The pathophysiological role of TDP-43 in amyotrophic lateral sclerosis due to C9orf72 mutations

Jakub Scaber

St Anne's College

University of Oxford

Supervised by Professor Kevin Talbot

Submitted for the degree of *Doctor of Philosophy*, Trinity Term 2017

To

Agnieszka, Emilia and Zosia

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative condition that affects corticospinal and spinal motor neurons and leads to death within 30 months of symptom onset in half of all cases. It remains incurable and treatment is supportive. The genetic and molecular understanding of ALS has gone through a rapid expansion in recent years, notably with the discoveries of TDP-43, a heterogeneous ribonucleoprotein as a major component of neuronal inclusions in ALS, as well as the discovery of the *C9orf72* hexanucleotide expansion as the most common genetic cause of this disease.

This first part of this thesis addresses the question of which of the various pathological hallmarks of the *C9orf72* Hexanucleotide Repeat Expansion (HRE) in autopsy material correlates best with the clinical presentation. The main finding is that TDP-43 distribution, rather than *C9orf72* RNA foci or dipeptide aggregation in the brain, corresponds best with the areas relevant to the clinical subtype of ALS-FTD. Subsequently the role of TDP-43 was investigated in induced pluripotent stem cell derived motor neurons, and no evidence of the hallmarks of TDP-43 dysfunction, were seen in this model of the disease. No mislocalisation is found on immunofluorescence, and biochemical analysis shows no differences in insoluble species between the patient and control cell lines. In the final section, RNA sequencing was used to study the transcriptome of a BAC transgenic mouse carrying a human M337V transgene expressed at low levels, to identify early presymptomatic differences in gene expression. Interestingly, no changes were found in genes known to be

associated with ALS through mutations, and the constitutive nuclear functions of TDP-43 in the regulation of splicing was maintained, prior to the emergence of a clinical phenotype in the mouse. This favours a gain of function mechanism for TDP-43 mutations in ALS.

Declaration

Except where acknowledgement is made, all the work presented in this thesis was performed by the author in the Nuffield Department of Clinical Neurosciences, University of Oxford. The work reported in this thesis has not been submitted for any other degrees in this or any other University of Institute of learning.

Acknowledgements

The research presented here and its findings are part of an amazing journey. I am hugely thankful to the people I was able to work with during this time and for the patients and funders who made it all possible.

Most of all I would like thank Professor Kevin Talbot, for being a wonderful mentor who was always hugely supportive in all possible ways. I will always remain inspired by his dedicated care for patients with ALS, his passion for research, his sensitive and effective leadership of the lab. I am proud to have had the chance to work and study under his guidance and will take away many wonderful memories and experiences, which have changed me as a person, doctor and scientist.

I would like to thank the various PIs I have had the privilege of working with. Professor Martin Turner, for his always positive outlook, clinical guidance and honest opinion and support. Professor Olaf Ansorge, for his inspiring enthusiasm, his many ideas and his time to discuss ALS neuropathology however busy he was. Drs Andreas Heger and David Sims, for taking me on at CGAT and making me feel at home amongst a peer group of budding leaders of computational biology, and teaching me everything I know about it.

I could not have done this work on my own. The possibility to learn from, work with and discuss science with extraordinary scientists and doctors in the Talbot lab, Turner lab, Oxford MND Centre, Neuropathology and CGAT, as

well as the wider NDCN and FGU communities has been a dream come true for me. I am grateful for the many friendships that have emerged from this.

Finally, I would like to thank the patients and the ALS community, for their time, their enthusiasm and their relentless support for this and other research. This work would not have been possible without them, and the many experiences as a doctor looking after patients with ALS have shaped my personal determination to continue fighting this terrible disease with them. I would like to thank the MND Association, the Lady Edith Wolfson Fellowship Fund and the Medical Research Council for generously supporting this work.

Publications arising from this work

Scaber J, Douglas AGL, Bäumer D, Hofer M, Turner MR, Talbot K, et al. Neuropathology and epigenetics of the ALS-FTD spectrum in C9orf72 repeat expansion carriers. Manuscript in preparation.

Aoki Y, Manzano R, Lee Y, Dafinca R, Aoki M, Douglas AGL, Varela MA, Sathyaprakash C, **Scaber J** et al. C9orf72 and RAB7L1 regulate vesicle trafficking in amyotrophic lateral sclerosis and frontotemporal dementia. Brain 2017; 140

Scaber J, Talbot K. What is the role of TDP-43 in C9orf72-related amyotrophic lateral sclerosis and frontemporal dementia? Brain 2016; 139: 3057–3059.

Dafinca R, **Scaber J**, Ababneh N, Lalic T, Weir G, Christian H, et al. C9orf72 Hexanucleotide Expansions Are Associated with Altered Endoplasmic Reticulum Calcium Homeostasis and Stress Granule Formation in Induced Pluripotent Stem Cell-Derived Neurons from Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Demen. Stem Cells 2016; 34: 2063–78.

Table of Contents

List of Abbreviations

ALS	Amyotrophic Lateral Sclerosis
BAC	Bacterial artificial chromososme
BDNF	Brain derived neurotrophic factor
BMP	Bone morphogenic protein
ChAT	Choline acetyl transferase
ddNTP	Dideoxynucleotide
DAPI	4′,6-Diamidino-2-Phenylindole
DNA	Deoxyribonucleic acid
dNTP	Dexoynucleotide
ER	Endoplasmic reticulum
ERCC	External RNA Controls Consortium
FGF	Fibroblast growth factor
FISH	Fluorescence *in situ* hybridisation
FTD	Frontotemporal dementia
GDNF	Glial-cell derived neurotrophic factor
GEF	GTPase exchange factor
GO	Gene ontology
GTF	Genome transfer file
GWAS	Genome wide association study
hnRNP	Heterogenous nuclear ribunucleoprotein
HRE	Hexanucleotide repeat expansion
iPSC	Induced pluripotent stem cell

iPSMN	Induced pluripotent stem cell derived motor neuron
KEGG	Kyoto Encyclopedia of Genes and Genomes
LMN	Lower motor neuron
MGI	Mouse Genome Informatics
NIV	Non-invasive ventilation
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PLS	Primary lateral sclerosis
PI	Propidium iodide
pTDP-43	Phosphorylated TDP-43
qPCR	Quantitative real-time polymerase chain reaction
qRT-PCR	Quantitative reverse transcription polymerase chain reaction
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RRM	RNA recognition motif
RT-PCR	Reverse transcription polymerase chain reaction
SSC	Sodium-saline citrate
SSH	Sonic hedgehog
TBS	Tri-buffered saline
UMI	Unique molecular identifier
UMN	Upper motor neuron
UTR	Untranslated region
WT	Wild-type

List of Figures

List of Tables

Chapter 1: Introduction

Amyotrophic Lateral Sclerosis (ALS) is neurodegenerative disease that predominantly affects motor neurons in the motor cortex and spinal cord. It remains one of the most devastating acquired neurological disorders, resulting in progressive paralysis and death within 1-5 years. There is still no cure for the condition. Clinically, ALS is on a spectrum with Frontotemporal Dementia, with which it shares TDP-43 positive pathological neuronal inclusions and considerable genetic overlap.

1.1 ALS is a neurodegenerative disease of the motor system

1.1.1 ALS is incurable

Epidemiological studies have consistently shown that ALS is not an uncommon disease. ALS has a relatively homogenous incidence rate of 1.81/100,000 person-years of follow up in Europe and North America. The lifetime risk of ALS in the United Kingdom is 1:400 for women and 1:350 for men (Johnston *et al.*, 2006). Incidence is lower in East and South Asia (Marin *et al.*, 2017). ALS has slightly higher incidence in men (2.03 per 100,000 person years) than women (1.45 per 100,000 person years). No differences have been observed in the demographic status or the disease progression between men and women. The highest incidence of ALS is observed between the ages of 58 and 63, and

falls sharply after the age of 80 (Logroscino *et al.*, 2010). Median survival is around 30 months from symptom onset and less than two years from diagnosis (Chio *et al.*, 2002; del Aguila *et al.*, 2003).

ALS remains incurable. Riluzole is the only disease modifying treatment available to date (Bensimon *et al.*, 1994; Cheah *et al.*, 2010). Its mechanism of action remains unclear, but is thought to modulate excitotoxicity through interaction with GABA receptors and calcium channels. Riluzole has a modest survival benefit of about two to three months, but offers no symptomatic relief or improvement in muscle strength (Miller *et al.*, 2007). Beyond riluzole, ALS treatment is principally symptomatic in a multidisciplinary care setting. Treatments can be offered to improve nutrition, respiration, salivation and spasticity amongst others. While non-invasive ventilation (NIV) for respiratory failure increases quality of life, evidence also exists for a survival benefit (Pinto *et al.*, 1995; Bourke *et al.*, 2006).

1.1.2 ALS is clinically heterogeneous

Neurodegeneration in ALS is predominantly seen in corticospinal tract neurons (upper motor neurons) as well as alpha motor neurons in the spinal cord innervating muscles (lower motor neurons). Clinically this gives rise to a syndrome of mixed upper and lower motor neuron signs. Ocular muscles are relatively spared. Within the spectrum of presentation, significant variation is observed.

1.1.2.1 Site of onset

ALS usually starts in one limb or the bulbar muscles and then subsequently spreads to other areas. Bulbar onset patients make up around 25% of all patients (Caroscio *et al.*, 1987; Roche *et al.*, 2012), with limb onset accounting for most of the remainder and respiratory onset occurring in less than 5% of patients (Pinto *et al.*, 1995). Both bulbar and respiratory onset have a significant impact on survival, compared to limb onset disease (del Aguila *et al.*, 2003; Chio *et al.*, 2009). Although differences exist between patients with limb or bulbar onset, there is also clearly significant variation within these subgroups, with significant numbers of limb onset cases following a very slowly progressive course, so-called flail-limb variants of ALS (Vucic and Kiernan, 2007; Wijesekera *et al.*, 2009; Turner *et al.*, 2010). A significant variation in survival has also been described in respiratory onset patients (Shoesmith *et al.*, 2007), and 20% of bulbar onset patients survive beyond 5 years.

1.1.2.2 Variation in upper versus lower motor neuron involvement

Other than the site of onset, clinically ALS can also be distinguished by the relative contribution of upper and lower motor neuron pathology. The differentiation of ALS by clinical subtype has been recognised early in the history of the disease, as has the fact that both upper motor neuron (UMN) as well as lower motor neuron (LMN) predominant ALS are both more slowly progressive and less common than the mixed type (Swank and Putnam, 1943).

The lower motor neuron predominant form of ALS, often known as progressive muscular atrophy, has been demonstrated to have subclinical corticospinal tract involvement at autopsy with pathology typical of ALS (Ince *et al.*, 2003). Ongoing uncertainty persists about the relationship of isolated UMN degeneration with survival measured in decades, known as primary lateral sclerosis (PLS), to ALS. Over time subclinical LMN abnormalities in nerve conduction studies are seen in some, but not all patients, arguing that it is part of the ALS spectrum (Le Forestier *et al.*, 2001). Others however have proposed that PLS could be a separate pathophysiological entity (Gordon *et al.*, 2006).

1.1.3 ALS is on a spectrum with frontotemporal dementia

Further heterogeneity can be detected in ALS on cognitive testing. As motor degeneration proceeds, half of all patients can be found to have some form of cognitive impairment on neuropsychological testing, especially affecting executive function. Around 15% of all ALS patients meet the formal diagnostic criteria for frontotemporal dementia (Ringholz *et al.*, 2005). Similarly, occurrence of ALS in patients initially diagnosed with FTD is well described (Lomen-Hoerth *et al.*, 2002). The clinical evidence for overlap is supported by the common pathological features of ALS with a proportion of FTD cases, both of which are characterised by neuronal TDP-43 positive inclusions (Neumann *et al.*, 2006). FTD-TDP accounts for about half of all FTD cases (the remainder being mostly positive for Tau inclusions). Finally, genetic evidence of overlap comes from family studies in which mixed inheritance of FTD and ALS syndromes (Morita *et al.*, 2006), which are mostly explained by the

hexanucleotide expansion in the gene *C9orf72* (DeJesus-Hernandez *et al.*, 2011; Renton *et al.*, 2011), but other genes such as *Ubiquilin2* have also been implicated in the ALS-FTD overlap (Deng *et al.*, 2011).

1.1.4 Familial ALS is caused by a heterogeneous set of genes and sporadic ALS has a high heritability

In ALS clinic populations at least 10% to 15% of patients report a family history of ALS or dementia. Prospective studies with active investigation of family trees have yielded higher estimates for family history in ALS patients, closer to 20% (Chiò *et al.*, 2011). Familial disease is similarly phenotypically heterogeneous and is not distinguishable from sporadic disease on clinical grounds (Li *et al.*, 1988; Andersen *et al.*, 1997). In FTD, the proportion of patients with familial disease is even higher, with 30-50% of patients reporting at least one relative with a similar presentation (Goldman *et al.*, 2005). Twin studies estimate the heritability of ALS at the population level as 61% (Al-Chalabi *et al.*, 2010), indicating that there is complex genetic inheritance of ALS beyond the known ALS genes. Only 21% of heritability of sporadic ALS is explained by common SNP variants, and only 1% of heritability is explained by loci identified by genome wide association studies (GWAS) to date (Keller *et al.*, 2014).

Over the last two decades, 32 genes have been linked to familial ALS-FTD (Guerreiro *et al.*, 2015), and these explain approximately two thirds all familial cases. There is ongoing debate about what genes are relevant for

typical ALS, and many of the genes associated with ALS, such as *alsin* and *sentaxin*, are thought to cause rare atypical forms. A significant proportion of the Mendelian inheritance of typical ALS-FTD is explained by mutations in five genes. In descending order of frequency, they are: *C9orf72* (ALS and FTD), *SOD1* (ALS), *GRN* (FTD), *FUS* (ALS), *TARDBP* (ALS). Because of small family sizes and reduced penetrance of some mutations, many ALS gene mutations are also found in up to an estimated 11% of apparently sporadic ALS patients (Renton *et al.*, 2014). Given that two ALS mutations are found together in ALS patients more commonly than expected by chance, some have argued that in some cases ALS is an oligogenic disease, with the requirement for two or more genes to be mutated for disease development (van Blitterswijk *et al.*, 2012). Indeed, there is epidemiological evidence to support an assertion of a multistep process occurring in ALS, similar to carcinogenesis (Al-Chalabi *et al.*, 2014). Others have argued that cooccurrence of two ALS mutations are chance findings confounded by founder effects or that one of the two variants may be a benign polymorphism (Chio *et al.*, 2012).

1.1.5 Familial ALS and FTD-TDP are caused by genetic mutations in genes involved in RNA metabolism and protein catabolism

The genes in which mutations are known to cause ALS-FTD fall into two broad categories, according to their known function: genes involved in RNA processing and genes involved in protein degradation and aggrephagy (Thomas *et al.*, 2013; Guerreiro *et al.*, 2015). These mutations collectively provide initial evidence for the dysregulation of these two crucial cellular

functions in ALS. ALS genes linked to RNA processing include the heterogeneous nuclear ribonucleoproteins (hnRNPs) TDP-43, hnRNPA1, hnRNPA2B1, as well as members of the FET protein family FUS and TAF15. Matrin 3 is another RNA binding protein that interacts with TDP-43 and has been found mutated in ALS. ALS genes involved in autophagy and lysosomal function include *TBK1*, *Sequestosome 1* (p62), *Ubiquilin 2*, *VCP* and *Optineurin*. Many of these genes are involved in cargo-specific autophagy of ubiquitinated proteins, and p62 is a major component of inclusions in sporadic and familial ALS (Thomas *et al.*, 2013).

1.2 TDP-43 pathology is the molecular hallmark of ALS and FTD-TDP

On a molecular level, multiple pathological signatures of ALS have been observed. The ubiquitin positive inclusions seen in motor neurons of all ALS patients can stain positively for TDP-43, SOD1 or FUS. However, the latter two only account for less than three percent of all cases. 97% of all ALS patients have TDP-43 inclusions in the anterior horn cells and the motor cortex, making TDP-43 inclusions the molecular hallmark of ALS (Neumann *et al.*, 2006; Mackenzie *et al.*, 2007).

1.2.1 TDP-43 protein structure

TDP is a 414-amino-acid protein encoded by the *TARDBP* gene. It is an heterogenous nuclear ribunucleoprotein (hnRNP) that has a predominantly

nuclear localisation but shuttles to the cytoplasm. It contains a folded N-terminal domain that promotes oligomerisation (Chang *et al.*, 2012; Mompean *et al.*, 2016) as well two tandem RNA recognition motifs (RRMs) motifs that bind to UG-rich sequences (Lukavsky *et al.*, 2013). The C-terminal region is glycine rich and predominantly disordered, and is commonly referred to as low-complexity domain and sometimes also as a prion-like domain. The domain is essential for protein-protein interactions (Buratti *et al.*, 2005). The significance of low-complexity domains in health and disease is only beginning to emerge. The domains permit self-assembly of ribonucleoproteins into membraneless organelles (Kato *et al.*, 2012) via liquid-liquid phase transition. Like other ribonucleoproteins, TDP-43 is able to undergo liquid-liquid phase separation mediated by its C-terminal domain (Conicella *et al.*, 2016).

1.2.2 TDP-43 Function

As a member of the hnRNP family TDP-43 is unsurprisingly involved in RNA processing, but also in several other cellular processes. TDP-43 is expressed ubiquitously and essential for embryogenesis (Wu *et al.*, 2010). TDP-43 binding targets have been identified in thousands of pre-mRNA species. Binding sites are found in introns, but also in 3' untranslated regions (UTRs). (Polymenidou *et al.*, 2011; Tollervey *et al.*, 2011).

1.2.2.1 Nuclear functions of TDP-43

The best described function of TDP-43 is its involvement in splicing. TDP-43 acts as a negative splicing regulator binding to UG sequences and has been

most extensively studied using the *CTFR* reporter gene (Buratti and Baralle, 2001; Wang *et al.*, 2004). Unbiased screens of differential splicing in mice show many other differential splicing events upon TDP-43 knockdown (Polymenidou *et al.*, 2011). There are many more TDP-43 binding sites than there are splicing alterations, indicating that TDP-43 binding may alter splicing but must also have a role in pre-mRNA stabilisation. TDP-43 is not an independent splicing factor and requires binding to other hnRNPs to perform its splicing function (Buratti *et al.*, 2005). Other than binding to and regulating a large number of transcripts across the genome, TDP-43 also autoregulates its own mRNA by inducing alternative polyadenylation, promoting the subsequent nuclear retention and degradation of this alternatively polyadenylated *TARDBP* mRNA (Avendaño-Vázquez *et al.*, 2012). A further nuclear role of TDP-43 is the regulation of microRNAs (Gregory *et al.*, 2004; Chendrimada *et al.*, 2005), and its absence leads to alteration in the levels of multiple microRNAs (Buratti *et al.*, 2010).

1.2.2.2 Extranuclear functions of TDP-43

Even at rest a significant proportion of TDP-43 can be found in the cytoplasm, where TDP-43 is involved in RNA transport. This is supported by its binding to the 3' UTRs of mature mRNAs as well as the fact that TDP-43 co-purifies with proteins involved in RNA transport (Freibaum *et al.*, 2010). This is particularly important in neurons where it is actively transported along axons and dendrites (Fallini *et al.*, 2012), and is a key component of transport granules (Liu-Yesucevitz *et al.*, 2014). Furthermore TDP-43 can be found in the synaptic

compartment , where it may have a role in local translation (Narayanan *et al.*, 2012). Finally, TDP-43 is known to move from the nucleus into the cytoplasm during activity and stress (Moisse *et al.*, 2009; Sato *et al.*, 2009). During stress, TDP-43 is a component of stress granules, which are transient membraneless structures that repress translation during stress (Colombrita *et al.*, 2009; McDonald *et al.*, 2011).

1.2.3 Evidence for pathogenicity of TDP-43

The discovery of TDP-43 in the cytoplasmic inclusions of ALS and FTD-TDP prompted a scientific debate on the relevance of the protein for the disease. There is mounting of evidence for the pathogenicity of TDP-43.

1.2.3.1 Genetics

Pathological mutations in TDP-43 have been found in 3% of familial and 1.5% of sporadic ALS cases, and their significance is supported by appropriate co-segregation (Sreedharan *et al.*, 2008; Van Deerlin *et al.*, 2008). Importantly, mutations are functionally clustered: of 52 mutations found to date, all but three localise to the C-terminal low complexity domain (Buratti, 2015).

1.2.3.2 Abnormalities of TDP-43 protein in ALS

TDP-43 shares important features with other proteins implicated in neurodegeneration, α-synuclein and tau. TDP-43 is mislocalised, aggregated, ubiquitinated, phosphorylated and C-terminally truncated (Neumann *et al.*, 2006). It remains unclear if these alterations in TDP-43 distribution and

properties are protective or if they contribute to ALS pathophysiology (Ling *et al.*, 2013).

1.2.3.2.1 Posttranslational modifications

In post-mortem ALS motor neurons, TDP-43 is tagged by polyubiquitinated chains that are K48 and K63-linked (Hebron *et al.*, 2013). Only larger inclusions are labelled by ubiquitin (Giordana *et al.*, 2010), suggesting that ubiquitination may be a late event during the maturation of inclusions. Phosphorylation, on the other hand, appears to predate ubiquitination (Kraemer *et al.*, 2010). Phosphorylation may modulate degradation and extends the half life of TDP-43 in experimental models (Zhang *et al.*, 2010).

1.2.3.2.2 TDP-43 fragmentation

In human post-mortem tissue a 25kDa TDP-43 C-terminal fragment is specific for ALS and not seen in controls (Neumann *et al.*, 2006). 35kDa fragments have also been found in patient non-central nervous system (CNS) tissues (Kabashi *et al.*, 2008), but are usually much less abundant in ALS brain. The role of 35kDa fragments is therefore not clear, but they are observed at significant levels in TDP-43 mutant mouse models of ALS (Wegorzewska and Baloh, 2011). Both 25kDa and 35kDa TDP-43 fragments have been shown to be products of caspase 3 cleavage (Zhang *et al.*, 2007), but recent evidence also suggests the possibility of alternative splicing to produce the 35kDa species (Xiao, Sanelli, *et al.*, 2015). C-terminal fragments lack the nuclear localisation motifs found close to the N-terminus of TDP-43 and are therefore

thought to localise to the cytoplasm. Interestingly enrichment of C-terminal TDP-43 fragments is not seen uniformly throughout the central nervous system. C-terminal fragments are seen particularly in brain-derived TDP-43 inclusions but less in spinal cord inclusions, suggesting different compositions of inclusions in brain and spinal cord (Igaz *et al.*, 2008). Interestingly, in FTD-TDP there are multiple C-terminal fraction signatures which correlate closely with neuropathological subtypes of FTD-TDP and the associated clinical syndromes (Tsuji *et al.*, 2012).

1.2.3.2.3 Mislocalisation

Mislocalisation of TDP-43 from the nucleus to the cytoplasm is a key pathological hallmark of ALS (Neumann *et al.*, 2006). In neuropathological tissue, a concept of pre-inclusions has been postulated; in cells in which these small granular pre-aggregates can be observed, there is already characteristic loss of nuclear TDP-43 (Giordana *et al.*, 2010). Consequently, increased TDP-43 in the cytoplasm as well as depletion of TDP-43 from the nucleus have been hypothesised to play a role in the pathophysiology of the disease. There is evidence from animal models that both loss of TDP-43, produced by knockdown (Polymenidou *et al.*, 2011), or excess TDP-43 produced by overexpression (Kabashi *et al.*, 2009) are toxic to the cell. Overexpression of TDP-43 with a mutant nuclear localisation signal resulting in cellular redistribution is also toxic (Winton *et al.*, 2008). While most mutant TDP-43 animal models do have mislocalisation, it is not clear that this is essential for pathogenesis: the Q331K and M337V mutations in mice result in a motor

phenotype without any evidence of TDP-43 mislocalisation or aggregation (Arnold *et al.*, 2013; D'Alton *et al.*, 2014).

1.2.3.2.4 Aggregation

Together with mislocalisation, TDP-43 aggregation in the cytoplasm is a consistent neuropathological feature of ALS (Mackenzie *et al.*, 2007). Truncated, phosphorylated and ubiquitinated TDP-43 is mainly found in the sarkosyl insoluble fraction. There is ongoing debate as to whether the sequestration of TDP-43 is toxic or neuroprotective, with no correlation between TDP-43 inclusion burden and cell death in several mouse models (Stallings *et al.*, 2010; Igaz *et al.*, 2011; Arnold *et al.*, 2013). In vitro studies suggest that some TDP-43 mutations in the low complexity domain increase aggregation, while other mutations that cause severe disease in humans do not (Jiang *et al.*, 2016). Other mechanisms such as liquid-liquid phase transition may be equally important in explaining the pathogenicity of TDP-43 mutations (Conicella *et al.*, 2016). However, not only the C-terminal domain has been implicated in aggregations and some studies suggest that aggregation may be property of the N-terminal domain (Sasaguri *et al.*, 2016).

1.2.3.3 Disruption of TDP-43 function in ALS

There is increasing evidence for disruption of the mRNA splicing function of TDP-43 function in ALS. TDP-43 levels in a cell are tightly regulated (Avendaño-Vázquez *et al.*, 2012), and small differences in cytoplasmic and nuclear concentrations may have a significant impact on TDP-43 function

(Arnold *et al.*, 2013). High throughput technologies have elucidated the binding sites of TDP-43 and the effect of TDP-43 depletion on splicing in a mouse model (Polymenidou *et al.*, 2011; Tollervey *et al.*, 2011). In post-mortem tissue a shift in the pattern of alternative splicing in sporadic and familial ALS brains has also been observed, though in the absence of TDP-43 pathology in the areas studied (Prudencio *et al.*, 2015). Further evidence of RNA processing derangement in human post mortem tissues comes from the expression of human specific cryptic exons in ALS but not healthy controls (Ling *et al.*, 2015). A detailed discussion of TDP-43 splicing alterations in ALS will be discussed below in the context of a review of the impact of high throughput sequencing technologies on ALS.

TDP-43 is known to be recruited to stress granules which stabilise and triage RNA in response to oxidative or other stressors (Colombrita *et al.*, 2009). Other than the constitutive nuclear function in splicing there is some evidence for disruption in stress granule formation in ALS. While some studies have detected putative pathological stress granules in post-mortem ALS tissue (Liu-Yesucevitz *et al.*, 2010), others have not (Colombrita *et al.*, 2009).

Finally, studies have suggested alterations in the binding profile of TDP-43 with other proteins could contribute to TDP-43 pathology. Some TDP-43 mutants have been shown to have elevated binding to FUS (Ling *et al.*, 2010), potentially providing a further mechanism linking the two proteins, though the relevance to sporadic disease is unclear.

1.3 Mutant TDP-43 mouse models increase our understanding of TDP-43 biology

The best known mouse model of ALS is the *SOD1* mouse model with the G93A mutation, which develops progressive hindlimb tremor and weakness at three months, resulting in paralysis and death after four months (Turner and Talbot, 2008). While providing important insights into familial ALS with *SOD1* mutations, models of *SOD1* ALS suffer from two major shortcomings. Firstly, *SOD1* mouse models do not show the typical TDP-43 positive inclusions found in 97% of all patients with ALS at post-mortem, suggesting that SOD1-related ALS may be biologically distinct from most other cases (Turner *et al.*, 2008). Secondly, therapies successful in *SOD1* mutant mouse models of the disease, have not generally translated into positive human trials (Benatar, 2007).

The discovery of TDP-43 in 2006 therefore encouraged the development of new mouse models based on mutations of TDP-43 to study TDP-43 biology and in the hope that they could provide more relevant insights to sporadic ALS. While a comprehensive review of TDP-43 mouse models is not possible here, the most significant findings and limitations of currently available TDP-43 rodent models are discussed below. Given the increasing complexity of ALS genetics, mouse models based on other ALS-related mutations have also been generated, including mice overexpressing transgenic FUS and the *C9orf72* HRE and will not be discussed further.

1.3.1 Neuron specific depletion of TDP-43 recapitulates an age-dependent motor phenotype

It is known that TDP-43 is essential in embryogenesis. TDP-43 knockout in mice is lethal at E7.5 (Kraemer *et al.*, 2010). Conditional Cre-dependent ubiquitous knockdown of TDP-43 is similarly toxic, leading to death within 9 days of TDP-43 knockdown (Chiang *et al.*, 2010), underlining the importance of TDP-43 to cellular function. Longer survival was observed in two neuronal specific conditional knock-out mouse models. The first of these was generated by using flanking loxP sites driven by a motor neuron specific Hb9 promoter. These mice had a normal motor phenotype up to 10 weeks (Wu *et al.*, 2012) and had weight loss at birth. Another, less aggressive conditional knock out model was generated using an Frt-flanked neomycin cassette and loxP sites flanking the second exon of *TARDBP*. This model was crossed to VAChT-Cre.Fast mice to achieve motor neuron specificity and remained symptom-free for 50 weeks (Iguchi *et al.*, 2013). Both models develop an age-dependent progressive motor phenotype, subsequently with loss of motor neurons and astrocytosis and offer an opportunity to study the TDP-43 loss of function hypothesis in a motor neuron specific manner.

1.3.2 Ubiquitous overexpression of wild-type and mutant TDP-43 results in motor system degeneration in a dose dependent fashion

Mouse models of both mutant and wild-type TDP-43 overexpression recapitulate features of ALS. The most prominent TDP-43 mutant mouse model is the prion promoter driven *A315T* mouse model that overexpresses

human mutant TDP-43 three-fold compared to the endogenous mouse TDP-43 (Wegorzewska and Baloh, 2011). The mouse develops motor problems at three months, though the cause of death in this mouse is due to paralytic ileus caused by the degeneration of the myenteric plexus. Feeding with jellified food overcomes the gut dysfunction and allows motor pathology and motor neuron degeneration to progress further (Herdewyn *et al.*, 2014). As this mouse model lacks a wild-type control, it is impossible to ascertain if the pathology arises from transgene insertion, TDP-43 overexpression or biological disruption due to the A315T mutation itself. Another A315T mouse model with expression driven by the prion promoter did include a wild-type overexpressing control which does not exhibit the mortality found in mutants (Stallings *et al.*, 2010). Unfortunately, lower expression of TDP-43 in the wild-type mouse than in the mutant again prohibits a conclusive appraisal of the effect of the TDP-43 mutation in isolation.

The importance of a wild-type TDP-43 dose effect was demonstrated in a model in which the human wild-type protein overexpression is driven by a prion promoter. TDP-43PrP mice showed reactive gliosis, axonal and myelin degeneration, gait abnormalities, and early lethality (Xu *et al.*, 2010). Another transgenic mouse model expresses human TDP-43 with a Q331K mutation under the prion promoter at levels 1.5- to 3-fold above endogenous mouse TDP-43 levels. This mouse model develops adult onset motor weakness, but does not have a survival deficit (Arnold *et al.*, 2013). The mouse models generated in this study exhibited a complex relationship between transgene

dose and genotype with some pathology being dose-dependent and some mutation-dependent.

To avoid overexpression of TDP-43 via a heterologous promoter such as PrP, another study resorted to multiple insertion of human native-promoter driven bacterial artificial chromosome (BAC) A315T or WT transgenes into the mouse genome. These mice are also characterised by significant overexpression of TDP-43, and both wild-type and mutant A315T mice share clinical motor and cognitive abnormalities. Most significantly, TDP-43 overexpression induces the NF-κB proinflammatory pathway through direct interaction of TDP-43 with NF-κB, resulting in death via terminal gliosis rather than neurodegeneration (Swarup *et al.*, 2011). Interestingly, the motor neuron transcriptome of this mouse model has been studied using translating ribosome affinity purification followed by microarray analysis revealing a small number of differentially expressed genes despite the severe phenotype (Macnair *et al.*, 2016).

1.3.3 Tissue-specific overexpression provides opportunities to study autonomous and non-cell autonomous toxicity of TDP-43

To overcome some of the off-target effects of overexpression of TDP-43 as well as to study cell-autonomous and non-cell autonomous pathways in ALS, several mouse models overexpressing TDP-43 in neurons or glia have been studied. Neuron specific overexpression has been achieved using the Thy1.2 promoter overexpressing wild-type TDP-43 with a range of overexpression relative to mouse TDP-43 and establishing a dose-dependent motor phenotype

(Wils *et al.*, 2010). The same group also used similar constructs to overexpress M337V mutant TDP-43, showing a modest worsening of phenotype with the mutation. Interestingly, end-stage TDP-43^{M337V} mice had fewer ubiquitinated inclusions and fewer 25-kDa C terminal fragments compared to wild-type overexpressing mice (Janssens *et al.*, 2013). Another study using the Thy1.2 promoter resulted in 2.3 to 4.6-fold overexpression of TDP-43. Although the mice develop motor axon loss, they die prematurely of an unknown cause (Shan *et al.*, 2010). Several mouse models have also been created using the CaMK2 promoter, which overexpresses TDP-43 in frontal cortex and hippocampus, but not the spinal cord. Consequently, these mice develop a cognitive but not motor phenotype and thus are more relevant to modelling frontotemporal dementia (Tsai *et al.*, 2010).

To study non-cell autonomous toxicity, human wild-type TDP-43 has also been overexpressed selectively in rat and mouse astrocytes. While one induced astrocytic TDP-43 overexpression model was found to have astrocyte loss and motor system symptoms (Tong *et al.*, 2013), a transplantation study of TDP-43 overexpressing glial precursor cells found no pathology (Haidet-Phillips *et al.*, 2013).

1.3.4 Limitations of currently available mouse models

The TDP-43 rodent models used up till now have yielded important insights into disease biology. They have demonstrated a dose dependent toxicity of TDP-43, whereby too much or too little TDP-43 can cause cellular imbalances.

Motor neurons appear to be sensitive to these fluctuations, however the relevance of reduction or overexpression of TDP-43 as a primary driver of the pathogenesis of ALS is completely unclear. In this context, the severe overexpression and young onset phenotype in many of these models makes them a resource with limited usefulness for translational work. In some of the mouse models, there is mutation specific-toxicity, but this is often difficult to disentangle from toxicity due to overexpression, and more physiologically relevant models will be able to provide more information on mutation-specific toxicity. There is a fundamental question if a disease that arises after decades in humans can be truly modelled in an organism with a much shorter lifespan. Nonetheless, neurodegenerative disease modelling should at least consider the pathways involved in aging and maturation of the model. Mouse models with very early onset of symptoms during adolescence are therefore conceivably suboptimal for disease modelling.

1.4 The *C9orf72* HRE is the most common genetic cause of ALS and FTD-TDP

The discovery of the *C9orf72* hexanucleotide repeat expansion as the most common genetic cause for ALS and FTD (DeJesus-Hernandez *et al.*, 2011; Renton *et al.*, 2011), opened an important new chapter in ALS research.

The linkage to chromosome 9 was originally reported in ALS cohorts from Boston and Chicago (Hosler *et al.*, 2000). This was further delineated to chromosome 9q 12-21 in a large Dutch kindred autosomal dominant

inheritance of ALS-FTD (Vance *et al.*, 2006). Genome-wide association studies narrowed the region to 240 kb (Laaksovirta *et al.*, 2010; Shatunov *et al.*, 2010), but a further year passed before the hexanucleotide expansion was found: the repeat is located in an intron and repeat sequences could not be characterised by short read sequencing.

1.4.1 Epidemiology and clinical characteristics

C9orf72 is inherited in an autosomal dominant manner with likely incomplete penetrance (Galimberti *et al.*, 2014). In European and North American populations, approximately 40% of familial cases and 10% of sporadic cases of ALS have been found to have a *C9orf72* HRE (Majounie, Renton, *et al.*, 2012; Smith *et al.*, 2013; Cooper-Knock, Kirby, *et al.*, 2015). There is wide geographic variation of the mutation amongst familial ALS cases. In European populations studies have reported frequencies ranging from 11 to 55%, whereas in Asian populations between 0 and 16% of familial ALS cases have been linked to the *C9orf72* HRE (Zou *et al.*, 2017). A founder effect with the mutation possibly arising from a Scandinavian population has been proposed given the geographical distribution of *C9orf72* (Majounie, Renton, *et al.*, 2012; Smith *et al.*, 2013).

The most common clinical presentations associated with the mutation are ALS and behavioural variant FTD, or an overlap between the two (Boeve *et al.*, 2012). Compared to other FTD patients, patients with *C9orf72* are more likely to have delusions and hallucinations, which have been descried to occur in a

third to half of all *C9orf72* patients and may sometimes lead to psychiatric diagnoses in the early stages of the disease (Rohrer, Isaacs, *et al.*, 2015). *C9orf72* ALS patients are otherwise generally indistinguishable from other ALS patients clinically and have similar frequencies of limb and bulbar onset disease to sporadic patients. On a population level, the most obvious difference is the presence of cognitive impairment or a family history of dementia in up to 50% of patients (Chiò *et al.*, 2012; Cooper-Knock *et al.*, 2012). Extrapyramidal symptoms have been noted in *C9orf72* patients in multiple studies (Boeve *et al.*, 2012; Cooper-Knock *et al.*, 2012), but the expansion has not been identified in large Parkinson's disease cohorts, indicating that the clinical presentation does not overlap (Majounie, Abramzon, *et al.*, 2012; Harms *et al.*, 2013). C9orf72 patients also show a distinct pattern of non-motor cortex involvement on 3T magnetic resonance imaging (MRI) and may have more aggressive disease as evidenced by shorter disease duration from symptom onset in some studies (Byrne *et al.*, 2012; Cooper-Knock *et al.*, 2012).

1.4.2 Genetic and molecular properties of the *C9orf72* HRE

The C9orf72 hexanucleotide expansion is found in the first intron of the *C9orf72* gene and is flanked by the two alternative exons 1a and 1b. The *C9orf72* has three known transcript variants. Variant 1 (V1, NM_145005.6) and Variant 3 (V3, NM_001256054.2) both contain exon 1a and therefore their pre-mRNA contains the HRE. Variant 2 (V2, NM_018325.4) contains exon 1b and has the HRE in its promoter region. Both V2 and V3 encode the same protein isoform a and differ only in their 5'UTR. V1 encodes protein

isoform b which is considerably shorter ('Database resources of the National Center for Biotechnology Information', 2016). An antisense transcript, which contains the antisense intronic sequence and hexanucleotide repeat sequence, can also be detected (Gendron *et al.*, 2013). The site of the HRE normally contains less than 30 GGGGCC repeats with most normal individuals carrying 2, 5 or 8 repeats (van der Zee *et al.*, 2013). While no controls have been documented with more than 100 repeats, there is ongoing debate about the intermediate region ranging from 20-100 repeats as incomplete co-segregation of familial FTD has even been seen in cases with 20-22 repeats (Gómez-Tortosa *et al.*, 2013).

Given its intronic location, the presence of the HRE in *C9orf72* does not result in *C9orf72* protein sequence abnormalities. Instead the HRE results in three distinct molecular phenomena, all of which have been postulated to have pathophysiological significance:

1. Hapolinsufficiency: the presence of the repeat sequence in the promoter region of the most abundant transcript V2 results in overall reductions in *C9orf72* mRNA and protein levels (Xiao, MacNair, *et al.*, 2015).
2. RNA aggregation: the transcription of the HRE in exon 1a containing transcripts as well as antisense transcription results in the formation of RNA sense and antisense foci which can be detected using fluorescence in situ hybridisation (Gendron *et al.*, 2013).
3. Repeat associated non-ATG translation: nuclear export of sense and antisense repeat containing mRNA results in the initiation of

unconventional translation in all three reading frames, resulting in the production of five species of repetitive dipeptide proteins, which can be detected in the TDP-43 negative, p62 positive inclusions characteristic of *C9orf72* HRE ALS-FTD (Gendron *et al.*, 2013).

1.4.3 Neuropathological features of *C9orf72* positive ALS-FTD

The insights into the molecular biology of the *C9orf72* HRE gained through the analysis of brain tissue have been critical to the current understanding of the disease and its characterisation within the wider context of sporadic ALS. *C9orf72* HRE ALS and FTD are positive for typical TDP-43 inclusions seen in ALS and FTD-TDP, but also contain TDP-43 negative, p-62 positive inclusions which stain for dipeptide proteins.

1.4.3.1 Sense and antisense foci are present throughout the CNS

1.4.3.1.1 Sense and antisense foci do not correlate with clinical phenotype at post-mortem

The *C9orf72* hexanucleotide expansion gives rise to of both sense (DeJesus-Hernandez et al., 2011; Renton et al., 2011) as well as antisense (Gendron et al., 2013) RNA foci. The distribution of foci has been studied extensively, and foci are present in throughout the CNS as well as non-CNS tissues including white blood cells and fibroblasts (Lagier-Tourenne *et al.*, 2013). In the CNS, RNA sense and antisense foci can be found in neurons, but also in oligodendrocytes and astrocytes, albeit at lower frequencies (Mizielinska *et al.*,

2013). There are important differences in the subcellular distribution of sense versus antisense RNA foci. Sense foci are more commonly found in the cytoplasm than antisense foci (Mizielinska *et al.*, 2013), and antisense foci have been shown to co-localise with nucleolin and nuclear speckle components (Cooper-Knock, Higginbottom, *et al.*, 2015). A perinucleolar distribution of RNA foci was also observed in a recent study of two *C9orf72* frontotemporal dementia cases, the significance of this remains unclear (Vatsavayai *et al.*, 2016). Regional differences between sense and antisense foci distributions have been studied to a limited extent and antisense foci are more frequently seen in motor neurons and Purkinje neurons in the cerebellum whereas sense foci are more frequent in cerebellar granule cells (Cooper-Knock, Higginbottom, *et al.*, 2015). Finally, multiple studies have also suggested that antisense foci may be more numerous than sense foci (Lagier-Tourenne *et al.*, 2013; Mizielinska *et al.*, 2013; Cooper-Knock, Higginbottom, *et al.*, 2015). Foci burden has rarely been linked to clinical phenotype in post-mortem studies. One study suggested that foci burden in the frontal cortex significantly inversely correlated with age at onset (Mizielinska *et al.*, 2013). A similar trend, though not significant, was observed for sense as well as antisense foci in both the frontal cortex and hippocampus.

1.4.3.1.2 Sense and antisense foci interact with similar RNA binding proteins

The presence of foci prompted important questions about their involvement in pathogenesis via RNA mediated gain-of-function. Sequestration of the splicing factor muscleblind by repeat RNAs is known to result in splicing dysfunction

and disease pathology in myotonic dystrophy (Konieczny *et al.*, 2014). In *C9orf72* carriers, pulldown experiments using hexanucleotide repeat RNA have revealed multiple RNA binding proteins. Some of these results have been validated in post-mortem studies. Sense foci were shown to colocalise with ADARB2 (Donnelly *et al.*, 2013), nucleolin (Haeusler *et al.*, 2014), Pur-alpha (Sareen *et al.*, 2013), SRSF2, hnRNP H1/F, hnRNP A1, and ALYREF (Cooper-Knock *et al.*, 2014). Interestingly, despite their different confirmation, antisense foci shared the same pulldown signature and colocalisation as sense foci in one study, colocalising with SRSF2, hnRNP H1/F, hnRNPA1 and ALYREF (Cooper-Knock, Higginbottom, *et al.*, 2015).

1.4.3.2 Distribution of dipeptide protein pathology

RNA foci undergo repeat associated non-ATG translation, giving rise to five dipeptide species: three in each direction but including one overlapping dipeptide produced by both sense and antisense strands (poly-GP). The frequency of dipeptide repeats and their concentration in brain tissue has now been assessed systematically by multiple groups.

1.4.3.2.1 Relative abundances of dipeptide species are conserved across all areas of the CNS

The concentrations and distributions of dipeptide proteins vary considerably across dipeptide species. GA dipeptide has been shown the most abundant protein using measures of inclusion density as well as biochemical methods such as dot blot analysis. While there are differences in overall dipeptide

abundance across tissues, GA is relatively the most common dipeptide in all areas of the CNS, followed by poly-GP, poly-GR and only rare poly-PR and poly-PA inclusions (Gendron *et al.*, 2013; Islam *et al.*, 2014; Gomez-Deza *et al.*, 2015; Mackenzie *et al.*, 2015; Davidson *et al.*, 2016). This could be explained by different translational efficiency across the various reading frames (Wojciechowska *et al.*, 2014) or potentially different solubility (Kwon *et al.*, 2014).

1.4.3.2.2 Dipeptide proteins are nearly absent from lower motor neurons

Dipeptide inclusions are found in significant quantities in all neocortical areas as well as the hippocampus. They are also found in similarly high concentrations in areas not affected in ALS or FTD such as the occipital cortex and the cerebellum (Gendron *et al.*, 2013; MacKenzie *et al.*, 2013; Mackenzie *et al.*, 2015). Lower motor neurons, a collective term used for the motor nuclei of the brainstem as well as anterior horn cells, have been found to contain no or little dipeptide pathology. Little dipeptide is seen in pure *C9orf72* ALS cases with typical TDP-43 positive spinal cord pathology (Gomez-Deza *et al.*, 2015), as well as in patients with ALS-FTD and FTD (Mackenzie *et al.*, 2015; Davidson *et al.*, 2016), leading some authors to question whether they are involved in lower motor neuron pathogenesis.

1.4.3.2.3 There are no reproducible differences in dipeptide protein distribution between ALS and FTD patients

Dipeptide distribution patters and abundances do not differ between clinical phenotype in *C9orf72* HRE carriers in the majority of studies (MacKenzie *et al.*, 2013; Gomez-Deza *et al.*, 2015; Mackenzie *et al.*, 2015; Schludi *et al.*, 2015). A few exceptions have been reported in single studies. GA dipeptide inclusions were found more commonly in the frontal cortex in *C9orf72* positive FTD patients than ALS patients and correlated with age at onset (Mackenzie *et al.*, 2015), whereas another study showed an higher GA dipeptide load in the cerebellum of FTD versus ALS patients (Schludi *et al.*, 2015). The same study also reported an increased frequency of GR-dipeptide inclusions in the CA3/4 regions of the hippocampus, but these inclusions remained rare. Finally, one study suggested that reduced GP-dipeptide burden is more common in FTD patients and correlates with cognitive dysfunction (Gendron *et al.*, 2015).

1.4.3.3 TDP-43 correlates best with clinical presentation

C9orf72 HRE positive ALS-FTD is characterised by typical TDP-43 pathology also seen in sporadic ALS and FTD (Murray *et al.*, 2011; Hsiung *et al.*, 2012). The most common subtype of TDP-43 pathology in FTD and ALS-FTD patients is type B pathology (Simón-Sánchez *et al.*, 2012; Stewart *et al.*, 2012), which is characterised by moderate numbers of neuronal cytoplasmic inclusions throughout all cortical layers but very few dystrophic neurites (Mackenzie *et al.*, 2011). *C9orf72* HRE carriers with pure FTD can also have more heterogeneous TDP-43 pathology with type A pathology, which is

characterised by abundant dystrophic neurites and the presence of neuronal intranuclear inclusions. Type A pathology has been reported on its own in some cases (Murray *et al.*, 2011) but also has been shown to coexist with type B pathology (MacKenzie *et al.*, 2013). There is now substantial evidence that the burden and distribution of TDP-43 pathology correlates very well with neurodegeneration (MacKenzie *et al.*, 2013) as well as clinical phenotype (Mackenzie *et al.*, 2015; Davidson *et al.*, 2016) in the substantial majority of cases.

1.4.3.4 RNA foci and dipeptide protein pathology precedes TDP-43 pathology

1.4.3.4.1 Evidence for temporal antecedence of C9orf72-specific pathology to TDP-43 pathology

Evidence for the temporal sequence of the pathologic changes in *C9orf72* positive cases comes from rare neuropathological studies where premorbid tissue has been available. The most compelling evidence for premorbid accumulation of foci and dipeptide proteins comes from a patient who suffered from temporal lobe epilepsy and underwent temporal lobe resection at the age of 61 years. At the time of resection only minor language problems were evident on neuropsychological testing. The patient developed typical frontotemporal dementia six years later (Vatsavayai *et al.*, 2016). Material from the resection demonstrated abundant RNA foci and dipeptide proteins but no TDP-43 pathology. The post-mortem on the other hand showed a typical TDP-43 positive dementia as well as established dipeptide pathology

and foci. Another insight into early *C9orf72* ALS came from a premature death of a young *C9orf72* carrier with apparently stable learning difficulties, who developed a rapidly progressive neurodegenerative condition in his twenties. He had dipeptide pathology but no TDP-43 pathology at post-mortem (Proudfoot *et al.*, 2014).

1.4.3.4.2 An emerging phenotype of TDP-43 negative C9orf72 dementia?

There have now been multiple reports of cognitive impairment and progressive frontotemporal dementia in patients with the *C9orf72* HRE expansion who have very little or no TDP-pathology. This was first reported in three FTD cases (Gijselinck *et al.*, 2012). Another group gave similar reports of TDP-43 negative FTD patients and noted no or mild atrophy of the brain, which is uncharacteristic of FTD (Baborie *et al.*, 2015). A very well characterised case study of a *C9orf72* HRE positive patient with a typical history of FTD but not TDP-43 pathology reported RNA foci and dipeptide pathology throughout the brain and noted atrophy of the median pulvinar thalamic nucleus and the subgenual anterior cingulate cortex (Vatsavayai *et al.*, 2016). This particular pattern of atrophy without marked frontotemporal atrophy has previously been reported in an imaging study in a subset of *C9orf72* positive FTD patients (Lee *et al.*, 2014), with similar imaging findings early in presymptomatic subjects (Rohrer, Nicholas, *et al.*, 2015). While these changes provide interesting insights into *C9orf72* HRE biology, it is important to note that they represent a small subset of cases. It is noteworthy that while there are four independent reports of TDP-43 negative FTD, no such cases have been

reported in ALS, indicating that dipeptide toxicity could differentially affect the two main diseases caused by the *C9orf72* HRE.

1.4.4 Reduced *C9orf72* expression and the loss of function hypothesis

Multiple studies have now established the downregulation of *C9orf72* transcripts in HRE carriers in frontal cortex (DeJesus-Hernandez *et al.*, 2011; Gijselinck *et al.*, 2012; Waite *et al.*, 2014), cerebellum (van Blitterswijk *et al.*, 2015), motor cortex and spinal cord (Donnelly *et al.*, 2013; Haeusler *et al.*, 2014). Transcript variant 2, the only transcript containing exon 1b, has been identified as the main transcript in *C9orf72* HRE carriers (E. Y. Liu *et al.*, 2014), and as the transcript most consistently downregulated. Few clinical associations have been found that correlate haploinsufficiency with clinical outcomes, but variant 1 expression in frontal cortex and cerebellum were found to correlate with longer survival (van Blitterswijk *et al.*, 2015).

Given that RNA forms a template for dipeptide production, multiple studies have found an increase of potentially truncated intron 1a containing transcripts (Mori *et al.*, 2013; van Blitterswijk *et al.*, 2015). Using nuclear cytoplasmic fractionation one group has now demonstrated the cytoplasmic presence of mature mRNA with a retained intron 1a containing the expansion in its 5' UTR (Niblock *et al.*, 2016).

There have been few studies looking at *C9orf72* protein expression, and until recently no consistent antibody had been available. The *C9orf72* long isoform

was found to be decreased in frontal cortex but not cerebellum (Waite *et al.*, 2014). The results of this initial antibody study were subsequently confirmed with a novel long isoform antibody, showing downregulation of C9orf72 protein in the temporal and frontal lobe, but not in the cerebellum, motor cortex and lumbar spine (Xiao, MacNair, *et al.*, 2015). The same study revealed a significantly higher short isoform levels in *C9orf72* HRE carriers compared to controls.

1.4.5 Methylation

The downregulation of the exon 1b containing transcript raises questions about the mechanism of such downregulation, especially given the proximity of the repeat to the promoter region of this transcript.

1.4.5.1 An inverse relationship between methylation and expression is observed in other repeat disorders

Reduced transcript expression due to a DNA repeat sequence had previously been shown in other repeat disorders, in which hypermethylation induced by repeats has been proposed as the mechanism of transcriptional repression. This is observed in Friedreich ataxia, where hypermethylation upstream of the GAA repeat leads to decreased *FXN* transcript levels, which have been shown to correlate with disease severity (Greene *et al.*, 2007; Al-Mahdawi *et al.*, 2008). Similarly, methylation reduces expression in Fragile-X syndrome, but it has been observed that there is no simple relationship between the degree of methylation and either the level of expression or the severity of the clinical

phenotype (Bell *et al.*, 1991; Sutcliffe *et al.*, 1992). Repeat disorders can also cause hypermethylation in genetic promoters downstream of the disease gene, as is the case in myotonic dystrophy (Castel *et al.*, 2011).

1.4.5.2 DNA Methylation levels are variable at the CpG island but consistent at the repeat *C9orf72* HRE expansion carriers

Studies of *C9orf72* promoter methylation at the 5' end of the repeat expansion have consistently shown a high degree of variability in *C9orf72* HRE carriers. In studies of blood, DNA methylation of the CpG island upstream of the HRE only 21% to 41% of cases studied had methylation of four or more CpG sites (Xi *et al.*, 2013, 2014), the remainder having low or absent methylation. This heterogeneity was subsequently replicated in neuropathological studies from *C9orf72* HRE carriers. The subset of *C9orf72* HRE carriers with high methylation ranged from 10% in the cerebellum in a study using direct bisulphite sequencing (Belzil *et al.*, 2014) to 40% in cerebellum and frontal cortex in a study assessed by methylation specific restriction digestion (E. Y. Liu *et al.*, 2014). *C9orf72* promoter methylation in these studies was confined to *C9orf72* HRE carriers and not found in controls and sporadic ALS patients. Methylation of the repeat itself was also studied using restriction digestion and a methylation sensitive repeat primed polymerase chain reaction (PCR) assay. Repeat methylation was consistently seen in all but two of 71 blood samples and all 38 brain samples from four different cortical regions. No CCCCGG methylation was observed in blood from healthy controls or sporadic ALS patients.

1.4.5.3 DNA Methylation may be acquired during embryonal development and is conserved across generations and across tissues

An important observation regarding *C9orf72* promoter methylation came from a study employing methylation sensitive restriction digestion method that showed consistent methylation levels across 21 different CNS and non-CNS tissues in three patients, two with absent and one with high methylation levels. The same study confirmed that methylation levels were heritable and siblings as well as offspring were likely to have similar methylation levels to their relatives (Russ *et al.*, 2015). Another study provided evidence higher levels of methylation in offspring of patients and this correlated with shorter survival in the second generation (Gijselinck *et al.*, 2015).

Evidence for low methylation levels during embryonal human development came from human embryonal tissue obtained during preimplantation screening (Cohen-Hadad *et al.*, 2016). Importantly, the haploidentical fibroblasts from the patient donor also had absent methylation. Intriguingly, induced pluripotent stem cells (iPSCs) derived from these fibroblasts had high levels of methylation, in stark contrast to previous reports that DNA methylation may turn off during iPSC reprogramming and then re-emerge after neuronal differentiation (Esanov *et al.*, 2016).

1.4.5.4 Complex relationship between DNA methylation and clinicopathological characteristics of the *C9orf72* HRE

A complex picture emerges from current studies of DNA methylation in *C9orf72* carriers thus far. In the few studies conducted thus far, a relationship

between high methylation and low *C9orf72* expression (Xi *et al.*, 2013; Gijselinck *et al.*, 2015) has been suggested. Two studies have also reported reduced foci and fewer dipeptide protein inclusions in patients with C9orf72 inclusions (E. Y. Liu *et al.*, 2014; Bauer, 2016). In two contradictory studies, higher *C9orf72* CpG island methylation has been associated with both longer (Gijselinck *et al.*, 2015) as well as shorter HRE lengths (Russ *et al.*, 2015). The clinical significance of *C9orf72* promoter methylation in HRE carriers was studied by two groups, who demonstrated that *C9orf72* promoter methylation was correlated with slower disease progression in imaging studies (McMillan *et al.*, 2015) and associated with longer disease duration and later age at death in neuropathological studies (Russ *et al.*, 2015).

1.4.5.5 Other epigenetic events could explain reduced *C9orf72* gene expression

While DNA methylation has been the focus of most epigenetic studies in C9orf72, other epigenetic methods may be equally important in controlling gene expression. A study of histone methylation in *C9orf72* carriers using chromatin immunoprecipitation confirmed trimethylation of histones in ten *C9orf72* carriers but not in 18 controls (Belzil *et al.*, 2013). The same findings were replicated in fibroblasts from *C9orf72* patients. Treatment of fibroblasts with a demethylation agent restored *C9orf72* expression levels to those of healthy participants.

1.4.5.6 Global DNA methylation changes have been reported in neurodegenerative diseases including ALS

With the advent and greater accessibility of genomic techniques, it has been possible to use sequencing technologies and array-based methods for the investigation of global changes in methylation in the CNS. Previous studies in other neurodegenerative diseases have found small but significant changes in the differential methylation in neurodegeneration (Bakulski et al., 2012; De Jager et al., 2014). One such study has been carried out in sALS cases, which showed changes in methylation of genes that on pathway analysis were involved in calcium homeostasis (Morahan et al., 2009).

1.4.6 *C9orf72* function is beginning to be elucidated

Although the *C9orf72* HRE expansion is in a noncoding region of the *C9orf72* gene and no abnormal protein is produced, the reduction in exon 1b containing transcripts has led some to argue even a small decrease in transcript levels could result in a functional detriment that could become significant with age.

The function of C9orf72 has been difficult to elucudate given the lack of availability of reliable antibodies to the protein. Early *in silico* studies suggested the presence of a DENN domain in C9orf72, implying that it may act as a GTPase exchange factor (Marat *et al.*, 2011; Levine *et al.*, 2013). This hypothesis has been tested in various overexpression models, in which C9orf72 has been postulated to interact or colocalise with Rab1, Rab5, Rab7 and Rab11

(Farg *et al.*, 2014), Rab1a (Webster *et al.*, 2016) as well as Rab8a and Rab39b (Sellier *et al.*, 2016). The reason for the different interaction results is unclear, and could be explained by different antibodies and different *C9orf72* constructs and tags. The interaction with Rab proteins indicates a role for C9orf72 in autophagy and membrane trafficking. Mice lacking the C9orf72 mouse orthologue have been studied and are characterised by progressive splenomegaly and lymphadenopathy but no neuromuscular defects. The mice have abnormal lysosomal function in macrophages and microglia consistent with its postulated function as a Rab GTPase exchange factor (GEF) (ORourke *et al.*, 2016). The previously mentioned study of human C9orf72 protein isoforms demonstrated important differences between the two: The long isoform showed diffuse cytoplasmic staining, whereas the short isoform labelled the nuclear membrane. There was apparent relocalisation to the plasma membrane in sporadic ALS, but this result has not been replicated to date (Xiao, MacNair, *et al.*, 2015).

1.5 *C9orf72* HRE disease can be modelled using iPS-derived motor neurons

The discovery of transcription-factor based reprogramming of somatic cells into stem-cell like pluripotent cells (Takahashi and Yamanaka, 2006; Takahashi *et al.*, 2007), was a ground-breaking development in biomedical research. Induced pluripotent stem cells (iPSCs) can be differentiated into almost any tissue type and allow the study of disease specific mutations with

the genetic background *in situ*. The opportunity to study disease-specific human neurons in culture without the restrictions surrounding human embryonal cells, opened a wealth of opportunities in the field of neuroscience and particularly within the field of ALS research.

1.5.1 Induced Pluripotent Stem Cells from ALS patients can be successfully differentiated into motor neurons

1.5.1.1 Developmental biology provides the framework for motor neuron differentiation

The scientific background required for the differentiation of motor neurons from pluripotent stem cells was provided by decades of research in developmental biology into motor neuron differentiation in amphibian, fish, chick and murine embryos.

1.5.1.1.1 Neural Induction

In the developing embryo, the dorsal mesoderm and its midline product, the notochord, function as an organiser that is the source of signals resulting in neural induction of the overlying ectoderm. This is in part achieved through antagonising bone morphogenic proteins (BMPs) (Harland, 2000). There is ongoing debate about the role of this mesodermal organiser tissue (Klingensmith *et al.*, 1999), and other mechanisms for neural tube formation that have been proposed include neuralisation as a direct consequence of early dorsoventral patterning. Dorsoventral patterning in the embryo results in

stabilisation of beta-catenin levels induced by Wnt signalling, which in turn are sufficient for the formation of neural tissue (Baker *et al.*, 1999).

1.5.1.1.2 Rostrocaudal patterning of the nervous system

One of the key events in the specification of nervous system cell fate is the distinction between forebrain, midbrain, hindbrain and spinal cord. This occurs within 6 hours of gastrulation in the chick embryo, and is coordinated through fibroblast growth factors (FGFs), which acts indirectly via its effect on paraxial mesoderm, as well as retinoid signalling (Muhr *et al.*, 1999). Multiple molecules are present in the neural tube with a rostrocaudal gradient and are thought to be involved in rostrocaudal fate specification in the spinal cord. They include retinoic acid (Durston *et al.*, 1989), FGFs and growth differentiation factor 11 (Liu *et al.*, 2001), all of which result in the induction of segment-specific Hox expression in the developing spinal cord (Bel-Vialar *et al.*, 2002). The expression of specific Hox genes, such as Hox6 and Hox10 in brachial and lumbar segments result in the generation of limb-innervating motor neurons. This process is orchestrated by the transcription factor FOXP1 (Dasen *et al.*, 2008), which is in turn repressed in thoracic segments under the control of other Hox transcription factors (Jung *et al.*, 2010).

1.5.1.1.3 Dorsoventral patterning in the spinal cord

Upon formation of the neural tube from neurectoderm, dorsoventral patterning of the spinal cord commences. This is driven predominantly through the polarising activity of the notochord and the floorplate of the neural tube

(Yamada *et al.*, 1991), mediated predominantly through Sonic Hedgehog (SHH) signalling. Sonic Hedgehog signalling induces a molecular cascade that prevents the degradation of GLI-family proteins, which is counteracted by GLI repressor signals from the roof plate. The thus established GLI gradient forms the molecular basis for expression of various homeobox domain proteins that determine dorsoventral identity, which in the ventral spinal cord are NK6 homeobox 1 and 2 (NKX6.1/2), as well as Class-II NK2 homeobox 2 (NKX2.2). Ultimately dorsoventral patterning results in the generation of five pools of neuronal cells with a specific identity (V0-V3 interneurons and motor neurons) (Alaynick *et al.*, 2011). Spinal motor neurons arise from a specific population of ventral cells that expresses NKX6.1 and PAX6, as well as the helix-loop-helix protein OLIG2 (Tanabe *et al.*, 1998) which in turn activates neurogenin 2 (NEUROG2) signalling. Interestingly, acquisition of motor neuron identity requires dorsoventral retinoic acid signalling, which is independent of the earlier signalling seen in the specification of rostrocaudal fate (Novitch *et al.*, 2003). In a complex process, which involves interplay between SHH signalling and retinoic acid signalling, the generation of post-mitotic motor neurons is achieved, through activation of motor neuron and pancreas homeobox 1 (MNX1, also known as HB9). Importantly, MNX1/HB9 activates its own expression and thus makes motor neuron identity independent of morphogen signalling (Tanabe *et al.*, 1998). Interestingly, there is debate about the ongoing maintenance of spinal motor neuron fate, and the ability of developing motor neurons to fate-switch with V2 interneurons following MNX1/HB9 activation (Arber *et al.*, 1999). In prospective motor neurons the

final step in the commitment to motor neuron fate is achieved by the action of Islet1 transcription factor that is induced by SHH and results in direct activation of cholinergic signalling (Yamada *et al.*, 1991).

1.5.1.2 Motor neuron differentiation protocols mimic motor neuron development

ALS affects both corticospinal tract motor neurons, also known as "upper motor neurons" as well as anterior horn cells, known as "lower motor neurons." While there are cortical motor neuron protocols, that can generate large numbers of cortical projection neurons (Shi *et al.*, 2012), protocols that can generate corticospinal tract neurons specifically are lacking and the task is more complex and less well understood from a developmental perspective. Several protocols have now been proposed for the generation of alpha motor neurons, but given the multitude of subclasses of motor neurons that can be studied, a consensus for the generation of motor neurons has yet to emerge. The protocols are comprehensively summarised elsewhere (Sances *et al.*, 2016). All protocols, however, follow the pattern set out by normal embryology, which can be divided into neural induction, motor neuron specification and neural maturation.

1.5.1.2.1 Neural Induction

Early embryologic studies have shown that neural fate is the default developmental programme (Jessell and Sanes, 2000). Neural induction can be enhanced and accelerated by inhibition of alternative pathways, which was

demonstrated by dual SMAD inhibition by LDN-193189 and SB-431542 (Chambers *et al.*, 2009). The same study also investigated the effect of seeding density on the differentiation process, with low seeding densities also giving rise to melanocytes and neural crest cells as well as CNS cells. While some protocols use SB-431542 and LDN-193189 for SMAD inhibition, others have shown that compound C, also known as dorsomorphin, has similar success rates for neural differentiation. Compound C also causes SMAD inhibition through inhibition multiple BMP receptors but also through independent inhibition of activin A signalling (Zhou *et al.*, 2010).

1.5.1.2.2 Motor neuron specification

Motor neuron specification is achieved through caudalisation by retinoic acid and ventralisation by recombinant sonic hedgehog or small molecule sonic hedgehog agonists. Further insight into motor neuron specification came from a study that demonstrated the importance of Wnt pathway induction by CHIR-99021. In a key study, concentration dependent addition of CHIR-99021 was found to promote the acquisition of a hindbrain identity and induced *Hox* gene expression (Maury *et al.*, 2014). The same study also emphasised the importance of even small differences of the timing of the addition of factors for motor neuron specification. Interestingly, *in vitro*, motor neurons possess a predominantly cervical identity by expression of *Hoxc4* through *Hoxc8*, but not lower limb spinal cord segments (Amoroso *et al.*, 2013). Expression of lower limb innervating motor neurons from iPSCs has not been achieved to date.

1.5.1.2.3 Maturation

Maturation of induced pluripotent stem-cell derived motor neurons (iPSMNs) *in vitro* has conventionally been achieved by prolonged incubation in neural support media including growth factors (Hu and Zhang, 2009; Qu *et al.*, 2014). Acceleration of motor neuron maturation can be accelerated by inhibition of the notch pathway by γ-secretase inhibitors DAPT and compound E (Crawford and Roelink, 2007; Borghese *et al.*, 2010), though the differences in motor neuron subtypes arising from induced maturation have only begun to be explored (Maury *et al.*, 2014). Assessment of motor neuron maturation *in vitro* has been difficult because of lack of reliable motor neuron markers. As described above, the markers Islet 1 and MNX1/HB9 are key transcription factors that are involved in motor neuron fate specification. However, these markers have been shown to inactivate during maturation (Thaler *et al.*, 2004; Qu *et al.*, 2014), possibly leading to underestimation of actual motor neuron counts in culture. Another marker of motor neurons is the heavy chain neurofilament SMI32, which is relatively specific for motor neurons within the spinal cord, but cannot reliably distinguish between motor neurons and other large cortical neurons (Sternberger and Sternberger, 1983). The most specific marker for motor neurons in the spinal cord is choline acetyl transferase (ChAT), which is also expressed by other cholinergic neurons in the brainstem and cortex. ChAT is particularly useful as it is only expressed in more mature motor neurons (Sances *et al.*, 2016). Another way to assess the maturity of motor neurons is through transcriptomic analysis. Single cell sequencing of

iPSC-derived neurons from a cortical neuron differentiation protocol has shown that they resemble foetal rather than adult cortical neurons, irrespective of duration of culture (Handel *et al.*, 2016). Similarly, a gene expression chip analysis of iPSMNs, showed similar results for motor neurons with iPSMNs having a similar but distinct expression profile from early foetal motor neurons around day 50 (Ho *et al.*, 2016). A final and important way to assess motor neuron maturation functionally is through electrophysiological studies. Patch clamping experiments have repeatedly confirmed that iPSMNs are functional and can fire trains of action potentials (Kiskinis *et al.*, 2014; Devlin *et al.*, 2015). Electrophysiological properties of iPSMNs have also been studied successfully using calcium imaging (Dafinca *et al.*, 2016).

1.5.2 iPS-derived motor neurons from *C9orf72* HRE carriers recapitulate features of *C9orf72* pathology *in vitro*

The discovery of the hexanucleotide expansion as a major cause for familial and sporadic motor neuron disease coincided with the development of iPSMN differentiation protocols and widespread adoption of iPSMNs to model human nervous system disease. The first attempts at iPSMN modelling resulted the successful generation of iPSMNs as well as glial cells from an SOD1 patient (Dimos *et al.*, 2008).

1.5.2.1 iPSMNs from *C9orf72* HRE carriers have molecular features of the hexanucleotide expansion

iPSMNs derived from hexanucleotide expansion carriers share many of the molecular features found in neuropathological samples of the mutation, providing validation for the iPSMN model of the disease. Similar to post-mortem studies, *C9orf72* gene expression has also been found to be downregulated in iPSMNs from *C9orf72* patients. Some have found that all tested isoforms of *C9orf72* are downregulated (Donnelly *et al.*, 2013), whereas others have shown downregulation only in variant 2, albeit in cortical neurons (Almeida *et al.*, 2013). Both sense and antisense RNA foci are observed in iPSC-derived motor neurons (Sareen *et al.*, 2013; Zhang *et al.*, 2015; Dafinca *et al.*, 2016), but are also seen in iPSC-derived cortical neurons (Almeida *et al.*, 2013), iPSCs (Almeida *et al.*, 2013) as well as fibroblasts (Almeida *et al.*, 2013; Lagier-Tourenne *et al.*, 2013; Cooper-Knock *et al.*, 2014). For repeat associated dipeptides, there has been a more mixed picture. Some groups have reported dipeptide protein inclusions in iPS-derived motor neurons (Donnelly *et al.*, 2013; Westergard *et al.*, 2016), whereas others did not detect aggregated dipeptides but did detect increased dipeptide levels using dot blot analysis (Almeida *et al.*, 2013; Dafinca *et al.*, 2016).

Finally, an interesting scientific debate surrounds the methylation status of iPSMNs. As previously discussed, post-mortem studies have shown that *C9orf72* promoter DNA methylation is seen at high levels in a subset of cases only. In iPSMNs, one study showed that fibroblast methylation is lost during reprogramming and then regained during neural differentiation (Esanov *et al.*,

2016). Another study however found that while human embryonic stem cells from *C9orf72* HRE positive embryos obtained from preimplantation genetic screening were hypomethylated, the haploidentical iPSCs derived from maternal fibroblasts were hypermethylated at the *C9orf72* promoter, contradicting the previous results (Cohen-Hadad *et al.*, 2016). Whether these discrepancies between these studies are explained by dynamic changes or differences in reprogramming remains to be shown.

1.5.2.2 Evidence for a neurodegeneration phenotype in *C9orf72* HRE positive iPSMNs

While pathophysiological changes that directly relate to the hexanucleotide repeat expansion have been repeatedly found in iPSMNs, protein aggregation, a key feature of neurodegenerative disease, has been more challenging to model in this system. As discussed above, while most groups agree on the presence of dipeptide repeat proteins in iPSMNs derived from *C9orf72* HRE carriers, only some demonstrated the presence of dipeptide aggregates. Similarly, p62, a protein that targets cargoes for autophagy and is known to stain TDP-43 as well as dipeptide positive inclusions (MacKenzie *et al.*, 2013), has been shown to be elevated in *C9orf72* HRE positive motor neurons (Dafinca *et al.*, 2016) as well as cortical neurons (Almeida *et al.*, 2013). There is contradictory evidence on whether p62 aggregates in iPSMNs, with studies reporting an absence of p62 inclusions (Sareen *et al.*, 2013), while others demonstrated the presence of p62 inclusions in cortical neurons, but not in motor neurons (Dafinca *et al.*, 2016). Interestingly, no typical TDP-43

pathology has been observed in iPSMNs to date (Donnelly *et al.*, 2013). However, defects in nucleocytoplasmic transport due to the *C9orf72* HRE have been postulated to influence TDP-43 reimport (Khosravi *et al.*, 2016), and TDP-43 nuclear cytoplasmic ratios have been found to be altered in *C9orf72* positive iPSMNs (Zhang *et al.*, 2015). No differences in subcellular TDP-43 distribution were observed in cortical neurons (Almeida *et al.*, 2013). There appears to be an increased vulnerability of *C9orf72* HRE positive motor neurons *in vitro.* In one study differences in propidium iodide staining and apoptotic activity was observed in normal culture (Dafinca *et al.*, 2016), supported by increased caspase 3 activation. Other studies did not observe such changes at basal conditions, but observed increased susceptibility to glutamate (Donnelly *et al.*, 2013) and tunicamycin (Haeusler *et al.*, 2014).

1.5.3 Summary: Using iPSMNs to model ALS

Given the ability to study human neuronal cells that carry both the known mutations as well as the unknown genetic heritability of ALS, the iPSMN model provides important benefits over existing models of the disease. Important limitations include the inherent variability of differentiations as well as the variability of different iPSC clones. Such variability can be managed using novel gene editing technologies such as CRISPR/Cas9 permitting the study of the individual contribution of the mutation to the phenotype (Merkle and Eggan, 2013).

1.6 RNA Sequencing reveals abnormalities in RNA metabolism in ALS

The introduction of high-throughput sequencing has greatly advanced the rate of scientific discoveries in numerous fields of biology. The field of amyotrophic lateral sclerosis is no exception. This section will discuss the emergence of sequencing and analysis methods and provide an overview of high-throughput sequencing methodology in ALS research with a focus on RNA sequencing experiments to date.

1.6.1 Historical evolution of sequencing technology and comparison of second generation sequencing methods

1.6.1.1 History of first-generation sequencing

The discovery the molecular structure of nucleic acid (Watson and Crick, 1953) initiated a number of efforts to determine base sequence relying on 2 dimensional separation and restriction digestion. These laborious techniques that involved considerable analytical chemistry methods led to the sequencing of the first ribonucleic acid sequence (Holley *et al.*, 1965) and the MS-2 phage genome (Fiers *et al.*, 1976). Development of methods to separate nucleic acids by size using electrophoresis culminated in the conception of chain-termination method also known as Sanger sequencing (Sanger and Nicklen, 1977). Sanger sequencing relies on the addition of dideoxyribonucleotides (ddNTPs), which serve as chain terminating molecules in a DNA polymerase reaction. The absence of the 3'-hydroxyl group on ddNTPs means that the

molecules can be added to a growing DNA strand, but no further nucleotides can be added. The earliest implementation of the method relies on four separate polymerase reactions per sample. Each reaction contains a primer, DNA polymerase as well as four molecules of deoxyribonucleotides (dNTPs). One species of chain terminating ddNTPs (ddATP, ddCTP, ddGTP, ddTTP), is added to each reaction in smaller proportions and randomly inhibits the reaction at multiple sites of the occurrence of the particular base, due to competition with dNTPs. The resulting mixtures are then run on an acrylamide gel.

The addition of fluorescent terminators was an important advance that allowed the addition of all four ddNTPs to a single reaction (Smith *et al.*, 1985), allowing for automated detection by a fluorescent laser at the end of the gel. The development of Sanger sequencing culminated in the inception of capillary sequencing (Swerdlow *et al.*, 1990). The acrylamide gel was replaced by a thin cuvette containing gel with a laser detector at the end of the capillary. The small diameter of the capillary allowed for the application of much greater electric fields and faster separation of the bands. The technology was developed in subsequent years to handle many samples simultaneously enabling the sequencing of the human genome many years ahead of plan (Lander *et al.*, 2001). Capillary sequencing was also used to sequence millions of 300-700 base pair long partial clones of complementary DNA that provided a cheaper alternative to whole genome methods as well as information about exon composition of transcripts. Finally, capillary sequencing also encouraged

the emergence of methods that relied on using the overlap between reads for mapping instead of the laborious mapping by cloning – this shotgun approach to genome sequencing enabled the successful completion of an alternative genome project in a greatly reduced timeframe (Venter, 2001) and laid the foundations for second generation sequencing.

1.6.1.2 Second generation sequencing

The development of second generation sequencing technology began with the conception of pyrosequencing, which measures pyrophosphate, a by-product of dNTP incorporation, by making it a substrate for luciferase (Nyrén and Lundin, 1985). The detection of pyrophosphate synthesis in real time formed the conceptual framework for the first successful high throughput sequencing – 454 Sequencing, which was rapidly followed by other methods. All second generation sequencing methods rely on the parallelisation of sequencing by shredding the input into random short fragments of DNA and therefore rely on significant computational power to assemble the genome from overlaps or to align reads against a known reference genome. Because longer repetitive sequences cannot be aligned or mapped accurately, for higher organisms the method relies on the presence of a scaffold often provided by hierarchical cloning experiments in the first place. There are now multiple second generation sequencing methods, which will not be discussed here in detail. Broadly, the conceptual underpinnings of the techniques follow a similar course that can be divided into library preparation and sequencing.

1.6.1.2.1 Library preparation

Library preparation involves the creation of a sequencing library from DNA or RNA. It may contain the following subroutines:

- **Selection.** Firstly, steps can be undertaken to remove unwanted molecules from the experimental sample. In the case of DNA sequencing, the use of exome sequencing allows for the selection of the subset of expressed genes in the genome using prefabricated baits or primers, resulting in much lower sequencing costs. In the case of RNA sequencing, there is an abundance of ribosomal RNA and transfer RNA that can make up 95% of cellular RNA species and would overwhelm an expression study targeted at protein coding genes. Two principal methods exist to enrich for mRNA. Selection of poly-adenylated RNA selects for mature mRNAs while depletion of ribosomal RNA from the total RNA pool will also leave immature intron containing mRNAs and non-polyadenylated long noncoding RNAs in the sample (Sultan *et al.*, 2014).
- **Addition of reference sample (optional).** Reference samples can be added for a number of quantitative experiments, and the External RNA Controls Consortium (ERCC) spike-ins are most widely used reference samples for RNA sequencing experiments (Baker *et al.*, 2005). More recently, more sophisticated methods that allow assessment of differential splicing have been developed (Hardwick *et al.*, 2016).

- **Fragmentation.** High-throughput sequencing technology relies on the massive parallelisation of sequencing at the expense of read length. Fragmentation of the genome into random fragments of 100-1000 bases is therefore an important step usually performed using time-limited heating of the sample.
- **Reverse transcription (RNA only).** First and second strand cDNA translation is necessary for RNA sequencing, followed by end repair.
- **Adapter ligation.** A variety of adapters, also known as barcodes are now available. Once ligated onto each of the DNA fragments they provide a way of identifying the origin of each DNA fragments that can be recognised by the downstream sequencing technology. Other adapter molecules that have become prominent are unique molecular identifiers (UMIs), which enable the distinction between true duplication and PCR duplication, which is increasingly important in single cell sequencing and foot-printing methods (Islam *et al.*, 2014; Smith *et al.*, 2016).
- **PCR amplification.** Selective enrichment of DNA fragments containing adaptor molecules is currently required during many library preparation protocols. Following this step, libraries destined for multiplexing are normalised and pooled.

1.6.1.2.2 Sequencing

Several second generation sequencing methods have emerged that can sequence the fragmented samples generated for most library preparation. The

common property of these methods is the ability to detect signal from millions of short reads in real time. This is achieved by various methods, of which Illumina Sequencing has dominated the market**.** This sequencing method uses flow cells that capture the adapter ends of the DNA fragments. Each captured DNA fragment is then amplified locally via a method called bridge amplification, creating a cluster of molecules of the same sequence identity, greatly amplifying the signal coming from each DNA fragment. This is followed by primer addition and sequencing by cyclic reversible termination. All four fluorescently labelled nucleotides are provided in each reaction, but per DNA template only one is incorporated due to a blocked 3' hydroxyl group. The flow cell is then imaged using multiple fluorescent light sources and bases incorporated at each location are called and given quality scores based on the purity of the colour signal. The fluorescent nucleotides are then washed away and a cleavage step restores the 3' hydroxyl group for further nucleotide incorporation in the next cycle. The results of sequencing are sample files in the *fastq* format containing sequence and quality information for the captured reads.

1.6.2 Overview of RNA sequencing analysis methodology

For RNA sequencing, large numbers of reads from *fastq* files need to be computationally assigned to their relevant genes and transcripts to obtain information about differential gene expression, differential splicing and transcript usage. RNA sequencing analysis follows the model for large data analysis as described by the epicycle of analysis (Peng and Matsui, 2015):

	Examples of workflows for RNA sequencing analysis
Stating the question	Experimental design Power calculations
Exploratory data analysis	Quality control Principal component analysis Hierarchical clustering
Model building	Differential expression analysis Differential splicing analysis
Interpretation	Network analysis Gene ontology analysis
Communication	Data visualisation

Table 1.1: Epicycle of Analysis

Many elements of the epicycle of analysis have numerous implementations in RNA sequencing analysis, but follow similar principles. Others, such as experimental design are highly dependent on the experimental question and model. The most significant choices are to be made in the differential expression and splicing analysis and these options will be discussed further.

1.6.2.1 Alignment-based methods

Originally, software for differential expression analysis required the prior alignment of every DNA fragment to a reference genome or transcriptome. The aligned fragments are subsequently counted, normalised and then quantified on a gene or transcript level. A variety of alignment software programmes exist for alignment of reads from RNA sequencing experiments. These programmes are able to take spliced alignment into account and often perform similarly on simulated datasets (Engström *et al.*, 2013). All programmes contain methods to deal with gapped alignments, but also single nucleotide polymorphisms and indels.

Initially tools were developed as extensions of previous contiguous mappers that used a database of splice junctions to align split-read portions contiguously, with *Tophat* (Trapnell *et al.*, 2009) and *GSNAP* (Wu and Nacu, 2010) being prominent examples. *STAR* subsequently provided a different alignment algorithm that finds the maximal mappable portion of the read in the genome. Should the read be over a splice site, the second part of the read is used as a seed to find the mapping on the other side of the splice junction in an unbiased way. The use of an index in the form of an uncompressed suffix array increases speed but also memory requirements (Dobin *et al.*, 2013). To increase accuracy, *STAR* can be run again in a second mapping pass in which it will no longer find new junctions but instead will align spliced reads with short overhangs to the previously found splice junctions. These features have made *STAR* the method recommended by the genome analysis toolkit for variant detection in RNA sequencing data (Alkan *et al.*, 2011). One of the most recent gapped short read aligners is *HISAT* (Kim *et al.*, 2015), which uses a hierarchical indexing scheme based on the Burrows-Wheeler transform and the Ferragina-Manzini index. An important feature of this method is the ability of the small local indices to fit into processor cache, avoiding RAM access. Searches for splice acceptors are likely to be found in the local index, avoiding resource intensive whole genome searches for splice acceptor sites, greatly increasing speed. Following alignment, gene level counts can be assembled using programs such as *Feature Counts* from the *Rsubread* package (Liao *et al.*, 2013), *Gene Counts* from *CGAT* (Sims *et al.*, 2014) or *HTSeq* (Anders *et al.*, 2015). All these tools use the positional information from aligned reads to

assign the reads to the genes reducing them to counts. These counts can subsequently be used for differential expression analysis.

1.6.2.2 Alignment-free methods

Alignment free quantification, first introduced in the software packages *Sailfish, Salmon* and *Kallisto*, allows the skipping of the computationally intensive alignment step. These methods no longer perform read alignments against the genome but align fragments of the read against an index containing transcript information.

Kallisto (Bray *et al*., 2016) constructs an index by shredding the reference transcriptome into all possible subsequences of length k contained within the reference sequence, also known as k-mers. Differential splicing is implemented using a coloured transcriptome de Bruijn graph in which the different paths correspond to differential splicing events and the colours represent the different transcript variants. The reads are then broken into k-mers which are matched to the index. The authors call this pseudoalignment. The programme then walks along the de Bruijn graph and only performs further lookups if the composition of compatible transcripts changes or if the end of the read is reached, greatly reducing computational burden.

Salmon (Patro *et al.*, 2015) uses a similar procedure in which an FMD index (a modified version of the BWA index) is prepared from the transcriptome. During what the authors have termed lightweight alignment, *Salmon* uses chains of super-maximal exact matches and maximal exact matches to find the

highest scoring transcript locus without aligning the entire read. *Salmon* can also perform *quasi-mapping* in which a k-mer based index is employed, similar to *Kallisto*.

1.6.2.3 Methods for quantification of gene-level differential expression

Following alignment-based and alignment-free quantification, differential expression software is used to normalise samples, estimate dispersions and fit a model based on provided variables and co-variates. Gene-level differential expression will be discussed here, whereas transcript-level differential expression will be presented subsequently. There are multiple advantages in performing gene-level differential expression over transcript-level expression for short-read RNA sequencing data. It has been shown that many genes express one dominant transcript per gene (Gonzàlez-Porta *et al.*, 2013), despite often having multiple annotations. Both simulated and real datasets indicate that gene abundance estimates are more accurate than transcript abundance estimates (Soneson *et al.*, 2015). Finally, gene level differential expression analysis is more robust to sequence and position-specific biases (Love *et al.*, 2016).

The two most commonly used differential expression tools, *DESeq2* (Lov*e et al.*, 2014) and *edgeR* (Robinson *et al.*, 2009), use counts derived from alignment-based methods as their input. Both methods use similar scaling factor normalisations to estimate the variance between samples (Dillies *et al.*, 2013), and both methods model the counts distribution for each gene using the

negative binomial distribution. Within-group biological variation is modelled by sharing information across genes. Subsequently, between group differences are modelled and the resultant values are defined by their false discovery rate. There are important differences between the two software packages. *EdgeR* moderates the dispersion estimate using a weighted conditional likelihood model using a user-adjustable parameter (a user determined prior) for weighing the contribution of the gene estimate. *DESeq2* models the variability between replicates of the same condition and the variability between groups using maximal likelihood estimation followed by shrinkage using an empirical Bayes approach, reducing the dispersion of each gene based on the available information for this gene. Importantly, *EdgeR* and *DESeq2* can also analyse transcript abundance data generated from alignment-free methods on a gene-level basis. The use of scaled transcript abundance estimates from *Kallisto* or *Salmon* as pseudocounts improves the gene abundance estimates compared with the use of traditional counts obtained from the same programmes, especially in simulated datasets (Soneson *et al.*, 2015). Part of this effect, which is small in real RNA sequencing data, may by attributable to the fact that abundance estimates are more appropriately normalised to transcript lengths rather than to the average transcript length used in traditional count based methods.

Gene-level and isoform-level differential analysis on transcript abundance estimates from alignment free methods can also be performed using the software package *Sleuth* (Pimentel *et al.*, 2016). *Sleuth* similarly stabilizes

variance by shrinkage. However, it uses error estimates from bootstrapping data obtained from *Kallisto* or *Salmon* to decouple biological variance from inferential variance before shrinkage.

1.6.2.4 Methods to estimate differential splicing

Two major approaches exist to estimate differential splicing in RNA sequencing data. Inferences from both these methods are limited due to the short read lengths in RNA sequencing. The main determinant of the result, however, is the specific method used. This was demonstrated most clearly by a comparison of differential splicing results from five tools with representatives from both approaches on a dataset from *Arabidopsis thaliana*, a well-annotated plant organism. There was little overlap between any of the five tools, and no single result was consistent between the tools (R. Liu *et al.*, 2014). The study sheds important light on the different assumptions and models selected by each of these tools.

1.6.2.4.1 Differential Transcript Expression (DTE)

Transcript abundance quantification methods attempt to assign transcripts to the likely transcript of origin. This was first widely implemented through *Cufflinks*, which like *DESeq*2 and *EdgeR* requires prior read alignment to the genome (Trapnell *et al.*, 2010). *Cufflinks* implements reference based transcript assembly, allowing for prediction of novel isoforms not present in annotations, if reads cannot be modelled to conform to existing annotations. As a result, *Cufflinks* is relatively robust to poor transcript annotation (R. Liu *et*

al., 2014). However, studies have questioned the accuracy of the novel splicing predictions and gene models made by *Cufflinks* (Rogers *et al.*, 2012; Hayer *et al.*, 2015). The ability to predict *de novo* transcript variants also results in decreased accuracy compared to newer transcript differential expression methods (Kanitz *et al.*, 2015).

These newer methods include the alignment-free quantification methods such as *Kallisto* and *Salmon*, introduced previously. As they align to a reference transcriptome, no novel isoforms can be discovered. For well annotated organisms such as humans, *Kallisto* and *Sailfish* significantly outperform *Cufflinks* on simulated and real data (Patro *et al.*, 2014; Bray *et al.*, 2016).

An important limitation of transcript based methods is their ability to deal with multiple transcript isoforms, which are often very similar. While transcript differential expression algorithms perform well when there are only two transcripts, the accuracy decreases dramatically with increasing numbers of known transcripts. Accuracy is poor for four or more transcripts (Hayer *et al.*, 2015). In complex organisms, such as mouse and human, where most genes have more than five annotated transcript variants, transcript-level differential expression using short reads remains challenging.

1.6.2.4.2 Differential Exon Usage (DEU)

The alternative to transcript-level differential expression is the exon-based differential exon usage estimation. This approach skips isoform expression estimation and only compares the expression in exons instead. One of the most

popular tools, *DEXseq* (Anders *et al.*, 2012), tests for significant differences in read counts in exons and at exon junctions. Another software, *rMATS*, uses a Bayesian model to compare actual with expected exon usage based on junction reads (Shen *et al.*, 2014). These methods can model a variety of alternative splicing events including exon skipping, cassette exon usage or intron retention. While exon-based methods show a high level of validation experimentally (Shen *et al.*, 2012; R. Liu *et al.*, 2014), they cannot be easily translated to the transcript level.

1.6.3 RNA Sequencing has made important contributions to understanding of ALS biology

The involvement of TDP-43 in splicing regulation was the first function of TDP-43 discovered (Buratti *et al.*, 2001). Unsurprisingly, RNA processing and biology became one of the key areas of interest in ALS research following the discovery of TDP-43 in intracellular inclusions of ALS patients. The contemporaneous development of high throughput sequencing technology provided key methods for the study of RNA biology in ALS.

1.6.3.1 TDP-43 preferentially binds to UG-rich intronic motifs

Using high-throughput sequencing, RNA binding sites of TDP-43 were discovered using high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP) in mouse and human brain (Polymenidou *et al.*, 2011) and using individual-nucleotide resolution cross-linking and immunoprecipitation (iCLIP) in cell lines and human brain from TDP-43

positive FTD subjects (Tollervey *et al.*, 2011). TDP-43 was found to bind UG-rich intronic motifs and one or more binding sites were found in around 30% of genes. TDP-43 binding was uniformly distributed across the length of pre-mRNAs and the great majority being >500 bases distal to intron-exon boundaries. On nuclear-cytoplasmic fractionation the majority of binding sites were in the nuclear fraction, but there was a significant enrichment of 3'UTR binding in the cytoplasm. Similar results were also obtained from RNA immunoprecipitation followed by microarray analysis (RIP-chip) with further support for TDP-43 binding motifs (Narayanan *et al.*, 2012).

1.6.3.2 TDP-43 is a key regulator of splicing with dose dependent effects on splicing events

One of the first global studies of splicing upon TDP-43 knockdown in mouse brains was performed using both RNA-sequencing and a custom designed microarray (Polymenidou *et al.*, 2011), revealing 779 alternative splicing events, the majority of which had prior evidence for alternative splicing. The TDP-43 knockdown also had an effect on gene expression with 362 downregulated and 239 upregulated genes. One of the key observations of the study was that TDP-43 potentially stabilised pre-mRNAs containing long introns: this was supported by the observation that long intron containing pre-mRNAs were downregulated upon depletion of TDP-43.

The same splicing microarray was subsequently used to characterise splicing changes in mice expressing the human wild-type or human Q331K-mutant TDP-43 under the murine prion promoter. When compared to the

nontransgenic mouse, overexpression of human TDP-43 resulted in 824 differential splicing events in the wild-type mouse, 1195 in the Q331K mutant mouse with high overexpression and only 208 in the Q331K mutant mouse with lower overexpression level (Arnold *et al.*, 2013). Many of the changes were driven predominantly by overexpression and differential splicing events opposite to those observed upon TDP-43 knockdown (Polymenidou *et al.*, 2011), providing important evidence for conserved function of human TDP-43 in mouse cells. Furthermore, dose-dependent effects on splicing were observed in the two mice overexpressing TDP-43 variably. Further evidence for the conservation of TDP-43 function across species was provided by RNA sequencing of central nervous system tissue of *Drosophila melanogaster* larvae (Hazelett *et al.*, 2012). Both overexpression and knockdown of the fly TDP-43 orthologue, TBPH, resulted in differential expression changes and splicing changes that had significant overlap with mouse data, however analysis was conducted using a subset of results with putative TDP-43 binding sites derived from known TBPH motifs. Evidence of dysregulation of splicing at post-mortem came from laser capture microdissection of motor neurons from individuals with sporadic ALS followed by exon array analysis, showing alternative splicing in 4311 genes (24% of all genes studied), including many of the genes known to harbour ALS-causing mutations. There was significant overlap with differential splicing events observed in mutant TDP-43 fibroblasts.

1.6.3.3 The ALS-associated proteins TDP-43 and FUS have distinct binding profiles but have some functional overlap of differential expression and splicing changes

RNA sequencing and related microarray analysis has also been used to quantify the relationship and functional overlap between TDP-43 and FUS. The two RNA binding proteins were shown to have distinct binding profiles using HITS-CLIP (Lagier-Tourenne *et al.*, 2012). Upon antisense oligonucleotide mediated knockdown of the respective proteins in mouse brain, the two divergent RNA binding proteins had modest overlap in splicing alterations and differential expression. However, the overlap was enriched for long intron containing genes, potentially indicating a functional convergence. Comparative knockdown of TDP-43 and FUS was subsequently also performed in primary cortical neurons using shRNA delivered via lentivirus (Honda *et al.*, 2014). 25% of differentially expressed genes and 10% of differential splicing events were shared by the two knockdown experiments, similar to the earlier in vivo study, again indicating potential pathways that could be shared between TDP-43 and FUS, but also significant divergence.

1.6.3.4 TDP-43 is involved in repression of cryptic exons

A key discovery in ALS biology made through RNA sequencing was the discovery that TDP-43 represses non-canonical splice site usage. Non-canonical splice site usage leads to expression of exons that are not annotated by current genomic databases, also known as ‘cryptic exons’. The incorporation of these exons into transcripts often results in novel stop codons and nonsense

mediated decay. Depletion of TDP-43 in mouse embryonic stem cells using a Cre-inducible TDP-43 knockout and siRNA knockdown of TDP-43 in HeLa cells identified novel splice variants in non-conserved areas of the genome. Human and mouse cryptic exons shared no overlap. Importantly, human cryptic exon expression was validated in post-mortem ALS-FTD brains. The repression of cryptic exons in mice could be restored using a minimally chimeric protein bound to UG repeats, resulting in the prevention of cell death in this model of TDP-43 knockdown (Ling *et al.*, 2015).

1.6.3.5 Post mortem transcriptional signatures of sporadic ALS are different from familial ALS

A large RNA expression study using gene expression arrays on adult laser-captured spinal motor neurons from patients and controls provided particularly interesting insights on ALS biology. Weighted gene co-expression network analysis was used to find networks that correlated with age. These maturation and aging pathways had significant overlap with sALS modules in opposite directions with significant disruption of aging and maturation hub genes by sporadic ALS (Ho *et al.*, 2016).

RNA sequencing studies have tried to elucidate differences between sporadic ALS and *C9orf72* HRE positive ALS (Prudencio *et al.*, 2015) in post-mortem brain tissue. The two areas studied were the cerebellum and prefrontal cortex at the level of the nucleus accumbens, which were not affected by TDP-43 pathology in these patients. Both sporadic and *C9orf72* positive ALS patients had a significant amount of differential expression changes compared to

control (>100 in sALS and >300 in c9ALS, $\log_2$ fold change > 2), widespread misregulation of alternative splicing with a predominance for cassette exon events and large amounts of alternative polyadenylation events. However, only a small number of the differentially expressed genes and alternative splice sites overlapped between sporadic ALS and *C9orf72* positive ALS, indicating divergent pathways in these two conditions.

1.6.3.6 RNA sequencing of iPSC derived motor neuron models of ALS provides a detailed appraisal of the model as well as novel insights into disease biology

As discussed previously, induced pluripotent stem cell models of amyotrophic lateral sclerosis have provided a unique opportunity to study human motor neurons with disease causing mutations *in vitro*. Transcriptional studies have now repeatedly shown that these cells have a transcriptional profile that resembles foetal neurons more closely than mature human neurons (Handel *et al.*, 2016; Ho *et al.*, 2016), raising important questions about their ability to model diseases associated with ageing. The power of iPSC derived motor neuron technology when combined with isogenic control lines derived by CRISPR technology allowed for the dissection of mutation-specific molecular pathways in iPSMNs from patients with the *SOD1* mutation (Kiskinis *et al.*, 2014). Differentially upregulated transcripts were found to be enriched for cytoskeleton organisation, transcriptional regulation and motor proteins, while transcripts involved in mitochondrial function and protein translation were downregulated. These pathways were subsequently experimentally validated

using mitochondrial imaging and readouts of the unfolded protein response and endoplasmic reticulum (ER) stress.

1.7 Aims

The discovery of *C9orf72* as the most common genetic cause of ALS has resulted in novel approaches and ideas in the study of the pathophysiology of ALS. The typical TDP-43 pathology seen in *C9orf72* HRE carriers and the fascinating pathobiology of the mutation itself raise important questions about the importance of the key protein found in intracellular inclusions of ALS patients.

This thesis will examine the significance of TDP-43 pathology in *C9orf72* post-mortem tissue, examining its distribution and comparing it to other features of *C9orf72* pathology including sense RNA foci, GA dipeptide inclusions, *C9orf72* expression and both *C9orf72* promoter hypermethylation as well as a global methylation profile (Chapter 3).

The second part of this thesis will investigate TDP-43 abnormalities in induced pluripotent stem cell derived motor neurons from *C9orf72* HRE carriers, looking at the distribution and posttranslational modifications of TDP-43 protein (Chapter 4).

The final part of this thesis will investigate TDP-43 dysfunction in a BAC transgenic mouse model expressing the TDP-43 with and without the M337V

mutation under the control of its native promoter using RNA sequencing (Chapter 5).

Chapter 2: Methods

2.1 Analysis of Neuropathological Samples

2.1.1 Cases

12 cases with a *C9orf72* HRE as well as 5 sporadic ALS (sALS) cases and 5 controls who died from non-CNS disease were obtained from the Thomas Willis Brain Bank at the University of Oxford. Tissue bank approval for this study was granted by LREC 07/H0606/85. Controls and sALS cases were matched for age at death (Table 2.1).

	C9 ALS	C9 ALS-FTD	C9 FTD	Controls	sALS
Number of cases	6	3	3	5	4 (+1 with cognitive symptoms)
Age at Death (years)	58 [41-71]	59 [55-62]	69 [56-78]	60 [41-67]	58 [38-64]
Sex	4M, 2F	3M	2M, 1F	5F	3M, 2F
Disease Duration (years)	1.8 [1-3]	9.0 [8-10]	6.7 [3-9]	N/A	1.5 [1-2]

Table 2.1: Demographic characteristics of post-mortem cases included in the study. C9 = *C9orf72* positive cases.

Where available, paraffin-embedded blocks were obtained from non-motor frontal cortex, hippocampus at the level of the lateral geniculate nucleus, cerebellum, medulla at the level of the hypoglossal nucleus and the spinal cord. Frozen tissue from the same brain regions from the opposite hemisphere as well as frozen spinal cord was also retrieved from the brain bank.

2.1.2 Immunohistochemistry for TDP-43, p62 and GA dipeptide repeats

Immunohistochemistry was performed on 4µm sections from all available formalin fixed, paraffin-embedded tissue. One section from each block was stained with haematoxylin and eosin. Immunostaining was performed with TDP-43 (dilution 1:2000, rabbit, Proteintech), p62 (1:2000, mouse, BD Biosciences) and anti-GA dipeptide clone 5F2 (dilution 1:6000, mouse, gift). Slides were baked at 60°C for 10 minutes and dewaxed. Antigen retrieval was performed in citrate buffer (pH 6) in an autoclave. Staining was performed using the Envision kit (Dako) with overnight primary antibody incubation at 4ºC.

Immunohistochemistry was assessed using a semi-quantitative scoring system after looking at the entire slide:

0 = no abnormal staining

1 = rare abnormal inclusions

2 = occasional abnormal inclusions present on most fields of view

3 = frequent abnormal inclusions

4 = very frequent abnormal inclusions

2.1.3 Fluorescence in situ hybridisation (FISH) and immunofluorescence analysis

Fluorescence in situ hybridisation for sense foci was performed on 4µm sections from all available paraffin-embedded sections using a previously

published method (Mizielinska et al., 2013). In brief, slides were dewaxed using HistoClear (National Diagnostics). Antigen retrieval was performed in a microwave in citrate buffer (pH6) and sections were then dehydrated through serial alcohol dilutions and air dried. Following rehydration in phosphate buffered saline (PBS), sections were placed in pre-hybridisation buffer (2x saline-sodium citrate (SSC) , 50% formamide) and heated to 80°C. Hybridisation was performed for two hours using hybridisation solution (2x SSC, 50% formamide, 0.16% BSA, 0.8 mg/ml salmon sperm and yeast tRNA, 8% dextran sulfate, 1.6 mM vanadyl ribonucleoside, 5mM ethylenediaminetetraacetic acid (EDTA)) containing 0.2ng/µl of oligonucleotide (sense foci: (CCCCGG)4-Cy3, control: (CAGG)6-Cy3). Extensive washes were performed in 50% formamide and 2X SSC at 80°C followed by washes in 2x SSC at room temperature.

Following FISH, the sections were blocked in goat serum and primary antibody was applied overnight (pTDP clone 1D3, 1:1000, gift; GA dipeptide clone 5F2, 1:6000, gift) followed by washes and incubation in secondary antibody (Alexa Fluor 488 and 647, 1:2000, Life Technologies). Images were acquired on a Zeiss LSM 710 confocal microscope at 63X magnification using z-stacks. Four fields of view were acquired in the frontal cortex, one in the area of hippocampal granule cells, three in the area of CA4 neurons, one in the cerebellum as well as two fields of view each containing XII cranial nerve nucleus cells and anterior horn cells. Only neurons were counted, as judged by their morphology and position.

2.1.4 DNA and RNA extraction

DNA was extracted using the Quiagen QiaAMP DNA Mini kit following manufacturer's instructions and stored at 4°C in 0.5 mM EDTA for use within days. The DNA quantity and quality were tested using Nanodrop (ThermoFisher). RNA was extracted using the RNA mini kit and also quantified using a spectrophotometer (Nanodrop). All RNA samples were further quality assured using automated RNA gel electrophoresis to determine RNA integrity (Bioanalyzer 2100) and stored at -80°C. For *C9orf72* repeat carriers a Southern Blot analysis was carried out to confirm the presence of the repeat.

2.1.5 Bisulfite sequencing

Direct bisulfite sequencing was achieved by adapting a published assay (Baker *et al.*, 2005). 400 ng of genomic DNA was bisulfite-converted using the EpiTect bisulfite kit (QIAGEN). PCR was performed on bisulfite-treated DNA using the following primers to amplify the 5' CpG island of *C9orf72*: forward 5'-GGAGTTGTTTTTTATTAGGGTTTGTAGT-3', reverse 5'-TAAACCCACACCTACTCTTACTAAA-3' (E. Y. Liu *et al.*, 2014). GoTaq Hot Start polymerase (Promega) was used as per manufacturer's instructions under the following thermal cycling conditions: 95°C for 5 minutes, followed by 10 cycles of touchdown PCR with 30 seconds denaturation at 95°C, 30 seconds annealing reducing the annealing temperature from 70°C to 60°C by -1°C per cycle and with 1 minute extension at 72°C, followed by 30 cycles with the annealing temperature at 60°C and with a final extension step at 72°C for 10

minutes. PCR products of 466 bp were purified by agarose gel extraction and commercially Sanger sequenced by Source BioScience using the forward primer. Resulting chromatograms were analysed using Chromas v2.6 software and the percentage of methylcytosine (mC) determined for each CpG using relative peak heights whereby %mC = C/(C + T). In all samples over 95% of unmethylated C nucleotides were converted to T. **The direct bisulfite sequencing was designed and carried out by Dr Andrew Douglas on DNA provided by the author. The final analysis was carried out by the author.**

2.1.6 Restriction digest methylation analysis

For restriction digest methylation analysis 100ng of DNA was digested for 2 hours at 37°C in a total volume of 20µl. Three digestion conditions were performed per sample. For the quantification of CpG island methylation the digestions were performed with 0.5µl of Hae3 (20000U/ml) alone, Hha1 (20000U/ml) and Hae3 (20000U/ml) as well as Hpa2 (10000U/ml) and Hae3 (20000U/ml). Quantitative real-time PCR (qPCR) was performed using ThermoFisher Fast SYBR Green Master Mix on the digested samples using 10µl reactions in triplicate, each containing 1.25µl of digested DNA. qPCR was performed on a Roche LightCycler 480 using primers to assess promoter methylation (forward: 5' CAGTGTGAAAATCATGCTTGAGAGA 3', reverse: 5' TTTGTGCTTGGTAGGCAGTG 3') as well as a region not affected by methylation-specific digestion, to exclude sample degradation as an explanation for the results (forward: 5' GTGGCGAGTGGTGAGTGA 3'; reverse 5' GCTTCGGTCAGAGAAATGAG 3').

2.1.7 450k Methylation Analysis

For analysis of global methylation four control cases, four cases with sALS and four *C9orf72* positive cases were selected. Three of the *C9orf72* cases and three sporadic cases had ALS only. The remaining subject in each group had frontal cortex involvement pathologically as well as evidence of clinical frontal lobe involvement. DNA was extracted and quality assured as described above.

Quantification was performed using PicoGreen (ThermoFisher) according to the manufacturer's instructions. 600ng of DNA per sample was submitted to the Welcome Trust Centre for Human Genetics, University of Oxford. Following bisulphite conversion, the samples were processed on a 450K Illumina Methylation Chip.

The analysis was carried out using the RnBeads package on R 3.2.3 using a manual workflow. In brief, the sample identity was ascertained using SNP-dependent probes, resulting in the removal of two of twenty-four replicates due to sample errors. Subsequently low quality probes, all known SNP-dependent probes as well as all probes on sex chromosomes were filtered. Hierarchical clustering, principal component analysis and multidimensional scaling was performed as part of exploratory data analysis. Differential methylation analysis was carried out using the limma package from within RnBeads. The final differential methylation analysis was carried out on male samples only, to evade sex-dependent differential methylation on non-sex chromosomes.

2.1.8 Gene expression analysis of *C9orf72*

Following RNA extractions, cDNA libraries were generated from 2µg RNA using Superscript VILO, following manufacturer's protocol in a total volume of 20µl. Quantitative real-time PCR was performed using 20µl reactions with four TaqMan assays (ThermoFisher): A published custom made assay detecting the exon 1B containing variant 2 (NM_018325.3, DeJesus-Hernandez, et al., 2011), assay Hs00948764_m1 for variant 3 (NM_001256054.1), assay ID Hs00331877_m1 for variant 1 (NM_145005.5) and assay ID Hs00266705 for GAPDH. Reactions were carried out in triplicate on a Quiagen Rotor-Gene Q cycler (Qiagen) using TaqMan Gene Expression Mastermix (ThermoFisher). Relative abundance estimation was carried out using the $2^{-\Delta\Delta Ct}$ method.

2.2 iPSMN Differentiation and Experiments and RNA sequencing validation

2.2.1 Induced Pluripotent Stem Cell Culture

The human iPSC lines derived for this study were derived from human skin biopsy fibroblasts, following signed informed consent, with approval from the South East Wales Research Ethics Committee (REC ref no. 12/WA/0186) and the South Central – Berkshire Research Ethics Committee, UK (REC 10/H0505/71). Induced pluripotent stem cells, grown in mTesR medium (StemCell Technologies) on Matrigel (Corning) in 6 well plates. All iPSC lines were derived from skin biopsy fibroblasts in the James Martin Stem Cell

Facility, University of Oxford, under standardised protocols to minimise any potential variation attributable to laboratory differences and/or handling. Cells were either immediately differentiated using the neuronal differentiation protocol described below, or expanded in their original medium. Cells were grown to full confluence and passaged at a maximum dilution of 1:3. For passaging, cells were incubated with 0.5 mM EDTA (Life Technologies) for 5 minutes, which was subsequently aspirated. Cells were then detached and resuspended using prewarmed mTesR medium and grown on Matrigel coated plates. For freezing, cells were re-suspended in appropriately diluted 2X freeze mix containing 80% USDA-approved foetal bovine serum (Sigma Aldrich), 10% Knockout DMEM (Life Technologies) and 10% DMSO (Sigma Aldrich).

2.2.2 Motor Neuron Differentiation

Motor neuron differentiation was carried out using a protocol derived from previously published work (Maury *et al.*, 2014; Qu *et al.*, 2014). On becoming confluent, iPSCs were grown in DMEM/F12 media with HEPES supplemented with N2 (Life Technologies), B27 (Life Technologies), antibiotic-antimycotic (Life Technologies), β-mercaptoethanol (Sigma Aldrich) 0.5µM Ascorbic Acid (Sigma Aldrich), 3µM Chir99021 (Tocris) and 1μM Compound C (Tocris). On day 2, 1μM retinoic acid (Sigma Aldrich) and 500nM smoothened agonist (SAG) were added to the medium. On day 4 compound C and Chir99021 were removed from the medium. On day 9, cells were incubated in Accutase (Life Technologies) and split 1:3 to 1:6. Cells were resuspended in medium with the same composition and grown on 6-well plates pre-coated with Geltrex (Life

Technologies). Between day 16 and 20, the cells were split harshly (1:12) and resuspended in neuronal maturation medium containing the above factors as well as 10ng/ml brain-derived neurotrophic factor (BDNF, (Life Technologies), 10ng/ml Glial-cell derived neurotrophic factor (GDNF, Life Technologies), 0.5 µg/ml Laminin (Life Technologies) and 10μM DAPT (Tocris). Proliferation was observed and on reaching confluence cells were further passaged using the above methods as necessary. After passaging of cells the medium was supplemented with ROCK inhibitor (Tocris) for 24-48 hours.

2.2.3 Treatments and fixation

After day 30 differentiated motor neurons underwent treatment with MG-132. Cells were incubated for 24 hours in DMEM/F12 medium containing either N2 and B27 only (control medium) or the same medium containing 5μM MG-132. Following treatment cells were either immediately fixed in 4% paraformaldehyde or frozen at -80°C for further use.

2.2.4 Analysis of Protein Expression

2.2.4.1 Soluble and insoluble fractions

Extraction of soluble and insoluble protein were carried out in accordance with a published protocol (Volpicelli-Daley *et al.*, 2014) with modifications.

2.2.4.1.1 Soluble

Following extraction, pellets were optionally frozen and then resuspended in 200 µl of solution A: 1% Triton-X 100/tris-buffered saline (TBS) with protease and phosphatase inhibitors (Roche). The cells were sonicated using 10 pulses of 1s at 10% of power and incubated on ice for 30 minutes and transferred into microcentrifuge tubes (Beckmann-Coulter). Subsequently, the suspension was centrifuged at 4°C at 100,000*g* for 30 minutes. The supernatant was aspirated as the soluble fraction.

2.2.4.1.2 Insoluble

The pellet was again resuspended in 100 µl of the above solution A and sonicated 10 times using 1s pulses at 10% power. The suspension was centrifuged at 4°C at 100,000*g* for 30 minutes. The supernatant was discarded. The pellet was then resuspended in 100 µl 2% SDS/TBS solution with protease and phosphatase inhibitors. The solution was sonicated 15 times at 1s pulses and 10% power and the supernatant was then collected from all tubes as the insoluble fraction.

2.2.4.2 Western Blotting

For western blotting, 10µg of protein with Sample Reducing Agent (Life Technologies) and LDS Sample Buffer (Life Technologies) was denatured at 90°C for 5 minutes before returning on ice. Samples were run on 4-12% bis-tris gel (Life Technologies) in running buffer (Life Technologies) at 100V for 100 minutes. Gels were transferred onto PVDF membrane using the iBlot 2 dry

blotting system (Life Technologies). The membrane was blocked using 10% milk in containing 0.1% Tween-20 (PBS-T) for 1 hour at room temperature and primary antibodies (see Table 2.2**:** Primary antibody suppliers and concentrations used for immunofluorescence and FISH experiments

) were incubated overnight at 4°C in 1% milk/PBS-T. Following three 10-minute washes in PBS-T at room temperature, the membrane was incubated in secondary antibody (GE) in 1% milk/PBS-T for 1 hour at room temperature. Bands were detected using ECL and ECL Prime western blotting detection reagents (GE).

2.2.5 Analysis of RNA expression analysis using RT-PCR and qRT-PCR

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) reactions were performed in two steps.

2.2.5.1 Generation of cDNA

RNA was extracted using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions and stored at -80°C. On-column DNase digestion was performed. RNA was quantified using NanoDrop (ThermoScientific). cDNA was created using the High-Capacity Reverse Transcription Kit (Life Technologies) according to manufacturer's instructions.

2.2.5.2 Primer Design

Primers were designed using Primer3Plus (Untergasser *et al.*, 2007) and checked for specificity using NCBI Blast (Johnson *et al.*, 2008). Primers for qPCR were validated using control iPSC-derived motor neuronal cDNA with 5 serial 1:5 dilutions. All primer Sequences can be found in the Appendix.

2.2.5.3 qPCR

qPCR was performed using the Fast SYBR Green Master Mix (Life Technologies) using 0.2µl of cDNA and 250µM primers. 20µl samples were run in triplicate on the Lightcycler480 (Roche) using the enzyme manufacturers machine-specific instructions. Three reference genes were used for all reactions (*TOP1*, *RPL13A*, *ATP5B*) and the geometric mean of all three genes was used as the reference for gene expression estimation. Relative abundance estimation was carried out using the $2^{-\Delta\Delta Ct}$ method.

2.2.6 Immunofluorescence

For immunofluorescence cells were grown for a minimum of 48 hours on 12mm coverslips previously pre-coated for one hour with Geltrex (Life Technologies). Coverslips were fixed for 30 minutes in 4% paraformaldehyde and washed three times in phosphate buffered saline (PBS) for 10 minutes each. After blocking for one hour in 10% normal goat serum (Vector Laboratories) or donkey serum (Sigma Aldrich) in 0.2% Triton X-100/PBS primary antibodies were incubated overnight at 4°C in 1% serum/0.2% Triton X-100/PBS. Three 10-minute washes with 0.2% Triton X-100/PBS were

followed by a one hour incubation in secondary antibody at 1:2000 (all Life Technologies: Alexa Fluor) in 1% serum/0.2% Triton X-100/PBS. Three 10-minute washes in with 0.2% Triton X-100/PBS were carried out, the middle wash containing 20μg/ml 4′,6-Diamidino-2-Phenylindole (DAPI) (Sigma-Aldrich) and subsequently coverslips were mounted onto glass slides using hard-set mounting medium. Primary antibodies suppliers and concentrations used are outlined in Table 2.3.

Antibody	Concentration	Supplier
TDP-43(rb)	1:2000	Proteintech
TDP-43 (ms)	1:2000	Proteintech
Islet1	1:200	DHSB
Hb9	1:50	DHSB
Tuj1 (rb)	1:2000	Covance
Tuj1 (ch)	1:2000	Covance
pTDP (rt)	1:1000	Gift from Prof M Neumann
5F2 (ms)	1:6000	Gift from Prof D Edbauer

Table 2.2: Primary antibody suppliers and concentrations used for immunofluorescence and FISH experiments

2.2.7 Fluorescence In Situ Hybridisation

RNA fluorescence *in situ* hybridisation (FISH) was performed using modified methods previously published (Mizielinska *et al.*, 2013). Suppliers and catalogue numbers for FISH reagents are detailed in Table 2.3. Cells grown on 12mm RNase-free coverslips in 24 well plates were fixed using 4% paraformaldehyde for 20 min, washed in PBS and permeabilised using 0.2% Triton X-100 (Sigma) in PBS for 30 minutes. Cells were then dehydrated in a graded series of alcohols, air dried and rehydrated in PBS, briefly washed in 2xSSC and incubated for 30 minutes in pre-hybridisation solution containing 2xSSC and 50% formamide. Hybridisation was carried out for two hours in

250μl of pre-heated hybridisation solution (50% formamide, 2xSSC, 0.16% BSA, 0.8mg/ml salmon sperm, 0.8mg/ml tRNA, 8% dextran sulfate, 1.6 mM vanadyl ribonucleoside, 5mM EDTA, 0.2ug/ml probe) at 80°C. Coverslips were firstly washed three times for 30 minutes each at 80°C in 1ml high-stringency wash solution (50% formamide/0.5xSCC), and then washed three times for 10 minutes each at room temperature in 1ml of 0.5xSCC. This was followed by a brief wash in PBS and immunofluorescence as described above. Optional treatments using RNase A (0.1mg/ml; Life Technologies: 12091-021) or DNase (150U/ml, Quiagen: 79254) were performed at 37°C for 1 hour following permeabilisation. After treatment the cells were dehydrated and air dried and the hybridisation protocol was resumed as described above. 2'*O*-methyl RNA probes were conjugated with Cy3 (Integrated DNA Technologies) and probe sequences were $(CCCCGG)_4$ or $(CAGG)_6$.

Reagent	**Supplier**
tRNA	Sigma-Aldrich
Formamide	Life Technologies
BSA	Sigma-Aldrich
20xSSC	Life Technologies
Salmon sperm	Life Technologies
Dextran sulfate	VWR
Vanadyl ribonucleoside complex	New England Biolabs
EDTA	ThermoFisher

Table 2.3: Suppliers for FISH reagents

2.3 RNA sequencing experimets submission and analysis

2.3.1 Isolation of RNA from mouse spinal cords

Mice were sacrificed at 3 and 12 months of age, following laboratory procedures. Spinal cords were dissected and lumbar cords were isolated by an experienced scientist and subsequently frozen at -80°C in RNALater. For extraction, the sample was removed from RNALater and homogenised in 700µl Trizol reagent (Life Technologies) and transferred to a phase-lock tube (5prime). 300µl of chloroform was added followed by centrifugation at 12,000g for 15 minutes at 4°C. RNA was extracted from the aqueous layer using the miRNAeasy mini kit (QIAGEN), following manufacturer's instructions and eluted using 30µl nuclease-free water. RNA was quantified using spectrophotometry (Nanodrop, ThermoFisher) and quality was ascertained using automated RNA electrophoresis (Bioanalyzer 2100, Agilent). Only samples with RNA Integrity Number (RIN) above 8 were submitted for library prep and RNA sequencing. **All steps, other than RIN testing were carried out by Dr David Gordon.**

For iPSMN RNA sequencing 30 day old cells were directly lysed in the dish using lysis buffer from the RNA Easy Plus kit (QUIAGEN) and then following manufacturer's instructions. Elution and quality control was performed as above, but only samples with RIN above 9.5 were submitted.

2.3.2 RNA sequencing

Library preparation was carried out using the Illumina TruSeq RNA Sample Prep Kit v2 and was performed by Oxford Gene Technology (3-month data) and the Oxford Centre for Human Genomics (12-month data). Although the library prep was done by two different institutions, the actual sequencing was done at the same sequencing centre. The libraries were then run on the Illumina HiSeq2000 platform using TruSeq v3 chemistry. Sequences were paired-end and performed over 100 cycles. *Fastq* files were generated from the sequencing platform using the manufacturer's proprietary software.

iPSMN RNA sequencing was performed bye the Oxford Centre for Human Genomics on the Illumina HiSeq4500 platform over 75 cycles.

2.3.3 Addition of spike-ins

For the iPSMN RNA sequencing experiment *Sequin* spike-ins were added according to the manufacturer's instructions, with Mix A added to CRISPR samples and Mix B to *C9orf72* HRE positive iPSMNs.

2.3.4 Quality Control

Fastq files were received merged for the 3-month data and split by sequencing lane for 12-month data. Quality control was performed on individual *Fastq* files as part of the *CGATPipelines readqc* pipeline. In summary, *fastq* quality was evaluated using *FastQC* (0.11.2) and contamination by other organisms was excluded using *FastQ Screen* (0.4.4). Files were trimmed using

Trimmomatic (0.32): Sequences overrepresented in the data as judged by *FastQC* were also passed to *Trimmomatic* for trimming. Quality trimming was performed with the following options for paired end data (fa:2:40:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36).

2.3.5 Mapping

2.3.5.1 Generation of genome for analysis

For the purposes of this experiment the mouse genome was modified to contain human TDP-43. The genomic locus relevant to the experiment was determined by finding the largest overlapping fragment between the BAC and the human reference sequence. The annotations relevant to this sequence were extracted from the Ensembl build 80 human gene transfer format (*GTF*) file. The human annotations were added to *Mus musculus* Ensembl build 78 (also known as mm10) as an additional *in silico* chromosome named chrBAC.

For the experiments in chapter 4, a compound genome containing the sequins was obtained from Ensembl build 87 human was obtained by adding a sequin GTF obtained from the sequin website (www.sequin.xyz) as chrIS.

2.3.5.2 2-pass Mapping

Reads were mapped to the custom genome using *STAR* (2.4.2a) using the two-pass method. Maximum intron size was specified as 2,000,000. During the first run *STAR* was run with an index and junctions generated using the custom genome annotation, as well as the following options: --outFilterType BySJout -

-outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.04 --outFilterIntronMotifs RemoveNoncanonicalUnannotated. The junction file generated in the first run was used in the second run, without additional options. The mapping files from different sequencing lanes were merged at this stage.

2.3.5.3 Quality control

Mapping quality was assessed using methods included in the mapping pipeline from the CGATPipelines collection. Mapping quality control included *Picard* (1.106) metrics, and plotting of alignment statistics, mapping statistics, library complexity, splicing statistics as well as gene profile plots for assessment of coverage and fragmentation biases. Genotype of mice as well as their sex were confirmed by visual inspection of mapped reads at this stage using the *Integrated Genome Browser* (2.3.72).

2.3.6 Differential expression analysis

2.3.6.1 Generation of counts table

Reads mapping to genes were counted for each replicate using *featureCounts* (1.4.6) from the *Subread* software package. For this repetitive RNA as well as the following contigs were removed: chrM|chrMT|_random|chrUn. Minimum mapping quality was set at 10.

Transcripts per million (TPM) tables were also obtained using *Salmon* (0.8.2) with kmer size option 31 and index option fmd. Salmon results were imported into *R* using *tximport* (1.4.0).

2.3.6.2 Exploratory data analysis

The counts were read into *R* (3.3.0) and converted into a counts object for subsequent differential expression analysis. Following variance-stabilising transformation of the counts table, exploratory data analysis was performed using heat maps, hierarchical clustering and principal component analysis. Male mice were excluded from subsequent analysis due to divergent differential expression.

2.3.6.3 Differential Expression

The raw count data was used in *DESeq2* (1.12.2), and group-wise comparisons were performed on the results using a generalised linear model. Multiple testing correction was performed using the Benjamini-Hochberg correction. The significance of the differential expression results was ascertained using 1000 simulations using random permutations of the group label across all samples. The results were interpreted using annotations from the gene ontology (GO) database, Kyoto Encyclopaedia of Genes and Genomes (KEGG) database and the mouse genome informatics (MGI) database.

2.3.6.4 Sequin analysis

RNA Sequins were analysed by providing counts and differential expression results tables to *RAnaquin* (1.2.0).

2.3.7 Network Analysis

2.3.7.1 Network construction and data

Phenotypic linkage networks were constructed as described (Honti *et al.*, 2014), and a general network as well as a nervous system specific network was constructed using all nervous system terms from the mouse genome informatics database to re-score and integrate the network.

For all differential expression analyses genes from the mouse model were converted to human orthologues using the *Ensembl Biomart* data mining tool (Yates *et al.*, 2016). **Networks and scripts for network analysis were provided by Dr Katarina Vrlcelj and Dr Caleb Webber.**

2.3.7.2 Clustering analysis

Network analysis was carried out by selecting the top 10% of links from the respective networks. The sum of all links for the differentially expressed genes was calculated. The significance of clustering was tested by sampling genes at random from the network 1000 times, controlling for number of genes, CDS length and node degree and using these to compute a links score. Clustering was plotted using *R*.

2.3.7.3 Co-Clustering Analysis

A list of forty known ALS genes was assembled using information from the ALSoD online database (Abel *et al.*, 2012). To exclude atypical forms of ALS a second short list of typical ALS genes contained *TARDBP*, *FUS*, *SOD1*, *TBK1*,

OPTN, *TUBA4A*, *HNRNPA1*, *HNRNPA2B1*, *C9orf72*, *NEK1*. Co-clustering was tested by removing all links between the differentially expressed genes and the provided ALS genes, thus only considering links between the differentially expressed genes and ALS genes. 1000 simulations were performed to test the hypothesis that genes of interest were more similar to known ALS genes than to other genes in the network.

2.3.7.4 Subnetwork Analysis

Subnetwork analysis was carried out by identifying gene communities independently, using a recursive algorithm within the differentially expressed genes, and the resulting gene communities were tested for functional enrichment using GO, KEGG and MGI databases against a nervous system background.

2.3.8 Splicing Analysis

For splicing analysis, untrimmed reads were mapped using the 2-pass *STAR* method described above. The results were analysed using *rMATS* (3.2.5), with a predetermined power threshold set to detect only those alternative splicing events where sufficient power was available to detect 1% power in differential exon usage (Shen *et al.*, 2014).

2.3.9 TDP-43 binding data analysis

HITS-CLIP data from a previous experiment performed in mice (Polymenidou *et al.*, 2011) was obtained from *GEO* (Accession GSE27201). Quality control of

reads was performed as before, with the following Trimmomatic modified option: MINLEN:36. Trimmed reads were mapped using *STAR* and mapping quality control was performed as described previously. The output was deduplicated using *UMI-tools* with the option –ignore-umi.

Peak calling was performed with a modified CGAT iCLIP pipeline (Sudbery *et al.*, 2016). In brief, HITS-CLIP reads were represented by single coordinates, taking deletions into account. Cluster calling counted the number of read at each base and within 15 base pairs from this site and used a model to estimate peaks that favours reads on the same coordinate and therefore reads deletions. To correct for multiple testing, a Benjamini-Hochberg correction against the entire genome was subsequently performed. Low and high micrococcal nuclease peak results were merged to include peaks found only in both conditions. Peaks were analysed for their gene context. Motifs were searched using *MEME* (4.9.1). The number of TDP-43 binding peaks found in the differential expression results was determined, and significance was ascertained by permutation testing using a background gene dataset that was matched for expression distribution using the Mahalanobis distance as implemented in the software *MatchIt* (2.4-21).

Chapter 3: Neuropathology of *C9orf72* HRE carriers

3.1 Introduction and Aims

The anatomical distribution of RNA foci and dipeptide repeat inclusions, as well as *C9orf72* transcript downregulation has been extensively studied in post-mortem samples from *C9orf72* HRE carriers. Multiple groups have now shown that the burden of TDP-43 pathology and not the distribution of RNA foci or dipeptides correlates best with clinical presentation (Baborie *et al.*, 2015; Schipper *et al.*, 2015). Neuropathology has provided some insight into the temporal relationship of inclusions in *C9orf72* disease, suggesting that dipeptide pathology may predate TDP-43 pathology (Proudfoot *et al.*, 2014; Vatsavayai *et al.*, 2016). Other characteristic features of pathology consistently demonstrated in *C9orf72* HRE carriers are transcript downregulation, and RNA foci and dipeptide pathology in the cerebellum, an area that remains clinically grossly unaffected in ALS (van Blitterswijk *et al.*, 2015).

The mechanism of *C9orf72* transcript downregulation also remains under investigation. Upstream CpG island methylation (Belzil *et al.*, 2013; E. Y. Liu *et al.*, 2014; Xi *et al.*, 2014) as well as repeat methylation (Xi *et al.*, 2015) have been studied and proposed as potential mechanisms for transcript downregulation. It remains unclear if these methylation changes are only seen in proximity to the repeat or if they are a global characteristic of the disease, as

methylation changes have previously been linked to motor neuron cell death (Chestnut *et al.*, 2011).

This study sought to validate previously published distributions of GA dipeptide protein, sense RNA foci and TDP-43 pathology in a well-characterised cohort of 12 *C9orf72* patients using brain areas typically affected in ALS and FTD, as well as the cerebellum due to its prominent dipeptide pathology. The second aim was to investigate the geographical distribution of *C9orf72* transcript downregulation and methylation, and to correlate methylation with gene expression in the regions studied. Finally, the study aimed to determine if methylation is a local phenomenon related to the repeat or a global phenomenon characterising the disease.

The study cohort consists of 12 *C9orf72* HRE carriers, who had a clinical diagnosis in the ALS-FTD spectrum. Six patients were diagnosed with ALS during their lifetime, three had ALS-FTD and three patients had FTD only. Control samples were obtained from five patients who died of sporadic ALS and five age-matched controls who died of unrelated causes. All samples were obtained from the Oxford Brain Bank, and donated brains and spinal cord were sampled using a standardised sampling protocol, in which one hemisphere is fixed in formalin for immunohistochemical analysis and the other hemisphere is frozen for molecular investigations.

3.2 Results

3.2.1 TDP-43 pathology and not GA dipeptide or RNA sense foci correlates best with clinical phenotype in this cohort of *C9orf72* HRE carriers

Fluorescence in situ hybridisation (FISH) was used to successfully visualise RNA foci. Quantitative analysis of sense RNA foci showed no difference between the clinical subtypes of *C9orf72* in any of the brain regions studied (Figure 3.1A). Overall, foci were more frequent in hippocampal granule cells than in cerebellar granule cells ($p < 0.05$, Dunn's multiple comparison test), but no other regional differences in foci distribution were found. RNA sense foci were not seen in sALS or control cases, persisted with DNase treatment, and were eliminated by RNase treatment (Figure 3.2). GA dipeptide inclusions were not present in spinal cord and scarcely present in the hypoglossal nucleus, two areas characteristically affected in ALS (Figure 3.1B). In the frontal cortex, hippocampus and cerebellum they were found in all clinical subtypes of *C9orf72* disease. There were no statistical differences in the percentage of neurons with GA dipeptide inclusions on non-parametric testing between clinical subtypes in the tested brain areas. GA immunohistochemistry was also performed in all studied regions. GA inclusions were quantified using a semi-quantitative scoring method commonly used in neuropathological studies, in which the entire slide is analysed and given a score between 0 (no pathology) and 4 (very frequent pathology). Immunohistochemical analysis

Figure 3.1: Quantification of RNA foci in *C9orf72* HRE carriers at post-mortem

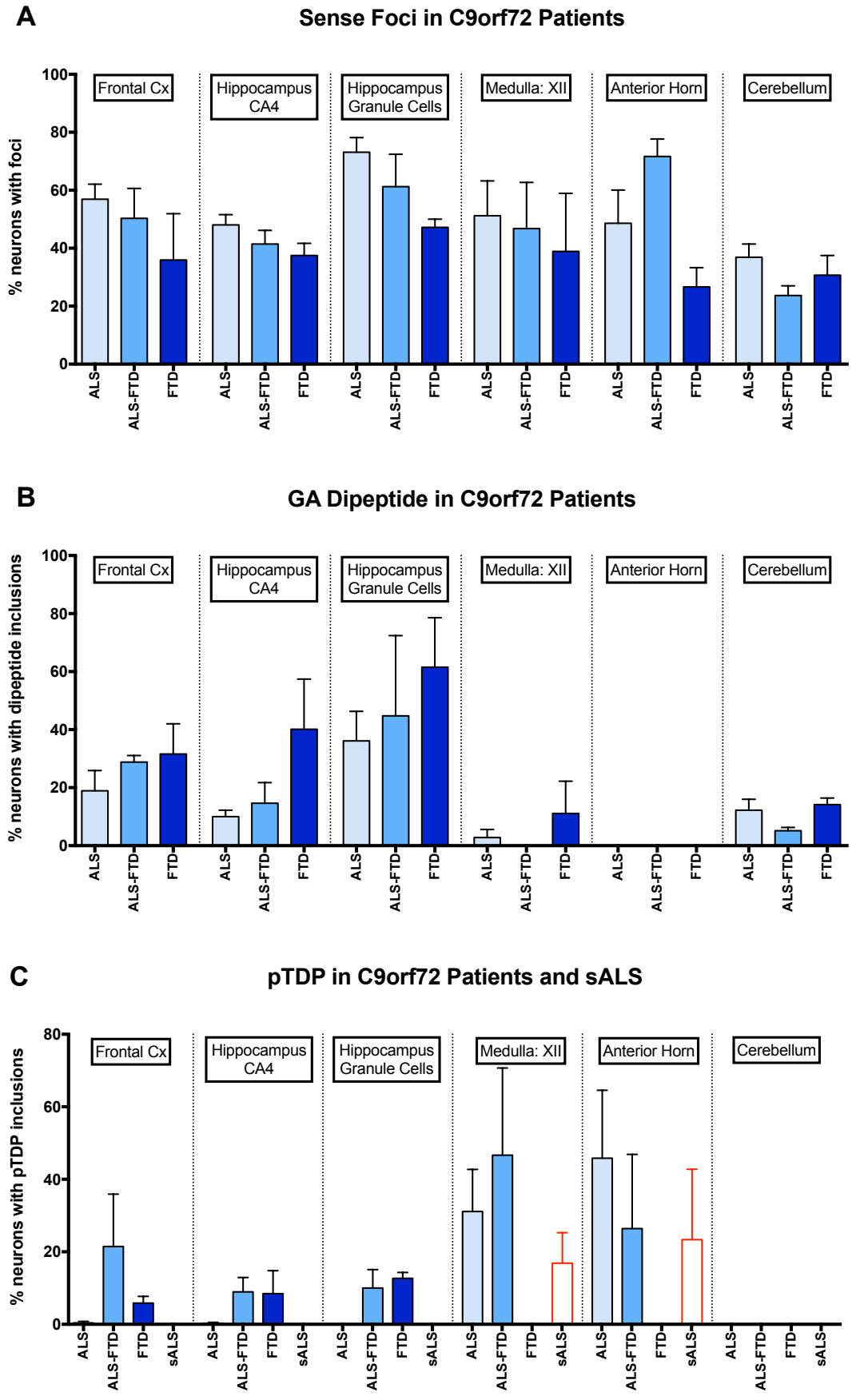

Figure 3.1 (continued)

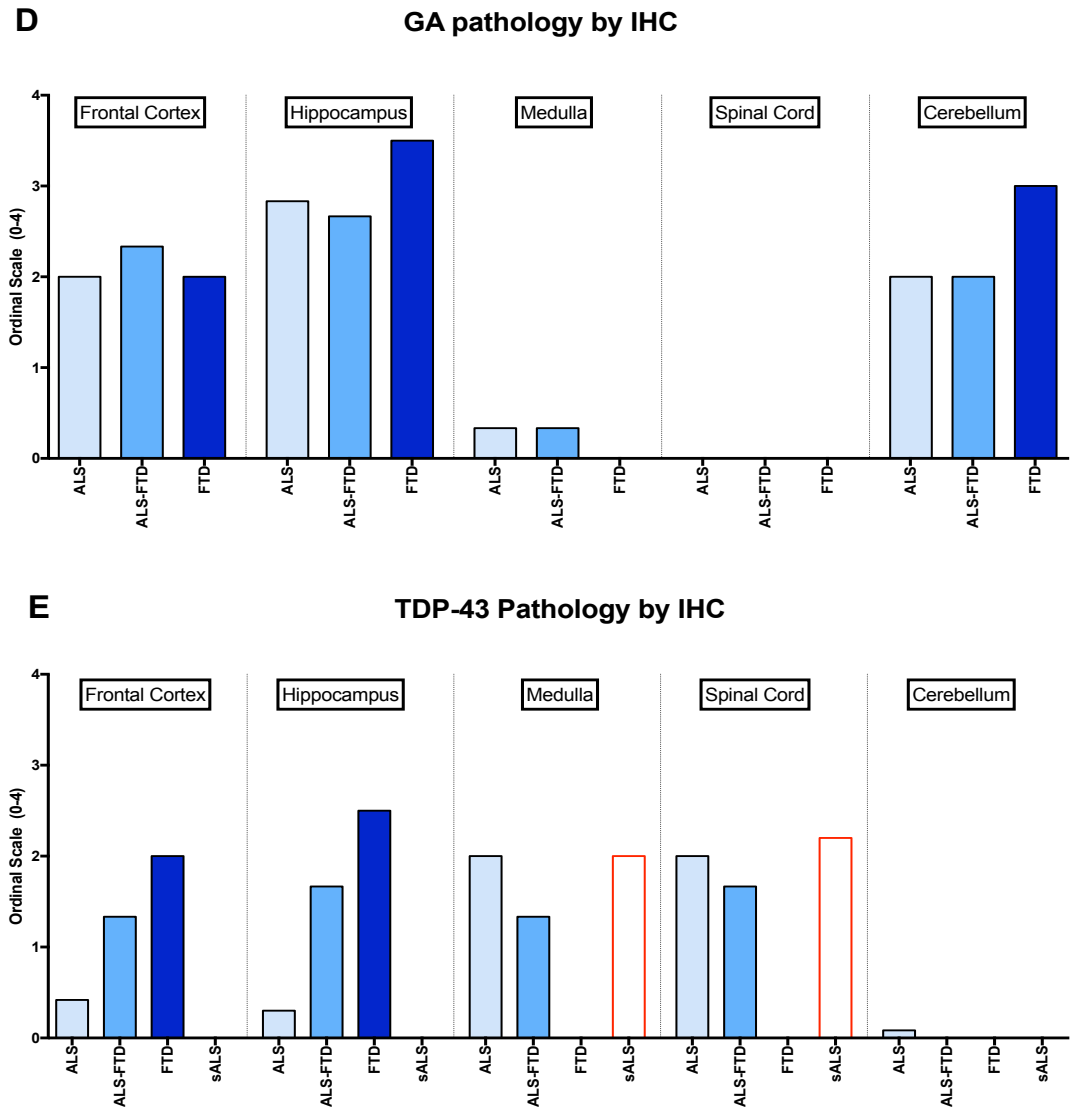

Figure 3.1: A: Frequency of sense RNA foci in neurons from six distinct regions from ALS, ALS/FTD and FTD cases carrying the *C9orf72* hexanucleotide expansion. No RNA foci were seen in controls or sporadic ALS cases (not shown in this figure). **B:** Frequency of GA dipeptide inclusions in neurons from six distinct regions from ALS, ALS/FTD and FTD cases carrying the *C9orf72* hexanucleotide expansion, visualised using immunofluorescence. No GA dipeptide inclusions were seen in controls or sporadic ALS cases (not shown in this figure). **C:** Frequency of pTDP-43 inclusions in neurons from six distinct regions from ALS, ALS/FTD and FTD cases carrying the *C9orf72* hexanucleotide expansion, as well as sporadic ALS cases, visualised using immunofluorescence. No pTDP-43 inclusions were seen in controls (not shown in this figure). **D:** Frequency of GA dipeptide inclusions in neurons from six distinct regions from ALS, ALS/FTD and FTD cases carrying the *C9orf72* hexanucleotide expansion. Immunohistochemical staining quantified using a semiquantitative scale ranging from 0-4. No staining seen in sALS and non-ALS controls (not plotted). **E:** Frequency of TDP-43 pathology in neurons from six distinct regions from ALS, ALS/FTD and FTD cases carrying the *C9orf72* hexanucleotide expansion as well as sporadic ALS controls. Immunohistochemical staining quantified using a semiquantitative scale ranging from 0-4. No staining seen in non-ALS controls (not plotted).

Figure 3.2: Control experiments for RNA foci

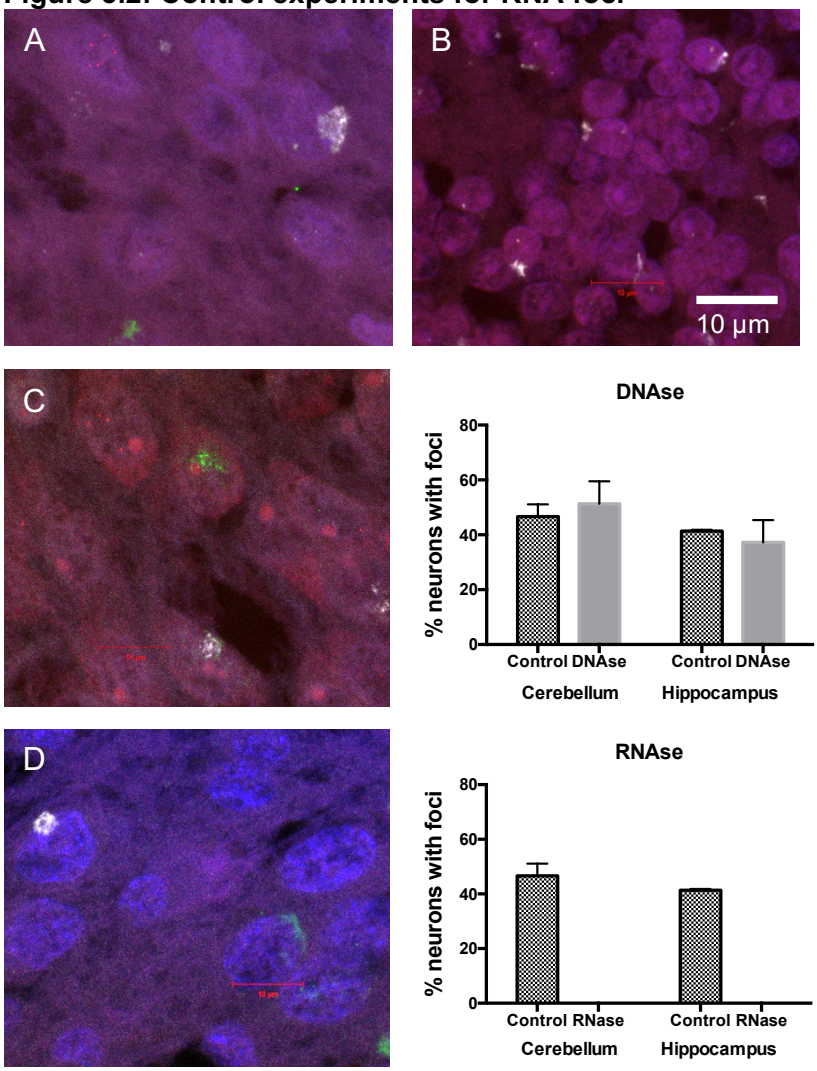

Figure 3.2: A: Immunofluorescence image showing RNA Foci in hippocampal granule cells (red), GA dipeptide pathology (white) and phospho-TDP-43 pathology (green). DAPI staining (blue) visibly staining nucleoli. **B:** Control CAGG probe staining cerebellar granule cells showing GA dipeptide pathology but no foci. **C:** Left panel demonstrates representative immunofluorescence image following DNase treatment with preserved foci (red) but no DAPI staining. Panel to right shows quantitative assessment of foci in non-treated sections compared to foci counts following DNase treatment (n = 2 cases, 6 areas each). **D:** Left panel demonstrates example immunofluorescence image following RNase treatment with preserved DAPI staining but no foci. Panel to right shows quantitative assessment of foci in non-treated sections and following RNase treatment (n = 2 cases, 6 areas each).

confirmed the findings on immunofluorescence (Figure 3.1D). Phosphorylated TDP-43 (pTDP-43) inclusions were seen in the frontal cortex and hippocampus of C9-positive ALS-FTD and FTD cases only. No pTDP-43 inclusions were seen in these areas in C9-ALS, sALS and control cases (Figure 3.1C). On the other hand, the hypoglossal nucleus and anterior horn cells were positive for pTDP-43 inclusions in C9-positive ALS and ALS-FTD cases as well as cases with sALS. No pathology in these lower motor neuron areas was seen in C9-positive FTD or in controls. The cerebellum was completely free of pTDP-43 inclusions in all cases studied. To validate findings, TDP-43 immunohistochemistry was performed on the same regions in all cases (Figure 3.1E).

3.2.2 *C9orf72* transcript downregulation is seen in transcript variant 2 in cerebellum and frontal cortex.

The transcriptional profile of *C9orf72* expression was evaluated in all areas for which frozen tissue was available. Only samples with RIN above 5 were included, in accordance with published recommendations (Fleige and Pfaffl, 2006). Variant 2 was the most abundant transcript in all areas studied in normal brains (Figure 3.3B) and in *C9orf72* HRE positive as well as sporadic ALS patients (Figure 3.3A). In the cerebellum and frontal cortex there was a significant reduction in *C9orf72* variant 2 expression between *C9orf72*-HRE positive cases and sALS cases (Figure 3.3C, $p < 0.05$, Dunn's multiple comparisons test). No significant differences were observed for intron 1a containing transcripts (Figures 3.3D and 3.3E).

Figure 3.3: *C9orf72* expression in brain and spinal cord samples from *C9orf72* HRE carriers, sALS and control patients

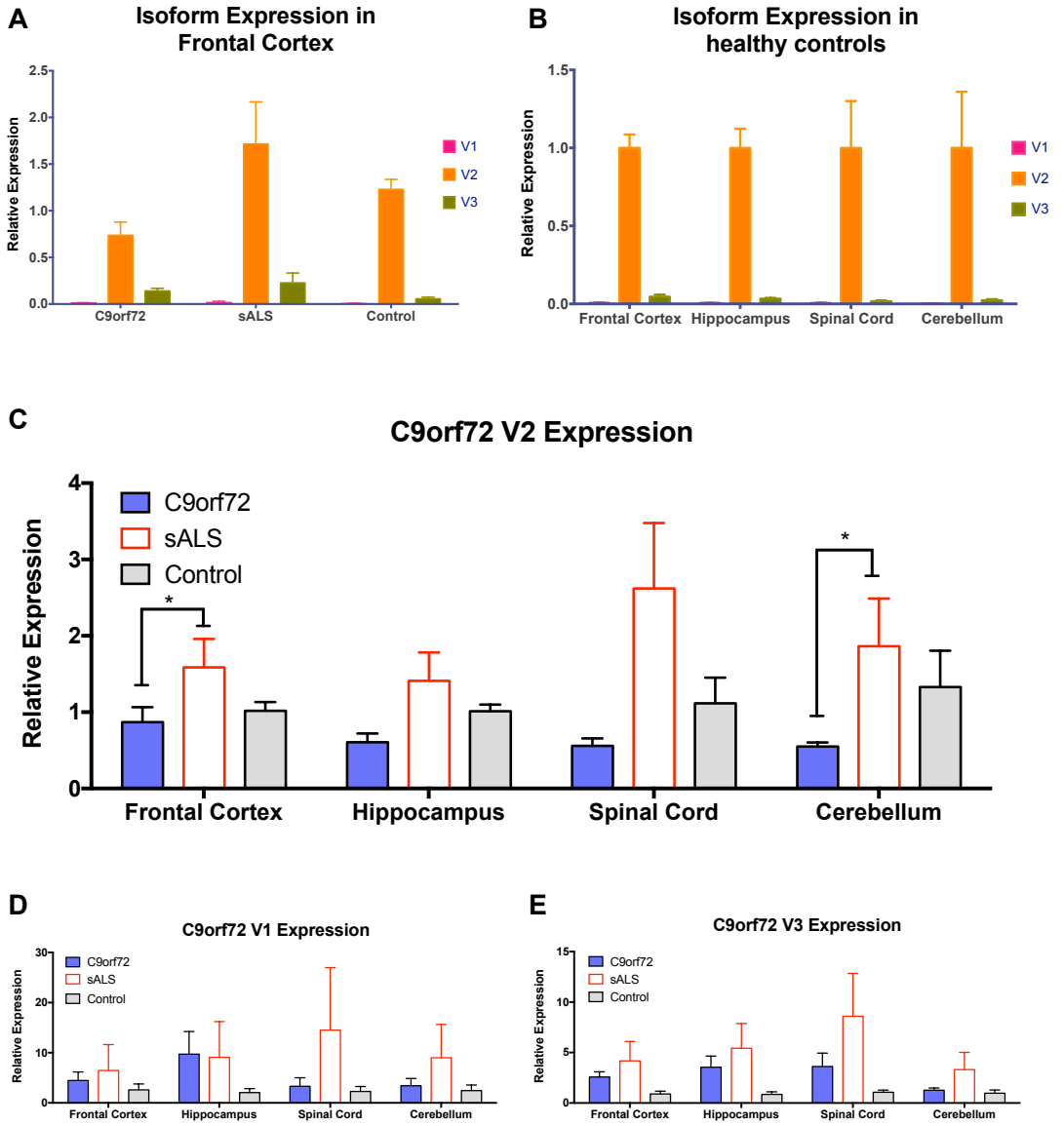

Figure 3.3: A: Comparison of *C9orf72* isoform expression in the frontal cortex in *C9orf72* HRE carriers, sALS patients and controls. In all cases isoform V2 is the dominant isoform. **B:** Comparison of *C9orf72* isoform expression in all tested areas in controls. V2 is also the dominant isoform in all brain regions tested. **C:** Taqman qRT-PCR showing significant downregulation of the dominant V2 isoform in C9orf72 carriers (of all clinical subtypes) compared to sALS patients in frontal cortex and cerebellum (* = $p<0.05$, Dunn's multiple comparisons correction). Nonsignificant trends were observed in the hippocampus and spinal cord. **D:** V1 isoform expression shows no statistical differences between *C9orf72* HRE carriers, sALS cases and controls in all assessed brain areas. **E:** V3 isoform expression shows no statistical differences between *C9orf72* HRE carriers, sALS cases and controls in all assessed brain areas.

3.2.3 *C9orf72* transcript downregulation does not require hypermethylation of CpG islands upstream of the repeat

Assessment of CpG island methylation upstream of the hexanucleotide repeat was carried out using a published technique that relies on methylation specific restriction enzymes (E. Y. Liu *et al.*, 2014). Restriction digestion was

performed with either Hha1 or Hpa2 restriction enzymes followed by quantitative PCR. The two enzymes gave highly concordant results (Figure 3.4C, $R2 = 0.83$, $p < 0.0001$). Samples with control methylation results above 20% were discarded. There was a high degree of variability in the results using this method (Figure 3.4D), with significant methylation in controls, sporadic ALS patients and *C9orf72* HRE carriers.

Given these variable results, *C9orf72* promoter methylation was also carried out using direct bisulphite sequencing (experiments carried out by Dr Andrew Douglas). The upstream CpG island region was amplified by PCR from bisulfite-treated genomic DNA and the purified product directly sequenced from the 5' end. Owing to technical limitations of this method, the number of CpG sites that could be analysed in each sample was between 9 and 25 out of the total 26 located within the amplicon. Methylation was highly correlated between CpG dinucleotides within each sample (standard deviation between 0.5 and 7% for lowly methylated samples and below 20% for highly methylated samples). If 50% methylation in an individual DNA sample equates

Figure 3.4: *C9orf72* upstream CpG island methylation in *C9orf72* HRE carriers, sALS and controls

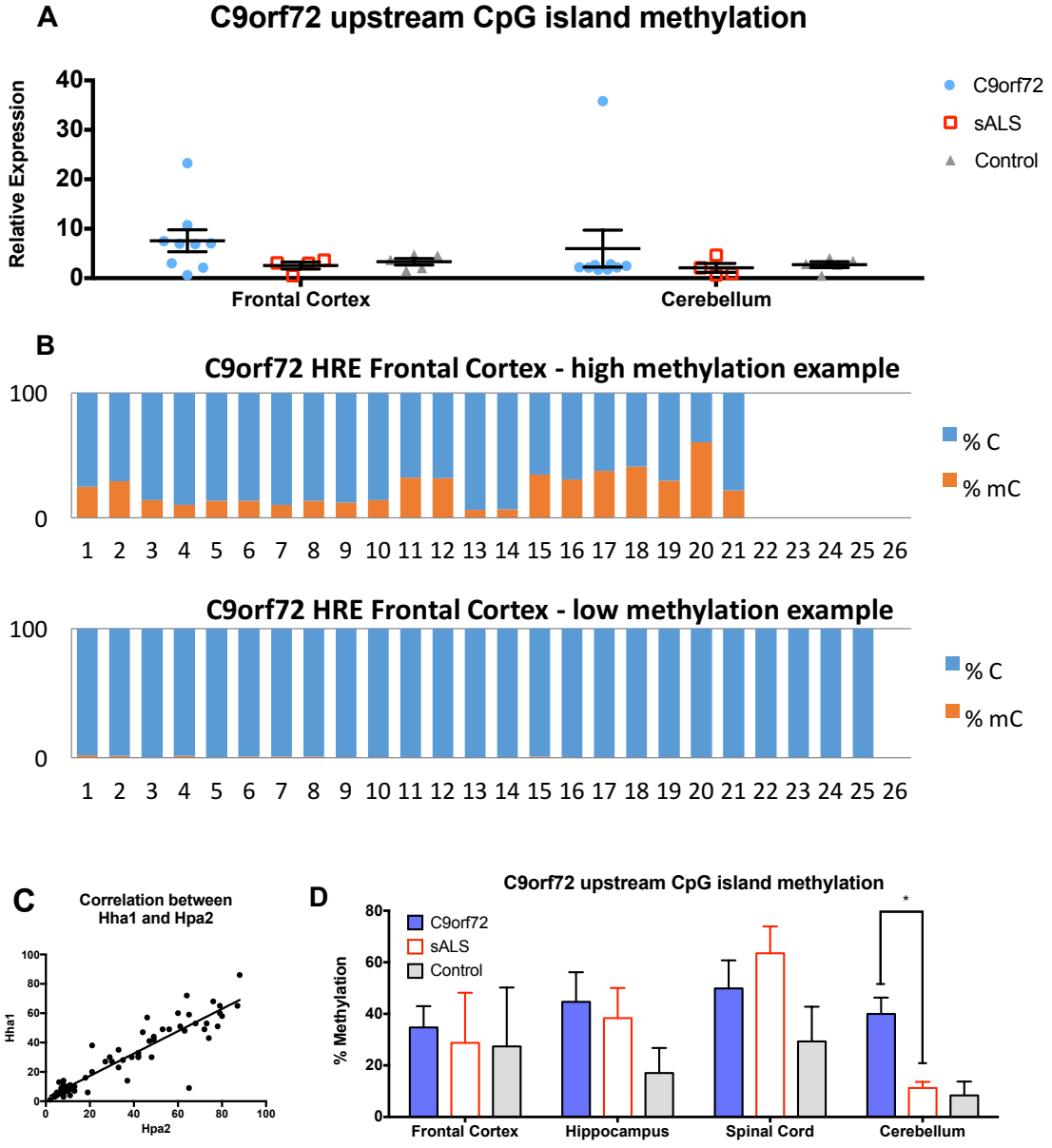

Figure 3.4: A: Representative traces of CpG island methylation of a highly methylated and lowly methylated *C9orf72* expansion carrier, **B:** Mean methylation of all CpG islands assessed by bisulphite sequencing shows a single hypermethylated *C9orf72* HRE carrier but no overall differences between the groups in either tissue. **C:** Significant correlation between two independent methylation-sensitive restriction assays (Hha1 and Hpa2) for the CpG island upstream of the *C9orf72* gene. R^2 = 0.83, $p < 0.0001$. **D:** Mean methylation from both Hha1 and Hpa2 in *C9orf72* expansion carriers, sporadic ALS patients and controls methylation values discordant with bisulphite sequencing. A significant difference in methylation of the cerebellum was found using this method when comparing sporadic ALS and *C9orf72* HRE carriers (* = $p<0.05$, Dunn's multiple comparison's correction).

to complete methylation of the *C9orf72* allele carrying the HRE, classified DNA methylation levels were classified into low (up to 12.5%), intermediate (12.5%-25%) and high (over 25%) categories. Using this classification, low levels of methylation were observed across all patients with one exception. This was an ALS patient carrying the *C9orf72* HRE, who had a high mean methylation level (35%) in the cerebellum and intermediate mean methylation level (23%) in the frontal cortex. However, overall there were no significant differences in mean CpG methylation level between *C9orf72* HRE, sALS or control patient groups (Figure 3.4A). Overall, there were no significant differences in methylation between *C9orf72* positive patients, sALS patients and controls in both brain areas studied. To further characterise methylation changes throughout the genome DNA from four sporadic ALS cases, four *C9orf72* HRE carriers with ALS and four controls was analysed using the 450K methylation platform (ThermoFisher) to compare methylation in the cerebellum and frontal cortex. The expected tissue (21604 differentially methylated probe sites) and sex differences (2416 differentially methylated probe sites) were observed, and accounted for the great majority of the variance in this dataset. The dominance of these two phenotypes is replicated in the multidimensional scaling (Figure 3.5A) and hierarchical clustering (Figure 3.5B), both of which show that the clear majority of the variance is accounted for by sex and tissue type. Once sex chromosome probes were discarded there were no major differences between controls, sporadic ALS cases and *C9orf72* hexanucleotide expansion carriers in both frontal cortex and cerebellum. To avoid any false positives due to differences on autosomal

chromosomes driven by sex, only male samples from *C9orf72* positive patients and sporadic ALS cases were used in the final analysis (Figure 3.4C). Six probe sites were differentially methylated between sALS cases and controls in the frontal cortex and 15 probe sites were differentially methylated in the cerebellum, with no differential methylation observed on chromosome 9 (Tables 3.1 and 3.2). All 6 differentially methylated sites in the frontal cortex were also differentially methylated in the cerebellum.

The results from the 450K analysis also supported the findings from the bisulphite sequencing study. The probes which fell into the bisulphite PCR amplicon were cg1436378, cg1161387 and cg2307474 and corresponded to CpGs 6, 4, 20 and were not methylated.

Figure 3.5: Analysis of global methylation changes in *C9orf72* HRE carriers and sALS patients using 450k bead-chip

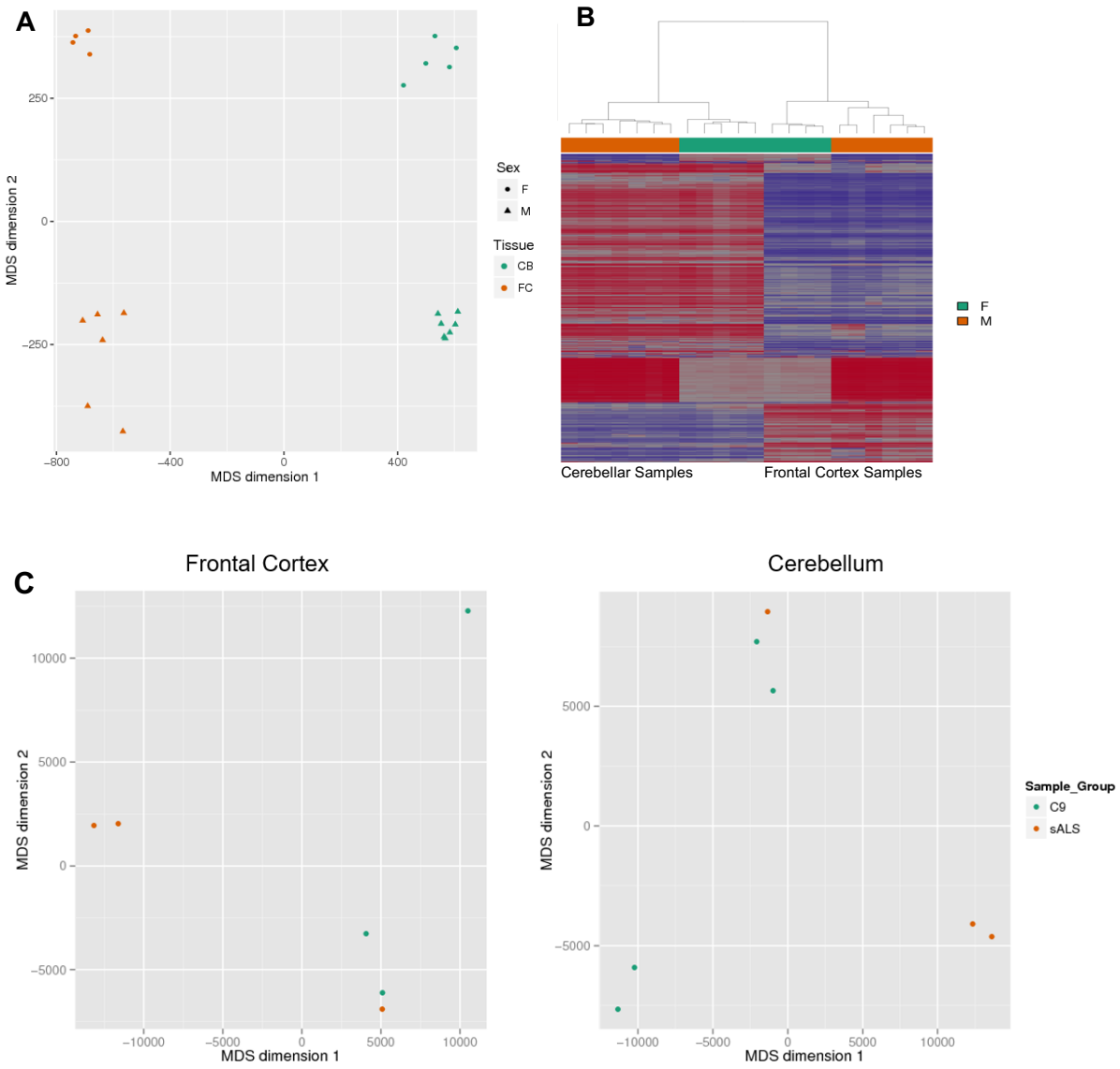

Figure 3.5: A: Multidimensional scaling of all 22 samples by CpG island methylation data shows separation of samples by tissue in the first dimension and by sex in the second dimension. **B:** The 1000 most differentially methylated CpG sites displayed in a heat map with unsupervised hierarchical clustering, show the predominance of tissue related variance in this cohort, followed by unsupervised segregation by sex. Each column represents a single sample from a patient. **C:** Multidimensional scaling of male *C9orf72* HRE compared with males sporadic ALS cases shows no convincing clustering of samples with inter-group differences being comparable to intra-group variance

Table 3.1: 450k results in cerebellum

Probe	Gene	Feature	Chromosome	Start	Strand	mean.C9	mean.sALS	mean.diff	diffmeth.p.val	max.C9	min.C9	sd.C9	max.sALS	min.sALS	sd.sALS	diffmeth.p.adj.fdr
cg01991743	HLA-DPB1	Body	chr6	33053608	-	0.53255012	0.93514526	-0.4025951	2.8205E-09	0.54645425	0.52196659	0.01161991	0.93736073	0.93106192	0.00354055	0.000657159
cg02479782			chr2	71033156	-	0.08616547	0.56757831	-0.4814128	6.94213E-08	0.10109941	0.06511116	0.01712859	0.58361453	0.54549181	0.0197684	0.004621346
cg05176970	NXN	Body	chr17	724273	-	0.32390579	0.86092051	-0.5370147	5.86097E-08	0.36672494	0.27280937	0.04042019	0.87860763	0.84919843	0.01558562	0.004551894
cg05757385	DMTF1	Body, 5'UTR,S_Shore	chr7	86782435	+	0.51586953	0.03775466	0.47811487	2.76795E-10	0.5243235	0.50232856	0.00946688	0.03889804	0.0366098	0.00114412	0.000128983
cg06855422	GRID2	Body	chr4	94566151	+	0.6567967	0.88744411	-0.2306474	1.01743E-06	0.671497	0.64590071	0.01203351	0.90374447	0.87252469	0.01565564	0.038296974
cg13824456			chr7	13837775	+	0.65913236	0.94498737	-0.285855	3.99844E-08	0.6841366	0.61781372	0.02935007	0.94729381	0.94043551	0.00394214	0.003726443
cg16963093	ATXN7L1	Body	chr7	105399252	-	0.9394996	0.52630154	0.41319807	3.8615E-08	0.95098179	0.92212853	0.01226695	0.5327807	0.51540237	0.00949494	0.003726443
cg18391209	CAPN8	Body, S_Shelf	chr1	223747670	+	0.2690462	0.88139326	-0.6123471	1.21878E-06	0.37519004	0.22148829	0.07141182	0.90552192	0.8332934	0.04165577	0.038296974
cg21859421	FAM63A	3'UTR	chr1	150969592	+	0.83669359	0.44255613	0.39413747	1.16428E-06	0.87387728	0.79729194	0.03141466	0.4614775	0.41630772	0.02345936	0.038296974
cg25385150	RADIL	Body	chr7	4870012	-	0.02803994	0.23824586	-0.2102059	2.78835E-07	0.03308603	0.02339106	0.00396993	0.28263267	0.188238	0.04744772	0.014437062
cg26207657			chr10	112207327	-	0.04712509	0.41508998	-0.3679649	3.44885E-07	0.04884511	0.04485146	0.00170554	0.49307538	0.31470228	0.09127233	0.01607118
cg26980937			chr14	56991775	-	0.804706	0.48785011	0.31685589	1.23277E-06	0.83846277	0.78772934	0.02319046	0.50329394	0.4755459	0.01413794	0.038296974
cg27084028	RSPH10B	Body	chr7	6836436	-	0.57655589	0.03070678	0.5458491	7.79253E-07	0.72950793	0.42038102	0.12653672	0.03463098	0.0278022	0.00352672	0.033011053
cg27481428		Island	chr7	155623690	+	0.49890796	0.91765589	-0.4187479	4.79864E-09	0.51976908	0.48125949	0.01584754	0.92091225	0.91347315	0.00380509	0.000745368
cg27519424	HLCS	5'UTR, Island	chr21	38362420	+	0.30999236	0.03223255	0.27775981	1.48397E-07	0.3387686	0.28498244	0.02752448	0.03830288	0.02280638	0.00827536	0.008643859

Table 3.1: Results from differential methylation analysis in the cerebellum, comparing samples from four male *C9orf72* HRE carriers and three sporadic ALS patients. A total of 15 samples were found to be differentially methylated (Benjamini-Hochberg correction for multiple testing, FDR < 0.05)

Table 3.2: 450k results in frontal cortex

Probe	Gene	Feature	Chromosome	Start	Strand	mean.C9	mean.sALS	mean.diff	diffmeth.p.val	max.C9	min.C9	sd.C9	max.sALS	min.sALS	sd.sALS	diffmeth.p.adj.fdr
cg01991743	HLA-DPB1	Body	chr6	33053608	-	0.53853642	0.93535105	-0.3968146	5.622E-08	0.54697482	0.52820414	0.00952757	0.93763844	0.93113235	0.00365784	0.013099071
cg05176970	NXN	Body	chr17	724273	-	0.34068042	0.86069218	-0.5200118	2.6323E-07	0.36695442	0.31236698	0.02735081	0.87828891	0.84969568	0.01539694	0.030666166
cg05757385	DMTF1	5'UTR, Body, S_Shore	chr7	86782435	+	0.51529318	0.03716687	0.47812631	7.9474E-09	0.52602879	0.50246358	0.01192134	0.03850293	0.03599819	0.00126073	0.003703437
cg13824456			chr7	13837775	+	0.65765728	0.94502923	-0.287372	5.19E-07	0.68875633	0.62144488	0.0339458	0.94745257	0.94030238	0.00409402	0.040308769
cg16963093	ATXN7L1	Body	chr7	105399252	-	0.9388168	0.52422997	0.41458684	4.9783E-07	0.95116487	0.9230509	0.01436524	0.53014049	0.51369185	0.00914879	0.040308769
cg27481428		Island	chr7	155623690	+	0.50004335	0.91856148	-0.4185181	9.2856E-08	0.5201825	0.48190016	0.01921906	0.92221442	0.91422788	0.00403654	0.014423379

Table 3.2: Results from differential methylation analysis in the frontal cortex, comparing samples from three male *C9orf72* HRE carriers and three sporadic ALS patients. A total of 6 samples were found to be differentially methylated (Benjamini-Hochberg correction for multiple testing, FDR < 0.05)

3.3 Discussion

This case series of patients with the *C9orf72* expansion supports many of the recent findings from neuropathological studies. This study demonstrates a correlation of the burden of TDP-43 pathology with clinical subtype within the ALS-FTD spectrum, both using semi-quantitative immunohistochemistry for total TDP-43 and quantitative immunofluorescence for pTDP-43. The results support the findings from other *C9orf72* clinicopathological studies, that have established a firm regional correlation between TDP-43 pathology and clinical phenotype in the *C9orf72* ALS/FTD spectrum (Baborie et al., 2015; Mackenzie et al., 2015).

This study also confirms the relative absence of dipeptide inclusions from lower motor neurons, as well as their distribution throughout the brain, which has previously been reported (Mizielinska *et al.*, 2013; Baborie *et al.*, 2015; Cooper-Knock, Higginbottom, *et al.*, 2015; Mackenzie *et al.*, 2015). While dipeptide species other than GA were not studied, previous publications have suggested GA to be correlated with other dipeptide species (Baborie et al., 2015; Mackenzie et al., 2015). Interestingly, a trend towards increasing GA dipeptide inclusions in *C9orf72* FTD cases compared with *C9orf72* ALS cases in hippocampus and frontal cortex was observed, but this was not significant, but an increased dipeptide load has previously been suggested in FTD patients (MacKenzie *et al.*, 2013). Supporting these results, the relative absence of dipeptide inclusions in lower motor neurons could signify a specific pathophysiological role of dipeptides in the development of cognitive impairment. Furthermore, *C9orf72* repeat carriers with FTD have been reported to

have a number of cognitive deficits and have extensive dipeptide pathology at post-mortem with little TDP-43 pathology (Proudfoot *et al.*, 2014; Vatsavayai *et al.*, 2016). There was no association between number of sense foci and the clinical phenotype. In previous studies, an inverse correlation of frontal cortex foci with age of onset has been observed (Mizielinska et al., 2013). Antisense foci, which have been linked to TDP-43 pathology (Cooper-Knock, Higginbottom, *et al.*, 2015), were not assessed in this study.

C9orf72 variant 2 mRNA downregulation, like dipeptide burden and foci, did not have a regional relationship to clinical pathology, with significant downregulation of *C9orf72* transcripts being seen in the cerebellum. Other studies have shown downregulation of the dominant *C9orf72* variant 2 in frontal cortex or cerebellum (DeJesus-Hernandez et al., 2011; Gijselinck et al., 2012; van Blitterswijk et al., 2015). Other isoforms have been less well studied, but an early study indicated downregulation of transcript variant 1 (Gijselinck et al., 2012) and a further study used qRT-PCR to look at short and long isoforms and showed that both were downregulated in frontal cortex (Waite et al., 2014). A more recent study suggested that transcript variant 1 may be undetectable by qPCR and used nanostring technology to assess the abundance of the transcript, correlating lower V1 values in the cerebellum with decreased survival (van Blitterswijk et al., 2015). The authors also suggested that exon 1a containing transcripts could compensate for the reduction in V2 transcripts. The current study did results do not indicate any significant patterns in V1 and V3 expression. It is notable that relative expression of these isoforms is so low, that meaningful compensation could only occur if there was a

marked upregulation of these isoforms, which was not seen here. Neuropathological studies such as this one can only provide associations of the described *C9orf72*-specific molecular mechanisms with TDP-43 pathology. Here, mechanisms of *C9orf72* pathology that have been implicated in loss of function (downregulation of transcripts) as well as those implicated in gain of function (foci and dipeptides) are both present throughout all areas studied. This study therefore does not lend support for any of the three postulated pathomechanisms leading to TDP-43 pathology and disease.

Promoter methylation has been studied both in relation to reduced *C9orf72* expression as well as a potentially independent phenomenon or marker. Hypermethylation of the promoter has been confirmed by multiple studies, but increasing evidence is emerging that within *C9orf72* cases there are examples of both low and high promoter methylation (Russ et al., 2015). Only a single case of *C9orf72* promoter hypermethylation was found in this cohort using bisulfite sequencing, confirming previous reports from a study using direct bisulfite sequencing which only found hypermethylation in one of ten cases(Belzil *et al.*, 2014). This same study also highlighted an apparent inconsistency between promoter DNA methylation and the presence of histone trimethylation in *C9orf72* patients. Although histone trimethylation was not studied here, information relating to histone modifications and chromatin conformation may shed further light on *C9orf72* regulation and, given the dynamic nature of chromatin, may prove valuable in offering potential therapeutic avenues for exploration.

Of interest is the presence of differences particularly in the cerebellum, where significant differences in dipeptide accumulation, sense foci and V2 transcript downregulation were observed between *C9orf72* HRE carriers and controls in this study. The cerebellum is grossly clinically unaffected in ALS and FTD (Tan et al., 2014). Interestingly, there has been an abundance of reports with more frequent significant results in the cerebellum (Tan et al., 2014; Russ et al., 2015; van Blitterswijk et al., 2015). It is not clear however, if this is due to the cerebellum playing a key role in the pathogenesis in *C9orf72* positive ALS/FTD or if this is simply a reflection of the fact that cerebellar granule cells predominate the signature in the cerebellum causing less interference from variations in tissue composition.

With the advent and greater accessibility of genomic techniques, it has been possible to use sequencing technologies and array-based methods for the investigation of global changes in methylation in the CNS. Previous studies in other neurodegenerative diseases have found small but significant changes in the differential methylation in ALS and Alzheimer's Dementia (Morahan *et al.*, 2009; Bakulski *et al.*, 2012; De Jager *et al.*, 2014). The results generated here do not show major differences in methylation between *C9orf72* cases, sALS controls and age-matched controls. This study did suffer from the limitation that only one of four sALS cases and one of four *C9orf72* cases had TDP-43 inclusions in the prefrontal cortex, though there were no striking differences between the cases with and without pathological prefrontal cortex involvement on exploratory data analysis, despite known divergent differential expression changes and splicing changes previously observed in these areas (Prudencio *et al.*, 2015).

Further neuropathological research into *C9orf72*-positive ALS would benefit from cell specific analysis of expression and methylation, as the cellular make up could affect the result in this and previous studies. It is likely that temporal variations in methylation and expression are highly dynamic and more research will be required to determine what determines high and low methylation cases as well as the wide differences in C9 expression seen in this and previous studies.

Chapter 4: Modelling ALS using iPSC derived motor neurons

4.1 Introduction and Aims

Studies of iPSMNs from ALS patients have thus far concentrated on the pathological features of the cultured cells through phenotypic characterisation as well as the investigation of pathways thought to be affected by the mutation (Almeida *et al.*, 2013; Kiskinis *et al.*, 2014; Zhang *et al.*, 2015; Dafinca *et al.*, 2016). Studies have consistently recapitulated RNA foci (Almeida *et al.*, 2013; Sareen *et al.*, 2013; Haeusler *et al.*, 2014). However more downstream products of the HRE have been more elusive. Dipeptide repeats have been detected on dot blot only (Dafinca *et al.*, 2016; Lopez-Gonzalez *et al.*, 2016), confirming their presence but not their aggregation. Aggregation of p62 has not been consistently demonstrated (Sareen *et al.*, 2013; Dafinca *et al.*, 2016). TDP-43 has been studied in iPSMNs, with one study suggesting that it may be mislocalised as a result of nucleo-cytoplasmic transport defects induced by the *C9orf72* HRE (Zhang *et al.*, 2015), but another study of cortical neurons did not observe such changes (Almeida *et al.*, 2013). Given the relative immaturity of iPSMNs even after prolonged culture (Handel *et al.*, 2016; Ho *et al.*, 2016), questions about the disease relevance of the model are important. The ability to demonstrate at least some of the pathological features of ALS including TDP-43 mislocalisation and alteration would be of great importance to the field.

The aim of this study was to investigate specific pathological features of ALS due to the *C9orf72* HRE in iPSMNs, building on findings from the neuropathological study in Chapter 3. For this, the iPSMN differentiation was optimised and characterised. Firstly, *C9orf72* specific features such as the occurrence of sense RNA foci and inclusions formed from GA dipeptide repeats and p62 were studied in *C9orf72* HRE positive iPSMNs. Secondly, the presence of pathological features of ALS was assessed. This included the presence of TDP-43 mislocalisation, aggregation, hyperphosphorylation and fragmentation. Finally, RNA sequencing analysis was performed comparing a *C9orf72* HRE line with its isogenic control derived using CRISPR-Cas9 gene editing.

4.2 Prior work

The work presented in this thesis builds extensively on detailed characterisation of C9orf72 *HRE* positive iPSMNs (Dafinca *et al.*, 2016) as well as the generation of a CRISPR-Cas9 corrected isogenic control line (Ababneh, 2017). Induced pluripotent stem cells were obtained through skin biopsy from patients with *C9orf72* positive ALS-FTD patients and age-matched controls. Reprogramming was performed at the James Martin Stem Cell Facility and the resulting cell lines passed quality control including pluripotency testing, karyotyping and SNP-chip analysis (Dafinca *et al.*, 2016).

After differentiation, iPSMNs derived from *C9orf72* HRE carriers have reproducible phenotypic differences compared to cell lines from healthy controls. *C9orf72* HRE positive iPSMN have increased susceptibility to apoptosis as evidenced by elevated

Figure 4.1: Prior phenotyping of iPSMNs (figures and work by Dr Ruxandra Dafinca)

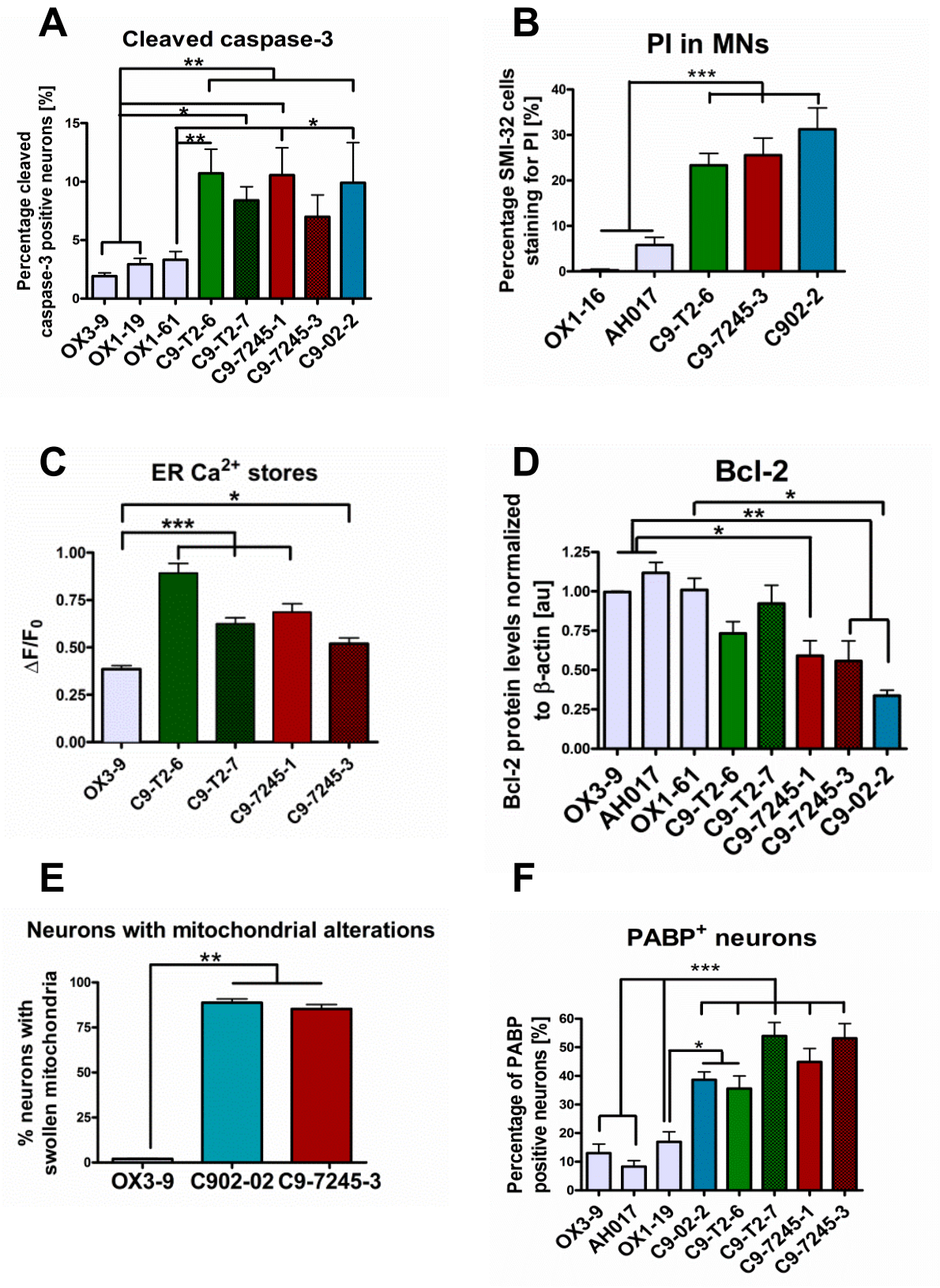

Figure 4.1: Data are represented as mean ± SEM.**A:** Immunostaining for cleaved caspase-3 shows increased frequency of apoptotic neurons in C9orf72 MNs compared to control MNs (*, $p < .05$; **, $p < .01$, one-way ANOVA with Dunnett's post hoc test); n = 3 independent differentiations; m = 400 cells per genotype analyzed for each individual experiment.

Figure 4.1 (continued):

B: Reduced cell viability was detected in *C9orf72* MNs by PI staining (***, p < .001, one-way ANOVA with Dunnett's *post hoc* test) n = 2 independent differentiations. **C:** Measurement of the TG-evoked ER Ca2+ release shows significantly higher ER Ca2+ levels in C9-T2 and C9-7245 lines compared to OX3-9 (***, p < .001; *, p < .05; one-way ANOVA with Dunnet's post hoc test); n = 3 independent differentiations, m = 150–200 neurons recorded per genotype. **D:** Immunoblotting for the antiapoptotic Bcl-2 protein shows downregulation of levels in patient lines (*, p < .05; **, p < .01, one-way ANOVA with Dunnett's post hoc test); n = 3 independent differentiations (n = 2 independent differentiations for AH017-13 and OX1-61), m = 3–4 technical replicates. **E:** Electron micrographs (EM) show mitochondrial alterations in 88% of C9orf72 neurons (**, p < .01, ANOVA with Dunnet's post hoc test). **F:** PolyA binding protein (PABP) positive stress granules form in C9 CNs (***, p < .001, one-way ANOVA with Dunnett's post hoc test).

cleaved caspase-3 levels (Figure 4.1A, Dr Ruxandra Dafinca) and propidium iodide (PI) staining (Figure 4.1B, Dr Ruxandra Dafinca). *C9orf72* patient derived iPSMNs also have significantly higher ER calcium stores measured through thapsigargin-induced calcium release (Figure 4.1C, Dr Ruxandra Dafinca). These changes correspond to a reduction in Bcl-2 levels in the cells (Figure 4.1D, Dr Ruxandra Dafinca). Bcl-2 is an important modulator of calcium signalling and an anti-apoptotic protein. There is a marked increase in abnormal mitochondrial morphology on electron microscopy in *C9orf72 HRE positive* iPSMNs compared to healthy controls (Figure 4.1E, Dr Ruxandra Dafinca). Finally, up to a fivefold increase in the frequency of neurons with PABP-positive stress granules was observed in *C9or72* HRE positive iPSMNs compared to controls (Figure 4.1F, Dr Ruxandra Dafinca).

These experiments were followed by the successful generation of an isogenic control line from one *C9orf72* patient line (Ababneh, 2017) using clustered regularly interspaced short palindromic repeats (CRISPR) technology. Excision of the repeat was done using a double nicking approach using two guide RNAs, ensuring increased genomic specificity. A donor template containing the 2 repeats was inserted together with a puromycin selection cassette flanked by LoxP sites. Following successful integration and screening using puromycin, the selection marker was removed using Cre recombinase (Figure 4.2A, Dr Nida'a Ababneh). The absence of the repeat in the edited line was confirmed using repeat primed PCR and direct capillary sequencing (Figure 4.2B, Dr Nida'a Ababneh).

Figure 4.2: Design and confirmation of CRISPR/Cas9 isogenic control line (figures and work by Dr Nida'a Ababneh)

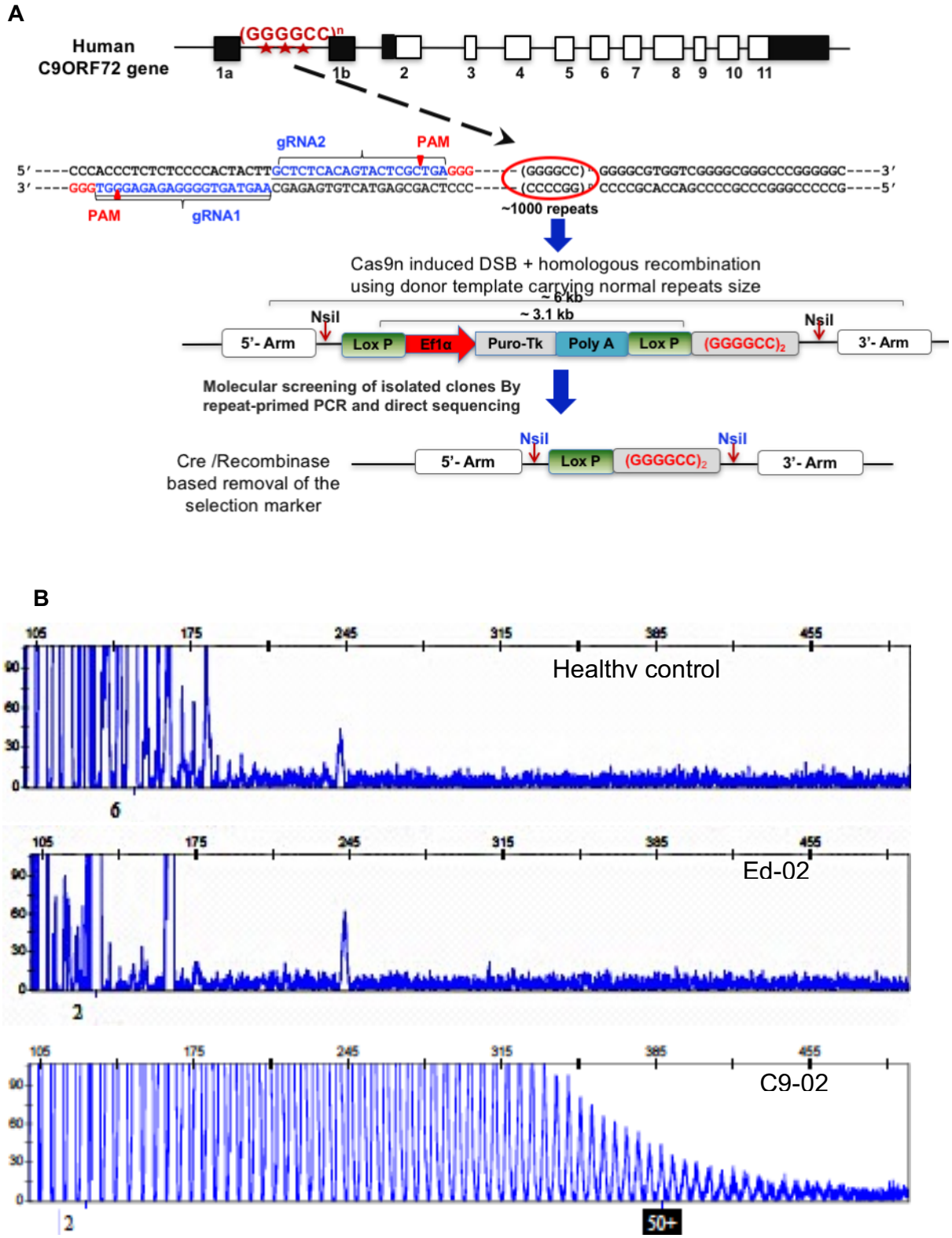

Figure 4.2: A: Experimental method for CRISPR/Cas9 excision of repeat. gRNA: guide RNA, DSB: double strand breaks PCR: polymerase chain reaction. **B:** Confirmation of absence of repeats in CRISPR/Cas9 edited lines using repeat primed PCR.

4.3 Results

4.3.1 Optimisation and characterisation of a protocol for motor neuron differentiation

The efficient, reliable and reproducible generation of motor neurons from iPSCs is a key prerequisite for ALS research. The starting point for this optimisation was a recently published motor neuron disease protocol (Qu *et al.*, 2014), which, while generating 50% motor neurons after 30 days of differentiation, had two major shortcomings. Firstly, significant cell death was observed in the early stages of the protocol, and secondly there was ongoing proliferation in the later stages of the protocol which was paradoxical given the postmitotic condition of motor neurons, implying that these cells were motor neuron precursors.

To improve on this, insights from a combinatorial and systematic approach to motor neuron differentiation (Maury *et al.*, 2014) were introduced, while maintaining cells in monolayer culture. The first step included the introduction of a new medium containing Neurobasal medium and β-mercaptoethanol. The timings of SAG and RA administration were adjusted and SMAD inhibition by compound C was supplemented with Wnt activation using Chir99021. Prolonged incubation with these modifications yielded the desired increases in OLIG2 concentration (Figure 4.3A). Given the near-ubiquitous OLIG2 expression on day 16, DAPT was added at this stage, accelerating the neuronal maturation and conversion to postmitotic neurons and achieving a static culture, with high neuronal content as visualised using fluorescent (Figure 4.3B) and light microscopy (Figure 4.3C).

Figure 4.3: Characterisation of iPSMN differentiation protocol

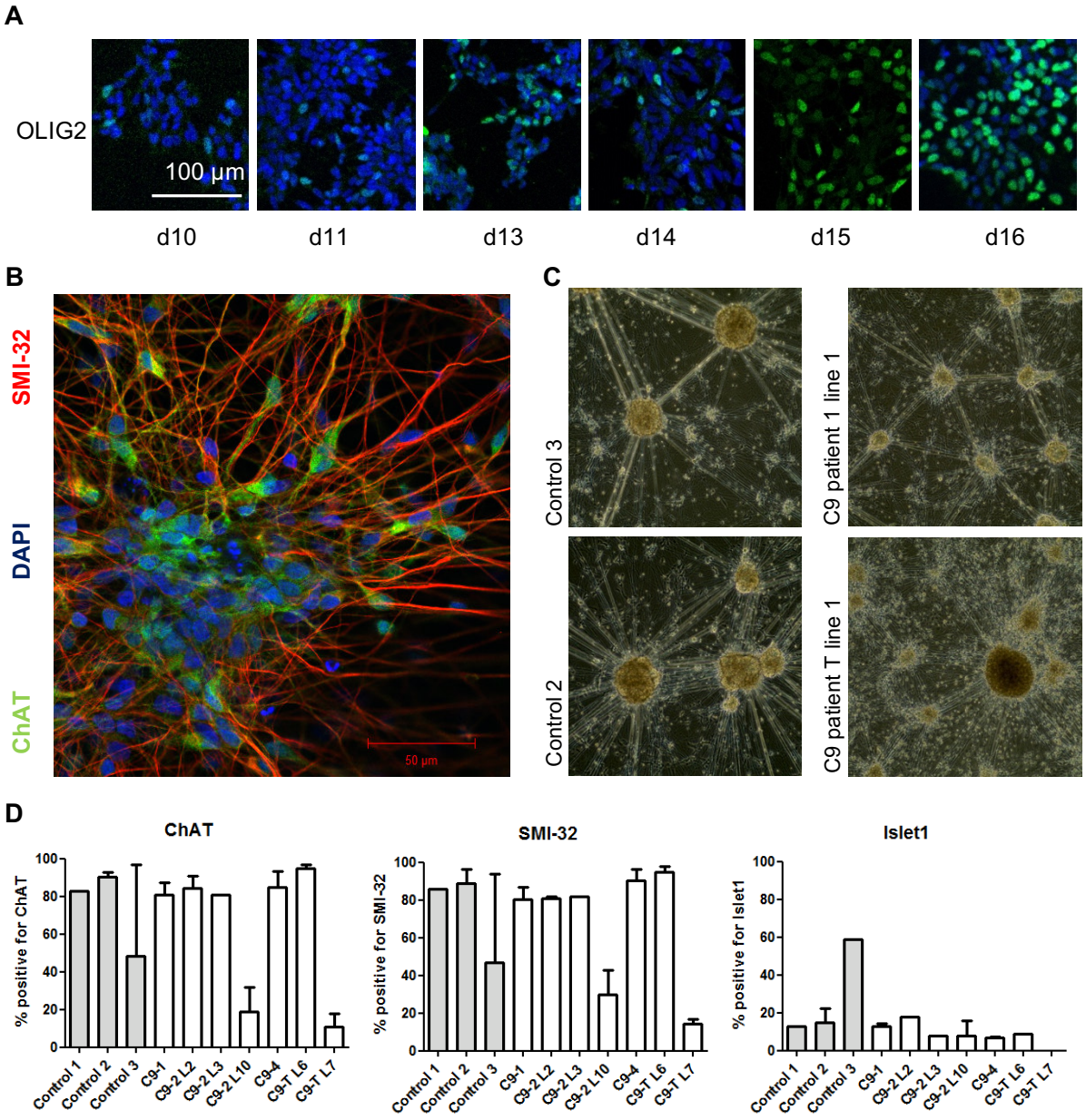

Figure 4.3: Motor neuron differentiation efficiency. **A**: Representative consecutive immunostaining for Olig2 staining from patient C9-2 line 2 showing increasing OLIG2 proportions with time. **B**: Example immunofluorescence staining at day 35 from line C9-4. **C**: Panel of four experiments showing light microscopic image of differentiation. **D:** Quantified differentiation efficiency at day 32-40 from four independent differentiations for ChAT, SMI-32 and Islet1.

The differentiation was quantified using the markers ChAT, SMI-32 and Islet-1, showing high proportions of the mature motor neuron markers ChAT and SMI-32, but low proportions of Islet-1. Between lines, differentiation was consistently successful with more than 80% of cells expressing ChAT, but there were both control and patient lines that consistently did not yield successful differentiations, and these lines were not used in subsequent experiments.

4.3.2 iPSMNs from *C9orf72* patients have sense RNA foci but no dipeptide aggregates

To assess the presence of pathological hallmarks of the *C9orf72* HRE expansion, the frequency of sense RNA foci was investigated using a Cy3 tagged repeat RNA oligonucleotide probe. Approximately 30% of patient iPSMNs had RNA foci, which were not seen in control iPSMNs (Figures 4.4A and 4.4B). Most cells with RNA foci had one focus, but some cells had up to three foci in this study (Figure 4.4C). RNA foci were confirmed to be sensitive to incubation with RNases, but were resistant to DNase treatment (Figure 4.4D). GA dipeptide inclusions, readily detected in hippocampal and cortical neurons at post-mortem, were not detected in iPSMNs, even after treatment with the proteasomal inhibitor MG-132. This was confirmed by the absence of p62 inclusions in the same cells (Figure 4.5B). p62 staining intensity increased in both control and patient iPSMNs, and occasional circumscribed areas of particulary high p62 staining were observed in all cell lines.

Figure 4.4: Sense RNA foci in *C9orf72* HRE iPSMNs

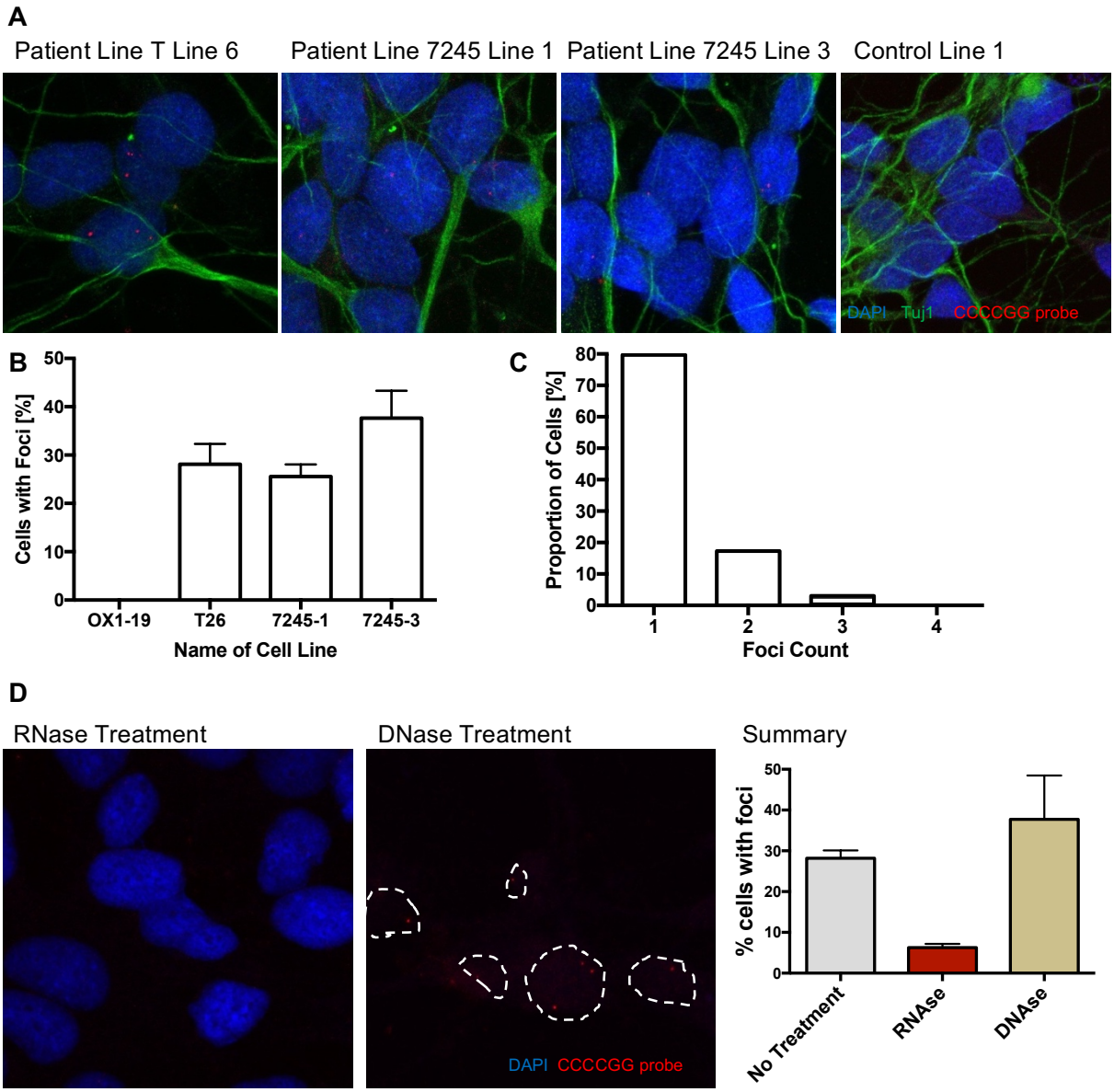

Figure 4.4: Sense RNA foci are present in *C9orf72* HRE positive motor neurons. A: Representative fluorescent in situ hybridisation images from three patient lines and one control line. B: Proportion of Tuj1-positive neurons with RNA foci in cell lines OX1-19 (control) and T26, 7245-1, 7245-3 (*C9orf72* positive cases). Cell count >120 cells, five fields of view per cell line. C: Frequency of foci per cell in Tuj1-positive neurons in all *c9orf72*-positive cell lines. Cell count: 126 cells. D: Representative images following RNase and DNase treatment. Summary shows proportion of cells with RNA foci in cell lines T26, 7245-1, 7245-3 (*C9orf72* positive cases) with and without treatment. Cell count >150 cells.

Figure 4.5: GA dipeptide and p62 immunofluorescence in *C9orf72* HRE iPSMN

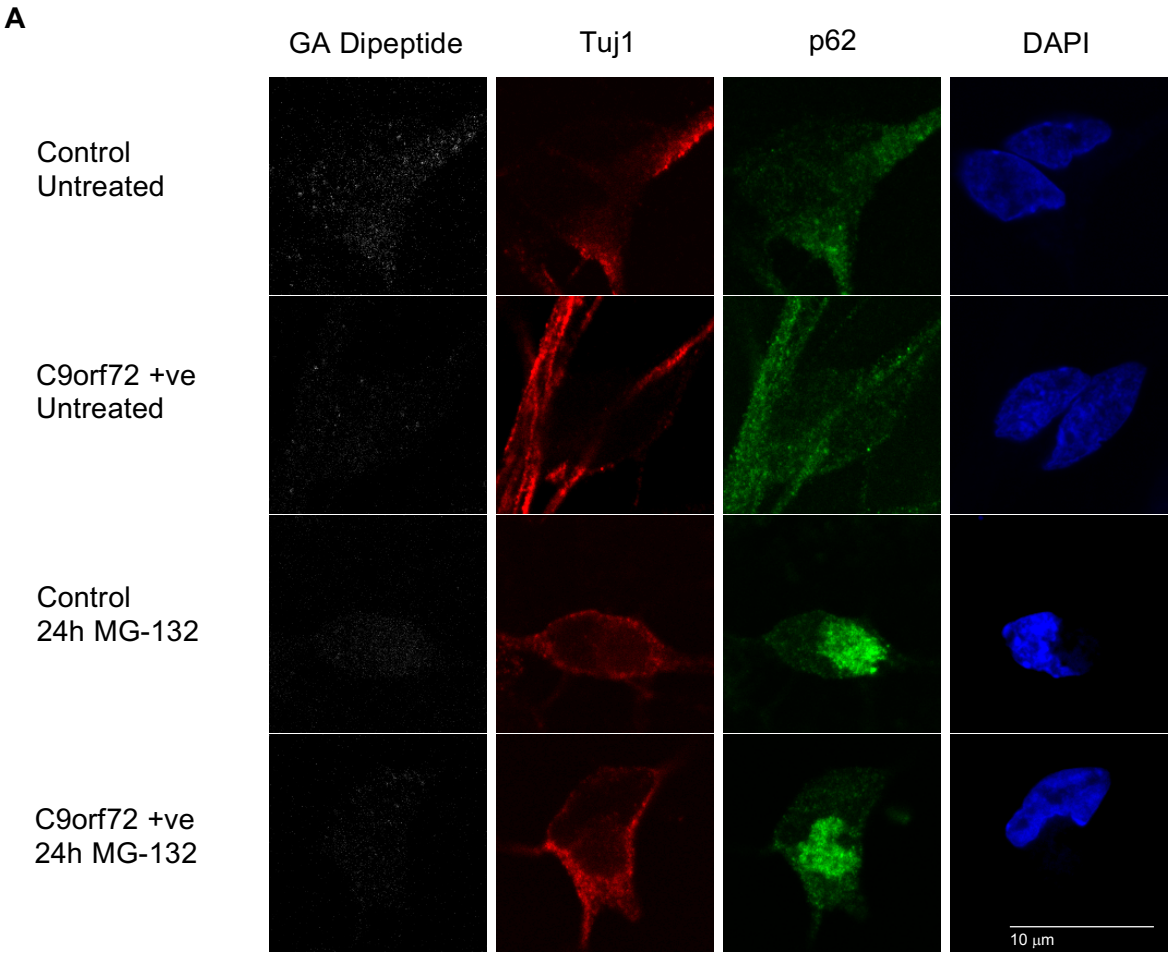

Figure 4.5: No GA dipeptide staining observed in iPSMNs at rest or following 24 hours of MG132 treatment. Following 24 hours of MG132 treatment, the intensity of cytoplasmic p62 staining increases both in control and patient cell lines. Areas of particularly high staining intensity can occasionally be seen in both conditions.

Figure 4.6: TDP-43 and pTDP-43 immunofluorescence in *C9orf72* HRE iPSMNs

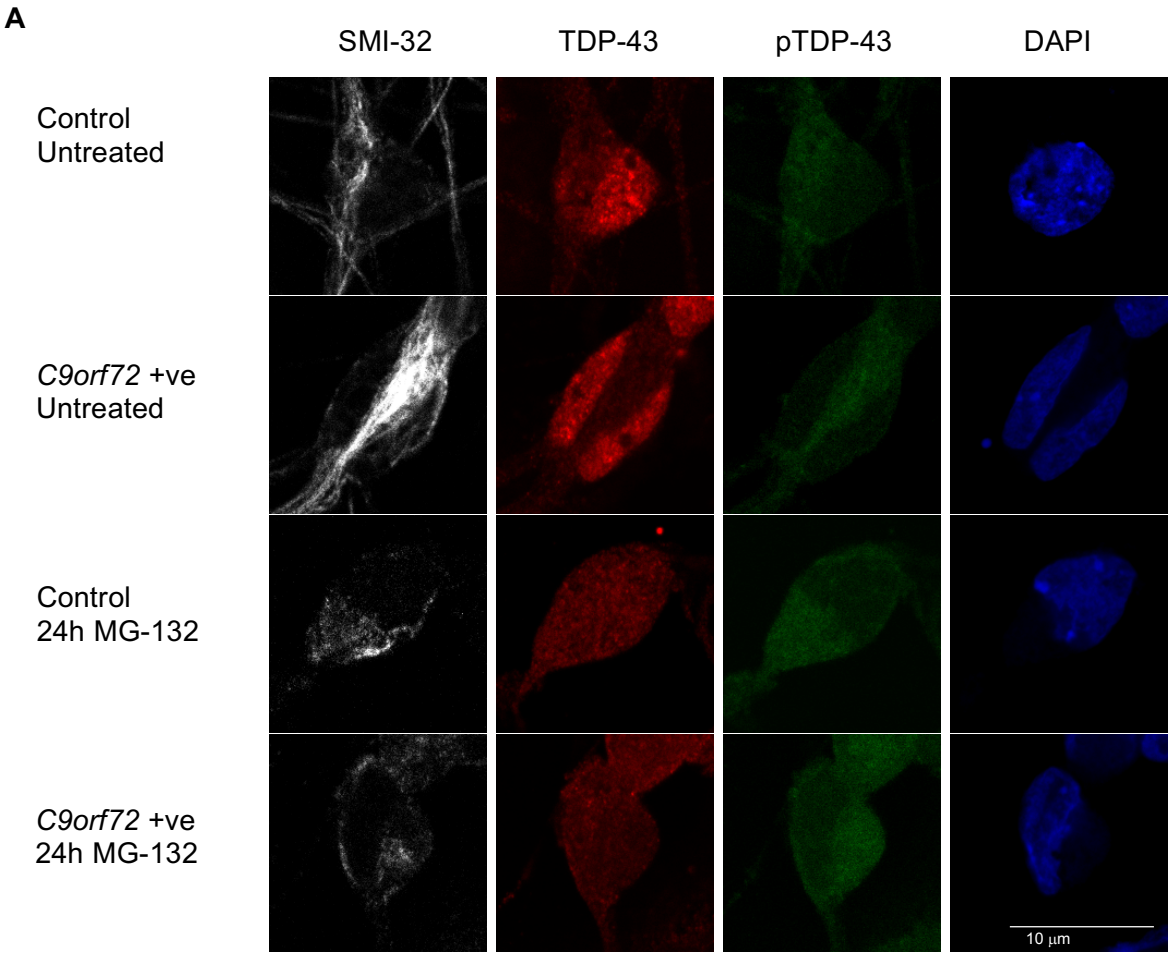

Figure 4.6 (continued)

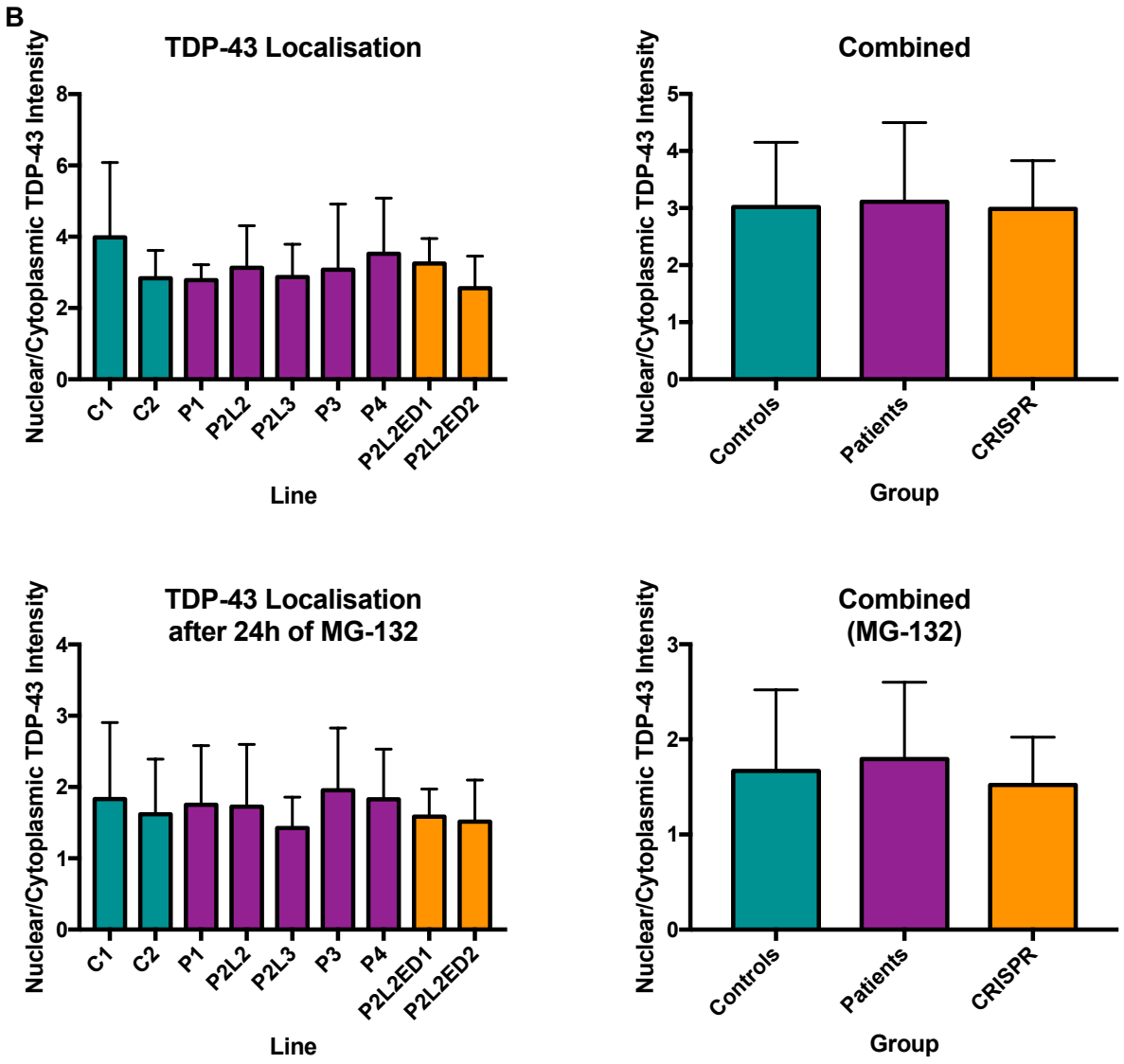

Figure 4.6: No abnormality of TDP-43 can be detected in iPSMNs at rest or following 24 hours of MG-132 treatment. **A**: Representative fluorescent images from control and patient lines showing TDP-43 and phospho-TDP-43 during control conditions and after 24 hours of MG-132 in SMI-32 positive neurons. No TDP-43 and phospho-TDP-43 inclusions are seen. Mislocalisation of TDP-43 from the nucleus to the cytoplasm is seen to an equal degree in control and patient neurons following 24 hours of MG-132 treatment. Image size: 135 micrometers **B**: Quantification of TDP-43 mislocalisation by a ratio of nuclear to cytoplasmic intensities shows no differences between control, patient and CRISPR corrected lines both at rest and following MG-132 treatment. Data was obtained through three independent differentiations. A mean of 20 neurons was analysed per line per condition (range 6 to 38). Combined plots contain >50 neurons per condition. Error bars: standard deviation.

4.3.3 TDP-43 is not mislocalised or aggregated in iPSMNs from *C9orf72* patients

iPSMNs were then analysed for the presence of TDP-43 inclusions, which are the hallmark of ALS and are readily seen in motor neurons of *C9orf72* HRE positive patients. No aggregation of TDP-43 was seen in the C9orf72 HRE positive lines. In line with this finding, there was also no visible phospho-TDP-43 staining in these cells, even after proteasomal inhibition with MG-132 (Figure 4.6A). Quantification of the nuclear cytoplasmic ratio of TDP-43? in *C9orf72* HRE positive cells showed no difference nuclear/cytoplasmic ratio in patient cells compared to controls (Figure 4.6B).

4.3.4 Insoluble TDP-43 is post-transcriptionally modified in *C9orf72* patients and controls

No significant abnormalities can be seen in the soluble (in Triton X-100) TDP-43 fraction of *C9orf72* positive iPSMN compared to control iPSMNs (Figure 4.5A). TDP-43 protein levels were increased following 24 hours of MG-132 treatment compared to untreated cells, but again no differences were observed between the two genotypes. An SDS-soluble fraction was obtained from the Triton-X insoluble fraction. Following treatment, control as well as patient lines contained a hyperphosphorylated TDP-43 band as well as 35 and 25 kDa bands (Figure 4.7A).

Figure 4.7: Soluble and insoluble TDP-43 in *C9orf72* iPSMNs

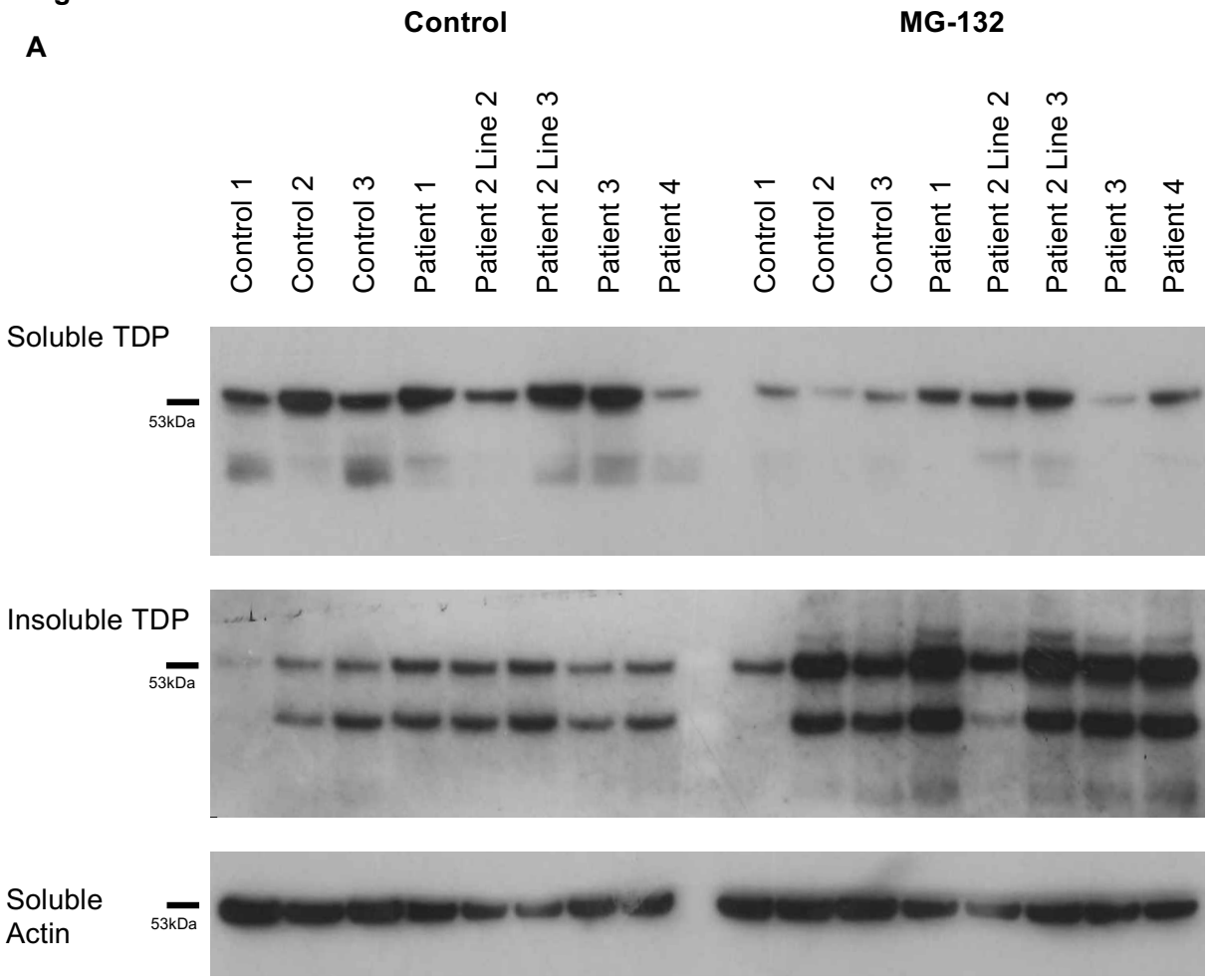

Figure 4.7 (continued)

B

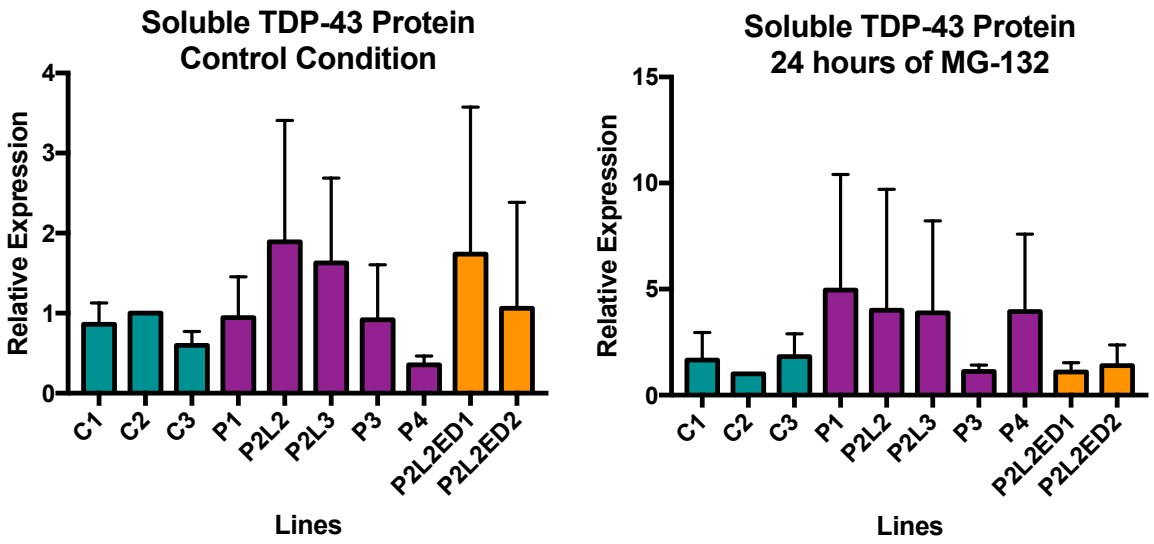

Figure 4.7: The SDS soluble fraction contains multiple TDP-43 species in iPSMNs. **A:** Example Western blot showing TDP-43 in control and patient cell lines with and without MG-132 treatment. In the soluble fraction TDP-43 levels are variable, the insoluble fraction demonstrates the presence of a 35kDa fraction at rest condition and the additional appearance of a hyperphosphorylated band and a faint 25kDa band following MG-132 treatment. The additional bands are seen both in control, patient and CRISPR corrected lines. **B:** Quantification of soluble TDP-43 in control conditions and following MG-132 treatment. Soluble actin was used as loading control. No significant differences were observed between conditions and there was high variability of TDP-43 levels across four differentiations studied. Error bars: standard deviation.

4.3.5 208 genes are differentially expressed between the CRISPR/Cas9 corrected line and its *C9orf72* HRE

To investigate pathways related specifically to the hexanucleotide expansion, the *C9orf72* patient line 2 and its isogenic CRISPR/Cas9 control were analysed by RNA sequencing. Cells were harvested at day 30 and replicates were generated by independent differentiations. RIN values of extracted RNA were above 9.5, and RNA sequin spike-ins were added to the RNA prior to library preparation. Five replicates of the patient line and three replicates each from two independently generated CRISPR/Cas9 corrected clones were successfully sequenced and passed quality control.

Abundance estimation with *Salmon* confirmed a high correlation between sequin concentrations and normalised transcripts per million (Figure 4.8A). No effects of the timing of the differentiation or the person performing it were observed on exploratory data analysis. Principal component analysis did not show separation of the two experimental groups (Figure 4.8B). Differential expression analysis using *DESeq2* found 208 genes differentially expressed between *C9orf72* HRE iPSMNs and their isogenic controls (FDR < 0.05, absolute log2-fold change > 1, Figure 4.8C). Accuracy of differential expression was ascertained using sequins and a limit of detection ratio for log2 fold changes of 1 and above was estimated to be 5 transcripts per million (Figure 4.8D).

Figure 4.8: RNA Sequencing analysis of C9orf72 HRE iPSMNs and isogenic controls

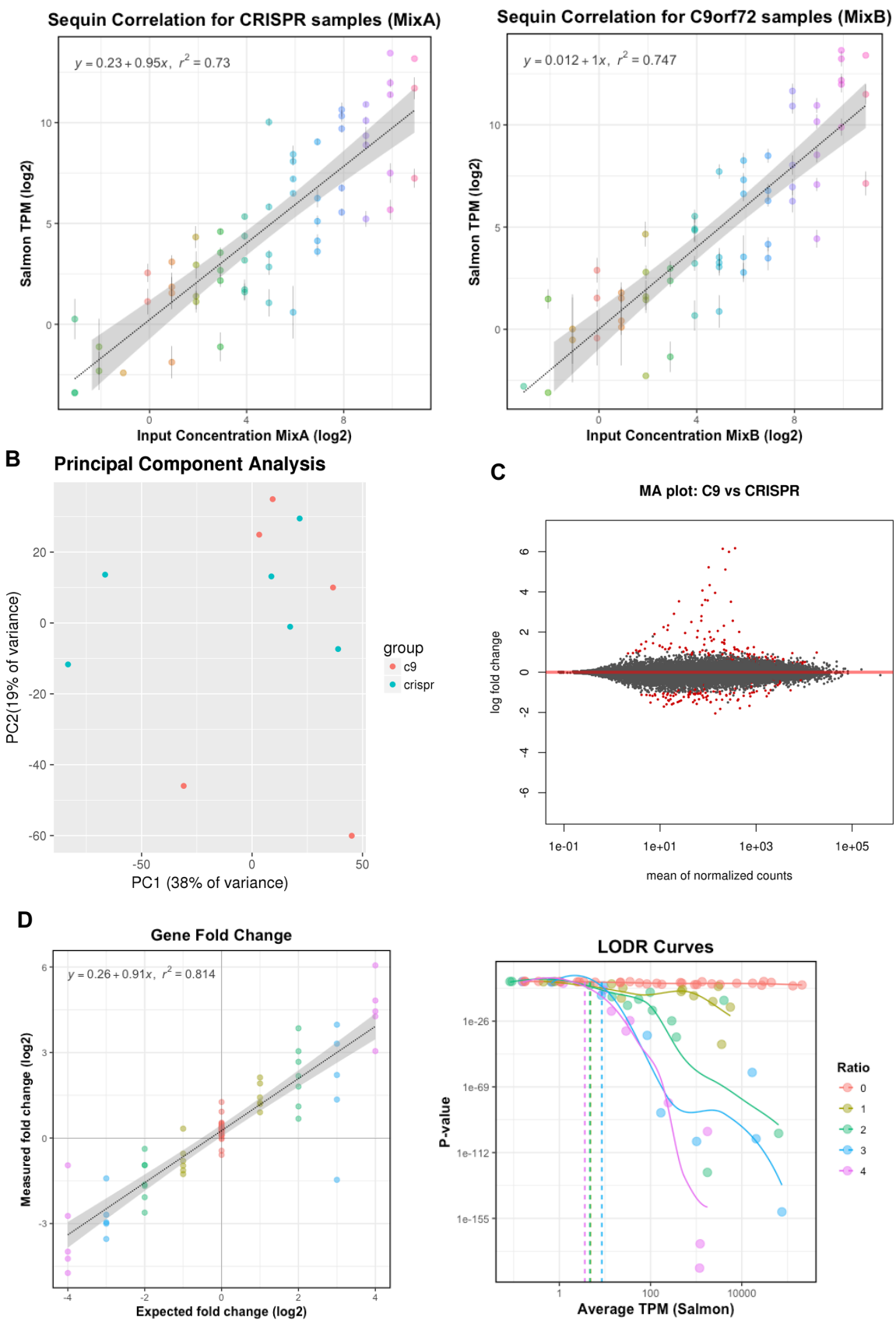

Figure 4.8 (continued)

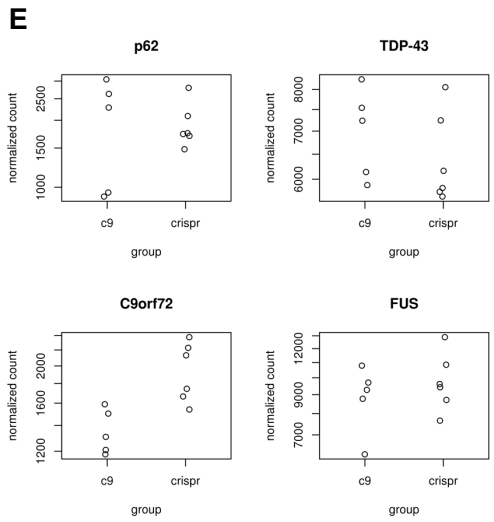

Figure 4.8 (continued): Analysis of RNA sequencing. C9 = *C9orf72* HRE positive iPSMN, CRISPR = isogenic control iPSMNs. **A:** Input concentrations of RNA sequin spike-ins averaged across all samples plotted against log2 scaled transcripts per million (TPM) derived from salmon. **B:** Principal component analysis for submitted samples. **C:** MA plot showing log2 fold change in differentially expressed genes versus normalised counts (measure of relative abundance). **D:** Measured log2 fold difference between the two samples for sequins with known concentrations and level of detection curves for sequins. Ratio: known log2 fold change between the two conditions. **E:** Normalised counts for ALS genes of interest in the two conditions studied.

On exploratory analysis of the results, two classes of genes were noted. Firstly, there was significantly lower expression of 32 zinc finger proteases in the CRISPR compared to the *C9orf72* positive line. Secondly class switching was observed for protocadherins: Protocadherin Gamma Subfamily A 1-3 and 5-7 being more expressed in *C9orf72* positive neurons, whereas PCDHGB 6 and 7 found more abundant in the isogenic controls.

There were some significant differences in key transcription factors involved in neuronal differentiation. *SOX6*, *Nkx6.1*, *PAX8* were more abundant in *C9orf72* HRE positive cells. *PAX8* was very lowly expressed overall. HoxC6 and HoxB13 were more highly expressed in CRISPR isogenic controls. However, *Islet1*, *Hb9* and *ChAT* were not significantly different between the conditions. As reported previously in iPSMNs (Sances *et al.*, 2016), more rostral genes such as *HoxA4-5* and *HoxC6-8* were expressed robustly, whereas lumbar Hox genes *HoxC9-10* were lowly expressed and *HoxC11-12* were not expressed at all.

C9orf72 expression was lower in *C9orf72* HRE positive iPSMNs by 29%, but this was not significant after FDR correction (Figure 4.8E).

4.3.6 Pathways related to transcription and to the extracellular domain particularly enriched in the differentially expressed genes

The robustness of differential expression between the two groups was ascertained using permutation analysis and this demonstrated the experimental and biological relevance of the differential expression result (Figure 4.9A).

Figure 4.9: Pathway analysis of differential expression between *C9orf72* positive iPSMNs and CRISPR controls

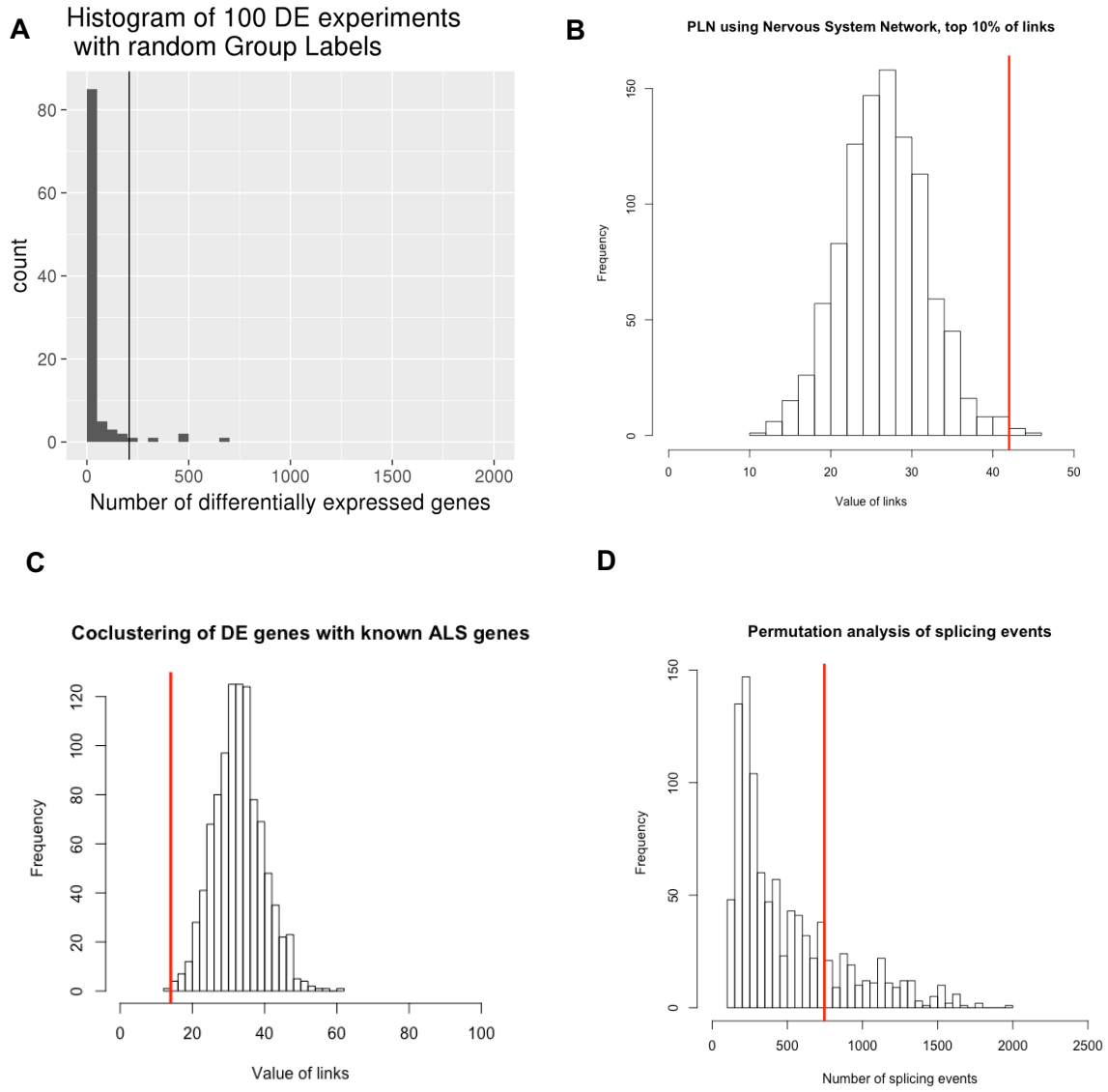

Figure 4.9: A: Number of differentially expressed genes after permutation of sample labels shows that higher differential expression was seen in only 4 of 100 permutations, underlining the biological relevance of the result (p = 0.04). **B:** The 208 differentially expressed genes are functionally related, as demonstrated by Phenotypic Linkage Network Analysis, using a nervous system specific network. The red line shows the strength of links in the dataset, compared to the distribution of link strength for samples randomly drawn from the nervous system network. **C:** The results dataset was not related or functionally similar to the dataset of known genes with ALS mutations. . **Abbreviation:** DE = differentially expressed/differential expression; PLN = Phenotypic Linkage Network

GO enrichment analysis of the results against a central nervous system background revealed only the one term to be significant against a neuronal background:

GO Term	OR	FDR	Type
cell-cell adhesion via plasma-membrane adhesion molecules	7.7683	0	BP

Table 4.1: GO terms for RNA sequencing results against central nervous system background (Hippocampus).
Types of GO term: BP = biological process, MF = molecular function.

To further explore subgroups within the dataset, phenotypic linkage network analysis was performed on the results dataset, using a nervous system specific network constructed from multiple ontology and protein interaction databases (Honti *et al.*, 2014). The network analysis showed that the differentially expressed genes were more strongly linked than a random geneset adjusted for gene length (Figure 4.9B). The differentially expressed genes were not found to be related to genes known to harbour ALS causing mutations (Figure 4.9C).

Subdivision of the differentially expressed genes in the network yielded communities that are more interlinked between themselves than with other members of the network (Modularity score = 0.45). The subdivision enabled ontological analysis of subgroups. Differential expression between the *C9orf72* positive iPSMN and the CRISPR controls was driven by dysregulation of transcription and DNA binding, but also caused by alteration of transcription in genes related to extracellular function, receptor binding and extracellular vesicles. The third notable group was the alteration of transcription in genes related to voltage gated potassium channel transport (Table 4.2).

Subnetwork	Number of Genes	GO enrichment (OR)
0	1	Nucleotide diphosphatase activity (Inf)
1	38	Transcription (7.2), DNA binding (6.9)
2	29	Extracellular region (14.7), receptor binding (10.7)
3	22	Voltage-gated potassium channel activity (67.5), potassium ion transmembrane transport (42.7)
4	2	No enrichment
5	26	extracellular vesicle (10.0)
6	12	No enrichment
7	28	Sequence-specific DNA binding (18.0)
8	6	No enrichment

Table 4.2: Results of Subnetwork Analysis.
OR = Odds Ratio

The dataset was also used to perform differential splicing analysis. Although 746 differential splicing events were discovered between the conditions, it was not possible to determine the significance of these splicing changes. Permutation analysis carried out by randomly swapping the group label for each sample revealed that random alteration of group labels yielded higher number of differential splicing events in a fifth of cases ($p=0.20$, Figure 4.9D).

4.3.7 qRT-PCR validation yields candidate genes and provides insight into CRISPR-specific downregulation of iPSMNs.

qRT-PCR validation of the sequencing was performed on an extended cohort of samples collected over two years of iPSMN differentiations, containing independent samples. A total of six C92-2 samples and 10 CRISPR corrected samples (five 59-1, four 59-2, one 59-3) were available. Eight genes were chosen based on the results

Figure 4.10: Validation of RNA sequencing in iPSMNs

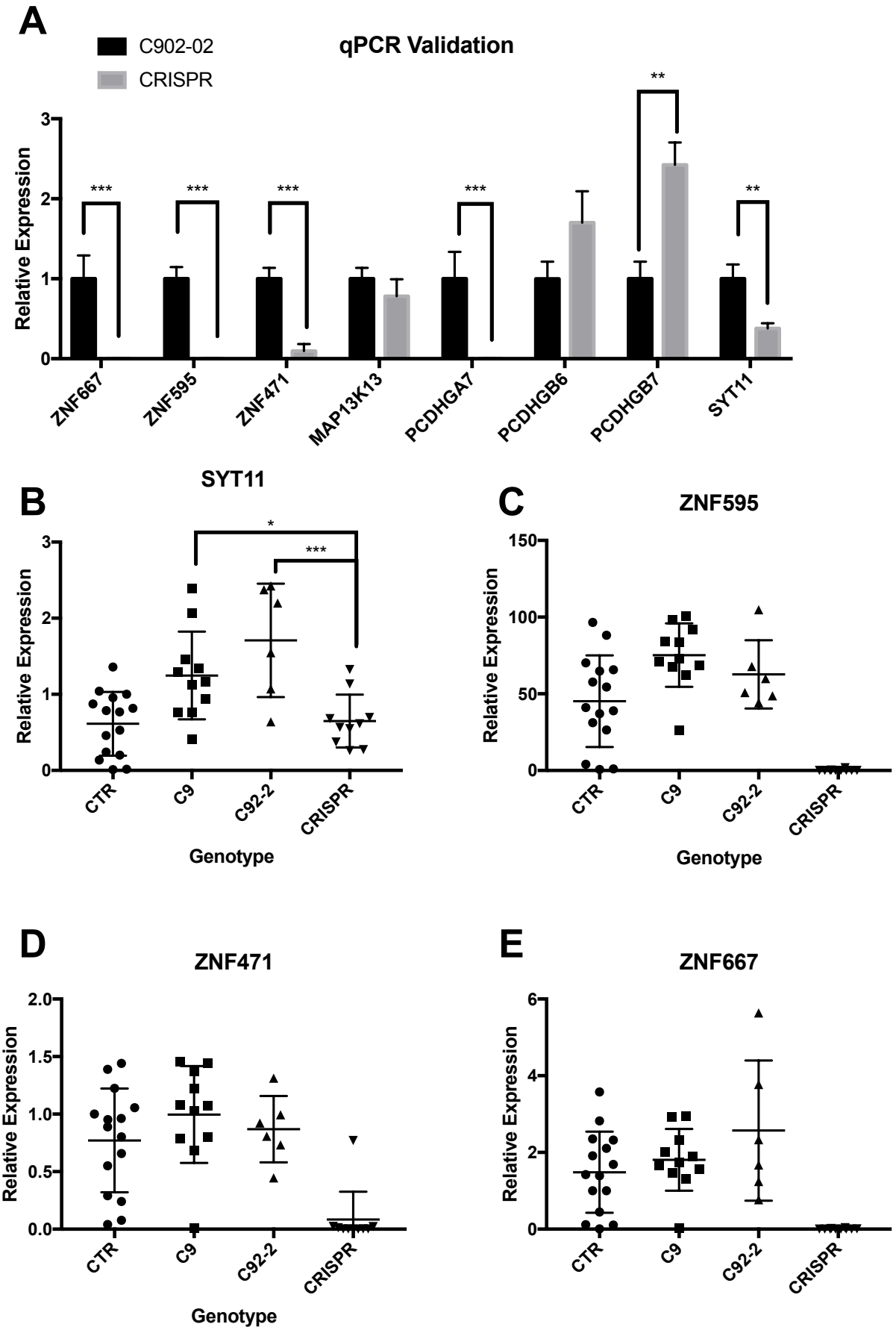

Figure 4.10: A: Results of qRT-PCR validation of RNA sequencing experiment in 8 genes selected on biological relevance and abundance. t-test. **B:** Synaptotagmin 11 (SYT11) is less abundant in CRISPR iPSMNs compared to C92-2(**) and other *C9orf72* HRE cell lines (*). Control iPSMNs also have reduced SYT11 levels compared to C92-2 (***) and to other *C9orf72* HRE positive samples (*), not shown on graph. ANOVA with Holm-Sidak *post hoc* test. **C-E:** CRISPR-specific downregulation of selected zinc finger gene expression. * = $p<0.05$, **p<

and their biological relevance, and six of these were validated by qRT-PCR (Figure 4.10A).

The results were also validated against 11 samples from four different *C9orf72* HRE cell lines as well as against 15 samples from four control lines, which provided important information on the generalisability of the results. Synaptotagmin 11 (SYT11) was less abundant in the CRIPSR and control cell lines compared to both the *C9orf72* HRE cell line of origin, but also to the remaining *C9orf72* HRE positive samples ($p < 0.05$ ANOVA, Holm-Sidak's *post hoc* comparisons test, Figure 4.10B).

Zinc finger proteins on the other hand, were found to be downregulated specifically in all isogenic CRISPR control lines, but none of the remaining lines, suggesting a persistent effect of CRISPR/Cas9 correction, which was replicated in three independent CRISPR cell lines (59-1, 59-2 and 59-3, Figures 4.10C-E).

4.4 Discussion

This study of iPSMNs derived from *C9orf72* HRE carriers concentrated on characterising and assessing for the presence of pathological signatures of C9orf72 ALS observed in post-mortem studies.

The study confirms that sense RNA foci, a direct product of *C9orf72* transcription are easily and consistently detected in iPSMNs, in line with previous studies. While some have reported similar results after foci quantification of 1-3 foci in around 20% of cells (Sareen *et al.*, 2013), others have detected a mean of 7 sense foci in iPSMNs (Donnelly *et al.*, 2013). These differences may result from different thresholds being

selected for quantification and small differences in hybridisation protocols. The results presented here are in line with neuropathological studies that quantified RNA foci using the same hybridisation method (Mizielinska *et al.*, 2013; Cooper-Knock, Higginbottom, *et al.*, 2015).

The absence of dipeptide inclusions in this study is in line with neuropathologic findings post-mortem studies including the earlier findings of this thesis, where spinal motor neurons only rarely show dipeptide accumulation. Other groups who have investigated cortical neurons have not observed dipeptide inclusions in these cells either, suggesting it may be difficult to model this feature of *C9orf72* pathology in a dish (Almeida *et al.*, 2013). Other studies in iPSMNs, presumably under different culture conditions or using different antibodies, have been able to demonstrate dipeptide staining in lower motor neurons (Donnelly *et al.*, 2013; Westergard *et al.*, 2016). The absence of p62 inclusions, which in neuropathologic tissue co-stain either with dipeptide antibodies or TDP-43 antibody is therefore not surprising in this cohort. Increased staining for p62 is indeed seen following treatment with the proteasomal inhibitor MG-132 as expected. However, there were no differences in staining morphology between control and patient lines observed in this study. This is consistent with previous observations of the same model, showing increased p62 aggregation in cortical neurons but not motor neurons (Dafinca *et al.*, 2016).

This study, using a wide range of cell lines from *C9orf72* carriers, did not reveal any TDP-43 mislocalisation in *C9orf72* patient lines, both compared to three control lines, as well as a CRISPR controls. All cell lines displayed similar levels of nuclear/cytoplasmic ratios of TDP-43 staining intensity, both in control conditions

and following 24 hours of stress. The absence of mislocalisation and aggregation of TDP-43 and the absence of differences in post-translational modification between control and patient lines, indicates that TDP-43 dysfunction may result at a later stage, and may not be captured by this iPSMN model. The results do not exclude the possibility of subtler TDP-43 abnormalities. Indeed, cytoplasmic dysfunction of TDP-43 does not require aggregation (Conicella *et al.*, 2016) and other pathophysiological mechanisms such as liquid phase-transition abnormalities as well as stress granule formation could be involved. It is however also possible that TDP-43 acts downstream of other pathways triggered by the hexanucleotide repeat expansion.

RNA sequencing of the iPSMN model yielded important findings. Cell-to-cell adhesion was enriched in the geneset even when controlling for neuronal subtype. Within this group the differential regulation of cadherin related genes appears to be important, but requires further validation. Protocadherin-gamma is known to be required for synaptic development and protocadherin gamma mutant mice have interneuron degeneration (Wang *et al.*, 2002; Weiner *et al.*, 2005). Protocadherin gamma expression is highest in neural tissues, and protocadherin gamma protein levels and its subcellular distribution has been postulated to undergo significant changes during development (Kallenbach *et al.*, 2003). The other important finding of this study is the finding that synaptotagmin 11 (*SYT11*) is lower in iPSMNs derived by CRISPR correction and healthy control iPSMNs compared to *C9orf72* HRE carriers. SYT11 is expressed mainly in the brain (Yeo *et al.*, 2012). One report suggested it may inhibit clathrin-mediated endocytosis (Wang *et al.*, 2015).

The findings of CRISPR-specific downregulation in the results is of equal importance. CRISPR-Cas9 system and zinc finger proteins are both nucleases. The zinc finger nucleases downregulated in this experiment are spread throughout the genome at multiple loci, so a direct off-target effect is very unlikely. The more likely explanation is a lasting effect that occurred during the CRISPR treatment, possibly at the epigenetic level as a response of the cell to nuclease attack. What is particularly interesting here, is that the downregulation of nucleases was persistently observed in samples collected over a 1 year period at different passages. Further work will be required to elucidate the mechanism.

The study is limited by the constraints of the iPSC model and its inherent variability (Merkle and Eggan, 2013; Sances *et al.*, 2016). This is shown clearly in Figure 4.1D, which demonstrates that certain lines will not successfully differentiate to motor neurons. The variability between the lines is not connected to the quality of the starting material as demonstrated by the C9-T clones one of which differentiates successfully while the other one does not. This may have to do with differences in reprogramming that are not detectable using conventional methods. As such, some of these lines may retain epigenetic features that prevent them from acquiring a motor neuron fate, while passing the tests for pluripotency. Lines that were not differentiating successfully were excluded from the experiments, but it is plausible that differences in reprogramming or in the original fibroblast clone ultimately contribute to the variation between clones of a single line. Other factors that could contribute to variance in differentiation include differences in passage number, or differences resulting from the batches of supplements used.

An important question arises as to how to appropriately label and classify cells in these experiments. In this experiment, most cells were ChAT and SMI-32 positive. In the context of high concentrations of retinoic acid the cells are cholinergic spinal neurons and this confirms their motor neuron identity, but the absence of Islet-1 from the neurons raises questions about possible subgroups that will require further research.

Importantly, iPSMNs are a model of ALS that allows the study of ALS-associated mutations in their native genetic background, without artificial overexpression of protein. CRISPR technology has allowed for dissection of the disease-causing mutations and therefore also allows study of the mutation itself with a control line with an identical genetic background. Despite the difficulties of working with iPSMNs and the large numbers of clones and replicates required for reproducible research, iPSMNs are therefore an important research tool which are likely to be used extensively in the future.

Chapter 5: Using RNASeq to investigate early transcriptional changes in ALS

This chapter describes experiments in which the transcriptome of TDP-43 transgenic mice was analysed to identify early changes in the pathophysiology of ALS. The mice were developed in collaboration with Professor Richard Wade-Martins' group in the Department of Physiology, Anatomy and Genetics, where the BAC construct was made. The mice were generated in a core transgenic facility at the Wellcome Centre for Human Genetics, Oxford, by Dr Ben Davies. The mice were fully characterised by Dr David Gordon in Professor Talbot's group in the Nuffield Department of Clinical Neurosciences.

5.1 Introduction: A TDP-43 BAC transgenic mouse model of ALS with physiologically relevant expression

To overcome the limitations of previous mouse models in ALS, a novel approach to mouse modelling was taken, in which an 80kb genomic fragment containing the human *TARDBP* locus and upstream regulatory regions was cloned into a BAC vector and engineered to contain the TDP-43 mutation M337V. Both the wild-type construct and M337V construct express a Y-pet tag (modified GFP) at the C-terminal of TDP-43. The BAC vector was stably inserted at single copy number into the mouse genome at the *Rosa26* locus on chromosome 6. Mice can be either homozygous or heterozygous for the human transgene (Figure 5.1A, Dr David Gordon). Unlike in previous models, human TDP-43 constitutes between 20% and 40% of total TDP in the brain and

between 10% and 20% in the spinal cord in these transgenic mice. In heterozygotes, transgenic TDP-43 expression is similar in wild type and M337V transgenic mice (Figure 5.1B, Dr David Gordon). At 3 months, the mouse is asymptomatic by motor testing and clinically indistinguishable from wild type and nontransgenic (Figure 5.1C, Dr David Gordon). Mutant, but not wild-type, human TDP-43 mice develop progressive motor dysfunction at 9-12 months. This is evident by a reduced time spent on accelerating rotarod testing as well as grip strength testing. Motor weakness is dose dependent in mutant mice as evidenced by worsening of the phenotype between heterozygous and homozygous mice.

Therefore, this mouse recapitulates the toxicity of the mutation, uncoupling it from the dose-dependent wild-type effects seen in previous models. While this mouse model also has no spinal cord pathology at 3 or 12 months, robust NMJ pathology is seen from 9 months onward. While a clinical and neuromuscular phenotype of the mutation emerges at 12 months in this mouse, the effects of the M337V mutation are observable earlier. TDP-43 cytoplasmic mislocalisation can be seen in primary motor neurons cultured from day E13.5 embryonic mice (not shown here), and the mouse shows deficits in axonal transport as early as three months. (Figure 5.1D, Dr James Sleigh). Taken together, this mouse has a phenotype of motor and neuropathological dysfunction which emerges late in adulthood, which allows for ageing and maturation prior to disease development. The presence of early molecular

Figure 5.1: Characterisation of BAC transgenic TDP-43^{M337V} mouse model of ALS (Figures and work by Dr David Gordon)

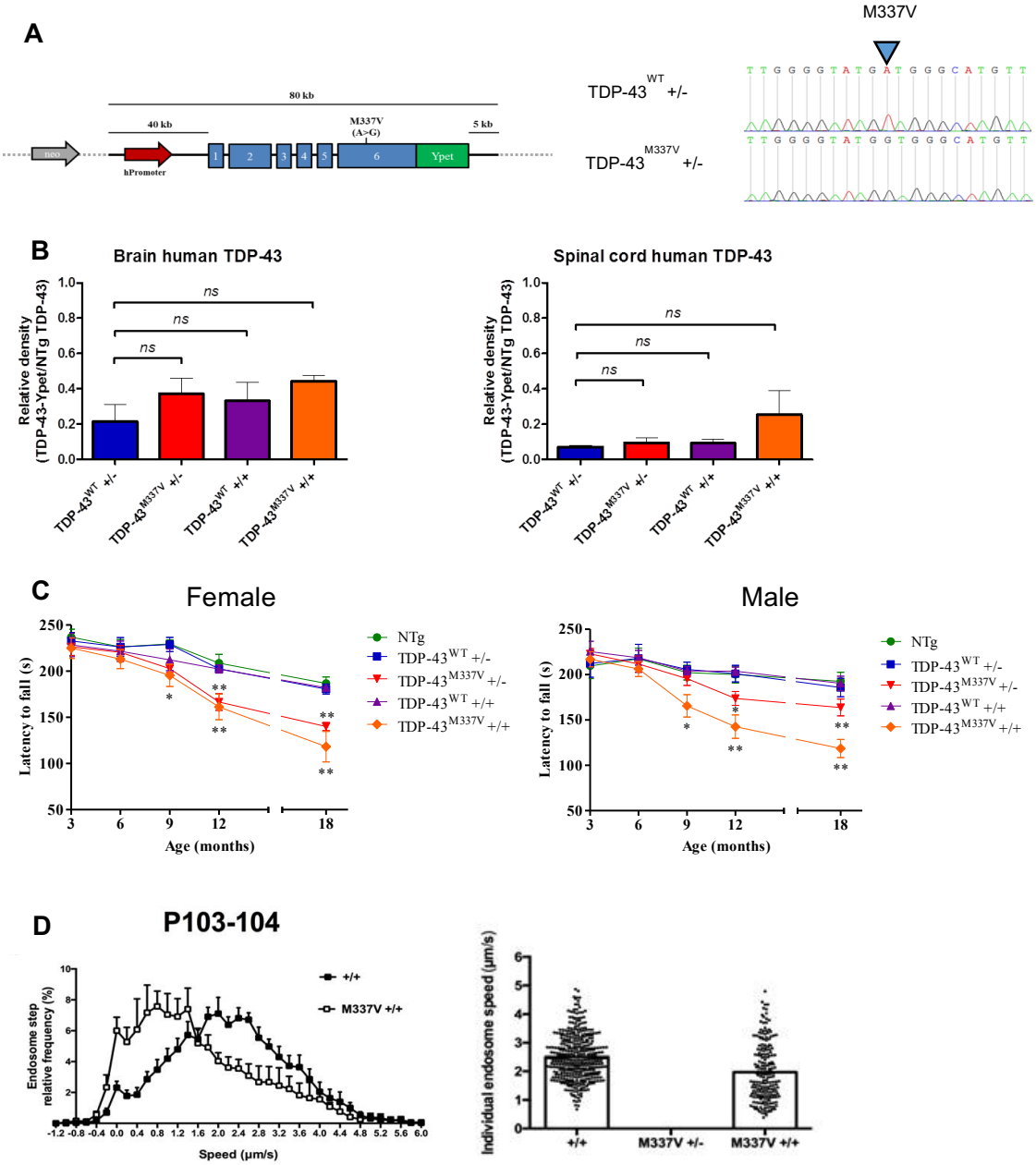

Figure 5.1: Design and characterisation of TDP-43^{M337V} mouse model. **A:** Design of BAC transgene containing complete human TDP-43 chromosomal sequence with 40kb upstream promoter region and incorporated Ypet tag. Right panel demonstrates sanger sequencing result from wild type and engineered M337V mutation. **B:** Human TDP-43 protein expression in the brain and spinal cord of TDP-43^{WT} and TDP-43^{M337V} mice displayed as a percentage of total TDP-43. No significant differences are seen between genotypes. **C:** Differences in rotarod testing showing significant differences between TDP-43^{M337V} mice and nontransgenic mice at 9 months in both male and female mice **D:** Differences in axonal transport at day 103 between homozygous TDP-43^{M337V} and nontransgenic mice.

dysfunction in mutant TDP-43 mice also allows for the study of the presymptomatic phenotype in adulthood.

5.2 Aims

The aim of this study was to characterise the transcriptional profile of the presymptomatic time point in the TDP-43^{M337V} mouse model. The first aim was to investigate early transcriptional dysregulation, and describe earliest pathways that are affected by the mutation prior to the development of the disease. The second aim was to investigate if the constitutive function of TDP-43 in modulating splicing was impaired at presymptomatic and symptomatic time points, and to therefore establish if splicing dysfunction is an early feature of the disease which drives pathology.

5.3 Results

5.3.1 Quality control

5.3.1.1 Quality and Mapping statistics reveal batch effects

Quality control was performed for all samples, which were processed and sequenced in two batches: 3-month data was sequenced in the first batch, and 9 and 12-month data in the second batch. All sequences were of good quality, and no reads were marked as poor quality by *ReadQC*. GC content was similar across all samples and lanes (Figure 5.2A). The read depth was similar across all three timepoints studied, with higher but also more variable coverage for 9 and 12 month samples (Figure

5.2B). Mapping efficiency was good to excellent with 98.3%, 89.4% and 90.3% mapped reads at the 3, 9 and 12-month time points respectively (Figure 5.2C). Only 13.2% of reads were duplicated in the 3 month samples, but read duplication was significant in the 9 and 12 month samples – 66.7% and 70.7% respectively (Figure 5.2D). Such low library complexity could be explained by significantly lower RNA input for the second sequencing batch.

5.3.1.2 9 and 12-month data was compromised by fragmentation bias

To further investigate the differences between the two sequencing batches, gene coverage biases were analysed in all samples and quantified using *Picard*. Again, there were marked differences in 3' to 5' biases batches, with minimal bias in 3-month data and variable bias in the samples derived from the second batch. The bias was however also found to be non-uniformly distributed within the 9 and 12-month samples, affecting wild-type and mutant transgenic samples more than nontransgenic samples (Figure 5.3A). The effect of the bias on the data was investigated using exploratory data analysis. Data exploration revealed a strong correlation between the median 5' to 3' bias and the first principal component, which represented 46.0% of the variance in the entire dataset (adjusted R^2=0.9512, $p < 2e\text{-}16$, Figure 5.3B). The bias was also visualised using *IGV* (Figure 5.3C). As the bias was distributed unevenly between conditions, this precluded the analysis of raw 9 and 12-month data, in which variance would be driven by uneven gene coverage that differentially affects long and short genes. The 3-month dataset was of high quality and was analysed further.

Figure 5.2: Quality control of the RNA sequencing experiment

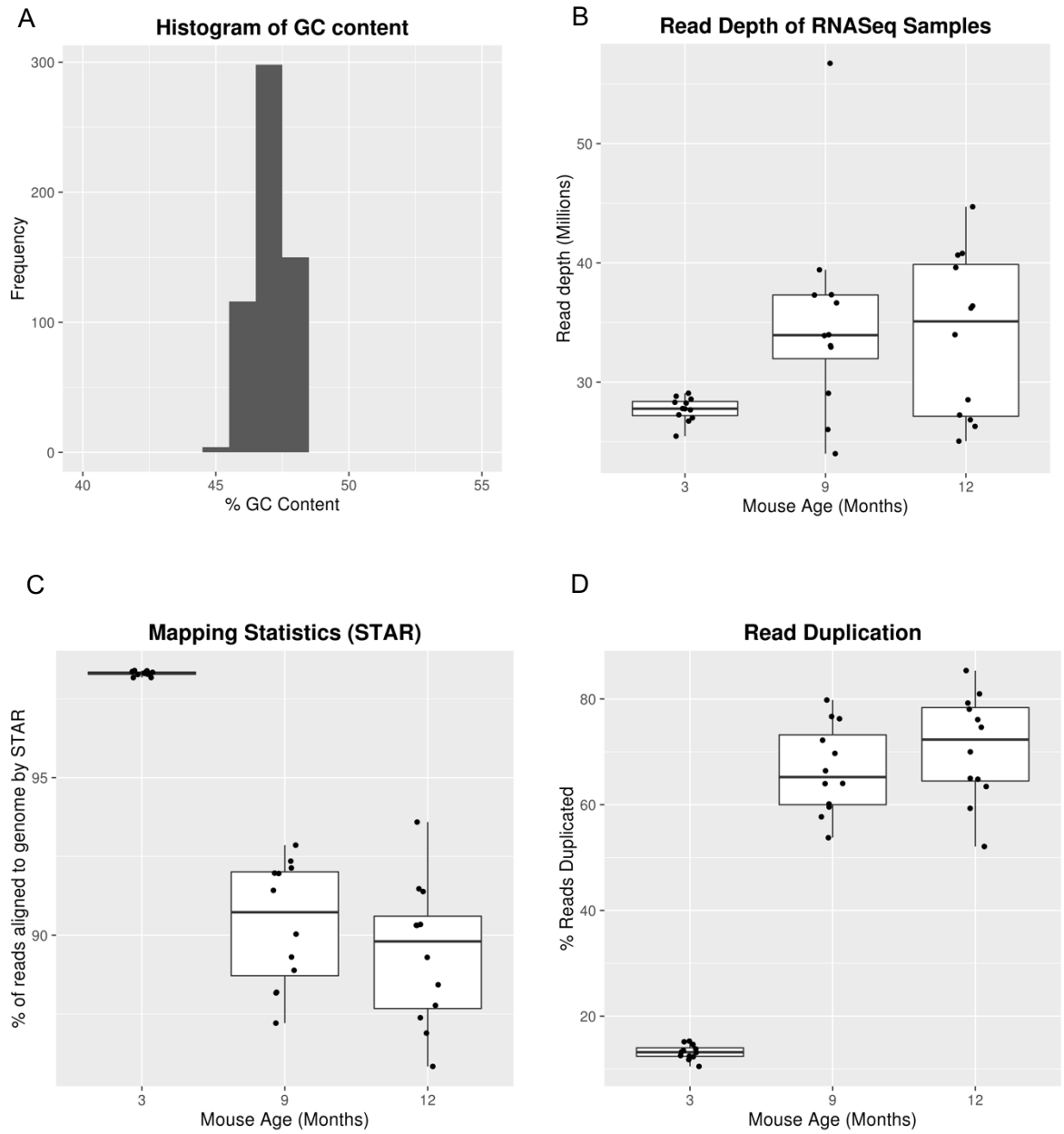

Figure 5.2: Quality control of sequencing samples from all three time points indicating some batch effects between 3 month data and 9 and 12 month data. All data files received from the sequencing centre from individual lanes and experiments is represented. **A:** GC fragment bias across all samples included in the experiment. The narrow distribution of the GC content histogram gives an indication of successful sequencing. **B:** Read depths were acceptable in all samples studied with a minimum read depth of 24 million reads. **C:** Mapping was performed by *STAR* and was good across all samples, with exceptional mapping percentages at 3 months. **D:** Significant levels of read duplication were detected for 9 and 12 month data

Figure 5.3: Analysis of fragmentation bias in RNA sequencing experiment

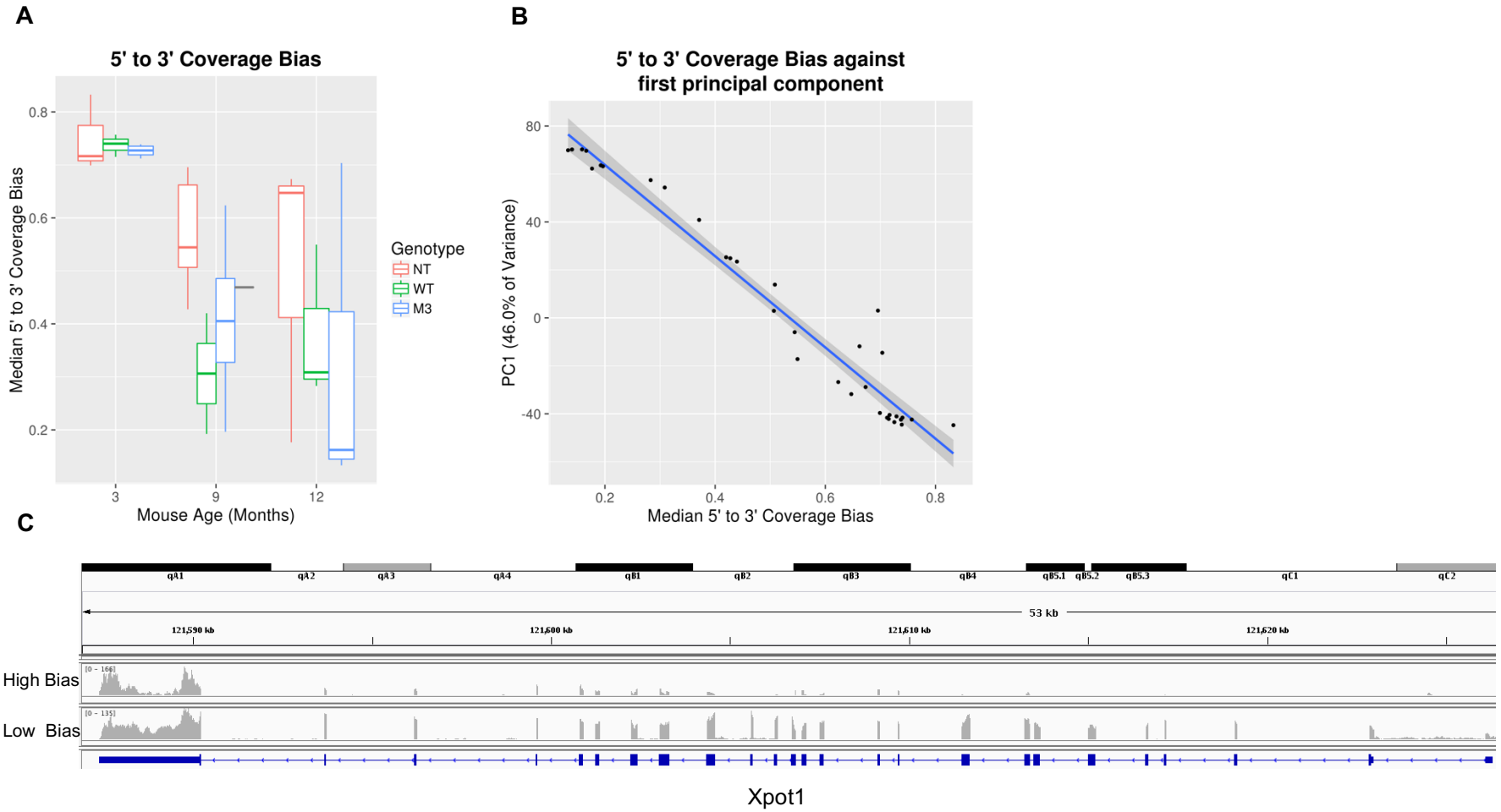

Figure 5.3: A: Gene context specific analysis reveals significant 5' to 3' bias as assessed by *Picard* using a ratio of coverage at the two gene ends. Values close to 1 represent high concordance of 3' and 5' values. Low values signify a poor coverage at the 5' end. **B:** Exploratory data analysis demonstrating the strong effect of 5' to 3' bias on the variance in the dataset. **C:** Gene coverage plot showing decreasing coverage toward the 3' end of the *Xpot* gene on chromosome 10 in the top panel (TDP-43^{M337V} 3 month timepoint) compared to the lower pane (TDP-43^{M337V} 12 month timepoint).

5.3.2 Differential expression analysis at 3 months

5.3.2.1 No significant group-wise differences were identified between wild-type and nontransgenic spinal cords

In keeping with the clinical differences observed between TDP-43^{M337V} mice and TDP-43^{WT} or nontransgenic animals, exploratory data analysis revealed separation of TDP-43^{M337V} mouse RNA from the other two genotypes (Figure 5.4A). Unsupervised hierarchical clustering showed clustering of all TDP-43^{M337V} mouse samples (Figure 5.4B). Two of the wild-type mice clustered closer to the M337V group than the remaining wild-type and nontransgenic samples. The similarity of nontransgenic and wild-type mouse spinal cords was reflected in a lack of differential expression between the two, whereas both showed robust differential expression compared to the mutant (Figure 5.4C). Given these results, together with the pathological and clinical results, nontransgenic and TDP-43^{WT} mice were grouped to increase the power of the analysis. For completeness, differential expression analysis between TDP-43^{M337V} and TDP-43^{WT} only was also performed. As compared to using NTg/WT combined as the control, a much smaller number of genes were found to be differentially expressed. However, as the vast majority of these (83%) were found with either control group, only the results from the more powerful NTg/WT-combined control comparison are reported here.

Figure 5.4: Exploratory data analysis of three months old mouse spinal cord RNA sequencing

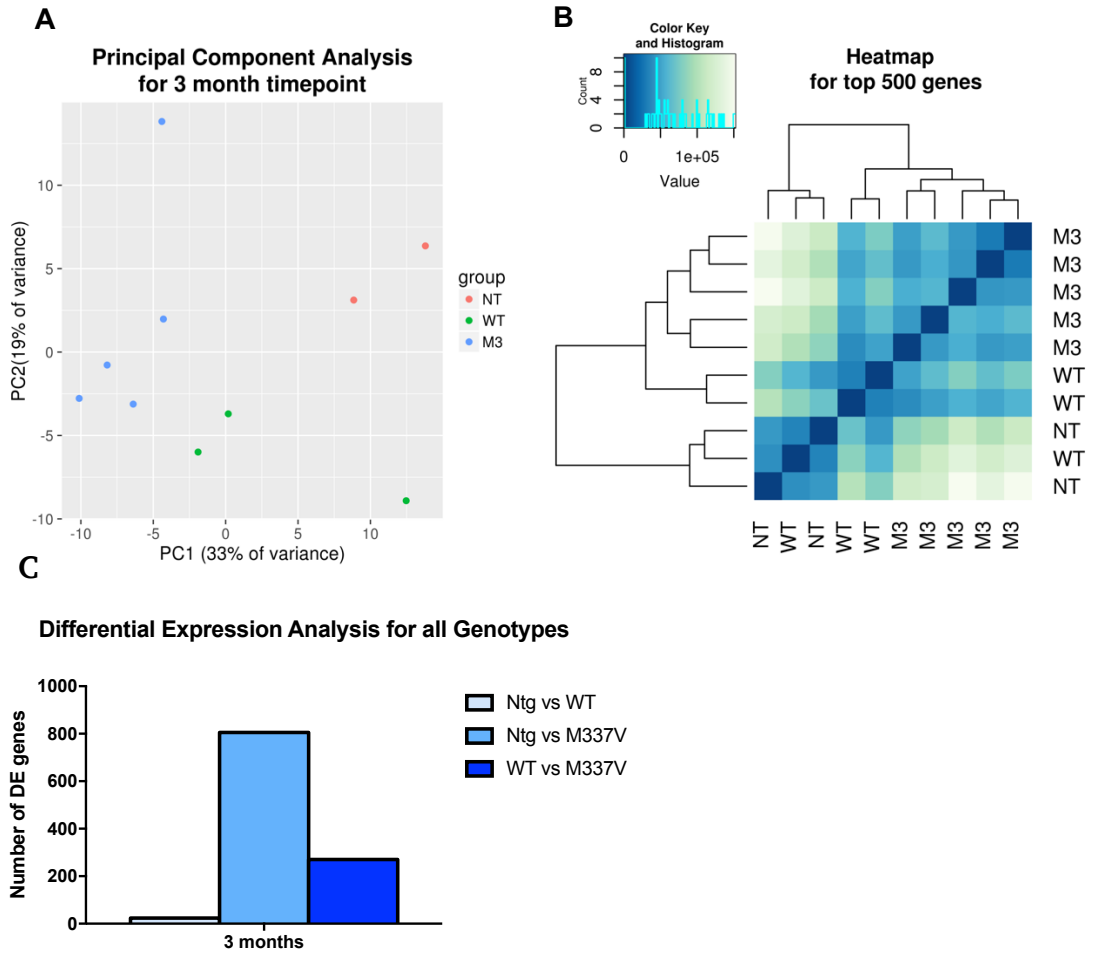

Figure 5.4: Results from differential expression analysis comparing all three genotypes in three-month-old mice**. A:** Principal component analysis shows the first principal component separating TDP-43^{M337V} mutant mice from TDP-43^{WT} and nontransgenic mice. **B:** Hierarchical clustering using transposed distance information derived from the 500 most expressed genes demonstrates unsupervised clustering of TDP-43^{M337V} mice. Nontransgenic and TDP-43^{WT} mice cluster separately, with two wild-type being mice more similar to the mutants than the remaining wild-type and nontransgenic mice. **C:** Differential expression results from three comparisons showing no differences between nontransgenic and wild-type mice, but differences between TDP-43^{M337V} mice and the other two groups.

5.3.2.2 Differential expression reveals a distinct transcriptional profile in presymptomatic TDP-43^{M337V} mouse spinal cords

5.3.2.2.1 633 genes are differentially expressed between TDP-43^{M337V} and TDP-43^{WT}/nontransgenic

Differential expression analysis was carried out on five female TDP-43^{M337V} mice versus five female controls, which were either TDP-43^{WT} or nontransgenic. A total of 1322 genes were differentially expressed. Permutation testing of all labels was performed to ascertain the specificity of the findings ($p=0.007$, Figure 5.5A). As the resulting differential expression changes in this preclinical model were modest (Figure 5.5B), a 0.2-fold change cut-off was selected, at which 633 genes were found to be differentially expressed. Of these, 185 were upregulated and 448 were downregulated.

5.3.2.2.2 Differentially expressed genes are functionally related

A Phenotypic Linkage Network Analysis was performed to inform the interpretation of the differentially expressed geneset. For this, a nervous system specific was constructed using relevant phenotypes from the MGI database. The network contained data from co-expression databases and ontological databases, which were normalised against data from the MGI phenotype database.

Figure 5.5: Results from differential expression analysis of 3 months old TDP-43^{M337V} versus nontransgenic and wild-type mouse spinal cords

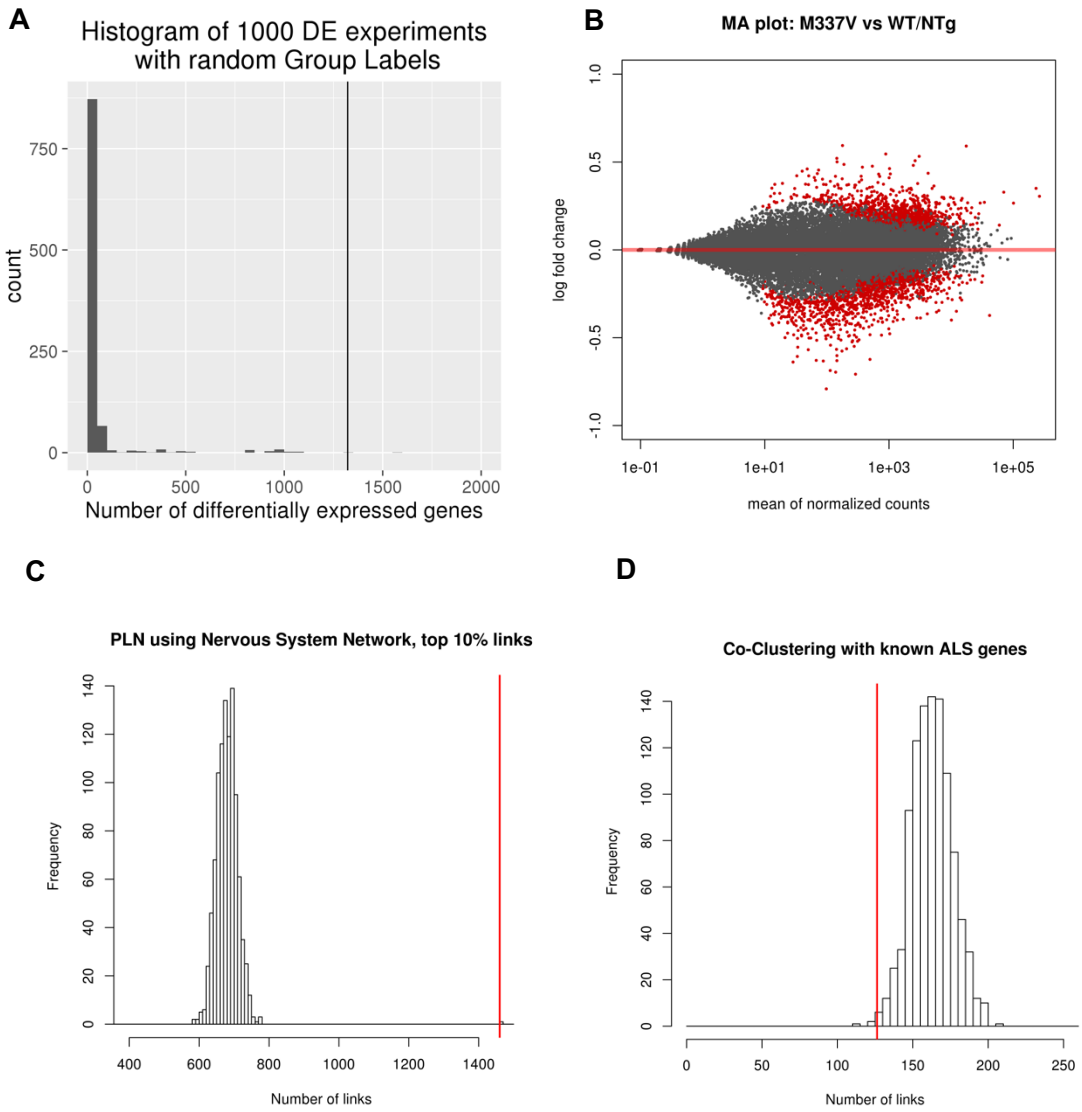

Figure 5.5: A: Differential expression between TDP-43^{M337V} mutants and wild-type/nontransgenic mice was highly robust as demonstrated by random permutation of genotype labels. Black line indicates the unpermuted result. **B:** Differential expression plotted against mean normalised counts, with all differentially expressed genes coloured in red. The data shows good normalisation. **C:** The 633 differentially expressed genes are functionally related, as demonstrated by Phenotypic Linkage Network Analysis, using a nervous system specific network. The red line shows the strength of links in the dataset, compared to the distribution of link strength for samples randomly drawn from the nervous system network. **D:** No co-clustering of differentially expressed genes with a set of 40 known genes harbouring ALS causing mutations was observed in the network analysis.

Using this approach, there was a more than two-fold enrichment of functional connections in the differentially expressed genes compared to genes randomly drawn from the network ($p<0.001$, Figure 5.5C). Differential expression in the dataset did not appear to be related to genes known to be mutated in ALS – a list of forty known ALS genes and their immediate network partners were not enriched in the differential expression results (Figure 5.5D). Interestingly, the only ALS/FTD related gene that was differentially expressed was progranulin, which was downregulated in the TDP-43^{M337V} mouse.

5.3.2.2.3 Sub-network analysis reveals important pathways involved in the development of ALS

Phenotypic network analysis allows for partitioning of the network using an unbiased approach, so that genes in sub-networks are more connected to other genes from the sub-network compared to genes outside the sub-network. The ratio of these connection strengths is represented by its modularity. For the geneset generated in this experiment, 9 subnetworks were obtained with a resulting modularity of 0.59. Genes in each sub-network were subsequently tested for similarity using gene ontology testing and were found to be related to cellular functions relevant to motor neuron biology (Table 5.1).

Subnetwork	Number of Genes	GO enrichment (OR)
0	91	Peptide cross-linking (30.2), extracellular matrix binding (25.6)
1	82	Myelination (38.6)
2	31	adenylate cyclase-inhibiting GPC AchR signaling pathway (216.6)
3	31	Cell-cell adhesion (84.4)
4	32	Maltose metabolic process (365.64), amino acid transport (123.9)
5	80	TGF beta-activated receptor activity (80.8)
6	5	Lipid transport and binding (Inf)
7	42	G1/S transition of mitotic cell cycle (60.5)
8	59	TLR4 signalling (64.3), TAP binding (50.0)

Table 5.1: Results of Subnetwork Analysis.
OR = Odds Ratio

5.3.2.3 No evidence of TDP-43 nuclear dysfunction

5.3.2.3.1 No splicing changes seen in TDP-43^{M337V} mice

Despite the robust differential expression changes, no differences in splicing between TDP-43^{M337V} mice and controls were observed using rMATS with a predetermined power threshold for detection of differential splicing events of at least 1% magnitude (Figure 5.6A).

5.3.2.3.2 Differentially expressed genes are not enriched for TDP-43 binding sites

To test the hypothesis of correlation of the differentially expressed dataset with TDP-43 binding, a previously published HITS-CLIP dataset (Polymenidou *et al.*, 2011) was analysed and peaks were called using a modified iCLIP pipeline that takes into account crosslinking deletions observed in HITS-CLIP data. As

Figure 5.6: No evidence of splicing dysfunction in TDP^{M337V} mice at three months

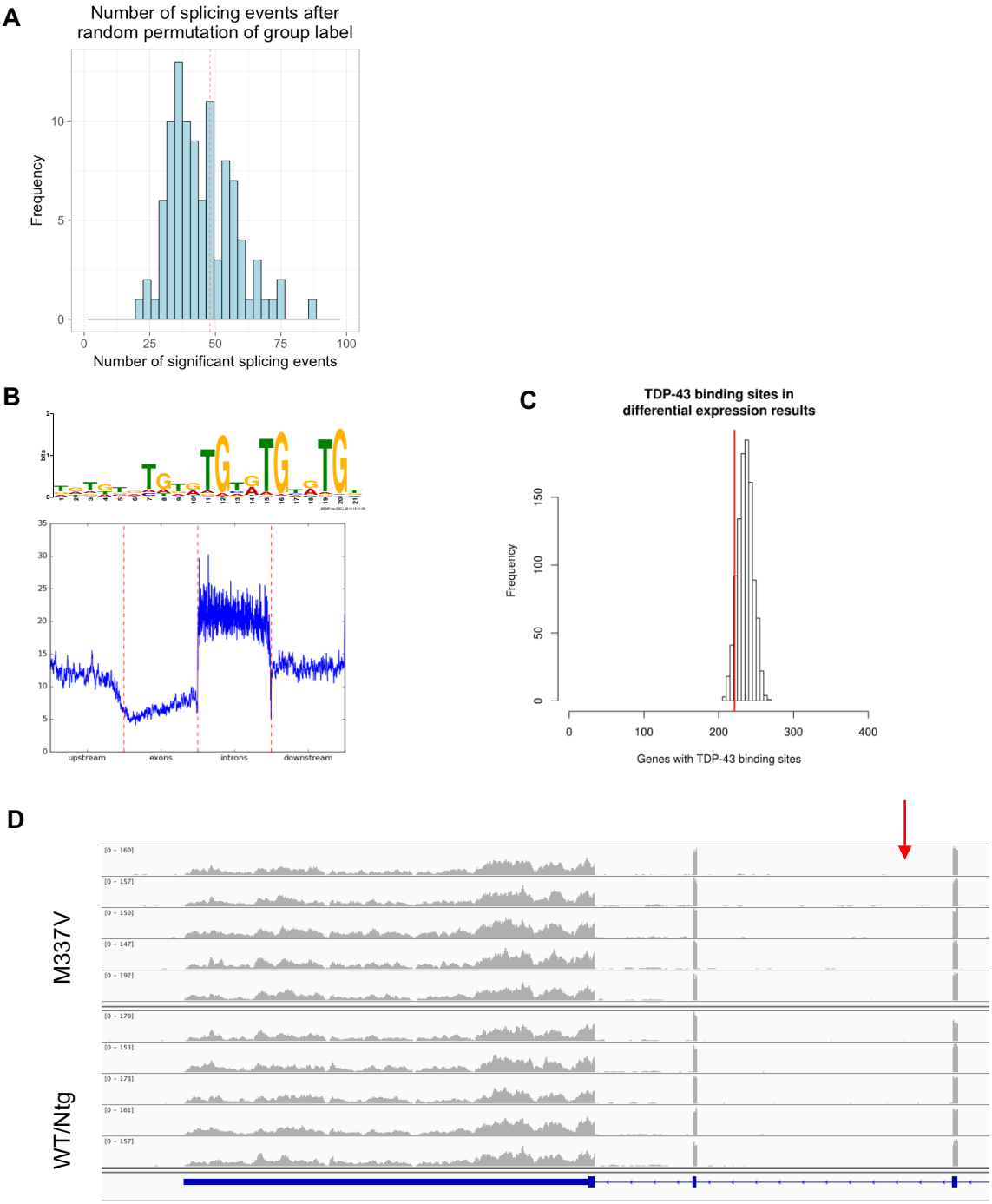

Figure 5.6: A: Assessment of TDP-43 nuclear function using RNA sequencing data. **A:** No evidence of abnormal splicing between M337V and WT/NTg on splicing analysis using *rMATS* on permutation analysis. Red line indicates unpermuted original result. **B:** Validation of reanalysis of HITS clip data showing predominantly intronic peaks with the typical UG-rich TDP-43 binding motif. **C:** TDP-43 binding sites were not enriched in the results dataset (red line). The simulated datasets were appropriately matched to the abundance distributions of the actual dataset. **D:** Cryptic exons remained suppressed in the M337V mice. An example from the *Synj2bp* gene is shown here, with the expected cryptic exon location marked by a red arrow.

previously reported, 97% of peaks were intronic (Figure 5.6B). Using only peaks from both experimental conditions, 26933 TDP-43 binding sites were found in 7468 genes and sequences were enriched for the UG-rich TDP motif. Amongst all 633 differentially expressed genes, 221 overlapped with the TDP-43 binding site dataset, and this was not found to be different from simulated data matched for abundance ($p=0.93$, Figure 5.6C).

5.3.2.3.3 Cryptic exons remain suppressed in the TDP-43^{M337V} mouse

Further evidence for conservation of nuclear function was provided by the analysis of known mouse-specific cryptic exon sites, which have previously been found in TDP-43 knockdown experiments (Ling *et al.*, 2015). There was no expression at these UG-rich sites in either the TDP-43^{M337V} or the control mice (Figure 5.6D).

5.4 Discussion

The RNA sequencing results presented here, give evidence for a presymptomatic transcriptional profile that is unique to TDP-43^{M337V} mutant mice. This result is concordant with other mutation-specific findings in this mouse model such as motor testing, and provides further evidence for a biological effect of the mutation itself.

The relatively low fold changes of the differentially expressed genes can be explained by multiple factors. Given this is a native promoter driven model with TDP-43 expression closer to physiologically relevance one might expect less

severe changes to the transcriptome. Furthermore, the RNA sequencing data comes from the presymptomatic stage of the mouse. Finally, whole spinal cords were used, and it is possible that the mutation only affects certain cell types in the spinal cord. Given these considerations a 20% threshold for minimal fold change was selected. The resulting geneset was robust as demonstrated by permutation analysis and phenotypic linkage analysis that showed functional connectivity. GO ontology analysis of the subnetwork analysis revealed alteration of nervous system specific pathways such as myelination and acetyl-choline receptor mediated adenylate cyclase repression, but also fundamental cellular functions such as cell-to-cell adhesion and amino acid and lipid transport.

There are multiple results from the differential expression analysis that could be pursued in future experiments. An obvious choice is progranulin, a known ALS/FTD gene found downregulated in the mutant mouse model. (*GRN*) mutations cause GRN haploinsufficiency and are known to account for up to 40% of cases of TDP-43 positive familial frontotemporal dementia (Rademakers *et al.*, 2012). Progranulin is an 88 kDa secreted glycoprotein, which is expressed by neurons and microglia in the brain. Once secreted, it can be cleaved into 6 kDa subunits called granulins that serve as neurotrophic factors. Progranulin has been implicated in multiple cellular functions, including lysosomal function, neuroinflammation, neurite outgrowth, synaptic stability and stress response (Petkau and Leavitt, 2014). Progranulin has also been studied in ALS. Upregulation of progranulin has been found in the later stages of the *SOD1*

mouse model (Philips *et al.*, 2010), and aberrant sortilin splicing due to TDP-43 dysfunction has been postulated to inhibit extracellular progranulin function (Prudencio *et al.*, 2012).

Other pathways of interest highlighted by the mouse model include the downregulation of *TAP1*, *TAP2* and *TAPBP* in the analysis. These genes are found in the MHCI locus and have an important role in antigen presentation (Leone *et al.*, 2013). Given the presence of whole spinal cord RNA in this experiment, one explanation would be the transcriptional changes in immune receptors in glial cells, although a neuroinflammatory phenotype was not observed on neuropathological analysis. Another explanation is the increasing evidence for functional importance of MHC I complexes on neuronal cells (Song *et al.*, 2016). Downregulation of MHCI expression could thus have significant consequences for motor neuron health and survival.

An important negative from this study is the absence of genes known to harbour ALS-causing mutations. This could be explained by the fact that these genes have divergent functions and may result in neurodegeneration through a number of different pathways (Al-Chalabi *et al.*, 2012). Another possible explanation, is that these genes predispose to molecular cascades of disease, and may be upstream of common pathways that lead to ALS. Finally, it is also possible that the differentially expressed geneset is specific to the TDP-43 mutation or even to the murine model studied. The absence of any dysfunction of nuclear RNA processing is a striking feature in this mutant TDP-43 murine model. This is further supported by the absence of neuropathological changes at three months.

The absence of RNA dysfunction indicates that other cellular functions of TDP-43 could play a more important role in the early stages of pathology in these mice. Examples of such other functions of TDP-43 include stress granule formation (Dewey *et al.*, 2012; Aulas and Vande Velde, 2015) and axonal mRNA transport (Alami *et al.*, 2014), which are both affected by TDP-43 mutations. There is some evidence for abnormalities in both these pathways in this M337V mouse model: These include the abnormal stress granule dynamics seen in primary motor neurons and the axonal transport defects.

There are important limitations of this study. Firstly, dose dependent effects cannot be entirely excluded in this model. As demonstrated in Figure 1C, transgenic protein levels are possibly higher in the M337V mouse spinal cords than in the WT mouse, although these differences did not reach statistical significance. Secondly, RNA sequencing has limited power to detect sequencing changes below a log2-fold threshold of 1. Although permutation analysis and network analysis provide support for the significance of the overall findings in this mouse model, the small differences between the model make it difficult to validate individual results, with high likelihood of false positives due to biological and experimental variability. Thirdly, the assessment of TDP-43 function in this experiment is limited by sequencing depth. Spike-ins that can assess the power of sequencing experiment to detect splicing changes have only recently become available (Hardwick *et al.*, 2016), and splicing changes only become detectable at 30 million read depth. Small alterations in splicing are unlikely to have been fully captured by this experiment.

A difficulty faced by all transcriptomic studies is the uncertainty about the contribution of transcriptomic changes to the disease. It is possible that the most significant and earliest changes in ALS occur at the post-transcriptional level by posttranslational modification (Dangoumau *et al.*, 2016) or protein templating (Nonaka *et al.*, 2013) and these would not be detected in transcriptomic studies. Nonetheless, the detection of a distinct transcriptional profile at the presymptomatic stage in the mouse model studied underlines the power of transcriptomic studies in providing evidence for pathway alterations without the inherent biases introduced by hypothesis driven research (Raoult, 2010). Given the mouse does not develop disease until much later, the differentially expressed gene pool represents a set that almost certainly represents early changes in the mouse, and will provide a useful resource for future study. This is truly independent of whether they are of pathophysiological relevance or merely represent early cellular responses and adaptations to the original insult.

While this study provides additional evidence for the scientific validity of the TDP-43^{M337V} mouse, follow up of the 622 identified differential expression changes is not trivial. As noted previously, the individual results with low-fold changes have limited positive predictive value and will require validation on independent samples or sequencing experiments. Furthermore, the choice of candidates from such a large gene pool will to some extent be driven by prior scientific knowledge and hypotheses, and is prone to subjective interpretation. Such difficulties may be remedied in part by selecting functional clusters of differentially expressed genes, such as the TAP complex in the case of this study.

Later time-points were originally planned but had to be discarded due to quality control issues. Longitudinal transcriptomic studies will provide important validation of the current findings.

Chapter 6: Conclusions

6.1 Main conclusions

This thesis confirms the central role of TDP-43 pathology in *C9orf72* ALS-FTD at post-mortem. TDP-43 is present in areas of the CNS that would be expected to be affected based on the clinical subtype of the patient (ALS, FTD or ALS-FTD), and absent in clinically unaffected areas. *C9orf72* HRE specific pathology, limited in this study to RNA foci, GA dipeptide and assessment of haploinsufficiency, does not exhibit such regional specificity. Thus, this study provides further indirect support for a likely pathophysiological role of TDP-43 in *C9orf72* HRE positive ALS and FTD, at least in the latter stages of the disease.

The work presented here provides further evidence that methylation upstream of the hexanucleotide repeat is variable in *C9orf72* HRE ALS-FTD. Only one of the 12 cases studied in our cohort had high methylation levels, in line with other studies using the similar techniques, indicating that other mechanisms must be causing the observed downregulation of transcription. Only very few changes in global methylation were found on 450k chip analysis of global methylation, though the findings may have been affected by sample heterogeneity.

No TDP-43 alterations in solubility or localisation were observed in an iPSMN model of *C9orf72* HRE ALS in this thesis. Cellular stress using the proteasomal inhibitor MG-132 increased cytoplasmic localisation in iPSMN from *C9orf72* HRE positive patients and healthy controls equally. RNA foci, but not dipeptide

inclusions could be observed, though others have previously demonstrated dipeptide protein presence using dot blotting. The absence of TDP-43 alterations has important implications for the generalisability of the model and its ability to model later stages of ALS.

RNA sequencing compared a *C9orf72* HRE cell line to its isogenic control and indicated derangements particularly in cell-to-cell adhesion pathways and differential expression of multiple protocadherins was observed. Validation in a wider, more general cohort had advantages for target selection, as demonstrated by identification of *SYT11*. But the approach also allowed the identification of unexpected downregulation of zinc finger proteins in the CRISPR isogenic control. This result has important implications for future CRISPR work

Finally, RNA sequencing also revealed significant differences in gene expression between the TDP-43^{M337V} and the TDP-43^{WT}/nontransgenic mouse spinal cords at 3 months, a time when mice a symptom free and show no pathology, except for slowing of axonal transport.

6.2 The pathophysiology of ALS is complex

Understanding of ALS has greatly evolved in recent years following molecular and genetic advances. ALS is remains a clinical syndrome defined by its clinical presentation only, and even electrophysiology is only a supporting biomarker (Talbot, 2009).

But within ALS, increasingly, heterogeneity is increasingly recognised at the molecular level. The genetics of ALS are increasingly complex, and the increasing genetic understanding of the condition has significantly shaped our understanding of the disease, causing re-evaluation of the impact of penetrance in a disease of older age (Murphy *et al.*, 2017) This understanding has had a significant impact on clinical management (Al-chalabi *et al.*, 2016), but others have warned about the over-inflation of the genetic contribution to sporadic disease (Gibson *et al.*, 2017). It remains unclear at which point the various genetic forms of ALS converge onto common pathways, such as TDP-43 pathology, which accounts for 98% of all sporadic and familial cases. Some disease pathways converge downstream of TDP-43: it is now well-established that both SOD1 and FUS do not harbour TDP-43 pathology but cause an indistinguishable clinical disease with the same specificity for upper and lower motor neuron pathology.

Most of ALS remains sporadic, and only few risk genes such as ataxin 2 have thus far been identified. Some groups have postulated an oligogenic hypothesis for ALS given that ALS mutations co-occur in patients more frequently than expected (van Blitterswijk *et al.*, 2012), however the low frequency of these events is unlikely to explain a large proportion of the prevalence. Epidemiological methods have postulated that ALS, like cancer, is a multistep process requiring six distinct events for pathology to occur (Al-Chalabi *et al.*, 2014), though it remains unclear what these events could be and how environmental factors interact with them.

6.3 Neuropathology provides an opportunity to study the final stages of ALS in humans

This thesis describes a series of experiments that investigate genetic forms of ALS, focussing on two important areas of research that can be viewed as occupying opposite ends of research methodology. On the one hand the neuropathological investigation of *C9orf72* illustrates the importance of post-mortem research and provides a context for relating various molecular phenotypes derived from models to the core of the disease itself. While mechanistic conclusions cannot be reached from the data available, the distribution and characterisation of the various terminal phenomena seen at post-mortem greatly informs the understanding of the role of *C9orf72* mutations in ALS. The confirmation in this study of the pivotal role of TDP-43 inclusions in *C9orf72* pathology is a very important finding that underlines the relevance of this key protein.

However the limitation of post-mortem studies are that they represent a terminal 'snapshot' of the disease processes. It was argued very early after the discovery of TDP-43 in ALS inclusions that the protein may merely be a bystander of neurodegeneration, a terminal marker with no pathophysiological significance (Rothstein, 2007). While there is now ample evidence of TDP-43 functional disruption in ALS, as outlined in this thesis, the question of the specific role of TDP-43 in ALS in general remains unanswered. The discovery of ALS-causing mutations in Matrin3 (Johnson *et al.*, 2014; Coelho *et al.*, 2015), as well as hNRNPA1 and hNRNPA2B1 (Kim *et al.*, 2013) expanded the spectrum of

ribonucleoproteins involved in ALS. This is further reinforced by increasing reports of a wide range of hNRNPs in the inclusions of ALS (Boeynaems *et al.*, 2016), indicating that TDP-43 is only one of many hNRNPs that may be dysregulated through abnormal nucleocytoplasmic transport in ALS.

A fundamental question that will need to be answered is whether the deficiencies in nucleocytoplasmic shuttling and the subsequent inclusion formation incorporating TDP-43 and other hNRNPs are a late disease-unifying phenomenon that is the ultimate step in a molecular cascade, or rather TDP-43 functional disturbance is a pivotal step in disease initiation? The finding of TDP-43 pathology in Alzheimer's Disease and Parkinson's Disease brains does lend some weight to the argment that TDP-43 pathology can be a general marker of cellular dysfunction (Higashi *et al.*, 2007), but TDP-43 is relatively scarce in these cases. Neuropathological studies are unlikely to be able to answer the question of TDP-43 involvement in early stages of pathology.

6.4 Relevant models of ALS can provide insight into early pathophysiological pathways

To understand changes that precede the pathophysiological stage of the disease, cellular and animal models are required in the case of CNS diseases such as ALS. These models not only provide insight into the molecular function of important ALS genes and proteins, but can also provide insight into the earlier stages of pathogenesis. The importance of ageing in modelling neurodegenerative disease as well as physiologically relevant expression of the transgenes are vitally

important to avoid the study of overexpression phenomena. The two models in used in this thesis are complementary examples of such approaches.

The potential to generate induced pluripotent stem cells and differentiate them into motor neuron has been particularly important in neurodegenerative disease research. iPSMNs and more recently iNeurons directly generated from the donor cell type offer the first opportunity to study human neurons in vitro, with human mutations in situ together with the relevant genetic background, which may well contribute to the mutation. The young age in culture of neurons (30 days) as well as their embryologic phenotype (Handel *et al.*, 2016) are important limitations here. Modelling a "disease in a dish", when that disease arises in humans after 50 or more years of normal function, will always have its limitations. This is particularly true for the *C9orf72* HRE, which is expressed in all cells in the human body, and the resultant RNA foci and dipeptide production is seemingly normally tolerated by neuronal and non-neuronal cells throughout a person's lifetime. The various markers of cellular stress found in iPSMNs to date in multiple studies, could be reflections of this ongoing dipeptide production. Importantly, this *C9orf72* HRE model does not exclude early TDP-43 involvement in the disease, and mechanisms other than aggregation and nucleo-cytoplasmic transport could still be abnormal.

Mouse models and can model more complex interactions at the network level of the transgene or environmental insult and the CNS. In well-designed models, that develop disease in adulthood, ageing of the animal can contribute to the pathophysiology. However, major neuroanatomic differences exist between the

human and mouse motor system (Nicholson *et al.*, 2000). The recent studies showing differences between non-conserved cryptic exon sites in mice and humans also raises specific questions about TDP-43 dysfunction in mice and its relevance to humans (Ling *et al.*, 2015).

Knowledge of the strengths and weakness of the various model systems used will be a key factor in successful study of ALS in model organism, and will be key to provide relevant systems for drug screening, for example.

6.5 High-throughput sequencing and its relevance to ALS

Recent advances in high-throughput sequencing have made fundamental contributions to our understanding of TDP-43 biology (Polymenidou *et al.*, 2011; Tollervey *et al.*, 2011; Arnold *et al.*, 2013; Lukavsky *et al.*, 2013; Rot *et al.*, 2017). The non-biased approach is an important complement to existing hypothesis-driven molecular biology approaches.

Other than cost, the method also has important limitations that need to be appreciated to correctly use the bioinformatics toolkits available to researchers now. The current short-read alignment offers robust differential expression, but biases such as the fragmentation bias observed in chapter 5 of this thesis have significant impact on the results, that can easily be interpreted as disease relevant changes. The most commonly available technique today, high throughput short read sequencing, has significant limitations in the assessment

of RNA splicing changes, with poor accuracy due to the length of the reads. This is not helped by likely over-annotation of the genome. New long read sequencing methods such as Nanopore technology are still in development, but will offer whole transcript resolution which will be of key importance. In the meantime, analysis of alternative polyadenylation, which is also affected by TDP-43 binding may offer a more accurate assessment of TDP-43 intranuclear functions (Rot *et al.*, 2017).

Single cell approaches to sequencing are increasingly now being used and will be especially useful in iPSMN models, where heterogeneity is a key factor. Using this approach, the heterogeneity of the cultures can be unravelled to study motor neuron specific changes and cell to cell differences in vulnerability.

6.6 Toward single-cell and approaches in ALS biology and personalised treatment in the ALS clinic

The results presented in this thesis must be seen in the context of a rapidly changing field of ALS research that has seen two decades of unprecedented discoveries and growth. The thesis showcases the power, as well as the limitations of iPSMNs and high throughput sequencing methods as tools to study ALS. In the case of the *C9orf72* iPSMN model, RNA sequencing provided a molecular overview that allowed for the detection of potential candidate genes such as SYT11, as well as identifying unexpected changes due to the CRISPR process.

The mouse RNA sequencing on the other hand provided support for early dysregulation due to mutant TDP-43 long before the development of symptoms or pathology. RNA sequencing here provides unbiased evidence for abnormal function of the mutant protein that is different from its preserved constitutive nuclear function. How exactly the TDP-43 mutation acts in this mouse model to cause disease is under investigation. It is interesting that other ALS-related genes are not enriched in the differential expression results at 3 months.

It is unclear what a TDP-43 mutation model can tell us about TDP-43 dysfunction in *C9orf72*-related ALS, and the disease pathways in the two genetic conditions could converge only in late stages of pathogenesis. Nonetheless, investigation of pathways through which TDP-43 mutations cause disease will have a significant impact on our understanding of other familial disease as well as sporadic disease.

Neuropathological, iPSMN and mouse studies in this thesis all suffered from lack of cellular specificity. Single cell molecular and imaging approaches will be key to understand processes that are cell-type specific, by utilising the ability to study subtypes and vulnerability at the single cell level.

The dissection of molecular pathophysiology will shed light on the fact that there may be multiple diseases in the ALS clinical syndrome, likely requiring better subtyping of clinical trials and supporting the development of personalised treatment based on the genetic mutation or molecular subtype.

Bibliography

Ababneh N. Modelling of amyotrophic lateral sclerosis (ALS) using induced pluripotent stem cells (iPSC), DPhil Thesis. University of Oxford. 2017

Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Hum. Mutat. 2012; 33: 1345–1351.

del Aguila MA, Longstreth WT, McGuire V, Koepsell TD, van Belle G. Prognosis in amyotrophic lateral sclerosis: a population-based study. Neurology 2003; 60: 813–9.

Al-chalabi A, Berg LH Van Den, Veldink J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. Nat. Rev. Neurol. 2016: 1–9.

Al-Chalabi A, Calvo A, Chio A, Colville S, Ellis CM, Hardiman O, et al. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. Lancet. Neurol. 2014; 13: 1108–13.

Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J. Neurol. Neurosurg. Psychiatry 2010; 81: 1324–6.

Al-Chalabi A, Jones A, Troakes C, King A, Al-Sarraj S, van den Berg LH. The genetics and neuropathology of amyotrophic lateral sclerosis. Acta Neuropathol. 2012; 124: 339–52.

Al-Mahdawi S, Pinto RM, Ismail O, Varshney D, Lymperi S, Sandi C, et al. The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. Hum. Mol. Genet. 2008; 17: 735–746.

Alami NH, Smith RB, Carrasco MA, Williams LA, Winborn CS, Han SSW, et al. Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. Neuron 2014; 81: 536–543.

Alaynick WA, Jessell TM, Pfaff SL. SnapShot: spinal cord development. Cell 2011; 146: 178–178.e1.

Alkan C, Coe BP, Eichler EE. GATK toolkit. Nat. Rev. Genet. 2011; 12: 363–76.

Almeida S, Gascon E, Tran H, Chou HJ, Gendron TF, Degroot S, et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. Acta Neuropathol. 2013; 126: 385–399.

Amoroso MW, Croft GF, Williams DJ, O'Keeffe S, Carrasco MA, Davis AR, et al.

Accelerated high-yield generation of limb-innervating motor neurons from human stem cells. J Neurosci 2013; 33: 574–586.

Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics 2015; 31: 166–169.

Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res. 2012; 22: 2008–2017.

Andersen PM, Nilsson P, Keranen ML, Forsgren L, Hagglund J, Karlsborg M, et al. Phenotypic heterogeneity in motor neuron disease patients with CuZn-superoxide dismutase mutations in Scandinavia. Brain 1997; 120 (Pt 1: 1723–1737.

Arber S, Han B, Mendelsohn M, Smith M, Jessell TM, Sockanathan S. Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. Neuron 1999; 23: 659–674.

Arnold ES, Ling S-C, Huelga SC, Lagier-Tourenne C, Polymenidou M, Ditsworth D, et al. ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. Proc. Natl. Acad. Sci. 2013; 110: E736-45.

Aulas A, Vande Velde C. Alterations in stress granule dynamics driven by TDP-43 and FUS: a link to pathological inclusions in ALS? Front. Cell. Neurosci. 2015; 9: 423.

Avendaño-Vázquez SE, Dhir A, Bembich S, Buratti E, Proudfoot N, Baralle FE. Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. Genes Dev. 2012; 26: 1679–84.

Baborie A, Griffiths TD, Jaros E, Perry R, McKeith IG, Burn DJ, et al. Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in C9ORF72 gene. Neuropathol. Appl. Neurobiol. 2015; 41: 601–12.

Baker JC, Beddington RSP, Harland RM. Wnt signaling in Xenopus embryos inhibits Bmp4 expression and activates neural development. Genes Dev. 1999; 13: 3149–3159.

Baker SC, Bauer SR, Beyer RP, Brenton JD, Bromley B, Burrill J, et al. The External RNA Controls Consortium: a progress report. Nat. Methods 2005; 2: 731–734.

Bakulski KM, Dolinoy DC, Sartor MA, Paulson HL, Konen JR, Lieberman AP, et al. Genome-wide DNA methylation differences between late-onset Alzheimer's disease and cognitively normal controls in human frontal cortex. J. Alzheimers. Dis. 2012; 29: 571–88.

Bauer PO. Methylation of C9orf72 expansion reduces RNA foci formation and dipeptide-repeat proteins expression in cells. Neurosci. Lett. 2016; 612: 204–209.

Bel-Vialar S, Itasaki N, Krumlauf R. Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxB genes in two distinct groups. Development 2002; 129: 5103–5115.

Bell M V., Hirst MC, Nakahori Y, MacKinnon RN, Roche A, Flint TJ, et al. Physical mapping across the fragile X: Hypermethylation and clinical expression of the fragile X syndrome. Cell 1991; 64: 861–866.

Belzil V V., Bauer PO, Gendron TF, Murray ME, Dickson D, Petrucelli L. Characterization of DNA hypermethylation in the cerebellum of c9FTD/ALS patients. Brain Res. 2014; 1584: 15–21.

Belzil V V., Bauer PO, Prudencio M, Gendron TF, Stetler CT, Yan IK, et al. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. Acta Neuropathol. 2013; 126: 895–905.

Benatar M. Lost in translation: Treatment trials in the SOD1 mouse and in human ALS. Neurobiol. Dis. 2007; 26: 1–13.

Bensimon G, Lacomblez L, Meininger V. A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. N. Engl. J. Med. 1994; 330: 585–91.

van Blitterswijk M, van Es MA, Hennekam EAM, Dooijes D, van Rheenen W, Medic J, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum. Mol. Genet. 2012; 21: 3776–84.

van Blitterswijk M, Gendron TF, Baker MC, DeJesus-Hernandez M, Finch NA, Brown PH, et al. Novel clinical associations with specific C9ORF72 transcripts in patients with repeat expansions in C9ORF72. Acta Neuropathol. 2015; 130: 863–76.

Boeve BF, Boylan KB, Graff-Radford NR, DeJesus-Hernandez M, Knopman DS, Pedraza O, et al. Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72. Brain 2012; 135: 765–83.

Boeynaems S, Bogaert E, Van Damme P, Van Den Bosch L. Inside out: the role of nucleocytoplasmic transport in ALS and FTLD. Acta Neuropathol. 2016; 132: 159–73.

Borghese L, Dolezalova D, Opitz T, Haupt S, Leinhaas A, Steinfarz B, et al. Inhibition of notch signaling in human embryonic stem cell-derived neural stem cells delays G1/S phase transition and accelerates neuronal differentiation in

vitro and in vivo. Stem Cells 2010; 28: 955–964.

Bourke SC, Tomlinson M, Williams TL, Bullock RE, Shaw PJ, Gibson GJ, et al. Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial. Lancet. Neurol. 2006; 5: 140–7.

Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 2016; 34: 525–7.

Buratti E. Functional Significance of TDP-43 Mutations in Disease. Adv. Genet. 2015; 91: 1–53.

Buratti E, Baralle FE. Characterization and Functional Implications of the RNA Binding Properties of Nuclear Factor TDP-43, a Novel Splicing Regulator of CFTR Exon 9. J. Biol. Chem. 2001; 276: 36337–36343.

Buratti E, Brindisi A, Giombi M, Tisminetzky S, Ayala YM, Baralle FE. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: An important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. J. Biol. Chem. 2005; 280: 37572–37584.

Buratti E, De Conti L, Stuani C, Romano M, Baralle M, Baralle F. Nuclear factor TDP-43 can affect selected microRNA levels. FEBS J. 2010; 277: 2268–2281.

Buratti E, Dörk T, Zuccato E, Pagani F, Romano M, Baralle FE. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. EMBO J. 2001; 20: 1774–1784.

Byrne S, Elamin M, Bede P, Shatunov A, Walsh C, Corr B, et al. Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: A population-based cohort study. Lancet Neurol. 2012; 11: 232–240.

Caroscio JT, Mulvihill MN, Sterling R, Abrams B. Amyotrophic lateral sclerosis. Its natural history. Neurol. Clin. 1987; 5: 1–8.

Castel AL, Nakamori M, Tomé S, Chitayat D, Gourdon G, Thornton CA, et al. Expanded CTG repeat demarcates a boundary for abnormal CpG methylation in myotonic dystrophy patient tissues. Hum. Mol. Genet. 2011; 20: 1–15.

Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. 2009; 27: 275–280.

Chang C ke, Wu TH, Wu CY, Chiang M hui, Toh EKW, Hsu YC, et al. The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. Biochem. Biophys. Res. Commun. 2012; 425: 219–224.

Cheah BC, Vucic S, Krishnan a V, Kiernan MC. Riluzole, neuroprotection and

amyotrophic lateral sclerosis. Curr. Med. Chem. 2010; 17: 1942–199.

Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 2005; 436: 740–744.

Chestnut B a., Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic Regulation of Motor Neuron Cell Death Through DNA Methylation. J. Neurosci. 2011; 31: 16619–16636.

Chiang P, Ling J, Jeong YH, Price DL, Aja SM, Wong PC. Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism. Proc. Natl. Acad. Sci. U. S. A. 2010; 107: 16320–4.

Chio a., Restagno G, Brunetti M, Ossola I, Calvo a., Canosa a., et al. ALS/FTD phenotype in two Sardinian families carrying both C9ORF72 and TARDBP mutations. J. Neurol. Neurosurg. Psychiatry 2012; 83: 730–733.

Chiò A, Borghero G, Pugliatti M, Ticca A, Calvo A, Moglia C, et al. Large proportion of amyotrophic lateral sclerosis cases in Sardinia due to a single founder mutation of the TARDBP gene. Arch. Neurol. 2011; 68: 594–598.

Chiò A, Borghero G, Restagno G, Mora G, Drepper C, Traynor BJ, et al. Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72. Brain 2012; 135: 784–793.

Chio A, Logroscino G, Hardiman O, Swingler R, Mitchell D, Beghi E, et al. Prognostic factors in ALS: A critical review. Amyotroph. Lateral Scler. 2009; 10: 310–323.

Chio A, Mora G, Leone M, Mazzini L, Cocito D, Giordana MT, et al. Early symptom progression rate is related to ALS outcome: a prospective population-based study. Neurology 2002; 59: 99–103.

Coelho MB, Attig J, Bellora N, König J, Hallegger M, Kayikci M, et al. Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. EMBO J. 2015; 34: 653–68.

Cohen-Hadad Y, Altarescu G, Eldar-Geva T, Levi-Lahad E, Zhang M, Rogaeva E, et al. Marked Differences in C9orf72 Methylation Status and Isoform Expression between C9/ALS Human Embryonic and Induced Pluripotent Stem Cells. Stem cell reports 2016

Colombrita C, Zennaro E, Fallini C, Weber M, Sommacal A, Buratti E, et al. TDP-43 is recruited to stress granules in conditions of oxidative insult. J. Neurochem. 2009; 111: 1051–1061.

Conicella AE, Zerze GH, Mittal J, Fawzi NL. ALS Mutations Disrupt Phase

Separation Mediated by ??-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. Structure 2016; 24: 1537–1549.

Cooper-Knock J, Hewitt C, Highley JR, Brockington A, Milano A, Man S, et al. Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. Brain 2012; 135: 751–764.

Cooper-Knock J, Higginbottom A, Stopford MJ, Highley JR, Ince PG, Wharton SB, et al. Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. Acta Neuropathol. 2015; 130: 63–75.

Cooper-Knock J, Kirby J, Highley R, Shaw PJ. The Spectrum of C9orf72-mediated Neurodegeneration and Amyotrophic Lateral Sclerosis. Neurotherapeutics 2015; 12: 326–339.

Cooper-Knock J, Walsh MJ, Higginbottom A, Highley JR, Dickman MJ, Edbauer D, et al. Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. Brain 2014; 137: 2040–2051.

Crawford TQ, Roelink H. The Notch response inhibitor DAPT enhances neuronal differentiation in embryonic stem cell-derived embryoid bodies independently of Sonic Hedgehog signaling. Dev. Dyn. 2007; 236: 886–892.

D'Alton S, Altshuler M, Cannon A, Dickson DW, Petrucelli L, Lewis J. Divergent phenotypes in mutant TDP-43 transgenic mice highlight potential confounds in TDP-43 transgenic modeling. PLoS One 2014; 9

Dafinca R, Scaber J, Ababneh N, Lalic T, Weir G, Christian H, et al. C9orf72 Hexanucleotide Expansions Are Associated with Altered Endoplasmic Reticulum Calcium Homeostasis and Stress Granule Formation in Induced Pluripotent Stem Cell-Derived Neurons from Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Demen. Stem Cells 2016; 34: 2063–78.

Dangoumau A, Marouillat S, Burlaud Gaillard J, Uzbekov R, Veyrat-Durebex C, Blasco H, et al. Inhibition of pathogenic mutant SOD1 aggregation in cultured motor neuronal cells by prevention of its SUMOylation on lysine 75. Neurodegener. Dis. 2016; 16: 161–171.

Dasen JS, De Camilli A, Wang B, Tucker PW, Jessell TM. Hox Repertoires for Motor Neuron Diversity and Connectivity Gated by a Single Accessory Factor, FoxP1. Cell 2008; 134: 304–316.

Davidson Y, Robinson AC, Liu X, Wu D, Troakes C, Rollinson S, et al. Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. Neuropathol. Appl. Neurobiol. 2016; 42: 242–54.

Van Deerlin VM, Leverenz JB, Bekris LM, Bird TD, Yuan W, Elman LB, et al.

TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. Lancet Neurol. 2008; 7: 409–416.

DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. Neuron 2011; 72: 245–256.

Deng H-X, Chen W, Hong S-T, Boycott KM, Gorrie GH, Siddique N, et al. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature 2011; 477: 211–5.

Devlin A-C, Burr K, Borooah S, Foster JD, Cleary EM, Geti I, et al. Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. Nat. Commun. 2015; 6: 5999.

Dewey CM, Cenik B, Sephton CF, Johnson BA, Herz J, Yu G. TDP-43 aggregation in neurodegeneration: Are stress granules the key? Brain Res. 2012; 1462: 16–25.

Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, et al. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief. Bioinform. 2013; 14: 671–83.

Dimos JT, Rodolfa KT, Niakan KK, Weisenthal LM, Mitsumoto H, Chung W, et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science 2008; 321: 1218–21.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013; 29: 15–21.

Donnelly CJ, Zhang PW, Pham JT, Heusler AR, Mistry NA, Vidensky S, et al. RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. Neuron 2013; 80: 415–428.

Durston AJ, Timmermans JP, Hage WJ, Hendriks HF, de Vries NJ, Heideveld M, et al. Retinoic acid causes an anteroposterior transformation in the developing central nervous system. Nature 1989; 340: 140–144.

Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, Rätsch G, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 2013; 10: 1185–91.

Esanov R, Belle KC, van Blitterswijk M, Belzil V V., Rademakers R, Dickson DW, et al. C9orf72 promoter hypermethylation is reduced while hydroxymethylation is acquired during reprogramming of ALS patient cells. Exp. Neurol. 2016; 277: 171–177.

Fallini C, Bassell GJ, Rossoll W. The ALS disease protein TDP-43 is actively transported in motor neuron axons and regulates axon outgrowth. Hum. Mol. Genet. 2012; 21: 3703–3718.

Farg MA, Sundaramoorthy V, Sultana JM, Yang S, Atkinson RAK, Levina V, et al. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. Hum. Mol. Genet. 2014; 23: 3579–3595.

Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 1976; 260: 500–507.

Fleige S, Pfaffl MW. RNA integrity and the effect on the real-time qRT-PCR performance. Mol. Aspects Med. 2006; 27: 126–139.

Le Forestier N, Maisonobe T, Piquard a, Rivaud S, Crevier-Buchman L, Salachas F, et al. Does primary lateral sclerosis exist? A study of 20 patients and a review of the literature. Brain 2001; 124: 1989–1999.

Freibaum BD, Chitta RK, High AA, Taylor JP. Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. J. Proteome Res. 2010; 9: 1104–1120.

Galimberti D, Arosio B, Fenoglio C, Serpente M, Cioffi SMG, Bonsi R, et al. Incomplete penetrance of the C9ORF72 hexanucleotide repeat expansions: Frequency in a cohort of geriatric non-demented subjects. J. Alzheimer's Dis. 2014; 39: 19–22.

Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PEA, Caulfield T, et al. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. Acta Neuropathol. 2013; 126: 829–844.

Gendron TF, van Blitterswijk M, Bieniek KF, Daughrity LM, Jiang J, Rush BK, et al. Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. Acta Neuropathol. 2015; 130: 559–573.

Gibson SB, Downie JM, Tsetsou S, Feusier JE, Figueroa KP, Bromberg MB, et al. The evolving genetic risk for sporadic ALS. Neurology 2017: 10.1212/WNL.0000000000004109.

Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, et al. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. Lancet. Neurol. 2012; 11: 54–65.

Gijselinck I, Van Mossevelde S, van der Zee J, Sieben A, Engelborghs S, De Bleecker J, et al. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol. Psychiatry 2015; 21: 1–13.

Giordana MT, Piccinini M, Grifoni S, De Marco G, Vercellino M, Magistrello M, et al. TDP-43 redistribution is an early event in sporadic amyotrophic lateral sclerosis. Brain Pathol 2010; 20: 351–360.

Goldman JS, Farmer JM, Wood EM, Johnson JK, Boxer A, Neuhaus J, et al. Comparison of family histories in FTLD subtypes and related tauopathies. Neurology 2005; 65: 1817–1819.

Gomez-Deza J, Lee Y-B, Troakes C, Nolan M, Al-Sarraj S, Gallo J-M, et al. Dipeptide repeat protein inclusions are rare in the spinal cord and almost absent from motor neurons in C9ORF72 mutant amyotrophic lateral sclerosis and are unlikely to cause their degeneration. Acta Neuropathol. Commun. 2015; 3: 38.

Gómez-Tortosa E, Gallego J, Guerrero-López R, Marcos A, Gil-Neciga E, José Sainz M, et al. C9ORF72 hexanucleotide expansions of 20-22 repeats are associated with frontotemporal deterioration. Neurology 2013; 80: 366–370.

Gonzàlez-Porta M, Frankish A, Rung J, Harrow J, Brazma A. Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. Genome Biol. 2013; 14: R70.

Gordon PH, Cheng B, Katz IB, Pinto M, Hays AP, Mitsumoto H, et al. The natural history of primary lateral sclerosis. Neurology 2006; 66: 647–53.

Greene E, Mahishi L, Entezam A, Kumari D, Usdin K. Repeat-induced epigenetic changes in intron 1 of the frataxin gene and its consequences in Friedreich ataxia. Nucleic Acids Res. 2007; 35: 3383–3390.

Gregory RI, Yan K-P, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. Nature 2004; 432: 235–240.

Guerreiro R, Brás J, Hardy J. SnapShot: Genetics of ALS and FTD. Cell 2015; 160: 798.e1.

Haeusler AR, Donnelly CJ, Periz G, Simko EAJ, Shaw PG, Kim M-S, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 2014; 507: 195–200.

Haidet-Phillips AM, Gross SK, Williams T, Tuteja A, Sherman A, Ko M, et al. Altered astrocytic expression of TDP-43 does not influence motor neuron survival. Exp. Neurol. 2013; 250: 250–9.

Handel AE, Chintawar S, Lalic T, Whiteley E, Vowles J, Giustacchini A, et al.

Assessing similarity to primary tissue and cortical layer identity in induced pluripotent stem cell-derived cortical neurons through single-cell transcriptomics. Hum. Mol. Genet. 2016; 25: 989–1000.

Hardwick SA, Chen WY, Wong T, Deveson IW, Blackburn J, Andersen SB, et al. Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat. Methods 2016; 13: 792–8.

Harland R. Neural induction. Curr. Opin. Genet. Dev. 2000; 10: 357–362.

Harms MB, Neumann D, Benitez BA, Cooper B, Carrell D, Racette BA, et al. Parkinson disease is not associated with C9ORF72 repeat expansions. Neurobiol. Aging 2013; 34

Hayer KE, Pizarro A, Lahens NF, Hogenesch JB, Grant GR. Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data. Bioinformatics 2015; 31: 3938–3945.

Hazelett DJ, Chang J-C, Lakeland DL, Morton DB. Comparison of parallel high-throughput RNA sequencing between knockout of TDP-43 and its overexpression reveals primarily nonreciprocal and nonoverlapping gene expression changes in the central nervous system of Drosophila. G3 (Bethesda). 2012; 2: 789–802.

Hebron ML, Lonskaya I, Sharpe K, Weerasinghe PPK, Algarzae NK, Shekoyan AR, et al. Parkin ubiquitinates tar-DNA binding protein-43 (TDP-43) and promotes its cytosolic accumulation via interaction with histone deacetylase 6 (HDAC6). J. Biol. Chem. 2013; 288: 4103–4115.

Herdewyn S, Cirillo C, Van Den Bosch L, Robberecht W, Vanden Berghe P, Van Damme P. Prevention of intestinal obstruction reveals progressive neurodegeneration in mutant TDP-43 (A315T) mice. Mol. Neurodegener. 2014; 9: 24.

Higashi S, Iseki E, Yamamoto R, Minegishi M, Hino H, Fujisawa K, et al. Concurrence of TDP-43, tau and alpha-synuclein pathology in brains of Alzheimer's disease and dementia with Lewy bodies. Brain Res. 2007; 1184: 284–94.

Ho R, Sances S, Gowing G, Amoroso MW, O'Rourke JG, Sahabian A, et al. ALS disrupts spinal motor neuron maturation and aging pathways within gene co-expression networks. Nat. Neurosci. 2016; 19: 1256–67.

Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, et al. Structure of a Ribonucleic Acid. Science (80-.). 1965; 147: 1462–1465.

Honda D, Ishigaki S, Iguchi Y, Fujioka Y, Udagawa T, Masuda A, et al. The ALS/FTLD-related RNA-binding proteins TDP-43 and FUS have common downstream RNA targets in cortical neurons. FEBS Open Bio 2014; 4: 1–10.

Honti F, Meader S, Webber C. Unbiased Functional Clustering of Gene Variants with a Phenotypic-Linkage Network. PLoS Comput. Biol. 2014; 10

Hosler BA, Siddique T, Sapp PC, Sailor W, Huang MC, Hossain A, et al. Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22. JAMA 2000; 284: 1664–9.

Hsiung G-YR, DeJesus-Hernandez M, Feldman HH, Sengdy P, Bouchard-Kerr P, Dwosh E, et al. Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p. Brain 2012; 135: 709–22.

Hu B-Y, Zhang S-C. Differentiation of spinal motor neurons from pluripotent human stem cells. Nat. Protoc. 2009; 4: 1295–304.

Igaz LM, Kwong LK, Lee EB, Chen-Plotkin A, Swanson E, Unger T, et al. Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. J. Clin. Invest. 2011; 121: 726–738.

Igaz LM, Kwong LK, Xu Y, Truax AC, Uryu K, Neumann M, et al. Enrichment of C-Terminal Fragments in TAR DNA-Binding Protein-43 Cytoplasmic Inclusions in Brain but not in Spinal Cord of Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Am. J. Pathol. 2008; 173: 182–194.

Iguchi Y, Katsuno M, Niwa J, Takagi S, Ishigaki S, Ikenaka K, et al. Loss of TDP-43 causes age-dependent progressive motor neuron degeneration. Brain 2013; 136: 1371–82.

Ince PG, Evans J, Knopp M, Forster G, Hamdalla HHM, Wharton SB, et al. Corticospinal tract degeneration in the progressive muscular atrophy variant of ALS. Neurology 2003; 60: 1252–8.

Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods 2014; 11: 163–166.

De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci 2014; 17: 1156–1163.

Janssens J, Wils H, Kleinberger G, Joris G, Cuijt I, Ceuterick-de Groote C, et al. Overexpression of ALS-associated p.M337V human TDP-43 in mice worsens disease features compared to wild-type human TDP-43 mice. Mol. Neurobiol. 2013; 48: 22–35.

Jessell TM, Sanes JR. Development: The decade of the developing brain. Curr. Opin. Neurobiol. 2000; 10: 599–611.

Jiang L-L, Zhao J, Yin X-F, He W-T, Yang H, Che M-X, et al. Two mutations

G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. Sci. Rep. 2016; 6: 23928.

Johnson JO, Pioro EP, Boehringer A, Chia R, Feit H, Renton AE, et al. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. Nat. Neurosci. 2014; 17: 664–6.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008; 36

Johnston CA, Stanton BR, Turner MR, Gray R, Blunt AH-M, Butt D, et al. Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. J. Neurol. 2006; 253: 1642–3.

Jung H, Lacombe J, Mazzoni EO, Liem KF, Grinstein J, Mahony S, et al. Global Control of Motor Neuron Topography Mediated by the Repressive Actions of a Single Hox Gene. Neuron 2010; 67: 781–796.

Kabashi E, Lin L, Tradewell ML, Dion PA, Bercier V, Bourgouin P, et al. Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. Hum. Mol. Genet. 2009; 19: 671–683.

Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Vande Velde C, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat. Genet. 2008; 40: 572–574.

Kallenbach S, Khantane S, Carroll P, Gayet O, Alonso S, Henderson CE, et al. Changes in subcellular distribution of protocadherin gamma proteins accompany maturation of spinal neurons. J. Neurosci. Res. 2003; 72: 549–556.

Kanitz A, Gypas F, Gruber ARJ, Gruber ARJ, Martin G, Zavolan M. Comparative assessment of methods for the computational inference of transcript isoform abundance from RNA-seq data. Genome Biol. 2015; 16: 150.

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, et al. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. Cell 2012; 149: 753–767.

Keller MF, Ferrucci L, Singleton AB, Tienari PJ, Laaksovirta H, Restagno G, et al. Genome-wide analysis of the heritability of amyotrophic lateral sclerosis. JAMA Neurol. 2014; 71: 1123–34.

Khosravi B, Hartmann H, May S, Möhl C, Ederle H, Michaelsen M, et al. Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72 ALS/FTLD. Hum. Mol. Genet. 2016

Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 2015; 12: 357–60.

Kim HJ, Kim NC, Wang Y-D, Scarborough EA, Moore J, Diaz Z, et al. Mutations

in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature 2013; 495: 467–73.

Kiskinis E, Sandoe J, Williams LA, Boulting GL, Moccia R, Wainger BJ, et al. Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. Cell Stem Cell 2014; 14: 781–795.

Klingensmith J, Ang S-L, Bachiller D, Rossant J. Neural Induction and Patterning in the Mouse in the Absence of the Node and Its Derivatives. Dev. Biol. 1999; 216: 535–549.

Konieczny P, Stepniak-Konieczna E, Sobczak K. MBNL proteins and their target RNAs, interaction and splicing regulation. Nucleic Acids Res. 2014; 42: 10873–10887.

Kraemer BC, Schuck T, Wheeler JM, Robinson LC, Trojanowski JQ, Lee VMY, et al. Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. Acta Neuropathol. 2010; 119: 409–19.

Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science (80-.). 2014; 345: 1139–1145.

Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai SL, Myllykangas L, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: A genome-wide association study. Lancet Neurol. 2010; 9: 978–985.

Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li H-R, et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. Proc. Natl. Acad. Sci. U. S. A. 2013; 110: E4530-9.

Lagier-Tourenne C, Polymenidou M, Hutt KR, Vu AQ, Baughn M, Huelga SC, et al. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat. Neurosci. 2012; 15: 1488–1497.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001; 409: 860–921.

Lee SE, Khazenzon AM, Trujillo AJ, Guo CC, Yokoyama JS, Sha SJ, et al. Altered network connectivity in frontotemporal dementia with C9orf72 hexanucleotide repeat expansion. Brain 2014; 137: 3047–3060.

Leone P, Shin EC, Perosa F, Vacca A, Dammacco F, Racanelli V. MHC class i antigen processing and presenting machinery: Organization, function, and defects in tumor cells. J. Natl. Cancer Inst. 2013; 105: 1172–1187.

Levine TP, Daniels RD, Gatta AT, Wong LH, Hayes MJ. The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics 2013; 29: 499–503.

Li TM, Alberman E, Swash M. Comparison of sporadic and familial disease amongst 580 cases of motor neuron disease. J. Neurol. Neurosurg. Psychiatry 1988; 51: 778–784.

Liao Y, Smyth GK, Shi W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. 2013; 41

Ling JP, Pletnikova O, Troncoso JC, Wong PC. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. Science 2015; 349: 650–5.

Ling S-C, Albuquerque CP, Seok J, Lagier-tourenne C, Tokunaga S, Han JS, et al. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proc. Natl. Acad. Sci. U. S. A. 2010; 107: 13318–13323.

Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in als and FTD: Disrupted RNA and protein homeostasis. Neuron 2013; 79: 416–438.

Liu-Yesucevitz L, Bilgutay A, Zhang Y-J, Vanderweyde T, Vanderwyde T, Citro A, et al. Tar DNA binding protein-43 (TDP-43) associates with stress granules: analysis of cultured cells and pathological brain tissue. PLoS One 2010; 5: e13250.

Liu-Yesucevitz L, Lin AY, Ebata A, Boon JY, Reid W, Xu Y-F, et al. ALS-Linked Mutations Enlarge TDP-43-Enriched Neuronal RNA Granules in the Dendritic Arbor. J. Neurosci. 2014; 34: 4167–4174.

Liu EY, Russ J, Wu K, Neal D, Suh E, McNally AG, et al. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol. 2014; 128: 525–41.

Liu JP, Laufer E, Jessell TM. Assigning the positional identity of spinal motor neurons: Rostrocaudal patterning of Hox-c expression by FGFs, Gdf11, and retinoids. Neuron 2001; 32: 997–1012.

Liu R, Loraine AE, Dickerson JA. Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. BMC Bioinformatics 2014; 15: 364.

Logroscino G, Traynor BJ, Hardiman O, Chiò A, Mitchell D, Swingler RJ, et al. Incidence of amyotrophic lateral sclerosis in Europe. J. Neurol. Neurosurg. Psychiatry 2010; 81: 385–90.

Lomen-Hoerth C, Anderson T, Miller B. The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. Neurology 2002; 59: 1077–9.

Lopez-Gonzalez R, Lu Y, Gendron TF, Karydas A, Tran H, Yang D, et al. Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and

Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron 2016; 92: 383–391.

Love MI, Anders S, Huber W. Differential analysis of count data - the DESeq2 package. 2014.

Love MI, Hogenesch JB, Irizarry RA. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. Nat. Biotechnol. 2016

Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuani C, Buratti E, et al. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. Nat. Struct. Mol. Biol. 2013; 20: 1443–9.

Mackenzie IRA, Bigio EH, Ince PG, Geser F, Neumann M, Cairns NJ, et al. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann. Neurol. 2007; 61: 427–434.

Mackenzie IRA, Frick P, Grässer FA, Gendron TF, Petrucelli L, Cashman NR, et al. Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathol. 2015; 130: 845–61.

Mackenzie IRA, Neumann M, Baborie A, Sampathu DM, Du Plessis D, Jaros E, et al. A harmonized classification system for FTLD-TDP pathology. Acta Neuropathol. 2011; 122: 111–113.

MacKenzie IR, Arzberger T, Kremmer E, Troost D, Lorenzl S, Mori K, et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: Clinico-pathological correlations. Acta Neuropathol. 2013; 126: 859–879.

Macnair L, Xiao S, Miletic D, Ghani M, Julien JP, Keith J, et al. MTHFSD and DDX58 are novel RNA-binding proteins abnormally regulated in amyotrophic lateral sclerosis. Brain 2016; 139: 86–100.

Majounie E, Abramzon Y, Renton AE, Keller MF, Traynor BJ, Singleton AB. Large C9orf72 repeat expansions are not a common cause of Parkinson's disease. Neurobiol. Aging 2012; 33

Majounie E, Renton AE, Mok K, Dopper EGP, Waite A, Rollinson S, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. Lancet Neurol. 2012; 11: 323–330.

Marat AL, Dokainish H, McPherson PS. DENN domain proteins: Regulators of Rab GTPases. J. Biol. Chem. 2011; 286: 13791–13800.

Marin B, Boumédiene F, Logroscino G, Couratier P, Babron M-C, Leutenegger

AL, et al. Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. Int. J. Epidemiol. 2017; 46: 57–74.

Maury Y, Côme J, Piskorowski RA, Salah-Mohellibi N, Chevaleyre V, Peschanski M, et al. Combinatorial analysis of developmental cues efficiently converts human pluripotent stem cells into multiple neuronal subtypes. Nat Biotechnol 2014; 33: 89–96.

McDonald KK, Aulas A, Destroismaisons L, Pickles S, Beleac E, Camu W, et al. TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. Hum. Mol. Genet. 2011; 20: 1400–1410.

McMillan CT, Russ J, Wood EM, Irwin DJ, Grossman M, Mccluskey L, et al. C9orf72 promoter hypermethylation is neuroprotective: Neuroimaging and neuropathologic evidence. Neurology 2015; 84: 1622–1630.

Merkle FT, Eggan K. Modeling human disease with pluripotent stem cells: From genome association to function. Cell Stem Cell 2013; 12: 656–668.

Miller RG, Mitchell JD, Lyon M, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). Cochrane Database Syst. Rev. 2007

Mizielinska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P, et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. Acta Neuropathol. 2013; 126: 845–857.

Moisse K, Volkening K, Leystra-Lantz C, Welch I, Hill T, Strong MJ. Divergent patterns of cytosolic TDP-43 and neuronal progranulin expression following axotomy: Implications for TDP-43 in the physiological response to neuronal injury. Brain Res. 2009; 1249: 202–211.

Mompean M, Romano V, Pantoja-Uceda D, Stuani C, Baralle FE, Buratti E, et al. The TDP-43 N-Terminal Domain Structure at High Resolution. FEBS J. 2016: 1–19.

Morahan JM, Yu B, Trent RJ, Pamphlett R. A genome-wide analysis of brain DNA methylation identifies new candidate genes for sporadic amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. 2009; 10: 418–429.

Mori K, Weng S-M, Arzberger T, May S, Rentzsch K, Kremmer E, et al. The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. Science (80-.). 2013; 339: 1335–1338.

Morita M, Al-Chalabi A, Andersen PM, Hosler B, Sapp P, Englund E, et al. A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology 2006; 66: 839–844.

Muhr J, Graziano E, Wilson S, Jessell TM, Edlund T. Convergent inductive signals specify midbrain, hindbrain, and spinal cord identity in Gastrula stage chick embryos. Neuron 1999; 23: 689–702.

Murphy NA, Arthur KC, Tienari PJ, Houlden H, Chiò A, Traynor BJ. Age-related penetrance of the C9orf72 repeat expansion. Sci. Rep. 2017: 1–7.

Murray ME, DeJesus-Hernandez M, Rutherford NJ, Baker M, Duara R, Graff-Radford NR, et al. Clinical and neuropathologic heterogeneity of c9FTD/ALS associated with hexanucleotide repeat expansion in C9ORF72. Acta Neuropathol. 2011; 122: 673–90.

Narayanan RK, Mangelsdorf M, Panwar A, Butler TJ, Noakes PG, Wallace RH. Identification of RNA bound to the TDP-43 ribonucleoprotein complex in the adult mouse brain. Amyotroph. Lateral Scler. Front. Degener. 2012; 8421: 1–9.

Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 2006; 314: 130–3.

Niblock M, Smith BN, Lee Y-B, Sardone V, Topp S, Troakes C, et al. Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. Acta Neuropathol. Commun. 2016; 4: 18.

Nicholson SJ, Witherden AS, Hafezparast M, Martin JE, Fisher EM. Mice, the motor system, and human motor neuron pathology. Mamm. Genome 2000; 11: 1041–52.

Nonaka T, Masuda-Suzukake M, Arai T, Hasegawa Y, Akatsu H, Obi T, et al. Prion-like Properties of Pathological TDP-43 Aggregates from Diseased Brains. Cell Rep. 2013; 4: 124–134.

Novitch BG, Wichterle H, Jessell TM, Sockanathan S. A requirement for retinoic acid-mediated transcriptional activation in ventral neural patterning and motor neuron specification. Neuron 2003; 40: 81–95.

Nyrén P, Lundin A. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. Anal. Biochem. 1985; 151: 504–509.

ORourke JG, Bogdanik L, Yanez A, Lall D, Wolf AJ, Muhammad AKMG, et al. C9orf72 is required for proper macrophage and microglial function in mice. Science (80-.). 2016; 351: 1324–1329.

Patro R, Duggal G, Kingsford C. Salmon: Accurate, Versatile and Ultrafast Quantification from RNA-seq Data using Lightweight-Alignment. bioRxiv 2015: 21592.

Patro R, Mount SM, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nat. Biotechnol.

2014; 32: 462–464.

Peng RD, Matsui E. The Art of Data Science. Skybrude Consulting, LLC; 2015.

Petkau TL, Leavitt BR. Progranulin in neurodegenerative disease. Trends Neurosci. 2014; 37: 388–398.

Philips T, De Muynck L, Thu HNT, Weynants B, Vanacker P, Dhondt J, et al. Microglial upregulation of progranulin as a marker of motor neuron degeneration. J. Neuropathol. Exp. Neurol. 2010; 69: 1191–1200.

Pimentel HJ, Bray N, Puente S, Melsted P, Pachter L. Differential analysis of RNA-Seq incorporating quantification uncertainty. bioRxiv 2016: 58164.

Pinto AC, Evangelista T, Carvalho M, Alves MA, Sales Luís ML. Respiratory assistance with a non-invasive ventilator (Bipap) in MND/ALS patients: Survival rates in a controlled trial. J. Neurol. Sci. 1995; 129: 19–26.

Polymenidou M, Lagier-tourenne C, Hutt KR, Huelga SC, Moran J, Liang TY, et al. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat. Neurosci. 2011; 14: 459–468.

Proudfoot M, Gutowski NJ, Edbauer D, Hilton DA, Stephens M, Rankin J, et al. Early dipeptide repeat pathology in a frontotemporal dementia kindred with C9ORF72 mutation and intellectual disability. Acta Neuropathol. 2014; 127: 451–458.

Prudencio M, Belzil V V, Batra R, Ross CA, Gendron TF, Pregent LJ, et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. Nat. Neurosci. 2015; 18: 1175–1182.

Prudencio M, Jansen-West KR, Lee WC, Gendron TF, Zhang Y-J, Xu Y-F, et al. Misregulation of human sortilin splicing leads to the generation of a nonfunctional progranulin receptor. Proc. Natl. Acad. Sci. U. S. A. 2012; 109: 21510–5.

Qu Q, Li D, Louis KR, Li X, Yang H, Sun Q, et al. High-efficiency motor neuron differentiation from human pluripotent stem cells and the function of Islet-1. Nat. Commun. 2014; 5: 3449.

Rademakers R, Neumann M, Mackenzie IR. Advances in understanding the molecular basis of frontotemporal dementia. Nat. Rev. Neurol. 2012; 8: 423–34.

Raoult D. Technology-driven research will dominate hypothesis-driven research: the future of microbiology. Future Microbiol. 2010; 5: 135–7.

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

Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al.

A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 2011; 72: 257–268.

Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE. Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology 2005; 65: 586–590.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2009; 26: 139–140.

Roche JC, Rojas-Garcia R, Scott KM, Scotton W, Ellis CE, Burman R, et al. A proposed staging system for amyotrophic lateral sclerosis. Brain 2012; 135: 847–52.

Rogers MF, Thomas J, Reddy AS, Ben-Hur A. SpliceGrapher: detecting patterns of alternative splicing from RNA-Seq data in the context of gene models and EST data. Genome Biol. 2012; 13: R4.

Rohrer JD, Isaacs AM, Mizlienska S, Mead S, Lashley T, Wray S, et al. C9orf72 expansions in frontotemporal dementia and amyotrophic lateral sclerosis. Lancet Neurol. 2015; 14: 291–301.

Rohrer JD, Nicholas JM, Cash DM, van Swieten J, Dopper E, Jiskoot L, et al. Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: A cross-sectional analysis. Lancet Neurol. 2015; 14: 253–262.

Rot G, Wang Z, Huppertz I, Modic M, Lenče T, Hallegger M, et al. High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43. Cell Rep. 2017; 19: 1056–1067.

Rothstein JD. TDP-43 in amyotrophic lateral sclerosis: pathophysiology or patho-babel? Ann. Neurol. 2007; 61: 382–4.

Russ J, Liu EY, Wu K, Neal D, Suh E, Irwin DJ, et al. Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. Acta Neuropathol. 2015; 129: 39–52.

Sances S, Bruijn LI, Chandran S, Eggan K, Ho R, Klim JR, et al. Modeling ALS with motor neurons derived from human induced pluripotent stem cells. Nat. Neurosci. 2016; 16: 542–53.

Sanger F, Nicklen S. DNA sequencing with chain-terminating. PNAS 1977; 74: 5463–5467.

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

Sasaguri H, Chew J, Xu YF, Gendron TF, Garrett A, Lee CW, et al. The extreme N-terminus of TDP-43 mediates the cytoplasmic aggregation of TDP-43 and associated toxicity in vivo. Brain Res. 2016; 1647: 57–64.

Sato T, Takeuchi S, Saito A, Ding W, Bamba H, Matsuura H, et al. Axonal ligation induces transient redistribution of TDP-43 in brainstem motor neurons. Neuroscience 2009; 164: 1565–1578.

Schipper LJ, Raaphorst J, Aronica E, Baas F, de Haan R, de Visser M, et al. Prevalence of brain and spinal cord inclusions, including dipeptide repeat proteins, in patients with the C9ORF72 hexanucleotide repeat expansion: A systematic neuropathological review. Neuropathol. Appl. Neurobiol. 2015: n/a-n/a.

Schludi MH, May S, Grässer FA, Rentzsch K, Kremmer E, Küpper C, et al. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. Acta Neuropathol. 2015; 130: 537–555.

Sellier C, Campanari M-L, Julie Corbier C, Gaucherot A, Kolb-Cheynel I, Oulad-Abdelghani M, et al. Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. EMBO J. 2016; 35: 1–22.

Shan X, Chiang P-M, Price DL, Wong PC. Altered distributions of Gemini of coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice. Proc. Natl. Acad. Sci. U. S. A. 2010; 107: 16325–30.

Shatunov A, Mok K, Newhouse S, Weale ME, Smith B, Vance C, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: A genome-wide association study. Lancet Neurol. 2010; 9: 986–994.

Shen S, Park JW, Huang J, Dittmar KA, Lu ZX, Zhou Q, et al. MATS: A Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucleic Acids Res. 2012; 40

Shen S, Park JW, Lu Z, Lin L, Henry MD, Wu YN, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc. Natl. Acad. Sci. U. S. A. 2014; 111: E5593-601.

Shi Y, Kirwan P, Livesey FJ. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat. Protoc. 2012; 7: 1836–1846.

Shoesmith CL, Findlater K, Rowe A, Strong MJ. Prognosis of amyotrophic lateral sclerosis with respiratory onset. J. Neurol. Neurosurg. Psychiatry 2007; 78: 629–31.

Simón-Sánchez J, Dopper EGP, Cohn-Hokke PE, Hukema RK, Nicolaou N, Seelaar H, et al. The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. Brain 2012; 135: 723–735.

Sims D, Ilott NE, Sansom SN, Sudbery IM, Johnson JS, Fawcett KA, et al. CGAT: Computational genomics analysis toolkit. Bioinformatics 2014; 30: 1290–1291.

Smith BN, Newhouse S, Shatunov A, Vance C, Topp S, Johnson L, et al. The C9ORF72 expansion mutation is a common cause of ALS+/−FTD in Europe and has a single founder. Eur. J. Hum. Genet. 2013; 21: 102–108.

Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. Nucleic Acids Res. 1985; 13: 2399–412.

Smith TS, Heger A, Sudbery I. UMI-tools: Modelling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. 2016.

Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 2015; 4: 1521.

Song S, Miranda CJ, Braun L, Meyer K, Frakes AE, Ferraiuolo L, et al. Major histocompatibility complex class I molecules protect motor neurons from astrocyte-induced toxicity in amyotrophic lateral sclerosis. Nat. Med. 2016; 22: 397–403.

Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, Rogelj B, et al. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science 2008; 319: 1668–72.

Stallings NR, Puttaparthi K, Luther CM, Burns DK, Elliott JL. Progressive motor weakness in transgenic mice expressing human TDP-43. Neurobiol. Dis. 2010; 40: 404–414.

Sternberger L a, Sternberger NH. Monoclonal antibodies distinguish phosphorylated and nonphosphorylated forms of neurofilaments in situ. Proc. Natl. Acad. Sci. U. S. A. 1983; 80: 6126–6130.

Stewart H, Rutherford NJ, Briemberg H, Krieger C, Cashman N, Fabros M, et al. Clinical and pathological features of amyotrophic lateral sclerosis caused by mutation in the C9ORF72 gene on chromosome 9p. Acta Neuropathol. 2012; 123: 409–417.

Sudbery IM, Scaber J, Sims D, Heger A. Pipeline_iCLIP - Branch JS-HITS-CLIP [Internet]. 2016[cited 2016 Nov 30] Available from: https://github.com/jscaber/Pipeline_iCLIP/tree/JS-HITS-CLIP

Sultan M, Amstislavskiy V, Risch T, Schuette M, Dökel S, Ralser M, et al.

Influence of RNA extraction methods and library selection schemes on RNA-seq data. BMC Genomics 2014; 15: 675.

Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, et al. DNA methylation represses FMR-1 transcription in fragile x syndrome. Hum. Mol. Genet. 1992; 1: 397–400.

Swank RL, Putnam T. Amyotrophic Lateral Sclerosis and Related Conditions. Arch. Neurol. Psychiatry 1943; 49: 151.

Swarup V, Phaneuf D, Dupré N, Petri S, Strong M, Kriz J, et al. Deregulation of TDP-43 in amyotrophic lateral sclerosis triggers nuclear factor κB-mediated pathogenic pathways. J. Exp. Med. 2011; 208: 2429–47.

Swerdlow H, Wu S, Harke H, Dovichi NJ. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. J. Chromatogr. A 1990; 516: 61–67.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 2007; 131: 861–872.

Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 2006; 126: 663–676.

Talbot K. Motor neuron disease: the bare essentials. Pract. Neurol. 2009; 9: 303–9.

Tan RH, Devenney E, Dobson-Stone C, Kwok JB, Hodges JR, Kiernan MC, et al. Cerebellar integrity in the amyotrophic lateral sclerosis-frontotemporal dementia continuum. PLoS One 2014; 9: e105632.

Tanabe Y, William C, Jessell TM. Specification of motor neuron identity by the MNR2 homeodomain protein. Cell 1998; 95: 67–80.

Thaler JP, Koo SJ, Kania A, Lettieri K, Andrews S, Cox C, et al. A Postmitotic Role for Isl-Class LIM Homeodomain Proteins in the Assignment of Visceral Spinal Motor Neuron Identity. Neuron 2004; 41: 337–350.

Thomas M, Alegre-Abarrategui J, Wade-Martins R. RNA dysfunction and aggrephagy at the centre of an amyotrophic lateral sclerosis/frontotemporal dementia disease continuum. Brain 2013; 136: 1345–1360.

Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, et al. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat. Neurosci. 2011; 14: 452–458.

Tong J, Huang C, Bi F, Wu Q, Huang B, Liu X, et al. Expression of ALS-linked TDP-43 mutant in astrocytes causes non-cell-autonomous motor neuron death in

rats. EMBO J. 2013; 32: 1917–26.

Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics 2009; 25: 1105–1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 2010; 28: 511–515.

Tsai K-J, Yang C, Fang Y, Cho K, Chien W, Wang W, et al. Elevated expression of TDP-43 in the forebrain of mice is sufficient to cause neurological and pathological phenotypes mimicking FTLD-U. J. Exp. Med. 2010; 207: 1661–1673.

Tsuji H, Arai T, Kametani F, Nonaka T, Yamashita M, Suzukake M, et al. Molecular analysis and biochemical classification of TDP-43 proteinopathy. Brain 2012; 135: 3380–3391.

Turner BJ, Bäumer D, Parkinson NJ, Scaber J, Ansorge O, Talbot K. TDP-43 expression in mouse models of amyotrophic lateral sclerosis and spinal muscular atrophy. BMC Neurosci. 2008; 9

Turner BJ, Talbot K. Transgenics, toxicity and therapeutics in rodent models of mutant SOD1-mediated familial ALS. Prog. Neurobiol. 2008; 85: 94–134.

Turner MR, Brockington A, Scaber J, Hollinger H, Marsden R, Shaw PJ, et al. Pattern of spread and prognosis in lower limb-onset ALS. Amyotroph. Lateral Scler. 2010; 11

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res. 2007; 35

Vance C, Al-Chalabi A, Ruddy D, Smith BN, Hu X, Sreedharan J, et al. Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. Brain 2006; 129: 868–76.

Vatsavayai SC, Yoon SJ, Gardner RC, Gendron TF, Vargas JNS, Trujillo A, et al. Timing and significance of pathological features in C9orf72 expansion-associated frontotemporal dementia. Brain 2016

Venter JC. The Sequence of the Human Genome. Science (80-.). 2001; 291: 1304–1351.

Volpicelli-Daley LA, Luk KC, Lee VM-Y. Addition of exogenous α-synuclein preformed fibrils to primary neuronal cultures to seed recruitment of endogenous α-synuclein to Lewy body and Lewy neurite-like aggregates. Nat. Protoc. 2014; 9: 2135–46.

Vucic S, Kiernan MC. Abnormalities in cortical and peripheral excitability in flail

arm variant amyotrophic lateral sclerosis. J. Neurol. Neurosurg. Psychiatry 2007; 78: 849–852.

Waite AJ, Bäumer D, East S, Neal J, Morris HR, Ansorge O, et al. Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. Neurobiol. Aging 2014; 35

Wang C, Wang Y, Hu M, Chai Z, Wu Q, Huang R, et al. Synaptotagmin-11 inhibits clathrin-mediated and bulk endocytosis. EMBO Rep. 2015: 1–17.

Wang H-Y, Wang I-F, Bose J, Shen C-KJ. Structural diversity and functional implications of the eukaryotic TDP gene family. Genomics 2004; 83: 130–139.

Wang X, Weiner JA, Levi S, Craig AM, Bradley A, Sanes JR. Gamma protocadherins are required for survival of spinal interneurons. Neuron 2002; 36: 843–854.

Watson JD, Crick FHC. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953; 171: 737–8.

Webster CP, Smith EF, Bauer CS, Moller A, Hautbergue GM, Ferraiuolo L, et al. The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. EMBO J. 2016; 35: 1656–76.

Wegorzewska I, Baloh RH. TDP-43-based animal models of neurodegeneration: New insights into ALS pathology and pathophysiology. Neurodegener. Dis. 2011; 8: 262–274.

Weiner JA, Wang X, Tapia JC, Sanes JR. Gamma protocadherins are required for synaptic development in the spinal cord. Proc. Natl. Acad. Sci. U. S. A. 2005; 102: 8–14.

Westergard T, Jensen BK, Wen X, Cai J, Kropf E, Iacovitti L, et al. Cell-to-Cell Transmission of Dipeptide Repeat Proteins Linked to C9orf72-ALS/FTD. Cell Rep. 2016; 17: 645–652.

Wijesekera LC, Mathers S, Talman P, Galtrey C, Parkinson MH, Ganesalingam J, et al. Natural history and clinical features of the flail arm and flail leg ALS variants. Neurology 2009; 72: 1087–94.

Wils H, Kleinberger G, Janssens J, Pereson S, Joris G, Cuijt I, et al. TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. Proc. Natl. Acad. Sci. U. S. A. 2010; 107: 3858–63.

Winton MJ, Igaz LM, Wong MM, Kwong LK, Trojanowski JQ, Lee VMY. Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. J.

Biol. Chem. 2008; 283: 13302–13309.

Wojciechowska M, Olejniczak M, Galka-Marciniak P, Jazurek M, Krzyzosiak WJ. RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. Nucleic Acids Res. 2014; 42: 11849–11864.

Wu L-S, Cheng W-C, Hou S-C, Yan Y-T, Jiang S-T, Shen C-KJ. TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. Genesis 2010; 48: 56–62.

Wu L-S, Cheng W-C, Shen C-KJ. Targeted depletion of TDP-43 expression in the spinal cord motor neurons leads to the development of amyotrophic lateral sclerosis-like phenotypes in mice. J. Biol. Chem. 2012; 287: 27335–44.

Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 2010; 26: 873–881.

Xi Z, Rainero I, Rubino E, Pinessi L, Bruni AC, Maletta RG, et al. Hypermethylation of the CpG-island near the C9orf72 G4C2-repeat expansion in FTLD patients. Hum. Mol. Genet. 2014; 23: 5630–5637.

Xi Z, Zhang M, Bruni AC, Maletta RG, Colao R, Fratta P, et al. The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. Acta Neuropathol. 2015; 129: 715–727.

Xi Z, Zinman L, Moreno D, Schymick J, Liang Y, Sato C, et al. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. Am. J. Hum. Genet. 2013; 92: 981–989.

Xiao S, MacNair L, McGoldrick P, McKeever PM, McLean JR, Zhang M, et al. Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. Ann. Neurol. 2015; 78: 568–583.

Xiao S, Sanelli T, Chiang H, Sun Y, Chakrabartty A, Keith J, et al. 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. 2015; 130: 49–61.

Xu Y-F, Gendron TF, Zhang Y-J, Lin W-L, D'Alton S, Sheng H, et al. Wild-Type Human TDP-43 Expression Causes TDP-43 Phosphorylation, Mitochondrial Aggregation, Motor Deficits, and Early Mortality in Transgenic Mice. J. Neurosci. 2010; 30: 10851–10859.

Yamada T, Placzek M, Tanaka H, Dodd J, Jessell TM. Control of cell pattern in the developing nervous system: Polarizing activity of the floor plate and notochord. Cell 1991; 64: 635–647.

Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al.

Ensembl 2016. Nucleic Acids Res. 2016; 44: D710–D716.

Yeo H, Kim H-W, Mo J, Lee D, Han S, Hong S, et al. Developmental expression and subcellular distribution of synaptotagmin 11 in rat hippocampus. Neuroscience 2012; 225: 35–43.

van der Zee J, Gijselinck I, Dillen L, Van Langenhove T, Theuns J, Engelborghs S, et al. A Pan-European Study of the C9orf72 Repeat Associated with FTLD: Geographic Prevalence, Genomic Instability, and Intermediate Repeats. Hum. Mutat. 2013; 34: 363–373.

Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. Nature 2015; 525: 56–61.

Zhang Y-J, Gendron TF, Xu Y-F, Ko L-W, Yen S-H, Petrucelli L. Phosphorylation regulates proteasomal-mediated degradation and solubility of TAR DNA binding protein-43 C-terminal fragments. Mol. Neurodegener. 2010; 5: 1–13.

Zhang Y-J, Xu Y, Dickey CA, Buratti E, Baralle F, Bailey R, et al. Progranulin mediates caspase-dependent cleavage of TAR DNA binding protein-43. J. Neurosci. 2007; 27: 10530–4.

Zhou J, Su P, Li D, Tsang S, Duan E, Wang F. High-efficiency induction of neural conversion in human ESCs and human induced pluripotent stem cells with a single chemical inhibitor of transforming growth factor beta superfamily receptors. Stem Cells 2010; 28: 1741–1750.

Zou Z-Y, Zhou Z-R, Che C-H, Liu C-Y, He R-L, Huang H-P. Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. J. Neurol. Neurosurg. Psychiatry 2017; 88: 540–549.

Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016; 44: D7–D19.

Appendix

Primers used for RNA sequencing validation

TARDBP	GTCGGGCAGTGGTTTTAATG	ACAACCCCACTGTCTACATTCC
C9orf72-all	CCCACTTCATAGAGTGTGTGTTG	TTCCATTCTCTCTGTGCCTTC
C9orf72-V2	GAAATCACACAGTGTTCCTGAAGAA	ATCTGCTTCATCCAGCTTTTATGA
PCDHGA7	TCGGAAGAGTCACCTGATCTTC	GCGGGGCTTGCTGAATAC
PCDHGB6	AAAATGCAGTGTGCCCCTAC	GGGGCTTGCGAAGTCAGAG
TOP1 (reference)	GTCCAAGCATAGCAACAGTGAAC	GCTCGAACCTTTTCCTCTTTTCG
RPL13A (reference)	AGCTCATGAGGCTACGGAAAC	TTTATTGGGCTCAGACCAGGAG
ATP5B (reference)	GCATTTGGGTGAGAGCACAG	TGGTGCACCAGAATCCAGTAC
PCDHGB7	TCATTGTCCAGCCACACAAG	CGGGGCTTGCTGTTTAAGAATC
ZNF471	CCTCCCAAGACACTGTTCTTC	TCACTAAGTCCTGGGGCATG
ZNF595	TCTCCCTGGGTTTTGTGATCTC	TGCACTGGTGAAAGGTCTTG
ZNF667	TCGCTTGGTCTTTCCTTTCG	TCTACCATCCAGGGTGCTTTC
MAP13K13	TGGCTATGAGAACCCCATGC	TTGGTCCTAACTGTGGCATCAG
SYT11	GACCTAGCTTTGATGTGTCACC	TGGTGGGTTCTTCTGCTTCTTC