# Characterizing RNA Targets of TDP-43

## Thesis

# Characterizing RNA Targets of TDP-43

Maureen V. Akinyi

Thesis submitted in fulfilment of the requirements of the Open University, (U.K) for the Degree of Doctor of Philosophy

Life Sciences


International Centre for Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy

Director of studies: Francisco E. Baralle, M.D. Ph.D

Internal Supervisor: Marco Baralle, Ph.D

External Supervisor: Colin Sharpe, Ph.D

November 2014

DATE OF SUBMISSION : 11 NOVEMBER 2014
DATE OF AWARD : 23 JANUARY 2015

*For Priscah and Richie*

## ACKNOWLEDGEMENTS

With the utmost sincerity I would like to thank:

-Professor F.E. Baralle for providing me with the opportunity to do my PhD under his mentorship and guidance; for his constant support in all aspects during my PhD and above all, for training me to think critically.

-Dr Marco Baralle and Dr Laura de Conti for their continued patience, understanding, support, enthusiasm and willingness to teach throughout the years.

-Dr Colin Sharpe for his invaluable contributions and enthusiastic discussions on the project.

-Dr E. Buratti for conceptualizing numerous aspects of the project and Dr Romano for his assistance with Bioinformatics.

-In general both the molecular pathology and RNA biology lab members for providing a collaborative and fun environment to work in.

-My lovely friends both in Trieste and Cape Town, for taking this journey with me and for their constant support and encouragement.

-Last but not least; my guardians/parents and the rest of my family for their never ending support and encouragement throughout all the years of my studies.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ABBREVIATIONS

| | |
|---|---|
| °C | degree Celcius |
| µg | micrograms |
| µl | microlitre |
| µm | micrometre |
| AD | Alzheimer's disease |
| ALS | Amyotrophic Lateral Sclerosis |
| BIM/BCL2-L11 | B-cell lymphoma 2 like protein 11 |
| bp | base pair |
| BRD8 | Bromodomain containing protein 8 |
| $CaCl_2$ | Calcium Chloride |
| cDNA | complementary DNA |
| DNA | Deoxyribonucleic acid |
| *DYRK1A,* | Dual specificity tyrosine phosphorylation regulated kinase 1A |
| $dH_2O$ | distilled water |
| dNTPs | deoxynucleotide triphosphates |
| DTT | Dithiothreitol |
| EDTA | Ethylenediamine tetra-acetic acid |
| EEF2 | Eukaryotic elongation factor 2 |
| ESE | Exonic splicing enhancer |
| ESS | Exonic splicing silencer |
| EtBr | Ethidium bromide |
| F | Phenyalanine |
| FUS/TLS | Fused in sarcoma/translocated in liposarcoma |
| FTLD | Fronto-temporal lobar degeneration |
| FNIP1 | Folliculin interacting protein 1 |
| Fwd | Forward |
| g | gram |
| GAPDH | Glyceraldehyde-3-phosphate dehydrogenase |
| Gly | Glycine |
| HD | Huntington's disease |
| HCl | Hydrochloric Acid |
| hnRNP | Heterogenous ribonucleoprotein |
| ISE | Intronic splicing enhancer |
| ISS | Intronic splicing silencer |
| kDa | kilo Dalton |
| kb | kilo base |
| L | Leucine |
| L | Litre |
| LAMC1 | Lammin subunint gamma-1 |
| MADD | Map-kinase activating death domain protein |
| MDH1 | Malate dehydrogenase 1 |
| $MgCl_2$ | Magnesium Chloride |
| mL | millilitre |
| mM | millimolar |
| mRNA | messenger RNA |

| | |
|---|---|
| N | Asparagine |
| NASP | Nuclear antigenic sperm protein |
| NAP1L1 | Nucleosome assembly protein-1 like-1 |
| NMD | Nonsense Mediated Decay |
| ng | nanogram |
| Oligo(dT) | Oligo (deoxythymidine |
| PBS | Phosphate Buffered Saline |
| PCR | Polymerase chain reaction |
| PD | Parkinson's disease |
| POLDIP3 | Polymerase delta-interacting protein 3 |
| Pre-mRNA | Precursor messenger RNA |
| qPCR | quantitative real-time PCR |
| Q | Glutamine |
| Rev | reverse |
| RNA | Ribonucleic acid |
| RRM | RNA recognition motif |
| RT-PCR | Reverse-transcriptase PCR |
| sec | seconds |
| SOD1 | Super-oxide dismutase 1 |
| ss | splice site |
| snRNP | small nuclear ribonucleoprotein |
| STAG2 | Stromal antigen 2 protein |
| *STX3,* | Syntaxin 3 |
| Ta | Annealing temperature |
| TBE | Tris Borate EDTA buffer |
| TDP-43 | Trans-active response DNA binding protein 43 |
| TEMED | NNN'N"-tetramethylethylenediamine |
| Tris | tris[hydroxymethyl]aminomethane |
| U | Units |
| U2AF1 | U2-auxillary factor 1 |
| UCSC | University of California Santa Cruz |
| UTR | Untranslated region |
| UV | Ultraviolet |
| V | volts |
| *VAMP3* | Vesicle-associated membrane protein 3 |

# ABSTRACT

TAR-DNA-binding 43kDa protein (TDP-43), is an RNA binding protein that has been linked to the pathology of neurodegenerative diseases such as ALS and FTLD. However, pathological mechanisms involving TDP-43, remain elusive. Current hypotheses have converged on altered RNA processing, due to a loss of function of TDP-43. To gain better insight into the pathogenic mechanisms linked to TDP-43, we hypothesised that a loss of function may lead to alterations in splicing and/or changes in protein expression in transcripts potentially linked to neurodegeneration. Thus, 2-dimensional electrophoresis (2-DE) gels and splice-junction microarrays were utilised to identify targets of TDP-43 whose expression exhibited association with altered TDP-43 levels.

The 2-DE analyses depicted changes in several proteins (spots), which, upon subsequent validation, did not reveal any correlations with differential spot intensities. In contrast, the splice-junction arrays identified TDP-43 dependent changes in 2371 genes from which candidate genes for validation were selected upon fulfilment of criteria for being altered when TDP-43 was depleted and reversible by overexpression of wild type TDP-43, but not with mutant (F4L) TDP-43. 162 genes were selected and further narrowed down based on a two-fold or higher splicing score. RT-PCR validation confirmed 6/19 transcripts, *POLDIP3*, *BCL2L11*, *MADD*, *STAG2*, *BRD8* and *FNIP1*, as undergoing TDP-43 dependent splicing. Protein level alterations, were also observed in *MADD* and *STAG2*, following which, TDP-43 binding sites were mapped. Lastly, using an inducible TDP-43 aggregation model, a loss of function effect for TDP-43 was recapitulated i.e. similar changes in splicing profiles as when TDP-43 was knocked down, for all six of the above genes, thus, supporting a loss of function hypothesis in TDP-43 aggregates.

Our results further contribute to the number of genes known to be affected by TDP-43, with the distinct observation of involvement in apoptotic and mitotic pathways, which may have implications for TDP-43 proteinopathies.

# 1. INTRODUCTION

The global increase in life expectancy, has led to an increase in the prevalence of neurodegenerative disorders, which are placing a huge burden on health care costs, and consequently an important economic and social concern (Banks et al. 2008). Neurodegenerative diseases are characterised by the progressive degeneration of neurons resulting in a wide range of symptoms including motor dysfunction, cognitive failure and dementia. Symptoms observed in patients tend to correlate with the type of neurons affected and their location, although it is not clear why only certain subsets of neurons are affected and not others (Przedborski et al. 2003). In addition, some cases also appear to have overlaps in symptoms or manifestation of the disease, e.g. dementia in patients with Amytrophic lateral Sclerois (ALS), Fronto-temporal dementia (FTD), Parkinson's disease (PD) and Alzheimer's (AD), suggesting a common molecular pathogenesis in these diseases (Przedborski et al. 2003; Lill & Bertram 2011). A common feature of neurodegenerative diseases is the deposition and accumulation of intracellular and or extracellular protein aggregates (Taylor et al. 2002; Ross & Poirier 2004; Skovronsky et al. 2006; Renoux & Todd 2012). These aggregates are often composed of different misfolded proteins, as identified by immunohistochemical and histological staining, and have further facilitated the grouping and identification (diagnostic) of neurodegenerative diseases, based on common molecular pathology (Przedborski et al. 2003; Skovronsky et al. 2006). Furthermore, the type of aggregates can be sub-classified based on the presence or absence of amyloid structures, filaments or skein-like inclusions (Ross & Poirier 2004; Robinson et al. 2013), although in cases such as Amyotrophic Lateral Sclerosis (ALS) heterogeneous properties of aggregates have been reported (Robinson et al. 2013).

Studies aiming to understand the pathology of neurodegenerative diseases signified by the accumulation of these aggregates have postulated that an initial seeding of misfolded

protein occurs, after which the wild type endogenous protein is recruited to the aggregates leading to the increased accumulation of protein (Aguzzi & Rajendran 2009; Polymenidou & Cleveland 2012). It is also thought that these intracellular aggregates can spread amongst adjacent cells (Desplats et al. 2009; Polymenidou & Cleveland 2011; Hansen et al. 2011), which incidentally, could account for the progressive nature of these diseases but not for the varied specificity of affected neurons. Thus, these studies seem to converge on the presence of a common molecular pathway i.e. cascade of events, beginning with the seeding of misfolded protein and subsequent spread in adjacent neurons. Despite the proposed mechanisms of disease progression, the exact causes or triggers of aggregate formation are still not well understood. Given that for each neurodegenerative disease, the major component of misfolded protein is different, or in some cases includes combined accumulations of several proteins, it is thought that the aggregates are an end stage manifestation of the disease, signifying a collapse in cellular maintenance of protein homeostasis (Polymenidou & Cleveland 2011).

In keeping with this hypothesis, several studies examining the pathological mechanisms of neurodegenerative diseases have analysed the identified misfolded proteins for mutations that could be relevant to disease. Advances in genome sequence analytical technologies such as next-generation sequencing, have enabled the wide-scale identification of such polymorphisms (Lill & Bertram 2011). Indeed, several sequence polymorphisms have been identified in several genes that appear to segregate with disease (Bertram & Tanzi 2005; Lill & Bertram 2011) and have been attributed to aberrant protein folding and production. These mutations are thought to play a role in the initial seed of misfolded protein (Polymenidou et al. 2011). However, in most cases of neurodegenerative diseases, familial inheritance i.e. Mendelian segregation within families, is in fact rare, and the majority of cases are sporadic or idiopathic (Bertram & Tanzi 2005; Lagier-Tourenne & Cleveland 2009; Lill & Bertram 2011). While mutations or sequence polymorphisms in relevant

genes could be responsible for encoding aberrantly folded proteins or increase propensity for aggregation, they do not account for cases, where no specific mutations have been identified. Indeed, most neurodegenerative diseases are classified as complex diseases that seem to be linked to both genetic risk and environmental factors (Przedborski et al. 2003; Sheikh et al. 2012). Possible environmental factors range from decreased efficiency in cellular processes attributed to aging, to harmful chemical exposures (Przedborski et al. 2003; Cannon & Greenamyre 2011).

Other avenues of neurodegenerative pathogenesis research have taken a more molecular approach, by attempting to understand the involvement of relevant proteins in various pathways, including specific targets of these proteins and how a loss of function could contribute to neurodegeneration. Such studies have identified pathological mechanisms such as tri-nucleotide repeat expansion disorders, in Huntington's disease (HD) and myotonic dystrophies (DM), that perturb the system through RNA toxicity and which, has also resulted in a patho-clinical sub classification of these neurological disorders (La Spada & Taylor 2010; Renoux & Todd 2012). Repeat expansion disorders subsequently paved the way for the identification of RNA as a toxic species, which expanded the neurodegenerative research field towards examining the role of RNA metabolism and RNA binding proteins in neurodegenerative diseases (Renoux & Todd 2012; Belzil et al. 2012).

Indeed, a major revolution in the field came when two RNA binding proteins TAR-DNA binding protein, 43kDa (TDP-43) (Arai et al. 2006; Neumann, Sampathu, Kwong, Truax, Micsenyi, Chou, Bruce, Schuck, Grossman, Clark, McCluskey, Miller, Masliah, Mackenzie, Feldman, Feiden, H. A. Kretzschmar, et al. 2006) and fused in sarcoma/translocated in sarcoma (FUS/TLS) were identified as the major components of pathological inclusions in patients with ALS and Frontotemporal lobar degeneration (FTLD) (Neumann et al. 2009). In the cases of ALS and FTLD, common pathological

pathways are beginning to emerge, implicating altered RNA metabolism that is enforced by the identification of common RNA binding proteins and perturbations in protein homeostasis (Ling et al. 2013). Based on these observations, it has been suggested that ALS and FTLD are representative of a clinicopathological spectrum of a single disease (Geser et al. 2009; Cohen et al. 2011; Ling et al. 2013).

Currently, TDP-43 is known to constitute the main protein (97%) sequestered within the aggregates of ALS patients with relatively minor contributions (cumulative 3%) from superoxide dismutase 1 (SOD1) and FUS (Ling et al. 2013). This unprecedented percentage of inclusion in ALS provides a unique opportunity for gaining insight into pathological mechanisms involving TDP-43, which can be extrapolated to other neurodegenerative diseases with TDP-43 inclusions (TDP-43 proteinopathies). Thus this study focused on examining TPD-43 targets with the aim of gaining better insight into pathological mechanisms involved in ALS, which ultimately could inform the development of therapeutic strategies.

## 1.1. Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS), also known as Lou Gherig's disease, or motor neuron disease is an adult onset (typically 60 years and above although younger ages of onset have been reported) neurodegenerative disease, characterized by the progressive degeneration and death of motor neurons in the brain and spinal cord, ultimately resulting in paralysis and death (Rothstein 2009; Neumann 2009; Mackenzie et al. 2010; Geser et al. 2011). The incidence of ALS is about 6 in 100 000 individuals affected with an average poor prognosis of about 2 to 5 years (Geser et al. 2011) (http://ghr.nlm.nih.gov/condition/amyotrophic-lateral-sclerosis). Symptoms generally begin with muscle fatigue and weakness, followed by wasting and fasciculation and are generally grouped into bulbar or spinal onset (Forbes et al. 2004; Ravits et al. 2013). The spinal onset of ALS is characterised by asymmetrical focal muscle weakness and atrophy.

Bulbar onset on the other hand, is characterised by difficulties in swallowing (dysphagia) and speech (dysarthia) and respiratory failure due to the atrophy of upper and lower motor neurons and has faster (1-year) progression (Forbes et al. 2004; Ravits et al. 2013).

Incidences of ALS vary among different ethnicities and gender, with higher incidences generally being observed in men than in women (Logroscino et al. 2011). As mentioned previously, prognosis of ALS is poor (average 2-5 years) (Banks et al. 2008; Lagier-Tourenne & Cleveland 2009; Da Cruz & Cleveland 2011), with some patients also exhibiting symptoms of FTLD suggesting a continuum for disease pathology (Baloh, 2011). Currently, no treatment exists that hinders the progression of the disease significantly. This is likely due to the fact that several factors are involved in the disease, as, since its approval in the 1990's, Riluzole by Sanofi-Aventis, a presynaptic glutamate release inhibitor, remains the only prescribed treatment for ALS patients offering a very modest survival benefit of 2 to 3 months (Miller et al. 2012).

### 1.1.1. *Aetiology of ALS*

The majority of ALS cases are sporadic, whereas 10% of cases are attributed to familial inheritance due to mutations in several genes (Da Cruz & Cleveland 2011; van Blitterswijk et al. 2012). The first gene found to be linked to ALS was superoxide dismutase-1 *(SOD1)* in which mutations on exons 2 and 4 were found to segregate in a dominant manner in families with ALS (Rosen et al. 1993). Since then, of the rare (10%) familial cases of ALS (f-ALS) reported, 20% are linked to mutations in *SOD1*, with over 150 different mutations identified with a dominant inheritance (Taylor et al. 2002; Mackenzie et al. 2007; Lagier-Tourenne & Cleveland 2009; van Blitterswijk et al. 2012). Up until recently, efforts to understand ALS pathogenesis had focused on the biological consequences of mutations in *SOD1*, an enzyme which catalyses the intracellular detoxification of superoxide anion $O_2^-$ radical into hydrogen peroxide that is eventually converted into water (Rosen et al. 1993;

Da Cruz & Cleveland 2011). Mutations in other genes have since been identified including fused in sarcoma/translocated in sarcoma (*FUS/TLS)* (Kwiatkowski et al. 2009; Vance et al. 2009), *TDP-43* (Kabashi et al. 2008; Yokoseki et al. 2008; Sreedharan et al. 2008), angiogenin (*ANG)* (Greenway et al. 2006; Fernández-Santiago et al. 2009) and more recently, in the Matrin 3 (*MATR3*) gene (Johnson et al. 2014) and a repeat expansion in chromosome 9 open reading frame 72 (*C9orf72*) (DeJesus-Hernandez et al. 2011; Renton et al. 2011). In addition, a recent study suggested that the development of ALS may be oligogenic i.e. dependent on more than one mutation in one or more unrelated genes (van Blitterswijk et al. 2012), supporting a hypothesis for the involvement of other modifying factors both genetic and non-genetic.

Fundamentally, research efforts into familial forms of ALS have been driven by the hypothesis that the observed symptomatic similarities amongst fALS and sporadic ALS (sALS) implied a common molecular pathway, which could provide insights into the pathogenesis of ALS (Mackenzie et al. 2007). However in most cases, even between the different forms of ALS the presence of identified proteins in pathogenic inclusions differs, for instance the identification of misfolded SOD1 in ALS-*SOD1* (mutations in *SOD1* identified) but not in other sporadic forms of ALS, which mostly contained TDP-43 suggesting a different mechanism of disease in fALS (Mackenzie et al. 2007; Neumann 2009; Mackenzie et al. 2010).

In 2006, the identification of TAR-DNA binding protein (TDP-43) as the major protein in proteinaceous ubiquitin-positive and tau-negative cytoplasmic inclusions in FTLD and sALS patients (Arai et al. 2006; Neumann, et al. 2006), led to a shift in research focus to this multi-functional RNA-binding protein. The presence of TDP-43 in the cytoplasmic inclusions of both FTLD and ALS supported the hypothesis of a common disease pathogenesis and the notion of the presence of a clinicopathological spectrum in a single disease (Arai et al. 2006). Consequently, current emerging hypotheses on ALS pathology

have converged on altered RNA processing linked to RNA binding proteins, focusing on the role of TDP-43 as an RNA binding protein.

## 1.2. TAR-DNA Binding Protein-43 (TDP-43)

The trans-active responsive DNA binding protein of 43kDa (TDP-43) is ubiquitously expressed and encoded by the *TARDBP* gene on chromosome 1p36. TDP-43 was first identified as a transcriptional regulator in the human immune-deficiency virus type 1 (HIV-1) genome where it was reported to bind to the trans-active response (TAR) element and repress transcription of the *HIV-1* transcript (Ou et al. 1995). Recently, however, the role of TDP-43 in repressing viral gene expression has come into question, as one study reported no repression of viral (HIV) expression in either early or late stages of infection (Nehls et al. 2014). Nonetheless, a seminal role for TDP-43 in splicing regulation was reported in 2001 by Buratti et al. , wherein TDP-43 was found to contribute to the inhibition of CFTR (Cystic fibrosis transmembrane conductance regulator) exon 9 recognition, resulting in the skipping of this exon. This was the first described role of TDP-43 as a splicing regulator.

Nonetheless, the aforementioned discovery of TDP-43 in pathogenic inclusions of patients with ALS, FTLD, Huntington's (HD), inclusion body myopathy (IBM) Alzheimer's (AD) cases have emphasised a role for this protein in neurodegeneration (Lagier-Tourenne et al. 2010; Chen-plotkin et al. 2010), thereby redirecting research into understanding the functions and targets of TDP-43 that go awry in pathology.

### 1.2.1. ***Structure and Function***

TDP-43 belongs to a highly conserved group of heterogeneous nuclear ribonucleoproteins (hnRNPs) that are known to be involved in multiple steps of gene expression regulation including transcription, splicing, mRNA stability, DNA replication/repair, protein translation and export or retention of nascent RNA (Krecic & Swanson 1999; Ayala et al.

2008; Buratti & Baralle 2010). Structurally, all hnRNPs contain RNA-binding domains also known as RNA recognition motifs (RRM) that determine specificity to target RNA or DNA as well as auxiliary domains that facilitate protein to protein interactions (Krecic & Swanson 1999). In addition, hnRNPs contain nuclear export and localisation signals that enable them to shuttle between the nucleus and cytoplasm depending on the required function, at specific time points in the cell (Krecic & Swanson 1999; Ayala et al. 2008; Buratti & Baralle 2010). This is thought to facilitate the multifunctional aspects of hnRNPs. Accordingly, TDP-43 is a 414 amino acid protein that has two RRM domains known to facilitate RNA binding, and a glycine-rich C-terminal domain (Figure 1-1) (Buratti & Baralle 2001; Wang et al. 2004; Buratti et al. 2005; Ayala et al. 2005; Ayala, Zago, et al. 2008). These structural domains are similarly highly conserved in *Drosophila melanogaster, Xenopus laevis and Caenorhabditis elegans* homologs (Wang et al. 2004; Ayala et al. 2005). In addition, within the N-terminal region, TDP-43 also has nuclear localisation (NLS) and nuclear export signals (NES) that facilitate shuttling between the nucleus and the cytoplasm (Ayala, Zago, et al. 2008).

TDP-43 has been shown to preferentially bind to UG/TG repeats in RNA and single stranded DNA (Buratti & Baralle 2001), although this is not always the case as has been demonstrated by TDP-43 binding to its own 3'UTR to regulate expression in a negative feedback loop mechanism (Ayala et al. 2011). Sequence specific binding of TDP-43 to UG/TG repeats, is thought to be facilitated by RRM1, proximal to the N-terminal domain (Buratti & Baralle 2001) (Figure 1-1). Recent structural studies have further supported this role for TDP-43 RRM1 (Kuo et al. 2009; Bhardwaj et al. 2013; Kuo et al. 2014) and proposed that RRM2 is also involved in recognition and binding of UG/TG sequences in RNA and single stranded DNA by mechanistic folding (Kuo et al. 2009).

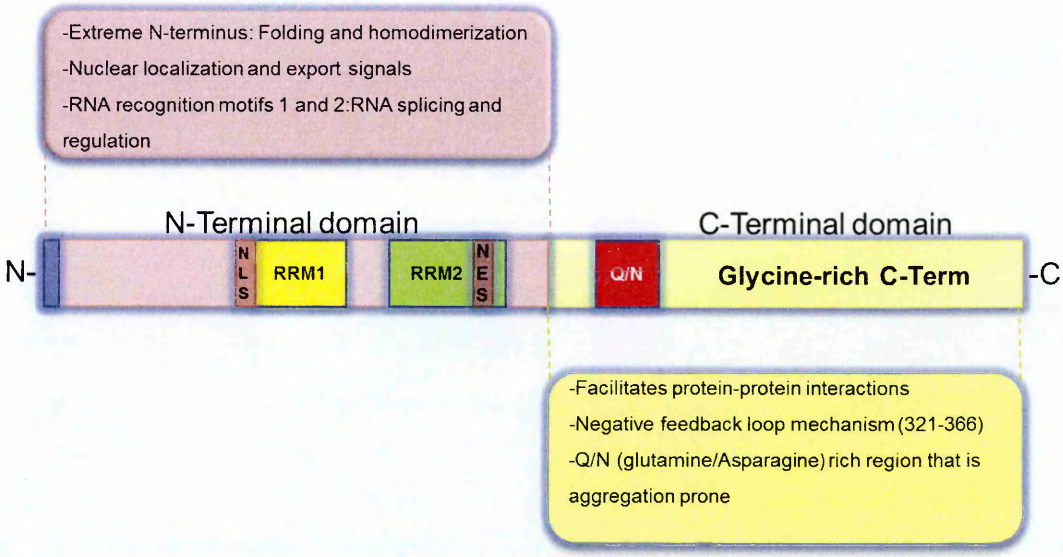



Figure 1-1: **Schematic diagram of the TDP-43 protein highlighting the structural domains and their functions**. Functions of the various domains are described including the N-terminal which contains, nuclear export and localisation signals known to facilitate shuttling in and out of the nucleus and the Glycine-rich C-terminal that facilitates protein-protein interactions.

On the other hand, the glycine-rich region of the C-terminal domain of TDP-43 has been demonstrated to be necessary for interaction with other proteins including other hnRNPs such as hnRNPA2/B1 and hnRNPA1 (Buratti et al. 2005; D'Ambrogio et al. 2009). Indeed, TDP-43 lacking the C-terminal is no longer able to regulate the skipping of exon 9 in the *CFTR* gene (Wang et al. 2004). Although the C-terminal domain of TDP-43 is not entirely required for localisation, TDP-43 lacking the C-terminal had more cytoplasmic presence despite the nuclear localisation signal (Ayala, Zago, et al. 2008). In addition, deleting or mutating the TDP-43 C-terminal significantly reduces the solubility of the protein and results in inclusion body formations, thought to be a result of the lack of interaction between TDP-43 and other protein factors that promote solubility (Ayala, Zago, et al. 2008). The decrease in solubility and presence of inclusion bodies could be extrapolated to patients, wherein, a non-functioning or misfolded protein could increase the aggregation propensity of TDP-43. Indeed, studies that conducted sequence analyses of TDP-43 have found that the C-terminal of TDP-43 is a 'hot spot' region for mutations and TDP-43 has thus been linked to both fALS, sALS and FTLD (Lagier-Tourenne &

Cleveland 2009; Pesiridis et al. 2009). Nonetheless, the role of TDP-43 as an RNA-binding protein (RBP) has becoming increasingly important as several aspects of gene expression mysregulation have been linked to disease pathogenesis.

## 1.3. TDP-43: A global transcript regulator

The involvement of RBPs in neurodegenerative disease has become an increasingly relevant field of research, following the identification and implication of several RBPs, such as TDP-43, FUS/TLS, hnRNP A1/A2, fragile X mental retardation protein (FMRP) amongst others (Hanson & Tibbetts 2012; Vanderweyde et al. 2013). RBPs are known to bind either RNA or DNA or both by forming either homodimers or heterodimers and perform specific functions within the cell. In general RBPs play numerous roles in the nucleus including pre-mRNA splicing, maturation and export, whereas in the cytoplasm they are able to regulate translation, transport and degradation (Vanderweyde et al. 2013). As discussed previously, TDP-43 is a typical hnRNP protein with characteristic features of RBPs that include RRMs and Glycine-rich domains. Currently, based on the role of TDP-43 and FUS/TLS, there is a convergence on the theory of altered RNA metabolism as a causative factor in neurodegenerative disease (Polymenidou et al. 2012) . Recently, another RBP, TATA box-binding protein associated factor 2N (TAF15) was identified and linked to ALS further supporting the functional role of this group of proteins in neurodegeneration (Ugras & Shorter 2012; Polymenidou et al. 2012).

Specifically, TDP-43 has been shown to be involved in several RNA processing pathways including splicing, transcription, localisation, microRNA (miRNA) biogenesis and stabilization (Lagier-Tourenne & Cleveland 2009a; Buratti & Baralle 2010; Polymenidou et al. 2012a). Similarly, FUS/TLS has been linked to transcriptional regulation (Buratti & Baralle 2010; Hanson & Tibbetts 2012; Lagier-Tourenne et al. 2012a), and additionally, a role in pre-mRNA splicing of growth factors has been reported, thereby influencing embryonic developmental pathways (Dichmann & Harland 2012). Both TDP-43 and

FUS/TLS are reported to be involved in the formation of transient stress granules containing RNA in complex with RBPs (Polymenidou et al. 2012; Colombrita et al. 2012), however, for the purposes of this study, only the role of TDP-43 in relation to gene expression regulation will be discussed.

Based on the hypothesis of altered RNA metabolism due to a loss of function of TDP-43, several studies aiming at identifying functions of TDP-43 and its targets have been conducted. Emergent from these studies, is that TDP-43 plays a major role in the regulation of numerous transcripts by acting on the afore-mentioned RNA processing mechanisms that are discussed in more detail below. Currently, the most defined role of TDP-43 is its role in alternative splicing, which provides a strong basis for the hypothesis of an altered RNA metabolism in TDP-43 proteinopathies.

### 1.3.1. *Gene expression regulation: Focus on Alternative Splicing*

Within the cell several modes of gene expression regulation exist, that determine the temporal-spatial expression of specific sets of genes in different cells. Gene expression regulation can occur both during transcription and post-transcriptionally, and involves several mechanisms that range from alternative splicing and poly-adenylation, mRNA stabilization and localisation to non-sense mediated degradation of transcripts (Glisovic et al. 2008; Ward & Cooper 2011).

With the advent of the genomics sequencing era, came the perplexing discovery that the number of coding genes in organisms was not an indicator of cellular complexity. This observation is perhaps more emphasised in the comparison between mammalians and *A. thaliana,* that have similar numbers of coding genes, approximately 25,000 (Blencowe 2006). Thus, it is evident that there are mechanisms in place, which are responsible for the observed differential cellular complexity in organisms notwithstanding the number of coding genes. One of the mechanisms responsible for this diversity and complexity is

alternative splicing, which is described as the mechanism by which components of primary transcripts (pre-mRNA), including 5' and 3' UTR, exons, introns and poly-A sites are spliced and rearranged to include various combinations, that result in different gene isoforms and consequently proteins (Faustino & Cooper 2003; Blencowe 2006; Tazi et al. 2009; Kalsotra & Cooper 2012). An alternative view is that the non-coding sequences within genes, that contain enhancer and repressor elements, might be key drivers of organismal complexity as they respond to a range of signals, including thresholds and combinations of trans-acting factors (Levine & Tjian 2003).

High-throughput sequencing technologies have revealed that approximately 90-95% of the human genome undergoes alternative splicing (Pan et al. 2008; Wang et al. 2008; Kalsotra & Cooper 2012). In addition, several factors have been shown to regulate alternative splicing, including *cis*-acting enhancer and silencer elements, the spliceosome and other *trans*-acting factors that are temporally expressed (Figure 1-2) (Tazi et al. 2009; Pagani & Baralle 2004; Baralle et al. 2009). A specific group of Uridine-rich small nuclear ribonucleproteins (U-snRNPS), collectively known as the spliceosome are responsible for the main splicing catalytic actions that remove introns and join exons together (Wahl et al. 2009). Each of these snRNPs contain small nuclear RNA (snRNA) that mediate binding to the intron through base-pair interactions, thereby ensuring fidelity in the splicing process (Wahl et al. 2009). The spliceosome machinery is further sub-categorised into major and minor spliceosomes.

The major spliceosome, consisting of the U1, U2, U4/U6 and U5, snRNPs recognizes specific consensus sequences in introns, including the dinucleotides GT-AG in the 5' and 3' splice sites (SS) of introns, the branch point sequence and the polypyrimidine tract adjacent to the 3' splice site, that facilitate the specific splicing of pre-mRNA (Wahl et al. 2009). The minor spliceosome on the other hand, consists of different but functionally similar set of U11/U12, U4atac/U6atac and U5 (common to both major and minor

spliceosomes) snRNPs, which recognize a different set of minor introns, known as 'U12-type introns' (Patel & Steitz 2003). Although initially thought to recognize specific AT-AC dinucleotides at the 5' and 3' SS of introns, the minor spliceosome is now known to recognize a set of longer and tightly constrained consensus sequences at the 5' SS and branch point of minor introns (Patel & Steitz 2003). The spliceosome machinery facilitates splicing of both constitutive and alternatively spliced exons in a step-wise manner, assisted by the numerous *trans*-acting factors that facilitate the recognition and definition of exon/intron boundaries (De Conti et al. 2013).

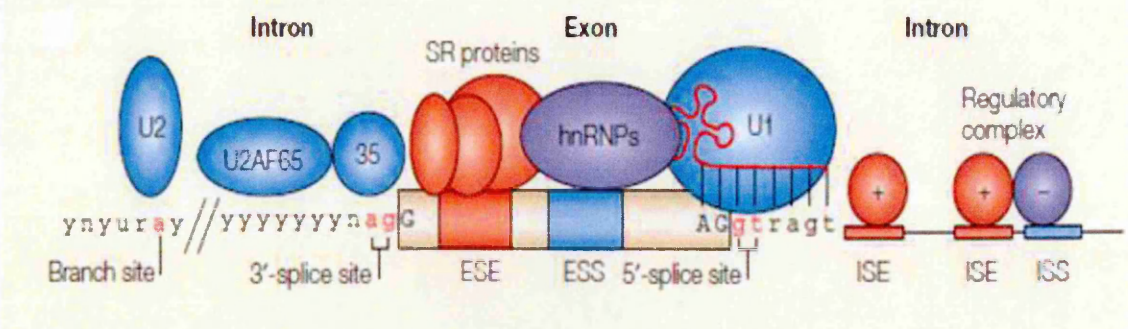



Figure 1-2: **Schematic representation of factors involved in splicing.** Consensus sequences found in the 5'/ 3' positions of introns also known as splice sites, and including the branch site, determine which splicing factors and components of the spliceosome bind to facilitate the splicing process. Enhancer or silencer elements are also found within the exons or introns and are specific for certain proteins. As such, Serine/Arginine (SR) proteins and hnRNPs are generally thought to bind to enhancer or silencer elements respectively, thereby influencing the recognition of splice sites by the spliceosomal complex. ISE-Intronic splicing enhancer; ISS-intronic splicing silence; ESE-Exonic splicing Enhancer; ESS-Exonic splicing silencer. Adapted from Pagani and Baralle, 2004.

In addition to the three core splicing sequences, additional elements within introns and exons, are present. These sequences known as intronic and exonic silencers or enhancers, bind *trans*-acting splicing factors that modulate the efficiency of splice site recognition, thereby determining which exons are constitutively or alternatively expressed (Figure 1-2) (Pagani & Baralle 2004; Ward & Cooper 2011). The combinatorial control and competition of these factors is thought to play an important role in the definition of introns/exons, including alternative poly-A site selection (De Conti et al. 2013).

The temporal expression of these *trans*-acting factors can determine how and where a pre-mRNA transcript is spliced, necessitating a tightly regulated homeostatic control of these factors (Tazi et al. 2009; Lutz & Moreira 2011; McManus & Graveley 2012). A perturbation in any of these components including the structure of mRNA could lead to mis-processing of transcripts thus resulting in disease (Pagani & Baralle 2004).

Furthermore, the differential complexity of gene isoforms can be achieved by different types of alternative splicing events, including cassette exon inclusion or exclusion, alternative 5' and 3' splice site selection, mutually exclusive exons, intron retention and alternative promoter and poly-adenylation (poly-A) site selection (Blencowe 2006). The most common of these types of alternative events are cassette exons, accounting for approximately one third of known alternative splicing (AS) events (Blencowe 2006).

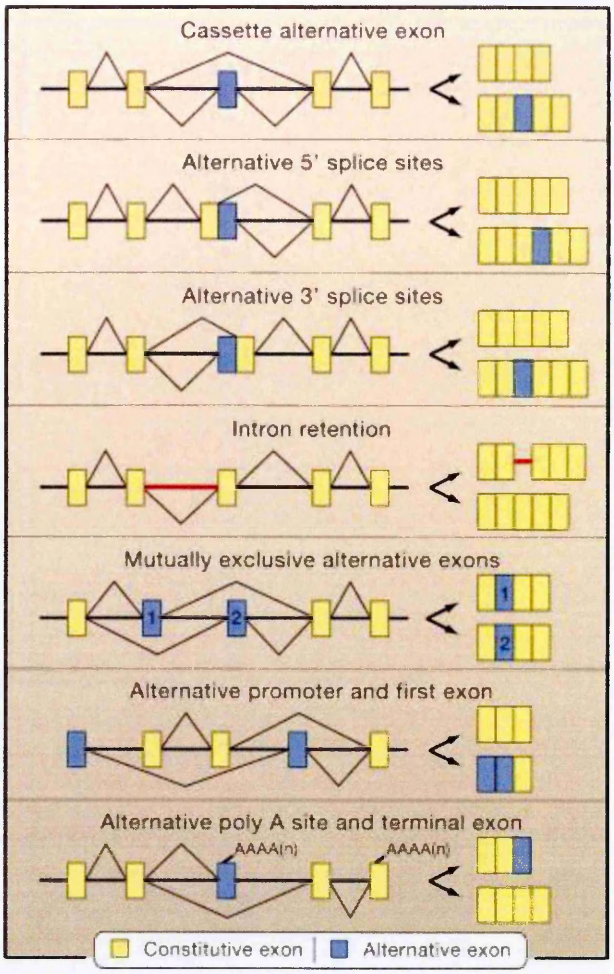



Figure 1-3: **Schematic representation of the various types of alternative splicing that can occur.** The most common type of alternative splicing decribed to date is cassette exon inclusion or exclusion. Within each pre-mRNA, different kinds of splicing events can occur depending on the concentration and type of *trans*-acting factors recruited to the sequence elements (Blencowe, 2006).

Apart from creating transcriptomic and consequently proteomic diversity, alternative splicing can also function to further regulate gene expression post-transcriptionally, such that aberrantly spliced transcripts are targeted for degradation. Aberrant splicing can lead to the introduction of premature termination codons (PTC) within transcripts, alter mRNA stability or even alter the localization of mRNA, resulting in degradation mechanisms such as non-sense mediated decay (NMD) (Soergel et al. 2006).

#### 1.3.1.1. *TDP-43 in alternative splicing*

A role for TDP-43 in splicing was first brought to light in a study by Buratti et al. (2001), in which TDP-43 was shown to be a negative regulator (promoting skipping/exclusion) of exon 9 in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. This study also identified a polymorphic TGmTn repeat in the 3' end of intron 8, to which TDP-43 bound and resulted in the skipping of exon 9 (Buratti et al. 2001). Subsequent work by the same group showed that TDP-43 through its RRM domains, preferentially binds to UG repeats in single stranded DNA or RNA, thus confirming the nucleic binding capacity of the RRM (particularly RRM 1) domains (Buratti & Baralle 2001). In addition, the role of TDP-43 in splicing was further validated in Mercado et al. (2005) where depletion of TDP-43 was found to result in the inclusion of exon 3 of the apolipoprotein A-II (*apoA-II*) pre-mRNA. It should be emphasised however, that TDP-43 in most splicing situations does not act independently and binds to other proteins including other hnRNPs and splicing factors. Indeed, further work by Buratti et al. (2005) showed that TDP-43 was able to bind to hnRNP A1 and hnRNP A2/B1 through its C-terminal and form a repressor complex that resulted in the skipping of exon 9 in *CFTR* pre-mRNA. This study also confirmed the protein-protein binding role of the glycine-rich C-terminal of TDP-43, which is also a characteristic feature of the hnRNP group of proteins (Buratti et al. 2005; Ayala et al. 2005).

TDP-43 was also reported to be a negative splicing regulator role in a more recent study that analysed the negative feedback loop of a splicing factor SC35, wherein TDP-43 was shown to competitively bind the terminal exon sequences in the 3'UTR of the *SC35* gene (Dreumont et al. 2010). The binding of TDP-43 to the terminal exon resulted in the inhibition of SC35-mediated skipping of the terminal intron (normally retained), thereby inhibiting un-productive splicing that serves as negative feedback mechanism (Dreumont et al. 2010).

In contrast, another study reported a positive regulatory role for TDP-43 in splicing. Specifically, TDP-43 was found to enhance the inclusion of exon 7 in survival of motor neuron (*SMN*) pre-mRNA, through a multimeric complex formation with another positive splicing factor, (Transformer-2 protein homolog beta) Htra2-β1, resulting in increased expression of the *SMN* transcript (Bose et al. 2008). This study was however, performed by overexpressing TDP-43 which may not recapitulate the natural situation *in vivo* where protein levels are tightly regulated, and other studies that are in agreement with this role are required. Indeed, in the same study no changes in the splicing pattern were observed when TDP-43 was depleted. Nonetheless, the concept of TDP-43 having a dual role in splicing regulation is not novel and has been reported for other proteins (Bose et al. 2008).

The link between TDP-43 and alternative splicing provides further support for the involvement of altered RNA processing in TDP-43 proteinopathies, since theoretically, a lack of TDP-43 function in both the nucleus and cytoplasm should disrupt cellular homeostatic control and expression of multiple transcripts.

#### 1.3.1.2. *TDP-43 in transcriptional regulation*

Apart from alternative splicing, other mechanisms involved in gene expression regulation include transcriptional repression or activation, mRNA stabilisation and localisation and TDP-43 has been shown to be involved in a few of these processes. In fact, TDP-43 was first identified as a transcriptional repressor in HIV as a result of binding to the trans-active response (TAR) element (Ou et al. 1995). The mechanisms involved in TDP-43 mediated transcriptional repression in this case remain unknown, and do not require TG repeats but is reported to involve pyrimidine-rich sequences (Ou et al. 1995). Another role for TDP-43 in transcriptional repression was reported for human intraacrosomal protein *SP-10* (*ACVR1*). In this case, TDP-43 was able to repress premature expression of SP-10 by

binding to a complementary strand containing TG repeats in the promoter (Acharya et al. 2006).

#### 1.3.1.3. *microRNA regulation, stress granules and RNA translation*

Other regulatory roles for TDP-43 have been reported. A previous study by Gregory et al. (2005) found TDP-43 to be in a complex with Dicer in the cytoplasm where it was reported to play a role in microRNA (miRNA) biogenesis, including being involved in target RNA cleavage. A subsequent study showed that TDP-43 was able to bind to and control expression levels of certain miRNAs; specifically let-7b which was down-regulated and miR-663 which was up-regulated (Buratti et al. 2010b). The authors were also able to show that several other mRNA transcript expression levels were altered as a result of these interactions; among them, *DYRK1A, STX3, VAMP3* and *LAMC1* which are known to be involved in neuronal processes such as growth, synapse formation and exocytosis (Buratti et al. 2010b). Further support for the involvement of TDP-43 in miRNA biogenesis and regulation was provided in recent work, which showed that TDP-43 was able to bind both the Drosha and dicer complexes through protein-protein interactions and directly by binding specific miRNAs (Kawahara & Mieda-Sato, 2012). The previously reported changes in expression levels of let-7b and miR-663 were however not observed, and this was attributed to the different cell lines used in the different studies (Kawahara & Mieda-Sato 2012).

Several other reports have shown that TDP-43 and FUS/TLS are present in transient cytoplasmic stress granules bound to RNA molecules, which is thought to be facilitated by their ability to shuttle between the nucleus and cytoplasm (Liu-Yesucevitz et al. 2011; Bentmann et al. 2012). It is still debated as to whether stress granules are involved (i.e. as seed) in aggregate formation, since they have been found in brains of ALS patients, however the fact that stress granules contain RBPs including TDP-43 provides a plausible

explanation for aggregate-independent toxicity due to depletion of the RNA molecules (Liu-Yesucevitz et al. 2011). In addition, TDP-43 has been found in RNA granules at neuron synapses where it is thought to respond to neuronal activity (Liu-Yesucevitz et al. 2011). The presence of TDP-43 in the RNA granules has been linked to stability and translation of RNA (Buratti & Baralle 2010; Freibaum et al. 2011). The study conducted by Freibaum (2011), a global proteomic analysis which identified proteins that co-immunoprecipitated with FLAG-tagged TDP-43, further supports a translation role for TDP-43 as it was found to be bound to numerous mRNA translation proteins in the cytoplasm.

With regards to mRNA stabilisation, TDP-43 was shown to bind to UG sequences in the 3'UTR of lower molecular weight neurofilament light polypeptide (*NF1*), thus stabilising the transcript (Volkening et al. 2009). TDP-43 also interacted with both SOD-1 and 14-3-3 proteins, known binders of NF1 for increased stability. Furthermore, the fact that NF1 mRNA was also found in stress granules of ALS neurons, supports a role for altered mRNA processing in disease (Volkening et al. 2009).

#### 1.3.1.4. *TDP-43 autoregulation: A negative feedback loop*

Aside from having a regulatory role in the processing of other transcripts, TDP-43 has also been shown regulate its own expression through a negative feedback loop. A negative feedback autoregulatory mechanism has been reported for a number of regulatory proteins including hnRNP L (Rossbach et al. 2009) and Arginine/Serine (SR) proteins (Lareau et al. 2007) and the splicing factor SC35 (Dreumont et al. 2010). These proteins are able to bind to sequences within their own pre-mRNA resulting in unproductive splicing products that are targeted for degradation through various mechanisms including NMD, thus establishing a post-transcriptional control of protein levels (Rossbach et al. 2009; Dreumont et al. 2010). In these instances the binding of 'self-proteins' results in

alternatively spliced transcripts, which depending on their composition determines the fate of the transcript.

In keeping with this mechanism, autoregulation through a negative feedback loop has been described for TDP-43. In the work done by Ayala et al. (2011), TDP-43 was reported to bind its 3'UTR, resulting in the down-regulation of nascent transcript under conditions of TDP-43 overexpression. In addition, the authors were able to prove that the capacity of TDP-43 to bind RNA was necessary for this autoregulation, as mutation of the phenylalanine (F) residues in RRM1 (reported to play the major role in RNA binding) led to a lack of regulation (Ayala et al. 2011). Interestingly, NMD was reported to not be involved in the degradation of the two major transcripts that utilised two different poly-adenylation (poly-A) sites in autoregulation (Ayala et al. 2011; Avendaño-Vázquez et al. 2012).

Subsequently, the same group identified the mechanisms involved in TDP-43 autoregulation, described to be a complex interplay between transcription, alternative splicing of an intron in the TDP-43 3'UTR and the alternative selection of poly-A cleavage sites (Avendaño-Vázquez et al. 2012). More specifically, they were able to show that the recognition of intron 7 for splicing within the TDP-43 3'UTR was dependent on RNA Polymerase II (RNA Pol II) stalling, which led to TDP-43 competitively binding to a CstF64 (Poly-A factor) binding site ($GU_{3-5}$, $U_{2-4}$). This interaction led to the splicing out of intron 7, which also contains the first poly-A in TDP-43, thus resulting in the use of an alternative second poly-A. The resultant transcript was found to be retained in the nucleus, similar to the transcript that utilises the fourth poly-A (Avendaño-Vázquez et al. 2012). A model depicting the autoregulation mechanism of TDP-43 is shown in Figure 1-4, below.

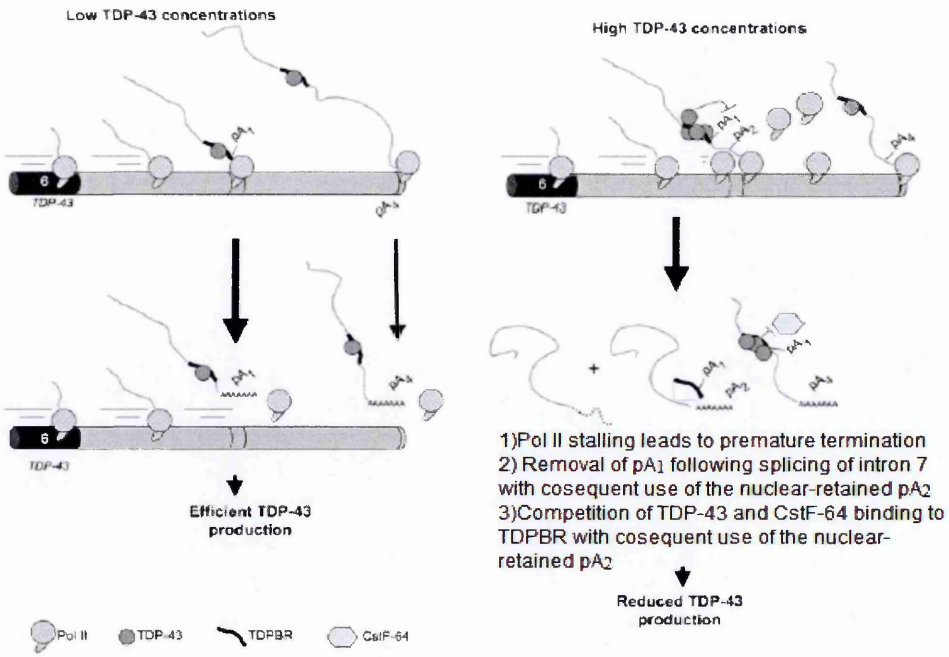



Figure 1-4: **Model of TDP-43 of autoregulation.** The binding of TDP-43 to non-canonical sites in its own 3'UTR results in unproductive splicing events, which are targeted for degradation. Under physiological levels TDP-43 can still bind to its 3'UTR, however, the number of bound molecules is determined through competition with the CstF-64. In contrast, high or increased cellular levels of TDP-43, tip the balance in favour of TDP-43 resulting in the recognition and definition of exon 7. The unproductive splicing of intron 7 in the TDP-43 3'UTR also leads to an alternative polyadenylation site selection, poly-$A_2$ which is retained in the nucleus.

*In vivo* studies using mice expressing transgenic human TDP-43 have further supported the notion of TDP-43 autoregulation, as transgenic TDP-43 was able to reduce endogenous levels of mouse TDP-43 transcript, proving that this mechanism is conserved amongst various vertebrate species (Avendaño-Vázquez et al. 2012).

TDP-43 autoregulation through alternative splicing and differential poly-A site selection further emphasizes the role of TDP-43 in gene expression regulation. This observation coupled with the fact that TDP-43 is involved in numerous post-transcriptional regulation mechanisms supports the concept of TDP-43 being a global transcript regulator. Thus, it is not difficult to imagine that at the transcript level, dysregulation in TDP-43 levels could lead to a dysregulation cascade, in several transcripts through both direct and indirect interactions. Indeed, TDP-43 could be at the centre of a complex regulatory network. The

exact role of TDP-43 in the cell, however, still remains to be defined conclusively. Indeed, a current spurt of research is looking into TDP-43 interactions at both RNA and protein level in an effort to provide a correlation with pathogenesis.

## 1.4. Characterization of TDP-43 proteinopathies

The newly coined term 'TDP-proteinopathies' refers to a group of neurodegenerative disorders in which, TDP-43 has been identified as a major protein in pathological aggregates/inclusions in patients (Cohen et al. 2011). As has been discussed previously, the identification of TDP-43 in pathological inclusions of patients with ALS and FTLD sparked a paradigm shift in neurodegenerative research resulting in the discovery of TDP-43 inclusions in other disorders such as Alzheimer's, Huntington's hippocampal sclerosis and corticobasal degeneration (Neumann 2009). Currently, ALS and FTLD fall within this clinico-pathological group of TDP-43 proteinopathies that affect different subsets of neurons and are viewed as broad spectrum of a single disease (Mackenzie et al. 2007; Geser et al. 2009; Mackenzie et al. 2010; Baloh 2011). TDP-43 proteinopathies are characterised by TDP-43 accumulation in round inclusions bodies or skeins that are associated with nuclear clearance, although some nuclear inclusions have been reported (Arai et al. 2006; Neumann, et al. 2006). Within these pathological inclusions TDP-43 is found highly ubiquitinated and phosphorylated, with cleaved C-terminal fragments perceived to be toxic, also present (Arai et al. 2006; Neumann et al. 2006; Neumann 2009).

In an effort to gain insight into the pathological mechanism involved in TDP-43 proteinopathies, several studies have performed sequence analyses of TDP-43 in patients compared to controls thus identifying several mutations segregating with disease (Rutherford et al. 2008; Pesiridis et al. 2009; Neumann 2009; van Blitterswijk et al. 2012). Currently, over 40 mutations have been identified in TDP-43 (Gendron et al. 2013). Remarkably, most of these mutations are concentrated on the glycine-rich C-terminal of the protein with a subset being found in the 5'/3' un-translated regions (UTR) and intronic

regions (Pesiridis et al. 2009). The high frequency of mutations in the glycine-rich region of TDP-43 implicates the protein's capacity to bind and interact with other proteins (C-terminal functionality) as a potentially key process in pathogenesis (Pesiridis et al. 2009). The numbers of mutations identified relative to the FTLD and ALS TDP proteinopathies vary, with the majority being linked to ALS. Although the frequency of TDP-43 mutations in FTLD is low (five identified thus far), most co-segregate with a sub-type of FTLD with motor neuron disease (FTLD-MND) (Gitcho et al. 2008; Gitcho et al. 2009; Pesiridis et al. 2009; Neumann et al. 2009) which emphasises the notion of a disease continuum in TDP-43 proteinopathies. Nonetheless, pathological accumulations of TDP-43 are still observed in FTLD linked to mutations in other genes, such as pro-granulin (*PGRN*), Valosin-containing protein (*VCP*) and charged multivesicular body protein 2B (*CHMP2B*) (Cairns et al. 2010; Neumann et al. 2009), which implies that TDP-43 accumulation might be a secondary pathological mechanism in these diseases.

On the other hand, approximately 30 mutations identified in TDP-43 thus far, as shown in Figure 1-5, have been linked to ALS (http://www.molgen.vib-ua.be/FTDMutations; ALS database also present) (Cruts et al. 2012), with the most commonly reported of these being the A382T and G348C mutations as reviewed in Kabashi et al. (2008) and Pesiridis et al. (2009). Phenotypically, no major differences exist between familial and sporadic forms of ALS with mutations in TDP-43 apart from the autosomal dominant inheritance patterns observed in familial ALS (Geser et al. 2009; Neumann 2009). Other studies have suggested an oligogenic effect of TDP-43 proteinopathies that is not only linked to mutations in TDP-43 but in other genes as well (van Blitterswijk et al. 2012).

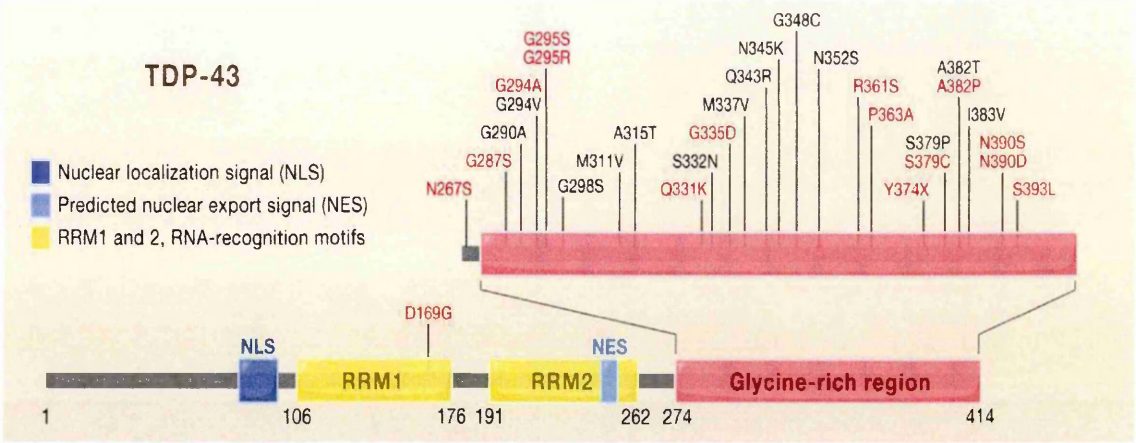



**Figure 1-5: Schematic diagram of the distribution of mutations in TDP-43.** The majority of mutations are found in the Glycine-rich terminal of TDP-43 with the exception of D169G found in RRM1. Mutations highlighted in red have been identified in sporadic ALS whereas those highlighted in black are familial (Lagier-Tourenne & Cleveland 2009).

With regards to understanding the biological relevance of these mutations, functional *in vivo* and *in vitro* analyses of some of these mutations have been shown to recapitulate the phenotypes observed in patients i.e. proteinaceous cytoplasmic inclusions and presence of cleaved C-terminal fragments, although exact mechanisms still need to be elucidated (Rutherford et al. 2008; Pesiridis et al. 2009; Neumann 2009). Currently, none of the identified mutations have been linked to altered splicing activity, mRNA instability or protein-protein interaction (Mackenzie et al. 2010), which implies an interplay between genetic risk factors and the involvement of other yet unknown factors in disease mechanisms. In complex diseases such as ALS and FTLD, where the environment and genetic risk factors are thought to play a role, functional biochemical assays may not provide a satisfactory answer. Indeed, it has been reported that most mutations in TDP-43 as observed in large pedigrees, tend to have low penetrance (moreso in sporadic ALS), strongly suggesting the involvement of other modifying factors (van Blitterswijk et al. 2012; Onodera et al. 2013). The involvement of these yet unknown modifying factors could provide clues as to the selective formation of these protein inclusions in specific subsets of neurons and why single mutation hit in TDP-43 does not always result in disease.

### 1.4.1. *TDP-43 pathology: Loss of function or Gain of Function?*

The presence of aggregates in many neurodegenerative diseases, including ALS, presents a conundrum. Several speculations and hypotheses exist as to whether the aggregates are the cause of neurodegeneration, or are an un-related epiphenomena signalling cellular distress or even a protective mechanism for the cell (Baloh, 2011). Typically, the cytoplasmic inclusions found in TDP-43 proteinopathies have been shown to consist mostly of cleaved C-terminal fragments of TDP-43 accompanied with a nuclear depletion of the protein (Arai et al. 2006; Neumann et al. 2006; Neumann 2009; Mackenzie et al. 2010). Two main theories, therefore, exist with regards to the pathological mechanism of TDP-43 proteinopathies. One theory proposes a role for the aggregates observed in patients to be a result of a gain of toxic function, in which misfolded and cleaved TDP-43 induce cellular toxicity. The other hypothesis, which correlates with the observed nuclear depletion of TDP-43, proposes that the sequestration of the wild-type full-length protein within the aggregates could lead to a loss of function in the nucleus.

With regards to the former, in which the aggregates are perceived to have a gain of function mechanism, several mechanisms could be at play, including the sequestration of functional endogenous protein, or the sequestration of other factors such as TDP-43 binding partners or RNA that could result in toxicity. Within these aggregates as has been described previously, TDP-43 C-terminal fragments of 25 kDa and 35 kDa (with truncated N-terminal domains) have been observed in brain cytoplasmic protein inclusion and are thought to induce cellular toxicity as a result of subsequent post-translational modifications or as precursors for aggregation formation (Rutherford et al. 2008; Geser et al. 2009; Zhang et al. 2009). In support of the theory of toxic C-terminal fragments, two studies by Zhang et al. have shown that caspase-mediated cleavage of TDP-43 is a prior event to cytoplasmic translocation (Zhang et al. 2007), and that the 25 kDa fragment that is hyper-phosphorylated in residues 409/410 induces cellular toxicity by mechanisms that are still

unknown (Zhang et al. 2009), implicating a role for cleavage and phosphorylation in the pathological process (Zhang et al. 2013). Furthermore, the C-terminal fragments of TDP-43 containing truncated RRM2 have been demonstrated to be aggregation prone, through the formation of β-sheet strands that are thought to abnormally bind wild-type TDP-43 protein (Winton et al. 2008; Igaz et al. 2009; Zhang et al. 2009; Wang et al. 2013). More importantly, the C-terminal domain of TDP-43 has been shown to contain a glutamine and asparagine (Q/N) prion-like region that could enhance aggregate formation (Fuentealba et al. 2010; Budini et al. 2012; King et al. 2012). Considering the known features of prion proteins, i.e. a misfolded seed can recruit proteins, it is not difficult to imagine that the presence of these domains in TDP-43 could contribute to the formation of aggregates among neurons that have encountered an injury or change in homeostasis (Cushman et al. 2010; Furukawa et al. 2011). Interestingly, aggregate formation may not only involve the TDP-43 C-terminal, as a recent study implicated the extreme N-terminus of TDP-43 in the formation of oligomers that in turn form aggregates which sequester the wild-type full-length protein into inclusion bodies (Zhang et al. 2013). In contrast, a new dynamic related to aggregates has recently been proposed in an *in vivo* study conducted in *D. melanogaster*, in which the aggregates were found to confer a protective role in the retina against overexpression of TBPH (a TDP-43 ortholog) that was toxic (Cragnaz et al. 2014).

On the other hand, in support of the loss of function hypothesis, a recent review by Budini and Buratti (2011) proposed that a slight perturbation in the delicate balance of TDP-43 regulation, could tip the scale towards an over expression of TDP-43 leading to increased aggregate formation that sequesters the full-length protein, thus creating a TDP-43 'sink' that culminates in nuclear depletion. A lack of TDP-43 in the nucleus could lead to a break-down in the negative feedback loop resulting in increased production of TDP-43 protein that could further drive aggregate formation (Budini & Buratti 2011; Polymenidou et al. 2012a). Since TDP-43 has a natural propensity to aggregate, it is possible that the

increased amount of TDP-43 in the cell leads to increased oligomerisation between TDP-43 molecules (Figure 1-6).

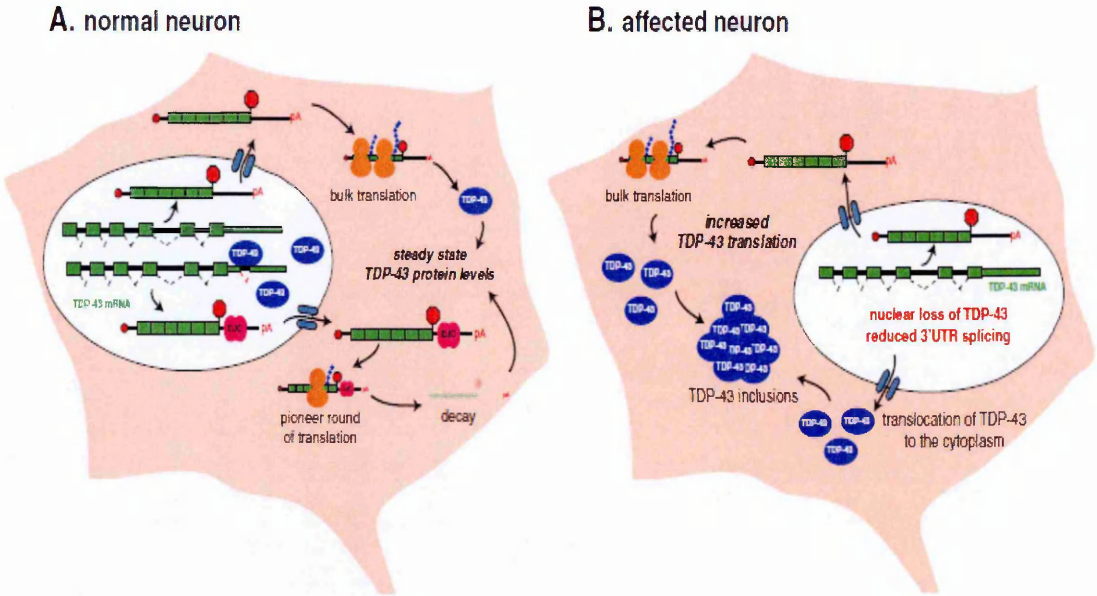



Figure 1-6: **Schematic diagram depicting the combinatorial effect of TDP-43 aggregation and loss of function in the cell.** Accumulation of TDP-43in the cytoplasm could create a perturbation in the autoregulatory loop mechanism of TDP-43 resulting in increased expression of TDP-43.TDP-43 inclusions may also serve as seed for increased recruitment of TDP-43 and other proteins (Polymenidou et al. 2012).

In keeping with efforts to understand triggers for aggregation, a study that examined the behavior of TDP-43 with mutated nuclear localization and export signals (NLS and NES), found that abolishing either one of these regions resulted in TDP-43 proteinopathy that was reminiscent of that observed in patients i.e, aggregate formation, hyper-phosphorylation and ubiquitination (Winton et al. 2008). However, up until recently, none of the previous studies had reported a functional effect as a result of endogenous TDP-43 sequestration within the aggregates. To this end, a follow up study building up on a previous cellular aggregation model by Budini; Romano et al (2014), found that when full length TDP-43 was fused to 12 repetitions of the Q/N region found in the C-terminal of TDP-43, the resulting aggregates were able to sequester endogenous wild-type protein, resulting in a

loss function, as determined by the altered splicing profile analyses of exons 9 and 3 of *CFTR* and Polymerase delta interacting protein 3 (*POLDIP3/SKAR*), respectively.

Interestingly, *in vivo* studies of TDP-43 knock-out showed that TDP-43 is essential for development as no embryos survived (Baloh, 2011; Budini et al. 2012). However, another study performed in mice, reported that when TDP-43 was conditionally knocked-down, progressive ALS-like symptoms were observed including motor-dysfunction and motor neuron loss, supporting a loss of function hypothesis and a role for TDP-43 in the survival of motor neurons (Wu et al. 2012).

Still, despite the evidence in support of both hypotheses, there is no consensus on the mechanisms that lead to TDP-43 pathology (Da Cruz & Cleveland 2011) and given the multi-functional role of TDP-43, it could very well be a combination of both mechanisms i.e. loss of function of the endogenous protein and gain of function of the subsequent aggregates, in addition to alterations in protein homeostatic mechanisms regulating other factors that interact with TDP-43. The fact that ALS and other TDP-43 proteinopathies are complex neurodegenerative diseases further provides challenges for researchers with regards to identifying the exact disease mechanisms that could inform the development of therapeutics.

### 1.4.2. ***TDP-43 proteinopathy disease models***

Efforts towards understanding disease mechanisms in TDP-43 proteinopathies have incorporated several disease models, each with their unique limitations. The use of *in vitro* systems such as cellular models could provide partial insight albeit an incomplete picture, into factors involved in pathogenesis. For instance, Budini et al. (2012) engineered a cellular model in which TDP-43 aggregation and consequent effects could be studied. In effect, they were able to map the TDP-43 Q/N region, corresponding to 321-369 amino acids which facilitated binding of TDP-43 to other proteins (hnRNPs) and played a role in

aggregation, as determined by the results of a synthetic construct containing 12 repeats (12X) of the Q/N region, fused to a GFP reporter (Budini et al. 2012). Although studies using this model were able to recapitulate the formation of aggregates, no toxicity and or cleaved C-terminal TDP-43 fragments were observed and only minimal amounts of endogenous TDP-43 were sequestered within the aggregates (non-functional), highlighting a temporal limit of this cellular model as suggested by the authors (Budini et al. 2012). Nonetheless, a build-up on this study by the same authors incorporating full length TDP-43 fused to the above mentioned 12Q/N repeats, was able to functionally ascertain the sequestration of the endogenous protein within the aggregates, thereby supporting a loss of function hypothesis disease mechanism involving TDP-43 (Budini; Romano et al. 2014).

*In vivo* disease models in part, are able to provide a solution for the above mentioned limits with the added advantage of a relatively easily manipulated system that closely resembles that of humans. For TDP-43 proteinopathies specifically, various systems based on protein homology have been identified, including Mouse, Rat and *D. melanogaster*. Other organisms such as yeast (*S. Cerevisiae*) (Johnson et al. 2008; 2009) and bacteria have also been used as host systems for expression of transgenic human TDP-43. TDP-43 bears a cumulative 70% similarity with *M. musculus* and *D. melanogaster* indicating a shared function of the orthologous proteins in their respective species (Wang et al. 2004).

The use of *D. melanogaster* as a model for studying neurodegenerative diseases is appealing due to the presence of high neuronal complexity (different cell types and neurotransmitters) and brain function (learning and memory), similar to humans (Romano et al. 2012). Furthermore, *D. melanogaster* have a relatively short life-cycle and are easily manipulated using inducible gene expression systems. Indeed, a study by Ayala et al (2005) showed that hTDP-43 and *D. melanogaster* TDP-43 also known as TBPH, bear functional similarities in RNA binding and roles in splicing. Functional similarity between TDP-43 and TBPH was further confirmed in a study that analysed TDP-43 regulation of

*HDAC6*, wherein knockout of *TBPH* resulted in the down-regulation of the *HDAC6* transcript (Fiesel et al. 2010).

With regards to TDP-43 pathology, several insights using the *D. melanogaster* model have been gained. For instance, over-expression of TBPH or hTDP-43 in flies has been reported to result in toxicity-induced neurodegeneration that is not linked to cytoplasmic inclusions (Feiguin et al. 2009; Li et al. 2010). On the other hand deletion or complete knockout of TBPH was shown to greatly reduce survival as embryos only survived to the second instar larval stage (Fiesel et al. 2010). In addition, partial deletion of TBPH was reported not to change viability of embryos but adult flies showed decreased motility and lifespan as a result of underdeveloped neuromuscular junctions (less axonal branches) (Feiguin et al. 2009). Another study confirming functional similarity between TDP-43 and TBPH showed that hTDP-43 was able to rescue/increase dendritic branching in genetic null mutants of TBPH (Lu et al. 2009). These phenotypes closely resemble those observed in patients and emphasise the suitability of using *D. melanogaster* as a disease model. In keeping with this, other studies have examined the effect of previously identified mutations in TDP-43 and reported enhanced neurotoxicity as well presence of aggregate and fibril formation (Estes et al. 2011; Guo et al. 2011). Nonetheless, the sequence similarity of TBPH and hTDP-43 (77% at the N-terminal and 22% at the C-terminal (Romano et al. 2012) could account for disease phenotypic differences observed in humans and other animal models such as rodents that have a much higher sequence similarity.

In fact several studies, conducted on mice (95% homology to hTDP-43) and rats (reviewed in Baloh, 2011 and Tsao, 2012) have taken advantage of this fact in an attempt to create TDP-43 disease models that can provide insight into the pathogenesis of the disease. Based on the previously discussed hypotheses on the pathogenicity of TDP-43, transgenic mice and rat lines that over-express and knock-down TDP-43 have been studied (Tsao et al. 2012). In particular, mice over-expressing transgenic hTDP-43 under heterologous

promoters that target specific subsets of cells, including neurons have been found to be consistently toxic across various studies in a dose-dependent manner (Da Cruz & Cleveland 2011; Baloh 2011; Tsao et al. 2012). Other studies aiming to overcome the issue of toxicity due to over-expressing TDP-43 used a bacterial artificial chromosome (BAC) gene expression system in mice and rats, and found that the lower levels of expression (almost 3X more than endogenous-still quite high), recapitulated better the physiological features observed in patients (Zhou et al. 2010; Swarup et al. 2011). On the other hand, as has been mentioned previously, knock-down of TDP-43 is embryonic lethal and a conditional knockout was shown to result in death soon after induction, highlighting that TDP-43 is not only important for development but also for survival (Sephton et al. 2010; Kraemer et al. 2010). In addition, conditional knockout of TDP-43 in post-natal mice did not show a decrease in protein expression levels nor exhibition of any ALS or FTLD-like symptoms, except for a lean phenotype, which proved un-related (Chiang et al. 2010).

Nonetheless, mice expressing transgenic human or mouse TDP-43 with mutations or altered levels of expression, are reported to exhibit similar phenotypic features of TDP-43 pathology as observed in patients including, cognitive impairment, selective loss of neurons, motor deficits, gait abnormality and lethality to varying extents depending on the type of promoter used in the study (Da Cruz & Cleveland 2011; Tsao et al. 2012). Line to line variability has also been observed making it difficult to reproduce results or make conclusive analyses. Still, expression of transgenic wild type or mutant TDP-43 in rodent models, broadly recapitulates pathological features as observed in patients, including nuclear depletion, C-terminal fragments and hyper-phosphorylation (Baloh 2011). In addition, it appears that the cellular inclusions containing phosphorylated TDP-43 aggregates are not an outstanding feature of these animal models (Baloh, 2011) unlike in patients where inclusions are a hallmark of disease.

The use of transgenic mice and rats as *in vivo* disease models however, highlight a few limitations. Firstly, human TDP-43 may be non-functional (viewed as mutant) in the rodent background resulting in toxicity, through a perturbation of endogenous mouse tdp43 homeostasis (Da Cruz & Cleveland 2011; Baloh 2011; Tsao et al. 2012). Secondly, the physiological variability of symptoms observed could point towards a necessity for controlled and matched mice genetic backgrounds, despite the use of identical promoters. Thirdly, the fact that the characteristic protein inclusions observed in patients are not regularly observed in the rodents could mean that the proposed toxicity of TDP-43 pathology is achieved in a manner that is independent of aggregate formation in these animals and does not provide clues into the perceived role of these aggregates in patients (Da Cruz & Cleveland 2011; Baloh 2011).

In parallel with mammalian disease models of TDP-43, other studies have looked at other eukaryotic organisms for TDP-43 pathology answers, including yeast. Although yeast lack a true TDP-43 homolog, over-expression of TDP-43 in yeast was shown to recapitulate the characteristic features observed in cell lines and patients, where cytoplasmic protein aggregation and toxic C-terminal fragments were observed (Johnson et al. 2008; Johnson et al. 2009). In addition, yeast studies have supported hypothesis that cellular toxicity may be independent of aggregates (despite their presence) and that the RRM or RNA binding capacity of TDP-43 is necessary for induction of toxicity (Johnson et al. 2008; 2009; Baloh 2011). Other models such as zebrafish of TDP-43 proteinopathy have also been reported to exhibit a motor neuron phenotype when TDP-43 was knocked down and these symptoms could be rescued upon overexpression of wild type TDP-43 (Kabashi et al. 2008). Notably, in invertebrate models of TDP-43 proteinopathies, there is a convergence on the theory of toxicity being independent of cytoplasmic inclusions and is thought to involve a modulation within TDP-43 RNA binding partners (Baloh 2011).

More recently, with the discovery that somatic cells could be reprogrammed, the use of induced pluripotent stem cells (iPS) as a disease model has gained momentum. iPS cells have provided a platform in which cells from patients could be used to understand pathogenesis, based on mutations (Mattis & Svendsen 2011; Patani et al. 2012). With regards to TDP-43 proteinopathies, a recent study that used iPS cells reported an increased vulnerability of motor neurons due to a M337V mutation in TDP-43 inherent in a patient, following stress induction through inhibition of key signalling pathways (Bilican et al. 2012). In addition, the authors also reported an increase in levels of protein including soluble and detergent-insoluble TDP-43, although no cytoplasmic inclusions were observed (Bilican et al. 2012). A similar study that examined phenotypic differences in a patient (A90V) and control iPS cells upon stress induction, reported a higher percentage of TDP-43 in the cytoplasm, lower levels of TDP-43 and down-regulated mir-9 (Zhang et al. 2013).

Despite being a more closely related model in terms of disease and species specificity, there are still some limits of using iPS cells as disease models. For instance, the fact that the cells undergo genetic reprogramming means that cells are reverted to a somewhat "younger" state which would make it difficult to study age-related factors in connection with complex disorders such as ALS that also have sporadic forms (Mattis & Svendsen 2011). Furthermore, in the context of the brain and other tissues, neuronal cells are normally surrounded by other differentiated cells and factors, and the lack thereof in cell culture conditions, may not recapitulate the cellular milieu that initially led to disease. Coupled with this, is the fact that in disease situations, specific subsets or populations of neurons are affected, and so the challenge would be to enrich iPS cells of these specific subsets in a reproducible and predictable manner (Patani et al. 2012). In addition, it has been reported that these somatic cells do not undergo complete reprogramming and could retain epigenetic memory of the original tissue (Kim et al. 2010; Mattis & Svendsen 2011).

With the rapid progression of advances in technology, it is possible that these limitations will soon be addressed.

Thus, lessons from disease models of TDP-43 proteinopathies, point towards a cellular toxicity whose exact mechanism remains unknown. The involvement of aggregates is still quite heavily debated and as such there is no consensus on whether TDP-43 pathology exerts a toxic gain of function or loss of function mechanism. The latter seems a more attractive candidate for pathogenesis given the nuclear clearance observed in TDP-43 proteinopathies, implicating a disturbance in the autoregulation of TDP-43, although, the involvement of both mechanisms could be at play. However, for plausibility of the loss of function theory, further studies looking into the physiological role of TDP-43 in the cell are required. To this end, there has been a flurry of research aimed at identifying targets of TDP-43 at a global scale, utilising advanced high-throughput technologies.

## 1.5. Identification of TDP-43 RNA targets

In the post-genomic era, several high-throughput technologies have been developed that enable global analyses of cellular changes in both the transcriptome and proteome in large data sets, including fold differences in expression (Buratti et al. 2012). With regards to analysing global transcriptomic changes, technologies such as microarrays (Johnson et al. 2003) and next-generation sequencing have been particularly useful in distinguishing between alternatively spliced transcripts, including expression levels of the various isoforms (Blencowe 2006; Buratti et al. 2012). Coupled to sequencing, technologies such as UV cross-linking immunoprecipitation (CLIP) are able to identify *trans*-acting factor binding sites *in vivo*, which provide insight on positional regulation by these factors (Ule et al. 2005). These methodologies have been applied with regards to TDP-43 and ALS (discussed below) in an effort to understand global changes within cells under physiological and pathological conditions.

Since microarrays were first described (Lipshutz et al. 1999), several adjustments have been made, including the development of splice junction arrays (Yeakley et al. 2002; Johnson et al. 2003), that are now routinely used to analyse global splicing changes within the cell. Splice-sensitive microarrays utilise probes positioned on exons and exon-exon junctions and are able to detect changes in splicing as well as changes in expression of isoforms due to these splice changes (Hallegger et al. 2010). In particular, a study that utilised splice-sensitive microarrays, RNA-seq and CLIP to analyse consequences of TDP-43 depletion in mice, found that transcripts with very long introns and which also encoded proteins involved in synaptic activity, were significantly depleted in mice brain and were significantly enriched for TDP-43 clusters (Polymenidou et al. 2011). In addition, the study also identified over 500 transcripts that underwent TDP-43 dependent alternative splicing based on combined data from RNA-seq and splice-sensitive microarrays (Polymenidou et al. 2011). Several similar studies have been conducted using different disease models, in an effort to determine whether there is a correlation of altered RNA processing in TDP-43 proteinopathies with toxicity and neurodegeneration as observed in patients. One such study examined the effects of TDP-43 depletion and overexpression in Drosophila and concluded that since there was no significant overlap in affected genes, both mechanisms of toxicity employed different cellular programs (Hazelett et al. 2012).

Other studies examining expression changes that were TDP-43 dependent such as Ayala et al (2008), employed microarray analysis in HeLa cells depleted of TDP-43 and found a dramatic increase in expression levels of Cyclin dependent kinase 6 (Cdk6) at both the mRNA and protein level. Cdk6 phosphorylates the retinoblastoma protein (pRb) involved in the control of cellular proliferation, thus hyper-phosphorylation due to depleted TDP-43 could possibly lead to an increased activation of pro-apoptotic pathways (Ayala et al. 2008; Buratti & Baralle 2010). Similarly, two separate studies using expression profiling and splice sensitive arrays to examine cellular changes upon TDP-43 depletion found histone

deacetylase 6 (HDAC6), a protein involved in the formation and degradation of cellular aggregates to be significantly down-regulated (Fiesel et al. 2010) and TDP-43 dependent inclusion of exon 3 in (S6 kinase 1 (S6K1) Aly/REF-like target or Polymerase delta interacting protein 3) (SKAR/POLDIP3) (Fiesel et al. 2012). These studies highlighted a role for TDP-43 in cellular toxicity, through HDAC6 and translation regulation through a much more active isoform of SKAR, thereby broadly affecting cellular proteomic homeostasis which could be extrapolated to disease (Fiesel et al. 2012).

Similar large scale analyses have since been ustilised involving UV cross-linking immunoprecipitation (UV-CLIP) and high throughput sequencing cross-linking immunoprecipitation (HITS-CLIP) techniques. CLIP, analyses involve UV-mediated crosslinking of protein bound to RNA in the cell, after which immunoprecipitation is carried out using a protein-specific antibody. Subsequently, the bound protein and RNA are digested, leaving just the peptide bound to RNA, which can then be sequenced, following amplification with adaptor sequences (Ule et al. 2005); various modifications exist for the downstream processes.

A study by Tollervey et al. (2011) conducted using iCLIP (modified version individual nucleotide resolution) in SH-SY-5Y cells, human embryonic stem cells and brain tissue from healthy controls and patients with FTLD-TDP-43, found TDP-43 clusters in deep intronic regions, long non-coding RNA (lncRNA), intergenic regions as well 3'UTRs of numerous mRNA sequences. In addition, this study confirmed that TDP-43 preferentially bound stretches of UG repeats and reported a position-dependant splicing regulatory role (Tollervey et al. 2011). Concurrently, in addition to finding that TDP-43 preferentially bound UG stretches, Xiao et al. (2011) showed that TDP-43 also bound poly-pyrimidine tracts (although at a much lower frequency than UG-rich sequences), in lumbar spinal cord RNA of controls and ALS patients. In addition, this study also confirmed TDP-43 position-dependent regulation of alternative spliced exons in five out eight genes analysed (Xiao et

al. 2011). A similar independent study using mouse models and iPS cells confirmed the above findings in addition to finding that while FUS/TLS and TDP-43 had distinct mRNA targets, they also shared common targets and functional roles in the processing of distinct sets of lncRNAs (Lagier-Tourenne et al. 2012b). This study also reported a general reduction in mRNA and proteins for transcripts that were bound by both TDP-43 and FUS/TLS in neuronal cells and spinal cords of patients (Lagier-Tourenne et al. 2012b), supporting a role for a loss of function mediated by TDP-43 aggregation.

Other methodologies that have been used to identify RNA targets of TDP-43 include RNA immunoprecipitation-microarray analysis (RIP-Chip) and RIP-sequencing. With this method, RNA bound to a protein of interest is immunoprecipitated using a protein-specific antibody followed by proteinase digestion to separate the protein from RNA (Keene et al. 2006). Once the RNA has been isolated from the ribonucleoprotein complex (RNP), cDNA conversion is perfomed and transcripts are hybridised on a microarray chip (Keene et al. 2006). Using RIP-Chip, performed on mouse neuronal cells (NSC-34) Colombrita et al. (2012) confirmed that TDP-43 and FUS/TLS bind distinct sets of mature RNAs and this was thought to facilitate their involvement in translation and stability. The authors were also able to confirm RNA targets that had previously been identified using other methods, such as *Taf15* and *Atxn1*, and that TDP-43 targets were enriched for transcripts associated with neuron-specific activities (Colombrita et al. 2012). Another study using RIP-analysis performed on mice confirmed that TDP-43 targets were enriched for synaptic function, neuronal development and RNA metabolism, in addition to finding other proteins that were in complex with TDP-43 (Sephton et al. 2011).

Collectively, these studies confirm that TDP-43 preferentially binds to UG stretches within long introns and 3'UTRs of various transcripts with varying consequences. Such global studies provide large sets of data that enable the identification of RNA targets of TDP-43 and in most cases the consequences of TDP-43 binding in these targets.

Effectively, most of the works described in these studies highlight a major role for TDP-43 in alternative splicing regulation, and to a minor extent, roles in transcription and translation regulation. Other, mechanisms of TDP-43 on its targets, if they exist, remain to be elucidated.

Still, some challenges exist, such as limitations presented by mice models that do not necessarily undergo the same splicing events as in humans. In addition, the heterogeneity within cellular models and technological differences could confound results, making it difficult to find consensus targets for TDP-43. Even more importantly, is the fact that despite the millions of data that are obtained from global scale analyses, these often need to be validated using other methods. In the case of TDP-43, since a major role in splicing has been identified, most validation strategies involve the use of RT-PCRs and minigenes, which are well-known and established assays for identifying and characterizing alternative splicing as well as binding sites. The characterization of TDP-43 targets and the mechanisms involved could provide a much needed insight into disease mechanisms including its actual role in proteinopathies associated with neurodegeneration, which to date, have not been clearly defined.

## 1.6. Study Rationale

All the approaches described above have yielded a vast number of genes whose expression/splicing levels become altered following changes in TDP-43 expression/localization. This result is not surprising if we consider the high number of molecular pathways in which TDP-43 seems to be involved, that include mRNA/lncRNA/miRNA processing, mRNA transport/stability, and mRNA translation (Buratti & Baralle 2012). However, this abundance of targets also suggests that not all of these changes can be ascribed to a direct consequence of just TDP-43 action (Romano & Buratti 2013). The reason being that most hnRNPs regulate mRNA splicing processes in a highly co-operative manner (Huelga et al. 2012; Chen et al. 2012; Wang et al. 2013) and it

is possible that many of TDP-43 functional roles will also be dependent on the presence of specific partners. Indeed, TDP-43 is certainly no exception to this situation, at least with regards to the role it plays in splicing regulation (Buratti et al. 2006). To this date, there are still very few endogenous exons/introns that are known to be directly affected by TDP-43 levels (Romano & Buratti 2013). In addition, substantial changes in the endogenous protein production in neuronal cell lines following TDP-43 depletion, has been shown only for SKAR/Poldip3 and the TDP-43 protein itself (Shiga et al. 2012; Fiesel et al. 2012; Bembich et al. 2014). From the point of view of understanding TDP-43 pathology, therefore, this could represent a crucial aspect of future therapeutic approaches.

### 1.6.1. *Aims and Objectives*

In order to gain better insight into the consequences of TDP 43 depletion in the cell a series of overlapping high-throughput screening approaches were utilised; taking advantage of previously constructed HEK-293 stable cell lines that can be induced to express a variety of TDP-43 isoforms (Ayala et al. 2011). As has been described in Ayala et al (2011), each of the stable cell lines were generated using the Flp-In recombinase system which ensures that only a single copy of the integrated transgene is expressed following tetracycline induction. The stable cell lines used in this study are described below;

- HEK-293-Flp-In: This cell line also known as the 'host' cell line, only contained the integrated Flp-In recognition sites and were used as control cells in all analyses.
- HEK 293-TDP-43 wild-type tagged protein: -This cell line inducibly expresses full length TDP-43 cDNA fused to a FLAG tag at the N-terminal. In addition, in these cells, the over-expression of wild type TDP-43 is able to down-regulate endogenous TDP-43 expression as shown in Figure 1-7a below.
- HEK 293-TDP-43-RRM-1-2 Mutant: -In this cell line TDP-43 cDNA containing both RRM-1 (F147L, F149L) and RRM-2 (F229L, F231L) mutations are combined in addition to an N-terminal FLAG. The substitution of Phenylalanine (F) for

Leucine (L) in the RRMs inhibits the RNA binding capacity of TDP-43, which is confirmed by the lack of TDP-43 autoregulation observed in cells expressing this mutant Figure 1-7b following tetracycline induction.

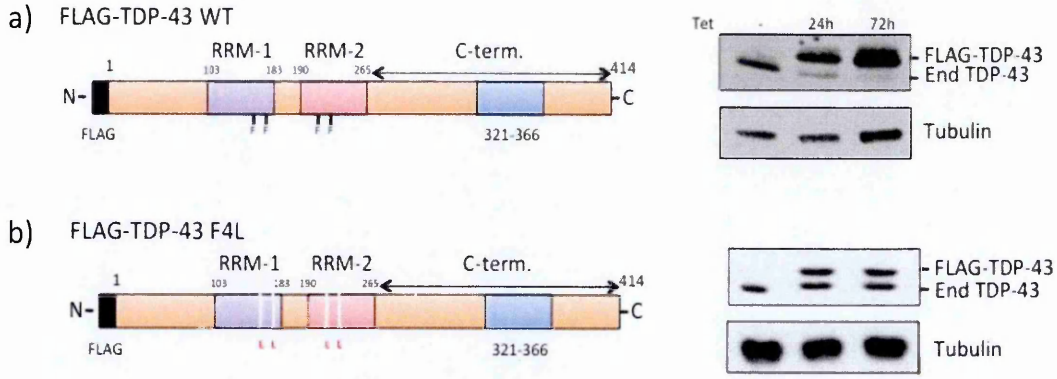



Figure 1-7: **Schematic representation of HEK 293 stable cell lines expressing wild-type and mutant TDP-43.** (a) The HEK 293 stable clone containing full length cDNA of TDP-43 is able to down-regulate expression of endogenous TDP-43 as early as 24 hr, as can be observed in the western blot analyses. (b) In contrast, mutating the Phenylalanine residues in both RRM domains inhibits the ability of TDP-43 to bind RNA, consequently impairing autoregulation as can be observed in the adjacent western blot, where both endogenous TDP-43 and flag-tagged TDP-43 are visible. (Reproduced from Ayala et al., 2011).

Due to the similar background between these lines, this approach has been particularly useful for identifying several novel genes where TDP-43 can be ascribed as playing a direct role in regulating protein isoform production. Therefore, in this study a two-pronged approach that utilised Affymetrix Genesplice arrays (that have previously been described in detail) and 2-Dimensional gel electrophoresis was used to elucidate a common network of TDP-43 RNA targets, in which expression changes could be observed, both at the RNA and protein level.

The use of two-dimensional gel electrophoresis (2-DE) to analyse global differential protein expression, is a popular and versatile method of protein resolution that combines isoelectric focusing (separates proteins according to their isoelectric point (pI)), and SDS-PAGE (separation according to molecular weight) (Garfin 2003). 2-DE separation permits the generation of protein maps of cells under different conditions including protein expression changes associated with the different cellular states.

By utilizing both splice-sensitive and 2-D electrophoresis, the observed global changes at the RNA and protein level will provide further insight into the role of TDP-43, including pathogenic processes that could lead to neurodegeneration.

# 2. MATERIALS AND METHODS

## 2.1. Cell culture and sample preparation for microarray and 2-DE analyses.

For both microarray and 2-DE analyses, Flp-In HEK293 (Invitrogen) with inducible siRNA resistant FLAG-tagged wild type TDP-43, F4L mutant (Ayala et al. 2011) cells were cultured in DMEM-Glutamax-I (GIBCO, Life technologies) supplemented with 10% fetal bovine serum (EuroClone) and 1% penicillin/streptomycin (Gibco) in 35 mm dishes or six well plates and incubated at 37°C and 5% $CO_2$. Induction of tagged TDP-43 expression was achieved with 1 μg/ml tetracycline (Sigma).

### 2.1.1. *RNA interference and RNA extraction*

Knockdown of endogenous TDP-43 was performed by RNA interference using HiPerFect Transfection Reagent (Qiagen) and siRNA specific for TDP-43 (Dharmacon; target sequence 5'-aagcaaagccaagaugagccu-3'). Shortly prior to transfection $5x10^5$ cells were seeded in 6-well plates in 1.4 ml of culture medium containing serum and no antibiotics. 3μl of 40μM siRNA TDP-43 was diluted in 91 μl of Opti-MEM (Gibco, Life Technologies) and 6 μl of HiPerFect Transfection Reagent was added to the diluted siRNA. Following 10 min of incubation, the complexes were added drop-wise onto the cells. After 24 hours, the same procedure was performed with the exception that cells were detached from the 6-well plates, centrifuged at low speed (1,000 x g) to remove previous media containing siRNA transfections mixture, and re-seeded shortly before transfection. Consequently, three siRNA transfections were performed for each experiment. The siRNA against firefly luciferase gene was used as a control (Dharmacon Non-Targeting siRNA #2). Total RNA from individual samples was purified using TRIzol reagent (Invitrogen) according to manufacturer's instructions. RNA quality control and microarray analysis was

performed at GenoSplice (France). A schematic representation of the experimental set-up for both analyses is shown below (Figure 2-1).

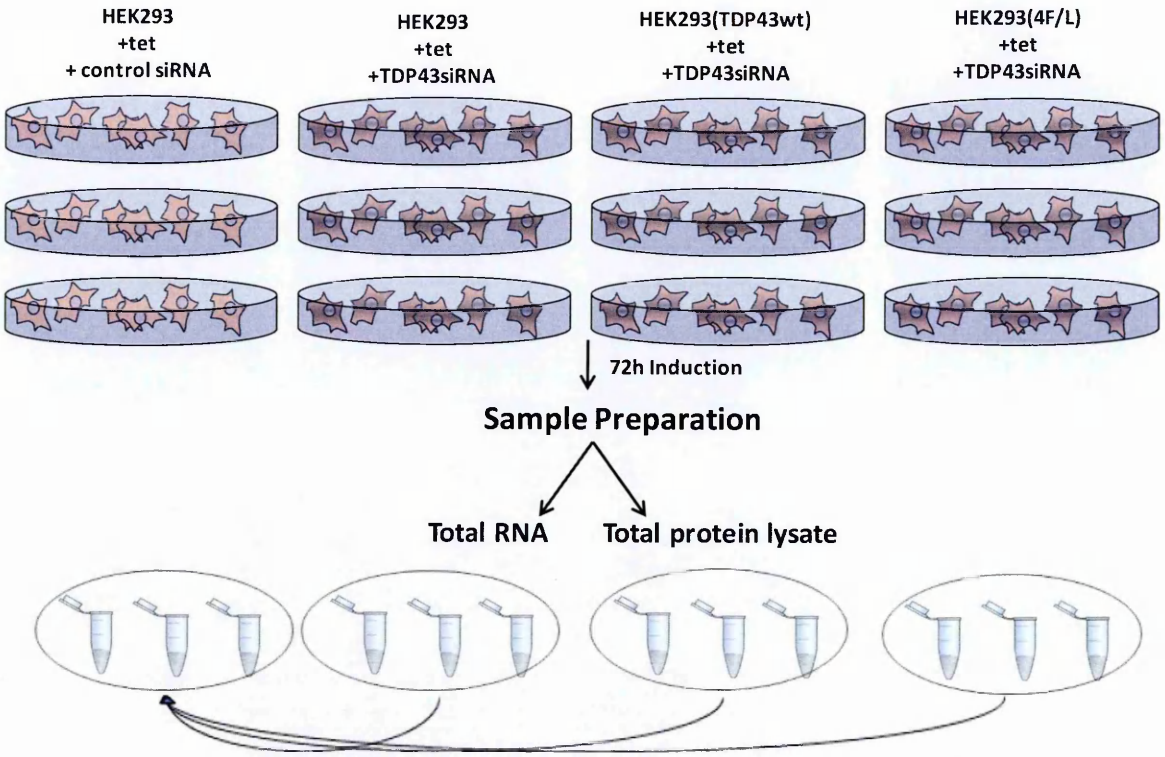



Figure 2-1: Schematic diagram of experimental set-up for both microarray and 2-dimensional gel analyses.

### 2.1.2. *Protein extraction and bi-dimensional separation.*

With regards to the 2-DE gels, the same culturing procedure and siRNA transfection protocol as described above was utilised. Similarly, after 72 hours of induction and siRNA treatment, total protein was harvested from the cells as follows. Protein isolation was performed as has been described previously (Tang et al. 2007) with a few exceptions. Approximately 1 x $10^7$ cells in triplicate estimated to contain 1 mg of protein were initially washed with 9% sucrose in $dH_2O$, followed by another wash containing 1X protease inhibitor cocktail (Roche) to remove salts. Cells were then pelleted by centrifugation at maximum velocity and the resulting pellet in re-supended in destreak rehydration solution (GE Healthcare) supplemented with immobilised pH gradient (IPG) buffer, sonicated, centrifuged and again re-suspended in the same solution in preparation for isoelectric

focusing. Protein samples were first separated according to their isoelectric point on immobilized pH-gradient strips (pH3-11 NL) 1% v/v as follows: proteins were first put onto the ceramic holder and the dry strip placed on top with the gel-side being in direct contact with the proteins. Cover fluid was put on top and left for at least 8 hr. Isoelectric focusing was carried out on an EttanIPGphorR IEF unit (Amersham Biosciences), at 20 °C, using the following program: 50 V for 4 h, 500 V for 1 h, 1000 V for 2 h, and 8000 V up to 48 000 V hr and 5500 V for 29 hr (to avoid diffusion). Each focused strip was dried of excess oil and gel melted at low voltage prior to treatment with equilibration buffer 2.5 mL (6 M urea, 30% glycerol, 50 mM Tris-HCl pH 8.8, 2% w/v SDS, 25 mg DTT, and bromophenol blue (tracking dye)), for 10 min, at room temperature. After removing this solution, a second equilibration was performed for 7 min with 2.5 mL of a solution containing the same components with the exception of 62.5 mg iodoacetamide used instead of DTT. Strips were then washed briefly in 1X running buffer.

Subsequently, the proteins were then separated based on molecular weight by placing the strips horizontally on 12% SDS-PolyAcrylamide gels with 30 µl of standard protein marker (Fermentas) loaded alongside strip and run at 40 mA per gel overnight. After running gels were washed once with $dH_2O$ and fixed for 1 hr in 10% (v/v) ethanol 7% (v/v) Acetic acid and stained overnight with 0.12% Colloidal G-250 Coomassie Blue and de-stained with $dH_2O$.

### 2.1.3. *In-gel digestion and peptide extraction*

Initially, spots identified to undergo differential intensity were excised and sent to Proteome Factory (Germany) for analysis. When performing the re-analysis the spots selected were prepared for analysis by an in-house mass spectrometry facility. Prior to this analysis in-gel digestion of spots identified to undergo differential intensity but had no genes identified, shown in Table 2-1, was performed essentially as described in Bhardwaj et al. (2013) with a few exceptions. All traces of Coomassie blue staining were removed

from the excised spots by washing with different concentration of methanol (MeOH), starting at 100% and then rehydrated with 30% MeOH and dH2O wash. The gel pieces were then subsequently washed 4-6X for 30 min each on a mixer with 50% MeOH containing 20 mM TEAB (TriethylammoniumBicarbonate) then with 1 ml $dH_2O$ and dried on a SpeedVac for 15 min at medium temperature. 10 µl of Trypsin solution (0.5 µg/ml; 98 µl of 20 mM TEAB) was added to the gel pieces, mixed and incubated at 37°C for 15 min followed by addition of 10 µl of $dH_2O$ and further incubation at 37°C overnight. Following trypsinization supernatant from the gel pieces was collected into a clean tube. 5% Formic Acid was then added to the gel pieces (just enough to cover) and sonicated. Supernatants from the trypsinization and sonication were combined, concentrated by drying on the Speed Vacuum and sent for mass spectrometry analysis (Mike Mayers, ICGEB).

## 2.2. Validation analyses

### 2.2.1. *Quantitative real-time PCR (qPCR) Analysis for genes identified in 2-DE analyses*

For all qPCR analyses, RNA extraction and cDNA conversion were performed as described above. qPCR was performed with the iQ SYBER Green Supermix on a CFX96 Real-time PCR Detection System (BioRad). Primer sequences used to amplify genes analysed in the 2-DE analyses are shown in Table 1 below:

| Gene/Primer Name | Forward | Reverse |
|---|---|---|
| *EEF2* | 5'- GCCAGATCATCCCCACAGC-3' | 5'- CCCGCTTCCTGTTCAAAACC-3' |
| *CCT8* | 5'-CCCCAGGTTCTCAGAGCTCAC-3' | 5'-GGATGACACACAAAGCATCG-3' |
| *U2AF1* | 5'-TGCTGCCGTCAGTATGAGATG-3' | 5'-TCTCGACCGCCTCCTGTC -3' |
| *MDH1* | 5'-TTGTTGTGGGTAATCCAG-'3 | 5-GACATTCTTTACATCATTAGC-3' |
| *NASP* | 5'-GAGTCCACAGCCACTGCC-'3 | 5'-GCATTGACAGCTGCTGGA-'3 |
| *HSPA9* | 5'-CAGAAGCAATCAAGGGAGCAG-3' | 5'-CACAACTGAAGGGGTGGTTC-3' |

**Table 2-1: Primer sequences used to amplify genes analysed in both microarray and 2-DE analyses.** Sequences are written in 5'-3' direction. Primers were mapped to exonic sequences and not exon-intron junctions.

PCRs were performed at 98°C for 30 sec, 95°C for 10 sec, Ta (relevant annealing temperature; range 55-57°C) for 30 sec for 40 cycles followed by a thermal denaturation step. Expression levels of target genes were quantified relative to either *GAPDH* or *HPRT* house-keeping genes using the $2^{-\Delta\Delta C_T}$ method. Student's *t*-test was used to determine significant differences between groups where a p-value $< 0.05$ was considered statistically significant.

### 2.2.2. ***Northern Blots***

Northern blot analyses were performed to quantify relative mRNA abundance. Total RNA was extracted from the cells by means of a single-step extraction method using EuroGold TriFast™ (Euroclone, SanBio) reagent, according to the manufacturer's guidelines. Subsequently, for each sample, 10-20 µg of RNA mixed with 1X sample buffer (50 µl of MOPS 10X, 2 µl of EtBr (10 mg/ml), 250 µl of deionised formamide, 37% formaldehyde, bromophenol blue and $dH_2O$) was denatured at 70°C for 5 min, cooled on ice, spun down and loaded on 1.2% formaldehyde Agarose gels and run at 85 V for 4 hr. After the run RNA was transferred onto Hybond $N^+$ nylon membranes (Amersham Biosciences) and UV-cross-linked. A 1 hr pre-hybridization of the membrane was performed in ULTRAhybaid® Ultrasensitive hybridization buffer (Ambion) at 42°C followed by hybridization at either 42/50 depending on the probe. The probes were generated by PCR using primers described in section 4.2.1 and PCR products labelled with Rediprime II DNA Labeling System (GE Healthcare) according to the manufacturer's instructions. Probes were denatured at 95-100 °C for 5 min prior to hybridization. Washes to remove un-hybridized signal were performed for 20 min each as follows: 2X SCC + 0.1% SDS, 1X SSC + 0.1% SDS, 0.5X SSC +0.1% SDS, 0.1 X SSC + 0.1% SDS. Visualization of signal was carried out with a Cyclone Plus Storage Phosphor Scanner and the included OptiQuant Software (Perkin Elmer).

### 2.2.3. *RT-PCRs*

Validation of transcripts depicted to be altered in a TDP-43 dependent manner was performed using the same cell lines, culture conditions and siRNA treatment as described in 4.1 above. Prior to reverse transcriptase PCR (RT-PCR) analyses, total RNA was extracted as in section 4.2.2 and converted to cDNA as follows: 1 μg of RNA together with 2 μl (100 ng/μl) of random hexameric primers (Promega) were denatured at 70°C and snap cooled on ice, after which, transcription mixture containing 1X First strand buffer (Invitrogen) 10 mM DTT, 1 mM dNTPs, and 1 μl (100 U/ μl) Moloney murine leukaemia virus (MMLV) (Invitrogen) reverse transcriptase added in a final volume of 20 μl. The reaction mixture was incubated at 37°C for one hour. 4 μl of cDNA was used in 50 μl PCR. For each gene, primers used in the endogenous RT-PCR assay are as shown in Table 2-2 below.

| GENE/ PRIMER | FORWARD | REVERSE |
|---|---|---|
| *BIM/BCL-2* | 5'-TCTGAGTGTGACCGAGAAGG-3' | 5'-TCTTGGGCGATCCATATCTC3' |
| *SKAR/POLDIP3* | 5'-GCTTAATGCCAGACCGGGAGTTG-3' | 5'-TCATCTTCATCCAGGTCATATAAATT-3' |
| *STAG2* | 5'-GTATGTTTACTTGGAAAAGTTCATG-3' | 5'-TGATTCATCCATAATTGAAGCTGGA-3' |
| *MADD* | 5'-GACCTGAATTGGGTGGCGAGTTCCCT-3' | 5'-CATTGGTGTCTTGTACTTGTGGCTC-3' |
| *FNIP1* | 5'-GCTACAAGATAGTCTTGAATTCATC-3' | 5'-CAGACCGTGCTATGCCACTGTCTCT-3' |
| *BRD8* | 5'-CAATTCTTGGCCACGCAGTTGATTA-3' | 5'-CTCAGAGAGAAAGTGGAGGAGGTTC-3' |
| *ERGIC3* | 5'-ATGGAGGCGCTGGGGAAGCTGAA-3' | 5-'CATGCCAGCTCAGCCTCTGAGCTC-3' |
| *CASK* | 5'-AGGGAAATGCGGGGGAGTATTAC-3' | 5'-CATTCCTCAAGTTCAGGAGAAGG-3' |
| *CRAMP1L* | 5'-AGAACACTGCTCCCTAGACCATCG-3' | 5'-CCCCACGGTGGGAGGGTATCTC-3' |
| *ZNF207* | 5'- GTATGCCCCCACCTGTTCCACGT-3' | 5'-CAGGGATATATCCTGATCTGGATGGA-3' |
| *BID* | 5'-GTCAACAACGGTTCCAGCCTCAGGA-3' | 5'-TCACCAGGCCCGGAGGGATGCTACG-3' |
| *NLGN2* | 5'-ATGTGACTCCTGGCGCTGTGTCT-3' | 5'-GCCTGGTCCCCCGTGCTGAGAAA-3' |
| *ALFY* | 5'-CGTCAGCAGAATGCCCTGAAGTA-3' | 5'-GTTTGACTTTAGCAATAGTATCT-3' |
| *TBX19* | 5'-ACTAATGAGATGATTGTGACCAA-3' | 5'-GCTGATCGTCCACAGGCACGCGT-3' |
| *FBOX018* | 5'-CTCGGGCTCCAGGTCCCGGCCAG-3' | 5'-GCCGAGACGTACTGTCCTGTCGT-3' |
| *ANKRD12* | 5'-AATCGCGGGGCGACGCTGTCCTG-3' | 5'-CTTTCTTCCATAGGTTTCTCTA-3' |
| *GANAB* | 5'-TTGCTGGTGCTAGAGCTTCAGGG-3' | 5'-CCAAATTGTAGAGCCGATATGGC-3' |
| *C20ORF24* | 5'-ATGAGCGGGGCGGCGGCGGAAGGA-3' | 5'-CAGTGTAAAAGATGATCCAAATG-3' |
| *C14ORF18* | 5'-GAGTAAAATTTCAAGAGACTGAA-3' | 5'-TCCAAAGGAAACTTGGAGTTGTTAC-3' |

**Table 2-2: Primer sequences used in endogenouse RT-PCR assays.**

RT-PCR analyses of the hybrid minigenes used in this study were similarly performed by conversion of total RNA extract in into cDNA as described above. For pTB based minigenes, cDNA was amplified in a 50 µl PCR using the *ALFA* Fwd 2–3α 5'-CAACTTCAAGCTCCTAAGCCACTGC-3' and *BRA*-2 Rev 5'-TAGGATCCGGTCACCAGGAAGTTGGTTAAATCA-'3 reverse primers. In the case of the pcDNA3 minigene, the universal primers T7 forward and SP6 reverse (Sigma-Aldrich) were used for amplification.

Except where the annealing temperatures have been specified (Table 4-2), PCR cycling conditions used for all RT-PCR assays were as follows; 94°C for 2 min, 35 cycles of 94°C for 30 sec, 50°C (+/-1) for 35 sec, 72°C for 40 sec and a final extension of 72°C for 7 min. Subsequently, PCR products were analysed on a 2% Agarose gel stained with EtBr (Sigma-Aldrich).

### 2.2.4. *Minigene constructs*

In order to determine the exact binding sites of TDP-43 in the alternatively spliced exons, minigenes were used. Minigenes are well-known assays used to study splicing and incorporate the use of an expression plasmid containing a global promoter and poly-A site (Cooper 2005). The presence of restriction endonuclease sites allow for the introduction of exogenous genomic sequence, which can be further analysed to determine intronic or exonic elements that bind *trans*-acting factors.

Mini-gene constructs for both *MADD* and *STAG2* genes were prepared by amplifying the relevant skipped/included exons as well as a few 100 base pairs of flanking intronic sequences. Specifically for the *MADD* mini-gene, exon 31 together with 278bp and 266 bp of 5' and 3' flanking intronic sequences were amplified using forward primer 5'-ccatatggAGGCCAGCAGCAGGGCCTTCGG-3' and reverse primers 5'-ccatatggTTCACTGAGCTGCTTCAGGACC-3'. Both primers contained NdeI restriction sites at their 5' ends that were used to clone the 609 bp amplified fragment into the pTB expression plasmid which is a modified version of the α-globin-fibronectin mini-gene (Baralle et al, 2003; Pagani et al, 2000). Unless otherwise stated, all subsequent primer sequences highlighted with bold letters indicate restriction sites for the primer specific enzymes. For the *BRD8* and *FNIP1* minigenes, exon 20 and exon 7 with 250 bp flanking intronic sequences respectively, were cloned into the pTB vector as described above. Primers used to amplify the relevant regions in *BRD8* and *FNIP1* were as follows; *BRD8* Intron 19 NdeI forward 5'-GGAATTC**CATATG**AATTATTAATTCCTCTGGGA-3',

*BRD8* Intron 20 NdeI reverse 5'-GGAATTC**CATATG**AGATCCTTCACCCCTAGAAT-3' and *FNIP1* Intron 6 NdeI forward 5'-GGAATTC**CATATG**ATTACATATAGATTTATTAG-3', *FNIP1* Intron 7 NdeI reverse 5'-GGAATTC**CATATG**TCATTATGCCACAAGAGAAATG-3'.

For *STAG2*, initial minigenes were constructed in the pTB minigene using the following primers: *STAG2* Intron 30-NdeI Fwd 5'-C**CATATG**GTTGATCATTTTCTGTACTATA-3' and 5'-C**CATATG**GTATTTAAACTTTTCCAACAGAT-3' reverse primers were used to amplify the 540 bp fragment containing exon 30b that was then cloned into the pTB mini-gene. No change in splicing profile was observed for STAG2 with this minigene therefore, a minor modification was performed to widen the context by including more sequences upstream of exon 30b which included exon 30 and 223 bp of upstream intronic sequence. In this modified minigene, only the forward primer was altered to *STAG2* Intron 29 Fwd 5'-C**CATATG**GTCTCCTTGCATTTACCACACTG-3' whereas the reverse primer remained the same. Similarly, no splicing profile changes were obtained with the above mini-gene and therefore, a three-exon mini-gene construct in pCDNA3 was constructed for *STAG2* using the following primers; *STAG2* Ex 30-KpnI-Fwd 5'-CGG**GGTACC**CTGAAGAAAGTAGTAGTAGTGACAG-'3 and *STAG2*-Intron 30b-NotI Rev 5'-ATAAGAAT**GCGGCCGC**CAATGGGGAGCCACAGA-3', was used to amplify the first fragment of 898bp (exon 30, exon 30b and 508bp downstream intronic sequence ) whereas the second set; *STAG2* Intron 30b-NotI-Fwd 5'-ATAAGAAT**GCGGCCG**CCTCTGTTTGCCATGGTAGAC-3' and *STAG2* Ex 31-ApaI Rev 5'-ATATAT**GGGCCC**CAAATCTATATCCATGGTGTCAAAATCC-3' was used to amplify the second fragment of 737 bp (505 bp of upstream intronic sequence and exon 31). The two fragments were sequentially cloned into the pcDNA3 vector (Invitrogen) and all constructs confirmed by standard Sanger sequencing (Macrogen, Netherlands) prior to large scale plasmid preparations and transfections.

### 2.2.5. *Transfections*

HEK 293 cells were grown in 35 mm dishes in standard Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and Pen/Strep and incubated at 37°C+ 5% $CO_2$. Prior to transfections cells were approximately 50-60% confluent. Transfections of all minigene constructs were performed using a modified calcium phosphate method (Kingston et al. 2001). For each transfection, approximately 3 μg plasmid DNA was mixed with 2M $CaCl_2$ and water added to make up the volume to 100 μl. The DNA/$CaCl_2$ mixture was then added to another microfuge tube containing equal volume (100 μl) 2X Hepes Buffered Saline (HBS) in a drop-wise manner while mixing the HBS vigorously with a pipette (to create bubbles) to facilitate calcium phosphate complex formation. The mixture is then added to the cell slowly in a drop-wise manner. In cases where cells were required to be silenced prior to transfection with minigenes, cells were washed gently with 1X warm PBS to remove siRNA mixture and fresh media added then minigenes were transfected on the evening of the third day (56hrs after siRNA transfection) after which cells were incubated for another 16-20 hr.

### 2.2.6. *Western Blots and Protein extraction*

Cells were pelleted by centrifugation at 1000 x g and total protein extracted in cell lysis buffer containing 15 mM Hepes (pH 7.5), 250 mM NaCl, 0.5% NP-40, 10% Glycerol, and 1X complete protease Inhibitor. The cell lysate was then sonicated for 10 min on high and centrifuged at 10,000 x g to remove cellular debris. The resulting supernatant was quantified using Bradford (Thermoscientific, Pierce). Approximately 30 μg of protein mixed with 2X sample buffer were loaded and ran on 10% SDS-PAGE with 1X running buffer (50 mM Tris, 0.38 M Glycine, 0.1% w/v SDS) at 25 mA. After electrophoresis, the proteins were transferred (wet) onto nitrocellulose membrane (Amersham, BioSciences) in 1X transfer buffer at 200 V for 1.5 hr. Antibodies used for protein detection in this study included anti-TDP-43 (Buratti et al, 2001), Tubulin (Sigma) anti-MADD (Abcam), anti-

MDH1 (Abcam), anti-NAP1L1 (Abcam) and anti-STAG2 (Cellular signalling technologies). Detection was conducted according to standard western blotting procedure using either secondary anti-rabbit or anti-mouse HRP conjugated antibodies and ECL (Amersham) for developing blots.

### 2.2.7. *Electrophoretic-mobility shift Assays (EMSA)*

A region of intronic *MADD* sequence, approximately 64 bp upstream of exon 31 (skipped exon) containing a stretch of TG-repeats (5'-GTGTGCTGTGT-3') was chosen for band shift analysis following the observation that the minigene splicing profile was similar to that of the endogenous transcript. The DNA oligo was synthesised by Sigma, Life Science and radio-labelled using [$\gamma$-$^{32}$P] ATP and T4 polynucleotide kinase (New England Biolabs). Briefly, 5µl of DNA (100 ng/ µl) oligo, 1X PNK buffer, 1 µl (10U/ µl) and 1 µl [$\gamma$-$^{32}$P] ATP (1000 µCi/ µl) were incubated for 1 hour at 37°C. The labelled oligo was then precipitated in three volumes ethanol and 3M NaAc at pH 5.2 for an hour on dry ice and subsequently washed with 70% ethanol. Following centrifugation, the air-dried pellet was re-suspended in 50 µl $dH_2O$ and 1µl used in the binding reaction.

For the *STAG2* 3-exon minigene EMSA analysis, the cloned region was first divided into five separate fragments by PCR amplification with primers that contained a T7 promoter sequence (highlighted in bold on primer sequence) on the forward strand for each pair. Primers used in the amplification of each fragment were as follows; *STAG2* Fragment 1: forward 5'-**TACGTAATACGACTCACTATAGG**CACGCAGGTAACATGGATGTTA-3' and reverse 5'-ATGGCATGCTGA-3', Fragment 2: forward 5'-**TACGTAATACGACTCACTATAGG**GTAAGTGAGAGTGCCTTATT-'3 and reverse 5'-AAGCTAATACAATA-3', Fragment 3: forward 5'-**TACGTAATACGACTCACTATAGG**TCTAACTGGTTTTCTTCCCTCAA-3' and reverse 5'-GTGTACCAGGCATG-3', Fragment 4: forward 5'-

**TACGTAATACGACTCACTATAGG**CTGTCACGTAGTAGGCATTGTGTGAGTGA GTGCGCGCA-3' and reverse 5'-GTGTACCAGGCATGCGCGCACT-3', fragment 5: forward 5'-**TACGTAATACGACTCACTATAGG**ATTAGGTACTGAATGAATGA-3' and reverse 5'- GAATTAAAGGTCAG-3'. PCR products obtained were then gel purified and used as templates for *in vitro* T7 (Stratagene) RNA transcription and labelling with [$\gamma$-$^{32}$P] UTP (800Ci/mol) (Perkin Elmer). Transcribed RNA was treated with DNAse I (Roche) according to the manufacturer's guidelines and purified on Nick columns (Amersham Pharmacia Biotech) according to the manufacturer's instructions. The labelled RNAs were then ethanol precipitated as described previously and re-suspended in RNAse-free water.

Binding reactions containing labelled DNA oligos (*MADD*) or labelled RNA probes (*STAG2*) together with purified recombinant TDP-43 protein (300 ng) were performed in 1X binding buffer (10 mM $NaCl_2$, 10 mM Tris pH 8.0, 2 mM $MgCl_2$, 5% Glycerol and 1 mM DTT) for 10-20 min at room temperature prior to electrophoresis on a 6% Polyacrylamide native gel at 100 V for 1.5 hours in 0.5X Tris borate/EDTA (TBE) buffer at 4°C. In the case of *STAG2* EMSA analyses where transcribed RNA was used in the binding assay, RNAse inhibitor (Roche) was also added to the reaction. A pre-run of the gel (approximately 10-20 min) was performed before samples were loaded. Following electrophoresis, the gels were then dried on 3 MM Whatmann paper and exposed on a Cyclone™ Phosphor screen (Packard). In addition, cold competition binding analyses against a known positive binder ($UG_6$) of TDP-43 were performed using 20-fold molar excess amounts of un-labelled *MADD* oligo, and un-labelled *in vitro* transcribed RNA for *STAG2*. Cold competitors were added 5 min prior to addition of labelled oligos or RNAs and band shifts were analysed by electrophoresis under conditions described above.

#### *2.2.7.1. Recombinant GST-TDP-43 purification*

Purified recombinant TDP-43 expressed in BL21 bacterial cells was obtained using Glutathione S-Sepharose 4B beads (Pharmacia) elution. Briefly, 5 ml of previously transformed BL21 cells were inoculated into 100 ml LB with Ampicillin and left to grow for approximately 4 hours in a shaker. The culture was then induced with 1M IPTG (Isopropyl β-D-1-thiogalactopyranoside) and left to grow for another 4 hours. The cells were then harvested by pelleting at 4,000 rpm for 20 min and pellet re-suspended in lysis buffer (1X PBS and 0.01% Triton X-100) and sonicated. The supernatant following centrifugation was mixed with 0.5 ml of resin or slurry and incubated on shaker at 4°C for 1 hr 30 min. Resin washes were performed in lysis buffer and subsequent elution in reduced L-Glutathione (Sigma) according to the manufacturer's guidelines.

#### *2.2.7.2. RNA In-vitro transcription*

Fragments used as templates for in vitro transcription were obtained by PCR amplification with primers containing T7 promoter sequence at the 5' end to facilitate T7 polymerase transcription as described in section 4.2.7 above. Gel extracted PCR fragments were quantified and used as templates in the *in vitro* transcription reaction as follows: 1 µg of DNA, 4 µl of transcription buffer (Stratagene) NTP mix (15 mM each of ATP, CTP, GTP and 1.5 mM UTP), 100 mM DTT (Dithithreitol) and 0.5 µl T7 (50 U/µl) RNA polymerase were mixed briefly in a microfuge before adding 2 µl of [alpha-$^{32}$P] UTP (1000 µCi/µl) (Perkin Elmer) in a final volume of 20.5 µl and incubated at 37°C for 2 hr. A DNAse I (Roche) digestion of DNA template was performed for 20 min prior to Nick Column purification and ethanol precipitation of transcribed RNA. In cases where 'cold' (un-labelled) RNA was transcribed, 5 mM UTP was added to the reaction instead of the radioactive isotope.

### 2.2.8. *Mutagenesis and deletion constructs*

Mutagenesis of the *MADD* intronic sequence (5'-GTGTGCTGTGT-3') was performed using the Quik-change® Site-Directed mutagenesis kit according to the manufacturer's guidelines. Two complimentary mutagenesis primers; *MADD* MUT Fwd: 5'-gggtggggctgtagctgggagaACAtAcCgCgaggggcaggggtggagcctgtgggc-3' and *MADD* MUT Rev: 5'-gcccacaggctccacccctgcccctcGcGgTaTGTtctcccagctacagccccaccc-3', (mutated nucleotides in capital letters) were used in the mutagenesis PCR and mutated sequences confirmed by sanger sequencing. Similarly for *STAG2*, deletion of fragments three and five which were shown to bind to TDP-43 in the EMSA analysis was achieved by a deletion PCR using cycling conditions outlined in the Quik-change® manufacturer's guidelines. Primers used for the deletion PCRs consisted of *STAG2* delta n3 forward 5'-CTATTGTATTAGCTTCATCATTTTCCATT-3' and reverse 5'-AATGGAAAATGATGAAGCTAATACAATAG-3' and *STAG2* delta n5 forward 5'-TGAGTGAGTGCGCGCATGCCTGGTACACCATCATTTTCCATTATACTTGAATATAG-3' reverse 5'-CTATATTCAAAGTATAATGGAAAATGATGGTGTACCAGGCATGCGCGCACTCACTCA-3'. Subsequently, DpnI digestion at 37°C was performed for all PCR products prior to transformation in DH5α *E-Coli* cells.

## 2.3. Immunoflourescence in the TDP-43 cellular aggregation model

Construction of the TDP43-12XQ/N aggregation model has been described in detail in Budini et al. (2014). For immunofluorescence, cells were seeded onto 22x22 mm, glass cover slips in 35 mm dishes induced with tetracycline for 72 hr, fixed with 4% PFA, blocked with 2% BSA/PBS and incubated with primary antibodies. The primary antibodies used were anti-Flag (Sigma, F1804) and anti-TDP-43 (Protein Tech, 10782-2-AP). Secondary antibodies: anti-mouse-AlexaFluor 594 (cat. A21203), anti-rabbit-AlexaFluor

488 (cat. A21200) and TO-PRO3 dye (cat. T3605) were all purchased from Life Technologies. Cells were analyzed on a Zeiss LSM510 Metaconfocal microscope.

## 2.4. General procedures

### 2.4.1. *Cell Culture Maintenance*

All cell lines were plated in 10 cm culture dishes at concentrations of 6.0 x $10^5$ and maintained in DMEM (Gibco) supplemented with 10% heat inactivated FBS, 1% Pen/Strep in a 37°C with 5% $CO_2$ atmosphere. Cells were passaged every 3 days at 80-100% confluence. Experimental culture plates were seeded at concentrations of 5 x $10^5$ in 35 mm dishes or six-well culture plates.

### 2.4.2. *Agarose Gel Electrophoresis*

Analysis of PCR products and DNA was performed on either 1-2% Agarose gels stained with EtBr (0.5µg/ml) in 1X TBE Buffer at 100 V. To estimate sizes a DNA ladder marker (Invitrogen) was loaded with each gel run alongside samples.The duration of the run was dependent on the type of analysis and length of fragments. DNA was visualised in a UV transilluminator and either printed or stored digitally.

#### 2.4.2.1. *Gel extraction*

Gel extraction was performed following generation of inserts with PCR for sub-cloning purposes or in preparation for RNA transcription. DNA samples were electrophoresed in 1% Agarose stained with EtBr at 100 V. Following visualization with UV, the desired bands were excised from the gel and purified using the Eurogold gel extraction kit (Euroclone). Briefly, 600 µl of binding buffer (1 g/ml) was added to gel slices in a 1.5 ml microfuge tube and incubated at 55°C for 10 min with vortexing every 2 min. The mixture with the dissolved gel was then loaded onto the column and centrifuged at maximum speed for 1 min. Flow-through was discarded and column washed 2X with 700 µl wash buffer, with the flow-through discarded each time. Elution of DNA was performed using 30 µl

elution buffer and column centrifuged at maximum speed for 2 min. Recovered DNA was quantified by approximating the minimum amount of EtBr intercalated DNA multiplied by the number or µl loaded onto the gel.

### 2.4.3. ***Cloning***

#### 2.4.3.1. *Competent cells*

Competent DH5α *Escherischia Coli* (*E.coli*) cells were prepared according to the Calcium Chloride ($CaCl_2$) method described in Sambrook, Fritsch and Manniatis (1989). Briefly, a single colony of bacteria was inoculated in 5 ml Luria Broth (LB) and incubated in a 37°C shaker overnight. A 1/100 dilution, i.e. 1 ml of the of the overnight culture was inoculated in 100 ml LB using aseptic technique and left to grow for 2-3 hr in a 37°C shaker until mid-log phase i.e. optical density of LB at 600 nm was between 0.5 and 0.8. Optical density was determined using a spectrophotometer. The 100 ml LB containing bacteria was then poured into two separate 50 ml tubes and pelleted at 3,000 x g for 10 min at room temperature. The pellet was then re-suspended using 1 ml ice cold 100 mM $CaCl_2$ and an extra 30 ml added before a second spin at 3,000 rpm for 10 min at 4°C. The pellet produced was re-suspended again in 1 ml $CaCl_2$ and 50% glycerol and stored at -80°C in 200 µl aliquots.

#### 2.4.3.2. *Klenow-Kinase Reactions*

Klenow-kinase reactions were used to prepare PCR products for blunt-ended ligation into the pUC vector. The Klenow fragment, a proteolytic product of *E.Coli* DNA Polyemerase I was used for removal of nucleotide overhangs introduced by the Taq Polymerase following PCR, whereas T4 Polynucleotide kinase was used to phosphorylate the 5'ends of the PCR products. First, PCR products were denatured at 98°C for 2 min, then 5 mM $MgCl_2$ and Klenow fragment (2.5 U), 1 µl (5 mM) dNTPs were added and the mixture incubated at 37°C for precisely 15 min. Subsequently, a mixture containing EDTA (0.2 mM), ATP (1

mM), T4 polynucleotide kinase (10 U), and kinase buffer were added to the previous mixture, which was then incubated for a further 30 min at 37°C. Heat inactivation of enzymes was conducted at 65°C for 20 min.

#### 2.4.3.3. *Ligation reactions*

Generation of recombinant plasmids was performed using T4 DNA ligase (Roche) in a ligation reaction as follows: 1 µl of vector (20 ng), X (calculated amount of insert), 1X ligase buffer (Roche), 1 U of T4 DNA ligase and $dH_2O$ in a final volume of 30 µl incubated at 25°C or room temperature. In most cases, 15 µl of the ligation reaction was used to transform cells. The vector to insert ratio was calculated as 5:1 (insert/vector). T4 DNA ligase was used for both sticky and blunt end ligations. In some cases, vectors were treated with Calf-intestinal alkaline phosphatise (CIP) to reduce the number of false positives due to plasmid recirculation.

#### 2.4.3.4. *Bacteria transformation*

Previously frozen competent cells were thawed quickly and placed on ice. Plasmid DNA was added to 60 µl of competent cells in a 1.5 ml microfuge tube and incubated on ice for 1 hr. Following the incubation, cells were heat shocked at 42°C for 90 sec and snap cooled on ice for another 90 sec. Antibiotic-free LB medium (60 µl) was subsequently added to the cells and incubated in a 37°C shaker for 30 min-1 hr. Subsequently, 200 µl of transfromed cells were plated in pre-warmed selective agar plates containing Ampicillin (50 µg/ml) and incubated overnight at 37°C.

#### 2.4.3.5. *Small-scale and large-scale plasmid preparations*

In general, small (mini-prep) preparations were used to purify small volumes of up to 20 µg of high-copy plasmid DNA in volumes ranging between 50-100 µl, whereas higher concentrations and large volumes of plasmid DNA were purified using to large-scale

preparation methods. Mini-preps, were performed using the he Wizard® **Plus** SV miniprep kit according to the manufacturer's guidelines (Promega) whereas large-scale (midi-preps) preparations were performed using the JETSTAR Plasmid Midi Kit (Genomed). For the midipreps inoculation of bacteria cells was performed in 50 ml LB with antibiotic. Purified plasmids obtained from the above preparations were either used in a restriction endocnuclease digest (to verify presence of insert) or in transient transfections (Midipreps) of cells.

#### 2.4.3.6. *Restriction enzymatic analysis*

In this study, restriction digests were used for recombinant plasmid analysis. In each case the appropriate enzyme and its specific buffer were selected. The incubation temperatures for each enzyme as given by the manufacturer were adhered to including the use of additives such as BSA. In all cases 100 ng of plasmid DNA was digested and in a 20 µl reaction volume. Enzymes were obtained from New England Biolabs.

### 2.4.4. *General Reagents and Chemicals*

-PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO4$, 1.8 mM $KH_2PO_4$, pH 7.4

-10X TBE: 108 g/l Tris , 55 g/l Boric Acid, 9.5g/l EDTA

-5X Loading dye: 100g Sucrose, 48g Urea, 100mL TBE, 1% Bromophenol blue in 200 mL final volume

-10X Transfer buffer: 30g Tris, 144g Glycine, ; 1X= 1/10 Dilution + 20 mL Methanol in 1 L final volume

-4X SDS Protein loading buffer: 0.2 M Tris pH 6.8, 40% Glycerol, 0.8% Beta-mercaptoethanol, 0.4% Bromophenol blue, 8% SDS in a final volume of 10 mL

-20X SSC (pH 7): 175.3 g NaCl, 88.2 g Tris Sodium Citrate in final volume of 1 L

-10X MOPS: 0.2 M MOPS (3-(N-morpholino)propanesulfonic acid), 0.01 M EDTA pH 8.0, 0.05 M Sodium Acetate (NaAc).

-Colloidal: 600 mL dH2O, 100 g $NH_4SO_4$, 70 mL Phosphoric Acid,1.2 g Coomassie Brilliant blue G-250 in a final volume of 1 L.

# 3. RESULTS

## 3.1. Analyses and validation of TDP-43 dependent differential protein expression using 2-Dimensional gel electrophoreses.

High throughput analyses of cellular processes have gained considerable momentum in the advent of the 'omics' era, enabling a more global view of changes within cells under various conditions. An example of such a method that offers a global analytical view of changes in protein expression under different conditions is 2-Dimensional (2-DE) gel electrophoresis. In this study, 2-DE was utilised to determine global protein expression changes that were dependent on the levels of expression of TDP-43, including its RNA binding capacity. The exact set-up of the experiments for 2-DE analyses has been described in section 2.1 of materials and methods. The cells used for this analysis consisted of previously constructed HEK 293 stable cell lines (described in introduction section 1.6.1), which could inducibly express wild-type or mutant transgenic TDP-43. Four triplicate groups of cells representing different cellular conditions were analysed. Group (A) HEK 293 Flp-In T-Rex cells expressing physiological levels of TDP-43 (siLuciferase treatment), group (B) HEK 293 Flp-In T-Rex cells depleted of endogenous TDP-43 (siTDP-43 treatment), Group (C) HEK 293 Flp-In-T-Rex cells depleted of endogenous TDP-43 with simultaneous tetracycline-induced expression of wild type FLAG-TDP-43, and lastly, group (D) HEK 293 FlpIn-T Rex cells depleted of endogenous TDP-43 with simultaneous tetracycline-induced expression of mutant (F4L: 4 Phenylalanines mutated) FLAG-TDP-43, that is unable to bind RNA. The western blot confirmation of the various levels TDP-43 can be seen in Figure 3-1.

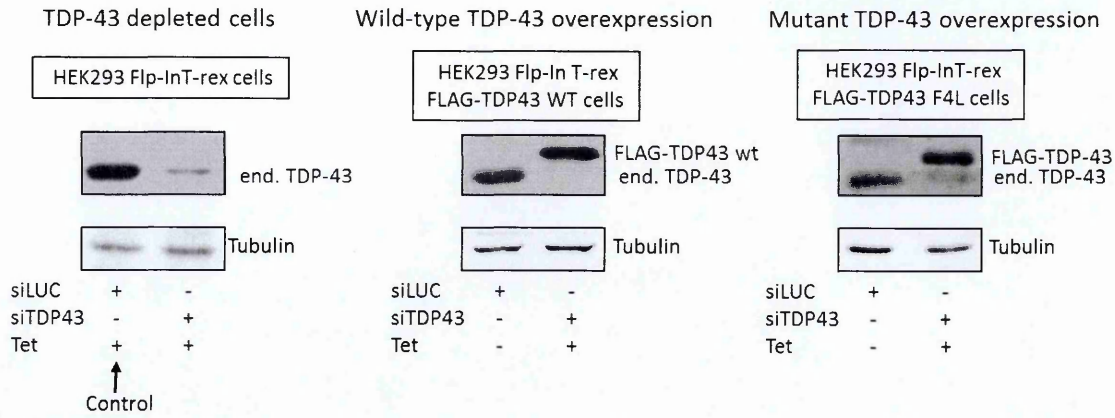



Figure 3-1: **Western blot analyses of the various levels of TDP-43 in HEK-293 stable cell lines used in both 2-DE and splice-sensitive microarrays.** Antibody against TDP-43 was used to detect both the endogenous and transgenic TDP-43 expression whereas tubulin was used as a loading control. In addition, for a uniform background, tetracycline was added to cells treated with control (siluciferase) and siTDP-43. Similarly, tetracycline inducible cell lines expressing wild-type and mutant TDP-43 were also treated with siTDP-43.

Additionally, tetracycline was added to all four groups of cells and the inducible cells lines were also treated with siTDP-43 for uniformity. Proteins extracted from these four groups of cells, were initially separated based on their isoelectric points, followed by molecular weight separation using 10% SDS-PAGE. For ease of reference, these four groups cells will be referred to as A, B, C and D in subsequent descriptions. Analysis of differential spot intensity resulting from the various cellular conditions was analysed and normalised automatically using algorithms from the REDFIN software from LUDESI (http://www.ludesi.com). Briefly, the normalisations were performed by matching pairs of spots (match ratio normalisation method) between a base image/gel (e.g. control luciferase gel) and study gel after which the spot-volume ratio was calculated. Individual spot volumes for each gel are then multiplied by the spot volume ratio (also known as the normalisation factor). Alternatively, the software calculates the ratio between the sum of the volumes of all spots in the base gel and the study gel; each spot volume is then multiplied by this ratio volume. Representative images of the gels obtained from the 2-DE

analysis are shown below (Figure 3-2). Specifically, differences in spot intensities were compared between the three groups (B, C, and D) and the control (A), and the most variable spots selected and sent for mass spectrometric analysis. Peptides were identified using nanoLC-ESI-MS/MS by the Proteome Factory (AG, Berlin, Germany).

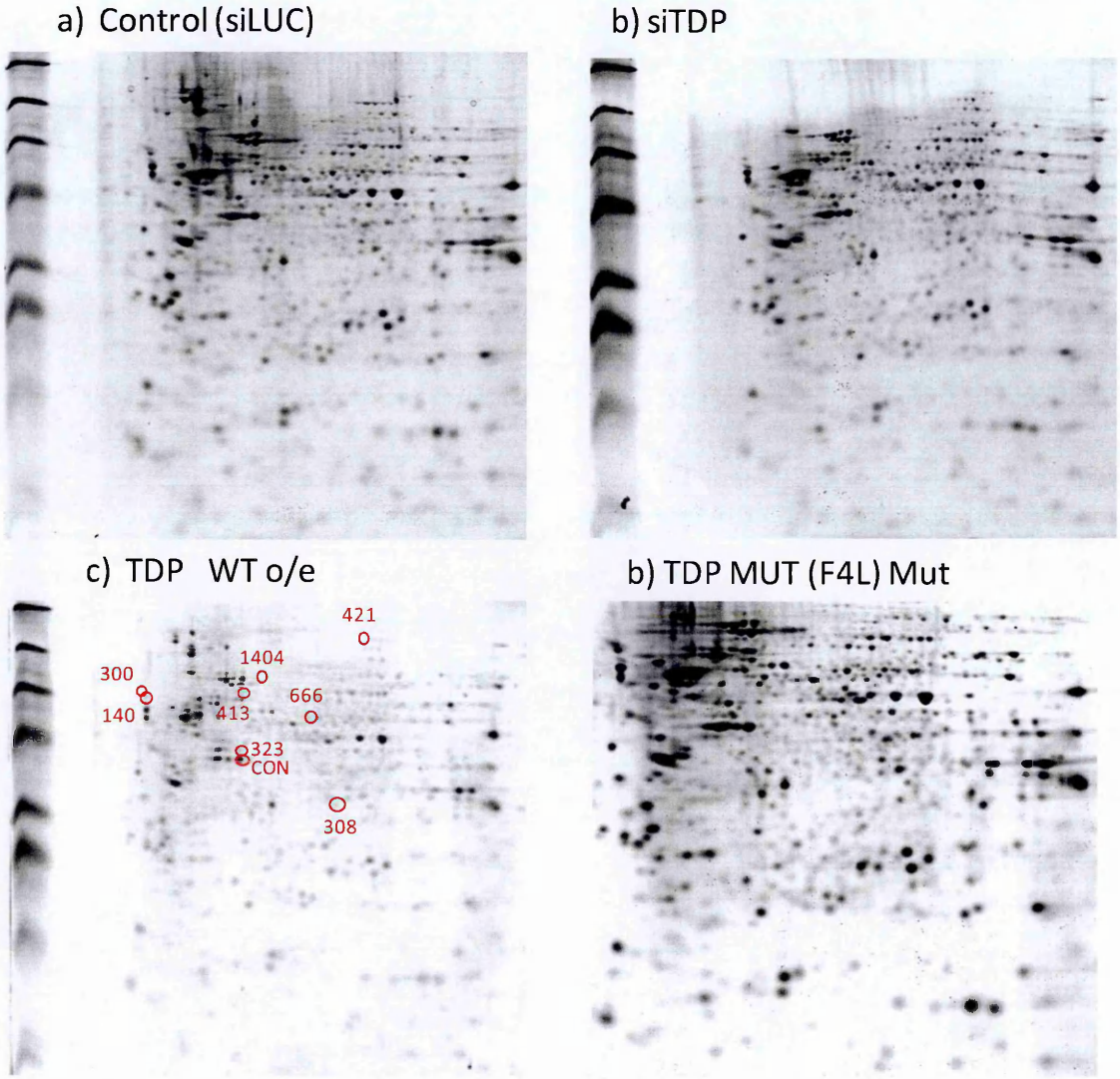



**Figure 3-2: (a-d) Representative 2-DE gel images of the relative conditions of TDP-43 levels that were analysed.** Red circles represent the most variable spots selected by the software and their identification number. The spot marked CON represents an overlapping spot to the 323, which was analysed to ensure that there were no overlapping between the two spots.

### 3.1.1. *Spot validation: Comparative analyses of transcripts and spot intensities*

Comparative analyses of differential intensity amongst spots under the various cellular conditions described previously, identified the most variable spots compared to the control (group A) to be 140, 1404, 300, 308, 323, 413, 421 and 666. Mass spectrometry analysis identified hits (peptides) within each protein spot that corresponded to either one or more genes (Table 2-1) and which were subsequently selected for validation based on their perceived relevance to TDP-43 related processes, mascot score and proximity to observed molecular weights and isoelectric points. This included genes involved in RNA metabolism or cellular survival/apoptotic processes. Interestingly, within the spots sent for mass spectrometric analysis, peptides matching TDP-43 were identified in spot 323, which validated the 2-DE analysis with regards to differential intensities, as this spot was shown to undergo an increase in staining intensity in group C gels which contained protein from cells overexpressing wild-type TDP-43.

To begin with, genes that were likely representative of matched peptides, *U2AF1* and *MDH1* in spot 308, *HSPA9* and *CCT8* in spot 413 and *EEF2* in spot 421 (Table 2-1), were chosen for subsequent follow-up analyses. Splicing factor U2 small nuclear RNA auxillary factor 1 (*U2AF1*) is a component of the spliceosome and was thus relevant for this study since TDP-43 is involved in splicing and RNA metabolism processes. Similarly, the eukaryotic elongation factor 2 (*EEF2*), is known to be involved in translation of RNA. The genes malate dehydrogenase 1 (*MDH1*), T-complex protein 1 subunit theta (*CCT8*) and mitochondria heat shock protein 75 (*HSPA9*) were chosen for validation on the basis of their involvement in cellular metabolism and or protein folding (chaperones) under cellular stress conditions.

| Spot No. | Protein name/s | Accession No. | Theoretical MW/pI | Observed MW/pI | Mascot score | Matched peptides |
|---|---|---|---|---|---|---|
| 140 | ALB | P02768 | 69kDa/5.9 | 72kDa/4.3 | 335 | 8 |
| 1404 | ALB | P02768 | 69kDa/5.9 | 58kDa/5.8 | 435 | 18 |
| 300 | N.S | - | | 63kDa/4.3 | - | - |
| 308 | U2AF1; MDH1 | Q01081;P40925 | 27kDa/9.1; 36kDa/6.9 | 36kDa/6.5 | 197;149 | 8;6 |
| 323 | TDP-43 | Q13148 | 45kDa/5.9 | 48kDa/5.8 | 90 | 5 |
| 323 Con | UQCRC1 | P31930 | 53kDa/5.9 | N.C. | 276 | 8 |
| 413 | HSPA9;CCT8 | P38646; P50990 | 74kDa/5.9; 60/5.4 | 68kDa/5.5 | 192 | 4 |
| 421 | EEF2 | Q6PK56 | 95kDa/6.4 | 72kDa/6.31 | 649 | 34 |
| 666 | UAP1, CCT2 | Q16222; P78371 | 59kDa/5.9; 58kDa/6.0 | 59kDa/6.1 | 85;79 | 2;2 |

Table 3-1: **List of the most variable spots identified from the 2-DE analyses and the number of matched peptides for each gene.** Spots with matched peptides corresponding to various genes were selected for subsequent secondary validation analyses. ALB-Albumin, U2AF1-U2 auxillary factor 1, MDH1-Malate dehydrogenase 1, TDP-43-TAR-DNA binding protein, HSPA9-Heat shock 70 kDa protein 9, CCT8-Chaperonin containing TCP 1 subunit theta, EEF2-Eukaryotic elongation factor, UQCRC1-Ubiqiunol cytochrome reductase complex core protein 1, UDP-GlcNac/UAP1- Udp-N-Acetylglucosamine Pyrophosphorylase 1, CCT2-Chaperonin containing TCP 1 subunit beta, N.S-No significant peptides matched N.C- Not calculated.

Different validation strategies were employed in an effort to confirm correlative changes between genes corresponding to matched peptides and spot intensities, and which, would also facilitate a quick and simple methodology for validation. To begin with, quantitative real-time PCR (q-PCR) analyses were performed to determine whether changes in transcript expression of the genes corresponding to matched peptides did indeed result in differential protein production due to variation in TDP-43 levels. Specifically, q-PCR analyses were performed for *U2AF1*, *MDH1*, *EEF2*, *HSPA9* and *CCT8* in control cells (group A) and cells overexpressing wild-type flag-TDP-43 (group C). Figure 3-3 shows fold differences in transcript expression analysed by qPCR (upper panel) compared to differential spot intensities (protein expression differences) within the three spots (lower panel histograms-Figure 3-3).

As can be observed in Figure 3-3, the genes *U2AF1*, *MDH1* and *EEF2* exhibited transcript expression changes that correlated with the difference in spot intensity observed for spots 308 and 421 i.e. an increase in protein expression (Figure 3-3 a-c). In contrast, *HSPA9* and *CCT8* in spot 413, exhibited a decrease in transcript expression when TDP-43 was overexpressed, which did not correlate with the increase in spot intensity (protein) observed (Figure 3-3 d-e).

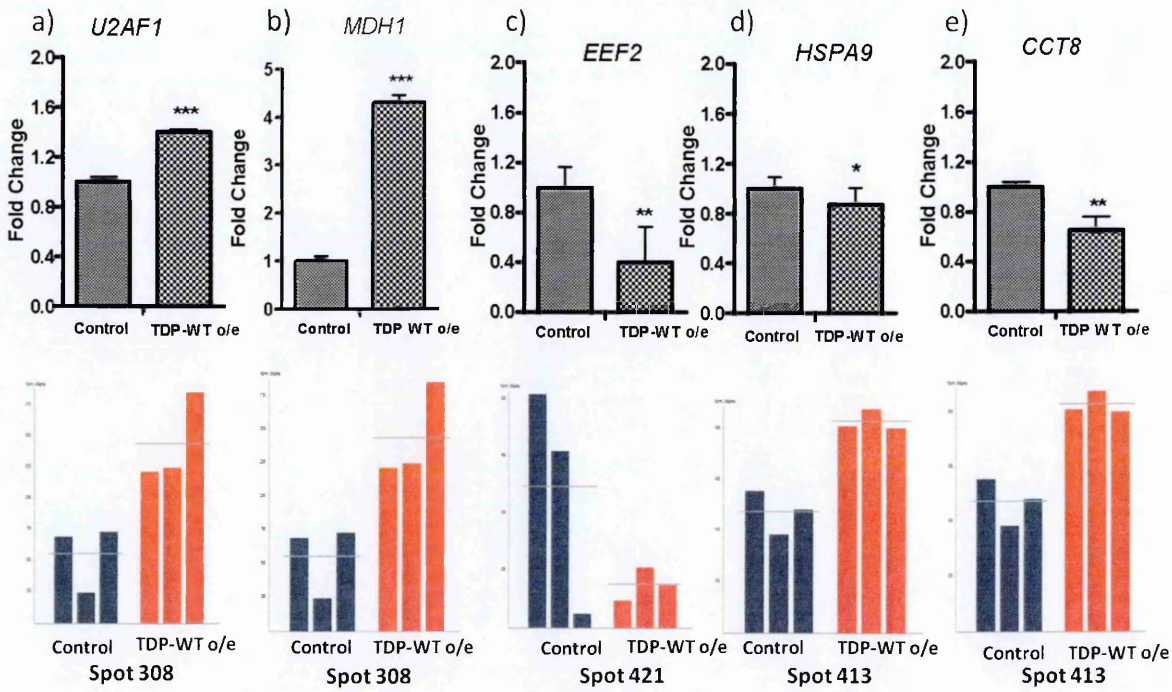



**Figure 3-3: Comparative analysis of transcript expression and relative spot intensities.** Changes in transcript expression (graphs in the upper panel) that correlated with the change in spot intensity (graphs in the bottom panel) were observed for the genes *U2AF1*, *MDH1* and *EEF2* (a-c) upon tetracycline induction of wild type FLAG-TDP-43 (TDP-WT o/e).For HSPA9 and CCT8 there was no correlation with the differential intensities observed within the spots (d-e). The y-axes on the bottom panel graphs represent normalized values of spot intensities, with the horizontal line representing mean value of the triplicates. qRT-PCRs are representative of three independent experiments in each cellular condition. *p-value≤ 0.05; **p-value≤0.01; ***p-value≤0.001. P-values are as follows: *U2AF1*-$3.36\times10^{-6}$; *MDH1*-$3.4\times10^{-6}$; *EEF2*-$4.2\times10^{-5}$; *HSPA9*-0.04; *CCT8*-0.006.

Subsequently, due to the seemingly positive correlation observed between protein and RNA expression in *U2AF1*, *MDH1* and *EEF2* (Figure 3-3 a-c), additional validation analyses were performed. No subsequent validations were performed for *HSPA9* and *CCT8* in spot 413, as no correlation was observed.

The highest ranking peptide match score within spot 421, corresponded to the *EEF2* transcript, which was previously shown to undergo a decrease in expression relative to the overexpression of TDP-43 that correlated with the 2-DE differential spot analysis (Figure 3-3c). To further validate this result *EEF2* transcript expression was analysed in cells depleted of TDP-43(group B) versus cells overexpressing TDP-43 (group C). In principle, if the overexpression of TDP-43 had a direct correlation with the decrease in expression of the *EEF2* transcript, then a TDP-43 knockdown should result in an increase in transcript expression.

As depicted by histograms representative of 2-DE differential spot intensity (Figure 3-4a, compare group B versus C), an increase in spot intensity at spot 421 was observed upon TDP-43 depletion. However, analysis of the expression levels of the *EEF2* transcript in TDP-43 depleted cells versus TDP-43 overexpression (Figure 3-4b), showed a 1.7 fold increase in expression of *EEF2* in cells overexpressing TDP-43, which was contrary to the observed decrease in spot intensity (Figure 3-4a).

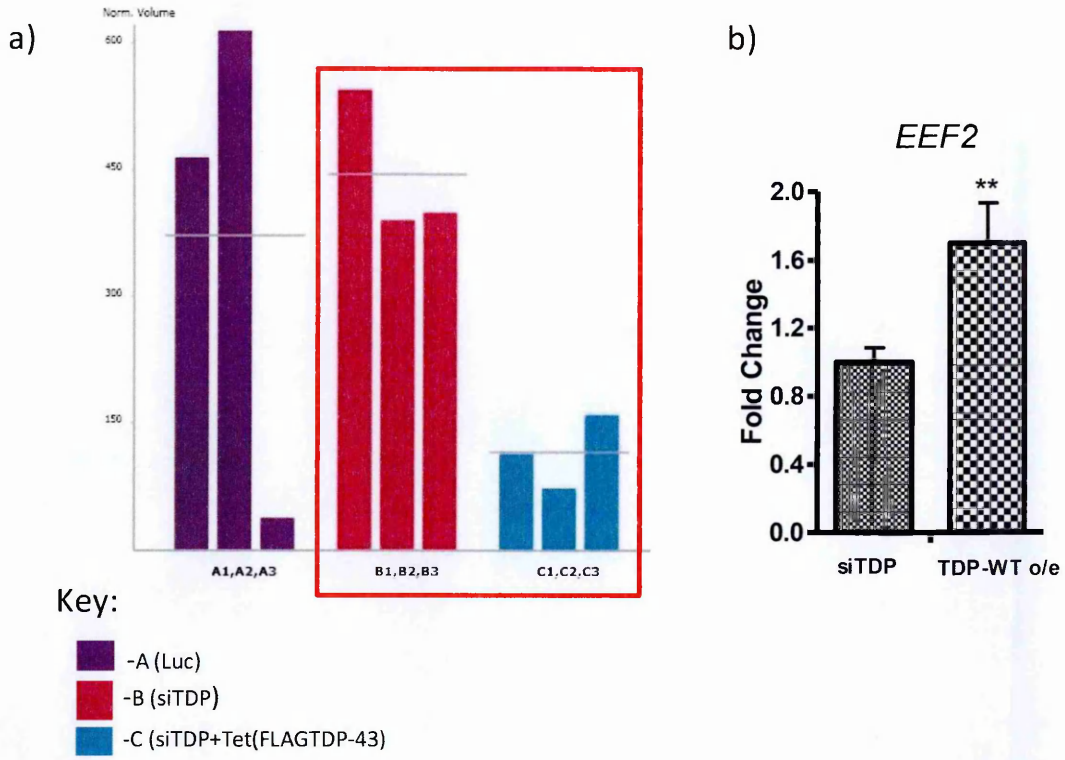



Figure 3-4: **A comparative analysis of spot 421 and EEF2 transcript expression following TDP-43 knockdown versus TDP overexpression.** (a) Triplicates of spot intensity differences detected in spot 421 under the various conditions pertaining to TDP-43: siRNA against TDP-43 treated cells of group B (red) versus FLAG-TDP-43 overexpression group C (blue). The mean of each triplicate is represented by the horizontal line across the bars. The knockdown of TDP-43 results in an overall increase in spot intensity whereas overexpression results in a decrease. (b) Fold difference in expression of the *EEF2* transcript TDP-43 depleted cells compared to overexpression of TDP-43. Fold differences in qPCR are representative of three independent experiments. *EEF2*-0.008. **p-value $\leq$ 0.01.

This lack of correlation indicated that either TDP-43 had an in-direct effect on the *EEF2* transcript, or another intervening gene/protein not idenetified by mass spectrometry was effecting the observed changes in spot intensity.

For spot 308, different validation analyses were performed for both *U2AF1* and/or *MDH1* that included analysis of transcript abundance using northern blots and western blots to detect differences in protein expression. In this case the choice of validation strategy was informed by the observation that for the previous analysis in spot 421, no correlation was observed using q-PCR, thus northern blots were utilised to better detect relative differences in abundance of mRNA transcript and validate the qPCR analyses which are known to sometimes have amplification biases. Northern blot analysis of the *U2AF1* mRNA transcript

showed a decrease in transcript abundance as quantified by Image J from three independent experiments (Figure 3-5, lower panel) upon overexpression of TDP-43 as compared to the control cells after normalisation with *GAPDH* mRNA (Figure 3-5a). This decrease in transcript abundance as determined by northern blot analysis was the opposite of the previously observed increase in transcript expression by qPCR and the increase in staining (protein) intensity at spot 308. Since northern blots are more reliable in detecting differences mRNA abundance (Ding et al. 2007) , this analysis was more likely to reflect the actual situation of transcript change following overexpression of TDP-43, thus no further analyses were performed for *U2AF1.*

On the other hand, transcript abundance of *MDH1* by means of a northern blot, another protein identified in spot 308 by mass spectrometry, was found to increase when TDP-43 was over-expressed (Figure 3-5). This was in line with the observed increase in differential spot intensity and q-PCR analysis (Figure 3-3b). Consequently, as a final validation effort and taking into consideration the correlation between all the analyses performed thus far, an analysis of protein expression differences by means of a Western blot was performed. Immunodetection of MDH1 protein levels (Figure 3-5c) detected two different isoforms of the protein (phosphorylated and mature), both of which did not show any changes in expression levels upon overexpression of wild type FLAG-TDP-43 that correlated with the previous analyses.

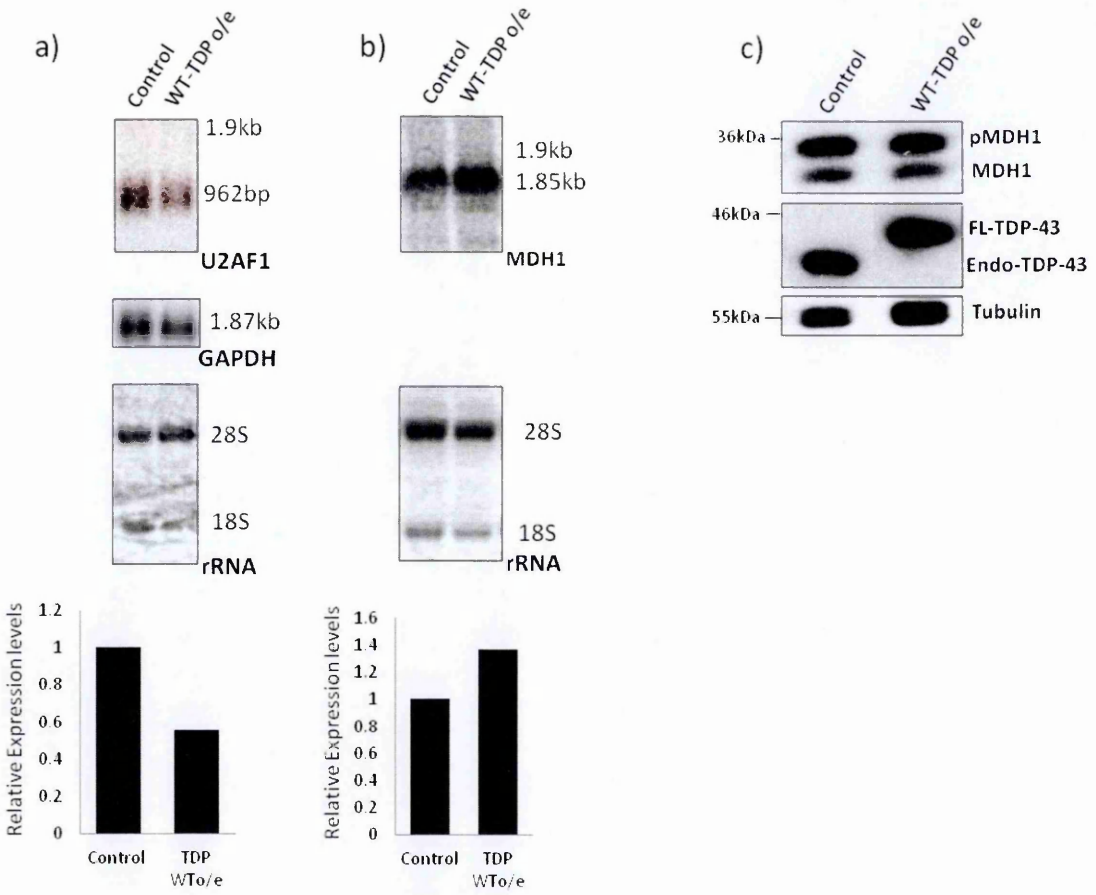



Figure 3-5: **Northern blot analyses of U2AF1 and MDH1 mRNA.** (a) In the upper panel *U2AF1* northern blot analysis in physiological levels of TDP-43 (control), and overexpression of the FLAG-TDP-43 protein. In the middle panel, a northern blot with GAPDH transcript signal which was used for the normalisation of the *U2AF1* signal. In the lower panel, EtBr staining of the ribosomal RNAs (rRNA) that were used as length markers and to check the integrity of total RNA. EtBr stained rRNAs were also used as alternative normalisers. The graph at the bottom represents Image J quantified relative expression levels of the mean values obtained from three independent experiments in the already described cellular conditions. (b) *MDH1* northern blot under physiological levels of TDP-43 and overexpression of the FLAG-TDP-43 protein with the EtBr stained rRNA gel below, that was used to normalise relative transcript abundances. The graph at the bottom represents Image J quantified relative expression levels of *MDH1*obtained from the mean of three independent experiments. (c) Western blot analysis of MDH1. Two isoforms were detected including phosphorylated (pMDH1) and un-phosphorylated (MDH1) forms of the protein. Endogenous levels of TDP-43 and the overexpression of the FLAG-TDP-43 were analysed using an anti TDP-antibody. Antibody against tubulin was used as loading control.

These results suggested two hypotheses; another protein, not identified by the mass spectrometry analysis, could be responsible for changes observed in the 2D gels or that the program used for normalisation (Ludesi image analysis software) and quantification of spots was not as accurate as had been perceived.

Consequently, since no succesful correlation had been observed in this initial validation analysis, a mass spectrometric re-analysis was performed for the rest of the spots shown to undergo significant variability, but for which no proteins had been identified (Table 3-1), namely, spots 140, 300 and 1404.

### 3.1.2. *Mass spectrometry re-analysis and validation*

Prior to the mass spectrometry re-analysis, an in-gel digestion and peptide extraction was performed for spots 140, 300 and 1404 according to the protocol specified in section 2.1.3 of materials and methods. In addition, other spots (308 and 323) that had previously been analysed were included as technical controls. From the re-analysis, 2/5 spots analysed had the same high-ranking peptides identified as in the previous mass spectrometry analysis performed by Proteome Factory (Berlin, Germany). Specifically, serum Albumin (ALB) and MDH1/U2AF1, were again identified as high-ranking peptide hits for spots 1404 and 308 respectively. For spot 323, Ubiqiunol cytochrome reductase complex core protein 1 (UQCRC1), was identified as a high-ranking peptide match, which was in contrast to the previously identified TDP-43 hit. In the previous mass spectrometry analysis for spot 323, other low-ranking matched peptides identified, did not include UQCRC1, however, within the spot '323 Con' an overlapping spot close to 323, TDP-43 had been identified as a low-ranking peptide. Therefore, since only UQCRC1 and not TDP-43 was identified in this re-analysis, it is highly likely that there may have been a shift within the spots probably due to gel-gel run variability that may have led to a misidentification of this spot. In fact, other than UQCRC1, none of the other low-ranking peptides were identified in the re-analysis. Furthermore, in the other two spots, 140 and 300, new high-ranking peptides were identified, i.e. nuclear autoantigenic sperm protein (NASP) and nucleosome assembly protein 1 like-1 (NAP1L1) respectively, in which previously, albumin and a non-significant result had been reported. The observed variability between the initial and

subsequent mass spectrometry analysis could points towards imperfect nature of image analysis.

Validation analysis of spot 308 (MDH1/U2AF1) has already been discussed in section 3 while the genes *UQCRC1* and *ALB*, identified in spots 323 and 1404 respectively were not investigated further since they were not relevant to the study. *UQCRC1* is a mitochondrial gene whereas *ALB* (bovine), a serum protein may have been a contaminant. With regards to spots 140 and 300, subsequent validation analyses of NASP and NAP1L1 are discussed hereafter.

#### 3.1.2.1. *Spots 140 and 300 validation: the case of NAP1L1 and NASP*

Validation analyses of spots 140 and 300 were performed initially by quantifying changes in transcript expression using q-PCR. The *NASP* transcript had a near five-fold increase in expression upon TDP-43 overexpression, which was the opposite of what was expected considering the decrease in spot intensity observed (Figure 3-6a, lower panel). On the other hand, the *NAP1L1* transcript could not be quantified accurately and further primer optimisation was required. Alternatively, a Western blot analysis was used to detect changes in protein expression relative to TDP-43 overexpression. Immunodetection of NAP1L1 protein identfied two isoforms, whose expression levels did not change upon TDP-43 overexpression and was thus not in agreement with the decrease in spot intensity observed for spot 300.

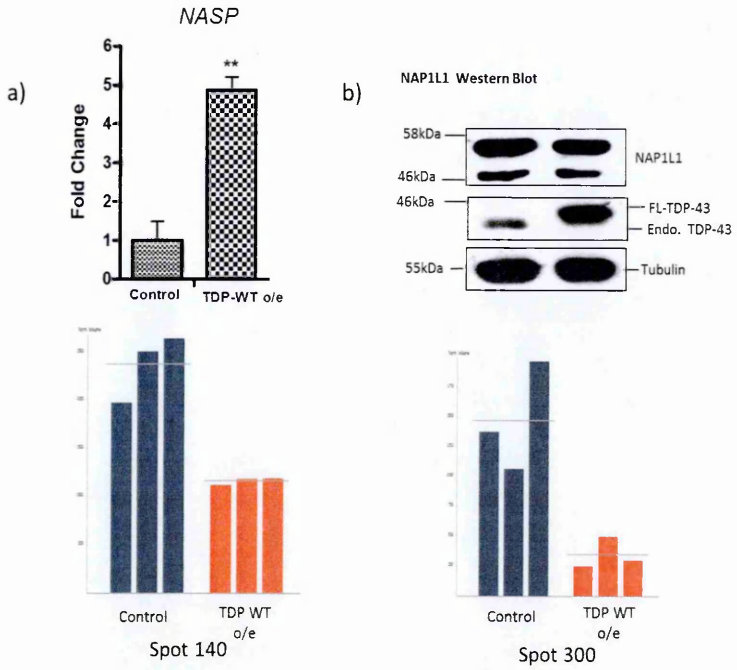



Figure 3-6: **Validation analyses for spots 140 and 300.** (a) Real time PCR analysis of *NASP* mRNA under physiological levels of TDP-43 versus overexpression are shown in the upper graph. Fold differences in expression are representative of three independent experiments. The lower panel graphs represent 2-DE differential spot intensities in the previously described cellular conditions. (b) Upper panel-Western blot analysis of NAP1L1 together with control checks of endogenous TDP-43 levels and FLAG-TDP-43 overexpression. Antibody against tubulin was used as a loading control. **p-value $\leq 0.01$.

Taken together, the validation analyses for spots 140 and 300 did not exhibit any correlation with the changes in spot intensities observed.

Fundamentally, the 2-DE analyses did not reveal any real correlation between the observed differential spot intensities and the various levels of TDP-43 expression. An independent mass spectrometry analysis of some of the spots confirmed that this lack of correlation was not related to technical processing. The utilities and limitations of 2-DE gels for global protein analyses have been discussed in depth in the discussion (Chapter 3). Several other more sensitive applications for 2-DE analyses exist and these could be employed in future work.

## 3.2. Splice Junction Microarray Analysis

### 3.2.1. *Splicing Data and Pathway Analyses*

Splice-sensitive microarrays have become a popular approach for global profiling of alternative splicing events and gene expression changes (Blencowe et al. 2009; Shen et al.

2010). To gain deeper insight into global transcriptome changes that are dependent on TDP-43, the splice-junction (H-JAY) Affymetrix microarray platform was utilised. Specifically, using HEK 293 stable cell lines previously described in the introduction section 1.6.1 and whose TDP-43 expression profile is shown in Figure 3-1, transcriptome changes resulting from four different cellular conditions were analysed: Group A in which cells were depleted of endogenous TDP-4 (siRNA-TDP-43), group B which consisted of cells that had endogeous TDP-43 silenced and simultaneously over-expressed the si-resistant form of wild-type TDP-43, and group C which similarly consisted of cells depleted of TDP-43 and simultaneously overexpressed si-resistant mutant (F4L) form of TDP-43, which is unable to bind RNA. The analysis of these cells under the various conditions were compared to a control group of cells that had been treated with siLuciferase (non-targeting siRNA) as has been shown in Figure 3-1. Furthermore, to ensure a uniform background induction media containing tetracycline was added to all groups of cells.

Total RNA was extracted from each set of samples prepared in triplicate and sent to Genosplice (Evry, France). Splice junction arrays contain probes on exons and exon-exon junctions which enable the identification of included and excluded exons in all known human transcripts. Splice junction arrays are also able to provide evidence of relative changes in transcript expression as determined by relative fluorescent intensities of the specific probes, with regards to different cellular conditions (Johnson et al. 2003). The analyses obtained from the Affymetrix data identified several genes that underwent TDP-43 dependent transcript level changes (Figure 3-7). Overexpression of wild type TDP-43 was observed to affect transcripts to a greater extent compared to depletion (1099 vs. 483 genes uniquely affected by either action, respectively). It is interesting to note that depletion, overexpression of TDP-43, and expression of the F4L mutant commonly affect a significantly large group of genes (2371), suggesting that alterations of TDP-43 levels act

on similar pathways, irrespective of the proteins capacity to bind mRNA. Furthermore, expression of the TDP-43 F4L mutant shows that only 251 genes are additionally affected by the knockout of RNA binding capacity. Expression of this mutant TDP-43 protein modifies genes in a similar manner to TDP43 wild-type overexpression than TDP-43 depletion (506 vs. 91) i.e. higher number of genes. Given the significant role reported thus far for TDP-43 in regulation of RNA metabolism, validation experiments were centred on data obtained for alternative splicing events that appeared to be dependent on TDP-43.

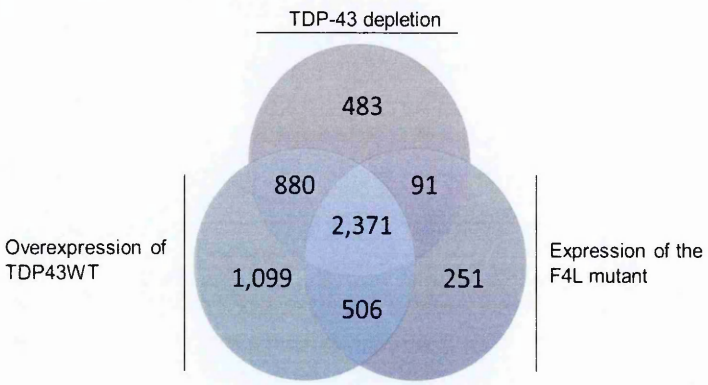



Figure 3-7: **Venn diagram showing the distribution of genes undergoing transcript level changes upon TDP-43 depletion, TDP-43 wild type overexpression and TDP-43 F4L overexpression.** Each group of genes was compared with the control (siLuciferase) reference set of genes.

In keeping with this role for TDP-43, hits obtained for the alternatively spliced exons were grouped according to three types of alternative splicing events; differential cassette exon splicing, mutually exclusive exon splicing, and differential usage of 5'/3' splice sites (Figure 3-8) below.

To improve the potential of identifying events that are directly dependent on TDP-43 levels, genes that satisfied all three of the following criteria were considered: events that were altered following depletion of TDP-43 (Group A), that could be rescued by the induction of the flagged si-resistant TDP-43 WT (Group B), but which could not be rescued following induction of the flagged si-resistant TDP-43 F4L mutant (Group C).

Principally, this included 134 cassette exons, 16 alternative 5'/3' splice site selections and 13 mutually exclusive exons. The full list of these genes can be seen in Figure 3-8 below.

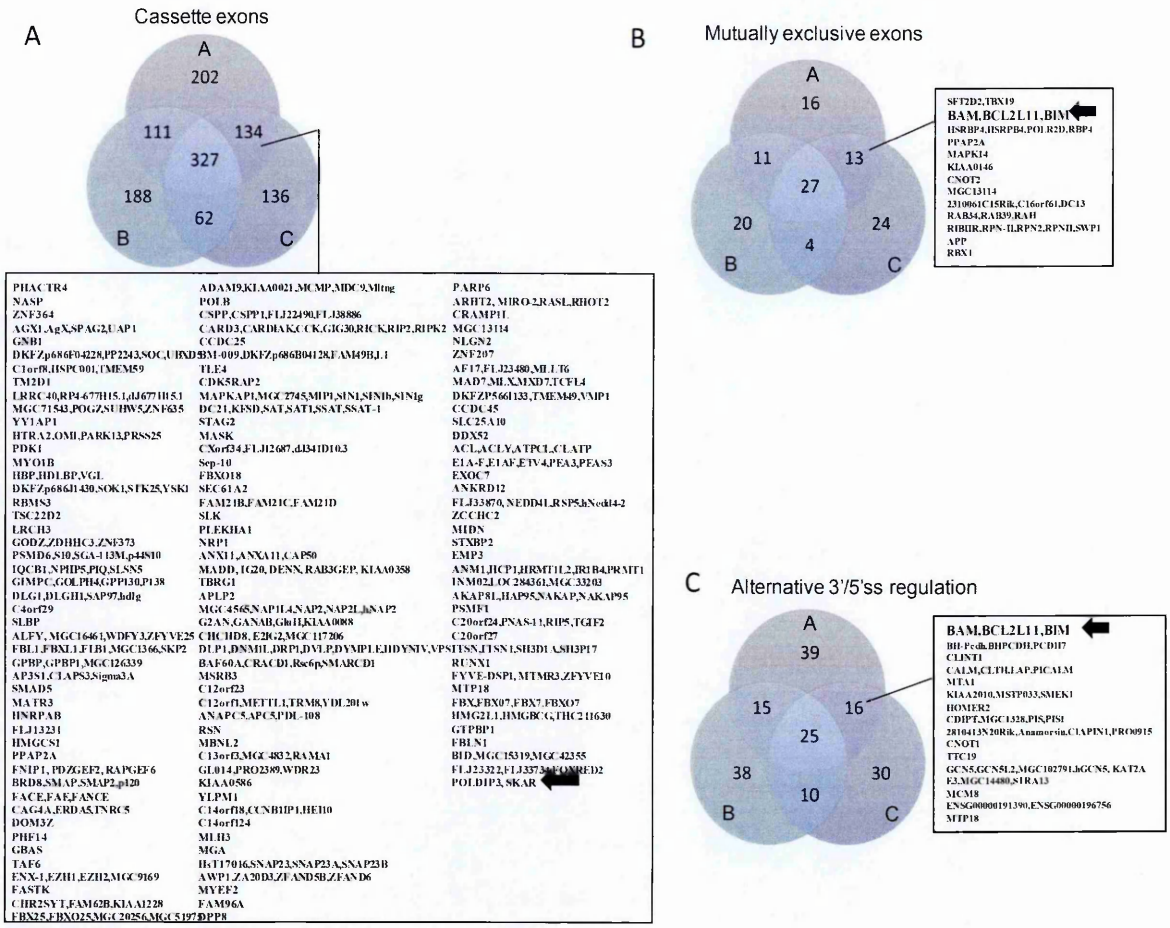



Figure 3-8: **Hits obtained from the splice sensitive microarray analysis.** Venn diagrams showing the distribution of genes undergoing: (a) differential cassette exon splicing, (b) mutually exclusive exons and (c) alternative 3'/5'splice site regulation in a TDP-43 dependent manner. Within the closed boxes are lists of genes that satisfy the criteria of being directly affected by TDP-43. Bold arrows indicate genes that were identified in this analysis, and that were already validated and shown to undergo alternative splicing following TDP-43 depletion by other studies (*BIM/Bcl-2* and *POLDIP3/SKAR*).

A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the 162 genes that fulfilled all three criteria as described previously, showed that they were mostly involved in regulating the alternative splicing profile of other factors involved in alternative splicing/RNA binding and phosphoproteins, (Figure 3-9). In addition, many genes present in this list were also involved in cell cycle control, Ubl (Ubiquitin) conjugation, and apoptotic pathways that have previously been associated with TDP-43.

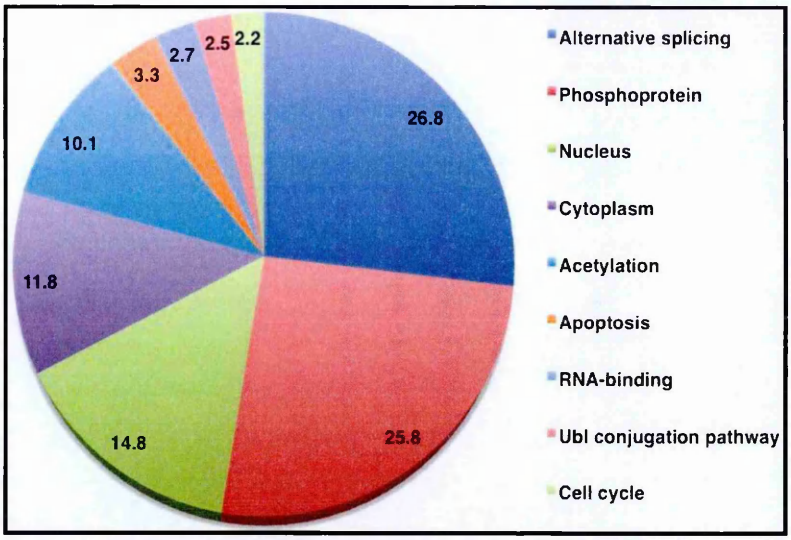



**Figure 3-9: Pie chart depicting percentages of genes involved in various pathways following a KEGG pathway analysis.** Most genes were depicted as being involved in alternative splicing and phosphoproteins, cumulatively 52.6%.

### 3.2.2. *Identifying genes whose splicing was directly affected by TDP-43 depletion*

From the 163 candidate list of genes corresponding to the previously described criteria i.e. appeared to be altered directly by TDP-43 depletion, two genes, *BIM/BCL-2* (Bcl-2 interacting protein/B-cell lymphoma 2-like 11) and *SKAR/POLDIP3,* had previously been shown to undergo alternative splicing following TDP-43 depletion by others (Tollervey et al. 2011; Shiga et al. 2012; Fiesel et al. 2012). It was therefore of interest, as further validation of the experimental procedure in this study, to perform RT-PCR analysis of *BIM/BCL-2L11* confirming that depleting TDP-43 shifted the splicing profile of *BIM/BCL-2L11* towards the increased expression of the pro-apoptotic BIMs isoform with respect to the anti-apoptotic isoforms $BIM_L/BIM_{EL}$ (Figure 3-10a, compare lanes 1 and 2). Most importantly, this shift in splicing isoforms could be reverted in the cells that expressed the si-resistant wild-type FLAG-TDP-43 (lane 3) and no rescue could be observed in the cell lines expressing the mutant FLAG-TDP43 (F4L) RNA-binding mutant (lane 4). In keeping with this pattern, the same type of splicing effects could also be observed for SKAR/POLDIP3, where inclusion of exon 3 was abolished following endogenous TDP-43

depletion (Figure 3-10b, lane 2), rescued by the expression of the si-resistant FLAG-TDP43 WT (lane 3), and not rescued by the FLAG-TDP43 F4L mutant (lane 4).

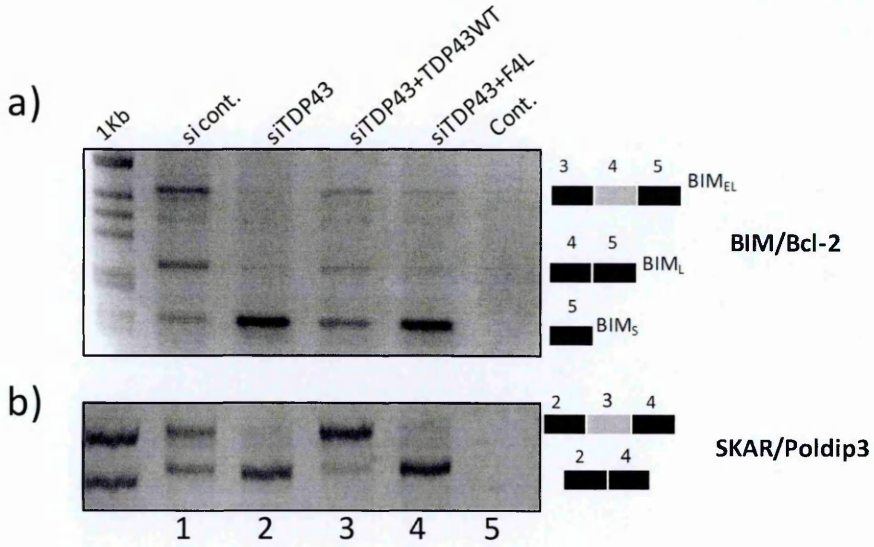



Figure 3-10: **Confirmation of previously reported TDP-43 dependent altered splicing profiles of *BIM/Bcl-2* and *SKAR/POLDIP3*.** a) RT-PCR image showing alternative splicing of exon 4 in *BIM/Bcl-2* and b) exon 3 in *POLDIP3/SKAR* under physiological levels of TDP-43 (lane 1, sicont.), that are altered following depletion of TDP-43 (lane 2, siTDP43), can be rescued by induction of si-resistant FLAG-TDP-43 WT (lane 3, siTDP43+TDP43-WT) but cannot be rescued with and overexpression of mutant FLAG-TDP-43 (F4L) (lane 4, siTDP43+F4L).

Having verified experimentally, that the data obtained from the microarray contained genes known to be affected by TDP-43 levels, from the initial candidate gene list of 162, a list of genes were selected based on their score (>2-fold change in splicing profile) for subsequent vadlidation. The splicing profiles of these selected genes was analysed by RT-PCR and are listed in table 3-2 below.

| GENE | EXON | AMPLIFICATION | CHANGES OBSERVED WITH RT-PCR |
|---|---|---|---|
| **SKAR/POLDIP3** | 3 | ex 2-4 | Yes |
| **BIM/BCL2L11** | 3 | ex 2-4 | Yes |
| TBX19 | 7 | ex 1-8 | No |
| C14orf18 | 5 | ex 3-7 | No |
| **STAG2** | 30b | ex 29-32 | Yes |
| GANAB | 6 | ex 4-9 | No |
| C20orf24 | 3 | ex 1-4 | No |
| NLGN2 | 2 | ex 1-4 | No |
| FBXO18 | 2 | ex 1-3 | No |
| **MADD** | 31 | ex 30-32 | Yes |
| ANKRD12 | 4 | ex 1-2 | No |
| **BRD8** | 20 | ex 19-21 | Yes |
| **FNIP1** | 7 | ex 6-8 | Yes |
| ALFY | 45 | ex 44-46 | No |
| ERGIC3 | 3 | ex 1-4 | No |
| CASK | 19 | ex 18-20 | No |
| CRAMP1L | 15 | ex 14-16 | No |
| ZNF207 | 9 | ex 8-10 | No |
| BID | 4 | ex 3-5 | No |

Table 3-2: **Genes chosen for validation of altered splicing profiles based on their predicted score.** The regions amplified and relevant exons are also shown below. Genes highlighted in red depict positive confirmations of changes in splicing profiles, whereas those highlighted in green indicate changes that were observed in our list and that were previously published.

TDP-43 dependent changes in splicing were confirmed in four (highlighted in red–Table 3-2) out of nineteen genes analysed: Stromal antigen 2 (*STAG2*), MAP-kinase activating death domain (*MADD*), Folliculin interacting protein 1 (*FNIP*) and Bromo-domain 8 containing protein (BRD8).

RT-PCR analyses were performed on endogenous transcripts using primers located in exons flanking the relevant skipped/included exons (highlighted in red-Figure 3-11 below).

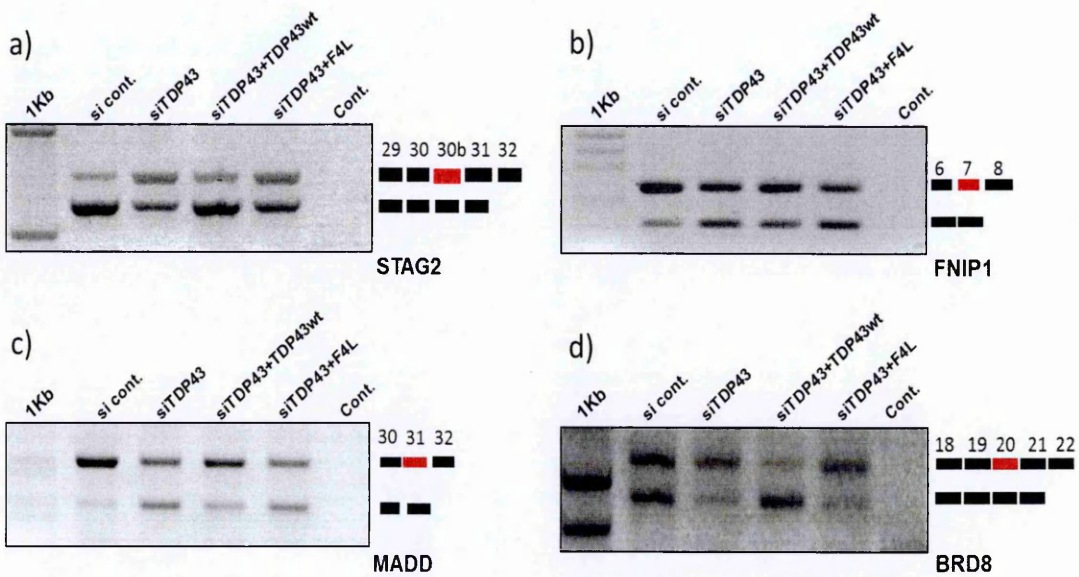



Figure 3-11: **Endogenous splicing profiles in HEK 293 cells of *STAG2*, *MADD*, *FNIP1* and *BRD8* that were found to be TDP-43 dependent.** (a-d) RT-PCR gel images depicting TDP-43 dependent alternative splicing in endogenous transcripts of *STAG2*, *MADD*, *FNIP1* and *BRD8*. The relevant exons (highlighted in red) as well as flanking exons are shown (a-d).

In two of the genes identified, *STAG2* and *BRD8*, depletion of TDP-43 led to an increased inclusion of the relevant exons, as observed by RT-PCR. Specifically, in *STAG2*, exon 30b was observed to undergo an increase in inclusion when TDP-43 was depleted in cells, which could be rescued when the FLAG-TDP-43 WT was over-expressed. Furthermore, over-expressing mutant TDP-43 (F4L) did not result in increased inclusion and in effect had a splicing profile similar to that observed when TDP-43 was silenced (Figure 3-11a) i.e. increased inclusion signifying that TDP-43 binding capacity was essential for

inhibiting recognition of the normally excluded exon 30b of *STAG2*. In the *BRD8* RT-PCR analysis, exon 20 was observed to undergo a significant increase in inclusion upon TDP-43 depletion Figure 3-11d, which could be rescued with over-expression of the WT TDP-43 but not the mutant form.

In contrast, an increase in exclusion of exons 7 and 31 of *FNIP1* and *MADD* respectively, was observed when TDP-43 was depleted from the cells. A partial rescue of these exons could be observed when FLAG-TDP-43 WT but not FLAG-TDP-43 F4L mutant was overexpressed.

Similar RT-PCR assays were performed in neuron-derived cell lines SH-SY-5Y and SK-N-AS, which are well-known *in vitro* models used to assess neuron function in studies of neurodegenerative diseases. As in HEK 293 cells, similar TDP-43 dependent alternative splicing profiles of exons 30b, 31, 7 and 20 of *STAG2*, *MADD*, *FNIP1* and *BRD8* respectively, were observed in these neuronal cell lines (Figure 3-12b). In particular, following TDP-43 depletion an increase in *STAG2* exon 30b and BRD8 exon 8 inclusion was observed, whereas for *MADD* exon 31 and *FNIP1* exon 7, there was an increase in exclusion.

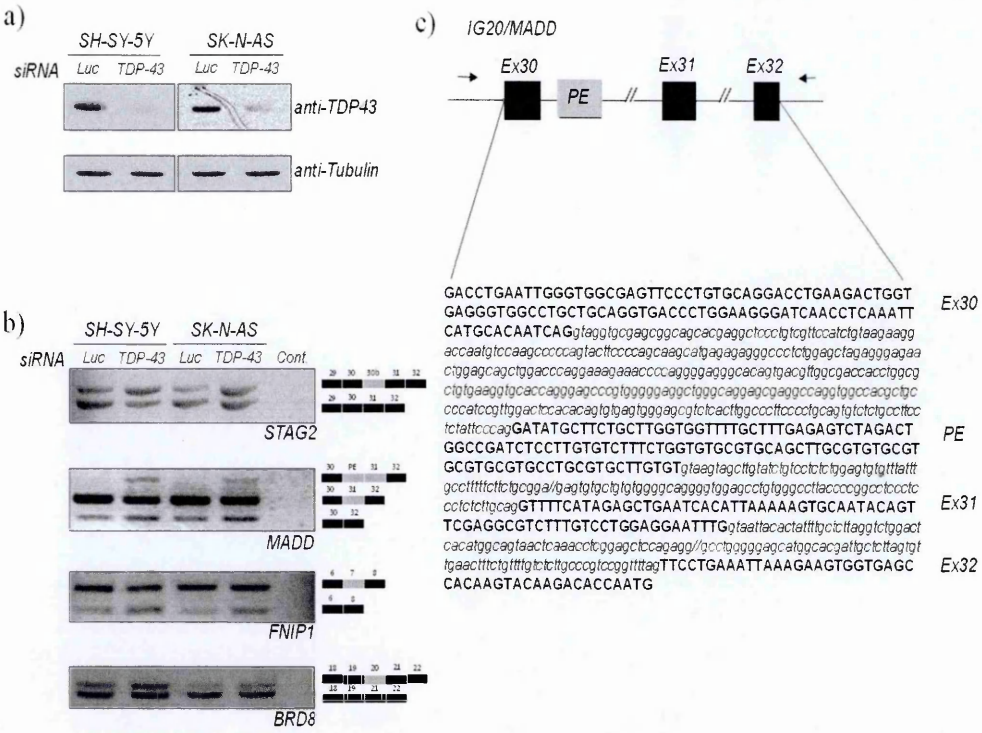



**Figure 3-12: Depletion of TDP-43 in neuroblastoma cell lines results in alterations in pre-mRNA splicing profiles that resemble those observed in HEK-293 cell lines.** (a) Western blot showing efficiency of TDP-43 silencing in SH-SY-5Y and SK-N-AS neuroblastoma cell lines. An antibody detecting tubulin was used as protein loading control. (b) RT-PCR analysis of all the selected genes performed in control (luciferase siRNA) and TDP-43 depleted cells with schematic representations of the amplicons on the right. (c) Schematic representation of the endogenous *MADD* gene spanning exon 30 to 32. The grey box shows the pseudoexon identified in the endogenous RT-PCR in these neuronal cells but not in the minigene analysis or in the endogenous profile of HEK 293 cells. The lines stemming from the exon scheme depict both the intronic (lower case) and exonic sequences (upper case) including the pseudoexon sequence.

Interestingly, analysis of MADD exon 31 in SK-N-AS and SH-SY-5Y neuroblastoma cell lines not only revealed a similar splicing profile as that observed previously in HEK 293, in which depletion of TDP-43 resulted in increased exclusion of this exon (Figure 3-11) but also showed an extra band (pseudo exon-PE), above the all-inclusive (inclusion of exon 31) band, that was present only when TDP-43 was depleted in the cell. Sequencing of this extra band revealed that it consisted of exon 30, 31 and 32, as well as an additional exon (identified as a pseudo exon (PE)) of 115 bp intronic sequence (Figure 3-12b, second panel) that was flanked by viable donor and acceptor splice sites. Under physiological levels of TDP-43, the pseudo exon donor and acceptor splices sites are not normally recognized which highlights a role for TDP-43 in repressing the recognition of this exon (Figure 3-12c). A possible explanation for the presence of the pseudo exon in these

neuronal cell lines and not in HEK 293 cells could be the difference in abundance levels of *trans*-acting factors that aid in the recognition of the pseudo exon splice sites. At the protein level, insertion of this sequence just upstream of exon 31 is predicted to introduce a premature stop codon (PTC) and could possibly lead to degradation of the transcript by nonsense mediated decay mechanisms.

Following the confirmation of TDP-43 dependent altered splicing profiles in *STAG2*, *MADD*, *BRD8* and *FNIP1*, in both neuronal and non-neuronal cell lines, further characterisation of TDP-43 binding in these genes was performed.

### 3.2.3. *MADD: Characterisation of TDP-43 dependent alternative splicing of Exon 31*

*MADD* plays an essential role in inhibiting the apoptotic pathway, and has been reported to have a neuroprotective role (Miyoshi & Takai 2004). In the splice-junction microarray analyses, *MADD* exon 31 was depicted to undergo TDP-43 dependent alternative splicing that was subsequently confirmed with an RT-PCR assay. In particular, depleting TDP-43 in the cell, led to increased exclusion of exon 31, which could be rescued with overexpression of FLAG TDP-43 WT, as shown in Figure 3-11c. In addition, using Western blot analyses, the TDP-43 dependent alternative splicing of MADD exon 31 could also observed at the protein level following immunodetection of MADD protein (Figure 3-13). In the western blot analysis, two bands are visible corresponding to inclusion or exclusion of exon 31 in the proteins, which differ in abundance in a manner identical to the splicing profile observed in the mRNA when TDP-43 levels are altered (Figure 3-13). Notably, in the control (siLUC) lane, the protein isoforms containing included (upper) and excluded (lower) exon 31 are expressed in almost equal proportions. In addition, there seemed to be an overall decrease in expression of the observed protein isoforms following TDP-43 depletion, which could not be restored even when wild type TDP-43 was

overexpressed (lanes 2-4, Figure 3-13). As has been mentioned previously, the skipping of exon 31 following depletion of TDP-43 introduces a pre-mature termination codon, which could lead to a reduction in expression of the resulting isoforms due to non-sense mediated mechanisms. A poosible explanation for the observed decrease in expression of protein isoforms could be that protein recovery levels are lagging behind mRNA recovery, as the MADD has a high molecular weight (200 kDa) .

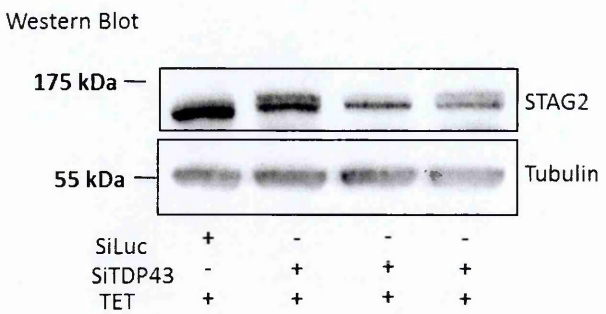



Figure 3-13: **TDP-43 dependent endogenous splicing profile of MADD exon 31 can be detected at the protein level.** Western blot performed with anti-MADD/IG20 detected altered isoform expression that correlated with the mRNA splicing isoforms (see figure 3-11). Protein isoform expression in the control cells appeared to be much higher and of equal ratio as opposed to the protein isoforms observed in the other three lanes that were depleted of endogenous TDP-43. Anti-tubulin was used as a protein loading control.

#### 3.2.3.1. *Analysis of MADD exon 31 alternative splicing in a heterologous context*

A closer examination of intronic sequences surrounding *MADD* exon 31 identified a short TG stretch approximately 53 bp in the upstream intron, a putative TDP-43 binding site. Since TDP-43 is known to preferentially bind to TG stretches (minimum of six required; Buratti and Baralle, 2001), this was the starting point for mapping the TDP-43 binding site involved in the alternative splicing of exon 31. In order to characterize this interaction that results in the usual inclusion of exon 31 under physiological levels of TDP-43, 278 bp of upstream and 263 bp of downstream intronic sequences together with 70 bp of exon 31 were cloned into the pTB minigene (Figure 3-14). The pTB minigene is a hybrid expression plasmid that contains exons from α-globin and fibronectin under the α-globin promoter. The intronic region between the two fibronectin exons contains an NdeI

restriction site that facilitates the insertion of a single exon together with short intronic sequences that would enable distinction between excluded and included exons. This is performed by RT-PCR analysis using primers that are specific for the globin and fibronectin exons (Pagani et al. 2000).

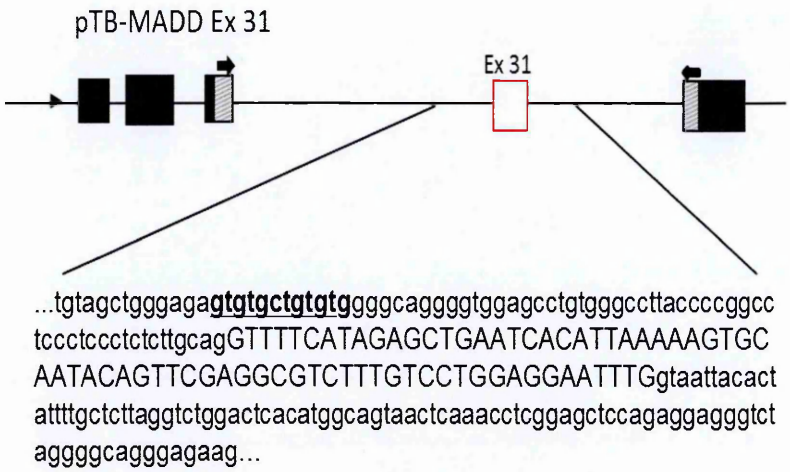



Figure 3-14: **Schematic diagram of a hybrid pTB-MADD-Ex 31 minigene containing exon 31 and flanking sequences.** The α-globin, fibronectin and human *MADD* exon 31 are shown as black, grey and red boxes, respectively. Intronic sequences are represented by lines connecting the exons. A zoomed-in view of intronic sequences flanking exon 31 are shown with the upstream intronic sequence of the short 12 nucleotide TG stretch highlighted in bold. Bold arrows indicate the location of primers (Alfa 2, 3 fwd and Bra rev) used for amplification in the RT-PCR assay.

Following transient transfections of the hybrid pTB-MADD-Ex 31 minigene under conditions in which TDP-43 was depleted and over-expressed, analysis of the splicing pattern resulted in a similar splicing profile as that observed in the endogenous transcript. Specifically, using primers specific for the hybrid minigene (bold arrows in Figure 3-14) three bands were visible from the RT-PCR analysis, two of which were relevant to the splicing profile (Figure 3-15a). The upper band, approximately 471 bp, labelled a hybrid exon (HE) (Figure 3-15 a&c) was identified as constituting of the EDB exon, *MADD* exon 31 (70 bp) and downstream MADD intronic sequence cloned into the minigene, and interpreted as an artefact of the minigene system. Nonetheless, the intensity of this band was not altered with the changes in the levels of TDP-43 and was thus not considered significant for the *MADD* splicing profile analysis. The two other bands consisted of a 310 bp band which was representative of exon 31 inclusion, and another of 239 bp, which was

representative of exon 31 exclusion. The 239 bp band was faintly visible in cellular conditions of physiological levels of TDP-43 i.e. siLUC of TDP-43 (Figure 3-15a) whereas when TDP-43 was depleted from the cell, the intensity of the 239 bp band increased, signifying an increase in exclusion of exon 31, consistent with the splicing profile observed for the endogenous transcripts (Figure 3-15b).

Upon overexpression of FLAG-TDP-43 WT, the splicing profile of the hybrid pTB-MADD-Ex 31 minigene exhibited a rescue of exon 31 inclusion resembling that of the minigene analysed under normal (siLUC) levels of TDP-43 (Figure 3-15c). In fact, from the minigene analysis, overexpression of TDP-43 seemed to result in an overall better inclusion of *MADD* exon 31 compared to physiological (siLUC) levels of TDP-43.

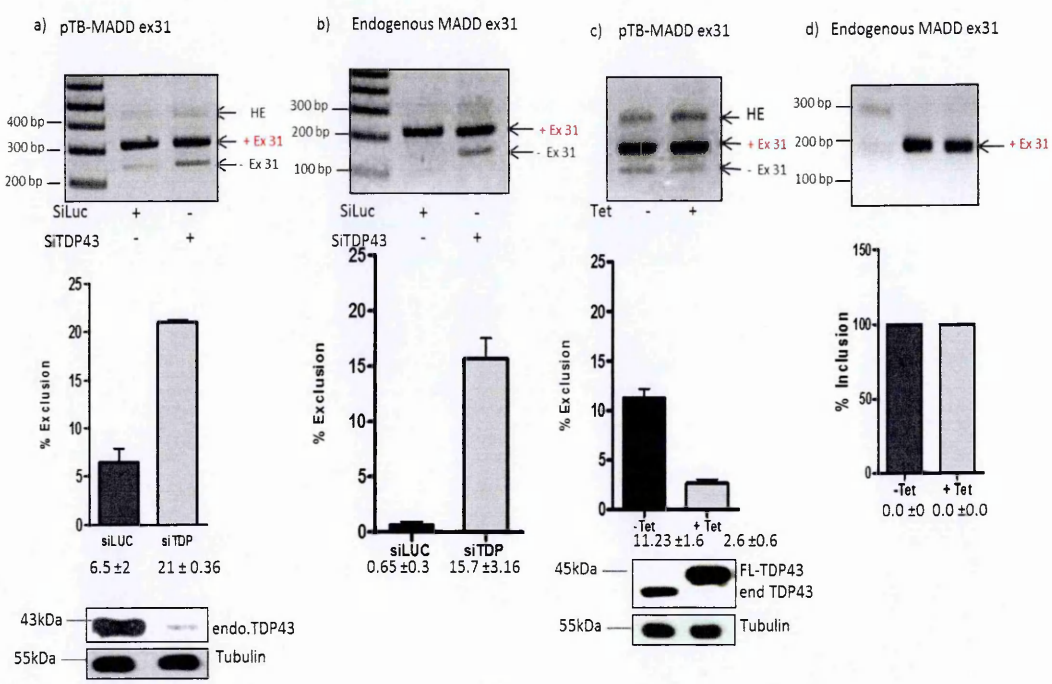



Figure 3-15: **The pTBP-MADD Ex 31 minigene recapitulates the endogenous MADD splicing profile.** a) RT-PCR analysis and relative Image J quantification (below) of the pTB-MADD-Ex 31 minigenes analysed under physiological levels of TDP-43 expression and depletion. b) Depletion of TDP-43 resulted in the increased exclusion of exon 31 as observed in the endogenous MADD splicing profile, and which could also be recapitulated in the pTB-MADD-Ex 31 minigene when TDP-43 was depleted in the cells. A rescue of exon 31 exclusion could similarly be observed in both the c) the pTB-MADD-Ex 31 minigene and in the d) endogenous transcript, when FLAG-TDP-43 WT was overexpressed. In fact, in the endogenous splicing profile exon 31 is 100% included upon TDP-43 overexpression. In addition, a hybrid exon (HE) that is not altered by the levels of TDP-43 and determined to be an artefact of the minigene system was observed only in the pTB-MADD-Ex 31 minigenes. Anti-TDP antibody was used to confirm the levels of TDP-43 in the same cells i.e normal (siLUC) and knock-down (siTDP-43) and transgenic WT TDP-43 overexpression (tet induced). Anti-tubulin was used as protein loading control for the western blots.

The confirmation of a TDP-43 dependent inclusion and exclusion of *MADD* exon 31 in the minigene indicated that the sequence to which TDP-43 bound was present in the sequences cloned into the pTB minigene. Thus subsequent analyses aimed to determine whether the TG stretch upstream of exon 31, a likely binding site of TDP-43 was involved in the TDP-43 dependent alternative splicing of MADD exon 31.

### 3.2.3.2. *Site directed mutagenesis of MADD TG stretch and EMSA analyses*

The Human Splice finder (HSF) (http://www.umd.be/HSF/) (Desmet et al. 2009), was used to analyse and identify sequences that would not create new splicing factor binding sites upon mutating the TG stretch. Following this analysis, the 5'-GTGTGCTGTGTG-3' sequence present in the upstream intronic sequence of MADD exon 31(bold and underlined in Figure 3-14) was mutated to 5'-ACATACCGCGAG-3' in the minigene. A comparison of the wild type pTB-MADD Ex 31 minigene and pTB-MADD Ex 31 mutant minigene is shown in Figure 3-16 below.

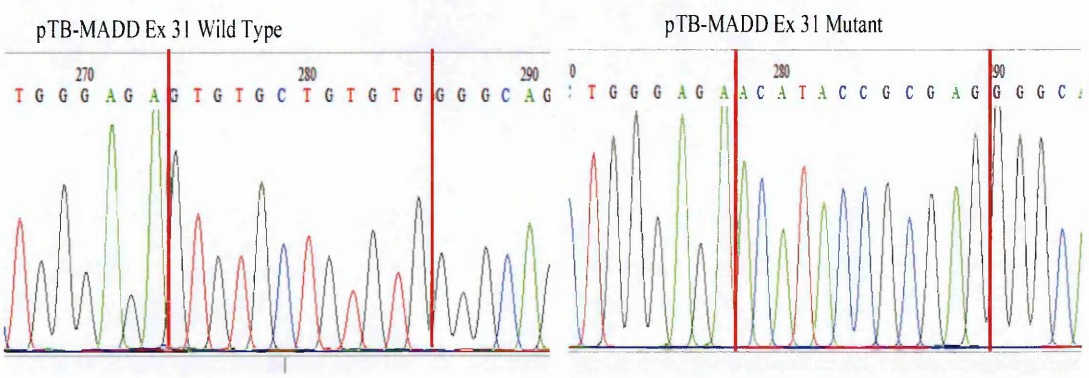



**Figure 3-16: Chromatogram comparison of wild type and mutant pTB-MADD Ex 31 minigenes showing mutation of TG stretch found in the upstream intronic region.** The sequences of interest are delineated with red lines.

Analysis of both wild type and mutant minigenes splicing profiles was performed using RT-PCR analysis following transient transfections into HEK 293 cells. Compared to the pTB-MADD Ex-31 (wild-type) WT minigene, the splicing profile observed for the mutant

minigene exhibited increased levels of exon 31 skipping (approximately 12%) that were independent of the levels of TDP-43 (compare lanes 2 and 4, Figure 3-17a).

When comparing wild type and mutant pTB-MADD following knockdown of TDP-43, the mutant minigene had higher levels of exon 31 skipping (12% to 6%) (Figure 3-17a). This indicated that mutating the TG stretch in the upstream intronic region was more efficient at inhibiting TDP-43 mediated recognition of MADD exon 31, thereby resulting in increased levels of skipping. In contrast, the depletion of TDP-43 by siRNA targeted sequences did not result in complete depletion as observed in the Western blot analysis (lower panel, Figure 3-17) and thus the residual TDP-43 may have still been able to elicit an effect towards the recognition of exon 31, resulting in slightly lower levels (6%) of exclusion of exon 31 as compared to the mutant pTB-MADD Ex-31 minigene.

Taken together, the observed lack of change in the splicing profile of the mutant pTB-MADD Ex-31 minigene i.e. constant higher levels of exon 31 exclusion that were independent of physiological or depleted levels of TDP-43 in the cell, strongly indicated that the TG stretch previously identified, may be involved in binding TDP-43 and aid in the recognition and inclusion of *MADD* exon 31. Similarly, overexpression of FLAG-TDP-43 WT TDP-43 did not lead to inclusion of exon 31 in the mutant minigene and there were still high levels (16%) of exon 31 exclusion compared to the wild type pTB-MADD Ex-31 minigene (Figure 3-17b).

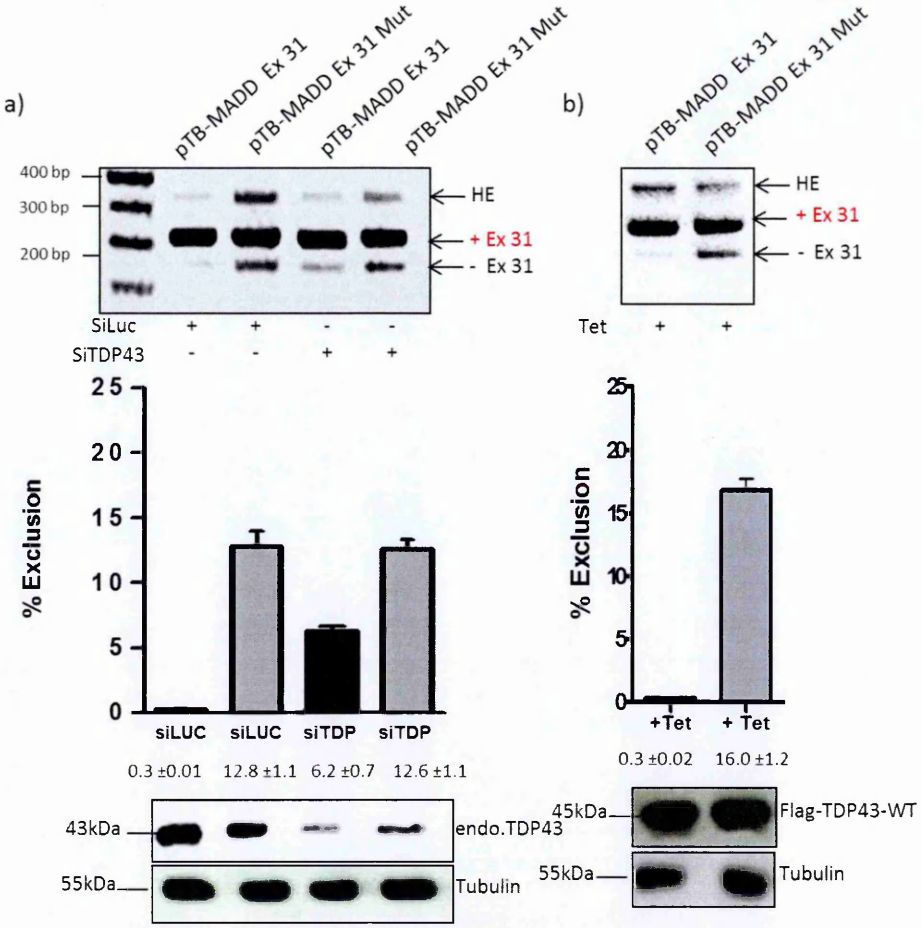



Figure 3-17: **Comparative analysis of pTB-MADD Ex 31 wild type and mutant minigenes indicates involvement of TG stretch in binding TDP-43** (a) RT-PCR analysis of wild type and mutant pTB-MADD Ex 31 minigenes under physiological and depleted levels of TDP-43, showed that there was increased exclusion of *MADD* exon 31 in the mutated minigene, irrespective of TDP-43 levels. The splicing profile of the wild type pTB-MADD Ex 31 minigene in lane 1 showed that under physiological levels of TDP-43 exon 31 is recognized and included, whereas depletion of TDP-43 resulted in increased exclusion of this exon. (b) Overexpression of FLAG-TDP-43 WT in mutant pTB-MADD Ex 31 minigene did not significantly alter exclusion of exon 31 indicating that the TG stretch is involved in binding TDP-43 to result in the usual recognition of this exon. This observation was in contrast to the adjacent splicing profile observed for wild type pTB-MADD Ex 31, which contains the TG stretch and in which exon 31 was included. The hybrid exon (HE), an artefact of this minigene is indicated with the arrows labelled HE. Anti-TDP-43 was used to confirm levels of TDP-43 and anti-tubulin was used as a loading control.

To further demonstrate that the effect of TDP-43 on MADD exon 31 splicing was indeed occurring due to the protein binding to this TG stretch, an electrophoretic-mobility shift assay (EMSA) was performed. The EMSA analysis was performed using an oligo that consisted of the twelve nucleotides of the TG stretch (MADD TG) (see underlined sequence in Figure 3-14) in the upstream intronic sequence of MADD exon 31.

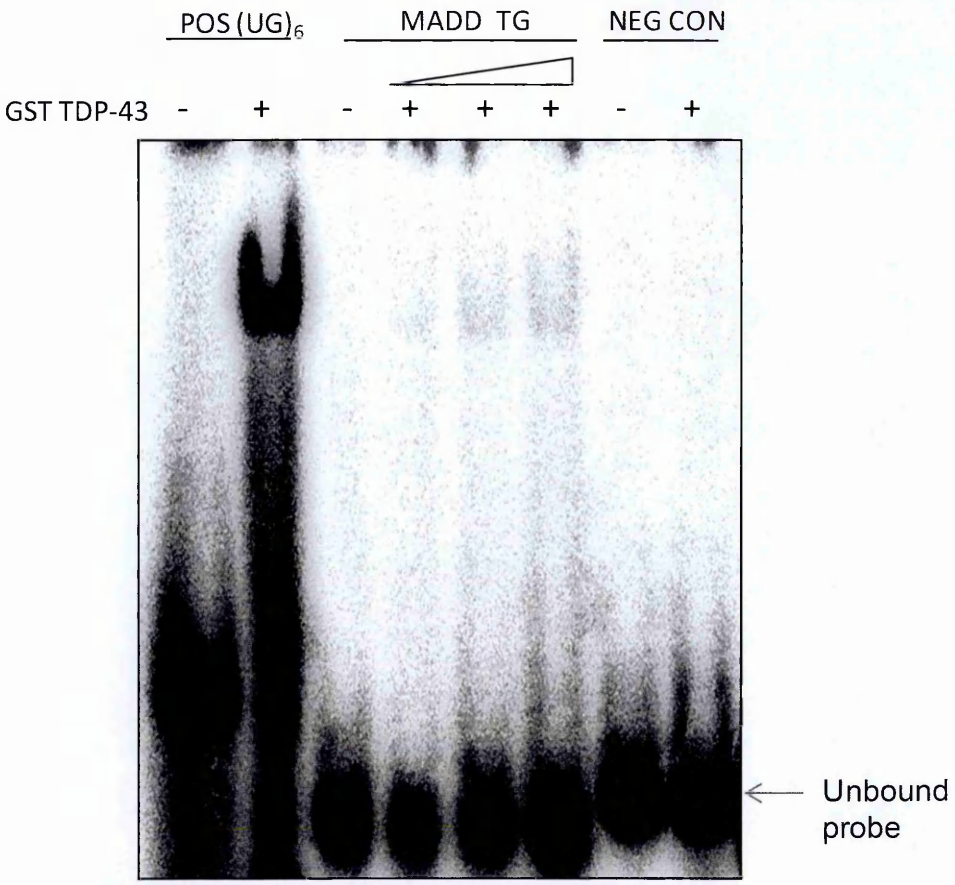



Figure 3-18: **MADD EMSA analyses using two different oligos and a known binder of TDP-43 ($UG_6$) used as a positive control.** EMSA performed with this oligo containing just the TG stretch revealed an increase in signal intensity with increasing amounts of GST-TDP-43.

The MADD TG oligo as well as an oligo consisting of six UG repeats ($UG_6$) that is known to bind TDP-43 (Bhardwaj et al. 2013) were kinase end-labelled and used in a binding reaction together with purified recombinant TDP-43, (see materials and methods section 2.2.7). EMSA performed with the MADD TG oligo, revealed a gradual increase in signal intensity of the observed band shift, with increasing amount of GST-TDP-43. This analysis confirmed that TDP-43 did indeed bind to this TG stretch. Furthermore, binding of TDP-43 to the TG stretch was validated by cold competition analyses, in which 10-20 fold molar excess concentrations of cold (un-labelled) MADD TG oligo were used in direct competition with the labelled MADD TG oligo (Figure 3-19a) or the $UG_6$ oligo (Figure 3-19b). Particularly, cold competition between labelled and cold MADD TG oligo revealed a decrease in signal intensity with increasing concentrations of the cold oligo, however,

since the signal was quite weak, the disappearance of the signal appeared to be immediate ('cold madd TG'-second lane) (Figure 3-19a). Similarly, binding competition between the labelled $UG_6$ (positive control) and un-labelled MADD TG oligo revealed a gradual increase in displacement, with increasing concentrations of the cold oligo (Figure 3-19b) as determined by the increase in signal intensity of un-bound oligo.

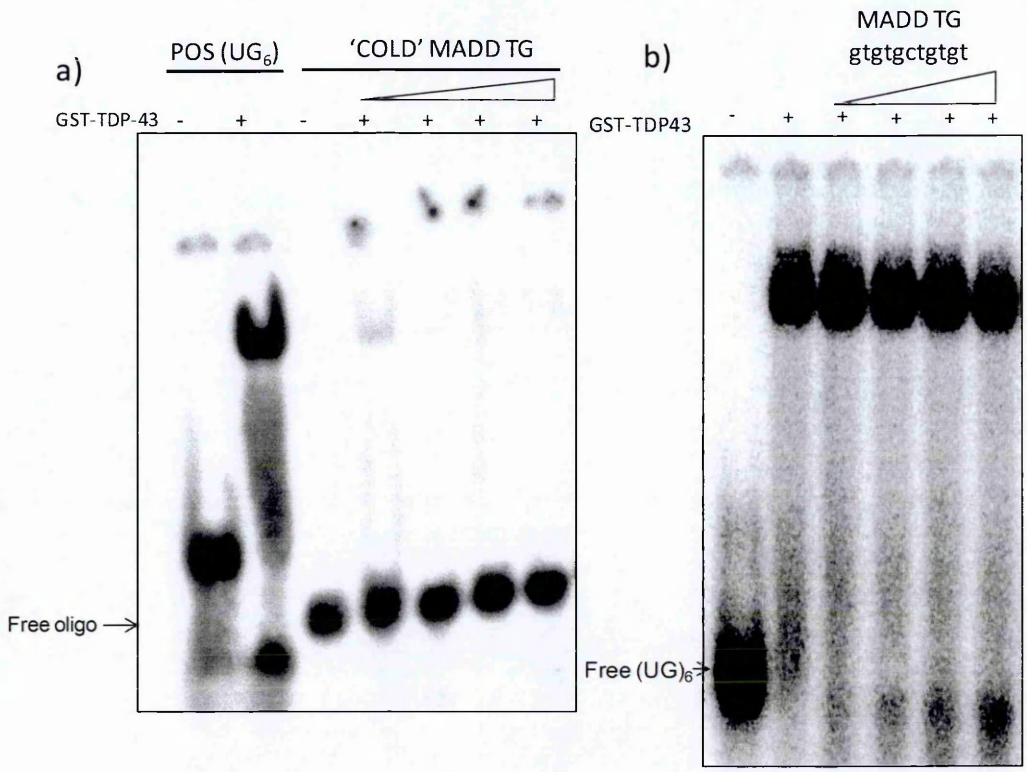



Figure 3-19: **Cold competition EMSA analysis confirms TDP-43 binding to MADD TG stretch**. Cold competition EMSA with (a) labelled and un-labelled MADD TG oligo showed an immediate decrease in signal intensity with increasing concentrations of the un-labelled oligo. (b) In addition, in competition EMSA with labelled $UG_6$, which is known to bind TDP-43, there was a gradual increase in displacement of un-bound $UG_6$ with increasing concentrations of un-labelled MADD TG oligo, further confirming that the TG sequence did indeed bind TDP-43.

Consequently, TDP-43 interaction with the alternatively spliced MADD exon 31 was mapped to a stretch of TG repeat sequences in the upstream intronic region of MADD exon 31, using a mutation analysis which revealed that the TG stretch upstream of MADD exon 31 was involved in aiding recognition of this exon, and an EMSA analysis that confirmed direct TDP-43 binding to this sequence.

### 3.2.4. *STAG2: Characterization of TDP-43 dependent alternative splicing of exon 30b*

Another gene observed to undergo TDP-43 dependent alternative splicing was *STAG2*, in which exon 30b was observed to undergo increased inclusion upon depletion of TDP-43. This increase in inclusion of *STAG2* exon 30b could similarly be reversed with overexpression of si-resistant wild type FLAG-TDP-43 as depicted in lane 3, Figure 3-11a. Furthermore, analysis of STAG2 exon 30b TDP-43 dependent alternative splicing profile in neuron-derived cell lines (SK-N-AS/SK-SY-5Y) (Figure 3-12a) revealed splicing profiles similar to those observed in the HEK 293 cells.

Analysis of the STAG2 protein isoforms by immunodetection showed that this variation in mRNA levels could also be observed at the protein level (Figure 3-20). Indeed, the increased inclusion of exon 30b, upon depletion of TDP-43 resulted in the appearance of an extra protein band, which was not visible in the control (Figure 3-20). As with the mRNA (Figure 3-11a), overexpression of wild type FLAG-TDP-43 but not mutant FLAG-TDP-43 (F4L) in the endogenous TDP-43 depleted background, reverted the ratio of both protein bands (compare lanes 3 and 4, Figure 3-20) to levels comparable to the ratio observed in the splicing profile of the endogenous transcript under the same conditions of TDP-43 expression in the cells.


Figure 3-20: **TDP-43 dependent alternative splicing of STAG2 in HEK 293 cells can be detected at the protein level.** Immunodetection of STAG2 detected changes in protein isoforms similar to those observed for the mRNA (see **Figure 3-11**a) that were dependent on the levels of TDP-43. At the protein level depletion of TDP-43 led to the appearance of an extra band that was abolished when wild type FLAG-TDP-43 was overexpressed. Expressing mutant F4L FLAG-TDP-43 resulted in a similar protein isoform expression as that observed in TDP-43 depleted cells. Cells were treated with siTDP-43 and tetracycline to provide a uniform background. Tubulin was used as a protein loading control. It should be noted that the STAG2 antibody is specific for total STAG2 and does not cross-react with other STAG (STAG1/STAG3) proteins.

As TDP-43 dependent alternative splicing of *STAG2* exon 30b was visible both at the RNA and protein level, further characterization was performed to map the region in which TDP-43 was binding. As an initial step towards the characterization of the TDP-43 interaction with exon 30b of STAG2, it was necessary to determine whether the endogenous alternative splicing profile was reproducible in a heterologous context.

#### 3.2.4.1. *Analysis of STAG2 exon 30b alternative splicing in a heterologous context*

In order to identify which sequences proximal to exon 30b of *STAG2* bound TDP-43 to elicit a preferential exclusion (under physiological levels of TDP-43), I constructed a hybrid pTB minigene consisting of only exon 30b, and subsequently both exons 30 and 30b. However, following transfection and RT-PCR, both minigenes could not recapitulate the endogenous splicing profile, and in both cases exon 30b was not recognized and excluded (Appendix I-6). A possible explanation for this observation may be that in many

minigenes, for successful reproduction of splicing patterns observed in the endogenous transcript, it is necessary to maintain as much as is possible of the original genomic environment. RT-PCR analyses from these minigenes can be found in the Appendix Figure I-6. Thus, a three-exon minigene (Figure 3-21a) was constructed that included exons 30, 30b and 31 together with significant portions of intronic sequences, as depicted in the schematic representation, cloned into a pcDNA3 vector backbone as described in section 2.2.4 of materials and methods.

Transient transfection in HEK 293 cells followed by RT-PCR analysis of the pcDNA-*STAG2*-Ex 30-30b-31 minigene showed that although exon 30b (seen as a band of 461 bp) inclusion was much higher than that observed in the endogenous transcript (compare lanes 1 and 2 Figure 3-21b and c) the knockdown of TDP-43 recapitulated the endogenous scenario with an increase in the inclusion of the exon. Indeed, TDP-43 depletion resulted in the 350 bp band corresponding to exon exclusion to decrease in intensity from 10%-2.8% (Figure 3-21b) signifying an increase in inclusion of exon 30b. In the endogenous transcript (Figure 3-21c), depletion of TDP-43 resulted in a 23% increase (19%-42%) of the transcript including exon 30b. Overexpression of TDP-43 resulted in a 6% (7.8%-13.8%) increase in exclusion of exon 30b in the pcDNA-*STAG2*-Ex 30-30b-31 minigene (Figure 3-21d) as determined by quantification of the lower 350 bp band, indicating that TDP-43 promotes exclusion of exon 30b. An increase in exclusion of exon 30b was also observed in the endogenous transcript (Figure 3-21e), where overexpression of wild type FLAG-TDP resulted in lower (19%-11%) levels of the upper 506 bp band that included exon 30b.

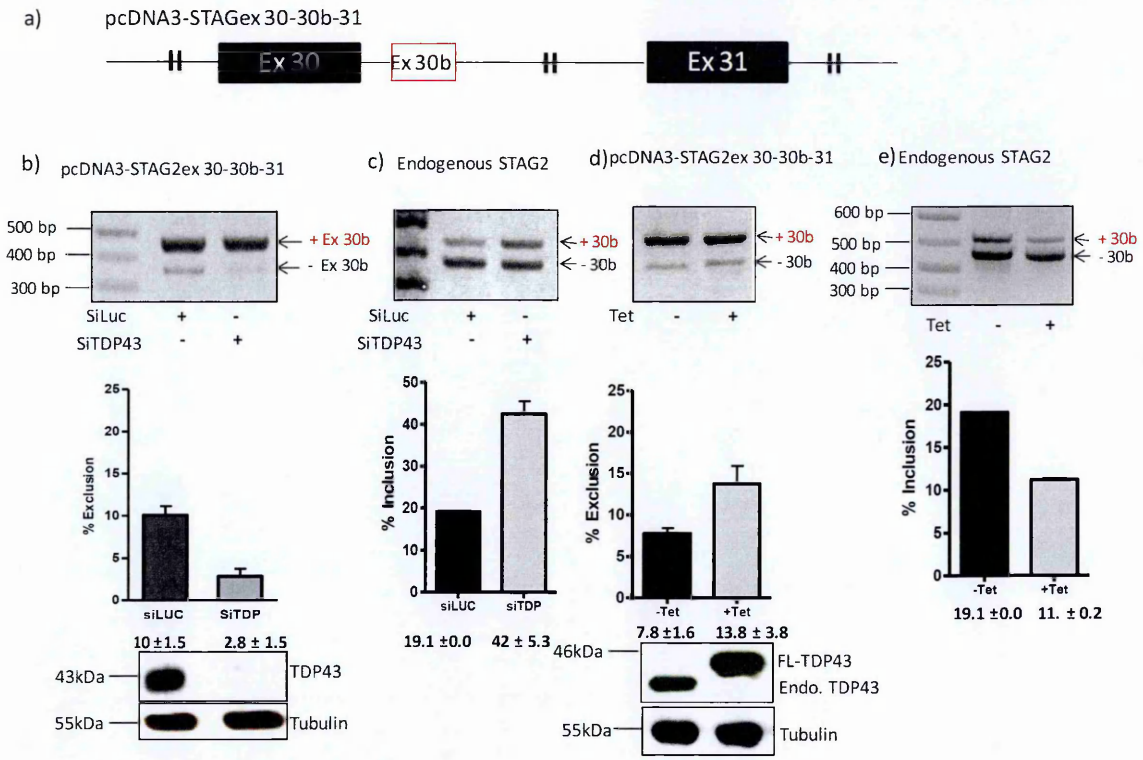



**Figure 3-21: Depletion of TDP-43 results in an increased inclusion of exon 30b.** (a) Schematic diagram of the *STAG2* sequence cloned into pcDNA3. (b) STAG2 RT-PCR gel images in different cellular contexts of TDP-43 levels. The three-exon minigene showed higher starting levels of included exon 30b and therefore levels of TDP-43 dependent inclusion of exon 30b were measured by estimating exclusion levels. On the other hand, the endogenous splicing ratio was referred to as percentage of exon inclusion. Quantification was performed using Image J program on three independent experiments. Depletion of TDP-43 in cells transfected with the three-exon-minigene led to a 7% increase in inclusion (10%-2.8%) of exon 30b. (c) In the endogenous transcript, TDP-43 dependent inclusion of exon 30b increased from 19%-42%. Overexpression of WT-TDP-43 resulted in higher exclusion levels of both the (d) minigene (7%-13%) and the (e) endogenous transcript (reflected in decreased inclusion from 19% to 11%) thereby determining a TDP-43 dependent alternative splicing of exon 30b. Western blot analysis using anti-TDP antibodies were used to verify depletion and overexpression of TDP-43 whereas anti-tubulin was used as a protein loading control.

The recapitulation of the endogenous STAG2 splicing profile in a TDP-43 dependent manner indicated that within the pcDNA-STAG2 Ex 30-30b-31 minigene, there were sequences that could bind TDP-43, thus necessitating further characterization.

### 3.2.4.2. *STAG2 EMSA and deletion constructs analyses*

In contrast to *MADD*, in which the strategy for mapping the TDP-43 binding site involved analysis of a TG stretch, the sequences (both exonic and intronic) surrounding exon 30b of *STAG2*, did not contain any obvious TG repeats. As such, in order to identify the TDP-43 binding site, the entire region spanning exon 30 to exon 31 that was cloned into pcDNA3

was divided into five fragments, that were subsequently used as probes in EMSA binding analyses (Figure 3-22). The fragments were created by PCR amplification and subsequent products used as templates for *in vitro* RNA transcription as described in section 2.1.1.2 (materials and methods).

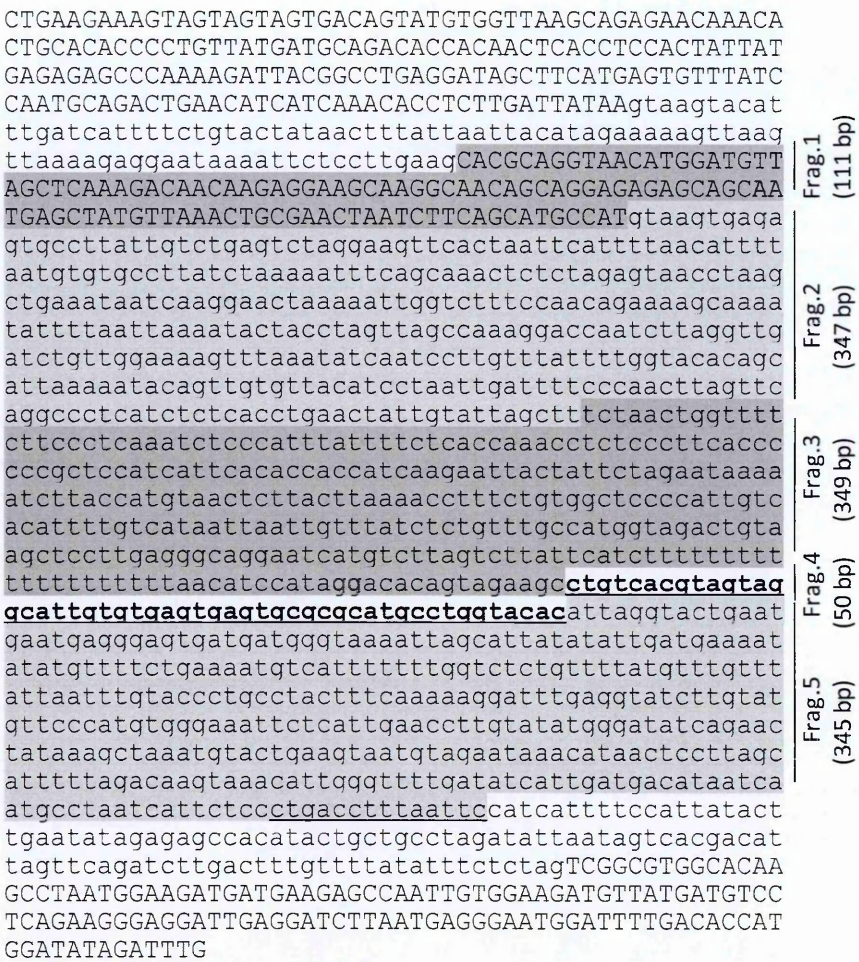



**Figure 3-22: STAG2 fragmentation for EMSA analysis with the various fragments highlighted in greyscale.** The shortest fragment (4), highlighted in bold underlined letters was obtained from part of the sequence in fragment 3 and has not been highlighted for visibility.

In order to determine which fragments (Figure 3-22) were able to bind TDP-43, cold competition EMSA analyses using cold *STAG2* RNA fragments were performed against a known binder of TDP-43, the $UG_6$ oligo, as previously described in the MADD EMSA analysis. The competition EMSA revealed TDP-43 binding in fragments three and five (Figure 3-23, red rectangles) as determined by an increase in intensity of free (un-bound) probe, which indicated increased displacement of the $UG_6$ oligo.

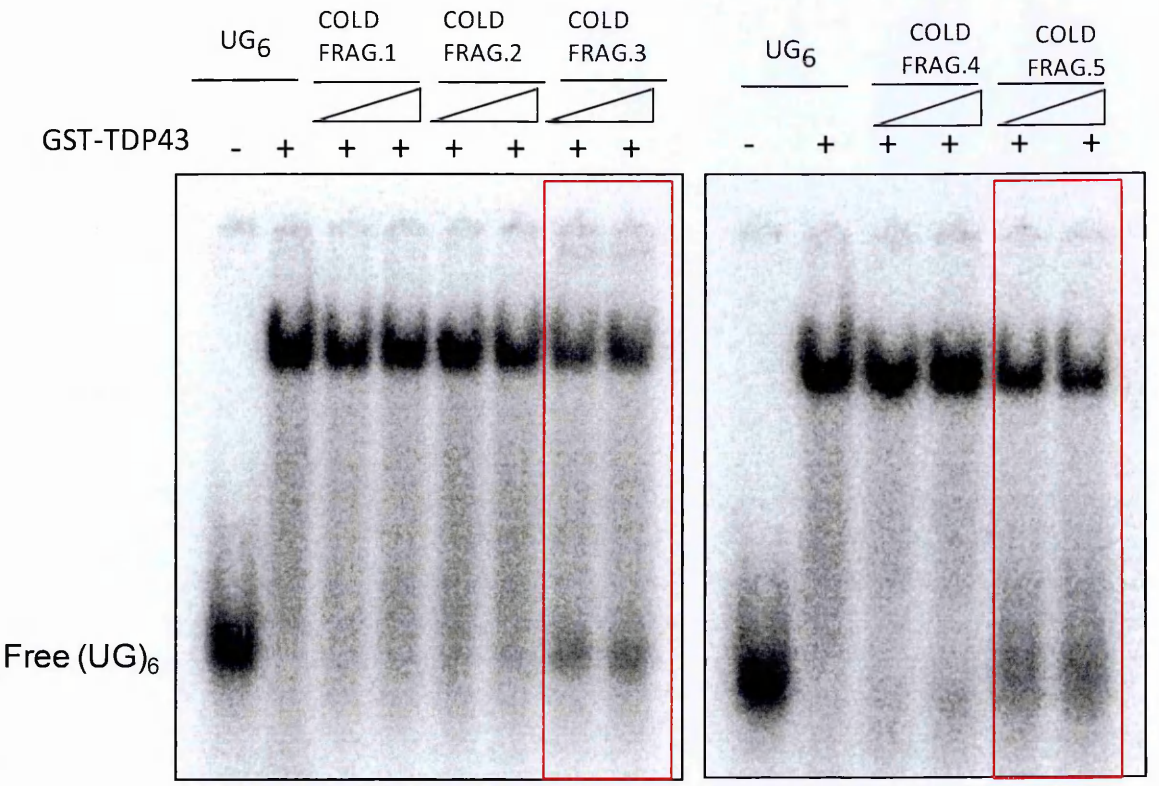



Figure 3-23: **STAG2 fragments three and five bind to TDP-43.** Cold competition EMSA against $UG_6$ showed increased displacement for fragments three and five (red rectangles) as a consequence of binding to TDP-43.

To further validate the capacity of fragments 3 and 5 to bind TDP-43 EMSA analysis was performed using radioactively labelled fragments three and five and recombinant GST-TDP-43. As can be seen in Figure 3-24 a shift was only observed in fragment five. In this case, the EMSA showed an increase in intensity of the band shift signal with increasing amounts of TDP-43 as shown in Figure 3-23a. In addition, the band shift signal could be eliminated with 20-fold excess molar amounts of cold RNA fragment five (Figure 3-24b).

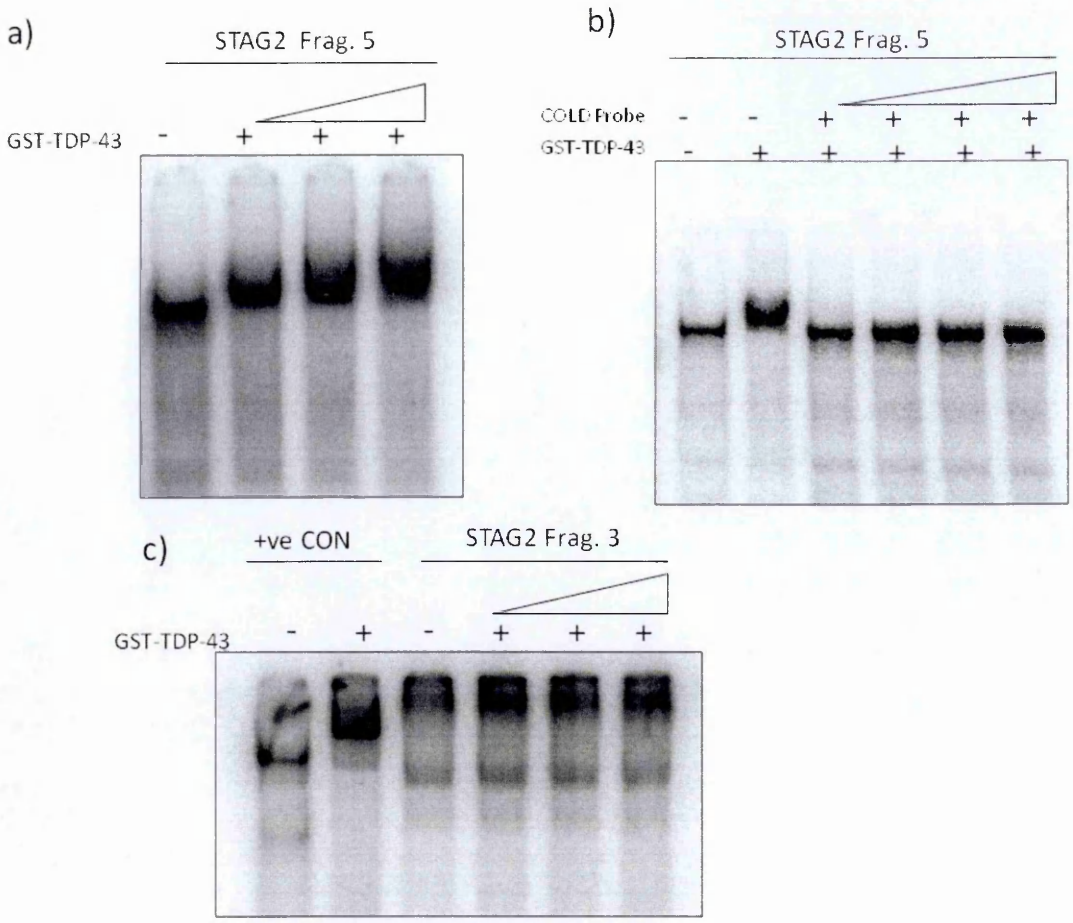



Figure 3-24: **Fragment five of the STAG2 three-exon minigene binds to TDP-43 but not to fragment three.** (a) EMSA performed with radio-labelled fragment five showed an increase in band shift intensity with increasing concentrations of TDP-43, and vice versa in a (b) competition EMSA performed with the cold fragment. (c) EMSA performed with radio-labelled fragment three in which no discernible band shift could be observed.

In contrast, for fragment three (Figure 3-24c), no distinct band shift was observed with increasing concentrations of TDP-43, despite the displacement observed previously in the cold competition EMSA against the $UG_6$ oligo.

To further investigate if fragments three and five played a role in the observed effect of TDP-43 levels on STAG2 exon 30b alternative splicing, a series of minigenes were constructed in which both these regions were deleted. These minigenes were then transfected and analysed by RT-PCR in HEK 293 cells under varying conditions of TDP-43 levels i.e. physiological, depletion or overexpression of TDP-43 (Figure 3-25 & Figure 3-26).

RT-PCR analyses of the pcDNA3-*STAG2* minigene with fragment five deleted schematically depicted in Figure 3-25a , resulted in increased inclusion of exon 30b (Figure 3-25b) which was independent of the levels of TDP-43 in the cell. In fact, the increase in inclusion of exon 30b (as determined by the decrease in intensity of the lower band) of the *STAG2* deletion mutant, was similar (approximately 2%) to that of the WT minigene (2.6%) when TDP-43 was silenced or depleted in the cells. Furthermore, overexpressing TDP-43 did not result in increased exclusion (Figure 3-25c) with respect to the wild type minigene. Consequently, the above observations further indicated that fragment five was involved in binding TDP-43.

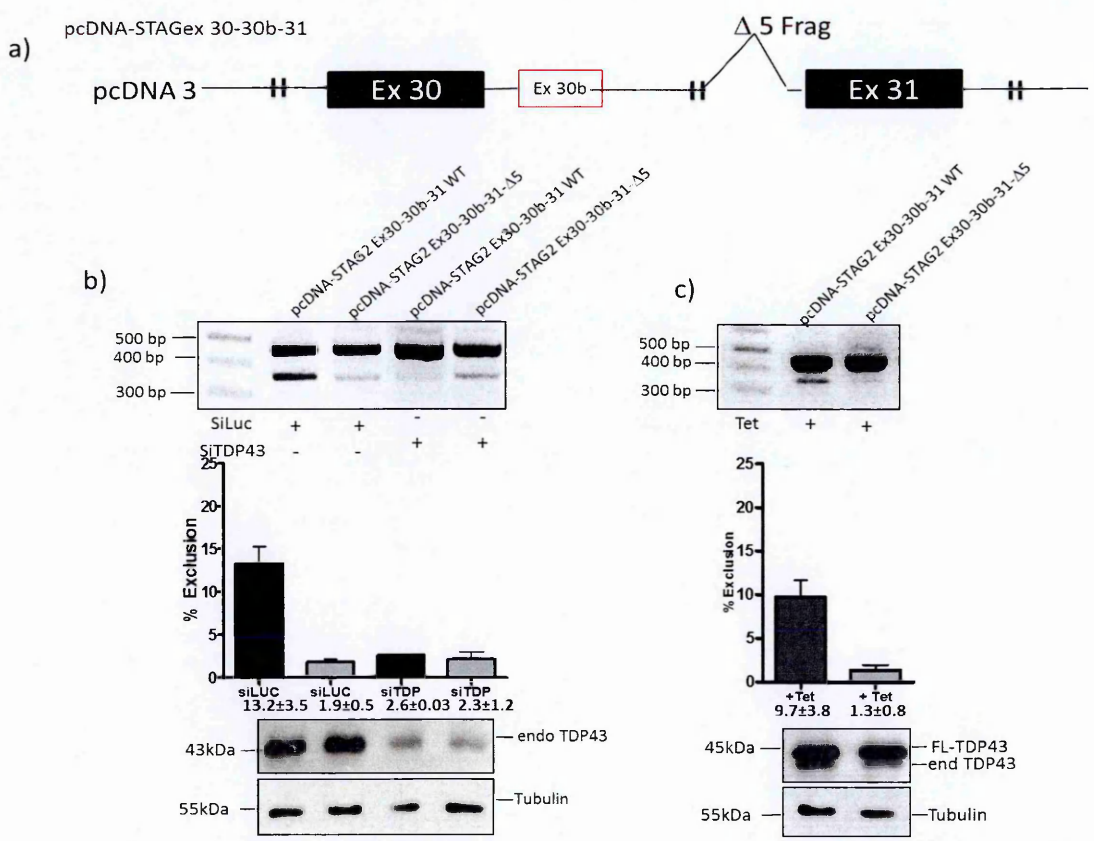



Figure 3-25: **Deleting fragment five from the STAG2 three-exon minigene results in increased inclusion of exon 30b that is not altered with varying TDP-43 levels.** (a) Schematic representation of the pcDNA-STAG2-Ex30-30b-31Δ5 minigene. (b) RT-PCR comparison analysis of the wild type and deletion mutant minigenes showed similar levels (2.6-2.3%) of exon 30b exclusion, when TDP-43 was depleted in the cells. Furthermore, analysis of the deletion mutant minigene showed increased inclusion of exon 30b that was independent of TDP-43 levels. (c) On the other hand, overexpressing TDP-43 did not result in increased exclusion of exon 30b, as there was no increase in intensity of the lower band in the deletion mutant *STAG2* minigene.

As EMSA analysis of fragment three was found to have binding capacity through displacement of the $UG_6$ oligo, but not with the direct EMSA analysis with labelled fragment three oligo, assessment of whether this fragment might also be playing a role in the inclusion of exon 30b was performed by constructing a minigene lacking fragments three to five (pcDNA-STAG2Ex 30-30b-31-Δ3-Δ5) (Figure 3-26).

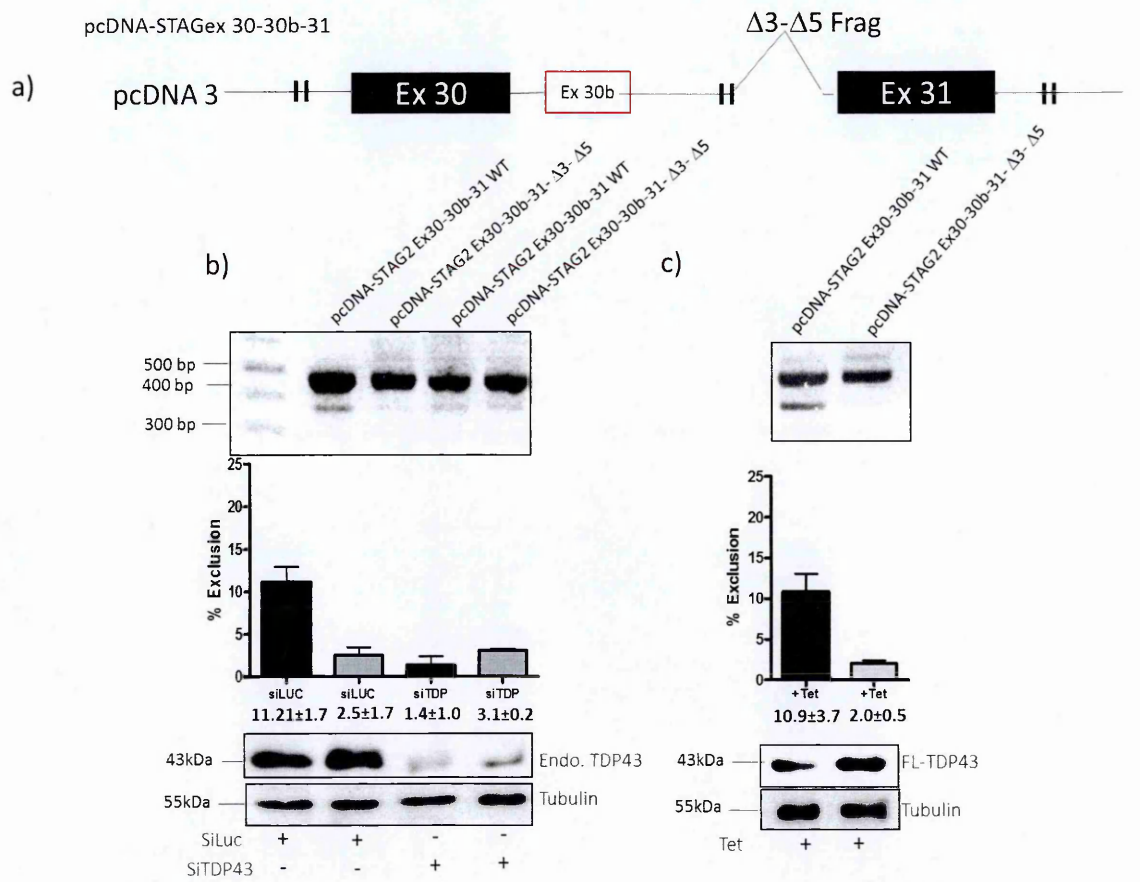



**Figure 3-26: Deletion of fragments three and five in the STAG2 three-exon minigene results in an increased inclusion of exon 30b.** (a) Schematic representation of pcDNA3-STAG2Δ3-Δ5 minigene depicting the deleted regions. (b) RT-PCR comparison analyses of wild type and deletion minigenes under phyisiological and depleted levels of TDP-43 show a constant increase in inclusion in the deletion minigene indicating that fragments three and five do play a role in the recognition of this exon (c) In fact, overexpressing TDP-43 did not result in increased exclusion of exon 30b, as there was no increase in intensity of the lower band in the deletion mutant pcDNA3-STAG2Δ3-Δ5 minigene. In the bottom panels, image J quantifications and verification of cellular TDP-43 levels by western blot analysis are shown. Tubulin was used as a loading control.

RT-PCR analysis of this double deletion minigene resulted in a further increase in inclusion of exon 30b, as determined by the decrease in intensity of the 350 bp band from 11%-1.2% compared to the wild type plasmid (Figure 3-26a). Comparison of the splicing

profiles between the pcDNA3-*STAG2*Δ5 and pcDNA3-*STAG2*Δ3-Δ5 results in a better inclusion of exon 30b for the minigene lacking fragment three and five suggesting that the ameliorated effect was conferred by the additional deletion of fragment 3. Notwithstanding the conflicting data regarding the ability of this fragment to bind TDP-43, and since both deleted fragments were quite large (Figure 3-23) with no typical TDP-43 binding sites (TG stretches) overtly visible, TDP-43 binding in this case may be linked to the secondary structure of the RNA. Similar levels of a decrease in band intensity were also observed in the wild type minigene (1.6%) when TDP-43 was depleted. Furthermore, overexpressing FLAG-TDP-43 wild type did not change the exclusion levels in the double deletion mutant minigene (Figure 3-26c). Analysis of a construct lacking only fragment three was not performed since direct binding could not be determined in EMSA, and since the overall difference in exclusion levels of exon 30b observed in these minigenes are quite small, it is likely that the contributory effect of fragment three was catered for in the minigene with both fragments deleted.

### 3.2.5. ***BRD8 and FNIP1: TDP-43 dependent alternative splicing of exons 20 and 7 respectively***

In addition to *MADD* and *STAG2*, *BRD8* and *FNIP1* were also found to undergo TDP-43 dependent splicing. Specifically for BRD8, exon 20 (45 bp) was found to undergo increased inclusion when TDP-43 was depleted, which could also be reverted with overexpression of wild type FLAG- TDP-43 but not with the mutant form of TDP-43 that is unable to bind RNA. At the protein level and unlike *MADD* and *STAG2*, western blot analyses could not sufficiently detect protein expression and was thus not useful in analysing altered splicing changes at the protein level.

Interestingly, within both the upstream and downstream intronic sequences surrounding exon 20 of BRD8, TG stretches were present, with the downstream intron containing the

longest stretch Figure 3-27a. As had been performed for the other genes, to enable the characterisation of the interaction between TDP-43 and exon 20 in *BRD8*, it was first necessary to determine whether the endogenous splicing was reproducible in a heterologous context. Consequently, a region consisting of exon 20 together with 250 base pairs each of upstream and downstream intronic sequence was cloned into the pTB hybrid backbone (Figure 3-27a) and analysed by RT-PCR analysis following transient transfection in HEK 293 cells.

RT-PCR analysis of the pTB-*BRD8* minigene in HEK 293 cells revealed an increase in inclusion of exon 20 (Figure 3-27b) when TDP-43 was depleted. This increase in inclusion, indicated by an approximately 50% (7%-4%) decrease in the intensity of the lower band was similar, (although to a lower extent) to that observed in the endogenous *BRD8* transcript when TDP-43 was depleted. Similarly, overexpressing TDP-43 resulted in increased exclusion of exon 20, in both the minigene and endogenous transcript (Figure 3-27d-e). However, in the endogenous transcript, the recovery was more pronounced (approximately 80%) (Figure 3-27e) compared to the profile observed with the minigenes. This could be due to the insufficient inclusion of *BRD8* exon 20 flanking sequences cloned in the minigene. It is also likely that in this minigene, some other TDP-43 binding sites responsible for finer definition and recognition of *BRD8* exon 20, are not included.

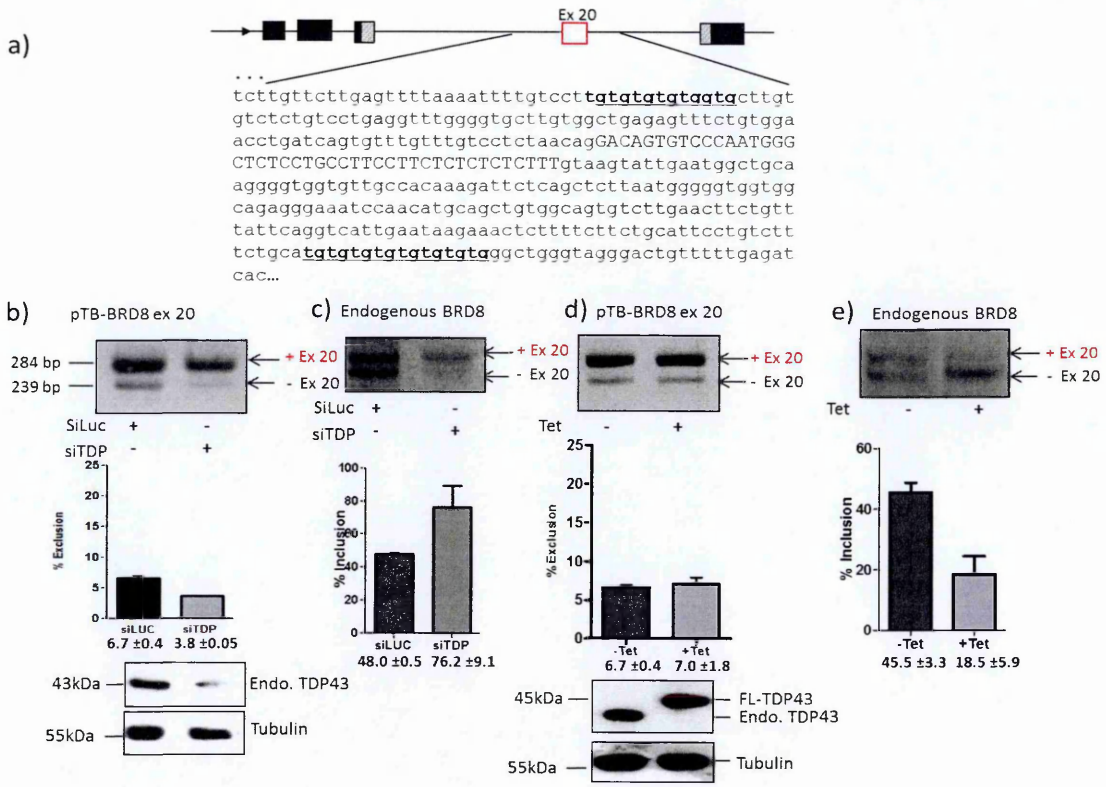



**Figure 3-27: Depletion of TDP-43 results in the increased inclusion of BRD8 exon 20 that can be rescued with overexpression of WT TDP-43.** (a) Schematic diagram of the pTB-*BRD8* minigene construct. The α-globin, fibronectin EDB and human *BRD8* exon 20 are shown as black, grey and red boxes respectively. Putative TDP-43 binding sites are highlighted in bold and underlined. (b) RT-PCR analysis showing increased inclusion of exon 20 (50% (7-4%) decrease in lower exon) following depletion of TDP-43, resembling the (c) endogenous splicing profile of BRD8 showing increased inclusion of exon 20 upon TDP-43 depletion (d) Overexpression of TDP-43 results in exclusion of exon 20, reverting the splicing profile i.e similar to the profile observed when there are normal levels of TDP-43. (e) In the endogenous transcript, a more pronounced (ratio 80:20%) exclusion of exon 20 is visible.Western blots detecting anti-TDP-43 were performed to verify physiological levels of TDP-43, depletion and overexpression of the transgenic WT TDP-43. Anti-tubulin was used as a loading control

Thus, future work in characterising TDP-43 dependent exon 20 inclusion in *BRD8* will include mutagenesis or deletion of the observed TG stretches to distinguish which sequence interacts with TDP-43.

Similarly, the *FNIP1* gene was analysed for TDP-43 dependent alteration in the splicing of exon 7 in a heterologous context. Typically, exon 7 (84 bp) together with a 241 bp of upstream and 251 of downstream intronic sequences were cloned into the pTB hybrid minigene in an effort to reproduce the alternative splicing profile observed in the endogenous transcript. Following transient transfection, RT-PCR analysis of this minigene

showed a single band of 239 bp representative of exon 7 exclusion, which was independent of the levels of TDP-43 in the cell (Figure 3-28b).

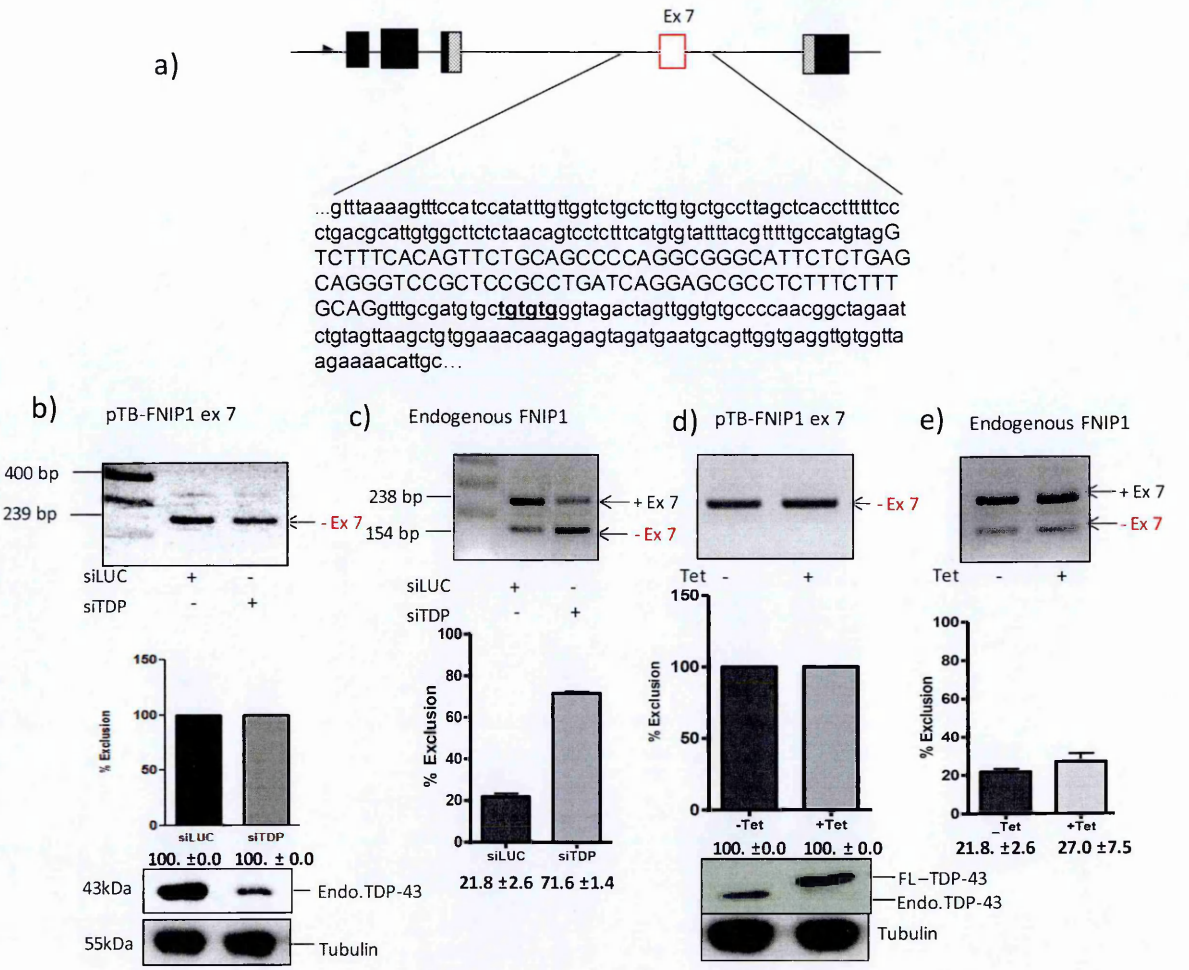



**Figure 3-28: *FNIP1* Exon 7 is not recognised in the pTB-FNIP1 minigene.** (a) Schematic diagram of the pTB-FNIP1 minigene, depicting sequences cloned into the minigene. α-globin, fibronectin and human *FNIP1* exon 7 are shown as black, grey and red boxes, respectively. Putative TDP-43 binding sites are highlighted in bold and underlined.(b) RT-PCR analysis of the minigene depicted 100% exclusion of exon 7 (239 bp) in the minigene in both siLUC and siTDP compared to the (c) endogenous wherein depletion of TDP-43 resulted in 50% increase. (d) Overexpressing WT-TDP-43 did not change the splicing profile of the minigene, contrary to the endogenous transcript wherein there was a rescue of exclusion to increased inclusion i.e similar levels (27%) to siLUC. Verification of TDP-43 levels in the cells were analysed using anti-TDP-43 western blots anti tubulin as a loading control.

In contrast, in the endogenous transcript, depletion of TDP-43 resulted in increased exclusion of TDP-43 (approximately 71%). Overexpressing TDP-43 did not have any effect on the pTB-*FNIP1* splicing profile i.e. 100% exclusion of *FNIP1* exon 7 compared to the endogenous transcript, wherein there was a rescue to increased inclusion of exon 7 to similar levels (decrease in exclusion band from 71% to 27%) as observed in siLUC (normal levels of TDP-43).

An in depth analysis of the flanking intronic sequences, identified a short six base pair TG stretch (3 TG repeats) (Figure 3-28a) in the downstream intron that were seemingly not sufficient to bind TDP-43 in the context of the minigene. This corresponds with previous work, in which a minimum of six repeats was found to be required for TDP-43 binding (Buratti and Baralle, 2001). Thus, similar to the *STAG2* gene, the reproduction of the TDP-43 dependent endogenous splicing of *FNIP1* exon 7 requires a broader context, which will be addressed in future work. Further investigation of whether this altered splicing profile is also translatable to the protein level will also be assessed.

## 3.3. Analysis of altered splicing of the MADD, STAG2, BRD8 and FNIP1 genes using a TDP-43 loss of function model.

Currently, the two main schools of thought on the pathogenesis of TDP-43 are centred on the loss of function or gain of function hypotheses. As described previously, in patients, TDP-43 is reported to be sequestered in the cytoplasm and is the main component of the prototypical intra-cellular inclusions or aggregates (see chapter 1; section 1.2).

Considering the involvement of TDP-43 in neuropathology it was of interest to investigate whether the same alterations found in the stable HEK-293 cell lines could also be observed in a cellular aggregation model that results in TDP-43 loss-of-function developed in our laboratory (Budini, Romano et al. 2014) . Briefly, this system is based on the expression of 12 tandem repeats of the Glutamine/Asparagine (Q/N) rich region of TDP-43 (residues 342-366) fused to the full length TDP-43 protein sequence tagged with FLAG. This transgene was stably transfected into HEK 293 cells using the Flp-In T-Rex recombination system. Expression of this effector following tetracycline induction results the formation of aggregates that co-localize with endogenous TDP-43 Figure 3-29a. In addition, cells expressing a mutant form of TDP-43, with the four phenylalanines (F4L) in RRMs 1 and 2 mutated, has also been constructed. The F4L mutations render TDP-43 unable to bind

RNA, thus allowing the discernment of whether the loss of function is a result of aggregation or RNA binding capacity of the protein. As can be seen in Figure 3-29, expression of this mutant form of TDP-43 fused to the 12Q/N repeats, is also able to form aggregates that sequester the endogenous TDP-43.

The aggregates formed by expressing wild type TDP-43-12Q/N (Figure 3-29a) appeared to be smaller and diffuse in nature whereas those formed by mutant (F4L) TDP-43-12Q/N appeared bigger and more localised (Figure 3-29b). Since the only difference between the two forms of TDP-43 is the mutation of the four phenylalanines of the mutant, it could point towards the involvement of the RNA binding capacity of the protein in aggregates formation.

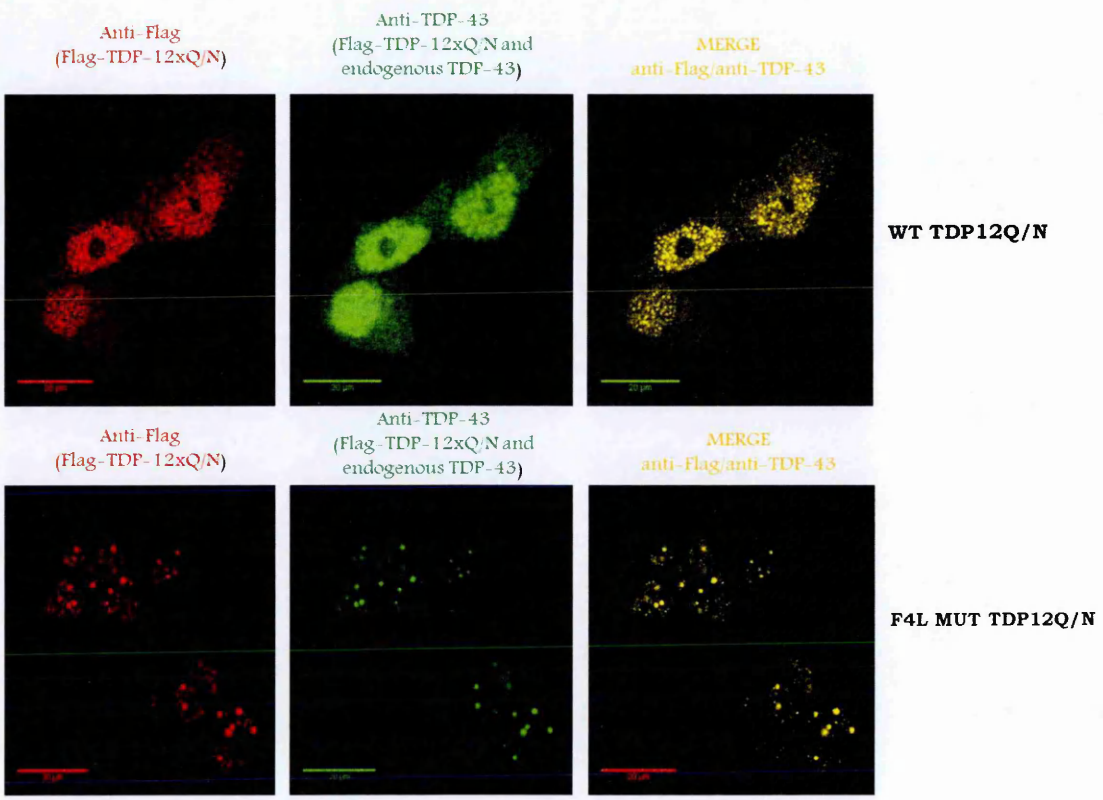



Figure 3-29: **Induced expression of transgenic TDP-43-12XQ/N results in aggregate formation that co-localizes with endogenous TDP-43.** Expression of aggregate inducing transgenic TDP-4312XQ/N was detected using immunoflourescence after 72 hr induction. Flag-tagged TDP-43 was labelled with Alexa (Red) and endogenous TDP-43 labelled with FITC (green). In the upper panel, aggregates formed from expression of (a) WT FL-TDP-12XQ/N are shown whereas those formed by expression of the mutant form (b) F4L TDP-4312XQ/N are shown in the lower panel. Scale bar 20μm.

Taking advantage of these cellular aggregate models, RT-PCR assays were used to analyse the genes identified from the splice-sensitive microarray study, to determine whether the sequestration of TDP-43 within the aggregates had a functional effect on the RNA targets of TDP-43.

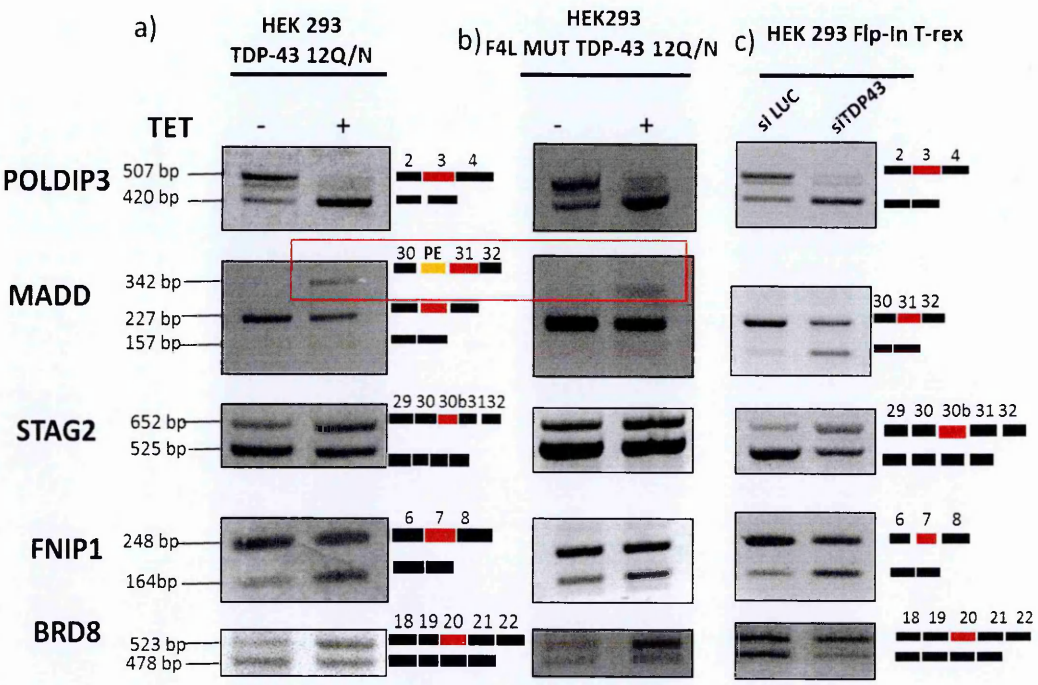



Figure 3-30: **The TDP-43 cellular aggregate model has a similar effect to silencing TDP-43 on genes identified from the Affymetrix analysis**. RT-PCR of endogenous transcripts analysed in the (a) WT12XQ/N and (b) mutant (F4L12XQ/N) aggregation models compared to (c) cells in which TDP-43 was depleted. The red rectangle highlights the pseudoexon in both WT and mutant F4L aggregation model cells and which was also detected in neuronal ells (See figure 3-12b).

Specifically, for the genes validated from the Affymetrix study, the TDP-43 dependent increase in inclusion of exons 30b and 20 in *STAG2* and *BRD8* was observed, as well as exclusion of exon 3 in *POLDIP3* and exon 7 in *FNIP1* (Figure 3-31). However, for the *MADD* alternative splicing profile, in both the wild type and mutant cellular aggregate models, apart from the exclusion of exon 31, an extra band (Figure 3-30 a&b, red rectangle) above the normally observed 227 bp band representative of the inclusion of exons 30, 31 and 32 was observed. This extra band that was 115 bp larger had not been

observed previously in HEK293 cells depleted of TDP-43, but was present in the SH-SY-5Y and SK-N-AS cell lines (Figure 3-12c).

As previously determined for *MADD* and *STAG2* (Figure 3-13 & Figure 3-20), TDP-43 dependent alternative splicing profiles were analysed at the protein level in these cells. Immunodetection of STAG2 and MADD was performed in un-induced and induced (plus tetracycline) cells (Figure 3-31).

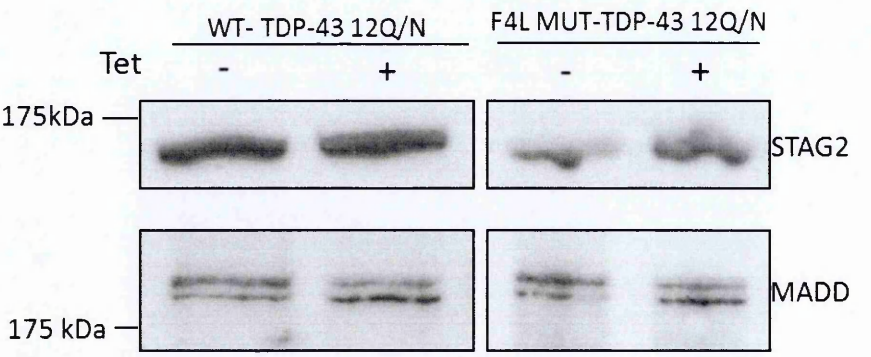



Figure 3-31: **TDP-43 dependent alternative splicing in STAG2 and MADD is detectable at the protein level in the TDP-43 aggregate model.** Antibodies against STAG2 (155 kDa) and MADD (200 kDa) western blots in (-Tet) and (+Tet) induced cellular aggregate models with WT TDP-43-12XQ/N and F4L TDP-43-12XQ/N, showing the appearance of an extra band upon aggregate induction.

Western blot analyses showed changes in the ratio of the proteins in the lanes with induced (+ Tet) samples as had previously been observed. For the STAG2 protein, depletion or loss of TDP-43 results in increased presence of the larger protein isoform in both cell lines correlating with the increase in the inclusion of exon 30b observed at the RNA level and the density ratio (Figure 3-31).

Similarly, for the MADD protein, two isoforms were visible in both induced and un-induced cells. In un-induced WT-TDP-4312XQ/N cells, the upper band appeared to be denser, and upon induction, there was a density shift to the lower band having slightly higher intensity. The same shift in density was also observed in the F4L-TDP-43-12XQ/N cells. Protein analyses of both STAG2 and MADD in the TDP-43 cellular aggregate

models reflect those observed in HEK 293 cells that were transiently silenced using siTDP-43.

Consequently, the TDP-43 cellular aggregate models confirmed TDP-43 dependent alteration in splicing by recapitulating a 'knock-down' effect in the genes identified from the Affymetrix microarray study. In the cellular aggregate models, TDP-43 was shown to be sequestered within aggregates and was no longer able to perform its nucleo-cytoplasmic functions, thereby supporting a role for the loss of function hypothesis in TDP-43 proteinopathies. In addition, the similarity in alterations of splicing profiles observed in both wild-type and mutant TDP-43 models emphasizes the significance of aggregate formation rather than the RNA binding capacity of the protein in pathogenesis.

### 3.3.1. *MADD alternative splicing in the TDP-43 cellular aggregate model*

As described previously (results section 3.3), RT-PCR analyses of the *MADD* splicing profile in the cellular aggregate models revealed an extra band approximately 115 bp higher (342 bp) than the all-inclusive (including exons 30,31 and 32) band of 227 bp. A comparative analysis of the *MADD* splicing profile in HEK 293 cells depleted of TDP-43 and in the TDP-43 aggregate cell lines, revealed this band to be present only in the aggregate cell lines, suggesting a link with the aggregate formation.

To better understand the reasons for the appearance of this extra amplicon, this band was extracted from the gel and sequenced. This showed that it resulted from the inclusion of a pseudo exon, located in the intronic region between exons 30 and 31, 340 bp downstream of exon 30 (Figure 3-32b). An *in silico* splice site strength analysis using MaxEntScan for both the 5'and 3' splice sites of the pseudoexon scored the 5' splice site at 10.28 and 3'splice site at 10.0 (Yeo & Burge 2003). Interestingly, from the RT-PCR comparative analysis in the aggregate cell lines, there was a preferential inclusion of the pseudo exon

rather than the exclusion of exon 31 (157 bp), since the presence of the lower band signifying the exclusion of exon 31 was very faint (Figure 3-32c). Furthermore, analysis of the *MADD* exon 31 splicing profile in a preceding aggregate model i.e cells expressing GFP12XQ/N upon tetracycline induction, did not detect an extra band (Figure 3-32d). In fact, in these cells, only the un-spliced band (exons 30, 31 and 32) of 227 bp is observed, independent of the presence or absence of aggregates (Figure 3-32d).

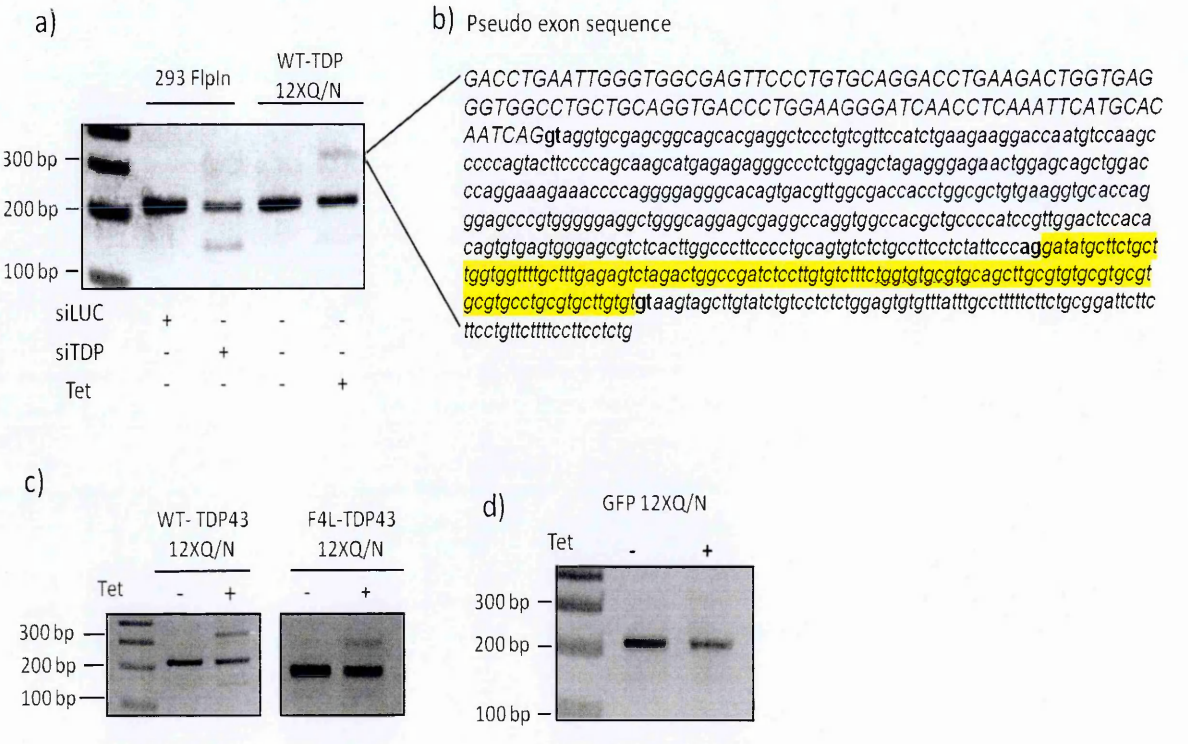



Figure 3-32: **The inclusion of a pseudo exon in the MADD alternative splicing profile in TDP-43 linked aggregates.** Comparative RT-PCR analysis of the *MADD* splicing profile in 'empty' HEK 293 cells and HEK 293 cells stably expressing TDP-12XQ/N. (b) The sequence shown in yellow represents the pseudo exon band in (a). (c) RT-PCR analysis of the MADD splicing profile in HEK293 cells inducibly expressing wild type TDP-43 12XQ/N (left panel) and mutant F4L TDP-43-12XQ/N (right panel). (d) MADD splicing profile in a prototypical aggregation model, in which no changes were observed.

Similar to the TDP-43-12XQ/N aggregate model, cells expressing GFP-12XQ/N cells form aggregates, with some co-localization with the endogenous TDP-43 protein observed (Budini et al. 2012), and as such, not sufficient to elicit a functional effect. Therefore, as expected, there was no alteration in the MADD exon 31 splicing profile, which indicated that the inclusion of the pseudo-exon previously observed was not an artefact of

aggregation caused by the 12Q/N repeats, but rather linked to the combination of full length TDP-43 and aggregates in the TDP-4312XQ/N aggregate models.

Further analysis is required to try and understand whether this differential MADD splicing profile between TDP-43 depleted cells and cells with the TDP-43 12XQ/N aggregates is due to a finer depletion of the protein from the cellular environment or due to the sequestration of other yet un-identified factors. Since no appearance of the pseudoexon was observed in HEK293 cells following treatment with two different and very efficient siRNAs against TDP-43 (data not shown), the second hypothesis described seems to be the most likely. Under normal conditions, the un-identified factors together with TDP-43 could be playing a role in the definition of the pseudo exons. Furthermore, since the pseudo exon was also detectable in the neuronal cell lines (Figure 3-12b), this may indicate that these yet un-identified sequestered factors could be naturally less abundant in these cells.

# 4. DISCUSSION

The reclassification of amyotrophic lateral sclerosis (ALS) and a sub-set of fronto-temporal lobar degeneration (FTLD-U) as TDP-43 proteinopathies, and the subsequent discovery of TDP-43 in other neurodegenerative diseases (although to a lower extent) such as Alzheimer's and hippocampal sclerosis, revolutionised the field in terms of understanding the pathological mechanisms underlying these diseases (Da Cruz & Cleveland 2011; King et al. 2012). In ALS specifically, over 90% of the accumulated protein observed within the aggregates is TDP-43 (Ling et al. 2013), which presents a unique opportunity for gaining insight into pathological mechanisms, without other confounding factors e.g. the presence of other proteins. In addition, it has become evident that RNA-binding proteins (RBP) are central to disease pathogenesis in cases such as Spinal Muscular Atrophy (SMA), Fragile-X Syndrome (FXS) and Spinocerebellar Ataxia type II (SCAII) amongst others, and given the cellular roles played by these proteins, current pathogenesis hypotheses have converged on altered RNA metabolism as a key process in disease (King et al. 2012; Hanson & Tibbetts 2012).

In keeping with this view, efforts aimed at elucidating pathological mechanisms involving TDP-43 have confirmed that alterations in TDP-43 expression levels, both in a positive and a negative manner, can lead to several changes in the general RNA and protein expression profile (Buratti et al. 2010; Tollervey et al. 2011; Fiesel et al. 2012). Although many of these reported events/targets have shown a high degree of variability, depending on which tissue/cells were analyzed (Buratti et al. 2013), they have all supported the hypothesis that defects at the RNA processing level, may considerably explain the role played by TDP-43 in disease (Polymenidou, Lagier-Tourenne, Hutt, et al. 2012; Ramaswami et al. 2013). Still lacking, however, is a clear understanding of which of these changes are directly connected with TDP-43, as opposed to simply reflecting secondary changes due to a general dysregulation. Consequently, one of the main issues in current pathogenesis research

efforts is to determine the major RNA metabolism altering events that are directly regulated by TDP-43 and whether these events, could similarly be affected at the neuronal level in addition to aggregation. At present, a well described example of a target of TDP-43 in neuronal and non-neuronal cells that is modified according to TDP-43 expression levels in a significant manner is the SKAR/POLDIP3 gene (Shiga et al. 2012; Fiesel et al. 2012). Work emanating from these studies has shown that a knockdown of TDP-43 can significantly shift the expression of SKAR-α to SKAR-β, which results in a more efficient activation of ribosomal protein S6 kinase 1 (S6K1), and consequently an increase in global protein translation. Thereby, suggesting that when TDP-43 is sequestered within insoluble aggregates, a general alteration in protein homeostasis may also contribute to disease pathogenesis.

Nonetheless, although TDP-43 has broadly been linked to the altered splicing of several genes, the full extent or consequence of TDP-43 expression or lack thereof, in these genes is still not known. In addition, the majority of current research is skewed towards the role of TDP-43 in relation to altered splicing profiles and only few studies have focused on protein-protein interactions or changes in protein expression that are affected by the levels of TDP-43. In fact, these few studies include the discovery that TDP-43 interacts with other hnRNPs (Buratti et al. 2005), regulates HDAC6 levels (Fiesel et al. 2010), and broadly, other splicing factors, as well as proteins involved in the translational machinery (Freibaum et al. 2011). Thus, given that TDP-43 interacts with several other proteins, a feature typical of other hnRNPs, it is likely that changes in the expression levels of TDP-43 can alter the expression of these interactors, directly or indirectly (Buratti & Baralle 2012).

In an effort to gain insight into cellular events that were directly regulated by TDP-43, this study utilised two main approaches, one at the proteomic and one at the transcriptomic level, incorporating 2-DE analyses and Affymetrix splice junction arrays, to determine TDP-43 dependent global changes.

## 4.1. Identification of differential protein expression linked to TDP-43

Two-dimensional electrophoresis (2-DE) analyses are well-known and established assays for separating proteins based on their isoelectric points and subsequently molecular weight. Coupled with protein stains for visualisation, differential protein expression can be determined based on spot intensity that has a linear correlation with the abundance of protein (Ong & Pandey 2001; Gauci et al. 2011). In this manner, 2-DE analyses provide an un-matched methodology for the separation and detection of complex proteins (Ong & Pandey 2001; Gauci et al. 2011). In this study, global differential protein expression that was dependent on TDP-43 expression levels was analysed using 2-DE, in four different cellular conditions consisting of a control group (siLUC), cells depleted of TDP-43 (siTDP-43), cells overexpressing transgenic wild-type FLAG-TDP-43 with the endogenous protein silenced and a fourth group expressing a mutant (F4L) transgenic form of FLAG-TDP-43 that cannot bind RNA with the endogenous protein silenced.

Following densitometric analyses that determined fold differences in spot intensities, a list of the most variable spots was obtained, which were selected for mass-spectrometry and subsequent secondary validation analyses (Table 2-1). The initial strategy for the secondary validation consisted of quantifying mRNA expression levels of genes identified from the mass spectrometry, by means of qPCRs, to determine if the observed expression changes correlated with the differential spot intensities at each of these spots. This strategy was further reinforced by the observation that current literature supports a role for TDP-43 in mRNA regulation through a variety of mechanisms, including alternative splicing and translation regulation (Emanuelle & Baralle 2010). In addition, a quick and simplistic analysis was performed by comparing transcript changes in control (siLUC) and FLAG-TDP-43 overexpressing cells.

To begin with, genes matched to peptides within spots 308 (*U2AF1 & MDH1*), 413 (*HSPA9 & CCT8*) and 421 (*EEF2*), were analysed for changes in transcript expression. This analysis revealed transcript expression changes that correlated with spot intensity, in spots 308 and 421. Interestingly, in spot 308, the two genes *U2AF1* and *MDH1*, with high-ranking matches from the mass spectrometry analysis, both exhibited an increase in transcript expression, under conditions of wild-type FLAG-TDP-43 overexpression. Both U2AF1 and MDH1 were particularly interesting candidates as U2AF1 is a known splicing factor that forms part of the U2-snRNP complex (Wahl et al. 2009) and MDH1 (a metabolic enzyme) was previously identified in two separate studies as being relatively up-regulated in the prefrontal cortex of FTLD-U patients (Martins-de-Souza et al. 2012), and motor neurons of presymptomatic and symptomatic mice with a *SOD-1* fALS mutation (Ferraiuolo et al. 2007). In spot 421, the gene *EEF2* was found to undergo a decrease in transcript expression that correlated with the decrease in spot intensity observed at this spot, when FLAG-TDP-43 was overexpressed.

For spot 421, the validation strategy involved analysis of the *EEF2* transcript in cells depleted of TDP-43 versus cells overexpressing TDP-43. Contrary to the previous qPCR analyses, *EEF2* was found to undergo an increase in expression, when FLAG-TDP-43 was overexpressed relative to when TDP-43 was depleted in cells. Despite the overall trend in differential spot intensity suggesting a direct correlation with TDP-43 levels, i.e. knockdown of TDP-43 resulted in an increase in protein expression (spot intensity) whereas overexpression resulted in a decrease in spot intensity, *EEF2* expression did not reflect these changes at the mRNA level. Thus, it was speculated that another gene, not identified by the mass spectrometry analysis may have effected the observed changes.

Secondary validation of spot 308 was performed using northern blot analysis, which ruled out U2AF1 as a candidate target of TDP-43, as no correlation to spot intensity was observed. In fact there appeared to be a decrease in transcript abundance (Figure 3-5). In

contrast, *MDH1* mRNA showed an increase in transcript abundance. However, this increase in transcript expression could not be detected at the protein level with western blots. In fact, MDH1 antibody detected two isoforms of the protein, presumably the native and phosphorylated forms, none of which corresponded to an increase in expression as determined by differential spot intensity analysis.

Taking into consideration the fact that no positive correlation had been observed with the genes previously identified, a mass-spectrometric re-analysis of spots 140, 300 and 1404 initially identified as undergoing significant variability, and for which no proteins had been identified (only albumin contaminants), was performed. In addition, previously analysed spots (308 and 323) were included to verify the reproducibility of the analysis. This re-analysis identified two new high-ranking peptides matching NASP and NAP1L1 in spots 140 and 300 respectively. Both NAP1L1 and NASP are nuclear proteins involved in cellular proliferation and have been reported in previous high-throughput analyses, reviewed in Buratti et al. (2013). In addition, within spot 308 and 1404, U2AF1/MDH1 and ALB were identified again as high-ranking matched peptides, which corroborated the previous mass spectrometry analysis. However in spot 323, in which peptides matching TDP-43 had previously been identified, UQCRC1 came up as high-ranking peptide. UQCRC1 had been identified in a proximal overlapping spot (323 Con) and may reflect a shift in this region, possibly due to run-run variation. The results obtained from the mass spectrometric re-analysis reflect an inherent inconsistency in the gel analysis that may have skewed the correct identification of spots and subsequent analyses.

As with the previous validation analyses, a positive correlation could not be observed for *NASP* mRNA analysed by qPCR or the NAP1L1 protein analysed by Western blot.

The lack of correlation between the 2-DE results and the validation analyses reflect potential errors that may have been inherent in the analytical processing of the gels.

Indeed, despite being a useful tool for global proteomic analyses that can relatively be automated by the use of analytical software, operator-dependent steps, such as gel-scanning and image warping present opportunities for errors (Ong & Pandey 2001; Berth et al. 2007). In addition, errors introduced by the limited sensitivity of the protein-staining dye Coomassie Brilliant Blue (CBB), may have resulted in inaccurate normalisation, and consequently inaccurate fold differences in spot intensities. Coomassie staining is reported to be less sensitive as a densitometric dye compared to its counterparts (Fluorescent dyes) (Gauci et al. 2011). This limitation was further supported by the observation that the Coomassie stain was only able to detect the transgenic wild-type TDP-43 and not the endogenous, which is a highly abundant protein. Given this observation, it is a distinct possibility that the normalisation of spots was inaccurate, as overly abundant protein (high concentrations) may also exceed the limits of the linear dynamic range (linear relationship between protein quantity and staining intensity) of the dye (Gauci et al. 2011). In this study, silver staining was not used as an alternative to Coommassie, as it is reportedly not compatible with downstream mass spectrometry analyses and has a low dynamic range (Gauci et al. 2011).

In addition, gel-gel variability introduced by the different runs, lysate compositions, or differences in resolution of the immobilized pH gradients (IPG) cannot be ruled out, notwithstanding that these differences were in part accounted for by triplicate gels of each specific cellular condition, and image analysis software. It could also be possible that in some cases, the effect of TDP-43 on target proteins was limited to translation level changes that could not be detected at the mRNA level. Nonetheless, since this would have involved the extensive use of various anti-bodies, the validation strategy was limited to changes that directly correlated at the mRNA and protein level.

Consequently, no meaningful correlation was obtained from the 2-DE analyses and in future, more sensitive methods, incorporating fluorescent dyes or pulse-labelling could be

employed to depict more accurate global proteomic changes that are dependent on TDP-43.

## 4.2. Altered mRNA splicing profiles of several genes are dependent on TDP-43

In an additional and complementary manner, analysis of global changes in gene expression that were dependent on levels of TDP-43 was performed using Affymetrix splice junction arrays. Aside from detecting general changes in transcript expression, splice-sensitive arrays are also able to detect alternative splicing events that occur under various cellular conditions. In this study, splice-sensitive arrays were particularly relevant, given the well-characterised role of TDP-43 in splicing (Buratti et al. 2001; Tollervey et al. 2011; Fiesel et al. 2012).

Consequently, several genes (2371) were shown to be influenced by TDP-43 levels in the cell, with the interesting observation that overexpression of TDP-43 uniquely affected a higher number (1099) of genes compared to the depletion (483) of TDP-43. To our knowledge this is the first study to include an analysis of the consequences of TDP-43 over-expression in cells (HEK 293) using microarray studies. Moreover, this observation may have been related to the fact that overexpression of TDP-43 in several disease models is reported to be toxic (Tsao et al. 2012; Romano et al. 2012) and more recently in cells (SH-SY5Y) (Yamashita et al. 2014). We focused on alternative splicing events that exhibited a clear correlation with TDP-43 cellular levels i.e. events that were altered upon TDP-43 depletion, could be rescued/reverted by overexpression of wild-type TDP-43, but not with the mutant (F4L) TDP-43, thereby signifying a direct consequence of TDP-43 RNA binding capacity. The number of genes that matched these criteria was 162. Interestingly, when compared to other studies that have examined the effects of a loss of TDP-43, there was some overlap, albeit little, in the number of genes identified (Figure 4-1), which appears to be representative of these kinds of screens. Furthermore, amongst

these four studies analysed, there was also very little or no overlap. This discrepancy seems to be a confounding trend, not only across microarray studies but in other high-throughput methodologies that analyse global cellular phenomena, including methods such as CLIP and RIP-Chip which examine single RBPs (Buratti et al. 2013). Several reasons exist for this observation, such as differences in tissues analysed and technical processing of samples, which also includes analytical methods (Buratti et al. 2013).

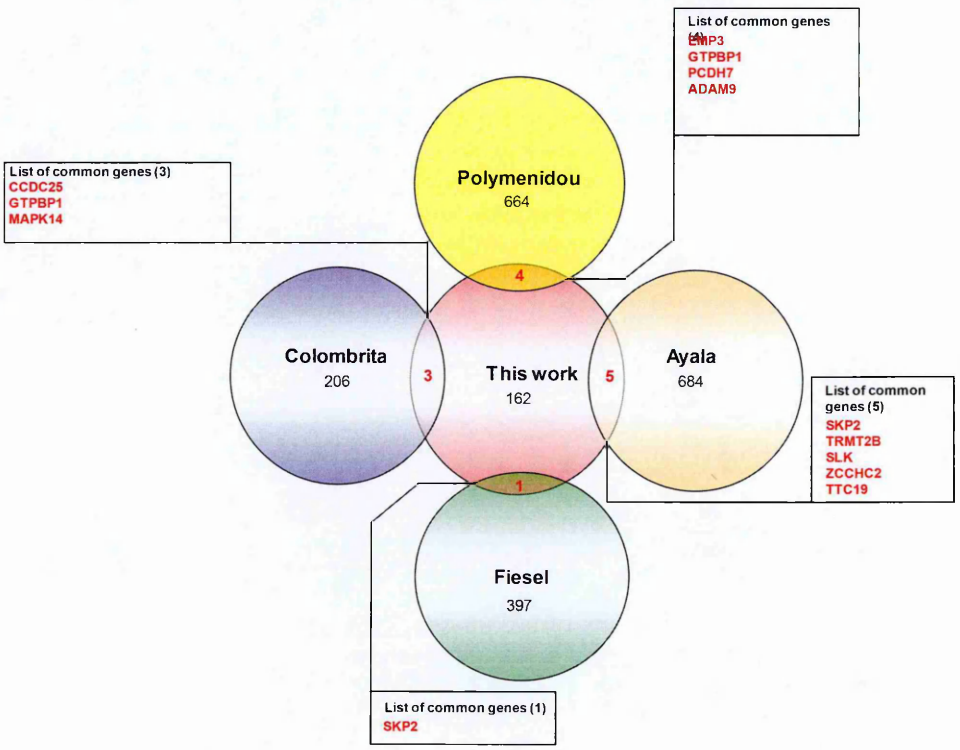



**Figure 4-1: Venn diagrams showing the overlap between the genes detected in this screening and the ones reported in other studies.** Each coloured circle represents a set of genes detected in different studies in which TDP-43 was depleted in different cell types (Ayala and Fiesel circles), in mouse brain (Polymenidou circle) or immunoprecipitated for a RIP-chip analysis (Colombrita circle). In the closed boxes are listed the overlapping genes with this Affymetrix splicing array.

Of the 162 genes, genes predicted to have a two-fold or higher splicing score were chosen for validation analyses. Thus, 19 genes were analysed for TDP-43 dependent changes using RT-PCR assays, of which 6/19 genes were confirmed to undergo TDP-43 dependent alternative splicing in non-neuronal (HEK 293) and neuronal (SK-N-SH/SK-SY5Y) cells alike. These six genes, included *POLDIP3*, *BCL2L11*, *STAG2*, *MADD*, *BRD8* and *FNIP1*. Two genes, *POLDIP3* and *BCL2L11* were already known and validated examples of genes

shown to undergo TDP-43 dependent alternative splicing (Tollervey et al. 2011; Fiesel et al. 2012). For the four other genes, depletion of TDP-43 was found to result in the increased inclusion of exons 30b and 20 in *STAG2* and *BRD8* respectively, and the exclusion of exons 31 and 7 in *MADD* and *FNIP1*. This is consistent with TDP-43 promoting both exon inclusion and exclusion of alternative exons, depending on the position in which it binds. In addition, TDP-43 dependent alternative splicing of exons 30b and 31 in *MADD* and *STAG2*, could also be observed at the protein level (see figure 2-13 & 2-20).

At the translation level, TDP-43 dependent inclusion of exon 30b in *STAG2* and exon 20 in *BRD8* result in the in-frame insertion of extra amino-acids in their respective proteins (Appendix, Figure I-1a; Figure I-4a), as well as in *FNIP1* where the exclusion of exon 7 was also predicted to be in-frame. On the other hand, TDP-43 dependent exclusion of exon 31 in *MADD*, results in an out-of-frame insertion of seven amino acids and the introduction of a premature termination codon (Appendix, Figure I-2a). A deeper look at the regions surrounding the alternatively spliced exons and flaking introns, high nucleotide conservation could be observed for the eight vertebrate species analysed, with long introns separating exons.

In the characterization of the interaction of TDP-43 with the above genes, it was interesting to note that unlike in *MADD*, where a short TG stretch upstream of exon 31 was identified, no TG repeats were present in the mapped TDP-43 binding region of *STAG2*. As has been mentioned previously, TDP-43 preferentially binds TG repeats and it was therefore not un-expected that the short TG repeat upstream of *MADD* exon 31 was involved in enhancing the inclusion of exon 31 under physiological levels of TDP-43 expression. On the other hand, the intronic region mapped to TDP-43 binding in *STAG2*, was located much further in the downstream intron, in a region that was relatively well conserved, suggesting a functional significance in this region (Figure I-1b). TDP-43 binding in transcripts with

long introns, has previously been described in Polymenidou et al. (2011) and has been reported to maintain the levels of long pre-mRNAs. Therefore, a similar mechanism could be at play for the *STAG2* exon 30b identified in this study. In addition, relative to the positioning of the TDP-43 binding site in STAG2, Tollervey et al. (2011) have shown that TDP-43 binding further downstream of alternatively spliced exons, is associated with silencing/repression of these exons. Indeed, under physiological levels of TDP-43, exon 30b of *STAG2* is normally silenced or excluded, and a shift towards increased inclusion is observed when TDP-43 is depleted in cells. Interestingly, in the same iCLIP study performed by Tollervey et al (2011), *MADD*, *STAG2* and *BRD8* were also identified, although the identified alternatively spliced exons were different. These previously reported alternative exons were not analysed in this study.

For the *BRD8* gene, despite TDP-43 dependent inclusion of exon 20 recapitulated in the pTB-minigene, mapping of the TDP-43 binding site could not be performed due to time constraints and this will be followed up in future work. Lastly, for *FNIP1*, TDP-43 dependent exclusion of exon 7 could not be reproduced in the minigene and further work is required in terms of expanding the context that would enable better recognition of this exon in the minigene.

The TDP-43 dependent changes in the four genes identified and validated from the microarrays analysis is summarised in Table 4-1 below.

| Gene | Alternatively Spliced Exon | Loss of TDP-43 consequence | Proteomic consequence |
|---|---|---|---|
| ***STAG2*** | Exon 30b | Inclusion | In-frame insertion of 36 extra AA |
| ***MADD*** | Exon 31 | Exclusion | Frameshift insertion of 7 AA and a PTC |
| ***BRD8*** | Exon 20 | Inclusion | In-frame insertion of 15 AA |
| ***FNIP1*** | Exon 7 | Exclusion | In-frame insertion of 24 AA |

Table 4-1: **Summary of genes identified from the microarray studies confirmed to undergo TDP-43 dependent alternative splicing.** AA (amino acid) and PTC (Premature termination codon).

### 4.2.1. *TDP-43 dependent alternative splicing in STAG2, MADD, BRD8 and FNIP1: link to pathological mechanisms*

Of the four genes analysed in detail in this study, it was of specific interest and relevance that *MADD* and *BIM/BCL2L11,* known to be involved in apoptotic pathways were found to undergo alternative splicing in a TDP-43 dependent manner. In fact, together with *BIM/BCL2L11* initially described by Tollervey et al. (2011) (and confirmed in this work), these two genes are important in ensuring cellular survival. *BIM/BCL2L11* is reported to be important in the induction of neuronal apoptotic mechanisms in response to nerve growth factor deprivation (Putcha et al. 2001; Kristiansen et al. 2011). On the other hand, *MADD* is implicated in cancer cell survival and apoptosis (Kurada et al. 2009), and can be spliced to yield at least six different splice variants, of which four are expressed more ubiquitously (MADD, DENN-SV, IG20pa, IG20-SV2) (Miura et al. 2012). Furthermore, MADD/DENN is thought to play a dual role in neurotransmission and neuroprotection (Miyoshi & Takai 2004), down-regulation of this protein has been reported in hippocampal regions of Alzheimer's patients (Villar & Miller 2004). The connection with Alzheimer's is particularly interesting given that TDP-43 proteinopathy has been observed in patients (Amador-Ortiz et al. 2007). In addition, the expression of these splicing variants has previously been found to be affected in neuronal cell lines exposed to high concentrations of the Amyloid-beta peptide (Mo et al. 2012), further supporting the importance of this

gene in neuronal survival. Accordingly, in this study, the changes observed in *MADD* expression suggest that there is not only a change in the level of expression of these isoforms, but also that the TDP43-induced depletion can lead to a general decrease in protein expression protein level (see figure 2.13). Further investigation is required to determine the correlation between depletion of TDP-43 and reduction in MADD protein expression, since overexpression of wild type FLAG TDP-43 could not restore protein expression levels to those observed in control cells with physiological levels of TDP-43 expression. A possible explanation could be that since MADD is a high molecular weight protein, recovery of exon 31 inclusion at the protein level lags behind that of the mRNA.

In a similar manner, the functional role of STAG2 (also known as SA2/SCC3) in the mitotic cell cycle, as a component of the cohesin complex involved in the cohesion of sister chromatids during replication (Hagstrom & Meyer 2003), could be important for cellular survival as perturbations of check points could induce cellular apoptosis. Interestingly, both overexpression and silencing of TDP-43 expression have been reported to be capable of affecting cell cycle progression. Initial observations by Ayala et al. (2008) showed that TDP-43 silenced in U2OS cells displayed alterations in cell cycle progression, with a 60% decrease in cells at the G0/G1 phase, and a corresponding increase in cells at the S and G2/M phases (Ayala, Misteli, et al. 2008). More recently, Yamashita et al. (2014) have reported that in SH-SY-5Y cells overexpressing full length wild type TDP-43 fused to GFP, there was an accumulation of cells in the G2/M and sub-G1 phase, reflecting growth arrest and apoptosis respectively. In addition, Lee et al. (2012) showed that 2- to 5-fold overexpression of wild type TDP-43 in HeLa cells was also capable of inducing partially, p53-dependent G2/M arrest and p53-independent cell death. Nonetheless, in none of these cases, was the direct connection between TDP-43 and cell cycle affecting proteins identified. In this work, some evidence that at least some of the observed effects of TDP-43 could directly be mediated by altering the pre-mRNA splicing profile of *STAG2*, an

important component of the cohesin complex, is provided.

The finding that TDP-43 can affect *STAG2* splicing also offers a novel connection towards reports that implicate a possible role of TDP-43 in tumorigenesis, as recent cancer genomics analyses have reported a rather large number of somatic mutations at the level of the core cohesin factors (SMC1A, SMC3, RAD21, and STAG1/2), in a particular subset of human tumors including Ewing sarcoma (Solomon et al. 2014). It is also interesting to note that a recent genome-wide association study (GWAS) study aimed at identifying risk factors for Ewing sarcoma has found a significant decrease in TDP-43 expression in tumor material from patients (Postel-Vinay et al. 2012). With regards to neurons, although known to be post-mitotically quiescent, a mechanism of apoptosis that involves re-entry into the cell cycle has been proposed in Alzheimer's (Currais et al. 2009), and could similarly apply to other neurodegenerative diseases. Furthermore, mechanisms involved in the maintenance of gene expression of neural specific factors after differentiation, could contribute to pathological mechanisms if their expression is perturbed by age-related cellular inefficiency (Deneris & Hobert 2014).

Finally, in the cases of *FNIP1* and *BRD8*, whose splicing profile was considerably altered by TDP-43, literature that could suggest a connection with observed consequences of TDP-43 depletion from cells is limited. *BRD8* appears to have a number of functions ranging from thyroid hormone co-activator to being an accessory protein in histone-modifying complexes, as well as playing a role in cancer survival, where elevated levels of the protein have been reported (Yamada & Rao 2009). On the other hand, FNIP1 and FNIP2, together with folliculin have been shown to physically interact with adenosine monophosphate kinase (AMPK), and affect mammalian target of rapamycin (mTOR) activities during cellular differentiation, thus representing a possible connection with cancer (Rosner et al. 2008). A role for mTOR in neuronal atrophy has been described in Alzheimer's disease (Chano et al. 2007) as well as in fragile-X syndrome through a perturbation in FMRP (and

RNA binding protein) translational repression of mTOR (Darnell & Klann 2013).

## 4.3. Analysis of TDP-43 dependent alternative splicing in a TDP-43 cellular aggregate model; Evidence for loss of function

In this study, the depletion of TDP-43 was found to alter the splicing of *STAG2*, *MADD*, *BRD8* and *FNIP1*. In an effort to gain better insight into the mechanisms underlying TDP-43 pathology, we subsequently hypothesised that if indeed, aggregates were sequestering endogenous TDP-43, as had been reported in a recent cellular model of aggregation (Budini, Romano et al. 2014), then the splicing profiles of the above genes, would be similar to those observed when TDP-43 was silenced or depleted.

Thus, the endogenous splicing profiles of previously identified and validated genes from the microarray study were analysed in a TDP-43 cellular aggregation model, following induction of the aggregation effector i.e. transgenic full length FLAG-TDP-43 fused to 12Q/N repeats. As expected, the splicing profiles of all four genes was similar to the profiles observed when TDP-43 cellular levels were depleted using siRNA in HEK 293 cells (Figure 3-11), with the exception of *MADD*, in which a pseudo-exon was also identified. This similarity in splicing profiles observed provided conclusive evidence for a loss of function of endogenous TDP-43 and consequently substantiating the loss of function hypothesis in TDP-43 proteinopathies.

Therefore, the observation that changes in the alternative splicing of *MADD*, *STAG2*, *BRD8* and *FNIP1* in the aggregate model mimic those observed in cells transiently depleted of TDP-43, suggests that an important mechanism of disease involves the full length TDP-43 protein. Indeed, Budini et al. (2014) have shown that the N-terminal domain of TDP-43, although not necessarily involved in the formation of aggregates, is necessary for the sequestration of endogenous TDP-43 that results in loss of function. In relation to aggregation, and given the involvement of TDP-43 in stress granule formation, it is plausible that the initial trigger of aggregation results from a lack of dissipation of

stress granules that transform to mature aggregates (Parker et al. 2012; Vanderweyde et al. 2013). Furthermore, stress granules are known to be associated with several RBPs (Hanson & Tibbetts 2012; Vanderweyde et al. 2013), which could contribute to a general loss of function of these proteins, leading to a perturbation of homeostatic mechanisms in both the proteome and RNA metabolome.

Interestingly in this work, the *MADD* splicing profile observed in the TDP-43 cellular aggregate model included the appearance of a pseudo-exon, a unique feature in these cells. The appearance of this pseudo exon following TDP-43 depletion was also observed in neuronal (SK-N-SH/SK-SY5Y) cells. A possible explanation for this could be that the aggregates sequester other factors, probably other hnRNPs, in addition to TDP-43 that are normally important for the definition of this exon. This is plausible given that RNA granules and aggregates contain numerous RBPs (Hanson & Tibbetts 2012). In addition, the observation that this splicing profile was present in neuronal cells and not HEK 293 highlights a role for the combinatorial definition of exons that could involve several factors and which provides other avenues for future research. Furthermore, RBPs are normally present in differing abundances that are cell-type specific, thus, it is also possible that within the neuronal cell lines, the relative abundance of these factors is naturally lower than in HEK 293 cells where this pseudo-exon was not detected.

# 5. CONCLUSIONS AND FUTURE PROSPECTS

The main aim of this study was to identify and characterize RNA targets of TDP-43 using a variety of global analytical techniques; 2-DE and microarray analyses that were designed to provide a global and complementary picture of changes that were dependent on TDP-43. With regards to the 2-DE studies, several limits of processing and analytical methods were highlighted from the results, as no positive correlation was achieved with the genes identified, and the observed differential spot intensities. Going forward, such a study would require the use of much more sensitive dyes such as fluorescent labelling, which apart from being more sensitive could partially reduce user-dependent analytical processing through automated analyses, thereby reducing opportunities for errors such as 2-D difference gel electrophoresis (DIGE), that allow for multiple samples to be run in a single gel. However, fluorescent labelling as a means of detecting differential protein intensity is also in part limited by the higher signal to noise ratio. Thus, in these types of analyses, a careful consideration of the possible limits of each reagent in relation to the research question should be performed.

On the other hand, validation and characterization of data obtained from the splice-sensitive arrays revealed that in conditions of TDP-43 depletion, the splicing profile of several important genes could be altered, in both neuronal and non-neuronal cell lines. Mapping of the TDP-43 binding site could be achieved in two of the genes (*MADD* and *STAG2*), however, due to time constraints could not be performed for *FNIP1* and *BRD8*. Thus, future work will aim to map the TDP-43 binding sites in *BRD8* and widen the minigene context of *FNIP1* to enable the recognition of exon 7, using methods described in this study.

Essentially, the genes identified and validated from the microarray in this study, indicate a general trend towards TDP-43 dependent altered splicing in genes linked to cellular

apoptotic and survival mechanisms. This could have interesting implications in terms of understanding neurodegenerative mechanisms, especially in relation to apoptosis.

The observation that these alterations can also occur in non-neuronal cell lines, and the recapitulation thereof in cellular models mimicking TDP-43 aggregation, could present a strategic starting point for biomarker analysis in monitoring disease onset/progression. Furthermore, these neuronal cells could be used as screening tools for therapeutic efficacy in the human context, for drugs aimed at restoring TDP-43 functional levels within cells.

With regards to the inclusion of a pseudo-exon in the *MADD* splicing profile that was observed in the neuronal cell lines and the aggregate cell model, it was evident that, other than the depletion of TDP-43, other unidentified factors also play a role in enhancing the recognition of this exon. Thus, future work would aim to identify these factors using techniques such as pull-down assays as well as determine whether these factors are also sequestered within aggregates. In a similar vein, the relative depletion (sequestration) of TDP-43 between these cell lines could be analysed using soluble/insoluble fractionation studies to determine if finer levels of depletion, contribute to the inclusion of this exon.

In conclusion, the data obtained in this study shows that a loss of TDP-43 can have several consequences, related to its role as a regulator of RNA metabolism. In addition, the roles played by the affected genes highlight complex, interacting pathways, that when perturbed could result in increased vulnerability to disease. Nonetheless, a clear pathological mechanism linked to TDP-43, still remains to be elucidated, as it appears, that an altered RNA metabolism, may not be the only 'hit' for the development of TDP-43 proteinopathies.

# 6. REFERENCES:


Acharya, K.K., Govind, C.K., Shore, A.N., Stoler, M.H. & Reddi, P.P., 2006. Cis-Requirement for the Maintenance of Round Spermatid-Specific Transcription. *Developmental biology*, 295(2), pp.781–90.

Aguzzi, A. & Rajendran, L., 2009. The transcellular spread of cytosolic amyloids, prions, and prionoids. *Neuron*, 64(6), pp.783–90.

Amador-Ortiz, C., Wen-Lang, L., Ahmed, Z., Davies, P., Duara, R., Graff-radford, N.R., Bbch, M., Lond, F., Hutton, M.L. & Dickson, D.W., 2007. TDP-43 IMMUNOREACTIVITY IN HIPPOCAMPAL SCLEROSIS AND ALZHEIMER'S DISEASE. *Annals of neurology*, 61(5), pp.435–445.

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y. & Oda, T., 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.

Avendaño-Vázquez, S.E., Dhir, A., Bembich, S., Buratti, E., Proudfoot, N. & Baralle, F.E., 2012. Autoregulation of TDP-43 mRNA levels involves interplay between transcription, splicing, and alternative polyA site selection. *Genes & development*, 26(15), pp.1679–84.

Ayala, Y.M., De Conti, L., Avendaño-Vázquez, S.E., Dhir, A., Romano, M., D'Ambrogio, A., Tollervey, J., Ule, J., Baralle, M., Buratti, E. & Baralle, F.E., 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 of the United States of America*, 105(10), pp.3785–9.

Ayala, Y.M., Pantano, S., D'Ambrogio, A., Buratti, E., Brindisi, A., Marchetti, C., Romano, M. & Baralle, F.E., 2005. Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *Journal of molecular biology*, 348(3), pp.575–88.

Ayala, Y.M., Zago, P., D'Ambrogio, A., Xu, Y.-F., Petrucelli, L., Buratti, E. & Baralle, F.E., 2008. Structural determinants of the cellular localization and shuttling of TDP-43. *Journal of cell science*, 121(Pt 22), pp.3778–85.

Baloh, R.H., 2011. TDP-43: the relationship between protein aggregation and neurodegeneration in amyotrophic lateral sclerosis and frontotemporal lobar degeneration. *The FEBS journal*, 278(19), pp.3539–49.

Banks, G.T., Kuta, A., Isaacs, A.M. & Fisher, E.M.C., 2008. TDP-43 is a culprit in human neurodegeneration, and not just an innocent bystander. *Mammalian genome : official journal of the International Mammalian Genome Society*, 19(5), pp.299–305.

Baralle, D., Lucassen, A. & Buratti, E., 2009. Missed threads. he Impact of pre-mRNA splicing defects on clinical practice. *EMBO reports*, 10 (8), pp 810-816.

Belzil, V. V, Gendron, T.F. & Petrucelli, L., 2012. RNA-mediated toxicity in neurodegenerative disease. *Molecular and cellular neurosciences*.

Bembich, S., Herzog, J.S., De Conti, L., Stuani, C., Avendaño-Vázquez, S.E., Buratti, E., Baralle, M. & Baralle, F.E., 2014. Predominance of spliceosomal complex formation over polyadenylation site selection in TDP-43 autoregulation. *Nucleic acids research*, 42(5), pp.3362–71.

Bentmann, E., Neumann, M., Tahirovic, S., Rodde, R., Dormann, D. & Haass, C., 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.

Berth, M., Moser, F.M., Kolbe, M. & Bernhardt, J., 2007. The state of the art in the analysis of two-dimensional gel electrophoresis images. *Applied microbiology and biotechnology*, 76(6), pp.1223–43.

Bertram, L. & Tanzi, R.E., 2005. Review series The genetic epidemiology of neurodegenerative disease. , 115(6), pp.1449–1457.

Bhardwaj, A., Myers, M.P., Buratti, E. & Baralle, F.E., 2013. Characterizing TDP-43 interaction with its RNA targets. *Nucleic acids research*, 41(9), pp.5062–74.

Bilican, B. Serio, A., Barmada, S.J., Nishimura, A.L., Sullivan, G.J., Carrasco,M., Phatnani, H.P., Puddifoot, C.A., Story, D., Fletcher, J., Park, I-H., Friedman, B.A., Dalej, G.Q., Wyllie, D.J.A., Hardingham, G.E., Wilmut, I., Finkbeiner, S., Maniatis, T., Shaw, C.E., and Chandran, S.,, 2012. Mutant induced pluripotent stem cell lines recapitulate aspects of TDP-43 proteinopathies and reveal cell-specific vulnerability. *Proceedings of the National Academy of Sciences of the United States of America*, 109(15), pp.5803–8.

Blencowe, B.J., 2006. Alternative splicing: new insights from global analyses. *Cell*, 126(1), pp.37–47.

Blencowe, B.J., Ahmad, S. & Lee, L.J., 2009. Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. *Genes & development*, 23(12), pp.1379–86.

van Blitterswijk, M., van Es, M. a, Hennekam, E. a M., Dooijes, D., van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., van der Kooi, A.J., de Visser, M., de Bakker, P.I.W., Veldink, J.H. & van den Berg, L.H., 2012. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Human molecular genetics*, 21(17), pp.3776–84.

Bose, J.K., Wang, I.-F., Hung, L., Tarn, W.-Y. & Shen, C.-K.J., 2008. TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. *The Journal of biological chemistry*, 283(43), pp.28852–9.

Budini, M. & Buratti, E., 2011. TDP-43 autoregulation: implications for disease. *Journal of molecular neuroscience*, 45(3), pp.473–9.

Budini, M., Buratti, E., Stuani, C., Guarnaccia, C., Romano, V., De Conti, L. & Baralle, F.E., 2012. Cellular model of TAR DNA-binding protein 43 (TDP-43) aggregation based on its C-terminal Gln/Asn-rich region. *The Journal of biological chemistry*, 287(10), pp.7512–25.

Budini, M., Romano, V., Quadri, Z., Buratti, E. & Baralle, F.E., 2014. TDP-43 loss of cellular function through aggregation requires additional structural determinants beyond its C-terminal Q/N prion-like domain. *Human molecular genetics*, pp.1–12.

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.

Buratti, E. & Baralle, F.E., 2012. TDP-43: gumming up neurons through protein-protein and protein-RNA interactions. *Trends in biochemical sciences*, 37(6), pp.237–47.

Buratti, E. & Baralle, F.E., 2010. The multiple roles of TDP43 in premRNA processing and gene expression regulation . *RNA Biology*, 7(4), pp.420–4.

Buratti, E. & Baralle, F.E., 2010. The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation. *RNA biology*, 7(4), pp.420–9.

Buratti, E., Baralle, M. & Baralle, F.E., 2006. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. *Nucleic acids research*, 34(12), pp.3494–510.

Buratti, E., Baralle, M. & Baralle, F.E., 2012. From single splicing events to thousands: the ambiguous step forward in splicing research. *Briefings in functional genomics*.

Buratti, E., Brindisi, A., Giombi, M., Tisminetzky, S., Ayala, Y.M. & Baralle, F.E., 2005. TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *The Journal of biological chemistry*, 280(45), pp.37572–84.

Buratti, E., De Conti, L., Stuani, C., Romano, M., Baralle, M. & Baralle, F., 2010. Nuclear factor TDP-43 can affect selected microRNA levels. *The FEBS journal*, 277(10), pp.2268–81.

Buratti, E., Dörk, T., Zuccato, E., Pagani, F., Romano, M. & Baralle, F.E., 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., Romano, M. & Baralle, F.E., 2013. TDP-43 high throughput screening analyses in neurodegeneration: advantages and pitfalls. *Molecular and cellular neurosciences*, 56, pp.465–74.

Cairns, N.J., Perrin, R.J., Schmidt, R.., Gru, A., Green, K.., Carter, D., Taylor-Reinwald, L., Morris, J.., Gitcho, M.A. & Baloh, R.., 2010. TDP-43 proteinopathy in familial motor neuron disease with TARDBP A315T mutation: a case report. *Neuropathology Applied Neurobiology*, 36(7), pp.673–679.

Cannon, J.R. & Greenamyre, J.T., 2011. The role of environmental exposures in neurodegeneration and neurodegenerative diseases. *Toxicological sciences : an official journal of the Society of Toxicology*, 124(2), pp.225–50.

Chano, T., Okabe, H. & Hulette, C.M., 2007. RB1CC1 insufficiency causes neuronal atrophy through mTOR signaling alteration and involved in the pathology of Alzheimer's diseases. *Brain research*, 1168, pp.97–105.

Chen, M., David, C.J. & Manley, J.L., 2012. Concentration-dependent control of pyruvate kinase M mutually exclusive splicing by hnRNP proteins. *Nature structural & molecular biology*, 19(3), pp.346–54.

Chen-plotkin, A.S., Lee, V.M. & Trojanowski, J.Q., 2010. TAR DNA-binding protein 43 in neurodegenerative disease. *Nature Reviews Neurology* , 6(4), pp.211–220.

Chiang, P.-M., Ling, J., Jeong, Y.H., Price, D.L., Aja, S.M. & Wong, P.C., 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 of the United States of America*, 107(37), pp.16320–4.

Cohen, T.J., Lee, V.M.Y. & Trojanowski, J.Q., 2011. TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. *Trends in molecular medicine*, 17(11), pp.659–67.

Colombrita, C., Onesto, E., Megiorni, F., Pizzuti, A., Baralle, F.E., Buratti, E., Silani, V. & Ratti, A., 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. *The Journal of biological chemistry*, 287(19), pp.15635–47.

de Conti, L., Baralle, M. & Buratti, E., 2013. Exon and intron definition in pre-mRNA splicing. *Wiley interdisciplinary reviews. RNA*, 4(1), pp.49–60.

Cooper, T. A, 2005. Use of minigene systems to dissect alternative splicing elements. *Methods (San Diego, Calif.)*, 37(4), pp.331–40.

Cragnaz, L., Klima, R., Skoko, N., Budini, M., Feiguin, F. & Baralle, F.E., 2014. Aggregate formation prevents dTDP-43 neurotoxicity in the Drosophila melanogaster eye. *Neurobiology of Disease*, 71, pp.74–80.

Cruts, M., Theuns, J. & Van Broeckhoven, C., 2012. Locus-specific mutation databases for neurodegenerative brain diseases. *Human mutation*, 33(9), pp.1340–4.

da Cruz, S. & Cleveland, D.W., 2011. Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. *Current opinion in neurobiology*, (Table 1), pp.1–16.

Currais, A., Hortobágyi, T. & Soriano, S., 2009. The neuronal cell cycle as a mechanism of pathogenesis in Alzheimer's disease. *AGING*, 1(4), pp.363–371.

Cushman, M., Johnson, B.S., King, O.D., Gitler, A.D. & Shorter, J., 2010. Prion-like disorders: blurring the divide between transmissibility and infectivity. *Journal of cell science*, 123(Pt 8), pp.1191–201.

D'Ambrogio, A., Buratti, E., Stuani, C., Guarnaccia, C., Romano, M., Ayala, Y.M. & Baralle, F.E., 2009. Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic acids research*, 37(12), pp.4116–26.

Darnell, J.C. & Klann, E., 2013. The translation of translational control by FMRP: therapeutic targets for FXS. *Nature neuroscience*, 16(11), pp.1530–6.

DeJesus-Hernandez, M. et al., 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, 72(2), pp.245–56.

Deneris, E.S. & Hobert, O., 2014. Maintenance of postmitotic neuronal cell identity. *Nature neuroscience*, 17(7), pp.899–907.

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M. & Béroud, C., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic acids research*, 37(9), p.e67.

Desplats, P., Lee, H., Bae, E., Patrick, C., Rockenstein, E., Crews, L. & Spencer, B., 2009. Inclusion formation and neuronal cell death through neuron-to-neuron transmission of alpha-synuclein. *Proceedings of the National Academy of Sciences*, 106(41), pp.17606–17606.

Dichmann, D.S. & Harland, R.M., 2012. fus/TLS orchestrates splicing of developmental regulators during gastrulation. *Genes & development*, 26(12), pp.1351–63.

Ding, Y., Xu, L., Jovanovic, B.D., Helenowski, I.B. & Kelly, D.L., 2007. The Methodology Used to Measure Differential Gene Expression Affects the Outcome. *Journal of biomolecular techniques*, 18(5), pp.321–330.

Dreumont, N., Hardy, S., Behm-Ansmant, I., Kister, L., Branlant, C., Stévenin, J. & Bourgeois, C.F., 2010. Antagonistic factors control the unproductive splicing of SC35 terminal intron. *Nucleic acids research*, 38(4), pp.1353–66.

Estes, P.S., Boehringer, A., Zwick, R., Tang, J.E., Grigsby, B. & Zarnescu, D.C., 2011. Wild-type and A315T mutant TDP-43 exert differential neurotoxicity in a Drosophila model of ALS. *Human molecular genetics*, 20(12), pp.2308–21.

Faustino, N.A. & Cooper, T. A, 2003. Pre-mRNA splicing and human disease. *Genes & development*, 17(4), pp.419–37.

Feiguin, F., Godena, V.K., Romano, G., D'Ambrogio, A., Klima, R. & Baralle, F.E., 2009. Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior. *FEBS letters*, 583(10), pp.1586–92.

Fernández-Santiago, R., Hoenig, S., Lichtner, P., Sperfeld, A.-D., Sharma, M., Berg, D., Weichenrieder, O., Illig, T., Eger, K., Meyer, T., Anneser, J., Münch, C., Zierz, S., Gasser, T. & Ludolph, A., 2009. Identification of novel Angiogenin (ANG) gene missense variants in German patients with amyotrophic lateral sclerosis. *Journal of neurology*, 256(8), pp.1337–42.

Ferraiuolo, L., Heath, P.R., Holden, H., Kasher, P., Kirby, J. & Shaw, P.J., 2007. Microarray analysis of the cellular pathways involved in the adaptation to and progression of motor neuron injury in the SOD1 G93A mouse model of familial ALS. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(34), pp.9201–19.

Fiesel, F.C., Voigt, A., Weber, S.S., Vanden Haute, C., Waldenmaier, A., Gorner, K., Walter, M., Anderson, M.L., Kern, J.V., Rasse, T.M., Schmidt, T., Springer, W., Roland Kirchner, R., Bonin, M., Neumann, M., Baekelandt, V., Alunni-Fabbroni, M.., Schulz, J.B., and Kahle, P.J., 2010. Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6. *The EMBO journal*, 29(1), pp.209–21.

Fiesel, F.C., Weber, S.S., Supper, J., Zell, A. & Kahle, P.J., 2012. TDP-43 regulates global translational yield by splicing of exon junction complex component SKAR. *Nucleic acids research*, 40(6), pp.2668–82.

Forbes, R.B., Colville, S. & Swingler, R.J., 2004. The epidemiology of amyotrophic lateral sclerosis (ALS/MND) in people aged 80 or over. *Age and ageing*, 33(2), pp.131–4.

Freibaum, B.D., Chitta, R., High, A.A. & Taylor, J.P., 2011. Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. *Journal of Proteome Research*, 9(2), pp.1104–1120.

Fuentealba, R. a, Udan, M., Bell, S., Wegorzewska, I., Shao, J., Diamond, M.I., Weihl, C.C. & Baloh, R.H., 2010. Interaction with polyglutamine aggregates reveals a Q/N-rich domain in TDP-43. *The Journal of biological chemistry*, 285(34), pp.26304–14.

Furukawa, Y., Kaneko, K., Watanabe, S., Yamanaka, K. & Nukina, N., 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.

Garfin, D.E., 2003. Two-dimensional gel electrophoresis: an overview. *Trends in Analytical Chemistry*, 22(5), pp.263–272.

Gauci, V.J., Wright, E.P. & Coorssen, J.R., 2011. Quantitative proteomics: assessing the spectrum of in-gel protein detection methods. *Journal of chemical biology*, 4(1), pp.3–29.

Gendron, T.F., Rademakers, R. & Petrucelli, L., 2013. TARDBP mutation analysis in TDP-43 proteinopathies and deciphering the toxicity of mutant TDP-43. *Journal of Alzheimer's disease*, 33 Suppl 1, pp.S35–45.

Geser, F., Martinez-Lage, M., Kwong, L.K., Lee, V.M.-Y. & Trojanowski, J.Q., 2009. Amyotrophic lateral sclerosis, frontotemporal dementia and beyond: the TDP-43 diseases. *Journal of neurology*, 256(8), pp.1205–14.

Geser, F., Prvulovic, D., O'Dwyer, L., Hardiman, O., Bede, P., Bokde, a L.W., Trojanowski, J.Q. & Hampel, H., 2011. On the development of markers for pathological TDP-43 in amyotrophic lateral sclerosis with and without dementia. *Progress in neurobiology*, 95(4), pp.649–62.

Gitcho, M. A, Bigio, E.H., Mishra, M., Johnson, N., Weintraub, S., Mesulam, M., Rademakers, R., Chakraverty, S., Cruchaga, C., Morris, J.C., Goate, A.M. & Cairns, N.J., 2009. TARDBP 3'-UTR variant in autopsy-confirmed frontotemporal lobar degeneration with TDP-43 proteinopathy. *Acta neuropathologica*, 118(5), pp.633–45.

Gitcho, M.A., Baloh, R.H., Chakraverty, S., Norton, J.B., Levitch, D., Hatanpaa, K.J., Iii, C.L.W., Bigio, E.H., Caselli, R., Baker, M., Al-lozi, M.T., Morris, J.C., Pestronk, A., Goate, A.M. & Cairns, N.J., 2008. TDP-43 A315T Mutation in Familial Motor Neuron Disease. *Annals of Neurology*, 63(4), pp.535–538.

Glisovic, T., Bachorik, J.L., Yong, J. & Dreyfuss, G., 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS letters*, 582(14), pp.1977–86.

Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., Morrison, K.E., Green, A., Acharya, K.R., Brown, R.H. & Hardiman, O., 2006. ANG mutations segregate with familial and "sporadic" amyotrophic lateral sclerosis. *Nature genetics*, 38(4), pp.411–3.

Gregory, R.I., Chendrimada, T.P., Cooch, N. & Shiekhattar, R., 2005. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell*, 123(4), pp.631–40.

Guo, W. Chen, Y., Zhou, X., Kar, A., Ray, P., Chen, X., Rao, E.J., Yang, M.,Ye, H., Zhu, L., Liu, J., Xu, M.,Yang, Y., Wang, C., Zhang, D., Bigio, E.H., Mesulam, M., Shen, Y., Xu, Q., Fushimi, K., & Wu, J.Y. 2011. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nature structural & molecular biology*, 18(7), pp.822–30.

Hagstrom, K. A & Meyer, B.J., 2003. Condensin and cohesin: more than chromosome compactor and glue. *Nature reviews. Genetics*, 4(7), pp.520–34.

Hallegger, M., Llorian, M. & Smith, C.W.J., 2010. Alternative splicing: global insights. *The FEBS journal*, 277(4), pp.856–66.

Hansen, C., Angot, E., Bergström, A., Steiner, J.A., Pieri, L., Paul, G., Outeiro, T.F., Melki, R., Kallunki, P. & Fog, K., 2011. α-Synuclein propagates from mouse brain to grafted dopaminergic neurons and seeds aggregation in cultured human cells. *Journal of clinical investigation*, 121(2), pp.715–725.

Hanson, K.A. & Tibbetts, R.S., 2012. RNA binding proteins in neurodegenerative disease: TDP-43and beyond. *Wiley Interdiscilinary Review RNA*, 3(2), pp.265–285.

Hazelett, D.J., Chang, J.-C., Lakeland, D.L. & Morton, D.B., 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 (Bethesda, Md.)*, 2(7), pp.789–802.

Huelga, S.C., Vu, A.Q., Arnold, J.D., Liang, T.Y., Patrick, P., Yan, B.Y., Donohue, J.P., Shiue, L., Hoon, S., Jr, M.A. & Yeo, G.W., 2012. of alternative splicing by hnRNP proteins. *Cell Reports*, 1(2), pp.167–178.

Igaz, L.M., Kwong, L.K., Chen-Plotkin, A., Winton, M.J., Unger, T.L., Xu, Y., Neumann, M., Trojanowski, J.Q. & Lee, V.M.-Y., 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.

Johnson, B.S., McCaffery, J.M., Lindquist, S. & Gitler, A.D., 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., Snead, D., Lee, J.J., McCaffery, J.M., Shorter, J. & Gitler, A.D., 2009. TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. *The Journal of biological chemistry*, 284(30), pp.20329–39.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. & Shoemaker, D.D., 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science (New York, N.Y.)*, 302(5653), pp.2141–4.

Johnson, J.O. Pioro, E.P., Boehringer, A., Chia, R., Feit, H., Renton, A.E., Pliner, H.A., Abramzon, Y., Marangi, G., Winborn, B.J., Gibbs, J.R., Nalls, M.A., Morgan, S., Shoai, M., Hardy, J., Pittman, A., Orrell, R.W., Malaspina, A., Sidle, K.C., Fratta, P., Harms, M.B., Baloh, R.H., Pestronk, A., Weihl, C.C., Rogaeva, E., Zinman, L., Drory, V.E., Borghero, G., Mora, G., Calvo, A., Rothstein, J.D., ITALSGEN, Drepper, C., Sendtner, M., Andrew B Singleton, A.B., Taylor, J.P., Cookson, M.R., Restagno, G., Sabatelli, M., Bowser, R., Chiò, A & Traynor, B.J. 2014. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nature neuroscience*, 17(5), pp.664–6.

Kim K, Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L., Yabuuchi1, A., Takeuchi, A., Cunniff, K.C., Hongguan, Hg., Mckinney-Freeman, S., Naveiras, O., Yoon, T.J., Irizarry, R.A., Jung, N., Seita, J., Hanna, J., Murakami, P., Jaenisch, R., Weissleder, Orkin, S.H., Weissman, I.L., Feinberg, A.P., and Daley.G.Q., 2010. Epigenetic Memory in induced pluripotent stem cells. *Nature*, 467(7313), pp.285–290.

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.-P., Lacomblez, L., Pochigaeva, K., Salachas, F., Pradat, P.-F., Camu, W., Meininger, V., Dupre, N. & Rouleau, G. a, 2008. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nature genetics*, 40(5), pp.572–4.

Kalsotra, A. & Cooper, T., 2012. Functional consequences of developmentally regulated alternative splicing. *Nature Review Genetics*, 12(10): , 12(10), pp.715–729.

Kawahara, Y. & Mieda-Sato, A., 2012. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), pp.3347–52.

Keene, J.D., Komisarow, J.M. & Friedersdorf, M.B., 2006. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature protocols*, 1(1), pp.302–7.

King, O.D., Gitler, A.D. & Shorter, J., 2012. The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain research*, 1462, pp.61–80.

Kingston, R.E., Chen, C.A. & Okayama, H., 2001. Calcium phosphate transfection. *Current protocols in immunology / edited by John E. Coligan ... [et al.]*, Chapter 10, p.Unit 10.13.

Kraemer, B.C., Schuck, T., Wheeler, J.M., Robinson, L.C., Trojanowski, J.Q., Lee, V.M.Y. & Schellenberg, G.D., 2010. Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta neuropathologica*, 119(4), pp.409–19.

Krecic, A. & Swanson, M., 1999. hnRNP complexes: Composition, Structure and function. *Current Opinion in Cell Biology*, pp.363–371.

Kristiansen, M., Menghi, F., Hughes, R., Hubank, M. & Ham, J., 2011. Global analysis of gene expression in NGF-deprived sympathetic neurons identifies molecular pathways associated with cell death. *BMC genomics*, 12(1), p.551.

Kuo, P.-H., Chiang, C.-H., Wang, Y.-T., Doudeva, L.G. & Yuan, H.S., 2014. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic acids research*, 42(7), pp.4712–22.

Kuo, P.-H., Doudeva, L.G., Wang, Y.-T., Shen, C.-K.J. & Yuan, H.S., 2009. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic acids research*, 37(6), pp.1799–808.

Kurada, B.R.V.V.S.N., Li, L.C., Mulherkar, N., Subramanian, M., Prasad, K. V & Prabhakar, B.S., 2009. MADD, a splice variant of IG20, is indispensable for MAPK activation and protection against apoptosis upon tumor necrosis factor-alpha treatment. *The Journal of biological chemistry*, 284(20), pp.13533–41.

Kwiatkowski, T.J., Bosco, D. A., LeClerc, A. L., Tamrazian, E., Vanderburg, C. R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E. J., Munsat, T. Valdmanis, P., Rouleau, G. A., Hosler, B. A., Cortelli, P., . de Jong, P. J., Yoshinaga, Y., Haines, J. L., Pericak-Vance, M. A., Yan, J., Ticozzi, N., Siddique, T., McKenna-Yasek, D., Sapp, P. C., Horvitz, H. R., Landers, J. E., Brown Jr, R. H.,, 2009. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science (New York, N.Y.)*, 323(5918), pp.1205–8.

Lagier-Tourenne, C., Polymenidou, M., Hutt , K.R., Vu, A.Q, Baughn, M., Huelga,S.C., Clutario, K.M., Ling, S-C., Liang,T.Y., Mazur, C., Wancewicz, E., Kim, A.S., Watt, A., Freier, S., Hicks, G.G., Donohue, J.P., Shiue, L., Frank, C. B., Ravits, J., Cleveland, D.W., and Yeo, G.W., 2012. Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nature neuroscience*, 15(11), pp.1488–97.

Lagier-Tourenne, C. & Cleveland, D.W., 2009. Rethinking ALS: the FUS about TDP-43. *Cell*, 136(6), pp.1001–4.

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.

Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C. & Brenner, S.E., 2007. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*, 446(7138), pp.926–9.

Lee, K., Suzuki, H., Aiso, S. & Matsuoka, M., 2012. Overexpression of TDP-43 causes partially p53-dependent G2/M arrest and p53-independent cell death in HeLa cells. *Neuroscience letters*, 506(2), pp.271–6.

Levine, M. & Tjian, R., 2003. Transcription regulation and animal diversity. *Nature*, 424(6945), pp.147–51.

Li, Y., Ray, P., Rao, E.J., Shi, C., Guo, W., Chen, X., Woodruff, E. a, Fushimi, K. & Wu, J.Y., 2010. A Drosophila model for TDP-43 proteinopathy. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), pp.3169–74.

Lill, C.M. & Bertram, L., 2011. Towards unveiling the genetics of neurodegenerative diseases. *Seminars in neurology*, 31(5), pp.531–41.

Ling, S.-C., Polymenidou, M. & Cleveland, D.W., 2013. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron*, 79(3), pp.416–38.

Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. & Lockhart, D.J., 1999. High density synthetic oligonucleotide arrays. *Nature Genetics* , 21(january).

Liu-Yesucevitz, L., Bassell, G.J., Gitler, A.D., Hart, A.C., Klann, E., Richter, J.D., Warren, S.T. & Wolozin, B., 2011. Local RNA translation at the synapse and in disease. *The Journal of neuroscience* 31(45), pp.16086–93.

Logroscino, G., Traynor, B.J., Hardiman, O., Chiò, A., Swingler, R.J., Millul, A., Benn, E. & Eurals, E.B., 2011. Incidence of Amyotrophic Lateral Sclerosis in Europe. , 81(4), pp.385–390.

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.

Lutz, C. & Moreira, A., 2011. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip Rev RNA*, 2(1), pp.23–31.

Mackenzie, I.R. Bigio, E.H., Ince, P.G., Felix Geser, F., Neumann, M., Cairns, N.J., Kwong, L.K., Forman, M.S., Ravits, J., Stewart, H., Eisen, A., McClusky, L., Kretzschmar, H.A., Monoranu, C.M., Highley,J.R., Kirby, J., Siddique, T., Shaw,P.J., Lee, V.M-Y., and Trojanowski, J.Q.,, 2007. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Annals of neurology*, 61(5), pp.427–34.

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

Martins-de-Souza, D., Guest, P.C., Mann, D.M., Roeber, S., Rahmoune, H., Bauder, C., Kretzschmar, H., Volk, B., Baborie, A. & Bahn, S., 2012. Proteomic analysis identifies dysfunction in cellular transport, energy, and protein metabolism in different brain regions of atypical frontotemporal lobar degeneration. *Journal of proteome research*, 11(4), pp.2533–43.

Mattis, V.B. & Svendsen, C.N., 2011. Induced pluripotent stem cells: a new revolution for clinical neurology? *Lancet neurology*, 10(4), pp.383–94.

McManus, J. & Graveley, B., 2012. RNA structure and the mechanisms of alternative splicing. *Curr Opin Genet Dev.* 2, 21(4), pp.373–379.

Mercado, P.A., Ayala, Y.M., Romano, M., Buratti, E. & Baralle, F.E., 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.6000–10.

Miller, R.., Mitchell, J.. & Moore, D.., 2012. Riluzole for amyotrophic lateral sclerosis (ALS )/ motor neuron disease (MND). *Cochrane Database of Systematic Reviews*, (3).

Miura, K., Fujibuchi, W. & Unno, M., 2012. Splice Variants in Apoptotic pathway. *Experomental Oncology*, 2012 (34-3), pp.212–217.

Miyoshi, J. & Takai, Y., 2004. Dual role of DENN/MADD (Rab3GEP) in neurotransmission and neuroprotection. *Trends in molecular medicine*, 10(10), pp.476–80.

Mo, Y., Williams, C. & Miller, C.A., 2012. DENN/MADD/IG20 alternative splicing changes and cell death in Alzheimer's disease. *Journal of molecular neuroscience : MN*, 48(1), pp.97–110.

Nehls, J., Koppensteiner, H., Brack-Werner, R., Floss, T. & Schindler, M., 2014. HIV-1 Replication in Human Immune Cells Is Independent of TAR DNA Binding Protein 43 (TDP-43) Expression. *PloS one*, 9(8), p.e105478.

Neumann, M., 2009. Molecular neuropathology of TDP-43 proteinopathies. *International journal of molecular sciences*, 10(1), pp.232–46.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., McCluskey, L.F., Miller, B.L., Masliah, E., Mackenzie, I.R., Feldman, H., Feiden, W., Kretzschmar, H.A., et al., 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (New York, N.Y.)*, 314(5796), pp.130–3.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., McCluskey, L.F., Miller, B.L., Masliah, E., Mackenzie, I.R., Feldman, H., Feiden, W., Kretzschmar, H. a, et al., 2006. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science (New York, N.Y.)*, 314(5796), pp.130–3.

Neumann, M., Rademakers, R., Roeber, S., Baker, M., Kretzschmar, H. a & Mackenzie, I.R. a, 2009. A new subtype of frontotemporal lobar degeneration with FUS pathology. *Brain : a journal of neurology*, 132(Pt 11), pp.2922–31.

Ong, S.E. & Pandey, A., 2001. An evaluation of the use of two-dimensional gel electrophoresis in proteomics. *Biomolecular engineering*, 18(5), pp.195–205.

Onodera, O., Sugai, A., Konno, T., Tada, M. & Koyama, A., 2013. What is the key player in TDP-43 pathology in ALS : Disappearance from the nucleus or inclusion formation in the cytoplasm ? *Neurology and Clinical Neuroscience*, 1, pp.11–17.

Ou, S.H., Wu, F., Harrich, D., García-Martínez, L.F. & Gaynor, R.B., 1995. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *Journal of virology*, 69(6), pp.3584–96.

Pagani, F. & Baralle, F.E., 2004. Genomic variants in exons and introns: Identifying the splicing spoilers. *Nature reviews. Genetics*, 5(May).

Pagani, F., Buratti, E., Stuani, C., Romano, M., Zuccato, E., Niksic, M., Giglio, L., Faraguna, D. & Baralle, F.E., 2000. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *The Journal of biological chemistry*, 275(28), pp.21041–7.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, 40(12), pp.1413–5.

Parker, S.J., Meyerowitz, J., James, J.L., Liddell, J.R., Crouch, P.J., Kanninen, K.M. & White, A.R., 2012. Endogenous TDP-43 localized to stress granules can subsequently form protein aggregates. *Neurochemistry international*, 60(4), pp.415–24.

Patani, R., Sibley, C.R., Chandran, S. & Ule, J., 2012. Using human pluripotent stem cells to study post-transcriptional mechanisms of neurodegenerative diseases. *Brain research*, 1462, pp.129–38.

Patel, A. A & Steitz, J. A., 2003. Splicing double: insights from the second spliceosome. *Nature reviews. Molecular cell biology*, 4(12), pp.960–70.

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. & Cleveland, D.W., 2012. Prion-like spread of protein aggregates in neurodegeneration. *The Journal of experimental medicine*, 209(5), pp.889–93.

Polymenidou, M. & Cleveland, D.W., 2011. The seeds of neurodegeneration: prion-like spreading in ALS. *Cell*, 147(3), pp.498–508.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Bennett, C.F., Cleveland, D.W. & Yeo, G.W., 2012. Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain Research*, 1462, pp.3–15.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Bennett, C.F., Cleveland, D.W. & Yeo, G.W., 2012. Misregulated RNA processing in amyotrophic lateral sclerosis. *Brain research*, 1462, pp.3–15.

Polymenidou, M., Lagier-tourenne, C., Hutt, K.R., Stephanie, C., Moran, J., Liang, T.Y., Ling, S., Sun, E., Wancewicz, E., Mazur, C., Kordasiewicz, H., Sedaghat, Y., Paul, J., Shiue, L., Bennett, C.F., Yeo, G.W. & Cleveland, D.W., 2011. Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nature neuroscience*, 14(4), pp.459–468.

Postel-Vinay, S. et al., 2012. Common variants near TARDBP and EGR2 are associated with susceptibility to Ewing sarcoma. *Nature genetics*, 44(3), pp.323–7.

Przedborski, S., Vila, M. & Jackson-lewis, V., 2003. Neurodegeneration : What is it and where are we ? *Neurodegeneration*, 111(1), pp.3–10.

Putcha, G. V, Moulder, K.L., Golden, J.P., Bouillet, P., Adams, J.A., Strasser, A., Johnson, E.M., Louis, S. & Walter, T., 2001. Induction of BIM , a Proapoptotic BH3-Only BCL-2 Family Member , Is Critical for Neuronal Apoptosis. *Neuron*, 29, pp.615–628.

Ramaswami, M., Taylor, J.P. & Parker, R., 2013. Altered ribostasis: RNA-protein granules in degenerative disorders. *Cell*, 154(4), pp.727–36.

Ravits, J., Appel, S., Baloh, R.H., Barohn, R., Brooks, B.R., Elman, L., Floeter, M.K., Henderson, C. & Lomen-hoerth, C., 2013. Deciphering amyotrophic lateral sclerosis: What phenotype, neuropathology and genetics are telling us about pathogenesis. *Amytrophic lateral Scle. Frontotemoral Degeneration*, 14(01), pp.5–18.

Renoux, A.J. & Todd, P.K., 2012. Neurodegeneration the RNA way. *Progress in neurobiology*, 97(2), pp.173–89.

Renton, A.E. Véron, A.S , Tirode, F., Pierron, G., Reynaud,S., Kovar, H., Oberlin, O., Lapouble, E., Ballet, S., Lucchesi, C., Kontny, U., González-Neira, A., Picci, P., Alonso, J., Patino-Garcia, A., de Paillerets, B.B., Laud, K., Dina, C., Froguel, P., Clavel-Chapelon, F., Doz, F, Michon, J., Chanock, S.J., Thomas, G., Cox, D.G., & Delattre, O., Majounie, E., Waite, A., Simon-Sanchez, J., Sara Rollinson, S., Gibbs, R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas,L., Kalimo,H., et al. 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, 72(2), pp.257–68.

Robinson, J.L., Geser, F., Stieber, A., Umoh, M., Kwong, L.K., Van Deerlin, V.M., Lee, V.M.-Y. & Trojanowski, J.Q., 2013. TDP-43 skeins show properties of amyloid in a subset of ALS cases. *Acta neuropathologica*, 125(1), pp.121–31.

Romano, M. & Buratti, E., 2013. Targeting RNA binding proteins involved in neurodegeneration. *Journal of biomolecular screening*, 18(9), pp.967–83.

Romano, M., Feiguin, F. & Buratti, E., 2012. Drosophila Answers to TDP-43 Proteinopathies. *Journal of amino acids*, 2012, p.356081.

Rosen, D.R., Siddique, T., David, P., . Figlewicz, D. A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Reganparallel, J.P., Deng, H-X., Rahmani, Z., Krizus, A., McKenna-Yasek, D., Cayabyab, A., Gaston, S.M., Berger, R., Tanzi, R.E., Halperin, J.J., Herzfeldt, B., van den Bergh, R., Hung, W-Y., Bird, T., Deng, G., Mulder, D.W., Smyth, C., Laing, N.G., Soriano, E., Pericak-Vance, M.A., Haines, J., Rouleaou, G.A., Gusella, J.S., Horvitz, R.H., Brown Jr, R.H., 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature*, 362, pp.59–62.

Rosner, M., Hanneder, M., Siegel, N., Valli, A., Fuchs, C. & Hengstschläger, M., 2008. The mTOR pathway and its role in human genetic diseases. *Mutation research*, 659(3), pp.284–92.

Ross, C.A & Poirier, M. A, 2004. Protein aggregation and neurodegenerative disease. *Nature medicine*, 10 Suppl(July), pp.S10–7.

Rossbach, O., Hung, L.-H., Schreiner, S., Grishina, I., Heiner, M., Hui, J. & Bindereif, A., 2009. Auto- and cross-regulation of the hnRNP L proteins by alternative splicing. *Molecular and cellular biology*, 29(6), pp.1442–51.

Rothstein, J.D., 2009. Current hypotheses for the underlying biology of amyotrophic lateral sclerosis. *Annals of neurology*, 65 Suppl 1, pp.S3–9.

Rutherford, N.J. Zhang, Y-J., Baker, M, Gass, J .M., Finch, N.C., Xu, Y-F., Stewart, H., Kelley, B.J., Kuntz, K., Crook, R.J.P., Sreedharan, J., Vance, C., Sorenson, E., Lippa, C., Bigio, E.H., Geschwind, D.H., Knopman, D.S., Mitsumoto, H., Petersen, R.C., Cashman, N.R., Hutton, M., C.E., Boylan, K.B., Boeve, B., Graff-Radford, N.R., Wszolek, Z.K., Caselli, R.J., Dickson, D.W., Mackenzie, I.R., Petrucelli, L., Rademakers, R., 2008. Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. *PLoS genetics*, 4(9), p.e1000193.

Sephton, C.F., Cenik, C., Kucukural, A., Dammer, E.B., Cenik, B., Han, Y., Dewey, C.M., Roth, F.P., Herz, J., Peng, J., Moore, M.J. & Yu, G., 2011. Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *The Journal of biological chemistry*, 286(2), pp.1204–15.

Sephton, C.F., Good, S.K., Atkin, S., Dewey, C.M., Mayer, P., Herz, J. & Yu, G., 2010. TDP-43 is a developmentally regulated protein essential for early embryonic development. *The Journal of biological chemistry*, 285(9), pp.6826–34.

Sheikh, S., Safia, Haque, E. & Mir, S., 2012. Neurodegenerative Diseases : Multifactorial Conformational Diseases and Their Therapeutic Interventions. *Journal of Neurodegenerative diseases*, 2013. doi.org/10.1155/2013/563481

Shen, S., Warzecha, C.C., Carstens, R.P. & Xing, Y., 2010. MADS+: discovery of differential splicing events from Affymetrix exon junction array data. *Bioinformatics (Oxford, England)*, 26(2), pp.268–9.

Shiga, A., Ishihara, T., Miyashita, A., Kuwabara, M., Kato, T., Watanabe, N., Yamahira, A., Kondo, C., Yokoseki, A., Takahashi, M., Kuwano, R., Kakita, A., Nishizawa, M., Takahashi, H. & Onodera, O., 2012. Alteration of POLDIP3 splicing associated with loss of function of TDP-43 in tissues affected with ALS. *PloS one*, 7(8), p.e43120.

Skovronsky, D.M., Lee, V.M.-Y. & Trojanowski, J.Q., 2006. Neurodegenerative diseases: new concepts of pathogenesis and their therapeutic implications. *Annual review of pathology*, 1, pp.151–70.

Soergel, D.A.W., Lareau, L.F. & Brenner, S.E., 2006. Regulation of Gene Expression by Coupling of Alternative Splicing and NMD. In *Nonsense-mediated mRNA Decay*.

Solomon, D.A., Kim, J. & Waldman, T., 2014. Cohesin gene mutations in tumorigenesis : from discovery to clinical significance. *BMB reports*, 47(6), pp.299–310.

la Spada, A.R. & Taylor, J.P., 2010. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nature reviews. Genetics*, 11(4), pp.247–58.

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., Baralle, F., de Belleroche, J., Mitchell, J.D., Leigh, P.N., Al-Chalabi, A., Miller, C.C., Nicholson, G. & Shaw, C.E., 2008. TDP-43 mutations in

familial and sporadic amyotrophic lateral sclerosis. *Science (New York, N.Y.)*, 319(5870), pp.1668–72.

Swarup, V., Phaneuf, D., Bareil, C., Robertson, J., Rouleau, G. a, Kriz, J. & Julien, J.-P., 2011. Pathological hallmarks of amyotrophic lateral sclerosis/frontotemporal lobar degeneration in transgenic mice produced with TDP-43 genomic fragments. *Brain : a journal of neurology*, 134(Pt 9), pp.2610–26.

Tang, L.-J., De Seta, F., Odreman, F., Venge, P., Piva, C., Guaschino, S. & Garcia, R.C., 2007. Proteomic analysis of human cervical-vaginal fluids. *Journal of proteome research*, 6(7), pp.2874–83.

Taylor, J.P., Hardy, J. & Fischbeck, K.H., 2002. Toxic proteins in neurodegenerative disease. *Science (New York, N.Y.)*, 296(5575), pp.1991–5.

Tazi, J., Bakkour, N. & Stamm, S., 2009. Alternative splicing and disease. *Biochimica et biophysica acta*, 1792(1), pp.14–26.

Tollervey, J.R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A.L., Zupunski, V., Patani, R., Chandran, S., Rot, G., Zupan, B., Shaw, C.E. & Ule, J., 2011. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nature neuroscience*, 14(4), pp.452–8.

Tsao, W., Jeong, Y.H., Lin, S., Ling, J., Price, D.L., Chiang, P.-M. & Wong, P.C., 2012. Rodent models of TDP-43: recent advances. *Brain research*, 1462, pp.26–39.

Ugras, S.E. & Shorter, J., 2012. RNA-Binding Proteins in Amyotrophic Lateral Sclerosis and Neurodegeneration. *Neurology research international*, 2012, p.432780.

Ule, J., Jensen, K., Mele, A. & Darnell, R.B., 2005. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods (San Diego, Calif.)*, 37(4), pp.376–86.

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Al-saraj, S., Al-chalabi, A., Leigh, P.N., Blair, I.P. & Nicholson, G., 2009. Mutations in FUS, an RNA processing prtein, cause Familial Amyotrophic Lateral Sclerosis Type 6. *Science (New York, N.Y.)*, 323(February), pp.1208–1211.

Vanderweyde, T., Youmans, K., Liqun, L.-Y. & Wolozin, B., 2013. The Role Stress Granules and RNA Binding Proteins in Neurodegeneration. *Gerontology*, 59(6), pp.1–13.

Villar, K. Del & Miller, C.A., 2004. Down-regulation of DENN ¯MADD , a TNF receptor binding protein , correlates with neuronal cell death in Alzheimer ' s disease brain and hippocampal neurons. *Proceedings of the National Academy of Sciences*.

Volkening, K., Leystra-Lantz, C., Yang, W., Jaffee, H. & Strong, M.J., 2009. Tar DNA binding protein of 43 kDa (TDP-43), 14-3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS). *Brain research*, 1305, pp.168–82.

Wahl, M.C., Will, C.L. & Lührmann, R., 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell*, 136(4), pp.701–18.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I. & Zhang, L., 2008. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature*, 456(7221), pp.470–476.

Wang, H.-Y., Wang, I.-F., Bose, J. & Shen, C.-K.J., 2004. Structural diversity and functional implications of the eukaryotic TDP gene family. *Genomics*, 83(1), pp.130–139.

Wang, Y., Xiao, X., Zhang, J., Choudhury, R., Robertson, A., Li, K., Ma, M., Burge, C.B. & Wang, Z., 2013. A complex network of factors with overlapping affinities represses splicing through intronic elements. *Nature structural & molecular biology*, 20(1), pp.36–45.

Wang, Y.-T., Kuo, P.-H., Chiang, C.-H., Liang, J.-R., Chen, Y.-R., Wang, S., Shen, J.C.K. & Yuan, H.S., 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.

Ward, A.J. & Cooper, T.A., 2011. The Pathobiology of Splicing. *Journal of Pathology*, 220(2), pp.152–163.

Winton, M.J., Igaz, L.M., Wong, M.M., Kwong, L.K., Trojanowski, J.Q. & Lee, V.M.-Y., 2008. Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. *The Journal of biological chemistry*, 283(19), pp.13302–9.

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. *The Journal of biological chemistry*, 287(33), pp.27335–44.

Xiao, S., Sanelli, T., Dib, S., Sheps, D., Findlater, J., Bilbao, J., Keith, J., Zinman, L., Rogaeva, E. & Robertson, J., 2011. RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Molecular and cellular neurosciences*, 47(3), pp.167–80.

Yamada, H.Y. & Rao, C. V, 2009. BRD8 is a potential chemosensitizing target for spindle poisons in colorectal cancer therapy. *International Journal of Oncology*, pp.1101–1109.

Yamashita, M., Nonaka, T., Hirai, S., Miwa, A., Okado, H., Arai, T., Hosokawa, M., Akiyama, H. & Hasegawa, M., 2014. Distinct pathways leading to TDP-43-induced cellular dysfunctions. *Human molecular genetics*, 23(16), pp.4345–56.

Yeakley, J.M., Fan, J., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S. & Fu, X., 2002. Profiling alternative splicing on fiber-optic arrays. *Nature Biotechnology*, 20(April), pp.353–358.

Yeo, G. & Burge, C.B., 2003. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Proceedings of the seventh annual international conference on Computational molecular biology - RECOMB '03*, pp.322–331.

Yokoseki, A., Shiga, A., Tan, C.-F., Tagawa, A., Kaneko, H., Koyama, A., Eguchi, H., Tsujino, A., Ikeuchi, T., Kakita, A., Okamoto, K., Nishizawa, M., Takahashi, H. & Onodera, O., 2008. TDP-43 mutation in familial amyotrophic lateral sclerosis. *Annals of neurology*, 63(4), pp.538–42.

Zhang, Y.-J., Caulfield, T., Xu, Y.-F., Gendron, T.F., Hubbard, J., Stetler, C., Sasaguri, H., Whitelaw, E.C., Cai, S., Lee, W.C. & Petrucelli, L., 2013. The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Human molecular genetics*, pp.1–30.

Zhang, Y.-J., Xu, Y., Dickey, C. a, Buratti, E., Baralle, F., Bailey, R., Pickering-Brown, S., Dickson, D. & Petrucelli, L., 2007. Progranulin mediates caspase-dependent cleavage of TAR DNA binding protein-43. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(39), pp.10530–4.

Zhang, Y.-J., Xu, Y.-F., Cook, C., Gendron, T.F., Roettges, P., Link, C.D., Lin, W.-L., Tong, J., Castanedes-Casey, M., Ash, P., Gass, J., Rangachari, V., Buratti, E., Baralle, F., Golde, T.E., Dickson, D.W. & Petrucelli, L., 2009. Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. *Proceedings of the National Academy of Sciences*, 106(18), pp.7607–7612.

Zhang, Z., Almeida, S., Lu, Y., Nishimura, A.L., Peng, L., Sun, D., Wu, B., Karydas, A.M., Tartaglia, M.C., Fong, J.C., Miller, B.L., Farese, R. V, Moore, M.J., Shaw, C.E. & Gao, F.-B., 2013. Downregulation of microRNA-9 in iPSC-derived neurons of FTD/ALS patients with TDP-43 mutations. *PloS one*, 8(10), p.e76055.

Zhou, H., Huang, C., Chen, H., Wang, D., Landel, C.P., Xia, P.Y., Bowser, R., Liu, Y.-J. & Xia, X.G., 2010. Transgenic rat model of neurodegeneration caused by mutation in the TDP gene. *PLoS genetics*, 6(3), p.e1000887.

# 7. APPENDIX

## A. Analysis of the consequences of alternatively spliced exons at the protein level and conservation status of surrounding regions

Secondary validation analysis of the genes identified in the Affymetrix microarray analyses confirmed *POLDIP3*, *BCL2L11*, *MADD*, *STAG2*, *FNIP1* and *BRD8* as undergoing TDP_43 dependent alternative splicing (Table 2-2). As has been mentioned previously, two genes, *POLDIP3* and *BCL2L11* were already identified in previous work as undergoing TD-43 dependent alternative splicing and for the purposes of this study, were not explored further. Following the confirmation of altered splicing profiles by means of RT-PCR, an analysis of the consequences of exon skipping or inclusion in the already identified genes was performed.

***STAG2***: Exon 30b of *STAG2* was found to undergo increased inclusion when TDP-43 was depleted in cell. From the translation point of view, the inclusion of exon 30b results in the in-frame insertion of 36 amino acid residues near the C-terminus of the protein. Analysis of exon 30b from the UCSC genome browser, revealed that exon 30b is only included in two transcripts (UCSC; GRCh37/hg19) with the region encompassing exon 30, 30b being highly conserved (Figure I-1). Interestingly the intronic regions proximal to exon 30b and 31 also appear to be well conserved with the mid-section having a lower conservation of nucleotides (Figure I-1b), which could suggest functional significance for this intron. The region in which TDP-43 was found to bind (intronic region closer to exon 31) is also well conserved.

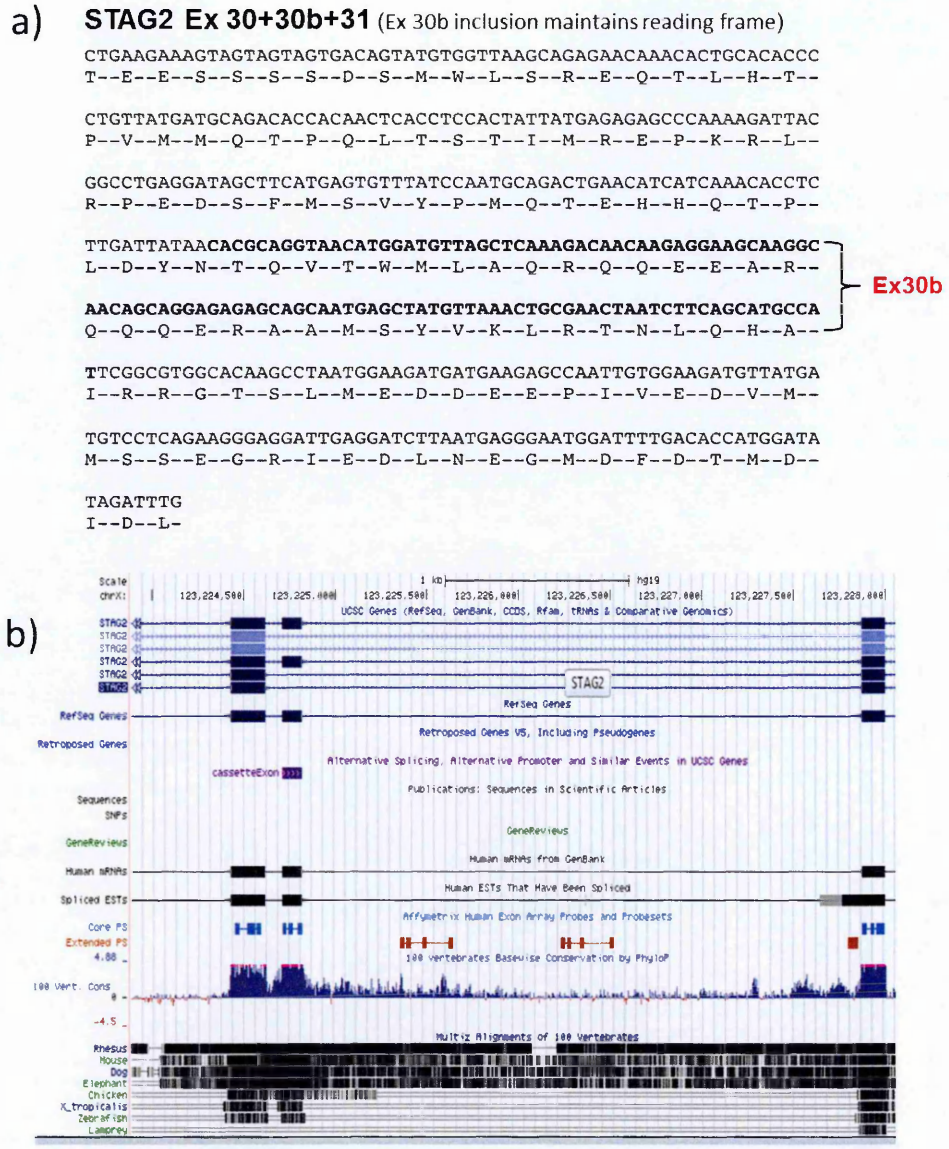



Figure I-1: (a) Inclusion of exon 30b results in the addition of 36 more residues closer to the C-terminus of the protein. (b) A screen shot from the UCSC genome browser analysis of the region surrounding *MADD* exon 30b showing high conservation of both exons 30 and 30b. The proximal downstream intronic region also appears to be well conserved suggesting a functional significance.

***MADD***: As determined by RT-PCR analysis, exon 31 of the *MADD* gene was found to undergo increased inclusion upon TDP-43 depletion. At the protein level, the skipping of exon 31 introduces seven residues followed by a stop codon (Figure I-2a). Furthermore, exon 31 is highly conserved with islands of highly conserved nucleotides in both the upstream and downstream intronic sequences (Figure I-2b).

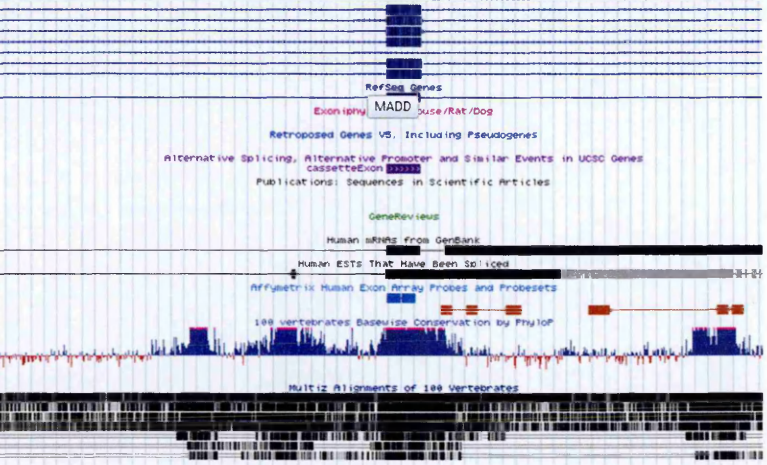
a)

**MADD Ex 30+31+32** (Ex 31 skipping introduces a stop codon)

```
GACCTGAATTGGGTGGCGAGTTCCCTGTGCAGGACCTGAAGACTGGTGAGGGTGGCCTGC
G--P--E--L--G--G--E--F--P--V--Q--D--L--K--T--G--E--G--G--L--   Ex30

TGCAGGTGACCCTGGAAGGGATCAACCTCAAATTCATGCACAATCAGGTTTTCATAGAGC
L--Q--V--T--L--E--G--I--N--L--K--F--M--H--N--Q--V--F--I--E--
                                                               Ex31
TGAATCACATTAAAAAGTGCAATACAGTTCGAGGCGTCTTTGTCCTGGAGGAATTTGTTC
L--N--H--I--K--K--C--N--T--V--R--G--V--F--V--L--E--E--F--V--

CTGAAATTAAAGAAGTGGTGAGCCACAAGTACAAGACACCAATG
P--E--I--K--E--V--V--S--H--K--Y--K--T--P--M-                   Ex32
```



```
GACCTGAATTGGGTGGCGAGTTCCCTGTGCAGGACCTGAAGACTGGTGAGGGTGGCCTGC
G--P--E--L--G--G--E--F--P--V--Q--D--L--K--T--G--E--G--G--L--   Ex30

TGCAGGTGACCCTGGAAGGGATCAACCTCAAATTCATGCACAATCAGTTCCTGAAATTAA
L--Q--V--T--L--E--G--I--N--L--K--F--M--H--N--Q--F--L--K--L--   Ex32

AGAAGTGGTGAGCCACAAGTACAAGACACCAATG
K--K--W--*-
```

b)




**Figure I-2 : (a) Skipping of exon 31 results in the creation of seven new amino acids and stop codon. (b) A screen shot from the UCSC genome browser of *MADD* exon 31 analysis showing high conservation for this exon, together with the proximal upstream and downstream intronic that similarly appear to be highly conserved.**

***FNIP1***: *FNIP1* exon 7 was found to undergo increased exclusion when TDP-43 was depleted. The skipping of this exon is in-frame (Figure I-3a). Similar to previously analysed exons, exon 7 of *FNIP1* is highly conserved. In addition, both immediate upstream and downstream intronic regions surrounding exon 7 are similarly highly conserved (Figure I-3b). The proximal upstream intronic region also has another area (further upstream) of high conservation. In the analysis of TDP-43 dependent splicing in FNIP1 using minigenes, no

change in splicing profile was obtained. Furthermore, within the intronic regions surrounding FNIP1, there were no obvious TG stretches.

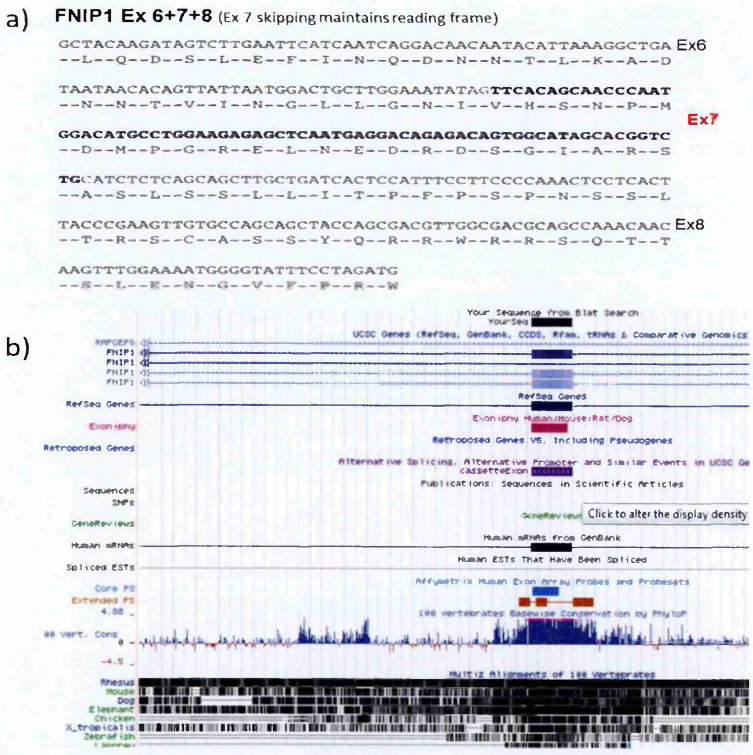



Figure I-3: (a) Translation of the FNIP1 amino acid sequence with exon 7 skipped is in-frame. (b) A screen shot from the UCSC genome browser depicting the highly conserved exon 7 and surrounding intronic regions.

***BRD8***. Exon 20 of BRD8 was also shown to undergo TDP-43 dependent inclusion i.e depletion of TDP-43 resulted in increased inclusion (Figure I-4a). Interestingly, the region encompassing exon 20 and proximal intronic area is highly conserved, despite exon 20 being quite short (45 bp) (Figure I-4b).

a) **BRD8 Ex19+20+21** (Ex 20 inclusion maintains reading frame)

```
CAATTCTTGGCCACGCAGTTGATTATGCAAACATCCGAGTCTGGGATCAGTGCTAAAAGT Ex19
-Q--F--L--A--T--Q--L--I--M--Q--T--S--E--S--G--I--S--A--K--S-

CTTCGAGGGAGAGATTCTACCCGCAAACAGGATGCTTCAGAGAAGGACAGTGTCCCAATG Ex20
-L--R--G--R--D--S--T--R--K--Q--D--A--S--E--K--D--S--V--P--M-

GGCTCTCCTGCCTTCCTTCTCTCTCTCTTTATGGGACACGAGTGGGTTTGGCTGGATTCT
-G--S--P--A--F--L--L--S--L--F--M--G--H--E--W--V--W--L--D--S-

GAACAAGATCATCCCAATGACTCTGAGTTGAGCAATGACTGCAGGTCCCTCTTCAGCTCA Ex21
-E--Q--D--H--P--N--D--S--E--L--S--N--D--C--R--S--L--F--S--S-

TGGGACTCCAGTCTGGATCTTGATGTGGGCAACTGGAGGGAAACTGAGGATCCAGAGGCT
-W--D--S--S--L--D--L--D--V--G--N--W--R--E--T--E--D--P--E--A-

GAGGAACTAGAGGAAAGCAGCCCGGAGAGAGAACCTAGTGAACTGCTTGTTGGGGATGGA
-E--E--L--E--E--S--S--P--E--R--E--P--S--E--L--L--V--G--D--G-

GGCAGTGAGGAATCTCAGGAAGCGGCAAGGAAAGCCAGCCACCAGAACCTCCTCCACTTT
-G--S--E--E--S--Q--E--A--A--R--K--A--S--H--Q--N--L--L--H--F-

CTCTCTGAG
-L--S--E-
```

b)

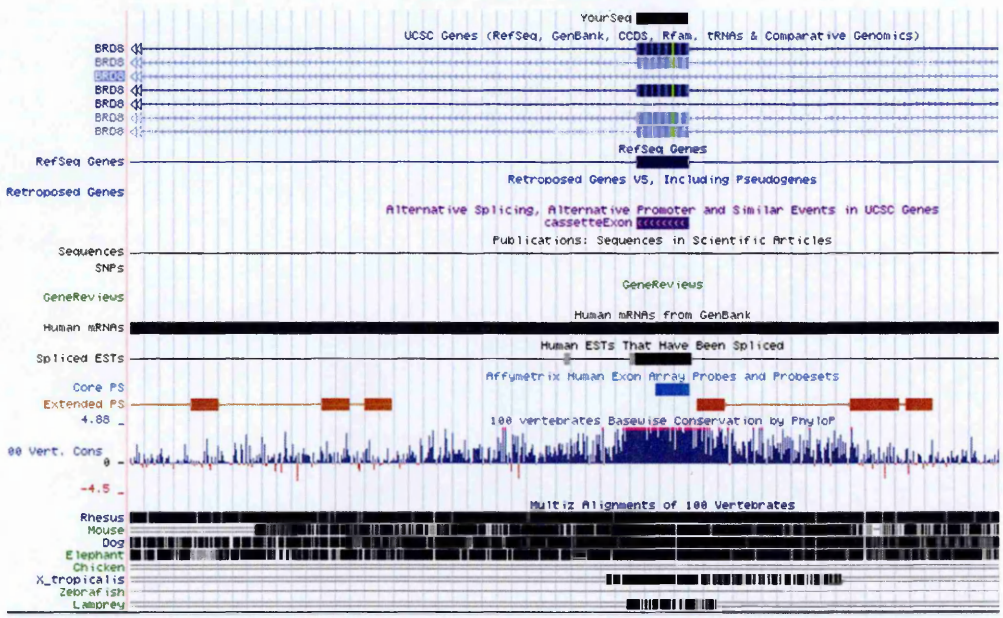



Figure I-4: (a) Inclusion of *BRD8* exon 20 does not change the reading frame and results in a slightly bigger protein. (b) A screen shot from the UCSC genome browser depicting a highly conserved region encompassing exon 20 with upstream and downstream proximal intronic regions.

## B. Supplementary Figures

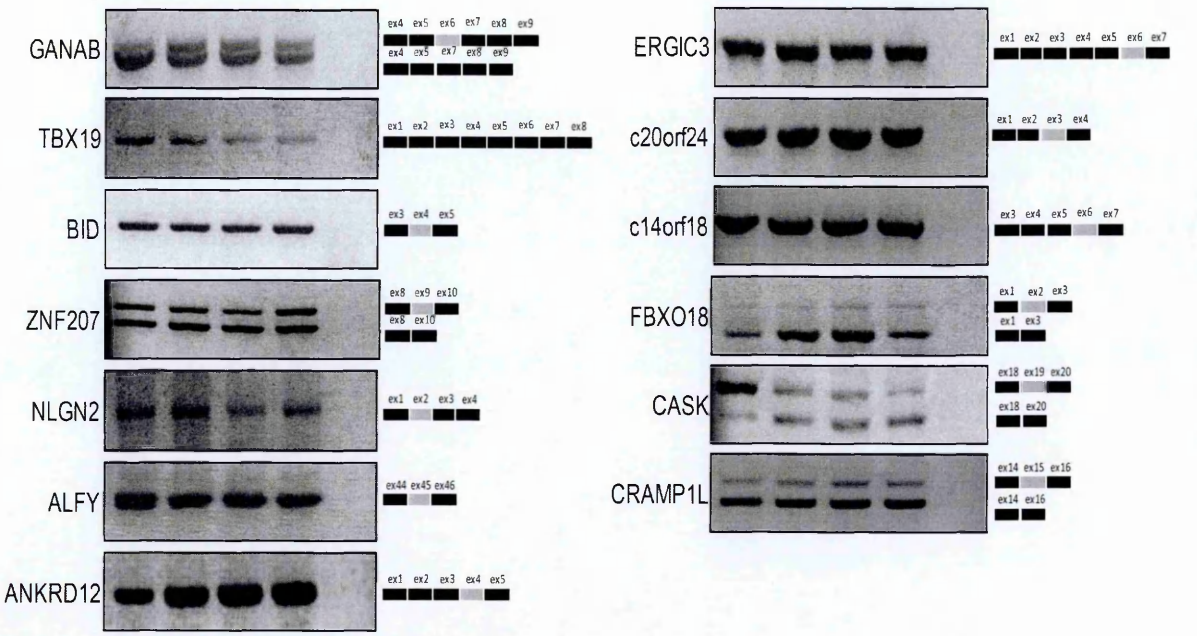



Figure I-5: RT-PCR analysis of the rest of genes validated by RT-PCR and not found to undergo TDP-43 dependent alternative splicing in the relevant exons. Genes presenting more than a 2-fold change in their splicing profile were selected for subsequent validation. The putative alternative exon for each gene is highlighted in grey and indicated according to Ensembl nomenclature as well as the amplified exons in the RT-PCR analysis. Only 4 out of the 19 genes analysed showed a change in alternatively spliced exons that matched the selection criteria.

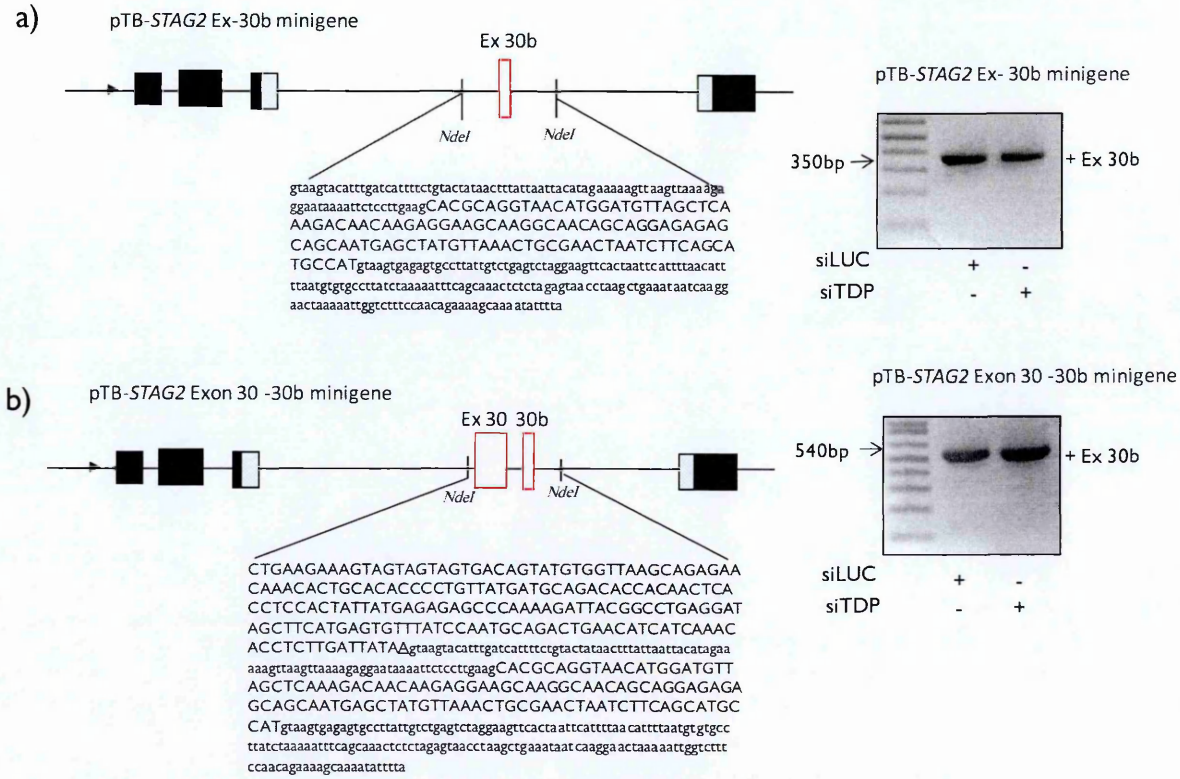



Figure I-6: Initial pTB-*STAG2* minigenes used to analyse TDP-43 dependent inclusion of exon 30b. No changes were observed with these minigenes and in fact in both cases exon 30b was 100% recognized and included. From these minigenes it was evident that a wider context of the sequences flanking exon 30b was necessary to determine where TDP-43 was binding to result in the exclusion of exon 30b under physiological levels of TDP-43 expression.