RNA PROCESSING DYSFUNCTION IN TDP-ALS

Prasanth Sivakumar

Institute of Neurology, University College London

MRC Centre for Neuromuscular Diseases, London

Ph.D. Thesis

October 2018

DECLARATION

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

ACKNOWLEDGEMENTS

First of all, I would like thank Pietro Fratta for his supervision throughout my study. From providing me with the opportunity to join this investigation, all the way through to its conclusion, thank you for your ever-present assistance, advice and ideas. I would also like to express my gratitude to Lizzy Fisher for giving me the chance to join her lab, and when necessary helping to fund my attendance to a number of conferences around the world. These have greatly aided both the studies detailed in this thesis and also my all-round knowledge of the fields in TDP-43, ALS and transcriptomics.

I am grateful to Aga Ule and Nicol Birsa of the Fratta lab for their guidance and advice. They also ensured that my progress in learning R did not falter, and made sure to test me on my knowledge frequently with a wide variety of questions. Alongside them, I would like to acknowledge the rest of the Fisher and Fratta labs for their assistance.

Thank you to Jack Humphrey for continued help in my learning of bioinformatics, particularly in wrangling the unwieldy UCL computer cluster, and our combined efforts during this TDP-43 mouse model investigation. Thank you also to the rest of the Plagnol lab whom I had the pleasure of working with.

The work summarised in this thesis is the result of several collaborative efforts; the same was also true for the development of my PhD. Therefore, I thank members of the MRC Prion Unit, the Isaacs lab, the Tabrizi lab and the Sobell Department for their help, friendship and bothering throughout. Thank you too to the MRC Neuromuscular Diseases Centre for the opportunity to study for a PhD, the studentship and funding over the course of this process.

Finally, I thank my family for their love and support at all times during my study, I am truly grateful. அன்பும் ஆதரவும் ஆசிகளும் தந்து என்னை ஊக்குவித்த அம்மா, அப்பா, பிரணி மூவருக்கும் எனது மனமார்ந்த நன்றிகள்.

ABSTRACT

Background: RNA processing dysfunction has been implicated in the pathology of the neurodegenerative disease amyotrophic lateral sclerosis (ALS), not only due to causative TDP-43 disease mutations (TDP-ALS), but further the characteristic mislocalisation of crucial RNA-binding protein TDP-43 regardless of genetic background. This implicates the importance of investigating the widespread TDP-43 dysfunction-mediated changes in RNA processing, and the need to identify differential gene expression and alternative splicing changes that may underlie neurodegenerative disease.

Methods: Two mouse models of TDP-43 were investigated, each containing a single substitution within the coding region of the *Tardbp* gene: one mutation in the RNA binding domain (RBD), the other in the low complexity domain hotspot region for ALS-causative mutations. RNA sequencing was used to examine resulting differential gene expression and alternative splicing events, and integrated iCLIP analysis highlighted changes in RNA binding patterns.

Results and discussion: We identified severe molecular dysregulation in both models. The RBD mutant displayed dose-dependent preferential exon inclusion, including the retention of cryptic exons. Alongside this was downregulation of long intron-containing genes, typically related to neuronal activity. These features, alongside other RNA processing changes, were similar to those resulting from TDP-43 loss, and thus indicated the RBD mutation to cause TDP-43 loss of function. Contrastingly, we found that the LCD mutation caused greater levels of exon skipping, including novel gain of TDP-43 function in splicing which resulted in mutant-specific 'skiptic' transcripts. A subset of these was then demonstrated to be conserved in a TDP-ALS mouse model, and further in human TDP-ALS patient settings. Importantly, this gain of function mouse model developed a neurodegenerative, ALS-like phenotype. This indicated that TDP-43 gain of function, even in the absence of loss of function effects, is sufficient to cause a neurodegenerative disease phenotype.

IMPACT STATEMENT

Amyotrophic lateral sclerosis (ALS) is a severe, rapidly progressive and fatal disease that affects thousands of people worldwide. Research into the underlying biology has progressed quickly, however the development of resulting therapeutics remains inadequate. The protein TDP-43 has been heavily implicated in ALS through both causative mutation as well as dysregulated localisation regardless of disease aetiology. The aims of this study revolved around the characterisation and analysis of two disease models at the transcriptomic, cellular and phenotypic level in the context of ALS. The primary focus of this investigation was to understand how the mutations of crucial RNA binding protein TDP-43 disrupted RNA processing using high-throughput RNA sequencing analysis. Following on from this was the aim of further identifying specific molecular signatures which could be linked towards understanding the precise mechanism by which TDP-43 mutation and dysfunction could culminate in neurodegeneration and ALS. Alongside study of the transcriptome, other features of cellular dysfunction were also investigated, particularly those previously described in ALS settings. The results of this study have now been published in (Fratta et al. 2018 and Sivakumar et al. 2018), answering key questions in the field on the relative importance of loss of function and gain of function of TDP-43 in disease pathology. Several patterns and specific targets of differential gene expression and alternative splicing have been identified, and further validated to occur in ALS patient tissues; therefore this work also provides a foundation for further investigation and development of drug and biomarker targets in ALS.

CONTENTS

FIGURES

TABLES

ABBREVIATIONS

ALS – Amyotrophic lateral sclerosis

CLIP – Crosslinking and immunoprecipitation

ENU – *N*-ethyl-*N*-nitrosurea

FTD – Frontotemporal dementia

GO – Gene ontology

HnRNP – Heterogeneous ribonucleoprotein

KD/KO – Knock-down/Knock-out

LCD – Low complexity domain

MEF – Mouse embryonic fibroblasts

NLS – Nuclear localisation signal

PSI – Percentage spliced in

RBD – RNA binding domain

RBP – RNA binding protein

RNA-Seq – RNA sequencing

RRM – RNA recognition motif

TDP-ALS – TDP-43 mutation-mediated ALS

TDPBR – TDP-43 binding region

UTR – Untranslated region

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF TDP-43

1.1.1 PROTEIN STRUCTURE

TDP-43 is a ubiquitously expressed RNA binding protein (RBP) of the heterogeneous nuclear ribonucleoprotein (hnRNP) family. It consists of 414 amino acid residues (Figure 1.1), and contains an N-terminal nuclear localisation signal (NLS), an RNA binding domain (RBD) consisting of two RNA recognition motifs (RRMs), and a glycine-rich, prion-like, low complexity domain (LCD). Within RRM2 lies a nuclear export sequence (NES) (Figure 1.2). This domain structure is conserved between humans and mice.

```
Human     1    MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRLVEGI  60
               MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRLVEGI
Mouse     1    MSEYIRVTEDENDEPIEIPSEDDGTVLLSTVTAQFPGACGLRYRNPVSQCMRGVRLVEGI  60

Human     61   LHAPDAGWGNLVYVVNYPKDNKRKMDETDASSAVKVKRAVQKTSDLIVLGLPWKTTEQDL  120
               LHAPDAGWGNLVYVVNYPKDNKRKMDETDASSAVKVKRAVQKTSDLIVLGLPWKTTEQDL
Mouse     61   LHAPDAGWGNLVYVVNYPKDNKRKMDETDASSAVKVKRAVQKTSDLIVLGLPWKTTEQDL  120

Human     121  KEYFSTFGEVLMVQVKKDLKTGHSKGFGFVRFTEYETQVKVMSQRHMIDGRWCDCKLPNS  180
               K+YFSTFGEVLMVQVKKDLKTGHSKGFGFVRFTEYETQVKVMSQRHMIDGRWCDCKLPNS
Mouse     121  KDYFSTFGEVLMVQVKKDLKTGHSKGFGFVRFTEYETQVKVMSQRHMIDGRWCDCKLPNS  180

Human     181  KQSQDEPLRSRKVFVGRCTEDMTEDELREFFSQYGDVMDVFIPKPFRAFAFVTFADDQIA  240
               KQS DEPLRSRKVFVGRCTEDMT +EL++FF QYG+V+DVFIPKPFRAFAFVTFADD++A
Mouse     181  KQSPDEPLRSRKVFVGRCTEDMTAEELQQFFCQYGEVVDVFIPKPFRAFAFVTFADDKVA  240

Human     241  QSLCGEDLIIKGISVHISNAEPKHNSNRQLERSGRFGGNPGGFGNQGGFGNSRGGGAGLG  300
               QSLCGEDLIIKGISVHISNAEPKHNSNRQLERSGRFGGNPGGFGNQGGFGNSRGGGAGLG
Mouse     241  QSLCGEDLIIKGISVHISNAEPKHNSNRQLERSGRFGGNPGGFGNQGGFGNSRGGGAGLG  300

Human     301  NNQGSNMGGGMNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQNQSGPSGNNQNQGNMQ  360
               NNQG NMGGGMNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQNQSGPSGNNQ+QG+MQ
Mouse     301  NNQGGNMGGGMNFGAFSINPAMMAAAQAALQSSWGMMGMLASQQNQSGPSGNNQSQGSMQ  360

Human     361  REPNQAFGSGNNSYSGSNSGAAIGWGSASNAGSGSGFNGGFGSSMDSKSSGWGM  414
               REPNQAFGSGNNSYSGSNSGA +GWGSASNAGSGSGFNGGFGSSMDSKSSGWGM
Mouse     361  REPNQAFGSGNNSYSGSNSGAPLGWGSASNAGSGSGFNGGFGSSMDSKSSGWGM  414
```

Figure 1.1. Amino acid sequences in human and mouse TDP-43. Central sequence highlights conserved residues. "+" indicates non-conserved residues with similar chemical properties.

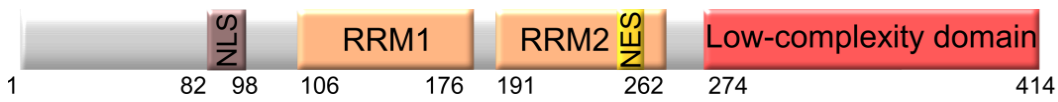


Figure 1.2. TDP-43 domains with numbered amino acid residues, accurate for human and mouse TDP-43 isoforms.

The N-terminal of TDP-43 is involved in protein interaction, and enables the formation of TDP-43 homodimers (Xiao et al. 2011), and further the disulphide bridges that lead

to tetramerisation. This regulation of TDP-43 formation is likely to play roles in TDP-43's RNA processing functions, as well as preventing insoluble aggregation of the protein in the cytoplasm (Jiang et al. 2017).

The structure of TDP-43 closely resembles other hnRNP proteins, particularly those linked to amyotrophic lateral sclerosis (ALS), such as FUS, hnRNP A1 and hnRNP A2B1 (Figure 1.3) (Purice & Taylor 2018). However, while these proteins contain a proline-tyrosine NLS (PY-NLS), which contains a proline-tyrosine pairing allowing binding to the transportin protein for nuclear shuttling, TDP-43 contains a classical bipartite NLS of K82RK84 and K95VKR98 (Winton et al. 2008). Mutation of these regions has been demonstrated as capable of leading to TDP-43 mislocalisation.

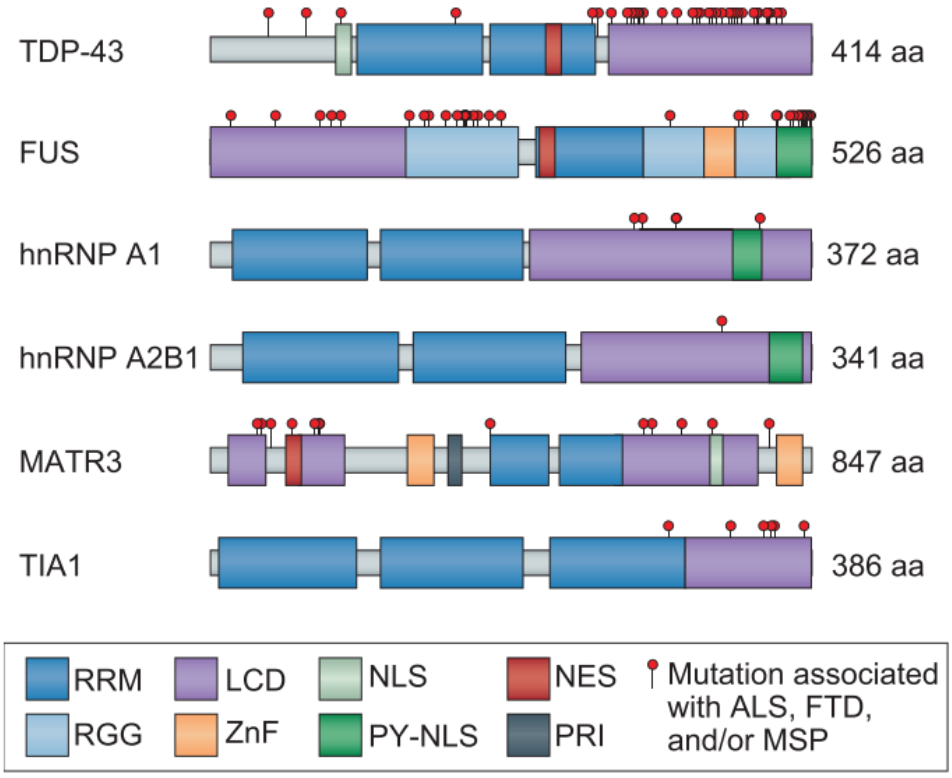


Figure 1.3. TDP-43 domains compared to related hnRNPs and RBPs. RGG – Arg-Gly-Gly box, ZnF – Zinc finger, PRI – Polypyrimidine tract binding RRM interaction domain, MSP – Multisystem proteinopathy (Purice & Taylor 2018).

TDP-43 further contains two RRMs, both involved in controlling TDP-43 DNA and RNA binding function (Kuo et al. 2009). TDP-43 was first discovered as a possible transcriptional repressor binding to the transactivation response element of HIV-1 DNA, hence it's full name: transactive response DNA binding protein 43 kDa (Ou et al. 1995). However, due to its interaction with a large number of transcript targets (Polymenidou et al. 2011), a strong focus on its role in RNA processing control has developed. The role of both RRMs in controlling aspects of RNA regulation is not equally distributed; RRM1 is known to be both necessary and sufficient for nucleic acid binding in TDP-43 (Buratti & Baralle 2001). For RNA binding in particular, mutation of just two amino acid residues – F147L and F149L – is enough to completely abolish TDP-43 binding, and remove control of splicing in target transcripts (Buratti & Baralle 2001; D'Ambrogio et al. 2009). RRM2 has an unusual structure in comparison to typical RRM sequences –it contains an additional β strand which would imply roles in protein interaction (Kuo et al. 2009). This structure is thought to lead to self-association, and results in a high level of stability of the RRM2 to act as the 'stability core' of TDP-43 (Morgan et al. 2017). However, its role in RNA binding is reduced relative to RRM1, with its affinity for RNA two orders of magnitude lower (Kuo et al. 2009). Nonetheless, it partially controls overall binding ability of TDP-43 to RNA, likely performing a more regulatory role (Furukawa et al. 2016).

The LCD of TDP-43 contains a central hydrophobic element surrounded on either side by two prion-like domains. This means it is intrinsically disordered, but possesses the ability to form β-rich dynamic oligomers at high concentrations, resulting in dynamic self-assembly and formation of intracellular, membrane-less organelles such as stress granules, a behaviour known as liquid-liquid phase separation (LLPS) (Kato et al. 2012; Nott et al. 2015; Mitrea & Kriwacki 2016; Lim et al. 2016). The LCD is primarily involved in protein interaction, and contains the residues 321-366 which are required for recruitment of a variety of other hnRNPs (D'Ambrogio et al. 2009). Tied to this, the

LCD of TDP-43 is necessary for its control of RNA splicing, and this function is conserved across several species (Ayala et al. 2005). Each of the domains of TDP-43 are also involved in overall inter-domain interaction within the protein, modulating its activity by causing conformational changes which result in the protein forming open and closed states. This altering of overall protein folding is primarily controlled by the LCD (Wei et al. 2016).

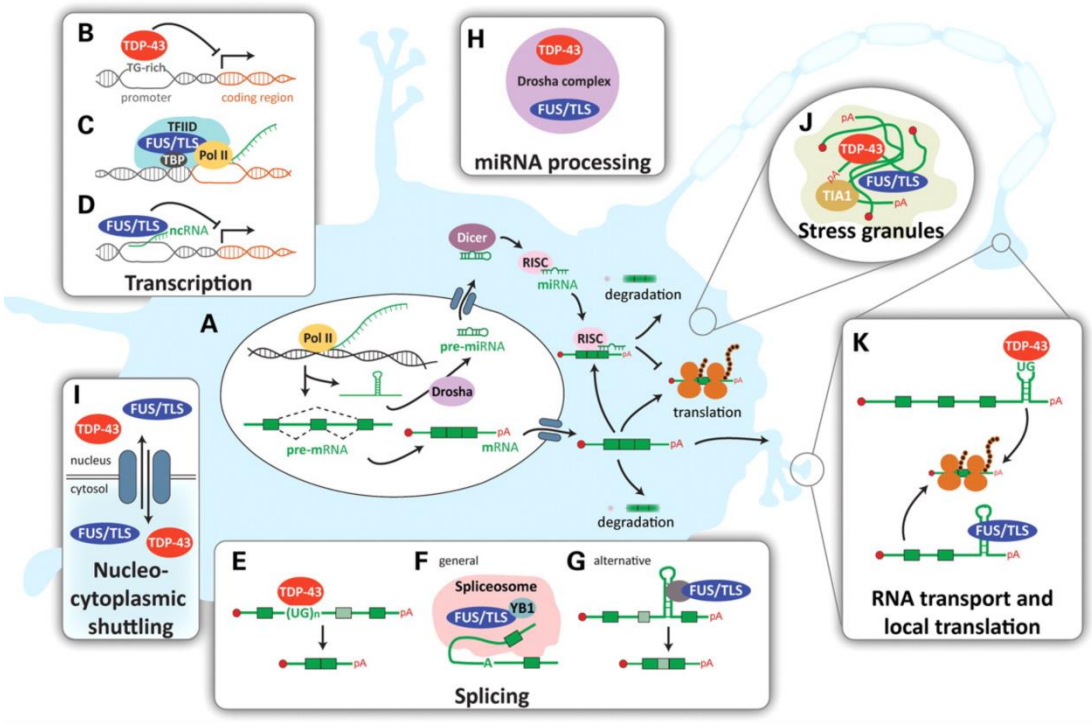


Figure 1.4. Schematic overview of TDP-43 cellular functions in RNA processing and regulation (Lagier-Tourenne et al. 2010).

Outside of regulating its own transcript levels, TDP-43 functions to control several aspects of RNA processing, such as alternative splicing (Buratti et al. 2001; Buratti & Baralle 2001), turnover (Ayala et al. 2008; Fiesel et al. 2010; Godena et al. 2011), stability (Strong et al. 2007; Colombrita et al. 2009), transport (Dewey et al. 2011; Fallini et al. 2012) and translation (Wang et al. 2008) in target transcripts (Figure 1.4). These functions are not limited to mRNA, with roles in ncRNA regulation also reported (Kawahara & Mieda-Sato 2012). Several RBPs exist, but TDP-43 is thought to be particularly vital given that it is reported to bind to 30 % of the mouse transcriptome (Polymenidou et al. 2011), and its RBD is notably conserved across species ranging from *C. elegans* all the way to humans (Ayala et al. 2005). In particular, TDP-43 is known to have high affinity for UG repeat motifs in RNA targets (Buratti & Baralle 2001), whether in the form of long stretches of UG repeats or as short clusters of UG

interspersed with other sequences. In determining this affinity, RRM1 plays a primary role, with a low dissociation constant of 14.2 nM for RNA sequences containing at least six UG repeats (Kuo et al. 2009). RRM2 affinity for UG repeats is markedly lower, but the affinity for this motif is highest with the RRMs in conjunction as the total RBD of TDP-43 (Kuo et al. 2009).

The role of TDP-43 in alternative splicing of target transcripts has been especially well documented. It was initially attributed to repression of exon inclusion, such as of exon 9 of the *CFTR* gene (Buratti et al. 2001; Mercado et al. 2005), but contrary events describing TDP-43-mediated exon skipping have also now been established (Fiesel et al. 2012). Currently, several strongly TDP-43-mediated splicing events have been identified, including in *POLDIP3*, *SORT1*, *EIF4H* and *MADD* (Fiesel et al. 2012; Prudencio et al. 2012; Yang et al. 2014; De Conti et al. 2015), and these provide a subset of primary targets for investigating presence of TDP-43 dysfunction. TDP-43 functions by interacting with nascent pre-RNA transcripts and co-ordinates the processing of these targets via interaction with other RBPs, hnRNPs and the spliceosome. Importantly, this regulatory splicing control is also likely to be conserved across several species (Ayala et al. 2005). The changes in TDP-43-RNA processing in various diseases has been investigated in several models, further discussed in 1.3 Models of TDP-43 dysfunction,, and reviewed in Nussbacher, Batra, Lagier-Tourenne, & Yeo (2015).

The structure of TDP-43 enables LLPS, and the formation of transient membrane-less stress granules designed to protect RNA transcripts from degradation as part of the cellular stress response (Liu-Yesucevitz et al. 2010; Kato et al. 2012). It has been hypothesised that these structures may be dysfunctional in disease, with protein mutations causing unwanted stability of the organelles to form toxic aggregates (Ratti & Buratti 2016).

1.1.3 *TARDBP* ISOFORMS AND AUTOREGULATION

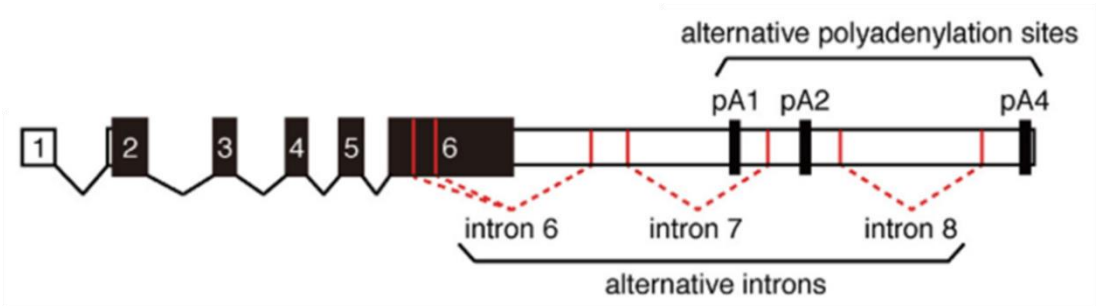


Figure 1.5. Human *TARDBP* transcript. UTR sequences represented by white boxes, exons represented by black boxes. Key introns and alternative polyadenylation sites annotated. Adapted from Koyama et al. (2016).

TDP-43 is coded for by the *TARDBP* gene, found on chromosome 1 in humans and chromosome 4 in mice. In both species, several isoforms of the transcript exist to enable regulation of TDP-43 protein levels. TDP-43 is tightly autoregulated in a mechanism involving the alternative splicing and polyadenylation of *TARDBP* transcripts in response to changes in TDP-43 binding. Figure 1.5 illustrates the most common *TARDBP* transcript isoforms and the alternative introns which can be spliced to result in usage of different polyadenylation sites (pAs).

TDP-43 is known to directly bind to its own transcripts to influence processing of the RNA, and the integral region of binding for this control lies within intron 7, 5' of polyadenylation site pA_1 (Ayala et al. 2011). This 34 nt region, termed the TDP-43 Binding Region (TDPBR), contains UG repeats but, unlike the archetypal TDP-43 target sequences, this sequence consists of short interspersed UG motifs as opposed to long stretches of UG. Despite the resulting lower affinity for TDP-43 to the TDBPR in comparison to long UG sequences, TDP-43 binding to this region was confirmed by the presence of high cluster density in TDP-43 CLIP studies (Ayala et al. 2011).

TDP-43 binding to the TDPBR has been shown to be crucial to autoregulation through transfection studies of the *TARDBP* 3'UTR isoforms into HeLa cell lines by Polymenidou et al. (2011). Either the long, unspliced *TARDBP* 3'UTR form containing the TDPBR, or the short spliced 3'UTR form lacking the TDPBR, was cloned adjacent to a *Renilla* luciferase gene and transfected into HeLa cells. Upon TDP-43 overexpression, cells transfected with the unspliced 3'UTR isoform (containing the TDPBR) showed significantly less Relative Fluorescence Units (RFUs) produced relative to the spliced isoform with no TDPBR, suggesting that TDP-43 was able to restrict expression of the unspliced 3'UTR isoform but not the spliced isoform lacking the TDPBR.

There has been a great deal of debate surrounding the precise splicing and polyadenylation processes undergone by *TARDBP* transcripts and their respective consequences on *TARDBP* transcript fate. Specifically how the splicing of introns 6 and 7, and the binding of TDP-43 throughout the TDPBR, influence the choice of polyadenylation sites pA_1, pA_2 and pA_4 is not fully understood.

In early studies, computational analysis predicted the presence of four possible polyadenylation signals in the *TARDBP* gene. However, Expressed Sequence Tag (EST) evidence from the UCSC Genome Browser database suggested two sites, pA_1 and pA_4, accounted for the vast majority of transcript polyadenylation sites (Ayala et al. 2011). Current consensus is that *TARDBP* transcripts with pA_1 are stable, can be exported into the cytoplasm, and translated into protein. Transcripts with pA_4, a polyadenylation site 1.5 kb further 3', are instead retained in the nucleus and are unavailable for translation (Koyama et al. 2016). In HEK293 cells, the ratio of transcripts using pA_1:pA_4 was approximated to be 3:1 (Ayala et al. 2011). Alongside these polyadenylation sites is another, pA_2, and *TARDBP* transcripts with pA_2 were identified using Rapid Amplification of cDNA Ends (RACE) (Eréndira Avendaño-Vázquez et al. 2012). Like pA_4-containing transcripts, *TARDBP* transcripts using pA_2 are also not translated into

protein. However, whether they are exported from the nucleus into the cytoplasm and degraded (Koyama et al. 2016), or whether they are retained in the nucleus (Eréndira Avendaño-Vázquez et al. 2012), is still a topic of debate. Importantly, all three of the aforementioned polyadenylation sites are conserved between mice and humans, showing near identical arrangement (Eréndira Avendaño-Vázquez et al. 2012).

The polyadenylation site to be used is partly dependent on the splicing of the *TARDBP* transcript, particularly the retention or splicing of introns 6 and 7. Intron 7 is known to be critical to the autoregulation mechanism as it contains pA_1, the polyadenylation site typical required for *TARDBP* transcripts to undergo translation, as well as the TDPBR.

Typically, both intron 6 and intron 7 are retained, leading to use of pA_1 and a stable isoform to be exported and translated. However, some transcripts can retain both introns but as a result of TDP-43 binding around the pA_1 region, do not use pA_1 but instead pA_4 and are therefore not exported.

Upon TDP-43 overexpression, intron 7 splicing is enhanced. This removes the pA_1 site, leading instead to a switch to the next polyadenylation site, pA_2. As transcripts with pA_2 are not translated, this means that TDP-43 overexpression would reduce the level of *TARDBP* transcripts to be translated and behave as the required feedback mechanism. Concordantly, nullification of intron 7 splice sites can prevent the autoregulatory mechanism from functioning (Eréndira Avendaño-Vázquez et al. 2012).

Intron 6 is of interest as it contains ALS-causative mutations in TDP-43 (Koyama et al. 2016). It has also recently been identified as possibly important in TDP-43 autoregulation, as retention of intron 6 alongside splicing of intron 7 reduces the severity of autoregulation in comparison to splicing of both introns (Koyama et al. 2016). Studies disagree on the importance of this transcript, with some stating that it is very rarely seen under basal conditions (Koyama et al. 2016), whereas others state that

it is the primary transcript that displays pA_2 usage (Eréndira Avendaño-Vázquez et al. 2012).

More simply, there is unanimous agreement on the non-existence of *TARDBP* transcripts with spliced intron 6 and retained intron 7 (Eréndira Avendaño-Vázquez et al. 2012; Koyama et al. 2016). Indeed, Koyama et al. (2016) state that intron 7 splicing is necessary for, and precedes, intron 6 splicing.

The final combination, splicing of both intron 6 and intron 7, leads to a pA_2-utilising transcript (Koyama et al. 2016), which was also the transcript described to originally be involved in autoregulation by Polymenidou et al. (2011). As it contains the pA_2 polyadenylation site, this transcript is not translated. This isoform is also detected, albeit at low levels and termed "V2", in Ayala et al. (2011), however they determine that as it does not change in expression levels upon either TDP-43 overexpression or nonsense-mediated decay (NMD) inhibition, it is not involved in autoregulation.

Further disagreement in the field exists on whether the autoregulatory mechanism results in a reduction in *TARDBP* transcript translation through NMD (Polymenidou et al. 2011; Koyama et al. 2016), or another pathway (Ayala et al. 2011; Eréndira Avendaño-Vázquez et al. 2012).

From the aforementioned Polymenidou et al. (2011) investigation of *TARDBP* 3'UTR isoforms fused with the luciferase gene and transfected into HEK293 cells, the role of NMD was tested through concomitant treatment with siRNA targeting UPF1, a component required for NMD. They found that inhibition of NMD resulted in increased RFU produced by the unspliced 3'UTR isoform, suggesting that NMD played a role in the degradation of the mRNA. Further experiments revealed that blocking NMD in this manner caused increased expression of the *TARDBP* isoform with both introns 6 and 7 spliced out in HeLa cells; a similar effect to the transient overexpression of TDP-43 in the same model. Koyama et al. (2016) also report the involvement of NMD in TDP-43

autoregulation, with pA_2-containing transcripts with both introns spliced out showing an increase upon cyclohexamide treatment (CHX) and UPF1 phosphorylation inhibition in HEK293 cells.

Conversely, Ayala et al. (2011) found that upon CHX treatment of HEK293 cells, there was no change in the autoregulation mechanism. However, in this case they report no difference in *TARDBP* isoforms containing pA_1 or pA_4, both of which result from retention of introns 6 and 7. They do suggest that the isoform with both intron 6 and 7 spliced out ("V2") would undergo NMD, having ruled that it was not involved in autoregulation. Eréndira Avendaño-Vázquez et al. (2012) look into pA_2 isoforms and maintain that CHX treatment in HEK293 cells had no effect on pA_2 isoform abundance. Instead, these studies propose alternative methods by which translation is prevented, including a nuclear retention of the transcripts (since contradicted in Koyama et al. [2016]) or exosome-mediated degradation.

The precise mechanism by which TDP-43 binding to the TDPBR leads to intron splicing is also unclear. One likely possibility is that TDP-43 binding to this region is required in order to recruit other hnRNPs and splicing factors to enable the splicing of the intron, but the identity of these factors is not yet known. The importance of spliceosome assembly in intron 7 in particular has been suggested to be the key feature for subsequent polyadenylation site selection, over and above other factors such as the intrinsic quality of the polyadenylation sites themselves (Bembich et al. 2014). Another mechanism posits that TDP-43 binding causes stalling of RNA polymerase II (Pol II). It has previously been suggested that stalling of Pol II can lead to favouring of suboptimal splicing (Kornblihtt 2007). Furthermore, ChIP analysis has revealed that TDP-43 overexpression results in increased Pol II signals throughout the *TARDBP* 3'UTR, stretching from the TDPBR to pA_4 (Eréndira Avendaño-Vázquez et al. 2012). Therefore, overexpression of TDP-43 may lead to sufficient stalling of Pol II to enable recognition

of the suboptimal splice sites, resulting in intron removal. Additionally, CstF-64, a cleavage factor involved in 3' end cleavage and polyadenylation, targets a similar consensus motif to TDP-43, $GU_{3\text{-}5}U_{2\text{-}4}$ (Takagaki & Manley 1997). It is possible that TDP-43 overexpression results in it outcompeting CstF-64, blocking the latter recognising pA_1 and therefore resulting in the nascent transcript forming an isoform with a polyadenylation site that will not be translated. Evidence for this hypothesis was provided from RNA immunoprecipitation (RIP) experiments which showed that when TDP-43 and CstF-64 were in competition, TDP-43 overexpression caused increased pA_1 RNA binding and reduced CstF-64-pA_1 RNA binding as a result (Eréndira Avendaño-Vázquez et al. 2012).

The autoregulation of TDP-43 may be yet more complicated. One possibility was that the length of the 3'UTR itself may play a role in *TARDBP* transcript instability, but results from Koyama et al. (2016) suggest that this is not the case. Intriguingly, the entire ~3 kb *TARDBP* 3'UTR long isoform is well conserved between mice and humans, despite TDP-43 itself only binding to a relatively small region. This raises the possibility that TDP-43 is not alone in regulating *TARDBP* levels, and other proteins possess their own conserved regulatory binding sites in the *TARDBP* 3'UTR to control *TARDBP* transcript fate (Ayala et al. 2011). Post-translational control may also be involved, as TDP-43 protein half-life results have been described as highly variable depending on the cell line used (Ling et al. 2010). The presence of changes in TDP-43 autoregulation in disease settings is also unclear, and so identifying the precise intricate measures by which TDP-43 synthesis is modulated is crucial for understanding how otherwise tightly-controlled expression levels may be heavily disrupted in ALS.

1.2 AMYOTROPHIC LATERAL SCLEROSIS

1.2.1 DISEASE OVERVIEW

First described in 1869 by Jean-Martin Charcot, amyotrophic lateral sclerosis (ALS) is a rapidly progressive neurodegenerative disease characterised by the degeneration of both the upper motor neurons in the motor cortex and the lower motor neurons of the brainstem and spinal cord. The degeneration of the latter leads to inactivity and atrophy of no longer-stimulated voluntary muscles, resulting in muscle weakness and eventual paralysis. ALS is typically lethal within 2-3 years of onset, often through respiratory failure (Rowland & Shneider 2001). It is often a late-onset disease, and region of onset is variable, but early-onset and even juvenile forms of ALS have been described (Orban et al. 2007). It has an incidence rate of 2.16 in 100,000 in Europe (Logroscino et al. 2010), and is 20 % more common in men than women (Manjaly et al. 2010).

It is also strongly related to another neurodegenerative disease, frontotemporal dementia (FTD), and the two are now thought to coexist on an ALS/FTD spectrum (Mackenzie et al. 2010). FTD is a form of dementia that results in behavioural and personality changes, as well as impairment of brain functions such as attention and memory through neurodegeneration of the frontal and temporal brain lobes. 50 % of ALS patients are additionally diagnosed with additional cognitive impairment (Gordon et al. 2010). The precise mechanisms by which mutation in these genes leads to either of these diseases is unclear, with research suggesting the presence of overlap but also distinct pathways disrupted in each disease (Hardy & Rogaeva 2014). One broad distinction of the summary is that RNA processing genes such as TDP-43 are typically more associated with the ALS variant of the ALS/FTD spectrum.

ALS has no cure, and development of therapeutics has been poor, with only two drugs approved by the FDA in the US – Riluzole and Edaravone – following over two decades of translational research. Even then, the latter is not yet approved in the UK, and the former has only modest effects on extended lifespan (Miller et al. 2002).

1.2.2 GENETICS OF ALS

ALS is a primarily sporadic disease, with only 5-20 % of cases classified as familial (fALS) (Byrne et al. 2011; He et al. 2015). Phenotypes resulting from either cause of the disease are difficult to distinguish, and substantial overlap exists (Turner et al. 2017). In fALS, the majority of mutations are classed as autosomal dominant, and primary cause is mutation of *C9ORF72*, thought to be pathologically mutated in 40 % of cases (Figure 1.6). Mutations in *SOD1* were the first identified cause of ALS, and these underlie approximately 20 % of fALS. Causative variants in *TARDBP* and *FUS* further account for 4 % of fALS cases each. Genes with even rarer ALS-causative mutations include *OPTN*, *VCP*, *p62/SQSTM1*, *PFN1*, *UBQLN2* and *NEK1* (Renton et al. 2013). Roughly 70 % of fALS can be attributed to these known ALS genes, and 10 % of sporadic ALS cases can also be traced back to these. Clinical symptoms resulting from mutation in each of these genes are also variable, but broad patterns have emerged linking mutation of each of the most common ALS genes to the ALS/FTD spectrum of disease phenotype (Ling et al. 2013).

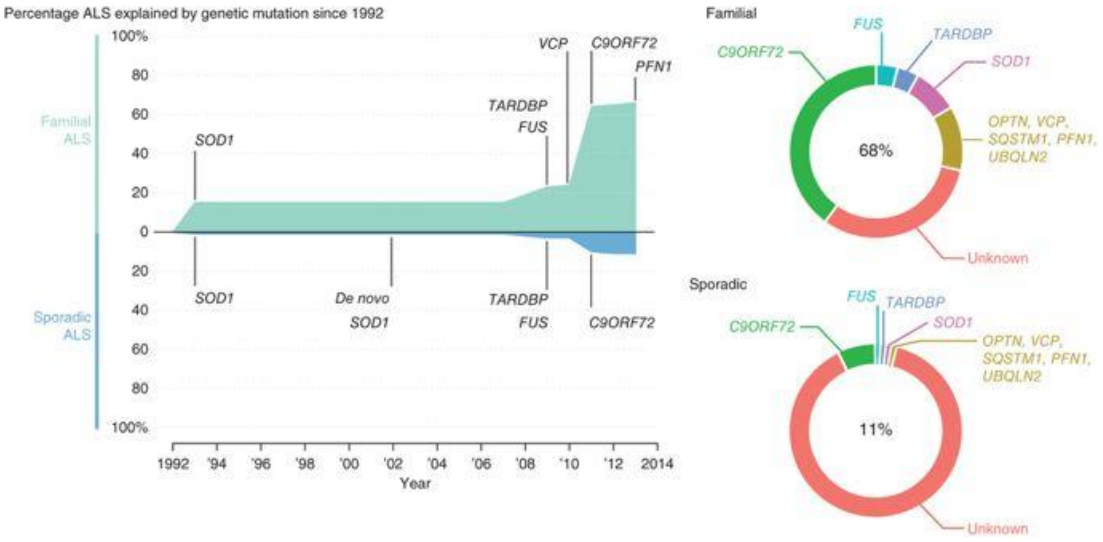


Figure 1.6. Left: Progression of ALS-causative gene identification with cumulative tally of percentage of explained fALS and sALS cases. Right: ALS-causative genes in fALS and sALS represented by proportion of respective cases explained by gene mutation. Central number represents total number of cases explained by known ALS-causative gene mutations (Renton et al. 2013).

Intriguingly, several environmental risk factors have also been posited, including smoking, alcohol consumption, military service, lead exposure and traumatic brain injury (reviewed in Martin et al. 2017). However, these effectors in general have been poorly studied, and even where correlation to disease onset exists it is typically weak and highly variable. Therefore, currently these risk factors do not substantially add to the existing knowledge on causes of ALS.

1.2.3 TDP-43 IN ALS

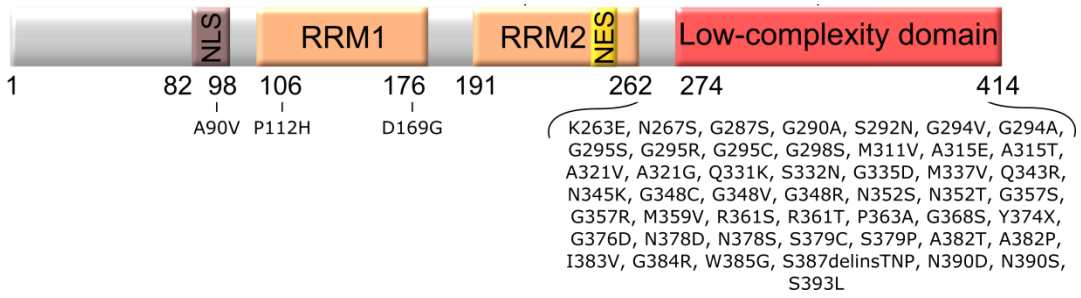


Figure 1.7. TDP-43 domains with all TDP-ALS mutations annotated, adapted from Emanuele Buratti (2015).

Mutations in *TARDBP* are known to cause ALS (Sreedharan et al. 2008), and as of 2015, 52 different mutations have been identified as ALS-causative in *TARDBP* (TDP-ALS) (Buratti 2015). Intriguingly, all but three of these mutations are found in the LCD of TDP-43 (Figure 1.7), implicating dysfunction of protein interaction as key to TDP-ALS (Pesiridis et al. 2009). Importantly, however, although *TARDBP* mutation is responsible for a minority of ALS cases, almost all ALS patients exhibit dysregulation of the TDP-43 protein regardless of genetic cause. This dysregulation is in the form of nuclear depletion of TDP-43, and its subsequent mislocalisation and aggregation juxtaposed to the nucleus, in the cytoplasm (Figure 1.8) (Arai et al. 2006; Neumann et al. 2006). This TDP-43 pathology is a hallmark of ALS, occurring in approximately 97 % of all cases; absent in only SOD1- and FUS-caused ALS (Ling et al. 2013). 50 % of FTD cases also show similar TDP-43 mislocalisation, and interestingly, other diseases can also exhibit this type of pathology, such as Alzheimer's disease (Amador-Ortiz et al. 2007; Higashi et al. 2007; Uryu et al. 2008; Josephs et al. 2014). Together, these diseases can be described as TDP-43 proteinopathies (Geser et al. 2010). Alongside pathology, elevated *TARDBP* transcript and TDP-43 protein levels have been detected in ALS patients, further suggesting TDP-43 dysfunction (Mishra et al. 2007; Gitcho et al. 2009; Kasai et

al. 2009; Swarup, Phaneuf, Dupré, et al. 2011; Verstraete et al. 2012; Iguchi et al. 2016; Koyama et al. 2016).

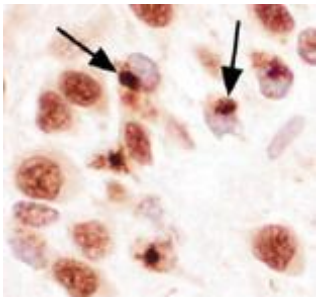


Figure 1.8. Example TDP-43-DAB staining of TDP-43 proteinopathy. TDP-43 is primarily nuclear, however some cells show nuclear depletion and cytoplasmic aggregation of TDP-43, indicated by arrows. Adapted from Vatsavayi et al. (2014).

The TDP-43 aggregates formed in ALS have been heavily studied, given the strong link between pathology and disease. The nuclear depletion and cytoplasmic mislocalisation precede the formation of eventual aggregates (Giordana et al. 2010). Rather than the full length form, these aggregates often consist of C-terminal fragments of TDP-43. A 35 kDa fragment exists, containing all but the N-terminal region, whereas the 25 kDa fragment lacks any of the RRM1 and consists of a truncated RRM2 as well as the LCD. Both of these forms have been linked to enhanced aggregation propensity and toxicity of the protein (Johnson et al. 2008), and in particular the 25 kDa fragment is thought to recapitulate the pathological features of TDP-43 proteinopathy (Zhang et al. 2009; Wei et al. 2017). Whether these aggregates display amyloid negative or positive staining properties is still debated, however (Chen et al. 2010; Guo et al. 2011; Wang et al. 2013).

It has been hypothesised that the reason the 25 kDa fragment is enriched in TDP-43 aggregates is that the cleavage of the RRM2 at R208 drastically reduces its solubility, and further eliminates its folding structure, resulting in increased risk of aggregation

(Zhang et al. 2009; Wei et al. 2016). Concomitant with this, the loss of the N-terminal in both TDP-43 fragment forms also prevents dimerisation, leaving unassembled RRM2 β strands exposed and prone to oligomerisation into high-order inclusions (Wang et al. 2013). The initial seed inclusions are also capable of sequestering wildtype full-length TDP-43, propagating the aggregation (Furukawa et al. 2011). Degradation of these cytoplasmic aggregates is also markedly slow, one-fifth of the speed of nuclear TDP-43 clearance, enabling their persistence (Sugai et al. 2018).

Importantly, these TDP-43 aggregates in ALS are also hyperphosphorylated (Arai et al. 2006). This phosphorylation is linked to enhance formation of the aggregates, as well as the resulting toxicity, in vivo (Igaz et al. 2009; Zhang et al. 2009), but this notion has been contradicted by in vitro study (Wei et al. 2016). Specifically, the phosphorylation of serine residues S409 and S410 have been pinpointed as signs of pathological TDP-43 aggregation (Neumann et al. 2009). The phosphorylation itself may not be the cause of aggregation however; rather it could demarcate attempts by the cell to degrade the inclusions (Mompeán et al. 2016). Acetylation of TDP-43 aggregates has also been reported in ALS patient spinal cords but not in FTD-TDP brain, suggesting a degree of variability in this feature (Cohen et al. 2015). Oxidative stress may also play a role in the formation of TDP-43 inclusions (Rabdano et al. 2017), displayed by cysteine oxidation, and markers of oxidative stress can be found elevated in ALS patient spinal cords (Pedersen et al. 1998), but a concrete link has not yet been established. TDP-43 aggregates are also often ubiquitinated, and show p62 co-localisation in ALS patient CNS tissue (Arai et al. 2006; Tanji et al. 2012). These are factors involved in protein degradation and so they further demonstrate cellular mechanisms attempting to clear the TDP-43 proteinaceous aggregates.

1.3 MODELS OF TDP-43 DYSFUNCTION

1.3.1 LOSS OF FUNCTION MODELS

To fully understand the role of TDP-43 in cellular homeostasis, and the subsequent effects of mutation or dysfunction that could underlie ALS phenotype, several cellular and animal models have been created. Due to the joint presence of nuclear TDP-43 depletion alongside cytoplasmic increase and aggregation, it is not clear whether the pathological effects of TDP-43 result from a loss of its normal function or a gain of function, either its normal function to excessive levels or a toxic gain of function with novel effects in the cytoplasm.

To understand the effects of TDP-43 loss of function, primary method of study has involved knock-out (KO) or knock-down (KD) of *Tardbp*, reducing the levels of TDP-43 and investigation the consequential effects. KO of TDP-43 homologue *TBPH* in *Drosophila melanogaster* resulted in severe motor phenotype, with the neuromuscular junction in particular regarded to be malformed (Feiguin et al. 2009). These flies also exhibited semi-lethality, and a reduced lifespan, and further these effects could be reversed using ectopic expression of a human wildtype TDP-43 construct. In zebrafish, two forms of *Tardbp* exist, *Tardbp* and *Tardbpl*. KD of *Tardbp* alone in zebrafish was sufficient to cause motor phenotype in one investigation (Kabashi et al. 2009), and while KD of *Tardbpl* did not result in a phenotype, KD of both culminated in a more severe motor phenotype than KD of *Tardbp* alone (Hewamadduma et al. 2013). However, another study determined that *Tardbp* loss alone did not result in motor phenotype (Schmid et al. 2013).

Three mouse models were also simultaneously developed with TDP-43 KO, and in all cases homozygosity resulted in early embryonic lethality with substantial growth deficiency (Wu et al. 2009; Kraemer et al. 2010; Sephton et al. 2010). Inducible KO

developmental models showed similar rapidly lethal results, highlighting the importance of TDP-43 function throughout development (Chiang et al. 2010). Heterozygous KO of TDP-43 instead resulted in viable mice, however, likely due to the autoregulation mechanism of TDP-43 compensating for the loss of one allele as outlined in 1.1.2 *TARDBP* isoforms and autoregulation. Importantly, these adult heterozygous KO models did eventually develop variable motor phenotype, but no neurodegeneration (Kraemer et al. 2010), suggesting TDP-43 was important even in mature models, and also implied a reduction in the capabilities of the TDP-43 autoregulatory mechanism in adult mice. This was further supported by TDP-43 KO in adult mouse motor neurons resulting in axonal and muscular atrophy alongside progressive motor neuron dysfunction (Iguchi et al. 2013). TDP-43 KD in the adult mouse striatum was further revealed to cause mass RNA dysregulation (Polymenidou et al. 2011), further discussed in Chapter 3 F210I model transcriptome dysregulation.

Finally, cell death has also been shown to result in human HeLa cells as a result of TDP-43 loss, as well as in neuron-like Neuro-2a cells, in which neurite growth was diminished (Iguchi et al. 2009).

1.3.2 OVEREXPRESSION MODELS

Given the elevated *TARDBP* and TDP-43 levels detected in ALS patients, alongside the increased proportion of protein localised in the cytoplasm, the effects of TDP-43 overexpression have also been similarly studied in several models. In *C. elegans*, overexpression of TDP-43 orthologue *Tdp-1* was not found to be toxic (Ash et al. 2010). However, highlighting the differences between *C. elegans* and human systems, complete *Tdp-1* loss was also not described as toxic in *C. elegans*, and bizarrely even associated with a slight increase in lifespan (Vaccaro et al. 2012). Overexpression of *TBPH* in *Drosophila melanogaster* lead to axonopathy (Lu et al. 2009), whereas

overexpression of wildtype human TDP-43 in *Drosophila melanogaster* resulted in motor neuron degeneration, with neuromuscular junction formation again pinpointed (Li et al. 2010; Hanson et al. 2010). However, despite the overexpression, mislocalisation of TDP-43 was not triggered. Interestingly, overexpression of the C-terminal TDP-43 fragments alone were found to cause reduced phenotype in comparison to the full length TDP-43 (Li et al. 2010; Voigt et al. 2010).

In mice, transgenic overexpression of either mouse or human wildtype TDP-43 is toxic, and results in motor phenotype in the form of gait abnormalities. Mouse lifespan is also typically reduced (Shan et al. 2010; Wils et al. 2010; Xu et al. 2010). Wils et al. (2010) also reported neuronal inclusion formation and subsequent muscle paralysis in their model, while Xu et al. (2010) described mitochondrial aggregation alongside motor deficits. FTD-like symptoms of progressive learning and memory defects have also been reported, in conjunction with degeneration of the frontal cortex and hippocampus (Tsai et al. 2010; Igaz et al. 2011). However, lower motor neuron degeneration is often not seen in these models (Joyce et al. 2011). TDP-43 overexpression phenotype and lifespan reduction has also been reported as dose-dependent (McGoldrick et al. 2013). It is thought that overexpression of TDP-43 leads to incomplete oligomeric complexes (Xu 2012), but whether this replicates the outcome of TDP-ALS mutation is not clear.

1.3.3 TDP-ALS MODELS

Although the TDP-43 KD, KO and transgenic wildtype overexpression models all provide insight into the varying roles of TDP-43 mislocalisation in subsequent molecular, cellular and model phenotype, it is not clear precisely how these changes relate to those caused by TDP-ALS mutations. To examine this further, several mouse models have now been developed expressing transgenic TDP-ALS mutated human *TARDBP* constructs under the control of various promoters. These include models with

the A315T mutation from Wegorzewska, Bell, Cairns, Miller, & Baloh, (2009), Stallings, Puttaparthi, Luther, Burns, & Elliott, (2010), Swarup, Phaneuf, Bareil, et al. (2011) and Stribl et al. (2014), the M337V from Stallings et al. (2010), Y.-F. Xu et al. (2011), Arnold et al. (2013), Janssens et al. (2013), and the Q331K from Arnold et al. (2013), and Mitchell et al. (2015). Phenotypes across these models vary, and between them several features of the typical ALS phenotype are covered, including TDP-43 mislocalisation and aggregation, gliosis, mitochondrial defects, behavioural changes, motor neuron degeneration and loss, and neuromuscular junction loss.

However, a key drawback of all of these models is the use of transgenic *TARDBP* construct expression for study of TDP-ALS mutation effects. This results in TDP-43 overexpression, and as outlined earlier, overexpression of even wildtype TDP-43 can cause severe motor phenotype. Therefore, it is not possible to distinguish the phenotype that results from mutation of the protein rather than overexpression. Furthermore, in several of these models, the effect of mutation on RNA processing is also not studied for the same reason. TDP-43 overexpression, and subsequent autoregulation, would likely have an impact on its control of target transcript processing, again preventing investigation of how the TDP-ALS mutation specifically affects TDP-43 functions. Overcoming the shortcomings of these ALS mouse models is vital for downstream therapeutic development, as RNA dysregulation may be a key area of impairment to be targeted (Ferraiuolo et al. 2011). In one study, noted to have *Tardbp* expression levels comparable to wildtype animals despite transgenic expression, RNA splicing dysfunction was studied (Arnold et al. 2013). This investigation reported some splicing changes resembling TDP-43 loss of function, and others following the pattern expected of TDP-43 gain of function.

The work documented in this thesis is based on two mouse models with single point mutations inside the endogenous *Tardbp* gene through *N*-ethyl-*N*-nitrosurea (ENU)

mutagenesis (Acevedo-Arozena et al. 2008; Georgel et al. 2008), developed in order to study the effect of TDP-43 mutations without the drawback of transgenic TDP-43 overexpression. These mutations were selected upon testing of several mutated TDP-43 constructs using the CFTR minigene assay (detailed in Chapter 3 – F210I model transcriptome dysregulation) which revealed notable changes in TDP-43 splicing activity. In this method, the mutagen was administered to adult male mice via a series of intraperitoneal injections and allowed to spread throughout the body, including to the testes, where it caused mutations in sperm cells as they were formed from the spermatagonial stem cells. Approximately 6 weeks after ENU injection, male mice were mated to wildtype females. In the sperm of mice from the MRC Harwell programme, the result of mutagen exposure was a point mutation per million base pairs of genomic DNA. The ENU treated male mice were bred to produce F1 progeny, culled and two storage banks, of DNA and sperm from the same animal, were created, ready for multiplex analysis. The DNA bank was screened in a high-throughput manner for mutations in genes of interest, and the gametes of models containing desired mutations were revived and used, via in vitro fertilisation of female mice of the desired background, to create a mouse model containing the mutation of interest. These models were further crossed in order to eliminate other undesired mutations as a result of the ENU mutagen, eventually resulting in a mouse model with a wildtype background aside from the mutation of interest.

Using this method, two mouse models of mutant TDP-43 have been developed, each containing a single A>T missense mutation in the *Tardbp* gene resulting in a single amino acid change in either RRM2 (F210I) or LCD (M323K) of TDP-43 (Figure 1.9). Both models have been extensively studied for molecular, cellular and developmental phenotype to further elucidate the early molecular mechanisms underlying pathology as a consequence of TDP-43 dysfunction and the findings have been recently been published (Fratta et al. 2018; Sivakumar et al. 2018).

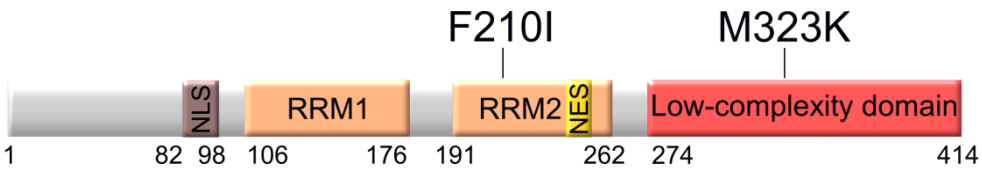


Figure 1.9. Mouse TDP-43 domains with the two ENU mutagenesis-derived mutations primarily investigated in this thesis annotated, F210I and M323K.

1.3.4 TDP-43 LOSS VERSUS GAIN OF FUNCTION

Distinguishing the loss and gain of function mechanisms that result from TDP-43 mutation, dysfunction, and mislocalisation is made difficult due to the processes occurring simultaneously in models with TDP-43 proteinopathy. Therefore, it has so far been difficult to attribute the molecular dysregulation that results from TDP-43 change in activity to one specific mechanism, although previous investigations provide evidence for the importance of both loss and gain of function.

One hypothesis is that loss of function of basal TDP-43 activity is primarily responsible for the RNA processing dysregulation upon TDP-ALS mutation, and possibly the phenotype that results. This hypothesis posits that the key event is the nuclear depletion of TDP-43. As it is lost from the nucleus, its roles in the transcription, splicing, and transport of pre-mRNA and mRNA transcripts are compromised. This leads to the failure to process transcripts for a large number of different genes, some of which may be of particular importance to motor neuron function. The loss of the subsequent mRNA translation, and therefore protein, leads to disturbance in cellular homeostasis that proves toxic to the cell. Alongside this, the theory infers that nuclear TDP-43 depletion is also the key trigger for TDP-43 protein aggregation. As nuclear TDP-43 is lost, the synthesis of TDP-43 is increased via the autoregulatory mechanism of *TARDBP* transcripts, increasing *TARDBP* transcript and TDP-43 protein levels. However, TDP-43 continues to be mislocalised towards the cytoplasm, with the

excessive concentration forming aggregates, while the continued nuclear depletion maintains the upregulation of TDP-43 production, causing a continuous cycle of increased aggregation.

Several lines of evidence exist for the role of loss of function of TDP-43 in cellular toxicity. Study of TDP-43 KO and KD models, as previously outlined, are often lethal, and even when not can result in several variations of developmental and motor phenotype, showing that standard TDP-43 function is vital at the cellular level. TDP-43 is also known to be an RBP that interacts with an exceptionally high number of transcripts (Polymenidou et al. 2011), therefore its loss is likely to affect many different pathways and a number of key genes. Previous study of TDP-43 loss of function has also revealed widespread RNA misregulation, including mass changes in RNA expression and splicing patterns (Polymenidou et al. 2011; Ling et al. 2015). In regards to ALS, the detection of nuclear depletion is widespread and evident in more than 97 % of all ALS patients (Ling et al. 2013), providing a strong link between the loss of function of TDP-43 that would result and the possibly consequent phenotype. Although the patients show cytoplasmic mislocalisation and TDP-43 aggregation, tying this to gain of function of the protein is more difficult to infer.

Gain of function effects of TDP-43 have also been previously investigated, both in terms of excessive levels of normal function and possible toxic gain of new function. These hypotheses also suggest that cytoplasmic TDP-43 is the key driver of TDP-43 pathology; perhaps the mutant toxic TDP-43 protein seeds aggregate formation, sequestering wildtype TDP-43 protein too. This toxic effect may reduce the synthesis of stable *TARDBP* transcripts and TDP-43 protein further and further until it is completely depleted in the nucleus and remains only in the form of aggregates in the cytoplasm. Like TDP-43 KD and KO models, models of TDP-43 overexpression are also typically toxic and documented to cause motor phenotype. Specifically, overexpression of

wildtype TDP-43 was found to enable TDP-43 pathology and neurological symptoms in mice, and the cessation of the overexpression reversed both the TDP-43 aggregate formation and the concomitant phenotype (Walker et al. 2015). Overexpression of transgenic human TDP-43 with the NLS removed was also found to cause neurodegeneration in mouse models despite the presence of endogenous TDP-43 in the nucleus, suggesting loss of function effects were not taking place (Igaz et al. 2011).

One study developed mouse models with human TDP-43 constructs harbouring either the M337V or Q331K mutation and showed TDP-43 expression levels comparable to wildtype animals (Arnold et al. 2013). In these models, the nuclear depletion and subsequent TDP-43 mislocalisation and pathology were not detected, suggesting no loss of function or toxic gain of new function. However, the mouse models did show substantial RNA dysregulation with several examples of excessive TDP-43-mediated exon repression, and developed progressive motor axon degeneration and motor neuron death, and this phenotype was not detected upon transgenic expression of wildtype human TDP-43 in a similar manner. Therefore, this study would strongly suggest the excessive gain of normal TDP-43 function as a result of TDP-43 mutation as a primary mechanism. Further evidence for this method is the knowledge that TDP-43 overexpression toxicity is well-documented to rely on the presence of the RBD (Ash et al. 2010; Voigt et al. 2010; Ihara et al. 2013), and several other RBPs have also been implicated in both ALS and a variety of other diseases, implying that the typical protein-RNA interaction is crucial and its alteration is central to toxic effects (Kapeli et al. 2017).

In simple organisms, the study of TDP-43 loss vs overexpression has been investigated by controlled dysregulation of the protein specifically in either the nuclear or cytoplasmic compartments. In flies, using the *TBPH* homologue, studies have contrastingly found that either both the nuclear depletion and cytoplasmic

overexpression are toxic (Miguel et al. 2011), or that only the cytoplasmic overexpression alone is harmful (Ritson et al. 2010). Other studies examining both mechanisms have reported both as detrimental, and further that both result in a similar phenotype (Diaper et al. 2013), even at the transcriptome level (Vanden Broeck et al. 2013). It seems likely that loss of normal function and gain of function are both likely to prove inimical to normal TDP-43 activities; further work is required to understand precisely how each mechanism contributes to the overall pathogenicity that results.

1.3 STUDY OF RNA

The work in this thesis primarily focuses on the RNA dysregulation that results from the F210I and M323K TDP-43 mutations in mouse models. In order to study this in a high-throughput manner, RNA sequencing (RNA-Seq) of these models was utilised. RNA-Seq is an example of an application of next-generation sequencing (NGS), a technology which uses the sequencing-by-synthesis method. First notably utilised by Margulies et al. in 2005, NGS enables the investigation of a substantial and sizeable number of target genetic or genomic features in parallel, providing large scale datasets for downstream analysis and enabling focus on features of interest in the context of the complete genetic setting. As a result, several ambitious projects detailing complete analysis of features such as transcription and chromatin structure across the entire genome (ENCODE Project Consortium 2012), and in tissue-specific settings (Melé et al. 2015), have been conducted.

Previous high-throughput methodology, such as hybridisation-based microarrays, still required the prior selection of targets, introducing selection bias, and further needed post-experiment removal of background hybridisation levels where RNA-Seq enables accurate identification of expression levels. Drawbacks with RNA-Seq exist, including

batch effects, but these mostly result from prior sample processing and library preparation and methods exist to incorporate this variation into downstream analysis. The use of RNA-Seq is particularly important in the context of ALS, for which focus on transcriptomics is not new (Malaspina et al. 2001), given the growing understanding of the importance of RBPs and subsequent mass RNA processing changes that result from their mutation and dysregulation.

1.5 AIMS

The aims of this study revolved around the characterisation and analysis of the F210I and M323K TDP-43 mutant models at the transcriptomic, cellular and phenotypic level in the context of ALS. The primary focus of this investigation was to understand how the mutations of TDP-43 disrupted RNA processing resulting from mutation of either the RBD or LCD of TDP-43 using high-throughput RNA-Seq analysis. Following on from this was the aim of further identifying specific molecular signatures which could be linked towards understanding the precise mechanism by which TDP-43 mutation and dysfunction could culminate in neurodegeneration and ALS. Alongside study of the transcriptome, overall phenotypes was also investigated, particularly those previously described in ALS. Initiated at the MRC Mammalian Genetics Unit, Harwell, by Abraham Acevedo-Arozena, the complete characterisation of the models was completed by me and several others across a number of institutions. The results of this study have now been published in (Fratta et al. 2018; Sivakumar et al. 2018). Summarised aims involved:

- RNA sequencing and bioinformatics analysis of the F210I and M323K mouse models at multiple time points to form complete picture of RNA dysregulation.
- Identification of molecular signatures that provide clues as to how the two mutations disrupt TDP-43 activity, and how they may link to ALS.

- Study of phenotype developed by the TDP-43 mutant models, with a particular focus on ALS or neurodegeneration-linked symptoms.
- Incorporation of previously published transcriptomic analysis to provide a more complete understanding of TDP-43 dysfunction in ALS.
- Validation of key RNA disruption events in these models that remain conserved in TDP-ALS patient settings.

CHAPTER 2

MATERIALS AND METHODS

2.1 TRANSCRIPTOMICS

2.1.1 RNA ISOLATION

Dissected tissues were homogenised using the TissueRuptor (Qiagen) in 1 ml QIAzol (Qiagen) until uniformly homogenous. 350 µl of the lysate was transferred to LoBind microcentrifuge tubes (Sigma-Aldrich) for each RNA extraction and left to rest at room temperature for 5 min to allow dissociation of nucleoprotein complexes. After addition of 70 µl 1-bromo-3-chloropropane (BCP), the tubes were shaken vigorously for 15 s to mix, and from this stage the miRNeasy Mini Kit (Qiagen) was used according to the manufacturer's protocol. The mix was incubated at room temperature for 3 min, and then centrifuged at 4 °C at 12,000 xg for 15 min. The upper aqueous phase was then transferred to a new LoBind tube and the RNA was precipitated with 1.5 volumes of absolute ethanol. The samples were then transferred to RNeasy Mini spin columns and centrifuged at 8,500 xg for 15 s. 350 µl Buffer RWT was then added to the columns and they were further centrifuged at 8,500 xg for 15 s. RNase-free DNase (Qiagen) was made by adding 70 µl Buffer RDD to 10 µl DNase I stock solution (solubilised in ddH_2O), and 80 µl of this solution was added to the spin columns to remove DNA contaminants. The columns were incubated with the DNase solution for 15 min, before being washed with 350 µl Buffer RWT at 8,500 xg for 15 s. The columns were then washed twice with 500 µl Buffer RPE, first at 8,500 xg for 15 s and then at 8,500 xg for 2 min. Final RNA extract was eluted in 30 µl RNase-free water by centrifugation at 8,500 xg for 1 min. RNA concentrations were measured using the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) at absorption wavelengths of 260 nm and 280 nm.

2.1.2 RNA QUALITY CONTROL

RNA integrity was tested by electrophoresing RNA samples through RNA ScreenTape (Agilent Technologies) using the Agilent 2200 Tapestation System (Agilent Technologies). The electrophoretic trace of the RNA was used to determine the RNA Integrity Number (RIN), which enabled quantification of RNA integrity. In order to conduct these measures, the RNA ScreenTape, RNA ScreenTape Sample Buffer and RNA ScreenTape Ladder (all Agilent Technologies) were allowed to equilibrate to room temperature for 30 min. Then, 1 µl of each RNA sample, and also 1 µl RNA ScreenTape Ladder, was added to 5 µl RNA ScreenTape Sample Buffer. Each mix was vortexed using the IKA vortexer and adaptor (Agilent Technologies) at 2,000 rpm for 1 min. The samples were then heated to 72 °C for 3 min, immediately cooled on ice for 2 min, and loaded into the Agilent 2200 Tapestation System alongside the RNA ScreenTape. Gel electrophoresis was performed using the RNA ScreenTape programme. Completion of the programme provided the RIN score for each RNA sample, and RNA samples with scores of 8 or higher were used for experiments.

2.1.3 REVERSE TRANSCRIPTION POLYMERASE CHAIN REACTION

Reverse transcription polymerase chain reaction (RT-PCR) of the RNA to complementary DNA (cDNA) was achieved using the SuperScript III Reverse Transcriptase kit (Thermo Fisher Scientific) following the manufacturer's instructions. Typically, the desired volume of RNA was diluted in ddH_2O to give a volume of 13 µl. To this, 1 µl oligo dT_{20} (50 µM) and 1 µl dNTP Mix (10 mM each of dATP, dGTP, dCTP and dTTP, all Thermo Fisher Scientific) were added. The samples were heated at 65 °C for 5 min and then incubated on ice for 2 min. To the cooled samples, 4 µl 5X First-Strand Buffer, 1 µl 0.1 M DTT, 1 µl RNaseOUT and 1 µl SuperScript III RT (all Thermo Fisher Scientific) were added. The samples were then incubated at 50 °C for 1 h, and the

reaction was inactivated by heating at 70 °C for 15 min, as in Table 2.1. The cDNA was further diluted between 1:1 and 1:5 with ddH_2O as required to give the final working concentration of cDNA. For post-mortem tissue-derived RNA, the desired volume of RNA was diluted in ddH_2O to give a lesser volume of 12 µl. To this, 1 µl oligo dT_{20} (50 µM) and 1 µl random hexamers were added, alongside the 1 µl dNTP Mix. The heating to 65 °C, cooling and addition of further reagents were carried out as above. Before the 50 °C heating step, the samples were heated to 37 °C for 10 min. Further steps were carried out as above, also outlined in Table 2.1.

	RT-PCR		RT-PCR post-mortem tissue	
Step	**Temp**	**Time**	**Temp**	**Time**
1	65 °C	5 min	65 °C	5 min
2	Further addition of reagents		Further addition of reagents	
3			37 °C	10 min
4	50 °C	1 h	50 °C	1 h
5	70 °C	15 min	70 °C	15 min

Table 2.1. PCR programmes for RT-PCR.

2.1.4 POLYMERASE CHAIN REACTION

Polymerase chain reactions (PCRs) were performed using the 2X PCR Master Mix (Thermo Fisher Scientific). A 10 µl reaction was set up, consisting of 5 µl 2X PCR Master Mix, 1 µl of the appropriate forward and reverse primers (10 µM), and the remaining volume of cDNA diluted in ddH_2O. For certain alternative splicing events, an alternative reaction mix was set up consisting of three primers – either two forward primers and one reverse, or one forward primer and two reverse. In this case the total primer volume constituted 3 µl, but the total reaction volume was maintained at 10 µl. PCR cycling conditions for each experiment are described in Table 2.2, Table 2.3. Typically, the PCR consisted of an initial heating step to 95 °C for 3 min, and then cycling repeats between 35-40 times as required for final product volume. In each cycling step, DNA denaturation was conducted at 95 °C for 20 s, primer annealing at

between 58-62 °C for 30 s as recommended for specific primer pairs, and primer extension at 72 °C for 30 s. Following the PCR, samples were allowed to cool prior to further experiments or analysis.

	Cryptic exon PCR		Skiptic exon PCR		Skiptic exon human PCR	
Step	**Temp**	**Time**	**Temp**	**Time**	**Temp**	**Time**
1	95 °C	3 min	95 °C	3 min	95 °C	3 min
2	95 °C	20 s	95 °C	20 s	95 °C	20 s
3	60 °C	30 s	58-62 °C	30 s	60 °C	30 s
4	72 °C	30 s	72 °C	30 s	72 °C	30 s
	Steps 2-4 repeated 37X		Steps 3-4 repeated 35-37X		Steps 3-4 repeated 40X	

Table 2.2. PCR programmes used in this thesis (1/2).

	Skiptic exon human 3 primer PCR		3 primer 3'UTR PCR		Long gene qPCR primer test	
Step	**Temp**	**Time**	**Temp**	**Time**	**Temp**	**Time**
1	95 °C	3 min	95 °C	3 min	95 °C	3 min
2	95 °C	20 s	95 °C	20 s	95 °C	20 s
3	60 °C	30 s	58 °C	30 s	60 °C	30 s
4	72 °C	30 s	72 °C	30 s	72 °C	30 s
	Steps 3-4 repeated 35-40X		Steps 2-4 repeated 35X		Steps 3-4 repeated 40X	

Table 2.3. PCR programmes used in this thesis (2/2).

To visualise the PCR results, the product was electrophoresed on either the Agilent 2200 Tapestation System or on 1.5 % agarose gels. The agarose gels were made by dissolving agarose (Thermo Fisher Scientific) in 1X Tris-borate-EDTA (TBE) buffer (National Diagnostics) in a microwave for 3 min. The solution was then cooled and 0.05 µl/ml ethidium bromide (Sigma-Aldrich) was added. The solution was then poured into a gel cast and left to set for 20 min. Samples were loaded into the wells, with HyperLadder 100 bp (Bioline) used as the molecular weight marker. Gel electrophoresis was conducted in 1X TBE buffer at 120 V for 45 min, and the gel was visualised with UV light to acquire the resulting images.

To visualise using the Agilent 2200 Tapestation System, the PCR products were electrophoresed on D1000 ScreenTape (Agilent Technologies). The D1000 ScreenTape,

D1000 Ladder and D1000 Sample Buffer (all Agilent Technologies) were allowed to equilibrate to room temperature for 30 min. 3 µl D1000 Sample Buffer was then added to 1 µl of each PCR product sample, and also 1 µl D1000 Ladder. All samples were then mixed using the IKA vortexer and adaptor at 2,000 rpm for 1 min, and then loaded into the Agilent 2200 Tapestation System along with the D1000 ScreenTape. The D1000 ScreenTape programme was executed, and the results of the gel electrophoresis were visualised to analyse the size and quantity of the PCR product samples.

2.1.5 QUANTITATIVE PCR

Quantitative PCR (qPCR) was used to investigate mRNA expression. Primers were tested for dimer formation or non-specific products by electrophoresing the products of the qPCR programme on an agarose gel (see 'Polymerase Chain Reaction: Long gene qPCR primer test'). The 15 µl qPCR mix consisted of 7.5 µl PowerUp SYBR Green Master Mix (Thermo Fisher Scientific), 1 µl appropriate forward and reverse primers (10 µM), and the remaining volume of cDNA in ddH_2O. The qPCR was performed on a 7500 Fast Real-Time PCR System (Thermo Fisher Scientific) to be quantified via the comparative CT method. The reaction involved heating to 50 °C for 20 s, 95 °C for 10 min, and then 40 cycles of 95 °C for 15 s followed by 60 °C for 1 min, as outlined in Table 2.4. An optional melting curve was also subsequently used as described in steps 5-8 of Table 2.4. 7500 software was used to calculate the baseline and Ct values of each sample. Relative gene expression was calculated by subtracting the mean Ct value of reference genes from the Ct value of the gene of interest, to give ΔCt. The ΔCt for one sample of interest was then subtracted from the ΔCt of a control to give the ΔΔCt value, and by using the calculation $2^{-\Delta\Delta Ct}$, the relative gene expression change between two samples was determined.

			40 cycles		Optional melting curve			
Temp	50 °C	95 °C	95 °C	60 °C	95 °C	60 °C	95 °C	60 °C
Time	20 s	10 min	15 s	1 min	15 s	1 min	30 s	15 s

Table 2.4. QPCR programme with optional melting curve step.

2.1.6 RNA-SEQUENCING

Following RNA quality control, RNA samples to be sequenced were normalised to the desired concentration, typically consisting of 500-1500ng of RNA diluted into 30 µl ddH_2O. For standard poly-A-purified or ribo-depleted RNA sequencing libraries, the library preparation and sequencing stages were outsourced to the Wellcome Trust Centre for Human Genetics, Oxford (WTCHG). Library preparation was conducted using the TruSeq Stranded mRNA HT Sample Prep Kit (Illumina) with poly-A purification or ribo-depletion as required. Paired-end multiplex sequencing of the cDNA libraries was conducted on either the HiSeq 2500 or HiSeq 4000 Sequencing System (both Illumina).

Model	Tissue	Genotypes and sample sizes	Read length (bp)	Average reads per sample (~millions)
F210I	E14.5 head	3 WT, 3 HET, 3 HOM	35 x2	20
	E14.5 forebrain	4 WT, 4 HOM	150 x2	60
	Adult spinal cord	5 WT, 7 HET	75 x2	45
	MEFs	3 WT, 3 HET, 3 HOM	50 x2	10
M323K	E14.5 head	3 WT, 3 HET, 3 HOM	35 x2	20
	Adult spinal cord	4 WT, 4 HET, 5 HOM	75 x2	45
	MEFs	3 WT, 3 HET, 3 HOM	50 x2	10
F210I / M323K compound HET	E14.5 head	3 WT, 3 COMPOUND HET	35 x2	20
TDP-43-KD	MEFs	3 WT, 3 KD	50 x2	10
TDP-ALS patient	Fibroblasts	4 WT, 4 HET	35 x2	25

Table 2.5. RNA-Seq experiments conducted for this thesis. Read lengths given x2 in paired-end datasets.

2.1.7 BIOINFORMATICS ANALYSIS

RNA-Seq results were provided from the WTCHG in the FASTQ format and data processing was conducted by me or a member of Dr. Vincent Plagnol's lab. All samples were first examined using FastQC (Andrews 2010) to check read quality and for the presence of adapters. Any reads determined to contain base calls of low quality (phred score < 20), or adaptor sequences, were trimmed with Trim Galore! 0.4.1 (Krueger 2012). Reads were then aligned to the mouse genome mm10 or human genome hg38 as appropriate with STAR aligner v2.4.2a using the two-pass alignment method (Dobin et al. 2013). The aligned reads were sorted and duplicate reads marked with NovoSort 1.03.09 (Novocraft).

For gene expression analysis, reads were grouped by gene to form a count matrix using HTSeq (Anders et al. 2015). Differential gene expression was determined using DESeq2 (Love et al. 2014). Briefly, sample library sizes were normalised for sequencing depth and library composition using scaling factors calculated from the median of gene count ratios between each sample and the average across all samples. Genes with low adjusted counts were filtered from the analysis. Variability of each gene was estimated using maximum likelihood estimate, and a curve was fitted to the distribution to provide an approximate value of dispersion across the range of normalised counts. The gene-wise dispersion estimates were then shrunk towards the curve using an empirical Bayes approach. A negative binomial generalised linear model was used to derive maximum likelihood estimates for the $\log_2$ fold change across two conditions and another empirical Bayes approach was used to shrink the $\log_2$ fold changes towards 0. The Wald test was used to test the shrunken $\log_2$ fold changes for statistical significance of differential gene expression. P-values were corrected for multiple-testing using the Benjamini-Hochberg method with a false discovery rate of 10% as

suggested by the authors of DESeq2 for exploratory differential gene expression analysis (Love et al. 2014).

Alternative splicing analysis was conducted using SGSeq (Goldstein et al. 2016). Briefly, SGSeq identifies exons and splice junctions present across all samples from spliced reads, and forms a splice graph which contains this information in the form of splicing events. Novel splicing events were included in downstream analysis. Events were classified appropriately as cassette exons, retained introns, alternative 5' junctions, alternative 3' junctions, alternative first exons, alternative last exons, mutually exclusive exons or some combination of the above. Quantification and differential use of these splice sites were conducted with DEXSeq (Anders et al. 2012). Normalisation and dispersion estimation was carried out as outlined above in differential gene expression analysis. Statistical analysis was conducted by fitting a generalised linear model accounting for the condition of interest and using chi-squared distribution to test whether the usage of each splicing event was statistically significant. P-values were corrected for multiple-testing using the Benjamini-Hochberg method with a false discovery rate of 1% to minimise false positives and identify reliably altered individual splicing events.

2.1.8 GENE ONTOLOGY ANALYSIS

To determine specific biological processes significantly affected in particular disease conditions, Gene Ontology (GO) analysis was conducted using g:Profiler (Reimand et al. 2016). A statistical over-representation test was used in order to identify biological processes most represented by genes dysregulated in disease condition datasets. Over-representation was determined by comparing the set of significantly dysregulated genes to the overall expressed gene set.

2.2 INDIVIDUAL NUCLEOTIDE RESOLUTION CROSS-LINKING AND IMMUNOPRECIPITATION

Individual nucleotide resolution cross-linking and immunoprecipitation (iCLIP) was conducted to assess direct binding of TDP-43 to its RNA targets (Huppertz et al. 2014). Several solutions were prepared for this protocol – protocol is depicted in Figure 2.10, solutions are outlined in Table 2.6 - Table 2.10.

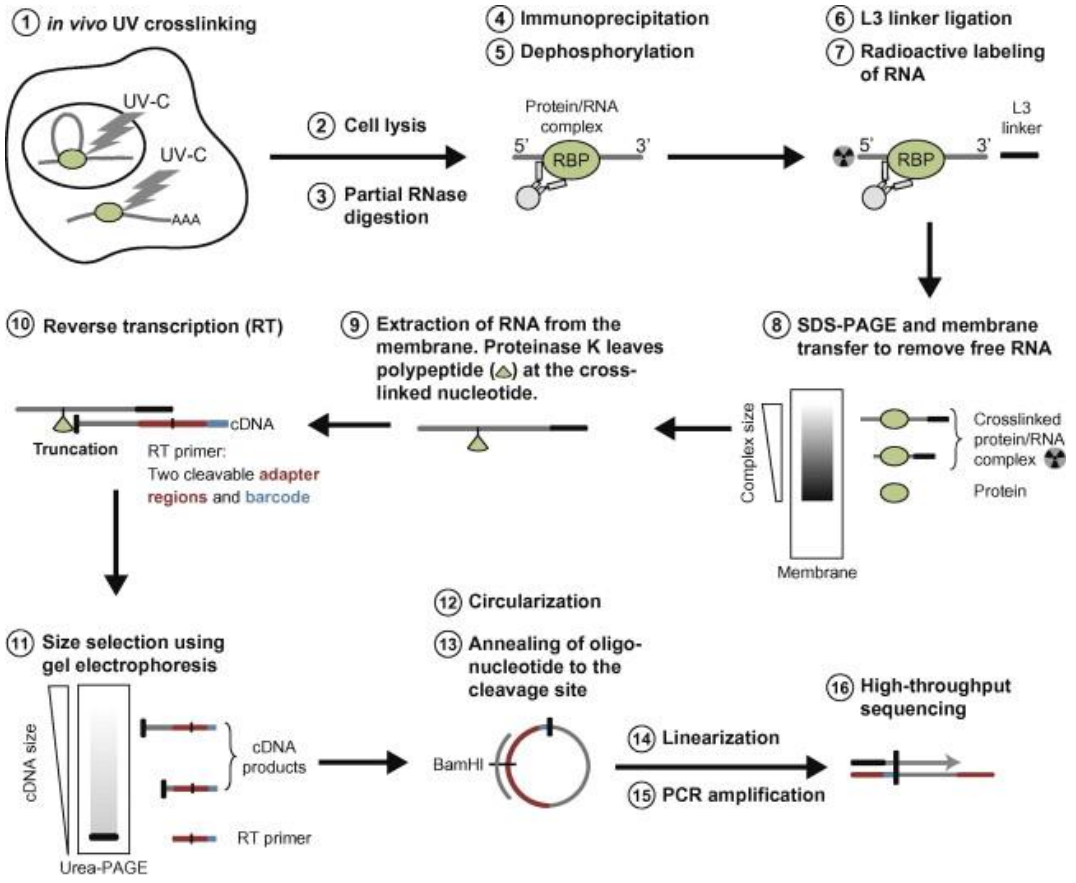


Figure 2.10. Schematic overview of iCLIP protocol from sample crosslinking through to libraries ready for sequencing (Huppertz et al. 2014).

Lysis buffer	High-salt wash	PNK buffer	5x PNK pH 6.5 buffer
50 mM Tris-HCl pH 7.4	50 mM Tris-HCl pH 7.4	20 mM Tris-HCl pH 6.5	350 mM Tris-HCl pH 6.5
100 mM NaCl	1 M NaCl	10 mM $MgCl_2$	50 mM $MgCl_2$
1 % Igepal CA-630 (Sigma-Aldrich)	1 % Igepal CA-630 (Sigma-Aldrich)	0.2 % **Tween**-20	5 mM dithiothreitol
0.1 % SDS	0.1 % SDS		
0.5 % sodium deoxycholate	0.5 % sodium deoxycholate		

Table 2.6. Table of buffers utilised in iCLIP experiment (1/2).

4x Ligation buffer	PK buffer	PK buffer + 7M urea	StrepBead wash buffer
200 nM Tris-HCl pH 7.8	100 mM Tris-HCl pH 7.4	100 mM Tris-HCl pH 7.4	100 mM Tris-HCl pH 7
40 mM $MgCl_2$	50 mM NaCl	50 mM NaCl	1 M NaCl
4 mM dithiothreitol	10 mM EDTA	10 mM EDTA	10 mM EDTA
		7 M urea	0.1 % Tween-20

Table 2.7. Table of buffers utilised in iCLIP experiment (2/2).

3' end RNA dephosphorylation		3' adaptor ligation		5' end labelling	
Reagent	**µl**	**Reagent**	**µl**	**Reagent**	**µl**
5X PNK pH 6.5 buffer	4	4X Ligation buffer	5	T4 PNK (NEB)	0.2
T4 PNK (NEB)	0.5	RNA ligase (NEB)	1	^{32}P-γ-ATP	0.5
RNasin (Promega)	0.5	RNasin (Promega)	0.5	10X PNK buffer (NEB)	0.4
ddH_2O	15	Pre-adenylated L3-App 20 µM	1.5	ddH_2O	2.9
		PEG400	4		

Table 2.8. Table of reaction mixes utilised in iCLIP experiment (1/3).

Reverse transcription		5’ end primer ligation		cDNA precipitation	
Reagent	**µl**	**Reagent**	**µl**	**Reagent**	**µl**
iCLIP primer (HPLC purified)	1	10X CircLigase buffer II (Epicentre)	0.8	Glycoblue (Ambion)	1
dNTP mix (10 mM)	0.5	CircLigase (Epicentre)	0.4	Sodium acetate (3 M) pH 5.5	40
Heat to 70°C for 5 min		$MnCl_2$ (50 mM)	0.3	Absolute ethanol	1000
5X RT buffer	2	ddH_2O	6.5		
0.1 M DTT	0.5	Heat to 60 °C for 1 h			
RNasin	0.25	FastDigest buffer (Fermentas)	3		
SuperScript III	0.25	10 µM Cut_oligo	1		
		ddH_2O	26		

Table 2.9. Table of reaction mixes utilised in iCLIP experiment (2/3).

Preparatory PCR amplification		PCR amplification	
Reagent	**µl**	**Reagent**	**µl**
cDNA	1	cDNA	10
P5Solexa/P3Solexa (10 µM)	0.25	P5Solexa/P3Solexa (10 µM)	1
Accuprime Supermix 1 (Thermo Fisher Scientific)	5	Accuprime Supermix 1 (Thermo Fisher Scientific)	20
ddH_2O	3.75	ddH_2O	9

Table 2.10. Table of reaction mixes utilised in iCLIP experiment (3/3).

2.2.1 UV CROSS-LINKING

Dissected tissues were homogenised at a ratio of 10 µl phosphate buffered saline (PBS) (VWR International) per mg of tissue by thoroughly passing tissue through sequentially smaller pipette tips. The lysate was transferred to an appropriately sized tissue culture plate, placed on ice, and irradiated four times with UV light at 100 mJ/cm^2 using the Stratagene Stratalinker 2400 (Agilent Technologies) at a wavelength of 254 nm. The lysate was then snap frozen on dry ice until further use.

2.2.2 IMMUNOPRECIPITATION AND LABELLING

Immunoprecipitation was conducted using antibodies conjugated to magnetic beads to pull out the protein of interest. 100 µl Dynabeads Protein A (Thermo Fisher Scientific) was washed twice with 900 µl lysis buffer (Table 2.6) before being conjugated with 5 µg of the desired antibody and further washed with 900 µl lysis buffer. A small proportion of the beads was not conjugated to antibody and kept to be used as a negative control. Separately, the cell lysate protein concentrations were measured and normalised using the DC Protein Assay (Bio-Rad Laboratories). To determine concentration, five dilutions BSA protein standards were prepared in lysis buffer at 2 mg/ml, 1 mg/ml, 0.5 mg/ml, 0.25 mg/ml and lysis buffer only, and 5 µl were plated into a 96-well plate in duplicate. 5 µl protein samples were plated in duplicate undiluted. 20 µl reagent S, a surfactant solution for colorimetric assays, was added to 1 ml reagent A, an alkaline copper tartrate solution, and 25 µl this mix was added to each standard and protein sample. 200 µl reagent B, a dilute Folin reagent, was then added to each sample, and the samples were put on a gentle shake for 10 min before being analysed using a Tecan Sunrise Reader (Tecan) to measure the absorbance signal intensity at 562 nm using XFluor software (Tecan).

Once protein concentration had been standardised, samples were sonicated for ten cycles with alternating 30 s on/off at low intensity using the Bioruptor Sonicator (Diagenode). 0.1 units/ml RNase I (Thermo Fisher Scientific) and 2 µl Turbo DNase (Thermo Fisher Scientific) were then added to shorten RNA transcripts and remove bound DNA. A sample was treated with a higher concentration of 0.25 units/ml RNase I as a negative control. Both variants were incubated at 37 °C for 3 min shaking at 1,100 rpm. The samples were then centrifuged at 4 °C at 21,800 xg for 10 min. These lysates were then mixed with the antibody-conjugated beads (or no-antibody beads as negative control) at 4 °C for 16 h. The beads were then washed twice with high-salt

wash and further twice with PNK buffer (both in Table 2.6), before being resuspended in 1 ml PNK buffer. The RNA transcripts attached to the immunoprecipitated protein were then dephosphorylated at the 3' end using the 3' end dephosphorylation mix (Table 2.8) incubated at 37 °C for 20 min in an Eppendorf Thermomixer (Eppendorf) shaking at 1,100 rpm. The samples were washed with PNK buffer, then high-salt wash, then twice more with PNK buffer. To ligate the adaptor to the 3'end of the transcripts, the samples were resuspended in 20 µl 3' adaptor ligation mix (Table 2.8) and incubated at 16 °C for 16 h in an Eppendorf Thermomixer shaking at 1,100 rpm. The samples were then washed twice with PNK buffer, then high-salt wash, then again twice with PNK buffer. 20 % of each sample was collected for radiolabelling. In a lab designated to radioisotope experimentation, 4 µl 5' end labelling mix (Table 2.8) was added to these samples and incubated at 37 °C for 5 min shaking in an Eppendorf Thermomixer at 1,100 rpm. Samples were then washed with 200 µl PNK buffer before 20 µl NuPAGE loading buffer (Invitrogen) was added. These samples were then added to the remaining 80 % of samples not radiolabelled.

2.2.3 SDS-PAGE AND NITROCELLULOSE TRANSFER

The samples were loaded onto a 4-12 % NuPAGE Bis-Tris 1.5 mm 10-well gel (Invitrogen). All 20 µl each sample was loaded onto the gel, alongside 4 µl the protein size marker PAGE ruler plus (Fermentas). Gels were used in conjunction with the XCell SureLock Mini-Cell Electrophoresis System (Thermo Fisher Scientific) filled with 0.5 l 1X NuPAGE MOPS running buffer (Invitrogen) for 50 min at 180 V. Once completed, the dye front was removed, and the samples transferred to 0.2 µm Protan BA85 Nitrocellulose Membrane (Whatman) using the XCell II Blot Module (Thermo Fisher Scientific) filled with 1X Tris-Glycine Electroblotting Buffer (National Diagnostics) with 10 % methanol added for 2 h at 30 V. After transfer, the membrane was washed with

PBS buffer and wrapped in saran wrap, before being exposed to Fuji film (Thermo Fisher Scientific) at -80 °C for either 1 h, 4 h or 16 h.

2.2.4 RNA ISOLATION

Using the high-RNase sample as a guide protein marker in the absence of RNA binding, the protein-RNA complexes were isolated by using the autoradiograph created in step 2.2.3 as a mask to cut the desired product sizes from the nitrocellulose membrane. The membrane fragments were then placed into LoBind tubes and 200 µl PK buffer was added to each fragment. 10 µl proteinase K (Roche) was further added, and the mix was incubated at 37 °C for 20 min shaking at 1,100 rpm. 200 µl PK buffer and 7 M urea were then added and the mix was again incubated at 37 °C for 20 min shaking at 1,100 rpm. The supernatant was collected and added, along with 400 µl Phenol:Chloroform:Isoamyl Alcohol (Sigma-Aldrich) to a 2 ml Phase Lock Gel Heavy tube (VWR). These were incubated at 30 °C for 5 min shaking at 1,100 rpm, and the phases were separated by then centrifuging for 5 min at 13,000 rpm. The aqueous layer was then transferred to a new tube and precipitated cDNA precipitation mix for 16 h at -20 °C. Once precipitated, the samples were centrifuged at 4 °C for 20 min at 15,000 rpm to collect as pellets. The pellets were resuspended in 5.5 µl ddH_2O.

2.2.5 REVERSE TRANSCRIPTION

Reverse transcription of the RNA to cDNA was carried out with the reverse transcription mix (Table 2.9). Once all reagents had been added, the samples were incubated at 25 °C for 5 min, 42 °C for 20 min, 50 °C for 40 min, 80 °C for 5 min and then cooled to 4 °C. 1.65 µl 1 M NaOH was then added and samples were incubated at 98 °C for 20 min to hydrolyse remaining RNA. 20 µl 1 M Hepes-NaOH pH 7.3 was added to eliminate radioactivity. At this point, the biological replicate samples were

multiplexed. 350 µl TE buffer and cDNA precipitation mix (Table 2.9) were both added to the samples to precipitate the cDNA at -20 °C for 16 h.

2.2.6 GEL PURIFICATION

Precipitate was collected by centrifuging samples at 4 °C for 15 min at 15,000 rpm. The subsequent pellet was washed with 500 µl 80 % ethanol (diluted in ddH_2O), centrifuged briefly to re-collect pellet, and resuspended in 6 µl ddH_2O. 6 µl 2X TBE-urea loading buffer (Invitrogen) was added to the cDNA, and also to the Low molecular weight protein marker (NEB). The XCell SureLock Mini-Cell Electrophoresis System was used for the gel electrophoresis, and the chamber was filled with 800 ml TBE running buffer. The full 12 µl of each sample was loaded to each well of a 6 % TBE-urea gel and electrophoresed for 40 min at 180 V. 2 µl SYBR green II stock solution was diluted in 10 ml TBE buffer, and the TBE-urea gel was immersed in this solution and visualised by UV transillumination.

Using a scalpel blade, two bands were cut from each sample, one between 75-95 nucleotides (nt) and the other between 95-200 nt. The isolated gel pieces were each placed into a 0.5 ml LoBind tube with three holes punctured into the bottom using a 21 G needle, and each of these tubes was then placed inside a 1.5 ml LoBind tube. 400 µl TBE was added, and the samples were centrifuged for 5 min at 8,000 rpm. The samples were incubated at 37 °C for 1 h shaking at 1,100 rpm, cooled on dry ice for 2 min and further incubated again at 37 °C for 1 h shaking at 1,100 rpm. The liquid supernatant was transferred to a Costar SpinX column (Corning Incorporated) with two 1 cm diameter glass pre-filters (Whatman) placed inside. The columns were centrifuged for 1 min at 13,000 rpm, and the resulting solution was added together with 400 µl RNA phenol/chloroform (Sigma-Aldrich) to a 2 ml Phase Lock Gel Heavy tube. These were incubated at 30 °C for 5 min shaking at 1,100 rpm, with phase separation subsequently

conducted through centrifugation for 5 min at 13,000 rpm. The aqueous portion was again added together with 400 µl RNA phenol/chloroform (Sigma-Aldrich) to a 2 ml Phase Lock Gel Heavy tube, incubated at 30 °C for 5 min shaking at 1,100 rpm, and phase separated by centrifugation for 5 min at 13,000 rpm. The aqueous layer was then transferred to a new 1.5 ml LoBind tube and precipitated with cDNA precipitation mix for 16 h at -20 °C.

2.2.7 5' END LIGATION OF PRIMER TO CDNA

Precipitate was collected by centrifuging samples at 4 °C for 15 min at 15,000 rpm. The subsequent pellet was washed with 500 µl 80 % ethanol (diluted in ddH_2O), centrifuged briefly to re-collect pellet, and resuspended in the 8 µl 5' end primer ligation mix (Table 2.9). The samples were incubated at 60 °C for 1 h. and the remaining 30 µl oligo annealing mix was added. The oligonucleotides were annealed by heating samples to 95 °C for 2 min, and then successively decreasing temperature by 1 °C every 20 s down to 25 °C. Following annealing, 2 µl BamHI (Fast Fermentas) restriction enzyme was incubated at 37 °C for 30 min and then at 80 °C for 5 min. Precipitation of the cDNA was conducted using the cDNA precipitation mix for 16 h at -20 °C.

2.2.8 PCR AMPLIFICATION

Prior to PCR amplification of the entire cDNA, a preparatory PCR was conducted on a small portion of cDNA to identify the appropriate number of cycles to be used for amplification. The preparatory PCR mix (Table 2.10) was processed using the PCR programme in Table 2.11. Samples were heated to 94 °C for 2 min, before undergoing 20-30 cycles of 94 °C for 15 s, 65 °C for 15 s and 68 °C for 45 s, with cycle number being approximated by PCR product concentration. Finally, samples were heated to 68 °C for 3 min and then cooled.

		20-30 cycles			
Temp	94 °C	94 °C	65 °C	68 °C	68 °C
Time	2 min	15 s	15 s	45 s	3 min

Table 2.11. PCR programme for preparatory PCR amplification.

The PCR product was then mixed with 2 µl 5X TBE loading buffer and loaded onto a 6 % TBE gel. The XCell SureLock Mini-Cell Electrophoresis System was used for the gel electrophoresis, and the chamber was filled with 800 ml TBE running buffer. The samples were electrophoresed at 180 V for 25 min. 2 µl SYBR green II stock solution was diluted in 10 ml TBE buffer, and the TBE-urea gel was immersed in this solution and visualised by UV transillumination. If over-amplification was present – in the form of smears extending above a solid band – then the cycle number was re-adjusted until over-amplification was no longer present. Once the desired cycle number had been established, the remaining cDNA was amplified. This was carried out in batches of two to guard against PCR failure ending the experiment, and the cycle number was reduced by one to adjust for the higher concentration of cDNA being used. The PCR amplification mix (Table 2.10) was used, and the PCR programme was the same as for the preparatory PCR amplification with the additional slight cycle number adjustment. All PCR products were then multiplexed. The ratio for the multiplexing of the shorter and longer bands was determined by the intensity of the signal from the TBE gel under UV light.

2.2.9 PURIFICATION OF LIBRARIES

Purification of the multiplexed libraries is conducted to remove remaining PCR primers. To do this, 1.8 µl paramagnetic AMPure XP beads (Beckman Coulter) were added per 1 µl PCR product and mixed. The samples were left to bind for 5 min, and the beads were then collected using a magnetic stand. The beads were washed twice with 500 µl 70 % ethanol (diluted in ddH_2O) and left to ensure the complete evaporation of the ethanol. The purified product was eluted in 40 µl TE buffer and vortexed for 30 s.

2.2.10 LIBRARY QUANTIFICATION AND SEQUENCING

To establish the concentration of the cDNA libraries using the Agilent 2200 Tapestation System, the PCR products were electrophoresed on D1000 ScreenTape. The D1000 ScreenTape, D1000 Ladder and D1000 Sample Buffer were allowed to equilibrate to room temperature for 30 min. 3 µl D1000 Sample Buffer was then added to 1 µl the multiplexed library, and also 1 µl D1000 Ladder. All samples were then mixed using the IKA vortexer and adaptor at 2,000 rpm for 1 min, and then loaded into the Agilent 2200 Tapestation System along with the D1000 ScreenTape. The D1000 ScreenTape programme was performed, and the results of the gel electrophoresis produced the size range and nanomolar concentration of the multiplexed library. Sequencing of the library was kindly conducted by the Ule Lab on the HiSeq 2500 using the standard Illumina protocol for 50bp single-end sequencing (Huppertz et al. 2014).

2.2.11 BIOINFORMATICS ANALYSIS

Processing of raw iCLIP data was conducted using the iCount pipeline (Curk 2010). Briefly, reads were mapped to the UCSC mm10/GRCm38 genome assembly, and only reads mapped uniquely to a single region were used for further analysis. Crosslink peaks were identified using the iCount peak finding algorithm, which compared enrichment levels in identified binding sites in comparison to shuffled data. Crosslink clusters were defined as peaks which passed a false discovery rate threshold of 0.05, and peaks five nucleotides apart or closer were merged into the same cluster. Similar methods of iCLIP analysis are explored in Haberman et al. (2017).

Kmer analysis was further conducted using the kmer enrichment branch of the iCount pipeline (Curk 2010). Pentamer (5-nt kmer) frequencies were determined for each possible pentamer in the datasets. Permutation analysis was used to determine the mean frequencies of each pentamer by chance, and pentamer enrichment involved

determining the magnitude and significance of the increase in actual pentamer frequency in comparison to mean random permutation frequency.

2.3 QUANTSEQ LIBRARY PREPARATION AND SEQUENCING

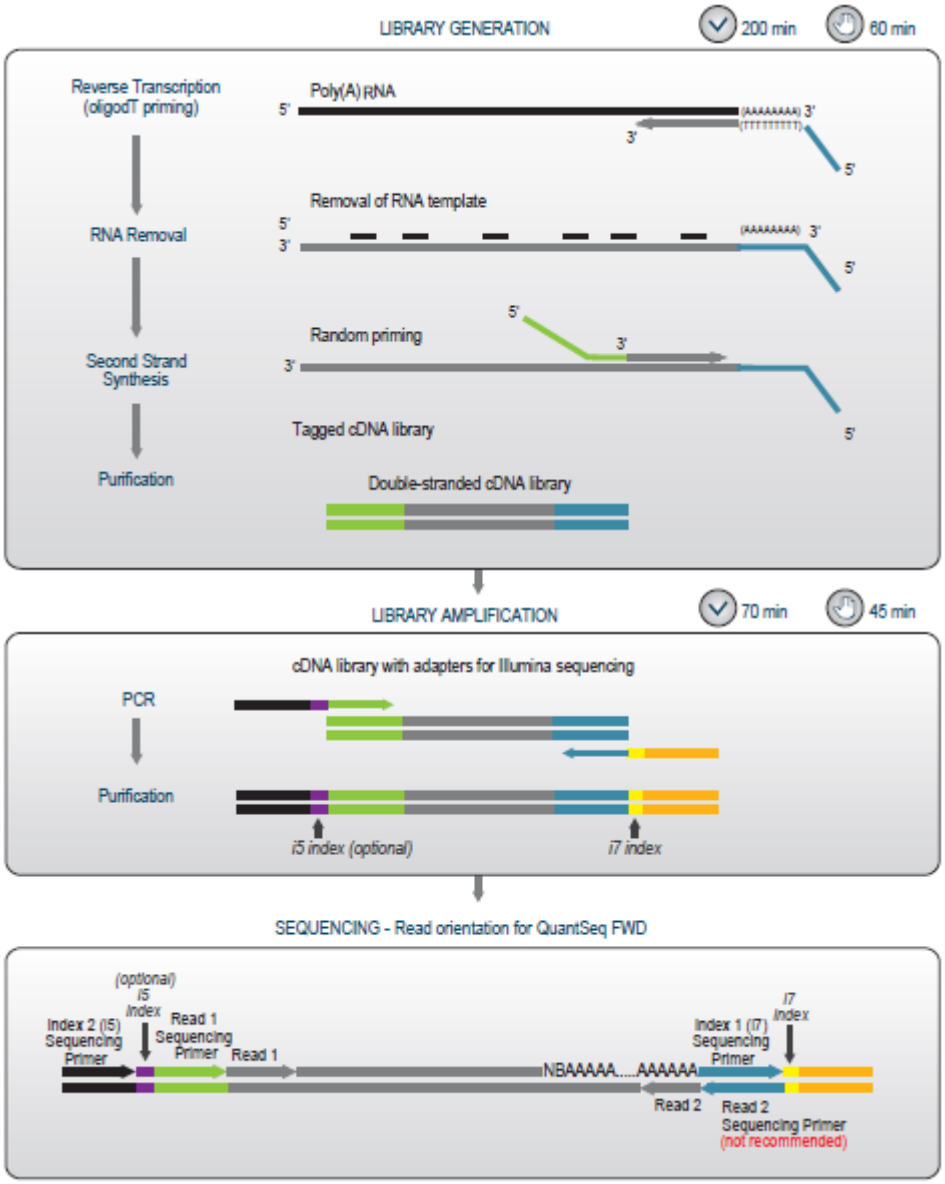


Figure 2.11. Schematic overview of QuantSeq protocol used (Lexogen).

2.3.1 FIRST STRAND CDNA SYNTHESIS

QuantSeq library preparation was conducted using the 3' mRNA-Seq Library Prep Kit FWD (Lexogen) following manufacturer's instructions. 150 ng of each RNA sample was diluted in 5 µl ddH_2O. 5 µl First Strand cDNA Synthesis Mix 1 (FS1) was added to each sample, and the mixture was incubated at 85 °C for 3 min, and then cooled to 42 °C. A mastermix containing 9.5 µl First Strand cDNA Synthesis Mix 2 (FS2) and 0.5 µl Enzyme Mix 1 (E1) per sample was prepared, and 10 µl of this mastermix was added to each RNA / FS1 mixture. Each mixture was further incubated at 42 °C for 15 min. Remaining RNA was removed by adding 5 µl RNA Removal Solution (RS) to each sample and aspirating. Each sample was incubated at 95 °C for 10 min and then cooled to 25 °C.

2.3.2 SECOND STRAND SYNTHESIS

To create double stranded DNA (dsDNA), 10 µl Second Strand Synthesis Mix 1 (SS1) was added to each sample. The mixture was incubated at 98 °C for 1 min and then slowly cooled to 25 °C at a rate of 0.5 °C/s. The mixture was further incubated at 25 °C for 30 min. A mastermix containing 4 µl Second Strand Synthesis Mix 2 (SS2) and 1 µl Enzyme Mix 2 (E2) was prepared, and 5 µl of the mastermix added to each sample. This mixture was incubated at 25 °C for 15 min.

2.3.3 PURIFICATION

To remove residual reaction components, samples were purified. 16 µl resuspended Purification Beads (PB) were added to each reaction, and incubated at room temperature for 5 min. The samples were then placed onto a magnetic rack for 5 min, and while the beads were attached to the side of the tube, all supernatant was removed and discarded. The samples were then removed from the magnetic rack, each sample was resuspended in 40 µl Elution Buffer (EB) and the mixture was incubated at room

temperature for 2 min. 56 µl Purification Solution (PS) was added to the mixture, mixed thoroughly and incubated at room temperature for 5 min. The samples were again placed onto a magnetic rack, the beads were allowed to collect for 5 min, and the supernatant was removed and discarded. Whilst remaining on the magnetic rack, 120 µl 80 % ethanol was added to the samples and incubated for 30 s. The supernatant was then removed and discarded, and this wash step was repeated once more. The samples were then left on the magnetic rack for 5 min to let the beads air dry, after which they were removed from the magnetic rack and rinsed with 20 µl EB and incubated at room temperature for 2 min. The samples were placed on the magnetic rack once more and this time 17 µl of the supernatant was collected and transferred to a fresh tube.

2.3.4 LIBRARY AMPLIFICATION

A mastermix of 7 µl PCR Mix (PCRM) and 1 µl Enzyme Mix 3 (E3) was prepared for each sample. 8 µl of the mastermix was added to each 17 µl library sample. 5 µl i7 index was added per sample, ensuring a unique i7 index was used for each library. The library was amplified by initially heating to 98 °C for 30 s, then 26 cycles consisting of each of 98 °C for 10 s, 65 °C for 20 s and 72 °C for 1 min, and then a final heating of 72 °C for 1 min followed by cooling to 10 °C, as outlined in Table 2.12 below.

		26 cycles			
Temp	98 °C	98 °C	65 °C	72 °C	72 °C
Time	30 s	10 s	20 s	30 s	1 min

Table 2.12. PCR programme for QuantSeq library amplification.

2.3.5 LIBRARY PURIFICATION

30 µl resuspended PB was added to each amplified library, and incubated at room temperature for 5 min. Samples were then placed on the magnetic rack and the beads were allowed to dry for 5 min. The supernatant was removed and discarded. Samples were removed from the magnetic rack and washed with 30 µl EB, beads resuspended

and incubated for 2 min. 30 µl PS was added to each sample and incubated at room temperature for 5 min. Samples were then placed back on the magnetic rack for 5 min, and the supernatant was removed and discarded. Samples were kept on the magnetic rack and washed with 120 µl 80 % ethanol with 30 s incubation twice. Beads were kept on the magnetic rack to dry for 5 min, and then 20 µl EB was added. Samples were removed from the magnetic rack and resuspended and incubated for 2 min. Samples were again placed onto the magnetic rack for 5 min, and this time the 17 µl supernatant was transferred to a fresh tube as final libraries.

2.3.6 LIBRARY QUANTIFICATION AND SEQUENCING

To establish the concentration of the cDNA libraries using the Agilent 2200 Tapestation System, the PCR products were electrophoresed on D1000 ScreenTape. The D1000 ScreenTape, D1000 Ladder and D1000 Sample Buffer were allowed to equilibrate to room temperature for 30 min. 3 µl D1000 Sample Buffer was then added to 1 µl of the multiplexed library, and also 1 µl D1000 Ladder. All samples were then mixed using the IKA vortexer and adaptor at 2,000 rpm for 1 min, and then loaded into the Agilent 2200 Tapestation System along with the D1000 ScreenTape. The D1000 ScreenTape programme was performed, and the results of the gel electrophoresis produced the size range and nanomolar concentration of the multiplexed library. Sequencing of the library was conducted by UCL Genomics on the MiSeq using the standard Illumina protocol for 50bp single-end sequencing. Processing of the data was conducted by Gregor Rot using his published pipeline (Rot et al. 2017).

2.4 CELL CULTURE

Standard cell culture practice was used to grow the immortalised MEFs and human fibroblasts. Cells were cultured in DMEM (Gibco) supplemented with 10 % FBS (Invitrogen), 1 % L-glutamine (Invitrogen) and 1 % penicillin / streptomycin

(Invitrogen) in 75 cm^2 culturing flasks in a 37 °C incubator with 5 % CO_2. Cells were split upon 80-90% visual confluence, typically after 48-72 hrs. Splitting involved washing cells with PBS (Invitrogen) and then adding 4 ml 0.05 % Trypsin (Invitrogen) and placing cells in the incubator for 2-5 min. 5 ml of warmed culture medium was then added to deactivate the trypsin. Cells were then collected and centrifuged at 300 xg for 5 min to pellet, and resuspended in 2 ml warmed culture medium. Cells were then re-plated in fresh culture medium in new 75cm^2 flasks. Typically cells were split 1:5. When required, cells were frozen into 1ml aliquots with 90 % cell culture medium and 10 % DMSO (Invitrogen) and stored at -80 °C short term, or in a liquid nitrogen tank long term.

2.5 IMMUNOHISTOCHEMISTRY

2.5.1 TISSUE FIXATION

Tissue fixation was outsourced to the MRC Mammalian Genetics Unit, Harwell. Briefly, mice were exsanguinated with 4 % PBS and then fixed with 4 % paraformaldehyde (PFA). The brain and spinal cord were dissected out and immersion fixed in 10 % buffered formal saline (BFS) (Pioneer Research Chemicals).

2.5.2 TISSUE PROCESSING

The brain and spinal cord were detached, and the spinal cord was further dissected into the cervical, thoracic and lumbar components. The cervical and lumbar regions were further equally subdivided into three regions, and the thoracic was subdivided into two regions, giving eight pieces of spinal cord tissue per animal. Each piece of tissue was placed into a plastic biopsy cassette in a pre-determined orientation and immersed in BFS. For paraffin embedding, samples were then placed in the Leica ASP300 S Automated Vacuum Tissue Processor (Leica Biosystems) and processed with

by immersion in 70 % ethanol for 1 hr once, 100 % ethanol for 1 hr five times, xylene for 1 hr three times and molten paraffin wax for 1 hr three times, as outlined in Table 2.13:

Reagent	Time	Repeat
70 % Ethanol	1 h	X1
100 % Ethanol	1 h	X5
Xylene	1 h	X3
Molten paraffin wax	1 h	X3

Table 2.13. Tissue processing programme for paraffin embedding.

Following the processing, samples were embedded in molten paraffin wax using the Leica EG1150H Tissue Embedding Center (Leica Biosystems) and then cooled to solidify using the Leica EG1150C cold plate (Leica Biosystems).

2.5.3 TISSUE SECTIONING

Tissue blocks were sectioned using the Leica RM2135 Rotary Microtome (Leica Biosystems) at -15 °C to obtain 4 µm-thick sections. Sections were then quickly mounted onto Superfrost Plus Microscope slides (Thermo Fisher Scientific). Sections were stored at 4 °C.

2.5.4 DAB STAINING

3,3′-Diaminobenzidine (DAB) staining was conducted by the Division of Neuropathology, UCL Institute of Neurology. Briefly, slides were pre-treated with SCC1, and the DAB-MAP kit was then used for 8 min with the Discovery XT (Ventana Medical Systems) before using staining protocols detailed in Table 2.14.

Primary antibody					Secondary antibody			
Target	**Species**	**Dilution**	**Epitope**	**Time**	**Target**	**Species**	**Dilution**	**Time**
P62	Rabbit	1:200	Polyclonal	6 h	Anti-rabbit	Swine	1:500	1 h
TDP-43	Rabbit	1:1000	Monoclonal	12 h	Anti-rabbit	Donkey	1:500	32 min
Ubiquitin	Rabbit	1:200	Polyclonal	4 h	Anti-rabbit	Donkey	1:500	1 h

Table 2.14. Antibodies used for immunohistochemistry using DAB staining.

2.5.5 FLUORESCENT STAINING

Antigen retrieval was required prior to immunohistochemical staining of the paraffin embedded sections. Antigen retrieval was conducted manually with the protocol in Table 2.15. This process involved immersing samples in xylene for 5 min three times, 100 % ethanol for 5 min twice, 90 % ethanol for 5 min once, 70 % ethanol for 5 min once and ddH_2O three times, submerging the samples in containers containing the required solution each time.

Reagent	Time	Repeat
Xylene	5 min	X3
100 % Ethanol	5 min	X2
90 % Ethanol	5 min	X1
70 % Ethanol	5 min	X1
ddH_2O	5 min	X3

Table 2.15. Section processing protocol for immunohistochemistry using fluorescent staining.

Sections were then placed in a pressure cooker (Criterion) filled with 1.5 l 10 mM sodium citrate buffer pH 6 and heated for 10 min at the highest setting. The sections were then gradually cooled and washed with PBS before blocking with 10 % normal horse serum diluted in PBS with 0.2 % Triton X-100 for 1 h. All primary and secondary antibodies were diluted in 10 % normal horse serum in PBS with 0.2 % Triton X-100 (Table 2.16). All sections were incubated with primary antibody for 16 h, and

secondary antibody for 1 h. Three PBS washes were conducted between the blocking and primary antibody stage, and primary antibody and secondary antibody stage.

Primary antibody					
Target	**Species**	**Dilution**	**Epitope**	**Brand**	**Catalogue no.**
P62	Guinea pig	1:250	Polyclonal	Abcam	Ab91526
TDP-43 (C-terminal)	Rabbit	1:1000	Polyclonal	Proteintech	12892-1-AP
TDP-43 (monoclonal)	Mouse	1:1000	Monoclonal	Abcam	Ab104223
TDP-43 (phosphorylated)	Rabbit	1:1000	Polyclonal	Proteintech	22309-1-AP
Tuj1	Rabbit	1:1000	Monoclonal	Sigma	T2200
Ubiquitin	Rabbit	1:200	Polyclonal	Santa Cruz	Sc-8017

Table 2.16. Primary antibodies used for fluorescence immunohistochemistry.

Secondary antibody				
Target	**Raised in**	**Dilution**	**Brand**	**Catalogue no.**
Anti-guinea pig	Goat	1:500	Abcam	Ab150188
Anti-rabbit	Donkey	1:500	Thermo Fisher	A-21206
Anti-mouse	Donkey	1:500	Thermo Fisher	A-21202

Table 2.17. Secondary antibodies used for fluorescence immunohistochemistry.

Following staining, sections were further washed three times with PBS and then mounted to Thickness No. 1.5 cover glasses (VWR) with Prolong Anti-fade Mountant with DAPI (Thermo Fisher Scientific). Imaging was conducted using an Inverted Zeiss LSM 710 confocal microscope (Carl Zeiss).

2.6 STATISTICAL ANALYSIS

Statistical hypothesis testing between two independent groups conducted with the two sample t-test where samples can be assumed to be normally distributed, and with the non-parametric Mann-Whitney test where normal distribution cannot be assumed and sample sizes were sufficiently low. Statistical hypothesis testing between multiple groups conducted with a one-way ANOVA where samples can be assumed to be normally distributed, and with the Kruskal-Wallis test where normal distribution cannot be assumed and sample sizes were sufficiently low.

P-values graphically denoted by stars: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Unless otherwise stated, error bars denote the mean and standard error of the mean.

For untargeted RNA-Seq data analysis, p-value refers to final p-value adjusted for multiple testing using the Benjamini-Hochberg method.

CHAPTER 3

F210I MODEL TRANSCRIPTOME DYSREGULATION

3.1 INTRODUCTION

3.1.1 RNA RECOGNITION MOTIFS IN ALS

The RBD of TDP-43 has been poorly studied in the context of ALS, particularly in the form of mouse models (Lutz 2018). This is to be expected, given the prominence of LCD mutations in TDP-ALS, but this does not lessen the need for understanding of how dysfunction of the RBD contributes to ALS pathology. Importantly, a mutation in the RBD of TDP-43, D169G in RRM1, has been found to cause ALS (Kabashi, Paul N. Valdmanis, et al. 2008). This mutation is found between α2 and β4 within loop 5 of RRM1 (Austin et al. 2014).

The effect of this mutation, and the reason it results in symptoms similar to the numerous LCD mutations, is unclear. Study of the mutation's effect on TDP-43 binding found that neither TDP-43-RNA binding nor TDP-43-hnRNP binding was substantially altered (Ayala et al. 2011; McDonald et al. 2011; Bentmann et al. 2012). It does, however, cause higher thermal stability of the protein (Austin et al. 2014), and this mutant TDP-43 is cleaved by caspase 3 more efficiently, resulting in greater production of the 35 kDa fragments (Chiang et al. 2016). This fragment has been associated with TDP-43 aggregation and toxicity, and the increase in cleavage efficiency was not caused by other TDP-ALS mutations outside the RBD, such as A90V, G287S, G294A, G298S and A315T (Chiang et al. 2016), suggesting that perhaps RBD mutations have a unique mechanism of pathology. The RBD has also been reported as important in the pathogenicity of the 35 kDa fragment and its role in sequestration of wildtype full-length TDP-43 (Che et al. 2015).

Even in study of the RBD of TDP-43, it is RRM1 which receives the most attention, due to the minor role RRM2 plays in direct binding to RNA (Lukavsky et al. 2013). RRM2 also plays a lesser role in TDP-43 binding to its target UG repeat motifs (Kuo et al. 2009). Furthermore, there exists a strong correlation between RNA binding affinity and inhibitory splicing function in RRM1, but the same is not true for RRM2 (Kuo et al. 2009). Double point mutation in two residues of the RRM2, K192A and E261A, had minimal effects in RNA binding affinity, whereas similar mutations in RRM1 had more substantial impact (Lukavsky et al. 2013). However, despite the lack of direct binding, RRM2 is still likely to have a strong influence in overall binding of TDP-43 with RNA. Interaction between RRM1 and RRM2 increases affinity for UG repeat motifs in comparison to RRM1 alone (Lukavsky et al. 2013). It may be involved in indirect control of TDP-43 dimerisation prior to target binding (Lukavsky et al. 2013), and additionally play a role in directing of splicing factors with less specific target motifs, such as hnRNP A2 (D'Ambrogio et al. 2009). Interestingly, mutation of RRM2 of TDP-43 in NSC-34 cells was associated with cytoplasmic TDP-43- and ubiquitin-positive inclusions, demonstrating that RRM2 mutation could incite TDP-43 pathology (Maurel et al. 2018).

RRM2 is also an area of focus in regards to TDP-43 pathology due to the cleavage of TDP-43. TDP-43 is cleaved to form 35 kDa and 25 kDa C-terminal fragments, both of which have been implicated in TDP-43 aggregation and toxicity (Johnson et al. 2008). In particular, the formation of the 25 kDa fragment involves cleavage of RRM2, at arginine residue 208. Study of this fragment has shown that it is able to recapitulate the pathological features of TDP-43 as seen in TDP-43 proteinopathies (Zhang et al. 2009; Wei et al. 2016). RRM2 is typically of stable structure, but cleavage at R208 leaves the remaining RRM2 region of the C-terminal fragment highly disordered and prone to aggregation (Wei et al. 2016). Its solubility is also lost (Zhang et al. 2009; Wang et al. 2013; Morgan et al. 2017), and the exposed β strands may assemble into amyloid-like

inclusions (Wang et al. 2013; Guenther et al. 2018). Alongside this, RRM2 cleavage also removes inter-domain interaction in TDP-43 to expose a large proportion of the LCD, making this domain too more prone to aggregation (Wei et al. 2016). Therefore, RRM2 cleavage may have dual action underlying TDP-43 proteinopathy.

In relation to the F210I mutation, phenylalanine residues have been identified as vital for RNA binding properties in protein-RNA interactions (Jones et al. 2001). For TDP-43 in particular, phenylalanine residues are crucial for UG repeat binding (Lukavsky et al. 2013). The F210I mutation lies in the helix α1 of TDP-43, and is unlikely to be involved in direct RNA binding (Ramos, 2015, personal communication). However, it may be important for indirect modulation of TDP-43-RNA processing functionality, especially as it is conserved between mice and humans. This also means that study of the F210I mouse model may be able to recapitulate the effects of similar mutations in human settings.

3.1.2 PREVIOUS STUDY OF THE F210I MOUSE MODEL

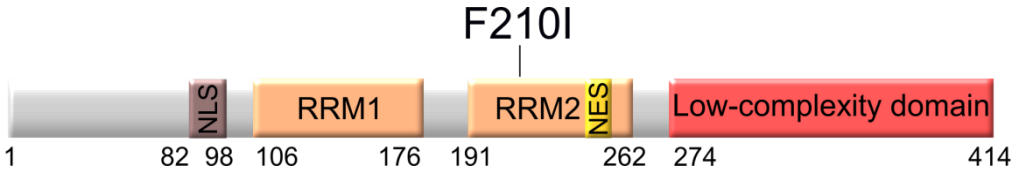


Figure 3.12. TDP-43 domains and location of the F210I mutation within the RRM2 domain.

Original characterisation of the F210I mutation (Figure 3.12) was conducted in vitro, with initial insight into the effects of the mutation on splicing investigated using the cystic fibrosis transmembrane conductance regulator (*CFTR*) minigene assay. TDP-43 has been well characterised in its ability bind to the UG repeats at the 3' splice site of exon 9 of the *CFTR* gene and repress inclusion of the exon (Niksic et al. 1999; Buratti & Baralle 2001; Buratti et al. 2001; Pagani, Stuani, et al. 2003; Pagani, Buratti, et al. 2003;

Buratti et al. 2004; Ayala et al. 2006; Buratti et al. 2007). As *CFTR* exon 9 is not spliced by TDP-43 in mice (Buratti et al. 2001), a minigene construct has been designed based on the exon and surrounding sequences, in order to test TDP-43 splicing function in mouse models. Wildtype TDP-43 typically splices exon 9 of this minigene at a ratio of 50 %, resulting in exon skipping in half of the minigene constructs, whereas siRNA-mediated TDP-43 KD in HeLa cells results in almost complete levels of exon 9 inclusion in the minigene. Further tests established that both human and mouse wildtype TDP-43 expression are able to rescue the KD effects and restore exon 9 splicing levels to 50 % (Figure 3.13) (Ricketts 2012).

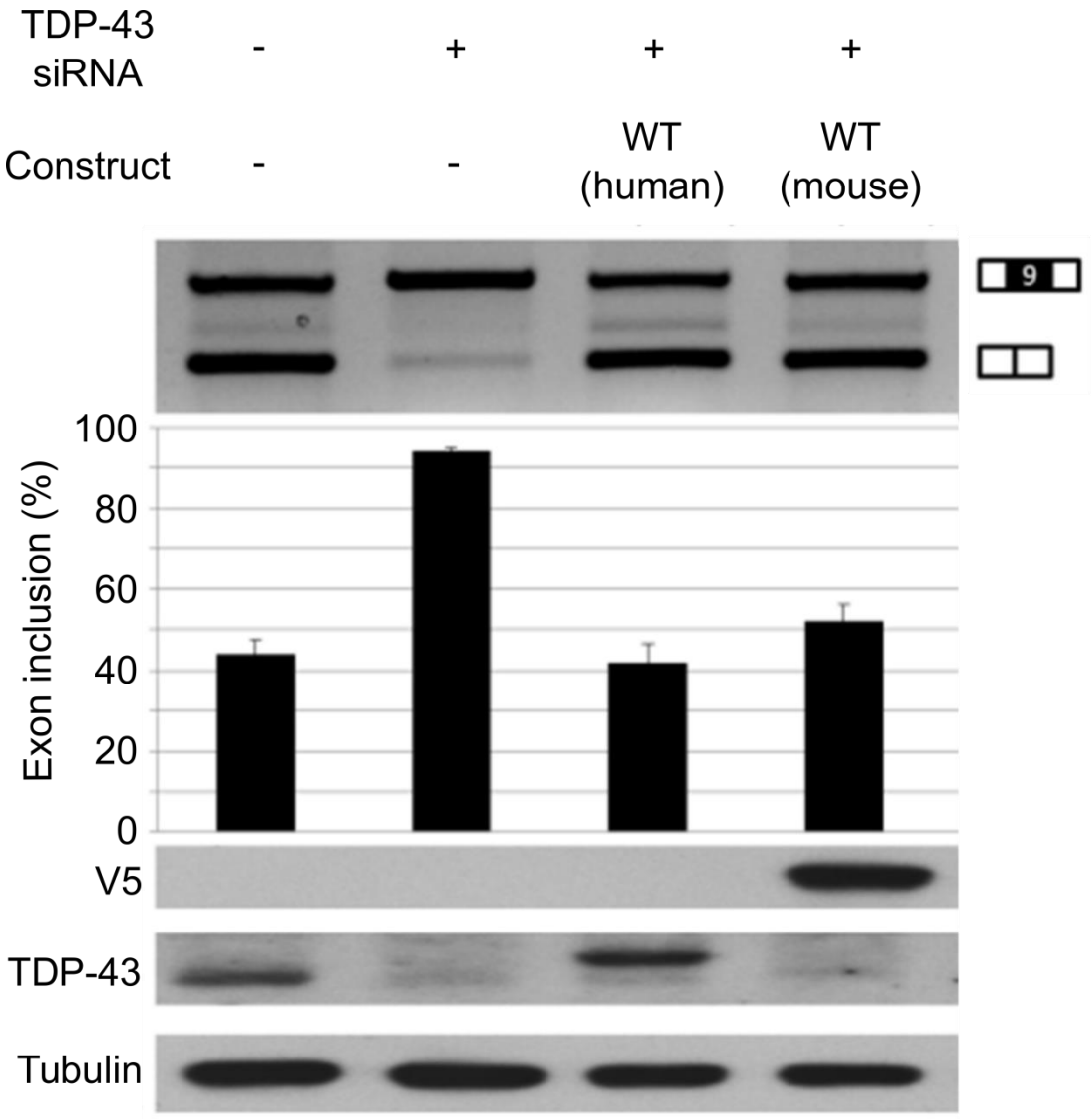


Figure 3.13. *CFTR* minigene assay showing RT-PCR results and % exon 9 inclusion in (left) wildtype TDP-43, (middle left) TDP-43 KD, (middle right) TDP-43 rescue with human TDP-43, (right) TDP-43 rescue with V5-tagged mouse TDP-43 construct in HeLa cells. Presence of TDP-43 targetting siRNA indicated by '+'. Exon 9 inclusion at approximately 50 % in wildtype conditions, substantially increased upon TDP-43 KD and rescured by both human and mouse TDP-43 expression. Minigene isoforms for exon 9 inclusion and skipping labelled. Cryptic splice site resulting in isoform between full exon 9 inclusion and skipping not included in measurements. Example bands for V5, TDP-43 and Tubulin labelled. Error bars display SEM (Ricketts, 2012).

Having characterised wildtype TDP-43 regulation of the *CFTR* minigene, and the effect of TDP-43 KD, the effect of the F210I mutation was next investigated by transfecting mutated TDP-43 into HeLa cells post-siRNA-mediated TDP-43 KD and measuring splicing rescue levels. The mutation was found to be unable to completely rescue the effects of TDP-43 KD, and cause significantly greater levels of exon inclusion in the *CFTR* minigene in comparison to equivalent levels of wildtype TDP-43 (Figure 3.14). This suggested a deficiency in ability of F210I TDP-43 to repress the splicing of the exon (Ricketts 2012).

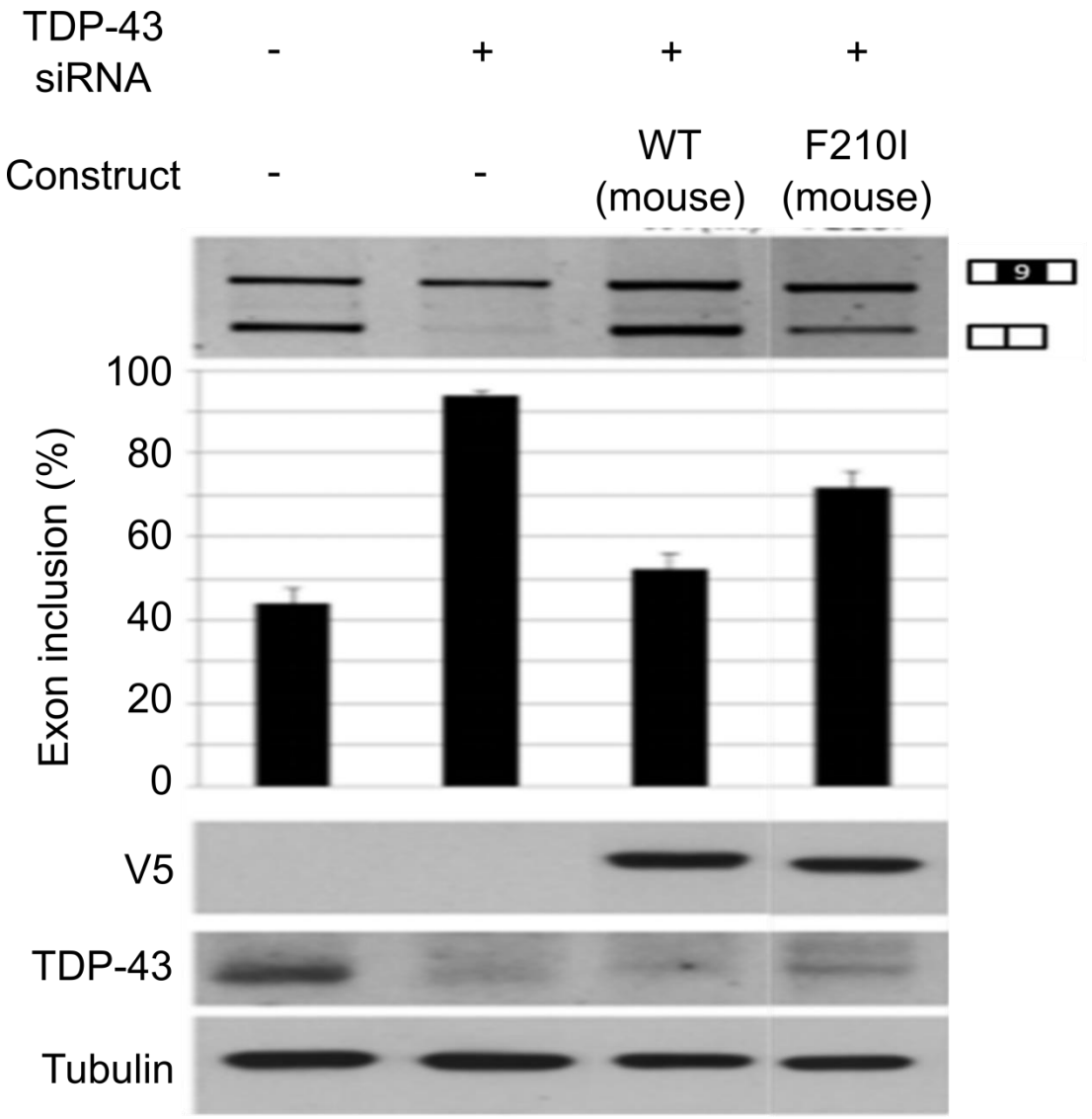


Figure 3.14. *CFTR* minigene assay showing RT-PCR results and % exon 9 inclusion in (left) wildtype TDP-43, (middle left) TDP-43 KD, (middle right) TDP-43 rescue with V5-tagged mouse TDP-43 construct, (right) TDP-43 rescue with V5-tagged mouse F210I-TDP-43 in HeLa cells. Presence of TDP-43 siRNA indicated by '+'. Exon 9 inclusion at approximately 50 % in wildtype conditions, substantially increased upon TDP-43 KD, rescured by wildtype mouse TDP-43 expression, but only partially rescued by F210I mutant TDP-43 expression. Minigene isoforms for exon 9 inclusion and skipping labelled. Example bands for V5, TDP-43 and Tubulin labelled. Error bars display SEM (Ricketts 2012).

There was significant reduction in F210I TDP-43 ability to repress splicing of *CFTR* exon 9 in comparison to wildtype in HeLa cells upon transfection of the mutated mouse construct. Following this, the splicing of the *CFTR* minigene was tested ex vivo in wildtype, F210I heterozygous or F210I homozygous MEFs derived from E14.5 mice. After electroporation of MEFs with the minigene construct, results matched the effects seen in HeLa cells. F210I mutation resulted in an increase in exon 9 inclusion in the *CFTR* minigene in comparison to wildtype TDP-43, and the effect was dose dependent (Figure 3.15). In homozygous F210I MEFs, exon inclusion levels were close to 100 %, indicating almost complete failure of normal TDP-43 function, exon repression.

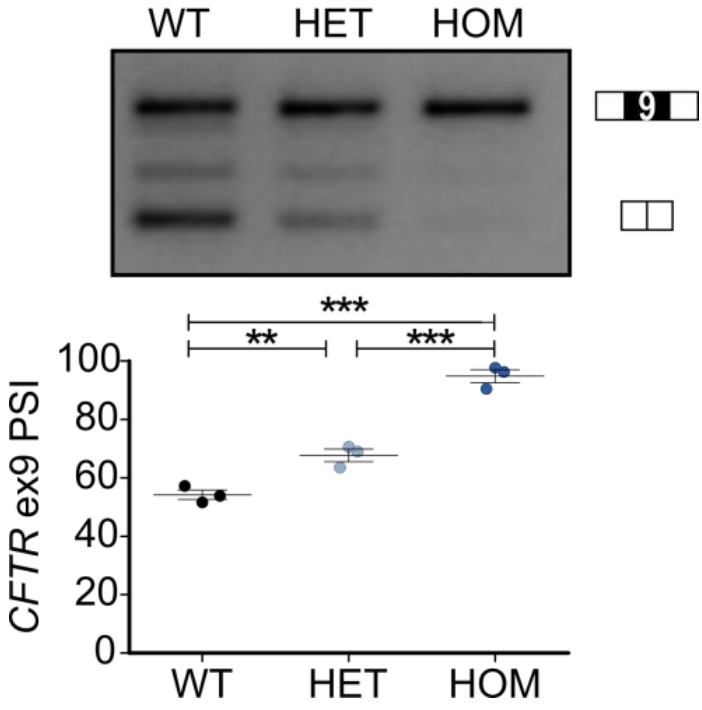


Figure 3.15. *CFTR* minigene assay showing RT-PCR results and % exon 9 inclusion in (left) wildtype TDP-43, (middle) heterozygous F210I TDP-43, (right) homozygous F210I TDP-43 in MEFs. Significant and dose dependent increase in exon 9 inclusion as a result of F210I mutation. Minigene isoforms for exon 9 inclusion and skipping labelled. Error bars show SEM. Results from Abraham Acevedo Arozena.

Having established the apparent loss of function in TDP-43 splicing activity on the *CFTR* exon 9 minigene, known targets of TDP-43 alternative splicing regulation were explored to elucidate whether the reduced splicing control occurred in endogenous transcript targets. RT-PCR was conducted for *Sort1* exon 18, an alternative splicing event in which TDP-43 functions to repress exon inclusion (Prudencio et al. 2012). In the F210I MEFs, an analogous, significant, dose-dependent reduction in exon repression in the *CFTR* exon 9 minigene was again observed (Figure 3.16). Comparable results were also obtained for other TDP-43-mediated alternative splicing events in *Eif4h* and *Dnajc5*. This effect was concordant with expected results in the context of loss of function in TDP-43 splicing.

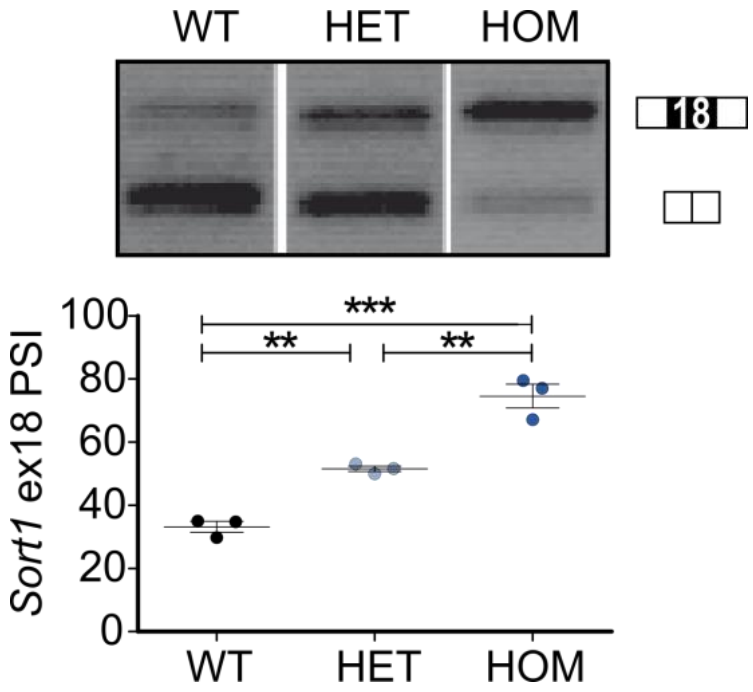


Figure 3.16. RT-PCR results and quantification for inclusion levels of *Sort1* exon 18 in wildtype, heterozygous F210I and homozygous F210I MEFs. Significant and dose dependent increase in exon 18 inclusion of *Sort1* as a result of F210I mutation. Isoforms for exon 18 inclusion and skipping labelled. Error bars show SEM. Results from Abraham Acevedo Arozena.

Alongside the loss of splicing regulation in TDP-43 targets, the F210I mutation also resulted in a change in RNA binding. An electrophoretic mobility shift assay (EMSA) was conducted with wildtype and F210I TDP-43 of increasing concentrations, and the binding of TDP-43 to oligonucleotides containing six UG repeats was measured, a sequence to which TDP-43 binds with high affinity. Intensity for TDP-43 binding to the target motif was severely reduced in the F210I mutant in comparison to wildtype at equivalent protein concentrations (Figure 3.17). This suggested a loss of TDP-43 binding ability to UG repeats, and, in conjunction with the loss of splicing activity, raised the possibility that the F210I mutation resulted in a hypomorphic allele and severe loss of function in TDP-43 RNA processing activity.

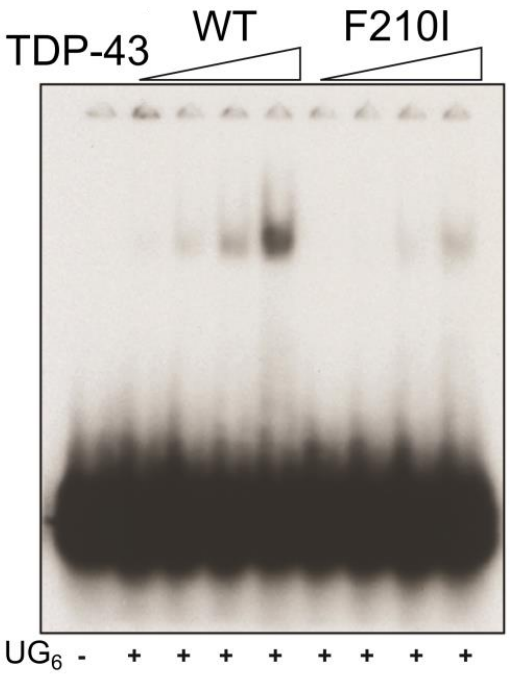


Figure 3.17. EMSA for increasing concentrations of wildtype and F210I TDP-43 binding to radiolabelled oligonucleotide sequences containing known TDP-43 target motifs of six UG repeats, indicated by '+'. Reduction in UG repeat binding as a result of F210I mutation. Results from Abraham Acevedo Arozena.

The F210I mutation disrupted the ability of TDP-43 to regulate aspects of RNA processing, leading to instances of molecular dysregulation. Alongside the changes at the RNA level, TDP-43 protein localisation in MEFs was also investigated. However, in this aspect there were no differences evident in the F210I heterozygous or homozygous models, with TDP-43 primarily nuclear-localised with no sign of cytoplasmic mislocalisation or aggregates.

The F210I mouse model was also re-derived following characterisation of the ENU mutagenesis-induced mutation, as described in 1.3.3 TDP-ALS models, in order to study the phenotype that would result from the effects the mutation. Mice were backcrossed from the founder C57BL/6J / DBA/2J mixed background to a pure C57BL/6J background (Ricketts 2012). As previously outlined, developmental knockout of TDP-43 is rapidly lethal in early embryonic life (Kraemer et al. 2010; Wu et al. 2009; Sephton et al. 2010). In keeping with this, the embryos homozygous for the F210I mutation were not viable, with the mutations proving perinatally lethal, suggesting that the F210I mutation may have considerable impact on development of the mouse embryo. This did, however, provide sufficient time to study the embryonic model for molecular and phenotypic changes, as well as obtain primary cells for culture, such as MEFs.

Prior to lethality, homozygous F210I embryos showed developmental delay in size and weight, as well as eye, bone and cleft palate development, according to investigation by Erwin Pauws (Figure 3.18). This further suggested a widespread impact of the dysfunctional F210I TDP-43 on overall development. Heterozygous mice were broadly healthy and survived until adulthood with no change in lifespan or phenotype, implicating the autoregulatory ability of TDP-43 in ameliorating the effects of a hypomorphic F210I allele.

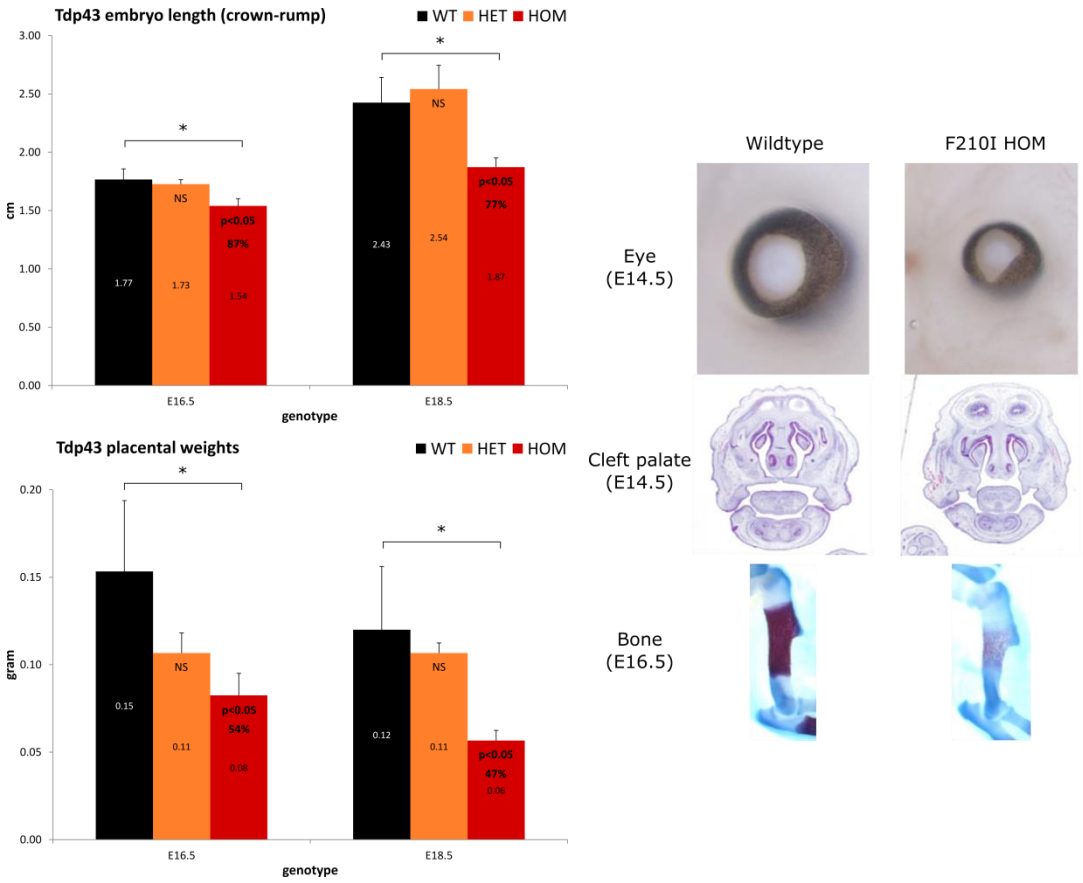


Figure 3.18. Phenotype in F210I embryonic mice. (Top left) Significant reduction in embryo size at E16.5 and E18.5 in homozygous F210I mice, (bottom left) significant reduction in placental weight at E16.5 and E18.5 in homozygous F210I mice, (right) developmental delay in the eye at E14.5, cleft palate at E14.5 and bone at E16.5, identified by Erwin Pauws. Error bars show SEM.

3.2 RESULTS AND DISCUSSION

3.2.1 RNA SEQUENCING OF F210I OVERVIEW

The initial low-throughput investigation into the consequences of the F210I mutation on TDP-43 function suggested effects analogous to TDP-43 loss of function, and dysregulation of downstream RNA targets such as *Sort1*. Alongside this, the severe developmental phenotype in the homozygous F210I models suggested a substantial

impact of the mutation in TDP-43 function. To further explore this possibility and gain a more complete understanding of the overall landscape of F210I effects on RNA processing, high throughput RNA sequencing studies were conducted on several F210I time points and tissue types. The data was processed with specific investigation into differential gene expression and alternative splicing changes in mind.

To establish the extent of RNA processing change as a result of the mutation, poly-A purified RNA sequencing was conducted on F210I mouse embryonic heads. Age E14.5 was chosen to allow study of homozygous as well as heterozygous models, enabling identification of dose-dependent effects. Sequencing was conducted as in 2.1.6 RNA sequencing, with short 35 bp paired-end reads at an approximate sequencing depth of 20 M unique reads per sample (n = 3). Differential gene expression and alternative splicing analyses of the homozygous E14.5 model in comparison to wildtype, as outlined in 2.1.7 Bioinformatics analysis, determined 1184 differentially expressed genes were present in this model, alongside 67 splicing changes. Therefore it was evident that the F210I mutation caused widespread dysregulation in gene expression and splicing control even at the early developmental stages, in accordance with the early phenotype development in this model.

In stark contrast to the homozygous model, the heterozygous F210I datasets suggested minimal differences in comparison to wildtype in alternative splicing. This was fitting with the absence of phenotype during development in heterozygous F210I models. Additionally, previous studies into mouse models hemizygous for *Tardbp* unanimously agree that there is minimal phenotype in comparison to wildtype controls at the early developmental stages (Wu et al. 2009; Kraemer et al. 2010; Sephton et al. 2010). Given the lack of impact from complete loss of function in one *Tardbp* allele, it follows that a single missense mutation in only one *Tardbp* allele would also have minor effects on mouse development. However, thorough sequencing and analysis of the adult

heterozygous model (n = 5), showed that signs of molecular dysregulation remained minimal at ages at which hemizygous *Tardbp* models develop motor performance deficits (Kraemer et al. 2010). This suggested that although the F210I mutation was capable of behaving as a hypomorphic allele and causing severe phenotype through loss of function effects, it was not equivalent to full knockout of a *Tardbp* allele.

Upon more thorough investigation into the splicing events detected in the RNA-Seq datasets, it was apparent that a substantial number of exons contained few aligned reads, due to the relatively low read depth per sample. This meant that any inferring differential exon usage in these exons would be very difficult to do with high confidence. The very short reads used to conduct this sequencing, 35 bp, also meant the chance of a read overlapping an exon-exon boundary, and therefore containing information on the splice junctions present, was minimal, making splicing detection through junction analysis difficult.

To enable a more thorough investigation into the evidently widespread alternative splicing changes present in the F210I model, another RNA-Seq dataset was generated from F210I E14.5 forebrain (n = 4). The sequencing for this dataset was conducted using much longer 150 bp paired-end reads, with a read depth of approximately 60 M per sample. Heterozygous models were omitted from this sequencing dataset given the absence of molecular changes seen previously. As with the prior sequencing dataset analysis, widespread RNA changes were again detected. Remarkably, 2,730 genes were judged to be differentially expressed, and alternative splicing analysis revealed 230 splicing events differentially used in the homozygous F210I embryonic mouse brain. The marked increase in differential gene expression and alternative splicing changes detected is likely due to the increased power of the follow-up RNA-Seq experiment. As a result, this was the primary RNA-Seq dataset used to investigate RNA processing dysfunction in the homozygous F210I mouse.

RNA-Seq dataset	Samples	Read length (bp)	Average reads per sample (~millions)	Differentially expressed genes	Alternatively spliced regions
F210I embryonic head	3 WT, 3 HOM	35 x2	20	1184	67
F210I embryonic brain	4 WT, 4 HOM	150 x2	60	2730	230
F210I adult spinal cord	5 WT, 7 HET	75 x2	45	16	2

Table 3.18. Overview of F210I RNA-Seq datasets and analysis as described in Chapter 2 – Materials and Methods.

3.2.2 F210I ICLIP OVERVIEW

RNA sequencing was conducted to understand the scope of RNA processing disruption as a result of TDP-43 loss of function mutation. However, the initial analysis of F210I-mutated TDP-43 interaction with target sequences through EMSA also revealed an indiscriminate reduction in RNA-binding. Therefore, in order to understand whether the loss of function effect was due to a change in transcripts targeted by F210I TDP-43, a change in target sequence from the high wildtype affinity to UG repeats, or a general reduction in RNA binding ability, we conducted individual nucleotide resolution crosslinking and immunoprecipitation (iCLIP) (König et al. 2010) on homozygous F210I E18.5 spinal cords. Libraries were successfully developed from two wildtype samples and two homozygous embryonic F210I forebrains. Alongside answering questions in general TDP-43-RNA interaction, the ability to detect TDP-43 binding sites on target transcripts to a single nucleotide resolution enabled us to implicate changes in direct TDP-43 binding in the context of patterns of RNA dysregulation.

A total of 3,097,210 crosslinked sites were detected in the wildtype embryonic forebrain samples, in comparison to 2,575,533 across the homozygous embryonic F210I mutants. Upon filtering of these peaks as suggested in the iCount pipeline, 4,479

genes were detected as targets in both embryonic wildtype replicates, and 4,950 genes were common in both embryonic homozygous F210I samples. Interestingly, 4,690 genes were shared targets in each of one wildtype and one homozygous F210I iCLIP dataset, suggesting substantial similarity in the two groups and no large scale change in gene targeting as a result of F210I mutation (Table 3.19).

Sample	Peaks	Peaks FDR < 0.05	Genes	Shared genes within genotype	Shared genes between genotypes
WT 1	1,782,003	224,668	5,521	4,479	4,690
WT 2	1,582,011	292,857	6,431		
HOM 1	1,658,355	368,525	7,227	4,950	
HOM 2	1,161,790	201,477	5,917		

Table 3.19. Overview of embryonic iCLIP datasets for wildtype and homozygous F210I forebrain. Analysis conducted as described in 2.2.11 iCLIP Bioinformatics. Shared genes within genotype represents number of genes targeted in both replicates of one genotype. Shared genes between genotypes represents number of genes targeted in at least one replicate of each genotype.

As previously outlined, TDP-43 has been shown to target UG repeat sequences with high affinity (Buratti & Baralle 2001). Wildtype TDP-43 binding to UG repeats was also established in our hands with the EMSA using UG repeat oligonucleotides. F210I TDP-43 also bound to the UG repeat oligonucleotides, but with reduced intensity. To check whether TDP-43 binding to UG repeats was enriched in our iCLIP dataset, we analysed the pentamer enrichment levels for every possible pentamer in each iCLIP replicate. In both wildtype and F210I datasets, the two replicates showed exceptionally similar enrichment levels for each pentamer, with Pearson correlation coefficients > 0.95 (Figure 3.19). Given the similarity between replicates, downstream analysis was conducted on grouped replicates for each dataset.

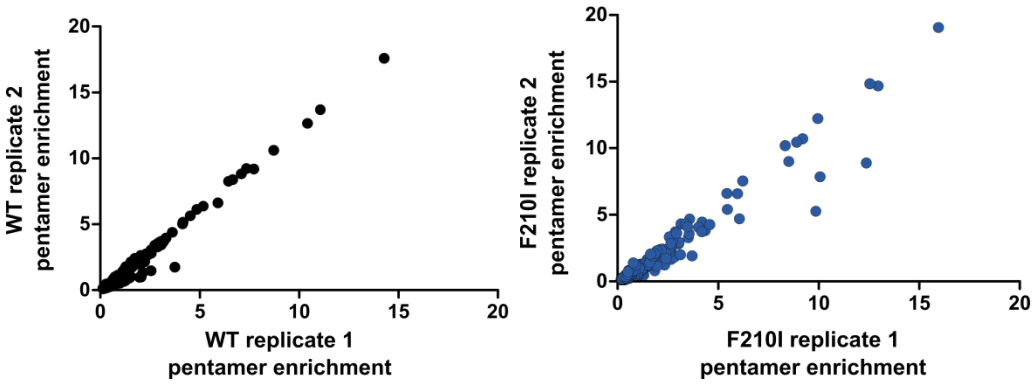


Figure 3.19. Scatter plot of enrichment for every pentamer in replicate iCLIP datasets for wildtype (left) and homozygous F210I E18.5 forebrain (right) in comparison to mean random predicted enrichment. Each dot represents one of the 1024 possible pentamer combinations. Strong correlation in pentamer enrichment in both wildtype and F210I datasets.

A subset of pentamers showed much greater levels of enrichment in both wildtype and F210I datasets. We next analysed whether in our iCLIP data, TDP-43 preferentially directly bound to UG repeats as had been previously reported. We determined that the two most enriched pentamers were the two possible variants of UG repeats, or TG repeats as the case would be for cDNA – TGTGT and GTGTG. Also heavily enriched were all pentamers containing the 4 nt of TGTG or GTGT, and additionally many pentamers with a 3 nt TGT or GTG sequence (Figure 3.20). This result tallied well with previous reports that TDP-43 targets UG regions, and has higher affinity for longer length UG repeats (Ayala et al. 2005). Having observed the reduced binding of F210I TDP-43 for UG repeats in the EMSA, we also sought to determine whether the mutation altered TDP-43 targeting of UG repeats in favour of other sequences. We found that in the F210I iCLIP data there was enrichment for UG repeats, with the TGTGT and GTGTG pentamers being most heavily enriched. Furthermore, the enrichment levels for each pentamer were very similar between the wildtype and F210I datasets, with a Pearson correlation coefficient of 0.95 (Figure 3.20).

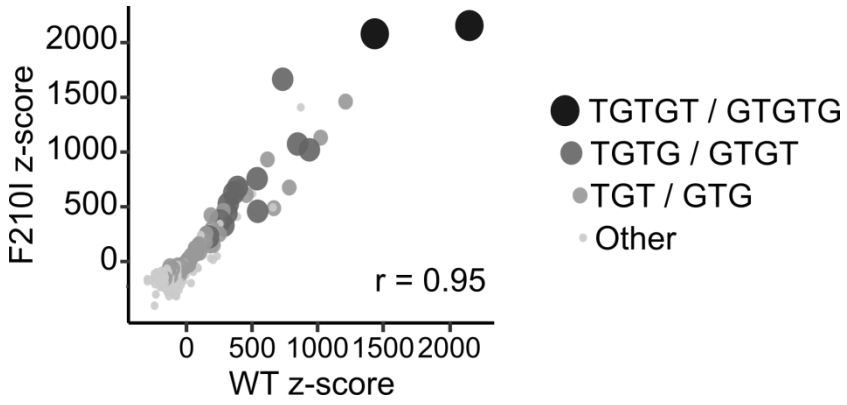


Figure 3.20. Pentamer enrichment for WT and homozygous F210I iCLIP datasets. Z-score measures enrichment of pentamer above mean predicted random enrichment. Pentamer sequences containing TG repeats highlighted by colour and size, with these pentamers enriched. Strong pentamer enrichment correlation with Pearson coefficient of 0.95.

This suggested that our iCLIP data not only agreed with previous findings, but additionally that the F210I mutation had no substantial effect on TDP-43 targeting of UG repeats. Instead, the reduced binding to UG repeat nucleotides seen in the EMSA was likely due to a global reduction in TDP-43 binding ability to RNA due to RRM2 mutation. This effect was further observed during iCLIP library preparation, with F210I TDP-43 showing reduced radio-labelled RNA intensity in comparison to wildtype (Figure 3.21).

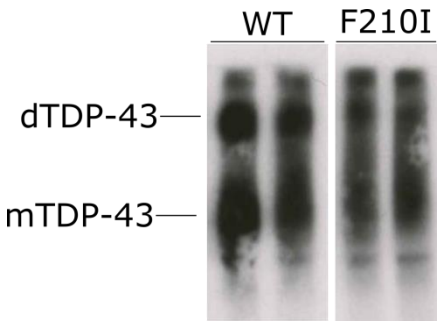


Figure 3.21. Autoradiograph of TDP-43-RNA complexes for two wildtype and two homozygous F210I samples. "dTDP-43" represents dimerised TDP-43, "mTDP-

43" indicates monomeric TDP-43. Lower RNA signal intensity in F210I samples in comparison to wildtype at these markers, with protein levels having been normalised.

Alongside TDP-43 target motifs, another aspect of global TDP-43 binding site analysis of interest was the region binding preferences of TDP-43 on target transcripts. TDP-43 has previously been reported to bind abundantly in intronic regions of protein coding genes, with particularly noteworthy binding in deep intronic regions (Tollervey et al. 2011). We determined the comparative frequency of significant TDP-43 binding peaks in all target transcripts, split into sub-regions (5'UTR, 3'UTR, open reading frame [ORF], non-coding RNA [ncRNA], intron, intergenic and antisense). We found that in wildtype embryonic mouse forebrain, TDP-43 overwhelmingly bound to introns of target transcripts (79 %). 10 % of binding was restricted to intergenic regions, with a minority of peaks assigned to the remaining sub-regions (Figure 3.22). Previous CLIP data suggested TDP-43 binding to introns constituted 58 % of all binding sites in mouse brain, and instead 74 % in cell lines (Tollervey et al. 2011), but this discrepancy may be due to the study being conducted on adult mouse brain as opposed to our embryonic forebrain model.

Having determined the transcript region binding profile of wildtype TDP-43 in embryonic mouse forebrain, we next analysed whether the F210I mutation would have any effect on target regions. Once again, we found very little difference between the RRM mutant and wildtype TDP-43. Intron binding was again the far and away the most common target region (78 %) and intergenic regions the second (11 %) (Figure 3.22). No sub-region showed a greater than 1 % difference in proportion of TDP-43 binding. This, alongside the maintained targeting of UG repeat regions, suggested that F210I TDP-43 did not affect targeting of specific transcript regions or sequences.

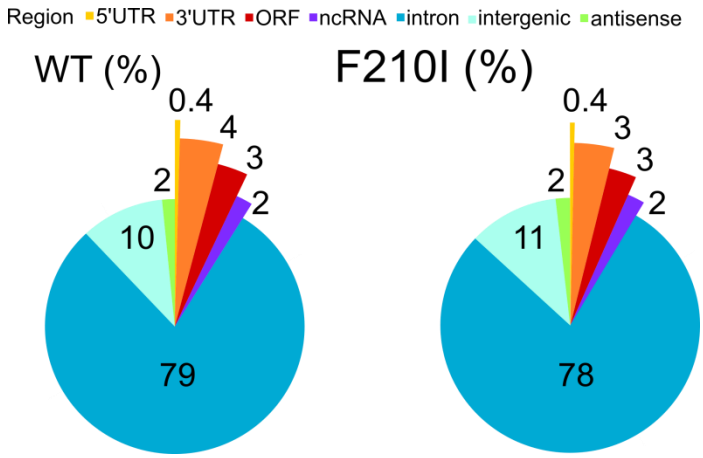


Figure 3.22. Percentage of wildtype and homozygous F210I iCLIP clusters which mapped to each sub-region of transcripts. Similar proportions of sub-region binding in F210I in comparison to wildtype.

3.2.3 F210I TDP-43 AUTOREGULATION

As previously outlined in 1.1.2 *TARDBP* isoforms and autoregulation, TDP-43 levels are tightly autoregulated through a feedback mechanism involving the alternative splicing and polyadenylation of *Tardbp* transcripts as a consequence of TDP-43 binding.

First, we confirmed that TDP-43 directly bound to *Tardbp* transcripts in our iCLIP datasets. In the wildtype mouse, TDP-43 peaks were detected throughout the entirety of intron 7 of the *Tardbp* transcript, which is the intron containing polyadenylation site pA_1 and the region previously described to show high TDP-43 CLIP cluster density, the TDPBR (Figure 3.23). Alongside substantial binding in intron 7, peaks were also discernible in introns 3 and 6, but with far fewer individual total binding sites. This binding pattern was consistent with previous reports highlighting the TDPBR as the primary region for TDP-43 binding to *Tardbp* transcripts (Ayala et al. 2011; Polymenidou et al. 2011).

We also sought to determine whether the F210I mutation had any effect on TDP-43 binding to its own transcripts. The iCLIP data from the homozygous F210I model

showed strikingly similar binding patterns for the mutated TDP-43 protein in comparison to the wildtype. Once again, the majority of peaks were aligned to intron 7 in *Tardbp*, across the TDPBR (Figure 3.23). Binding sites were again also detected in introns 3 and 6 to a lesser degree. This suggested that the F210I mutation had no noticeable effect on TDP-43 binding regions in *Tardbp* transcripts. This result was also fitting with global iCLIP analysis that suggested TDP-43 target motifs, sequences, and target regions were not altered in the F210I model.

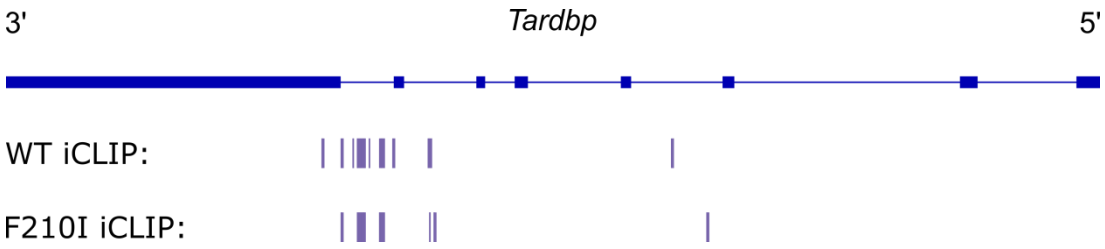


Figure 3.23. *Tardbp* transcript with iCLIP peak sites denoting TDP-43 binding in wildtype and homozygous F210I iCLIP datasets. Similar binding sites in F210I iCLIP dataset in comparison to wildtype.

Alongside the UG repeat oligonucleotide EMSA experiment on wildtype and F210I TDP-43, another EMSA had previously been conducted targeting sequences based off of the 34 nt TDPBR. Results from this EMSA showed similarity to the UG repeat EMSA, with F210I TDP-43 showing a reduction in binding ability to the TDPBR sequence oligonucleotides in comparison to wildtype (Figure 3.24). In combination with our iCLIP results, this experiment adds further evidence to the change in RNA binding in F210I TDP-43 being as a result of a general reduction in binding ability as opposed to specific differences in TDP-43 targeting.

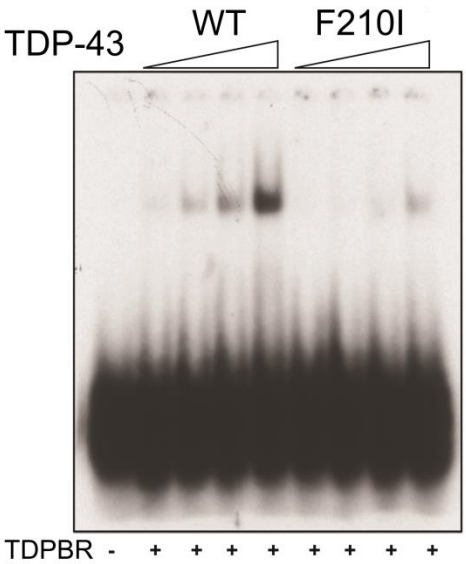


Figure 3.24. EMSA for increasing concentrations of wildtype and F210I TDP-43 binding to radiolabelled oligonucleotide sequences derived from the TDPBR of *Tardbp* transcripts, indicated by '+'. Lower signal intensity for TDPBR sequence binding in F210I TDP-43 in comparison to wildtype at equivalent protein concentrations (Ricketts 2012).

We established that TDP-43 targeting of the TDPBR was not substantially altered by the F210I mutation. However, as iCLIP is not a quantitative technique, we could not comment on the relative levels of TDP-43 binding in the wildtype and mutant models at this region. To investigate the consequences of a possible change in the autoregulation mechanism, we checked for changes in the level of *Tardbp* expression in the F210I model using the differential gene expression output of our RNA-Seq analysis. *Tardbp* was found to be significantly upregulated in the homozygous F210I embryonic forebrain dataset, with a $\log_2$ fold change of +0.57 and an adjusted p-value of $8.53e^{-19}$. This upregulation was in line with the expected outcome of the F210I mutation given its loss of function effects. An increase in *Tardbp* transcripts would enable a greater number of transcripts to be translated, leading to higher TDP-43 levels. This would in turn ideally compensate to some extent for the reduction in RNA binding of the F210I-

mutant TDP-43 and the resulting RNA processing dysregulation. Importantly, increase in total *TARDBP* transcript levels is seen in post-mortem ALS spinal cord motor neurons (Mishra et al. 2007; Gitcho et al. 2009; Swarup, Phaneuf, Dupré, et al. 2011; Koyama et al. 2016).

However, the increase in *Tardbp* translation would be dependent on the increase in *Tardbp* transcripts with the pA_1 polyadenylation site; a mere increase in transcripts with pA_2 and pA_4 may be detectable in overall transcript expression levels but would not lead to an increase in protein. To understand whether the TDP-43 autoregulation mechanism itself was modulated to compensate for the F210I loss of function, the alternative splicing analysis from the RNA-Seq output was further investigated, with a focus around intron 7 splicing, known to be the key feature of the autoregulatory mechanism.

As outlined in 1.1.2 *TARDBP* isoforms and autoregulation, splicing of intron 7 leads to usage of the pA_2 polyadenylation site and subsequently the transcript is not translated. Retention of the intron can lead to usage of pA_1 and formation of an isoform which can be translated to the TDP-43 protein, but it can also lead to the use of pA_4 and an isoform that is retained in the intron and therefore not translated. When looking at levels of intron 7 retention, wildtype Percentage Spliced In (PSI) levels were on average at 95.7 %, meaning 4.3 % of sequenced transcripts contained spliced intron 7 and would not be translated. In comparison, in the F210I model, intron 7 PSI levels for intron 7 retention amounted to 99.3 %, suggesting that just 0.7 % of sequenced transcripts would result in pA_2 usage. This ΔPSI of 3.6 % would lead to changes in a substantial number of transcripts given the high expression levels of *Tardbp*, and as a result the difference was highly significant (adjusted p-value = $2.78e^{-15}$). Therefore, it is likely that a change in the autoregulation mechanism in the form of reduced intron 7 splicing caused the increase in *Tardbp* transcript levels. If indeed the sequenced

transcripts are reflective of the spliced isoforms present prior to any degradation, then this would suggest that in the F210I model the autoregulatory response has shifted to produce almost complete levels of stable *Tardbp* transcript production. This would further imply that the ability of this mechanism to stabilise TDP-43 levels is close to its maximum capacity, and would therefore not be able to correct the molecular dysregulation that persists in the F210I model despite the autoregulation. However, other studies have reported far greater levels of splicing in *Tardbp* transcript pre-mRNA (Sugai et al. 2018), suggesting the sequenced transcript isoforms do not represent an accurate snapshot of the true relative levels of each.

It is important to note that although a proportion of transcripts were detected with intron 7 splicing, previous investigation into the effect of this alternative splicing has suggested that the resulting transcript would be unstable and lost through NMD, or another degradation mechanism (Eréndira Avendaño-Vázquez et al. 2012; Koyama et al. 2016). If this is the case, then the majority of transcripts which displayed pA_2 usage would have very short half-lives, and the proportion of sequenced transcripts which showed intron 7 splicing would be severely under-representative of the total number of transcripts undergoing this splicing. This means that although the difference in splicing in the F210I model is substantial and significant, the true change may be much greater, but would not be detectable without incorporating a technique enabling sequencing of unstable transcripts, which was not conducted here.

However, while pA_2-utilising transcripts are not stable, another transcript isoform which is not translated, containing pA_4, may be. Current studies suggest that this transcript is not translated either, but is not even exported into the cytoplasm and is instead retained in the nucleus (Eréndira Avendaño-Vázquez et al. 2012; Koyama et al. 2016). In the nucleus, it is not apparent that the transcript is unstable or degraded in any manner. Therefore, in our RNA-Seq datasets, it is possible to detect this isoform

and compare levels with total transcript levels to gain and understanding of the number of *Tardbp* pA_4 isoforms which are retained in the nucleus. As both pA_1 and pA_4 result from the same pattern of splicing, this information cannot be obtained from the splicing output of the RNA-Seq analysis. However, by comparing the total number of reads in the *Tardbp* 3'UTR, it is evident that there is also a substantial reduction in reads aligning to the long transcript isoform, which would represent pA_4 usage and therefore no translation (Figure 3.25). This would suggest a greater proportion of transcripts display pA_1 usage. As a result there would an increase in proportion of transcripts being translated, in line with the expression and splicing data and the expected outcome given the apparent loss of function effects of the F210I mutation.

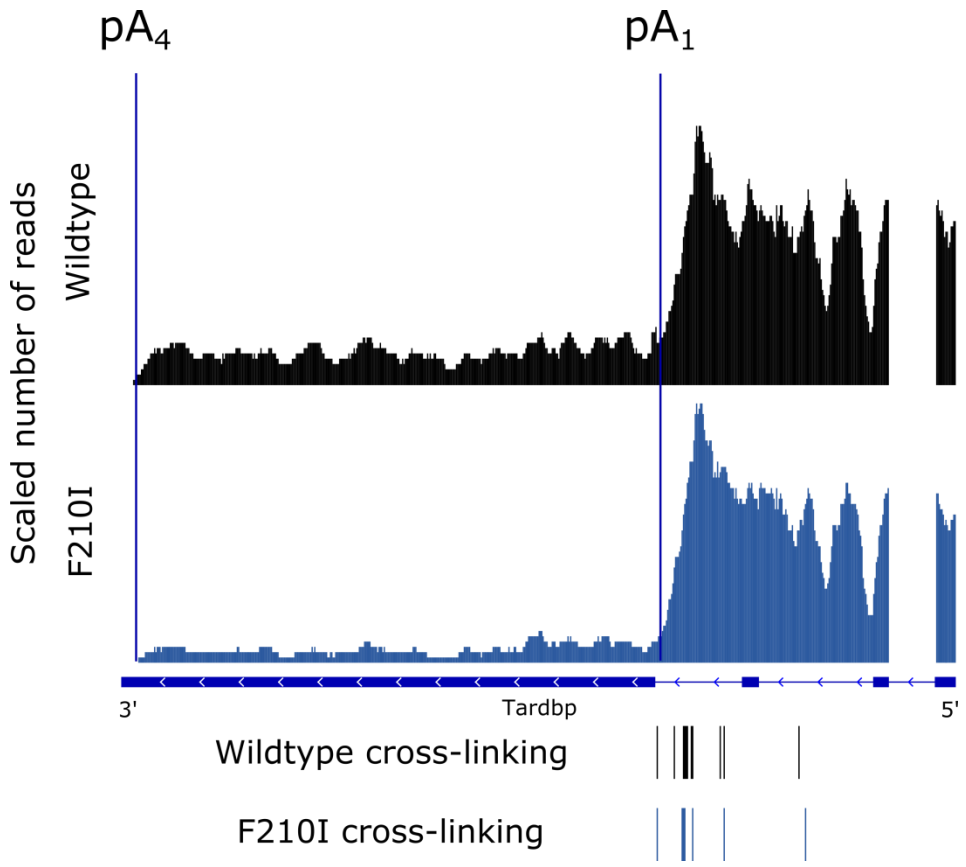


Figure 3.25. Polyadenylation sites pA_1 and pA_4 annotated. Summed read count tracks for the 3'UTR of *Tardbp* transcripts in wildtype and homozygous F210I RNA-Seq datasets. Reduction in proportional read counts between pA_1 and pA_4 in

homozygous F210I RNA-Seq. Below are the iCLIP peaks denoting TDP-43 binding sites in wildtype and homozygous F210I iCLIP datasets, showing similarity in TDP-43 binding.

The heterozygous F210I model showed no substantial molecular dysregulation, and it seemed likely that this was due to the autoregulation of TDP-43 producing greater levels of the wildtype transcripts and protein. However, in the low-depth RNA-Seq dataset, the heterozygous models showed no signs of this autoregulation in action. There was no detectable difference in intron 7 splicing, *Tardbp* 3'UTR long isoform usage, or overall *Tardbp* expression levels. The reason for the apparent lack of autoregulation is unclear, but the results in this case come with the notable caveat of low-depth, short read sequencing. In the adult heterozygous spinal cord dataset, again intron 7 splicing differences were not present, and long 3'UTR isoform usage was difficult to detect. However, unlike the embryonic heterozygous mouse, in the adult models there was a significant increase in overall *Tardbp* expression levels (log_2 fold change +0.26, adjusted p-value $6.57e^{-3}$). It may be that the F210I mutation is not so severe as to warrant exceptional changes in autoregulation, and the subtle differences may not have been detected in the embryonic heterozygous dataset. In conjunction with effects of aging, and as TDP-43 levels naturally lower, the autoregulation may have become stronger, hence the significant increase in transcript expression in adult models.

At the RNA level, the TDP-43 autoregulation mechanism shift as a result of the homozygous F210I loss of function mutation behaved exactly as expected, with a reduction in unstable and retained isoforms, a reduction in non-productive splicing, and an increase in overall *Tardbp* expression, all in conjunction with no obvious change in TDP-43 binding at the TDPBR. To examine whether the effect of this autoregulation was seen at the protein level, western blots for wildtype and F210I mutant TDP-43

protein were conducted. It would not be possible to obtain sufficient protein from embryonic forebrain, so protein levels were previously tested in heterozygous and homozygous F210I MEFs by Tom Ricketts at MRC Harwell. In contrast to the expected increase in protein levels given the autoregulation behaviour seen, significant reduction in TDP-43 protein levels in the homozygous MEFs was detected in comparison to wildtype, and also a trend towards lower TDP-43 levels in the heterozygous MEFs (Figure 3.26), although results varied by experiment (Ricketts 2012). There would be understandable differences in the brain of the F210I mouse model in comparison to MEF cell tissue, but this work provides evidence that a reduction in TDP-43 levels does occur, alongside TDP-43 autoregulation, in the context of the F210I model. TDP-43 protein levels in young adult F210I heterozygous mice were also examined by western blot, and here a non-significant trend towards TDP-43 downregulation (n = 3) was found (Figure 3.26), again with variable experimental results (Ricketts 2012). Although the adult F210I model did not show significant differences in *Tardbp* isoforms, upregulated *Tardbp* expression was identified, and so the trend towards TDP-43 protein downregulation is unexpected and surprising.

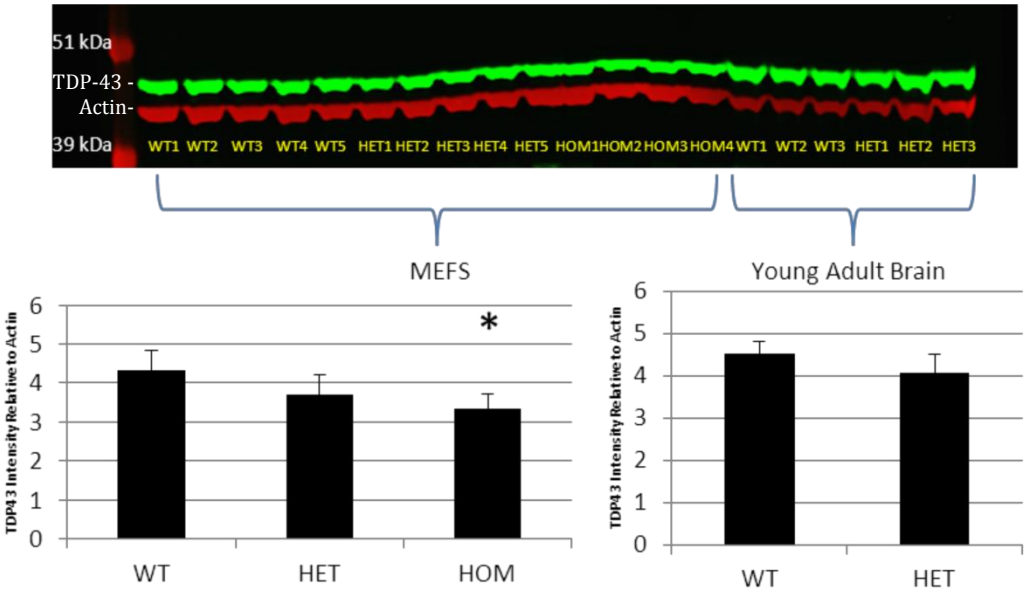


Figure 3.26. Western blot gel and quantification for TDP-43 in MEFs and young adult F210I mouse brain showing variable levels of TDP-43 in F210I models but significantly reduced in MEFs and trending towards a reduction in adult brain (Ricketts 2012). TDP-43 staining indicated in green, Actin in red. Quantification measured by signal intensity relative to Actin.

Although the autoregulation change in homozygous F210I models tallies with the expected correction at the RNA level as a result of the loss of function effects of the mutation, the clear nature of the shift may come as somewhat of a surprise given previous study of similar mutations. Ayala et al. (2011) investigated the effects of mutations in the different domains of TDP-43 on the ability of the protein to autoregulate. To the RRM2 domain, in which the F210I mutation lies, they introduced F229L and F231L mutations in conjunction. In further tests, they separately introduced F147L and F149L mutations to the RRM1 domain. They transfected these mutated and tagged cDNA constructs into HEK293 cells, and measured the levels of endogenous TDP-43 24 hrs and 72 hrs after transfection. They found that mutation of the two phenylalanine residues in RRM1 completely prevented TDP-43 autoregulation, with no

change in endogenous TDP-43 protein levels even 72 hrs after overexpression was caused with the RRM1 mutant protein.

In contrast, the results for RRM2 mutant expression closely resembled overexpression of the wildtype TDP-43, in that there was a strong autoregulatory response and, as the cDNA constructs could not be autoregulated, by 72 hrs endogenous TDP-43 levels had almost disappeared. Northern blotting for TDP-43 isoforms supported these findings, showing that the RRM1 mutations prevented the increase of non-productive and unstable isoforms, the RRM2 mutations did not, and allowed the autoregulatory mechanism to produce greater levels of these transcripts as was the case with wildtypes. These results suggested that phenylalanine mutations in the RRM2 domain did not cause any change in TDP-43 autoregulation, in contrast to mutation of RRM1. Combining this result with the possibility that the F210I mutation resides in a region of the RRM2 unlikely to be involved in direct RNA binding (Ramos, 2015, personal communication), it highlights the strong nature of the loss of function effects caused by the F210I mutation. Importantly, it also emphasises that mutations in RRM2 can too lead to changes in *Tardbp* autoregulation.

3.2.4 DIFFERENTIAL GENE EXPRESSION IN F210I

As outlined in the overview of RNA-Seq results for the F210I models, substantial differential gene expression was detected in the homozygous embryonic forebrain of these mice (2,730 genes with adjusted p-value < 0.1) (Figure 3.27), whereas heterozygous adults showed minimal significant changes (16 genes). Approximately equal instances of significant gene upregulation (1,335) and downregulation (1,395) were detected in the homozygous model. Similarly large scale RNA dysregulation has been previously reported upon TDP-43 loss of function (Polymenidou et al. 2011), highlighting the importance of TDP-43 in expression control of numerous genes. The

substantial dysregulation also caused by the F210I change, alongside the previously identified examples of loss of function effects in splicing and RNA binding as a result, further suggests that the F210I is a powerful hypomorphic mutation.

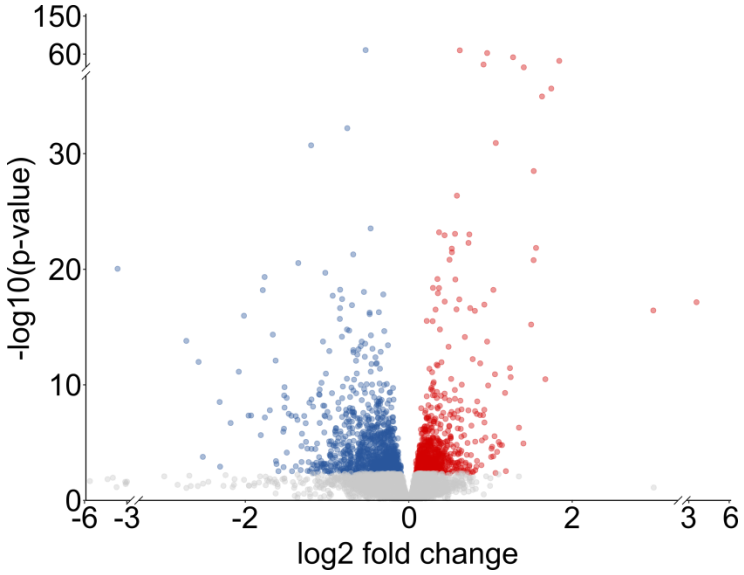


Figure 3.27. Volcano plot for differential gene expression changes in homozygous F210I forebrain vs wildtype RNA-Seq dataset. Each dot represents one gene. Substantial significant changes in gene expression levels. Non-significant changes in grey, significant upregulation in red, significant downregulation in blue. Axes scaled at extremes for convenient visualisation.

Given that approximately 13,500 genes contained reads in every sample of wildtype and F210I embryonic forebrain RNA-Seq, this would suggest about 20 % of the consistently expressed genes in this dataset showed dysregulation. Although the sheer number of gene expression changes may seem unreasonable, it is plausible due to the wide influence of TDP-43 on processing of the transcriptome. It is estimated from CLIP studies that TDP-43 binds to approximately 30 % of the entire mouse transcriptome (Polymenidou et al. 2011). To further understand the sizeable number of gene expression changes and identify specific pathways that may have been affected as a

result, gene ontology (GO) term enrichment analysis was performed on the significant changes, as described in 2.1.8 Gene ontology analysis.

A total of 36 pathways and processes were found to be significantly affected by the gene expression changes. The majority of these were GO terms, with a particular focus on developmental pathways (anatomical structure development, system development, multicellular organism development and developmental process). Dysregulation in these pathways may go some way towards explaining the array of developmental problems in the early embryonic stages of the F210I homozygous mice, with specific deficits in structural development of the eye, bone and cleft palate. Also highlighted were pathways involved in ion channelling and transport (ion gated channel activity, passive transmembrane transporter activity, channel activity, ion channel activity, metal ion transmembrane transporter activity, transmembrane signalling receptor activity, cellular potassium ion transport, potassium ion transmembrane transport, signalling receptor activity, ion channel complex, cation channel activity, transmembrane transporter complex, extracellular ligand-gated ion channel activity, postsynapse). Unlike the developmental phenotype, no obvious link between the dysregulation of these pathways and the F210I phenotype has been identified, and given the number and specificity of these terms, it would be an avenue of study worth pursuing in the future. A connection between ion channel deficits and TDP-43 loss of function would be particularly interesting if it affected neuronal activity, which many of these terms could suggest.

Intriguingly, alongside the variety of GO terms, one significantly affected pathway was identified from each of the KEGG and Reactome pathway databases. From the Reactome database, the term ‘Neuronal system’, was found to be significantly affected, furthering the suggestion that although a neuronal phenotype was not detected in the embryonic homozygous F210I mice, underlying neuronal changes may have occurred.

In addition to this is the term 'Glutamatergic synapse' that is highlighted from the KEGG database. This term neatly links the ion transport pathways affected in the GO terms to the neuronal system pathway affected in the Reactome, given the long-established reliance on glutamate transmission and recognition for motor neuron excitatory signalling (Baughman & Gilbert 1981). Glutamate signalling dysfunction has also been heavily implicated in ALS, with the prolonged excitation and slow clearance of the excitatory neurotransmitter proving toxic to motor neurons in ALS (Rothstein et al. 1992; Urushitani et al. 1998; Kruman et al. 1999). The importance of this pathway to ALS is emphasised by the fact that for a long time Riluzole, a drug that modulates glutamate levels, was the only FDA-approved drug for treatment of ALS (Bensimon et al. 1994; Lacomblez et al. 1996). Glutamatergic phenotype was not studied in the F210I model, but the dysregulation of genes key to glutamatergic neurons adds weight to the requirement of further study of the F210I nervous system.

3.2.5 DOWNREGULATION OF LONG GENES

The dysregulation of such a large number of genes enabled detection of underlying patterns by which the genes were linked. Alongside the GO term analysis, this was also investigated in terms of the physical properties of the genes, not only the functional relevance. One pattern to emerge was the notable downregulation of the longest genes in particular. Analysis by Jack Humphrey determined that when ranked by level of dysregulation, with the most downregulated genes and most upregulated genes at opposite ends of the spectrum, the average intron length of the most downregulated genes was substantially longer than the average intron length of unchanged or upregulated genes (Figure 3.28). Analysis of gene regulation as a result of antisense oligonucleotide (ASO) – mediated KD of TDP-43 in adult mouse brain by Polymenidou et al. (2011) showed markedly similar results (Figure 3.28), once again highlighting the

similarity in effects of the F210I mutation and overall TDP-43 loss of function. The downregulation of long genes was not seen in adult heterozygous F210I models.

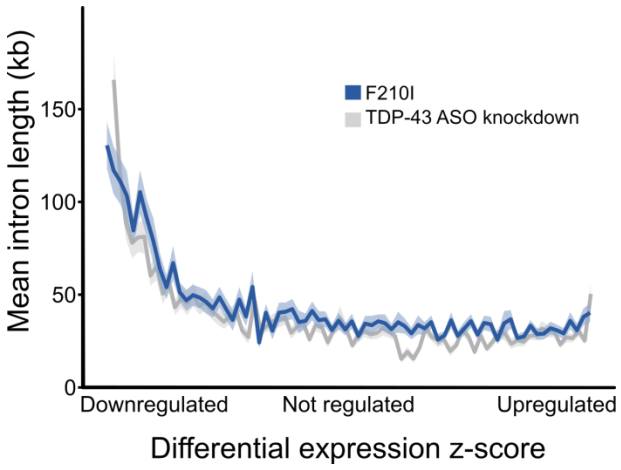


Figure 3.28. Comparison of mean intron lengths in genes binned by level of dysregulation into equal groups in the F210I embryonic forebrain vs wildtype RNA-Seq dataset performed by Jack Humphrey. Mean intron length substantially higher in downregulated genes in the F210I dataset in comparison to wildtype. Similar analysis and results for TDP-43 ASO-mediated KD in grey (Polymenidou et al. 2011).

To confirm the downregulation of long genes, a subset of these were selected for validation by qPCR, as described in 2.1.5 Quantitative PCR. The expression levels of 8 long genes was shown to be significantly and substantially downregulated in the embryonic F210I homozygous forebrain in comparison to wildtype tissue (Figure 3.29).

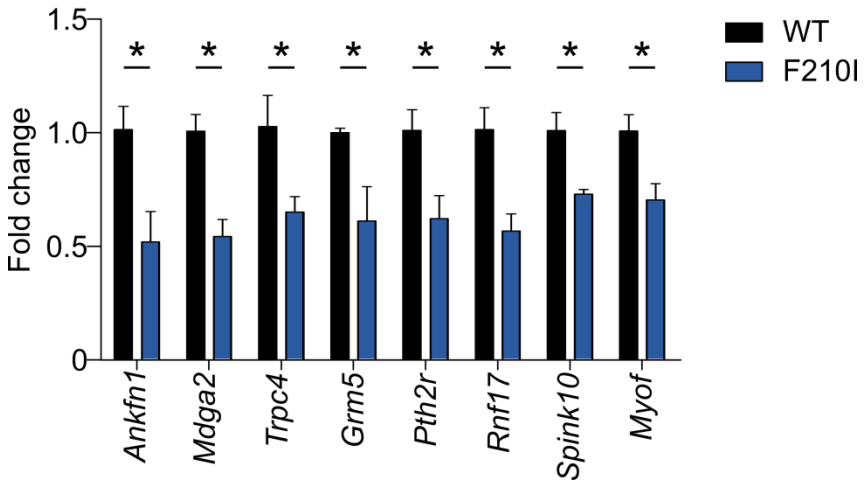


Figure 3.29. Fold changes in 8 long genes in homozygous embryonic F210I samples vs wildtype, each showing a significant, validated downregulation in the F210I model as measured by qPCR.

Comparison of the magnitude of expression change in these genes showed that in 5 of the 8 genes showed very similar levels of change, whereas the other 3 showed greater downregulation in the qPCR validation. Given the technical stringency of RNA-Seq, it is likely that in some cases qPCR may overestimate the true level of expression change (Figure 3.30).

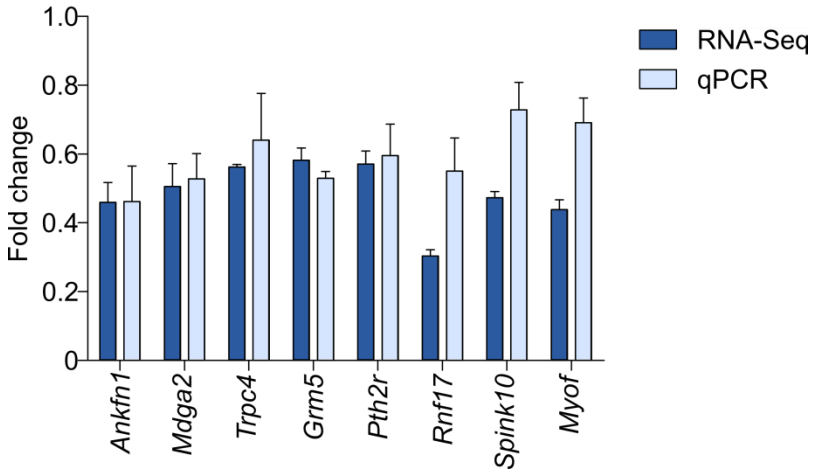


Figure 3.30. Comparison of size of expression change in 8 long genes in homozygous embryonic F210I samples vs wildtype as measured by RNA-Seq and

qPCR. Magnitude of downregulation broadly equivalent in the two techniques, with qPCR in some cases resulting in overestimation of change.

Study of TDP-43 KD resulting in long gene downregulation had also previously indicated that the downregulated long genes exhibited much greater levels of TDP-43 binding in comparison to unchanged or upregulated genes. However, this was largely due to the increased length of the genes themselves, and the level of TDP-43 binding per base of transcript was not found to be substantially different in the downregulated genes compared to unchanged or upregulated (Polymenidou et al. 2011).

To understand how TDP-43 binding interacted with the differently regulated genes in the F210I dataset, we examined the average iCLIP cluster coverage for TDP-43 binding in all transcripts, split by dysregulation and also gene length. Analysis by Jack Humphrey on genes in which the longest intron exceeded 100 kb in comparison to all other transcripts revealed that, in transcripts in which expression levels were unchanged, TDP-43 binding density was also unchanged (Figure 3.31). However, when these long transcripts were downregulated, they exhibited significantly higher levels of TDP-43 binding compared to both expression-unchanged long genes, and also compared to downregulated non-long genes. Intriguingly, it was also determined that the few long genes which displayed upregulation in the F210I model showed reduced TDP-43 binding density in comparison to unchanged long genes and also in comparison to upregulated non-long genes. This latter finding was not previously reported in the adult mouse TDP-43 KD study, and implies that as increased TDP-43 binding may lead to repressed expression of long genes, an absence of TDP-43 binding can result in increased expression of some long genes too. The reason for this discrepancy in the two TDP-43 loss of function datasets is not clear, and may be related to the loss of full protein in TDP-43 KD as opposed to the mutation of the RBD of TDP-43, with the LCD unaffected, in the F210I mutant model.

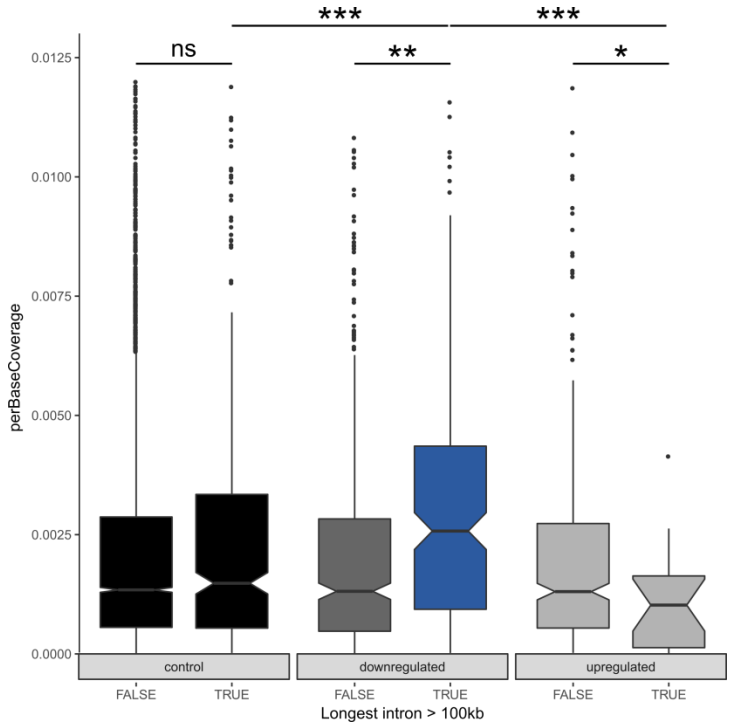


Figure 3.31. Comparison of per-base TDP-43 iCLIP cluster coverage in long and non-long genes also split by significant dysregulation. Long gene is here defined as a gene in which the longest intron exceeds 100kb in length. Genes which do not meet these criteria will be described as non-long genes. From left to right: expression-unchanged non-long genes, expression-unchanged long genes, downregulated non-long genes, downregulated long genes, upregulated non-long genes, upregulated long genes. Significantly higher levels of TDP-43 binding in downregulated long genes, and significantly lower levels in upregulated long genes. Boxplots plotted to show measurement of TDP-43 iCLIP cluster coverage in each group.

The downregulation of these long genes as a result of TDP-43 binding has implications for the neurological aspect of ALS. It has been reported that many of the longest genes are involved in neuronal functioning (Polymenidou et al. 2011), and TDP-43 has been demonstrated to regulate transcripts that regulate neuronal development or have been

linked to neurodegenerative disease (Tollervey et al. 2011; Sephton et al. 2011). Therefore, the importance of TDP-43 regulation in these long genes linked to neuronal function has a strong basis. In addition to their increased length, the transcripts from these genes also exhibit complex splicing patterns (Zheng & Black 2013), and TDP-43 is also known to bind to distal sites within long introns to control splicing patterns (Tollervey et al. 2011).

The precise mechanism by which TDP-43 binding regulates gene expression in general is unclear, as is the further, apparently specific, binding activity in relation to long gene regulation. One possibility is that TDP-43 binding may result in alternative splicing in these long intron genes, leading to unproductive transcripts. An example of this type of unproductive splicing-mediated expression control can be seen in the TDP-43 autoregulation mechanism. In turn, a reduction in TDP-43 binding may mean alternative splicing shifts towards more stable isoforms and therefore an upregulation in these sequenced transcripts. The dysregulation of these long genes is also likely to account for some of the pathways identified from the general GO term analysis, and may therefore contribute to the neuronal pathways found to be most affected, namely glutamatergic neurons and ion channel transport. These results further indicate the need to study the nervous system development and functioning of the homozygous F210I model. Intriguingly, dysregulation in a similar subset of genes has also been found to result from FUS depletion, offering a connecting explanation for how mutually exclusive pathology of the two RBPs may both lead to similar phenotypes in ALS (Lagier-Tourenne et al. 2012).

3.2.6 ALTERNATIVE SPLICING IN F210I

Substantial changes in alternative splicing were also detected in the F210I embryonic homozygous RNA-Seq dataset. Of the 230 significantly altered splicing events (adjusted

p-value < 0.01), 81 showed a reduction in exon (or retained intron) inclusion, whereas 85 showed increased exon or intron retention (Figure 3.32), but the splicing events with the largest ΔPSI values were exon inclusions.

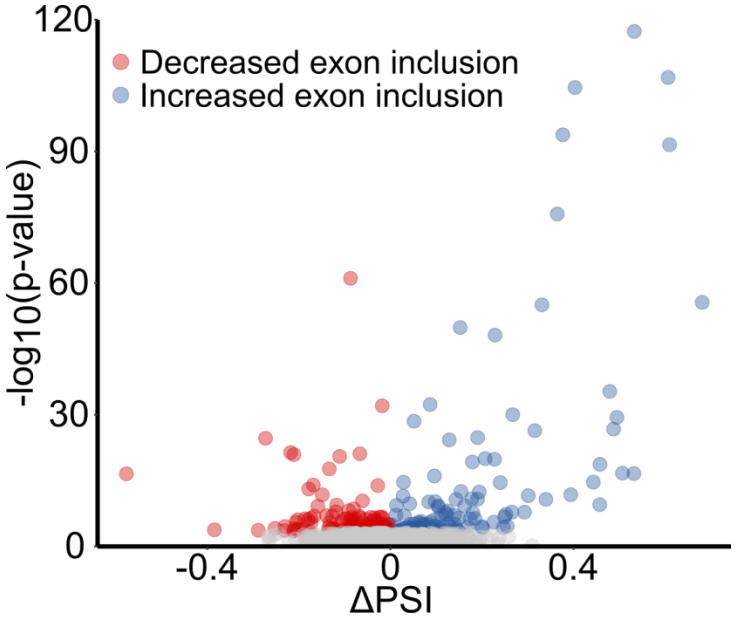


Figure 3.32. Volcano plot for cassette exon alternative splicing changes in homozygous F210I forebrain vs wildtype RNA-Seq dataset. Each dot represents one splicing event. Larger magnitude of change seen in exon inclusion events in comparison to exon skipping. Non-significant changes in grey, significant decrease in exon inclusion in red, significant increase in exon inclusion in blue.

A further 48 events represented differences in 5'UTR and 3'UTR start sites ,with 35 events exhibiting an increase in proximal splice site usage, and the other 13 showing more distal splice site use. The remaining number of splicing events were characterised as alternative first exon, alternative last exon or mutually exclusive exon. Interestingly, 74 of the 230 significantly alternatively spliced genes also showed significant changes in expression levels. In contrast to homozygous F210I models, heterozygous adult mice showed just two significantly altered splicing events.

As reported previously, TDP-43 KD in adult mouse striatum resulted in the dysregulation of 601 transcripts, compared to 2,730 in our F210I loss of function dataset, but 965 alternative splicing changes, versus 230 in the F210I (Polymenidou et al. 2011). Although it is not possible to strictly compare these numbers given the different methods of data processing and statistical analysis, the large discrepancy in number of differentially expressed and alternatively spliced genes warrants further comment, particularly given the similarity of the F210I and TDP-43 KD effects in specific long gene dysregulation. The shift away from changes in alternative splicing in particular may be due to the loss of function in F210I TDP-43 specifically affecting RNA binding. The C-terminal domain, required for protein-protein interaction, remains perfectly functional, and may be able to compensate for RNA binding deficits through modulation of splicing factor recruitment and interaction. In contrast, knockdown of TDP-43 levels in general would result in loss of both RNA binding and splicing factor interaction, which may have a more severe effect on alternative splicing control. In relation to modelling the TDP-43 loss in ALS, nuclear depletion of TDP-43 in motor neuron-like NSC-34 cells was also found to result in widespread changes in alternative splicing (Highley et al. 2014), and splicing changes have also been demonstrated in ALS patient tissue (Xiao et al. 2011).

Importantly, despite a strong level of conservation in the TDP-43 protein structure and function, and shared targets for splicing control in mouse and human such as the *Tardbp*/*TARDBP* gene itself, there remain key differences in the two settings. One previous study identified five splicing changes validated to be altered upon TDP-43 KD in a human setting: *POLDIP3*, *STAG2*, *FNIP1*, *BRD8* and *MADD* (De Conti et al. 2015). In the F210I mouse model, a similar feature of exon skipping was seen in the *Poldip3*, and this event had previously been reported in other studies (Shiga et al. 2012; Yang et al. 2014), demonstrating a similar function for TDP-43 in mice and humans. However, the four other splicing changes identified and confirmed in the HEK293 cells were not

changed in the F210I model. Given the thorough sequencing conducted in the embryonic homozygous F210I RNA-Seq dataset, it seems likely that the lack of change in these splicing targets points may point to a difference in importance of TDP-43 in these splicing events in the two species.

Unlike differential gene expression analysis, GO enrichment for significant changes in alternative splicing revealed no specifically affected pathways.

3.2.7 CRYPTIC EXON INCLUSION

As was the case with differential gene expression analysis, an intriguing pattern of change was detected in the alternative splicing behaviour in homozygous F210I models, namely the inclusion of "cryptic exons" in a number of transcripts. Cryptic exons here refer to typically intronic regions of transcripts that showed increased levels of splicing into processed transcripts. Unlike retained intron events, also linked to ALS (Luisier et al. 2018), cryptic exons are spliced by activation of unfavoured cryptic splice sites (Kapustin et al. 2011). Cryptic exons were first described in relation to TDP-43 loss of function in an investigation into TDP-43 knockout (KO) effects on mouse embryonic stem cells (mESCs) (Ling et al. 2015). Upon conditional KO of TDP-43 and subsequent RNA-Seq, the authors noticed that several regions annotated as intronic contained read profiles similar to neighbouring exons. When they further analysed previously published datasets, they found the same phenomena occurring elsewhere also. These cryptic exons typically resembled cassette exons, but also involved alternative UTR sequences, exon extensions, and alternative polyadenylation sites. It was predicted that the majority of these cryptic exons would result in a premature stop codon (PTC) and cause NMD of the transcript. Concluding that TDP-43 loss of repression was responsible for the appearance of these exons, forced repression of the cryptic exons ameliorated the cell-death phenotype that occurred as a result of

TDP-43 loss. As the cryptic exons occurred primarily in deep intronic regions, they were not predicted to be conserved in human settings. However, the phenomena of cryptic exons could be conserved, even if precise location was not, and a different set of cryptic exons were detected in HeLa cells upon TDP-43 loss. Crucially, two of these exons were further found to be increased in ALS patient tissue, suggesting a disease-relevant role for this type of alternative splicing. Since the original study, cryptic exon inclusion upon TDP-43 knockdown has been further studied and corroborated, emphasising the role of TDP-43 loss of repression in cryptic exon inclusion (Tan et al. 2016).

Given the clear molecular signature of cryptic exon inclusion as a result of TDP-43 loss of function, we expected that the F210I mutation would cause a similar pattern of splicing dysregulation. Indeed, several cryptic exons were discovered in the homozygous F210I embryonic forebrain RNA-Seq dataset, with a sizeable overlap with the cryptic exons detected in mESCs (Ling et al. 2015). An example gene with a cryptic exon as a result of F210I mutation, *Adnp2*, is shown in Figure 3.33. This cryptic exon was also identified in mESC TDP-43 KO (Ling et al. 2015).

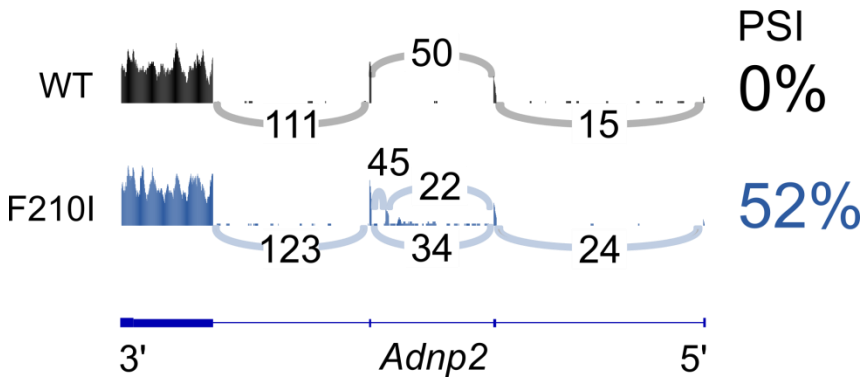


Figure 3.33. RNA-Seq read counts from wildtype and homozygous F210I datasets for part of the *Adnp2* gene aligned to a reference. Overlaid with sashimi plot showing splicing patterns in each dataset, with number of each splice event annotated. PSI levels for the cryptic exon are calculated and listed to the right. No

detected instance of alternatively spliced exon inclusion in wildtype samples, but approximately half of all transcripts showed exon inclusion in the F210I dataset.

Although the methods of analysis are not clear, it seems as though cryptic exon detection by Ling et al. (2015) was conducted in a manual manner. We adopted a more statistical approach to enable discovery of cryptic exons in our alternative splicing analysis datasets. Cryptic exons were defined as exons repressed in wildtype tissue, with a PSI of less than 0.05, or 5 %. Further, they were required to be significantly upregulated in test datasets ($p < 0.01$), and also required a ΔPSI increase of at least 0.05, with prior annotation of the exon or isoform not taken into consideration. Using these selection criteria, 33 cryptic exons were identified (Figure 3.34).

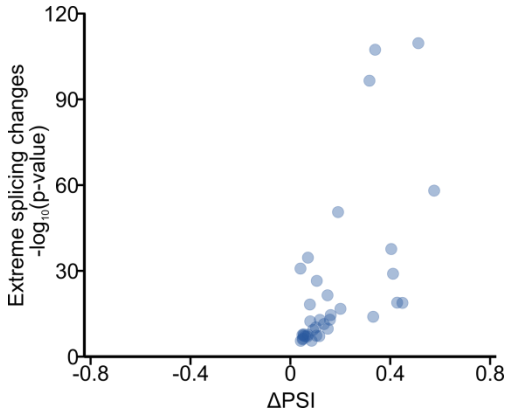


Figure 3.34. Cryptic exon (extreme splicing change) ΔPSI and level of significance for each of the 33 identified cryptic exons in homozygous embryonic F210I samples vs wildtype RNA-Seq dataset.

A subset of these was then validated by RT-PCR, with all showing significant and substantial increases in inclusion levels in the F210I model (Figure 3.35). Neither these, nor any other, cryptic exons were identified in the heterozygous adult F210I dataset.

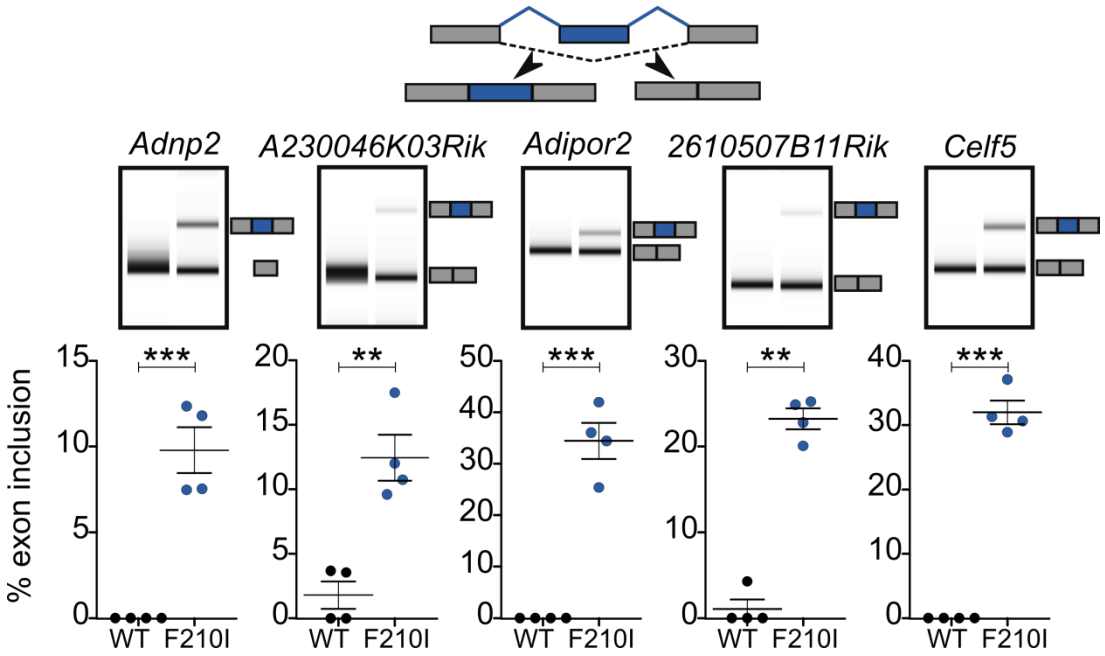


Figure 3.35. Top: Diagram of two possible isoforms from RT-PCR test, with cryptic exon denoted in blue. Middle: Example RT-PCR results visualised on Tapestation gels for five targets, with cryptic exon-containing transcripts and wildtype transcripts labelled. Cryptic exon transcripts only clearly evident in F210I samples. Bottom: Graphical and statistical analysis of cryptic exon inclusion levels for each target in wildtype and F210I samples.

Similar to the cryptic exons identified by Ling et al. (2015), analysis by Jack Humphrey suggested that the vast majority of cryptic exons would result in a PTC and eventual NMD in the containing transcripts. To understand the consequence of this shift in transcript fate, the differential expression levels of cryptic exon-containing genes were investigated. 12 of the 33 cryptic exon genes showed significant differences in expression levels in the F210I dataset (Figure 3.36). Although this is a sizeable portion, it is not significant to cryptic exon-type alternative splicing when compared to overall levels of overlap between alternatively spliced transcripts and their differential expression (hypergeometric test, $p = 0.35$); 74 out of 230 alternative spliced transcripts showed significant dysregulation. However, a curious finding was that of

the differentially expressed cryptic exon transcripts, there were equal levels of up- and downregulation, whereas one would expect primarily downregulation given the predicted NMD fate of most of these transcripts. The reason for this discrepancy is unclear. One possibility is that some of these transcripts are additionally otherwise regulated, or themselves capable of autoregulation, and an increase in total transcript number has been induced to compensate for the greater proportion of transcripts being degraded.

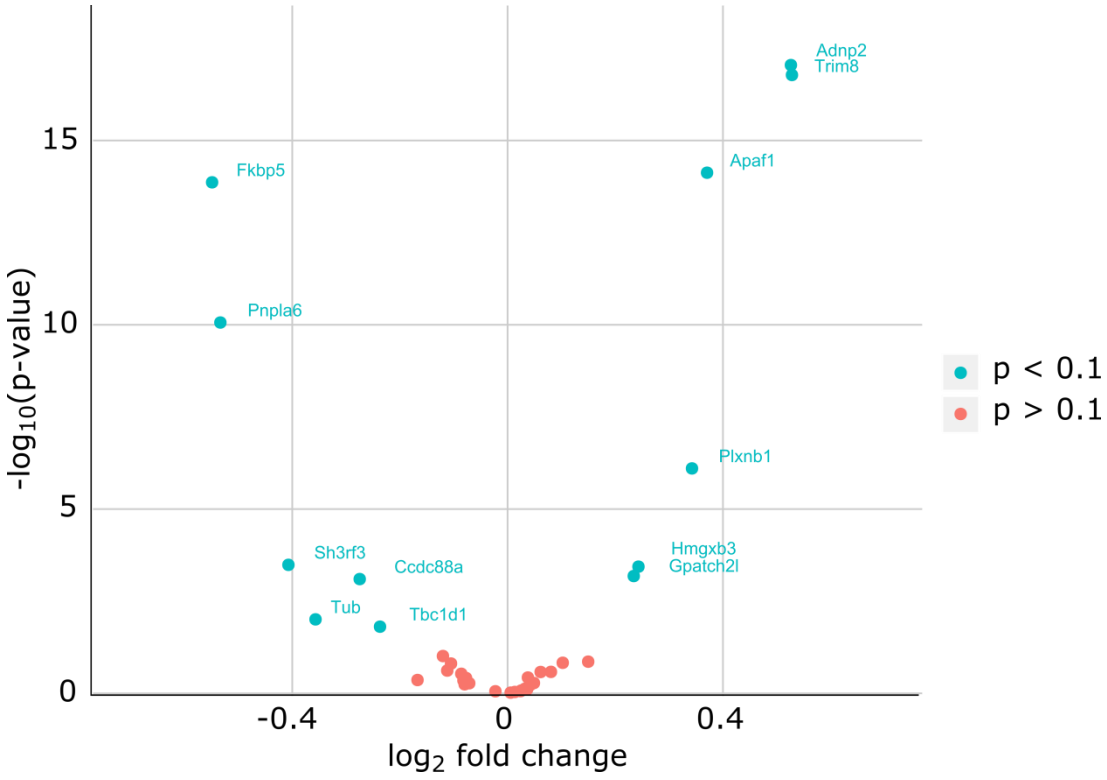


Figure 3.36. Volcano plot of differential gene expression in transcripts containing cryptic exons in homozygous embryonic F210I samples vs wildtype. Cryptic exon transcripts with significant changes in gene expression are highlighted by blue dots and labelled.

Further understanding of NMD involvement in cryptic exon transcript processing is also required to garner a clearer picture of the number and prevalence of cryptic exons

as a result of TDP-43 loss of function. The cryptic exons seen in our and other RNA-Seq datasets are often in transcripts which should be degraded by NMD. NMD can be a somewhat inefficient process (Cusack et al. 2011), but it is not known whether these transcripts are being sequenced due to NMD inefficiency in general, or whether cryptic exon transcripts in particular possess properties which enables them to escape NMD. Tagging and sequencing of nascently formed transcripts upon NMD block would provide considerable insight into this question, and warrants further investigation.

Ling et al. (2015) also noted that the majority of cryptic exons which resulted from TDP-43 loss were surrounded by UG repeats, and using previously published HITS-CLIP data demonstrated that these regions were directly bound by TDP-43. Jack Humphrey performed similar analysis and found that a large subset of cryptic exons identified in the F210I RNA-Seq dataset showed proximal TDP-43 binding from the F210I iCLIP datasets. Binding of TDP-43 was particularly common within the cryptic exon or proximally to it, with binding in some introns around cryptic exons occurring more distally (Figure 3.37).

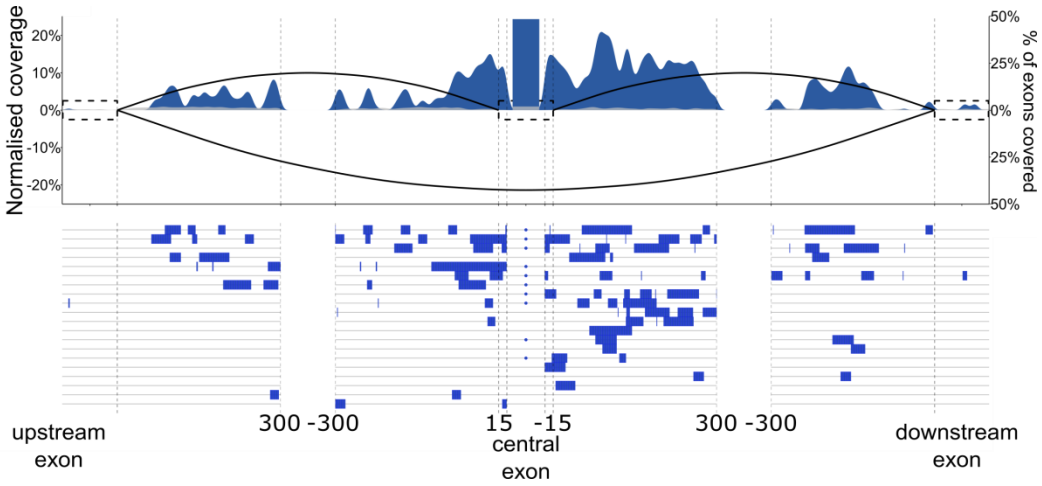


Figure 3.37. TDP-43 binding sites in and around cryptic exons and flanking intronic regions. Each track represents TDP-43 binding sites from F210I iCLIP

datasets for TDP-43-bound transcripts, with summed coverage displayed on top. The central cryptic exon length has been normalised to enable analysis proximity of TDP-43-intron binding. Results from Jack Humphrey.

The direct link between TDP-43 loss of function and cryptic exon inclusion is clear, and these results add yet more weight to the F210I as a model for TDP-43 loss of function in splicing activity. However, the link between these extreme splicing changes and a resulting disease phenotype is much less apparent, despite the discovery of two cryptic exon transcript isoforms seemingly specific to ALS/FTD (Ling et al. 2015). Specifically, the reasons behind cryptic exon transcripts leading to cell death need to be disentangled.

One possibility is that it is not the specific transcripts containing the cryptic exons which are important, but the nature of cryptic exons themselves. Perhaps cryptic exon inclusion as a result of TDP-43 loss is actually a widespread phenomenon, not yet identified due to the degradation of most cryptic exon isoforms, and that the existence of such a large number of mis-spliced transcripts leads to overuse and an eventual loss of control of the cell's degradation machinery. If the specific transcripts themselves are not vital to the final death of the cell, then cryptic exon processing could be further studied in mouse models to understand their role in human settings.

Another hypothesis may be that there exists a specific subset of transcripts crucial for neuronal function in which cryptic exons are repressed by TDP-43, and in which, upon TDP-43 loss, the resulting cryptic exon inclusion and dysregulation would lead to neuronal cell death. In this case, where the cryptic exons themselves are more directly implicated in death of the cell, there are a number of problems with studying this mechanism. The most crucial hurdle is that, as a result of their primarily intronic location, cryptic exons are not conserved between mice and humans. Therefore, attempting to understand specific pathways in which cryptic exon transcripts are

involved in the F210I model is of limited use. Although the method by which cryptic exons occur could be further studied, the link between this and human motor neuron cell death will not be clarified. Even if human models are used to investigate TDP-43 loss, there may still exist a lack of concordance between these models and patient motor neurons, as cryptic exons have been shown to be cell-specific as well as species specific (Jeong et al. 2017). For these reasons, the method for future investigation of these cryptic exons must utilise human motor neuron tissue. Modelling with iPSCs may be possible; it is not yet clear whether cryptic exon occurrence is also age-dependent.

Nonetheless, despite these obstacles, the strong link between cryptic exons and TDP-43 loss means that further investigation into their precise effects is vital. This is particularly true as cryptic exons as a result of TDP-43 loss are not ALS specific, having been found in the even more common neurodegenerative disease, Alzheimer's (Sun et al. 2017). Excitingly, in this case the same two isoforms found in ALS/FTD patients but not controls by Ling et al. (2015) are found in the Alzheimer's cases exhibiting nuclear TDP-43 loss, suggesting a common link between neurodegenerative diseases. This does however raise questions about cryptic exon specificity to the motor neuron changes in ALS. Additionally, as well as other diseases, other RBP loss has also been demonstrated to cause cryptic exons (Ling et al. 2016; McClory et al. 2018). That this feature is neither disease, nor protein specific may mean it is not the key feature underlying any one of these diseases or protein dysfunctions, but on the other hand it raises the possibility that it contributes to the molecular dysregulation and phenotype in a variety of disorders.

3.2.8 F210I CONCLUSIONS

The hypothesis that the F210I mutation resulted in a hypomorphic allele, and therefore a model of TDP-43 loss of function, was initiated through low-throughput splicing

analysis of known targets *CFTR* exon 9 and *Sort1* exon 18, alongside a global reduction in RNA binding. Analysis of transcriptomic and iCLIP data from the homozygous F210I mouse model has confirmed this prediction, with numerous parallels in molecular dysregulation between the homozygous mice and various TDP-43 knockdown studies. Intriguingly, some mutations that disrupt RNA binding of TDP-43 have actually been reported to ameliorate TDP-43-mediated toxicity (Voigt et al. 2010). The substantial differential gene expression caused by the mutation affected pathways previously identified as being impacted by TDP-43 loss, and linked the dysregulation to neuronal functions. This was furthered by the specific disruption in both TDP-43 KD and F210I mutation in regulation of the longest genes, again linked to neuron processes. Both aspects of gene expression changes suggest F210I TDP-43 can replicate some of the effects of nuclear TDP-43 loss seen in ALS, and therefore acts as a useful model for understanding the molecular dysregulation that can result.

The ability of TDP-43 to deal with the consequences of its own depletion is also explored in the F210I model, revealing a strong autoregulatory response due to the F210I loss of function mutation in a similar manner again to TDP-43 KD. However, the strong shift towards more transcripts resulting in productive *Tardbp* isoforms to be translated in the F210I animal, and the substantial molecular dysregulation that occurs despite this, may provide insight into the limitations of the autoregulatory mechanism, and the widespread consequences when this limit is reached. The alternative splicing changes, whilst not specifically related to neurons, highlighted the key feature of cryptic exons, for which the F210I model RNA-Seq and iCLIP strengthened the direct link to TDP-43 loss of function. Although the connection to resulting phenotypes is poorly understood, the link to cell death suggests that cryptic exon transcripts have significant consequences on cell survival.

A key finding of two cryptic exon isoforms which so far seem to be neurodegenerative disease-specific further suggests these often novel transcripts may be involved in disease progression, and it may be progression as opposed to initiation in which cryptic exons prove relevant. In summary, the F210I model provides novel insight into the effect of TDP-43 mutation-mediated loss of function in vivo, and highlights precise molecular signatures for which TDP-43 is crucial in development and the dysregulation that occurs when this control is lost.

CHAPTER 4

M323K MODEL TRANSCRIPTOME DYSREGULATION AND PHENOTYPE

4.1 INTRODUCTION

4.1.1 TDP-43 LOW COMPLEXITY DOMAIN IN ALS

Unlike the RRM mutation studied in the F210I mouse, several models for LCD mutations in TDP-43 exist (reviewed in Y.-C. Liu, Chiang, & Tsai, 2013), and this is primarily due to the vast majority of TDP-ALS mutations occurring in the disordered LCD. The broad region of interest is the amyloidogenic core region, comprised of residues 311-360, which is a hotspot for ALS-causing mutations in TDP-43 (Buratti 2015). Within this region is a particularly important, completely conserved 319-341 residue region. This highly conserved amino acid sequence contains the 321-330 residue region which results in an α helical folding, surrounded by regions of disorder (Lim et al. 2016). This folding structure was found to be disrupted by TDP-ALS mutations (Conicella et al. 2016); the altered phase separation abilities that resulted may indicate a possible mechanism underlying TDP-43 dysfunction in ALS.

Studying pathology, or other molecular dysregulation caused by TDP-ALS mutations, however, has proven difficult. Mouse models have, until recently, almost universally used transgenic expression of human TDP-43 constructs incorporating TDP-ALS mutations alongside the endogenous mouse TDP-43 expression to study the effects of the mutation (De Giorgio et al. 2018). These models often display a motor phenotype (Wegorzewska et al. 2009; Zhou et al. 2010). However, as alluded to in 1.3.3 TDP-ALS mouse models, such mice have the key drawback of not allowing distinction between the effects of the mutation, and the effects of TDP-43 overexpression, which is itself

known to be toxic (Tsai et al. 2010; Wils et al. 2010; Xu et al. 2010). Study of these mutations is further complicated by the differences of effects even in the TDP-ALS mutations found in similar regions of the LCD. Investigation into the results of four mutations in the hotspot region of the LCD, Q331K, G335D, M337V and Q343R, found that each mutation had different effects on RNA processing dysregulation and aggregation formation (Jiang et al. 2016). The first TDP-ALS mutations discovered – Q331K and M337V (Sreedharan et al. 2008) – were not found to cause TDP-43 aggregation, despite the evident inclusions seen in TDP-ALS patients with the same mutations.

One intriguing study by Arnold et al. (2013) developed transgenic mouse models based on the aforementioned two TDP-ALS mutations, Q331K and M337V, but importantly determined the overall TDP-43 levels to be similar to endogenous TDP-43 in wildtype controls. This model showed substantial changes in alternative splicing of TDP-43 targets, implicating the LCD domain in controlling TDP-43 splicing ability. The mouse also developed motor axon degeneration and displayed signs of motor neuron death, making it a model for neurodegeneration. However, it displayed no TDP-43 pathology, a hallmark of ALS, and therefore still failed to fully recapitulate the ALS phenotype (Arnold et al. 2013). As well as struggling to imitate patient phenotype, mouse models in general have also failed to accurately show changes in the transcriptome which relate to the changes seen in ALS patients, such as, for example, the opposite effects in expression changes of genes involved in stress response (Burns et al. 2015). Although differences would be expected, the fact that transcriptomic changes upon processes such as aging are well conserved between mouse and human settings (Burns et al. 2015) gives hope that similar correlation for disease modelling is possible.

4.1.2 PREVIOUS STUDY OF THE M323K MOUSE MODEL

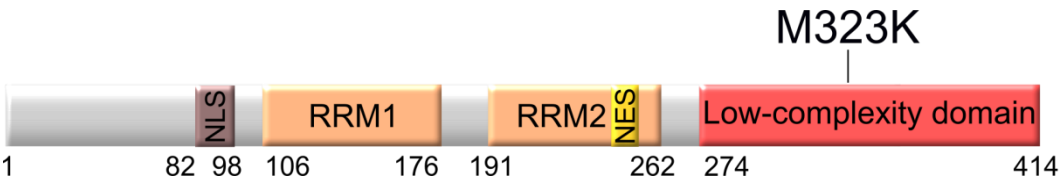


Figure 4.38. TDP-43 domains and location of the M323K mutation within the LCD.

The M323K mutation (Figure 4.38), although not yet found as a causative mutation in ALS patients, lies within the 20 amino acid stretch which influences TDP-43 α helix structure, and therefore could have an impact in the ability of TDP-43 to phase separate and lead to aggregation (Conicella et al. 2016). Turbidity tests undertaken by Nicolas Fawzi's lab revealed this decreased propensity for phase separation in M323K mutant TDP-43 in comparison to wildtype, and similar to the effects of TDP-ALS mutation Q331K (Figure 4.39).

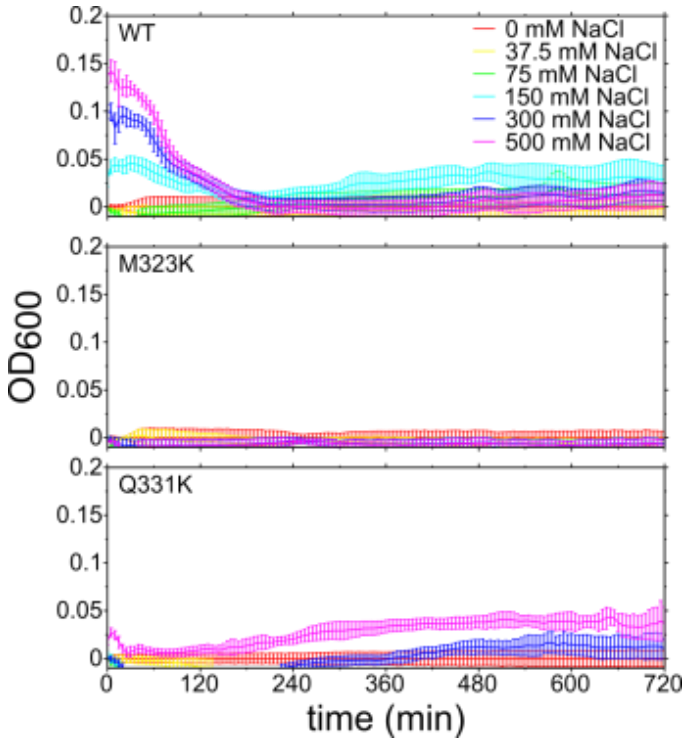


Figure 4.39. Turbidity of wildtype and M323K and Q331K mutated TDP-43 C-terminal fragments from residue 267-414 in the presence of 0-500 mM NaCl

quantified by optical density at 600 nm wavelength light. Measurements taken at 5 min intervals over 12 hrs, experiment by Nicolas Fawzi.

Initial characterisations of the effects of the M323K mutation were carried out similarly to the F210I mutation. To determine its impact upon TDP-43 splicing function, the *CFTR* minigene assay was again employed (Buratti et al. 2001). As outlined in Chapter 3 F210I model transcriptome dysregulation, the F210I loss of function mutation replicated the effects of TDP-43 KD, in that it caused increased levels of exon inclusion in the *CFTR* minigene from the 50 % level of inclusion brought about by wildtype TDP-43. Similar testing of V5-tagged M323K TDP-43 expression in HeLa cells with siRNA-mediated endogenous TDP-43 KD revealed that in contrast to the increased exon inclusion of the F210I and TDP-43 KD, M323K TDP-43 showed no change in exon 9 inclusion levels (Figure 4.40).

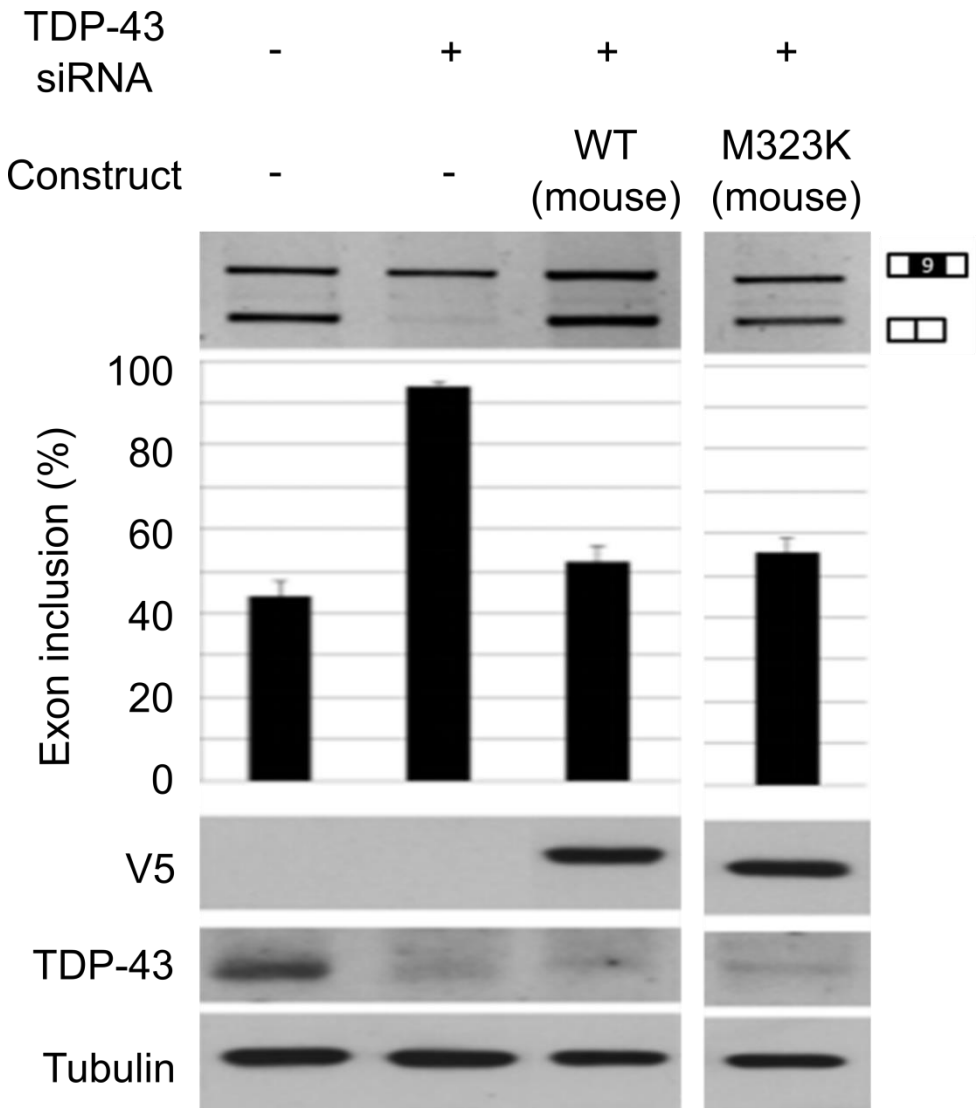


Figure 4.40. *CFTR* minigene assay showing RT-PCR results and % exon 9 inclusion in (left) wildtype TDP-43, (middle left) TDP-43 KD, (middle right) TDP-43 rescue with V5-tagged mouse TDP-43 construct, (right) TDP-43 rescue with V5-tagged mouse M323K-TDP-43 in HeLa cells. TDP-43 siRNA presence in experiment indicated by '+'. Exon 9 inclusion at approximately 50 % in wildtype conditions, substantially increased upon TDP-43 KD, rescured by wildtype mouse TDP-43 expression, and similarly rescued by M323K mutant TDP-43 expression. Minigene isoforms for exon 9 inclusion and skipping labelled. Example bands for V5, TDP-43 and Tubulin labelled. Error bars show SEM (Ricketts 2012).

However, when this effect was also tested ex vivo, in MEFs derived from wildtype, heterozygous or homozygous M323K E14.5 mice, M323K TDP-43 showed a surprising dose-dependent decrease in exon inclusion. Exon inclusion levels were approximately 50 % as expected for wildtype TDP-43-mediated splicing of *CFTR* exon 9, but were significantly reduced to 35 % in homozygous M323K MEFs (Figure 4.41). As this was the opposite to the loss of function effects modelled by the F210I mutation and TDP-43 KD, this *CFTR* minigene assay in MEFs suggested that the M323K resulted in an unexpected gain of function in TDP-43 splicing activity.

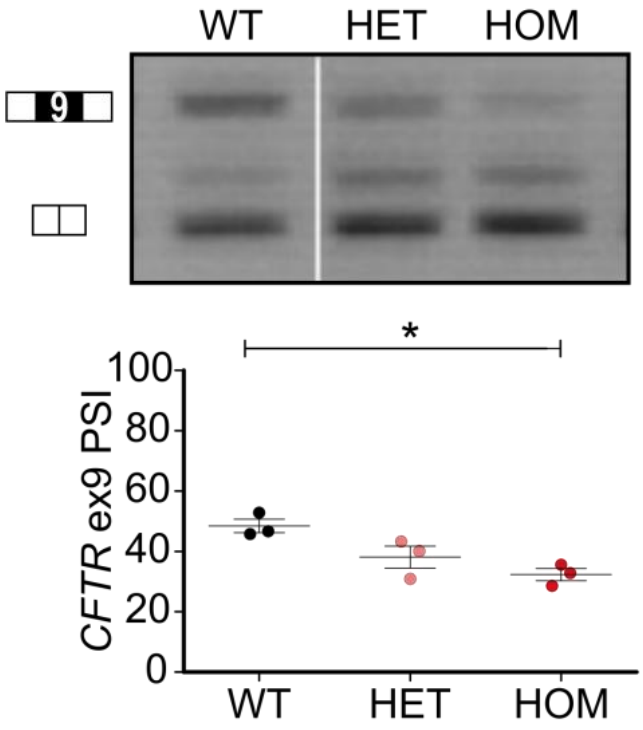


Figure 4.41. *CFTR* minigene assay showing RT-PCR results and % exon 9 inclusion in (left) wildtype TDP-43, (middle) heterozygous M323K TDP-43, (right) homozygous M323K TDP-43 in MEFs. Significant and dose dependent increase in exon 9 inclusion as a result of M323K mutation. Minigene isoforms for exon 9 inclusion and skipping labelled. Results from Abraham Acevedo Arozena. Error bars show SEM.

Analysis of known splicing targets of TDP-43 furthered the notion of hyperactive splicing of TDP-43. Exon 18 of *Sort1* is repressed by TDP-43 (Prudencio et al. 2012), and RT-PCR experiments revealed a loss in repression in F210I settings. Concordantly, the gain of function effect of M323K mutation resulted in increased repression of exon 18 in the heterozygous mice and even more so in the homozygous models (Figure 4.42). Splicing effects opposite to the F210I mutation were also seen by RT-PCR testing of alternative splicing events in *Eif4h* and *Dnajc5*, fitting with the respective loss of function versus gain of function splicing effects of the F210I and M323K mutations.

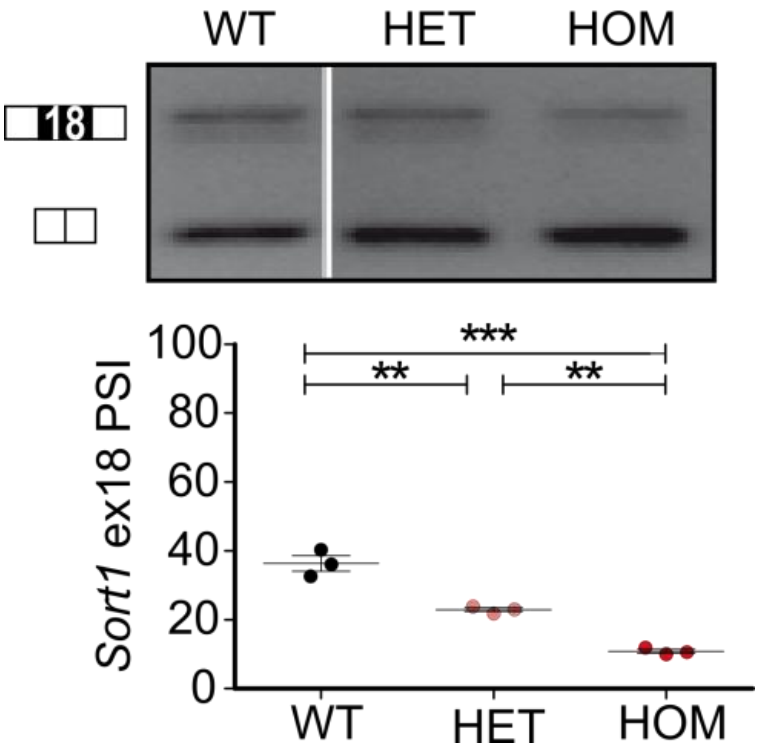


Figure 4.42. RT-PCR results and quantification for inclusion levels of *Sort1* exon 18 in wildtype, heterozygous M323K and homozygous M323K MEFs. Significant and dose dependent decrease in exon 18 inclusion of *Sort1* as a result of M323K mutation. Isoforms for exon 18 inclusion and skipping labelled. Results from Abraham Acevedo Arozena. Error bars show SEM.

The F210I mutation, as well as acting as a hypomorphic allele in TDP-43 splicing, also showed a global reduction in RNA binding ability. The effect of a gain of function in TDP-43 activity on RNA binding was not known, and testing using EMSA experiments for TDP-43 binding to oligonucleotides containing six UG repeats, the target motif of TDP-43, showed no difference in M323K mutant TDP-43 compared to wildtype (Figure 4.43). It was therefore unclear whether the M323K mutation only caused a gain of function in the splicing activity of TDP-43 in particular, or if it was a general gain of function in TDP-43 activity mirroring the varied loss of function effects of the F210I mutation.

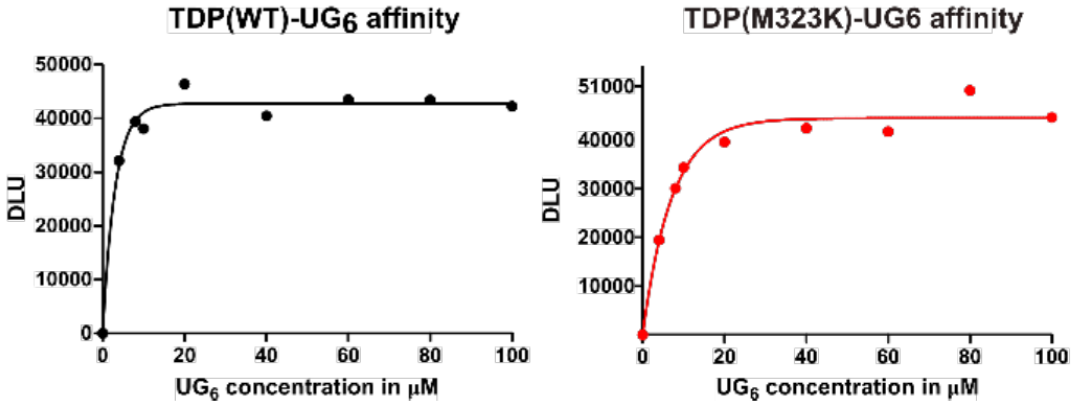


Figure 4.43. Quantification of EMSA for wildtype and M323K TDP-43 binding to oligonucleotide sequences containing known TDP-43 target motifs of six UG repeats. No change in DLU (Digital Light Units) suggesting no change in TDP-43 affinity for these sequences as a result of M323K mutation. Results from Abraham Acevedo Arozena.

To study the effects of the M323K mutation in vivo, the M323K mouse model was re-derived after characterisation of the ENU mutagenesis-generated mutation. Mice were initially backcrossed from the founder C57BL/6J / DBA/2J mixed background to a pure C57BL/6J background. On the pure background homozygous embryos were perinatally lethal, similar to the F210I mutation, while heterozygous models were viable. MEFs

were obtained for study from these models, and embryonic study was also conducted at this stage. However, when maintained on the mixed background, homozygous M323K mice not only survived to adulthood, but showed no reduction in lifespan. This effect of different mouse backgrounds has previously been reported (LeCouter et al. 1998). Therefore, in order to study the effects of homozygous gain of function mutation of TDP-43, study of the adult M323K model was conducted on mice of mixed C57BL/6J / DBA/2J background.

4.2 RESULTS AND DISCUSSION

4.2.1 RNA SEQUENCING OF M323K OVERVIEW

Investigation of splicing changes as a result of the M323K mutation in the LCD of TDP-43 revealed a surprising counter-effect to the modulation seen as a result of TDP-43 KD or loss of function in the F210I, suggesting an unexpected gain of function in TDP-43 splicing activity, in known splicing targets such as *Sort1*, *Eif4h* and *Dnajc5*. To investigate whether this gain of function held true across the transcriptome of TDP-43 targets, and contrast the effects to the loss of function in splicing caused by the F210I mutation, we conducted RNA-Seq on the M323K model at the embryonic and adult time points, once again focusing on differential gene expression and alternative splicing.

Initial sequencing of the M323K model was conducted on heterozygous and homozygous M323K E14.5 mouse heads, in keeping with the original F210I sequencing. Sequencing was conducted as described in 2.1.6 RNA Sequencing, using 35 bp paired-end reads with sequencing depth of approximately 20 M unique reads per sample (n = 3). Differential gene expression and alternative splicing analyses were conducted as outlined in 2.1.7 Bioinformatics analysis. Similarly to the F210I model, heterozygous mice showed minimal changes in RNA dysregulation, with two significant changes in gene expression and a further two significant splicing alterations. Furthermore, however, comparatively few changes were also detected in the homozygous M323K embryonic model in comparison to the homozygous F210I, with 53 differentially expressed genes and 11 splicing changes. This suggested that the M323K mutation had a reduced impact in RNA regulation during development of the mouse, in keeping with the ability of the homozygous M323K mice, but not the homozygous F210I mice, to survive until adulthood with no change in lifespan on a

mixed background. This difference also may be due to the absence of the global change in RNA binding ability in the M323K models as outlined in 4.1.2 Previous study of the M323K mouse model, in contrast to the reduced RNA binding resulting from the F210I mutation.

Importantly, whereas only the heterozygous F210I models could be studied at adult time points, the homozygous equivalents could also be investigated in the M323K. As was the case with every other heterozygous model, adult heterozygous M323K mice showed minimal instances of RNA processing disruption. However, by 6 months, adult homozygous M323K spinal cord showed severe changes in both gene expression (1649 genes significantly dysregulated) and splicing (295 splicing events significantly changed). The magnitude of these changes mirrored the substantial modifications seen in the homozygous F210I forebrain, and therefore suggested that the homozygous M323K mutation culminated in a serious effect on TDP-43 function in adulthood in mice.

RNA-Seq dataset	Samples	Read length (bp)	Average reads per sample (~millions)	Differentially expressed genes	Alternatively spliced regions
M323K embryonic head	3 WT, 3 HOM	35 x2	20	53	11
M323K adult spinal cord	4 WT, 5 HOM	75 x2	45	1649	295

Table 4.20. Overview of M323K RNA-Seq experiments and analysis as described in Chapter 2 – Materials and Methods.

4.2.2 M323K ICLIP OVERVIEW

Given that the substantial molecular dysregulation in the M323K model only developed in adulthood, iCLIP was conducted on adult M323K tissue in contrast to the embryonic tissue of the F210I iCLIP dataset. Libraries were successfully developed from

homozygous M323K 6 month brain (wildtype n = 1, homozygous M323K n = 2). Once again, the aim of the experiment was to understand how TDP-43 binding in general was altered as a result of the M323K model in relation to the expression and splicing changes detected at this time point.

The wildtype dataset revealed a total of 55,802 crosslinking peaks, compared to 45,863 sites across the M323K homozygous model replicates. Once filtered for peaks with a false discovery rate of less than 0.05, as suggest by data processing in iCount, a total of 1,795 genes were found to contain these crosslinked sites, and whereas a total of 1,390 genes exhibiting binding of TDP-43 across the two homozygous M323K mutant datasets, only 331 genes were detected to be shared targets in the latter replicates. It is not clear to what degree the variation across replicate gene targets, and also in peak numbers in general, is due to technical variability as opposed to true biological differences. Similarly, the large drop in total peak number in these adult mouse datasets compared to the embryonic wildtype and F210I data may be due to a combination of technical variation in experiments as well as changes in TDP-43 activity with aging. Although the M323K datasets may not capture the full spectrum of TDP-43 binding in the adult mouse models, the binding sites that were detected are likely to be accurately representative of true binding, as 473 of the 513 gene targets detected in either the wildtype or homozygous M323K mutant datasets were also found to be targets of TDP-43 in the embryonic wildtype and homozygous F210I mutant iCLIP datasets too.

Sample	Peaks	Peaks FDR < 0.05	Genes	Shared genes within genotype	Shared genes between genotypes
WT 1	55,802	12,705	1,795	-	513
HOM 1	16,576	2,920	643	331	
HOM 2	29,287	6,439	1,078		

Table 4.21. Overview of embryonic iCLIP datasets for wildtype and homozygous adult M323K forebrain. Analysis conducted as described in 2.2.11 iCLIP Bioinformatics. Shared genes within genotype represents number of genes targeted in both replicates of one genotype. Shared genes between genotypes represents number of genes targeted in at least one replicate of each genotype.

As described in 1.1.2 TDP-43 functions in RNA processing, TDP-43's target motif typically consists of UG repeats (Buratti & Baralle 2001). Using EMSA experiments, it was determined that the binding to UG repeats was true for wildtype TDP-43, and also for M323K TDP-43 with no apparent difference. Pentamer enrichment analysis was conducted as outlined in 2.2.11 Bioinformatics analysis, with the UG (TG) contents of each pentamer annotated. In both wildtype and M323K datasets, the greatest pentamer enrichment was seen in pentamers containing only TG repeats: GTGTG and TGTGT. Further strong enrichment levels were seen for 4 nt sequences of TGTG and GTGT, and several variations of pentamers containing 3 nt sequences of TGT or GTG. In general, pentamer enrichment was very similar for wildtype and M323K TDP-43, with a Pearson correlation coefficient of 0.93 (Figure 4.44).

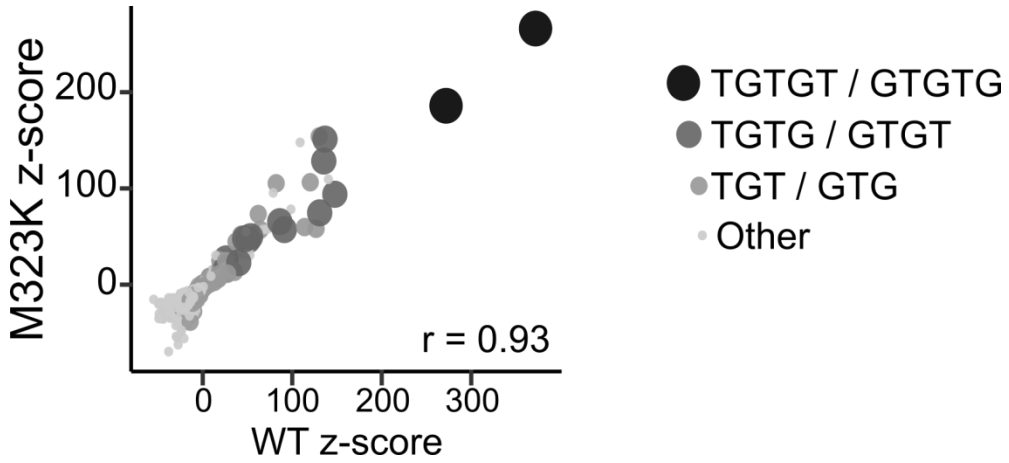


Figure 4.44. Pentamer enrichment for wildtype and homozygous M323K iCLIP datasets. Z-scores measure pentamer enrichment in comparison to mean predicted random enrichment. Pentamer sequences containing TG repeats

highlighted by colour and size. Pentamers containing TG repeats heavily enriched. Strong pentamer enrichment correlation: Pearson correlation coefficient of 0.93.

Similarly, embryonic F210I iCLIP data showed markedly comparable pentamer enrichment to the respective wildtype dataset. Given the similarity of the UG repeat pentamer enrichment in both mouse model datasets, the two were compared to highlight any subtle differences. This analysis revealed that although correlation remained strong, with UG pentamers the most enriched in both, there were a few pentamers which showed greater levels of enrichment in one mutant model compared to the other. However, this marginal difference was replicated in comparison of the two wildtype iCLIP datasets too, suggesting that it may be a difference in the binding motifs of embryonic TDP-43 versus adult TDP-43 (Figure 4.45).

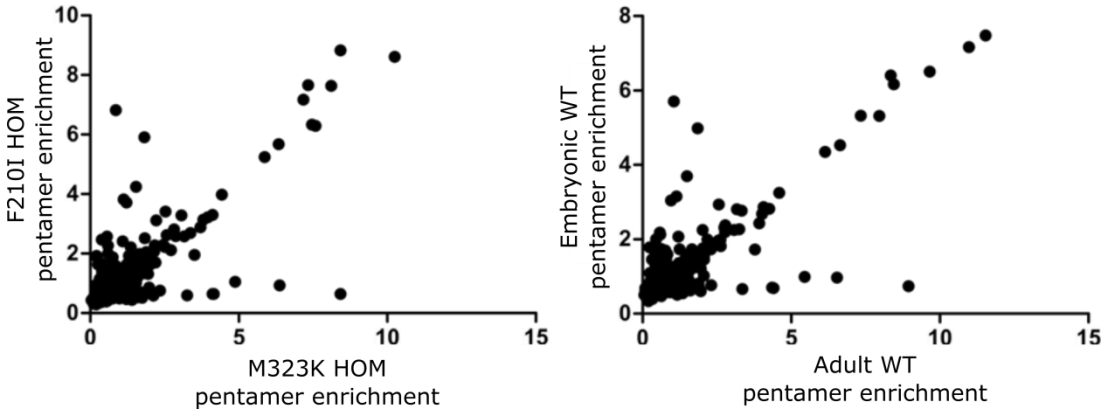


Figure 4.45. Pentamer enrichment comparisons for (left) homozygous F210I and homozygous M323K iCLIP datasets and (right) their respective wildtype controls, with enrichment measuring pentamer enrichment in comparison to mean predicted random enrichment. Greater differences seen in comparison to iCLIP datasets of the same time point.

In summary, the iCLIP data for the M323K mouse suggested that, like the F210I model, the mutation had no obvious effect on TDP-43 targeting of UG repeats.

The F210I mutation was also not found to have any great change in sub-region binding of TDP-43. Splitting of transcripts into sub-regions of 5'UTR, 3'UTR, open reading frame (ORF), non-coding RNA (ncRNA), intron, intergenic and antisense, and analysing TDP-43 binding sites, showed equivalent levels of binding in the F210I model to wildtype in each. Similar analysis for the M323K model revealed the same result. Again, TDP-43 binding sites in each sub-region for the M323K mutant strongly matched wildtype, suggesting the M323K mutation also did not change TDP-43 target sub-regions (Figure 4.46). Instead, there was again a greater apparent difference in embryonic versus adult TDP-43 binding, similar to the pentamer motif enrichments. Although the hierarchy of TDP-43 binding to sub-regions remained the same, the proportions of binding seen in adults were noticeably changed. The primary difference between the two time points was the reduction in intron binding, from 79 % to 53 %, and the increase in intergenic binding from 10 % to 29 %. The shift in the former fits with previous CLIP studies of TDP-43, which investigated binding in the adult mouse striatum and found intron binding accounted for 58 % of total (Tollervey et al. 2011), matching our adult mouse brain results but different to the embryonic mouse brain. However, the increase in intergenic region binding is as of yet unexplained.

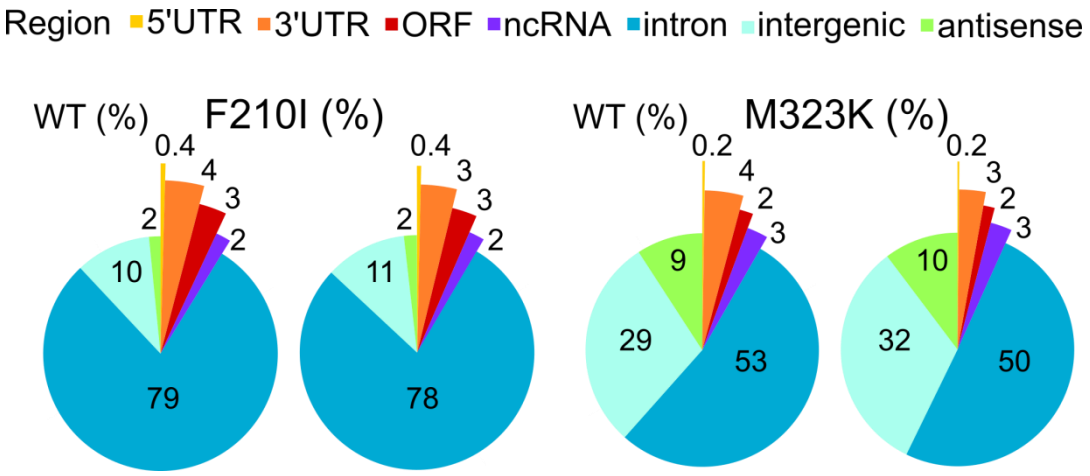


Figure 4.46. Percentage of iCLIP clusters which mapped to each sub-region of transcripts in (left) wildtype and homozygous F210I and (right) wildtype and

homozygous M323K. Similar proportions of sub-region binding in mutants compared to their respective wildtypes, but reduction in intron binding in the adult datasets versus embryonic.

4.2.3 M323K TDP-43 AUTOREGULATION

TDP-43 possesses an autoregulatory mechanism to control its own levels, as discussed in 1.1.2 *TARDBP* isoforms and autoregulation, and requires alternative splicing and polyadenylation of the *Tardbp* transcript.

Given the apparent gain of function in TDP-43 splicing activity, and the resulting spectrum of RNA processing dysfunction, it was expected that the autoregulatory mechanism would shift to reduce TDP-43 levels towards wildtype equilibrium. However, differential gene expression output for the adult homozygous M323K mice revealed a clear increase in *Tardbp* levels instead (Figure 4.47), with a $\log_2$ fold change of +0.38 and an adjusted p-value of $1.44e^{-4}$ obtained from general differential gene expression analysis. It is similar, but of lower magnitude, to the upregulation of *Tardbp* seen in the loss of function homozygous F210I model. Whereas in F210I this would represent the appropriate compensation to the resulting loss of function effects, in the M323K model this upregulation implies an increase in TDP-43 protein levels, which in theory would worsen the molecular phenotype resulting from the gain of function in splicing activity of the protein. The reason for this increase in expression, which seems to be contrary to the desired restoration of wildtype TDP-43 activity, is not apparent. Drawing parallels with other mouse models of LCD mutations is made difficult by the artificial overexpression resulting from transgenic TDP-43 expression, no matter how slight. However, similar increases in *TARDBP* transcript levels have been established in post-mortem ALS spinal cord motor neurons (Mishra et al. 2007; Gitcho et al. 2009; Swarup, Phaneuf, Dupré, et al. 2011; Koyama et al. 2016).

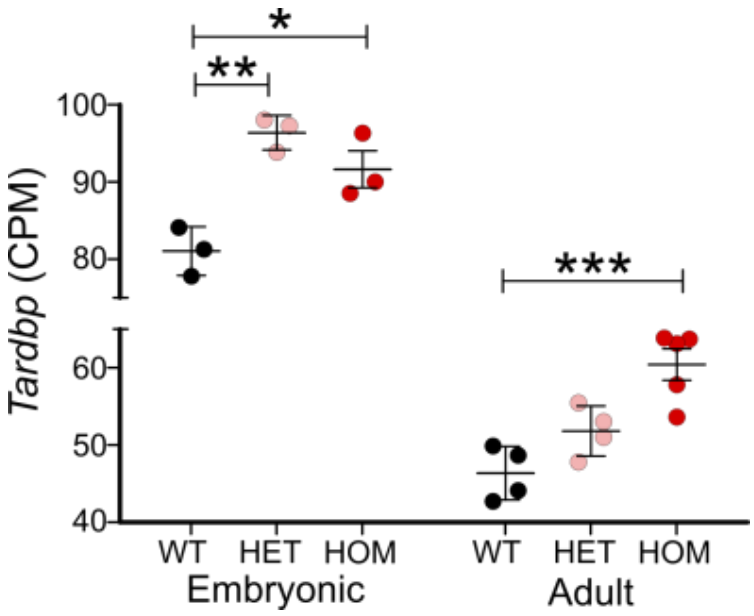


Figure 4.47. DESeq2-normalised *Tardbp* counts (counts per million) in embryonic and adult M323K RNA-Seq data in heterozygous and homozygous models in comparison to wildtype. Significant difference in *Tardbp* transcripts in both homozygous model time points upon targeted analysis.

The influence of TDP-43's autoregulatory mechanism underlying this increase in *Tardbp* expression was confirmed by splicing analysis of the *Tardbp* transcript. Intron 7 retention, was increased from a PSI of 89.9 % in sequenced *Tardbp* transcripts of wildtype mouse spinal cord to 95.6 % in the M323K mouse, resulting in a Δ PSI of 5.7 %. Incredibly, this represents a stronger shift in the autoregulatory mechanism than the 3.6 % PSI in the homozygous F210I mouse.

A point of note is the higher level of intron splicing, and therefore unstable transcript production, in the adult in comparison to the embryonic wildtype. Indeed, the increased intron 7 retention in the M323K adult mouse only reaches levels equivalent to the wildtype mouse embryo. This suggests that, provided the level of unstable *Tardbp* transcript sequencing remained constant across the two experiments, a greater proportion of mouse *Tardbp* transcripts are spliced to result in unproductive isoforms. This result tallies with previous investigations into the area, that found a significant

reduction in TDP-43 protein levels in the CNS of adult mice with no change in *Tardbp* transcript expression, implying a change in number of unproductive transcripts (Sephton et al. 2010; Huang et al. 2010). TDP-43 has been reported as particularly crucial in development (Wu et al. 2009; Kraemer et al. 2010; Sephton et al. 2010), and levels reduce with age (Huang et al. 2010; Sephton et al. 2010), but it is still of great importance in mature cells (Wu et al. 2012; Iguchi et al. 2013), and so the increase in transcripts targeted for degradation in adult versus embryonic models may aid the explanation of late onset of molecular dysregulation in the M323K models.

Concordant with the increased *Tardbp* levels, and the increased intron 7 retention, and again contradicting the initially expected downregulation of *Tardbp*, a small decrease in *Tardbp* transcripts with the long isoform 3'UTR was also observed (Figure 4.48). This isoform represents isoforms utilising pA_4, a polyadenylation site associated with transcript retention in the nucleus and no protein being translated. Therefore, a decrease in proportion of *Tardbp* transcripts with long 3'UTR isoforms would suggest an increase in TDP-43 protein production and a worsening of the gain of function effects brought about by the M323K mutation.

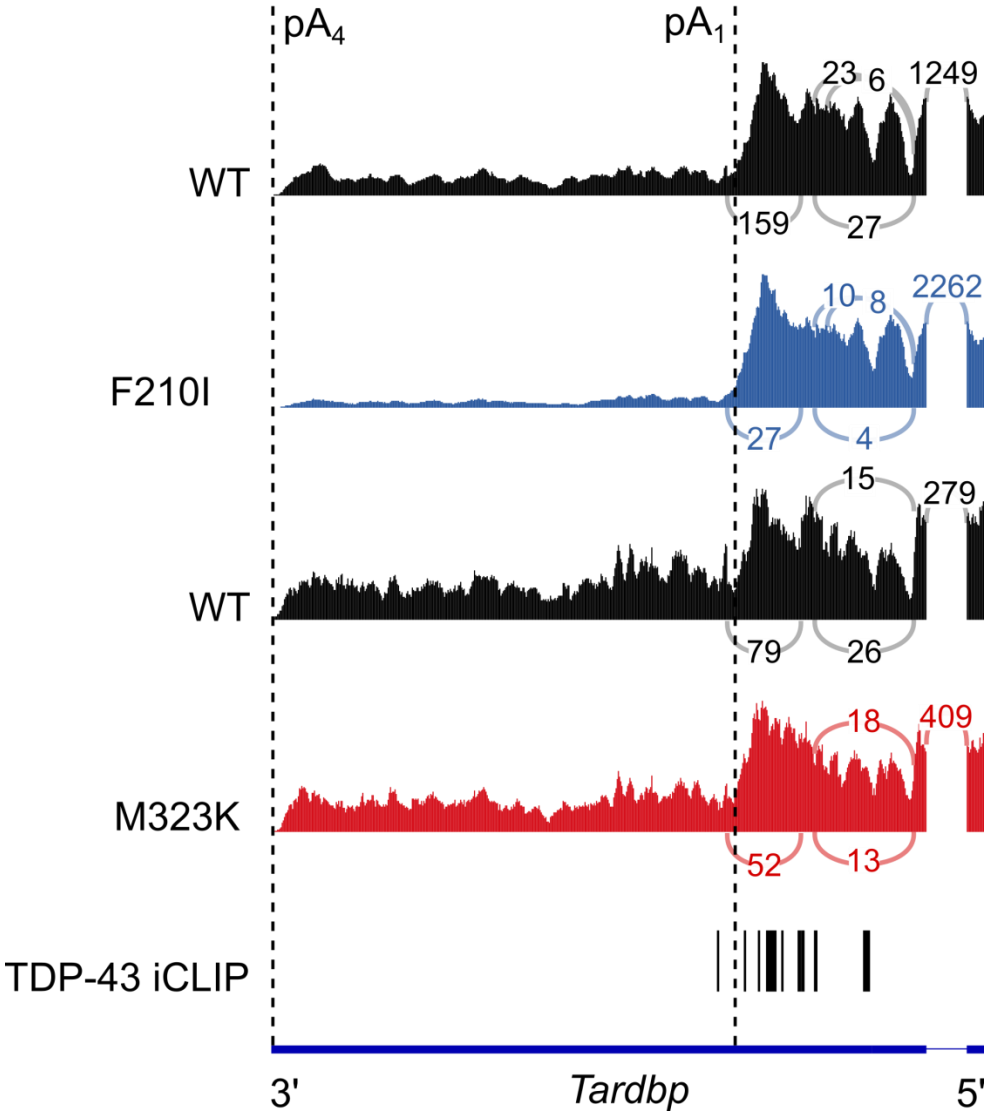


Figure 4.48. Polyadenylation sites pA_1 and pA_4 annotated. Read count tracks for the 3'UTR of *Tardbp* transcripts in embryonic wildtype and homozygous F210I RNA-Seq datasets, and adult wildtype and homozygous M323K RNA-Seq datasets. Sashimi plots show splice junction presence in each dataset. Reduction in proportional read counts between pA_1 and pA_4 in F210I and M323K mutants relative to respective wildtypes. Below are the iCLIP peaks denoting TDP-43 binding sites in the wildtype F210I iCLIP dataset showing regions of TDP-43 binding.

Given the lack of molecular dysregulation seen in the heterozygous M323K models at both the developmental and adult stages, it was thought that the autoregulatory ability

of TDP-43 was able to suppress the effects of the mutation present in just one allele. However, as was the case with the heterozygous embryonic F210I model, general differential gene expression and alternative splicing analysis revealed no changes that would indicate the presence of autoregulation. TDP-43 levels were not significantly changed, intron 7 splicing was not altered, and differences in usage of the long 3'UTR isoform were not visible. However, targeted investigation specifically into *Tardbp* expression revealed significant increases in expression levels in the heterozygous and homozygous embryonic models. This result suggests that the increase in *Tardbp* expression is not restricted to a joint-effect of the M323K mutation and aging, and that the mutation is sufficient to trigger *Tardbp* expression increase.

In the F210I model, the maintenance of high *Tardbp* levels at the RNA level to high TDP-43 protein levels was variable, and protein levels were not significantly higher in F210I mice. In a similar manner, the same was true for M323K mice. Normalised to wildtype TDP-43 levels, both the embryonic and adult M323K mice showed trends suggesting a possible increase in TDP-43, but this was highly variable and not a significant difference (Figure 4.49). The suspected reason for this is the comparatively low level of autoregulation required to compensate for the mutation in one allele culminating in minor changes in the aforementioned mechanisms that are not evident in low throughput protein analysis. TDP-43 levels are kept very stable, and so any autoregulatory change would only likely result in small changes in protein levels, even if the mutation of the protein has substantial impact on its function. Therefore, in order to identify this minor change, it is likely that higher sample sizes are required, and the 3-4 samples used per genotype in this current study is insufficient, even more so given the apparent increase in variation in levels of TDP-43 in M323K models in comparison to wildtype.

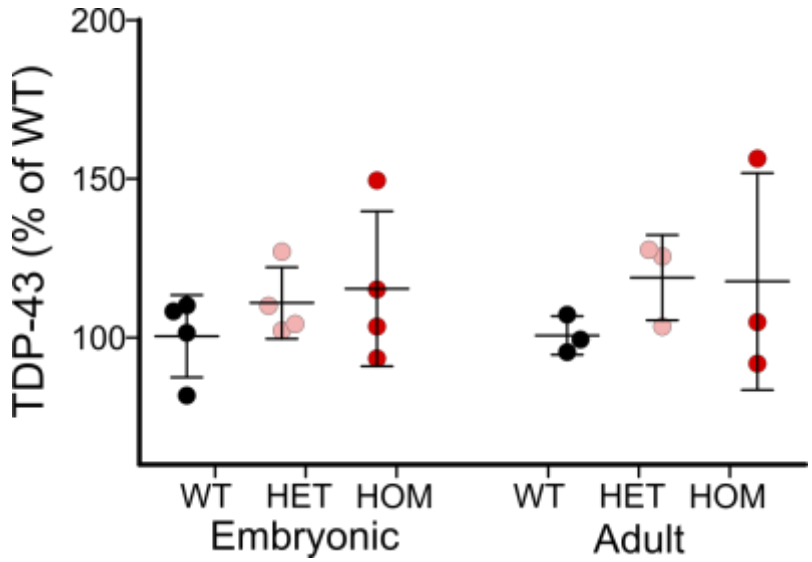


Figure 4.49. TDP-43 protein expression levels in embryonic and adult M323K models normalised to wildtype controls. Quantification from western blot experiments for TDP-43 in comparison to Actin controls. No significant change in any mutant model. Results from Nicol Birsa.

The finding that requires further clarification is the shift in autoregulation towards an increase in translatable *Tardbp* transcripts despite the gain of function effects of the M323K mutation. Whilst the underlying reason is unclear, the change does indicate a dysfunction of the ability of TDP-43 to appropriately autoregulate itself upon mutation of the LCD. This may provide the backdrop behind the increase in *Tardbp* levels, subsequent increase in cytoplasmic protein levels, and resulting protein aggregates induced by LCD mutations in TDP-ALS. Previous study of LCD mutation has shown that large changes can completely prevent the ability of TDP-43 to autoregulate itself, showing no downregulation of endogenous *Tardbp* expression or endogenous TDP-43 protein levels despite overexpression due to induced expression of mutant LCD *TARDBP* cDNA (Ayala et al. 2011). However, this mutation was a substantial deletion of a key regulatory region of TDP-43. The M323K mutation lies within this large deletion, but the single missense mutation alone was not sufficient to replicate the effect. Crucially, the altered regulation also replicates the effect of the F210I mutation in

causing increased *Tardbp* transcript levels, providing a common feature of TDP-43 dysregulation culminating from both loss of function and gain of function in TDP-43 splicing activity. The fact that this phenomenon is seen in ALS patient settings further suggests that both loss of function and gain of function of TDP-43 could play a role in disease development.

4.2.4 DIFFERENT GENE EXPRESSION IN M323K

As summarised in 4.2.1 RNA Sequencing of M323K overview, heterozygous M323K models showed minimal changes in differential gene expression at both the embryonic and adult time points. The same was true for homozygous M323K embryos, but in contrast to this in the adult homozygous M323K spinal cords 1,649 genes were found to show changes in gene expression with an adjusted p-value < 0.1 (Figure 4.50). As with the homozygous F210I model, there were broadly equal levels of gene upregulation (756 genes) and downregulation (893 genes). A level of gene expression dysregulation on par with mutation of the RBD as with the F210I mutation was unexpected, and highlights the key role the LCD also plays in the ability of TDP-43 to control expression levels of gene targets.

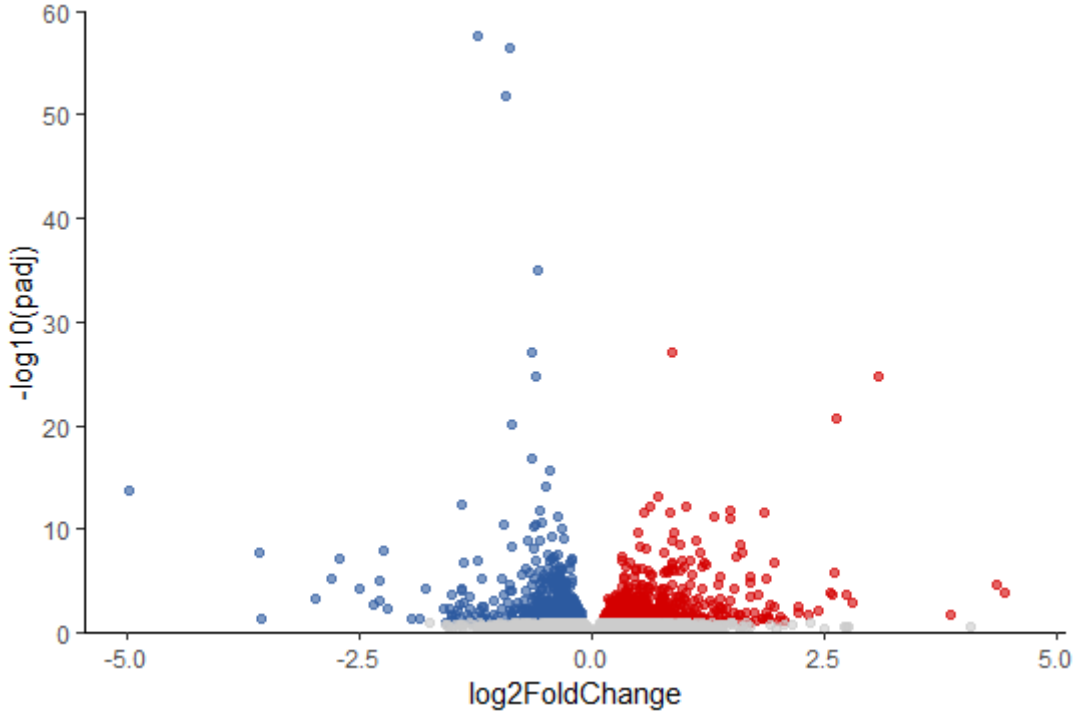


Figure 4.50. Volcano plot for differential gene expression changes in homozygous M323K spinal cord RNA-Seq dataset vs wildtype. Each dot represents one gene. Substantial significant changes in gene expression levels. Non-significant changes in grey, significant upregulation in red, significant downregulation in blue.

Unlike the F210I gene expression changes, which clustered around pathways involved in development and ion transport, the mass changes in the M323K adult spinal cord were not enriched for any pathways in the GO, KEGG or Reactome databases. Analysis by Jack Humphrey further revealed that the specific downregulation of long genes detected in the F210I dataset was also absent in M323K mice (Figure 4.51).

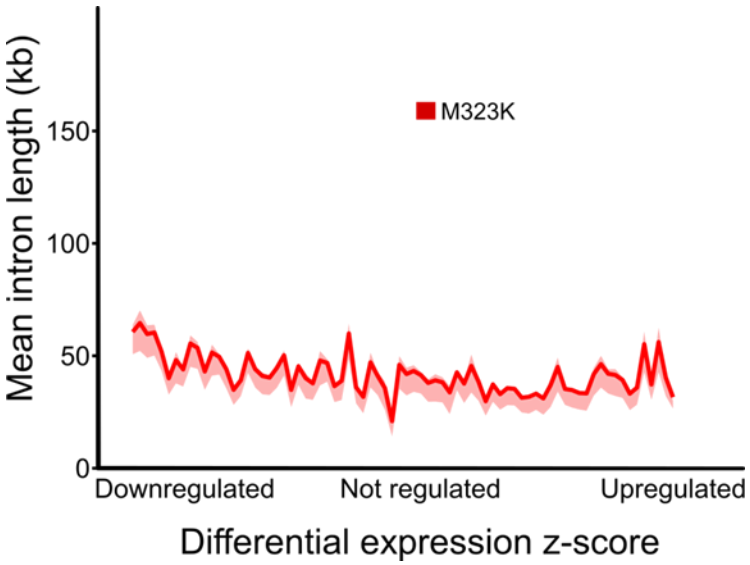


Figure 4.51. Comparison of mean intron lengths in genes binned by level of dysregulation in the M323K adult spinal cord RNA-Seq dataset performed by Jack Humphrey. No change in mean intron length in relation to changes in gene expression in the M323K dataset compared to wildtype.

Despite both the homozygous F210I and M323K models showing expression changes in 15-20 % of the expressed transcriptome, the overlap between datasets was only 322 genes. This is particularly surprising considering the target motif of TDP-43, UG repeat sequences, was not changed in the two models. They also showed strong correlation in iCLIP pentamer enrichment, suggesting the two mutant TDP-43 variants still bound to the same regions. The reason for the relatively low level of consistency in genes dysregulated in the two models is not clear, but a number of variables may contribute towards it. A key factor is will be the difference in model age, with substantial differences in gene expression likely to occur between embryonic development and mature adult CNS tissues independent of TDP-43 activity. The locations of the mutations is also likely to play a role, but segregating the specific effects of these is difficult, especially considering the mystery around how TDP-43 controls gene expression of targets in general. Finally, comparison between TDP-43 loss and

overexpression effects on gene expression regulation in flies has previously shown low levels of overlap in genes dysregulated by the two conditions (Hazelett et al. 2012).

4.2.5 ALTERNATIVE SPLICING IN M323K

Similarly to the F210I model, the homozygous M323K mouse RNA-Seq analysis revealed several differentially spliced genes alongside the changes in gene expression. 295 splicing events were significantly changed in total, and of these, 98 showed an increase in exon or retained intron inclusion, compared to 127 with an increase in exon or retained intron skipping. However, opposite to the F210I, in the M323K model the largest Δ PSI values were exon skipping changes (Figure 4.52).

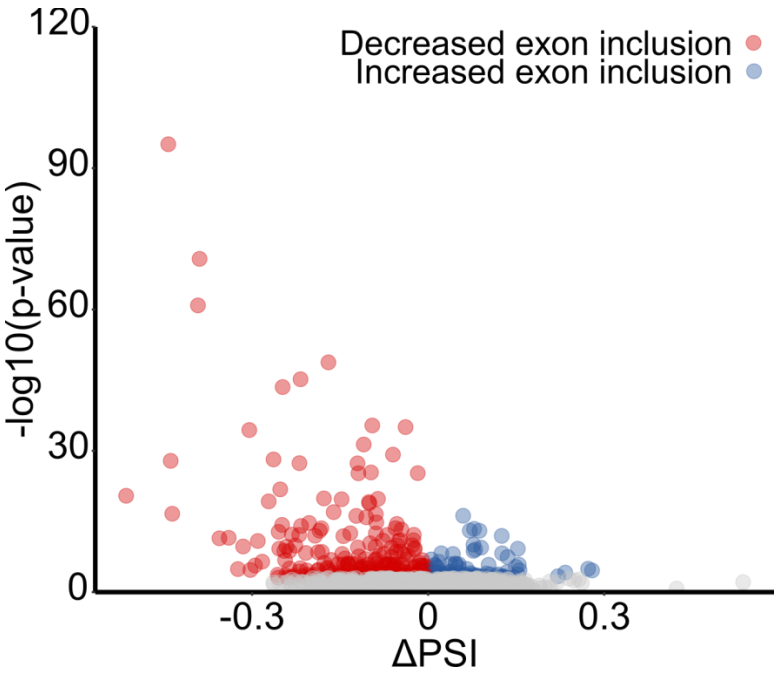


Figure 4.52. Volcano plot for cassette exon alternative splicing changes in homozygous M323K spinal cord RNA-Seq dataset vs wildtype. Each dot represents one splicing event. Larger magnitude of change seen in exon skipping events in comparison to exon inclusion. Non-significant changes in grey, significant decrease in exon inclusion in red, significant increase in exon inclusion in blue.

Alongside the cassette exons and retained introns, 30 splicing events were shifted towards usage of proximal splice sites, and 14 showed increase in distal splice sites. A small number of events were then characterised as alternative first exons, alternative last exons or mutually exclusive exons. Curiously, exactly the same number of changed alternative splicing events was present in genes that also showed significant dysregulation, 74, in the homozygous M323K and F210I models. Once again, both adult heterozygous models and the embryonic homozygous M323K model all showed minimal changes in alternative splicing.

Although the differences are not great, the adult homozygous M323K spinal cord shows reduced gene expression changes but more alternative splicing changes than the embryonic homozygous F210I forebrain. Although many variables are present, this difference may be partially explained by the different domains of TDP-43 the two mutations are present in. This is particularly true for the increase splicing changes in the M323K, where the mutation lies in a region critical for interaction with other hnRNPs and other splicing factors that work concurrently with TDP-43 to control splicing of nascent transcripts (D'Ambrogio et al. 2009). This mutation may inhibit the interaction with these splicing factors, resulting in the greater levels of splicing dysregulation. Currently, it is not known which protein interactions would be affected by the M323K mutation in TDP-43, but the spliceosome in general is thought to be defective in ALS settings (Tsuiji et al. 2013).

The idea that the M323K mutation caused a gain of function in TDP-43 splicing activity was based on the low throughput RT-PCR experiments on known TDP-43 splicing targets of *Sort1*, *Eif4h* and *Dnajc5*, which showed that while the F210I mutation caused increased exon inclusion comparable with TDP-43 KD, M323K TDP-43 resulted in converse increases in exon skipping. Having obtained RNA-Seq datasets for both the TDP-43 loss of function and gain of function mouse models, TDP-43 transcript targets

could be compared on a large scale, to analyse whether the opposite effects of the two mutations held true throughout the transcriptome. First we validated that the splicing events detected as opposite by RT-PCR held true in the RNA-Seq data (Figure 4.53). Interestingly, previous study of *Eif4h* alternative splicing, known to be controlled by TDP-43, had demonstrated that increased exon inclusion of the regulated exon occurred upon TDP-43 loss (Yang et al. 2014). This finding was replicated by both the adult mouse TDP-43 KD data (Polymenidou et al. 2011), and our F210I loss of function model. However, the previous study found no effect of TDP-43 overexpression on alternative splicing of this *Eif4h* exon. In our M323K data, the increase in exon skipping was confirmed by both RT-PCR and in RNA-Seq analysis. This would be expected from a gain of function in TDP-43 splicing, and the discrepancy to the overexpression study may be due to interfering variables arising from TDP-43 overexpression; such variation is minimised in our gain of function model with endogenous gene mutation.

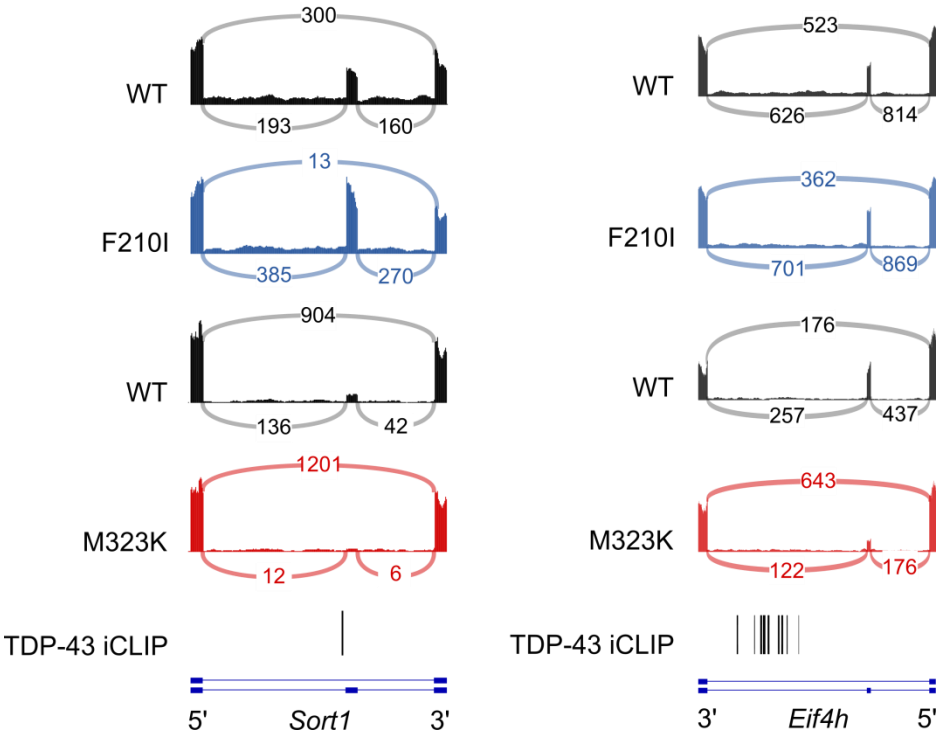


Figure 4.53. RNA-Seq read counts and overlaid splicing sashimi plots for *Sort1* and *Eif4h* alternative splicing in homozygous F210I and M323K datasets with

respective wildtype controls. Embryonic wildtype iCLIP binding displayed underneath RNA-Seq reads. Increased exon inclusion in F210I model versus embryonic wildtype, and increased exon skipping in M323K model versus adult wildtype, in both *Sort1* and *Eif4h* alternative splicing.

Also incorporating alternative splicing data from the TDP-43 KD dataset in adult mouse striatum (Polymenidou et al. 2011), several other key splicing events were identified to be changed in opposite directions in TDP-43 loss and gain of function (Figure 4.54). These added further weight to the hypothesis that the M323K caused gain of function in TDP-43 splicing activity across a number of regulated splicing events in the transcriptome.

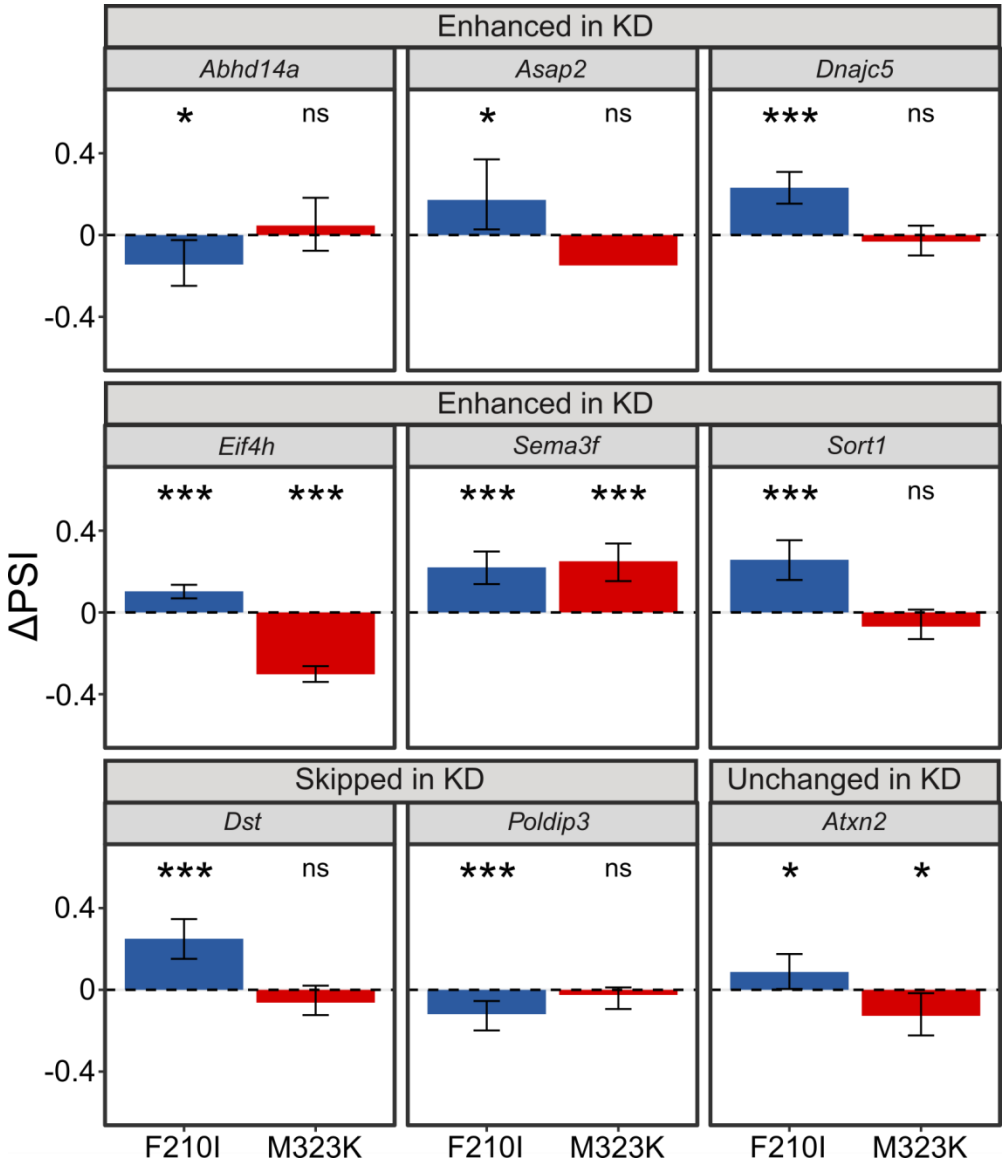


Figure 4.54. Alternative splicing of the same splicing event measured by Δ PSI across homozygous embryonic F210I and homozygous adult M323K RNA-Seq in key TDP-43 splicing targets, with banners above the graphs identifying how the splicing event was changed in TDP-43 KD in adult mouse brain (Polymenidou et al. 2011).

Alongside the opposite direction in splicing events in certain genes, a small number of genes also contained splicing events which were strongly changed in the same direction in both the F210I and the M323K models, such as *Sema3f*. It is likely that

these changes rely on a role of TDP-43 control of splicing distinct from the majority of targets.

As was the case with the splicing changes in the homozygous F210I forebrain, no GO term enrichment was found in the alternative splicing analysis for the homozygous M323K spinal cord. This may be due to the widespread role of TDP-43 in regulating transcripts from a number of different varied pathways, and that no one pathway is specifically the root connection between the splicing dysregulation and subsequent disease pathology.

4.2.6 SKIPTIC EXONS

The F210I mutation resulted in a loss of repression of several intronic splice sites, resulting in the inclusion of a number of cryptic exons in the mutant mouse model which were completely absent from the wildtype. Given that this was a loss of function phenotype, similar cryptic exon splicing in the M323K model was expected to be absent. In concordance with this, cryptic exon presence in the M323K model was scarce, with no instances of cryptic splicing present in validated F210I cryptic exon transcripts such as *Adnp2* (Figure 4.55). Only one cryptic exon from the F210I model was also found in M323K, a splicing event in *Reep3*. As with the rare splicing events modulated in the same direction by the two mutations, it is likely that the regulation of *Reep3* splicing involves a separate mechanism to the bulk of splicing changes analysed here.

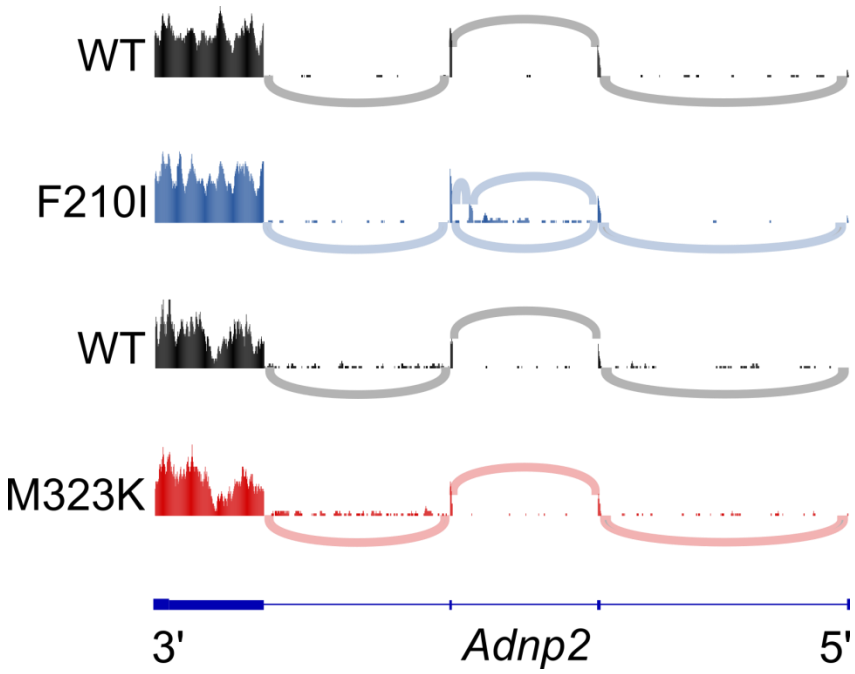


Figure 4.55. RNA-Seq read counts from homozygous F210I and M323K and respective wildtypes in the 3' end of *Adnp2*. Overlaid sashimi plot showing splicing patterns in each dataset. Cryptic exon in *Adnp2* gene found in F210I model, but absent from both wildtypes and also the homozygous M323K model.

Whilst the absent of cryptic exons was predicted, the M323K mutation still showed large scale splicing changes, and so the presence of extreme splicing events resulting in transcripts specific to the mutant model was possible. TDP-43 often functions to repress exon inclusion, and loss of function through F210I mutation or TDP-43 KD lead to increased inclusion levels. Considering this, we examined the alternative splicing output of homozygous M323K models for extreme skipping events. These would be the opposite of cryptic exons, instead showing skipping of supposedly constitutively spliced exons, again resulting in novel transcripts in the M323K model but not wildtype. We adopted thresholds mirroring those set for cryptic exons, looking for alternative splicing events with an adjusted p-value of < 0.01, an average control PSI of greater than 0.95, and a Δ PSI of at least 0.05. This meant these extreme skipping events would be present in less than 5 % of wildtype transcripts, and show a significant

increase in skipping of 5 % or more in the M323K mice. Using these filters, we discovered 47 of these extreme skipping events in 44 genes, with an example in *Herc2* (Figure 4.56). Taking into account the opposite in effects to cryptic exons, and the skipping nature of the splicing, we termed these events 'skiptic exons' in further analysis.

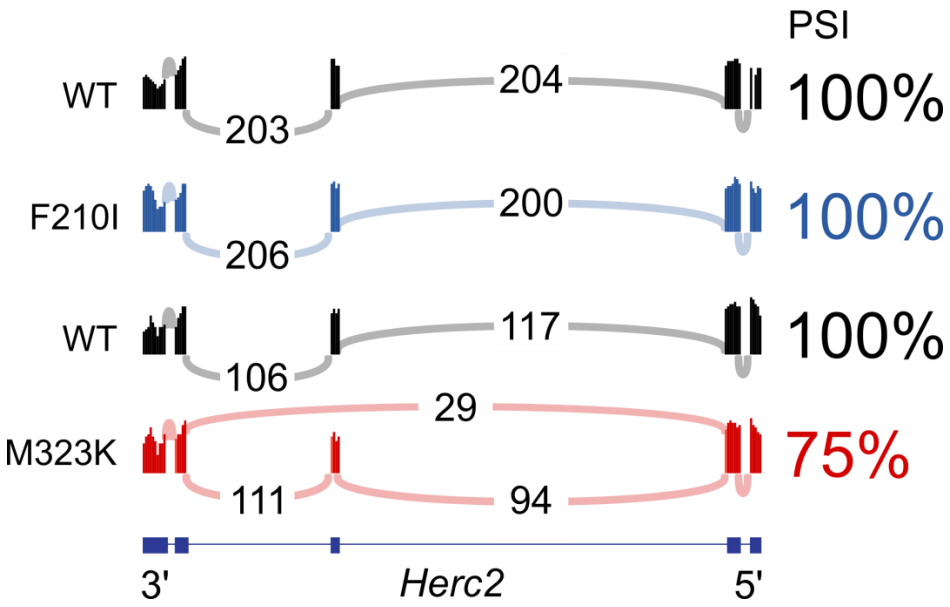


Figure 4.56. RNA-Seq read counts from homozygous F210I and M323K datasets along with respective wildtypes for part of the *Herc2* gene aligned to a reference. Overlaid with sashimi plot showing splicing patterns in each dataset, with number of each splice event annotated. PSI levels for the skiptic exon are calculated and listed to the right. No detected instance of alternatively spliced exon skipping in wildtype samples or the F210I dataset, but a quarter of all transcripts showed exon skipping in the M323K dataset.

A select number of these skiptic exons were then validated by RT-PCR, with all showing significant and substantial decreases in inclusion levels in the M323K model (Figure 4.57).

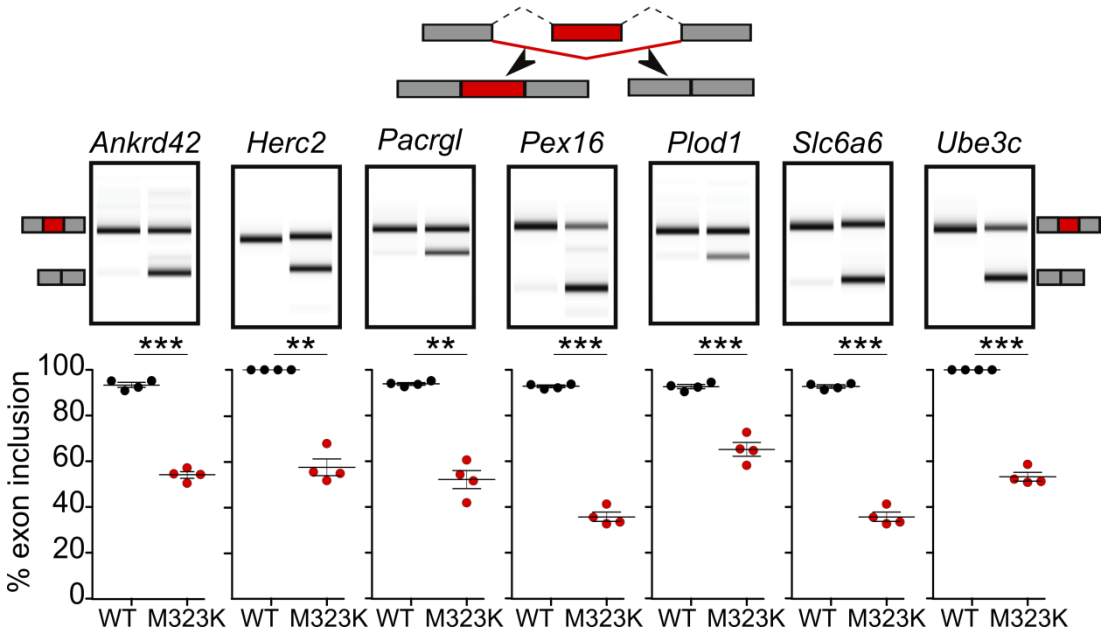


Figure 4.57. Top: Diagram of two possible isoforms from RT-PCR test, with skiptic exon denoted in red. Middle: Example RT-PCR results visualised on Tapestation gels for seven targets, with skiptic exon-containing transcripts and wildtype transcripts labelled. Skiptic exon transcripts only clearly evident in M323K samples. Bottom: Graphical and statistical analysis of skiptic exon inclusion levels for each target in wildtype and M323K samples.

As the phenomenon of skiptic exons was linked to the opposite effect of TDP-43 loss of function, the homozygous F210I embryonic dataset was reanalysed to investigate the presence of skiptic exons. In keeping with this hypothesis, fewer splicing events meeting skiptic exon criteria were detected in this F210I dataset, further suggesting that they were specifically linked to the gain of function effects of the M323K mutation (Figure 4.58). As with the *Reep3* cryptic exon being the only instance present in both models, one skiptic exon was detected in both datasets too, within the *Tsn* gene. Again, this outlier is likely to result from a different method of TDP-43 regulation to the majority of skiptic exon splicing events. In general, the contrasting presence of cryptic and skiptic exons not only further demonstrated the opposite effects of F210I and M323K mutations on general TDP-43 splicing activity in the transcriptome, but also

highlighted similarly opposite roles in extreme splicing events leading to novel transcripts in the mutant models.

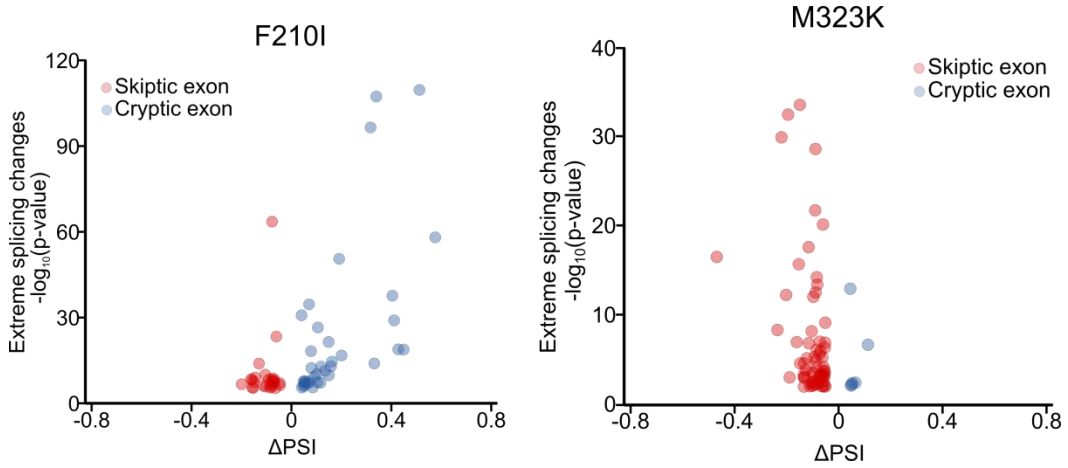


Figure 4.58. Δ PSI and level of significance for each extreme splicing event (cryptic or skiptic exon) present in the homozygous F210I embryonic or homozygous M323K adult RNA-Seq datasets vs respective wildtype. Cryptic exons typically seen in the loss of function F210I model, skiptic exons seen more in the gain of function M323K model.

A point of note is that alongside the common cryptic exon in *Reep3*, and common skiptic exon in *Tsn*, there was a cryptic exon in the embryonic F210I model, in *Ppfibp1*, in which the exon then became well expressed in adult wildtype mouse spinal cord and actually showed increased levels of skipping in the adult M323K spinal cord. Mirroring this, a skiptic exon in the adult M323K model, in *Klhdc10*, showed high levels of skipping in the embryonic wildtype forebrain, and even increased levels of inclusion in the embryonic F210I forebrain. These unique splicing events are unexplained, and highlight the complexity of splicing activity regulated by TDP-43, much of which cannot be simply explained by opposite effects in TDP-43 loss and gain of function, alongside some of the difficulties in comparing and contrasting mouse models at different ages.

Given the mirrored effects of F210I and M323K mutations for even the extreme splicing events, we further analysed respective F210I and M323K iCLIP data to understand how TDP-43 binding differed to result in these different effects. Proximal versus distal binding of TDP-43 to target splicing regions is known to influence level of exon inclusion/repression (Tollervey et al. 2011). Analysis by Jack Humphrey in F210I TDP-43 iCLIP clusters around cryptic exons in the F210I RNA-Seq showed marked similarity to M323K TDP-43 iCLIP clusters around skiptic exons in the M323K RNA-Seq (Figure 4.59). This suggested that both extreme splicing varieties culminated from similar binding proximity of TDP-43 to the alternatively spliced exon, and strengthened the hypothesis that the F210I and M323K mutations changed the results of TDP-43 splicing specifically through a loss and gain of function in activity respectively.

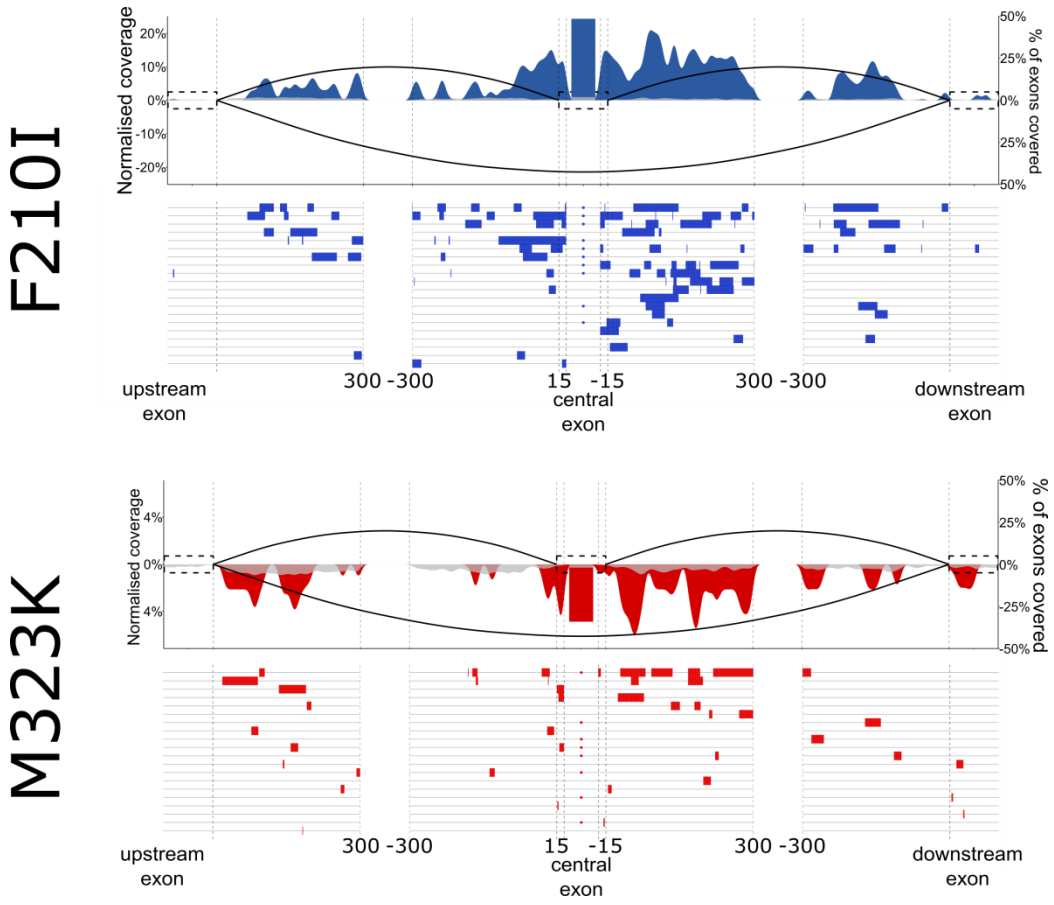


Figure 4.59. RNA maps show binding distribution of respective TDP-43 iCLIP clusters (positive and negative values on y-axis for increased and decreased ΔPSI

respectively) in cryptic exons in F210I (A) and skiptic exons in M323K (B); red and blue indicate when cluster coverage is increased above background set of non-regulated cassette exons (grey). Below, binding sites for the top 20 most significant cryptic exons (A) and skiptic exons (B) are plotted. Results from Jack Humphrey.

Cryptic exons present in the F210I model were also predicted to cause NMD of the transcripts. However, differential gene expression analysis of cryptic exon transcripts showed equal levels of upregulation and downregulation, questioning the ability of the NMD mechanism to degrade the transcripts as predicted, and also raising the possibility of other regulatory mechanisms at play to compensate for the increase in unstable transcripts. Similar analysis was conducted on the skiptic exons by Jack Humphrey, and revealed that significantly downregulated genes were particularly enriched for skiptic exons, which was not true for upregulated genes (Figure 4.60). Furthermore, these downregulated genes contained a significantly greater proportion of skiptic exons which were predicted to trigger NMD. These findings together implicated skiptic exons in causing transcript degradation and downregulation.

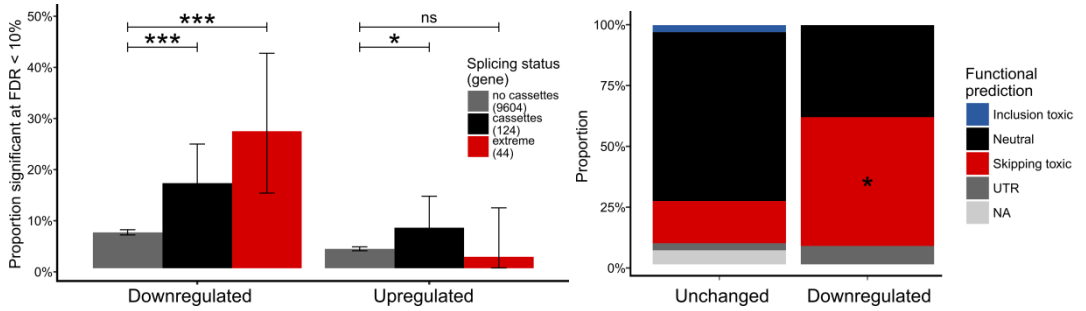


Figure 4.60. (Left) Dysregulated genes split by direction of regulation. Subsets then further split into those containing no cassette exons (grey), cassette exons (black), and cassette exons defined as skiptic exons (red). Downregulated genes significantly enriched most for skiptic exons. (Right) Downregulated or unchanged genes with extreme splicing events split by type of extreme splicing

event contained. Toxic NMD-predicted skiptic exons enriched in downregulated genes compared to unchanged genes. Results from Jack Humphrey.

Skiptic exons were demonstrated to be linked to gain of function in TDP-43 splicing, resulting in novel transcripts that were dysregulated and predicted to be degraded as a result. This was similar to the effect of cryptic exons occurring from TDP-43 loss of function. However, one key difference is that while cryptic exons occur in typically intronic regions and show low levels of conservation between mice and humans (Ling et al. 2015), skiptic exons are typically constitutively expressed exons and therefore should generally be very well conserved. Analysis by Jack Humphrey confirmed this, with cryptic exons from the F210I dataset showing noticeably low PhyloP conservation scores, whereas M323K skiptic exons were very well conserved (Figure 4.61).

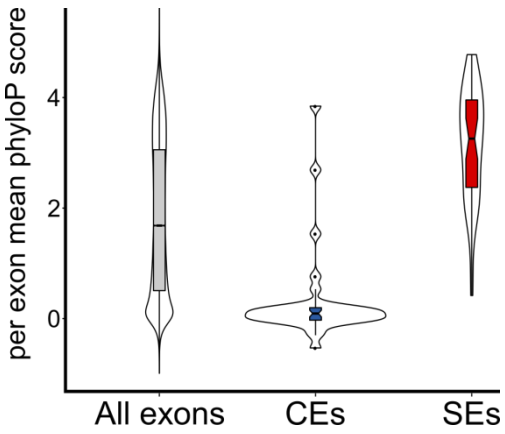


Figure 4.61. Per-exon mean PhyloP conservation scores of F210I cryptic exons (CEs) and M323K skiptic exons (SEs) in comparison to all annotated mouse exons. Violin plots showing distribution of data with overlaid boxplots to show the median and quartiles. Outliers are plotted as black dots. Cryptic exons show low conservation levels, whereas skiptic exons show high levels of conservation. Results from Jack Humphrey.

The difference in conservation level is crucial, given the aim of using mouse models to enable therapeutics for disease in patients, an aim for which models of

neurodegenerative disease have a particularly poor track record (Eaton & Wishart 2017). Importantly, although the high level of skiptic exon conservation suggests that a similar splicing mechanism could take place were gain of function in TDP-43 splicing to occur in human settings, the mechanism by which TDP-43 mediates this splicing is often through binding to intronic regions around the exon rather than the exon itself, as shown by M323K iCLIP analysis specific to skiptic exons (Figure 4.59). To enable a similar function in humans, the introns surrounding the skiptic exons could also be conserved. Such conservation of non-coding regions of genes often provide indication of regulatory importance of the sequence, notably for the autoregulation of TDP-43 itself at the TDPBR found within intron 7 of the *Tardbp* transcript (Ayala et al. 2011). To test this possibility, the conservation levels of intron regions around the skiptic exons between mouse and humans were also investigated. There is no consensus for quantification of sequence alignment, and several methods exist (Raghava & Barton 2006). Variables include incorporating gaps in sequences, and the total length of sequence to average over. The approach used in this investigation calculated percent sequence identity by dividing the number of identical matches in sequence by the total number of aligned positions plus any internal gaps. To determine whether the introns were conserved over and above levels for typical exons, introns flanking constitutive exons which were always expressed in both control and mutant datasets were used as controls. Using this method, it was determined that the majority of introns surrounding skiptic exons were substantially more conserved than the introns flanking the control constitutive exons. (Figure 4.62).

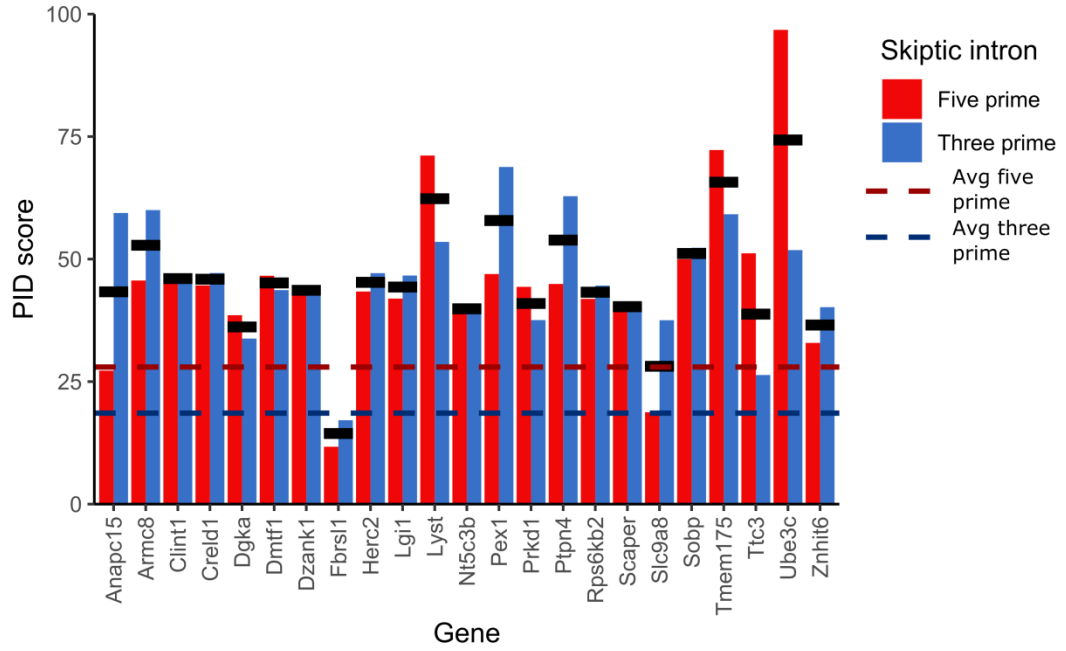


Figure 4.62. Percentage identity sequence (PID) plotted for all aligned bases in mouse and human 5' and 3' introns for conserved skiptic exons. PID calculated by: 100 * (identical positions) / (aligned positions + internal gap positions). Black crossbar indicates mean intron PID score for each gene. Dashed lines indicate PID scores for introns around 100 most constitutively expressed exons (PSI > 0.99 in both wildtype and homozygous M323K RNA-Seq datasets, FDR > 0.1) 5' (red) and 3' (blue). Most introns surrounding skiptic exons substantially more conserved than introns surrounding constitutively expressed exons.

Another important measure prior to investigating the presence of skiptic exons in humans was the skiptic exon presence in heterozygous models. TDP-ALS mutations are typically autosomal dominant, and so often the disease stems from a heterozygous genotype. As previously discussed, much of the molecular dysregulation seen in the homozygous M323K model was not present in heterozygous models to a significant extent, and this was also true for most of the skiptic exons identified. However, the lack of significance does not necessarily exclude their presence – they may occur but not sufficiently to pass confidence thresholds. If this was the case, then we would expect to

see skipping in the same splicing events, but lower Δ PSI levels. Analysis of the adult heterozygous mice confirmed exactly this, all but one of the detected skiptic exons also showed increased skipping, and all but one these showed a lower Δ PSI in comparison homozygous counterparts (Figure 4.63).

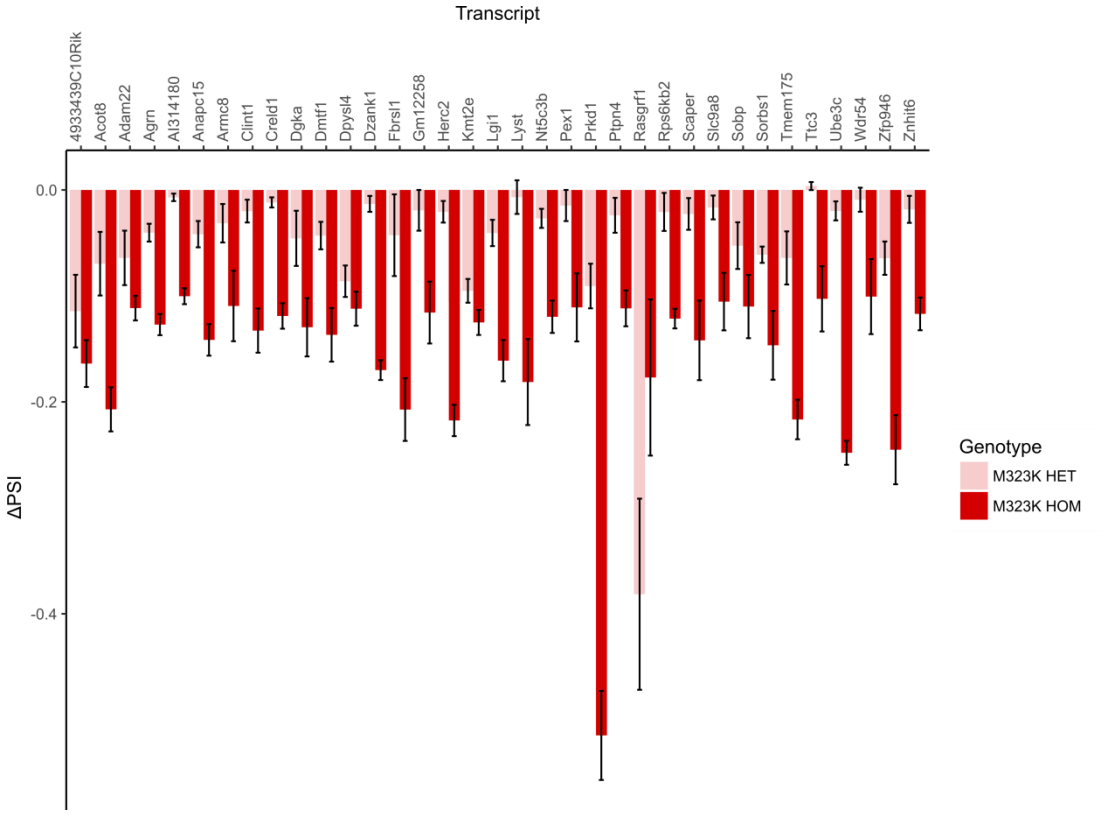


Figure 4.63. Δ PSI levels for skiptic exons from the homozygous adult M323K dataset also shown for heterozygous M323K adult mice vs respective wildtypes.

Intriguingly, even within the relatively small subset of 47 skiptic exons, there seemed to be an enrichment of genes involved in the ubiquitin pathway, in particular E3 ubiquitin ligases. Six genes – *Herc2, Ube3c, Ube2e2, Wwp1, Ttc3 and Wsb1* – fit this description. E3 ubiquitin ligases function to either directly catalyse the ubiquitination of a protein, or as adaptors to facilitate the ubiquitination (Berndsen & Wolberger 2014). Quantifying the precise enrichment of this subset of ubiquitin ligase genes containing skiptic exons is difficult, as the label describes a diverse group of proteins and the exact number with this function is very unclear (Ardley & Robinson 2005).

However, the link between extreme splicing events occurring in genes related to ubiquitin pathway function, and the subsequent ubiquitin pathology developed by the homozygous M323K mouse at the 2 year time point (4.2.8 M323K phenotype), cannot be ignored. Further work is required to examine this group of E3 ubiquitin ligases for further indication of RNA dysregulation, and to dissect precisely how the splicing changes in these genes could result in ubiquitin pathology.

Given the link between skiptic exons and LCD mutation of TDP-43-mediated gain of function, we suspected that similar splicing could occur in mouse models with mutations found in TDP-ALS patients. This was confirmed in the Q331K mouse model, in which three of the validated skiptic exons were found to show increased skipping in the mutant models versus wildtype. This is further discussed in 5.2 Q331K mouse model. The high level of skiptic exon conservation between mice and humans of both the exons and surrounding introns, and their reduced presence in heterozygous TDP-43 mutation, meant that it seemed entirely possible that similar splicing events would also occur in TDP-ALS patients. This theory was confirmed in fibroblasts from TDP-ALS patients, in which two of the validated skiptic exons, in *PLOD1* and *SLC6A6*, were found to be increasingly skipped in TDP-ALS patients compared to controls. This result is discussed further in 5.4 TDP-ALS patient fibroblasts.

4.2.7 POLYADENYLATION CHANGES IN M323K

Although TDP-43 typically binds to intronic regions overall, in the cytoplasm TDP-43 is well known to bind to the 3'UTR of transcripts (Colombrita et al. 2012), suggesting it plays a role in the post-transcriptional regulation of the transcript too, likely controlling its translocation in the cytoplasm (Bramham & Wells 2007). We hypothesised that the gain of function effects of the M323K mutation may also influence the polyadenylation of TDP-43 target transcripts, perhaps through

controlling alternative splicing of cassette exons of the 3'UTR itself indirectly resulting in use of different polyadenylation sites, as is this case with the TDP-43 autoregulatory mechanism with *Tardbp* transcripts itself.

To investigate this, the alternative splicing output from the RNA-Seq experiment was of limited used. This is because alternative polyadenylation site usage cannot necessarily be inferred by splicing events, and the RNA-Seq data spreads reads across the entire transcript, resulting in a limited number of reads being allocated to polyA sequences. We suspected these reads alone would be insufficient for study of alternative polyadenylation site usage. Therefore, we conducted QuantSeq on adult M323K homozygous mouse spinal cord, using the same RNA used for the initial RNA-Seq experiment. QuantSeq is a technique which focuses specifically on sequencing the 3' end of transcripts, and therefore a much greater portion of reads can be used for study of polyA site usage. Processing of the QuantSeq data was conducted by Gregor Rot, using his previously published pipeline (Rot et al. 2017). Results from the QuantSeq experiment suggested alternative polyadenylation was occurring, but determining statistical differences between genotypes proved difficult. In order to obtain the most evident polyadenylation site usage changes, RNA-Seq data was also incorporated using visual analysis of cumulative reads, enabling identification of regions in which polyA site usage is so strongly changed that a noticeably different number of reads mapped to the respective isoforms resulting from the change. An example of this can be seen in the *Brwd1* gene in Figure 4.64.

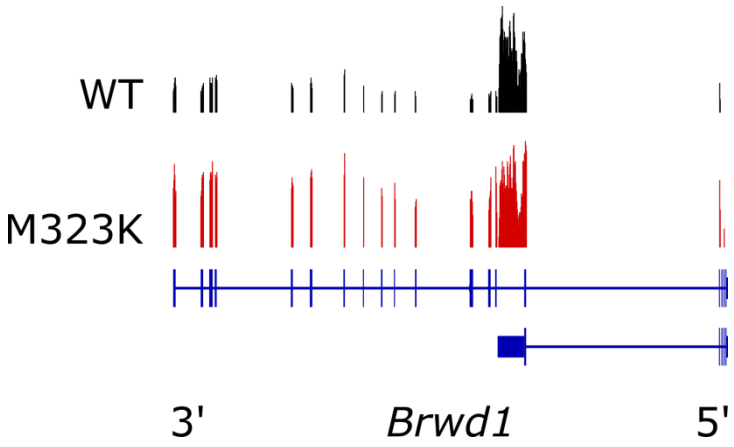


Figure 4.64. RNA-Seq reads for the 5' end of *Brwd1* transcripts in homozygous adult M323K spinal cord and wildtype. Two isoforms of *Brwd1* annotated below. Summed read peaks normalised to height of reads for 3'UTR of short *Brwd1* isoform. Reads 5' of this are aligned to the longer isoform, and there is a noticeable visual reduction in reads in wildtype compared to M323K data.

By combining the QuantSeq and RNA-Seq data, a subset of ten targets were identified to show differences in polyadenylation. Four of these showed increased use of a proximal polyA site in the M323K models compared to wildtype in the RNA-Seq data, and another six showed preferential use of a more distal polyA site. In the QuantSeq data, these changes were broadly reflected, with the occasional contrary result in one example (Figure 4.65).

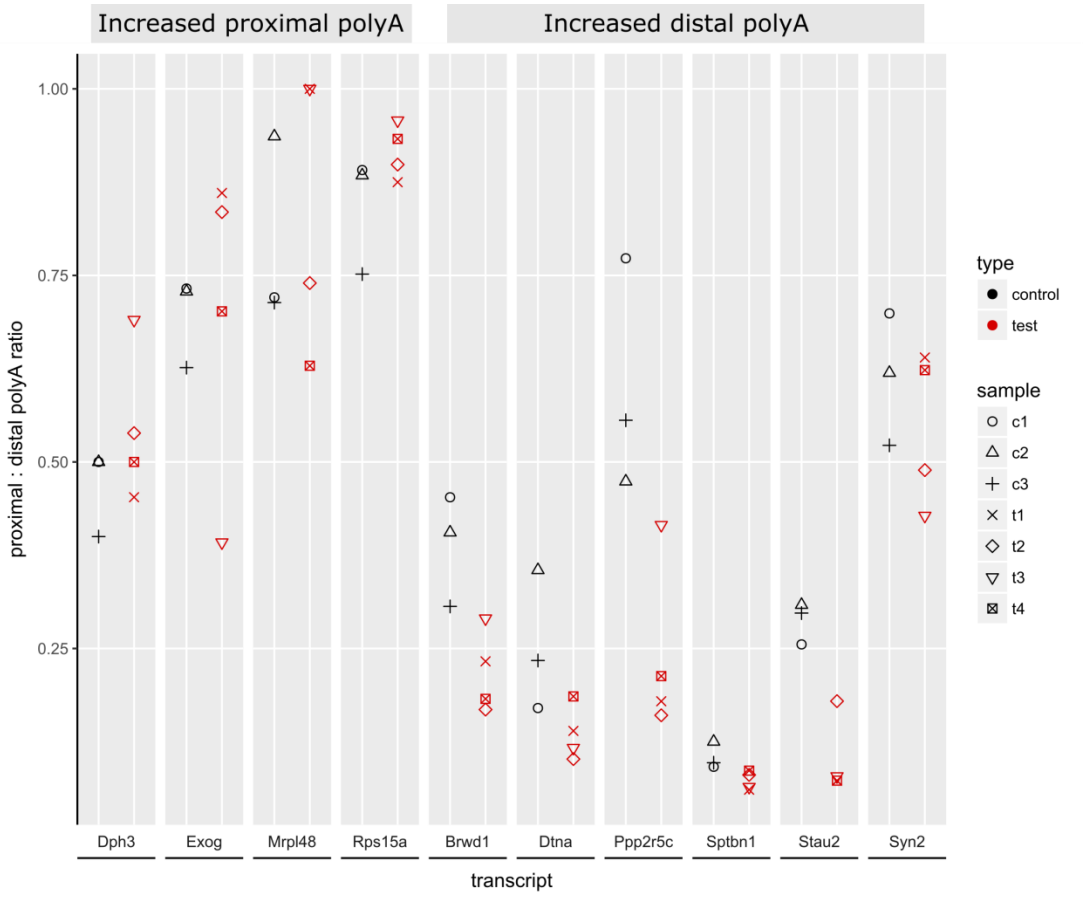


Figure 4.65. QuantSeq ratios for proximal vs distal polyA site usage in ten targets identified as likely showing alternative polyadenylation in the homozygous M323K RNA-Seq data. Direction of change in RNA-Seq data labelled at the top. QuantSeq data points for each sample shown to display no different effects caused by one sample. QuantSeq data broadly matches RNA-Seq data for the predicted changes in polyA site usage in the M323K model in comparison to wildtype.

To ensure the polyadenylation changes were true, validation was attempted for the selected targets. Validation was performed as outlined in 2.1.4 Polymerase chain reaction, using the three primer method. Although this would not strictly confirm polyadenylation differences, it would confirm the 3'UTR usage changes that would result. Using this method, the ratio of isoforms using the long 3'UTR isoform, and therefore the distal polyA site, to total isoforms with any 3'UTR would be compared.

However, the attempted validation failed to prove the 3'UTR use changes seen in the sequencing datasets. Of the ten attempted validations, two targets failed to produce PCR products as expected. Another two provided too many isoforms to enable quantification. Of the remaining six, two were expected to show increased proximal polyA site usage, but upon testing showed no changes. Similarly, three of the four targets thought to show increased distal polyA site usage failed to replicate this upon testing with PCR. Only the change in *Ppp2r5c* showed the direction of change as expected (Figure 4.66).

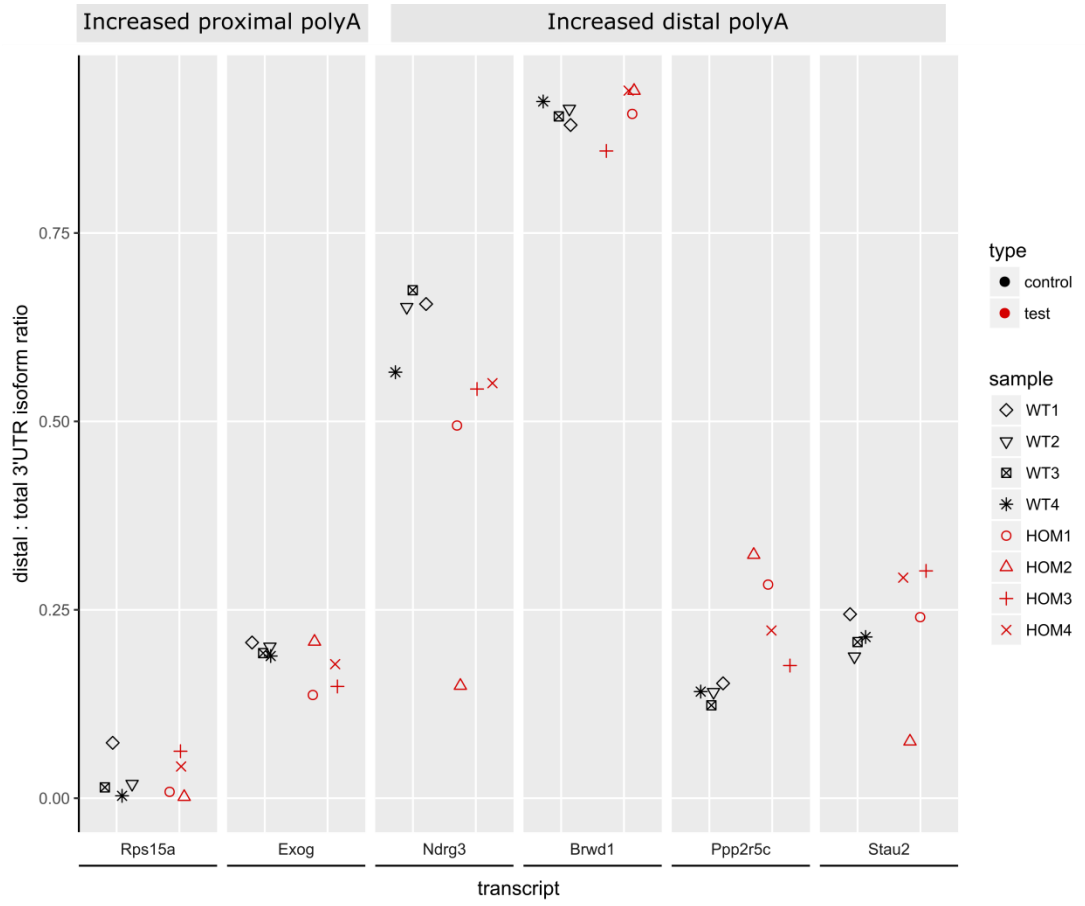


Figure 4.66. Validation of 3'UTR isoform use changes in the M323K model. Ratios for distal vs total polyA site usage, inferred from 3'UTR isoform measurement, in six targets identified as likely showing alternative polyadenylation in the homozygous M323K sequencing data. Direction of expected change labelled at

the top. Validation failed to replicate the expected changes in all but one target, *Ppp2r5c*.

Despite one target seemingly successfully replicated upon PCR validation, the general failure of this method raised several questions. The inference of alternative polyA site usage from RNA-Seq summed read counts is not ideal, as this is an indirect measure seeing the effects of the supposed change. The QuantSeq analysis, which is a direct measure, requires further refinement before polyadenylation changes can be adequately and reliably detected. The validation method employed also relies on measurement of distal versus total 3'UTR expression, rather than proximal versus distal. Although the latter ratio can be inferred, the method likely requires validation in settings where the change is more confidently known before attempted measurements of differential 3'UTR isoform usage. It seems likely that 3'UTR and polyadenylation changes are indeed occurring in the homozygous M323K adult mouse spinal cord, but further work is required to identify and validate these changes appropriately. It is an important avenue of research in neurodegenerative models especially, as in neurons the systems regulating mRNA transcript trafficking and translation are vital for conferring the expression of the transcripts in axonal and dendritic compartments (Bramham & Wells 2007). TDP-43 is known to regulate this, particularly translation through alternative 3'UTR usage (Wang et al. 2008; Liu-Yesucevitz et al. 2011; Alami et al. 2014). TDP-ALS mutations have been demonstrated to disrupt the transport of these transcripts in axons (Alami et al. 2014), it remains to be seen whether the M323K mutation results in similar effects.

4.2.8 M323K PHENOTYPE

The homozygous M323K mouse lived until adulthood with no change in lifespan, and this enabled study of pathology, and motor and behavioural phenotype, which was not possible in the embryonic homozygous F210I model.

As the adult homozygous M323K spinal cord showed substantial molecular dysregulation, with changes in both differential gene expression and alternative splicing, it seemed possible that an underlying cause could be through mislocalisation of the TDP-43 protein. To investigate this, immunohistochemical analysis was conducted on the adult homozygous M323K spinal cord at 1 year and 2 year time points. However, staining for TDP-43 revealed no change in TDP-43 localisation (Figure 4.67). It was not depleted from the nucleus, not excessively mislocalised in the cytoplasm, and did not form aggregates upon visual assessment. Staining for C-terminal fragments, 25 kDa and 35 kDa pathological fragments resulting from TDP-43 cleavage (Johnson et al. 2008), also revealed no change. TDP-43 hyperphosphorylation, reported to occur alongside aggregation (Arai et al. 2006), was also not detected (Figure 4.67).

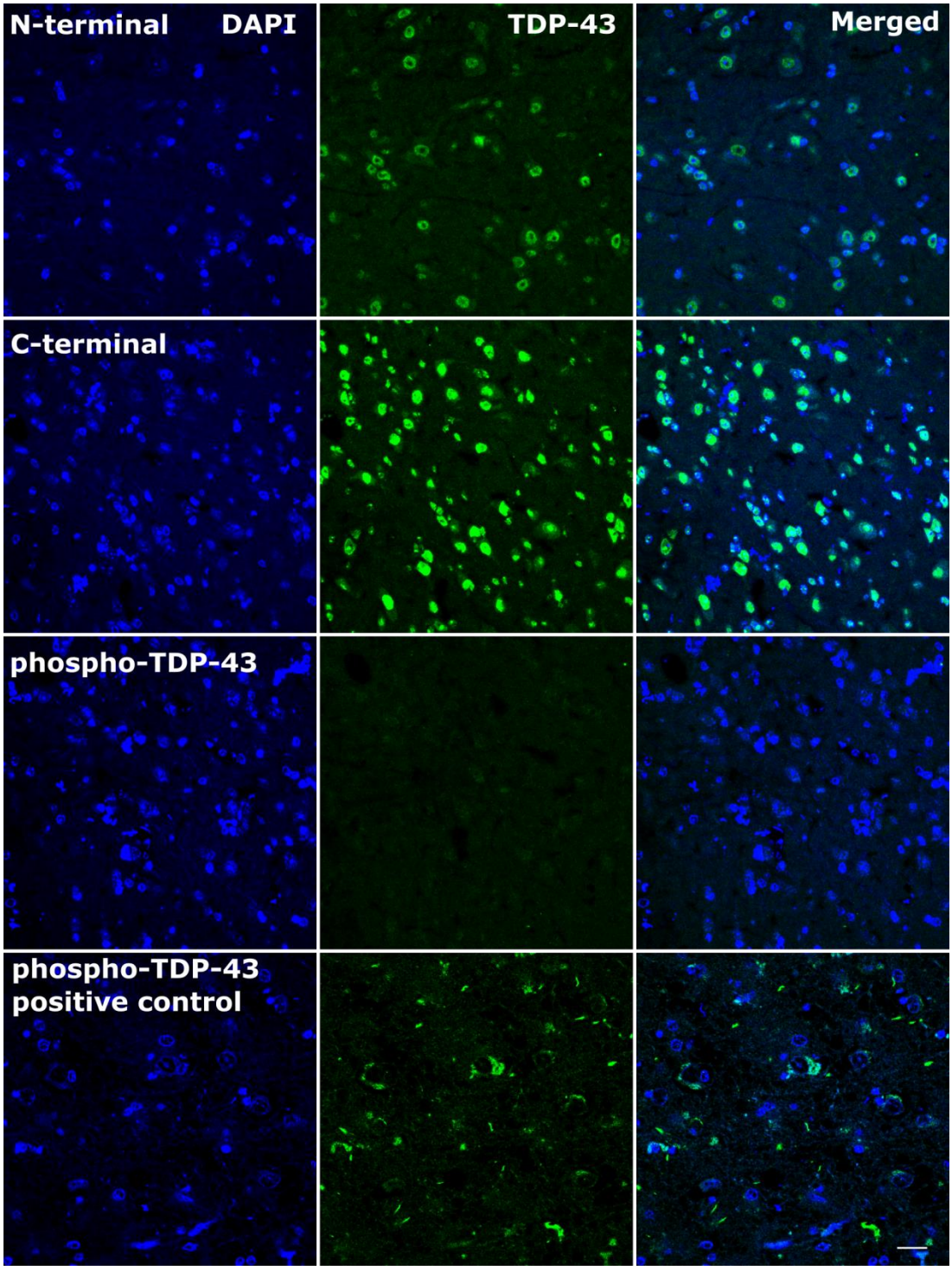


Figure 4.67. Staining panel for TDP-43 in adult homozygous M323K ventral horn of the lumbar spinal cord at the 2 year time point. Staining for TDP-43 N-terminal full length form showed no change in localisation or any aggregation. Staining for C-terminal TDP-43 to enable detection of pathological fragments showed no change in localisation or any aggregation. Staining for phosphorylated TDP-43

showed that hyperphosphorylation of TDP-43 was not present, while the positive staining for TDP-43 hyperphosphorylation in the positive control proved that technical failure was not the reason for the absence of TDP-43 hyperphosphorylation staining in the M323K. Scale bar indicates 20 µm.

Alongside TDP-43 pathology, ALS patients also typically display aggregation of components of the protein degradation pathway, such as p62 and ubiquitin, and so to understand whether protein degradation dysfunction was occurring in the absence of TDP-43 mislocalisation, these targets were also stained for. In contrast to TDP-43 itself, the M323K spinal cord was positive for both p62 and ubiquitin aggregation at the 2 year time point (Figure 4.68). Aggregates were limited to the ventral horns of the spinal cord, around the motor neurons but not within them. They were absent from dorsal grey matter regions, which contain sensory neurons.

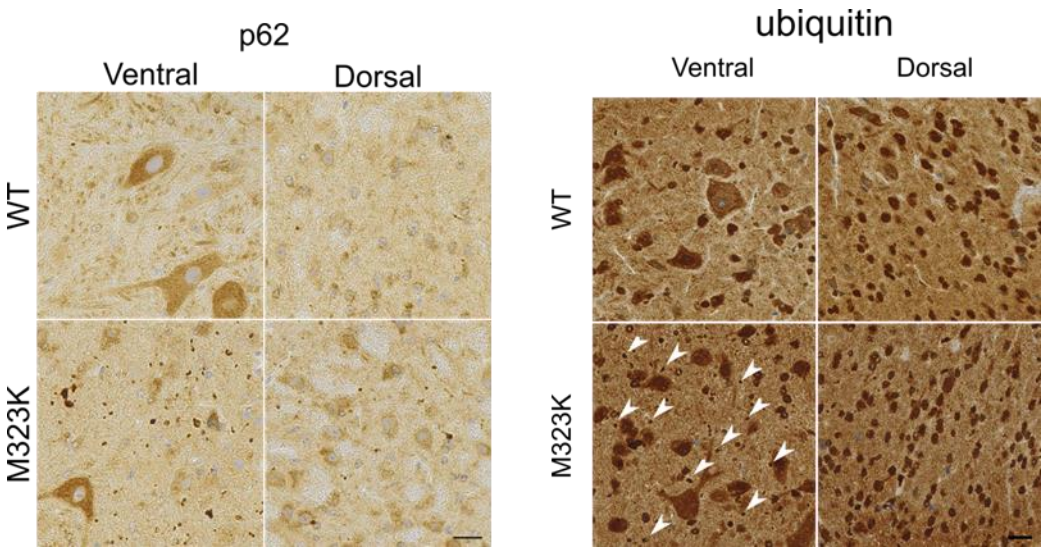


Figure 4.68. DAB staining for p62 and ubiquitin in adult homozygous M323K ventral horn lumbar spinal cord at the 2 year time point. P62 and ubiquitin pathology (the latter labelled by white arrows) detected in the ventral horns of the M323K model but were absent from dorsal regions, and also absent from any regions in the wildtype spinal cord. Scale bar indicates 20 µm.

The p62 pathology was not restricted to the spinal cord in the CNS; p62 pathology was also identified in the brainstem regions of the adult homozygous M323K brain, but absent from non-motor regions such as the hippocampus (Figure 4.69).

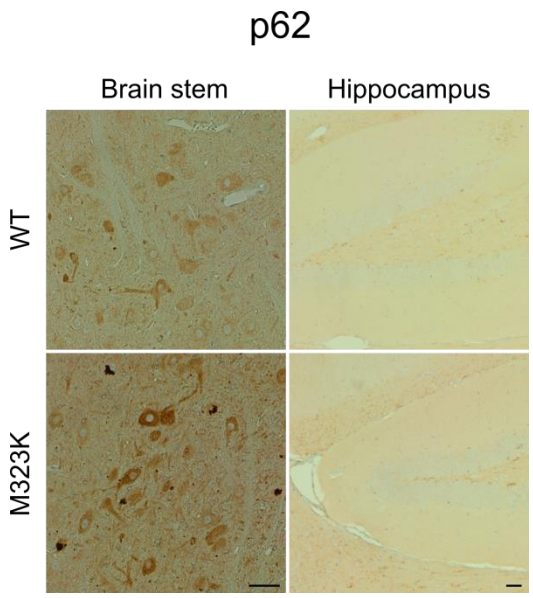


Figure 4.69. P62 pathology detected by DAB staining in the brainstem of homozygous M323K brain at the 2 year time point, but absent in the hippocampus and any wildtype region.

Alongside the molecular dysregulation and pathology, the motor phenotype of M323K mice at the 2 year time point was also tested. Investigation by Bernadett Kalmar and Alan Mejia Maza revealed significant motor deficits in the M323K mice, in grip strength, muscle force, motor unit number and motor neuron number (Figure 4.70).

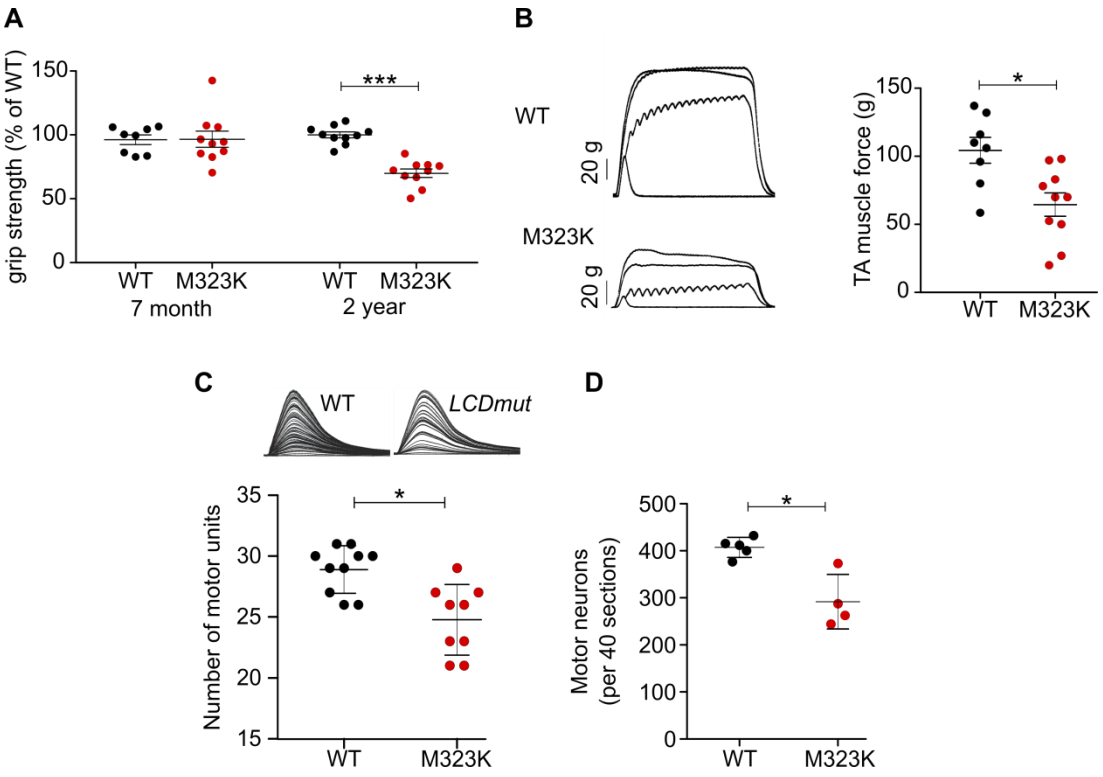


Figure 4.70. Motor phenotype investigation in homozygous M323K mice at 2 years. (A) Female mice show significant 30 % grip strength deficit at 2 years, but no difference at 7 months. (B) Muscle strength deficit in the tibialis anterior (TA) muscle in M323K mice, with representative trace. (C) Motor unit number reduction in M323K mice in the extensor digitorum (EDL), with representative trace from one motor unit shown above. (D) 28 % reduction in motor neuron counts from the sciatic pool in L4-L5 regions of the lumbar spinal cord. Results from Bernadett Kalmar and Alan Mejia Maza.

4.2.9 M323K CONCLUSIONS

The M323K two year old mice showed substantial RNA processing dysfunction, with numerous differential gene and exon expression changes. At the same time point, investigation of the mouse phenotype revealed no TDP-43 pathology. The absence of massive TDP-43 nuclear depletion could be predicted by the lack of RNA dysregulation resembling either the F210I TDP-43 loss of function model, or TDP-43 KD data. There

was no specific downregulation of long genes, and examples of TDP-43 loss of function splicing and cryptic exon inclusion were minimal. Given that these were evident in our TDP-43 loss of function models it could be deduced that no significant loss of function was occurring in the M323K model and therefore no nuclear depletion would be seen. However, alongside the absence of loss of function effects was the gain of function in TDP-43 splicing activity and skiptic exons, for which the resulting phenotype could not be predicted as no prior models with similar changes were available. Added to this was the seemingly dysfunctional autoregulatory mechanism of TDP-43, resulting in increased *Tardbp* levels where reduced levels would restore normal TDP-43 function, and it was plausible that this would lead to an increase in protein levels in the cytoplasm. Overexpression of TDP-43 has been demonstrated to result in excessive TDP-43 in the cytoplasm and subsequent aggregation (Tsai et al. 2010; Wils et al. 2010; Xu et al. 2010).

The detection of p62 and ubiquitin aggregates limited to the motor regions of the spinal cord and brain in the aged M323K models would further the presence of TDP-43 aggregation, as they function in the protein degradation pathway (Bjørkøy et al. 2005; Komatsu et al. 2007; Pankiv et al. 2007). Proteins to be degraded are ubiquitinated and p62 acts as an adaptor to facilitate the degradation via either the proteasome or autophagy pathway (W. J. Liu et al. 2016) using interaction with the microtubule-associated protein 1A/1B-light chain 3 protein (LC3) (Pankiv et al. 2007). Aggregations of both p62 and ubiquitin are themselves well-associated with ALS (Arai et al. 2006; Neumann et al. 2006; Mizuno et al. 2006), and mutation of the p62/SQSTM1 gene itself can cause ALS (Fecto et al. 2011).

Impaired liquid-liquid phase separation is also implicated in TDP-ALS mutations and thought to cause TDP-43 pathology (Johnson et al. 2009; Zhang et al. 2009; Guo et al. 2011), and the turbidity assay for M323K TDP-43, as described in

4.1.2 Previous study of the M323K mouse model (Figure 4.39), demonstrated a reduced ability of the mutated TDP-43 to phase separate. However, despite all of these indications, TDP-43 pathology was completely absent in the M323K model. This is in contrast to other mouse models with LCD mutations in TDP-43 (De Giorgio et al. 2018), however the type and extent of pathology varies even in these, and as most of the models involve transgenic expression of TDP-43, it is not possible to disentangle how much of the pathology results from the mutation alone.

It has been established that TDP-43 aggregation is not necessary for cellular, or neuronal, toxicity to occur (Barmada et al. 2010; Arnold et al. 2013). Highlighting this point, a recent study on a mouse model with the Q331K TDP-ALS mutation knocked in via CRISPR was found to at no stage to display TDP-43 pathology, despite not having a truncated lifespan and developing a phenotype resembling FTD (M. A. White et al. 2018). The only protein difference detected was a 45 % increase in nuclear TDP-43; this was not tested in the M323K mice and it would be of interest to see if it replicated the protein changes of a mouse with a TDP-ALS mutation. Studies of other ALS-related proteins such as SOD1 and FUS have conveyed the same results: transgenic overexpression of the protein can lead to a severe phenotype but mutation of the endogenous protein does not (Gurney et al. 1994; Mitchell et al. 2013; Joyce et al. 2015; Devoy et al. 2017; Scekic-Zahirovic et al. 2017). Such a link would also explain the absence of reduced lifespan in the models harbouring endogenous mutations.

Intriguingly, the presence of p62 and ubiquitin aggregates suggests that a protein is indeed aggregated in the M323K model, just not TDP-43. The identity of this protein is unknown; further tests for hnRNP A1 and hnRNP K, both themselves linked to TDP-43 pathology in ALS (Moujalled et al. 2017; Deshaies et al. 2018), have also proven negative. The idea that another RBP may be aggregated before TDP-43 itself upon TDP-43 mutation is especially curious, and establishing the identity of this protein could

provide strong clues on the early pre-symptomatic changes undergone in TDP-ALS. The detection of similar pathology in both the brain and spinal cord resembles the pattern of change in ALS, in which both upper and lower motor neurons are affected, and the motor phenotype of the M323K mouse strongly demonstrates neurodegeneration and an ALS-like phenotype too.

It is important to consider that at the embryonic and developmental stages, the homozygous M323K model showed no effects of the mutation, at the RNA level, in exhibiting any pathology or in any behavioural or phenotypic changes. However, the adults show mass changes in all of these aspects, suggesting they may all be linked. Crucially, the changes at the RNA level precede the development of motor phenotype, with the sizeable changes in differential gene expression and alternative splicing occurring in the six month old spinal cord on homozygous M323K animals, whereas a seven months these mice did not show a grip strength deficit, a phenotype that the older models proceeded to develop. This suggests the RNA changes may underlie the eventual phenotype of the mice, or at least provide the earliest indicator of the eventual neurodegeneration. The finding of conserved, novel, toxic, mutant specific skiptic exons that develop pre-symptomatically in adult models are likely to be of particular importance, although again whether that is through underlying causational changes or as early biomarkers is yet to be seen. In particular, the specific apparent enrichment of skiptic exons in genes coding for E3 ubiquitin ligases is of interest. These proteins function to directly or indirectly facilitate the ubiquitination of proteins for degradation, and a link between this RNA dysregulation and subsequent ubiquitin pathology developed by the 2 year old M323K mice implies a link between the two changes, although the exact mechanism behind this is not clear.

Incorporating the array of dysregulation as a result of M323K mutation, it is clear that the M323K model provides novel insight into the importance of TDP-43 loss of function

versus gain of function debate. The adult homozygous M323K mouse shows no loss of function effects, including those that would result from nuclear depletion if it fully replicated TDP-43 pathology in ALS, only the effects of a gain of function in TDP-43 splicing activity. Despite this, the mouse develops neurodegenerative pathology and a motor phenotype. Therefore, crucially, this model shows that LCD mutations in TDP-43 can lead to a neurodegenerative phenotype through gain of function in TDP-43 activity, in the absence of any loss of function effects.

CHAPTER 5

TRANSCRIPTOME DYSREGULATION IN OTHER MODELS OF ALS

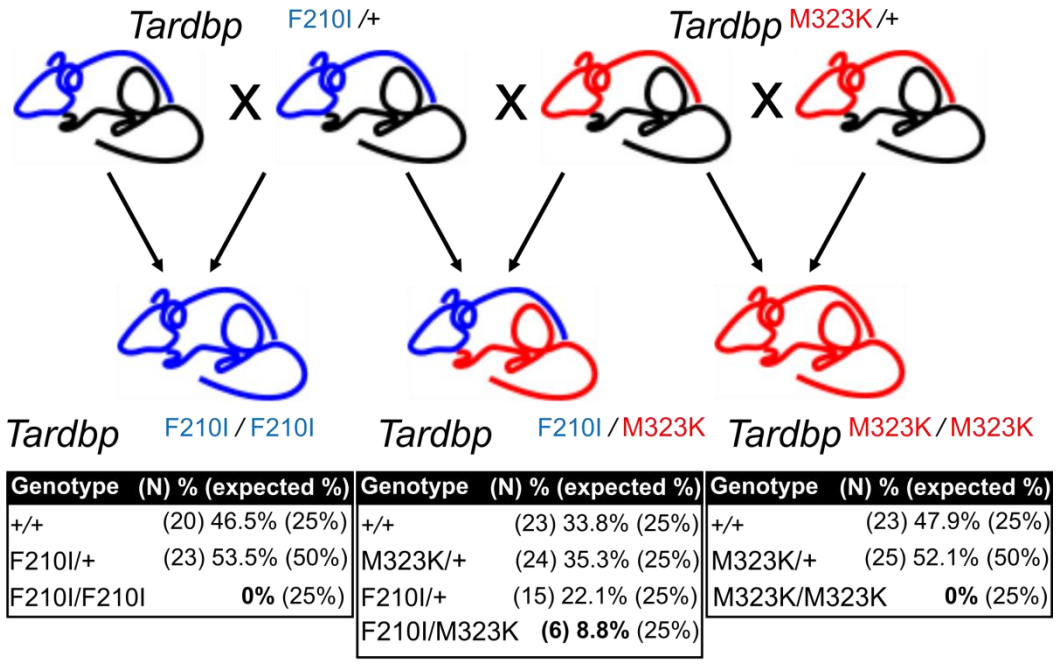
5.1 COMPOUND HETEROZYGOUS F210I / M323K MOUSE MODEL

The TDP-43 F210I and M323K mouse models showed opposite effects on TDP-43 splicing activity. This was demonstrated on a number of shared targets, such as the *CFTR* minigene assay in both HeLa cells and MEFs, as well as endogenous targets such as *Sort1*, *Eif4h* and *Dnajc5*. Comparing these effects to the dysregulation caused by TDP-43 KD established the F210I as causing a loss of function in TDP-43 splicing activity, whereas the M323K mutation caused the opposite, a gain of function. Therefore, we speculated that crossing these two models to produce compound heterozygous mice, with one *Tardbp* allele harbouring each mutation, may in effect cancel out some of the effects of the two mutations and result in a rescue of the RNA dysregulation.

Both the homozygous F210I and M323K mice were not viable after birth when raised on a pure C57BL/6J background, indicating that both mutations had substantial impact on early development. In the case of the F210I, this meant only the embryonic stages of homozygous models could be studied to investigate phenotype resulting from the mutation. However, the homozygous M323K model was able to survive on a mixed C57BL/6J / DBA/2J background, and showed no change in lifespan. This allowed study of the homozygous M323K mutation at developmental and adult time points, and this adult model was used as the primary focus of study.

To examine how the compound heterozygous mutation would affect embryonic viability, mice carrying one allele of each mutation were produced. While embryos

homozygous for either mutation were not viable, compound heterozygous F210I/M323K embryos were viable, suggesting a possible rescue effect of the two contrasting mutations. They did not follow expected Mendelian ratios, however, with just 8.8 % of pups possessing this genotype as opposed to the expected 25 %, suggesting a partial rescue (Figure 5.71).



Genotype	(N) % (expected %)
+/+	(20) 46.5% (25%)
F210I/+	(23) 53.5% (50%)
F210I/F210I	**0%** (25%)

Genotype	(N) % (expected %)
+/+	(23) 33.8% (25%)
M323K/+	(24) 35.3% (25%)
F210I/+	(15) 22.1% (25%)
F210I/M323K	**(6) 8.8%** (25%)

Genotype	(N) % (expected %)
+/+	(23) 47.9% (25%)
M323K/+	(25) 52.1% (50%)
M323K/M323K	**0%** (25%)

Figure 5.71. Diagram illustrating the F210I/M323K compound heterozygous model viability. Significant increase in survival of compound heterozygous model compared to homozygous F210I and homozygous M323K models ($P < 0.05$). (N) represents numbers of animals produced per genotype; % the percentage of mice observed per genotype and (expected %) the percentage of mice expected per genotype, calculated by Abraham Acevedo Arozena.

However, it was important to remember that both embryonic heterozygous F210I and embryonic heterozygous M323K mice themselves showed minimal RNA processing changes, so it was not yet clear whether this was a true rescue of the effects with the two mutations counteracting each other.

To investigate this aspect further, splicing of known TDP-43 targets was examined. RNA-Seq was conducted on the compound heterozygous models at E14.5, the same time point used for our previous embryonic RNA-Seq studies. At this stage, the homozygous F210I model showed mass RNA dysregulation, with 1,184 genes showing changes in expression (adjusted p-value < 0.1) and 67 significant shifts in alternative splicing (adjusted p-value < 0.01). Although the homozygous M323K model primarily showed RNA processing changes at the later stage adult time point, embryonic models still displayed several examples of dysregulation, with 53 differentially expressed genes and 11 differentially used splicing events. In keeping with the posited rescue effect, compound heterozygous mice showed almost no RNA dysregulation, with just 9 differentially expressed genes and no significant alternative splicing events, again suggesting that the mutations could be cancelling each other out.

Crucially, this was also true when specific splicing events were examined. *Eif4h* was significantly differentially spliced in opposite directions in both embryonic homozygous models, with the F210I model displaying greater levels of exon inclusion and the M323K model greater levels of exon skipping. Exemplifying the counteracting effect of the compound heterozygosity for the F210I/M323K mutations, in this model *Eif4h* was not significantly differentially spliced from wildtype, showing a middling level of exon inclusion levels in between the effects of the F210I and M323K mutations alone (Figure 5.72).

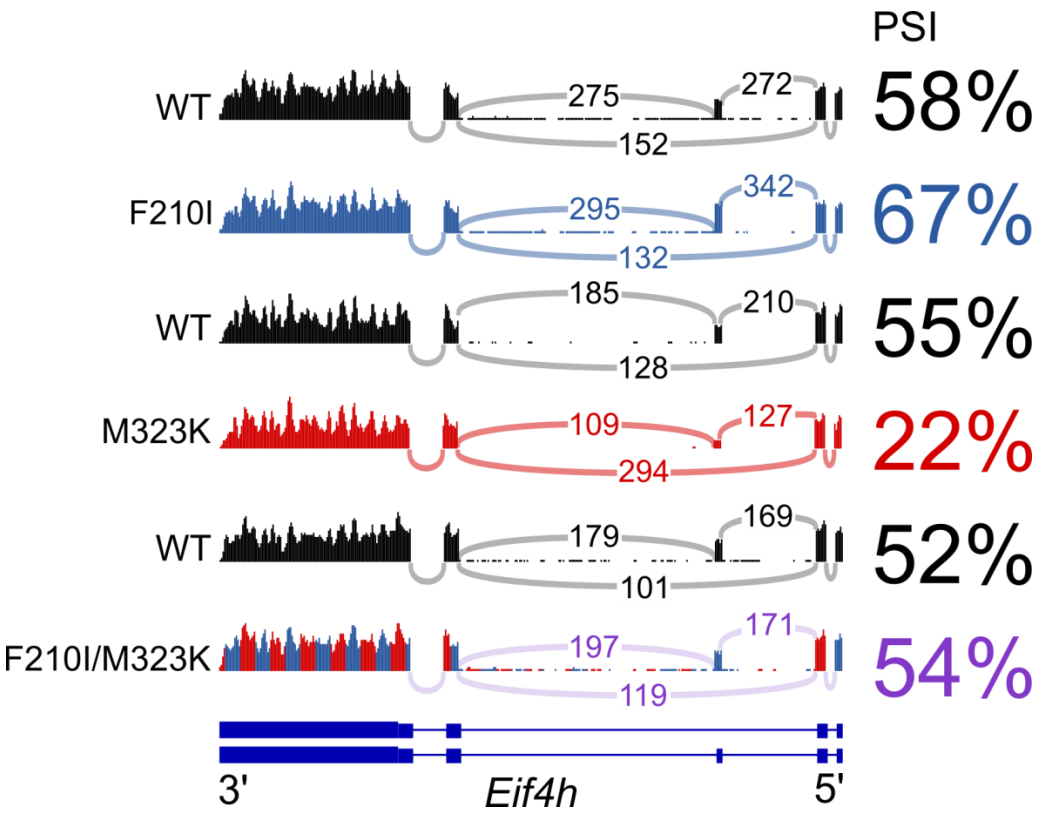


Figure 5.72. RNA-Seq reads for each embryonic model with sashimi plots demonstrating splicing patterns overlaid. Number of reads annotating each splicing event annotated on sashimi plot, with Δ PSI values for the splicing event from SGSeq alternative splicing output displayed on the right. RefSeq annotated *Eif4h* transcript isoforms below. Compound heterozygous F210I/M323K model showed alternative splicing pattern similar to wildtype, in contrast with the greater exon inclusion levels in F210I and greater skipping levels in M323K models.

Also absent from this model were any of the extreme splicing events from each homozygous model, neither cryptic exons nor skiptic exons were detected. Importantly, there was also no evidence for any autoregulation of TDP-43 occurring, with no increase in *Tardbp* transcript levels and no splicing change to indicate a shift in the mechanism from the RNA-Seq data. Therefore, the results for this compound mouse model followed on from previous results that suggested the two mutations had

opposite effects on TDP-43 splicing activity but shared splicing targets and binding regions, with the opposite effects caused specifically by changes in TDP-43 functioning activity levels.

5.2 Q331K MOUSE MODEL

The homozygous M323K model contained a mutation in the LCD of TDP-43, within the hotspot region in which several TDP-ALS mutations occur (Buratti 2015). As a result of the mutation, the mouse develops mass RNA dysregulation, novel skiptic exon skipping, protein degradation pathology and a neurodegenerative disease-like motor phenotype. However, the key drawback of this mouse model is that although similar, the M323K mutation itself has not yet been found to cause ALS in patients. Even within the LCD, different mutations have a variety of different effects on overall TDP-43 dysfunction (Jiang et al. 2016), therefore it is not clear to what extent the M323K mimics the molecular dysregulation upon LCD mutation in TDP-ALS. To further investigate the level of this correlation, we investigated a mouse model with a knock-in Q331K mutation in the endogenous mouse *Tardbp* gene.

The Q331K mutation (Figure 5.73) was amongst the first ALS-causative mutations identified in TDP-43 (Sreedharan et al. 2008). Since then, mouse models have been developed harbouring Q331K-mutated TDP-43 (Arnold et al. 2013; Mitchell et al. 2015), however until recently these have involved transgenic expression of the mutated protein and therefore suffer from the previously described issues of this method. However, a study published earlier this year investigated a mouse model with the Q331K mutation introduced into the endogenous mouse *Tardbp* gene via CRISPR (M. A. White et al. 2018). This represented the first published knock-in mouse model of TDP-ALS. A similar model to this, also with the Q331K mutation CRISPR-induced into the mouse *Tardbp* gene, was concurrently investigated by Abraham Acevedo Arozena

at MRC Mammalian Genetics Unit, Harwell, Oxford, and then Unidad de Investigación, Hospital Universitario de Canarias, Fundación Canaria de Investigación Sanitaria and Instituto de Tecnologías Biomédicas (CIBICAN), La Laguna. To understand whether the RNA processing changes in the M323K model could be replicated by TDP-ALS mutations, a selection of the alternative splicing, and skiptic exon, changes were tested for in the homozygous adult Q331K mouse brain using RT-PCR.

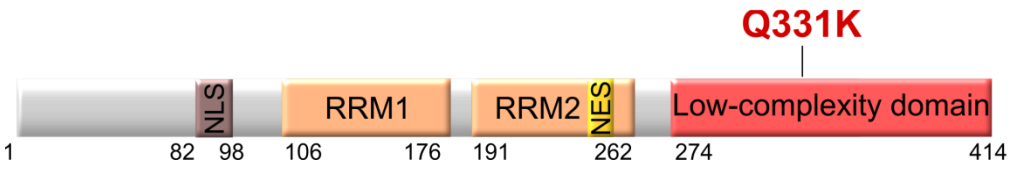


Figure 5.73. TDP-43 domains and location of the Q331K mutation with the LCD domain.

The first investigation was to determine whether the Q331K mutation displayed a similar gain of function in TDP-43 splicing activity to the M323K mutation, and for this the *CFTR* minigene assay was again employed. In homozygous M323K MEFs, the mutant TDP-43 caused greater levels of exon 9 skipping in the *CFTR* minigene. This characterised it as a gain of TDP-43 splicing function mutant, as it opposed the loss of function exon 9 inclusion increase seen in F210I MEFs and also upon TDP-43 KD. In concordance with the M323K model, the Q331K MEFs also displayed a dose dependent increase in exon skipping, identifying the Q331K mutation too as causing a gain of function in TDP-43 splicing (Figure 5.74).

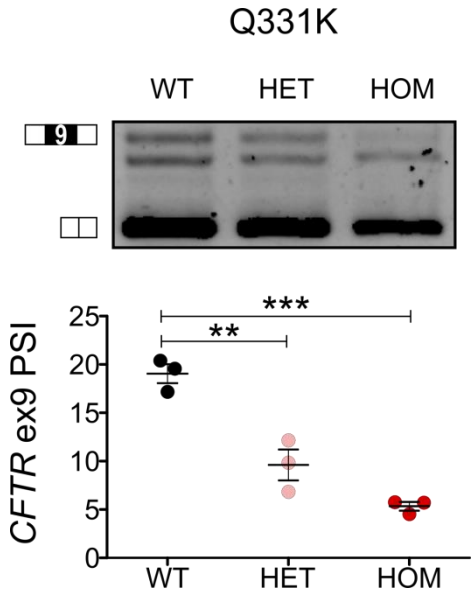


Figure 5.74. *CFTR* minigene assay showing RT-PCR results and % exon 9 inclusion in (left) wildtype TDP-43, (middle) heterozygous Q331K TDP-43, (right) homozygous Q331K TDP-43. Significant and dose dependent increase in exon 9 inclusion as a result of Q331K mutation. Minigene isoforms for exon 9 inclusion and skipping labelled. Results from Abraham Acevedo Arozena.

The definition of the Q331K mutation as causing similar, gain of TDP-43 function-mediated changes in alternative splicing to the M323K was strengthened by RT-PCR study into known TDP-43 alternative splicing targets such as *Sort1*, *Eif4h* and *Kcnip2*. These alternative splicing events also showed a dose dependent decrease in exon skipping in the Q331K model, parallel with the M323K mouse (Figure 5.75).

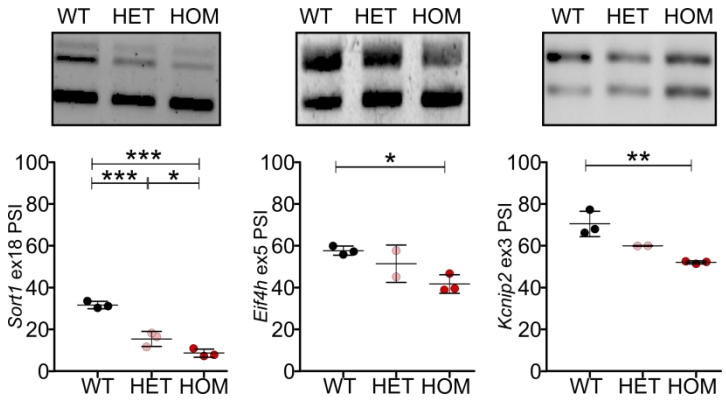


Figure 5.75. RT-PCR results and quantification for inclusion levels of *Sort1* exon 18, *Eif4h* exon 5 and *Kcnip2* exon 3 in wildtype, heterozygous Q331K and homozygous Q331KI MEFs. Significant and dose dependent decrease in exon inclusion as a result of Q331K mutation. Results from Abraham Acevedo Arozena.

As the well-defined splicing events showed similar changes in both the M323K and Q331K models, we further tested the Q331K transcriptome for the novel skiptic events described in 4.2.6 Skiptic exons. These events had not previously been described in relation to changes in TDP-43 functioning, but in the Q331K model Abraham Acevedo Arozena identified a subset of the validated skiptic exons from the M323K model which also showed significant increases in exon skipping (Figure 5.76).

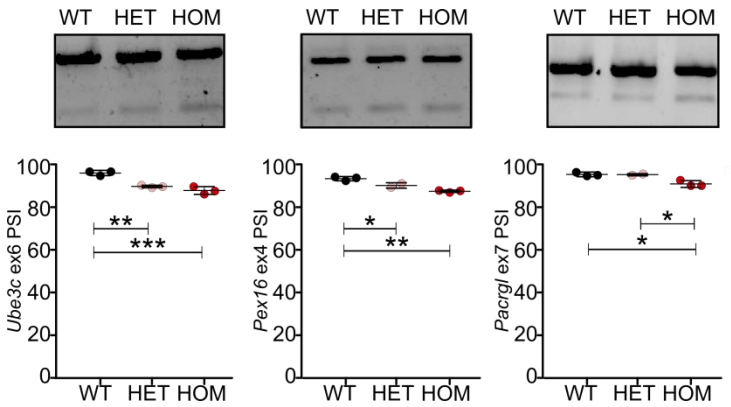


Figure 5.76. RT-PCR results and quantification for inclusion levels of M323K skiptic exons *Ube3c* exon 6, *Pex16* exon 4 and *Pacrgl* exon 7 in wildtype,

heterozygous Q331K and homozygous Q331K MEFs. Significant and dose dependent decrease in skiptic exon inclusions as a result of Q331K mutation. Results from Abraham Acevedo Arozena.

It is important to note that while the Q331K mouse replicates the effect of TDP-43 gain of splicing function seen in the M323K model, the magnitude of the changes are lower in the Q331K mutant. The skiptic exons in particular seen by RT-PCR showed significant skipping, but size of the change was far less than the equivalent RT-PCR tests for the M323K (Figure 4.57). One explanation for this could be that the Q331K mutation is simply 'weaker' than the M323K mutation, and therefore inducing a less severe TDP-43 gain of function phenotype. This theory is supported by the homozygous Q331K mice proving viable on a pure C57BL/6J background, unlike either the homozygous F210I or M323K embryos. Additionally, phenotype testing by Abraham Acevedo Arozena is currently underway, but initial results do not suggest a motor phenotype of the same severity as in the adult M323K model. Therefore, if the Q331K mutation is weaker, then it is to be expected that the splicing dysregulation that results will also be dampened in comparison to the M323K.

Testing of a subset of targets in the Q331K model already known to be significantly changed in the M323K mouse demonstrated some level of correlation, but to truly understand the similarity of the effects of the two LCD mutations on transcriptome regulation, RNA sequencing was required. For this, we utilised RNA-Seq experiments conducted on the Q331K model for the previously mentioned study by White et al. (2018). In this investigation, the C>A mutation resulting in the Q331K amino acid change seen in TDP-ALS patients was incorporated into the mouse *Tardbp* gene using CRISPR. As with the Q331K mouse in the hands of Abraham Acevedo Arozena, the homozygous Q331K model on a pure C57BL/6J background was viable and survived with no apparent change in lifespan. These mice displayed limited changes in

behaviours such as rearing and hanging, but initially nothing concrete to suggest a disease phenotype. Similarly, they performed worse on the rotarod test, but rather than a motor phenotype, this was determined to be due to an increase in Q331K model weight brought about by increased feeding tendency. These mice also showed no TDP-43 mislocalisation or pathology like the M323K model, but additionally also displayed no degenerative pathology, no motor neuron degeneration and motor unit loss. Thus, the M323K mutant more closely replicated the ALS disease phenotype than the mouse model harbouring the TDP-ALS mutation. However, further cognitive testing of the Q331K model revealed deficits in attention and memory, resembling a disease closer to dementia than motor neuron disease on the ALS-FTD spectrum.

To understand the molecular dysregulation underlying this neurodegenerative disease phenotype, White and colleagues conducted RNA-Seq analysis on heterozygous and homozygous Q331K frontal cortex at 5 months and then 20 months of age (Gene Expression Omnibus reference GSE99354). They described a handful of specific differentially expressed genes and alternative splicing events but, outside of the previously detailed changes mentioned in the M323K model, such as in *Tardbp*, *Sort1* etc., none of the changes they mentioned were significantly altered in the M323K models.

To gauge the true level of similarity between the Q331K and M323K models, the published RNA-Seq data was obtained and analysed using the same pipeline as the M323K analysis, removing a degree of technical variation. Broad analysis was then conducted on the correlation of changes seen in the heterozygous and homozygous models of both, with FDR correction thresholds modified to 0.05 for both differential gene expression and alternative splicing. Dataset names are abbreviated as follows:

6 month homozygous M323K spinal cord – m323k_adult_hom,
6 month heterozygous M323K spinal cord – m323k_adult_het,

5 month homozygous Q331K frontal cortex – q331k_5m_fc_hom,
5 month heterozygous Q331K frontal cortex – q331k_5m_fc_het,
20 month homozygous Q331K frontal cortex – q331k_20m_fc_hom,
20 month heterozygous Q331K frontal cortex – q331k_20m_fc_het.

In terms of differential gene expression, correlation analysis was conducted by comparing the log_2 fold changes of genes differentially expressed in either dataset, and calculation of the Pearson correlation coefficient. For reference, the correlation plot and r value for the M323K homozygous model in comparison to the M323K heterozygous model was also included. The homozygous and heterozygous M323K RNA-Seq datasets showed strong positive correlation, demonstrating that despite only a small number of expression changes occurring in the heterozygous model, the direction of change was mostly replicated (Figure 5.77). This correlation was highlighted by an r value of 0.783. Comparison of the homozygous M323K RNA-Seq data, from 6 month spinal cord, with homozygous Q331K frontal cortex RNA-Seq at 5 months and 20 months, also showed strong correlation in expression changes in both. The r values, although understandably smaller than comparison of the two M323K models, still suggested a high level of correlation, with values of 0.317 and 0.338 respectively. Specific gene expression levels were not tested experimentally in the M323K and Q331K models, and so these analyses provided strong new evidence that expression changes in the two models were highly similar across the transcriptome. Similar analysis was also conducted for the heterozygous models at each time point, and comparison between each of these as well as to each homozygous model showed strong levels of correlation, with comparison of no two datasets resulting in an r value of below 0.23 (Figure 5.77).

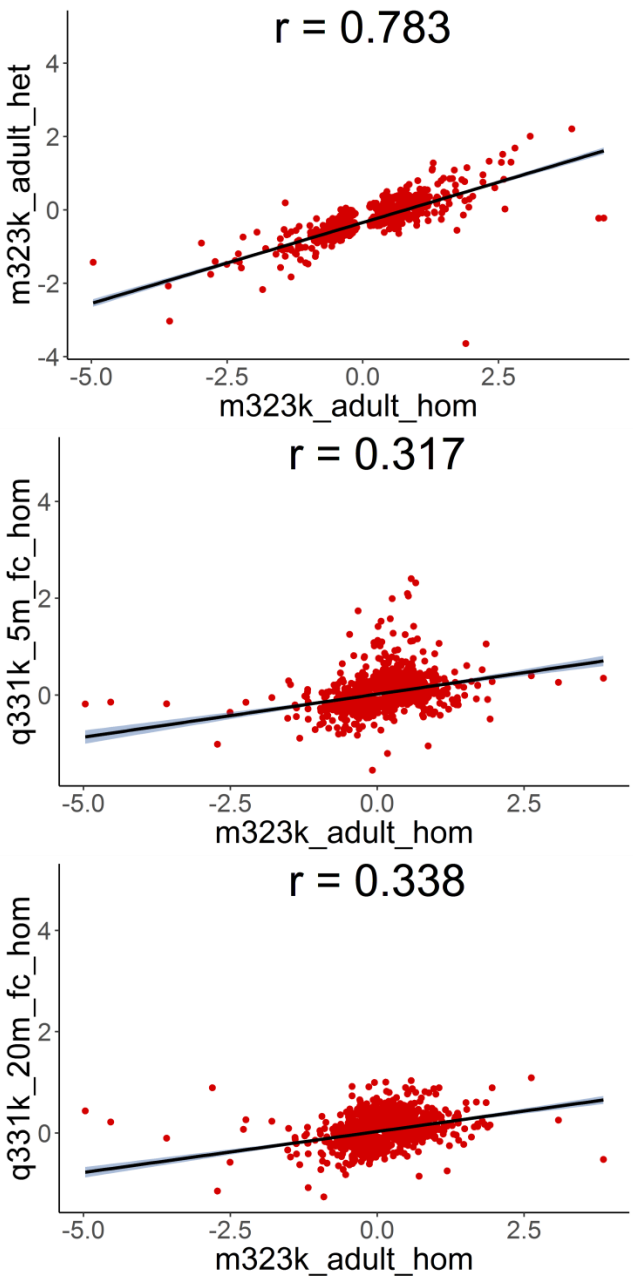


Figure 5.77. Scatter plots comparing significantly differentially expressed genes in either of the two compared RNA-Seq datasets measured by log_2 fold change of mutant transcript vs wildtype. Comparisons are homozygous adult M323K vs each of heterozygous adult M323K, homozygous 5 month Q331K frontal cortex, and homozygous 20 month Q331K frontal cortex. Linear model fitted for each comparison. Pearson correlation coefficient annotated for each comparison. Strong positive correlation seen in each comparison.

Given the number of RNA-Seq datasets from the two models, the significant hits from each were also compared to determine the level of overlap. Two of the six datasets showed no differentially expressed genes – the heterozygous M323K adult dataset and the heterozygous Q331K 5 month frontal cortex. Comparison was analysed graphically using the UpSetR package (Conway et al. 2017).

The vast majority of differentially expressed genes were unique to any one dataset (Figure 5.78). However, 29 genes showed changes in expression in every homozygous model dataset and 5 genes were differentially expressed in these three and also the aged heterozygous 20 month Q331K data. These five genes were *Spop*, *Gid4*, *Gstm7*, *D2hgdh* and *Gm37277*. Intriguingly, *Spop*, downregulated in each dataset, codes for a component of the E3 ubiquitin-protein ligase complex (Furukawa et al. 2003), the same category which was found to be enriched for skiptic exon transcripts in the homozygous M323K model. GID4 protein is also part of a ubiquitin ligase complex as highlighted by its gene ontology, and this gene is also downregulated in each dataset. The downregulation of two genes from this specific pathway, alongside the skiptic exons and the ubiquitin pathology in the more severe phenotype of the M323K model, further suggests that ubiquitin ligase complex activity is a key area for further study.

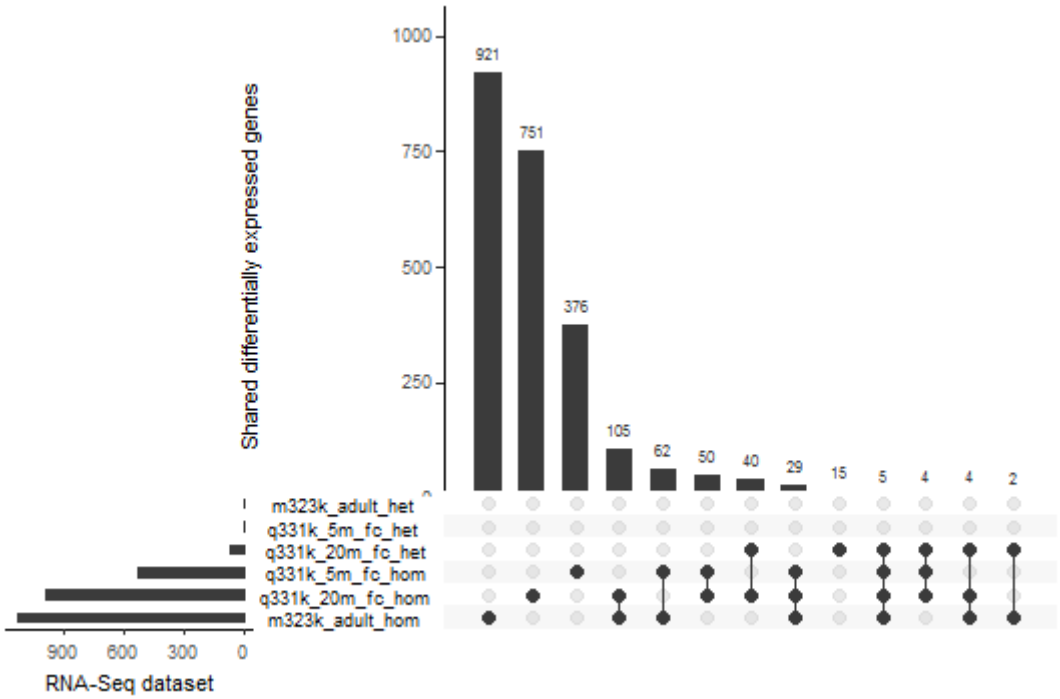


Figure 5.78. Comparison and quantification of overlap between the significantly differentially expressed gene sets from each of the analysed M323K and Q331K RNA-Seq datasets. Bar graph at the top indicates number of shared differentially expressed genes, dot matrix below indicates the datasets compared, and columns to the left show total size of differentially expressed genes set in each RNA-Seq dataset. Most differentially expressed genes unique to one RNA-Seq dataset, but five genes dysregulated in every dataset showing significant changes.

Similar analyses were also conducted on the alternative splicing events for each M323K and Q331K dataset. Scatter plot comparison again showed strong correlation first between the M323K homozygous and heterozygous models, with an r value of 0.737 (Figure 5.79). Comparison of the homozygous M323K alternative splicing to the events in the 5 month and 20 month homozygous Q331K models also showed sizeable correlation, with the r values of 0.522 and 0.49 noticeably higher than similar testing of the differential gene expressions correlations (Figure 5.79). Again, similar comparisons

were conducted with each heterozygous model, and every two way comparison showed positive correlation, with all r values above 0.38.

The strong levels of correlation in significant alternative splicing events, even more so than the similarity seen in gene expression changes, is a further indicator of the similarity in transcriptome dysregulation in the two LCD mutants. Specific transcripts were already shown to modulate splicing in the same manner (Figure 5.75, Figure 5.76); now the general transcriptome-wide splicing pattern was demonstrated as markedly similar between the two models across mutant copy number and time point differences. As it was the splicing changes that classed the M323K as a gain of function, the splicing pattern similarities further enabled the characterisation of the TDP-ALS mutation Q331K as also resulting in gain of TDP-43 splicing function. The enrichment of strong exon skipping changes as a result of Q331K TDP-43 mutation had also previously been described in a transgenic mouse model (Arnold et al. 2013). Notably, the Q331K splicing changes tended to exhibit smaller Δ PSI values than in M323K model in general, from which it could be inferred that the Q331K mutation may be the weaker mutation. This supports previous assertions given the reduced impact of the Q331K on known TDP-43 splicing targets, and the weaker phenotype in the homozygous mouse model.

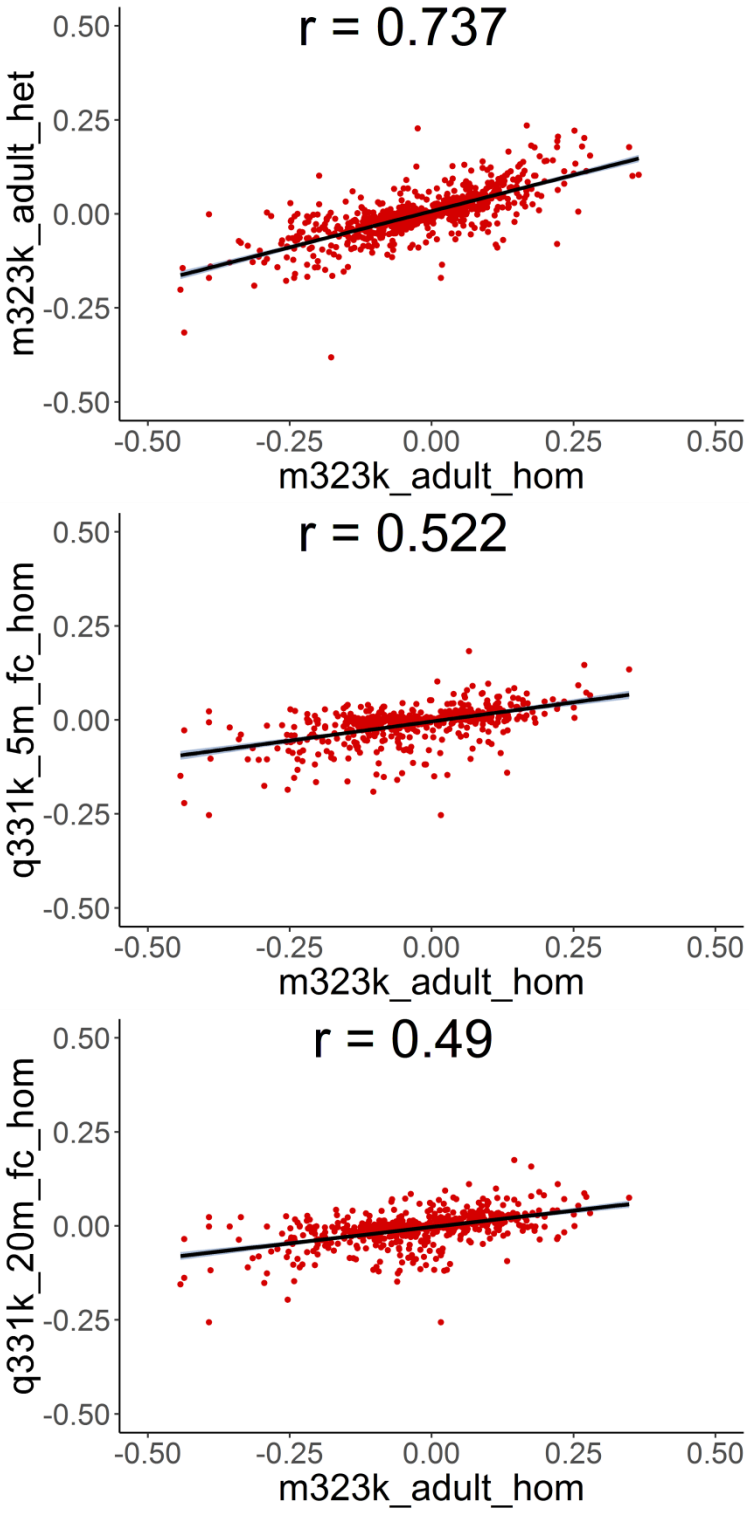


Figure 5.79. Scatter plots comparing significantly differential splicing events in either of the two compared RNA-Seq datasets measured by log_2 fold change of transcript in mutant vs respective wildtype. Comparisons are homozygous adult M323K vs each of heterozygous adult M323K, homozygous 5 month Q331K

frontal cortex, and homozygous 20 month Q331K frontal cortex. Linear model fitted for each comparison. Pearson correlation coefficient annotated for each comparison. Strong positive correlation seen in each comparison.

In a similar manner to the differential expression analysis, overlap of significantly differently alternatively spliced genes from each dataset was also tested. Visualisation with the UpSetR package (Conway et al. 2017) again revealed that the majority of splicing changes were unique to the models showing the most changes overall (Figure 5.80). 23 splicing events were significantly changed in each of the homozygous datasets but not in any of the heterozygous models. Further, 10 splicing events were detected to be changed in four of the six datasets.

Intriguingly, one event was alternatively spliced differently in all datasets but the heterozygous aged Q331K model, and this was the previously discussed splicing of *Eif4h*. The alternatively spliced exon in this gene was increasingly skipped in each dataset, corroborating the gain of TDP-43 splicing function in the two mutations. However, remarkably there was also one event showing splicing changes in all six datasets (Figure 5.80), and this was a splicing change in *Ttc3*. Unlike *Eif4h*, *Ttc3* has not been previously discussed, despite consistently showing splicing changes. This is due to several alternative splicing events being involved in the *Ttc3* gene; the same is not true for *Eif4h*. The multiple isoforms of *Ttc3* also renders validation of the splicing changes considerably more difficult. However, the *Ttc3* gene is of great interest as it belongs to the same group of proteins already heavily implicated in the dysregulation caused by the LCD mutants – it is another E3 ubiquitin ligase. As previously mentioned in 4.2.6 Skiptic exons, E3 ubiquitin ligases are a diverse group of proteins involved in protein ubiquitination, either directly or indirectly as adaptors to facilitate the ubiquitination (Berndsen & Wolberger 2014). The changes in this subgroup of proteins in both gene expression and alternative splicing, occasionally resulting in novel

transcripts due to skiptic exons, and the subsequent ubiquitin pathology in the M323K model, together strongly implicate this pathway as dysfunctional in these TDP-43 LCD mutant models of neurodegeneration.

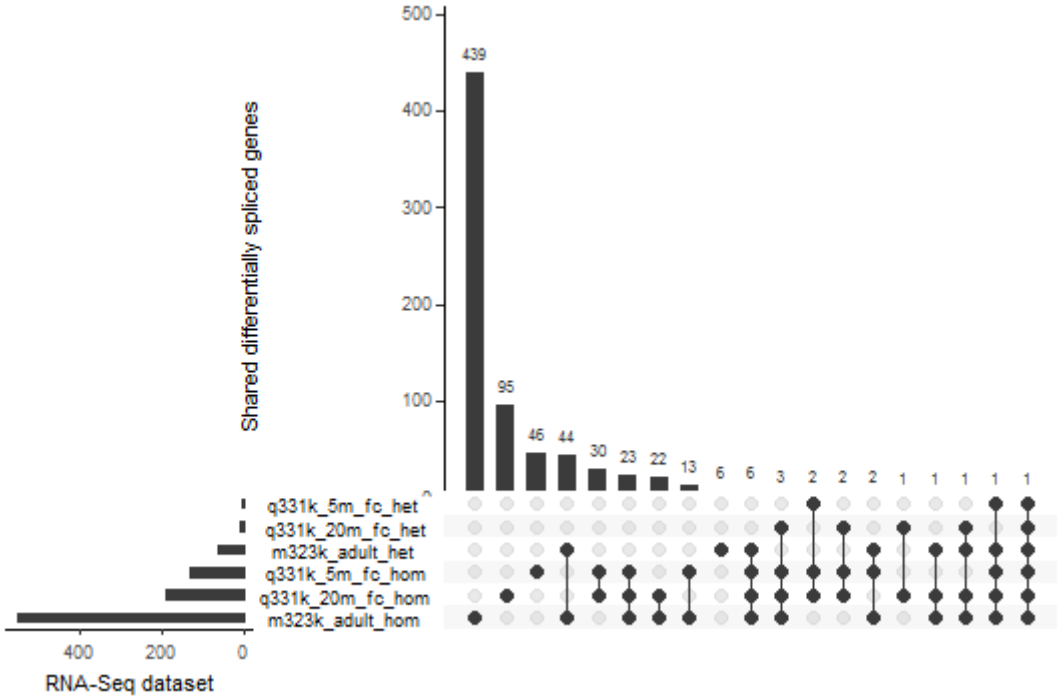


Figure 5.80. Comparison and quantification of overlap between the significantly differentially spliced gene sets from each of the analysed M323K and Q331K RNA-Seq datasets. Bar graph at the top indicates number of shared differentially spliced events, dot matrix below indicates the datasets compared, and columns to the left show total size of the differentially spliced event set in each RNA-Seq dataset. Most differential splicing events were unique to one RNA-Seq dataset, but one splicing event was dysregulated in every dataset.

Given the comparable changes in alternative splicing in the Q331K model compared to the M323K, extreme splicing events were also analysed in this dataset. Cryptic exons, shown to be as a result of loss of function in the F210I model, were minimal as expected. However, whereas 47 skiptic exons were identified in the M323K model, only 5 skiptic events were detected in each of the homozygous Q331K 5 month and 20

month datasets. Two of these skiptic events were in both Q331K datasets, in *Stt3b* and *Hsdl2*. Interestingly, the skiptic exon in *Hsdl2* was also detected as a skiptic exon in the M323K dataset, however overall level of overlap was poor. A reason for the lack of skiptic exons in the Q331K may have been technical, as the number of reads per sample in the Q331K RNA-Seq datasets were far less than in the M323K, and the skiptic exons were often rare splicing events. Therefore, although they may have shown a similar pattern of skipping, they may have not reached a sufficient level of significance to be classed as a skiptic exon. To test if this was true, the skiptic exon events from the homozygous M323K 6 month spinal cord dataset were also examined in each of the heterozygous and homozygous Q331K datasets in 5 month and 20 month frontal cortex. Analysis revealed that, with a couple of exceptions, the skiptic exons in the M323K tended to show increased skipping in the Q331K datasets too (Figure 5.81). However, the Δ PSI values for these events were generally much smaller than in the M323K. This is concordant with previous assertion that the Q331K is a weaker mutation than M323K. The relatively small magnitude of splicing changes prevents the identification of the M323K skiptic exons as skiptic in the Q331K dataset, but they do nonetheless typically show increased levels of skipping in each Q331K dataset.

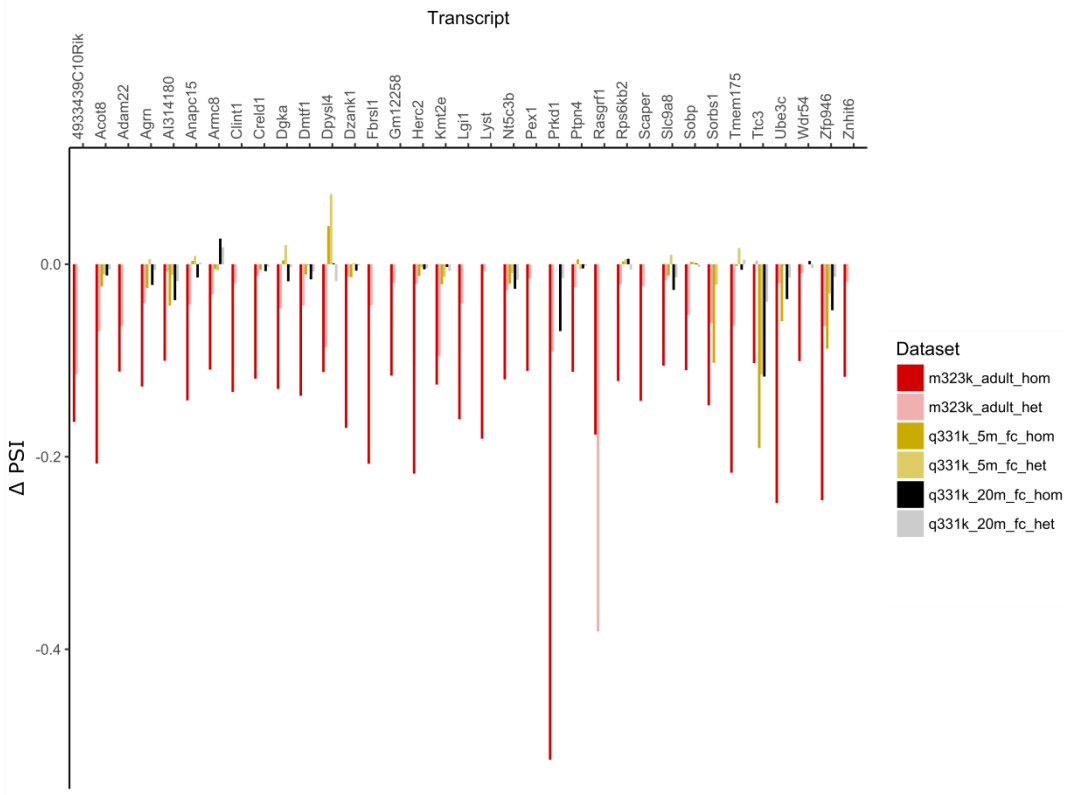


Figure 5.81. Comparison of Δ PSI levels of the skiptic exons identified from the homozygous adult M323K RNA-Seq dataset in all other analysed LCD mutant datasets. In most splicing events, Δ PSI skipping was greatest in the homozygous adult M323K dataset, but showed reduced levels of skipping in other datasets. Several skiptic exon splicing events were not seen in the Q331K datasets.

Both of the above comparison methods – the correlation analysis and the dataset overlaps – provide information on the similarity of the two datasets being analysed. However, they both have their drawbacks. The correlation analysis conveys the overall similarity of the two datasets, and the overlaps analysis provides specific instances of interesting change in the compared datasets, but neither makes use of the potential increase in statistical power to detect changes which would not be detected as significant in one dataset alone. These changes may show higher levels of variability, low levels of expression, or small but consistent changes, but may still prove of interest if occurring across the different models.

To find these changes, we also conducted a meta-analysis of the differential gene expression RNA-Seq datasets from the M323K and Q331K mutants. There are a number of methods by which this could have been conducted, and the first method trialled was using the Fisher method of p-value combination using the metaRNA-Seq package (Rau et al. 2014). With the modified FDR threshold of 0.05, the homozygous M323K 6 month spinal cord dataset showed 757 changes in gene expression. With the same threshold, the homozygous Q331K 5 month frontal cortex dataset showed 370 differentially expressed genes. Using a simple overlap method, the number of significant hits in both datasets amounts to just 90 genes, despite the clear strong correlation in gene expression level changes in the two datasets as shown in Figure 5.77.

However, using the Fisher method of p-value combination, 1,165 genes can be identified as significantly changed in terms of expression levels. The changes in significant hits is illustrated in Figure 5.82 below, in which the original FDR values of each gene for either dataset are plotted to enable visualisation. In red are genes which were identified as significant in either of the two datasets in the initial analysis, and remain significant upon Fisher combination. Black dots however represent genes which were classed as significant in either original analysis but not upon p-value combination. These genes are typically those that display significant change in one dataset but no evidence of any change in the other. In grey are genes not significant in any analysis. The genes in pink are those that were not originally classed as significant in either dataset but showed adjusted p-values approaching significance in both; these are the hits obtained with the increased power upon p-value combination. From the M323K dataset, 97 of the 757 genes lost significance upon p-value combination, but 505 new genes were classed as significant. For the Q331K dataset just 16 genes were lost and 811 gained (Figure 5.82). The remarkably high number of new significant hits using this Fisher p-value combination method highlights the strong level of similarity

in the two datasets, with read depth and sample numbers in individual experiments the likely limiting factors due to which these genes are not detected upon simple dataset overlap.

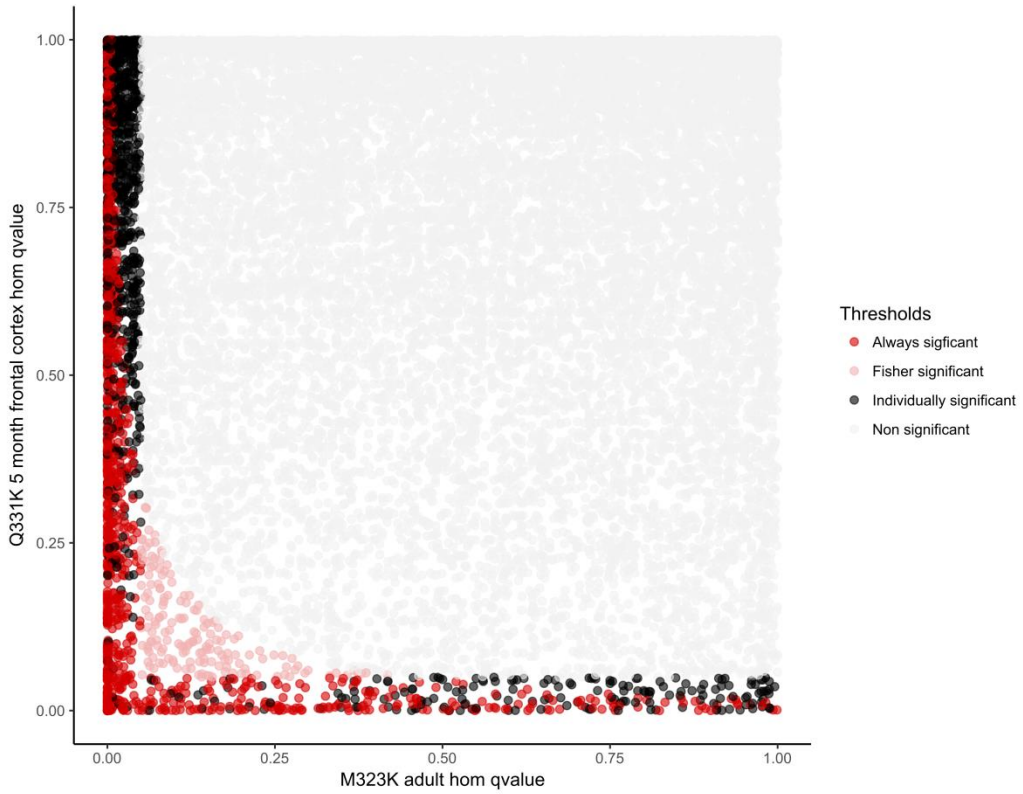


Figure 5.82. Adjusted p-values of all genes expressed in the homozygous adult M323K and homozygous 5 month Q331K RNA-Seq datasets compared, and the outcome of different tests for statistical significant illustrated. Red dots represent genes classed as significantly differently expressed in either individual dataset, and also upon p-value combination with the Fisher method. In black are genes classed as significant in either individual dataset but not upon p-value combination, in pink are genes not classed as significant in either individual dataset but significant upon p-value combination, and in grey are genes which show no significant differential expression upon either method.

Although the Fisher method of p-value combination incorporates the increased power of two datasets in its meta-analysis, it does not take other information into account,

namely the direction of the change. This was artificially incorporated in the above analysis, but the removal of genes showing expression changes in opposite directions reduces the power of the analysis. Furthermore, features of the gene expression itself, such as normalised gene expression levels, are not used in the statistical test.

This metric, alongside others such as gene-specific expression variability between samples, is incorporated into the modelling and statistical testing used in the initial differential gene expression analysis conducted through DESeq2, as outlined in 2.1.7 Bioinformatics analysis. As both expression datasets come from similar RNA-Seq experiments, this gene-level information could be incorporated by conducting the meta-analysis on both datasets as one experiment, comparing wildtype to LCD-mutants, with the addition of a study-specific covariate to identify the original dataset of each sample and take into account the technical variation resulting from the two different experiments. DESeq2 expression analysis of the M323K/Q331K mutants compared to wildtype with the covariate accounted for enabled the detection of 1,703 significant events in total, substantially more than in either individual dataset.

Figure 5.83 illustrates the changes in significant hits below, comparing the original FDR values from each dataset as above. Similarly to the Fisher p-value combination method, expression analysis of the two mutants with covariate addition again resulted in the loss of significance for some genes showing differences in expression patterns in the two models, shown in black. The genes detected as significant only upon covariate analysis, in pink, are also again mostly those approaching significance in both datasets. However, the clustering of these new hits is much more spread than upon p-value combination, and this is likely due to the additional incorporation of gene-level information in this analysis (Figure 5.83).

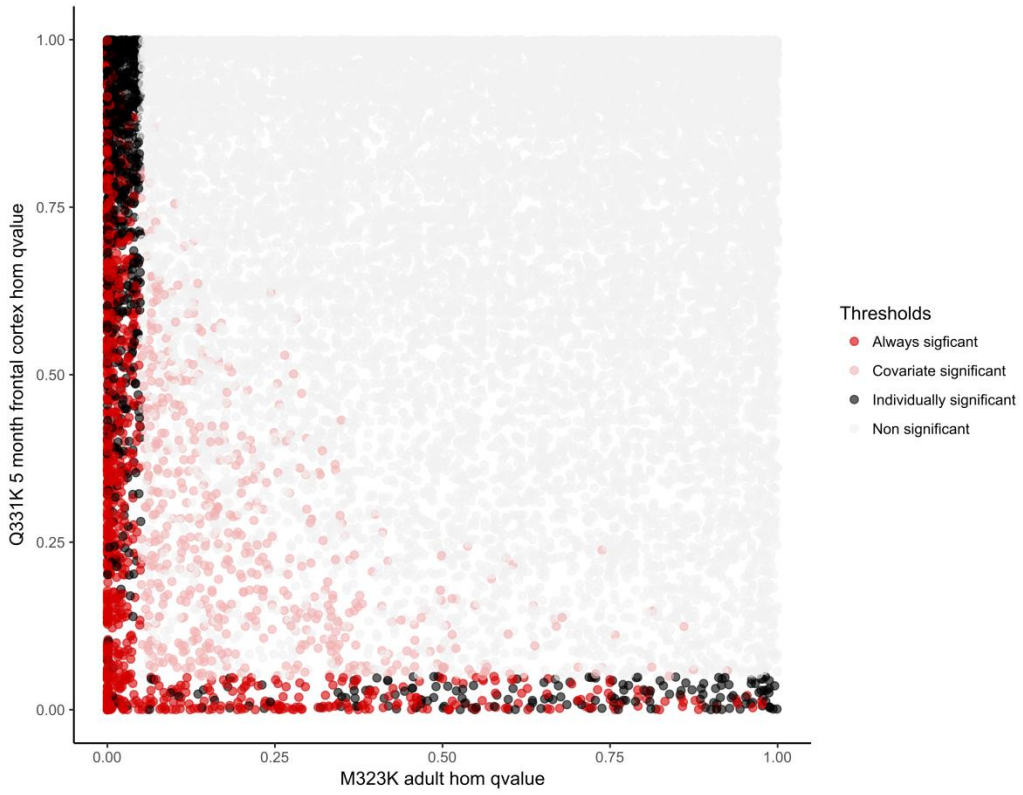


Figure 5.83. Adjusted p-values of all genes expressed in the homozygous adult M323K and homozygous 5 month Q331K RNA-Seq datasets compared, and the outcome of different tests for statistical significant illustrated. Red dots represent genes classed as significantly differently expressed in either individual dataset, and also upon joint analysis modelling with a covariate. In black are genes classed as significant in either individual dataset but not upon joint analysis, in pink are genes not classed as significant in either individual dataset but significant upon joint analysis, and in grey are genes which show no significant differential expression upon either method.

Overall the two meta-analysis methods clearly demonstrate the advantage of combining datasets to increase statistical power of differential gene expression detection in the M323K and Q331K models, most likely resulting from the overall similarity in gene expression changes in the two datasets. The new hits in each method are slightly different; a safe approach may be to focus on genes highlighted by both

methods. This work is currently in its early stages, further work is required to validate whether the newly identified genes by either method truly exhibit changes in expression, and also to understand whether the gene hits lost through combination are really false positives in the context of TDP-43 LCD mutation-caused gene expression changes. If confirmed, the above methods provide ability to study a much more substantial set of gene expression changes, which may aid the targeting of specific niche pathways otherwise missed if datasets were compared through simple overlap.

Following steps would also involve understanding the relevance of these non-model specific expression changes to disease, especially as gene expression dysregulation in the M323K model alone did not reveal any particularly affected pathways. The above methods would arguably prove more useful for the alternative splicing changes, given the stronger levels of correlation between the two datasets individually and the confirmed similarity in effect on a subset of TDP-43 targets. Perhaps this would also enable the detection of skiptic exons in both datasets, for which the statistical power of the Q331K dataset alone, due to the comparatively weaker sizes of change, may be insufficient.

5.3 ALS PATIENT POST-MORTEM TISSUE

Using the M323K and then the Q331K models, we established a pattern of alternative splicing in a subset of known TDP-43 targets which showed dysregulation upon LCD mutation. The direction of splicing changes in general was consistent between the two datasets, and specific events in genes such as *Sort1* and *Eif4h* were validated to be significantly changed in both models. However, whilst this demonstrated the impact of LCD and TDP-ALS mutation in mouse models, it was not yet clear whether these changes would be translated into patients. Neurodegenerative disease mouse models typically have a poor record of leading to therapeutics for patients (Eaton & Wishart

2017). Specifically looking at the RNA, transcriptomic changes in mouse models also can show poor levels of conservation when looking into patient tissue with similar settings (Burns et al. 2015). A specific example of this is the alternative splicing of *Sort1*. It is well characterised to be controlled by TDP-43 to produce two stable isoforms in mice, but the repression of the analogous splicing event in human *Sort1* is toxic (Prudencio et al. 2012). Therefore, it was of great importance to define precisely which of the disrupted RNA processing features from the M323K and Q331K mouse models were conserved in a human patient context.

Currently, there is no high quality published data of mRNA sequencing in post-mortem TDP-ALS patient tissue, although several transcriptomic investigations into post-mortem ALS tissue of other aetiologies have been conducted (Malaspina et al. 2001; Dangond et al. 2004; Jiang et al. 2005; Wang et al. 2006; Brockington et al. 2013; Highley et al. 2014). The design of these studies was not optimised for analysis of splice junctions, and as a result the presence of skiptic exons could not be analysed at a transcriptome-wide level. However, one of the key properties of skiptic exons is the conservation across mice and humans; therefore in theory the same skiptic exons detected in the adult diseased M323K mice could be occurring in post-mortem patient tissue.

Therefore, the seven skiptic exons validated by RT-PCR in the M323K spinal cord RNA – *Ankrd42, Herc2, Pacrgl, Pex16, Plod1, Slc6a6, Ube3c* (Figure 4.57) – were examined for in post-mortem ALS tissue. In total, post-mortem tissue-derived RNA was tested from 15 ALS, FTD or ALS-FTD patients alongside 3 controls. The tissue used for RNA extraction was the frontal cortex in each case bar one, in which the temporal lobe tissue was used. Patients had varied aetiology, with the disease cause of most not found in the most common ALS-linked genes, and *C9ORF72* mutation the primary culprit of the identified causative genes. Patients were also variable for TDP-43 pathology, in

keeping with the estimation that 50 % of FTD patients show TDP-43 pathology in comparison to the greater than 95 % of ALS patients. 4 further patients diagnosed with Pick's disease, who also displayed TDP-43 pathology, were included in the study. Of the seven targets – *ANKRD42*, *HERC2*, *PACRGL*, *PEX16*, *PLOD1*, *SLC6A6* and *UBE3C*, only two showed presence of skiptic isoforms: *SLC6A6* and *UBE3C*. *SLC6A6* skiptic isoforms were present in both controls and patient tissue however, and there was no apparent difference in skiptic isoform prevalence upon quantification. Contrastingly, only one sample of all tested displayed the skiptic isoform of *UBE3C*, but in this patient the skiptic isoform was very prominent (Figure 5.84). The shortened isoform in this sample fitted with the expected size of skiptic exon transcripts, but sequencing of the PCR product is required to confirm the finding. This RNA sample was obtained from the frontal cortex a 69 year old male patient diagnosed with ALS, with no further information available to determine why this sample alone would exhibit this skiptic exon.

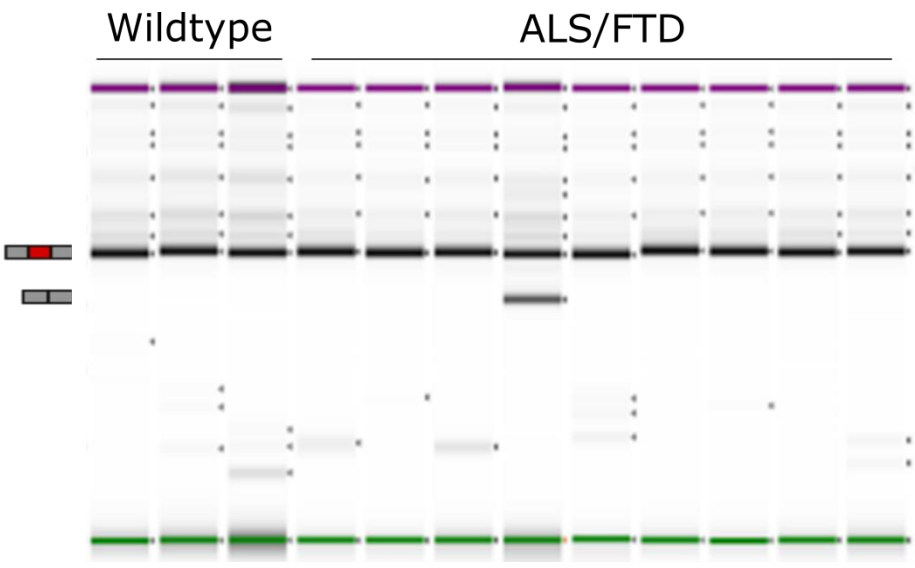


Figure 5.84. Tapestation gel showing RT-PCR products for tested skiptic exon in *UBE3C*. First three samples are controls, the next nine are ALS/FTD patient samples with various aetiology. Wildtype and skiptic isoforms annotated. Skiptic exon isoform prominently seen in one ALS sample alone.

The variability in skiptic exon expression in *UBE3C* in comparison to the M323K model may be partially accounted for by the much greater level of natural variation present in the general population versus inbred mouse models. Although TDP-43 splicing activity is strongly linked to control of *Ube3c* splicing in our M323K mice, several other regulators are also like to be simultaneously involved, as is common for certain splicing events (Appocher et al. 2017). Each of these factors would possess their own level of variability, therefore consistent levels of skiptic exon presence would be unexpected. Additionally, the nature of RNA preservation in post-mortem tissue can be inconsistent, although this is not typically the case provided appropriate storage protocols are followed (K. White et al. 2018). However, this could be a problem in particular for unstable transcript isoforms, such as those with skiptic exons as previously demonstrated by the predicted NMD fate in several examples. Any skiptic exon transcripts present may have been degraded prior to tissue preservation and not available for amplification by PCR.

Furthermore, it is important to note the heterogeneity of the samples also continues into the cause of disease. In the M323K model TDP-43 mutation was responsible for the splicing changes seen, but in these patients the causative mutation in most cases was unknown, and not found in the most common ALS-linked genes. Similarly, presence of TDP-43 pathology in these cases was not always clear. Therefore, it was not fully apparent what the expected splicing profile in these samples should be. However, the presence of an apparent skiptic exon isoform similar to the alternative splicing seen in the M323K model suggested that a similar disruption in TDP-43 splicing activity could be conserved in ALS patient tissue.

5.4 TDP-ALS PATIENT FIBROBLASTS

To gain a greater understanding for how precisely TDP-ALS mutations could influence splicing activity in patient tissue, fibroblasts were obtained from TDP-ALS patient skin biopsies. Fibroblasts were obtained from four TDP-ALS patients, three of whom possessed the A382T mutation (Kabashi, Paul N Valdmanis, et al. 2008), and one person the G298S mutation (Van Deerlin et al. 2008). Patients of both sexes were included, and one patient showed early onset ALS in comparison to the other three who had the more typical late onset (Figure 5.85). Four control lines were used for comparison.

Fibroblast line	*Mutation*	*Diagnosis*	*Age at onset*	*Site of onset**	*Gender*	*Age at biopsy*
TARDBP 1	G298S	ALS	62	LL	M	64
TARDBP 2	A382T	ALS	59	UL	F	62
TARDBP 3	A382T	ALS	25	LL	F	31
TARDBP 4	A382T	ALS	67	B	M	69
CTRL 1	-	Healthy	-	-	F	67
CTRL 2	-	Healthy	-	-	M	64
CTRL 3	-	Healthy	-	-	M	67
CTRL 4	-	Healthy	-	-	F	69

Figure 5.85. Sample details of TDP-ALS fibroblasts used. Site of onset, LL – lower limb, UL – upper limb, B – bulbar.

Fibroblasts were cultured as described in 2.4 Cell culture, and splicing patterns were tested as outlined in 2.1.4 Polymerase chain reaction. To first understand whether the gain of TDP-43 splicing function that occurred as a result of LCD mutation in the investigated mouse models held true in TDP-ALS patients with similar mutations, the splicing of validated mouse targets, such as *SORT1*, *EIF4H, ATXN2* and *DNAJC5*, was tested. Alternative splicing was confirmed in three of these targets, and two of the patient samples showed substantial increase in exon skipping in each of these three

transcripts (Figure 5.86). However, the other two patients' splicing profiles resembled the markedly consistent controls. This again showed the variability in true disease settings that is not replicated in inbred mouse models, but implied that in some patients the gain of TDP-43 splicing may hold true in a manner similar to the LCD mutant mouse models.

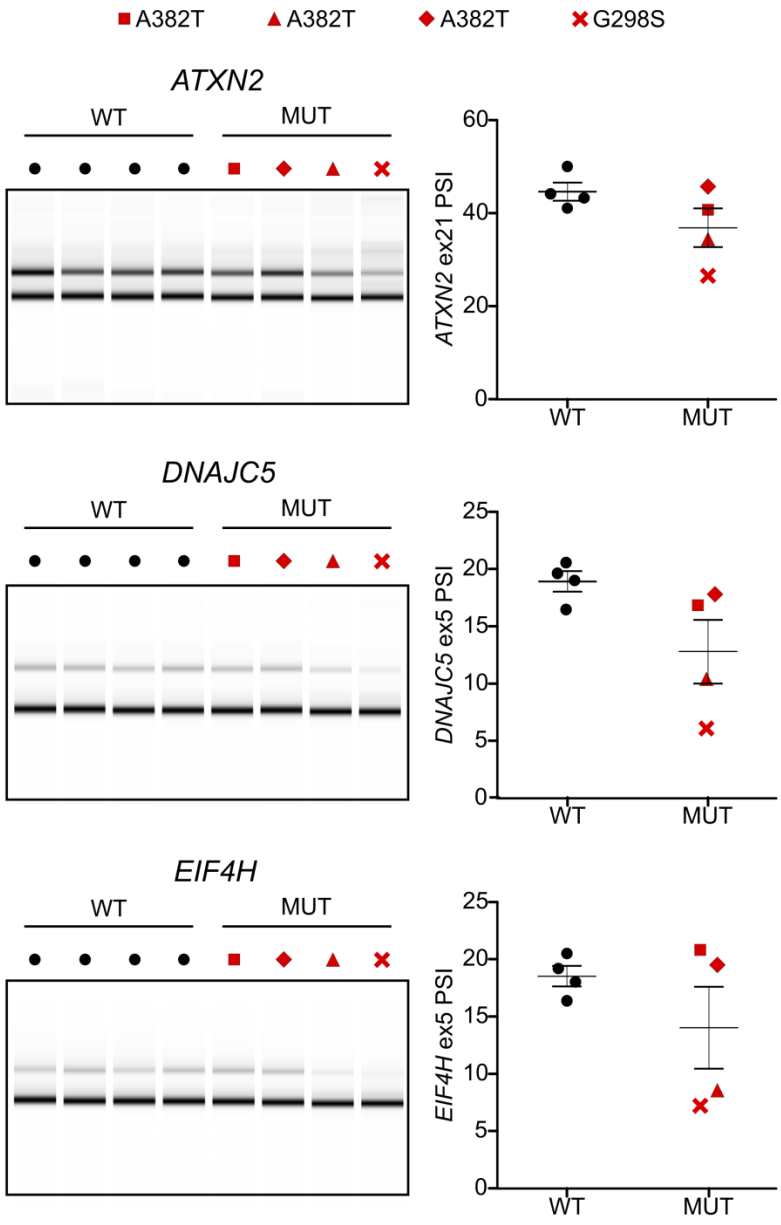


Figure 5.86. Tapestation gel and quantification of RT-PCR tests for alternative splicing event changes in *ATXN2*, *DNAJC5* and *EIF4H* in TDP-ALS fibroblast RNA

compared to wildtype. Two TDP-ALS samples showed reduced exon inclusion, but two showed wildtype levels, for each splicing event.

Following on from this finding, the presence of the extreme splicing skiptic exon events were also examined. Each of the seven initially validated targets were tested for, and of these, *ANKRD42, PACRGL, PLOD1* and *SLC6A6* showed the presence of the skiptic exon isoform. Crucially, upon quantification, two of these targets – *PLOD1* and *SLC6A6* – further showed increased presence of the skiptic exon isoform in all 4 patient samples in comparison to controls, with both changes statistically significant (Figure 5.87). These findings, although still somewhat preliminary, encouragingly demonstrate that a subset of the novel splicing events identified as skiptic exons in LCD mutant mouse models show similar, although reduced, skipping changes in TDP-ALS patients. Furthermore, these results also implicate gain of function in TDP-43 splicing activity as being caused by LCD mutations in both mouse and human settings, thereby encouraging further study of the mouse model transcriptome to find similarly conserved RNA processing changes in ALS patients.

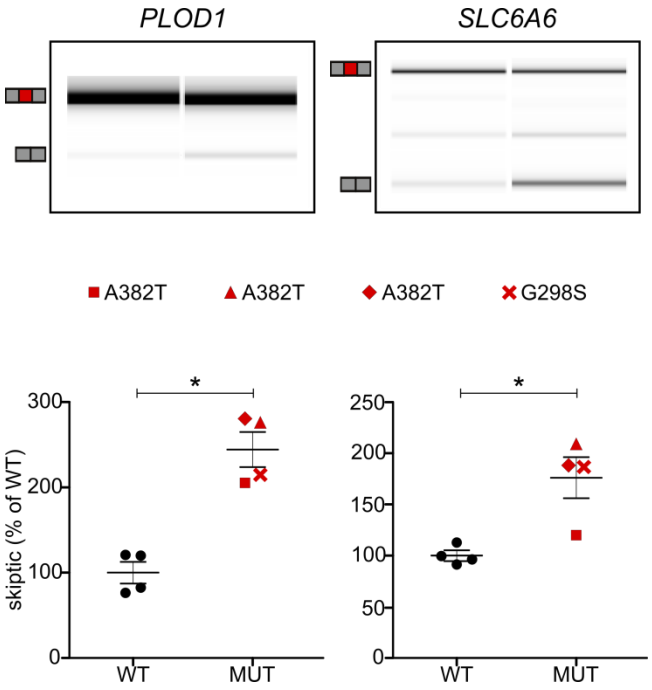


Figure 5.87. Tapestation gel and quantification of skiptic exon splicing in *PLOD1* and *SLC6A6* in TDP-ALS samples compared to wildtype. In both splicing events, all TDP-ALS samples showed significantly increased presence of the skiptic isoform in comparison to wildtypes.

The finding of conserved changes in alternative splicing of LCD-mutant mouse models also in TDP-ALS patient settings provides tremendous promise for the use of these mice as true models of disease. However, there remain many limitations to the translation of the mouse findings to a patient context. A key limitation of the aforementioned experiment is the use of fibroblasts for the splicing event profiling, a non-neuronal setting. ALS patients typically show impact on motor neurons specifically, in both the upper and lower motor neurons, suggesting either a motor neuron-specific change, or a change which for an unknown reason affects motor neurons above other cell types. In either case, studying motor neuron-specific changes is crucial to understanding the underlying mechanism. Additionally, if the extreme splicing changes resulting in novel transcripts like skiptic exons are of particular importance, then neuronal study is also vital due to transcriptome splicing showing

substantial cell-specific changes (Melé et al. 2015). Tied to this, the activity of hnRNPs, like TDP-43, are also changed in different cell types (McKee et al. 2005; Ling et al. 2010). It is important to also consider that the skiptic isoforms tested for, and those identified as conserved in human patients, are not necessarily implicated in the disease itself, rather they serve to highlight that the conservation exists.

Further work is required to examine the profile of all skiptic events identified in the M323K model with conserved exon skipping phenotype in the Q331K model. This is especially true for the E3 ubiquitin ligase linked genes *Herc2*, *Ube3c*, *Ube2e2*, *Wwp1*, *Ttc3* and *Wsb1*, of which only two have so far been tested. Perhaps the simultaneous dysregulation of several genes involved in the ubiquitination and subsequent protein degradation pathways would provide a clear link between the shared ubiquitin phenotypes in the disease models and patient settings. To investigate this further, high-throughput sequencing of the TDP-ALS mutant fibroblasts is required. Previous investigation into similar models has been conducted with microarray analysis (Highley et al. 2014), however high-depth RNA-Seq may be required for detection of the novel, rare, often unannotated skiptic exon events.

5.6 HNRNP A1 CASE STUDY

The overarching aim throughout this investigation has been to identify features of RNA processing which are dysregulated by TDP-43 mutation, and then further that the changes are conserved in TDP-ALS patients. Ideally, the changes should also themselves provide clues as to how the dysregulation could lead to a disease phenotype, rather than just a symptom of the mutation. Thus far, several examples of gene expression and splicing changes have been studied, and a number of these showed various levels of conservation in other models of loss or gain of TDP-43 function, in the Q331K mouse model, and in patient settings of post-mortem tissue or

non-neuronal cell types. The skiptic exons, a novel subset of transcripts shown to link LCD mutation effects in TDP-43 between the studied mouse models and patient settings, nevertheless do not fulfil the aforementioned aim entirely. Although the link to E3 ubiquitin ligases exists, and warrants further investigation, the precise connection between skiptic isoforms and how this could even lead to a phenotype at the cellular level remains unclear. Furthermore, although the skipping phenomenon persists, the reduction in Δ PSI in the Q331K model and in patient tissues raises the question of how impactful the changes would really be. However, one splicing change has met the stated aim, in its consistency and its presence in an ALS-related gene: the alternative splicing of *Hnrnpa1*.

Like TDP-43, hnRNP A1 is an RBP involved in several aspects of RNA processing (Mayeda et al. 1994; Bekenstein & Soreq 2013; Jean-Philippe et al. 2013), it also contains two RRM domains, and a C-terminal LCD domain, and it is primarily found in the nucleus. Crucially, in disease it is mislocalised in a similar manner to TDP-43 too, showing cytoplasmic mislocalisation (Kim et al. 2013). Finally, as with TDP-43, mutations in the LCD of *HNRNPA1* can cause ALS (Kim et al. 2013), although these mutations are exceptionally rare (Calini et al. 2013; Seelen et al. 2014). Therefore it is apparent that hnRNP A1 is a similar protein to TDP-43.

A recent study by Deshaies et al. (2018) investigated the splicing of *HNRNPA1* resulting in the formation of the protein isoform termed hnRNP A1-7B, a long isoform of hnRNP A1 that contained an extension of the LCD, elongated by 52 amino acids. This lengthening suggested a possible increased propensity for aggregation, and concordantly this hnRNP A1-7B was present in post-mortem ALS patient spinal motor neuron inclusions, although they were found distinct to TDP-43 aggregates, concordant with previous findings (Honda et al. 2015). Intriguingly, the transcript isoform that lead to translation of this longer protein was linked to loss of TDP-43 function.

Specifically, Deshaies et al. (2018) suggested that loss of TDP-43 reduced the repression of exon 7B, leading to its increased inclusion upon TDP-43 KD. These longer transcripts were then translated to produce greater levels of the aggregation prone hnRNP A1-7B protein.

The direct binding of TDP-43 to *HNRNPA1* transcripts in HeLa cells was determined by Deshaies et al. (2018) through immunoprecipitation of TDP-43 and subsequent RT-PCR of the RNA attached. They further predicted computationally that the binding site of TDP-43 to the *HNRNPA1* transcript would be surrounding the alternatively spliced exon 7B. To investigate this more thoroughly, we analysed our iCLIP datasets to determine the precise region of binding of TDP-43 to *Hnrnpa1* in mice. In our wildtype mouse iCLIP data, TDP-43 was bound to the 3'UTR of *Hnrnpa1* transcripts (Figure 5.88). Importantly, this binding region was also conserved in human TDP-43 iCLIP data (Tollervey et al. 2011; Rogelj et al. 2012), although in humans an additional binding region was identified in exon 8. There was, however, no TDP-43 binding detected in the intronic regions surrounding exon 7B, or in exon 7B itself.

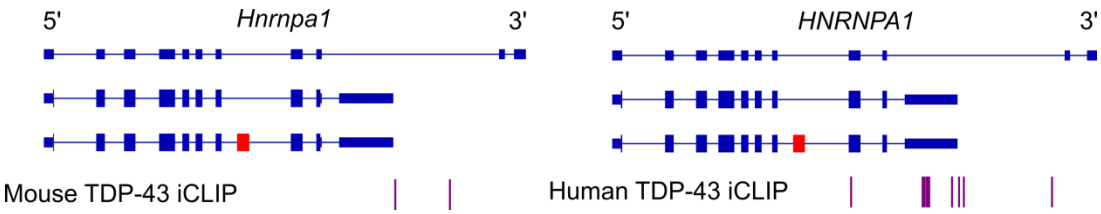


Figure 5.88. Mouse and human *Hnrnpa1/HNRNPA1* transcript isoforms. Alternatively spliced exon 7B highlighted in red. Respective mouse and human TDP-43 iCLIP data showing TDP-43 binding sites on *Hnrnpa1/HNRNPA1* shown below.

Furthermore, we thoroughly examined the hypothesis that loss of TDP-43 would lead to increased exon 7B inclusion in *HNRNPA1*. Two TDP-43 loss of function datasets were first utilised, our own homozygous embryonic F210I forebrain RNA-Seq dataset, and

the TDP-43 KD dataset from adult mouse striatum (Polymenidou et al. 2011). These two datasets have already been demonstrated to be highly concordant in terms of TDP-43 splicing of known targets, thereby allowing classification of the F210I mutation as resulting in a hypomorphic allele. Intriguingly, loss of TDP-43 splicing activity in these our F210I dataset was more typically linked with lost repression, and therefore increased exon inclusion levels, which would fit with the finding by Deshaies et al. However, in this case the opposite was true – in both the F210I and TDP-43 KD datasets, loss of TDP-43 function resulted in significantly lower levels of exon inclusion (Figure 5.89), the opposite of the findings by Deshaies et al.

We further analysed our RNA-Seq datasets for the adult M323K spinal cord, and the datasets for adult Q331K frontal cortex (M. A. White et al. 2018). These mutations have been demonstrated to cause a gain of function in TDP-43 splicing activity. Fittingly, in both the heterozygous and homozygous models, a dose dependent increase in exon 7B inclusion was detected in both mutants (Figure 5.89). Together, these six datasets comprehensively revealed that in mouse, *Hnrnpa1* exon 7B inclusion was enhanced by TDP-43, even more so upon TDP-43 gain of function, and repressed by TDP-43 loss of function.

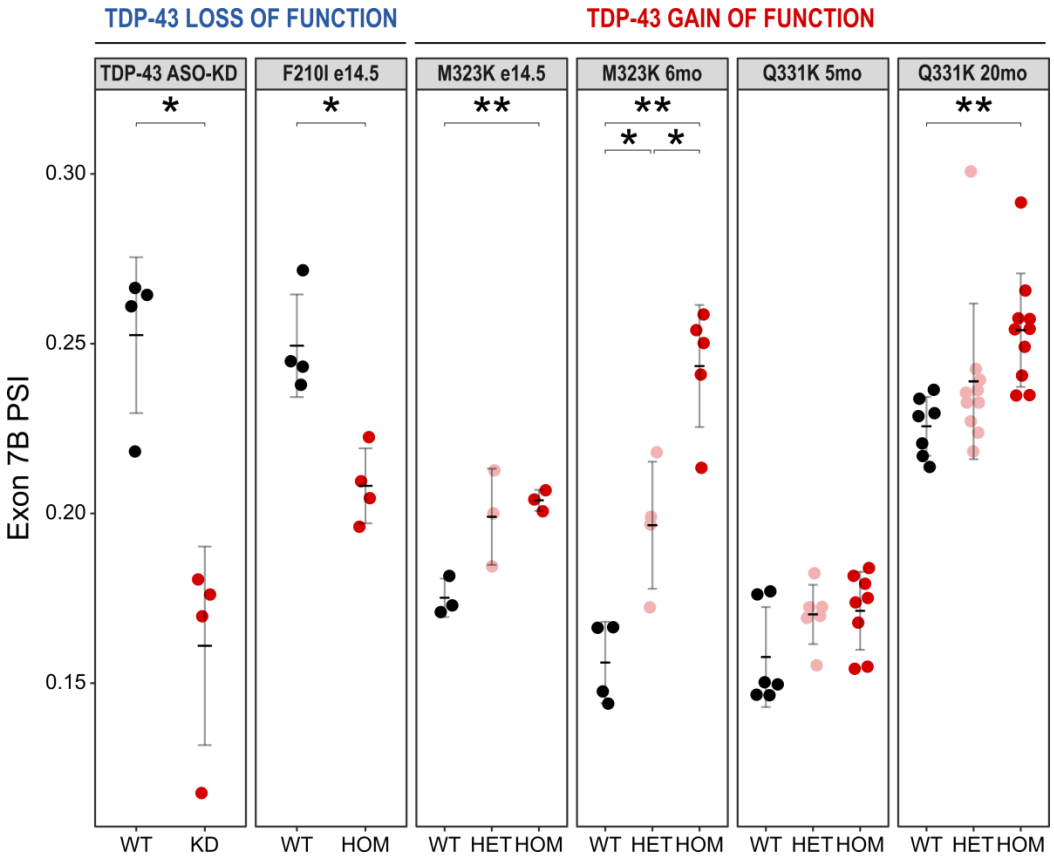


Figure 5.89. PSI values for *Hnrnpa1* exon 7B in six published RNA-Seq datasets: TDP-43 ASO-KD (Polymenidou et al. 2011), F210I embryonic day 14.5 frontal cortex, M323K embryonic day 14.5 head, M323K 6-month spinal cord, Q331K 5-month spinal cord and Q331K 20-month spinal cord (M. A. White et al. 2018), * p < 0.05, ** p < 0.01. Decreased exon 7B inclusion in TDP-43 loss of function models, in contrast to dose-dependent increase in exon 7B inclusion in M323K and Q331K TDP-43 gain of splicing function models.

Importantly, the work by Deshaies et al. was primarily in HeLa cells, a human setting, whereas the examined RNA-Seq datasets were in mice. Although unlikely given the substantial sequence conservation in both TDP-43 and *Hnrnpa1*, it was possible that in humans TDP-43 would function to have the opposite effect in *HNRNPA1* splicing. In order to test this, we additionally examined an RNA-Seq dataset from shRNA-mediated TDP-43 knockdown in motor neurons generated from wildtype human iPSCs (Kapeli et al. 2016). Concordant with the rest of our findings, the loss of TDP-43 resulted in

substantial reduction in *HNRNPA1* exon 7B inclusion (Figure 5.90), refuting the possibility that regulation of this splicing event by TDP-43 was opposite in mouse and human tissues.

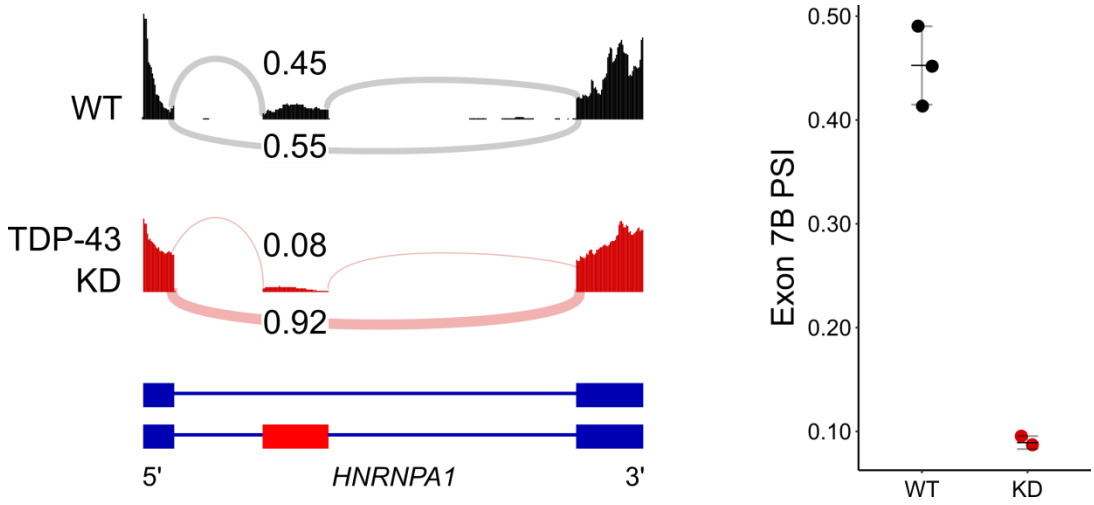


Figure 5.90. Left: representative RNA-Seq trace of wildtype and TDP-43 KD samples for *HNRNPA1* exon 7b, highlighted in red. Sashimi splicing schematic overlaid. Proportion of exon inclusion annotated above exon, proportion of junctions annotating skipping annotated below. Right: PSI quantification of *HNRNPA1* exon 7b in the human wildtype and TDP-43 KD dataset. Substantial increase in exon skipping as a result of TDP-43 KD illustrated in both cases.

Therefore, in summary the investigation of *HNRNPA1* splicing by TDP-43 convincingly contradicted the results published by Deshaies et al. In seven RNA-Seq datasets across mice and humans, loss of TDP-43 led to reduced inclusion of exon 7B in *Hnrnpa1*, and increased inclusion upon gain of TDP-43 splicing function. However, notably these findings agreed with Deshaies et al in that the increased prevalence of *HNRNPA1-7B* isoform was in the disease mouse models, M323K and Q331K, where Deshaies et al. found the increase primarily at the protein level in post-mortem ALS spinal cord motor neurons. Although the mechanism by which TDP-43 dysregulation leads to hnRNP A1-

7B increase may differ in the two studies, the link between the increased long isoform and disease is comparable (Sivakumar et al. 2018).

The details of the publicly available data can be found in these repositories:

Dataset	Repository	Samples	Read length	Sequencing depth
F210I/M323K	NCBI Sequence Read Archive reference SRP133158	See Chapter 2.1.6 RNA-Sequencing		
Mouse TDP- 43 KD	NCBI Sequence Read Archive reference SRP005860	4 WT, 4 KD	75 x2	~45M reads
Q331K	Gene Expression Omnibus reference GSE99354	5 months : 6 WT, 6 HET, 8 HOM 20 months: 7 WT, 10 HET, 10 HOM	100 x2	5 months: ~80M reads 20 months: ~100M reads
Human TDP-43 KD	Gene Expression Omnibus reference GSE77707	3 WT, 2 KD	50 x2	~30M reads

Table 5.22. Datasets used in analysis of *HNRNPA1* exon 7B splicing by TDP-43 with sequencing details and repository for each.

The high relevance of *HNRNPA1* to ALS heavily implicates this splicing event in particular to the mass RNA dysregulation seen upon TDP-43 dysregulation. The increased formation of the aggregate prone hnRNP A1-7B protein isoform may act as a secondary hit, leading to disrupted RNA processing of hnRNP A1 targets as well as the preceding gene expression and splicing changes caused by TDP-43 dysregulation. HnRNP A1 itself is known to be a vital protein, and like TDP-43 its transcript levels are tightly controlled through an autoregulatory mechanism (Suzuki & Matsuoka 2017). Complete loss of hnRNP A1 through KO in mice is embryonically lethal (Liu et al. 2017), while even low level of overexpression is toxic too (Suzuki & Matsuoka 2017), as is the case with TDP-43, demonstrating that this protein has functions which cannot be fully

compensated for. ALS-causative mutations in *HNRNPA1* can also occur in regions of the LCD only present in isoforms detected exclusively in CNS tissue (Q. Liu et al. 2016), providing a link between the RBP dysfunction and motor neuron-specific effects in ALS. Further study of TDP-43 interaction with other hnRNPs and RBPs is necessary to understand whether a possible 'domino' effect could be underlying the substantial RNA dysregulation culminating from TDP-43 mutation.

This finding is also highly relevant for the use of mouse models to study molecular dysregulation in human settings. The splicing of *HNRNPA1* exon 7B provides a perfect example of a splicing event that is so strongly controlled solely by TDP-43 that opposite changes in TDP-43 splicing activity alone are sufficient to significantly change the strength of the splicing. The splicing also not only held true in mouse models with different mutation, and all the technical variations that come with a different study, but remarkably also substantially so in human neuronal cells. This represents powerful evidence that TDP-43 splicing activity strongly determines *HNRNPA1* isoform production and eventual translation of an aggregate-prone protein variant, and therefore justifies the use of mouse models with mutations of the endogenous *Tardbp* gene for study of molecular signatures of neurodegeneration and ALS that can result from TDP-43 dysfunction (Fratta et al. 2018).

CHAPTER 6

DISCUSSION AND CONCLUSIONS

6.1 TDP-43 LOSS OF FUNCTION VS GAIN OF FUNCTION

In this thesis, the transcriptomic dysregulation in several models of TDP-43 dysfunction has been discussed. Primarily, analysis has focused on the F210I and M323K mutations, summarising previous research on the mutated proteins conducted in vitro and ex vivo, and outlining the current results from molecular characterisation of the mouse models. Initial investigation of the effects of the two mutations was carried out using the *CFTR* minigene assay, in order to elucidate the effects of either mutation on splicing activity. This assay, when conducted in MEFs obtained from homozygous mouse models with each of these mutations, revealed that the F210I mutation caused a loss of TDP-43-mediated repression of exon 9 in the minigene, suggesting it caused a loss of function in TDP-43 splicing. Conversely, the LCD mutation M323K caused an increase in exon 9 repression in the same assay, thereby implicating a surprising, opposite, gain of function in TDP-43 splicing activity. This characterisation of opposite effects in TDP-43 splicing due to the two mutations was solidified with further evidence of contrasting splicing changes in endogenous mouse TDP-43 splicing targets such as *Sort1* exon 18 and *Eif4h* exon 5.

Development of the mouse models with each mutation revealed the strong effects of both mutations in the developing embryonic stages of the mouse, with neither homozygous model viable on a pure C57BL/6J background. RNA sequencing of the homozygous embryo heads at E14.5 demonstrated that both mutations caused RNA dysregulation. In the F210I model, 1,184 genes were differentially expressed, and 67 splicing events were significantly changed. With the M323K model, the dysregulation was less, as 53 genes showed expressed changes and 11 events altered splicing.

However, with the ability of the homozygous M323K mice to survive until adulthood on a mixed background, the effect of LCD mutation in all TDP-43 protein in mature models could be studied. By the 6 month time point, the homozygous M323K spinal cord showed greater levels of RNA processing disruption, with 1,649 genes and 295 splicing events significantly altered. High-depth sequencing of the embryonic homozygous F210I forebrain revealed more substantial levels of gene expression and splicing changes too, with 2,730 and 230 features changed respectively. The large scale of change upon mutation of either the RBD or LCD followed previous findings which suggested both were important in TDP-43 RNA processing activity (Buratti & Baralle 2001; D'Ambrogio et al. 2009; Ayala et al. 2011).

Utilising the RNA-Seq datasets showing the greatest extent of RNA processing disruption in each model, the loss of function effects of F210I TDP-43, and the opposite gain of function effects of M323K TDP-43, were confirmed again in a subset of known TDP-43 splicing events, primarily resulting in increased exon inclusion in the F210I model and increased exon skipping in the M323K. This pattern was also found when exploring the extreme splicing changes caused by TDP-43, defined here as novel splicing events which resulted from the mutation and were absent in wildtype controls, and hence culminated in the formation of mutant-specific transcripts. The F210I-specific transcripts primarily showed inclusion of intronic regions, termed cryptic exons, whereas the M323K showed skipping of otherwise constitutive exons, here termed skiptic exons. Importantly, the loss of function characterisation of F210I was supported by similarly disturbed RNA processing upon TDP-43 KD in the *CFTR* minigene assay, splicing of endogenous mouse TDP-43 targets (Polymenidou et al. 2011), and the occurrence of cryptic exons (Ling et al. 2015). Similarly, the M323K gain of function in TDP-43 splicing activity was supported by similar results from the TDP-ALS mutant Q331K model, again in the same *CFTR* minigene assay, target genes such as

Sort1 and *Eif4h*, and skipping of the skiptic exons identified in M323K mice (M. A. White et al. 2018).

The reason for the loss of function and gain of function effects of the two mutations are difficult to fully explain. The F210I mutation affected the RNA binding domain of TDP-43, therefore the obvious expectation would be a change in RNA binding ability. Consistent with this, a reduction in TDP-43 binding to validated RNA targets – its target UG repeat motifs (Buratti & Baralle 2001), and its own *Tardbp* transcript binding region (Ayala et al. 2011) – demonstrated that the loss of TDP-43 function including RNA binding. As the F210I mutation is unlikely to affect a region directly binding to RNA itself (Ramos, 2015, personal communication), it is possible that a change in folding conformation of TDP-43 may result from the mutation. Linked to this, the regulatory role of RRM2 in TDP-43 RNA binding is likely to have been impaired. In particular, the alteration of a phenylalanine residue, often crucial for RNA binding, may be at the centre of the disturbed TDP-43 RNA processing function (Jones et al. 2001; Lukavsky et al. 2013). Currently, however, it is not yet clear as to what extent the general loss of TDP-43 binding to RNA targets is responsible for the loss of function splicing patterns seen.

As well as supporting the F210I mutation as being a hypomorphic allele, the comparison to TDP-43 KD or KO data also demonstrated that the F210I TDP-43 still retained some function and the F210I allele was not equivalent to the severity of full allele KO. This is shown by the development of a degree of motor phenotype in adult mouse models hemizygous for *Tardbp* (Wu et al. 2009; Kraemer et al. 2010; Sephton et al. 2010), but no phenotype in adult heterozygous F210I mice. Additionally, although homozygous F210I mice were not viable, they typically showed perinatal lethality, whereas complete KO of TDP-43 is lethal at the very early stages of embryonic development (Wu et al. 2009; Kraemer et al. 2010; Sephton et al. 2010). The use of this

less severe model of TDP-43 loss of function therefore allowed a more complete level of in vivo study into the resulting effects. Intriguingly, the resemblance of previously reported effects of TDP-43 KD in adult mouse striatum seen in the embryonic mutant F210I model implies that a large degree of TDP-43 function is consistent throughout development and into the adult mouse.

The gain of function effect of M323K is even more difficult to explain. Previously, the idea that LCD mutation of TDP-43 resulted in splicing changes opposite to those caused by loss of function was not known. However, during this investigation, this has become apparent across a number of targets and was shown to be true in the Q331K TDP-ALS mutation known to cause ALS in patients (Sreedharan et al. 2008). The mutation of the LCD domain suggests an impairment of protein-protein interaction (D'Ambrogio et al. 2009). The functions of TDP-43 affected by this would therefore likely be its interaction with splicing factors and other RBPs to control splicing events, and also its ability to dimerise, phase separate and form stress granules during the stress response. However, the impairment of TDP-43 regulation of RNA targets has previously been reported in the context of LCD mutation, and sequencing of TDP-ALS models with LCD mutation has also shown substantial RNA dysregulation (Arnold et al. 2013), although as previously alluded to, the proportion of these changes resulting from the transgenic overexpression of TDP-43 alone is not possible to determine.

Why the mutation would specifically cause the splicing role of TDP-43 to gain function is unclear. Perhaps the mutation causes TDP-43 to form a longer lasting interaction with target transcripts, resulting in greater inhibition of RNA Pol II splicing of the exons and culminating in greater exon skipping. However, no change in TDP-43 binding to target RNA sequences was seen in the same EMSA method that revealed the loss of binding due to F210I mutation. The mutation may be linked to a change in the recruitment of other hnRNPs to modulate splicing function; TDP-43 has been posited to

control directing of other hnRNPs with less high-affinity target motifs (D'Ambrogio et al. 2009).

Given the possibly different mechanisms by which the F210I and M323K mutations caused the respective loss of function and gain of function effects of TDP-43 activity, the global change of TDP-43 binding activity to target transcripts was tested in both models using iCLIP. However, analysis revealed that the well-established TDP-43 target UG motifs remained the sequences for which the mutant TDP-43 proteins held the highest affinity. In both models, correlation for pentamer sequence enrichment compared to control animals remained strong. Alongside this, TDP-43 is known to bind to different sub-regions, particularly depending upon protein localisation (Tollervey et al. 2011; Colombrita et al. 2012), but proportion of TDP-43 binding sites in each sub-region type closely resembled wildtype models. However, there was a noticeable reduction in intronic binding in the adult mouse forebrain compared to embryonic tissue. Currently, quantification of changes in TDP-43 binding at specific sites is difficult to determine. If an appropriate method can be found, then this data can be used to understand whether the loss of function splicing effects in the F210I mutant are caused by reduced binding of the protein to RNA targets, or whether the mutant TDP-43 exhibits reduced activity even when bound to targets. No difference in M323K TDP-43-RNA binding has been detected, but similar investigation could be useful to indicate how this mutation triggers the gain of function in TDP-43.

6.2 TDP-43 AUTOREGULATION

The TDP-43 autoregulatory mechanism by which it regulates translation levels of its own *Tardbp* transcripts is an intricate and complex process, involving differences in binding, splicing, polyA usage and expression level changes (Polymenidou et al. 2011; Ayala et al. 2011; Eréndira Avendaño-Vázquez et al. 2012; Koyama et al. 2016). In the F210I model, the loss of function effects were somewhat compensated for by TDP-43 autoregulation. The increased retention of intron 7 resulted in the upregulated formation of stable transcripts capable of being exported from the nucleus and eventually translated (Ayala et al. 2011). Similarly, the reduction in transcripts with long 3'UTR isoforms reduced the proportion of transcripts unavailable for translation, as these isoforms are thought to be retained in the nucleus (Koyama et al. 2016).

This forcing of long 3'UTR usage is thought to be at least partially due to excessive TDP-43 binding to the TDPBR and blocking of the pA_1 polyA site required to form translatable transcripts, therefore quantifiable iCLIP data would further the detection of TDP-43 binding changes to reduce the production of this long isoform in the F210I model, a feature that currently can only be detected visually by examined summed RNA reads. This dual mechanism for production of stable and exported *Tardbp* transcripts resulted in the expected increase in *Tardbp* transcript levels overall, suggesting the autoregulation mechanism was functioning to negate the loss of function effects of the F210I TDP-43. Although evidence of the autoregulatory mechanism in action was difficult to detect in the heterozygous F210I models, it remained the likely reason for the presence of one hypomorphic allele not causing any noticeable phenotype in the animal model.

In the case of the M323K TDP-43, the mutation caused the opposite expected effect. Given the gain of function in TDP-43 splicing activity, regulation of TDP-43 function back towards wildtype levels should require a reduction of *Tardbp* transcript levels, as

well as shifts in *Tardbp* splicing to enable this. Instead, the homozygous M323K model showed similar characteristics to the F210I model, with intron 7 retention again increased. The greater use of pA_1 instead of pA_2 and pA_4 would result in higher *Tardbp* levels, and concordantly significantly raised *Tardbp* expression was detected. The reason for the autoregulatory mechanism behaving in this way was not apparent, as this would theoretically cause a worsening of the gain of function molecular phenotype that the M323K model displayed. That the mutation of the LCD of TDP-43 can cause an apparent disturbance in the autoregulatory mechanism may provide clues as to how this process is impaired in TDP-ALS patients. The Q331K patient mutation has already been characterised as causing gain of function in TDP-43 splicing activity, like M323K, and the molecular environment of patient tissue is also known to harbour raised *TARDBP* levels (Mishra et al. 2007; Gitcho et al. 2009; Swarup, Phaneuf, Dupré, et al. 2011; Koyama et al. 2016), suggesting a similar counterproductive TDP-43 regulation process could be taking place.

The unwarranted increase in functional TDP-43 mRNA levels could lead to increased TDP-43 protein levels, a feature which was occasionally detected but inconsistently so in the LCD-mutant mouse tissues. The increased presence of TDP-43 may result in an accumulation, overwhelming a cell's systems to deal with this problem through protein degradation and culminating in the formation of TDP-43-positive aggregates in the cytoplasm, also positive for degradation factors such as p62 and ubiquitin. The aggregated TDP-43 would not be functional, therefore it may lead to a vicious cycle in which TDP-43 levels remain unnecessarily upregulated and lead to the aggregation of more and more TDP-43 protein. Although the M323K model developed a motor phenotype and neurodegeneration, it did not display any TDP-43 pathology. Therefore, the disrupted autoregulation does not immediately result in TDP-43 accumulation; the pathology hallmark of ALS may be not be a feature of early stages of the disease.

6.3 DIFFERENTIAL GENE EXPRESSION

Gene expression changes were widespread in the homozygous F210I model even at the early embryonic stages, while in the homozygous M323K model they were only evident in the adult mouse. Furthermore, the homozygous F210I mouse exhibited the specific downregulation of genes containing the longest introns. These genes were previously shown to be downregulated upon TDP-43 KD in the adult mouse striatum (Polymenidou et al. 2011), and so provided further evidence of the loss of function effects of the F210I mutation.

This feature of TDP-43 loss of function is particularly interesting in the context of neurodegenerative disease as many of the longest genes have been implicated in neuronal functions (Polymenidou et al. 2011), which follows from the demonstrated importance of TDP-43 in regulating transcripts involved in neuronal development (Sephton et al. 2011; Tollervey et al. 2011). Fittingly, GO enrichment analysis implicated pathways involved in neuronal function and ion transport. Therefore, this phenomenon provides a link between TDP-43 dysfunction and impact specific to neurons, including the upper and motor neurons most evidently affected in ALS. Furthermore, neuronal genes often require complex splicing (Zheng & Black 2013), and impaired TDP-43 splicing behaviour in the F210I model has been well established. Therefore, the loss of function of TDP-43 could impact both the expression and splicing of this subset of long genes enriched for genes involved in neuronal function.

No neuronal phenotype was detected in the F210I model, however a major caveat in the study of phenotype resulting from homozygosity for this mutation is that the models were perinatally lethal and the range of phenotypes that could be either displayed or studied was limited. It remains to be seen whether the long gene downregulation could underlie neuronal dysfunction and motor phenotype.

Although the homozygous M323K model did not show many changes in gene expression during development, by six months there were several hundred genes that could be detected by differential gene expression analysis. However, whereas the F210I model showed specific patterns of gene dysregulation, such as the enrichment of long genes, no such pattern was evident in the M323K model, and long gene dysregulation itself was not enriched or reduced in comparison to expected levels. GO enrichment analysis also failed to reveal any particularly effected pathways. Despite this, the presence of a subset of genes dysregulated by TDP-43 gain of function is likely to exist, as significant correlation was observed in the change in expression levels of genes dysregulated in either the M323K or Q331K models (M. A. White et al. 2018). Meta-analysis of these two models was further able to detect a substantially higher number of significantly differentially expressed genes, suggesting that gene dysregulation not identified as significant in either original model still showed similar direction of change. Initial GO enrichment analysis of the genes identified via meta-analysis using either the Fisher p-value combination method or the re-analysis with DESeq2 of both datasets jointly with the addition of the study-specific covariate to the model still failed to identify notably affected pathways. However, this analysis is still in the early stages, and further work is required to verify the accuracy of hits obtained, and lost, by either meta-analysis method. The derivation of an LCD-mutant specific subset of dysregulated genes would also provide higher accuracy in the search for candidate genes similarly dysregulated in TDP-ALS patients.

Intriguingly, the overlap in genes differentially expressed in both TDP-43 loss and gain of function in the F210I and M323K models was markedly poor. This phenomenon has previously been reported, in Drosophila, in which TDP-43 KD and overexpression primarily caused the dysregulation of a mostly different group of genes (Hazelett et al. 2012). In the case of these mouse models, this could also be as a result of the likely different mechanisms by which the mutations have induced functional change. The

F210I mutation shows overall reduction in RNA binding, suggesting all target transcripts could be affected. The M323K mutation, meanwhile, is likely to cause changes in TDP-43 interaction with other proteins, and in this case it is plausible that the interaction with some cofactors is affected more than others. However, this explanation is countered by the analysis of the compound heterozygous F210I/M323K model showing very few significantly dysregulated genes, implying that the mutations did indeed have an apparent counteracting effect. One feature that could be further investigated is the finding of genes that show matching direction of dysregulation upon both F210I and M323K mutation. These genes may represent failure of TDP-43 function upon LCD mutation, similar to the misregulation of the *Tardbp* transcript itself, and pinpoint candidate genes for analysis in patient settings.

6.4 ALTERNATIVE SPLICING

Alternative splicing analysis was the primary method by which both the F210I loss of function and M323K gain of function effects were characterised, firstly using the *CFTR* minigene assay and later using RT-PCR analysis of mutant TDP-43 splicing changes in known target transcripts. However, it was not until transcriptome-wide analysis of the splicing changes that the extent of changes in TDP-43 function was understood. Firstly, the increased exon inclusion representing loss of TDP-43 repression was further identified in the transcriptomic data for the F210I model in targets such as *Sort1* and *Eif4h*, and iCLIP data for TDP-43 confirmed the direct binding of TDP-43 at these splicing events. Importantly, the correlation between the F210I model and TDP-43 KD was further strengthened by the presence of the extreme splicing events resulting in cryptic exons in both models. The TDP-43 KD study by Ling, Pletnikova, Troncoso, & Wong, 2015, revealed the presence of supposedly intronic regions of transcripts being spliced into a subset of transcripts due to loss of TDP-43 repression upon KD, and a similar phenomenon was observed and validated in the F210I model. These cryptic

exons are predicted to be toxic, often causing frameshifts and a PTC, suggesting they are eventually degraded, however expression changes in these transcripts showed equal levels of up- and downregulation.

The M323K model RNA-Seq datasets also confirmed the mis-splicing of *Sort1* and *Eif4h*, in this case causing increased exon skipping compared to controls. Extreme splicing events were also identified in these models, but rather than cryptic exons, it was skipping of supposedly constitutive exons that were predominantly detected, here termed skiptic exons. These skiptic exon events, like the cryptic exons of the F210I, were often predicted to cause NMD of the transcript, but in this case enriched downregulation of skiptic exon transcripts was also detected. Whilst the Q331K model did not exhibit skiptic exons to the same extent as the M323K model, the majority of skiptic events from the M323K did at least show some level of skipping in the Q331K. This again suggests that the difference in TDP-43 splicing in the two mutations may be due to a weaker effect of the Q331K mutation.

Importantly, comparison of TDP-43 binding proximity to both cryptic and skiptic exons in each of the F210I and M323K models showed notably similar patterns, again suggesting that the difference in TDP-43 splicing of these events in the two mutants was not caused by changes in binding targets, but changes in functioning when bound.

The progression of study of these splicing events of interest in human settings provided mixed results. As well as the differences in TDP-43 function in difference species, the heterogeneity of TDP-ALS patients in comparison to inbred mouse models added another level of variability. Splicing of confirmed mouse TDP-43 targets in humans revealed similar patterns of exon skipping as a result of LCD mutation in some patients, but others with the same mutation showed splicing levels comparable to controls. However, despite this heterogeneity, skiptic exons isoforms were investigated in human settings due to the high level of conservation of these exons between mice

and humans, in comparison to the negligible levels of correlation in cryptic exons between species. Indeed, skiptic exon isoforms were consistently detected, and further, increased skipping identified, in TDP-ALS fibroblasts. This validated the use of mouse models with endogenous mutations, such as the M323K, for study of molecular dysregulation caused by LCD mutations of TDP-43.

Of the seven mouse skiptic targets tested, four of the splicing events resulting in skiptic isoforms were detected in humans, and two showed significantly increased skipping in TDP-ALS patients. This represents a remarkably high level of consistency between mouse and human TDP-43-LCD mutation effects, and further work to test the presence of similar skiptic exon events in patient settings is warranted.

This is particularly true for the set of genes coding for E3 ubiquitin ligases. This diverse group of genes, involved in both the direct and indirect ubiquitination of proteins for degradation (Berndsen & Wolberger 2014), have repeatedly been implicated in this study. Six of the 47 skiptic events were in genes involved in this pathway -- *Herc2, Ube3c, Ube2e2, Wwp1, Ttc3* and *Wsb1* – and alongside this the top hits for the overlap analysis of features dysregulated by both M323K and Q331K mutation were genes involved in similar ubiquitin ligase functions – *Spop* and *Gid4* by differential gene expression and *Ttc3* by alternative splicing changes. The dysregulation of a number of these genes, combined with the development of ubiquitin pathology in the M323K CNS tissue by the 2 year time point, heavily implicates this pathway as dysregulated. Previous transcriptomic meta-analysis of pathways affected in ALS has also identified related affected pathways, such as protein turnover (Cooper-Knock et al. 2012). The regulation of this group of genes next requires analysis in TDP-ALS patient CNS tissues, which also typically show ubiquitin pathology (Arai et al. 2006; Neumann et al. 2006).

The ability to find a splicing change as a result of TDP-43 mutation that showed expected effects in both loss of function and gain of function models was hampered by

substantial variation in every aspect of study, in both the technical variation caused by differences in experimental procedures, library preparation and sequencing specifications, and the biological variation attributed to different species, different tissues and different ages of models. Incredibly, despite all of these confounding variables, the splicing of exon 7B of *Hnrnpa1* was found to be consistently altered as expected by TDP-43 dysfunction. In loss of function models such as TDP-43 KD and F210I mutation in mouse models, a reduction in *Hnrnpa1* exon 7B inclusion was detected, suggesting that TDP-43 functions to enhance exon inclusion. Concordantly, in gain of function models such as the M323K mouse at the developing embryonic and adult time points, and the Q331K model at the adult and aged time points, exon 7B was increasingly included, and the effect was dose dependent, suggesting the effect was not eradicated by any TDP-43 autoregulation. This link between TDP-43 and *HNRNPA1* splicing was further established in human loss of function models, where TDP-43 KD once again caused significant and substantial reduction in exon 7B inclusion in *HNRNPA1* transcripts (Sivakumar et al. 2018).

Together these results presented a splicing event in an ALS-linked protein, hnRNP A1, which is strongly linked to TDP-43 function and dysregulation upon change in TDP-43 function (Kim et al. 2013; Deshaies et al. 2018). Further investigation of the splicing and gene expression levels of other hnRNPs is warranted, to gain a more complete understanding of the interplay involved and how TDP-43 dysfunction may indirectly affect the RNA processing of several more transcripts. Importantly, this *HNRNPA1* splicing also represents clear validation in the use of mouse models for study of patient disease, including the use of those not even containing TDP-ALS patient mutations, such as the M323K model.

6.5 PHENOTYPE

Unlike the molecular and transcriptomic analysis, comprehensive phenotype comparison between the effects of TDP-43 loss and gain of function could not be completed. This was because the homozygous F210I model, in which the TDP-43 loss of function in splicing was so evident, was perinatally lethal, and so only the developmental phenotype could be studied. In this context, substantial developmental defects were identified, including size and weight deficiency, alongside a two day developmental delay in the eye, bone and cleft palate. This added to the well-supported idea that TDP-43 is crucial for normal development. However, no CNS or neuronal specific deficiencies were identified in the embryonic model, therefore commenting on the phenotype of TDP-43 loss of function that may be linked to ALS is difficult. Importantly, the substantial effects of TDP-43 loss are not restricted to developmental stages, with KO of TDP-43 in mature neurons causing toxicity and neurodegeneration (Iguchi et al. 2013). This implies that, were the homozygous F210I adult models available for study, they could show a similar phenotype.

The survival of the homozygous M323K model on a mixed C57BL/6J / DBA/2J mixed background enabled investigation of phenotype resulting from homozygous LCD mutation, the ENU model more closely resembling *TARDBP* mutation in ALS than the F210I. Developmental phenotypes in this model were not evident, suggesting the gain of function effects may not have had the same severity as the F210I loss of function effects. However, the six month M323K model showed substantial RNA dysregulation, and this preceded the development of a motor phenotype in the two year model. In this aged model, grip strength deficits, muscle force reduction, motor unit count reduction, and motor neuron count reduction were all detected, along with p62 and ubiquitin pathology in the ventral horns of the spinal cord and also the brainstem.

These are all signs of neurodegeneration and linked to an ALS phenotype, however absent in this model were the TDP-43 pathology and reduced lifespan also associated with ALS in human disease, raising questions as to how applicable the underlying molecular changes are in true cases of ALS. However, not all ALS cases display TPD-43 pathology, such as those caused by mutations in SOD1 or FUS, and phenotype as in the disease can also prove variable, with a small subset of patients surviving far longer than the typical 3-5 years post-disease onset. Additionally, other mouse models of disease have shown similar differences to ALS pathology. The development of a motor phenotype in the absence of TDP-43 nuclear depletion and mislocalisation has been previously reported in a transgenic TDP-ALS model (Arnold et al. 2013). Importantly, the signs of neurodegeneration were evident but slight, with grip strength deficit the only detected motor change and the motor neuron loss size below the threshold predicted to cause substantial behavioural changes. These features, alongside the lack of change in lifespan, suggest that the M323K model represents the early stages of neurodegeneration in human disease. Linked with this, perhaps the subsequent TDP-43 mislocalisation and TDP-43 loss of function provides the 'second hit' to result in the phenotype associated with latter stages of neurodegenerative disease. Review of the published work documented in this thesis has led to similar conclusions (Rouaux et al. 2018).

If TDP-43 pathology would eventually result, plausible given the counterproductive functioning of the TDP-43 autoregulatory mechanism, then this may represent the late stage of the disease in human patients. The nuclear depletion that accompanies TDP-43 mislocalisation and aggregation would also then cause loss of function effects of TDP-43, giving rise to the long gene downregulation, mis-splicing, and cryptic exon inclusions all associated with the loss of function F210I model. This may be a 'second hit', worsening the molecular dysregulation and consequential disease phenotype. The effects of this loss of function may also be more severe, hence the severity of the F210I

model above the M323K counterpart. The suggestion from this sequence is that the F210I model may demonstrate the late stage effects of TDP-43 proteinopathies, however the vital conclusion from this data is that even in the absence of the TDP-43 loss of function effects, TDP-43 gain of function without TDP-43 aggregation, is sufficient to cause a neurodegenerative, ALS-like phenotype.

6.6 FUTURE WORK

The outcomes of this study have provided several insights into TDP-ALS. The misregulation of RNA processing upon TDP-43 mutation, the disentagling of distinct effects of TDP-43 loss and gain of function, and the impact of both forms of dysregulation in disease. However, much of the work described is preliminary, and requires building upon to form a more complete understanding of the specific mechanisms involved.

One area of key interest is the identification of extreme splicing events in the form of skiptic exons, resulting from LCD mutation of TDP-43. That these events form mutant-specific transcripts could allow them to be tested for as potential biomarkers, particularly as a subset of these was shown to occur in TDP-ALS patient fibroblasts. How these, or cryptic exon splicing events, could result in pathogenicity is unclear, it may be the mis-splicing of particular transcripts that is important, or the general pattern of splicing resulting in increased transcripts for degradation which may prove toxic at the cellular level. If the former is true, then in particular the E3 ubiquitin ligases should be studied further. Although the classification of this group is not precise, several transcripts coding for proteins of similar function have been identified as dysregulated in the LCD-mutant models. A more comprehensive study of proteins involved in the ubiquitination and protein degradation pathways in human tissues exhibiting ubiquitin pathology may show similar signs of dysregulation in many other ubiquitin ligase transcripts. The proteins translated from these skiptic exon transcripts have also thus far been poorly studied, the impact of predicted downregulation in a number of these transcripts at a protein level would provide further evidence implicating this pathway in the consequential ubiquitin pathology.

A more conclusive subset of skiptic exon transcripts could also be formed through meta-analysis of the M323K model with other LCD-mutant models as was conducted

with the Q331K, once they become available. In the Q331K mice, a select few skiptic exons from the M323K model showed increased levels of inclusion rather than the expected skipping. These skiptic exons may not be of interest in human settings given their variation even at the mouse level, and this refinement would enable a more streamlined analysis of skiptic exons in TDP-ALS tissues.

The meta-analysis of LCD-mutant mouse models demonstrated that both differential gene expression and alternative splicing changes showed substantial similarity, suggesting a common subset of features dysregulated regardless of the technical variations accompanying different studies. Currently, the identity of most of the genes and splicing events contributing to this correlation have not been identified, selection of the most well-correlated targets would present a priority set of targets to be tested in TDP-ALS settings in comparison to changes from the M323K or Q331K models alone. The meta-analysis method would also enable greater statistical power to detect small or variable differences, allowing the detection of several features of RNA processing dysregulation which would have been otherwise missed, although the importance of small or variable changes in the context of disease can be debated.

One area which could be improved upon is the settings in which these analyses were conducted. Each model had its drawbacks, the mouse models being a different species, the post-mortem tissue disadvantaged by possible RNA degradation and loss of unstable transcripts, and the tissues available being of mixed ALS aetiology, and the TDP-ALS fibroblasts not being a neuronal cell type and therefore not being able to fully recapitulate the most important changes in ALS patients.

The model that may to be closest to living ALS patient systems are motor neurons differentiated from iPSCs reprogrammed from TDP-ALS patients. These would negate the aforementioned drawbacks of models currently used. However, this setting too has flaws, with the likely use of immature motor neurons rather than the mature, aged

neurons of disease, and also the inconsistent presence of other cell types, such as astrocytes and oligodendrocytes, also thought to mediate motor neuron toxicity in ALS (Meyer et al. 2014; Ferraiuolo et al. 2016; Hall et al. 2017). Each of the RNA processing features described here could then be investigated in this setting, and if present could be proposed with higher confidence of occurring in TDP-ALS patients. A plan to analyse exactly this sequencing, using multiple differentiations from multiple clones of TDP-ALS iPSC-derived motor neurons from four patients and grown in two different institutions, is currently underway. The consequences of RNA dysregulation changes conserved between the mouse models studied here and the heterogeneous group of iPSC-derived motor neurons would be at the forefront of protein changes to be tested for in patient tissue. It remains to be seen whether any other splicing changes show remarkable levels of conservation similar to the *HNRNPA1* exon 7B splicing of TDP-43.

The aims of this investigation were primarily focused on understanding the different roles of TDP-43 dysfunction on RNA processing, and how this may be linked to disease. However, the overarching aim of this work, and translational study in general, would be to enable some level of development of therapeutics for disease. In regards to this, the highlighting of RNA processing dysfunction in disease, and the burgeoning importance of RBPs in several diseases, have led to the founding of biotechnology start-ups prioritising the task of correcting the RNA expression and splicing changes with the goal of finding possible therapeutic targets. Once such biotech founded earlier this year, Skyhawk Therapeutics, specifically lists correction of exon skipping dysregulation in neurological disorders among its aims, and has received a sizeable $8 M in seed funding to proceed with this. The venture of pharmaceutical companies into this field highlights the growing understanding of importance of RNA homeostasis disruption in disease. Ideally future patients unfortunate enough to develop ALS will be able to benefit from published research, such as the findings discussed in this thesis (Fratta et al. 2018; Sivakumar et al. 2018), and the investment that follows.

BIBLIOGRAPHY

Acevedo-Arozena, A. et al., 2008. ENU Mutagenesis, a Way Forward to Understand Gene Function. *Annual Review of Genomics and Human Genetics*, 9(1), pp.49–69.

Alami, N.H. et al., 2014. *Axonal Transport of TDP-43 mRNA Granules Is Impaired by ALS-Causing Mutations*,

Amador-Ortiz, C. et al., 2007. TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Annals of Neurology*, 61(5), pp.435–445.

Anders, S., Pyl, P.T. & Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), pp.166–169.

Anders, S., Reyes, A. & Huber, W., 2012. Detecting differential usage of exons from RNA-seq data. *Genome research*, 22(10), pp.2008–17.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Appocher, C. et al., 2017. Major hnRNP proteins act as general TDP-43 functional modifiers both in Drosophila and human neuronal cells. *Nucleic acids research*, 45(13), pp.8026–8045.

Arai, T. et al., 2006. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochemical and biophysical research communications*, 351(3), pp.602–11.

Ardley, H.C. & Robinson, P.A., 2005. E3 ubiquitin ligases. *Essays in biochemistry*, 41, pp.15–30.

Arnold, E.S. et al., 2013. ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *Proceedings of the National Academy of Sciences*, 110(8).

Ash, P.E.A. et al., 2010. Neurotoxic effects of TDP-43 overexpression in C. elegans. *Human Molecular Genetics*, 19(16), pp.3206–3218.

Austin, J.A. et al., 2014. Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proceedings of the National Academy of Sciences of the United States of America*, 111(11), pp.4309–14.

Ayala, Y.M. et al., 2005. Human, Drosophila, and C.elegans TDP43: Nucleic Acid Binding Properties and Splicing Regulatory Function. *Journal of Molecular Biology*, 348(3), pp.575–588.

Ayala, Y.M. et al., 2011. TDP-43 regulates its mRNA levels through a negative feedback loop. *The EMBO journal*, 30(2), pp.277–88.

Ayala, Y.M., Misteli, T. & Baralle, F.E., 2008. TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. *Proceedings of the National Academy of Sciences*, 105(10), pp.3785–3789.

Ayala, Y.M., Pagani, F. & Baralle, F.E., 2006. TDP43 depletion rescues aberrant CFTR

exon 9 skipping. *FEBS Letters*, 580(5), pp.1339–1344.

Barmada, S.J. et al., 2010. Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *The Journal of Neuroscience*, 30(2), pp.639–49.

Baughman, R.W. & Gilbert, C.D., 1981. Aspartate and glutamate as possible neurotransmitters in the visual cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 1(4), pp.427–39.

Bekenstein, U. & Soreq, H., 2013. Heterogeneous nuclear ribonucleoprotein A1 in health and neurodegenerative disease: from structural insights to post-transcriptional regulatory roles. *Molecular and cellular neurosciences*, 56, pp.436–46.

Bembich, S. et al., 2014. Predominance of spliceosomal complex formation over polyadenylation site selection in TDP-43 autoregulation. *Nucleic acids research*, 42(5), pp.3362–71.

Bensimon, G., Lacomblez, L. & Meininger, V., 1994. A Controlled Trial of Riluzole in Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, 330(9), pp.585–591.

Bentmann, E. et al., 2012. Requirements for stress granule recruitment of fused in sarcoma (FUS) and TAR DNA-binding protein of 43 kDa (TDP-43). *The Journal of biological chemistry*, 287(27), pp.23079–94.

Berndsen, C.E. & Wolberger, C., 2014. New insights into ubiquitin E3 ligase mechanism. *Nature Structural & Molecular Biology*, 21(4), pp.301–307.

Bjørkøy, G. et al., 2005. p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *The Journal of Cell Biology*, 171(4), pp.603–614.

Bramham, C.R. & Wells, D.G., 2007. Dendritic mRNA: transport, translation and function. *Nature reviews. Neuroscience*, 8(10), pp.776–89.

Brockington, A. et al., 2013. Unravelling the enigma of selective vulnerability in neurodegeneration: motor neurons resistant to degeneration in ALS show distinct gene expression characteristics and decreased susceptibility to excitotoxicity. *Acta Neuropathologica*, 125(1), pp.95–109.

Vanden Broeck, L. et al., 2013. TDP-43 loss-of-function causes neuronal loss due to defective steroid receptor-mediated gene program switching in Drosophila. *Cell reports*, 3(1), pp.160–72.

Buratti, E., 2015. Functional Significance of TDP-43 Mutations in Disease. In pp. 1–53.

Buratti, E. et al., 2001. Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *The EMBO journal*, 20(7), pp.1774–84.

Buratti, E. et al., 2004. Nuclear Factor TDP-43 Binds to the Polymorphic TG Repeats in CFTR Intron 8 and Causes Skipping of Exon 9: A Functional Link with Disease Penetrance. *The American Journal of Human Genetics*, 74(6), pp.1322–1325.

Buratti, E. et al., 2007. SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer. *Nucleic Acids Research*,

35(13), pp.4359–4368.

Buratti, E. & Baralle, F.E., 2001. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *The Journal of biological chemistry*, 276(39), pp.36337–43.

Burns, T.C. et al., 2015. Mouse models rarely mimic the transcriptome of human neurodegenerative diseases: A systematic bioinformatics-based critique of preclinical models. *European journal of pharmacology*, 759, pp.101–17.

Byrne, S. et al., 2011. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry*, 82(6), pp.623–627.

Calini, D. et al., 2013. Analysis of hnRNPA1, A2/B1, and A3 genes in patients with amyotrophic lateral sclerosis. *Neurobiology of Aging*, 34(11), p.2695.e11-2695.e12.

Che, M.-X. et al., 2015. TDP-35 sequesters TDP-43 into cytoplasmic inclusions through binding with RNA. *FEBS letters*, 589(15), pp.1920–8.

Chen, A.K.-H. et al., 2010. Induction of Amyloid Fibrils by the C-Terminal Fragments of TDP-43 in Amyotrophic Lateral Sclerosis. *Journal of the American Chemical Society*, 132(4), pp.1186–1187.

Chiang, C.-H. et al., 2016. Structural analysis of disease-related TDP-43 D169G mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation. *Scientific reports*, 6, p.21581.

Chiang, P.-M. et al., 2010. Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism. *Proceedings of the National Academy of Sciences*, 107(37), pp.16320–16324.

Cohen, T.J. et al., 2015. An acetylation switch controls TDP-43 function and aggregation propensity. *Nature Communications*, 6(1), p.5845.

Colombrita, C. et al., 2012. TDP-43 and FUS RNA-binding Proteins Bind Distinct Sets of Cytoplasmic Messenger RNAs and Differently Regulate Their Post-transcriptional Fate in Motoneuron-like Cells. *Journal of Biological Chemistry*, 287(19), pp.15635–15647.

Colombrita, C. et al., 2009. TDP-43 is recruited to stress granules in conditions of oxidative insult. *Journal of Neurochemistry*, 111(4), pp.1051–1061.

Conicella, A.E. et al., 2016. ALS Mutations Disrupt Phase Separation Mediated by α-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain. *Structure*, 24(9), pp.1537–1549.

De Conti, L. et al., 2015. TDP-43 affects splicing profiles and isoform production of genes involved in the apoptotic and mitotic cellular pathways. *Nucleic acids research*, 43(18), pp.8990–9005.

Conway, J.R. et al., 2017. UpSetR: an R package for the visualization of intersecting sets and their properties J. Hancock, ed. *Bioinformatics*, 33(18), pp.2938–2940.

Cooper-Knock, J. et al., 2012. Gene expression profiling in human neurodegenerative disease. *Nature Reviews Neurology*, 8(9), pp.518–530.

Curk, T., 2010. iCount. , p.http://icount.biolab.si.

Cusack, B.P. et al., 2011. Preventing Dangerous Nonsense: Selection for Robustness to Transcriptional Error in Human Genes J. Zhang, ed. *PLoS Genetics*, 7(10), p.e1002276.

D'Ambrogio, A. et al., 2009. Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic acids research*, 37(12), pp.4116–26.

Dangond, F. et al., 2004. Molecular signature of late-stage human ALS revealed by expression profiling of postmortem spinal cord gray matter. *Physiological Genomics*, 16(2), pp.229–239.

Van Deerlin, V.M. et al., 2008. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. *The Lancet. Neurology*, 7(5), pp.409–16.

Deshaies, J.E. et al., 2018. TDP-43 regulates the alternative splicing of hnRNP A1 to yield an aggregation-prone variant in amyotrophic lateral sclerosis. *Brain*, 141(5), pp.1320–1333.

Devoy, A. et al., 2017. Humanized mutant FUS drives progressive motor neuron degeneration without aggregation in 'FUSDelta14' knockin mice. *Brain*, 140(11), pp.2797–2805.

Dewey, C.M. et al., 2011. TDP-43 Is Directed to Stress Granules by Sorbitol, a Novel Physiological Osmotic and Oxidative Stressor. *Molecular and Cellular Biology*, 31(5), pp.1098–1108.

Diaper, D.C. et al., 2013. Loss and gain of Drosophila TDP-43 impair synaptic efficacy and motor control leading to age-related neurodegeneration by loss-of-function phenotypes. *Human Molecular Genetics*, 22(8), pp.1539–1557.

Dobin, A. et al., 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, 29(1), pp.15–21.

Eaton, S.L. & Wishart, T.M., 2017. Bridging the gap: large animal models in neurodegenerative research. *Mammalian genome : official journal of the International Mammalian Genome Society*, 28(7–8), pp.324–337.

ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), pp.57–74.

Eréndira Avendaño-Vázquez, S. et al., 2012. Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes and Development*, 26(15), pp.1679–1684.

Fallini, C., Bassell, G.J. & Rossoll, W., 2012. The ALS disease protein TDP-43 is actively transported in motor neuron axons and regulates axon outgrowth. *Human Molecular Genetics*, 21(16), pp.3703–3718.

Fecto, F. et al., 2011. <emph type="ital">SQSTM1</emph> Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis. *Archives of Neurology*, 68(11), p.1440.

Feiguin, F. et al., 2009. Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior. *FEBS letters*, 583(10), pp.1586–92.

Ferraiuolo, L. et al., 2011. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nature Reviews Neurology*, 7(11), pp.616–630.

Ferraiuolo, L. et al., 2016. Oligodendrocytes contribute to motor neuron death in ALS via SOD1-dependent mechanism. *Proceedings of the National Academy of Sciences*, 113(42), pp.E6496–E6505.

Fiesel, F.C. et al., 2010. Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6. *The EMBO Journal*, 29(1), pp.209–221.

Fiesel, F.C. et al., 2012. TDP-43 regulates global translational yield by splicing of exon junction complex component SKAR. *Nucleic acids research*, 40(6), pp.2668–82.

Fratta, P. et al., 2018. Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis. *The EMBO Journal*, p.e98684.

Furukawa, M. et al., 2003. Targeting of protein ubiquitination by BTB–Cullin 3–Roc1 ubiquitin ligases. *Nature Cell Biology*, 5(11), pp.1001–1007.

Furukawa, Y. et al., 2016. A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Scientific reports*, 6, p.20576.

Furukawa, Y. et al., 2011. A seeding reaction recapitulates intracellular formation of Sarkosyl-insoluble transactivation response element (TAR) DNA-binding protein-43 inclusions. *The Journal of biological chemistry*, 286(21), pp.18664–72.

Georgel, P. et al., 2008. ENU mutagenesis in mice. *Methods in molecular biology (Clifton, N.J.)*, 415, pp.1–16.

Geser, F., Lee, V.M.-Y. & Trojanowski, J.Q., 2010. Amyotrophic lateral sclerosis and frontotemporal lobar degeneration: a spectrum of TDP-43 proteinopathies. *Neuropathology : official journal of the Japanese Society of Neuropathology*, 30(2), pp.103–12.

Giordana, M.T. et al., 2010. TDP-43 Redistribution is an Early Event in Sporadic Amyotrophic Lateral Sclerosis. *Brain Pathology*, 20(2), pp.351–360.

De Giorgio, F. et al., 2018. *Transgenic, and physiological targeted and ENU mouse models, give insight into different aspects of amyotrophic lateral sclerosis*,

Gitcho, M.A. et al., 2009. TARDBP 3′-UTR variant in autopsy-confirmed frontotemporal lobar degeneration with TDP-43 proteinopathy. *Acta Neuropathologica*, 118(5), pp.633–645.

Godena, V.K. et al., 2011. TDP-43 Regulates Drosophila Neuromuscular Junctions Growth by Modulating Futsch/MAP1B Levels and Synaptic Microtubules Organization J. Treisman, ed. *PLoS ONE*, 6(3), p.e17808.

Goldstein, L.D. et al., 2016. Prediction and Quantification of Splice Events from RNA-Seq Data. *PloS one*, 11(5), p.e0156132.

Gordon, P.H. et al., 2010. A prospective cohort study of neuropsychological test performance in ALS. *Amyotrophic Lateral Sclerosis*, 11(3), pp.312–320.

Guenther, E.L. et al., 2018. Atomic-level evidence for packing and positional amyloid polymorphism by segment from TDP-43 RRM2. *Nature Structural & Molecular*

Biology, p.1.

Guo, W. et al., 2011. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nature Structural & Molecular Biology*, 18(7), pp.822–830.

Gurney, M.E. et al., 1994. Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science (New York, N.Y.)*, 264(5166), pp.1772–5.

Haberman, N. et al., 2017. Insights into the design and interpretation of iCLIP experiments. *Genome biology*, 18(1), p.7.

Hall, C.E. et al., 2017. Progressive Motor Neuron Pathology and the Role of Astrocytes in a Human Stem Cell Model of VCP-Related ALS. *Cell Reports*, 19(9), pp.1739–1749.

Hanson, K.A. et al., 2010. Ubiquilin Modifies TDP-43 Toxicity in a Drosophila Model of Amyotrophic Lateral Sclerosis (ALS). *The Journal of Biological Chemistry*, 285(15), p.11068.

Hardy, J. & Rogaeva, E., 2014. Motor neuron disease and frontotemporal dementia: sometimes related, sometimes not. *Experimental neurology*, 262 Pt B, pp.75–83.

Hazelett, D.J. et al., 2012. 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: Genes|Genomes|Genetics*, 2(7), pp.789–802.

He, J. et al., 2015. Amyotrophic Lateral Sclerosis Genetic Studies. *The Neuroscientist*, 21(6), pp.599–615.

Hewamadduma, C.A.A. et al., 2013. Tardbpl splicing rescues motor neuron and axonal development in a mutant tardbp zebrafish. *Human Molecular Genetics*, 22(12), pp.2376–2386.

Higashi, S. et al., 2007. Concurrence of TDP-43, tau and α-synuclein pathology in brains of Alzheimer's disease and dementia with Lewy bodies. *Brain Research*, 1184, pp.284–294.

Highley, J.R. et al., 2014. Loss of nuclear TDP-43 in amyotrophic lateral sclerosis (ALS) causes altered expression of splicing machinery and widespread dysregulation of RNA splicing in motor neurones. *Neuropathology and Applied Neurobiology*, 40(6), pp.670–685.

Honda, H. et al., 2015. Loss of hnRNPA1 in ALS spinal cord motor neurons with TDP-43-positive inclusions. *Neuropathology*, 35(1), pp.37–43.

Huang, C., Xia, P.Y. & Zhou, H., 2010. Sustained expression of TDP-43 and FUS in motor neurons in rodent's lifetime. *International journal of biological sciences*, 6(4), pp.396–406.

Huppertz, I. et al., 2014. iCLIP: protein-RNA interactions at nucleotide resolution. *Methods (San Diego, Calif.)*, 65(3), pp.274–87.

Igaz, L.M. et al., 2011. Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. *The Journal of clinical investigation*,

121(2), pp.726–38.

Igaz, L.M. et al., 2009. Expression of TDP-43 C-terminal Fragments in Vitro Recapitulates Pathological Features of TDP-43 Proteinopathies. *The Journal of biological chemistry*, 284(13), pp.8516–24.

Iguchi, Y. et al., 2016. Exosome secretion is a key pathway for clearance of pathological TDP-43. *Brain*, 139(12), pp.3187–3201.

Iguchi, Y. et al., 2013. Loss of TDP-43 causes age-dependent progressive motor neuron degeneration. *Brain*, 136(5), pp.1371–1382.

Iguchi, Y. et al., 2009. TDP-43 Depletion Induces Neuronal Cell Damage through Dysregulation of Rho Family GTPases. *Journal of Biological Chemistry*, 284(33), pp.22059–22066.

Ihara, R. et al., 2013. RNA binding mediates neurotoxicity in the transgenic Drosophila model of TDP-43 proteinopathy. *Human Molecular Genetics*, 22(22), pp.4474–4484.

Janssens, J. et al., 2013. Overexpression of ALS-associated p.M337V human TDP-43 in mice worsens disease features compared to wild-type human TDP-43 mice. *Molecular neurobiology*, 48(1), pp.22–35.

Jean-Philippe, J., Paz, S. & Caputi, M., 2013. hnRNP A1: the Swiss army knife of gene expression. *International journal of molecular sciences*, 14(9), pp.18999–9024.

Jeong, Y.H. et al., 2017. Tdp-43 cryptic exons are highly variable between cell types. *Molecular Neurodegeneration*, 12(1), p.13.

Jiang, L.-L. et al., 2017. The N-terminal dimerization is required for TDP-43 splicing activity. *Scientific Reports*, 7(1), p.6196.

Jiang, L.-L. et al., 2016. Two mutations G335D and Q343R within the amyloidogenic core region of TDP-43 influence its aggregation and inclusion formation. *Scientific reports*, 6, p.23928.

Jiang, Y.-M. et al., 2005. Gene expression profile of spinal motor neurons in sporadic amyotrophic lateral sclerosis. *Annals of Neurology*, 57(2), pp.236–251.

Johnson, B.S. et al., 2008. A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(17), pp.6439–44.

Johnson, B.S. et al., 2009. TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *Journal of Biological Chemistry*, 284(30), pp.20329–20339.

Jones, S. et al., 2001. Protein-RNA interactions: a structural analysis. *Nucleic acids research*, 29(4), pp.943–54.

Josephs, K.A. et al., 2014. TDP-43 is a key player in the clinical features associated with Alzheimer's disease. *Acta Neuropathologica*, 127(6), pp.811–824.

Joyce, P.I. et al., 2015. A novel SOD1-ALS mutation separates central and peripheral effects of mutant SOD1 toxicity. *Human molecular genetics*, 24(7), pp.1883–97.

Joyce, P.I. et al., 2011. SOD1 and TDP-43 animal models of amyotrophic lateral sclerosis: recent advances in understanding disease toward the development of clinical treatments. *Mammalian Genome*, 22(7–8), pp.420–448.

Kabashi, E. et al., 2009. Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Human Molecular Genetics*, 19(4), pp.671–683.

Kabashi, E., Valdmanis, P.N., et al., 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature Genetics*, 40(5), pp.572–574.

Kabashi, E., Valdmanis, P.N., et al., 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature Genetics*, 40(5), pp.572–574.

Kapeli, K. et al., 2016. Distinct and shared functions of ALS-associated proteins TDP-43, FUS and TAF15 revealed by multisystem analyses. *Nature Communications*, 7, p.12143.

Kapeli, K., Martinez, F.J. & Yeo, G.W., 2017. Genetic mutations in RNA-binding proteins and their roles in ALS. *Human genetics*, 136(9), pp.1193–1214.

Kapustin, Y. et al., 2011. Cryptic splice sites and split genes. *Nucleic acids research*, 39(14), pp.5837–44.

Kasai, T. et al., 2009. Increased TDP-43 protein in cerebrospinal fluid of patients with amyotrophic lateral sclerosis. *Acta Neuropathologica*, 117(1), pp.55–62.

Kato, M. et al., 2012. Cell-free Formation of RNA Granules: Low Complexity Sequence Domains Form Dynamic Fibers within Hydrogels. *Cell*, 149(4), pp.753–767.

Kawahara, Y. & Mieda-Sato, A., 2012. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Pnas*, 109(9), pp.3347–3352.

Kim, H.J. et al., 2013. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature*, 495(7442), pp.467–473.

Komatsu, M. et al., 2007. Homeostatic Levels of p62 Control Cytoplasmic Inclusion Body Formation in Autophagy-Deficient Mice. *Cell*, 131(6), pp.1149–1163.

König, J. et al., 2010. iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. *Nature Structural & Molecular Biology*, 17(7), pp.909–915.

Kornblihtt, A., 2007. Coupling transcription and alternative splicing. *Adv Exp Med Biol.*, 623, pp.175–89.

Koyama, A. et al., 2016. Increased cytoplasmic *TARDBP* mRNA in affected spinal motor neurons in ALS caused by abnormal autoregulation of TDP-43. *Nucleic Acids Research*, 44(12), pp.5820–5836.

Kraemer, B.C. et al., 2010. Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta neuropathologica*, 119(4), pp.409–19.

Krueger, F., 2012. Trim Galore!

Kruman, I.I. et al., 1999. ALS-Linked Cu/Zn–SOD Mutation Increases Vulnerability of Motor Neurons to Excitotoxicity by a Mechanism Involving Increased Oxidative Stress and Perturbed Calcium Homeostasis. *Experimental Neurology*, 160(1), pp.28–39.

Kuo, P.-H. et al., 2009. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic acids research*, 37(6), pp.1799–808.

Lacomblez, L. et al., 1996. Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. *Lancet (London, England)*, 347(9013), pp.1425–31.

Lagier-Tourenne, C. et al., 2012. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nature Neuroscience*, 15(11), pp.1488–1497.

Lagier-Tourenne, C., Polymenidou, M. & Cleveland, D.W., 2010. TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Human molecular genetics*, 19(R1), pp.R46-64.

LeCouter, J.E. et al., 1998. Strain-dependent embryonic lethality in mice lacking the retinoblastoma-related p130 gene. *Development (Cambridge, England)*, 125(23), pp.4669–79.

Li, Y. et al., 2010. A Drosophila model for TDP-43 proteinopathy. *Proceedings of the National Academy of Sciences*, 107(7), pp.3169–3174.

Lim, L. et al., 2016. ALS-Causing Mutations Significantly Perturb the Self-Assembly and Interaction with Nucleic Acid of the Intrinsically Disordered Prion-Like Domain of TDP-43 G. A. Petsko, ed. *PLOS Biology*, 14(1), p.e1002338.

Ling, J.P. et al., 2016. PTBP1 and PTBP2 Repress Nonconserved Cryptic Exons. *Cell reports*, 17(1), pp.104–13.

Ling, J.P. et al., 2015. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science*, 349(6248), pp.650–655.

Ling, S.-C. et al., 2010. ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30), pp.13318–23.

Ling, S.-C., Polymenidou, M. & Cleveland, D.W.W., 2013. Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis. *Neuron*, 79(3), pp.416–438.

Liu-Yesucevitz, L. et al., 2011. Local RNA Translation at the Synapse and in Disease. *Journal of Neuroscience*, 31(45), pp.16086–16093.

Liu-Yesucevitz, L. et al., 2010. Tar DNA Binding Protein-43 (TDP-43) Associates with Stress Granules: Analysis of Cultured Cells and Pathological Brain Tissue A. I. Bush, ed. *PLoS ONE*, 5(10), p.e13250.

Liu, Q. et al., 2016. Whole-exome sequencing identifies a missense mutation in *hnRNPA1* in a family with flail arm ALS. *Neurology*, 87(17), pp.1763–1769.

Liu, T.-Y. et al., 2017. Muscle developmental defects in heterogeneous nuclear Ribonucleoprotein A1 knockout mice. *Open biology*, 7(1).

Liu, W.J. et al., 2016. p62 links the autophagy pathway and the ubiqutin-proteasome system upon ubiquitinated protein degradation. *Cellular & molecular biology letters*, 21, p.29.

Liu, Y.-C., Chiang, P.-M. & Tsai, K.-J., 2013. Disease Animal Models of TDP-43 Proteinopathy and Their Pre-Clinical Applications. *International Journal of Molecular Sciences*, 14(10), pp.20079–20111.

Logroscino, G. et al., 2010. Incidence of amyotrophic lateral sclerosis in Europe. *Journal of neurology, neurosurgery, and psychiatry*, 81(4), pp.385–90.

Love, M.I., Huber, W. & Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), p.550.

Lu, Y., Ferris, J. & Gao, F.-B., 2009. Frontotemporal dementia and amyotrophic lateral sclerosis-associated disease protein TDP-43 promotes dendritic branching. *Molecular Brain*, 2, p.30.

Luisier, R. et al., 2018. Intron retention and nuclear loss of SFPQ are molecular hallmarks of ALS. *Nature Communications*, 9(1), p.2010.

Lukavsky, P.J. et al., 2013. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nature Structural & Molecular Biology*, 20(12), pp.1443–1449.

Lutz, C., 2018. Mouse models of ALS: Past, present and future. *Brain Research*, 1693, pp.1–10.

Mackenzie, I.R., Rademakers, R. & Neumann, M., 2010. TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *The Lancet Neurology*, 9(10), pp.995–1007.

Malaspina, A., Kaushik, N. & de Belleroche, J., 2001. Differential expression of 14 genes in amyotrophic lateral sclerosis spinal cord detected using gridded cDNA arrays. *Journal of neurochemistry*, 77(1), pp.132–45.

Manjaly, Z.R. et al., 2010. The sex ratio in amyotrophic lateral sclerosis: A population based study. *Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*, 11(5), pp.439–442.

Margulies, M. et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), pp.376–380.

Martin, S., Al Khleifat, A. & Al-Chalabi, A., 2017. What causes amyotrophic lateral sclerosis? *F1000Research*, 6, p.371.

Maurel, C. et al., 2018. Mutation in the RRM2 domain of TDP-43 in Amyotrophic Lateral Sclerosis with rapid progression associated with ubiquitin positive aggregates in cultured motor neurons. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 19(1–2), pp.149–151.

Mayeda, A. et al., 1994. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *The EMBO journal*, 13(22), pp.5483–95.

McClory, S.P., Lynch, K.W. & Ling, J.P., 2018. HnRNP L represses cryptic exons. *RNA*, p.rna.065508.117.

McDonald, K.K. et al., 2011. TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. *Human Molecular Genetics*, 20(7), pp.1400–1410.

McGoldrick, P. et al., 2013. Rodent models of amyotrophic lateral sclerosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1832(9), pp.1421–1436.

McKee, A.E. et al., 2005. A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Developmental Biology*, 5(1), p.14.

Melé, M. et al., 2015. Human genomics. The human transcriptome across tissues and individuals. *Science (New York, N.Y.)*, 348(6235), pp.660–5.

Mercado, P.A. et al., 2005. Depletion of TDP 43 overrides the need for exonic and intronic splicing enhancers in the human apoA-II gene. *Nucleic Acids Research*, 33(18), pp.6000–6010.

Meyer, K. et al., 2014. Direct conversion of patient fibroblasts demonstrates non-cell autonomous toxicity of astrocytes to motor neurons in familial and sporadic ALS. *Proceedings of the National Academy of Sciences*, 111(2), pp.829–832.

Miguel, L. et al., 2011. Both cytoplasmic and nuclear accumulations of the protein are neurotoxic in Drosophila models of TDP-43 proteinopathies. *Neurobiology of Disease*, 41(2), pp.398–406.

Miller, R. et al., 2002. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). In R. Miller, ed. *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd, p. CD001447.

Mishra, M. et al., 2007. Gene expression analysis of frontotemporal lobar degeneration of the motor neuron disease type with ubiquitinated inclusions. *Acta neuropathologica*, 114(1), pp.81–94.

Mitchell, J.C. et al., 2013. Overexpression of human wild-type FUS causes progressive motor neuron degeneration in an age- and dose-dependent fashion. *Acta Neuropathologica*, 125(2), pp.273–288.

Mitchell, J.C. et al., 2015. Wild type human TDP-43 potentiates ALS-linked mutant TDP-43 driven progressive motor and cortical neuron degeneration with pathological features of ALS. *Acta Neuropathologica Communications*, 3(1), p.36.

Mitrea, D.M. & Kriwacki, R.W., 2016. Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling*, 14(1), p.1.

Mizuno, Y. et al., 2006. Immunoreactivities of p62, an ubiqutin-binding protein, in the spinal anterior horn cells of patients with amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, 249(1), pp.13–18.

Mompeán, M. et al., 2016. An Amyloid-Like Pathological Conformation of TDP-43 Is Stabilized by Hypercooperative Hydrogen Bonds. *Frontiers in Molecular Neuroscience*, 9, p.125.

Morgan, B.R., Zitzewitz, J.A. & Massi, F., 2017. Structural Rearrangement upon Fragmentation of the Stability Core of the ALS-Linked Protein TDP-43. *Biophysical Journal*, 113(3), pp.540–549.

Moujalled, D. et al., 2017. TDP-43 mutations causing amyotrophic lateral sclerosis are associated with altered expression of RNA-binding protein hnRNP K and affect the Nrf2 antioxidant pathway. *Human Molecular Genetics*, 26(9), pp.1732–1746.

Neumann, M. et al., 2009. Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. *Acta neuropathologica*, 117(2), pp.137–49.

Neumann, M. et al., 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (New York, N.Y.)*, 314(5796), pp.130–3.

Niksic, M. et al., 1999. Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Human molecular genetics*, 8(13), pp.2339–49.

Nott, T.J. et al., 2015. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Molecular Cell*, 57(5), pp.936–947.

Nussbacher, J.K. et al., 2015. RNA-binding proteins in neurodegeneration: Seq and you shall receive. *Trends in Neurosciences*, 38(4), pp.226–236.

Orban, P. et al., 2007. Chapter 15 Juvenile amyotrophic lateral sclerosis. In *Handbook of clinical neurology*. pp. 301–312.

Ou, S. et al., 1995. Cloning and Characterization of a Novel Cellular Protein, TDP-43, That Binds to Human Immunodeficiency Virus Type 1 TAR DNA Sequence Motifs. *J. Virol.*, 69(6), pp.3584–3596.

Pagani, F., Buratti, E., et al., 2003. Missense, Nonsense, and Neutral Mutations Define Juxtaposed Regulatory Elements of Splicing in Cystic Fibrosis Transmembrane Regulator Exon 9. *Journal of Biological Chemistry*, 278(29), pp.26580–26588.

Pagani, F., Stuani, C., et al., 2003. Promoter architecture modulates CFTR exon 9 skipping. *The Journal of biological chemistry*, 278(3), pp.1511–7.

Pankiv, S. et al., 2007. p62/SQSTM1 Binds Directly to Atg8/LC3 to Facilitate Degradation of Ubiquitinated Protein Aggregates by Autophagy. *Journal of Biological Chemistry*, 282(33), pp.24131–24145.

Pedersen, W.A. et al., 1998. Protein modification by the lipid peroxidation product 4-hydroxynonenal in the spinal cords of amyotrophic lateral sclerosis patients. *Annals of Neurology*, 44(5), pp.819–824.

Pesiridis, G.S., Lee, V.M.-Y. & Trojanowski, J.Q., 2009. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Human molecular genetics*, 18(R2), pp.R156-62.

Polymenidou, M. et al., 2011. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nature neuroscience*, 14(4), pp.459–68.

Prudencio, M. et al., 2012. Misregulation of human sortilin splicing leads to the generation of a nonfunctional progranulin receptor. *Proceedings of the National Academy of Sciences*, 109(52), pp.21510–21515.

Purice, M.D. & Taylor, J.P., 2018. Linking hnRNP Function to ALS and FTD Pathology.

Frontiers in Neuroscience, 12, p.326.

Rabdano, S.O. et al., 2017. Onset of disorder and protein aggregation due to oxidation-induced intermolecular disulfide bonds: case study of RRM2 domain from TDP-43. *Scientific Reports*, 7(1), p.11161.

Raghava, G. & Barton, G.J., 2006. Quantification of the variation in percentage identity for protein sequence alignments. *BMC Bioinformatics*, 7(1), p.415.

Ratti, A. & Buratti, E., 2016. Physiological Functions and Pathobiology of TDP-43 and FUS/TLS proteins. *Journal of neurochemistry*.

Rau, A., Marot, G. & Jaffrézic, F., 2014. Differential meta-analysis of RNA-seq data from multiple studies. *BMC Bioinformatics*, 15(1), p.91.

Reimand, J. et al., 2016. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research*, 44(W1), pp.W83–W89.

Renton, A.E., Chiò, A. & Traynor, B.J., 2013. State of play in amyotrophic lateral sclerosis genetics. *Nature Neuroscience*, 17(1), pp.17–23.

Ricketts, T., 2012. *Assessment of an allelic series of mouse TDP43 mutations*. University College London.

Ritson, G.P. et al., 2010. TDP-43 mediates degeneration in a novel Drosophila model of disease caused by mutations in VCP/p97. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(22), pp.7729–39.

Rogelj, B. et al., 2012. Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Scientific reports*, 2, p.603.

Rot, G. et al., 2017. High-Resolution RNA Maps Suggest Common Principles of Splicing and Polyadenylation Regulation by TDP-43. *Cell Reports*, 19(5), pp.1056–1067.

Rothstein, J.D., Martin, L.J. & Kuncl, R.W., 1992. Decreased Glutamate Transport by the Brain and Spinal Cord in Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, 326(22), pp.1464–1468.

Rouaux, C., Gonzalez De Aguilar, J.-L. & Dupuis, L., 2018. Unmasking the skiptic task of TDP-43. *The EMBO journal*, p.e99645.

Rowland, L.P. & Shneider, N.A., 2001. Amyotrophic Lateral Sclerosis. *New England Journal of Medicine*, 344(22), pp.1688–1700.

Scekic-Zahirovic, J. et al., 2017. Motor neuron intrinsic and extrinsic mechanisms contribute to the pathogenesis of FUS-associated amyotrophic lateral sclerosis. *Acta neuropathologica*, 133(6), pp.887–906.

Schmid, B. et al., 2013. Loss of ALS-associated TDP-43 in zebrafish causes muscle degeneration, vascular dysfunction, and reduced motor neuron axon outgrowth. *Proceedings of the National Academy of Sciences of the United States of America*, 110(13), pp.4986–91.

Seelen, M. et al., 2014. No mutations in hnRNPA1 and hnRNPA2B1 in Dutch patients with amyotrophic lateral sclerosis, frontotemporal dementia, and inclusion body myopathy. *Neurobiology of Aging*, 35(8), p.1956.e9-1956.e11.

Sephton, C.F. et al., 2011. Identification of Neuronal RNA Targets of TDP-43-containing Ribonucleoprotein Complexes. *Journal of Biological Chemistry*, 286(2), pp.1204–1215.

Sephton, C.F. et al., 2010. TDP-43 is a developmentally regulated protein essential for early embryonic development. *The Journal of biological chemistry*, 285(9), pp.6826–34.

Shan, X. et al., 2010. Altered distributions of Gemini of coiled bodies and mitochondria in motor neurons of TDP-43 transgenic mice. *Proceedings of the National Academy of Sciences*, 107(37), pp.16325–16330.

Shiga, A. et al., 2012. Alteration of POLDIP3 Splicing Associated with Loss of Function of TDP-43 in Tissues Affected with ALS Y. Nagai, ed. *PLoS ONE*, 7(8), p.e43120.

Sreedharan, J. et al., 2008. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science (New York, N.Y.)*, 319(5870), pp.1668–72.

Stallings, N.R. et al., 2010. Progressive motor weakness in transgenic mice expressing human TDP-43. *Neurobiology of Disease*, 40(2), pp.404–414.

Stribl, C. et al., 2014. Mitochondrial Dysfunction and Decrease in Body Weight of a Transgenic Knock-in Mouse Model for TDP-43. *Journal of Biological Chemistry*, 289(15), pp.10769–10784.

Strong, M.J. et al., 2007. TDP43 is a human low molecular weight neurofilament (hNFL) mRNA-binding protein. *Molecular and Cellular Neuroscience*, 35(2), pp.320–327.

Sugai, A. et al., 2018. Robustness and Vulnerability of the Autoregulatory System That Maintains Nuclear TDP-43 Levels: A Trade-off Hypothesis for ALS Pathology Based on in Silico Data. *Frontiers in Neuroscience*, 12, p.28.

Sun, M. et al., 2017. Cryptic exon incorporation occurs in Alzheimer's brain lacking TDP-43 inclusion but exhibiting nuclear clearance of TDP-43. *Acta Neuropathologica*, pp.1–9.

Suzuki, H. & Matsuoka, M., 2017. hnRNPA1 autoregulates its own mRNA expression to remain non-cytotoxic. *Molecular and Cellular Biochemistry*, 427(1–2), pp.123–131.

Swarup, V., Phaneuf, D., Dupré, N., et al., 2011. Deregulation of TDP-43 in amyotrophic lateral sclerosis triggers nuclear factor κB–mediated pathogenic pathways. *The Journal of Experimental Medicine*, 208(12), pp.2429–2447.

Swarup, V., Phaneuf, D., Bareil, C., et al., 2011. Pathological hallmarks of amyotrophic lateral sclerosis/frontotemporal lobar degeneration in transgenic mice produced with TDP-43 genomic fragments. *Brain*, 134(9), pp.2610–2626.

Takagaki, Y. & Manley, J.L., 1997. RNA recognition by the human polyadenylation factor CstF. *Molecular and Cellular Biology*, 17(7), p.3907.

Tan, Q. et al., 2016. Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. *Human Molecular Genetics*, 25(23), p.ddw337.

Tanji, K. et al., 2012. p62/sequestosome 1 binds to TDP-43 in brains with frontotemporal lobar degeneration with TDP-43 inclusions. *Journal of Neuroscience Research*, 90(10), pp.2034–2042.

Tollervey, J.R. et al., 2011. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nature neuroscience*, 14(4), pp.452–8.

Tsai, K.-J. et al., 2010. Elevated expression of TDP-43 in the forebrain of mice is sufficient to cause neurological and pathological phenotypes mimicking FTLD-U. *The Journal of experimental medicine*, 207(8), pp.1661–73.

Tsuiji, H. et al., 2013. Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. *EMBO molecular medicine*, 5(2), pp.221–34.

Turner, M.R. et al., 2017. Genetic screening in sporadic ALS and FTD. *Journal of neurology, neurosurgery, and psychiatry*, 88(12), pp.1042–1044.

Urushitani, M. et al., 1998. Mechanism of selective motor neuronal death after exposure of spinal cord to glutamate: Involvement of glutamate-induced nitric oxide in motor neuron toxicity and nonmotor neuron protection. *Annals of Neurology*, 44(5), pp.796–807.

Uryu, K. et al., 2008. Concomitant TAR-DNA-binding protein 43 pathology is present in Alzheimer disease and corticobasal degeneration but not in other tauopathies. *Journal of neuropathology and experimental neurology*, 67(6), pp.555–64.

Vaccaro, A. et al., 2012. TDP-1/TDP-43 Regulates Stress Signaling and Age-Dependent Proteotoxicity in Caenorhabditis elegans B. Lu, ed. *PLoS Genetics*, 8(7), p.e1002806.

Vatsavayi, A.V. V et al., 2014. TAR DNA-binding protein 43 pathology in Alzheimer's disease with psychosis. *International psychogeriatrics*, 26(6), pp.987–94.

Verstraete, E. et al., 2012. TDP-43 plasma levels are higher in amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis*, 13(5), pp.446–451.

Voigt, A. et al., 2010. TDP-43-Mediated Neuron Loss In Vivo Requires RNA-Binding Activity M. B. Feany, ed. *PLoS ONE*, 5(8), p.e12247.

Walker, A.K. et al., 2015. An insoluble frontotemporal lobar degeneration-associated TDP-43 C-terminal fragment causes neurodegeneration and hippocampus pathology in transgenic mice. *Human Molecular Genetics*, 24(25), p.7241.

Wang, I.F. et al., 2008. TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor. *Journal of Neurochemistry*, 105(3), pp.797–806.

Wang, X. et al., 2006. Differential expression of genes in amyotrophic lateral sclerosis revealed by profiling the post mortem cortex. *Amyotrophic Lateral Sclerosis*, 7(4), pp.201–216.

Wang, Y.-T. et al., 2013. The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. *The Journal of biological chemistry*, 288(13), pp.9049–57.

Wegorzewska, I. et al., 2009. TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, 106(44), pp.18809–14.

Wei, Y. et al., 2017. *ALS-causing cleavages of TDP-43 abolish its RRM2 structure and unlock CTD for enhanced aggregation and toxicity*,

Wei, Y. et al., 2016. *Inter-domain interactions of TDP-43 as decoded by NMR*,

White, K. et al., 2018. Effect of Postmortem Interval and Years in Storage on RNA Quality of Tissue at a Repository of the NIH NeuroBioBank. *Biopreservation and biobanking*, 16(2), pp.148–157.

White, M.A. et al., 2018. TDP-43 gains function due to perturbed autoregulation in a Tardbp knock-in mouse model of ALS-FTD. *Nature Neuroscience 2018*, p.1.

Wils, H. et al., 2010. TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), pp.3858–63.

Winton, M.J. et al., 2008. Disturbance of Nuclear and Cytoplasmic TAR DNA-binding Protein (TDP-43) Induces Disease-like Redistribution, Sequestration, and Aggregate Formation. *Journal of Biological Chemistry*, 283(19), pp.13302–13309.

Wu, L.-S. et al., 2009. TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. *genesis*, 48(1), p.n/a-n/a.

Wu, L.-S., Cheng, W.-C. & Shen, C.-K.J., 2012. Targeted Depletion of TDP-43 Expression in the Spinal Cord Motor Neurons Leads to the Development of Amyotrophic Lateral Sclerosis-like Phenotypes in Mice. *Journal of Biological Chemistry*, 287(33), pp.27335–27344.

Xiao, S. et al., 2011. RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Molecular and Cellular Neuroscience*, 47(3), pp.167–180.

Xu, Y.-F. et al., 2011. Expression of mutant TDP-43 induces neuronal dysfunction in transgenic mice. *Molecular Neurodegeneration*, 6, p.73.

Xu, Y.-F. et al., 2010. Wild-type human TDP-43 expression causes TDP-43 phosphorylation, mitochondrial aggregation, motor deficits, and early mortality in transgenic mice. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(32), pp.10851–9.

Xu, Z.-S., 2012. Does a loss of TDP-43 function cause neurodegeneration? *Molecular Neurodegeneration*, 7(1), p.27.

Yang, C. et al., 2014. Partial loss of TDP-43 function causes phenotypes of amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences*, 111(12), pp.E1121–E1129.

Zhang, Y.-J. et al., 2009. Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), pp.7607–12.

Zheng, S. & Black, D.L., 2013. Alternative pre-mRNA splicing in neurons: growing up and extending its reach. *Trends in Genetics*, 29(8), pp.442–448.

Zhou, H. et al., 2010. Transgenic Rat Model of Neurodegeneration Caused by Mutation in the TDP Gene G. A. Cox, ed. *PLoS Genetics*, 6(3), p.e1000887.