodegenerative diseases such as Parkinson's disease (PD) without or with dementia (PDD), and dementia with LBs (DLB) alone or in association with Alzheimer's disease (AlzD) (Nakashima-Yasuda et al., 2007, Amador-Ortiz et al., 2007, Tziortzouda et al., 2021). Moreover, the presence of TDP-35, a truncated fragment of TDP-43, beyond certain thresholds in humans is suggested to affect neuronal health and may aggravate the course of TDP-43 proteinopathies (Chhangani and Rincon-Limas, 2022).

The formation of TDP-43 into stress granules is a natural and reversible process that enables cells to conserve energy by reducing mRNA translation during stressful periods. In vivo stress granules are a prime example of liquid-liquid phase separation (LLPS) and are held together by protein-protein interactions through TDP-43's low complexity domain and protein-RNA interactions via TDP-43 RRM. The presence of mutant TDP-43 has been linked to impaired stress granule formation and an elevated depletion of TDP-43 in the nucleus of the mammalian nervous system, which may be significant in the development of ALS/FTD pathology (Dubinski et al., 2023). However, TDP-35 can form irreversible aggregates in the cytoplasm, which could represent stress granules that failed to disassemble. The presence of TDP-35 is always considered

pathological, and my observations in induced astrocyte (iAstrocyte/iA) suggest that the mutations modelled have caused TDP-43 proteinopathy, a key feature of ALS. Additionally, these irreversible RNA granules containing TDP-35 may result in the mislocalization of endogenous TDP-43.

TDP-35 inclusions containing TDP-43 or other RNPs formed by RNA-assisted sequestration may be relevant to a gain-of-function pathology. The ability of TDP-35 to form inclusions in the cytoplasm, while possessing the most essential domains of TDP-43, may impact the inherent function of TDP-43 in mRNA processing (Che et al., 2011). In summary, TDP-35's ability to sequester normal TDP-43 and other RNA-binding proteins into cytoplasmic inclusions via RNA binding may be relevant to the pathogenesis of neurodegenerative diseases which exhibit TDP-proteinopathy, and the TDP-35 inclusions formed by RNA-assisted sequestration may be similar to cytoplasmic RNP granules. Understanding the mechanisms underlying TDP-35's pathogenicity may provide insights into developing new therapies for these devastating diseases.

1.3.1.2. Mitochondrial Dysfunction & Oxidative Stress

Mitochondria, which are the energy producers of the cell, generate ATP via oxidative phosphorylation (OXPHOS) and are involved in responding to oxidative stress and apoptosis. (Federico et al., 2012). Efficient mitochondrial function is particularly essential for large motor neurons because of their high energy demand. As mitochondria contain many redox enzymes, malfunction of mitochondria may lead to

oxidative phosphorylation generating reactive oxygen species (ROS) (Federico et al., 2012). Although ROS produced in normal cells is not thought to cause ALS, excessive production may contribute to the development of the disease due to energy failure (Federico et al., 2012, Zhao et al., 2022). ALS-associated mutations within both SOD1 and TARDBP proteins cause mitochondrial dysfunction leading to oxidative stress in ALS (Vande Velde et al., 2011, Magrane et al., 2014, Motataianu et al., 2022). Postmortem and biopsy samples from the spinal cord, nerves and muscles of both SALS and FALS show abnormalities in mitochondrial structure, number and localization (Lin and Beal, 2006) as well as increased oxidative stress (Stein et al., 2021). Mitochondrial dysfunction has been associated with many pathological changes in ALS and is therefore believed to be commonly involved in the pathogenesis of the disease (Zhao et al., 2022).

1.3.1.3. Excitotoxicity, Glutamate, its Transport and Receptors

In CNS, the main excitatory neurotransmitter is glutamate. Excitatory amino acid transporter 2 (*EAAT2*) acts as a glutamate reuptake transporter removing glutamate from the synaptic cleft and ending the excitatory signal. Thus alterations in glutamate receptor expression or neuronal energy homeostasis may cause increased sensitivity of the postsynaptic neuron to glutamate or synaptic levels of glutamate leading to excitotoxicity (Van Damme et al., 2005). The key components of excitotoxicity include abnormal mitochondrial function and ATP production, and intracellular calcium homeostasis (Arundine and Tymianski, 2003). Another excitatory glutamatergic

neurotransmission regulating α-Amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor is also found to mediate excitotoxicity through calcium permeability. Elevated CSF glutamate levels are found in ALS patients which might be consistent with excessive neuronal stimulation leading to excitotoxicity (Shaw et al., 1995). Riluzole, the first drug approved for the treatment of ALS, is thought to act via the prevention of excitotoxicity (Kyllo et al., 2022).

UNC13A (Unc-13 Homolog A) is a key player in the process of vesicle maturation during exocytosis and is targeted by the diacylglycerol second messenger pathway (Diekstra et al., 2012b). Specifically, it participates in neurotransmitter release by facilitating synaptic vesicle priming before vesicle fusion and contributes to the refilling of the readily releasable vesicle pool in an activity-dependent manner (Diekstra et al., 2012b).

TDP-43 has an important role as a repressor of cryptic exon inclusion during RNA splicing (Ling et al., 2015). Although human genome-wide association studies have identified single nucleotide polymorphisms in *UNC13A* as being strongly associated with FTD and ALS (Van Es et al., 2009, Diekstra et al., 2014), it is not clear how these variants increase the risk of disease. Recent studies by Ma et al. and Brown et al. shed light on this mechanism by revealing that loss of TDP-43 function leads to the inclusion of cryptic exons in specific mRNAs, including the synaptic gene *UNC13A* (Ma et al., 2022, Brown et al., 2022). Loss of TDP-43 from the nucleus in various human brain and neuronal cell lines, as well as motor neurons derived from induced pluripotent stem cells, resulting in the inclusion of a cryptic exon in *UNC13A* mRNA and reduced expression of the UNC13A protein.

Moreover, the top variants associated with FTD or ALS risk in humans are located in the intron-containing cryptic exon, and they increase *UNC13A* cryptic exon splicing in the presence of TDP-43 dysfunction. These findings provide evidence of a genetic association between UNC13A and the loss of nuclear TDP-43 function and excitotoxicity (Liu et al., 2023). Additionally, the research sheds light on the mechanisms by which UNC13A variants worsen the effects of reduced TDP-43 function (Ma et al., 2022, Brown et al., 2022, Liu et al., 2023).

1.3.1.4. Immunity, Inflammation & Cytokines

The pathology of ALS shows evidence of neuroinflammation such as activated microglia and infiltrating lymphocytes in the CNS (Henkel et al., 2004). In the early stage of ALS reduced monocytes ($CD14^+$ cells) and $CD4^+CD25^+$ regulatory T (T_{REG}) cells are detected (Henkel et al., 2013). Other studies reported a reduced amount of T_{REG} being associated with disease severity (Mantovani et al., 2009, Rentzos et al., 2012) and disease duration (Rentzos et al., 2012). In ALS altered immune responses have been linked to the degeneration and death of motor neurons (Hardiman et al., 2017). Neuroinflammation is also connected to the activation of immune response that can be observed in imaging studies in ALS patients, rodent models of ALS and human post-mortem samples (Corcia et al., 2012, Brites and Vaz, 2014). For example, deleting mutant *SOD1* from astrocytes and microglial cells in a mouse model of ALS increases survival by slowing the disease progression (Wang et al., 2011); it is thought that this may be via a direct effect on the inflammatory modulation performed by these glial cells.

1.3.2. Genetic factors involved in ALS

Both genetic and environmental factors are thought to be responsible for ALS. Multiple genetic factors causing ALS have been identified with mutations in more than 120 genes found to be associated with ALS (alsod.ac.uk), accounting for about 70% of familial and 15% of apparently sporadic cases (**Table 1.2**) (Nguyen et al., 2018a, Klim et al., 2019, Cooper-Knock et al., 2021a). Among them, the most frequently mutated ALS risk genes are chromosome 9 open reading frame 72 (*C9ORF72*), superoxide dismutase 1 (*SOD1*), TAR DNA Binding Protein (*TARDBP*) and Fused in Sarcoma (*FUS*) (Morgan and Orrell, 2016, Lattante et al., 2013, Paez-Colasante et al., 2015).

1.3.2.1. *C9ORF72*

The C9ORF72 protein is thought to be involved in the initiation of autophagy however, it is still unknown whether ALS-associated mutation changes the normal function of C9ORF72. The *C9ORF72* mutation is a GGGGCC (G4C2) hexanucleotide repeat expansion (HRE) within intron 1. The exact length of expansion necessary to cause the disease is debated but expansions of longer than 30 repeats are associated with ALS (Majounie et al., 2012). The expansion in *C9ORF72* is proposed to cause disease through 1) haploinsufficiency causing loss-of-function of *C9ORF72*, 2) RNA toxicity due to the construction of irregular HRE-containing RNA, and 3) accumulation of toxic dipeptide repeat proteins (DPR) translated from HRE RNA (Babic Leko et al., 2019). Around 39% of familial and 7% of sporadic cases of ALS of European ancestry carry the *C9ORF72* HRE; notably, the frequency varies significantly between populations (Saudagar et al.,

2019). The *C9ORF72* HRE is the most common mutation linking familial ALS and FTD (Gossye et al., 2020). For instance, the *C9ORF72* HRE is linked to 25% of familial FTD cases (Smeyers et al., 2021). FALS and SALS are largely indistinguishable clinically and pathologically. Therefore, studying, *C9ORF72* may be helpful for the diagnosis and design of therapies for all ALS patients as it is the largest single genetic cause of ALS.

1.3.2.2. *SOD1*

Cu–Zn superoxide dismutase 1 (*SOD1*) mutation was the first designated genetic cause for FALS (Rosen et al., 1993). Around 20% of familial and 1% of sporadic cases are estimated to have mutations in the *SOD1* gene (Saudagar et al., 2019). Around 200 mutations in the *SOD1* gene have been reported to cause ALS. In ALS, *SOD1* acquires a toxic, unknown gain of function (Bunton-Stasyshyn et al., 2015). However, the normal functions of *SOD1* and the cellular pathways that depend on this protein are still unclear, despite its abundance and ubiquity. Although it is known to function as an antioxidant in the cytosol, the levels of SOD1 protein are much higher than what is required for this role alone. Recent research has revealed new functions of *SOD1*, including its involvement in RNA metabolism as a nuclear transcription factor, and in metabolic signalling. Interestingly, given the roles of other genes that cause ALS, these functions shed new light on *SOD1*'s potential involvement in the disease (Bunton-Stasyshyn et al., 2015). The role of *SOD1* in ALS is affected by various factors such as excitotoxicity (via AMPA receptor and sodium channel activation), ER stress, and cell-specific impacts from astrocytes and microglia. Investigating these functions provides potential avenues for developing much-needed treatments to slow down disease

progression in *SOD1*-FALS, and perhaps even SALS, along with anti-sense oligonucleotide therapies (Musarò, 2013, Yang et al., 2013). While the majority of *SOD1*-related familial ALS cases exhibit autosomal dominant inheritance, there have been reports of rare cases displaying autosomal recessive inheritance. In certain families, a "tightly linked protective factor" has been identified, which modifies the toxic impact of the mutant *SOD1* and leads to the manifestation of recessive inheritance (Al-Chalabi et al., 1998). However, our understanding of *SOD1*'s functions is still limited, and many aspects of this small protein remain enigmatic.

1.3.2.3. *TARDBP*

The *TARDBP* gene encodes a protein called transactive response DNA binding protein 43 kDa (TDP-43). *TARDBP* encodes several isoforms of which a heterogeneous nuclear ribonucleoprotein (hnRNP), TDP-43 is considered to be the most predominant and a major component of the cytoplasmic ubiquitinated inclusions seen in the majority of ALS cases (Corrado et al., 2009). Mutations in the TDP-43 protein mostly affect the glycine-rich domain, the protein region involved in mRNA processing, likely interrupting the production of other proteins (Lagier-Tourenne et al., 2012). Around 5% of FALS patients are found to have *TARDBP* mutations (Saudagar et al., 2019). (Section 1.3.1.1 and 1.3.1.2).

1.3.2.4. *FUS*

The *FUS* gene encodes a protein involved in microRNA processing, the transcription process, transport, trafficking, and alternative splicing. At least 85 mutations causing ALS have been reported in the FUS gene, and account for 5% of FALS cases (Saudagar et al., 2019). Mutations mostly affect the protein region involved in DNA binding and messenger RNA (mRNA) processing (Fujii and Takumi, 2005, Fujii et al., 2005). These ultimately result in interfering with the transportation of mRNA from the nucleus of cells. Consequently, mutant FUS protein and mRNA are aggregated within the motor neuron cells of ALS patients. A younger onset of ALS and a decreased life expectancy have been observed in cases carrying mutations in the *FUS* gene compared to ALS triggered by mutations in other genes (Bäumer et al., 2010). In a study by López-Erauskin et al., humanized *FUS* mice were created to investigate the role of human wild-type and ALS/FTD-linked *FUS* mutants in complementing the functions of murine *FUS*. The research shows that when mutant human *FUS* is expressed at a level and subcellular distribution that mimics endogenous *FUS*, it leads to progressive motor and cognitive deficits in mice, accompanied by RNA alterations due to a gain of toxicity (López-Erauskin et al., 2020). The study reveals that even a slight increase in *FUS* is sufficient to trigger age-dependent motor deficits, indicating that neurodegeneration is not primarily caused by reduced *FUS* activity. The research suggests that RNA-binding proteins play a critical role in regulating local protein translation and synaptic function.

Table 1.2. ALS Genes. Abbreviations are Autosomal Dominant (AD), Autosomal Recessive (AR), X-linked (XD), *Extremly Rare.

ALS types	Gene	Abbreviation	Associated with Loci	Type of inheritance
Known ALS loci and genes				
ALS1	Superoxide dismutase 1	*SOD1*	21q22	AD/AR*
ALS2	Alsin Rho Guanine Nucleotide Exchange Factor	*ALS2*	2q33	AR
ALS4	Senataxin	*SETX*	9q34	AD
ALS5	Spastic paraplegia-11	*SPG11*	15q15-21	AR
ALS6	RNA binding protein Fused in Sarcoma	*FUS*	16q12	AD/AR
ALS8	Vesicle-associated membrane protein-associated protein B	*VAPB*	20q13.33	AD
ALS9	Angiogenin/ ribonuclease 5	*ANG*	14q11.2	AD
ALS10	TAR DNA Binding Protein	*TARDBP*	1p36.22	AD
ALS11	FIG4 Phosphoinositide 5-Phosphatase	*FIG4*	6q21	AD
ALS12	Optineurin	*OPTN*	10p15-p14	AD/AR
ALS13	Ataxin 2	*ATXN2*	12Q24.12	AD
ALS14	Valosin Containing Protein	*VCP*	9p13-p12	AD
ALS15	Ubiquilin 2	*UBQLN2*	Xp11-q12	XD
ALS16	Sigma Non-Opioid Intracellular Receptor 1	*SIGMAR1*	9p13.3	AD/AR
ALS17	Charged Multivesicular Body Protein 2B	*CHMP2B*	3p12.1	AD
ALS18	Profilin 1	*PFN1*	17p13.3	AD
ALS19	Receptor tyrosine-protein kinase erbB-4	*ERBB4*	2q33.3-q34	AD
ALS20	Heterogeneous Nuclear Ribonucleoprotein A1	*hnRNPA1*	12q13.1	AD
ALS21	Matrin 3	*MATR3*	5q31.2	AD
ALS22	Tubulin Alpha 4A	*TUBA4A*	2q35	AD

ALS23	Annexin A11	*ANXA11*	10q22.2	AD
ALS24	NIMA Related Kinase 1	*NEK1*	4q33	AD
ALS25	Kinesin Family Member 5A	*KIF5A*	12q13.3	AD
ALS26	Tia1 Cytotoxic Granule-Associated Rna-Binding Protein	*TIA1*	2p13.3	AD
Known ALS loci with unknown causative gene				
ALS3	Unknown	-	18q21	AD
ALS7	Unknown	-	20ptel-p13	AD
Known ALS & FTD loci and genes				
ALS-FTD1	Chromosome 9 open reading frame 72	*C9ORF72*	9q21.2	AD
ALS-FTD2	Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10	*CHCHD10*	22q11	AD
ALS-FTD3	Sequestosome 1	*SQSTM1*	5q35	AD
ALS-FTD4	TANK Binding Kinase 1	*TBK1*	12q14.2	AD
ALS-FTD5	Cyclin F	*CCNF*	16p13.3	AD
ALS-FTD6	Valosin-Containing Protein	*VCP*	9p13.3	AD
ALS-FTD7	Charged Multivesicular Body Protein 2B	*CHMP2B*	3p11.2	AD
ALS-FTD8	Cyld Lysine-63 Deubiquitinase	*CYLD*	16q12.1	AD
Potentially other ALS loci with known genes include				
ALS	Neurofilament Heavy	*NEFH*	22q12.1	AD/AR
ALS	Dynactin	*DCTN1*	2p13	AD/AR
ALS	D-Amino Acid Oxidase	*DAO*	12q24	AD
ALS	TATA-Box Binding Protein Associated Factor 15	*TAF15*	17q11.1-q11.2	AD
ALS	EWS RNA Binding Protein 1	*EWSR1*	22q12.2	AD
ALS	Rho Guanine Nucleotide Exchange Factor 28	*ARHGEF28*	5q13.2	AD
ALS	Mitofusin-2	*MFN2*	1p36.22	AD
ALS-Paskinsonism	Transient Receptor Potential Cation Channel, Subfamily M, Member 7	*TRPM7*	15q21.2	AD

Although the cause of sporadic amyotrophic lateral sclerosis (SALS) in most cases is not known, mutations in many of the genes causing familial amyotrophic lateral sclerosis (FALS) have been identified in apparently sporadic cases. ALS gene frequencies in **Figure 1.3** shows that the *C9ORF72* gene is the most frequent gene causing amyotrophic lateral sclerosis (ALS) (Balendra and Isaacs, 2018).

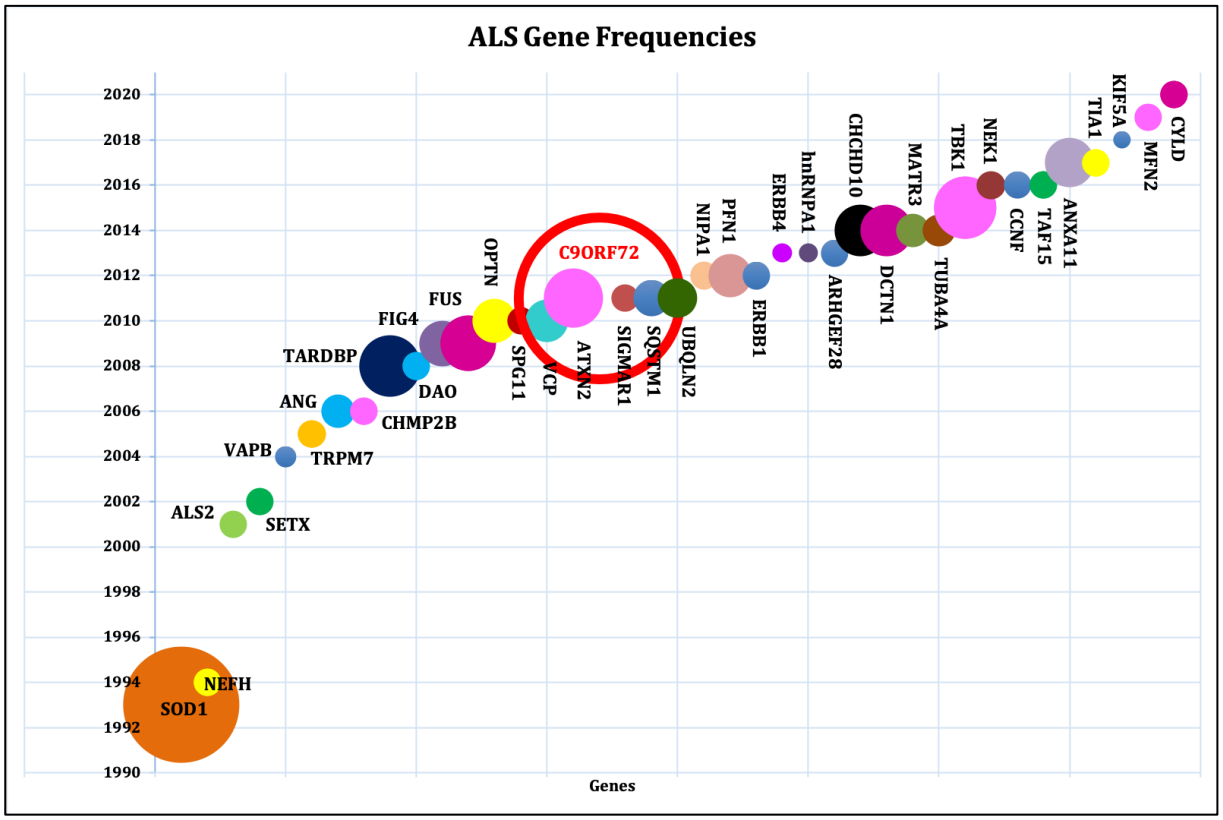


Figure 1.3. Gene frequencies in ALS were adapted and updated from Alsultan et al. (Alsultan et al., 2016).

Notes: The genes are plotted against the year they were found; the dimension of the circles indicates the frequency of mutations in FALS or ALS as cited in the literature. For illustrative purposes, a circle size equivalent to 1% is used where gene frequencies were not available (For abbreviation see Table 1.2).

1.3.3. Environmental factors involved in ALS

The environmental risk factors involved in ALS are less understood than the genetic factors (Armon, 2001, Belbasis et al., 2016). Smoking (Gallo et al., 2009), lifetime intensive sport or physical exertion (Harwood et al., 2009, Julian et al., 2021) and active service in the armed forces (Kasarskis et al., 2009) have been suggested as some of the environmental risk factors for ALS.

A hypothesis of gene-environment-time interaction in neurodegenerative disease suggested that genetic predisposition interacts with environmental exposures over time leading to ALS (Al-Chalabi and Hardiman, 2013, Bradley et al., 2018). There are likely to be hundreds of individual gene-environment interactions, each of which has its particular time course and mechanism of toxicity (Bradley et al., 2018). The gene-time-environment hypothesis of ALS has a "multistep" model to account for the environmental effect on disease onset and progression (Al-Chalabi and Hardiman, 2013). The gene-environment interaction promotes disease in up to 6 discrete steps in European and East Asian populations, whereas it has fewer steps in patients having known monogenic, penetrant mutations, e.g., *C9ORF72, SOD1, TARDBP, FUS* (Chio et al., 2018, Vucic et al., 2020, Goutman et al., 2022) suggesting that fewer environmental insults may be required in the context of a 'highly penetrant' mutation.

1.4. Somatic genetic variation

Although multiple genes have been associated with ALS the genetic basis of sporadic ALS is not yet fully understood. Mutation in known ALS genes has been identified in

almost 70% of familial and approximately 11% of sporadic ALS cases (Renton et al., 2014, Camu et al., 2022). The identification of *de novo* mutations as the cause of neurologic diseases, including ALS, has received some attention (Renton et al., 2014, Kim et al., 2015). A *de novo* mutation is a type of mutation that occurs spontaneously during the process of DNA replication during cell division in a fetus whose close, biological relatives don't have the mutation (Acuna-Hidalgo et al., 2016). As a result, only a subset of cells within the individual carry the mutation; in comparison, a germline mutation is replicated in all cells within an individual.

Somatic mutations are mutations that occur after fertilization and can result in mosaicism, where an organism has cells with genetic differences (Leija-Salazar et al., 2018). Mosaicism is common in the human body, including the brain, and its potential involvement in neurodegenerative diseases has been suggested for some time. Recent technological advancements have enabled a more detailed exploration of this phenomenon (Proukakis, 2020). Different types of somatic mutations, such as single nucleotide variants (SNVs), copy number variants (CNVs), and retrotransposon insertions, may occur during development or as a result of the disease process and could serve as initiators or risk factors for disease progression (Proukakis, 2020). Relevant mutations have been identified in sporadic neurodegenerative disorders, such as synucleinopathies involving somatic gains of alpha-synuclein (*SNCA*) in Parkinson's disease and multiple system atrophy, and in Alzheimer's disease, where a novel recombination mechanism has been identified, resulting in somatic variants of amyloid precursor protein (APP) and an excess of somatic SNVs affecting tau

phosphorylation (Proukakis, 2020). Mosaicism due to somatic instability has gained prominence in Mendelian repeat expansion disorders, first observed 25 years ago. Brain somatic SNVs have been observed in DNA repair disorders, and several genes associated with ALS have been implicated in DNA repair (Proukakis, 2020). Despite challenges and the need for further validation, this emerging area of research has the potential to revolutionize our understanding of neurodegeneration (Leija-Salazar et al., 2018). However, most genetic studies of neurodegenerative diseases, including ALS, have focused on DNA extracted from whole blood rather than disease-relevant central nervous system tissue, potentially missing a substantial role for somatic mutations.

Recent studies have explored the presence of somatic mutations in apparently healthy nervous system tissues and their potential role in non-neoplastic neurological diseases (Proukakis, 2020). However, the role of somatic mutations in neurodevelopmental disorders is well-established (D'Gama and Walsh, 2018). The possibility of somatic mutations contributing to neurodegenerative diseases has been suggested in multiple studies (Frank, 2010, Pamphlett, 2004, Van Broeckhoven, 2010, Proukakis et al., 2013), and early research in this area has been reviewed (Leija-Salazar et al., 2018, Lodato and Walsh, 2019, Nicolas and Veltman, 2019, Verheijen et al., 2018). Advanced DNA sequencing techniques have shown that mosaicism is more prevalent in humans than previously thought (Biesecker and Spinner, 2013, Macosko and McCarroll, 2012, Campbell et al., 2015, Forsberg et al., 2017), whether arising in development or ageing, with the term "somatic evolutionary genomics" used to describe the study of the

accumulation of somatic mutations in the body (Frank, 2010). Southern blotting has revealed somatic heterogeneity in C9ORF72-related ALS (Cooper-Knock et al., 2014).

1.4.1. How do neurons acquire somatic mutations?

Mutations can occur either during mitosis or post-mitotically. Mitotic mutations are present in a significant proportion of cells in one or multiple tissues, while post-mitotic mutations are limited to the single cell in which they occur. In mammalian neurogenesis, various types of mutations, such as SNVs, CNVs, long interspersed nuclear elements (LINE-1) insertions, and aneuploidy, are possible (Bae et al., 2018, Faulkner and Billon, 2018, Rohrback et al., 2018a). As neurons cannot be renewed, they tend to accumulate more DNA damage over an individual's lifetime compared to replicating cells. A neuron with private somatic mutations may survive for a considerable period if the damage is not severe enough to cause cell death (Rutten et al., 2007). Besides DNA damage, differential somatic mutation rates are also affected by variations in DNA repair (Supek and Lehner, 2019).

In most cases, DNA repair defects for example “DNA damage response” (DDR) triggered by DNA damage might be associated with neurological disorders (Coon and Benarroch, 2018, Iyama and Wilson, 2013, McKinnon, 2017). The types of DNA damage and repair that occur and are associated with neurodegeneration are shown in **Table 1.3**.

Table 1.3. DNA damage and repair in neurodegeneration arise from germline mutations (Proukakis, 2020). Abbreviations are Ataxia with Oculomotor Apraxia (AOA1, 4, and 5), Ataxia-telangiectasia (A-T), Cockayne Syndrome (CS) and Xeroderma Pigmentosum (XP).

Primary DNA break?	DNA lesion	Type of repair	Possible resulting mutation	Related Neurodegenerative disease
Yes	Single strand breaks (SSB)	Single strand breaks repair (SSBR)	Single nucleotide variant (SNV)	AOA1, SCAN1, AOA4, AOA5
	Double-strand breaks (DSBs)	Non-homologous end-joining (NHEJ)	Copy number variants (CNVs/SV, LINE-1 or gencDNA insertion	A-T, ATLD (familial ALS: FUS, TDP-43, C9ORF72, NEK1) (AOA2) (PD: possible GWAS associations)
		Homologous recombination (HR)	Copy number variants (CNV)/V	
No	Mismatch	Mismatch repair (MMR)	Further repeat expansion	(Repeat expansion disorders)
	Bulky adduct	Nucleotide excision repair (NER)	Single nucleotide variant (SNV)	CS, XP
	Base damage	Base-excision repair (BER)	Single nucleotide variant (SNV)	

There are two types of DNA damage: those that result from a primary "break" to the DNA phosphodiester backbone and those that do not, affecting one or both strands of DNA (Tiwari and Wilson, 2019) (**Table 1.3**). Double-strand breaks (DSBs) are the most harmful type of DNA damage in neurons (Alt and Schwer, 2018). Homologous recombination (HR) and non-homologous end-joining (NHEJ) are two ways to repair

DSBs. HR requires DNA synthesis and may not be possible in post-mitotic cells without aberrant cell cycle re-entry. NHEJ rejoins broken DNA ends directly, without requiring a nucleic acid template, and is more error-prone, often leading to deletions. DSBs may also have a physiological function in the normal biology of the brain, aligning with the concept of genomic mosaicism (Rohrback et al., 2018b, Weissman and Gage, 2016). In neurogenesis, DSBs were suspected of having a role after the requirement of DNA end-joining proteins was demonstrated (Gao et al., 1998), leading to the proposition that DSBs allow selection during apoptotic programmed cell death, making developing neurons susceptible to genomic rearrangements at a particular period (Chun and Schatz, 1999). Schwer et al. (2016) provided evidence that DSBs can impact neuronal genes during mouse neurodevelopment (Schwer et al., 2016), which aligns with the emergence of copy number variations (CNVs) in mouse neurons (Rohrback et al., 2018a).

Post-mitotically, DSBs may also occur, potentially playing a role in learning and memory (Madabhushi et al., 2015). DSBs can result in the generation of large copy number variants (CNVs) due to Topoisomerase 1 deletion, thereby preserving genomic stability during transcription doubling in the neurons (Fragola et al., 2020). Additionally, various other types of DNA damage can give rise to small-scale mutations, such as single nucleotide variants (SNVs) (discussed in the next section) (Table 1.2). Nucleotide excision repair (NER) repairs bulky lesions that distort the helix, while base-excision repair (BER) fixes base modifications caused by different processes, including oxidative stress. The correction of mismatched complementary bases is

carried out by the mismatch repair (MMR) mechanism, while primary single-strand breaks (SSBs) are repaired via a pathway similar to the base excision repair (BER) mechanism (Tiwari and Wilson, 2019).

1.4.2. Where and how should we look for relevant somatic mutations?

In both sporadic and inherited neurodegenerative disorders, somatic mutations are highly diverse in clinical and pathological features. In addition, mosaicism levels vary across brain regions and cell types, causing adverse effects in the cells carrying them (Proukakis, 2020).

Detecting somatic mutations requires methods that can detect them at low levels in a given tissue, and various mutation types may require different laboratory or bioinformatic analyses. The primary choice is "bulk" DNA sequence analysis extracted from "bulk" tissue, including data from glia and blood (D'Gama and Walsh, 2018). Although single-cell genomic analysis has revolutionized the ability to infer molecular changes within individual cells, there are still technical and analytical challenges, especially in the analysis of single-cell DNA sequencing data (Lahnemann et al., 2020, Rohrback et al., 2018b). The number of cells required to achieve the necessary sequencing depth can be determined through straightforward calculations, depending on the expected mosaicism (Davis et al., 2019). However, most studies have examined only a few brains with a relatively small number of cells. "Whole genome" amplification (WGA) results heavily depend on the technology used, and WGA cannot recover the entire genome, resulting in locus and allele dropouts. Single base changes and chimeric

structural variants can also cause positive FALS results. As a result, it is impossible to validate a variant present in a single cell orthogonally. Furthermore, the lack of standardized bioinformatic pipelines may result in variations in results from the same dataset, as shown in early control neuronal datasets (Garvin et al., 2015).

1.4.3. Evidence for somatic and *de novo* mutations in Neurodegenerative Diseases

New technologies hold the potential to assess somatic mutations in motor neurons (Blum and Gitler, 2022). An early study of cerebral cortex DNA did not find any somatic *SOD1* mutations which was the first paper specifically looking for somatic mutations in ALS (Shaw et al., 1997). *De novo* mutations occur in the germline and therefore do not result in somatic mosaicism although the basic mechanism is the same i.e. a new mutation which is replicated in all daughter cells. As a result, the observation of de novo mutations has been taken as evidence that somatic mutations may also occur. *De novo* mutations have been linked to various disorders such as autism and schizophrenia (Epi et al., 2013, Gratten et al., 2013, Neale et al., 2012, Sanders et al., 2012, Vissers et al., 2010, Xu et al., 2011), and have also been found in known ALS genes (Alexander et al., 2002, DeJesus-Hernandez et al., 2010, Chio et al., 2011, Conte et al., 2012, Zou et al., 2013, Calvo et al., 2014). By using exome sequencing on parent-child trios, investigators can search for de novo mutations across the entire genome, not just restricted to ALS genes. Recently, a study using this method identified *CREST* as a potentially new ALS gene. However, it's worth noting that individuals in the general population may carry up to four de novo coding mutations, which means that

identifying a de novo variant in a gene is not sufficient to prove its pathogenicity (Marangi and Traynor, 2015). Further genetic evidence is necessary to confirm that mutations in the designated gene are truly causative, which is currently lacking in published studies.

While there is no known causal link between somatic mutations and ALS, a study analyzing the whole-genome sequencing of peripheral blood leukocytes in monozygotic twins discordant with ALS found no significant somatic mutations (Meltz Steinberg et al., 2015). However, epigenetic modifications were recently reported (Young et al., 2017). In one sporadic case of Creutzfeldt–Jakob disease, a known disease-causing mutation was absent in the parents and appeared to be a somatic event, estimated to be present at a level of 97% in both peripheral blood leukocytes and brain tissue, suggesting that it occurred in the earliest postzygotic cell divisions and underwent positive selection (Alzualde et al., 2010). However, it is still possible that the mutation arose in the parental germline (Leija-Salazar et al., 2018).

Studies have shown that germline repeats can expand somatically in adult brain cells, providing support for the hypothesis that somatic repeat expansions may contribute to the prevalence of neurodegenerative diseases (Chong et al., 1995, Hubers et al., 2014, Malik et al., 2021, Kacher et al., 2021, Nordin et al., 2015). This is demonstrated by research conducted on various neurodegenerative diseases, including ALS (Beck et al., 2013, Lokanga et al., 2013), and Huntington's disease (Swami et al., 2009).

1.4.4. New insight regarding Next Generation Sequence analysis

The majority of the pathological abnormalities and clinical features in ALS are associated with damage to the corticospinal tract including upper and lower motor neurons with their cell bodies located within the motor cortex and anterior horn respectively (Eisen and Weber, 2001). However, the majority of genetic insights have come from sequencing DNA extracted from blood or cerebellum due to the accessibility and availability of sufficient DNA. We know that somatic heterogeneity is a feature of human genetics (Sturmberg, 2019) and specifically in the context of ALS genes (Buchman et al., 2013). Potentially, a new genetic mutation occurring within developing neurons of the motor cortex, termed somatic mutation, could adversely affect motor neuron function but would be undetectable in peripheral blood. To explore this possibility, whole genome sequencing was done for DNA extracted from the primary motor cortex of ALS patients where we already have sequencing data available from DNA extracted from blood and/or cerebellum. The Discovery of somatic heterogeneity would have important implications for the treatment and diagnosis of genetic forms of ALS.

1.5. Global variation of prevalence and incidence of ALS

Amyotrophic lateral sclerosis (ALS) is the third most frequent neurodegenerative disease after Alzheimer's and Parkinson's disease (Renton et al., 2014). ALS genetic

studies are more frequently available for developed countries in Europe and the USA compared to developing ones like India and Bangladesh.

Based on the ancestral origin, the incidence of ALS varies. The rate of incidence is higher in the European population (>3 cases per 100,000 individuals) compared to East and South Asia (Hardiman et al., 2017). The study by Chio et al. summarizes the incidence and prevalence of ALS cases in the total population by region that has been presented in **Table 1.4** where the rates are prospective in Europe (Ireland, Italy, Netherlands, Faroe Islands) but retrospective in Asia (Chiò et al., 2013).

Table 1.4. Country-specific crude prevalence (per 100,000) of ALS.

Region	Country	Date	Prevalence	References
EUROPE	Switzerland	30-Dec-90	3.9	(Huber and Henn, 1995)
	Yugoslavia	31-Dec-91	1.1	(Alčaz et al., 1996)
	Reggio Emilia, Italy	31-Dec-92	5.4	(Guidetti et al., 1996)
	Modena, Italy	31-Dec-99	4.0	(Mandrioli et al., 2003)
	Ireland	31-Dec-03	6.2	(O'Toole et al., 2008)
	Piemonte/Valle D-Aosta, Italy	31-Dec-04	7.9	(Chiò et al., 2009)
	Northern Ireland	30-Jun-05	4.9	(Donaghy et al., 2010)
	Northern Ireland & Rol	30-Jun-05	6.3	(Donaghy et al., 2009)
	England/Wales	30-Jun-06	4.9	(Abhinav et al., 2007)
	Norway	31-Dec-07	4.1	(Gundersen et al., 2011)
	The Netherlands	31-Dec-08	8.0	(Huisman et al., 2011)
	Faroe Islands	31-Dec-09	8.2	(Joensen, 2012)
	Modena, Italy	31-Dec-09	8.0	(Georgoulopoulou et al., 2011)
NORTH AMERICA	United States	31-Dec-02	4.2	(Turabelidze et al., 2008)
	United States	31-Dec-05	2.9	(Stickler et al., 2011)
ASIA/ PACIFIC	China	31-Dec-92	1.0	(Fong et al., 1996)
	China	31-Jan-02	3.1	(Fong et al., 2005)
	Japan	31-Dec-02	11.3	(Kihira et al., 2005)
	Isfahan, Iran	31-Mar-06	1.6	(Sajjadi et al., 2010)
SOUTH AMERICA	Uruguay	31-Dec-03	1.9	(Vazquez et al., 2008)

	Retrospective
	Prospective

All these studies show that there are differences in prevalence which are caused mostly by demographic composition differences (Chiò et al., 2013).

Previous epidemiological investigations of ALS have reported discrepancies in the prevalence and incidence of ALS among various populations, particularly in relation to age and gender (Xu et al., 2020). Other studies revealed regional and temporal variations for example in Norway the ALS prevalence increased from 3.67 per 1,00,000 population in 1988 (Tysnes et al., 1991) to 4.10 per 1,00,000 population in 2015 (Benjaminsen et al., 2018). Between 1978 and 1988, the incidence of ALS increased from 1.60 per 100,000 person-years (Tysnes et al., 1991) to 2.10 per 100,000 person-years between 2000 and 2015 (Benjaminsen et al., 2018). On the other hand, in Taiwan, the ALS prevalence and incidence in 2008 were 2.07 per 100,000 population and 0.53 per 100,000 person-years, respectively (Tsai et al., 2015).

ALS cases in India are estimated to be 5 in 100,000 by the Foundation for Research on Rare Diseases and Disorders [FRRDD] (Sahai and Ghosh, 2021). The survival rates observed among Indian ALS patients show similarities to those found in developing countries, especially in Africa (Nalini et al., 2008). In accordance with previous reports from India, compared to Europe and North America they have younger-onset ages, longer durations of symptoms, and slower disease progression (Sahai and Ghosh, 2021). In 2018, the first comprehensive genetic analysis in the ethnically diverse population of India was done that identified known pathogenic mutations along with 7 potentially pathogenic missense variants that have not been previously reported in ALS patients; this includes 3 novel variants (*OPTN*: p.Lys489Glu, *DAO*: p.Glu121Lys,

and *SETX*: p.Leu2163Val) that are not reported in large population databases and 4 rare variants (*CHMP2B*: p.Glu45Lys, *SQSTM1*: p.Gly262Arg and p.ro438Leu, *ERBB4*: p.Arg103His) with a minor allele frequency of <0.01 in large population databases (Narain et al., 2018). All known pathogenic, novel, and rare variants were detected in only 1 ALS patient each except for the *OPTN* (p.Lys489Glu) variant that was detected in 2 patients in that cohort (Narain et al., 2018). In 2016, Mukherjee et al. reported that *C9ORF72* hexanucleotide repeat expansions are rare in FTLDs in India (Mukherjee et al., 2016). *C9ORF72* pathogenic repeat expansions were not observed in ALS patients and healthy individuals from North India (Vats et al., 2017). However, a recent study found that 3.2% (19/593) of patients with ALS in India had a C9 expansion, of which 47.4% (9/19) were in the eastern region (Shamim et al., 2020). The common haplotype was found in 11 of the 19 cases (Shamim et al., 2020). Although pathogenic C9 expansions appear to be most prevalent in European ancestry, recent work supports the need for increased testing in more diverse populations (Nel et al., 2019, Shamim et al., 2020, Roggenbuck et al., 2020). Interestingly only a few studies are conducted on Bangladeshi MND patients, and almost all are demographic analyses rather than genetic studies (Raknuzzaman and Habib, 2020).

To interrogate the human genome for ALS genetic variation an independent large-scale whole genome research project named Project MinE has begun. At present, it is the largest genetic study for ALS, an international association that seeks to explore whole-genome sequence data of at least 15000 ALS patients and 7500 controls (Project MinE, 2018b). Many rare genetic variations absent in public databases have been found

through this project. The observation of Project MinE reflects the growing number of rare variants and population-specific variants as a result of sequencing larger samples around the globe such as identifying more novel genes like *NIPA1* (NIPA Magnesium Transporter 1) (Blauw et al., 2012, Tazelaar et al., 2019), *KIF5A* (Kinesin Family Member 5A) (Nicolas et al., 2018), *KANK1* (KN motif and ankyrin repeat domains 1) (Zhang et al., 2022) and other new risk loci (van Rheenen et al., 2021) associated with ALS.

Research has shown that the incidence of ALS is higher in European regions and lower in Asian regions (Xu et al., 2020). However, the prevalence in West Asia, including Cyprus and Israel, was much higher and second only to West Europe (Xu et al., 2020, Demetriou et al., 2017, Kahana et al., 1976). This may be due to the fact that West Asia is adjacent to Europe, suggesting a shared ancestry and more population movement (Taskent and Gokcumen, 2017). The modern humans who replaced Neanderthals in Europe likely originated from West Asia, and the genetic distances between West Asia and Europe are shorter than in other Asian regions (Bowcock et al., 1991, Cavalli-Sforza, 1997). The hexanucleotide GGGGCC repeat expansion in the *C9ORF72* gene, located in the 9p21.2 region, is the most frequent genetic cause of ALS in the white population (**section 1.3.2.1**) (Van Es et al., 2009, Diekstra et al., 2012a, Uyan et al., 2013, Fogh et al., 2014, Du et al., 2018, Shatunov et al., 2010, Balendra and Isaacs, 2018). However, this repeat expansion is relatively rare in Asia, and early population migration might explain this phenomenon (Balendra and Isaacs, 2018).

Interestingly, a few studies of other populations show significant differences. For example, in the case of *C9ORF72*, the expansions reported for FALS and SALS are correspondingly 39% and 7% in the European population while in China it is 3.5% for FALS and less than 1% for SALS patients. Moreover, a study from Japan identified the expansion of *C9ORF72* in 0.4% of SALS and 0% of FALS patients (**Figure 1.4**) (Saudagar et al., 2019, Wei et al., 2019). The repeat expansion has been detected in 18.2% of FALS and 2% of SALS patients in Taiwan (Tsai et al., 2012). Thus, it is important to understand the genetics of ALS in other populations, especially with the emergence of therapeutic strategies based on genetic mutation.

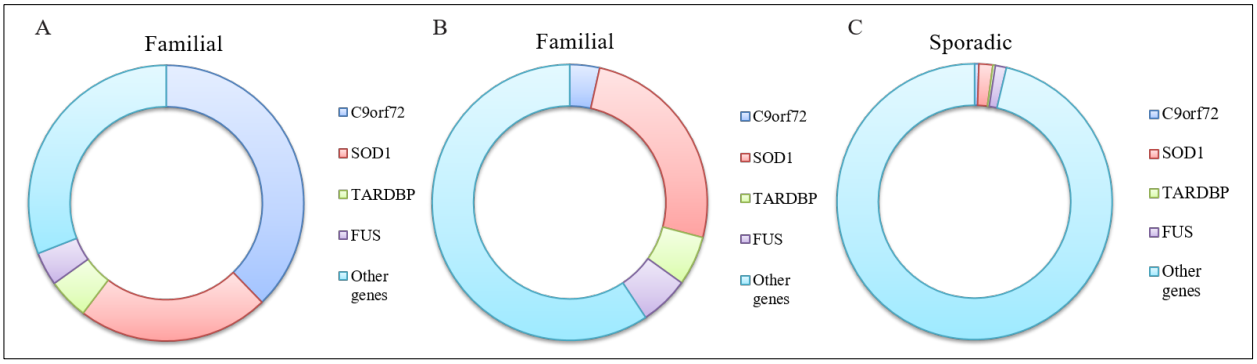


Figure 1.4. Comparison between the genetic architecture.

A. Percentage of the genes responsible for FALS in the European population (Alsultan et al., 2016). B. Percentage of the genes responsible for FALS in a Chinese population. C. Percentage of the genes responsible for SALS in the Chinese population (Wei et al., 2019).

1.5.1. Sex and age distribution of prevalence and incidence

In the case of heritability, another important matter in ALS is disease penetrance. It is defined as the frequency of individuals who develop a certain disease at a definite age as the individual carries a specified genotype. This issue is not well addressed. Being practically unsuitable for genetic studies, families with moderate and reduced disease penetrance remain unreported along with the gender ratio. It is unknown if some observed apparently sporadic ALS cases are actually familial but with reduced penetrance. The key difficulty is in the recruitment and longitudinal study of non-diseased individuals; this raises significant cost and ethical issues (Benatar and Wuu, 2012).

1.5.2. The spectrum of Neurological Disorders in Bangladesh

Over the last 50 years, Bangladesh has made great progress in healthcare. Life expectancy has risen from 47 to 72 years since Independence in 1971. Yet Bangladesh has remained on the Least Developed Countries list since 1975, and only reached the Lower Middle-Income Country threshold in 2015. With increased life expectancy, largely resulting from improved management of communicable disease and perinatal and maternal health, neurodegenerative disorders are now emerging as a major healthcare burden, reflecting demographic and epidemiologic transition. Stroke and other neurological disorders are now two of the top six causes of mortality and

morbidity (World Bank, 2013). The full impact of disabling, fatal conditions such as motor neurone disease (MND), especially in rural areas, remains unknown. Chronic, incurable neurodegenerative motor system disorders are increasing worldwide, and the prevalence of Parkinson's Disease (PD) is expected to double between 2005-2030 (Dorsey et al., 2007).

In Bangladesh, it is very common to note that neurological disorders are more prevalent than other diseases, such as headaches and migraine, stroke and seizures among different age groups and genders (Uddin et al., 2018). Still, the opportunities for genetic studies are very limited. A recent study showed that out of 1,684 patients with neurological disease genetic causes for the disorders were seen only among 12.35% of patients (Awan et al., 2019).

1.6. What is happening right now?

The usual medical expense for someone with ALS is anticipated at US$3436 per month in the USA (Oh et al., 2015). According to Gowland et al., it can be predicted that between the years 2015 and 2040, the number of persons living with ALS will increase by 20% in Europe (Gowland et al., 2019). This may lead to an increase of at least an additional US$95m in current UK healthcare costs (Gowland et al., 2019).

There is no specific diagnostic test available for ALS. Clinicians make the diagnosis of ALS when the patient has physical symptoms of the disease and the progression of these over time. A family history of ALS and/or FTD may also provide supporting evidence for diagnosis. Electrodiagnostic studies, neuroimaging and laboratory studies

are also included in the diagnosis of ALS. A novel amyotrophic lateral sclerosis (ALS) diagnostic index (ALSDI) has been developed to reliably differentiate ALS from mimicking disorders at an early stage in the disease course (Geevasinga et al., 2019).

The UK Genetic Testing Network (UKGTN, http://www.ukgtn.nhs.uk/gtn/Home) was established in 2001 and closed in 2018. At present, most of its functions moved to the Genomics Unit at National Health Service (NHS) in the UK. The clinical sensitivity of 'Familial Amyotrophic Lateral Sclerosis (FALS) with or without Frontotemporal Dementia 42 gene panel' test is estimated to be approximately 10% for sporadic cases and >40% in case of clear dominant family history while the specificity is close to 100%. The positive test result benefits developing prognostication supporting long-term care plans to be introduced.

A positive test result may play a role in initiating long-term care plans for ALS patients, particularly with the development of therapies targeted against specific mutations (Miller et al., 2022). Detection of pathogenic variants also facilitates genetic counselling and informing the family members about the risk for others and their offspring. The probability of developing symptoms of FTD may also be interpreted through genetic testing.

Despite extensive research, the complex molecular mechanism associated with ALS (section 1.3.1) is still not fully understood (Al-Chalabi and Hardiman, 2013, Morgan and Orrell, 2016, Brenner and Freischmidt, 2022). This has significantly challenged the search for a new biomarker for ALS which could inform diagnosis and prognosis in the clinic and could facilitate early intervention. There are different types of biomarkers

for measuring disease progression including lower motor neuron loss, upper motor neuron loss, imaging biomarkers, and wet biomarkers. All these techniques rely on equipment, expertise, time, and substantial resources. More fundamentally proposed biomarkers are downstream of motor neuron death which may limit their utility and specificity for the disease. So, a reliable and simple ideal biomarker is yet to be discovered. Several studies showed that phosphorylated neurofilament heavy subunit (pNfH) and neurofilament light chain (NfL) are promising biomarkers for ALS (Ganesalingam et al., 2013, Sturmey and Malaspina, 2022).

1.7. Current and future therapeutic interventions

1.7.1. Riluzole, Edaravone and Relyvrio

Riluzole, edaravone and Relyvrio are the only three drugs approved by the United States Food and Drug Administration for ALS patients as they showed a positive effect in ALS (Valko and Ciesla, 2019, Aschenbrenner, 2023). However, edaravone has not been licensed in the UK or Europe yet. Talampanel and tamoxifen are the other two drugs that are being investigated for having antiepileptic and neuroprotective properties respectively (Valko and Ciesla, 2019, Zhang et al., 2019, Wong et al., 2021).

Riluzole can increase 12-month survival by about 35% (Bensimon et al., 1994, Fang et al., 2018). Riluzole acts as an anti-glutamatergic agent using several mechanisms, for example, reducing glutamate release, preventing glutamate receptor hypofunction, activating glutamate transporters that increase glutamate uptake and reducing the release of excitotoxic glutamate (Kretschmer et al., 1998, Fumagalli et al., 2008, Wang

et al., 2004). The predominant mechanism of inhibiting the release of glutamate from nerve terminals, preventing “downstream”, resultant excitotoxic neuronal injury makes riluzole the most widespread neuroprotective treatment option for ALS (Calabrese et al., 2021). Additionally, riluzole acts as a sodium channel blocker. It reduces intracellular increases in sodium ion concentration as well as reverses the operation of axonal sodium calcium exchangers (Schwartz and Fehlings, 2002). It may prevent the disruption of the axonal sodium hydrogen antiporter system to preserve spinal cord white matter.

On the other hand, edaravone works as an antioxidant that slows the functional decline in ALS for an Iranian subgroup of 20 patients (10 cases, 10 controls) (Writing and Edaravone, 2017). As the first known free radical scavenger edaravone offers neuroprotection under oxidative stress (Cha and Kim, 2022, Mead et al., 2022).

On September 29, 2022, the U.S. Food and Drug Administration (FDA) approved the combination drug sodium phenylbutyrate–taurursodiol (Relyvrio) to treat ALS (Brown, 2022). The exact mechanism of action Relyvrio is not fully known. Sodium phenylbutyrate increases nitrogen excretion in the kidneys and affects protein folding to prevent misfolded proteins and reduce protein aggregation (Lowe, 2022). On the other hand, taurursodiol is a bile acid used for biliary cirrhosis. It also seems to affect the apoptosis pathways (Lowe, 2022). With this FDA approval, Relyvrio joins riluzole (Rilutek) and edaravone (Radicava) as therapeutic options for ALS patients (Aschenbrenner, 2023).

1.7.2. Stem cell therapy

The incomplete knowledge of the pathogenesis leads to the lack of therapeutic tools for ALS. Whilst it is not yet possible to regrow a new motor neuron in an adult, stem cell therapy is being used to tackle ALS pathogenesis through multiple mechanisms, for example introducing factors such as glial-derived neurotrophic factor (GDNF), brain-derived neurotrophic factor (BDNF), insulin-like growth factor-I (IGF-I), and vascular endothelial growth factor (VEGF) (Suzuki et al., 2008, Lunn et al., 2011) for neuroprotective effects, moderating the pathways resulting in a toxic environment (Faravelli et al., 2014). Because of the blood-brain barrier (BBB), there have been several difficulties to establish an efficient administration route for stem cell therapy (Meamar et al., 2013). Stem cells hold huge potential for the treatment of ALS but no stem cell-based therapy has been approved yet as more safety and efficacy trials are needed to be conducted in this field (Mazzini et al., 2018, Baloh et al., 2018). The path to stem cell-based therapy is still long and complex.

1.7.3. Antisense oligonucleotides (ASOs)

At present targeted gene therapies for ALS are also in development (Al-Chalabi and Brown, 2018). In the case of ALS caused by a genetic mutation, gene or allele-specific antisense oligonucleotides (ASOs) targeting is one of the effective methods (Klim et al., 2019). ASOs are single-stranded, short, synthetic molecules. They are designed to complement a target RNA and either sterically inhibit the binding of proteins associated with RNA maturation or translation or modify gene expression as the

transcript goes through the enzymatic cleavage. The endonuclease RNase H cleaves the target mRNA silencing the gene when oligo RNA duplexes are formed (Lima et al., 2007). This is an effective approach for ALS that is caused by toxin gain-of-function of either the RNA or protein because it selectively removes the mutant gene product but not the wild-type one where a genetic mutation is a reason (Klim et al., 2019). On the other hand, non-allele-specific ASOs are also playing an important role in ALS therapy in the broader patient base. Currently, three ASO-based therapies are in clinical trials with several new ones in progress for ALS (**Table 1.5**). Most of the ASOs activate RNase H cleavage, of either both wildtype and mutant transcripts or of the mutant transcript only. Their function is to moderate gain-of-toxic-function; *C9ORF72* RNA foci formation, *SOD1* or *C9ORF72* dipeptide repeat (DPR) polymers, or unusual recruitment of TDP-43 to stress granules by *ATXN2*. Auto-regulatory FUS protein binding is blocked by ASO through *FUS* splice-switching (Klim et al., 2019). A premature stop codon is produced in exon 8 by frameshifting which causes nonsense-mediated decay of *FUS* transcript and reduced FUS protein.

However, most of the ALS cases are sporadic with no identified 'mutant' target. Therefore, another route to using oligonucleotide-based therapies will be to regulate pathways associated with the disease pathogenesis of ALS. These pathways are also very common in other neurodegenerative diseases which offer an additional budget of the scale of synthesis along with the power to determine the efficacy. Currently, investigation of combining ASOs with presently available drugs may become safer long-term therapy.

Table 1.5. The selected ASO therapies are in trial or progress for ALS. (Klim et al., 2019, Fang et al., 2022)

Agent	ASOs Type	Target Gene	Trail Design	Site of Study	Status	Primary Outcome
BIIB067 or Tofersen (VALOR Trial)	Non-allele-specific	*SOD1*	Phase 3	USA, Canada, Europe	Complete (Extension of Phase III is still active)	N/A
ISIS 333611	Non-allele-specific	*SOD1*	Phase 1	USA	Complete	No AE, Well tolerated, dose-dependent CSF and plasma concentrations
BIIB078	Expanded repeat-specific	*C9ORF72*	Phase 1	USA, Canada, Europe	Complete	N/A
ION363	Non-allele-specific splice-switch ASO	*FUS*	Phase 1-3	USA, Canada, Belgium, UK	Active	N/A

1.8. Genetic profiling for prognosis and diagnosis

Being multifaceted and with numerous variants within each gene responsible for ALS, it has historically been both time-consuming and expensive to genetically screen all patients diagnosed with the disease. Shepheard and colleagues conducted targeted sequencing of genes relevant to ALS in 100 patients with ALS who were identified prospectively and attended a major ALS centre in Northern England (Shepheard et al., 2021). The study found genetic changes that could be clinically reported in 21% of patients, and 15 of these patients had *C9ORF72* or *SOD1* mutations that could make them eligible for participation in an ongoing genetic-therapy trial. As new genetic-therapy approaches for ALS are expected to be developed, the number of clinically

actionable results is likely to increase. Previous genetic studies of ALS have mostly been retrospective, making it impossible to determine the usefulness of genetic screening in clinical settings. However, this present study by Shepheard and colleagues strongly suggests that genetic testing should be routinely offered to patients with both familial and sporadic ALS.

Next-generation sequencing (NGS) shows promise for the diagnosis of both familial and sporadic ALS. Variants in *SOD1, TARDBP, C9ORF72,* and *FUS* genes are more likely to have a large effect in increasing the probability of ALS. Although gene variants with low penetrance may only slightly increase an individual's risk, they are still important to consider as they contribute to the overall genetic risk variance. Furthermore, studying these variants can enhance our understanding of the ALS pathway. NGS is not suitable to detect the most common genetic cause of ALS, *C9ORF72* repeat expansions (Morgan et al., 2015) or repeats in *ATXN2*. Therefore, the GGGGCC HRE mutations in *C9ORF72* and CAG repeat in *ATXN2* have been screened by repeat-primed PCR and standard fragment length analysis, respectively (Pulst et al., 1996, DeJesus-Hernandez et al., 2011). Pathogenic mutations in *SOD1, TARDBP,* and *FUS* mostly are found in patients with a family history (Morgan et al., 2017). A rare homozygous variant (*MFN2*) found in ALS patients in some cases described the rapid clinical deterioration as well as the early age at onset (Goldstein et al., 2019). **Table 1.6** shows the genetic factors affecting the prognosis of ALS.

Table 1.6. Genes as prognostic factors in ALS. (van Es et al., 2017)

Associated with long survival	Associated with short survival
• Mutations in *SOD1* • Reduced *EPHA4* expression	• Repeat expansions in *C9ORF72* or *ATXN2* • Mutations in *SOD1* • Homozygosity for the C allele of rs12608932 in *UNC13a* • Mutations in *FUS* (also associated with early onset)

Hence, genetic screening is important not only for developing the diagnosis of ALS but also predict the prognosis of the disease and better management required for ALS patients. A lack of understanding of somatic heterogeneity means that patients likely exist who carry an ALS-associated mutation within the CNS but they go unrecognised because the mutation is absent in peripheral blood. These patients may therefore miss out on potential benefits and treatments which are specific to recognised genetic cases of ALS.

1.9. Monomelic Amyotrophy

Monomelic Amyotrophy (MMA) is a rare, inherited motor neuron disorder that results in weakness and wasting of specific muscles, usually in the arm or leg. The genetics of MMA is complex and not well understood, but multiple factors have been implicated in the development of the disease. Some forms of MMA have been linked to mutations in specific genes, including the *HSPB1* gene, which codes for a heat shock protein, and the

ALS2 gene, which is involved in the regulation of nerve cell function. Other forms of MMA have been associated with genetic changes in chromosomes, including deletions and duplications. In addition to these specific genetic causes, MMA can also be caused by more complex genetic interactions, such as those involving multiple genes and environmental factors. Further research is needed to fully understand the genetics of MMA and to identify new therapeutic targets for the treatment of this debilitating disorder.

Given that MMA has a less severe prognosis than ALS, the clinical overlap between these two diseases has led to the idea that they may represent the same pathogenic process and that ALS could be therapeutically converted into MMA. This makes it important to understand the extent to which the molecular mechanisms of these two diseases overlap.

1.10. Aims and objectives

The overall aim of this thesis is to explore sources of genotype-phenotype heterogeneity in MND. Investigating somatic mutations in MND can help uncover the molecular mechanisms driving disease progression and identify potential therapeutic targets. Somatic mutations also contribute to a better understanding of MND heterogeneity, accounting for variations in disease onset, severity, and progression. By studying these mutations, researchers can unravel the genetic landscape of MND. Although the global population is comprised of over 50% Asians, there are few epidemiological studies on motor neuron diseases conducted beyond Europe and

North America (Shahrizaila et al., 2016). There is evidence suggesting that MND genetic architecture may differ among different ethnic groups. Understanding the genetic factors specific to the Asian population can provide insights into the unique susceptibility risk variants and pathways contributing to MND in this population. Moreover, genetic research into MND can contribute to global efforts to understand the disease as a whole. A large number of the Bangladeshi samples collected came from patients with MMA. The MMA samples have variants in the traditional genes for various motor neuron disorders which leads us to the last part of this study which involves finding out the genetic architectural similarities between MMA and MND/ALS.

The aims of this study can be summarised as:

a. Project MinE motor cortex analysis of ALS gene
    - To establish if somatic mutations occur in the motor cortex compared to the blood of the ALS patients of Project MinE.
    - To establish if there are differences between pathogenic mutations found in the motor cortex and blood or cerebellum samples.
b. Screening MND cases from Bangladesh
    - To establish the genetic component contributing to disease in individuals with
        i. Amyotrophic lateral sclerosis (ALS),
        ii. Monomelic amyotrophy (MMA),
        iii. Other motor system disorder variants

- To compare MMA patients from Bangladesh and the UK for mutations in 153 genes associated with motor system disorders to determine whether MMA shares a genetic architecture with other disorders including MND.
- To screen for TDP-43 proteinopathy by western blot and direct conversion of MMA patients fibroblast to induced neural stem/progenitor cells (NPCs).

Chapter 2. Methods and materials

2.1. Genetic analysis of WGS of ALS patients from Project MinE

2.1.1. Subjects

From Project MinE (https://www.projectmine.com/) whole genome sequencing has been completed on a set of 69 DNA samples extracted from the post-mortem motor cortex of ALS patients. I used these data for targeted analysis of ALS-associated genetic variation. Post-mortem motor cortex tissue was obtained from the Sheffield Brain Tissue Bank (Request 18/001), which is governed by REC 08/MRE00/103+5.

2.1.2. Motor cortex tissue sample collection for whole genome sequencing (WGS)

2.1.2.1. UK DNA Bank for Motor Neuron Disease (MND)

The Motor Neurone Disease Association (MNDA) and the Wellcome Trust created the collaborative project, UK MND DNA Bank that was adopted from a 'hub and spoke' model (Smith et al., 2015). Three regional Hub Centres were established that linked 16 'spoke centres' spread across England, Scotland, Wales and Northern Ireland (Smith et al., 2015). The hub centres were King's College Hospital in London, Queen Elizabeth Hospital in Birmingham and Royal Hallamshire Hospital in Sheffield (Smith et al., 2015). The DNA Bank has over 3000 high-quality DNA samples donated by people

living with MND/ALS (including their clinical phenotype data), family members and non-related controls (Smith et al., 2015). In 2014, these samples were submitted for sequencing to Project MinE, an international collaboration to sequence ALS patients from across the world (MinE).

2.1.2.2. The Sheffield Brain Tissue Bank (SBTB)

The Sheffield Brain Tissue Bank (SBTB) has a major local user community within the Sheffield Institute for Translational Neuroscience (SITraN) (HRA). It also supports national UK neuroscience research as part of the Medical Research Council (MRC) Brain Bank Network (HRA). It holds more than 800 brains donated from diseased patients mainly of degenerative disease (MND, Alzheimer's) (SITraN, SBTB). According to the data collection arrangements the tissue samples are stored in the secure joint Sheffield Teaching Hospitals/University of Sheffield Biorepository facilities within the Royal Hallamshire Hospital (RHH) accommodation (HRA). Donors are transported to the RHH mortuary and the tissues are removed by practised Neuropathologists supported by Anatomical Pathology Technicians (HRA). University computers and secure servers hold the Brain bank data (HRA). The clinical team is based in the SITraN and the Neurology Department of the RHH (HRA). The family consent forms are taken either in RHH accommodation, at home or via telephonic communication (HRA).

2.1.2.3. Sample collection and storage

The post-mortem samples of the motor cortex used in this study were from patients with either sporadic ALS or familial ALS. The tissues were snap-frozen and stored as unfixed frozen brain slabs at -80 °C. DNA was extracted by Matt Wyles from SITraN and the samples were sent to Project MinE. A total of 69 samples passed the quality control (QC) and were considered suitable for next-generation sequencing.

2.1.2.4. DNA Extraction and Sequencing

Matt Wyles extracted the genomic DNA using AllPrep DNA/RNA Kit from samples of the snap-frozen motor cortex that had been stored at −80°C. This kit uses a guanidinium thiocyanate-based lysis with column purification and retrieves RNA fragments >200 nucleotides. The following flowchart shows a summary of the procedure of DNA extraction.

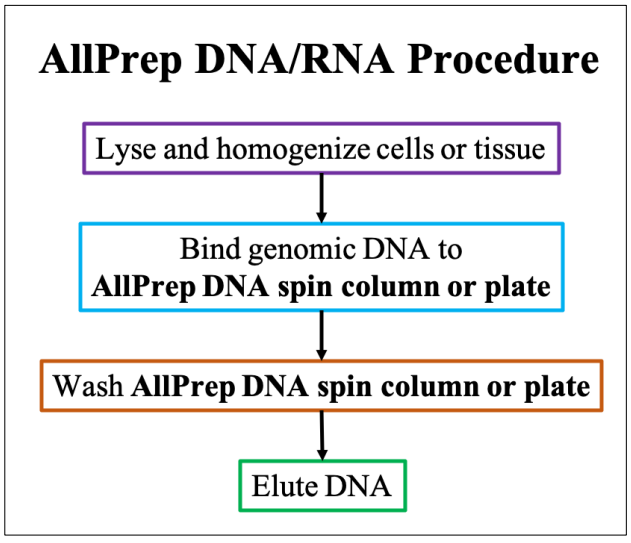


Figure 2.1. Flowchart of DNA extraction from motor cortex samples.

The frozen tissue was lysed with guanidine-isothiocyanate–containing Buffer RLT Plus in a 2.0mL microcentrifuge tube and homogenized by using TissueLyser II (Qiagen) with 5mm stainless steel beads. Tissue lysate was continued with the AllPrep protocol for the simultaneous extraction of genomic DNA using RNeasy Mini spin column technology. DNA concentration and purity (A260/A280 ratio) were measured using the QIAxpert System (Qiagen).

In conventional bulk sequencing, genetic material is mixed across cells and report average readouts such that genetically distinct subpopulations (e.g. subclones) cannot be identified. Genotypic measurements (e.g. variant allele frequencies, VAF) therefore represent a population average. Single-cell data is very sparse per cell and relies on inference between cells with similar properties.

2.1.2.5. Whole-genome sequencing

For sequencing 150 microliters of DNA at a concentration >30ng/microliter of each sample were sent for sequencing using Illumina (San Diego, USA). Using PCR-free library preparation and paired-end (100bp) sequencing on the HiSeq 2500 platform (Illumina®, San Diego, Illumina) to yield ~40X coverage (Van Rheenen et al., 2018). The Isaac pipeline (Raczy et al., 2013) was used to align the reads with the hg19 reference genome in addition to calling single nucleotide variants (SNVs), insertions and deletions (indels), and larger structural variants (SVs). The output of the aligned and unaligned reads was in binary sequence alignment/map format (BAM) along with variant call format (gVCF) files containing the SNVs, indels and SVs (Raczy et al., 2013).

gVCF files were generated per individual (Van Rheenen et al., 2018). Variants that failed the Isaac-based quality filter were set to missing on an individual basis (Van Rheenen et al., 2018).

2.1.2.6. Quality control of Sequencing Data (QC)

The data underwent quality control (QC) at individual and variant levels (Van Rheenen et al., 2018). The total set of genomic variant call format (gVCF) files obtained from sequencing the samples was merged by converting them to Plink format and then merging them, resulting in a single dataset containing all variants across all individuals. Non-autosomal chromosomes and multi-allelic variants were excluded based on pilot analyses. Sample and single nucleotide polymorphism (SNP) QC were performed using PLINK (Purcell et al., 2007, Chang et al., 2015) and VCFtools (Danecek et al., 2011).

No dataset is flawless, and as the number of samples and variants increases, missing data becomes more likely. To begin the process of sample QC, the missingness of each sample was assessed on a per-chromosome basis, and all samples had missingness < 10% across all 22 chromosomes. No samples were excluded at this stage (Van Rheenen et al., 2018). Further sample QC was conducted on a subset of high-quality biallelic SNPs with specific criteria: minor allele frequency (MAF) > 10%, missingness < 0.1%, LD-pruned at an r2 threshold of 0.2, not A/T or C/G SNPs, not located in the major histocompatibility complex (MHC) or LCT (Lactase) locus, and not present in the inversions on chromosome 8 or chromosome 17 (Van Rheenen et al., 2018, Cooper-Knock et al., 2021b). It is important to note that linkage disequilibrium (LD), which

measures the squared correlation between phased alleles (r2), can only reach a value of 1 when both loci have the same count of the minor allele (Calus and Vandenplas, 2018).

Principal components were calculated using ~30000 SNPs from the overlap between the set of high-quality biallelic SNPs and HapMap 3 (HM3), the third phase of the International HapMap project (Gibbs et al., 2003, Thorisson et al., 2005). ALS cases and controls were projected onto the HM3 samples to ensure an ancestrally homogeneous group of samples for association testing. Non-European samples, defined as those further than 10 standard deviations from the European-ancestry populations in HM3 (CEU, people of Northern and Western European ancestry living in Utah; TSI, Tuscans in Italy), were excluded from the analysis.

Samples that had an inbreeding coefficient greater than 3 standard deviations from the mean of the distribution were excluded from the analysis. This step was taken because relatedness between samples could potentially cause an artificial overrepresentation of certain rare genetic variants. Additionally, samples with discordant sex information, as determined by comparing chromosome X genotypes with the phenotype information, were also excluded.

To ensure the high quality of variants, several QC steps were applied. Variants with missingness > 5% were removed, as well as variants that violated Hardy-Weinberg equilibrium in controls ($p < 1 \times 10\text{-}6$) or were monomorphic (due to sample exclusions). Variants with significant differential missingness between cases and controls ($p < 1 \times 10\text{-}6$) and those with extreme depth of coverage (> 6 s.d. from the

mean of the total depth distribution) were also eliminated (Van Rheenen et al., 2018).. The analysis excluded mitochondrial, X, and Y chromosomes. However, some samples were lost during variant QC, and the final data set consisted of 69 patients.

2.1.3. Data analysis

2.1.3.1. BAM Files of Whole Genome Sequencing

The Single Nucleotide Polymorphisms (SNPs) data from whole-genome sequencing has been received in Binary Alignment/Map (BAM) format which is a binary format for storing sequence data. BAM files keep the same information as human-readable text format Sequence Alignment/Map (SAM) and show alignments but in a compressed format that is only machine readable.

2.1.3.2. BAM file to vcf file

BAM files were converted to vcf format and divided per chromosome using plink (Slifer, 2018).

2.1.3.3. Galaxy (SnpEff eff: annotate variants)

The public Galaxy online data platform (homepage: https://galaxyproject.org, main public server: https://usegalaxy.org) was used for the bioinformatics analysis (Afgan et al., 2018). The vcf files were uploaded in Galaxy for variant annotation.

SnpEff is one of the genetic variant annotation and functional prediction toolboxes available in Galaxy Version 4.3+T.galaxy1 (https://usegalaxy.eu/) which is a web-based platform for reproducible computational analysis. Using SnpEff we can annotate and predict the effects of genetic variants (such as amino acid changes) (Cingolani et al., 2012). Genetic variant means the difference between a genome and a "reference" genome. For our data analysis, the reference genome is “Home sapiens: hg19”. The annotated mutation vcf file contains the number of high, moderate, or low modifier effects by impact and by functional class, such as whether a variant is missense, nonsense or silent (**Table 2.1**). Typically, variants are categorized as follows:

SNP (Single-Nucleotide Polymorphism) Reference = 'A', Sample = 'G'

Ins (Insertion) Reference = 'A', Sample = 'ATG'

Del (Deletion) Reference = 'AC', Sample = 'A'

MNP (Multiple-nucleotide polymorphism) Reference = 'ATA', Sample = 'CTG'

MIXED (Multiple-nucleotide and an InDel) Reference = 'ATA', Sample = 'CTGAGT'

And the impact levels are pre-defined categories based on the ‘Effect’ of the variant, to help users find more significant variants making it easier to categorize and prioritize. **Table 2.1** shows the meaning of different impacts that are predicted in an annotated file.

Table 2.1. SnpEff Impact Prediction (Copied from https://pcingola.github.io/SnpEff/)

Impact	Meaning	Example
HIGH	The variant is assumed to have a high (disruptive) impact on the protein, probably causing protein truncation, loss of function or triggering nonsense-mediated decay.	stop_gained, frameshift_variant
MODERATE	A non-disruptive variant that might change protein effectiveness.	missense_variant, inframe_deletion
LOW	Assumed to be mostly harmless or unlikely to change protein behaviour.	synonymous_variant
MODIFIER	Usually, non-coding variants or variants affect non-coding genes, where predictions are difficult or there is no evidence of impact.	exon_variant, downstream_gene_variant

2.1.3.4. Variant Analysis Pipeline

2.1.3.4.1. Tools used

2.1.3.4.1.1. iTabixit

The vcf files of the variant calling generated from SnpEff analysis were downloaded and compressed using software called Tabix.

2.1.3.4.1.2. RStudio: Code verification for single gene

a. Part 1: Sorting out specific types of variants

Initially, RStudio Version 1.2.1335 (RStudio, Inc. https://www.rstudio.com/products/RStudio/) was used to verify the code of a single gene. The *SOD1* gene was selected to write the pipeline of the code. The full coding for the processing of *SOD1* is provided in the **Appendix B** and the key commands are explained in section **Appendix B iv** to make clear the command line.

The genomic locations in **Table 2.2** were used to determine the location of the gene and to only analyse that particular region of the chromosome to avoid identifying variants in any unrelated MND gene sequences.

Table 2.2. The ALS-related Gene location in chromosome and size (bases).

Chr	Gene	Loci	Chr Location	Genomic Locations (GRCh37/hg19)	Size (bases)
1	*TARDBP*	ALS10	1p36.22	chr1:11,072,414-11,085,796	13,383
2	*ALS2*	ALS2	2q33.1	chr2:202,564,986-202,645,912	80,636
2	*TUBA4A*	ALS22	2q35	chr2:220,114,433-220,142,892	28,460
2	*DCTN1*		2p13.1	chr2:74,588,281-74,619,214	30,934
2	*CYP27A1*		2q35	chr2:219,646,472-219,680,016	33,545
2	*ERBB4*	ALS19	2q34	chr2:212,240,442-213,403,565	1,163,120
2	*SPAST*		2p22.3	chr2:32,288,680-32,382,706	94,027
2	*VPS54*		2p14	chr2:64,119,280-64,246,214	126,927
3	*CHMP2B*	ALS17	3p11.2	chr3:87,276,413-87,304,698	28,278
4	*NEK1*	ALS24	4q33	chr4:170,314,421-170,533,780	219,355
5	*MATR3*	ALS21	5q31.2	chr5:138,609,441-138,667,366	57,920
5	*ARHGEF28*		5q13.2	chr5: 72,921,983-73,237,818	315,836
5	*SQSTM1*	FTDALS3	5q35.3	chr5:179,233,388-179,265,078	31,691

6	*FIG4*	ALS11	6q21	chr6:110,012,424-110,146,634	134,133
7	*HNRNPA2B1*		7p15.2	chr7:26,229,547-26,241,149	11,603
9	*C9ORF72*	FTDALS1	9p21.2	chr9:27,546,544-27,573,864	27,321
9	*VCP*	ALS14	9p13.3	chr9:35,056,061-35,073,246	17,186
9	*SETX*	ALS4	9q34.13	chr9:135,136,743-135,230,372	93,630
9	*GBA2*		9p13.3	chr9:35,736,863-35,749,983	13,121
9	*SIGMAR1*	ALS16	9p13.3	chr9:34,634,719-34,637,806	3,088
10	*OPTN*	ALS12	10p13	chr10:13,141,449-13,180,291	38,843
10	*ANXA11*	ALS23	10q22.3	chr10: 81,910,645-81,965,328	54,684
12	*TBK1*	FTDALS4	12q14.2	chr12:64,845,660-64,895,899	50,229
12	*KIF5A*	ALS25	12q13.3	chr12:57,943,781-57,980,415	36,635
12	*HNRNPA1*	ALS20	12q13.13	chr12:54,673,977-54,680,872	6,896
12	*ATXN2*		12q24.12	chr12: 111,890,018-112,037,480	147,463
12	*DAO*		12q24.11	chr12:109,252,708-109,294,819	42,112
12	*PRPH*		12q13.12	chr12:49,687,035-49,692,481	5,431
13	*SPG20/SPART*		13q13.3	chr13: 36,875,775-36,944,317	68,543
14	*ANG*	ALS9	14q11.2	chr14: 21,152,336-21,167,130	14,795
14	*VRK1*		14q32.2	chr14:97,263,641-97,398,059	134,419
15	*SPG11*	ALS5	15q21.1	chr15:44,854,894-44,955,876	100,983
16	*FUS*	ALS6	16p11.2	chr16:31,191,431-31,206,192	11,697
17	*PFN1*	ALS18	17p13.2	chr17:4,848,947-4,852,356	3,410
17	*GRN*		17q21.31	chr17:42,422,491-42,430,470	7,857
17	*MAPT*		17q21.31	chr17:43,971,748-44,105,700	133,953
17	*TAF15*		17q12	chr17:34,136,459-34,191,619	37,759
20	*VAPB*	ALS8	20q13. 32	chr20:56,964,175-57,026,157	61,980
20	*SS18L1*		20q13.33	chr20:60,718,822-60,757,566	38,719
21	*SOD1*	ALS1	21q22.11	chr21:33,031,935-33,041,244	9,310
22	*NEFH*		22q12.2	chr22:29,876,181-29,887,379	11,161
22	*CHCHD10*	FTDALS2	22q11.23	chr22:24,108,021-24,110,630	2,610
22	*EWSR1*		22q12.2	chr22:29,663,998-29,696,515	32,518

b. Part 2: Extract_clinvar

The variants of specific known 42 genes (Table 2.2) were extracted from ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/).

The variants reported in ClinVar are compared with the variant-specific csv file of 69 patients (Section 2.1.3.4.1.2.a) and saved as a csv file.

The interpretation of sequence variants is guided by the American College of Medical Genetics and Genomics (ACMG) (Richards et al., 2008, Fleischer and Lockwood, 2014), which recommends the use of specific terminology, including pathogenic, likely pathogenic, uncertain significance, likely benign, and benign (Richards et al., 2015). The criteria consist of two sets, one for the classification of Pathogenic or Likely Pathogenic variants (**Table 2.3**) and another for the classification of Benign or Likely Benign variants (**Table 2.4**) (Richards et al., 2015). Pathogenic criteria are assigned a weight of very strong (PVS1), strong (PS1-4), moderate (PM1-6), or supporting (PP1-5), while benign criteria are assigned a weight of stand-alone (BA1), strong (BS1-4), or supporting (BP1-6) (Richards et al., 2015). The numbering system within each category is only used for reference purposes and does not reflect differences in weight (Richards et al., 2015).

Table 2.3. Pathogenic Variants Classification Criteria (copied from (Richards et al., 2015)) (Arbustini et al., 2022).

Very strong evidence of pathogenicity		
PVS1	Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease	
Strong evidence of pathogenicity		
PS1	Same amino acid change as a previously established pathogenic variant regardless of nucleotide change	
	Example:	Val->Leu caused by either G>C or G>T in the same codon
PS2	*De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history	
PS3	Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product	
PS4	The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls	
Moderate evidence of pathogenicity		
PM1	Located in a mutational hot spot and/or critical and well-established functional domain (*e.g*. active site of an enzyme) without benign variation	
PM2	Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes or ExAC	
PM3	For recessive disorders, detected in *trans* with a pathogenic variant	
PM4	Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants	
PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been reported before	
	Example: p.Arg156His is pathogenic; now you observe p.Arg156Cys	
PM6	Assumed *de novo*, but without confirmation of paternity and maternity	

Supporting evidence of pathogenicity	
PP1	Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease
PP2	Missense variant in a gene that has a low rate of benign missense variation in which missense variants are a common mechanism of disease
PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc)
PP4	Patient's phenotype or family history is highly specific for a disease with a single genetic aetiology
PP5	Reputable source recently reports variant as pathogenic but the evidence is not available to the laboratory to perform an independent evaluation

Table 2.4. Benign Variants Classification Criteria (copied from (Richards et al., 2015)) (Arbustini et al., 2022).

Stand-Alone evidence of benign impact	
BA1	Allele frequency above 5% in Exome Sequencing Project, 1000 Genomes, or ExAC
Strong evidence of benign impact	
BS1	Allele frequency greater than expected for the disorder
BS2	Observed in healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder with full penetrance expected at an early age
BS3	Well-established functional studies (*in vitro* or *in vivo*) shows no damaging effect on protein function or splicing
BS4	Shortage of segregation in affected members of a family
Supporting evidence of benign impact	
BP1	Missense variant in a gene resulting truncating variants are known to cause disease
BP2	Observed in *cis* with a pathogenic variant in any inheritance pattern or observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder
BP3	In-frame deletions/insertions in a repetitive region without a known function
BP4	Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc)
BP5	Variant found in a case with an alternate molecular basis for disease
BP6	Reputable source recently reports variant as benign but the evidence is not available to the laboratory to perform an independent evaluation
BP7	A synonymous variant (silent) for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

c. Part 3: Identifying the patient's ID

From the whole chromosome .vcf.gz file where the gene is present, the patient's ID are sorted out when a variant is found to have significant importance (e.g. found by comparing with ClinVar data in section 2.1.3.4.2.b)

2.1.3.5. Rank on CADD score

CADD (Combined Annotation Dependent Depletion) is a tool used in genomics research to predict the deleteriousness of genetic variants. The CADD score is a metric that takes into account a variety of factors, such as functional annotations, conservation, and evolutionary constraints, to provide a score that reflects the likelihood that a particular genetic variant is deleterious. It is widely used for scoring the deleteriousness of single nucleotide variants as well as insertion/deletions (InDels) variants in the human genome that can effectively prioritize causal variants in genetic analyses, particularly highly penetrant contributors to severe Mendelian disorders (Rentzsch et al., 2019). A CADD score can quantitatively prioritize functional, deleterious, and disease-causal variants across a wide range of functional categories, effect sizes and genetic architectures and can be used to prioritize causal variation in both research and clinical settings. Because it correlates variants with annotations of functionality, allelic diversity, pathogenicity, disease severity, experimentally measured regulatory effects and complex trait associations, and they highly rank known pathogenic variants within individual genomes (Kircher et al., 2014a). CADD scores, software and documentation are available at: https://cadd.gs.washington.edu. The CADD score was determined for

each variant that was classified in ClinVar as conflicting interpretation of pathogenicity or uncertain significance to prioritize these according to the CADD score (Rentzsch et al., 2019). Those with a CADD score >15 were considered to be reported for further analysis.

A CADD score of 15 is commonly used as a threshold for filtering genetic variants in research studies. This threshold is based on the fact that variants with CADD scores greater than 15 are considered to be among the top 10% of most deleterious variants in the human genome. Variants with CADD scores greater than 15 are therefore more likely to be associated with a disease or other phenotypic effects and are thus more likely to be of interest in research studies.

It's important to note that the choice of CADD score threshold may vary depending on the specific research question and the population being studied. Some studies may use a higher or lower threshold depending on their specific needs. Additionally, it's important to use multiple lines of evidence, such as functional studies and population genetics, to validate the significance of any variant identified through genetic analysis.

2.1.3.6. Allele frequency

The allele frequencies of the reported variants from the 69 motor cortex samples were extracted from Project MinE (https://www.projectmine.com/) and the Genome Aggregation Database (https://gnomad.broadinstitute.org/) containing whole genome sequencing (ge) and whole exome sequencing (ex) and evaluated in comparison with the ClinVar information. Developed by an international coalition of investigators,

gnomAD has a goal of aggregating and harmonizing both exome and genome sequencing data from a wide variety of large-scale sequencing projects. This database is also making summary data available to the wider scientific community. Both versions v2.1.1 (125,748 WES and 15,708 WGS samples from unrelated individuals) and v3.1.2 (76,156 samples of diverse ancestries) were checked to get the frequencies of the variants.

2.1.3.7. *C9ORF72*

C9ORF72 is the most commonly known gene associated with ALS, explaining about 40% of familial cases and 8% of sporadic (Konno et al., 2013) in European and American populations. This most frequent genetic cause falls into the category of hexanucleotide repeat expansion mutation (HREM) which involves a hexanucleotide GGGGCC repeat expansion in the *C9ORF72* gene. As a result, the SNP data of the 69 motor cortex samples were not ideal for analysing the *C9ORF72* gene mutation. Screening of the HRE in DNA extracted from cerebellar samples from this sample set has been described previously (Cooper-Knock et al., 2012).

2.2. Genetic Screening of MND patients from Bangladesh

2.2.1. Subjects

Thirty-three clinically definite MND patients from the National Institute of Neurosciences & Hospital, Bangladesh (Org Code: 10017210) enrolled in this study. From each MND patient, peripheral blood samples were collected.

2.2.2. Ethics, consent, and permission

This study was approved by the University of Sheffield (TUoS) and the National Institute of Neurosciences & Hospital (NINS&H), Bangladesh. Documents such as the protocol, participant information sheets and consent forms used by the clinical team in Sheffield (under NHS ethics approval) were translated into Bengali and approved by the NINS&H Ethics Committee. Each participant in the study signed an informed consent form before enrolment. The collection of this data has been viewed to be ethically robust and the application (no. 025945) for UoS ethics approval has been approved too. The gender and diagnosis of the patient samples are listed in table 2.5. The sample collection of MND patients from Bangladesh did not involve a targeted population. All available MND patients were included in the study during the sample collection process.

Table 2.5. The list of Bangladeshi MND patient samples selected for DNA sequencing. (BD-MND = Bangladesh MND sample; ALS = amyotrophic lateral sclerosis; MMA = Monomelic; PMA = primary muscular atrophy). The empty boxes represent information that is not known and the bold-marked ones are MMA samples.

Sample ID	Gender	Diagnosis	Age
BD-MND-001	Male	ALS	32
BD-MND-002	Male	MMA- leg	38
BD-MND-003	Female	PMA	36
BD-MND-004	Female	ALS	45
BD-MND-005	Male	ALS	44
BD-MND-006	**Male**	**MMA-arm/Hirayama**	**27**
BD-MND-007	Male	ALS	40
BD-MND-008	Male	ALS	40
BD-MND-009	**Male**	**MMA-arm/Hirayama**	**35**
BD-MND-010	**Male**	**MMA-arm/Hirayama**	**25**
BD-MND-011	**Male**	**MMA-arm/Hirayama**	**31**
BD-MND-012	Male	MMA - leg	43
BD-MND-013	Female		
BD-MND-014			
BD-MND-015	Female	ALS	52
BD-MND-016	Male	ALS	30
BD-MND-017	Male	ALS	50
BD-MND-018			
BD-MND-019	**Male**	**MMA-arm/Hirayama**	**30**
BD-MND-020	**Male**	**MMA-arm/Hirayama**	**33**
BD-MND-021	Male	ALS	37
BD-MND-022	Male	ALS	43
BD-MND-023	Male	PMA	47
BD-MND-024	**Male**	**MMA-arm/Hirayama**	**33**
BD-MND-025	Female	ALS	39
BD-MND-026	Male	ALS	61
BD-MND-027	Male	PMA	40
BD-MND-028			
BD-MND-029	**Male**	**MMA-arm/Hirayama**	**21**
BD-MND-030	**Male**	**MMA-arm/Hirayama**	**35**
BD-MND-031			
BD-MND-032	**Male**	**MMA-arm/Hirayama**	**20**
BD-MND-033	**Male**	**MMA-arm/Hirayama**	**20**

The observed higher prevalence of disease onset at a younger age is consistent with previous findings in the Asian population (Sahai and Ghosh, 2021). However, due to the absence of a prospective cohort, it was not possible to perform a comprehensive analysis of the age of onset in the Bangladesh population. Consequently, there may be some selection bias present.

All patients diagnosed locally with MMA were recruited for this study. As for MND, this was not a comprehensive prospective recruitment and it is impossible to know whether some selection bias may have been present.

2.2.3. The cohort of MMA with Arm onset samples

Out of all the patients observed, only three had a potential family history of a motor neuron disorder, and notably, they all exhibited monomelic phenotypes. Specifically, the BD-MND-009 patient's father had experienced progressive wasting in the arm for 15 years, while the BD-MND-010 patient's brother had arm wasting and displayed a monomelic arm phenotype with Hirayama-type radiological changes. Additionally, the BD-MND-030 patient's father had slowly progressive arm wasting, and they also displayed a monomelic arm phenotype. Consanguinity was not detected in any of these cases, which sparked our curiosity to investigate the BD MMA cohort.

Eleven Bangladesh (BD) patients among 28 MND patients were diagnosed with MMA with arm onset disorder. Genetic screening results from eight UK MMA with arm onset patients (**Bold in Table 2.6**) available (previously analysed by Sophie Cadden) were included in this study to make a larger cohort of MMA samples and to also compare the

genetics of UK and Bangladesh MMA patients. In order to ensure consistency, the UK MMA results were re-analysed using the same pathway as that used for the Bangladesh MMA patients.

Table 2.6. Summary of the 8 UK participants with monomelic amyotrophy (MMA).

Patient	Gender	Age	Duration of MA (months)
SHF-067	Male	57	81
SHF-070	Male	33	138
SHF-072	Male	48	100
SHF-073	Male	24	72
SHF-089	Male	35	204
SHF-091	Male	80	140
SHF-099	Male	19	24
SHF-104	Male	48	260

2.2.4. Collection and processing of blood specimens

Two sets of peripheral blood samples (~100 ml each set) were obtained by venepuncture from each patient. Blood samples were collected into EDTA tubes to obtain whole blood and then stored at -80°C. One set of samples was couriered back to Sheffield on dry ice.

2.2.5. DNA extraction

DNA from the Bangladeshi MND patient's samples was extracted using the Maxwell System, available at the Sheffield Institute for Translational Neuroscience (SITraN) at the University of Sheffield (Henry, 2001). This is a cassette-based extraction protocol which minimizes the exposure of the researcher to the blood sample.

Four EDTA tubes with blood samples were taken out from the -80 °C freezer and kept in the laminar hood at room temperature to thaw. Once thawed, the tube was inverted several times to ensure the sample was well-mixed. The required number of cartridges was put on the tray and the foil lids were removed. From the samples, 500µL blood was pipetted into well number 1 of the cartridge, as shown in **figure 2.2**, which contains lysis buffer.

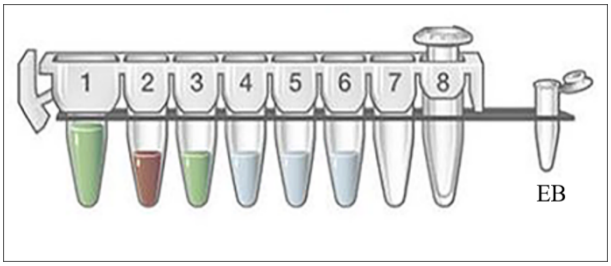


Figure 2.2. Cartridge for the Maxwell® RSC DNA or RNA extraction.

Plungers were put into well number 8 and 200µL of elution buffer was added into the elution tubes which were labelled with sample ID.

The cartridges containing the sample and pre-loaded buffers in tubes 2 to 6 were placed within the instrument. Each of the 4 samples was run in duplicate, so 8 cartridges were loaded.

The 'Whole blood DNA' protocol was selected from the main menu of the Maxwell application which allowed running the automated protocol to extract DNA from whole blood cells. This protocol takes approximately 35 mins to run.

When the run was complete, the eluted samples were transferred to the corresponding matrix tubes and labelled accordingly. The remaining blood samples were stored at -80 for longer-term storage in case additional extractions were required.

2.2.6. Evaluation of extracted DNA quality and quantity

The quality of the extracted DNA can be assessed using various methods including absorbance (optical density), agarose gel electrophoresis or the use of fluorescent DNA-binding dyes. Although all three methods are convenient, they differ in terms of equipment needed, calculation to consider and ease of use.

Measurement of absorbance is the most common technique for detecting DNA yield and purity. For our study, we used absorbance for confirming the concentration and purity of the BD MND DNA samples.

The absorbance was assessed using NanoDrop® ND-1000 UV-Vis Spectrophotometer, a spectrophotometer that enables highly accurate analyses of 1ul samples for DNA,

RNA, protein, pigments, whole UV-visible (220-700nm) spectrum and cell density tests with remarkable reproducibility. So, the quality of the extracted DNA was determined using a NanoDrop UV/Vis spectrophotometer, looking for a single absorbance peak at 260 nm, 260/280 absorbance ratio of 1.8-2.0, and no evidence of substantial band shearing or contamination (either RNA or polysaccharide).

2.2.7. Targeted NGS of ALS Genes at Sheffield Diagnostic Genetics Service

From the duplicates of each sample, the better quality one (260/280 absorbance ratio) was sent to the Sheffield Diagnostic Genetics Service (SDGS) for the targeted sequencing with Neurogenetic Motor Disorder NGS Panels. SDGS is sited within the Sheffield Children's NHS Foundation Trust, one of the only three dedicated children's hospital trusts in the UK. It provides both cytogenetic and molecular genetic analyses maintained through a series of technical, administrative and IT teams. Their Next Generation Sequencing (NGS) facility currently has three NGS systems: an Illumina MiSeq and HiSeq 2500, and the ThermoFisher Scientific Ion GeneStudio S5 Prime System.

The targeted sequencing of our samples was performed with HiSeq 2500 System which is used to sequence multiplexed libraries prepared by the SureSelectXT target enrichment protocol (Chen et al., 2015), enabling the simultaneous sequencing of 96 samples at a time covering custom-designed gene panels that contains a total of 153 genes mentioned in Table 2.7.

Table 2.7. The 153 genes are associated with neurological motor disorders.

AARS	*ABCD1*	*AFG3L2*	*ALDH18A1*	*ALS2*	*AMPD2*	*ANG*
ANXA11	*AP1S2*	*AP4B1*	*AP4E1*	*AP4M1*	*AP4S1*	*AP5Z1*
APP	*ARG1*	*ARHGEF28*	*ARL6IP1*	*ARSI*	*ASAH1*	*ATL1*
ATP2B4	*ATP7A*	*ATXN2*	*B4GALNT1*	*BICD2*	*BSCL2*	*C12orf65*
C19orf12	*CCT5*	*CHCHD10*	*CHMP2B*	*CPT1C*	*CSF1R*	*CYP27A1*
CYP2U1	*CYP7B1*	*DAO*	*DCTN1*	*DDHD1*	*DDHD2*	*DNAJB2*
DNM2	*DNMT1*	*DYNC1H1*	*EIF2B5*	*ENTPD1*	*ERBB4*	*ERLIN1*
ERLIN2	*EWSR1*	*EXOSC3*	*EXOSC8*	*FA2H*	*FARS2*	*FBXO38*
FIG4	*FLRT1*	*FUS*	*GAD1*	*GAN*	*GARS*	*GBA2*
GCH1	*GJA1*	*GJC2*	*GRN*	*HEXA*	*HNRNPA1*	*HNRNPA2B1*
HSPB1	*HSPB3*	*HSPB8*	*HSPD1*	*HTRA1*	*IBA57*	*IGHMBP2*
ITM2B	*KDM5C*	*KIAA0196*	*KIF1A*	*KIF1C*	*KIF5A*	*KLC4*
L1CAM	*LAS1L*	*LYST*	*MAG*	*MAPT*	*MARS*	*MARS2*
MATR3	*MT-ATP6*	*MTPAP*	*NEFH*	*NEK1*	*NIPA1*	*NOTCH3*
NT5C2	*OPTN*	*PFN1*	*PGAP1*	*PLEKHG5*	*PLP1*	*PNPLA6*
PRNP	*PRPH*	*PSEN1*	*PSEN2*	*RAB3GAP2*	*REEP1*	*REEP2*
RTN2	*SACS*	*SCO2*	*SETX*	*SIGMAR1*	*SLC16A2*	*SLC2A1*
SLC33A1	*SLC5A7*	*SMN1*	*SOD1*	*SPAST*	*SPG11*	*SPG20*
SPG21	*SPG7*	*SQSTM1*	*SS18L1*	*TAF15*	*TARDBP*	*TBK1*
TECPR2	*TFG*	*TREM2*	*TRPV4*	*TUBA4A*	*TYROBP*	*UBA1*
UBQLN2	*USP8*	*VAMP1*	*VAPB*	*VCP*	*VPS37A*	*VPS54*
VRK1	*WDR45*	*WDR48*	*ZFR*	*ZFYVE26*	*ZFYVE27*	

The S5 prime and Ion Chef, a semiconductor-based (works on the principle of detection of hydrogen ion release during incorporation of a new nucleotide in the growing DNA template) next-generation sequencing system were used for the rapid preparation and NGS sequencing of custom and ready-made Ampliseq gene panels, with minimal hands-on time. The 153 genes listed in Table 2.7 are subdivided into 4 different Neurogenetic Motor Disorder NGS Panels which are shown in Table 2.8.

Table 2.8. Neurogenetic Motor Disorders NGS Panels. (Table improvised from (SDGS)).

NGS Panel	Genes
Hereditary Spastic Paraplegia (HSP) 88 gene panel	*ABCD1, AFG3L2, ALDH18A1, ALS2, AMPD2, AP1S2, AP4B1, AP4E1, AP4M1, AP4S1, AP5Z1, ARG1, ARL6IP1, ARSI, ATL1, ATP2B4, B4GALNT1, BSCL2, C12orf65, C19orf12, CCT5, CPT1C, CYP27A1, CYP2U1, CYP7B1, DDHD1, DDHD2, DNM2, EIF2B5, ENTPD1, ERLIN1, ERLIN2, FA2H, FARS2, FIG4, FLRT1, GAD1, GAN, GBA2, GCH1, GJA1, GJC2, HEXA, HSPD1, IBA57, KDM5C, KIAA0196, KIF1A, KIF1C, KIF5A, KLC4, L1CAM, LYST, MAG, MARS, MARS2, MT-ATP6, MTPAP, NIPA1, NT5C2, PGAP1, PLP1, PNPLA6, PSEN1, RAB3GAP2, REEP1, REEP2, RTN2, SACS, SIGMAR1, SLC16A2, SLC2A1, SLC33A1, SPAST, SPG11, SPG20, SPG21, SPG7, TECPR2, TFG, USP8, VAMP1, VPS37A, WDR45, WDR48, ZFR, ZFYVE26, ZFYVE27*
Familial Amyotrophic Lateral Sclerosis with or without Frontotemporal Dementia 42 gene panel	*ALS2, ANG, ANXA11, ARHGEF28, ATXN2, CHCHD10, CHMP2B, CYP27A1, DAO, DCTN1, ERBB4, EWSR1, FIG4, FUS, GBA2, GRN, HNRNPA1, HNRNPA2B1, MAPT, MATR3, NEFH, NEK1, OPTN, PFN1, PRPH, SETX, SIGMAR1, SOD1, SPAST, SPG11, SPG20, SQSTM1, SS18L1, TAF15, TARDBP, TBK1, TUBA4A, UBQLN2, VAPB, VCP, VPS54, VRK1*
Dementia 27 gene panel	*APP, CHCHD10, CHMP2B, CSF1R, CYP27A1, DCTN1, DNMT1, FUS, GRN, HNRNPA2B1, HTRA1, ITM2B, MAPT, MATR3, NOTCH3, PRNP, PSEN1, PSEN2, SPG21, SQSTM1, TARDBP, TBK1, TREM2, TUBA4A, TYROBP, UBQLN2, VCP*
Spinal Muscular Atrophy 29 gene panel	*AARS, ASAH1, ATP7A, BICD2, BSCL2, CHCHD10, DCTN1, DNAJB2, DYNC1H1, EXOSC3, EXOSC8, FBXO38, GARS, HEXA, HSPB1, HSPB3, HSPB8, IGHMBP2, LAS1L, PLEKHG5, REEP1, SCO2, SIGMAR1, SLC5A7, SMN1, TRPV4, UBA1, VAPB, VRK1*

2.2.8. Variant Analysis Pipeline for 153 genes

VCF and FASTQ files were obtained from SDGS analysis. The pipeline described in section 2.1.3.4 was modified to analyse the 153 genes in Neurogenetic Motor Disorder NGS Panels shown in Table 2.7.

2.2.8.1. Part 1: Sorting out the variants

The full coding is provided in **Appendix B**. However, the key commands and changes (for one patient sample: BD-MND-001) are explained in section **Appendix B v** to make clear the command line.

2.2.8.2. Part 2: Extract_clinvar

The variants reported in CliVar were extracted with the command line as explained in section 2.1.3.4.1.2.b and compared with the csv file found in the previous section 2.2.8.1.

2.2.9. Rank on CADD score

The CADD score (section 2.1.3.5) was determined for each of the variants with the conflicting interpretation of pathogenicity or uncertain significance to prioritize them.

2.2.10. Mutation Taster Prediction

Mutation taster software was used to evaluate DNA sequence variants for their disease-causing potential (Schwarz et al., 2014). MutationTaster (MT) score is the probability that predicts an alteration as one of four possible types **(Table 2.9)**.

Table 2.9. Mutation Taster Predictions (MT-documentation).

Types	Prediction	Comments
D	Disease-causing	probably deleterious
A	Disease-causing automatic	known to be deleterious, see section dbSNP / TGP / ClinVar / HGMD for details
N	Polymorphism	probably harmless
P	Polymorphism automatic	known to be harmless, see section dbSNP / TGP / ClinVar / HGMD for details

2.2.11. Allele frequency

The allele frequencies of the reported variants of 36 samples were also included from Project MinE (MinE) (ALS-specific genome sequence database) and the Genome Aggregation Database (gnomAD) databases containing whole genome sequencing (ge) and whole exome sequencing (ex) and assessed along with ClinVar information (as explained in section 2.1.3.6).

2.3. TDP-43 Analysis of MMA Samples

TDP-43 pathology analysis was performed on patient-derived CNS-relevant cells from selected patients with variants in a gene related to MND to determine whether MMA shares a molecular basis with other disorders including MND.

2.3.1. Sample selection

The genetic testing of eight UK MMA patients showed that there were some potentially significant pathogenic and/or novel variants present in four patients. Three patients had *MAG* variants listed in table 4.5. Sample SHF-067 and SHF-087 had the same *MAG* variant c.1849C>G (p.Leu617Val) whereas the third sample SHF-070 had a different *MAG* variant c.711G>C (p.Lys237Asn). Neither variant is reported in ClinVar and both have high CADD scores, of 17.72 (p.Leu617Val) and 14.55 (p.Lys237Asn). One sample, SHF-072 was reported with a non-coding VUS *OPTN* c.1402-7C>T variant which is not reported in ClinVar. One of the samples carrying the MAG variant with the higher CADD score was selected (**SHF-067**) along with the *OPTN* variant (**SHF-072**) of unknown significance. To have a clear picture of the analysis two samples without these variants (**SHF-073** and **SHF-091**) were also selected to observe the difference in TDP-43 pathology. Two healthy controls (**155 and 161**) were included in this study as negative controls. Two positive controls are also included that were found positive for TDP-43 pathology analysis. One (**78**) was a FALS patient with *C9ORF72* expansion and another (**009**) was diagnosed with SALS.

2.3.2. Fibroblast Samples

Human skin fibroblast samples collected from different sources were reprogrammed into induced neural progenitor cells (iNPCs). The fibroblast samples were obtained from the AMBRoSIA Project, at the University of Sheffield (Integrated Research Application System (IRAS) Project ID: 204405, Research Ethics Committee (REC) reference: 16/LO/2136). An overview of the cell lines used for the TDP-43 pathology analysis is shown in Table 2.10. Due to time constraints regarding the reprogramming of the SHF iNPCs, cell lines were not passaged matched during this study. The SHF lines were used from passages 4-8 and the Control lines from passages 16-20.

Table 2.10. The Fibroblast Control and Sample cell lines information.

Cell Line	Ethnicity	Gender	Diagnosis	Mutation	Age at collection (y)
155 (-)	European	Male	Healthy Control	-	40
161 (-)	European	Male	Healthy Control	-	31
78 (+)	European	Male	FALS	C9ORF72	66
009 (+)	European	Female	SALS	-	61
SHF-067	European	Male	MMA	-	57
SHF-072	European	Male	MMA	-	48
SHF-073	European	Male	MMA	-	24
SHF-091	European	Male	MMA	-	80

2.3.3. Sample collations

All the patients and controls in this study were recruited by AMBRoSIA. The collected fibroblast are stored in Sheffield Biorepository - Human Tissue Storage (Storage). Here all the materials consisting of or including human cells are covered by the Human Tissue Authority (HTA) and include the initial isolation of cells for primary culture.

2.3.4. Reprogramming: Fibroblast to iNPC

Adult human fibroblasts were directly converted into induced neural progenitor cells using a modified version of the protocol outlined in Meyer et al 2014 (Figure 2.4A) (Meyer et al., 2014). These aspects of the project received assistance from Allan C Shaw, Raquel Rua Martins, Thomas R Marlow, and Katie M Bowden from the research group in SITraN under the supervision of Professor Laura Ferraiuolo. Specific input will be highlighted where appropriate.

2.3.4.1. Materials

1. PBS
2. Tryosin (Lonza Biowhittaker, Basel, Switzerland)
3. Fibronectin (Merck, Burlington, MA, USA)
4. Dulbecco's Modified Eagles Medium (DMEM) Media or known as DMEM/F-12 with GlutaMAX™ (Gibco Life Tech, Waltham, MA, USA)
5. Fetal bovine serum (FBS) (Life Science Production, Bedford, UK)
6. B27 Supplement (Gibco Life Tech, Waltham, MA, USA)

7. N2 Supplement (Gibco Life Tech, Waltham, MA, USA)
8. Epidermal growth factor (EGF) (Peprotech, Rocky Hill, NJ, USA)
9. Fibroblast growth factors (FGF) (Peprotech, Rocky Hill, NJ, USA)
10. Retroviral vectors (ALSTEM, Richmond, CA, USA):
    a. OCT3/4: transcription factor protein encoded by Pou5f1
    b. SOX2: SRY (sex determining region Y)-box 2
    c. KLF4: KLF Transcription Factor 4
    d. C-MYC: cellular Myc (MYC Proto-Oncogene, BHLH Transcription Factor)
11. Accutase (Invitrogen, A11105)
12. Normal Horse Serum (NHSer)

2.3.4.2. Media preparation

2.3.4.1.1. Fibroblast Medium

Fibroblast media contained DMEM/F12, Glutamax with 1% B27-supplement, 1% N2-supplement, 40 ng/mL EGF and 20 ng/mL FGF.

2.3.4.1.2. Conversion medium/ pre-iNPC Media

The conversion medium was prepared by adding 50µl Epidermal growth factor (EGF) and 2.5µl Fibroblast growth factors (FGF) in 500ml iNPC Media.

2.3.4.1.3. iNPC Media

The media prepared for iNPC contained DMEM/F12, Glutamax, 1% B27-supplement, 1% N2-supplement, and 40 ng/mL FGF.

2.3.4.1.4. iAstrocyte Media

DMEM supplemented with 10% FBS, 1% Penicillin-Streptomycin and 0.2% N2-supplement was used as iAstrocyte media.

2.3.4.3. Preparing the Fibroblast for transduction

The fibroblast sample vials retrieved from the Biorepository contain the information on the cell line name, passage number and date of storage. The samples were stored in a mixture of 10% DMSO and 90% FBS. DMSO prevents ice crystal formation while freezing. As DMSO is toxic the first steps involve removing the DMSO from the samples. The vials were kept in the water bath to thaw and the cell suspension was taken out using a pipette and transferred into a 15ml falcon tube. The tube was centrifuged at 400 RCF for 4 min and the supernatant was discarded without disturbing the palette. The palette was resuspended using 2ml of fibroblast media and transferred to a T75 flask that already had sufficient fibroblast media for growth.

It was made sure that the fibroblast lines of the samples were growing well which is important for successful conversion. Direct conversion is usually more efficient with younger passage fibroblast (P5-7). The passage number of a cell culture is a record of the number of times the culture has been sub-cultured. The fibroblast samples were grown in a T75 flask.

When the cells were ~ 80% confluent (every 4-6 days), subculture cells (as described above under subculturing) into new vessels. For each sample, a 6-well plate was coated with 1:400 fibronectin. After coating the fibronectin was replaced with 2ml iAstrocyte

media. Cells were counted and plated in 3 different densities 75,000, 125,000 and 175,000 in 2 wells each.

2.3.4.4. Transduction

The day after plating, a single well was chosen for transduction (~70% confluence with cells spread evenly throughout the well). The fibroblasts were transduced with Retroviral vectors. For making the master mix of the retroviral vectors 5 µl of each 4 vectors: OCT3/4, SOX2, KLF4 and C-MYC were mixed and 3 µl/well Transduction Agent (PEG) was added. The next day, 18 hours after transduction, the well was washed with PBS and 2 ml of fresh fibroblast media was added.

2.3.4.5. Conversion

The following day the medium was changed to conversion medium (iNPC media+ EGF+FGF) after washing the wells with 1x PBS to get rid of the residual fibroblast medium. For the first 2-3 days 0.5-1.0ml media was added daily. Then the media was replenished to maintain the volume and availability of fresh media. The cells needed observation every day as they started changing morphology very soon (within 2-3 days) (Figure 2.4 B, C and D).

2.3.4.6. Pre-iNPC

When the cells became very dense (after about 5-6 days) they were split into 1:2 using accutase (600 uL per well). For each sample, 2 wells of a new 6-well plate (6wp) were

coated using 1:100 fibronectin and left for a minimum of 5 minutes. During this time 600µl accutase was added into each cell containing well removing the media and incubating for 5 minutes. After incubation, the cells were collected by adding 1 ml fresh medium using P1000 to transfer them to a falcon. The falcon was centrifuged at 200g for 4 minutes. The pellet was resuspended in 6 ml conversion media. Fibronectin was removed from the new plate and 3 ml of the pellet resuspended conversion media was added to each well. The cultures were maintained in conversion media, with daily replenishment, keeping an amount of conditioned media in the well. After nearly 10 days the transduction feeding can switch to every two days a week.

2.3.4.7. iNPC

A good workflow of transduction normally followed 1x6wp well -> 2x6wp wells -> 3x6wp wells -> 1x10 cm petri dish -> 1x10cm and 1-2 vials frozen (usually in full iNPC media by this stage). When all the cells from the first 10 cm petri dish were split the media was changed to full iNPC media. If the pre-iNPC culture was in full iNPC media for at least one split without any gross changes to proliferation rate or morphology, it can be reclassified as iNPC before confirmation.

2.3.4.8. Immunocytochemistry (ICC): PAX6 & Nestin

Immunocytochemistry (ICC) was used to visualize the localization of a specific protein or antigen in the cells. A specific primary antibody bonds the protein or antigen, while

a secondary antibody with a conjugated fluorophore binds with the primary antibody and allows visualization (Figure 2.3 E, F and G).

For confirmation of iNPC conversion, 10,000 cells were plated per well in a minimum of 5 wells per sample in a 96-well plate (96wp) coated with 1:200 fibronectin. The cells were stained for a nuclear marker Hoechst (2'-[4-ethoxyphenyl]-5-[4-methyl-1-piperazinyl]-2,5'-bi-1H-benzimidazole trihydrochloride trihydrate), a transcription factor marker PAX6 (paired box protein Pax-6) and a cytoskeletal intermediate filament marker nestin also known as neuroepithelial stem cell protein (nestin).

First, the cells in 96wp were fixed using 4% Paraformaldehyde Solution (50µl/well) for 10 minutes at room temperature followed by PBS washes (2 times for 5 minutes). Then they can be stored at 4^0C until immunocytochemistry staining.

To begin with, the ICC procedure blocking solution was prepared using 5% Normal Horse Serum (NHSer) in PBS. A blocking solution was used to block any unspecific binding sites. The cells were permeabilised and blocked at room temperature for 1 hour adding 50µl of blocking solution containing 0.5% triton per well.

The Hoechst working solution used for this study was prepared by 1:6,000 dilution of the Stock Solution (10 mg/mL) in Assay Buffer (0.1 M NaCl, 10 mM EDTA, 10 mM Tris, pH 7.0). Table 2.11 showed details of the primary antibodies used for this analysis. The antibody mixture (50µl/well) was added removing the blocking solution after the incubation and kept at 4^0C for overnight (O/N) incubation.

Table 2.11. Primary and Secondary Antibody used for confirming the transformation of iNPC.

Marker	Primary Antibody	Concentration	Secondary Antibody	Concentration
PAX6	Rabbit polyclonal, Abcam, ab5790	1:400	AlexaFluor 568 goat polyclonal anti-rabbit IgG (H+L), Invitrogen, A-11011	1:1000
Nestin	2C1.3A11, Mouse monoclonal, Abcam, ab18102	1:400	AlexaFluor 488 donkey polyclonal, anti-mouse IgG (H+L), Invitrogen, A-21202	1:1000

The next day the primary antibodies were aspirated and the wells were washed once with PBS containing 0.2% Tween (approximately 100 µl/well) and twice with PBS (5 minutes for each wash).

The secondary antibody mixture (50µl/well) was added to the cell wells and incubated for 1 hour at room temperature. The plate was covered with foil paper to protect it from light. A cell-permeant nuclear counterstain Hoechst 33342 was used that emits blue fluorescence when bound to dsDNA. It was diluted in PBS (1:6000), added into the cell wells and incubated for 5 minutes at room temperature protected from light.

After incubation, the solutions were discarded and the wells were washed three times: once with 0.2% Tween in PBS and twice with PBS only. The plate was placed on the shaker for 5 minutes during each wash. Fresh PBS (100-150µl/well) was added before imaging.

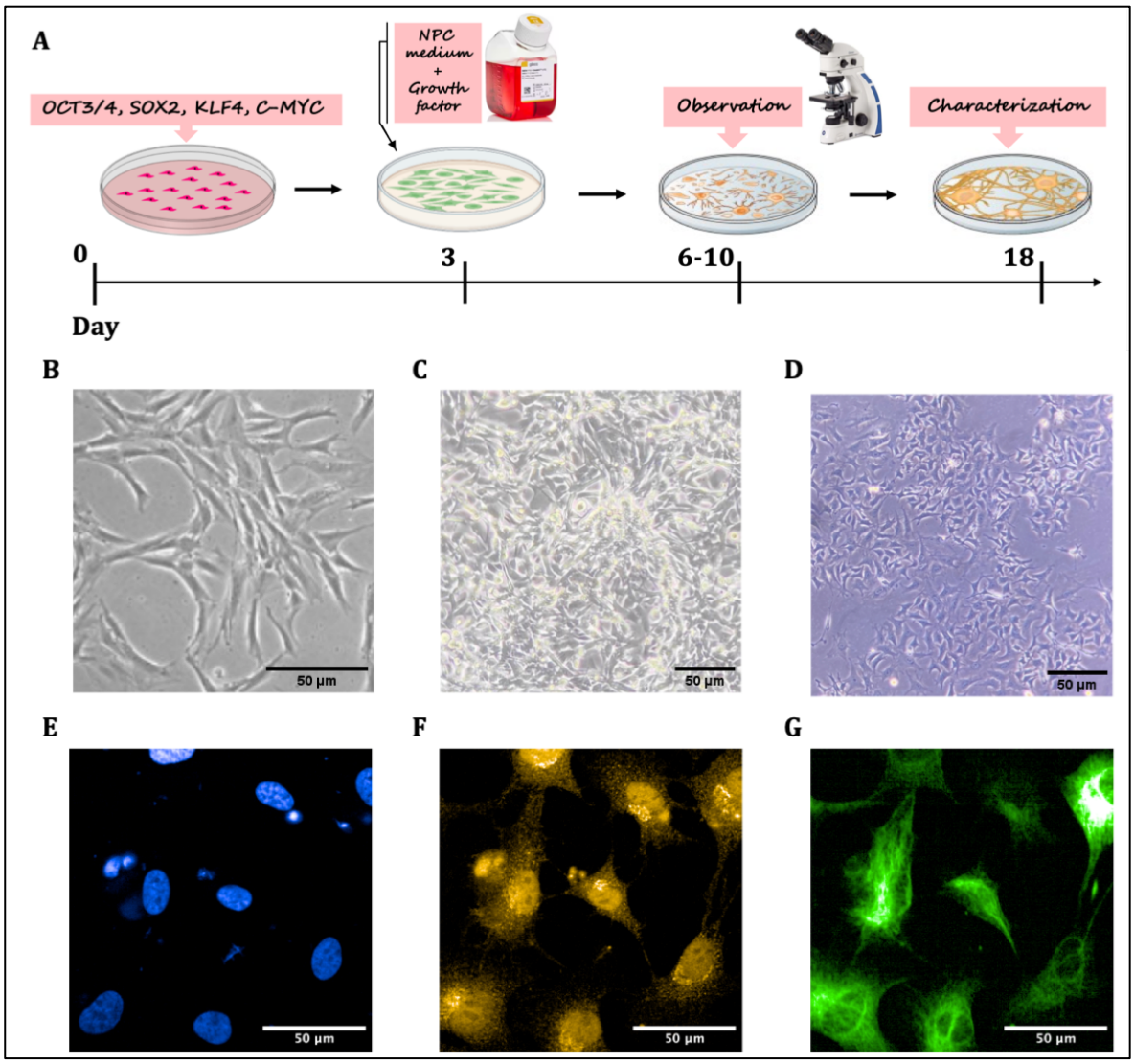


Figure 2.3. Direct conversion of sample fibroblasts to iNPCs. (A) Schematic of the conversion process from fibroblasts to induced neuronal progenitor cells (iNPCs). Fibroblasts were transduced with retroviral vectors containing four reprogramming factors (Oct3/4, Sox2, KLF4, c-Myc). (B, C and D) Within 6–10 days, cells underwent marked morphological changes from a fibroblastic spindle-like shape (B) to a sphere-like form commonly seen with NPCs (C and D). The cultures were expanded from 6wp to 10cm before confirmation of Pax6/Nestin. Hoechst staining (blue) was used to visualize nuclei (E). Immunofluorescence of cultures at passages 4-5 (minimum day 18) reveals the expression of the NPC markers Pax6 (F) and Nestin (G), as shown in yellow and green respectively.

The cells were imaged using the Opera® High Content Screening System from PerkinElmer (Figure 2.4).

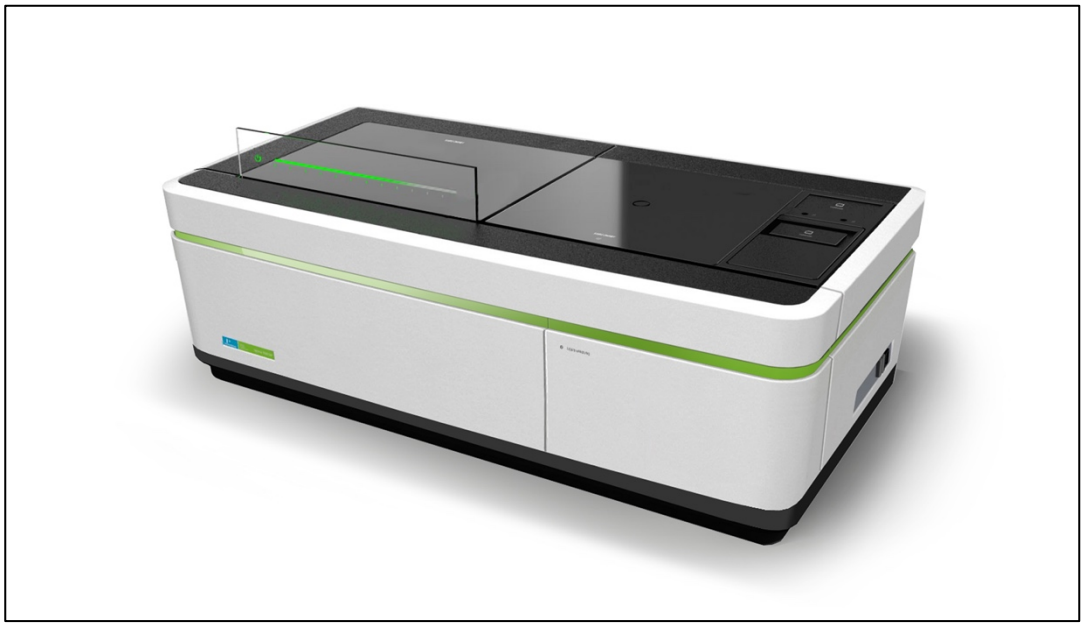


Figure 2.4. Opera Phenix High-Content Screening System from PerkinElmer.

2.3.5. TDP-43 Pathology Analysis

Cytoplasmic transactive response DNA-binding protein 43 kDa (TDP-43) inclusion was identified as a pathological marker of patients with amyotrophic lateral sclerosis (ALS) (Arai et al., 2006, Kawakami et al., 2019, Neumann et al., 2006a). So, the samples were subjected to western blot (WB) and immunocytochemistry (ICC) to detect pathological TDP-43.

2.3.5.1. Western blot

The western blot procedure followed for this study is described below in steps.

2.3.5.1.1. Mini-PROTEAN® Tetra System

The Bio-Rad Mini-PROTEAN® Tetra Cell was used for the western blot. The component of this Cell is listed in Table 2.12.

Table 2.12. Components of Bio-Rad Mini-PROTEAN® Tetra Cell.

Component	Description
Spacer Plate	The spacer plate is the taller glass plate with permanently bonded gel spacers. Spacer plates are available in 0.75 mm, 1.0mm, and 1.5 mm thicknesses, which are marked directly on each spacer plate.
Short Plate	The short plate is the shorter, flat glass plate that combines with the spacer plate to form the gel cassette sandwich.
Casting Frame	The casting frame, when placed on the benchtop, evenly aligns and secures the spacer plate and the short plate together to form the gel cassette sandwich before casting.
Gel Cassette Assembly	One casting frame, a spacer plate, and a short plate form one gel cassette assembly.
Casting Stand	The casting stand secures the gel cassette assembly during gel casting. It contains pressure levers that seal the gel cassette assembly against the casting gaskets.
Gel Cassette Sandwich	A spacer plate and a short plate with a polymerized gel form a gel sandwich.
Buffer Dam	The moulded, one-piece buffer dam is used when running only one or three gels.

Firstly the spacer plates (Bio-Rad) and short plates (Bio-Rad) and clamps (Bio-Rad) with gaskets (Bio-Rad) were assembled. After the gel is set in it the plates were set on the Mini-PROTEAN® Tetra Cell Casting Stand.

2.3.5.1.2. SDS-Polyacrylamide Gel Preparation

The 12% resolving gel was used for this analysis and it was prepared by mixing reagents listed in Table 2.13.

First, the water, acrylamide and buffer were mixed then the catalysts (APS and TEMED) which catalyse the polymerisation of acrylamide and cause the solution to set were added. The resolving gel solution was poured into the glass plates.

Stacking gel (5%) was prepared by mixing reagents listed in Table 2.13 (water, acrylamide and buffer first, followed by APS and TEMED). The stacking gel solution was poured onto the set resolving gels in the glass plates filling to the brim. Mini-PROTEAN® Combs (Bio-Rad) was inserted into the stacking gels and left for at least 15 mins to set.

Table 2.13. Materials for preparing resolving gel.

Reagents	5 % stacking gel	12 % resolving gel
dH_2O	5.8 mL	3.5 mL
30 % (w/v) Acrylamide	1.7 mL	4.0 mL
Resolving buffer (1.5 M Trizma®, 13.9 mM SDS, pH 8.8, filtered)	-	2.5 mL
Stacking buffer (0.5 M Trizma®, 13.9 mM SDS, pH 6.8, filtered)	2.5 mL	-
10 % (w/v) Ammonium Persulfate (APS)	50 µL	50 µL
Tetramethylethylenediamine (TEMED)	20 µL	20 µL

2.3.5.1.3. Cell Lysis

The media was removed from the petri dish with the iAstrocyte cells of corresponding samples and washed with ice-cold PBS. The cells were scraped and collected in an Eppendorf tube using PBS. The Eppendorf was centrifuged at 17,000 x g for 1 min followed by removing all PBS. The cell pellet can be stored in a -80°C freezer until further use.

For cell lysis ice-cold IP lysis buffer (150mM NaCl, 50mM HEPES, 1mM EDTA, 1mM DTT, 0.5% (v/v) Triton™ X-100, PIC, pH 8.0) + PIC (protease inhibitor cocktail) was added. The volume of lysis buffer needed was dependent on the size of the cell pellet, but 20-50 µL was usually sufficient. The tube was incubated for 15mins on ice.

The lysate was centrifuged at 17,000 x g for 5 mins at 4°C. The lysate supernatant which contains soluble proteins was reserved. The pellet of debris was discarded.

2.3.5.1.4. Bicinchoninic acid (BCA) Assay

The total protein concentration in cell lysates was determined using a BCA assay that utilizes a detergent-compatible formulation containing bicinchoninic acid (BCA) for colourimetric detection and quantification (Kumar, 2016).

This assay relies on reducing Cu^{+2} to Cu^{+1} by protein in an alkaline medium, known as the biuret reaction. It uses a unique reagent containing bicinchoninic acid (BCA) to selectively detect the resulting cuprous cation (Cu^{+1}) (Smith et al., 1985). The chelation of two molecules of BCA with one cuprous ion produces the purple-coloured product, a water-soluble complex that exhibits a strong absorbance at 562 nm that is nearly

linear with increasing protein concentrations over a broad working range (20–2000 µg/mL). The macromolecular structure of the protein, the number of peptide bonds, and the presence of specific amino acids (cysteine, cystine, tryptophan, and tyrosine) are believed to contribute to the colour formation observed with BCA (Wiechelman et al., 1988). The protein concentration is typically determined and reported using a common protein standard such as bovine serum albumin (BSA), with a series of dilutions of known concentration (0.1-2.0 µg/µl) prepared and assayed (Table 5.1 and 5.2) alongside unknown samples to generate a standard curve (Figure 5.2) for determination of unknown concentrations.

2.3.5.1.5. SDS-Polyacrylamide Gel Electrophoresis

The cell lysates in lysis buffer were diluted such that equal amounts (20µg) of protein were loaded per well and mixed with 4µl of 4x Laemmli buffer to make a total volume of 20µL. The samples were boiled for 5 mins at 95°C to denature proteins and coat them in a negative charge. The proteins' migration rate was inversely proportional to their molecular weight through the gels via electrophoresis.

The Mini-PROTEAN® Tetra Vertical Electrophoresis Cell from Bio-Rad was used to load the SDS-Polyacrylamide gels, filled with running buffer of 25mM Tris, 3.5mM SDS, and 20mM glycine. In the case of general immunoblotting techniques, 20-30 µg of denatured protein samples were loaded into each well of the SDS-polyacrylamide gels. Additionally, a 2 µL PageRuler Prestained Protein Ladder, from ThermoFisher was loaded in one well per gel as a molecular weight marker.

The gels needed to run at 50V for 30-40 mins, then 150V-180V for approximately 1.5h (or until the dye front reaches the bottom of the gel).

2.3.5.1.6. Transfer onto nitrocellulose membrane

The transfer conditions are dependent on the gel type, the immobilization membrane, the transfer apparatus used as well as the protein itself (Kurien and Scofield, 2015). SDS gels, urea gels (Alwine et al., 1977), lithium dodecyl sulfate-containing gels, nondenaturing gels, two-dimensional gels and agarose gels have been used for protein blotting (electrophoretic) (Gershoni and Palade, 1983). The electric charge of the protein should be determined and the membrane should be placed on the appropriate side of the gel. When using urea gels the membrane should be placed on the cathode side of the gel (Alwine et al., 1977). Proteins from SDS PAGE gels are eluted as anions and therefore the filter should be placed on the anode side of the gel (Kurien and Scofield, 2015).

In the wet transfer procedure, the sandwich is placed in a buffer tank with platinum wire electrodes (Kurien and Scofield, 2015) based on the design of Towbin et al. (Towbin et al., 1979): i.e., they have vertical stainless steel/platinum electrodes in a large tank.

In “Semi-dry” transfer, the gel-membrane sandwich is placed between carbon plate electrodes (Kurien and Scofield, 2015). “Semi-dry” or “horizontal” blotting uses two plate electrodes (stainless steel or graphite/carbon) for a uniform electrical field over a short distance, and sandwiches between these up to six gel/membrane/filter paper

assemblies, all well-soaked in transfer buffer as shown in Figure 2.5. The assembly is clamped or otherwise secured on its side, and electrophoretic transfer is effected in this position, using as transfer buffer only the liquid contained in the gel and filter papers or other pads in the assembly (Kurien and Scofield, 2015).

The proteins were transferred from the gels to the membranes at 0.15A / gel (i.e 0.3A if transferring two gels) for 1 hour.

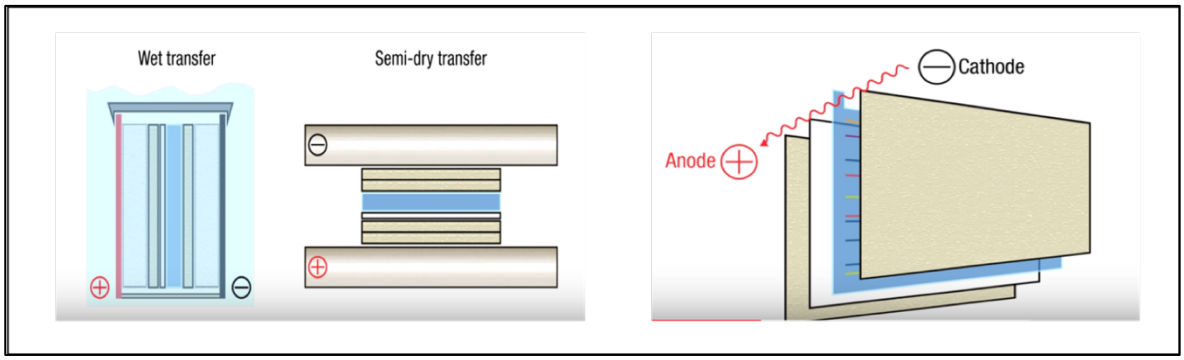


Figure 2.5. Transferring the protein onto the nitrocellulose membrane.

2.3.5.1.7. Immunoblotting

The nitrocellulose membrane was placed carefully in a 50 mL falcon tube and blocked in 5% (w/v) milk/Tris Buffered Saline, with Tween® 20 (TBST) (20mM Tris, 137mM NaCl, 0.2% (v/v) Tween® 20, pH 7.6) for 1 hour at room temperature on a roller.

The membranes were incubated with primary antibody (C-terminal TDP43, Rabbit polyclonal, Preprotech/Proteintech 12892-1-AP in 1:1000) diluted in 5% milk in TBST overnight at 4°C on a roller. Around 5 mL antibody in the blocking agent was enough for a 50 mL falcon. The membranes were washed 3 times in TBST for 15mins (5 min per wash) at room temperature. Then the membranes were incubated with a

secondary antibody conjugated to horseradish peroxidase (HRP) in 5% milk/TBST for 1 hour at room temperature on a roller. The membranes were washed 3 times in TBST for 15mins (5 min per wash) at room temperature.

Finally, the membranes were incubated with ECL for 1mins and images were taken using Odyssey® Fc Imaging System by LI-COR Biosciences (Dentlinger, 2018).

All the interest protein (TDP-43) was normalised for Bactin (β-actin). Thus, Beta-actin (42 kDa) was used as the loading control to normalize the levels of protein detected by confirming that protein loading is the same across the gel (Liao et al., 2000).

2.3.5.2. Immunocytochemistry (ICC)

For immunocytochemistry (ICC) analysis of TDP-43 pathology, the procedure explained in section 2.3.4.8 was followed. The antibodies are listed in table 2.14.

Table 2.14. Marker/Antibodies used to detect TDP-43 pathology.

Name	Primary Antibody	Concentration	Secondary Antibody	Concentration
TDP-43	C-terminal TDP-43, Mouse monoclonal, Newmarket Scientific, AX17-10010	1:200	AlexaFluor 488 donkey anti-mouse IgG (H+L), Invitrogen, A-21202	1:1000
CD44	Rabbit polyclonal, from Abcam, ab157107	1:300	AlexaFluor 568 goat polyclonal anti-rabbit IgG (H+L), Invitrogen, A-11011	1:1000
Vimentin	Chicken polyclonal, Millipore, AB5733	1:1000	Goat polyclonal Anti-Chicken IgY H&L (Cy5 ®) preadsorbed, Abcam, ab97147	1:1000

Chapter 3. Genetic analysis of WGS of ALS patients from Project MINE

3.1. Introduction

Somatic heterogeneity is a feature of human genetics and new genetic mutations occurring within developing neurons of the motor cortex could adversely affect motor neuron function but would be undetectable in peripheral blood, the source of most DNA for genetic screening. Therefore, could somatic mutations contribute to the cause of sporadic disease in ALS? With the advent of next-generation sequencing, many genetic causes and risk factors for ALS have been identified which has been a significant driver of translational research. The use of this technology in a large cohort of motor cortex samples from ALS patients could therefore determine whether somatic mutations cause sporadic ALS.

3.2. Aims

This study aims to screen a large cohort of motor cortex DNA samples from ALS patients to establish if mutations in ALS genes arise somatically in the CNS. Where mutations are found, the results will be compared with sequencing data corresponding to blood samples (where available) to establish their somatic or germline status.

3.3. Results

The Single Nucleotide Polymorphisms (SNPs) from the whole genome sequencing of 69 samples of the 69 motor cortex DNA sample set have been annotated using SnpEff (Galaxy Version 4.3+T.galaxy1). The annotated mutation vcf file contains the number of variants categorised as high, moderate, low, or modifier effects by impact and the number of missense, nonsense, or silent variants according to the functional class (Table 2.1). The following flowchart shows the steps of the analysis described in **section 2.1.3.3**.

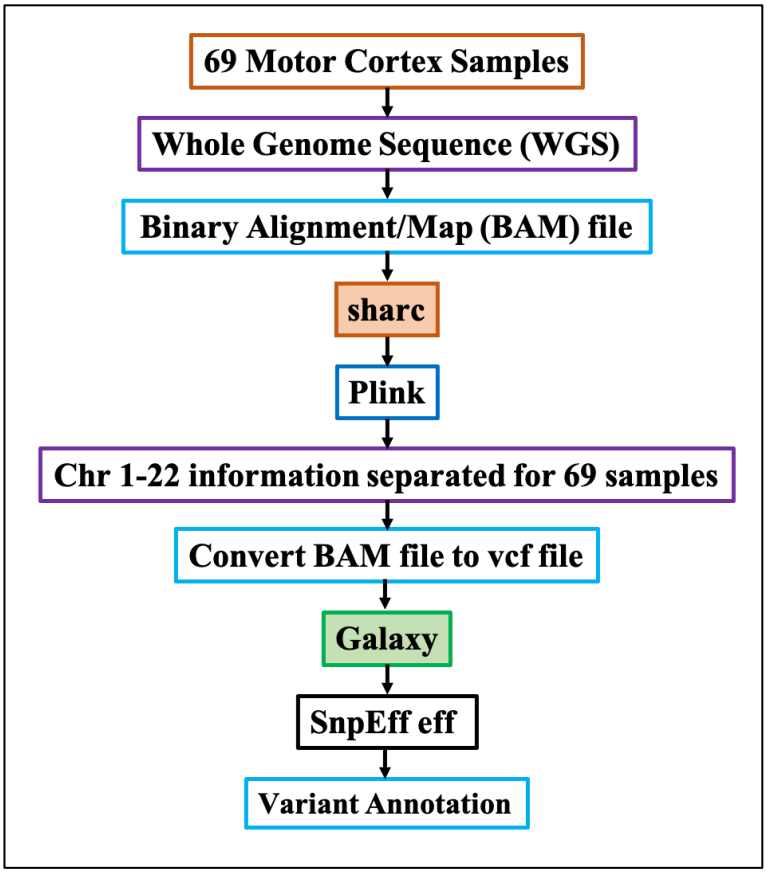


Figure 3.1. Flowchart of variant annotation of WGS data of 69 motor cortex samples.

3.3.1. Annotate variants

Initially, the specific chromosomes that contained the known 42 ALS genes were selected for running the variant calling as all 69 samples are from ALS patients.

The number of effects from the resulting annotated vcf file of 69 ALS patients SNPs from Project MinE is shown in **Table 3.1**. As mentioned in section 2.1.3.3 the annotated mutation contained four different "SNP effects" (SnpEff) by impact. The variants having a high impact on protein are classified as HIGH. MODERATE impact indicates the change in protein effectiveness. LOW impact is predicted as no change of protein behaviour whereas the MODIFIER indicates the variants affecting the non-coding gene.

As mentioned in **Table 2.1** the high and moderate ones were our variants of interest as they were assumed to have a high impact on protein and might affect the protein effectiveness.

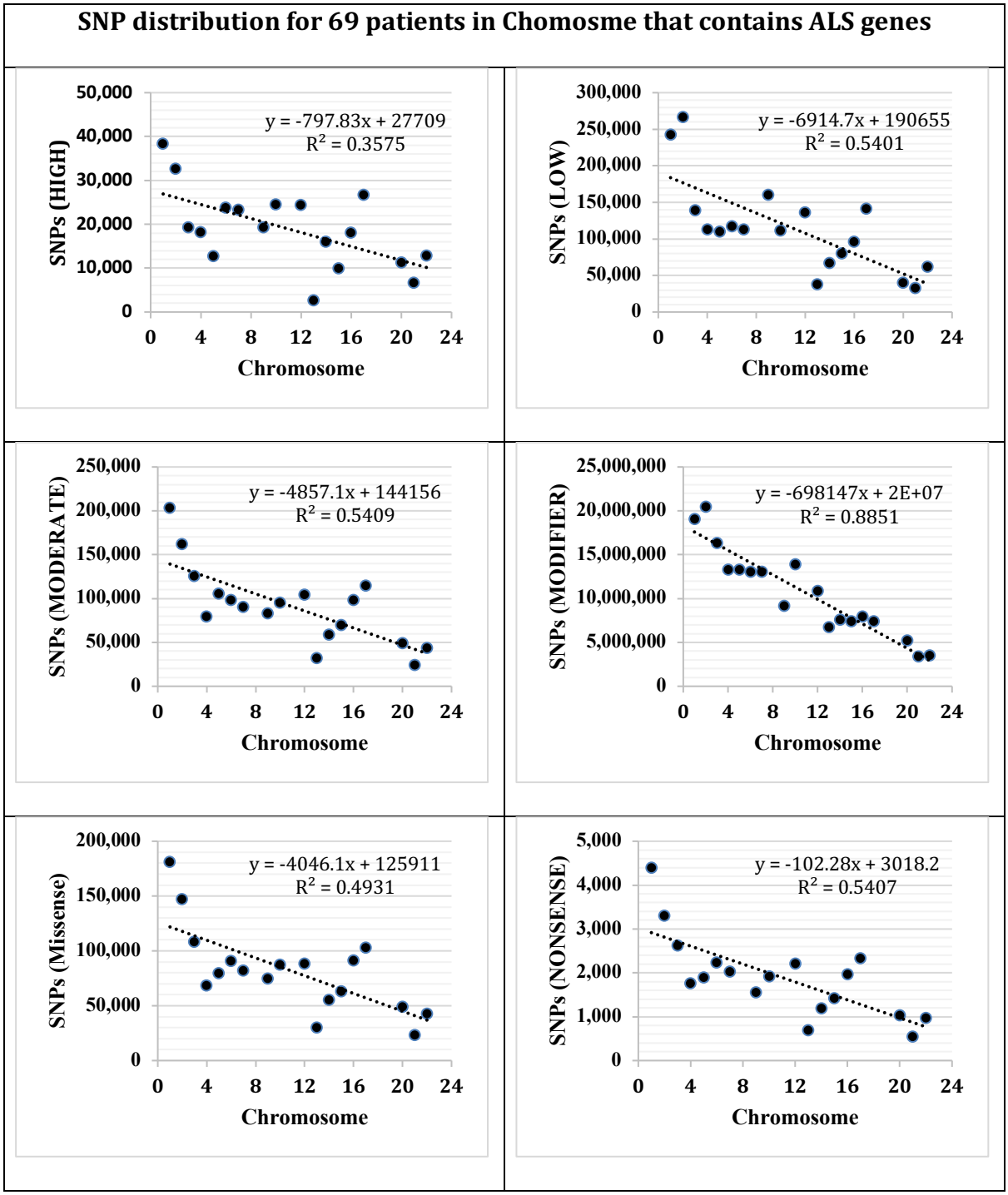
Only the chromosomes associated with the known ALS genes (Table 2.2) are listed in the table for minimizing the data analysis.

Table 3.1. The number of variants in each chromosome that contains an ALS gene within the 69 patients.

Chromosome	Number of effects on the chromosome by impact				Number of effects on the chromosome by functional class			Related ALS genes
	HIGH	LOW	MODERATE	MODIFIER	MISSENSE	NONSENSE	SILENT	
Chr1	38,359	242,783	203,389	19,072264	181,290	4,396	98,932	*TARDBP*
Chr2	32,674	266,683	161,938	20,445,911	147,516	3,304	78,254	*ALS2, TUBA4A, DCTN1, CYP27A1, ERBB4, SPAST, VPS54*
Chr3	19,304	139,553	125,590	16,310,945	108,532	2,632	59,459	*CHMP2B*
Chr4	18,236	113,340	79,286	13,317,037	68,738	1,767	36,724	*NEK1*
Chr5	12,797	110,221	105,555	13,282,441	79,668	1,895	45,292	*MATR3, ARHGEF28, SQSTM1*
Chr6	23,890	117,530	98,195	13,037,441	90,827	2,244	49,180	*FIG4*
Chr7	23,379	112,908	90,264	13,052,536	82,417	2,038	47,654	*HNRNPA2B1*
Chr9	19,300	160,439	82,844	9,189,163	74,798	1,561	44,218	*C9ORF72, GBA2, SETX, SIGMAR1, VCP*

Chr10	24,597	111,491	95,507	13,903,497	87,388	1,929	50,195	*ANXA11 , OPTN*
Chr12	24,443	136,399	104,364	10,843,438	88,636	2,217	50,886	*ATXN2, DAO, PRPH, TBK1, KIF5A, HNRNPA1*
Chr13	2,715	38,051	32,221	6,729,199	30,226	696	16,809	*SPG20/SPART*
Chr14	16,034	67,130	58,776	7,599,683	55,663	1,200	30,475	*ANG, VRK1*
Chr15	9,977	80,300	69,494	7,418,765	63,293	1,431	35,554	*SPG11*
Chr16	18,147	96,117	98,427	7,965,443	91,332	1,968	53,747	*FUS*
Chr17	26,808	141,261	114,911	7,384,387	103,408	2,335	61,934	*GRN, MAPT, PFN1, TAF15*
Chr20	11,325	40,274	48,909	5,230,034	49,143	1,040	29,736	*SS18L1, VAPB*
Chr21	6,729	33,173	24,412	3,401,489	23,602	552	14,375	*SOD1*
Chr22	12,867	61,936	43,890	3,485,218	42,840	973	25,695	*CHCHD10, EWSR1, KEFH*

The following figures represent the SNP found with different types of effects for 69 patients (Figure 3.2) in our study.



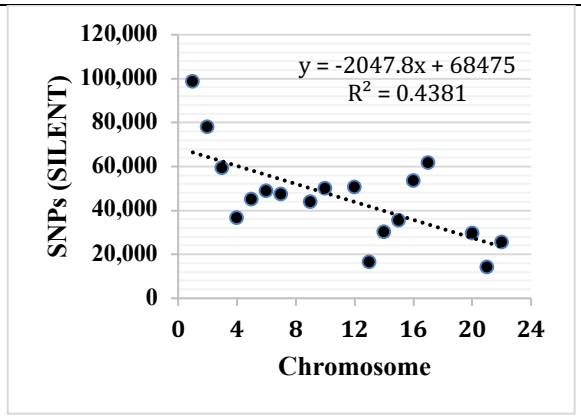


Figure 3.2. SNP distribution by chromosome in the case of 69 samples.

The analysis of various chromosomes revealed a correlation between the number of variants and the corresponding chromosome size. However, this correlation is not perfect. For example, chromosome 5 displayed 79,668 missense variants, while chromosomes 4 and 6 exhibited 68,738 and 90,827 missense variants, respectively. These findings indicate that the size of a chromosome does not serve as a reliable predictor of variant abundance. This inconsistency is also expected because we are capturing true functional genetic material not purely the amount of genetic material. To substantiate these unusual observations, a comprehensive literature mining was conducted for QC analysis. Specifically, the international SNP Map Working Group, led by Sachidanandam et al. (2001), mapped the sequence flanking each SNP through alignment to the genome sequence of the largest-insert clones in the Genbank database. Their comprehensive mapping efforts resulted in a description of 1.42 million SNPs distributed throughout the human genome (Sachidanandam et al., 2001). Although examining the total SNPs present in each chromosome, along with their respective lengths (as depicted in Table 3.2), revealed a coefficient of determination (R^2) of 0.8561 in Figure 3.3, the literature provided validation of our findings, citing 117,882 SNPs for chromosome 5, and

84,426 and 96,317 SNPs for chromosomes 4 and 6, respectively. These findings corroborate our data and the reliability of the code used to identify variants.

Table 3.2. SNPs distributed in each chromosome (Sachidanandam et al., 2001).

Chromosome	Length (bp)	All SNPs
1	214,066,000	129,931
2	222,889,000	103,664
3	186,938,000	93,140
4	169,035,000	84,426
5	170,954,000	117,882
6	165,022,000	96,317
7	149,414,000	71,752
8	125,148,000	57,834
9	107,440,000	62,013
10	127,894,000	61,298
11	129,193,000	84,663
12	125,198,000	59,245
13	93,711,000	53,093
14	89,344,000	44,112
15	73,467,000	37,814
16	74,037,000	38,735
17	73,367,000	34,621
18	73,078,000	45,135
19	56,044,000	25,676
20	63,317,000	29,478
21	33,824,000	20,916
22	33,786,000	28,410
X	131,245,000	34,842
Y	21,753,000	4,193
RefSeq	15,696,674	14,534
Totals	2,710,164,000	1,419,190

The following figure represents the SNP distribution by chromosome in the literature (Figure 3.2).

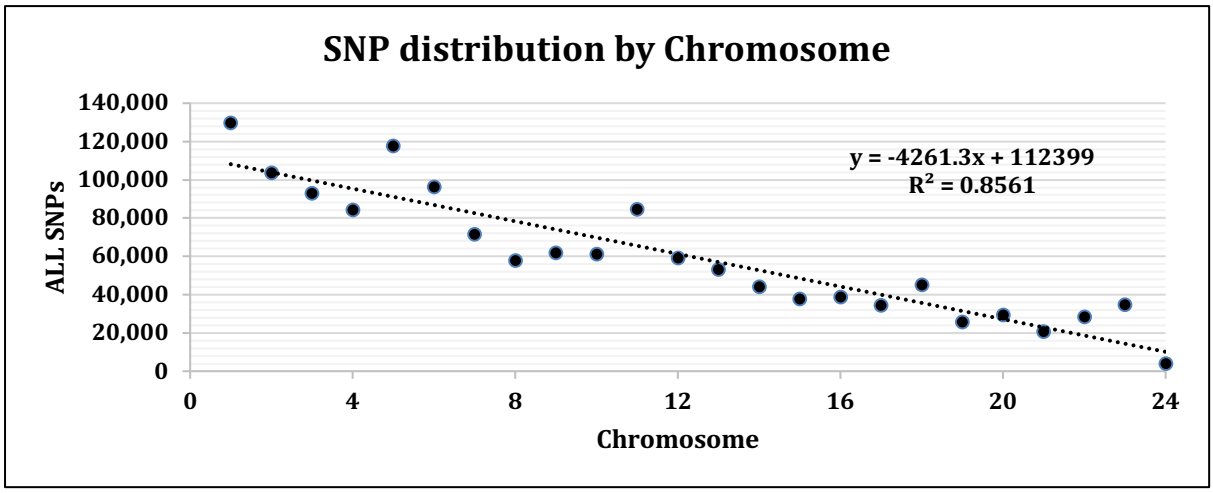


Figure 3.3. SNP distribution by Chromosome.

3.3.2. R code extract_clinvar: Extract the variants related to ALS

The R code *extract_clinvar* extracts the variants and genes related to a known genetic disorder or a clinical feature from the ClinVar database of NCBI. It annotates the phenotypes from GeneReview, MedGen, and OMIM. The alias of a disease/phenotype is considered in the HPO database. Furthermore, the variants on a use-defined gene list can be captured at the same time. First, missense_variants were sorted out from the individual chromosome data acquired from the whole SNP of 69 patients described in section 2.1.3.2. The codes were verified through RStudio for a single gene e.g. *SOD1* for missense_variants (Section 2.1.3.4.1.2). These variants were then compared with the ClinVar database to find out the pathogenic variant in the data set. The same analysis was done for high, low, moderate and modifier variants.

3.3.3. Code running in the iceberg (currently ShARC) and troubleshooting

3.3.3.1. Part 1: Specific variant containing csv file

The code was changed in the grep command line for sorting out different variants (Section 2.1.3.4.1.2.a).

For example,

```
sod1vcf.Ranges11 <- sod1vcf.Ranges[,12]


sod1indexmm <- grep ("missense_variant", sod1vcf.Ranges11)
sod1vcf.mis <- sod1vcf.Ranges[c(sod1indexmm),]
write.csv(sod1vcf.mis, file = "sod1misvcf.csv")
```

After sorting out the csv file with all the missense_variant for the *SOD1* gene range, the specific MODERATE impact missense_variants were sorted out using the following command:

```
sod1indexmm <- grep ("MODERATE", sod1vcf.Ranges11)
sod1vcf.mm <- sod1vcf.mis[c(sod1indexmm),]
write.csv(sod1vcf.mm, file = "sod1mmvcf.csv")
```

Or, HIGH impact missense_variants were sorted out using the following command:

```
sod1indexmh <- grep ("HIGH", sod1vcf.Ranges11)
sod1vcf.mh <- sod1vcf.mis[c(sod1indexmm),]
write.csv(sod1vcf.mh, file = "sod1mhvcf.csv")
```

3.3.3.2. Part 2: Extract_clivar

The localPDB downloads the necessary files from the public databases including ClinVar. The variants associated with 'Amyotrophic lateral sclerosis' were extracted separately to compare with the specific variant containing csv file and compiled in a new csv file. The new csv file contains variants from 69 patient samples that are reported in ClinVar (Table 3.3, 3.4 and 3.5).

3.3.3.3. Part 3: Identifying Patients ID

For identifying the patient's ID the general code verified in RStudio worked fine for chromosomes 4-22. But the command line started to break for chromosomes with bigger sizes chromosomes 1-3). The vcf file of Chromosome 1-3 with variants of 69 patients was divided into 2 smaller files to solve the problem.

For example,

```
chr1Gvcf <- fread("Galaxy-chr1.vcf.gz")
chr1Gvcf1 <- chr1Gvcf[,3]

chr1Gvcf2 <- data.frame(chr1Gvcf1)
nrow(chr1Gvcf1)
[1] 9501799
```

So, the chromosome was divided into two files: one containing row 1-4750899 (half of 9501799) and another containing row 4750900-951799.

```
chr1Gvcf1a <- chr1Gvcf1[1:4750899,2]
```

```
chr1Gvcf2a <- data.frame(chr1Gvcf1a)

chr1Gvcf1b <- chr1Gvcf1[4750900:9501799,]
chr1Gvcf2b <- data.frame(chr1Gvcf1b)
```

The csv file with variants found compared with ClinVar (Section 2.1.3.4.1.2.b) was matched with both parts of the chromosome files (vcf2a and vcf2b) to find out the positive patient's ID.

3.3.4. ALS gene mutations in WGS of motor cortex sample

The annotated variants reported in ClinVar found within the 69 patients' samples were also checked with the variant information in Project MinE and the gnomAD browser to confirm if there were any novel variants (MinE). Within the 42 genes (Table 2.2) related to ALS that have been analysed, 63 variants out of the 1,469,317 missense variants identified were reported in ClinVar (output of section 3.3.3.2). There are 48 Benign/Likely benign, 8 Uncertain significance, 4 Conflicting interpretations of pathogenicity, and 3 pathogenic variants (last updated analysis date: 1 July 2020).

3.3.4.1. Pathogenic variants

Three pathogenic variants in four patients' motor cortex samples were identified through data analysis: Two variants in the *SOD1* gene and one from the *FUS* gene (Table 3.3). The *SOD1*(c.341T>C;p.Ile114Thr) pathogenic variant was found in 2 patients samples while the *SOD1* (c.302A>G;p.Glu101Gly) was only present in one patient sample. The SOD1 (c.341T>C;p.Ile114Thr) variant was first reported as a pathogenic in 1993 (Rosen et al.,

1993), first reported in ClinVar on July 1, 2017, and last evaluated on December 18, 2021, as Pathogenic and Likely pathogenic on July 07, 2022 (VCV000197145.13, VCV000197145.14). The second *SOD1* (c.302A>G;p.Glu101Gly) variant was found in one sample. This variant was first reported in ClinVar on April 30, 2016, and last evaluated on October 14, 2021, as Pathogenic (VCV000014761.5). Finally, the *FUS* mutation (c.1520G>A;p.Gly507Asp) was last evaluated as a pathogenic variant on April 1, 2010, and was first reported in ClinVar on August 4, 2017 (Corrado et al., 2010, Hewitt et al., 2010b, VCV000016226.1).

Table 3.3. Pathogenic variants were identified for 69 motor cortex samples. The table shows the number of patients with each of the variants as well as the allele frequencies (AF) from Project MinE (MinE) and the gnomAD databases containing whole genome sequencing (ge) and whole exome sequencing (ex). comparing with ClinVar information. NA = Frequency hasn't been reported yet.

Gene	Variants	ClinSigSimple	Number of patients with variant	AF.all	AF.cases	AF.controls	gnomAD. ge.AF	gnomAD. ex.AF
FUS	c.1520G>A; p.Gly507Asp	Pathogenic	1	0.000080	0.000115	0	NA	NA
SOD1	c.302A>G; p.Glu101Gly	Pathogenic	1	NA	NA	NA	NA	NA
SOD1	c.341T>C; p.Ile114Thr	Pathogenic	2	0.000484	0.000687	0	NA	0.0000487

The pathogenic variants were compared with prior sequencing from different tissue within the same patients to determine whether mutations were examples of somatic mutations.

3.3.4.2. Conflicting interpretations of pathogenicity, Uncertain Significance

Along with 4 variants of 'conflicting interpretations of pathogenicity', 8 variants of 'uncertain significance' were identified and are listed in the following **Table 3.4**. Interestingly some of the variants show a high CADD score that indicated the variants are more likely to functionally impactful (Kircher et al., 2014b). Higher scores are more likely to be deleterious. Scores are determined as 10 * -log of the rank. So, variants with scores above 20 are predicted to be among the 1.0% most deleterious possible substitutions in the human genome (Rentzsch et al., 2019).

Table 3.4. Summary of 4 variants of conflicting interpretation and 8 uncertain significance variants identified within the 69 motor cortex samples. The table shows the number of patients with each of the variants as well as the allele frequencies (AF) from Project MinE (MinE) and the gnomAD databases containing whole genome sequencing (ge) and whole exome sequencing (ex). comparing with ClinVar information. NA = Frequency hasn't been reported yet.

Gene	Variants	ClinSigSimple	Number of patients with variant	CADD score	AF.all	AF.cases	AF. controls	gnomAD .ge.AF	gnomAD. ex.AF
GRN	c.1253G>A; p.Arg418Gln	Conflicting interpretations of pathogenicity Likely benign(2); Uncertain significance(2)	1	**11.31**	0.00040	0.00057	0	0.00019	0.00025
MAPT	c.1280C>T; p.Ser427Phe	Conflicting interpretations of pathogenicity Likely benign(2); Uncertain significance(1)	1	**26.3**	0.00178	0.00172	0.00191	0.00165	0.00160
NEK1	c.782G>A; p.Arg261His	Conflicting interpretations of pathogenicity	1	**25.8**	0.00653	0.00779	0.00355	0.00217	0.00236

PRPH	c.421G>T; p.Asp141Tyr	Conflicting interpretations of pathogenicity Likely benign(1); Uncertain significance(2)	1	**26.3**	0.00613	0.00596	0.00655	0.00355	0.00318
CHCHD10	c.31C>G; p.Arg11Gly	Uncertain significance	1	**26.6**	0.00008	0.00011	0	NA	NA
GBA2	c.2220C>A; p.Ser740Arg	Uncertain significance	1	**14.13**	0.00064	0.00068	0.00054	0.00035	0.00047
MAPT	c.671T>G; p.Val224Gly	Uncertain significance	1	**21.7**	0.00387	0.00332	0.00519	0.00178	0.00219
SETX	c.1750C>G; p.Leu584Val	Uncertain significance	1	**22.4**	NA	NA	NA	NA	NA
SETX	c.6172A>C; p.Lys2058Gln	Uncertain significance	1	**22.4**	NA	NA	NA	NA	NA
SPG11	c.336A>C; p.Glu112Asp	Uncertain significance	1	**9.48**	NA	NA	NA	NA	NA
SQSTM1	c.332C>T; p.Pro111Leu	Uncertain significance	1	**18.19**	NA	NA	NA	NA	NA
SQSTM1	c.80C>T; p.Pro27Leu	Uncertain significance	1	**29.8**	NA	NA	NA	NA	NA

3.3.4.3. Benign/Likely benign variants

According to the guidance developed by the American College of Medical Genetics and Genomics (ACMG), benign and likely benign variants are two of the five categories in the variant classifications. This means the variants listed here are weighted as stand-alone (BA1), strong (BS1–4) or supporting (BP1–6) (Section 2.1.3.4.1.2b).

Table 3.5. Summary of the 48 Benign/Likely benign variants identified within the 69 motor cortex samples. The table shows the number of patients with each of the variants as well as the allele frequencies (AF) from Project MinE (MinE) and the gnomAD databases containing whole genome sequencing (ge) and whole exome sequencing (ex). comparing with ClinVar information. NA = Frequency hasn't been reported yet.

Gene	Variants	ClinSigSimple	Number of patients with variant	AF.all	AF.cases	AF.controls	gnomAD.ge.AF	gnomAD.ex.AF
CHCHD10	c.403T>C; p.Tyr135His	Benign/Likely benign	2	0.0003	0.0003	0.0002	0.0001	0.0003
CHCHD10	c.424T>C; p.Tyr142His	Benign/Likely benign	2	NA	NA	NA	NA	NA
FIG4	c.1090A>T; p.Met364Leu	Benign/Likely benign	1	0.0327	0.0323	0.0336	0.0500	0.0789

FIG4	c.1961T>C; p.Val654Ala	Benign	11	NA	NA	NA	NA	NA
GBA2	c.946G>A; p.Gly316Arg	Benign/	2	0.0123	0.0133	0.0098	0.0070	0.0076
KIF5A	c.2957C>T; p.Pro986Leu	Benign/Likely benign	6	0.0178	0.0193	0.0142	0.0124	0.0112
MAPT	c.1108C>T; p.Arg370Trp	Benign	19	NA	NA	NA	NA	NA
MAPT	c.1321T>C; p.Tyr441His	Benign	19	NA	NA	NA	NA	NA
MAPT	c.1339T>C; p.Ser447Pro	Benign	19	0.2250	0.2200	0.2360	0.1240	0.1460
MAPT	c.605C>T; p.Pro202Leu	Benign	19	0.2250	0.2200	0.2360	0.1240	0.1460
MAPT	c.637G>A; p.Gly213Arg	Benign	1	0.0012	0.0010	0.0016	0.0097	0.0033
MAPT	c.689A>G; p.Gln230Arg	Benign	9	NA	NA	NA	NA	NA
MAPT	c.853G>A; p.Asp285Asn	Benign	19	NA	NA	NA	NA	NA
MAPT	c.866T>C; p.Val289Ala	Benign	19	NA	NA	NA	NA	NA
MAPT	c.953C>T; p.Ser318Leu	Benign	1	NA	NA	NA	NA	NA
NEFH	c.1387G>A; p.Glu463Lys	Benign	8	0.0860	0.0859	0.0862	0.0687	0.0706
NEFH	c.2414A>C; p.Glu805Ala	Benign	19	NA	NA	NA	NA	NA
NEFH	c.745G>A; p.Gly249Ser	Benign/Likely benign	1	0.0036	0.0034	0.0040	0.0102	0.0122
NEK1	c.1313C>T; p.Ala438Val	Benign/Likely benign	16	NA	NA	NA	NA	NA
NEK1	c.1388C>T; p.Ala463Val	Benign/Likely benign	16	0.0596	0.0593	0.0603	0.0322	0.0356

NEK1	c.1585G>A; p.Ala529Thr	Benign/Likely benign	12	NA	NA	NA	NA	NA
NEK1	c.1660G>A; p.Ala554Thr	Benign/Likely benign	12	NA	NA	NA	NA	NA
NEK1	c.1744G>A; p.Ala582Thr	Benign/Likely benign	12	NA	NA	NA	NA	NA
NEK1	c.1792G>A; p.Ala598Thr	Benign/Likely benign	12	NA	NA	NA	NA	NA
NEK1	c.1876G>A; p.Ala626Thr	Benign/Likely benign	12	0.0791	0.0818	0.0726	0.0497	0.0513
NEK1	c.1964A>G; p.Glu655Gly	Benign	14	NA	NA	NA	NA	NA
NEK1	c.2039A>G; p.Glu680Gly	Benign	14	NA	NA	NA	NA	NA
NEK1	c.2123A>G; p.Glu708Gly	Benign	14	NA	NA	NA	NA	NA
NEK1	c.2171A>G; p.Glu724Gly	Benign	14	NA	NA	NA	NA	NA
NEK1	c.2255A>G; p.Glu752Gly	Benign	14	NA	NA	NA	NA	NA
OPTN	c.293T>A; p.Met98Lys	Benign	2	0.0284	0.0294	0.0259	0.0616	0.0436
PRPH	c.26G>A; p.Arg9Gln	Benign	1	NA	NA	NA	NA	NA
SETX	c.1979C>G; p.Ala660Gly	Benign	7	0.0650	0.0651	0.0650	0.1300	0.1160
SETX	c.2975A>G; p.Lys992Arg	Benign	1	0.0152	0.0150	0.0156	0.0167	0.0155
SETX	c.3455T>G ; p.Phe1152Cys	Benign	3	0.0369	0.0362	0.0388	0.0562	0.0646

SETX	c.4660T>G; p.Cys1554Gly	Benign/Likely benign	1	0.0029	0.0032	0.0021	0.0029	0.0055
SETX	c.5563A>G; p.Thr1855Ala	Benign	19	NA	NA	NA	NA	NA
SETX	c.7640T>C; p.Ile2547Thr	Benign/Likely benign	1	0.0058	0.0065	0.0040	0.0026	0.0039
SIGMAR1	c.5A>C; p.Gln2Pro	Benign	19	0.1670	0.1710	0.1560	0.1680	0.2020
SPG11	c.1108G>A; p.Glu370Lys	Benign/Likely benign	2	0.0186	0.0182	0.0194	0.0219	0.0174
SPG11	c.1388T>C; p.Phe463Ser	Benign	37	NA	NA	NA	NA	NA
SPG11	c.2083G>A; p.Ala695Thr	Benign	2	0.0130	0.0121	0.0150	0.0188	0.0131
SPG11	c.3037A>G; p.Lys1013Glu	Benign	1	0.0175	0.0173	0.0180	0.0061	0.0102
SQSTM1	c.460A>G; p.Lys154Glu	Benign/Likely benign	1	NA	NA	NA	NA	NA
SQSTM1	c.570G>C; p.Glu190Asp	Benign/Likely benign	2	NA	NA	NA	NA	NA
SQSTM1	c.712A>G; p.Lys238Glu	Benign/Likely benign	1	0.0058	0.0061	0.0051	0.0014	0.0027
SQSTM1	c.822G>C; p.Glu274Asp	Benign/Likely benign	2	NA	NA	NA	NA	NA
TBK1	c.1391T>C; p.Val464Ala	Benign/Likely benign	1	NA	NA	NA	NA	NA

3.3.4.4. Patients' ID

The last part of the R code involves finding the patient's reference ID that is associated with each of the identified pathogenic variants. The patient's reference IDs are listed in **Table 3.6** which are found for *SOD1* (n=3) and *FUS* (n=1).

Table 3.6. Clinically verified pathogenic variants found in 69 motor cortex samples from Project MinE.

Patients ID	Chr	Position	Ref	Gene	Format
LP6008464-DNA_C01	21	33039633	chr21:33039633:A:G	*SOD1*	p.Glu101Gly
LP6008464-DNA_F07	21	33039672	chr21:33039672:T:C	*SOD1*	p.Ile114Thr
LP6008465-DNA_E01	21	33039672	chr21:33039672:T:C	*SOD1*	p.Ile114Thr
LP6008464-DNA_G11	16	31202410	chr16:31202410:G:A	*FUS*	p.Gly507Asp

3.3.5. Genotype-Phenotype Correlations

A summary of the clinical features of the specific ALS patients retrieved with Sheffield Brain Tissue Bank (SBTB) number from the Brain Bank is shown in **Table 3.7**. The clinical information shows all the variants found in motor cortex samples were reported previously for each of the patients.

Table 3.7. Clinical information for the patients with pathogenic ALS variants. Abbreviations are Female (f), and Male (m).

Patients ID	Sex	Age at onset	Length of Illness months	Age at death	Site of disease onset	Variant of disease	Familial cases	Notes on family history	The mutation identified from blood/ cerebellum	Cognitive Impairment
LP6008464-DNA_C01	f	37	41	40	Limb	ALS	Familial	None	*SOD1* mutation (p.Glu101Gly)	None
LP6008464-DNA_F07	f	65	14	66	Limb	ALS	Sporadic	None	*SOD1* mutation (p.Ile114Thr)	None
LP6008465-DNA_E01	f	59	12	60	Limb	ALS	Familial	Mother had PMA	*SOD1* mutation (p.Ile114Thr)	None
LP6008464-DNA_G11	m	69	38	72	Limb	ALS, LMN	Sporadic	None	*FUS*/TLS mutation (p.Gly507Asp)	None

3.3.5.1. Case 1

This case had the *SOD1* p.Glu101Gly (previously identified as E100G) mutation. The patient was a female and had disease onset at 37 years. The site of disease onset was the limb and she died after a disease duration of 41 months at the age of 40. The clinical information showed her having familial ALS. From the blood/cerebellum sample of the patient *SOD1* mutation was identified. The motor cortex sample also showed the same pathogenic mutation as previously found and reported for the blood sample for this patient and therefore this pathogenic variant was not a somatic mutation (Shaw et al., 1997).

3.3.5.2. Cases 2 and 3

These 2 cases had the *SOD1* p.Ile114Thr (previously identified as I113T) mutation. Interestingly both patients were female and the site of onset was a limb. One patient had no family history and was diagnosed with SALS at an age of onset of 65 and died at the age of 66 (after a disease duration of 14 months) whereas the other patient had a mother who was diagnosed with PMA and was therefore designated as having FALS. The latter patient had an onset of the disease at the age of 59 and died at 60 (after a disease duration of 12 months). Therefore, both patients had a relatively short disease duration (Shaw et al., 1997). The pathogenic mutation detected in the blood samples of these patients was also observed in their motor cortex samples, indicating that it was not a somatic mutation.

3.3.5.3. Case 4

This case had *FUS* p.Gly507Asp substitution. This patient was a male diagnosed with SALS following the development of lower motor neuron (LMN) syndrome (Hewitt et al., 2010a). His disease duration was 38 months, having developed the disease at 69 and died at 72. The pathogenic mutation identified in the patient's blood sample was present in the motor cortex sample, suggesting that it was not a somatic mutation.

In summary, the screening of these genes has so far identified 4 individuals with mutations, for whom we knew the mutations from sequencing of DNA extracted from blood and/or cerebellum. This confirms our approach and validates the analysis pipeline being used. However, we have not identified any significant evidence for somatic heterogeneity in these 69 patients.

3.3.6. *C9ORF72*

In a previous study, *C9ORF72* HRE screening was done for these 69 samples (Cooper-Knock et al., 2012). To present the complete genetic picture of these cases, the results of the C9orf72 HRE screening that had been done previously (Cooper-Knock et al., 2012) were also included. Table 3.8 shows the *C9ORF72* intronic expansion was present in 8 cases among the 69 samples considered for this study.

Table 3.8. Patients with the hexanucleotide repeat expansion of *C9ORF72.*

Patients ID	Sex	Site_ of_Onset	*C9ORF72*_ Status_ consensus	Age_ at_onset_ years	Age_ at_death_ years
LP6008464-DNA_A01	m	spinal	expanded	45	46
LP6008464-DNA_A02	m	spinal	expanded	67	79
LP6008464-DNA_C02	m	bulbar	expanded	62	64
LP6008465-DNA_B03	f	spinal	expanded	72	76
LP6008465-DNA_D01	m	spinal	expanded	70	72
LP6008465-DNA_E04	m	bulbar	expanded	67	68
LP6008465-DNA_F02	m	spinal	expanded	63	64
LP6008465-DNA_G03	m	generalized	expanded	47	48

3.4. Discussion

Mutations in more than 120 genes have been shown to cause or increase the risk of ALS (Ryan and Deng, 2021). Germline mutations in several genes have also been identified. Patients who do not have germline mutations in ALS genes have been hypothesized to have somatic mutations as a possible cause of ALS (Leija-Salazar et al., 2018).

In 2021, Hisahara et al. provide genetic evidence of somatic mosaicism of a novel *FUS* frameshift mutation in an ALS patient. The patient (*FUS* case 1) transmitted the mutant *FUS* to her son (*FUS* case 2). The age at onset in case 1 was reported as 44 years of age. Case 1 had invasive ventilation marking the disease's endpoint 6 years later. On the

other hand, case 2 showed a much earlier disease onset around 11 years of age. Case 2 required invasive ventilation less than 2 years after disease onset due to rapid disease progression. The WES analysis of the DNA isolated from the case 2 blood sample revealed a novel heterozygous 4 nucleotide deletion in *FUS* (c.1542_1545delGGGT, p.Gly515Serfs13*). This explains the consistency of clinical features for case 2. Because patients with a truncated FUS protein generally show an early onset and rapid progression of the disease.

The study by Müller et al. has addressed the issue of somatic mosaicism which is linked to *de novo* mutations (Muller et al., 2022) (section 1.4). The patients described in this report were the product of germline mutations, although there is a possibility of similar *de novo* mutations at any stage of nervous system development (Cooper-Knock, 2022). There is evidence that *de novo* mutations in *SOD1* may be the cause of different cases of SALS (Muller et al., 2022). Genomic *de novo* mutations have also been reported in single SALS cases in *ATXN2, SS18L1, CHRM1, ERBB4, VCP* and *RAPGEF2* genes raising the question: how frequent are *de novo* mutations as a cause of ALS (Nicolas and Veltman, 2019)?

An old study of the epidemiology of sporadic ALS supported the conclusion that accrued somatic mutations might trigger the onset of ALS (Armon, 2016). Both developmental mutations and somatic mosaicism would play a role in ALS risk only if the originating events in the focal cells derived from genetic (or epigenetic) changes in those cells (Frank, 2010). There is no direct evidence that exists for the role of somatic mutation in initiating ALS. Inherited germ-line mutations do strongly predispose to

ALS, so genetics may play a key role in the disease (Armon, 2003, Kabashi et al., 2008, Sreedharan et al., 2008, Strong et al., 1991, Valdmanis and Rouleau, 2008, Van Deerlin et al., 2008). However, the direct role of somatic mutation, and what follows, remains a strong hypothesis at present.

Repeat sequences are interesting for somatic mutation analysis because their secondary structures can lead to the expansion or contraction of repeat lengths (Thys and Wang, 2015).

Currently, *C9ORF72* hexanucleotide repeat expansion (HRE) is one of the most prevalent and penetrant causes of ALS (Taylor et al., 2016). In the general population, *C9ORF72* contains less than 30 GGGGCC repeats in the first intron. On the other hand, in ALS cases the number of repeats ranges between hundreds to thousands (Taylor et al., 2016). Many aspects of *C9ORF72*-related ALS have not been thoroughly investigated as it is difficult to quickly and precisely size the repeat length above 30 (Pamphlett et al., 2013). Nevertheless, a Southern hybridisation-based protocol has been optimized to allow for confident detection of *C9ORF72* repeat expansions, as well as the independent assessment of their heterogeneity and the number of repeat units (Buchman et al., 2013). In their study, Buchman et al. (2013) found that somatic cell genomes of ALS patients contain a variable number of GGGGCC repeats in the *C9ORF72* locus. Additionally, Southern hybridisation analysis of genomic DNA extracted from peripheral blood, cerebellum, cortex, and a lymphoblastoid cell line of patients with GGGGCC repeat expansion revealed a different pattern of high molecular mass fragments for each DNA sample from the same patient (Buchman et al., 2013).

However, the study conducted by Cooper-Knock et al. did not demonstrate the presence of somatic mutations in these 69 patients (Cooper-Knock et al., 2012).

Some interesting candidate mutations within ALS genes have been identified for example *CHCHD10, NEK1* and *SETX* which have conflicting evidence in ClinVar, with higher CADD scores. Mutations in the *CHCHD10* gene have been considered to be a frequent cause of ALS. In 2018, three rare missense mutations in *CHCHD10* were identified specific to ALS cases. The previously unreported p.Arg11Gly variant also found in my study was identified earlier in a single female ALS case from the United States without cognitive involvement and a negative family history of ALS or dementia (Project MinE, 2018a). The novel ALS-specific variant, a heterozygous c.31C>G variant resulting in a p.Arg11Gly amino acid change, was found only in a single ALS case beforehand and is therefore of unknown significance.

NEK1 has previously been identified as a potential gene associated with ALS (Cirulli et al., 2015, Brenner et al., 2016). The *NEK1* p.Arg261His variant detected in this study is located near the protein kinase domain and has been classified as harmful by most bioinformatics prediction algorithms (CADD, GERP, LRT, Mutation Assessor, MutationTaster, PolyPhen, PROVEAN, SIFT, and SiPhy). In 2016, autozygosity mapping revealed a *NEK1* p.Arg261His variant as a potential risk factor in an isolated Dutch community (Kenna et al., 2016). Subsequent studies have also supported p.Arg261His as the only *NEK1* missense variant associated with increased ALS risk (Kenna et al., 2016, Nguyen et al., 2018b, Lattante et al., 2021). Previous research has reported p.Arg261His frequencies ranging from 0.6 to 1.7% in patients with ALS (Kenna et al.,

2016, Nguyen et al., 2018b, Lattante et al., 2021), while recent data has suggested a higher frequency of 3.7% (Riva et al., 2022). Further research is needed to evaluate the role of other *NEK1* missense variants in the pathogenesis of ALS.

Two rare deleterious variants have been found in Senataxin (*SETX*) gene that causes an autosomal dominant neuromuscular disorder, amyotrophic lateral sclerosis 4 (ALS4) (de Carvalho, 2023). The variants c.1750C>G; p.Leu584Val and c.6172A>C; p.Lys2058Gln were reported in 2017 for the first time (Cooper-Knock et al., 2017). The latter one of the *SETX* mutations lies within a helicase domain which contains several previously reported mutations (Hirano et al., 2011).

A total of 42 ALS genes have been screened for mutations in the motor cortex samples and three *SOD1* (p.Glu101Gly, p.Ile114Thr) and *FUS* (p.Gly507Asp) pathogenic mutations were identified. The clinical information shows all the variants found in motor cortex samples are reported earlier for the patients. Therefore, we have not identified evidence for somatic heterogeneity for these 42 genes.

Chapter 4. Genetic profiling for Bangladeshi MND patients and different MMA cohorts

4.1. Introduction

To establish the genetic basis of Bangladeshi patients presenting with MND at the National Institute of Neurosciences & Hospital (NINS), Bangladesh, DNA was extracted from blood samples sent to SITraN. The DNA was sent to the Sheffield Diagnostics Genetics Service (SDGS) to be sequenced according to the protocol used by the NHS diagnostic laboratory (https://www.sheffieldchildrens.nhs.uk/sdgs/laboratory-services/). Sequencing of 28 patients' samples was done using the Neurogenetic Motor Disorder NGS Panels (total 153 genes) that also include the Familial Amyotrophic Lateral Sclerosis with or without Frontotemporal Dementia 42 gene panel (24 male, 4 female). Five samples from a total of 33 samples shown in **Table 2.5** were excluded because of the low quality and quantity of the DNA and lack of information regarding the sample. The data were trimmed and mapped to the hg19 reference genome. Among the 28 samples, 11 monomelic atrophy (BD MMA) with arm onset patients were considered as a subset for MMA analysis and these were compared with eight UK MMA with arm onset patients' data, which had previously been sequenced using the same gene panel at the SDGS.

4.2. Aim

This study aims to use Next Generation Sequencing (NGS) to understand the proportion of Bangladesh MND patients carrying known pathogenic variants in 153

MND-related genes including *C9ORF72, TARDBP, SOD1* and *FUS*. Moreover, previously unpublished pathogenic variants within the known ALS gene along with novel genetic causes of ALS may be identified. Additionally, it also aims to screen MMA patients from Bangladesh and the UK for mutations in genes associated with motor system disorders, to determine whether MMA shares a genetic architecture with other disorders including ALS.

4.3. Results

4.3.1. Bangladeshi MND Samples

Samples from Bangladeshi MND (BD-MND) patients were processed for DNA extraction followed by sequence and variant analysis. The gender and diagnosis of the patient samples were listed in table 2.5.

4.3.2. Quality and quantity of extracted DNA

As the quality of sequencing results depend greatly on the purity and concentration of the template DNA, it is very important to evaluate the quality and quantity of the extracted DNA to be sequenced. Low DNA concentration leads to low raw data signals.

DNA alterations for example breaks, thymine dimers, or abasic sites prevent DNA synthesis during polymerase-dependent reactions needed for sequencing. Therefore, DNA damage reduces the robustness of results, increases standard deviation, and induces complications in interpreting data of genome analyses (Betge et al., 2015).

The ideal final yield was considered to be. 350-400µl with a concentration of 30-50ng/µl. The 260/280 ratio should be 1.7 or higher according to the manufacturer's guidelines. Anything less than these values were considered to have a problem with the DNA purification and the extraction procedure was repeated for those samples.

The Neurogenetic Motor Disorder NGS Panels sequencing required 3606ng (minimum amount 1300ng) DNA that was quantified using the Nanodrop spectrophotometer and the best of the duplicate samples was chosen to send for sequencing. Table 4.1 showed the absorbance ratio and the amount of DNA sent for sequencing with the required concentration of 3606ng. Those highlighted in yellow in Table 4.1 were sent for sequencing.

Table 4.1. The quantity and quality of the Bangladeshi MND DNA samples. Samples highlighted in yellow were sent for sequencing.

Sample ID		ng/µl	260/280 (Absorbance ratio)	The amount required for sequencing (concentration required 3606ng; minimum 1300ng)
BD-MND-001	i	44.42	1.91	81.18
	ii	60.05	1.82	60.05
BD-MND-002	i	54.47	1.80	66.20
	ii	60.41	1.78	59.69
BD-MND-003	i	68.83	1.80	52.39
	ii	76.69	1.75	47.02
BD-MND-004	i	45.4	1.78	79.43
	ii	52.79	1.43	68.31
BD-MND-005	i	31.62	1.94	114.04
	ii	34.24	1.80	105.32
BD-MND-006	i	58.12	1.86	62.04
	ii	82.28	1.94	43.83
BD-MND-007	i	61.86	1.87	58.29
	ii	33.27	1.69	108.39
BD-MND-008	i	88.12	1.91	40.92
	ii	42.77	1.82	84.31
BD-MND-009	i	33.95	1.68	106.22
	ii	37.35	1.79	96.55
BD-MND-010	i	56.14	1.90	64.23
	ii	68.69	1.84	52.50
BD-MND-011	i	63.08	1.82	57.17
	ii	48.9	1.83	73.74
BD-MND-012	i	60.98	1.88	59.13
	ii	69.46	1.83	51.91
BD-MND-013	i	30.95	1.85	116.51
	ii	26.98	2.07	133.65
BD-MND-014	i	48.14	1.67	74.91
	ii	22.28	1.75	161.85
BD-MND-015	i	74.04	1.84	48.70
	ii	53.7	1.84	67.15

BD-MND-016	i	45.1	1.96	79.96
	ii	64.65	1.85	55.78
BD-MND-017	i	16.22	1.68	222.32
	ii	23.3	1.68	154.76
BD-MND-018	i	54.81	1.87	65.79
	ii	62.41	1.89	57.78
BD-MND-019	i	64.94	1.95	55.53
	ii	32.27	1.86	111.74
BD-MND-020	i	46.23	1.77	78.00
	ii	41.61	1.88	86.66
BD-MND-021	i	71.17	1.85	50.67
	ii	72.38	1.87	49.82
BD-MND-022	i	88.31	1.93	40.83
	ii	41.93	1.68	86.00
BD-MND-023	i	100.97	1.87	35.71
	ii	48.31	1.86	74.64
BD-MND-024	i	29.48	1.90	122.32
	ii	30.09	1.92	119.84
BD-MND-025	i	45.68	1.91	78.94
	ii	34.95	1.88	103.18
BD-MND-026	i	32.35	1.84	111.47
	ii	49.97	1.95	72.16
BD-MND-027	i	114.1	1.87	31.60
	ii	103.78	1.83	34.75
BD-MND-028	i	20.89	1.93	172.62
	ii	49.55	1.76	72.77
BD-MND-029	i	73.37	1.86	49.15
	ii	33.05	2.01	109.11
BD-MND-030	i	57.04	1.92	63.22
	ii	47.2	1.90	76.40
BD-MND-031	i	21.96	1.97	164.21
	ii	22.68	1.88	158.99
BD-MND-032	i	28.03	2.00	128.65
	ii	16.74	1.69	215.41
BD-MND-033	i	53.16	1.96	67.83
	ii	49.87	1.92	72.31

4.3.3. Variants reported by Neurogenetic Motor Disorder NGS Panels

The samples were sent to SDGS in Sheffield Children's NHS Foundation Trust for sequencing in July 2020. The excel sequence result files were shared in a drive through Unidrive (uosfstore.shef.ac.uk) in November 2020. The SDGS results only contain the variants in genes on the Familial Amyotrophic Lateral Sclerosis with or without Frontotemporal Dementia 42 gene panel, *C9ORF72* and *ATXN2*.

Table 4.2 showed the 99 variants reported by SDGS for 28 BD MND samples from the sequencing of Neurogenetic Motor Disorder NGS Panels. The Neurogenetic Motor Disorder NGS Panels results reported a total of 99 variants in 32 genes with 3 deletions, 1 duplication, 14 indels, 80 SNV and 1 unknown (C9ORF72) that are listed in **Supplementary Table 1**.

Table 4.2. Reported variants for 28 BD-MND samples from the Neurogenetic Motor Disorder NGS Panels. The variants in red need to be Sanger confirmed for reporting and checked for known pseudogenes.

Sample ID	cons_preferred_transcripts	id	Chromosome	Position	ref	alt	Zygosity
BD-MND-001	DCTN1:NM_004082:24:c.2761-18C>T:NA	rs549475401	chr2	74593163	G	A	HET
	ERBB4:NM_005235:6:c.645A>G:p.E215E	.	chr2	212589897	T	C	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCG	HET
	GBA2:NM_020944:UTR3:NA:NA	rs186702075	chr9	35737165	C	T	HET
	ANXA11:NM_145869:10:c.905G>A:p.R302H	rs140133265	chr10	81923853	C	T	HET

	ATXN2:NM_002973:3:c.769-10A>T:NA	rs183468414	chr12	111992031	T	A	HET
	SOD1:NM_000454:4:c.341T>C:p.I114T	rs121912441	chr21	33039672	T	C	HET
	NEFH:NM_021076:4:c.1646_1687del:p.549_563del	rs773780196	chr22	29885274	AAGGCCAAGTCTCCAGCAAAGGAAGAGGCAAAGTCACCGCCTG	A	HET
BD-MND-002	FIG4:NM_014845:4:c.446+22T>-:NA	rs563921466	chr6	110048489	CT	C	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCG	HET
	HNRNPA1:NM_031157:5:c.491-7T>A:NA	rs562426518	chr12	54676171	T	A	HET
	ATXN2:NM_002973:21:c.3317-15A>G:NA	rs201513990	chr12	111902534	T	C	HET
BD-MND-003	CYP27A1:NM_000784:9:c.1514C>T:p.T505M	rs76822427	chr2	219679671	C	T	HET
	CHMP2B:NM_014043:6:c.532-4A>G:NA	rs755848821	chr3	87302858	A	G	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCGGCCCCG	HET
	SOD1:NM_000454:4:c.268G>A:p.A90T	COSM1030067	chr21	33039599	G	A	HET
	EWSR1:NM_013986:7:c.599+15T>C:NA	rs201941907	chr22	29678561	T	C	HET
BD-MND-004	GBA2:NM_020944:in_25_bp_from_exon:NA:NA	rs755266508	chr9	35740644	CTG	C	HET
	ATXN2:NM_002973:1:c.137C>A:p.A46D	rs770648742	chr12	112037182	G	T	HET
	C9ORF72:NM_001256054:intronic:NA:NA	rs71492753; rs74180757	chr9	27573521	C	CGCCCCG	HET

BD-MND-005	VCP:NM_007126:16:c.2214A>G:p.E738E	rs374391034	chr9	35057474	T	C	HET
	GBA2:NM_020944:5:c.787-18C>G:NA	.	chr9	35741079	G	C	HET
	NEFH:NM_021076:1:c.745G>A:p.G249S	rs60825978	chr22	29876996	G	A	HET
BD-MND-006	DCTN1:NM_004082:11:c.1049-21A>C:NA	rs538246858	chr2	74597692	T	G	HET
	NEK1:NM_001199397:10:c.695G>A:p.R232H	rs772747361	chr4	170506612	C	T	HET
	ARHGEF28:NM_001080479:25:c.3137A>G:p.N1046S	rs200651003	chr5	73181756	A	G	HET
	SQSTM1:NM_003900:2:c.222G>A:p.L74L	rs371268375	chr5	179249974	G	A	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCG	HET
	TAF15:NM_139215:15:c.1675_1698del:p.559_566del	rs537726014	chr17	34171977	TGGAGGAGACCGAGGTGGGGGCTAC	T	HET
BD-MND-007	DCTN1:NM_004082:22:c.2596C>G:p.L866V	rs527389133	chr2	74593618	G	C	HET
	ERBB4:NM_005235:28:c.3783G>C:p.E1261D	.	chr2	212248484	C	G	HET
	SQSTM1:NM_003900:1:c.85C>T:p.P29S	rs752506754	chr5	179248021	C	T	HET
	VCP:NM_007126:16:c.2214A>G:p.E738E	rs374391034	chr9	35057474	T	C	HET
	SS18L1:NM_198935:8:c.824-20G>A:NA	rs368502247	chr20	60740458	G	A	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCG	HET

BD-MND-008	NEFH:NM_021076:1:c.745G>A:p.G249S	rs60825978	chr22	29876996	G	A	HET
BD-MND-009	SPAST:NM_014946:13:c.1494-13T>-:NA	rs760322678	chr2	32366959	AT	A	HET
	C9ORF72:NM_001256054:intronic:NA:NA	rs71492753; rs74180757	chr9	27573521	C	CGCCCCG	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCG	HET
	DAO:NM_001917:9:c.704C>T:p.T235I	rs531892910	chr12	109292463	C	T	HET
	ATXN2:NM_002973:24:c.3913+24C>T:NA	rs370303034	chr12	111891457	G	A	HET
	VRK1:NM_003384:8:c.646G>C:p.D216H	rs531480402	chr14	97321630	G	C	HET
BD-MND-010	SPAST:NM_014946:1:c.131C>T:p.S44L	rs121908515	chr2	32289031	C	T	HET
	ARHGEF28:NM_001080479:37:c.5038T>C:p.S1680P	rs282414	chr5	73236680	T	C	HET
	SETX:NM_015046:24:c.7114G>A:p.D2372N	rs150673589; COSM1236144	chr9	135147182	C	T	HET
	SETX:NM_015046:10:c.1504C>T:p.R502W	rs534723946	chr9	135205481	G	A	HET
BD-MND-011	DCTN1:NM_004082:14:c.1482G>A:p.A494A	rs372546194	chr2	74596529	C	T	HET
	ARHGEF28:NM_001080479:3:c.54G>A:p.A18A	rs367962149	chr5	73045682	G	A	HET
	ARHGEF28:NM_001080479:30:c.3842-22T>-:NA	rs537565843	chr5	73193764	AT	A	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCG	HET
	EWSR1:NM_013986:7:c.599+15T>C:NA	rs201941907	chr22	29678561	T	C	HET

BD-MND-012	ERBB4:NM_005235:in_25_bp_from_exon:NA:NA	rs748883732	chr2	212578379	TAA	T	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCG	HET
	ATXN2:NM_002973:1:c.659C>T:p.S220F	rs562996744	chr12	112036660	G	A	HET
	SOD1:NM_000454:4:c.341T>C:p.I114T	rs121912441	chr21	33039672	T	C	HET
BD-MND-015	DCTN1:NM_004082:30:c.3597C>T:p.V1199V	rs767954436	chr2	74589789	G	A	HET
	DCTN1:NM_004082:24:c.2761-18C>T:NA	rs549475401	chr2	74593163	G	A	HET
	ALS2:NM_020919:25:c.3904C>T:p.R1302C	rs759509511	chr2	202580495	G	A	HET
	CYP27A1:NM_000784:8:c.1297C>T:p.R433W	rs143002163	chr2	219679301	C	T	HET
	ARHGEF28:NM_001080479:29:c.3673C>G:p.Q1225E	.	chr5	73190232	C	G	HET
	ARHGEF28:NM_001080479:37:c.5038T>C:p.S1680P	rs282414	chr5	73236680	T	C	HET
	C9ORF72:NM_001256054:intronic:NA:NA	rs71492753;rs74180757	chr9	27573521	C	CGCCCCG	HET
	ATXN2:NM_002973:1:c.137C>A:p.A46D	rs770648742	chr12	112037182	G	T	HET
	FUS:NM_004960:6:c.679_690del:p.227_230del	rs752712347	chr16	31196414	CGGCGGCGGTGGT	C	HET
	NEFH:NM_021076:4:c.2775G>A:p.K925K	.	chr22	29886404	G	A	HET
BD-MND-016	ANXA11:NM_145869:9:c.745-21T>A:NA	rs771194919	chr10	81925974	A	T	HET
	TBK1:NM_013254:18:c.1865A>G:p.K622R	rs199605037	chr12	64890945	A	G	HET

	SOD1:NM_000454:4:c.341T>C:p.I114T	rs121912441	chr21	33039672	T	C	HET
BD-MND-017	DCTN1:NM_004082:11:c.1049-21A>C:NA	rs538246858	chr2	74597692	T	G	HET
	SQSTM1:NM_003900:4:c.556C>T:p.R186W	rs756791819	chr5	179251206	C	T	HET
	SQSTM1:NM_003900:6:c.819_820del:p.P273fs	.	chr5	179260095	CAG	C	HET
	ATXN2:NM_002973:23:c.3708G>A:p.A1236A	rs143166155; COSM935058	chr12	111893869	C	T	HET
	NEFH:NM_021076:1:c.745G>A:p.G249S	rs60825978	chr22	29876996	G	A	HET
BD-MND-019	SQSTM1:NM_003900:2:c.240C>T:p.D80D	rs148366738	chr5	179249992	C	T	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCGGCCCCG	HET
	SIGMAR1:NM_005866:UTR3:NA:NA	rs4879809	chr9	34635598	T	A	HET
	SIGMAR1:NM_005866:4:c.622C>T:p.R208W	rs11559048	chr9	34635679	G	A	HET
	VCP:NM_007126:UTR3:NA:NA	rs766048346	chr9	35057100	C	T	HET
BD-MND-020	ARHGEF28:NM_001080479:9:c.964C>T:p.R322C	rs201283070; COSM1069963; COSM1069964	chr5	73091156	C	T	HET
	MATR3:NM_199189:16:c.2271T>C:p.D757D	rs753870064	chr5	138661251	T	C	HET
	SQSTM1:NM_003900:8:c.1277C>T:p.A426V	rs201239306; COSM1567750	chr5	179263547	C	T	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCG	HET
	ANXA11:NM_145869:6:c.322C>T:p.P108S	rs553424505	chr10	81928964	G	A	HET

	PRPH:NM_006262:1:c.26G>A:p.R9Q	rs57451017	chr12	49689009	G	A	HET
	ATXN2:NM_002973:intronic:NA:NA	rs570193403	chr12	111892868	T	C	HET
	TAF15:NM_139215:15:c.1427_1447del:p.476_483del	rs768963549	chr17	34171729	GGAGGTGGCTATGGAGGAGATC	G	HET
	NEFH:NM_021076:1:c.745G>A:p.G249S	rs60825978	chr22	29876996	G	A	HET
BD-MND-021	ALS2:NM_020919:28:c.4375G>C:p.D1459H	rs764523082	chr2	202572620	C	G	HET
	CYP27A1:NM_000784:2:c.393G>A:p.E131E	rs911534514	chr2	219674437	G	A	HET
	VCP:NM_007126:16:c.2277C>G:p.A759A	rs747361319	chr9	35057411	G	C	HET
	VCP:NM_007126:16:c.2214A>G:p.E738E	rs374391034	chr9	35057474	T	C	HET
	ATXN2:NM_002973:1:c.540delG:p.Q180fs	rs757862555	chr12	112036778	GC	G	HET
	ATXN2:NM_002973:1:c.522_538del:p.Q174fs	rs759304594	chr12	112036780	TGTTGCTGCTGCTGCTGC	T	HET
BD-MND-022	DCTN1:NM_004082:23:c.2695A>G:p.T899A	rs747285271	chr2	74593436	T	C	HET
	NEK1:NM_001199397:25:c.2140-14A>G:NA	rs775248564	chr4	170398662	T	C	HET
	NEK1:NM_001199397:15:c.1141-14T>C:NA	rs551259813	chr4	170482902	A	G	HET
	ATXN2:NM_002973:21:c.3317-15A>G:NA	rs201513990	chr12	111902534	T	C	HET
	SS18L1:NM_198935:10:c.1128G>A:p.P376P	rs115792477	chr20	60749664	G	A	HET
	ERBB4:NM_005235:22:c.2706C>T:p.D902D	rs146441207	chr2	212293146	G	A	HET

BD-MND-023	NEK1:NM_001199397:10:c.807+18T>C:NA	rs571426626	chr4	170506482	A	G	HET
	ANXA11:NM_145869:6:c.520C>T:p.P174S	rs144975299	chr10	81928766	G	A	HET
	FUS:NM_004960:12:c.1292+15->T:NA	rs572228309; rs886051937	chr16	31201734	A	AT	HET
BD-MND-024	ATXN2:NM_002973:1:c.548_549insACAGCAGCA:p.Q183delinsQQQQ	.	chr12	112036770	C	CTGCTGCTGT	HET
	SPG11:NM_025137:1:c.176C>T:p.A59V	rs552320263	chr15	44955670	G	A	HET
	GRN:NM_002087:10:c.1019A>T:p.H340L	rs775196555	chr17	42429003	A	T	HET
BD-MND-025	MATR3:NM_199189:16:c.2191T>C:p.L731L	rs149714542	chr5	138661171	T	C	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCGGCCCCG	HET
	VCP:NM_007126:16:c.2214A>G:p.E738E	rs374391034	chr9	35057474	T	C	HET
	GRN:NM_002087:7:c.626C>T:p.P209L	rs368995988	chr17	42428086	C	T	HET
BD-MND-026	SQSTM1:NM_003900:8:c.1207T>A:p.S403T	rs771657338	chr5	179263477	T	A	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCGGCCCCG	HET
	SIGMAR1:NM_005866:UTR3:NA:NA	rs4879809	chr9	34635598	T	A	HET
	DAO:NM_001917:2:c.149_154del:p.50_52del	rs546485721	chr12	109278930	GGCCTCT	G	HET
	ATXN2:NM_002973:21:c.3317-15A>G:NA	rs201513990	chr12	111902534	T	C	HET

	SPG11:NM_025137:26:c.4512A>T:p.G1504G	rs539384073	chr15	44887580	T	A	HET
	FUS:NM_004960:5:c.499_504del:p.167_168del	rs762260286	chr16	31195692	TGGTGGA	T	HET
	NEFH:NM_021076:1:c.745G>A:p.G249S	rs60825978	chr22	29876996	G	A	HET
BD-MND-027	DCTN1:NM_004082:32:c.3824G>A:p.R1275H	rs560344779	chr2	74588639	C	T	HET
	ANXA11:NM_145869:6:c.191C>T:p.T64I	rs560648660	chr10	81929095	G	A	HET
	ATXN2:NM_002973:1:c.137C>A:p.A46D	rs770648742	chr12	112037182	G	T	HET
	NEFH:NM_021076:4:c.1569G>C:p.E523D	rs138278265	chr22	29885198	G	C	HET
BD-MND-029	MATR3:NM_199189:in_25_bp_from_exon:NA:NA	rs779005896	chr5	138658266	GGT	G	HET
	SPG11:NM_025137:13:c.2377G>A:p.V793M	rs546601155	chr15	44914485	C	T	HET
	GRN:NM_002087:7:c.614C>T:p.S205L	rs777211749	chr17	42428074	C	T	HET
BD-MND-030	ALS2:NM_020919:14:c.2581-23->T:NA	rs567439355	chr2	202593929	G	GA	HET
	NEK1:NM_001199397:10:c.807+18T>C:NA	rs571426626	chr4	170506482	A	G	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCG	HET
	C9ORF72:NM_001256054:intronic:NA:NA	.	chr9	27573521	C	CGCCCCGGCCCCGGCCCCGGCCCCGGCCCCG	HET
	SETX:NM_015046:12:c.5502C>T:p.H1834H	rs117410554	chr9	135176063	G	A	HET
	OPTN:NM_001008211:14:c.1466A>G:p.K489R	.	chr10	13174131	A	G	HET

	FUS:NM_004960:12:c.1292+15->T:NA	rs572228309; rs886051937	chr16	31201734	A	AT	HET
BD-MND-032	ALS2:NM_020919:29:c.4580+22G>A:NA	rs550794958	chr2	202571547	C	T	HET
	SETX:NM_015046:24:c.7114G>A:p.D2372N	rs150673589; COSM1236144	chr9	135147182	C	T	HET
	SPG11:NM_025137:24:c.4052G>A:p.R1351K	rs564171625	chr15	44889091	C	T	HET
BD-MND-033	ANXA11:NM_145869:15:c.1311G>A:p.A437A	rs140115990	chr10	81917747	C	T	HET
	SPG11:NM_025137:19:c.3362C>G:p.T1121S	rs552471760	chr15	44900733	G	C	HET

4.3.4. Variants analysed through RStudio Pipeline

The variants reported by SDGS only contain the variants of Familial Amyotrophic Lateral Sclerosis with or without Frontotemporal Dementia 42 gene panel, *C9ORF72* and *ATXN2*. Our study group was from Bangladesh where no previous genetic study report is unavailable. Therefore, to explore the possibility of the other motor neuron disorder variants the sequenced data were analysed using the verified pipeline mentioned in section 2.2.8.

The vcf files were analysed using a purpose-built pipeline for variant calling. First using the "VariantAnnotation" in the R pipeline, all the variants for 153 genes (Table 2.3) were made as csv files for 28 samples. Secondly, a localPDB library was created to extract all the pathogenic, likely pathogenic, Conflicting interpretations of pathogenicity, and variants of the uncertain significance of MND-related genes (153) from ClinVar. Lastly, the previous sample variant files were compared with the ClinVar reported variants to find out our variants of interest for further analysis.

The genes highlighted in Table 4.3 have variants with uncertain significance or conflicting interpretations of pathogenicity. No pathogenic variants or variants of uncertain significance (VUS) were found in the genes depicted in the white boxes.

Table 4.3. Overview of the gene with VUS reported in this study within 153 genes associated with neurological motor disorders. Variants with uncertain significance or conflicting interpretations of pathogenicity have been identified for the highlighted genes.

AARS	*ABCD1*	*AFG3L2*	*ALDH18A1*	*ALS2*	*AMPD2*	*ANG*
ANXA11	*AP1S2*	*AP4B1*	*AP4E1*	*AP4M1*	*AP4S1*	*AP5Z1*
APP	*ARG1*	*ARHGEF28*	*ARL6IP1*	*ARSI*	*ASAH1*	*ATL1*
ATP2B4	*ATP7A*	*ATXN2*	*B4GALNT1*	*BICD2*	*BSCL2*	*C12orf65*
C19orf12	*CCT5*	*CHCHD10*	*CHMP2B*	*CPT1C*	*CSF1R*	*CYP27A1*
CYP2U1	*CYP7B1*	*DAO*	*DCTN1*	*DDHD1*	*DDHD2*	*DNAJB2*
DNM2	*DNMT1*	*DYNC1H1*	*EIF2B5*	*ENTPD1*	*ERBB4*	*ERLIN1*
ERLIN2	*EWSR1*	*EXOSC3*	*EXOSC8*	*FA2H*	*FARS2*	*FBXO38*
FIG4	*FLRT1*	*FUS*	*GAD1*	*GAN*	*GARS*	*GBA2*
GCH1	*GJA1*	*GJC2*	*GRN*	*HEXA*	*HNRNPA1*	*HNRNPA2B1*
HSPB1	*HSPB3*	*HSPB8*	*HSPD1*	*HTRA1*	*IBA57*	*IGHMBP2*
ITM2B	*KDM5C*	*KIAA0196*	*KIF1A*	*KIF1C*	*KIF5A*	*KLC4*
L1CAM	*LAS1L*	*LYST*	*MAG*	*MAPT*	*MARS*	*MARS2*
MATR3	*MT-ATP6*	*MTPAP*	*NEFH*	*NEK1*	*NIPA1*	*NOTCH3*
NT5C2	*OPTN*	*PFN1*	*PGAP1*	*PLEKHG5*	*PLP1*	*PNPLA6*
PRNP	*PRPH*	*PSEN1*	*PSEN2*	*RAB3GAP2*	*REEP1*	*REEP2*
RTN2	*SACS*	*SCO2*	*SETX*	*SIGMAR1*	*SLC16A2*	*SLC2A1*
SLC33A1	*SLC5A7*	*SMN1*	*SOD1*	*SPAST*	*SPG11*	*SPG20*
SPG21	*SPG7*	*SQSTM1*	*SS18L1*	*TAF15*	*TARDBP*	*TBK1*
TECPR2	*TFG*	*TREM2*	*TRPV4*	*TUBA4A*	*TYROBP*	*UBA1*
UBQLN2	*USP8*	*VAMP1*	*VAPB*	*VCP*	*VPS37A*	*VPS54*
VRK1	*WDR45*	*WDR48*	*ZFR*	*ZFYVE26*	*ZFYVE27*	

Variants not found in ClinVar were also considered and designated as variants of uncertain significance (VUS) and subject to further analysis as they may be novel variants. First, the relevant CADD score (https://cadd.gs.washington.edu/info) was calculated and variants with a CADD score <15 (Mather et al., 2016) were designated benign and excluded. Population frequency for the remaining VUS was obtained from gnomAD (gnomAD) and from Project MinE (MinE) where ALS-patient frequency was also obtained. Additionally, Mutation Taster predictions were also included to have a better understanding of the VUS. Eight UK samples with MMA-arm were included in the analysis to compare with the genetic structure of Bangladeshi MMA patients. This analysis was undertaken separately from that of the other BD-MND cases. Overall 43 heterozygous candidate pathogenic variants were identified in BD and UK MMA samples (section 4.3.5) whereas 17 heterozygous candidate pathogenic variants were identified for other BD MND samples (section 4.3.6). All the Indel variants found for 28 BD MND samples were reported as benign and were excluded from further analysis. A total of 20 Indel, deletions and duplications were found and are listed in Table 4.7. Most of them had not been reported in ClinVar.

The following flowchart shows the process of the data analysis from the vcf file and an overview of the subclass made from BD MND and UK MMA samples.

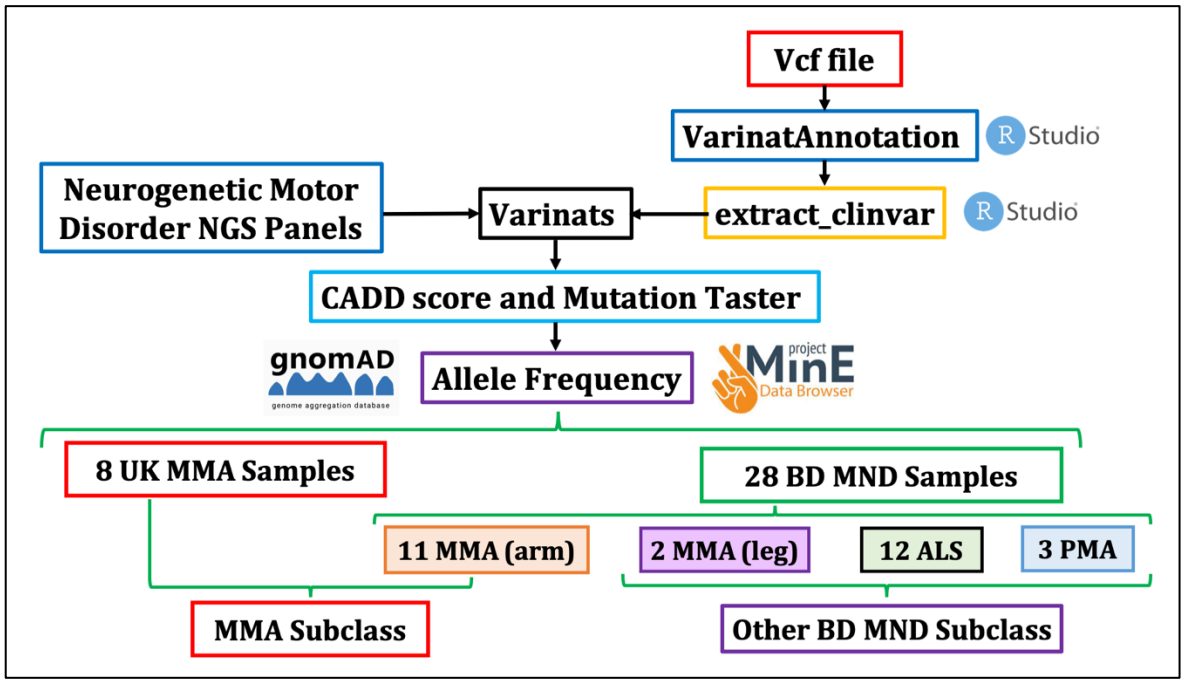


Figure 4.1. Flowchart of vcf data analysis process of the 153 Motor Neuron Disorder genes.

4.3.5. Variant identification in monomelic atrophy (MMA) samples

Eleven Bangladesh (BD) MMA patients and eight UK MMA patients from whom DNA samples were available underwent genetic testing. Following variant filtering, 43 heterozygous variants of uncertain significance were identified within genes associated with the 153 motor system disorders (**Table 4.4 and 4.5)**. No MMA patient carried a clinically validated variant within any gene. However, 2 VUS were present in MMA patient samples but rare in population-matched control databases. First, *SETX* c.1504C>T;(p.Arg502Trp) was present in two Bangladeshi MMA patients (BD-MND-010 and BD-MND-011) but not found in any UK-MMA samples. Second, a *MAG* c.1849C>G;(p.Leu617Val) was present in two UK patients (SHF-067 and SHF-089) and a *MAG* c.711G>C; p.Lys237Asn variant was found in a third UK MMA patient (SHF-070) while no MAG variant was reported for any BD samples. Two variants were identified

within *OPTN*, a known ALS gene (Ayaki et al., 2018): a novel missense variant *OPTN* c.1466A>G and p.Lys489Arg and a non-coding VUS *OPTN* c.1402-7C>T; these mutations occurred in one Bangladesh and one UK MMA patient, respectively. One of the Bangladeshi MMA patients (BD-MND-010) only had reported benign variants. Therefore, the data suggests that no clear pathogenic variants are causing MMA in the motor system genes screened. However, *MAG* variants may be associated with MMA in the UK, whilst *OPTN* variants may be a rare cause of MMA in both the UK and Bangladesh.

Table 4.4. List of variants found within ten Bangladesh (BD) patient subjects with CADD score >15 at <5% frequency in the control population excluding the benign ones. The empty boxes represent information that is not known.

Sample ID	Gene	Variants (HGVS nomenclature)	Frequency (%)		ClinVar Clinical Significance	Mutation Taster Prediction
			gnomAD v2.1.1	gnomAD v3.1.2		
BD-MND-006	*NEK1*	c.695G>A (p.Arg232His)	0.0000250	0.0000197	Conflicting interpretations of pathogenicity	Disease-causing
BD-MND-009	*ATXN2*	c.3913+24C>T:NA				
	VRK1	c.646G>C (p.Asp216His)	0.0001114	0.0000131		Disease-causing
BD-MND-010	*SETX*	c.1504C>T (p.Arg502Trp)	0.0004881	0.0002890	Conflicting interpretations of pathogenicity	Disease-causing
BD-MND-011	*SETX*	c.1504C>T (p.Arg502Trp)	0.0004881	0.0002890	Conflicting interpretations of pathogenicity	Disease-causing
BD-MND-020	*ANXA11*	c.322C>T (p.Pro108Ser)	0.0007470	0.0001446		Disease-causing
	LYST	c.7576C>T (p.Leu2526Phe)	0.0001596	0.0000395	Uncertain significance	Disease-causing
	SQSTM1	c.1277C>T (p.Ala426Val)	0.0001803	0.0001315	Uncertain significance	Disease-causing
BD-MND-024	*MARS1*	c.2180G>A (p.Arg727Gln)	0.0048963	0.0045541	Conflicting interpretations of pathogenicity	Polymorphism
	SPG11	c.176C>T (p.Ala59Val)	0.0017542	0.0002890	Conflicting interpretations of pathogenicity	Polymorphism
	ZFYVE26	c.4197C>T (p.Thr1399=)	0.0022456	0.0021223	Conflicting interpretations of pathogenicity	Disease-causing
BD-MND-029	*PLEKHG5*	c.34C>A (p.Pro12Thr)		0.0003810	Conflicting interpretations of pathogenicity	Disease-causing
	RAB3GAP2	c.2008C>G (p.Leu670Val)	0.0013761	0.0004930	Conflicting interpretations of pathogenicity	Polymorphism
	SPG7	c.*97G>A				
	WASHC5	c.2644T>G (p.Phe882Val)	0.0000849	0.0000263	Uncertain significance	Disease-causing
BD-MND-030	*OPTN*	c.1466A>G (p.Lys489Arg)				Disease-causing
BD-MND-032	*C19orf12*	c.391A>G (p.Lys131Glu)		0.0012225	Conflicting interpretations of pathogenicity	Disease-causing
	SPG11	c.4052G>A (p.Arg1351Lys)	0.0005491	0.0002040	Conflicting interpretations of pathogenicity	Polymorphism
BD-MND-033	*SPG11*	c.3362C>G (p.Thr1121Ser)	0.0000597	0.0000263		Polymorphism

Table 4.5. List of variants found within eight UK patient subjects with CADD score >15 at <5% frequency in the control population excluding the benign ones. The empty boxes represent information that is not known.

Sample ID	Gene	Variants (HGVS nomenclature)	Frequency (%)		ClinVar Clinical Significance	Mutation Taster Prediction
			gnomAD v2.1.1	gnomAD v3.1.2		
SHF-067	LYST	c.10941-7C>A	0.0069953	0.0062600	Conflicting interpretations of pathogenicity	
	MAG	c.1849C>G (p.Leu617Val)	0.0000040			Polymorphism
	MARS	c.617C>T (p.Pro206Leu)				Disease-causing
	MARS1	c.617C>T (p.Pro206Leu)	0.0006686	0.0006111	Conflicting interpretations of pathogenicity	Disease-causing
	SPG7	c.1529C>T (p.Ala510Val)	0.0028990	0.0035878	Conflicting interpretations of pathogenicity	Disease-causing
SHF-070	CCT5	c.377G>A (p.Arg126Gln)	0.0000080	0.0000066	Uncertain significance	Disease-causing
	IGHMBP2	c.2922T>G (p.Asp974Glu)	0.0010912	0.0014010	Conflicting interpretations of pathogenicity	Disease-causing
	MAG	c.711G>C (p.Lys237Asn)	0.0000107	0.0000205		Disease-causing
	WDR48	c.280T>G (p.Ser94Ala)	0.0030676	0.0026866	Uncertain significance	Disease-causing
SHF-072	LYST	c.10941-7C>A	0.0069953	0.0062600	Conflicting interpretations of pathogenicity	
	OPTN	c.1402-7C>T	0.0000796	0.0000197		
SHF-073	EIF2B5	c.-103C>G	0.0182300	0.0151300	Uncertain significance	
	ZFYVE26	c.4197C>T (p.Thr1399=)	0.0022456	0.0021223	Conflicting interpretations of pathogenicity	Disease-causing
SHF-089	ASAH1	c.126+517G>A				
	ATXN2	c.551_552ins (p.Gln185_Gln188dup)	0.0074625		Risk factor	
	EXOSC8	c.815G>C (p.Ser272Thr)	0.0039613	0.0039366	Conflicting interpretations of pathogenicity	Disease-causing
	MAG	c.1849C>G (p.Leu617Val)	0.0000040			Polymorphism
	PLEKHG5	c.882C>T (p.Phe294=)		0.0000789	Conflicting interpretations of pathogenicity	Disease-causing
	PNPLA6	c.2889C>T (p.Gly963=)				
SHF-091	AP5Z1	c.817A>G (p.Thr273Ala)				Polymorphism

	SPG11	c.-4C>A	0.0000241	0.0000394	Uncertain significance	
SHF-099	AP4B1	c.1244G>A (p.Cys415Tyr)	0.0001485	0.0002695	Uncertain significance	Polymorphism
	LYST	c.5518T>G (p.Ser1840Ala)	0.0043939	0.0047709	Conflicting interpretations of pathogenicity	Disease-causing
	PLEKHG5	c.1975G>A (p.Glu659Lys)	0.0000744	0.0000066	Uncertain significance	Disease-causing
	PLEKHG5	c.1738G>A (p.Glu580Lys)		0.0000920	Uncertain significance	Disease-causing
SHF-104	C19orf12	c.391A>G (p.Lys131Glu)		0.0012225	Conflicting interpretations of pathogenicity	Disease-causing
	DYNC1H1	c.4515G>A (p.Ser1505=)	0.0006469	0.0003746	Conflicting interpretations of pathogenicity	Disease-causing
	GAN	c.640A>G (p.Met214Val)	0.0000438		Uncertain significance	
	RAB3GAP2	c.1153C>T (p.Arg385Cys)	0.0000080		Uncertain significance	Disease-causing
	TBK1	c.964C>T (p.His322Tyr)	0.0004286	0.0003485	Conflicting interpretations of pathogenicity	Disease-causing

4.3.6. Variant identification in MND samples excluding MMA-arm

Twelve ALS samples, 3 PMA samples, and 2 monomelic-leg samples from the BD MND samples was subsequently analysed and 17 heterozygous variants of uncertain significance were identified within genes associated with motor system disorders (**Table 4.6)** for these 17 MND samples. Mutations in a well-known ALS gene, superoxide dismutase 1 (*SOD1*) (Orrell, 2000) (section 1.3.2.2) were detected in 4 patients. One pathogenic variant *SOD1* c.341T>C;(p.Ile114Thr) was present in 3 patients with sample numbers BD-MND-001, BD-MND-012 and BD-MND-016. All three patients are male. Two of them (BD-MND-001 and BD-MND-016) were diagnosed with ALS and had the age of onset of 32 (BD-MND-001) and 30 years (BD-MND-016). Another patient with this variant was diagnosed with monomelic-leg amyotrophy and had an age onset of 43 years, providing evidence of the genetic link between MND and MMA.

In one patient, BD-MND-003, a *SOD1* c.268G>A; (p.Ala90Thr) was reported as a VUS and was absent from the population database gnomAD and Project MinE (VCV000468253.3). This patient is a female diagnosed with PMA and age of onset at 36. The Mutation Taster predicted this variant as a Disease-causing variant. Moreover, the CADD score of the variant is 23.2 which represents that the variant is among the top 1% of deleterious variants in the human genome (Kircher et al., 2014a). Again, this illustrates the variability in clinical phenotype associated with the mutation of *SOD1*.

Table 4.6. List of variants found within Bangladesh (BD) patient subjects with CADD score >15 at <5% frequency in the control population excluding the benign ones and MMA sample. The empty boxes represent information that is not known.

Sample ID	Gene	Variants	Frequency (%)		ClinVar Clinical Significance
			gnomAD v2.1.1	gnomAD v3.1.2	
BD-MND-001	*ANXA11*	c.905G>A:p.Arg302His	0.00057987	0.00027604	
	SOD1	c.341T>C:p.Ile114Thr	0.00004772	0.00000657	Pathogenic
BD-MND-002	*ATXN2*	c.3317-15A>G:NA	0.00195537		
BD-MND-003	*SOD1*	c.268G>A:p.Ala90Thr			Uncertain significance
BD-MND-007	*DCTN1*	c.2596C>G:p.Leu866Val	0.00003982		
	ERBB4	c.3783G>C:p.Glu1261Asp			
	SQSTM1	c.85C>T:p.Pro29Ser	0.00009291	0.00008541	Uncertain significance
BD-MND-012	*SOD1*	c.341T>C:p.Ile114Thr	0.00004772	0.00000657	Pathogenic
BD-MND-015	*ALS2*	c.3904C>T:p.Arg1302Cys	0.00001604	0.00000657	
	ARHGEF28	c.3673C>G:p.Gln1225Glu		0.00000657	
	CYP27A1	c.1297C>T:p.Arg433Trp	0.00019123	0.00009199	Uncertain significance
BD-MND-016	*SOD1*	c.341T>C:p.Ile114Thr	0.00004772	0.00000657	Pathogenic
BD-MND-017	*SQSTM1*	c.556C>T:p.Arg186Trp	0.00003542	0.00003285	
BD-MND-021	*ALS2*	c.4375G>C:p.Asp1459His	0.00003615		
BD-MND-022	*ATXN2*	c.3317-15A>G:NA	0.00195537		
	DCTN1	c.2695A>G:p.Thr899Ala	0.00001591		
	NEK1	c.1141-14T>C:NA	0.00032564	0.00016427	
BD-MND-026	*ATXN2*	c.3317-15A>G:NA	0.00195537		
	SQSTM1	c.1207T>A:p.Ser403Thr			
BD-MND-027	*ANXA11*	c.191C>T:p.Thr64Ile	0.00072941	0.00020377	
	DCTN1	c.3824G>A:p.Arg1275His	0.00001592		Uncertain significance

Some of the variants reported through the NGS panel are not reported in ClinVar and as they are mostly Indel, deletion and duplication the CADD score couldn't be calculated for them. The CADD method can only generate predictive scores for single-nucleotide variants (SNVs) in all areas of the genome, including noncoding regions. These variants are listed in Table 4.7.

Table 4.7. List of Indel/Deletion/Duplication variants in Bangladeshi MND samples from NGS Panel results. The variants in red need to be Sanger confirmed for reporting and checked for known pseudogenes. The empty boxes represent information that is not known.

Sample ID	Gene	Variants	Chr	Position	ref	alt	id	Variant type	Molecular consequence	ClinVar
BD-MND-001	*NEFH*	NEFH:NM_021076:4:c.1646_1687del:p.549_563del	22	29885274	AAGGCCAAGTCTCCAGCAAAGGAAGAGGCAAAGTCACCGCCTG	A	rs773780196	Indel	Inframe Deletion	Not Reported in ClinVar
BD-MND-002	*FIG4*	FIG4:NM_014845:4:c.446+22T>-:NA	6	110048489	CT	C	rs563921466	Indel	Intron Variant	Not Reported in ClinVar
BD-MND-004	*GBA2*	GBA2:NM_020944:in_25_bp_from_exon:NA:NA	9	35740644	CTG	C	rs755266508	Indel	Intron Variant	Not Reported in ClinVar
BD-MND-006	*TAF15*	TAF15:NM_139215:15:c.1675_1698del:p.559_566del	17	34171977	TGGAGGAGACCGAGGTGGGGGCTAC	T	rs537726014	Indel	Inframe Deletion	Not Reported in ClinVar
BD-MND-009	*SPAST*	SPAST:NM_014946:13:c.1494-13T>-:NA	2	32366959	AT	A	rs760322678	Duplication		Benign
BD-MND-011	*ARHGEF28*	ARHGEF28:NM_001080479:30:c.3842-22T>-:NA	5	73193764	AT	A	rs537565843	Indel	Intron Variant	Not Reported in ClinVar
BD-MND-012	*ERBB4*	ERBB4:NM_005235:in_25_bp_from_exon:NA:NA	2	212578379	TAA	T	rs748883732			
BD-MND-015	*FUS*	FUS:NM_004960:6:c.679_690del:p.227_230del	16	31196414	CGGCGGCGGTGGT	C	rs752712347	Indel	Inframe Deletion	Not Reported in ClinVar
BD-MND-017	*SQSTM1*	SQSTM1:NM_003900:6:c.819_820del:p.P273fs	5	179260095	CAG	C				

BD-MND-020	*TAF15*	TAF15:NM_139215:15:c.1675_1698del:p.559_566del	17	34171977	TGGAGGAGACCGAGGTGGGGGCTAC	T	rs537726014	Indel	Inframe Deletion	Not Reported in ClinVar
BD-MND-020	*TAF15*	TAF15:NM_139215:15:c.1427_1447del:p.476_483del	17	34171729	GGAGGTGGCTATGGAGGAGATC	G	rs768963549	Indel	Inframe Deletion	Not Reported in ClinVar
BD-MND-021	*ATXN2*	ATXN2:NM_002973:1:c.540delG:p.Q180fs	12	112036778	GC	G	rs757862555	Deletion	Frameshift Variant	Not Reported in ClinVar
BD-MND-021	*ATXN2*	ATXN2:NM_002973:1:c.522_538del:p.Q174fs	12	112036780	TGTTGCTGCTGCTGCTGC	T	rs759304594	Deletion	Frameshift Variant	Not Reported in ClinVar
BD-MND-023	*FUS*	FUS:NM_004960:12:c.1292+15->T:NA	16	31201734	A	AT	rs572228309; rs886051937	Duplication		Likely benign
BD-MND-024	*ATXN2*	ATXN2:NM_002973:1:c.548_549insACAGCAGCA:p.Q183delinsQQQQ	12	112036770	C	CTGCTGCTGT				
BD-MND-026	*DAO*	DAO:NM_001917:2:c.149_154del:p.50_52del	12	109278930	GGCCTCT	G	rs546485721	Indel	Inframe Deletion	Not Reported in ClinVar
BD-MND-026	*FUS*	FUS:NM_004960:5:c.499_504del:p.167_168del	16	31195692	TGGTGGA	T	rs762260286			
BD-MND-029	*MATR3*	MATR3:NM_199189:in_25_bp_from_exon:NA:NA	5	138658266	GGT	G	rs779005896	Indel	Intron Variant	Not Reported in ClinVar
BD-MND-030	*ALS2*	ALS2:NM_020919:14:c.2581-23->T:NA	2	202593929	G	GA	rs567439355			
BD-MND-030	*FUS*	FUS:NM_004960:12:c.1292+15->T:NA	16	31201734	A	AT	rs572228309; rs886051937	Duplication		Likely benign

4.3.7. *C9ORF72* and *ATXN2*

The SDGS also provided the *C9ORF72* and ataxin 2 (*ATXN2*) testing results for all BD and UK cases that were screened for expansions using amplicon-length analysis. Tables 4.8 and 4.9 revealed that none of the samples tested positive for any expansion in either *C9ORF72* or *ATXN2*/Spinocerebellar ataxia 2 (*SCA2*), except for SHF-089 which exhibited an intermediate expansion for *ATXN2*.

Table 4.8. *C9ORF72* and *ATXN2*/*SCA2* testing results for BD samples. Abbreviations are Initial Repeat PCR Screen (STD) and Repeat-primed (RP) PCR.

Sample ID	*C9ORF72*-STD	*C9ORF72*-RP	*ATXN2*
BD-MND-001	2;7	No Exp	22
BD-MND-002	6;13	No Exp	22
BD-MND-003	2;8	No Exp	22
BD-MND-004	2;6	No Exp	22
BD-MND-005	2;4	No Exp	22
BD-MND-006	2;11	No Exp	22
BD-MND-007	2	No Exp	22
BD-MND-008	2;7	No Exp	22
BD-MND-009	4;7	No Exp	22
BD-MND-010	2;6	No Exp	22
BD-MND-011	2;8	No Exp	22
BD-MND-012	2;7	No Exp	22
BD-MND-015	4;6	No Exp	22
BD-MND-016	2	No Exp	22
BD-MND-017	2	No Exp	22
BD-MND-019	2;8	No Exp	22
BD-MND-020	7;9	No Exp	22
BD-MND-021	2	No Exp	17:22
BD-MND-022	2	No Exp	22
BD-MND-023	2	No Exp	22
BD-MND-024	2	No Exp	22:26
BD-MND-025	6;8	No Exp	22
BD-MND-026	2;12	No Exp	22
BD-MND-027	2	No Exp	22
BD-MND-029	2	No Exp	22
BD-MND-030	7;8	No Exp	22
BD-MND-032	2;10	No Exp	22
BD-MND-033	2	No Exp	22

Table 4.9. *C9ORF72* and *ATXN2*/*SCA2* testing results for UK samples.

Sample ID	Research *C9orf72*	*C9orf72* Call	Research *ATXN2*	*ATXN2* Call
SHF-067	2/2	No Exp	22/22	No Exp
SHF-070	2/4	No Exp	22/22	No Exp
SHF-072	2/2	No Exp	22/22	No Exp
SHF-073	2/2	No Exp	22/22	No Exp
SHF-089	4/5	No Exp	22/27	Intermediate Expansion
SHF-091	8/14	No Exp	22/22	No Exp
SHF-099	2/2	No Exp	22/22	No Exp
SHF-104	8/8	No Exp	22/22	No Exp

4.4. Discussion

Targeted next-generation sequencing of 28 BD samples identified no definite pathogenic variants in MMA with arm onset patients, although the *SOD1* c.341T>C (p.Ile114Thr) mutation was found in two ALS patients and one MMA patient with leg onset. One PMA patient has *SOD1* c.268G>A (p.Ala90Thr) variant.

Two Bangladeshi MMA patients BD-MND-010 (age of onset 17) and BD-MND-011 (age of onset 25) have the same variant of *SETX* c.1504C>T (p.Arg502Trp). *SETX* encodes a protein that contains a DNA/ribonucleic acid (RNA) helicase domain at its C-terminal end suggesting it may be involved in both DNA and RNA processing. Mutations may cause ataxia-ocular apraxia-2 or autosomal dominant juvenile ALS (ALS4) and this variant has also been reported to be pathogenic in a case showing compound heterozygosity (Tariq et al., 2021), and in heterozygous form in a patient suffering from hereditary motor neuropathy with pyramidal signs (Drew et al., 2015). This variant

was identified in two Bangladesh samples but no pathogenic, likely pathogenic or VUS variants of *SETX* were found in UK samples. The frequency of this variant in our patients is significantly higher than the background frequency (0.005 in ethnically matched control populations) (Tariq et al., 2021), meaning that this finding is likely clinically significant, and reflected in the high Combined Annotation Dependent Depletion (CADD) (Rentzsch et al., 2019) score of 27.2 which means the variant is amongst the top 1% deleterious variants in the human genome (Kircher et al., 2014a). The MutationTaster (MT) also predicts this variant as a Disease-causing variant.

Another variant of unknown significance, *MAG* c.1849C>G (p.Leu617Val) encodes myelin-associated glycoprotein, a transmembrane signalling molecule, which controls myelin formation and myelin-axon spacing (Pronker et al., 2016). Homozygous variants are associated with neurodegenerative conditions, including hereditary spastic paraplegia and Pelizaeus-Merzbacher disease (Lossos et al., 2015). Although MT predicts the p.Leu617Val variant as polymorphism it is present in an MND/ALS case included in Project MinE with a population frequency (%) of 0.000004 (http://databrowser.projectmine.com/). Interestingly, another patient included in this project harboured a separate heterozygous *MAG* variant: c.711G>C (p.Lys237Asn), also of unknown significance and population frequency of 0.0000107 (gnomAD v2.1.1). The MT predicts p.Lys237Asn as a disease-causing variant.

OPTN mutations causing ALS were first reported in 2010 (Maruyama et al., 2010b). Approximately, 26 different mutations in *OPTN* have been identified in sporadic and familial ALS patients of European and Asian descent (Li et al., 2015). The prevalence of

OPTN mutations in sporadic ALS patients is higher in Asian populations than in western countries (Li et al., 2015). In our study, two variants of *OPTN* c.1466A>G (p.Lys489Arg) and *OPTN* c.1402-7C>T were identified, corresponding to one Bangladeshi and one UK MMA patient. The intronic variant c.1402-7C>T has been reported in gnomAD and is a VUS. The p.Lys489Arg has not been reported previously. However, recently, a mutation affecting the same p.Lys489 (p.Lys489Glu) was reported as a novel, rare variant in two Indian patients with ALS (Narain et al., 2018). Therefore, given this report and the CADD score of 25 for the missense substitution, we propose that this novel p.Lys489Arg variant is also pathogenic. However, it is important to note that the precise variants discovered in MMA have not been reported in MND/ALS; in other genes, mutations within different protein domains may produce different phenotypes, for example, *KIF5A* (Nakamura et al., 2021).

The BD MND samples have very different frequencies of variants reported for the samples compared to those found in countries of white ancestry. While *C9ORF72* hexanucleotide repeat expansion is the most common genetic mutation identified in the European population no *C9ORF72* repeat expansion was reported for any BD samples. Among 17 samples, 4 of the patients had *SOD1* variants which were 14.29% of the sample population (total BD MND sample number 28). With the pathogenic variant *SOD1* c.341T>C (p.Ile114Thr) the patients with MMA with leg onset developed the disease comparatively later than the patients diagnosed with ALS. All the male patients having the pathogenic variant represent the previous data that states that ALS affects more males than females (McCombe and Henderson, 2010). About 60% of ALS-

affected individuals are males (*Web*MD). One female patient has *SOD1* c.268G>A (p.Ala90Thr) variant that was reported as having Uncertain significance in ClinVar. The population frequency is also unknown for these variants. With a CADD score of 23.2 and MT prediction of the Disease-causing variant, this *SOD1* variant took an interest for further studies in future.

Most of the variants reported for the BD MND samples excluding the MMA ones are not reported in ClinVar although they have a higher CADD score (>15) and low frequency (<5%) in the control population. The other variants reported with this criteria were for *ALS2*, *ANXA11*, *ARHGEF28*, *ATXN2*, *CYP27A1*, *DCTN1*, *ERBB4*, *NEK1* and *SQSTM1* genes.

My work suggests a number of conclusions:

- There is evidence for genetic overlap between MMA and MND including genetic variants in *ANXA11, ATXN2, NEK1, OPTN, SETX, SPG11, TBK1, and SQSTM1*.
- Profiling of the genetic basis of MND in Bangladesh; *SOD1* variants are present at a comparable frequency to the UK which contrasts with other global reports of *SOD1* mutation frequency. However, this may be due to sampling error as this screening was not conducted in a random cohort, but in cases presented to Dr Tom Jenkins when visiting Bangladesh. Moreover, several VUS were identified in BD MND patients which are not reported in other populations. If the genetic basis of BD MND is distinct then this will have implications for the design of therapies going forwards, particularly as we enter an age of personalised medicine and gene therapy for MND.

Chapter 5. TDP-43 Analysis of MMA Samples

5.1. Introduction

The samples from Bangladesh National Institute of Neuroscience & Hospital showed a large number (11) of MMA patients considering the small MND sample size of thirty-three. We decided to perform a specific analysis of MMA as detailed in Chapter 4, including both UK and Bangladesh MMA samples.

Brachial monomelic amyotrophy (MMA) is characterized by juvenile muscular atrophy of the distal upper extremities. Approximately 1,500 MMA patients are reported globally (Antonioni et al., 2020) with the majority in Asian populations. To date, mutations in only two genes (*KIAA1377* and *C5ORF42*) have been linked to an increased risk for MMA (Lim et al., 2012). Clinical phenotypic and genetic similarities between amyotrophic lateral sclerosis (ALS) and MMA suggest the two diseases may be part of a continuous spectrum with a shared molecular basis (genetic analysis in Chapter 4).

One of the key features in almost all cases of MND, including the most common genetic form related to changes to the *C9ORF72* gene, is the aggregation of a protein called TDP-43 in the motor neuron of the brain and the spinal cord.

Following the discovery of TDP-43 toxicity in ALS in 2006 (Neumann et al., 2006b, Arai et al., 2006), researchers have primarily concentrated on the cytoplasmic aggregation of TDP-43 and the disruption of its RNA-binding functions (Lalmansingh et al., 2011,

Polymenidou et al., 2011, De Conti et al., 2015, Costessi et al., 2014, Ishiguro et al., 2016, Kawahara and Mieda-Sato, 2012). In addition, TDP-43, which is a hallmark of degenerating motor neurons in ALS due to its nuclear clearance, has also demonstrated the ability to bind to DNA (Kuo et al., 2014, Lalmansingh et al., 2011). This categorizes TDP-43 as a member of a distinct subgroup of RNA-binding proteins known as RNA/DNA-binding proteins (RDBPs), which possess functional DNA-binding activity (Mitra et al., 2019).

TDP-43 has emerged as a key protein in understanding the pathogenesis of both sporadic and familial ALS. It was identified as a major component of cytosolic aggregates that are ubiquitinated and hyper-phosphorylated in post-mortem tissue of ALS patients (Neumann et al., 2006b, Arai et al., 2006), which is considered a hallmark of the disease and is observed in about 97% of cases, regardless of the mechanism of disease onset. However, exceptions to this pattern include familial ALS caused by mutations in Zn/Cu Superoxide Dismutase 1 (*SOD1*) and Fused in Sarcoma (*FUS*) (Hardiman et al., 2017, Kiernan et al., 2011, Rosen et al., 1993, Kwiatkowski Jr et al., 2009, Vance et al., 2009, Hergesheimer et al., 2019, Mackenzie et al., 2007).

Truncated forms of TDP-43, particularly the N-terminally truncated C-terminal fragments C-terminal fragments 35 kDa (CTF35) and 25 kDa (CTF25), are found in ALS aggregates, primarily in the cortex and to a lesser extent in the spinal cord (Berning and Walker, 2019, Tsuji et al., 2012, Smethurst et al., 2015, Hasegawa et al., 2008, Xiao et al., 2015). These fragments have been identified as the most prominent "species" of

TDP-43 in ALS, along with other C-terminal fragments (Neumann et al., 2006b, Arai et al., 2010, Neumann et al., 2009, Jeon et al., 2019).

As proposed in Chapter 4, if MMA is actually on a spectrum with MND, the hypothesis is that MMA-associated mutations will also manifest TDP-43 pathology. Since it is not possible to look inside people's brains during life to see if they have TDP-43 pathology, an alternative is to model the mutations *in vitro*. Therefore, a patient-derived iAstrocyte model was used, which has previously been shown to manifest TDP-43 pathology, when the iAstrocytes are derived from patients with MND-related mutations (Serio et al., 2013).

5.2. Aim

This study aims to screen for the presence of TDP-43 pathology in selected MMA patients with variants in MND-related genes. Therefore we aimed to convert the fibroblasts from SHF-067and SHF-072 with *MAG* (c.1849C>G; p.Leu617Val) and *OPTN* (c.1402-7C>T) variants respectively and SHF-073 and SHF-091, neither of which had any *MAG, OPTN* variants into iNPCs and then into iAstrocytes (iA) to see if the phenotype was consistent with those in ALS. The fibroblasts had been collected from UK MMA patients via the AMBRoSIA Biosampling Project. No fibroblast samples were available from the BD MMA cases.

5.3. Results

Direct conversion of fibroblast to iNPC was performed for positive and negative controls (155v2, 161, 78, 009) and MMA samples (SHF-067, SHF-072, SHF-073, SHF-091) mentioned in Table 2.10. During conversion due to bacterial contamination and a slower growth rate, UK MMA samples SHF-067 and SHF-072 failed to convert to iNPC. So, our further work on the characterization of the MMA patient samples was restricted to MMA samples SHF-073 and SHF-091 alongside FALS line 78, and SALS line 009 which served as a positive control and 161 as negative controls. Lines which converted successfully to iNPC were then converted to iAstrocytes.

5.3.1. Efficient differentiation of iA from donor fibroblast

To determine whether the donor fibroblast of the selected patients' samples transformed successfully into iNPC cells immunocytochemistry (section 2.3.4.8) was used to confirm markers of the iNPC. For example, Figure 5.1 showed the presence of Pax6 and Nestin in the transformed fibroblast of samples SHF-073 and SHF-091 selected for the transduction at the first stage.

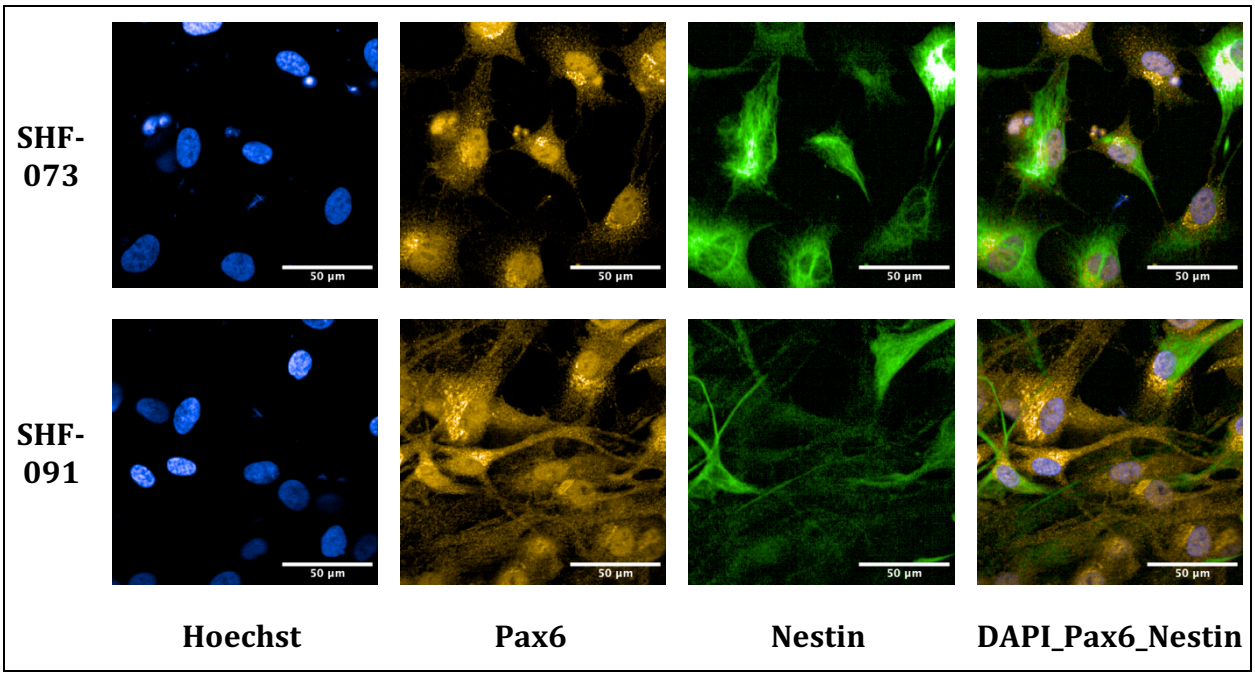


Figure 5.1. iNPCs express prototypic markers. Immunocytochemistry (ICC) assay to confirm the transduction of the fibroblast to iNPC for UK-073 and UK-091 samples.

The conversion efficiency ranges were observed between 60% and 95%, in correlation with the proliferative potential of the initial fibroblast cultures, as well as the quality of the viral vectors. Comparatively SHF-073 sample was converted more where the percentage is 72.22% on average and for SHF-091 the percentage is 64.13% (**Supplementary Table 2**). The other cell lines were grown by Allan C Shaw and Thomas R Marlow from the research group in SITraN under the supervision of Professor Laura Ferraiuolo.

5.3.2. TDP-43 proteinopathy

5.3.2.1. Western blot

Accumulation of TDP-43 fragments in motor neurons is a post-mortem hallmark of different neurodegenerative diseases (Chen-Plotkin et al., 2010). TDP-43 pathology has also been observed in postmortem astrocytes (Riku et al., 2021) and in iA derived from patient-derived fibroblasts (Franklin et al., 2021). TDP-43 contains two caspase 3 consensus cleavage sites leading to the formation of C-terminal fragments (CTFs) of 35 kDa and 25 kDa that are excluded from the nucleus as the nuclear localisation signal is not present (Chen-Plotkin et al., 2010). Chronic oxidative stress induced several features consistent with TDP-43 proteinopathies including loss of nuclear TDP-43, accumulation of diffuse TDP-43 in the cytosol, formation of a 35 kDa C-terminal fragment and accumulation of TDP-43 in stress granules (SGs), some of which revealed ubiquitylated (Meyerowitz et al., 2011). Therefore, we examined if the iAstrocytes derived from our MMA samples also were found to have C-terminal TDP-43 fragments (CTF-TDP-43), of 35 or 25 kDa in mass.

5.3.2.1.1. Loading calculations

A standard curve was made using Albumin Standard from BCA Protein Assay Kit (Thermo Scientific). Figure 5.2 showed the standard curve plotted for this analysis from which the concentration of the i-Astrocyte pellet from each of the MMA samples had been determined. The cell lysates were diluted such that equal amounts of protein

were loaded per well and mixed with 5µl of 4x Laemmli buffer to make a total volume of 20µL. The calculation was shown in Table 5.2.

Table 5.1. Absorbance of the Albumin Standard (BSA) in different concentrations.

Albumin Standard concentration (µg/µl)	Absorbance
0.1	0.0415
0.2	0.0880
0.4	0.1825
0.8	0.3320
1.2	0.4525
1.6	0.5890
2.0	0.6945

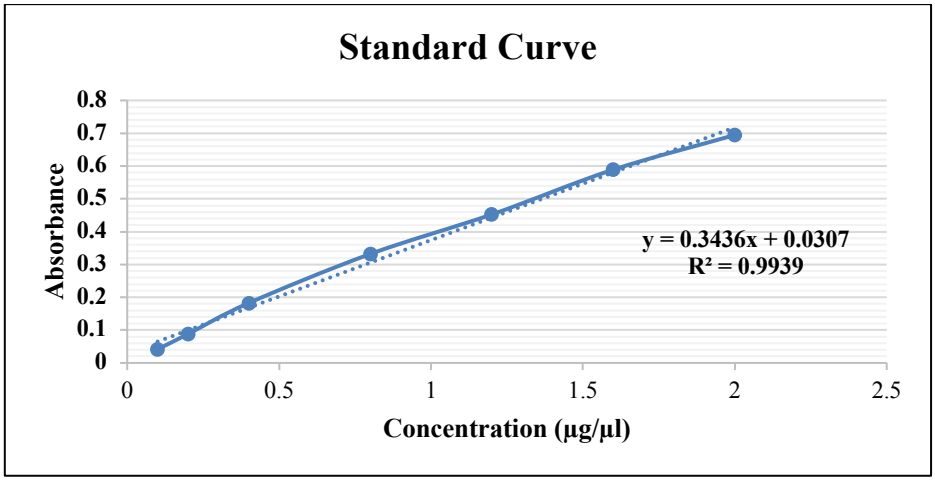


Figure 5.2. Standard Curve for protein concentration where y = concentration of protein, m = slope of the curve (x in the equation), m = absorbance, x = absorbance recorded in protein sample, c = value detected in the equation.

Table 5.2. The calculation for sample loading in SDS-Polyacrylamide gel.

Loading calculations for a total volume of 20µl						
Samples	**Absorbance**	**µg/µl**	**x4****	**µl to 20µg**	**IP lysis buffer (µl)**	**4x LB (µl)**
155v2	0.362	0.964203	3.86	5.16	9.83	5
161	0.790	2.209837	8.84	2.26	12.73	5
009 P21 iA	0.321	0.846333	3.4	5.9	9.1	5
SHF-073 P9 iA	0.205	0.507276	2.0	9.9	5.1	5
SHF-091 P7 iA	0.164	0.387951	1.6	12.9	2.1	5

5.3.2.1.2. Blot Image

Quantification of three technical replicates (randomly 3 were chosen in case of >3) was done using image studio for western blot analysis of sample iA cells. All values are given as mean ± SD. Ordinary one-way ANOVA was used to calculate the significance of the difference between the mean values of band samples. The analyses were conducted using PRISM Graph Pad (Version 9.5.0) software. The expression of the 35kDa CTF-TDP-43, termed TDP-35, in the case of SHF-073 and SHF-091 (Figure 5.3a), suggests that these MMA cases possibly showed a TDP-proteinopathy.

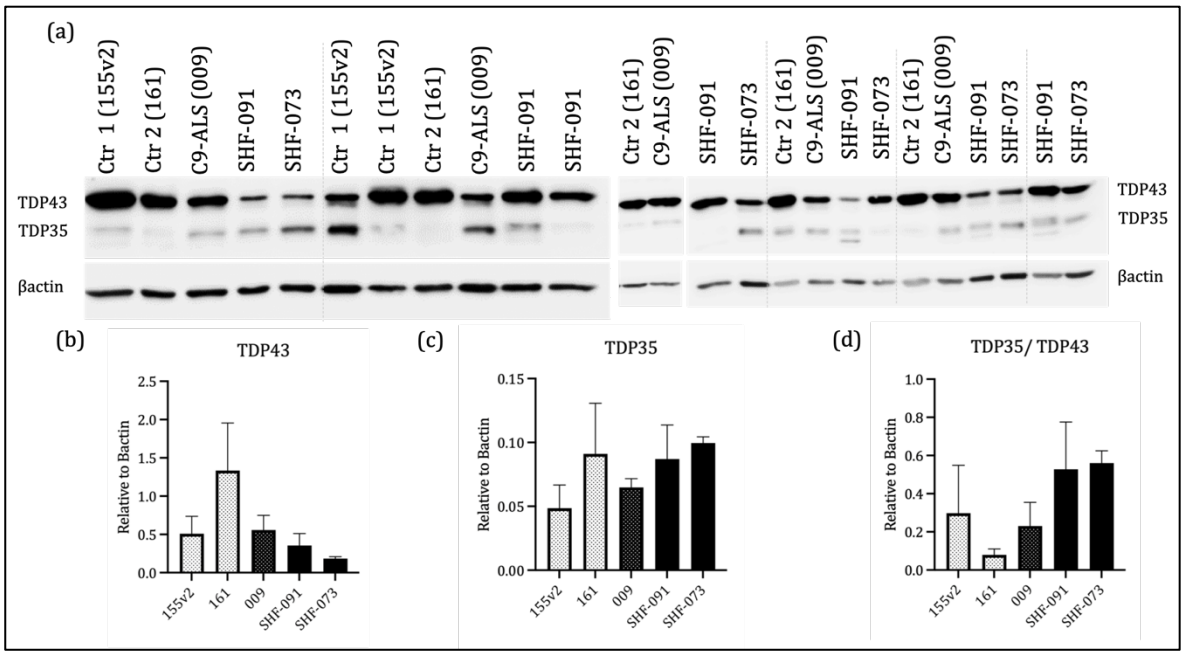


Figure 5.3. TDP-43 and 35 kDa band found in western blot analysis for negative controls 155 and 161, positive control 009 and MMA samples SHF-091 and SHF-073. (a) Representative images of western blot analysis of TDP-43, TDP-35 and Bactin. (b) Densitometric quantification of TDP-43 normalised to Bactin. (c) Densitometric quantification of TDP-35 normalised to Bactin. (d) The ratio of normalized TDP35 band compared to normalized TDP 43 in control and MMA samples.

5.3.2.2. Immunocytochemistry (ICC)

Neuropathologically, the TDP-43 proteinopathies are characterized by ubiquitin-immunoreactive (ub-ir) and TDP-43-ir neuronal cytoplasmic inclusions (NCIs), neuronal intranuclear inclusions (NIIs), dystrophic neurites (DNs) and, in cases of MND, glial cytoplasmic inclusions (GCIs) (Arai et al., 2006, Neumann et al., 2006b, Woulfe et al., 2001, Neumann et al., 2007b, Neumann et al., 2007a, Sampathu et al., 2006). These inclusions have also been observed in astrocytes (Velebit et al., 2020).

At the first attempt of the immunocytochemistry assay for TDP-43 pathology, the image showed the presence of a lot of cell debris that made it difficult to count TDP-43 pathology (**Supplementary Figure 1**). On the second attempt, the control 155v2 was unavailable for conversion which prohibited further analysis of sample 155v2. (**Figure 5.4**). Further replication is required to determine whether the expression of TDP-35 in the negative control is a reproducible finding.

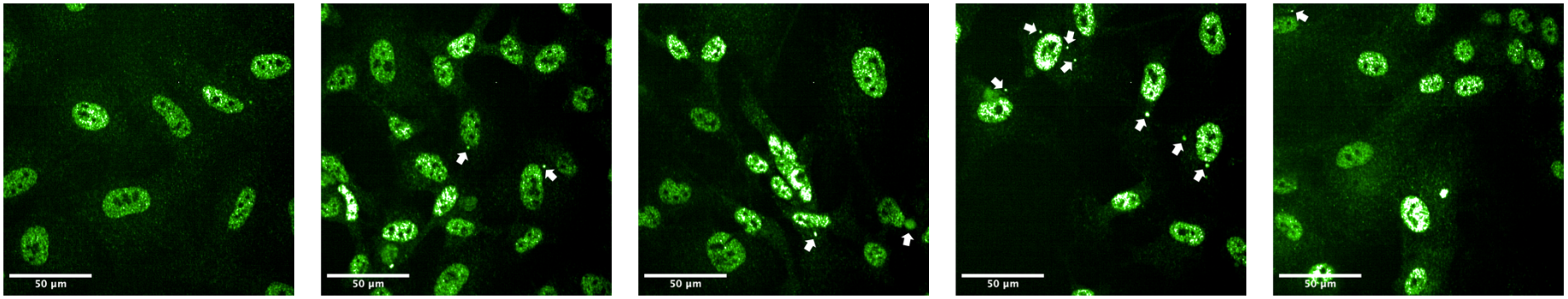


Figure 5.4. Spectrum of TDP-43 staining in MMA samples along with positive and negative controls. The white arrows mark the mislocalization of TDP-43. Detailed images of ICC have been shown in **Supplementary Table 3.**

The ICC sample images were obtained using Opera Phenix high-content imaging system and were anonymised by Allan Shaw. Manual quantification of TDP-43 spots was performed blindly for two technical replicated wells, each consisting of five fields. All values are presented as mean ± standard deviation (SD). To assess the significance of the differences in mean TDP-43 spot counts per cell between control subjects and patients with ALS, an ordinary one-way analysis of variance (ANOVA) was conducted. The statistical analysis was carried out using PRISM GraphPad software (Version 9.5.0). When comparing the average count of TDP-43 spots relative to the total number of cells, the MMA sample SHF-073 exhibited a higher range of TDP-43 aggregation in the cytoplasm compared to the control (CTR), as well as the two positive controls (009 and 78), and the other sample, SHF-091.

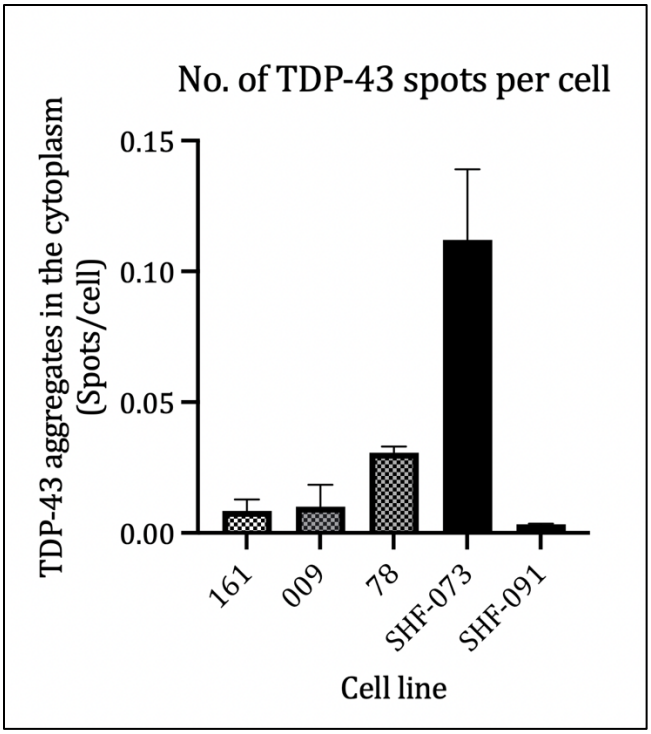


Figure 5.5. Manually counted TDP-43 spot number per cell.

5.4. Discussion

Since its discovery as a component of the ubiquitinated inclusions in the brains of patients with amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD or FTLD-TDP) diseases, TDP-43 has been recognized as a key player in disease pathogenesis (Arai et al., 2006, Neumann et al., 2006b). Abnormalities in the RNA/DNA binding protein TDP-43 have been found in approximately 95% of ALS patients, characterized by its nucleus-cytoplasmic mislocalization in spinal motor neurons, as well as in primary lateral sclerosis and progressive muscular atrophy, collectively known as TDP-proteinopathies (Dugger and Dickson, 2017). TDP-35 forms abnormal cytoplasmic inclusions in the affected neurons of patients with ALS, which contributes to the degeneration of these cells. TDP-35 is proven to be significantly more toxic than TDP-43 when expressed in the Drosophila eye, resulting in pupal lethality (Crippa et al., 2016). TDP-35 can sequester TDP-43 from a nuclear localization into cytoplasmic inclusions, and RNA binding plays an essential role in this process (Che et al., 2015).

The presence of TDP-35 is associated with pathological changes in ALS astrocytes (Serio et al., 2013). The identification of TDP-35 in the MMA patient samples suggests that the mutations under consideration might be causing a pathogenic process as is found in ALS. Further studies are needed to verify the result to support the idea that there is overlap in the pathogenesis of MMA and ALS, or even that MMA should be considered to be a mild form of ALS. This has significant implications for the design of therapeutics which may be able to convert ALS to a more mild phenotype similar to MMA.

Of the two UK MMA samples successfully converted to iNPC, SHF-073 had variants in *EIF2B5* and *ZFYVE26* and SHF-091 had variants in *AP5Z1* and *SPG11*. In ClinVar the *EIF2B5* c.-103C>G variant has been reported as uncertain significance (VCV000344328.5) and the *ZFYVE26* c.4197C>T (p.Thr1399=) variant has been reported as conflicting interpretations of pathogenicity (VCV000212640.25) through mutation taster predicted *ZFYVE26* c.4197C>T (p.Thr1399=) variant as disease-causing. This variant had not been previously evaluated by Illumina Laboratory Services, Illumina (ICSL) or reported in the Human Gene Mutation Database (HGMD) before June 1st, 2018 (ICSL, 2019). As a result, an automated scoring system was used to determine whether the variant's frequency, disease prevalence, penetrance estimates, and inheritance mode suggested it was too common to be responsible for the disease. According to the score, the possibility of this variant causing the disease could not be ruled out, and further investigation was necessary. A literature search was conducted on the gene, cDNA change, and amino acid change, but no relevant publications were found. As a result, this variant was classified as a variant of unknown significance for this disease. While SHF-091 had a variant of uncertain significance in *SPG11* (c.-4C>A), the other variant in *AP5Z1* (c.817A>G; p.Thr273Ala) was not reported either in gnomAD or ClinVar, as a result, the frequency is unknown. Two adjacent variants *AP5Z1* (c.815C>G; p.Ser272Cys) and *AP5Z1* (c.821T>G; p.Leu274Arg) were classified as a Variant of Uncertain Significance. These variants have not been reported in the literature in individuals affected with *AP5Z1*-related conditions. The serine residue (p.Ser272Cys) is highly conserved and leucine residue (p.Leu274Arg) is moderately conserved and there is a moderate physicochemical difference between the usual (serine and leucine) and alternative amino acid (cysteine and arginine). Algorithms developed to predict the effect of *AP5Z1* (c.815C>G; p.Ser272Cys) missense

changes on protein structure and function are either unavailable or do not agree on the potential impact of this missense change (SIFT: "Deleterious"; PolyPhen-2: "Probably Damaging"; Align-GVGD: "Class C0"). In summary, the available evidence is currently insufficient to determine the role of this variant in disease. On the other hand, algorithms developed to predict the effect of *AP5Z1* (c.821T>G; p.Leu274Arg) missense changes on protein structure and function (SIFT, PolyPhen-2, Align-GVGD) all suggest that this variant is likely to be disruptive. Therefore, these variants have been classified as a Variant of Uncertain Significance. A synonymous variant *AP5Z1* (c.819T>G; p.Thr273=) has been reported as likely benign.

Among the rest of the MMA samples, some have very crucial variants that might be very potential candidate genes related to ALS. For example, SHF-067 and SHF-089 have the same very rare *MAG* (c.1849C>G; p.Leu617Val) variant, SHF-070 has the *MAG* (c.711G>C; p.Lys237Asn), and SHF-072 has the *OPTN* (c.1402-7C>T) variant. Successful analysis might lead to more support for our hypothesis of MMA samples showing similar TDP-43 pathology like MND thus sharing genetic similarities as well.

The current study observed the presence of TDP-43 aggregations in both control samples, which is not a typical finding. The negative control, 155v2, exhibited a TDP-35 fragment band in Western blot, indicating cellular stress and its unavailability from immunocytochemistry (ICC) does not provide us any clear picture of the analysis. Conversely, ICC analysis showed TDP-43 spots in negative control 161, while the TDP-35 band in the Western blot was negligible. The positive control, 009, was positive in both WB and ICC analysis. The identification of TDP-43 aggregation in a control sample could complicate the differentiation

of individuals with and without underlying pathology, potentially limiting the accuracy of neurodegenerative disease diagnosis and classification. However, the implications of TDP-43 aggregation in a control sample may vary depending on factors such as the type and location of the aggregation, as well as age and gender. Therefore, further examination and assessment are necessary to fully comprehend the significance of TDP-43 aggregation in a control sample.

Reasons that may see TDP35 in control:

- Technical i.e. lack of specificity of the ICC and the antibody
- The control may truly have TDP-43 proteinopathy. This individual was relatively aged and a number of pathological processes have been associated with TDP-43 proteinopathy including Limbic-predominant age-related TDP-43 encephalopathy (LATE) (Nelson et al., 2019).

The SHF iNPCs were not passage matched in this study due to time limitations. For the study, SHF lines from passages 4-8 were utilized, while control lines from passages 16-20 were employed. If the SHF lines had been grown to the same passage as the control lines, the observed phenotypes might have been different. Thus, conducting this study again using cell lines of the same passage would be advantageous. Additionally, the examination of TDP-43 proteinopathy was carried out only once due to time constraints. It is recommended that future work incorporates biological duplicates and triplicate experiments.

Chapter 6. Discussion

MND is an incurable and relatively common neurodegenerative disease. In 2019, there were an estimated ~268,673 prevalent cases and 63,700 incident cases worldwide affected by this disease, with a 95% confidence interval of 213,893 to 310,663 (Park et al., 2022). These diseases caused 1,034,606 years of lost healthy life or disability-adjusted life-years (DALYs) and the global death count was 39081.23 in 2019 (Park et al., 2022). The global prevalence and deaths from MND have become higher (Park et al., 2022). The incidence of MND, however, has remained unchanged. More than half of the cases of MND and related deaths were reported in North America, Western Europe, and Australasia, which are high-income regions. In most areas with a high sociodemographic index, the prevalence, incidence, and burden of MND (measured in terms of DALYs) were significant. However, in high-income East Asia, these measures were relatively low compared to other areas with similar sociodemographic indices. Overall, the burden of MND has increased significantly. Knowing the changes in the frequency of occurrence, new cases, and deaths related to motor neuron diseases over time and by location is vital for uncovering the root causes of these conditions and for creating effective healthcare plans.

Distinct genetic architecture of ALS, the most prevalent form of MND, has been observed between Asian and European populations (Figure 1.4). With a growing Bangladeshi immigrant population in the UK (Garbin, 2005), this study investigating the genetic basis of ALS in Bangladesh could contribute to the development of the diagnostic process and therapeutic testing, benefiting both Bangladesh and the UK.

The genetic effects of pleiotropy (the genetic effect of a single gene on multiple phenotypic traits) and non-penetrance (the state in which a genetic trait fails to manifest itself in the phenotype) in ALS have yet to be fully understood. From a clinical perspective, the molecular cause of the majority of ALS is still largely unknown.

6.1. Somatic mutations as a cause of ALS

Somatic mutations (changes in the DNA sequence that occur after a person's conception and are present only in certain cells) in specific genes have been associated with an increased risk of developing MND, including ALS (Armon, 2016, Proukakis, 2020). Studying these mutations can help to shed light on the underlying biological mechanisms that contribute to the development of these diseases, which is important for developing new diagnostic and therapeutic approaches. Furthermore, understanding the genetic basis of MND can also inform genetic counselling and support for individuals and families affected by these diseases. Therefore, the aim was to discover evidence for somatic mutations within known ALS risk genes in DNA extracted from the motor cortex of ALS patients.

Determining somatic mutations is a complex task that presents various challenges. One major issue is the need for high-quality DNA samples from both the affected tissue and non-affected tissue, such as peripheral blood, which can be difficult to obtain, especially when the affected tissue is small or hard to access. Another challenge is that somatic mutations may be rare, occurring in only a subset of cells in the affected tissue, and can be difficult to detect against the background of the much more abundant germline

mutations. Additionally, the somatic mutations in different cells in the same tissue can be different, making it difficult to detect and confirm the presence of a mutation. The sequencing process can also generate artefacts that can be misinterpreted as somatic mutations, which highlights the importance of using multiple methods and filters to eliminate technical artefacts and confirm the authenticity of the mutations.

In the current study, the motor cortex samples of 42 ALS genes were screened for mutations. Out of these, three pathogenic mutations were identified in the *SOD1* and *FUS* genes (p.Glu101Gly in *SOD1*, p.Ile114Thr in *SOD1* and p.Gly507Asp in *FUS*). The clinical information available indicates that these variants had previously been reported in the patients from their blood sample analysis. As a result, there is no evidence to suggest somatic heterogeneity, i.e. the presence of different mutations in different cells of the same individual, for these 42 genes.

However, further studies are necessary to examine larger patient populations to confirm the presence or absence of somatic heterogeneity within ALS risk genes.

6.2. The genetic basis of MND in Bangladesh

MND is relatively uncommon in Bangladesh, however, they are significant public health issues in the country due to their debilitating nature and poor prognosis. The prevalence of MND in Bangladesh is not well documented, but studies suggest that the incidence is low compared to other countries. However, this could be due to underdiagnosis and limited access to medical resources. The lack of awareness about MND and the lack of specialised healthcare facilities in Bangladesh can also contribute

to the low detection and diagnosis rate of MND. To the best of my knowledge, no large-scale population-based studies have been conducted to determine the exact prevalence of MND in Bangladesh.

This study of BD MND samples has shown differences in the frequency of genetic variants compared to populations of white ancestry. The most common genetic mutation in the European population, *C9ORF72* hexanucleotide repeat expansion, was not found in any of the samples from Bangladesh. Out of 17 samples, 4 patients had *SOD1* variants, representing 14.29% of the total sample population. These patients with a pathogenic variant *SOD1* c.341T>C (p.Ile114Thr) developed the disease later than those diagnosed with non-*SOD1* ALS. All patients with the pathogenic variant were male, supporting the previous observation that ALS affects more males than females. One female patient had *SOD1* c.268G>A (p.Ala90Thr) variant that was reported as having "Uncertain significance" in ClinVar. Most of the variants reported for the Bangladesh MND samples were not reported in ClinVar and had a higher CADD score (>15) and low frequency (<5%) in the control population. This may reflect the low number of sequenced exomes and genomes from the South Asia region and the analysis of how the variants are associated with the disease. Variants were also reported in *ALS2*, *ANXA11*, *ARHGEF28*, *ATXN2*, *CYP27A1*, *DCTN1*, *ERBB4*, *NEK1* and *SQSTM1* among 153 genes related to MND. These findings suggest that further studies are needed to better understand the genetic basis of MND in Bangladesh but it is distinct from the MND population in the UK.

6.3. Overlap in the genetic and molecular basis of MMA and MND

The targeted sequencing of 28 BD samples with MMA did not find any definite pathogenic variants. The genetics of MMA is complex and not fully understood. To date, only *KIAA1377* and *C5ORF42* have been reported as susceptibility genes (Lim et al., 2012). However, multiple genetic factors have been implicated in the development of this disorder. This study showed evidence of genetic variants present in traditional MND genes including variants of unknown significance in *ANXA11, ATXN2, NEK1, OPTN, SETX, SPG11, TBK1, and SQSTM1*.

This suggests a possible genetic connection between MMA and MND, which led to the examination of TDP-43 pathology in patient-derived cells from selected MMA patients, as described in Chapter 5, to determine if these mutations are linked to TDP-43 pathology, similar to that seen in ALS.

Neurodegenerative diseases such as ALS, FTLD, and AD share a common pathological feature known as TDP-43 proteinopathy (Gao et al., 2018, Liao et al., 2022). This condition refers to the abnormal accumulation of TDP-43 in the cytoplasm, often through hyperphosphorylation, ubiquitination, and cleavage. TDP-43 is a protein involved in various cellular processes, including RNA metabolism, splicing, and transcription. In ALS, the TDP-43 protein is abnormally phosphorylated, and it mislocalizes from the nucleus to the cytoplasm, where it forms protein aggregates.

These aggregates can disrupt the normal function of neurons and contribute to the degeneration of motor neurons in ALS.

TDP-43 proteinopathy is a common pathological feature in the majority of ALS cases (Gao et al., 2018). TDP-35, a fragment of TDP-43, triggers cytoplasmic inclusions and alters RNA processing, leading to TDP-43 proteinopathy and potentially neurodegeneration. Sequestration of normal TDP-43 into inclusions by TDP-35, mediated by RNA, is associated with this proteinopathy (Che et al., 2015). RNA binding inhibits TDP-43 aggregation in vitro. TDP-35 inclusions mediated by RNA binding are relatively dynamic compared to insoluble protein aggregates, and they may be relevant to a gain-of-function pathology.

Previous research indicates that TDP-35 forms aggregates or inclusions that sequester full-length TDP-43 into cytoplasmic inclusions (Che et al., 2015). RNA binding facilitates this sequestration, which is enriched in the cytoplasmic inclusions. The RNA recognition motif 1 (RRM1) of TDP-43/TDP-35 is primarily responsible for nucleic-acid binding. Mutations in this domain prevent the formation of TDP-35 inclusions and their RNA-assisted association with TDP-43. Therefore, TDP-35 can move TDP-43 from the nucleus to cytoplasmic inclusions, and RNA binding is a crucial factor in this process.

Recent studies have shown that TDP-43 pathology is not limited to neurons, but can also occur in astrocytes (Serio et al., 2013). Astrocytes are a type of glial cell that plays a supportive role in the brain, providing nutrients and support to neurons, as well as contributing to the regulation of neuronal activity. Abnormal accumulation of TDP-43

in astrocytes has been implicated in the pathogenesis of neurodegenerative diseases, including ALS and FTLD.

The choice of a non-neuronal model to study TDP-43 pathology in astrocytes is justified for several reasons. Firstly, astrocytes are difficult to study *in vivo* due to their complex interactions with neurons and other glial cells. *In vitro* models of astrocytes are therefore valuable tools for studying the biology of these cells in isolation. Secondly, many studies of TDP-43 pathology have focused exclusively on neurons, even though TDP-43 pathology has been observed in astrocytes as well. By studying TDP-43 pathology in non-neuronal cells, we can gain a more comprehensive understanding of the mechanisms underlying these diseases. Moreover, this model is established in SITraN's lab and reproducibly produces TDP-43 pathology.

Using non-neuronal models such as cell lines or primary cultures of astrocytes provides the opportunity to manipulate and control experimental conditions more easily than *in vivo* studies. This can help to identify specific pathways and molecular mechanisms underlying TDP-43 pathology in astrocytes, which may be difficult to isolate *in vivo*.

One of the iA from MMA samples (SHF-073) in this study showed an increased level of TDP-43 aggregated count in the cytoplasm through ICC compared to the control. This gave an overview that further studies with a larger sample size are needed to confirm the potential genetic link between MMA and MND, especially with a focus on the variants found in genes related to MND. This would help determine the significance of these findings and provide a clearer understanding of the potential relationship between the two conditions.

6.4. Limitations

This study has several limitations that must be considered. The primary limitations are as follows:

1. **Somatic Mutation Study:**
    i. **X Chromosome:** The analysis of the Ubiquilin-2 (UBQLN2) gene in motor cortex samples was not possible in this study due to the unavailability of the X chromosome sequence in Project MinE.
    ii. **Sample size:** The original sample size was 169 but only 69 samples passed the QC, which became the main focus of this study. Additional extractions and QC would be able to expand the size of the cohort should additional funding be available.
    iii. **WGS:** Bulk WGS of DNA from the motor cortex will not be able to detect any genetic variants which are not universally present in the cells sampled. The variant calling pathway for heterozygous changes relies on determining that 50% of reads map to one or other bases. Bulk sequencing was chosen because of the ability to accurately genotype. The compromise is that there is almost no capability to recognise somatic heterogeneity within the sample. However, given that there were available matched blood samples and that motor cortex and peripheral blood have a different embryological origin then the comparison was thought to be worthwhile.
    iv. **VUS:** The variants identified as "variants of uncertain significance" couldn't be compared for somatic mutation as the WGS of the blood samples was not

available. WGS of blood samples could be included in future studies to have a better understanding of whether there is an impact of somatic mutations in motor cortex samples.

2. **Genetic Profiling of BD MND Samples**

   i. **Sample size:** Due to COVID-19 restrictions, only the first batch of 33 BD MND samples was transferred to SITraN for analysis, resulting in a small study sample. The sample size of Bangladeshi MND and UK MMA patients may be relatively small, which limits the generalizability of the findings.

   ii. **Controls:** Not having a control sample in disease genetics research poses significant limitations. Firstly, without a control group, it becomes challenging to distinguish disease-associated genetic variants from those present in the general population. A control sample serves as a reference point to assess the significance and frequency of genetic variants within the studied population. Secondly, the presence of genetic differences among different populations, known as population stratification, can lead to misleading associations between genetic variants and diseases. Without a control group from a similar population, it becomes difficult to account for this stratification and accurately identify disease-related variants. Additionally, the absence of a control group can introduce biases and confounding factors, potentially distorting the observed genetic associations.

   iii. **Genetic reference:** As stated in previous point ii, without a matched control population then significant variants present in Bangladeshi ALS patients may have been missed if these variants are not toxic in other genetic backgrounds

and therefore may form part of the reference genome. Ideally, a population-specific reference could be constructed but this is a significant practical undertaking. A more practical avenue might be to study trios whereby unaffected parents form the control reference genome used to identify de novo genetic changes in the affected child.

iv. **VUS:** Designating certain variants as "variants of uncertain significance" highlights the need for further validation and functional studies to understand their true impact on disease development.

v. ***C9orf72:*** None of the samples had *C9orf72* expansion; however, *C9orf72* was frequently found in several cases in India, with researchers reporting a unique haplotype (Shamim et al., 2020). The absence/presence of *C9orf72* expansion is usually evaluated by PCR-based methods. Most laboratories use a combination of amplicon length analysis and repeat-primed (RP)-PCR or simply RP-PCR (Akimoto et al., 2014). But it should be emphasised that a high risk of misclassification as either false positive or false negative still exists (Akimoto et al., 2014) which might explain the negative findings in this study. Southern blot analysis can be included for further confirmation of the results. However, this approach is time-consuming and unsuitable for exactly sizing expansions. On the other hand, PCR-based methods may lead to the difficult interpretation of some particular patterns, particularly in the case of intermediate-length alleles, that could resemble more expanded alleles (Akimoto et al., 2014). In 2017 Biasiotto et al. developed a new PCR-based approach for *C9orf72* genotyping, able to identify large GGGGCC expansions, undoubtedly known to be pathogenic

for ALS, sensitive enough to detect very low levels of mosaicism if they were present (Biasiotto et al., 2017).

The absence of *C9orf72* mutations in Bangladeshi MND cases suggests potential differences in the genetic landscape of MND between populations. However, further investigation is necessary to confirm this finding.

3. **Reprogramming:**
    i. **Replication:** In light of time limitations, the TDP-43 proteinopathy experiment was performed only once, which resulted in a lack of necessary repetitions to establish the experiment's robustness and reliability. Furthermore, the presence of observed aggregations in the control lines poses a significant drawback to the study's validity. Therefore, to validate the obtained findings, it is imperative to repeat this experiment with appropriate repetitions and controls.
    ii. **Cell line passages:** Due to time constraints, the SHF iNPCs were not passage matched during this study. The SHF lines were used from passages 4-8, and control lines were used from passages 16-20. If the SHF lines were grown to the same passage as the control lines, the observed phenotypes may have changed. Therefore, repeating this work with the same passage cell lines would be beneficial. Furthermore, the analysis of TDP-43 proteinopathy was only performed once due to time restrictions. Future work should include biological duplicates and triplicate regenerations of the results.

iii. **Control:** TDP-43 aggregations were observed in both control samples, an uncommon finding as they hadn't been found in these controls before. Negative control 155v2 was unavailable for ICC analysis. Negative control 161 displayed TDP-43 spots in ICC analysis, while the TDP-35 band on the Western blot was negligible. The presence of TDP-43 aggregation in a control sample could hinder distinguishing individuals with and without underlying pathology, although there was significantly more staining in the MMA cases. Further investigation is required to confirm the presence of TDP-43 aggregation in the non-diseased control sample.

6.5. Future work

There are several potential future works related to MND, including ALS and MMA:

1. Investigating the distinct genetic characteristics of MND between populations, such as between Asian and European populations, to improve diagnostic and therapeutic testing. Whole genome sequencing of the MND cases, including MMA cases from both BD and UK would provide more clarity on the genetic basis of both diseases.
2. Studying somatic mutations in specific genes associated with an increased risk for developing MND to shed light on underlying biological mechanisms and inform genetic counselling and support for individuals and families affected by these diseases.

3. Investigating the reliability and consistency of results obtained from motor cortex samples in understanding the underlying genetic factors involved in ALS.
4. Determining the significance of genetic variants found in the case of MMA, a combination of experiments can be employed. Genome-Wide Association Studies (GWAS) can be conducted to assess the variant's association with a specific trait or disease. Functional assays, such as protein expression and enzymatic activity tests, can evaluate the impact of the variant on protein function. In silico predictions using computational tools can help predict the variant's functional consequences. Transcriptomics and proteomics techniques can compare gene and protein expression profiles, respectively, between individuals with and without the variant. Animal models and population studies can further investigate the variant's effects on development, physiology, and disease susceptibility. Epidemiological studies can be performed to assess the variant's impact on disease progression in large populations. By employing these approaches, the significance of these genetic variants can be determined comprehensively.
5. Studying the potential similarities between MMA and ALS/MND is crucial to complete the experiment on TDP-43 proteinopathy. It is recommended to expand the number of biological samples and incorporate technical replicates. Additionally, rigorous statistical analysis should be applied to the data. These supplementary experimental and analytical measures are essential for acquiring a more comprehensive understanding of the relationship between MMA and ALS/MND, specifically in the context of TDP-43 proteinopathy, and

will facilitate the formulation of meaningful conclusions regarding their potential shared characteristics.

6. Testing whether MMA patient astrocytes display other motor neuron toxicity in the manner previously observed for ALS patient astrocytes (Meyer et al., 2014).

Furthermore, there are additional factors that can be taken into consideration to address the identified flaw and reduce time constraints. Initially, the transfer of whole blood samples from Bangladesh to the UK encountered delays in finding a suitable courier service capable of handling the samples. Subsequently, after securing the services of MARKEN for sample transportation, obtaining necessary permissions from the Ministry of Commerce and Ministry of Health and Family Welfare in Bangladesh for customs clearance documents was required. This entire process took approximately 7 months. Although sending DNA samples would have been a simpler solution, it was crucial to ensure that all samples underwent the same technical processing procedures. Hence, the decision was made to transfer the whole blood samples. However, establishing a collaboration with a laboratory in Bangladesh that possesses a similar setup could potentially minimize time wastage and expand the research scope beyond geographical boundaries.

Chapter 7. Conclusion

To summarise, Motor neuron diseases are severe conditions with high mortality rates and a growing global burden. The study of 69 motor cortex samples of ALS genes showed no evidence of somatic heterogeneity. Studies of Bangladeshi MND samples showed differences in genetic variants compared to populations of European ancestry but targeted sequencing of 28 Bangladeshi samples with MMA did not find any definite pathogenic variants. Figure 7.1 shows an overview of this study. Further studies with larger sample sizes are needed to confirm the potential genetic link between MMA and MND.

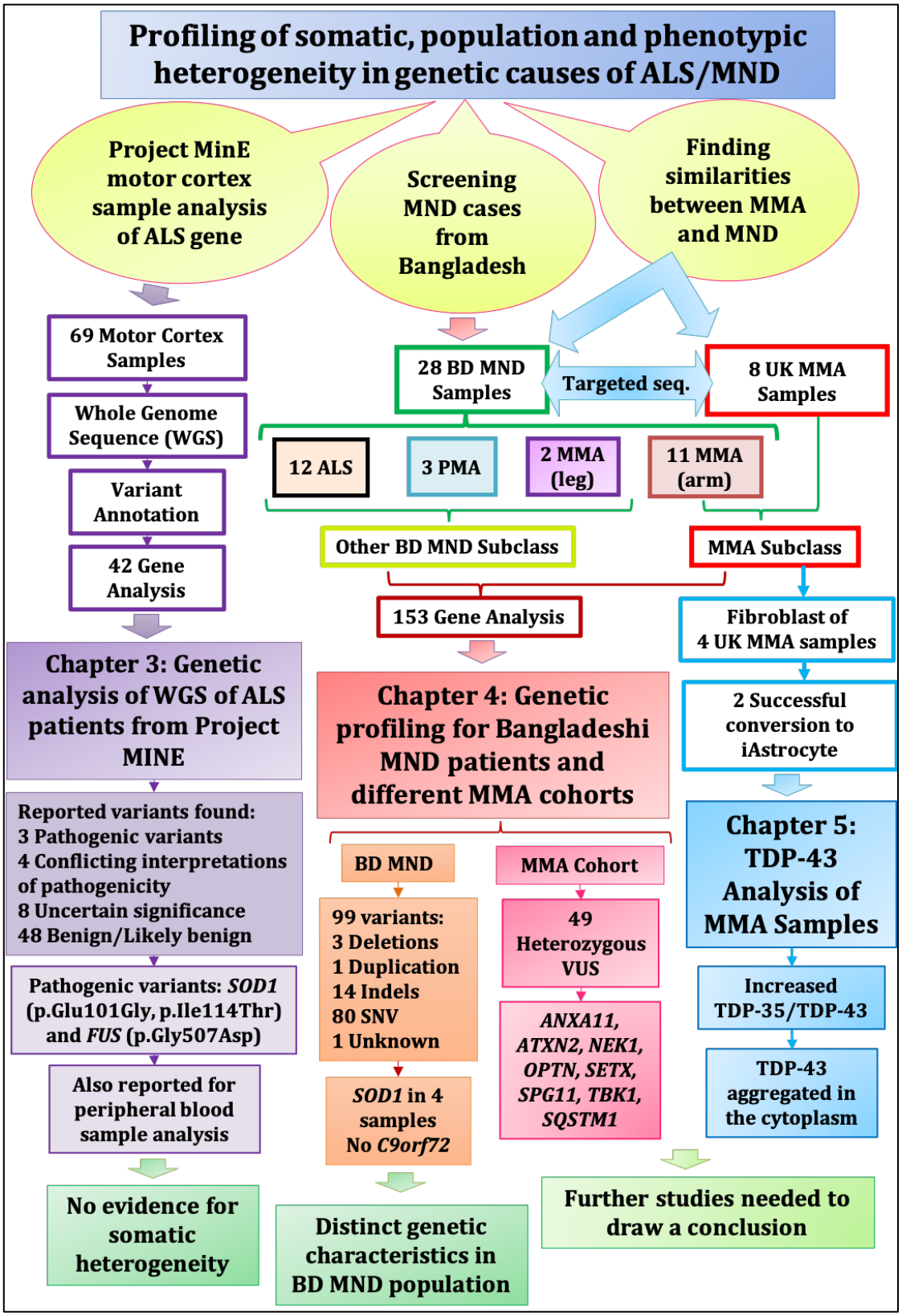


Figure 7.1. Overview of the study.

Appendices

A. Outcomes from the PhD

1. Ranganathan, R., **Haque, S.***, Coley, K., Shepheard, S., Cooper-Knock, J. & Kirby, J. 2020. Multifaceted Genes in Amyotrophic Lateral Sclerosis-Frontotemporal Dementia. Frontiers in Neuroscience, 14. **(Ranganathan et al., 2020)**
2. **Shaila Haque***, Sona Mistry, Johnathan Cooper-Knock, Sophie Cadden, Jenny Hosty, Amogh Patil, Pamela J Shaw, Quazi Deen Mohammad, Janine Kirby, Thomas M Jenkins, Laura Ferraiuolo. Hirayama syndrome is a TDP-43 proteinopathy, a model of arrested neurodegeneration. (**MANUSCRIPTS UNDER PREPARATION**)
3. **Shaila Haque***, Johnathan Cooper-Knock, Sophie Cadden, Quazi Deen Mohammad, Pamela J Shaw, Thomas M Jenkins, Janine Kirby (**2022**), Use of Next Generation Sequencing to Elucidate the Genetics of Monomelic Amyotrophy (MMA) in Bangladesh and UK patients. **ENCALS meeting 2022**. (**Poster Presentation**).

B. Genetic analysis of WGS of ALS patients from Project MinE

I. MODERATE missense_vatiant

From specific chromosome annotate variants vcf file of 69 patients from project MinE the MODERATE missense_vatiants are sorted out for each gene related to ALS.

```
setwd ("/Users/uos/df2")
library("VariantAnnotation")
#SOD1
sod1g.gr <- GRanges("21", IRanges(33031935, 33041244))
sod1params <- ScanVcfParam(which=sod1g.gr)
sod1vcf <- readVcf(TabixFile("/Users/uos/df2/Galaxy34-chr21.df2.snp_only.vcf.gz"), "hg19", sod1params)
sod1vcf.Ranges <- data.frame(rowRanges(sod1vcf), info(sod1vcf))
sod1vcf1 <- readVcf(TabixFile("/Users/uos/df2/Galaxy34-chr21.df2.snp_only.vcf.gz"), "GRCh37", sod1params)
sod1vcf1.Ranges <- data.frame(rowRanges(sod1vcf1), info(sod1vcf1))
sod1vcf.Ranges11 <- sod1vcf.Ranges[,12]
sod1indexmm <- grep ("missense_variant", sod1vcf.Ranges11)
sod1vcf.mis <- sod1vcf.Ranges[c(sod1indexmm),]
sod1vcf.Ranges11 <- sod1vcf.mis[,12]
sod1indexmm <- grep ("MODERATE", sod1vcf.Ranges11)
sod1vcf.mm <- sod1vcf.mis[c(sod1indexmm),]
sod1vcf.mm
write.csv(sod1vcf.mm, file = "sod1vcf.csv")
```

The IRange was determined by the genomic location from GeneCards listed in Table 1 for ALS genes.

II. Variants reported in ClinVar

```
library(VarfromPDB)
test <- extract_clinvar('Amyotrophic lateral sclerosis', localPDB.path = '/User/uos/localPDB/', type = 'both')
test[1:10,]
test1 <- test$variants
test1[1:10,]
test1[,5]
gene <- test1[,5]
table(gene)
#SOD1
sod1 <- which (gene == "SOD1")
sod1.table <- test1[c(sod1),]
sod1assembly <- sod1.table[,17]
sod1hg19 <- which(sod1assembly == "GRCh37")
sod1.table <- sod1.table[c(sod1hg19),]
sod1vars <- sod1.table[,c(19,20,22,23)]
sod1vars0 <- paste("chr", sod1vars[,1], sep="")
sod1vars1 <- paste(sod1vars0, sod1vars[,2], sod1vars[,3], sod1vars[,4], sep=":")
sod1.bb <- read.csv("/Users/uos/df2/xlsl.vcf df2 snp_only BrainBank copy/sod1vcf.csv", header=T)
sod1.bb.vars <- sod1.bb[,1]
sod1.bb.vars[1]
match <- which (sod1.bb.vars %in% sod1vars1)
sod1m1 <- sod1.bb.vars[c(match)]
sod1m2 <- unlist(strsplit(as.character(sod1m1),split = ":"))
sod1p1 <- sort(as.numeric(sod1m2))
sod1p2 <- which(sod1.table[,20] %in% sod1p1)
sod1p3 <- sod1.table[c(sod1p2),]
write.table(sod1.table[c(sod1p2),], file = "LP6008464-SOD1.csv")
```

III. Patients ID with positive pathogenic mutation

```
library(data.table)
install.packages('R.utils')
#SOD1
sod1chr21Gvcf <- fread("/Users/uos/df2/Galaxy34-chr21.df2.snp_only.vcf.gz")
sod1chr21Gvcf1 <- sod1chr21Gvcf[,3]
sod1chr21Gvcf2 <- data.frame(sod1chr21Gvcf1)
sod1chr21Gvcf3 <- unlist(strsplit(as.character(sod1chr21Gvcf2), split=" "))
sod1chr21Gvcf4 <- unlist(strsplit(as.character(sod1chr21Gvcf3), split=":"))
sod1chr21Gvcf5 <- sort(as.numeric(sod1chr21Gvcf4))
sod1pID1 <- which(sod1chr21Gvcf5 %in% sod1p1)
sod1chr21GpID <- sod1chr21Gvcf[c(sod1pID1),]
write.table(sod1chr21GpID, file="sod1chr21GpID.csv")
sod1pID2 <- as.matrix(sod1chr21GpID[,10:78])
sod1pID3 <- which(sod1pID2 == "0/1")
sod1pID4 <- rep(0,length (sod1pID2[1,]))
sod1pID4 <- rbind(sod1pID4, sod1pID4)
for (i in 1:length (sod1pID2[1,]))
{
 sod1Col <- sod1pID2[,i]
 sod1Col1 <- which (sod1Col == "0/1")
 if (length (sod1Col1)>0)
 {
  sod1pID4[1,i] <- 1
  sod1pID4[2,i] <- as.character(sod1chr21GpID[c(sod1Col1),8])
 }
}
colnames (sod1pID4) <- colnames (sod1chr21GpID)[10:78]
iID5 <- which (sod1pID4 == 1)
write.table( sod1pID4, file=" sod1pID4.csv")
```

IV. Key commands for Motor Cortex Samples

```
setwd ("/Users/uos/df2")
```

Setwd command-line was used to guide the right directory with the .vcf.gz and .vcf.gz.tbi files of data samples.

```
install.packages('BiocManager')
BiocManager::install("VariantAnnotation")
library("VariantAnnotation")
```

The *VariantAnnotation* package allows for annotating and filtering genetic variants. Sample data are in VCF Format which is a subset of chromosome 22 from 1000 Genomes. VCF text files contain meta-information lines, a header line with column names, data lines with information about a position in the genome, and optional genotype information on samples for each position.

```
sod1g.gr <- GRanges("21", IRanges(33031935, 33041244))
```

GRanges confirmed sorting out the chromosome site only to sort out the region coding the gene (*SOD1*). The genomic locations in **Table 2.2** were used to determine the location of the gene and to only analyse that particular region of the chromosome to avoid identifying variants in any unrelated MND gene sequences. In coding the chromosome number *(21),* start *(33031935)* and end *(33041244)* positions were used in the GRanges command.

`sod1params <- ScanVcfParam(which=sod1g.gr):` The specific MND-related gene containing sequence was then retrieved as a data subset using the *ScanVcfParam* command.

```
sod1vcf <- readVcf(TabixFile("/Users/uos/df2/Galaxy34-
chr21.df2.snp_only.vcf.gz"), "hg19", sod1params):
```

Command line *readVcf* imports records from bzip compressed or uncompressed VCF files.

```
sod1vcf.Ranges <- data.frame(rowRanges(sod1vcf), info(sod1vcf))
```

The function *data.frame()* creates data frames, tightly coupled collections of variables. The *rowRanges()* accessor is used to view the range information for data. In contrast to the genotype data, the info data are unique to the variant and the same across samples. Here all info variables are represented in a single *DataFrame*.

```
sod1vcf.Ranges11 <- sod1vcf.Ranges[,12]

sod1indexmm <- grep ("missense_variant", sod1vcf.Ranges11)
sod1vcf.mis <- sod1vcf.Ranges[c(sod1indexmm),]
write.csv(sod1vcf.mis, file = "sod1misvcf.csv")
sod1indexmm <- grep ("MODERATE", sod1vcf.Ranges11)
sod1vcf.mm <- sod1vcf.mis[c(sod1indexmm),]
write.csv(sod1vcf.mm, file = "sod1vcf.csv")
```

From a specific column, the interested variant type (e.g. missense variant, moderate etc) is separated using the grep command line and compiled in a csv file.

```
install.packages('VarfromPDB')

library(VarfromPDB)

localPDB()

test <- extract_clinvar('Amyotrophic lateral sclerosis',
localPDB.path = '/User/uos/localPDB/', type = 'both'
```

The VarfromPDB is a package capturing the genes and variants from public databases and PubMed abstracts automatically. It makes a local database, localPDB() that performs the localization of the necessary files in several databases, including ClinVar, HPO, OMIM, Orphanet, Uniprot and UCSC. Except for OMIM All the files can be freely accessed. Each database can be specified flexibly, which can be selected depending on the database update frequency.

V. Key commands for Bangladeshi MND samples

```
if (!requireNamespace("BiocManager", quietly = TRUE))

 install.packages("BiocManager")

BiocManager::install("biomaRt")

library (biomaRt)

symbol <- c('AARS', 'ABCD1', 'AFG3L2', 'ALDH18A1', 'ALS2', 'AMPD2',
 'ANG', 'ANXA11', 'AP1S2', 'AP4B1', 'AP4E1', 'AP4M1', 'AP4S1', 'AP5Z1',
 'APP', 'ARG1', 'ARHGEF28', 'ARL6IP1', 'ARSI', 'ASAH1', 'ATL1', 'ATP2B4',
 'ATP7A', 'ATXN2', 'B4GALNT1', 'BICD2', 'BSCL2', 'C12orf65', 'C19orf12',
 'CCT5', 'CHCHD10', 'CHMP2B', 'CPT1C', 'CSF1R', 'CYP27A1', 'CYP2U1',
 'CYP7B1', 'DAO', 'DCTN1', 'DDHD1', 'DDHD2', 'DNAJB2', 'DNM2',
 'DNMT1', 'DYNC1H1', 'EIF2B5', 'ENTPD1', 'ERBB4', 'ERLIN1', 'ERLIN2',
 'EWSR1', 'EXOSC3', 'EXOSC8', 'FA2H', 'FARS2', 'FBXO38', 'FIG4',
 'FLRT1', 'FUS', 'GAD1', 'GAN', 'GARS', 'GBA2', 'GCH1', 'GJA1', 'GJC2',
 'GRN', 'HEXA', 'HNRNPA1', 'HNRNPA2B1', 'HSPB1', 'HSPB3', 'HSPB8',
 'HSPD1', 'HTRA1', 'IBA57', 'IGHMBP2', 'ITM2B', 'KDM5C', 'KIAA0196',
 'KIF1A', 'KIF1C', 'KIF5A', 'KLC4', 'L1CAM', 'LAS1L', 'LYST', 'MAG',
 'MAPT', 'MARS', 'MARS2', 'MATR3', 'MT-ATP6', 'MTPAP', 'NEFH', 'NEK1',
 'NIPA1', 'NOTCH3', 'NT5C2', 'OPTN', 'PFN1', 'PGAP1', 'PLEKHG5', 'PLP1',
 'PNPLA6', 'PRNP', 'PRPH', 'PSEN1', 'PSEN2', 'RAB3GAP2', 'REEP1', 'REEP2',
 'RTN2', 'SACS', 'SCO2', 'SETX', 'SIGMAR1', 'SLC16A2', 'SLC2A1',
 'SLC33A1', 'SLC5A7', 'SMN1', 'SOD1', 'SPAST', 'SPG11', 'SPG20', 'SPG21',
 'SPG7', 'SQSTM1', 'SS18L1', 'TAF15', 'TARDBP', 'TBK1', 'TECPR2', 'TFG',
 'TREM2', 'TRPV4', 'TUBA4A', 'TYROBP', 'UBA1', 'UBQLN2', 'USP8',
 'VAMP1', 'VAPB', 'VCP', 'VPS37A', 'VPS54', 'VRK1', 'WDR45', 'WDR48',
 'ZFR', 'ZFYVE26', 'ZFYVE27')

grch37 =
 useEnsembl(biomart="ensembl",GRCh=37,dataset="hsapiens_gene_ensembl")

genes <-
 getBM(attributes=c('hgnc_symbol','chromosome_name','start_position','end
 _position'), filters ='hgnc_symbol', values =symbol, mart = grch37)
```

```
rem1 <- grep("PATCH", genes[,2])
rem2 <- grep("MT", genes[,2])
rem <- c(rem1,rem2)
genes <- genes[-c(rem),]
```

The BiomaRt (BioMart R package), a Bioconductor R package is a quick, easy and powerful way to access BioMart (databases) right from your R software terminal that enables the retrieval of large amounts of data uniformly without the need to know the underlying database schemas or write complex SQL queries. The most well-known examples of BioMart databases are maintained by Ensembl (which contains 63,677 annotated gene entries), which provides biomaRt users direct access to a diverse set of data and enables a wide range of powerful online queries from gene annotation to database mining. The chromosome number, start position and end position of 153 genes were sorted out through the biomaRt command line for our study.

```
result <- rep(0,100000)
for (i in 1:length(genes[,1]))
{
  chr = paste("chr", genes[i,2], sep="")
  start_position = genes[i,3]
  end_position = genes[i,4]
  g.gr <- GRanges(chr, IRanges(start=start_position, end=end_position))
  params <- ScanVcfParam(which=g.gr)
```

```
 vcf <- readVcf(TabixFile("/Users/uos/MND_Samples/vcf/ALS/MND01-S2011467.vcf.gz"), "hg19", params)

 vcf.Ranges <- data.frame(rowRanges(vcf), info(vcf))

 if (length(vcf.Ranges[,1])==0) next

 vcf.mis <- row.names (vcf.Ranges)

 print (vcf.mis)

 if (length(vcf.mis) ==0) next

 r <- which (result == 0)

 r <- min(r)

 result[r:(r+length(vcf.mis)-1)] <- vcf.mis

}

write.csv(result, file = "/Users/uos/csv/MND01v.csv")
```

The csv file contains all the variants found in one patient sample for 153 genes.

C. Genetic profiling for Bangladeshi MND patients and different MMA cohorts

Supplementary Table 1. Variants reported for 28 BD-MND samples from the Neurogenetic Motor Disorder NGS Panels. Variants in red must be Sanger confirmed if for reporting. The empty boxes represent information that is not known.

Gene	Variants	id	Chr	Position	ref	alt	Variant type	Molecular consequence	ClinVar	CADD GRCh37-v1.6	Sample ID
ALS2	ALS2:NM_020919:25:c.3904C>T:p.R1302C	rs759509511	2	202580495	G	A	SNV	Missense Variant	Not Reported in ClinVar	23.600	BD-MND-015
ALS2	ALS2:NM_020919:28:c.4375G>C:p.D1459H	rs764523082	2	202572620	C	G	SNV	Missense Variant	Not Reported in ClinVar	23.300	BD-MND-021
ALS2	ALS2:NM_020919:14:c.2581-23->T:NA	rs567439355	2	202593929	G	GA					BD-MND-030
ALS2	ALS2:NM_020919:29:c.4580+22G>A:NA	rs550794958	2	202571547	C	T	SNV	Intron Variant	Not Reported in ClinVar	0.562	BD-MND-032
ANXA11	ANXA11:NM_145869:10:c.905G>A:p.R302H	rs140133265	10	81923853	C	T	SNV	Missense Variant	Not Reported in ClinVar	32.000	BD-MND-001
ANXA11	ANXA11:NM_145869:9:c.745-21T>A:NA	rs771194919	10	81925974	A	T	SNV	Intron Variant	Not Reported in ClinVar	14.160	BD-MND-016
ANXA11	ANXA11:NM_145869:6:c.322C>T:p.P108S	rs553424505	10	81928964	G	A	SNV	Missense Variant	Not Reported in ClinVar	19.150	BD-MND-020
ANXA11	ANXA11:NM_145869:6:c.520C>T:p.P174S	rs144975299	10	81928766	G	A	SNV	Missense Variant	Not Reported in ClinVar	0.017	BD-MND-023
ANXA11	ANXA11:NM_145869:6:c.191C>T:p.T64I	rs560648660	10	81929095	G	A	SNV	Missense Variant	Not Reported in ClinVar	17.910	BD-MND-027

ANXA11	ANXA11:NM_145869:15:c.1311G>A:p.A437A	rs140115990	10	81917747	C	T	SNV	Synonymous Variant	Not Reported in ClinVar	10.420	BD-MND-033
ARHGEF28	ARHGEF28:NM_001080479:25:c.3137A>G:p.N1046S	rs200651003	5	73181756	A	G	SNV	Missense Variant	Not Reported in ClinVar	8.522	BD-MND-006
ARHGEF28	ARHGEF28:NM_001080479:3:c.54G>A:p.A18A	rs367962149	5	73045682	G	A	SNV	Synonymous Variant	Not Reported in ClinVar	8.938	BD-MND-011
ARHGEF28	ARHGEF28:NM_001080479:30:c.3842-22T>-:NA	rs537565843	5	73193764	AT	A	Indel	Intron Variant	Not Reported in ClinVar		BD-MND-011
ARHGEF28	ARHGEF28:NM_001080479:29:c.3673C>G:p.Q1225E		5	73190232	C	G				23.300	BD-MND-015
ARHGEF28	ARHGEF28:NM_001080479:37:c.5038T>C:p.S1680P	rs282414	5	73236680	T	C	SNV	Missense Variant	Benign	19.640	BD-MND-010; BD-MND-015
ARHGEF28	ARHGEF28:NM_001080479:9:c.964C>T:p.R322C	rs201283070; COSM1069963; COSM1069964	5	73091156	C	T	SNV	Missense Variant	Benign	5.221	BD-MND-020
ATXN2	ATXN2:NM_002973:3:c.769-10A>T:NA	rs183468414	12	111992031	T	A				14.740	BD-MND-001
ATXN2	ATXN2:NM_002973:21:c.3317-15A>G:NA	rs201513990	12	111902534	T	C	SNV	Intron Variant	Not Reported in ClinVar	18.580	BD-MND-002; BD-MND-022; BD-MND-026
ATXN2	ATXN2:NM_002973:1:c.137C>A:p.A46D	rs770648742	12	112037182	G	T	SNV	Non Coding Transcript Variant	Not Reported in ClinVar	1.032	BD-MND-004; BD-MND-015; BD-MND-027
ATXN2	ATXN2:NM_002973:24:c.3913+24C>T:NA	rs370303034	12	111891457	G	A	SNV	Intron Variant	Not Reported in ClinVar	21.000	BD-MND-009
ATXN2	ATXN2:NM_002973:1:c.659C>T:p.S220F	rs562996744	12	112036660	G	A	SNV	Missense Variant	Not Reported in ClinVar	19.830	BD-MND-012
ATXN2	ATXN2:NM_002973:23:c.3708G>A:p.A1236A	rs143166155; COSM935058	12	111893869	C	T	SNV	Synonymous Variant	Not Reported in ClinVar	7.480	BD-MND-017

ATXN2	ATXN2:NM_002973:intronic:NA:NA	rs570193403	12	111892868	T	C	SNV	Missense Variant	Not Reported in ClinVar	5.681	BD-MND-020
ATXN2	ATXN2:NM_002973:1:c.540delG:p.Q180fs	rs757862555	12	112036778	GC	G	Deletion	Frameshift Variant	Not Reported in ClinVar		BD-MND-021
ATXN2	ATXN2:NM_002973:1:c.522_538del:p.Q174fs	rs759304594	12	112036780	TGTTGCTGCTGCTGCTGC	T	Deletion	Frameshift Variant	Not Reported in ClinVar		BD-MND-021
ATXN2	ATXN2:NM_002973:1:c.548_549insACAGCAGCA:p.Q183delinsQQQQ		12	112036770	C	CTGCTGCTGT					BD-MND-024
C9ORF72	C9ORF72:NM_001256054:intronic:NA:NA		9	27573521							BD-MND-001; BD-MND-002; BD-MND-003; BD-MND-005; BD-MND-006; BD-MND-008; BD-MND-009; BD-MND-011; BD-MND-012; BD-MND-015; BD-MND-019; BD-MND-020; BD-MND-025; BD-MND-026; BD-MND-030
CHMP2B	CHMP2B:NM_014043:6:c.532-4A>G:NA	rs755848821	3	87302858	A	G	SNV	Intron Variant	Not Reported in ClinVar	0.534	BD-MND-003
CYP27A1	CYP27A1:NM_000784:9:c.1514C>T:p.T505M	rs76822427	2	219679671	C	T	SNV	Missense Variant	Uncertain significance	10.270	BD-MND-003

CYP27A1	CYP27A1:NM_000784:8:c.1297C>T:p.R433W	rs143002163	2	219679301	C	T	SNV	Missense Variant	Uncertain significance	27.900	BD-MND-015
CYP27A1	CYP27A1:NM_000784:2:c.393G>A:p.E131E	rs911534514	2	219674437	G	A	SNV	Missense Variant	Not Reported in ClinVar	3.110	BD-MND-021
DAO	DAO:NM_001917:9:c.704C>T:p.T235I	rs531892910	12	109292463	C	T	SNV	Missense Variant	Not Reported in ClinVar	12.530	BD-MND-009
DAO	DAO:NM_001917:2:c.149_154del:p.50_52del	rs546485721	12	109278930	GGCCTCT	G	Indel	Inframe Deletion	Not Reported in ClinVar		BD-MND-026
DCTN1	DCTN1:NM_004082:24:c.2761-18C>T:NA	rs549475401	2	74593163	G	A	SNV	Intron Variant	Not Reported in ClinVar	8.107	BD-MND-001; BD-MND-015
DCTN1	DCTN1:NM_004082:11:c.1049-21A>C:NA	rs538246858	2	74597692	T	G	SNV	Intron Variant	Not Reported in ClinVar	9.694	BD-MND-006; BD-MND-017
DCTN1	DCTN1:NM_004082:22:c.2596C>G:p.L866V	rs527389133	2	74593618	G	C	SNV	Missense Variant	Not Reported in ClinVar	20.200	BD-MND-007
DCTN1	DCTN1:NM_004082:14:c.1482G>A:p.A494A	rs372546194	2	74596529	C	T	SNV	Synonymous Variant	Likely benign	7.063	BD-MND-011
DCTN1	DCTN1:NM_004082:30:c.3597C>T:p.V1199V	rs767954436	2	74589789	G	A	SNV	Synonymous Variant	Likely benign	9.922	BD-MND-015
DCTN1	DCTN1:NM_004082:23:c.2695A>G:p.T899A	rs747285271	2	74593436	T	C	SNV	Missense Variant	Not Reported in ClinVar	24.100	BD-MND-022
DCTN1	DCTN1:NM_004082:32:c.3824G>A:p.R1275H	rs560344779	2	74588639	C	T	SNV	Missense Variant	Uncertain significance	23.800	BD-MND-027
ERBB4	ERBB4:NM_005235:6:c.645A>G:p.E215E		2	212589897	T	C	SNV		Not provided	11.150	BD-MND-001
ERBB4	ERBB4:NM_005235:28:c.3783G>C:p.E1261D		2	212248484	C	G				16.060	BD-MND-007
ERBB4	ERBB4:NM_005235:in_25_bp_from_exon:NA:NA	rs748883732	2	212578379	TAA	T					BD-MND-012

ERBB4	ERBB4:NM_005235:22:c.2706C>T:p.D902D	rs146441207	2	212293146	G	A	SNV	Synonymous Variant	Uncertain significance	8.297	BD-MND-023
EWSR1	EWSR1:NM_013986:7:c.599+15T>C:NA	rs201941907	22	29678561	T	C	SNV	Intron Variant	Not Reported in ClinVar	2.506	BD-MND-003; BD-MND-011
FIG4	FIG4:NM_014845:4:c.446+22T>-:NA	rs563921466	6	110048489	CT	C	Indel	Intron Variant	Not Reported in ClinVar		BD-MND-002
FUS	FUS:NM_004960:6:c.679_690del:p.227_230del	rs752712347	16	31196414	CGGCGGCGGTGGT	C	Indel	Inframe Deletion	Not Reported in ClinVar		BD-MND-015
FUS	FUS:NM_004960:12:c.1292+15->T:NA	rs572228309; rs886051937	16	31201734	A	AT	Duplication		Likely benign		BD-MND-023; BD-MND-030
FUS	FUS:NM_004960:5:c.499_504del:p.167_168del	rs762260286	16	31195692	TGGTGGA	T					BD-MND-026
GBA2	GBA2:NM_020944:UTR3:NA:NA	rs186702075	9	35737165	C	T	SNV	3 prime UTR	Not Reported in ClinVar	2.210	BD-MND-001
GBA2	GBA2:NM_020944:in_25_bp_from_exon:NA:NA	rs755266508	9	35740644	CTG	C	Indel	Intron Variant	Not Reported in ClinVar		BD-MND-004
GBA2	GBA2:NM_020944:5:c.787-18C>G:NA		9	35741079	G	C				9.361	BD-MND-005
GRN	GRN:NM_002087:10:c.1019A>T:p.H340L	rs775196555	**17**	42429003	A	T	SNV	Missense Variant	Not Reported in ClinVar	0.233	BD-MND-024
GRN	GRN:NM_002087:7:c.626C>T:p.P209L	rs368995988	**17**	42428086	C	T	SNV	Missense Variant	Not Reported in ClinVar	23.400	BD-MND-025
GRN	GRN:NM_002087:7:c.614C>T:p.S205L	rs777211749	**17**	42428074	C	T	SNV	Missense Variant	Not Reported in ClinVar	14.500	BD-MND-029
HNRNPA1	HNRNPA1:NM_031157:5:c.491-7T>A:NA	rs562426518	12	54676171	T	A				10.730	BD-MND-002
MATR3	MATR3:NM_199189:16:c.2271T>C:p.D757D	rs753870064	5	138661251	T	C	SNV	Synonymous Variant	Benign/Likely benign	9.341	BD-MND-020

MATR3	MATR3:NM_199189:16:c.2191T>C:p.L731L	rs149714542	5	138661171	T	C	SNV	Synonymous Variant	Benign	10.940	BD-MND-025
MATR3	MATR3:NM_199189:in_25_bp_from_exon:NA:NA	rs779005896	5	138658266	GGT	G	Indel	Intron Variant	Not Reported in ClinVar		BD-MND-029
NEFH	NEFH:NM_021076:4:c.1646_1687del:p.549_563del	rs773780196	22	29885274	AAGGCCAAGTCTCCAGCAAAGGAAGAGGCAAAGTCACCGCCTG	A	Indel	Inframe Deletion	Not Reported in ClinVar		BD-MND-001
NEFH	NEFH:NM_021076:1:c.745G>A:p.G249S	rs60825978	22	29876996	G	A	SNV	Missense Variant	Not provided	6.946	BD-MND-005; BD-MND-008; BD-MND-017; BD-MND-020; BD-MND-026
NEFH	NEFH:NM_021076:4:c.2775G>A:p.K925K		22	29886404	G	A				9.547	BD-MND-015
NEFH	NEFH:NM_021076:4:c.1569G>C:p.E523D	rs138278265	22	29885198	G	C	SNV	Missense Variant	Not Reported in ClinVar	10.420	BD-MND-027
NEK1	NEK1:NM_001199397:10:c.695G>A:p.R232H	rs772747361	4	170506612	C	T	SNV	Missense Variant	Conflicting interpretations of pathogenicity Likely pathogenic(1);Uncertain significance(1)	31.000	BD-MND-006
NEK1	NEK1:NM_001199397:25:c.2140-14A>G:NA	rs775248564	4	170398662	T	C	SNV	Non Coding Transcript Variant	Not Reported in ClinVar	13.670	BD-MND-022
NEK1	NEK1:NM_001199397:15:c.1141-14T>C:NA	rs551259813	4	170482902	A	G	SNV	Intron Variant	Not Reported in ClinVar	18.130	BD-MND-022

NEK1	NEK1:NM_001199397:10:c.807+18T>C:NA	rs571426626	4	170506482	A	G	SNV	Intron Variant	Not Reported in ClinVar	10.820	BD-MND-023; BD-MND-030
OPTN	OPTN:NM_001008211:14:c.1466A>G:p.K489R		10	13174131	A	G				25.000	BD-MND-030
PRPH	PRPH:NM_006262:1:c.26G>A:p.R9Q	rs57451017	12	49689009	G	A	SNV	Missense Variant	Not provided	23.200	BD-MND-020
SETX	SETX:NM_015046:12:c.5502C>T:p.H1834H	rs117410554	9	135176063	G	A	SNV	Missense Variant	Not Reported in ClinVar	0.023	BD-MND-030
SETX	SETX:NM_015046:24:c.7114G>A:p.D2372N	rs150673589; COSM1236144	9	135147182	C	T	SNV	Missense Variant	Benign	27.600	BD-MND-010; BD-MND-032
SETX	SETX:NM_015046:10:c.1504C>T:p.R502W	rs534723946	9	135205481	G	A	SNV	Missense Variant	Uncertain significance	27.200	BD-MND-010
SIGMAR1	SIGMAR1:NM_005866:UTR3:NA:NA	rs4879809	9	34635598	T	A	SNV	3 prime UTR	Uncertain significance	14.570	BD-MND-019; BD-MND-026
SIGMAR1	SIGMAR1:NM_005866:4:c.622C>T:p.R208W	rs11559048	9	34635679	G	A	SNV	Missense Variant	Benign	28.500	BD-MND-019
SOD1	SOD1:NM_000454:4:c.341T>C:p.I114T	rs121912441	21	33039672	T	C	SNV	Missense Variant	Pathogenic	26.700	BD-MND-001; BD-MND-012; BD-MND-016
SOD1	SOD1:NM_000454:4:c.268G>A:p.A90T	rs1568810660, COSM1030067	21	33039599	G	A	SNV	Missense Variant	Uncertain significance	23.200	BD-MND-003
SPAST	SPAST:NM_014946:13:c.1494-13T>-:NA	rs760322678	2	32366959	AT	A	Duplication		Benign		BD-MND-009
SPAST	SPAST:NM_014946:1:c.131C>T:p.S44L	rs121908515	2	32289031	C	T	SNV	Missense Variant	Benign/Likely benign	21.200	BD-MND-010

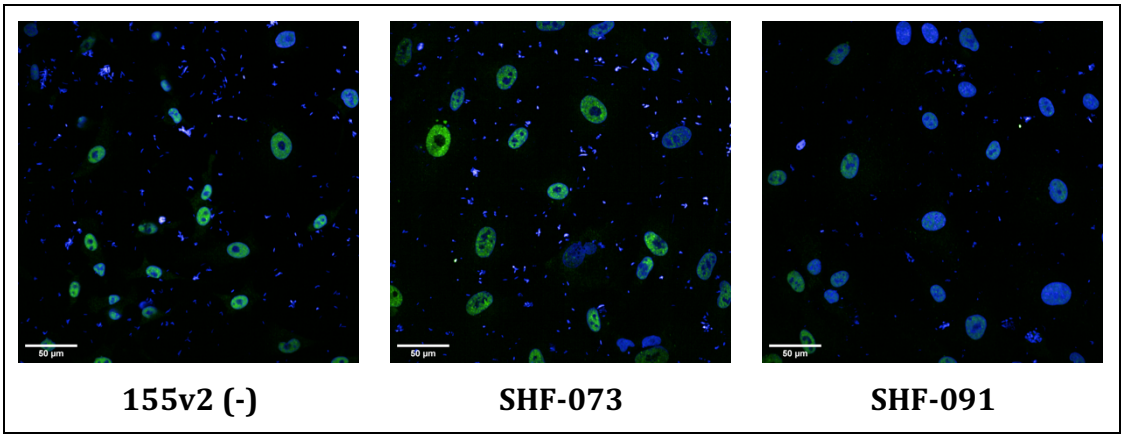
SPG11	SPG11:NM_025137:1:c.176C>T:p.A59V	rs552320263	15	44955670	G	A	SNV	Missense Variant	Conflicting interpretations of pathogenicity Benign(1);Likely benign(1);Uncertain significance(1)	21.100	BD-MND-024
SPG11	SPG11:NM_025137:26:c.4512A>T:p.G1504G	rs539384073	15	44887580	T	A	SNV	Synonymous Variant	Benign	3.554	BD-MND-026
SPG11	SPG11:NM_025137:13:c.2377G>A:p.V793M	rs546601155	15	44914485	C	T	SNV	Missense Variant	Likely benign	22.900	BD-MND-029
SPG11	SPG11:NM_025137:24:c.4052G>A:p.R1351K	rs564171625	15	44889091	C	T	SNV	Missense Variant	Not Reported in ClinVar	17.860	BD-MND-032
SPG11	SPG11:NM_025137:19:c.3362C>G:p.T1121S	rs552471760	15	44900733	G	C	SNV	Missense Variant	Not Reported in ClinVar	21.700	BD-MND-033
SQSTM1	SQSTM1:NM_003900:2:c.222G>A:p.L74L	rs371268375	5	179249974	G	A	SNV	Synonymous Variant	Not Reported in ClinVar	14.140	BD-MND-006
SQSTM1	SQSTM1:NM_003900:1:c.85C>T:p.P29S	rs752506754	5	179248021	C	T	SNV	Missense Variant	Uncertain significance	16.470	BD-MND-007
SQSTM1	SQSTM1:NM_003900:4:c.556C>T:p.R186W	rs756791819	5	179251206	C	T	SNV	Missense Variant	Not Reported in ClinVar	24.100	BD-MND-017
SQSTM1	SQSTM1:NM_003900:6:c.819_820del:p.P273fs		5	179260095	CAG	C					BD-MND-017
SQSTM1	SQSTM1:NM_003900:2:c.240C>T:p.D80D	rs148366738	5	179249992	C	T	SNV	Missense Variant	Not Reported in ClinVar	9.476	BD-MND-019
SQSTM1	SQSTM1:NM_003900:8:c.1277C>T:p.A426V	rs201239306; COSM1567750	5	179263547	C	T	SNV	Missense Variant	Uncertain significance	24.600	BD-MND-020

SQSTM1	SQSTM1:NM_003900:8:c.1207T>A:p.S403T	rs771657338	5	179263477	T	A	SNV	Missense Variant	Not Reported in ClinVar	26.600	BD-MND-026
SS18L1	SS18L1:NM_198935:8:c.824-20G>A:NA	rs368502247	**20**	60740458	G	A	SNV	Intron Variant	Not Reported in ClinVar	0.887	BD-MND-007
SS18L1	SS18L1:NM_198935:10:c.1128G>A:p.P376P	rs115792477	**20**	60749664	G	A	SNV	Synonymous Variant	Not Reported in ClinVar	7.760	BD-MND-022
TAF15	TAF15:NM_139215:15:c.1675_1698del:p.559_566del	rs537726014	**17**	34171977	TGGAGGAGACCGAGGTGGGGGCTAC	T	Indel	Inframe Deletion	Not Reported in ClinVar		BD-MND-006
TAF15	TAF15:NM_139215:15:c.1427_1447del:p.476_483del	rs768963549	**17**	34171729	GGAGGTGGCTATGGAGGAGATC	G	Indel	Inframe Deletion	Not Reported in ClinVar		BD-MND-020
TBK1	TBK1:NM_013254:18:c.1865A>G:p.K622R	rs199605037	12	64890945	A	G	SNV	Missense Variant	Not Reported in ClinVar	26.400	BD-MND-016
VCP	VCP:NM_007126:16:c.2214A>G:p.E738E	rs374391034	9	35057474	T	C	SNV	Synonymous Variant	Benign	12.070	BD-MND-005; BD-MND-007; BD-MND-021; BD-MND-025
VCP	VCP:NM_007126:UTR3:NA:NA	rs766048346	9	35057100	C	T	SNV	3 prime UTR	Not Reported in ClinVar	14.290	BD-MND-019
VCP	VCP:NM_007126:16:c.2277C>G:p.A759A	rs747361319	9	35057411	G	C	SNV	Synonymous Variant	Not Reported in ClinVar	10.860	BD-MND-021
VRK1	VRK1:NM_003384:8:c.646G>C:p.D216H	rs531480402	**14**	97321630	G	C	SNV	Missense Variant	Not Reported in ClinVar	30.000	BD-MND-009

D. TDP-43 Analysis of MMA Samples

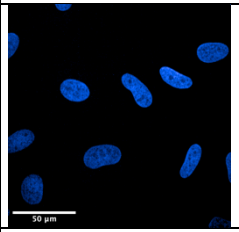
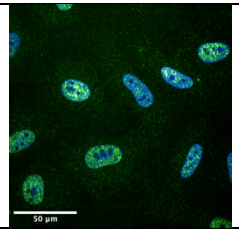
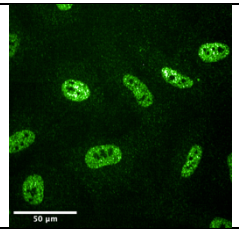
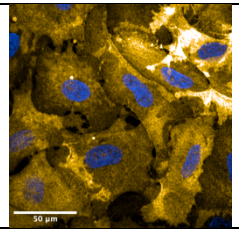
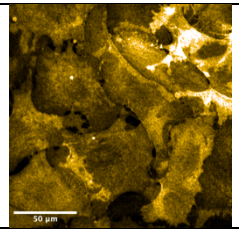
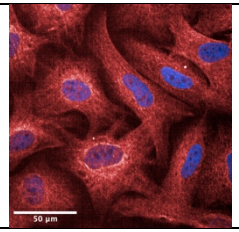
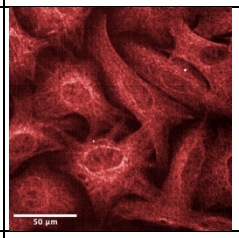
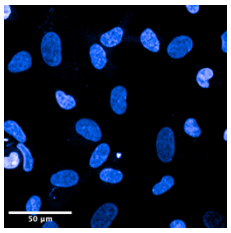
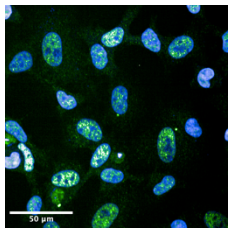
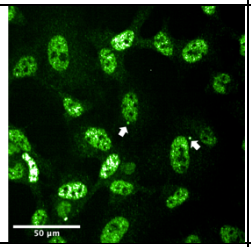
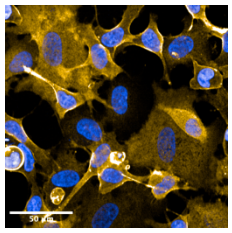
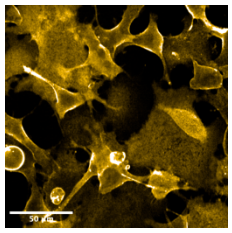
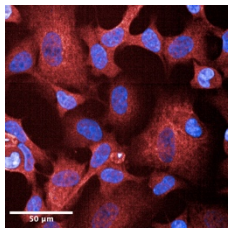
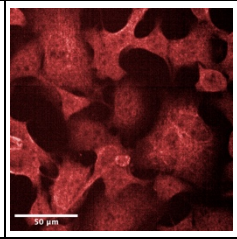
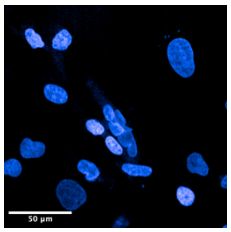
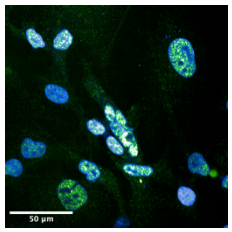
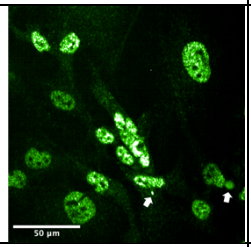
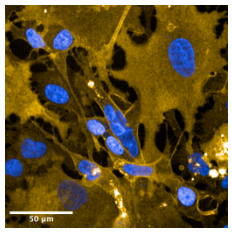
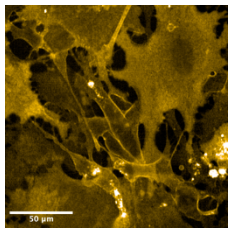
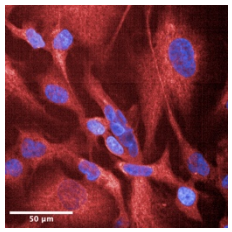
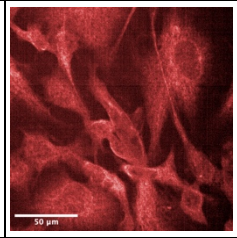
Supplementary Table 2. Percentage of cells converted to iNPC cells examined during passage 2 (P2).

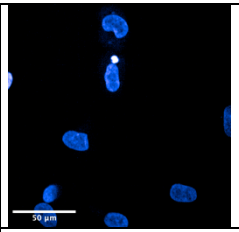
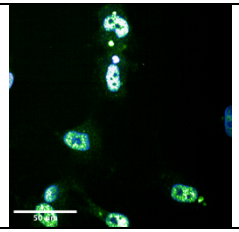
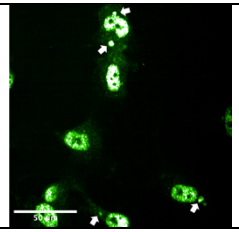
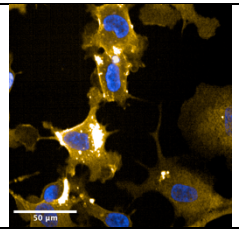
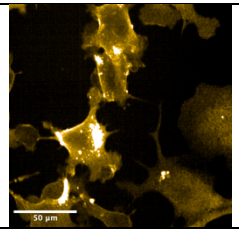
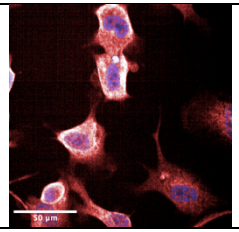
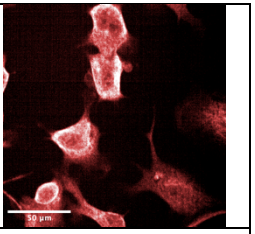
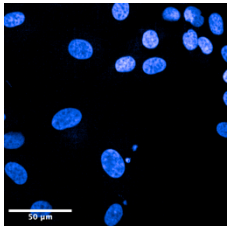
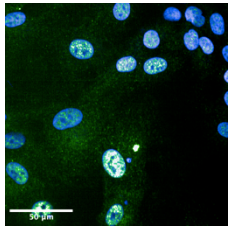
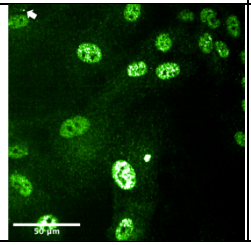
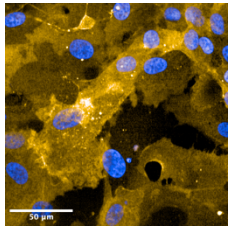
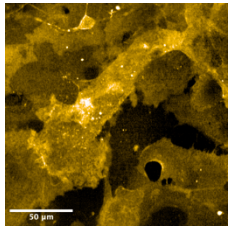
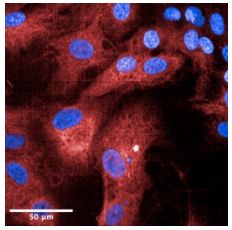
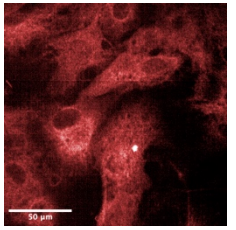
Sample	Nuclei - Number of Objects	PAX6+/Nestin+ Cells - Number of Objects	% Pax6+/Nestin+ cells	Average %
SHF-073 P2	515	363	70.48543689	72.2156448
	619	452	73.02100162	
	484	354	73.14049587	
SHF-091 P2	469	294	62.68656716	64.1339924
	421	247	58.66983373	
	373	265	71.04557641	



Supplementary Figure 1. The spectrum of DAPI_TDP-43 staining in MMA samples SHF-073 and SHF-091 along with negative control 155v2.

Supplementary Table 3. The spectrum of TDP-43 pathology in MMA samples along with positive and negative controls. The white arrows mark the mislocalization of TDP-43.

Sample / Staining	DAPI	DAPI_TDP43	TDP43	DAPI_CD44	CD44	DAPI_Vimentin	Vimentin
161 (-)							
009 (+)							
78 (+)							

SHF-073							
SHF-091							

References

Mini-PROTEAN® Tetra Cell [Online]. Available: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/10007296D.pdf [Accessed].

Sheffield Biorepository - Human Tissue Storage [Online]. Available: https://www.sheffield.ac.uk/medicine/facilities/sheffield-biorepository [Accessed].

ABHINAV, K., STANTON, B., JOHNSTON, C., HARDSTAFF, J., ORRELL, R., HOWARD, R., CLARKE, J., SAKEL, M., AMPONG, M.-A. & SHAW, C. 2007. Amyotrophic lateral sclerosis in South-East England: a population-based study. *Neuroepidemiology,* 29**,** 44-48.

ACUNA-HIDALGO, R., VELTMAN, J. A. & HOISCHEN, A. 2016. New insights into the generation and role of de novo mutations in health and disease. *Genome biology,* 17**,** 1-19.

AFGAN, E., BAKER, D., BATUT, B., VAN DEN BEEK, M., BOUVIER, D., CECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., GRUNING, B. A., GUERLER, A., HILLMAN-JACKSON, J., HILTEMANN, S., JALILI, V., RASCHE, H., SORANZO, N., GOECKS, J., TAYLOR, J., NEKRUTENKO, A. & BLANKENBERG, D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research,* 46**,** W537-W544.

AKIMOTO, C., VOLK, A. E., VAN BLITTERSWIJK, M., VAN DEN BROECK, M., LEBLOND, C. S., LUMBROSO, S., CAMU, W., NEITZEL, B., ONODERA, O. & VAN RHEENEN, W. 2014. A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories. *Journal of medical genetics,* 51**,** 419-424.

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. & POWELL, J. F. 1998. Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor. *Hum Mol Genet,* 7**,** 2045-50.

AL-CHALABI, A. & BROWN, R. H., JR. 2018. Finding a Treatment for ALS - Will Gene Editing Cut It? *N Engl J Med,* 378**,** 1454-1456.

AL-CHALABI, A., FANG, F., HANBY, M. F., LEIGH, P. N., SHAW, C. E., YE, W. & RIJSDIJK, F. 2010. An estimate of amyotrophic lateral sclerosis heritability using twin data. *Journal of Neurology, Neurosurgery & Psychiatry,* 81**,** 1324-1326.

AL-CHALABI, A. & HARDIMAN, O. 2013. The epidemiology of ALS: a conspiracy of genes, environment and time. *Nature Reviews Neurology,* 9**,** 617.

AL-SAIF, A., AL-MOHANNA, F. & BOHLEGA, S. 2011. A mutation in sigma-1 receptor causes juvenile amyotrophic lateral sclerosis. *Ann Neurol,* 70**,** 913-9.

ALČAZ, S., JAREBINSKI, M., PEKMEZOVIĆ, T., STEVIĆ-MARINKOVIĆ, Z., PAVLOVIĆ, S. & APOSTOLSKI, S. 1996. Epidemiological and clinical characteristics of ALS in Belgrade, Yugoslavia. *Acta neurologica scandinavica,* 94**,** 264-268.

ALEXANDER, M. D., TRAYNOR, B. J., MILLER, N., CORR, B., FROST, E., MCQUAID, S., BRETT, F. M., GREEN, A. & HARDIMAN, O. 2002. "True" sporadic ALS associated with a novel SOD-1 mutation. *Ann Neurol,* 52**,** 680-3.

ALSULTAN, A. A., WALLER, R., HEATH, P. R. & KIRBY, J. 2016. The genetics of amyotrophic lateral sclerosis: current insights. *Degener Neurol Neuromuscul Dis,* 6**,** 49-64.

ALT, F. W. & SCHWER, B. 2018. DNA double-strand breaks as drivers of neural genomic change, function, and disease. *DNA repair,* 71**,** 158-163.

ALWINE, J. C., KEMP, D. J. & STARK, G. R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National Academy of Sciences,* 74**,** 5350-5354.

ALZUALDE, A., MORENO, F., MARTINEZ-LAGE, P., FERRER, I., GOROSTIDI, A., OTAEGUI, D., BLAZQUEZ, L., ATARES, B., CARDOSO, S., MARTINEZ DE PANCORBO, M., JUSTE, R., RODRIGUEZ-MARTINEZ, A. B., INDAKOETXEA, B. & LOPEZ DE MUNAIN, A. 2010. Somatic mosaicism in a case of apparently sporadic Creutzfeldt-Jakob disease carrying a de novo D178N mutation in the PRNP gene. *Am J Med Genet B Neuropsychiatr Genet,* 153B**,** 1283-91.

AMADOR-ORTIZ, C., LIN, W. L., AHMED, Z., PERSONETT, D., DAVIES, P., DUARA, R., GRAFF-RADFORD, N. R., HUTTON, M. L. & DICKSON, D. W. 2007. TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann Neurol,* 61**,** 435-45.

ANTONIONI, A., FONDERICO, M. & GRANIERI, E. 2020. Hirayama disease: a case of an Albanian Woman clinically stabilized without surgery. *Frontiers in Neurology,* 11**,** 183.

ARAI, T., HASEGAWA, M., AKIYAMA, H., IKEDA, K., NONAKA, T., MORI, H., MANN, D., TSUCHIYA, K., YOSHIDA, M. & HASHIZUME, Y. 2006. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochemical and biophysical research communications,* 351**,** 602-611.

ARAI, T., HASEGAWA, M., NONOKA, T., KAMETANI, F., YAMASHITA, M., HOSOKAWA, M., NIIZATO, K., TSUCHIYA, K., KOBAYASHI, Z., IKEDA, K., YOSHIDA, M., ONAYA, M., FUJISHIRO, H. & AKIYAMA, H. 2010. Phosphorylated and cleaved TDP-43 in ALS, FTLD and other neurodegenerative disorders and in cellular models of TDP-43 proteinopathy. *Neuropathology,* 30**,** 170-81.

ARBUSTINI, E., BEHR, E. R., CARRIER, L., VAN DUIJN, C., EVANS, P., FAVALLI, V., VAN DER HARST, P., HAUGAA, K. H., JONDEAU, G., KAAB, S., KASKI, J. P., KAVOUSI, M., LOEYS, B., PANTAZIS, A., PINTO, Y., SCHUNKERT, H., DI TORO, A., THUM, T., URTIS, M., WALTENBERGER, J. & ELLIOTT, P. 2022. Interpretation and actionability of genetic variants in cardiomyopathies: a position statement from the European Society of Cardiology Council on cardiovascular genomics. *Eur Heart J,* 43**,** 1901-1916.

ARMON, C. 2001. Environmental risk factors for amyotrophic lateral sclerosis. *Neuroepidemiology,* 20**,** 2-6.

ARMON, C. 2003. Motor neuron disorders.

ARMON, C. 2016. Accrued somatic mutations (nucleic acid changes) trigger ALS: 2005-2015 update. *Muscle Nerve,* 53**,** 842-9.

ARUNDINE, M. & TYMIANSKI, M. 2003. Molecular mechanisms of calcium-dependent neurodegeneration in excitotoxicity. *Cell Calcium,* 34**,** 325-37.

ASCHENBRENNER, D. S. 2023. New Drug Approved For ALS. *Am J Nurs,* 123**,** 22-23.

AUNDHAKAR, S. C., MAHAJAN, S. K. & CHHAPRA, D. A. 2017. Hirayama's Disease: A Rare Clinical Variant of Amyotrophic Lateral Sclerosis. *Adv Biomed Res,* 6**,** 95.

AWAN, S., SIDDIQI, A. I., ASIF, A., AHMED, N., BROHI, H., JALBANI, S. & WASAY, M. 2019. Spectrum of neurological disorders in neurology outpatients clinics in urban and rural Sindh, Pakistan: a cross sectional study. *BMC neurology,* 19**,** 192.

AYAKI, T., ITO, H., KOMURE, O., KAMADA, M., NAKAMURA, M., WATE, R., KUSAKA, H., YAMAGUCHI, Y., LI, F., KAWAKAMI, H., URUSHITANI, M. & TAKAHASHI, R. 2018. Multiple Proteinopathies in Familial ALS Cases With Optineurin Mutations. *J Neuropathol Exp Neurol,* 77**,** 128-138.

BABIC LEKO, M., ZUPUNSKI, V., KIRINCICH, J., SMILOVIC, D., HORTOBAGYI, T., HOF, P. R. & SIMIC, G. 2019. Molecular Mechanisms of Neurodegeneration Related to C9orf72 Hexanucleotide Repeat Expansion. *Behav Neurol,* 2019**,** 2909168.

BAE, T., TOMASINI, L., MARIANI, J., ZHOU, B., ROYCHOWDHURY, T., FRANJIC, D., PLETIKOS, M., PATTNI, R., CHEN, B. J., VENTURINI, E., RILEY-GILLIS, B., SESTAN, N., URBAN, A. E., ABYZOV, A. & VACCARINO, F. M. 2018. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science,* 359**,** 550-555.

BALENDRA, R. & ISAACS, A. M. 2018. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol,* 14**,** 544-558.

BALOH, R. H., GLASS, J. D. & SVENDSEN, C. N. 2018. Stem cell transplantation for amyotrophic lateral sclerosis. *Curr Opin Neurol,* 31**,** 655-661.

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., AUGE, G., COCHAUD, C., LESPINASSE, F., N'GUYEN, K., DE SEPTENVILLE, A., BRICE, A., YU-WAI-MAN, P., SESAKI, H., POUGET, J. & PAQUIS-FLUCKLINGER, V. 2014. A mitochondrial origin for frontotemporal dementia and amyotrophic lateral sclerosis through CHCHD10 involvement. *Brain,* 137**,** 2329-45.

BÄUMER, D., HILTON, D., PAINE, S., TURNER, M., LOWE, J., TALBOT, K. & ANSORGE, O. 2010. Juvenile ALS with basophilic inclusions is a FUS proteinopathy with FUS mutations. *Neurology,* 75**,** 611-618.

BECK, J., POULTER, M., HENSMAN, D., ROHRER, J. D., MAHONEY, C. J., ADAMSON, G., CAMPBELL, T., UPHILL, J., BORG, A., FRATTA, P., ORRELL, R. W., MALASPINA, A., ROWE, J., BROWN, J., HODGES, J., SIDLE, K., POLKE, J. M., HOULDEN, H., SCHOTT, J. M., FOX, N. C., ROSSOR, M. N., TABRIZI, S. J., ISAACS, A. M., HARDY, J., WARREN, J. D., COLLINGE, J. & MEAD, S. 2013. Large

C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet,* 92**,** 345-53.

BELBASIS, L., BELLOU, V. & EVANGELOU, E. 2016. Environmental Risk Factors and Amyotrophic Lateral Sclerosis: An Umbrella Review and Critical Assessment of Current Evidence from Systematic Reviews and Meta-Analyses of Observational Studies. *Neuroepidemiology,* 46**,** 96-105.

BENATAR, M. & WUU, J. 2012. Presymptomatic studies in ALS: rationale, challenges, and approach. *Neurology,* 79**,** 1732-9.

BENJAMINSEN, E., ALSTADHAUG, K. B., GULSVIK, M., BALOCH, F. K. & ODEH, F. 2018. Amyotrophic lateral sclerosis in Nordland county, Norway, 2000-2015: prevalence, incidence, and clinical features. *Amyotroph Lateral Scler Frontotemporal Degener,* 19**,** 522-527.

BENSIMON, G., LACOMBLEZ, L. & MEININGER, V. 1994. A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. *N Engl J Med,* 330**,** 585-91.

BERNING, B. A. & WALKER, A. K. 2019. The Pathobiology of TDP-43 C-Terminal Fragments in ALS and FTLD. *Front Neurosci,* 13**,** 335.

BERTOLOTTI, A. 2018. Importance of the subcellular location of protein deposits in neurodegenerative diseases. *Curr Opin Neurobiol,* 51**,** 127-133.

BETGE, J., KERR, G., MIERSCH, T., LEIBLE, S., ERDMANN, G., GALATA, C. L., ZHAN, T., GAISER, T., POST, S., EBERT, M. P., HORISBERGER, K. & BOUTROS, M. 2015. Amplicon sequencing of colorectal cancer: variant calling in frozen and formalin-fixed samples. *PLoS One,* 10**,** e0127146.

BIASIOTTO, G., ARCHETTI, S., DI LORENZO, D., MEROLA, F., PAIARDI, G., BORRONI, B., ALBERICI, A., PADOVANI, A., FILOSTO, M., BONVICINI, C., CAIMI, L. & ZANELLA, I. 2017. A PCR-based protocol to accurately size C9orf72 intermediate-length alleles. *Mol Cell Probes,* 32**,** 60-64.

BIESECKER, L. G. & SPINNER, N. B. 2013. A genomic view of mosaicism and human disease. *Nat Rev Genet,* 14**,** 307-20.

BLAUW, H. M., VAN RHEENEN, W., KOPPERS, M., VAN DAMME, P., WAIBEL, S., LEMMENS, R., VAN VUGHT, P. W., MEYER, T., SCHULTE, C., GASSER, T., CUPPEN, E., PASTERKAMP, R. J., ROBBERECHT, W., LUDOLPH, A. C., VELDINK, J. H. & VAN DEN BERG, L. H. 2012. NIPA1 polyalanine repeat expansions are associated with amyotrophic lateral sclerosis. *Hum Mol Genet,* 21**,** 2497-502.

BLUM, J. A. & GITLER, A. D. 2022. Singling out motor neurons in the age of single-cell transcriptomics. *Trends Genet,* 38**,** 904-919.

BOWCOCK, A. M., KIDD, J. R., MOUNTAIN, J. L., HEBERT, J. M., CAROTENUTO, L., KIDD, K. K. & CAVALLI-SFORZA, L. L. 1991. Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci U S A,* 88**,** 839-43.

BRADLEY, W. G., ANDREW, A. S., TRAYNOR, B. J., CHIO, A., BUTT, T. H. & STOMMEL, E. W. 2018. Gene-Environment-Time Interactions in Neurodegenerative Diseases: Hypotheses and Research Approaches. *Ann Neurosci,* 25**,** 261-267.

BRENNER, D. & FREISCHMIDT, A. 2022. Update on genetics of amyotrophic lateral sclerosis. *Curr Opin Neurol,* 35**,** 672-677.

BRENNER, D., MULLER, K., WIELAND, T., WEYDT, P., BOHM, S., LULE, D., HUBERS, A., NEUWIRTH, C., WEBER, M., BORCK, G., WAHLQVIST, M., DANZER, K. M., VOLK, A. E., MEITINGER, T., STROM, T. M., OTTO, M., KASSUBEK, J., LUDOLPH, A. C., ANDERSEN, P. M. & WEISHAUPT, J. H. 2016. NEK1 mutations in familial amyotrophic lateral sclerosis. *Brain,* 139**,** e28.

BRITES, D. & VAZ, A. R. 2014. Microglia centered pathogenesis in ALS: insights in cell interconnectivity. *Front Cell Neurosci,* 8**,** 117.

BROTMAN, R. G., MORENO-ESCOBAR, M. C., JOSEPH, J. & PAWAR, G. 2021. Amyotrophic lateral sclerosis. *StatPearls [internet].* StatPearls Publishing.

BROWN, A. 2022. FDA new drug approvals in Q3 2022. *Nat Rev Drug Discov,* 21**,** 788.

BROWN, A. L., WILKINS, O. G., KEUSS, M. J., HILL, S. E., ZANOVELLO, M., LEE, W. C., BAMPTON, A., LEE, F. C. Y., MASINO, L., QI, Y. A., BRYCE-SMITH, S., GATT, A., HALLEGGER, M., FAGEGALTIER, D., PHATNANI, H., CONSORTIUM, N. A., NEWCOMBE, J., GUSTAVSSON, E. K., SEDDIGHI, S., REYES, J. F., COON, S. L., RAMOS, D., SCHIAVO, G., FISHER, E. M. C., RAJ, T., SECRIER, M., LASHLEY, T., ULE, J., BURATTI, E., HUMPHREY, J., WARD, M. E. & FRATTA, P. 2022. TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. *Nature,* 603**,** 131-137.

BUCHMAN, V. L., COOPER-KNOCK, J., CONNOR-ROBSON, N., HIGGINBOTTOM, A., KIRBY, J., RAZINSKAYA, O. D., NINKINA, N. & SHAW, P. J. 2013. Simultaneous and independent detection of C9ORF72 alleles with low and high number of GGGGCC repeats using an optimised protocol of Southern blot hybridisation. *Molecular neurodegeneration,* 8**,** 12.

BUNTON-STASYSHYN, R. K., SACCON, R. A., FRATTA, P. & FISHER, E. M. 2015. SOD1 Function and Its Implications for Amyotrophic Lateral Sclerosis Pathology: New and Renascent Themes. *Neuroscientist,* 21**,** 519-29.

CALABRESE, E. J., CALABRESE, V. & GIORDANO, J. 2021. Demonstrated hormetic mechanisms putatively subserve riluzole-induced effects in neuroprotection against amyotrophic lateral sclerosis (ALS): Implications for research and clinical practice. *Ageing Res Rev,* 67**,** 101273.

CALUS, M. P. L. & VANDENPLAS, J. 2018. SNPrune: an efficient algorithm to prune large SNP array and sequence datasets based on high linkage disequilibrium. *Genetics Selection Evolution,* 50.

CALVO, A., MOGLIA, C., CANOSA, A., BRUNETTI, M., BARBERIS, M., TRAYNOR, B. J., CARRARA, G., VALENTINI, C., RESTAGNO, G. & CHIO, A. 2014. De novo nonsense mutation of the FUS gene in an apparently familial amyotrophic lateral sclerosis case. *Neurobiol Aging,* 35**,** 1513 e7-11.

CAMPBELL, I. M., SHAW, C. A., STANKIEWICZ, P. & LUPSKI, J. R. 2015. Somatic mosaicism: implications for disease and transmission genetics. *Trends Genet,* 31**,** 382-92.

CAMU, W., DE LA CRUZ, E. & ESSELIN, F. 2022. Therapeutic tools for familial ALS. *Rev Neurol (Paris)*.

CAVALLI-SFORZA, L. L. 1997. Genes, peoples, and languages. *Proc Natl Acad Sci U S A,* 94**,** 7719-24.

CHA, S. J. & KIM, K. 2022. Effects of the Edaravone, a Drug Approved for the Treatment of Amyotrophic Lateral Sclerosis, on Mitochondrial Function and Neuroprotection. *Antioxidants (Basel),* 11.

CHANG, C. C., CHOW, C. C., TELLIER, L. C. A. M., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience,* 4.

CHE, M.-X., JIANG, L.-L., LI, H.-Y., JIANG, Y.-J. & HU, H.-Y. 2015. TDP-35 sequesters TDP-43 into cytoplasmic inclusions through binding with RNA. *FEBS letters,* 589**,** 1920-1928.

CHE, M. X., JIANG, Y. J., XIE, Y. Y., JIANG, L. L. & HU, H. Y. 2011. Aggregation of the 35-kDa fragment of TDP-43 causes formation of cytoplasmic inclusions and alteration of RNA processing. *FASEB J,* 25**,** 2344-53.

CHEN, R., IM, H. & SNYDER, M. 2015. Whole-exome enrichment with the agilent sureselect human all exon platform. *Cold Spring Harbor Protocols,* 2015**,** pdb. prot083659.

CHEN-PLOTKIN, A. S., LEE, V. M. & TROJANOWSKI, J. Q. 2010. TAR DNA-binding protein 43 in neurodegenerative disease. *Nat Rev Neurol,* 6**,** 211-20.

CHHANGANI, D. & RINCON-LIMAS, D. E. 2022. TDP-35, a truncated fragment of TDP-43, induces dose-dependent toxicity and apoptosis in flies. *Neural Regen Res,* 17**,** 2441-2442.

CHIÒ, A., CALVO, A., MOGLIA, C., MAZZINI, L., MORA, G. J. J. O. N., NEUROSURGERY & PSYCHIATRY 2011. Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population based study. 82**,** 740-746.

CHIO, A., CALVO, A., MOGLIA, C., OSSOLA, I., BRUNETTI, M., SBAIZ, L., LAI, S. L., ABRAMZON, Y., TRAYNOR, B. J. & RESTAGNO, G. 2011. A de novo missense mutation of the FUS gene in a "true" sporadic ALS case. *Neurobiol Aging,* 32**,** 553 e23-6.

CHIÒ, A., LOGROSCINO, G., TRAYNOR, B., COLLINS, J., SIMEONE, J., GOLDSTEIN, L. & WHITE, L. 2013. Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. *Neuroepidemiology,* 41**,** 118-130.

CHIO, A., MAZZINI, L., D'ALFONSO, S., CORRADO, L., CANOSA, A., MOGLIA, C., MANERA, U., BERSANO, E., BRUNETTI, M., BARBERIS, M., VELDINK, J. H., VAN DEN BERG, L. H., PEARCE, N., SPROVIERO, W., MCLAUGHLIN, R., VAJDA, A., HARDIMAN, O., ROONEY, J., MORA, G., CALVO, A. & AL-CHALABI, A. 2018. The multistep hypothesis of ALS revisited: The role of genetic mutations. *Neurology,* 91**,** e635-e642.

CHIÒ, A., MORA, G., CALVO, A., MAZZINI, L., BOTTACCHI, E. & MUTANI, R. 2009. Epidemiology of ALS in Italy: a 10-year prospective population-based study. *Neurology,* 72**,** 725-731.

CHONG, S. S., MCCALL, A. E., COTA, J., SUBRAMONY, S. H., ORR, H. T., HUGHES, M. R. & ZOGHBI, H. Y. 1995. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet,* 10**,** 344-50.

CHUN, J. & SCHATZ, D. G. 1999. Rearranging views on neurogenesis: neuronal death in the absence of DNA end-joining proteins. *Neuron,* 22**,** 7-10.

CINGOLANI, P., PLATTS, A., WANG, L. L., COON, M., NGUYEN, T., WANG, L., LAND, S. J., LU, X. Y. & RUDEN, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. *Fly,* 6**,** 80-92.

CIRULLI, E. T., LASSEIGNE, B. N., PETROVSKI, S., SAPP, P. C., DION, P. A., LEBLOND, C. S., COUTHOUIS, J., LU, Y. F., WANG, Q., KRUEGER, B. J., REN, Z., KEEBLER, J., HAN, Y., LEVY, S. E., BOONE, B. E., WIMBISH, J. R., WAITE, L. L., JONES, A. L., CARULLI, J. P., DAY-WILLIAMS, A. G., STAROPOLI, J. F., XIN, W. W., CHESI, A., RAPHAEL, A. R., MCKENNA-YASEK, D., CADY, J., VIANNEY DE JONG, J. M., KENNA, K. P., SMITH, B. N., TOPP, S., MILLER, J., GKAZI, A., CONSORTIUM, F. S., AL-CHALABI, A., VAN DEN BERG, L. H., VELDINK, J., SILANI, V., TICOZZI, N., SHAW, C. E., BALOH, R. H., APPEL, S., SIMPSON, E., LAGIER-TOURENNE, C., PULST, S. M., GIBSON, S., TROJANOWSKI, J. Q., ELMAN, L., MCCLUSKEY, L., GROSSMAN, M., SHNEIDER, N. A., CHUNG, W. K., RAVITS, J. M., GLASS, J. D., SIMS, K. B., VAN DEERLIN, V. M., MANIATIS, T., HAYES, S. D., ORDUREAU, A., SWARUP, S., LANDERS, J., BAAS, F., ALLEN, A. S., BEDLACK, R. S., HARPER, J. W., GITLER, A. D., ROULEAU, G. A., BROWN, R., HARMS, M. B., COOPER, G. M., HARRIS, T., MYERS, R. M. & GOLDSTEIN, D. B. 2015. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science,* 347**,** 1436-41.

CONTE, A., LATTANTE, S., ZOLLINO, M., MARANGI, G., LUIGETTI, M., DEL GRANDE, A., SERVIDEI, S., TROMBETTA, F. & SABATELLI, M. 2012. P525L FUS mutation is consistently associated with a severe form of juvenile amyotrophic lateral sclerosis. *Neuromuscular Disorders,* 22**,** 73-75.

COON, E. A. & BENARROCH, E. E. 2018. DNA damage response: Selected review and neurologic implications. *Neurology,* 90**,** 367-376.

COOPER-KNOCK, J. 2022. Implications of confirmed de novo pathogenic SOD1 mutations. *J Neurol Neurosurg Psychiatry,* 93**,** 118.

COOPER-KNOCK, J., HARVEY, C., ZHANG, S., MOLL, T., TIMPANARO, I. S., KENNA, K. P., IACOANGELI, A. & VELDINK, J. H. 2021a. Advances in the Genetic Classification of ALS. *Current opinion in neurology,* 34**,** 756.

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. & SHAW, P. J. 2012. Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. *Brain,* 135**,** 751-64.

COOPER-KNOCK, J., ROBINS, H., NIEDERMOSER, I., WYLES, M., HEATH, P. R., HIGGINBOTTOM, A., WALSH, T., KAZOKA, M., PROJECT MIN, E. A. L. S. S. C., INCE, P. G., HAUTBERGUE, G. M., MCDERMOTT, C. J., KIRBY, J. & SHAW, P. J. 2017. Targeted Genetic Screen in Amyotrophic Lateral Sclerosis Reveals Novel Genetic Variants with Synergistic Effect on Clinical Phenotype. *Front Mol Neurosci,* 10**,** 370.

COOPER-KNOCK, J., SHAW, P. J. & KIRBY, J. 2014. The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype. *Acta Neuropathol,* 127**,** 333-45.

COOPER-KNOCK, J., ZHANG, S., KENNA, K. P., MOLL, T., FRANKLIN, J. P., ALLEN, S., NEZHAD, H. G., IACOANGELI, A., YACOVZADA, N. Y. & EITAN, C. 2021b. Rare variant burden analysis within enhancers identifies CAV1 as an ALS risk gene. *Cell Reports,* 34.

CORCIA, P., PRADAT, P. F., SALACHAS, F., BRUNETEAU, G., FORESTIER, N., SEILHEAN, D., HAUW, J. J. & MEININGER, V. 2008. Causes of death in a post-mortem series of ALS patients. *Amyotroph Lateral Scler,* 9**,** 59-62.

CORCIA, P., TAUBER, C., VERCOULLIE, J., ARLICOT, N., PRUNIER, C., PRALINE, J., NICOLAS, G., VENEL, Y., HOMMET, C., BAULIEU, J. L., COTTIER, J. P., ROUSSEL, C., KASSIOU, M., GUILLOTEAU, D. & RIBEIRO, M. J. 2012. Molecular imaging of microglial activation in amyotrophic lateral sclerosis. *PLoS One,* 7**,** e52941.

CORRADO, L., DEL BO, R., CASTELLOTTI, B., RATTI, A., CEREDA, C., PENCO, S., SORARU, 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. & SILANI, V. 2010. Mutations of FUS gene in sporadic amyotrophic lateral sclerosis. *Journal of Medical Genetics,* 47**,** 190-194.

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. & D'ALFONSO, S. 2009. High frequency of TARDBP gene mutations in Italian patients with amyotrophic lateral sclerosis. *Hum Mutat,* 30**,** 688-94.

COSTESSI, L., PORRO, F., IACONCIG, A. & MURO, A. F. 2014. TDP-43 regulates beta-adducin (Add2) transcript stability. *RNA Biol,* 11**,** 1280-90.

COURATIER, P., LAUTRETTE, G., LUNA, J. A. & CORCIA, P. 2021. Phenotypic variability in amyotrophic lateral sclerosis. *Rev Neurol (Paris),* 177**,** 536-543.

CRIPPA, V., CICARDI, M. E., RAMESH, N., SEGUIN, S. J., GANASSI, M., BIGI, I., DIACCI, C., ZELOTTI, E., BARATASHVILI, M., GREGORY, J. M., DOBSON, C. M., CEREDA, C., PANDEY, U. B., POLETTI, A. & CARRA, S. 2016. The chaperone HSPB8 reduces the accumulation of truncated TDP-43 species in cells and protects against TDP-43-mediated toxicity. *Hum Mol Genet,* 25**,** 3908-3924.

D'GAMA, A. M. & WALSH, C. A. 2018. Somatic mosaicism and neurodevelopmental disease. *Nat Neurosci,* 21**,** 1504-1514.

DANECEK, P., AUTON, A., ABECASIS, G., ALBERS, C. A., BANKS, E., DEPRISTO, M. A., HANDSAKER, R. E., LUNTER, G., MARTH, G. T., SHERRY, S. T., MCVEAN, G., DURBIN, R. & GRP, G. P. A. 2011. The variant call format and VCFtools. *Bioinformatics,* 27**,** 2156-2158.

DAVIS, A., GAO, R. & NAVIN, N. E. 2019. SCOPIT: sample size calculations for single-cell sequencing experiments. *BMC Bioinformatics,* 20**,** 566.

DE CARVALHO, M. 2023. Advances in amyotrophic lateral sclerosis research in 2022. *Lancet Neurol,* 22**,** 21-22.

DE CONTI, L., AKINYI, M. V., MENDOZA-MALDONADO, R., ROMANO, M., BARALLE, M. & BURATTI, E. 2015. TDP-43 affects splicing profiles and isoform production of genes involved in the apoptotic and mitotic cellular pathways. *Nucleic Acids Res,* 43**,** 8990-9005.

DEJESUS-HERNANDEZ, M., 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., 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. & RADEMAKERS, R. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron,* 72**,** 245-56.

DEJESUS-HERNANDEZ, M., KOCERHA, J., FINCH, N., CROOK, R., BAKER, M., DESARO, P., JOHNSTON, A., RUTHERFORD, N., WOJTAS, A. & KENNELLY, K. 2010. De novo truncating FUS gene mutation as a cause of sporadic amyotrophic lateral sclerosis. *Human mutation,* 31**,** E1377-E1389.

DEMETRIOU, C. A., HADJIVASILIOU, P. M., KLEOPA, K. A., CHRISTOU, Y. P., LEONIDOU, E., KYRIAKIDES, T. & ZAMBA-PAPANICOLAOU, E. 2017. Epidemiology of Amyotrophic Lateral Sclerosis in the Republic of Cyprus: A 25-Year Retrospective Study. *Neuroepidemiology,* 48**,** 79-85.

DENG, H.-X., SHI, Y., FURUKAWA, Y., ZHAI, H., FU, R., LIU, E., GORRIE, G. H., KHAN, M. S., HUNG, W.-Y. & BIGIO, E. H. 2006. Conversion to the amyotrophic lateral sclerosis phenotype is associated with intermolecular linked insoluble aggregates of SOD1 in mitochondria. *Proceedings of the National Academy of Sciences,* 103**,** 7142-7147.

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. & SIDDIQUE, T. 2011. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature,* 477**,** 211-5.

DENTLINGER, M. 2018. Image Acquisition on the Odyssey Fc Imager.

DIEKSTRA, F. P., SARIS, C. G., VAN RHEENEN, W., FRANKE, L., JANSEN, R. C., VAN ES, M. A., VAN VUGHT, P. W., BLAUW, H. M., GROEN, E. J., HORVATH, S., ESTRADA, K., RIVADENEIRA, F., HOFMAN, A., UITTERLINDEN, A. G., ROBBERECHT, W., ANDERSEN, P. M., MELKI, J., MEININGER, V., HARDIMAN, O., LANDERS, J. E., BROWN, R. H., JR., SHATUNOV, A., SHAW, C. E., LEIGH, P. N., AL-CHALABI, A., OPHOFF, R. A., VAN DEN BERG, L. H. & VELDINK, J. H. 2012a.

Mapping of gene expression reveals CYP27A1 as a susceptibility gene for sporadic ALS. *PLoS One,* 7**,** e35333.

DIEKSTRA, F. P., VAN DEERLIN, V. M., VAN SWIETEN, J. C., AL-CHALABI, A., LUDOLPH, A. C., WEISHAUPT, J. H., HARDIMAN, O., LANDERS, J. E., BROWN, R. H., JR., VAN ES, M. A., PASTERKAMP, R. J., KOPPERS, M., ANDERSEN, P. M., ESTRADA, K., RIVADENEIRA, F., HOFMAN, A., UITTERLINDEN, A. G., VAN DAMME, P., MELKI, J., MEININGER, V., SHATUNOV, A., SHAW, C. E., LEIGH, P. N., SHAW, P. J., MORRISON, K. E., FOGH, I., CHIO, A., TRAYNOR, B. J., CZELL, D., WEBER, M., HEUTINK, P., DE BAKKER, P. I., SILANI, V., ROBBERECHT, W., VAN DEN BERG, L. H. & VELDINK, J. H. 2014. C9orf72 and UNC13A are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: a genome-wide meta-analysis. *Ann Neurol,* 76**,** 120-33.

DIEKSTRA, F. P., VAN VUGHT, P. W., VAN RHEENEN, W., KOPPERS, M., PASTERKAMP, R. J., VAN ES, M. A., SCHELHAAS, H. J., DE VISSER, M., ROBBERECHT, W. & VAN DAMME, P. 2012b. UNC13A is a modifier of survival in amyotrophic lateral sclerosis. *Neurobiology of aging,* 33**,** 630. e3-630. e8.

DONAGHY, C., CLARKE, J., PATTERSON, C., KEE, F., HARDIMAN, O. & PATTERSON, V. 2010. The epidemiology of motor neuron disease in Northern Ireland using capture-recapture methodology. *Amyotrophic Lateral Sclerosis,* 11**,** 374-378.

DONAGHY, C., O'TOOLE, O., SHEEHAN, C., KEE, F., HARDIMAN, O. & PATTERSON, V. 2009. An all-Ireland epidemiological study of MND, 2004–2005. *European journal of neurology,* 16**,** 148-153.

DONNELLY, C. J., GRIMA, J. C. & SATTLER, R. 2014. Aberrant RNA homeostasis in amyotrophic lateral sclerosis: potential for new therapeutic targets? *Neurodegener Dis Manag,* 4**,** 417-37.

DORSEY, E., CONSTANTINESCU, R., THOMPSON, J., BIGLAN, K., HOLLOWAY, R., KIEBURTZ, K., MARSHALL, F., RAVINA, B., SCHIFITTO, G. & SIDEROWF, A. 2007. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology,* 68**,** 384-386.

DREW, A. P., ZHU, D., KIDAMBI, A., LY, C., TEY, S., BREWER, M. H., AHMAD-ANNUAR, A., NICHOLSON, G. A. & KENNERSON, M. L. 2015. Improved inherited peripheral neuropathy genetic diagnosis by whole-exome sequencing. *Molecular Genetics & Genomic Medicine,* 3**,** 143-154.

DU, Y., WEN, Y., GUO, X., HAO, J., WANG, W., HE, A., FAN, Q., LI, P., LIU, L., LIANG, X. & ZHANG, F. 2018. A Genome-wide Expression Association Analysis Identifies Genes and Pathways Associated with Amyotrophic Lateral Sclerosis. *Cell Mol Neurobiol,* 38**,** 635-639.

DUBINSKI, A., GAGNÉ, M., PEYRARD, S., GORDON, D., TALBOT, K. & VANDE VELDE, C. 2023. Stress granule assembly in vivo is deficient in the CNS of mutant TDP-43 ALS mice. *Human Molecular Genetics,* 32**,** 319-332.

DUGGER, B. N. & DICKSON, D. W. 2017. Pathology of Neurodegenerative Diseases. *Cold Spring Harb Perspect Biol,* 9.

EISEN, A. & WEBER, M. 2001. The motor cortex and amyotrophic lateral sclerosis. *Muscle Nerve,* 24**,** 564-73.

EPI, K. C., EPILEPSY PHENOME/GENOME, P., ALLEN, A. S., BERKOVIC, S. F., COSSETTE, P., DELANTY, N., DLUGOS, D., EICHLER, E. E., EPSTEIN, M. P., GLAUSER, T., GOLDSTEIN, D. B., HAN, Y., HEINZEN, E. L., HITOMI, Y., HOWELL, K. B., JOHNSON, M. R., KUZNIECKY, R., LOWENSTEIN, D. H., LU, Y. F., MADOU, M. R., MARSON, A. G., MEFFORD, H. C., ESMAEELI NIEH, S., O'BRIEN, T. J., OTTMAN, R., PETROVSKI, S., PODURI, A., RUZZO, E. K., SCHEFFER, I. E., SHERR, E. H., YUSKAITIS, C. J., ABOU-KHALIL, B., ALLDREDGE, B. K., BAUTISTA, J. F., BERKOVIC, S. F., BORO, A., CASCINO, G. D., CONSALVO, D., CRUMRINE, P., DEVINSKY, O., DLUGOS, D., EPSTEIN, M. P., FIOL, M., FOUNTAIN, N. B., FRENCH, J., FRIEDMAN, D., GELLER, E. B., GLAUSER, T., GLYNN, S., HAUT, S. R., HAYWARD, J., HELMERS, S. L., JOSHI, S., KANNER, A., KIRSCH, H. E., KNOWLTON, R. C., KOSSOFF, E. H., KUPERMAN, R., KUZNIECKY, R., LOWENSTEIN, D. H., MCGUIRE, S. M., MOTIKA, P. V., NOVOTNY, E. J., OTTMAN, R., PAOLICCHI, J. M., PARENT, J. M., PARK, K., PODURI, A., SCHEFFER, I. E., SHELLHAAS, R. A., SHERR, E. H., SHIH, J. J., SINGH, R., SIRVEN, J., SMITH, M. C., SULLIVAN, J., LIN THIO, L., VENKAT, A., VINING, E. P., VON ALLMEN, G. K., WEISENBERG, J. L., WIDDESS-WALSH, P. & WINAWER, M. R. 2013. De novo mutations in epileptic encephalopathies. *Nature,* 501**,** 217-21.

FANG, T., AL KHLEIFAT, A., MEURGEY, J.-H., JONES, A., LEIGH, P. N., BENSIMON, G. & AL-CHALABI, A. 2018. Stage at which riluzole treatment prolongs survival in patients with amyotrophic lateral sclerosis: a retrospective analysis of data from a dose-ranging study. *The Lancet Neurology,* 17**,** 416-422.

FANG, T., JE, G., PACUT, P., KEYHANIAN, K., GAO, J. & GHASEMI, M. 2022. Gene Therapy in Amyotrophic Lateral Sclerosis. *Cells,* 11.

FARAVELLI, I., RIBOLDI, G., NIZZARDO, M., SIMONE, C., ZANETTA, C., BRESOLIN, N., COMI, G. P. & CORTI, S. 2014. Stem cell transplantation for amyotrophic lateral sclerosis: therapeutic potential and perspectives on clinical translation. *Cell Mol Life Sci,* 71**,** 3257-68.

FAULKNER, G. J. & BILLON, V. 2018. L1 retrotransposition in the soma: a field jumping ahead. *Mob DNA,* 9**,** 22.

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. & SIDDIQUE, T. 2011. SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol,* 68**,** 1440-6.

FEDERICO, A., CARDAIOLI, E., DA POZZO, P., FORMICHI, P., GALLUS, G. N. & RADI, E. 2012. Mitochondria, oxidative stress and neurodegeneration. *J Neurol Sci,* 322**,** 254-62.

FLEISCHER, J. A. & LOCKWOOD, C. M. 2014. Newborn screening by whole-genome sequencing: ready for prime time? *Clin Chem,* 60**,** 1243-4.

FOGH, I., RATTI, A., GELLERA, C., LIN, K., TILOCA, C., MOSKVINA, V., CORRADO, L., SORARÙ, G., CEREDA, C. & CORTI, S. 2014. A genome-wide association meta-analysis identifies a novel locus at 17q11. 2 associated with sporadic amyotrophic lateral sclerosis. *Human molecular genetics,* 23**,** 2220-2231.

FONG, G., CHENG, T., LAM, K., CHENG, W., MOK, K., CHEUNG, C., CHIM, C., MAK, W., CHAN, K. & TSANG, K. 2005. An epidemiological study of motor neuron disease in Hong Kong. *Amyotrophic Lateral Sclerosis,* 6**,** 164-168.

FONG, K., YU, Y., CHAN, Y., KAY, R., CHAN, J., YANG, Z., KWAN, M., LEUNG, K., LI, P. & LAM, T. 1996. Motor neuron disease in Hong Kong Chinese: epidemiology and clinical picture. *Neuroepidemiology,* 15**,** 239-245.

FORSBERG, L. A., GISSELSSON, D. & DUMANSKI, J. P. 2017. Mosaicism in health and disease - clones picking up speed. *Nat Rev Genet,* 18**,** 128-142.

FRAGOLA, G., MABB, A. M., TAYLOR-BLAKE, B., NIEHAUS, J. K., CHRONISTER, W. D., MAO, H., SIMON, J. M., YUAN, H., LI, Z., MCCONNELL, M. J. & ZYLKA, M. J. 2020. Deletion of Topoisomerase 1 in excitatory neurons causes genomic instability and early onset neurodegeneration. *Nat Commun,* 11**,** 1962.

FRANK, S. A. 2010. Evolution in health and medicine Sackler colloquium: Somatic evolutionary genomics: mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. *Proc Natl Acad Sci U S A,* 107 Suppl 1**,** 1725-30.

FRANKLIN, H., CLARKE, B. E. & PATANI, R. 2021. Astrocytes and microglia in neurodegenerative diseases: Lessons from human in vitro models. *Prog Neurobiol,* 200**,** 101973.

FUJII, R., OKABE, S., URUSHIDO, T., INOUE, K., YOSHIMURA, A., TACHIBANA, T., NISHIKAWA, T., HICKS, G. G. & TAKUMI, T. 2005. The RNA binding protein TLS is translocated to dendritic spines by mGluR5 activation and regulates spine morphology. *Current Biology,* 15**,** 587-593.

FUJII, R. & TAKUMI, T. 2005. TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. *J Cell Sci,* 118**,** 5755-65.

FUMAGALLI, E., FUNICELLO, M., RAUEN, T., GOBBI, M. & MENNINI, T. 2008. Riluzole enhances the activity of glutamate transporters GLAST, GLT1 and EAAC1. *Eur J Pharmacol,* 578**,** 171-6.

GALLO, V., BUENO-DE-MESQUITA, H. B., VERMEULEN, R., ANDERSEN, P. M., KYROZIS, A., LINSEISEN, J., KAAKS, R., ALLEN, N. E., RODDAM, A. W., BOSHUIZEN, H. C., PEETERS, P. H., PALLI, D., MATTIELLO, A., SIERI, S., TUMINO, R., JIMENEZ-MARTIN, J. M., DIAZ, M. J., SUAREZ, L. R., TRICHOPOULOU, A., AGUDO, A., ARRIOLA, L., BARRICANTE-GURREA, A., BINGHAM, S., KHAW, K. T., MANJER, J., LINDKVIST, B., OVERVAD, K., BACH, F. W., TJONNELAND, A., OLSEN, A., BERGMANN, M. M., BOEING, H., CLAVEL-CHAPELON, F., LUND, E., HALLMANS, G., MIDDLETON, L., VINEIS, P. & RIBOLI, E. 2009. Smoking and risk for amyotrophic lateral sclerosis: analysis of the EPIC cohort. *Ann Neurol,* 65**,** 378-85.

GANESALINGAM, J., AN, J., BOWSER, R., ANDERSEN, P. M. & SHAW, C. E. 2013. pNfH is a promising biomarker for ALS. *Amyotroph Lateral Scler Frontotemporal Degener,* 14**,** 146-9.

GAO, J., WANG, L., HUNTLEY, M. L., PERRY, G. & WANG, X. 2018. Pathomechanisms of TDP-43 in neurodegeneration. *J Neurochem*.

GAO, Y., SUN, Y., FRANK, K. M., DIKKES, P., FUJIWARA, Y., SEIDL, K. J., SEKIGUCHI, J. M., RATHBUN, G. A., SWAT, W. & WANG, J. 1998. A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell,* 95**,** 891-902.

GARBIN, D. Bangladeshi diaspora in the UK: some observations on socio-cultural dynamics, religious trends and transnational politics. Human Rights and Bangladesh Conference, SOAS, London, 2005.

GARVIN, T., ABOUKHALIL, R., KENDALL, J., BASLAN, T., ATWAL, G. S., HICKS, J., WIGLER, M. & SCHATZ, M. C. 2015. Interactive analysis and assessment of single-cell copy-number variations. *Nat Methods,* 12**,** 1058-60.

GEEVASINGA, N., HOWELLS, J., MENON, P., VAN DEN BOS, M., SHIBUYA, K., MATAMALA, J. M., PARK, S. B., BYTH, K., KIERNAN, M. C. & VUCIC, S. 2019. Amyotrophic lateral sclerosis diagnostic index: Toward a personalized diagnosis of ALS. *Neurology,* 92**,** e536-e547.

GEORGOULOPOULOU, E., VINCETI, M., BONVICINI, F., SOLA, P., GOLDONI, C. A., GIROLAMO, G. D., FERRARO, D., NICHELLI, P. & MANDRIOLI, J. 2011. Changing incidence and subtypes of ALS in Modena, Italy: A 10-years prospective study. *Amyotrophic Lateral Sclerosis,* 12**,** 451-457.

GERSHONI, J. M. & PALADE, G. E. 1983. Protein blotting: principles and applications. *Anal Biochem,* 131**,** 1-15.

GIBBS, R. A., BELMONT, J. W., HARDENBOL, P., WILLIS, T. D., YU, F. L., YANG, H. M., CH'ANG, L. Y., HUANG, W., LIU, B., SHEN, Y., TAM, P. K. H., TSUI, L. C., WAYE, M. M. Y., WONG, J. T. F., ZENG, C. Q., ZHANG, Q. R., CHEE, M. S., GALVER, L. M., KRUGLYAK, S., MURRAY, S. S., OLIPHANT, A. R., MONTPETIT, A., HUDSON, T. J., CHAGNON, F., FERRETTI, V., LEBOEUF, M., PHILLIPS, M. S., VERNER, A., KWOK, P. Y., DUAN, S. H., LIND, D. L., MILLER, R. D., RICE, J. P., SACCONE, N. L., TAILLON-MILLER, P., XIAO, M., NAKAMURA, Y., SEKINE, A., SORIMACHI, K., TANAKA, T., TANAKA, Y., TSUNODA, T., YOSHINO, E., BENTLEY, D. R., DELOUKAS, P., HUNT, S., POWELL, D., ALTSHULER, D., GABRIEL, S. B., QIU, R. Z., KEN, A., DUNSTON, G. M., KATO, K., NIIKAWA, N., KNOPPERS, B. M., FOSTER, M. W., CLAYTON, E. W., WANG, V. O., WATKIN, J., GIBBS, R. A., BELMONT, J. W., SODERGREN, E., WEINSTOCK, G. M., WILSON, R. K., FULTON, L. L., ROGERS, J., BIRREN, B. W., HAN, H., WANG, H. G., GODBOUT, M., WALLENBURG, J. C., L'ARCHEVEQUE, P., BELLEMARE, G., TODANI, K., FUJITA, T., TANAKA, S., HOLDEN, A. L., LAI, E. H., COLLINS, F. S., BROOKS, L. D., MCEWEN, J. E., GUYER, M. S., JORDAN, E., PETERSON, J. L., SPIEGEL, J., SUNG, L. M., ZACHARIA, L. F., KENNEDY, K., DUNN, M. G., SEABROOK, R., SHILLITO, M., SKENE, B., STEWART, J. G., VALLE, D. L., CLAYTON, E. W., JORDE, L. B., BELMONT, J. W., CHAKRAVARTI, A., CHO, M. K., DUSTER, T., et al. 2003. The International HapMap Project. *Nature,* 426**,** 789-796.

GNOMAD https://gnomad.broadinstitute.org/.

GOLDSTEIN, O., KEDMI, M., GANA-WEISZ, M., TWITO, S., NEFUSSY, B., VAINER, B., FAINMESSER, Y., ABRAHAM, A., NAYSHOOL, O., ORR-URTREGER, A. & DRORY, V. E. 2019. Rare homozygosity in amyotrophic lateral sclerosis suggests the contribution of recessive variants to disease genetics. *J Neurol Sci,* 402**,** 62-68.

GOSSET, P., CAMU, W., RAOUL, C. & MEZGHRANI, A. 2022. Prionoids in amyotrophic lateral sclerosis. *Brain Commun,* 4**,** fcac145.

GOSSYE, H., ENGELBORGHS, S., VAN BROECKHOVEN, C. & VAN DER ZEE, J. 2020. C9orf72 frontotemporal dementia and/or amyotrophic lateral sclerosis.

GOUTMAN, S. A., HARDIMAN, O., AL-CHALABI, A., CHIÓ, A., SAVELIEFF, M. G., KIERNAN, M. C. & FELDMAN, E. L. 2022. Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *The Lancet Neurology*.

GOWLAND, A., OPIE-MARTIN, S., SCOTT, K. M., JONES, A. R., MEHTA, P. R., BATTS, C. J., ELLIS, C. M., LEIGH, P. N., SHAW, C. E., SREEDHARAN, J. & AL-CHALABI, A. 2019. Predicting the future of ALS: the impact of demographic change and potential new treatments on the prevalence of ALS in the United Kingdom, 2020-2116. *Amyotroph Lateral Scler Frontotemporal Degener,* 20**,** 264-274.

GRAD, L. I., ROULEAU, G. A., RAVITS, J. & CASHMAN, N. R. 2017. Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS). *Cold Spring Harb Perspect Med,* 7.

GRATTEN, J., VISSCHER, P. M., MOWRY, B. J. & WRAY, N. R. 2013. Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. *Nat Genet,* 45**,** 234-8.

GROEN, E. J., VAN ES, M. A., VAN VUGHT, P. W., SPLIET, W. G., VAN ENGELEN-LEE, J., DE VISSER, M., WOKKE, J. H., SCHELHAAS, H. J., OPHOFF, R. A., FUMOTO, K., PASTERKAMP, R. J., DOOIJES, D., CUPPEN, E., VELDINK, J. H. & VAN DEN BERG, L. H. 2010. FUS mutations in familial amyotrophic lateral sclerosis in the Netherlands. *Arch Neurol,* 67**,** 224-30.

GROSSMAN, A. B., WOOLLEY-LEVINE, S., BRADLEY, W. G. & MILLER, R. G. 2007. Detecting neurobehavioral changes in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler,* 8**,** 56-61.

GUIDETTI, D., BONDAVALLI, M., SABADINI, R., MARCELLO, N., VINCETI, M., CAVALLETTI, S., MARBINI, A., GEMIGNANI, F., COLOMBO, A. & FERRARI, A. 1996. Epidemiological survey of amyotrophic lateral sclerosis in the province of Reggio Emilia, Italy: influence of environmental exposure to lead. *Neuroepidemiology,* 15**,** 301-312.

GUNDERSEN, M., YASEEN, R. & MIDGARD, R. 2011. Incidence and clinical features of amyotrophic lateral sclerosis in Møre and Romsdal County, Norway. *Neuroepidemiology,* 37**,** 58-63.

HARDIMAN, O., AL-CHALABI, A., CHIO, A., CORR, E. M., LOGROSCINO, G., ROBBERECHT, W., SHAW, P. J., SIMMONS, Z. & VAN DEN BERG, L. H. 2017. Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers,* 3**,** 17071.

HARRISON, J. E., WEBER, S., JAKOB, R. & CHUTE, C. G. 2021. ICD-11: an international classification of diseases for the twenty-first century. *BMC medical informatics and decision making,* 21**,** 1-10.

HARWOOD, C. A., MCDERMOTT, C. J. & SHAW, P. J. J. A. L. S. 2009. Physical activity as an exogenous risk factor in motor neuron disease (MND): a review of the evidence. 10**,** 191-204.

HASEGAWA, M., ARAI, T., NONAKA, T., KAMETANI, F., YOSHIDA, M., HASHIZUME, Y., BEACH, T. G., BURATTI, E., BARALLE, F., MORITA, M., NAKANO, I., ODA, T., TSUCHIYA, K. & AKIYAMA, H.

2008. Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann Neurol,* 64**,** 60-70.

HAVERKAMP, L. J., APPEL, V. & APPEL, S. H. J. B. 1995. Natural history of amyotrophic lateral sclerosis in a database population Validation of a scoring system and a model for survival prediction. 118**,** 707-719.

HENKEL, J. S., BEERS, D. R., WEN, S., RIVERA, A. L., TOENNIS, K. M., APPEL, J. E., ZHAO, W., MOORE, D. H., POWELL, S. Z. & APPEL, S. H. 2013. Regulatory T-lymphocytes mediate amyotrophic lateral sclerosis progression and survival. *EMBO Mol Med,* 5**,** 64-79.

HENKEL, J. S., ENGELHARDT, J. I., SIKLOS, L., SIMPSON, E. P., KIM, S. H., PAN, T., GOODMAN, J. C., SIDDIQUE, T., BEERS, D. R. & APPEL, S. H. 2004. Presence of dendritic cells, MCP-1, and activated microglia/macrophages in amyotrophic lateral sclerosis spinal cord tissue. *Ann Neurol,* 55**,** 221-35.

HENRY, J. B. 2001. Todd, Sanford, Davidsohn Clinical Diagnosis and Management by.

HERGESHEIMER, R. C., CHAMI, A. A., DE ASSIS, D. R., VOURC'H, P., ANDRES, C. R., CORCIA, P., LANZNASTER, D. & BLASCO, H. 2019. The debated toxic role of aggregated TDP-43 in amyotrophic lateral sclerosis: a resolution in sight? *Brain,* 142**,** 1176-1194.

HEWITT, C., KIRBY, J., HIGHLEY, J. R., HARTLEY, J. A., HIBBERD, R., HOLLINGER, H. C., WILLIAMS, T. L., INCE, P. G., MCDERMOTT, C. J. & SHAW, P. J. 2010a. Novel FUS/TLS Mutations and Pathology in Familial and Sporadic Amyotrophic Lateral Sclerosis. *Archives of Neurology,* 67**,** 455-461.

HEWITT, C., KIRBY, J., HIGHLEY, J. R., HARTLEY, J. A., HIBBERD, R., HOLLINGER, H. C., WILLIAMS, T. L., INCE, P. G., MCDERMOTT, C. J. & SHAW, P. J. 2010b. Novel FUS/TLS mutations and pathology in familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol,* 67**,** 455-61.

HIRANO, M., QUINZII, C. M., MITSUMOTO, H., HAYS, A. P., ROBERTS, J. K., RICHARD, P. & ROWLAND, L. P. 2011. Senataxin mutations and amyotrophic lateral sclerosis. *Amyotroph Lateral Scler,* 12**,** 223-7.

HIRAYAMA, K. 2008. [Juvenile muscular atrophy of unilateral upper extremity (Hirayama disease)--half-century progress and establishment since its discovery]. *Brain Nerve,* 60**,** 17-29.

HRA The Health Research Authority. website: https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/sheffield-brain-tissue-bank/.

HUBER, S. & HENN, V. 1995. Unchanged incidence and prevalence of amyotrophic lateral sclerosis in the canton of Zürich. *Schweizer Archiv fur Neurologie und Psychiatrie (Zurich, Switzerland: 1985),* 146**,** 52-54.

HUBERS, A., MARROQUIN, N., SCHMOLL, B., VIELHABER, S., JUST, M., MAYER, B., HOGEL, J., DORST, J., MERTENS, T., JUST, W., AULITZKY, A., WAIS, V., LUDOLPH, A. C., KUBISCH, C., WEISHAUPT, J. H. & VOLK, A. E. 2014. Polymerase chain reaction and Southern blot-based

analysis of the C9orf72 hexanucleotide repeat in different motor neuron diseases. *Neurobiol Aging,* 35**,** 1214 e1-6.

HUISMAN, M. H., DE JONG, S. W., VAN DOORMAAL, P. T., WEINREICH, S. S., SCHELHAAS, H. J., VAN DER KOOI, A. J., DE VISSER, M., VELDINK, J. H. & VAN DEN BERG, L. H. 2011. Population based epidemiology of amyotrophic lateral sclerosis using capture–recapture methodology. *Journal of Neurology, Neurosurgery & Psychiatry,* 82**,** 1165-1170.

ICSL, I. C. S. L. 2019. file:///Users/uos/Thesis%20Writing/Final%20Draft/Reference/v1.5.2/ICSL_Variant_Classification_Criteria_13_December_2019%20(2).pdf.

INCE, P. G., LOWE, J. & SHAW, P. J. 1998a. Amyotrophic lateral sclerosis: current issues in classification, pathogenesis and molecular pathology. *Neuropathol Appl Neurobiol,* 24**,** 104-17.

INCE, P. G., TOMKINS, J., SLADE, J. Y., THATCHER, N. M. & SHAW, P. J. 1998b. Amyotrophic lateral sclerosis associated with genetic abnormalities in the gene encoding Cu/Zn superoxide dismutase: molecular pathology of five new cases, and comparison with previous reports and 73 sporadic cases of ALS. *J Neuropathol Exp Neurol,* 57**,** 895-904.

ISHIGURO, A., KIMURA, N., WATANABE, Y., WATANABE, S. & ISHIHAMA, A. 2016. TDP-43 binds and transports G-quadruplex-containing m RNA s into neurites for local translation. *Genes to Cells,* 21**,** 466-481.

IYAMA, T. & WILSON, D. M., 3RD 2013. DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair (Amst),* 12**,** 620-36.

JEON, G. S., SHIM, Y. M., LEE, D. Y., KIM, J. S., KANG, M., AHN, S. H., SHIN, J. Y., GEUM, D., HONG, Y. H. & SUNG, J. J. 2019. Pathological Modification of TDP-43 in Amyotrophic Lateral Sclerosis with SOD1 Mutations. *Mol Neurobiol,* 56**,** 2007-2021.

JOENSEN, P. 2012. Incidence of amyotrophic lateral sclerosis in the Faroe Islands. *Acta neurologica Scandinavica,* 126**,** 62-66.

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. 2010a. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron,* 68**,** 857-864.

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., MONSURRO, M. R., BATTISTINI, S., SALVI, F., SPATARO, R., SOLA, P., BORGHERO, G., CONSORTIUM, I., GALASSI, G., SCHOLZ, S. W., TAYLOR, J. P., RESTAGNO, G., CHIO, A. & TRAYNOR, B. J. 2010b. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron,* 68**,** 857-64.

JOVICIC, A., PAUL, J. W., 3RD & GITLER, A. D. 2016. Nuclear transport dysfunction: a common theme in amyotrophic lateral sclerosis and frontotemporal dementia. *J Neurochem,* 138 Suppl 1**,** 134-44.

JULIAN, T. H., GLASCOW, N., BARRY, A. D. F., MOLL, T., HARVEY, C., KLIMENTIDIS, Y. C., NEWELL, M., ZHANG, S., SNYDER, M. P. & COOPER-KNOCK, J. 2021. Physical exercise is a risk factor for amyotrophic lateral sclerosis: Convergent evidence from Mendelian randomisation, transcriptomics and risk genotypes. *EBioMedicine,* 68**,** 103397.

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. & ROULEAU, G. A. 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet,* 40**,** 572-4.

KACHER, R., LEJEUNE, F. X., NOEL, S., CAZENEUVE, C., BRICE, A., HUMBERT, S. & DURR, A. 2021. Propensity for somatic expansion increases over the course of life in Huntington disease. *Elife,* 10.

KAHANA, E., ALTER, M. & FELDMAN, S. 1976. Amyotrophic lateral sclerosis: a population study. *J Neurol,* 212**,** 205-13.

KASARSKIS, E. J., LINDQUIST, J. H., COFFMAN, C. J., GRAMBOW, S. C., FEUSSNER, J. R., ALLEN, K. D., ODDONE, E. Z., KAMINS, K. A., HORNER, R. D. & ALS GULF WAR CLINICAL REVIEW, T. 2009. Clinical aspects of ALS in Gulf War veterans. *Amyotroph Lateral Scler,* 10**,** 35-41.

KAWAHARA, Y. & MIEDA-SATO, A. 2012. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proceedings of the National Academy of Sciences,* 109**,** 3347-3352.

KAWAKAMI, I., ARAI, T. & HASEGAWA, M. 2019. The basis of clinicopathological heterogeneity in TDP-43 proteinopathy. *Acta neuropathologica,* 138**,** 751-770.

KENNA, K. P., VAN DOORMAAL, P. T., DEKKER, A. M., TICOZZI, N., KENNA, B. J., DIEKSTRA, F. P., VAN RHEENEN, W., VAN EIJK, K. R., JONES, A. R., KEAGLE, P., SHATUNOV, A., SPROVIERO, W., SMITH, B. N., VAN ES, M. A., TOPP, S. D., KENNA, A., MILLER, J. W., FALLINI, C., TILOCA, C., MCLAUGHLIN, R. L., VANCE, C., TROAKES, C., COLOMBRITA, C., MORA, G., CALVO, A., VERDE, F., AL-SARRAJ, S., KING, A., CALINI, D., DE BELLEROCHE, J., BAAS, F., VAN DER KOOI, A. J., DE VISSER, M., TEN ASBROEK, A. L., SAPP, P. C., MCKENNA-YASEK, D., POLAK, M., ASRESS, S., MUNOZ-BLANCO, J. L., STROM, T. M., MEITINGER, T., MORRISON, K. E., CONSORTIUM, S., LAURIA, G., WILLIAMS, K. L., LEIGH, P. N., NICHOLSON, G. A., BLAIR, I. P., LEBLOND, C. S., DION, P. A., ROULEAU, G. A., PALL, H., SHAW, P. J., TURNER, M. R., TALBOT, K., TARONI, F., BOYLAN, K. B., VAN BLITTERSWIJK, M., RADEMAKERS, R., ESTEBAN-PEREZ, J., GARCIA-REDONDO, A., VAN DAMME, P., ROBBERECHT, W., CHIO, A., GELLERA, C., DREPPER, C., SENDTNER, M., RATTI, A., GLASS, J. D., MORA, J. S., BASAK, N. A., HARDIMAN, O., LUDOLPH, A. C., ANDERSEN, P. M., WEISHAUPT, J. H., BROWN, R. H., JR., AL-CHALABI, A., SILANI, V., SHAW, C. E., VAN DEN BERG, L. H., VELDINK, J. H. & LANDERS, J. E. 2016. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet,* 48**,** 1037-42.

KENNEDY, W. R., ALTER, M. & SUNG, J. H. 1998. Progressive proximal spinal and bulbar muscular atrophy of late onset: a sex-linked recessive trait. *Neurology,* 50**,** 583 and 10 pages following.

KIERNAN, M. C., VUCIC, S., CHEAH, B. C., TURNER, M. R., EISEN, A., HARDIMAN, O., BURRELL, J. R. & ZOING, M. C. 2011. Amyotrophic lateral sclerosis. *Lancet,* 377**,** 942-55.

KIHIRA, T., YOSHIDA, S., HIRONISHI, M., MIWA, H., OKAMATO, K. & KONDO, T. 2005. Changes in the incidence of amyotrophic lateral sclerosis in Wakayama, Japan. *Amyotrophic Lateral Sclerosis,* 6**,** 155-163.

KIM, J., HUGHES, E. G., SHETTY, A. S., ARLOTTA, P., GOFF, L. A., BERGLES, D. E. & BROWN, S. P. 2017. Changes in the Excitability of Neocortical Neurons in a Mouse Model of Amyotrophic Lateral Sclerosis Are Not Specific to Corticospinal Neurons and Are Modulated by Advancing Disease. *J Neurosci,* 37**,** 9037-9053.

KIM, Y. E., OH, K. W., KWON, M. J., CHOI, W. J., OH, S. I., KI, C. S. & KIM, S. H. 2015. De novo FUS mutations in 2 Korean patients with sporadic amyotrophic lateral sclerosis. *Neurobiol Aging,* 36**,** 1604 e17-9.

KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. 2014a. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet,* 46**,** 310-5.

KIRCHER, M., WITTEN, D. M., JAIN, P., O'ROAK, B. J., COOPER, G. M. & SHENDURE, J. 2014b. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics,* 46**,** 310-+.

KLIM, J. R., VANCE, C. & SCOTTER, E. L. 2019. Antisense oligonucleotide therapies for Amyotrophic Lateral Sclerosis: Existing and emerging targets. *Int J Biochem Cell Biol,* 110**,** 149-153.

KONNO, T., SHIGA, A., TSUJINO, A., SUGAI, A., KATO, T., KANAI, K., YOKOSEKI, A., EGUCHI, H., KUWABARA, S., NISHIZAWA, M., TAKAHASHI, H. & ONODERA, O. 2013. Japanese amyotrophic lateral sclerosis patients with GGGGCC hexanucleotide repeat expansion in C9ORF72. *J Neurol Neurosurg Psychiatry,* 84**,** 398-401.

KRETSCHMER, B. D., KRATZER, U. & SCHMIDT, W. J. 1998. Riluzole, a glutamate release inhibitor, and motor behavior. *Naunyn Schmiedebergs Arch Pharmacol,* 358**,** 181-90.

KUMAR, C. V. 2016. *Rational design of enzyme-nanomaterials*, Academic Press.

KUO, P. H., CHIANG, C. H., WANG, Y. T., DOUDEVA, L. G. & YUAN, H. S. 2014. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res,* 42**,** 4712-22.

KURIEN, B. T. & SCOFIELD, R. H. 2015. Western blotting: an introduction. *Methods Mol Biol,* 1312**,** 17-30.

KWIATKOWSKI JR, T., BOSCO, D., LECLERC, A., TAMRAZIAN, E., VANDERBURG, C., RUSS, C., DAVIS, A., GILCHRIST, J., KASARSKIS, E. & MUNSAT, T. 2009. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science,* 323**,** 1205-1208.

KYLLO, T., SINGH, V., SHIM, H., LATIKA, S., NGUYEN, H. M., CHEN, Y. J., TERRY, E., WULFF, H. & ERICKSON, J. D. 2022. Riluzole and novel naphthalenyl substituted aminothiazole derivatives prevent acute neural excitotoxic injury in a rat model of temporal lobe epilepsy. *Neuropharmacology***,** 109349.

LAGIER-TOURENNE, C., POLYMENIDOU, M., HUTT, K. R., VU, A. Q., BAUGHN, M., HUELGA, S. C., CLUTARIO, K. M., LING, S. C., LIANG, T. Y., MAZUR, C., WANCEWICZ, E., KIM, A. S., WATT, A., FREIER, S., HICKS, G. G., DONOHUE, J. P., SHIUE, L., BENNETT, C. F., RAVITS, J., CLEVELAND, D. W. & YEO, G. W. 2012. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat Neurosci,* 15**,** 1488-97.

LAHNEMANN, D., KOSTER, J., SZCZUREK, E., MCCARTHY, D. J., HICKS, S. C., ROBINSON, M. D., VALLEJOS, C. A., CAMPBELL, K. R., BEERENWINKEL, N., MAHFOUZ, A., PINELLO, L., SKUMS, P., STAMATAKIS, A., ATTOLINI, C. S., APARICIO, S., BAAIJENS, J., BALVERT, M., BARBANSON, B., CAPPUCCIO, A., CORLEONE, G., DUTILH, B. E., FLORESCU, M., GURYEV, V., HOLMER, R., JAHN, K., LOBO, T. J., KEIZER, E. M., KHATRI, I., KIELBASA, S. M., KORBEL, J. O., KOZLOV, A. M., KUO, T. H., LELIEVELDT, B. P. F., MANDOIU, II, MARIONI, J. C., MARSCHALL, T., MOLDER, F., NIKNEJAD, A., RACZKOWSKI, L., REINDERS, M., RIDDER, J., SALIBA, A. E., SOMARAKIS, A., STEGLE, O., THEIS, F. J., YANG, H., ZELIKOVSKY, A., MCHARDY, A. C., RAPHAEL, B. J., SHAH, S. P. & SCHONHUTH, A. 2020. Eleven grand challenges in single-cell data science. *Genome Biol,* 21**,** 31.

LALMANSINGH, A. S., UREKAR, C. J. & REDDI, P. P. 2011. TDP-43 is a transcriptional repressor: the testis-specific mouse acrv1 gene is a TDP-43 target in vivo. *Journal of Biological Chemistry,* 286**,** 10970-10982.

LATTANTE, S., DORONZIO, P. N., CONTE, A., MARANGI, G., MARTELLO, F., BISOGNI, G., MELEO, E., COLAVITO, D., DEL GIUDICE, E., PATANELLA, A. K., BERNARDO, D., ROMANO, A., ZOLLINO, M. & SABATELLI, M. 2021. Novel variants and cellular studies on patients' primary fibroblasts support a role for NEK1 missense variants in ALS pathogenesis. *Hum Mol Genet,* 30**,** 65-71.

LATTANTE, S., ROULEAU, G. A. & KABASHI, E. 2013. TARDBP and FUS mutations associated with amyotrophic lateral sclerosis: summary and update. *Hum Mutat,* 34**,** 812-26.

LEIJA-SALAZAR, M., PIETTE, C. & PROUKAKIS, C. 2018. Review: Somatic mutations in neurodegeneration. *Neuropathol Appl Neurobiol,* 44**,** 267-285.

LI, C. Y., JI, Y., TANG, L., ZHANG, N., HE, J., YE, S., LIU, X. L. & FAN, D. S. 2015. Optineurin mutations in patients with sporadic amyotrophic lateral sclerosis in China. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration,* 16**,** 485-489.

LIAO, J., XU, X. & WARGOVICH, M. J. 2000. Direct reprobing with anti-β-actin antibody as an internal control for Western blotting analysis. *Biotechniques,* 28**,** 216-218.

LIAO, Y. Z., MA, J. & DOU, J. Z. 2022. The Role of TDP-43 in Neurodegenerative Disease. *Mol Neurobiol,* 59**,** 4223-4241.

LIM, Y.-M., KOH, I., PARK, Y.-M., KIM, J.-J., KIM, D.-S., KIM, H.-J., BAIK, K.-H., CHOI, H.-Y., YANG, G.-S. & ALSO-RALLO, E. 2012. Exome sequencing identifies KIAA1377 and C5orf42 as susceptibility genes for monomelic amyotrophy. *Neuromuscular Disorders,* 22**,** 394-400.

LIMA, W. F., ROSE, J. B., NICHOLS, J. G., WU, H., MIGAWA, M. T., WYRZYKIEWICZ, T. K., SIWKOWSKI, A. M. & CROOKE, S. T. 2007. Human RNase H1 discriminates between subtle variations in the structure of the heteroduplex substrate. *Molecular pharmacology,* 71**,** 83-91.

LIN, M. T. & BEAL, M. F. 2006. Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. *Nature,* 443**,** 787-95.

LING, J. P., PLETNIKOVA, O., TRONCOSO, J. C. & WONG, P. C. 2015. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science,* 349**,** 650-655.

LIU, X., LI, X., QIAO, Q., LI, F. & WEI, G. 2023. ALS-Linked A315T and A315E Mutations Enhance β-Barrel Formation of the TDP-43307–319 Hexamer: A REST2 Simulation Study. *ACS Chemical Neuroscience*.

LODATO, M. A. & WALSH, C. A. 2019. Genome aging: somatic mutation in the brain links age-related decline with disease and nominates pathogenic mechanisms. *Hum Mol Genet,* 28**,** R197-R206.

LOKANGA, R. A., ENTEZAM, A., KUMARI, D., YUDKIN, D., QIN, M., SMITH, C. B. & USDIN, K. 2013. Somatic expansion in mouse and human carriers of fragile X premutation alleles. *Hum Mutat,* 34**,** 157-66.

LÓPEZ-ERAUSKIN, J., TADOKORO, T., BAUGHN, M. W., MYERS, B., MCALONIS-DOWNES, M., CHILLON-MARINAS, C., ASIABAN, J. N., ARTATES, J., BUI, A. T. & VETTO, A. P. 2020. ALS/FTD-linked mutation in FUS suppresses intra-axonal protein synthesis and drives disease without nuclear loss-of-function of FUS. *Neuron,* 106**,** 354.

LOSSOS, A., ELAZAR, N., LERER, I., SCHUELER-FURMAN, O., FELLIG, Y., GLICK, B., ZIMMERMAN, B. E., AZULAY, H., DOTAN, S., GOLDBERG, S., GOMORI, J. M., PONGER, P., NEWMAN, J. P., MARREED, H., STECK, A. J., SCHAEREN-WIEMERS, N., MOR, N., HAREL, M., GEIGER, T., ESHED-EISENBACH, Y., MEINER, V. & PELES, E. 2015. Myelin-associated glycoprotein gene mutation causes Pelizaeus-Merzbacher disease-like disorder. *Brain,* 138**,** 2521-2536.

LOWE, D. 2022. A new ALS drug. *Science*.

LUNN, J. S., SAKOWSKI, S. A., FEDERICI, T., GLASS, J. D., BOULIS, N. M. & FELDMAN, E. L. 2011. Stem cell technology for the study and treatment of motor neuron diseases. *Regen Med,* 6**,** 201-13.

LUTY, A. A., KWOK, J. B., DOBSON-STONE, C., LOY, C. T., COUPLAND, K. G., KARLSTROM, H., SOBOW, T., TCHORZEWSKA, J., MARUSZAK, A., BARCIKOWSKA, M., PANEGYRES, P. K., ZEKANOWSKI, C., BROOKS, W. S., WILLIAMS, K. L., BLAIR, I. P., MATHER, K. A., SACHDEV, P. S., HALLIDAY, G. M. & SCHOFIELD, P. R. 2010. Sigma nonopioid intracellular receptor 1 mutations cause frontotemporal lobar degeneration-motor neuron disease. *Ann Neurol,* 68**,** 639-49.

MA, X. R., PRUDENCIO, M., KOIKE, Y., VATSAVAYAI, S. C., KIM, G., HARBINSKI, F., BRINER, A., RODRIGUEZ, C. M., GUO, C., AKIYAMA, T., SCHMIDT, H. B., CUMMINGS, B. B., WYATT, D. W., KURYLO, K., MILLER, G., MEKHOUBAD, S., SALLEE, N., MEKONNEN, G., GANSER, L., RUBIEN, J. D., JANSEN-WEST, K., COOK, C. N., PICKLES, S., OSKARSSON, B., GRAFF-RADFORD, N. R., BOEVE, B. F., KNOPMAN, D. S., PETERSEN, R. C., DICKSON, D. W., SHORTER, J., MYONG, S., GREEN, E. M., SEELEY, W. W., PETRUCELLI, L. & GITLER, A. D. 2022. TDP-43 represses cryptic exon inclusion in the FTD-ALS gene UNC13A. *Nature,* 603**,** 124-130.

MACKENZIE, I. R., BIGIO, E. H., INCE, P. G., GESER, F., NEUMANN, M., CAIRNS, N. J., KWONG, L. K., FORMAN, M. S., RAVITS, J., STEWART, H., EISEN, A., MCCLUSKY, L., KRETZSCHMAR, H. A., MONORANU, C. M., HIGHLEY, J. R., KIRBY, J., SIDDIQUE, T., SHAW, P. J., LEE, V. M. & TROJANOWSKI, J. Q. 2007. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Ann Neurol,* 61**,** 427-34.

MACOSKO, E. Z. & MCCARROLL, S. A. 2012. Exploring the variation within. *Nat Genet,* 44**,** 614-6.

MADABHUSHI, R., GAO, F., PFENNING, A. R., PAN, L., YAMAKAWA, S., SEO, J., RUEDA, R., PHAN, T. X., YAMAKAWA, H., PAO, P. C., STOTT, R. T., GJONESKA, E., NOTT, A., CHO, S., KELLIS, M. & TSAI, L. H. 2015. Activity-Induced DNA Breaks Govern the Expression of Neuronal Early-Response Genes. *Cell,* 161**,** 1592-605.

MAGRANE, J., CORTEZ, C., GAN, W. B. & MANFREDI, G. 2014. Abnormal mitochondrial transport and morphology are common pathological denominators in SOD1 and TDP43 ALS mouse models. *Hum Mol Genet,* 23**,** 1413-24.

MAJOUNIE, E., RENTON, A. E., MOK, K., DOPPER, E. G., WAITE, A., ROLLINSON, S., CHIO, A., RESTAGNO, G., NICOLAOU, N., SIMON-SANCHEZ, J., VAN SWIETEN, J. C., ABRAMZON, Y., JOHNSON, J. O., SENDTNER, M., PAMPHLETT, R., ORRELL, R. W., MEAD, S., SIDLE, K. C., HOULDEN, H., ROHRER, J. D., MORRISON, K. E., PALL, H., TALBOT, K., ANSORGE, O., CHROMOSOME, A. L. S. F. T. D. C., FRENCH RESEARCH NETWORK ON, F. F. A., CONSORTIUM, I., HERNANDEZ, D. G., AREPALLI, S., SABATELLI, M., MORA, G., CORBO, M., GIANNINI, F., CALVO, A., ENGLUND, E., BORGHERO, G., FLORIS, G. L., REMES, A. M., LAAKSOVIRTA, H., MCCLUSKEY, L., TROJANOWSKI, J. Q., VAN DEERLIN, V. M., SCHELLENBERG, G. D., NALLS, M. A., DRORY, V. E., LU, C. S., YEH, T. H., ISHIURA, H., TAKAHASHI, Y., TSUJI, S., LE BER, I., BRICE, A., DREPPER, C., WILLIAMS, N., KIRBY, J., SHAW, P., HARDY, J., TIENARI, P. J., HEUTINK, P., MORRIS, H. R., PICKERING-BROWN, S. & TRAYNOR, B. J. 2012. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol,* 11**,** 323-30.

MALIK, I., KELLEY, C. P., WANG, E. T. & TODD, P. K. 2021. Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat Rev Mol Cell Biol,* 22**,** 589-607.

MANDRIOLI, J., FAGLIONI, P., MERELLI, E. & SOLA, P. 2003. The epidemiology of ALS in Modena, Italy. *Neurology,* 60**,** 683-689.

MANTOVANI, S., GARBELLI, S., PASINI, A., ALIMONTI, D., PEROTTI, C., MELAZZINI, M., BENDOTTI, C. & MORA, G. 2009. Immune system alterations in sporadic amyotrophic lateral sclerosis patients suggest an ongoing neuroinflammatory process. *J Neuroimmunol,* 210**,** 73-9.

MARANGI, G. & TRAYNOR, B. J. 2015. Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. *Brain Res,* 1607**,** 75-93.

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. & KAWAKAMI, H. 2010a. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature,* 465**,** 223-6.

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. & KAWAKAMI, H. 2010b. Mutations of optineurin in amyotrophic lateral sclerosis. *Nature,* 465**,** 223-U109.

MASRORI, P. & VAN DAMME, P. 2020. Amyotrophic lateral sclerosis: a clinical review. *Eur J Neurol,* 27**,** 1918-1929.

MATHER, C. A., MOONEY, S. D., SALIPANTE, S. J., SCROGGINS, S., WU, D., PRITCHARD, C. C. & SHIRTS, B. H. 2016. CADD score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. *Genetics in Medicine,* 18**,** 1269-1275.

MAZZINI, L., FERRARI, D., ANDJUS, P. R., BUZANSKA, L., CANTELLO, R., DE MARCHI, F., GELATI, M., GINIATULLIN, R., GLOVER, J. C., GRILLI, M., KOZLOVA, E. N., MAIOLI, M., MITRECIC, D., PIVORIUNAS, A., SANCHEZ-PERNAUTE, R., SARNOWSKA, A., VESCOVI, A. L. & NEUROLOGY, B. C. A. W. 2018. Advances in stem cell therapy for amyotrophic lateral sclerosis. *Expert Opin Biol Ther,* 18**,** 865-881.

MCALARY, L., PLOTKIN, S. S., YERBURY, J. J. & CASHMAN, N. R. 2019. Prion-Like Propagation of Protein Misfolding and Aggregation in Amyotrophic Lateral Sclerosis. *Front Mol Neurosci,* 12**,** 262.

MCCOMBE, P. A. & HENDERSON, R. D. 2010. Effects of gender in amyotrophic lateral sclerosis. *Gend Med,* 7**,** 557-70.

MCKINNON, P. J. 2017. Genome integrity and disease prevention in the nervous system. *Genes Dev,* 31**,** 1180-1194.

MEAD, R. J., SHAN, N., REISER, H. J., MARSHALL, F. & SHAW, P. J. 2022. Amyotrophic lateral sclerosis: a neurodegenerative disorder poised for successful therapeutic translation. *Nat Rev Drug Discov***,** 1-28.

MEAMAR, R., NASR-ESFAHANI, M. H., MOUSAVI, S. A. & BASIRI, K. 2013. Stem cell therapy in amyotrophic lateral sclerosis. *J Clin Neurosci,* 20**,** 1659-63.

MELTZ STEINBERG, K., NICHOLAS, T. J., KOBOLDT, D. C., YU, B., MARDIS, E. & PAMPHLETT, R. 2015. Whole genome analyses reveal no pathogenetic single nucleotide or structural differences between monozygotic twins discordant for amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration,* 16**,** 385-392.

MEYER, K., FERRAIUOLO, L., MIRANDA, C. J., LIKHITE, S., MCELROY, S., RENUSCH, S., DITSWORTH, D., LAGIER-TOURENNE, C., SMITH, R. A., RAVITS, J., BURGHES, A. H., SHAW, P. J., CLEVELAND, D. W., KOLB, S. J. & KASPAR, B. K. 2014. Direct conversion of patient fibroblasts demonstrates non-cell autonomous toxicity of astrocytes to motor neurons in familial and sporadic ALS. *Proc Natl Acad Sci U S A,* 111**,** 829-32.

MEYEROWITZ, J., PARKER, S. J., VELLA, L. J., NG, D., PRICE, K. A., LIDDELL, J. R., CARAGOUNIS, A., LI, Q. X., MASTERS, C. L., NONAKA, T., HASEGAWA, M., BOGOYEVITCH, M. A., KANNINEN, K. M., CROUCH, P. J. & WHITE, A. R. 2011. C-Jun N-terminal kinase controls TDP-43 accumulation in stress granules induced by oxidative stress. *Mol Neurodegener,* 6**,** 57.

MILLER, T. M., CUDKOWICZ, M. E., GENGE, A., SHAW, P. J., SOBUE, G., BUCELLI, R. C., CHIO, A., VAN DAMME, P., LUDOLPH, A. C., GLASS, J. D., ANDREWS, J. A., BABU, S., BENATAR, M., MCDERMOTT, C. J., COCHRANE, T., CHARY, S., CHEW, S., ZHU, H., WU, F., NESTOROV, I., GRAHAM, D., SUN, P., MCNEILL, M., FANNING, L., FERGUSON, T. A., FRADETTE, S., VALOR & GROUP, O. L. E. W. 2022. Trial of Antisense Oligonucleotide Tofersen for SOD1 ALS. *N Engl J Med,* 387**,** 1099-1110.

MINE. *The Project MinE* [Online]. Available: https://www.projectmine.com/ [Accessed].

MINE, T. P. website: https://www.projectmine.com/.

MITRA, J., GUERRERO, E. N., HEGDE, P. M., LIACHKO, N. F., WANG, H., VASQUEZ, V., GAO, J., PANDEY, A., TAYLOR, J. P. & KRAEMER, B. C. 2019. Motor neuron disease-associated loss of nuclear TDP-43 is linked to DNA double-strand break repair defects. *Proceedings of the National Academy of Sciences,* 116**,** 4696-4705.

MOLONEY, E. B., DE WINTER, F. & VERHAAGEN, J. 2014. ALS as a distal axonopathy: molecular mechanisms affecting neuromuscular junction stability in the presymptomatic stages of the disease. *Frontiers in neuroscience,* 8**,** 252.

MORGAN, S. & ORRELL, R. W. 2016. Pathogenesis of amyotrophic lateral sclerosis. *Br Med Bull,* 119**,** 87-98.

MORGAN, S., SHATUNOV, A., SPROVIERO, W., JONES, A. R., SHOAI, M., HUGHES, D., AL KHLEIFAT, A., MALASPINA, A., MORRISON, K. E., SHAW, P. J., SHAW, C. E., SIDLE, K., ORRELL, R. W., FRATTA, P., HARDY, J., PITTMAN, A. & AL-CHALABI, A. 2017. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. *Brain,* 140**,** 1611-1618.

MORGAN, S., SHOAI, M., FRATTA, P., SIDLE, K., ORRELL, R., SWEENEY, M. G., SHATUNOV, A., SPROVIERO, W., JONES, A. & AL-CHALABI, A. J. N. O. A. 2015. Investigation of next-generation sequencing technologies as a diagnostic tool for amyotrophic lateral sclerosis. 36**,** 1600. e5-1600. e8.

MOTATAIANU, A., SERBAN, G., BARCUTEAN, L. & BALASA, R. 2022. Oxidative Stress in Amyotrophic Lateral Sclerosis: Synergy of Genetic and Environmental Factors. *Int J Mol Sci,* 23.

MT-DOCUMENTATION. *https://www.mutationtaster.org/info/documentation.html* [Online]. [Accessed].

MUKHERJEE, O., DAS, G., SEN, S., DUTT, A., ALLADI, S. & GHOSH, A. 2016. C9orf72 mutations may be rare in frontotemporal lobar degeneration patients in India. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration,* 17**,** 151-153.

MÜLLER, H.-P., AGOSTA, F., RIVA, N., SPINELLI, E. G., COMI, G., LUDOLPH, A. C., FILIPPI, M. & KASSUBEK, J. 2018. Fast progressive lower motor neuron disease is an ALS variant: A two-centre tract of interest-based MRI data analysis. *NeuroImage: Clinical,* 17**,** 145-152.

MULLER, K., OH, K. W., NORDIN, A., PANTHI, S., KIM, S. H., NORDIN, F., FREISCHMIDT, A., LUDOLPH, A. C., KI, C. S., FORSBERG, K., WEISHAUPT, J., KIM, Y. E. & ANDERSEN, P. M. 2022. De novo mutations in SOD1 are a cause of ALS. *J Neurol Neurosurg Psychiatry,* 93**,** 201-206.

MUSARÒ, A. 2013. Understanding ALS: new therapeutic approaches. *The FEBS journal,* 280**,** 4315-4322.

NAKAMURA, R., TOHNAI, G., ATSUTA, N., NAKATOCHI, M., HAYASHI, N., WATANABE, H., YOKOI, D., WATANABE, H., KATSUNO, M., IZUMI, Y., TANIGUCHI, A., KANAI, K., MORITA, M., KANO, O., KUWABARA, S., ODA, M., ABE, K., AOKI, M., AIBA, I., OKAMOTO, K., MIZOGUCHI, K., HATTORI, N., NAKASHIMA, K., KAJI, R., SOBUE, G. & AMYOTROPHIC, J. C. 2021. Genetic and functional analysis of KIF5A variants in Japanese patients with sporadic amyotrophic lateral sclerosis. *Neurobiology of Aging,* 97.

NAKASHIMA-YASUDA, H., URYU, K., ROBINSON, J., XIE, S. X., HURTIG, H., DUDA, J. E., ARNOLD, S. E., SIDEROWF, A., GROSSMAN, M., LEVERENZ, J. B., WOLTJER, R., LOPEZ, O. L., HAMILTON, R., TSUANG, D. W., GALASKO, D., MASLIAH, E., KAYE, J., CLARK, C. M., MONTINE, T. J., LEE, V. M. & TROJANOWSKI, J. Q. 2007. Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. *Acta Neuropathol,* 114**,** 221-9.

NALINI, A., THENNARASU, K., GOURIE-DEVI, M., SHENOY, S. & KULSHRESHTHA, D. 2008. Clinical characteristics and survival pattern of 1153 patients with amyotrophic lateral sclerosis: experience over 30 years from India. *Journal of the neurological sciences,* 272**,** 60-70.

NARAIN, P., PANDEY, A., GUPTA, S., GOMES, J., BHATIA, R. & VIVEKANANDAN, P. 2018. Targeted next-generation sequencing reveals novel and rare variants in Indian patients with amyotrophic lateral sclerosis. *Neurobiology of Aging,* 71.

NEALE, B. M., KOU, Y., LIU, L., MA'AYAN, A., SAMOCHA, K. E., SABO, A., LIN, C. F., STEVENS, C., WANG, L. S., MAKAROV, V., POLAK, P., YOON, S., MAGUIRE, J., CRAWFORD, E. L., CAMPBELL, N. G., GELLER, E. T., VALLADARES, O., SCHAFER, C., LIU, H., ZHAO, T., CAI, G., LIHM, J., DANNENFELSER, R., JABADO, O., PERALTA, Z., NAGASWAMY, U., MUZNY, D., REID, J. G., NEWSHAM, I., WU, Y., LEWIS, L., HAN, Y., VOIGHT, B. F., LIM, E., ROSSIN, E., KIRBY, A., FLANNICK, J., FROMER, M., SHAKIR, K., FENNELL, T., GARIMELLA, K., BANKS, E., POPLIN, R., GABRIEL, S., DEPRISTO, M., WIMBISH, J. R., BOONE, B. E., LEVY, S. E., BETANCUR, C., SUNYAEV, S., BOERWINKLE, E., BUXBAUM, J. D., COOK, E. H., JR., DEVLIN, B., GIBBS, R. A., ROEDER, K., SCHELLENBERG, G. D., SUTCLIFFE, J. S. & DALY, M. J. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature,* 485**,** 242-5.

NEL, M., AGENBAG, G. M., HENNING, F., CROSS, H. M., ESTERHUIZEN, A. & HECKMANN, J. M. 2019. C9orf72 repeat expansions in South Africans with amyotrophic lateral sclerosis. *Journal of the neurological sciences,* 401**,** 51-54.

NELSON, P. T., DICKSON, D. W., TROJANOWSKI, J. Q., JACK, C. R., BOYLE, P. A., ARFANAKIS, K., RADEMAKERS, R., ALAFUZOFF, I., ATTEMS, J., BRAYNE, C., COYLE-GILCHRIST, I. T. S., CHUI, H.

C., FARDO, D. W., FLANAGAN, M. E., HALLIDAY, G., HOKKANEN, S. R. K., HUNTER, S., JICHA, G. A., KATSUMATA, Y., KAWAS, C. H., KEENE, C. D., KOVACS, G. G., KUKULL, W. A., LEVEY, A. I., MAKKINEJAD, N., MONTINE, T. J., MURAYAMA, S., MURRAY, M. E., NAG, S., RISSMAN, R. A., SEELEY, W. W., SPERLING, R. A., WHITE, C. L., 3RD, YU, L. & SCHNEIDER, J. A. 2019. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain,* 142**,** 1503-1527.

NEUMANN, M., KWONG, L. K., LEE, E. B., KREMMER, E., FLATLEY, A., XU, Y., FORMAN, M. S., TROOST, D., KRETZSCHMAR, H. A. & TROJANOWSKI, J. Q. 2009. Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. *Acta neuropathologica,* 117**,** 137-149.

NEUMANN, M., KWONG, L. K., SAMPATHU, D. M., TROJANOWSKI, J. Q. & LEE, V. M. 2007a. TDP-43 proteinopathy in frontotemporal lobar degeneration and amyotrophic lateral sclerosis: protein misfolding diseases without amyloidosis. *Arch Neurol,* 64**,** 1388-94.

NEUMANN, M., KWONG, L. K., TRUAX, A. C., VANMASSENHOVE, B., KRETZSCHMAR, H. A., VAN DEERLIN, V. M., CLARK, C. M., GROSSMAN, M., MILLER, B. L., TROJANOWSKI, J. Q. & LEE, V. M. 2007b. TDP-43-positive white matter pathology in frontotemporal lobar degeneration with ubiquitin-positive inclusions. *J Neuropathol Exp Neurol,* 66**,** 177-83.

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. 2006a. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science,* 314**,** 130-133.

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. & LEE, V. M. 2006b. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science,* 314**,** 130-3.

NGUYEN, H. P., VAN BROECKHOVEN, C. & VAN DER ZEE, J. J. T. I. G. 2018a. ALS Genes in the Genomic Era and their Implications for FTD. 34**,** 404-423.

NGUYEN, H. P., VAN MOSSEVELDE, S., DILLEN, L., DE BLEECKER, J. L., MOISSE, M., VAN DAMME, P., VAN BROECKHOVEN, C., VAN DER ZEE, J. & CONSORTIUM, B. 2018b. NEK1 genetic variability in a Belgian cohort of ALS and ALS-FTD patients. *Neurobiol Aging,* 61**,** 255 e1-255 e7.

NICOLAS, A., KENNA, K. P., RENTON, A. E., TICOZZI, N., FAGHRI, F., CHIA, R., DOMINOV, J. A., KENNA, B. J., NALLS, M. A. & KEAGLE, P. 2018. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron,* 97**,** 1268-1283. e6.

NICOLAS, G. & VELTMAN, J. A. 2019. The role of de novo mutations in adult-onset neurodegenerative disorders. *Acta Neuropathol,* 137**,** 183-207.

NISHIMURA, A. L., MITNE-NETO, M., SILVA, H. C., RICHIERI-COSTA, A., MIDDLETON, S., CASCIO, D., KOK, F., OLIVEIRA, J. R., GILLINGWATER, T. & WEBB, J. 2004. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *The American Journal of Human Genetics,* 75**,** 822-831.

NORDIN, A., AKIMOTO, C., WUOLIKAINEN, A., ALSTERMARK, H., JONSSON, P., BIRVE, A., MARKLUND, S. L., GRAFFMO, K. S., FORSBERG, K., BRANNSTROM, T. & ANDERSEN, P. M. 2015. Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. *Hum Mol Genet,* 24**,** 3133-42.

O'TOOLE, O., TRAYNOR, B. J., BRENNAN, P., SHEEHAN, C., FROST, E., CORR, B. & HARDIMAN, O. 2008. Epidemiology and clinical features of amyotrophic lateral sclerosis in Ireland between 1995 and 2004. *Journal of Neurology, Neurosurgery & Psychiatry,* 79**,** 30-32.

OH, J., AN, J. W., OH, S. I., OH, K. W., KIM, J. A., LEE, J. S. & KIM, S. H. 2015. Socioeconomic costs of amyotrophic lateral sclerosis according to staging system. *Amyotroph Lateral Scler Frontotemporal Degener,* 16**,** 202-8.

ORBAN, P., DEVON, R. S., HAYDEN, M. R. & LEAVITT, B. R. 2007. Juvenile amyotrophic lateral sclerosis. *Handbook of clinical neurology.* Elsevier.

ORRELL, R. W. 2000. Amyotrophic lateral sclerosis: copper/zinc superoxide dismutase (SOD1) gene mutations. *Neuromuscul Disord,* 10**,** 63-8.

PAEZ-COLASANTE, X., FIGUEROA-ROMERO, C., SAKOWSKI, S. A., GOUTMAN, S. A. & FELDMAN, E. L. 2015. Amyotrophic lateral sclerosis: mechanisms and therapeutics in the epigenomic era. *Nat Rev Neurol,* 11**,** 266-79.

PAMPHLETT, R. 2004. Somatic mutation: a cause of sporadic neurodegenerative diseases? *Med Hypotheses,* 62**,** 679-82.

PAMPHLETT, R., CHEONG, P. L., TRENT, R. J. & YU, B. 2013. Can ALS-associated C9orf72 repeat expansions be diagnosed on a blood DNA test alone? *PLoS One,* 8**,** e70007.

PARK, J., KIM, J. E. & SONG, T. J. 2022. The Global Burden of Motor Neuron Disease: An Analysis of the 2019 Global Burden of Disease Study. *Front Neurol,* 13**,** 864339.

PETERS, O. M. & BROWN JR, R. H. 2023. Amyotrophic lateral sclerosis. *Neurobiology of Brain Disorders.* Elsevier.

PIAO, Y. S., WAKABAYASHI, K., KAKITA, A., YAMADA, M., HAYASHI, S., MORITA, T., IKUTA, F., OYANAGI, K. & TAKAHASHI, H. 2003. Neuropathology with clinical correlations of sporadic amyotrophic lateral sclerosis: 102 autopsy cases examined between 1962 and 2000. *Brain pathology,* 13**,** 10-22.

POLYMENIDOU, M. & CLEVELAND, D. W. 2011. The seeds of neurodegeneration: prion-like spreading in ALS. *Cell,* 147**,** 498-508.

POLYMENIDOU, M., LAGIER-TOURENNE, C., HUTT, K. R., HUELGA, S. C., MORAN, J., LIANG, T. Y., LING, S. C., SUN, E., WANCEWICZ, E., MAZUR, C., KORDASIEWICZ, H., SEDAGHAT, Y., DONOHUE, J. P., SHIUE, L., BENNETT, C. F., YEO, G. W. & CLEVELAND, D. W. 2011. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci,* 14**,** 459-68.

PROJECT MINE, E. A. L. S. S. C. 2018a. CHCHD10 variants in amyotrophic lateral sclerosis: Where is the evidence? *Ann Neurol,* 84**,** 110-116.

PROJECT MINE, E. A. L. S. S. C. 2018b. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur J Hum Genet,* 26**,** 1537-1546.

PRONKER, M. F., LEMSTRA, S., SNIJDER, J., HECK, A. J. R., THIES-WEESIE, D. M. E., PASTERKAMP, R. J. & JANSSEN, B. J. C. 2016. Structural basis of myelin-associated glycoprotein adhesion and signalling. *Nature Communications,* 7.

PROUKAKIS, C. 2020. Somatic mutations in neurodegeneration: An update. *Neurobiol Dis,* 144**,** 105021.

PROUKAKIS, C., HOULDEN, H. & SCHAPIRA, A. H. 2013. Somatic alpha-synuclein mutations in Parkinson's disease: hypothesis and preliminary data. *Mov Disord,* 28**,** 705-12.

PRUDENCIO, M., DURAZO, A., WHITELEGGE, J. P. & BORCHELT, D. R. 2010. An examination of wild-type SOD1 in modulating the toxicity and aggregation of ALS-associated mutant SOD1. *Hum Mol Genet,* 19**,** 4774-89.

PULST, S. M., NECHIPORUK, A., NECHIPORUK, T., GISPERT, S., CHEN, X. N., LOPES-CENDES, I., PEARLMAN, S., STARKMAN, S., OROZCO-DIAZ, G., LUNKES, A., DEJONG, P., ROULEAU, G. A., AUBURGER, G., KORENBERG, J. R., FIGUEROA, C. & SAHBA, S. 1996. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet,* 14**,** 269-76.

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. & SHAM, P. C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics,* 81**,** 559-575.

RACZY, C., PETROVSKI, R., SAUNDERS, C. T., CHORNY, I., KRUGLYAK, S., MARGULIES, E. H., CHUANG, H. Y., KALLBERG, M., KUMAR, S. A., LIAO, A., LITTLE, K. M., STROMBERG, M. P. & TANNER, S. W. 2013. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics,* 29**,** 2041-2043.

RAKHIT, R., ROBERTSON, J., VANDE VELDE, C., HORNE, P., RUTH, D. M., GRIFFIN, J., CLEVELAND, D. W., CASHMAN, N. R. & CHAKRABARTTY, A. 2007. An immunological epitope selective for pathological monomer-misfolded SOD1 in ALS. *Nat Med,* 13**,** 754-9.

RAKNUZZAMAN, M. & HABIB, M. A. 2020. Demographic Pattern of Amyotrophic Lateral Sclerosis in Bangladesh Among Patient Admitted in a Tertiary Level Hospital.

RANGANATHAN, R., HAQUE, S., COLEY, K., SHEPHEARD, S., COOPER-KNOCK, J. & KIRBY, J. 2020. Multifaceted Genes in Amyotrophic Lateral Sclerosis-Frontotemporal Dementia. *Frontiers in Neuroscience,* 14.

RAVITS, J., APPEL, S., BALOH, R. H., BAROHN, R., BROOKS, B. R., ELMAN, L., FLOETER, M. K., HENDERSON, C., LOMEN-HOERTH, C., MACKLIS, J. D., MCCLUSKEY, L., MITSUMOTO, H., PRZEDBORSKI, S., ROTHSTEIN, J., TROJANOWSKI, J. Q., VAN DEN BERG, L. H. & RINGEL, S.

2013. Deciphering amyotrophic lateral sclerosis: what phenotype, neuropathology and genetics are telling us about pathogenesis. *Amyotroph Lateral Scler Frontotemporal Degener,* 1**,** 5-18.

RENTON, A. E., CHIÒ, A. & TRAYNOR, B. J. 2014. State of play in amyotrophic lateral sclerosis genetics. *Nature neuroscience,* 17**,** 17.

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. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron,* 72**,** 257-268.

RENTZOS, M., EVANGELOPOULOS, E., SERETI, E., ZOUVELOU, V., MARMARA, S., ALEXAKIS, T. & EVDOKIMIDIS, I. 2012. Alterations of T cell subsets in ALS: a systemic immune activation? *Acta Neurol Scand,* 125**,** 260-4.

RENTZSCH, P., WITTEN, D., COOPER, G. M., SHENDURE, J. & KIRCHER, M. 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research,* 47**,** D886-D894.

RICHARDS, C. S., BALE, S., BELLISSIMO, D. B., DAS, S., GRODY, W. W., HEGDE, M. R., LYON, E., WARD, B. E. & MOLECULAR SUBCOMMITTEE OF THE, A. L. Q. A. C. 2008. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med,* 10**,** 294-300.

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. & COMMITTEE, A. L. Q. A. 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. *Genet Med,* 17**,** 405-24.

RIKU, Y., SEILHEAN, D., DUYCKAERTS, C., BOLUDA, S., IGUCHI, Y., ISHIGAKI, S., IWASAKI, Y., YOSHIDA, M., SOBUE, G. & KATSUNO, M. 2021. Pathway from TDP-43-Related Pathology to Neuronal Dysfunction in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration. *Int J Mol Sci,* 22.

RITSON, G. P., CUSTER, S. K., FREIBAUM, B. D., GUINTO, J. B., GEFFEL, D., MOORE, J., TANG, W., WINTON, M. J., NEUMANN, M., TROJANOWSKI, J. Q., LEE, V. M., FORMAN, M. S. & TAYLOR, J. P. 2010. TDP-43 mediates degeneration in a novel Drosophila model of disease caused by mutations in VCP/p97. *J Neurosci,* 30**,** 7729-39.

RIVA, N., POZZI, L., RUSSO, T., PIPITONE, G. B., SCHITO, P., DOMI, T., AGOSTA, F., QUATTRINI, A., CARRERA, P. & FILIPPI, M. 2022. NEK1 Variants in a Cohort of Italian Patients With Amyotrophic Lateral Sclerosis. *Front Neurosci,* 16**,** 833051.

ROGGENBUCK, J., PALETTAS, M., VICINI, L., PATEL, R., QUICK, A. & KOLB, S. J. 2020. Incidence of pathogenic, likely pathogenic, and uncertain ALS variants in a clinic cohort. *Neurol Genet,* 6**,** e390.

ROHRBACK, S., APRIL, C., KAPER, F., RIVERA, R. R., LIU, C. S., SIDDOWAY, B. & CHUN, J. 2018a. Submegabase copy number variations arise during cerebral cortical neurogenesis as revealed by single-cell whole-genome sequencing. *Proc Natl Acad Sci U S A,* 115**,** 10804-10809.

ROHRBACK, S., SIDDOWAY, B., LIU, C. S. & CHUN, J. 2018b. Genomic mosaicism in the developing and adult brain. *Dev Neurobiol,* 78**,** 1026-1048.

ROSEN, D. R. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature,* 364**,** 362.

ROSEN, D. R., SIDDIQUE, T., PATTERSON, D., FIGLEWICZ, D. A., SAPP, P., HENTATI, A., DONALDSON, D., GOTO, J., O'REGAN, J. P. & DENG, H.-X. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature,* 362**,** 59-62.

ROWLAND, L. P. & SHNEIDER, N. A. J. N. E. J. O. M. 2001. Amyotrophic lateral sclerosis. 344**,** 1688-1700.

RUTTEN, B. P., SCHMITZ, C., GERLACH, O. H., OYEN, H. M., DE MESQUITA, E. B., STEINBUSCH, H. W. & KORR, H. 2007. The aging brain: accumulation of DNA damage or neuron loss? *Neurobiol Aging,* 28**,** 91-8.

RYAN, É. B. & DENG, H.-X. 2021. Does Somatic Mosaicism Account for Some Sporadic ALS? : AAN Enterprises.

SABATELLI, M., CONTE, A. & ZOLLINO, M. 2013. Clinical and genetic heterogeneity of amyotrophic lateral sclerosis. *Clin Genet,* 83**,** 408-16.

SACHIDANANDAM, R., WEISSMAN, D., SCHMIDT, S. C., KAKOL, J. M., STEIN, L. D., MARTH, G., SHERRY, S., MULLIKIN, J. C., MORTIMORE, B. J., WILLEY, D. L., HUNT, S. E., COLE, C. G., COGGILL, P. C., RICE, C. M., NING, Z. M., ROGERS, J., BENTLEY, D. R., KWOK, P. Y., MARDIS, E. R., YEH, R. T., SCHULTZ, B., COOK, L., DAVENPORT, R., DANTE, M., FULTON, L., HILLIER, L., WATERSTON, R. H., MCPHERSON, J. D., GILMAN, B., SCHAFFNER, S., VAN ETTEN, W. J., REICH, D., HIGGINS, J., DALY, M. J., BLUMENSTIEL, B., BALDWIN, J., STANGE-THOMANN, N. S., ZODY, M. C., LINTON, L., LANDER, E. S., ALTSHULER, D. & GRP, I. S. M. W. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature,* 409**,** 928-933.

SAHAI, S. P. & GHOSH, S. 2021. Motor neuron disease-The Indian Scenario. *Insights in Neurosurgery,* 5.

SAJJADI, M., ETEMADIFAR, M., NEMATI, A., GHAZAVI, H., BASIRI, K., KHOUNDABI, B., MOUSAVI, S., KABIRI, P. & MAGHZI, A. 2010. Epidemiology of amyotrophic lateral sclerosis in Isfahan, Iran. *European journal of neurology,* 17**,** 984-989.

SAMPATHU, D. M., NEUMANN, M., KWONG, L. K., CHOU, T. T., MICSENYI, M., TRUAX, A., BRUCE, J., GROSSMAN, M., TROJANOWSKI, J. Q. & LEE, V. M. 2006. Pathological heterogeneity of frontotemporal lobar degeneration with ubiquitin-positive inclusions delineated by ubiquitin immunohistochemistry and novel monoclonal antibodies. *Am J Pathol,* 169**,** 1343-52.

SANDERS, S. J., MURTHA, M. T., GUPTA, A. R., MURDOCH, J. D., RAUBESON, M. J., WILLSEY, A. J., ERCAN-SENCICEK, A. G., DILULLO, N. M., PARIKSHAK, N. N., STEIN, J. L., WALKER, M. F., OBER,

G. T., TERAN, N. A., SONG, Y., EL-FISHAWY, P., MURTHA, R. C., CHOI, M., OVERTON, J. D., BJORNSON, R. D., CARRIERO, N. J., MEYER, K. A., BILGUVAR, K., MANE, S. M., SESTAN, N., LIFTON, R. P., GUNEL, M., ROEDER, K., GESCHWIND, D. H., DEVLIN, B. & STATE, M. W. 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature,* 485**,** 237-41.

SAUDAGAR, R., GARGE, L. J. J. O. D. D. & THERAPEUTICS 2019. Amyotrophic Lateral Sclerosis: An Overview. 9**,** 613-616.

SBTB Sheffield Brain Tissue Bank (SBTB): https://directory.biobankinguk.org/Profile/Biobank/GBR-1-249.

SCHWARTZ, G. & FEHLINGS, M. G. 2002. Secondary injury mechanisms of spinal cord trauma: a novel therapeutic approach for the management of secondary pathophysiology with the sodium channel blocker riluzole. *Prog Brain Res,* 137**,** 177-90.

SCHWARZ, J. M., COOPER, D. N., SCHUELKE, M. & SEELOW, D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods,* 11**,** 361-362.

SCHWER, B., WEI, P. C., CHANG, A. N., KAO, J., DU, Z., MEYERS, R. M. & ALT, F. W. 2016. Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. *Proc Natl Acad Sci U S A,* 113**,** 2258-63.

SDGS Sheffield Diagnostic Genetics Service. *https://www.sheffieldchildrens.nhs.uk/sdgs/*.

SERIO, A., BILICAN, B., BARMADA, S. J., ANDO, D. M., ZHAO, C., SILLER, R., BURR, K., HAGHI, G., STORY, D., NISHIMURA, A. L., CARRASCO, M. A., PHATNANI, H. P., SHUM, C., WILMUT, I., MANIATIS, T., SHAW, C. E., FINKBEINER, S. & CHANDRAN, S. 2013. Astrocyte pathology and the absence of non-cell autonomy in an induced pluripotent stem cell model of TDP-43 proteinopathy. *Proc Natl Acad Sci U S A,* 110**,** 4697-702.

SHAHRIZAILA, N., SOBUE, G., KUWABARA, S., KIM, S. H., BIRKS, C., FAN, D. S., BAE, J. S., HU, C. J., GOURIE-DEVI, M., NOTO, Y., SHIBUYA, K., GOH, K. J., KAJI, R., TSAI, C. P., CUI, L., TALMAN, P., HENDERSON, R. D., VUCIC, S. & KIERNAN, M. C. 2016. Amyotrophic lateral sclerosis and motor neuron syndromes in Asia. *J Neurol Neurosurg Psychiatry,* 87**,** 821-30.

SHAMIM, U., AMBAWAT, S., SINGH, J., THOMAS, A., PRADEEP-CHANDRA-REDDY, C., SUROLIYA, V., UPPILLI, B., PARVEEN, S., SHARMA, P. & CHANCHAL, S. 2020. C9orf72 hexanucleotide repeat expansion in Indian patients with ALS: a common founder and its geographical predilection. *Neurobiology of aging,* 88**,** 156. e1-156. e9.

SHATUNOV, A., MOK, K., NEWHOUSE, S., WEALE, M. E., SMITH, B., VANCE, C., JOHNSON, L., VELDINK, J. H., VAN ES, M. A., VAN DEN BERG, L. H., ROBBERECHT, W., VAN DAMME, P., HARDIMAN, O., FARMER, A. E., LEWIS, C. M., BUTLER, A. W., ABEL, O., ANDERSEN, P. M., FOGH, I., SILANI, V., CHIO, A., TRAYNOR, B. J., MELKI, J., MEININGER, V., LANDERS, J. E., MCGUFFIN, P., GLASS, J. D., PALL, H., LEIGH, P. N., HARDY, J., BROWN, R. H., JR., POWELL, J. F., ORRELL, R. W., MORRISON, K. E., SHAW, P. J., SHAW, C. E. & AL-CHALABI, A. 2010. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol,* 9**,** 986-94.

SHAW, P. J., FORREST, V., INCE, P. G., RICHARDSON, J. P. & WASTELL, H. J. 1995. CSF and plasma amino acid levels in motor neuron disease: elevation of CSF glutamate in a subset of patients. *Neurodegeneration,* 4**,** 209-16.

SHAW, P. J., TOMKINS, J., SLADE, J. Y., USHER, P., CURTIS, A., BUSHBY, K. & INCE, P. G. 1997. CNS tissue Cu/Zn superoxide dismutase (SOD1) mutations in motor neurone disease (MND). *Neuroreport,* 8**,** 3923-3927.

SHEPHEARD, S. R., PARKER, M. D., COOPER-KNOCK, J., VERBER, N. S., TUDDENHAM, L., HEATH, P., BEAUCHAMP, N., PLACE, E., SOLLARS, E. S. A., TURNER, M. R., MALASPINA, A., FRATTA, P., HEWAMADDUMA, C., JENKINS, T. M., MCDERMOTT, C. J., WANG, D., KIRBY, J., SHAW, P. J., PROJECT, M. C. & PROJECT MIN, E. 2021. Value of systematic genetic screening of patients with amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry,* 92**,** 510-518.

SIDDIQUE, T., FIGLEWICZ, D. A., PERICAK-VANCE, M. A., HAINES, J. L., ROULEAU, G., JEFFERS, A. J., SAPP, P., HUNG, W. Y., BEBOUT, J., MCKENNA-YASEK, D. & ET AL. 1991. Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. *N Engl J Med,* 324**,** 1381-4.

SITRAN. *Sheffield Institute for Translational Neuroscience: https://sitran.org/facilities/neuropathology/* [Online]. [Accessed].

SLIFER, S. H. 2018. PLINK: Key Functions for Data Analysis. *Curr Protoc Hum Genet,* 97**,** e59.

SMETHURST, P., SIDLE, K. C. & HARDY, J. 2015. Review: Prion-like mechanisms of transactive response DNA binding protein of 43 kDa (TDP-43) in amyotrophic lateral sclerosis (ALS). *Neuropathol Appl Neurobiol,* 41**,** 578-97.

SMEYERS, J., BANCHI, E.-G. & LATOUCHE, M. 2021. C9ORF72: what it is, what it does, and why it matters. *Frontiers in Cellular Neuroscience,* 15**,** 661447.

SMITH, L., CUPID, B. C., DICKIE, B. G. M., AL-CHALABI, A., MORRISON, K. E., SHAW, C. E. & SHAW, P. J. 2015. Establishing the UK DNA Bank for motor neuron disease (MND). *Bmc Genetics,* 16.

SMITH, P. K., KROHN, R. I., HERMANSON, G. T., MALLIA, A. K., GARTNER, F. H., PROVENZANO, M. D., FUJIMOTO, E. K., GOEKE, N. M., OLSON, B. J. & KLENK, D. C. 1985. Measurement of protein using bicinchoninic acid. *Anal Biochem,* 150**,** 76-85.

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. & SHAW, C. E. 2008. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science,* 319**,** 1668-72.

STEIN, J., WALKENFORT, B., CIHANKAYA, H., HASENBERG, M., BADER, V., WINKLHOFER, K. F., RODERER, P., MATSCHKE, J., THEISS, C. & MATSCHKE, V. 2021. Increased ROS-Dependent Fission of Mitochondria Causes Abnormal Morphology of the Cell Powerhouses in a Murine Model of Amyotrophic Lateral Sclerosis. *Oxid Med Cell Longev,* 2021**,** 6924251.

STICKLER, D. E., ROYER, J. A. & HARDIN, J. W. 2011. Validity of hospital discharge data for identifying cases of amyotrophic lateral sclerosis. *Muscle & nerve,* 44**,** 814-815.

STORAGE, S. B.-H. T. Available: https://www.sheffield.ac.uk/medicine/facilities/sheffield-biorepository [Accessed].

STRONG, M. J., HUDSON, A. J. & ALVORD, W. G. 1991. Familial amyotrophic lateral sclerosis, 1850-1989: a statistical analysis of the world literature. *Can J Neurol Sci,* 18**,** 45-58.

STURMBERG, J. P. 2019. *Embracing Complexity in Health: The Transformation of Science, Practice, and Policy*, Springer.

STURMEY, E. & MALASPINA, A. 2022. Blood biomarkers in ALS: challenges, applications and novel frontiers. *Acta Neurol Scand,* 146**,** 375-388.

SUN, Y., CURLE, A. J., HAIDER, A. M. & BALMUS, G. 2020. The role of DNA damage response in amyotrophic lateral sclerosis. *Essays Biochem,* 64**,** 847-861.

SUPEK, F. & LEHNER, B. 2019. Scales and mechanisms of somatic mutation rate variation across the human genome. *DNA Repair (Amst),* 81**,** 102647.

SUZUKI, M., MCHUGH, J., TORK, C., SHELLEY, B., HAYES, A., BELLANTUONO, I., AEBISCHER, P. & SVENDSEN, C. N. 2008. Direct muscle delivery of GDNF with human mesenchymal stem cells improves motor neuron survival and function in a rat model of familial ALS. *Mol Ther,* 16**,** 2002-10.

SWAMI, M., HENDRICKS, A. E., GILLIS, T., MASSOOD, T., MYSORE, J., MYERS, R. H. & WHEELER, V. C. 2009. Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet,* 18**,** 3039-47.

SWINNEN, B. & ROBBERECHT, W. J. N. R. N. 2014. The phenotypic variability of amyotrophic lateral sclerosis. 10**,** 661.

TARIQ, H., TARIQ, I., BOURINARIS, T., HOULDEN, H. & NAZ, S. 2021. Some pathogenic SETX variants are partially conserved during evolution. *Gene,* 771.

TASKENT, R. O. & GOKCUMEN, O. 2017. The Multiple Histories of Western Asia: Perspectives from Ancient and Modern Genomes. *Hum Biol,* 89**,** 107-117.

TAYLOR, J. P., BROWN, R. H., JR. & CLEVELAND, D. W. 2016. Decoding ALS: from genes to mechanism. *Nature,* 539**,** 197-206.

TAZELAAR, G. H. P., DEKKER, A. M., VAN VUGT, J., VAN DER SPEK, R. A., WESTENENG, H. J., KOOL, L., KENNA, K. P., VAN RHEENEN, W., PULIT, S. L., MCLAUGHLIN, R. L., SPROVIERO, W., IACOANGELI, A., HUBERS, A., BRENNER, D., MORRISON, K. E., SHAW, P. J., SHAW, C. E., PANADES, M. P., MORA PARDINA, J. S., GLASS, J. D., HARDIMAN, O., AL-CHALABI, A., VAN DAMME, P., ROBBERECHT, W., LANDERS, J. E., LUDOLPH, A. C., WEISHAUPT, J. H., VAN DEN BERG, L. H., VELDINK, J. H., VAN ES, M. A. & PROJECT MIN, E. A. L. S. S. C. 2019. Association of NIPA1 repeat expansions with amyotrophic lateral sclerosis in a large international cohort. *Neurobiol Aging,* 74**,** 234 e9-234 e15.

THORISSON, G. A., SMITH, A. V., KRISHNAN, L. & STEIN, L. D. 2005. The International HapMap Project Web site. *Genome Research,* 15**,** 1592-1593.

THYS, R. G. & WANG, Y. H. 2015. DNA Replication Dynamics of the GGGGCC Repeat of the C9orf72 Gene. *J Biol Chem,* 290**,** 28953-62.

TIWARI, V. & WILSON, D. M., 3RD 2019. DNA Damage and Associated DNA Repair Defects in Disease and Premature Aging. *Am J Hum Genet,* 105**,** 237-257.

TOWBIN, H., STAEHELIN, T. & GORDON, J. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the national academy of sciences,* 76**,** 4350-4354.

TSAI, C.-P., SOONG, B.-W., TU, P.-H., LIN, K.-P., FUH, J.-L., TSAI, P.-C., LU, Y.-C., LEE, I.-H. & LEE, Y.-C. 2012. A hexanucleotide repeat expansion in C9ORF72 causes familial and sporadic ALS in Taiwan. *Neurobiology of aging,* 33**,** 2232. e11-2232. e18.

TSAI, C. P., WANG, K. C., HWANG, C. S., LEE, I. T. & LEE, C. T. 2015. Incidence, prevalence, and medical expenditures of classical amyotrophic lateral sclerosis in Taiwan, 1999-2008. *J Formos Med Assoc,* 114**,** 612-9.

TSUJI, H., ARAI, T., KAMETANI, F., NONAKA, T., YAMASHITA, M., SUZUKAKE, M., HOSOKAWA, M., YOSHIDA, M., HATSUTA, H., TAKAO, M., SAITO, Y., MURAYAMA, S., AKIYAMA, H., HASEGAWA, M., MANN, D. M. & TAMAOKA, A. 2012. Molecular analysis and biochemical classification of TDP-43 proteinopathy. *Brain,* 135**,** 3380-91.

TURABELIDZE, G., ZHU, B.-P., SCHOOTMAN, M., MALONE, J. L., HOROWITZ, S., WEIDINGER, J., WILLIAMSON, D. & SIMOES, E. 2008. An epidemiologic investigation of amyotrophic lateral sclerosis in Jefferson County, Missouri, 1998–2002. *Neurotoxicology,* 29**,** 81-86.

TURNER, M. R., BARNWELL, J., AL-CHALABI, A. & EISEN, A. J. B. 2012. Young-onset amyotrophic lateral sclerosis: historical and other observations. 135**,** 2883-2891.

TYSNES, O. B., VOLLSET, S. E. & AARLI, J. A. 1991. Epidemiology of amyotrophic lateral sclerosis in Hordaland county, western Norway. *Acta Neurol Scand,* 83**,** 280-5.

TZIORTZOUDA, P., VAN DEN BOSCH, L. & HIRTH, F. 2021. Triad of TDP43 control in neurodegeneration: autoregulation, localization and aggregation. *Nat Rev Neurosci,* 22**,** 197-208.

UDDIN, M. S., AL MAMUN, A., ASADUZZAMAN, M., HOSN, F., ABU SUFIAN, M., TAKEDA, S., HERRERA-CALDERON, O., ABDEL-DAIM, M. M., UDDIN, G. M. S., NOOR, M. A. A., BEGUM, M. M., KABIR, M. T., ZAMAN, S., SARWAR, M. S., RAHMAN, M. M., RAFE, M. R., HOSSAIN, M. F., HOSSAIN, M. S., ASHRAFUL IQBAL, M. & SUJAN, M. A. R. 2018. Spectrum of Disease and Prescription Pattern for Outpatients with Neurological Disorders: An Empirical Pilot Study in Bangladesh. *Ann Neurosci,* 25**,** 25-37.

UDINE, E., JAIN, A. & VAN BLITTERSWIJK, M. 2023. Advances in sequencing technologies for amyotrophic lateral sclerosis research. *Molecular Neurodegeneration,* 18**,** 1-15.

UYAN, O., OMUR, O., AGIM, Z. S., OZOGUZ, A., LI, H., PARMAN, Y., DEYMEER, F., OFLAZER, P., KOC, F., TAN, E., OZCELIK, H. & BASAK, A. N. 2013. Genome-wide copy number variation in

sporadic amyotrophic lateral sclerosis in the Turkish population: deletion of EPHA3 is a possible protective factor. *PLoS One,* 8**,** e72381.

VALDMANIS, P. N. & ROULEAU, G. A. 2008. Genetics of familial amyotrophic lateral sclerosis. *Neurology,* 70**,** 144-52.

VALKO, K. & CIESLA, L. 2019. Amyotrophic lateral sclerosis. *Prog Med Chem,* 58**,** 63-117.

VAN BROECKHOVEN, C. 2010. The future of genetic research on neurodegeneration. *Nat Med,* 16**,** 1215-7.

VAN DAMME, P., DEWIL, M., ROBBERECHT, W. & VAN DEN BOSCH, L. 2005. Excitotoxicity and amyotrophic lateral sclerosis. *Neurodegener Dis,* 2**,** 147-59.

VAN DEERLIN, V. M., LEVERENZ, J. B., BEKRIS, L. M., BIRD, T. D., YUAN, W., ELMAN, L. B., CLAY, D., WOOD, E. M., CHEN-PLOTKIN, A. S., MARTINEZ-LAGE, M., STEINBART, E., MCCLUSKEY, L., GROSSMAN, M., NEUMANN, M., WU, I. L., YANG, W. S., KALB, R., GALASKO, D. R., MONTINE, T. J., TROJANOWSKI, J. Q., LEE, V. M., SCHELLENBERG, G. D. & YU, C. E. 2008. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *Lancet Neurol,* 7**,** 409-16.

VAN DEN BERG-VOS, R. M., VISSER, J., FRANSSEN, H., DE VISSER, M., DE JONG, J. M., KALMIJN, S., WOKKE, J. H. & VAN DEN BERG, L. H. 2003. Sporadic lower motor neuron disease with adult onset: classification of subtypes. *Brain,* 126**,** 1036-47.

VAN ES, M. A., HARDIMAN, O., CHIO, A., AL-CHALABI, A., PASTERKAMP, R. J., VELDINK, J. H. & VAN DEN BERG, L. H. 2017. Amyotrophic lateral sclerosis. *Lancet,* 390**,** 2084-2098.

VAN ES, M. A., VELDINK, J. H., SARIS, C. G., BLAUW, H. M., VAN VUGHT, P. W., BIRVE, A., LEMMENS, R., SCHELHAAS, H. J., GROEN, E. J. & HUISMAN, M. H. 2009. Genome-wide association study identifies 19p13. 3 (UNC13A) and 9p21. 2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nature genetics,* 41**,** 1083-1087.

VAN RHEENEN, W., PULIT, S. L., DEKKER, A. M., AL KHLEIFAT, A., BRANDS, W. J., IACOANGELI, A., KENNA, K. P., KAVAK, E., KOOYMAN, M., MCLAUGHLIN, R. L., MIDDELKOOP, B., MOISSE, M., SCHELLEVIS, R. D., SHATUNOV, A., SPROVIERO, W., TAZELAAR, G. H. P., VAN DER SPEK, R. A. A., VAN DOORMAAL, P. T. C., VAN EIJK, K. R., VAN VUGT, J., BASAK, A. N., BLAIR, I. P., GLASS, J. D., HARDIMAN, O., HIDE, W., LANDERS, J. E., MORA, J. S., MORRISON, K. E., NEWHOUSE, S., ROBBERECHT, W., SHAW, C. E., SHAW, P. J., VAN DAMME, P., VAN ES, M. A., WRAY, N. R., AL-CHALABI, A., VAN DEN BERG, L. H., VELDINK, J. H. & SEQUENCING, P. M. A. 2018. Project MinE: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *European Journal of Human Genetics,* 26**,** 1537-1546.

VAN RHEENEN, W., VAN DER SPEK, R. A. A., BAKKER, M. K., VAN VUGT, J., HOP, P. J., ZWAMBORN, R. A. J., DE KLEIN, N., WESTRA, H. J., BAKKER, O. B., DEELEN, P., SHIREBY, G., HANNON, E., MOISSE, M., BAIRD, D., RESTUADI, R., DOLZHENKO, E., DEKKER, A. M., GAWOR, K., WESTENENG, H. J., TAZELAAR, G. H. P., VAN EIJK, K. R., KOOYMAN, M., BYRNE, R. P., DOHERTY, M., HEVERIN, M., AL KHLEIFAT, A., IACOANGELI, A., SHATUNOV, A., TICOZZI, N., COOPER-KNOCK, J., SMITH, B. N., GROMICHO, M., CHANDRAN, S., PAL, S., MORRISON, K. E., SHAW, P. J., HARDY, J., ORRELL, R. W., SENDTNER, M., MEYER, T., BASAK, N., VAN DER KOOI,

A. J., RATTI, A., FOGH, I., GELLERA, C., LAURIA, G., CORTI, S., CEREDA, C., SPROVIERO, D., D'ALFONSO, S., SORARU, G., SICILIANO, G., FILOSTO, M., PADOVANI, A., CHIO, A., CALVO, A., MOGLIA, C., BRUNETTI, M., CANOSA, A., GRASSANO, M., BEGHI, E., PUPILLO, E., LOGROSCINO, G., NEFUSSY, B., OSMANOVIC, A., NORDIN, A., LERNER, Y., ZABARI, M., GOTKINE, M., BALOH, R. H., BELL, S., VOURC'H, P., CORCIA, P., COURATIER, P., MILLECAMPS, S., MEININGER, V., SALACHAS, F., MORA PARDINA, J. S., ASSIALIOUI, A., ROJAS-GARCIA, R., DION, P. A., ROSS, J. P., LUDOLPH, A. C., WEISHAUPT, J. H., BRENNER, D., FREISCHMIDT, A., BENSIMON, G., BRICE, A., DURR, A., PAYAN, C. A. M., SAKER-DELYE, S., WOOD, N. W., TOPP, S., RADEMAKERS, R., TITTMANN, L., LIEB, W., FRANKE, A., RIPKE, S., BRAUN, A., KRAFT, J., et al. 2021. Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Nat Genet,* 53**,** 1636-1648.

VANCE, C., ROGELJ, B., HORTOBÁGYI, T., DE VOS, K. J., NISHIMURA, A. L., SREEDHARAN, J., HU, X., SMITH, B., RUDDY, D. & WRIGHT, P. 2009. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science,* 323**,** 1208-1211.

VANDE VELDE, C., MCDONALD, K. K., BOUKHEDIMI, Y., MCALONIS-DOWNES, M., LOBSIGER, C. S., BEL HADJ, S., ZANDONA, A., JULIEN, J. P., SHAH, S. B. & CLEVELAND, D. W. 2011. Misfolded SOD1 associated with motor neuron mitochondria alters mitochondrial shape and distribution prior to clinical onset. *PLoS One,* 6**,** e22031.

VATS, A., GOURIE-DEVI, M., SUROLIYA, V., VERMA, S., FARUQ, M., SHARMA, A., GANGULY, N. K., KUKRETI, R., WAJID, S. & TANEJA, V. 2017. Analysis of C9orf72 repeat expansion in amyotrophic lateral sclerosis patients from North India. *J Neurol Sci,* 373**,** 55-57.

VAZQUEZ, M., KETZOIAN, C., LEGNANI, C., REGA, I., SÁNCHEZ, N., PERNA, A., PENELA, M., AGUIRREZÁBAL, X., DRUET-CABANAC, M. & MEDICI, M. 2008. Incidence and prevalence of amyotrophic lateral sclerosis in Uruguay: a population-based study. *Neuroepidemiology,* 30**,** 105-111.

VCV000014761.5 National Center for Biotechnology Information. ClinVar; [VCV000014761.5], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000014761.5 (accessed March 4, 2023).

VCV000016226.1 National Center for Biotechnology Information. ClinVar; [VCV000016226.1], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000016226.1 (accessed Sept. 5, 2022).

VCV000197145.13 National Center for Biotechnology Information. ClinVar; [VCV000197145.13], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000197145.13 (accessed Sept. 5, 2022).

VCV000197145.14 National Center for Biotechnology Information. ClinVar; [VCV000197145.14], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000197145.14 (accessed Jan. 11, 2023).

VCV000212640.25 National Center for Biotechnology Information. ClinVar; [VCV000212640.25], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000212640.25 (accessed March 6, 2023).

VCV000344328.5 National Center for Biotechnology Information. ClinVar; [VCV000344328.5], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000344328.5 (accessed March 6, 2023).

VCV000468253.3 National Center for Biotechnology Information. ClinVar; [VCV000468253.3], https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000468253.3 (accessed Jan. 23, 2023).

VELEBIT, J., HORVAT, A., SMOLIC, T., PRPAR MIHEVC, S., ROGELJ, B., ZOREC, R. & VARDJAN, N. 2020. Astrocytes with TDP-43 inclusions exhibit reduced noradrenergic cAMP and Ca(2+) signaling and dysregulated cell metabolism. *Sci Rep,* 10**,** 6003.

VERHEIJEN, B. M., VERMULST, M. & VAN LEEUWEN, F. W. 2018. Somatic mutations in neurons during aging and neurodegeneration. *Acta Neuropathol,* 135**,** 811-826.

VISSERS, L. E., DE LIGT, J., GILISSEN, C., JANSSEN, I., STEEHOUWER, M., DE VRIES, P., VAN LIER, B., ARTS, P., WIESKAMP, N., DEL ROSARIO, M., VAN BON, B. W., HOISCHEN, A., DE VRIES, B. B., BRUNNER, H. G. & VELTMAN, J. A. 2010. A de novo paradigm for mental retardation. *Nat Genet,* 42**,** 1109-12.

VUCIC, S., HIGASHIHARA, M., SOBUE, G., ATSUTA, N., DOI, Y., KUWABARA, S., KIM, S. H., KIM, I., OH, K.-W. & PARK, J. 2020. ALS is a multistep process in South Korean, Japanese, and Australian patients. *Neurology,* 94**,** e1657-e1663.

WANG, L., DENG, H. X., GRISOTTI, G., ZHAI, H., SIDDIQUE, T. & ROOS, R. P. 2009. Wild-type SOD1 overexpression accelerates disease onset of a G85R SOD1 mouse. *Hum Mol Genet,* 18**,** 1642-51.

WANG, L., GUTMANN, D. H. & ROOS, R. P. 2011. Astrocyte loss of mutant SOD1 delays ALS disease onset and progression in G85R transgenic mice. *Hum Mol Genet,* 20**,** 286-93.

WANG, S. J., WANG, K. Y. & WANG, W. C. 2004. Mechanisms underlying the riluzole inhibition of glutamate release from rat cerebral cortex nerve terminals (synaptosomes). *Neuroscience,* 125**,** 191-201.

*WEB*MD https://www.webmd.com/brain/who-gets-als#:~:text=Sex%3A%20About%2060%25%20of%20people%20with%20ALS%20are%20male.

WEI, Q., CHEN, X., CHEN, Y., OU, R., CAO, B., HOU, Y., ZHANG, L. & SHANG, H. F. 2019. Unique characteristics of the genetics epidemiology of amyotrophic lateral sclerosis in China. *Sci China Life Sci,* 62**,** 517-525.

WEIDBERG, H. & ELAZAR, Z. 2011. TBK1 mediates crosstalk between the innate immune response and autophagy. *Sci Signal,* 4**,** pe39.

WEISSMAN, I. L. & GAGE, F. H. 2016. A Mechanism for Somatic Brain Mosaicism. *Cell,* 164**,** 593-5.

WHO 2018. ICD-11 for mortality and morbidity statistics (2018).

WIECHELMAN, K. J., BRAUN, R. D. & FITZPATRICK, J. D. 1988. Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. *Anal Biochem,* 175**,** 231-7.

WONG, C., STAVROU, M., ELLIOTT, E., GREGORY, J. M., LEIGH, N., PINTO, A. A., WILLIAMS, T. L., CHATAWAY, J., SWINGLER, R., PARMAR, M. K. B., STALLARD, N., WEIR, C. J., PARKER, R. A., CHAOUCH, A., HAMDALLA, H., EALING, J., GORRIE, G., MORRISON, I., DUNCAN, C., CONNELLY, P., CAROD-ARTAL, F. J., DAVENPORT, R., REITBOECK, P. G., RADUNOVIC, A., SRINIVASAN, V., PRESTON, J., MEHTA, A. R., LEIGHTON, D., GLASMACHER, S., BESWICK, E., WILLIAMSON, J., STENSON, A., WEAVER, C., NEWTON, J., LYLE, D., DAKIN, R., MACLEOD, M., PAL, S. & CHANDRAN, S. 2021. Clinical trials in amyotrophic lateral sclerosis: a systematic review and perspective. *Brain Commun,* 3**,** fcab242.

WOULFE, J., KERTESZ, A. & MUNOZ, D. G. 2001. Frontotemporal dementia with ubiquitinated cytoplasmic and intranuclear inclusions. *Acta Neuropathol,* 102**,** 94-102.

WRITING, G. & EDARAVONE, A. L. S. S. G. 2017. Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet Neurol,* 16**,** 505-512.

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., JR. & LANDERS, J. E. 2012. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature,* 488**,** 499-503.

XIAO, S., SANELLI, T., CHIANG, H., SUN, Y., CHAKRABARTTY, A., KEITH, J., ROGAEVA, E., ZINMAN, L. & ROBERTSON, J. 2015. Low molecular weight species of TDP-43 generated by abnormal splicing form inclusions in amyotrophic lateral sclerosis and result in motor neuron death. *Acta Neuropathol,* 130**,** 49-61.

XU, B., ROOS, J. L., DEXHEIMER, P., BOONE, B., PLUMMER, B., LEVY, S., GOGOS, J. A. & KARAYIORGOU, M. 2011. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet,* 43**,** 864-8.

XU, L., LIU, T., LIU, L., YAO, X., CHEN, L., FAN, D., ZHAN, S. & WANG, S. 2020. Global variation in prevalence and incidence of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol,* 267**,** 944-953.

YANG, Y. M., GUPTA, S. K., KIM, K. J., POWERS, B. E., CERQUEIRA, A., WAINGER, B. J., NGO, H. D., ROSOWSKI, K. A., SCHEIN, P. A. & ACKEIFI, C. A. 2013. A small molecule screen in stem-cell-derived motor neurons identifies a kinase inhibitor as a candidate therapeutic for ALS. *Cell stem cell,* 12**,** 713-726.

YOUNG, P. E., KUM JEW, S., BUCKLAND, M. E., PAMPHLETT, R. & SUTER, C. M. 2017. Epigenetic differences between monozygotic twins discordant for amyotrophic lateral sclerosis (ALS) provide clues to disease pathogenesis. *PLoS One,* 12**,** e0182638.

ZHANG, C., YANG, Y., LIANG, W., WANG, T., WANG, S., WANG, X., WANG, Y., JIANG, H. & FENG, H. 2019. Neuroprotection by urate on the mutant hSOD1-related cellular and Drosophila models of amyotrophic lateral sclerosis: Implication for GSH synthesis via activating Akt/GSK3beta/Nrf2/GCLC pathways. *Brain Res Bull,* 146**,** 287-301.

ZHANG, S., COOPER-KNOCK, J., WEIMER, A. K., SHI, M., MOLL, T., MARSHALL, J. N. G., HARVEY, C., NEZHAD, H. G., FRANKLIN, J., SOUZA, C. D. S., NING, K., WANG, C., LI, J., DILLIOTT, A. A., FARHAN, S., ELHAIK, E., PASNICEANU, I., LIVESEY, M. R., EITAN, C., HORNSTEIN, E., KENNA, K. P., PROJECT MIN, E. A. L. S. S. C., VELDINK, J. H., FERRAIUOLO, L., SHAW, P. J. & SNYDER, M. P. 2022. Genome-wide identification of the genetic basis of amyotrophic lateral sclerosis. *Neuron,* 110**,** 992-1008 e11.

ZHAO, J., WANG, X., HUO, Z., CHEN, Y., LIU, J., ZHAO, Z., MENG, F., SU, Q., BAO, W., ZHANG, L., WEN, S., WANG, X., LIU, H. & ZHOU, S. 2022. The Impact of Mitochondrial Dysfunction in Amyotrophic Lateral Sclerosis. *Cells,* 11.

ZOU, Z. Y., CUI, L. Y., SUN, Q., LI, X. G., LIU, M. S., XU, Y., ZHOU, Y. & YANG, X. Z. 2013. De novo FUS gene mutations are associated with juvenile-onset sporadic amyotrophic lateral sclerosis in China. *Neurobiol Aging,* 34**,** 1312 e1-8.