Mechanisms of Neurodegeneration in ALS and FTD

by

Brittany N. Flores

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2018

Doctoral Committee:

Assistant Professor Sami J. Barmada, Chair
Professor Robert S. Fuller
Professor Henry L. Paulson
Associate Professor Peter K. Todd

Brittany N. Flores

bnflores@umich.edu

ORCID iD: 0000-0002-5614-6901

Acknowledgements

First, I would like to thank my dissertation mentor Dr. Sami Barmada, who has been an excellent source of motivation, knowledge, and support throughout my training. I thank my dissertation committee members Dr. Robert Fuller, Dr. Henry Paulson, and Dr. Peter Todd for their support of my work with valuable suggestions during each committee meeting. The input received from each committee member has strengthened the quality of my work. I thank the members of the Barmada laboratory for their help over the years and for making the laboratory a positive and productive environment. I would like to give a special thanks to Xingli Li for teaching me techniques, and to Roberto Miguez for developing and optimizing computational tools for the automated analyses. I thank members of the Todd lab for giving me the opportunity to collaborate on several projects over the years. I thank the staff of the Cellular and Molecular Biology Graduate Program and the Program in Biomedical Sciences for all of their hard work and support over the years. Other collaborators, contributors, and sources of support are acknowledged throughout the dissertation.

Table of Contents

List of Figures

List of Tables

Abstract

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are devastating neurodegenerative conditions that share key clinical, pathologic, and genetic characteristics. Neuronal inclusions rich in the RNA binding protein TDP43 are found in the majority of ALS and FTD. Moreover, the most common cause of familial ALS and FTD is a hexanucleotide (G_4C_2) repeat expansion mutation within the first intron of chromosome 9 open reading frame 72, or *C9orf72*. Mutant *C9orf72* transcripts undergo repeat associated non-AUG (RAN) translation, generating five unique dipeptide repeat proteins (DPRs) that accumulate in degenerating neurons in *C9orf72*-associated ALS/FTD, but their significance in disease pathogenesis remains unclear. My dissertation investigates *C9orf72* RAN peptides, TDP43 deposition, and their respective contributions to neurodegeneration. My central hypothesis is that *C9orf72* RAN peptides disrupt TDP43 metabolism, leading to neurodegeneration via TDP43-dependent RNA misprocessing.

My thesis addresses the molecular pathways responsible for neurodegeneration in ALS and FTD. Chapter 1 reviews central features of ALS and FTD, including an overview of the proposed mechanisms of *C9orf72*-related neurodegeneration and aspects of TDP43 deposition. I first determined whether *C9orf72* RAN peptides, and more specifically which RAN peptides, are toxic to neurons. In collaboration with Dr. Magdalena Ivanova, we synthesized short polymers corresponding to the three sense-strand *C9orf72* RAN products, analyzed their structures by electron microscopy and

assessed their relative toxicity when applied to rodent primary cortical neurons. In doing so, we observed unique structural features for each dipeptide that correlated with their cellular internalization and relative toxicity. This work is described in further detail in Chapter 2.

I next began investigating the intrinsic properties of TDP43 that are critical for downstream neuronal toxicity in disease models. TDP43 binds thousands of transcripts, particularly UG-rich sequences, and TDP43-dependent toxicity is tightly tied to its ability to recognize RNA. Intramolecular interactions between TDP43's RNA binding domains, mediated by a salt bridge, are necessary for maintaining specificity for UG sequences. How sequence specificity of TDP43 binding to RNA affects TDP43 accumulation and survival remain unclear. Here, I show that genetically engineered mutations disrupting the TDP43 salt bridge reduce the affinity of nucleic acid binding and eliminate recognition of its native RNA targets. These same mutations dramatically destabilize TDP43, alter nuclear localization and abrogate toxicity upon overexpression in primary neurons. High-throughput RNA sequencing and splicing analyses indicated that TDP43 accumulation predominantly affects transcripts encoding components of the ribosome and oxidative phosphorylation pathways. These studies are illustrated in Chapter 3.

Chapter 4 describes relevant preliminary work with implications of a connection between the mutant *C9orf72* repeat expansion and TDP43 deposition. Briefly, I demonstrate that G_4C_2 oligonucleotides are recognized by TDP43 variants containing salt bridge-disrupting mutations, and co-expression of G_4C_2 and TDP43 enhance cytoplasmic mislocalization and neuronal toxicity. Chapter 5 concludes the dissertation outlining the

next steps moving forward with this work. Taken together, this dissertation uncovers novel disease pathways that can be targeted for therapy development.

Chapter 1

Introduction

1.1 Overview

Amyotrophic lateral sclerosis (ALS) is a devastating disease marked by progressive loss of motor neurons, muscle weakness, and eventually death. Approximately half of ALS patients present behavioral, personality, or language abnormalities associated with frontotemporal dementia (FTD)[1]. Not only do these diseases share overlapping clinical features, but also genetic and pathological characteristics[2-5]. Neuronal and glial cytoplasmic inclusions rich in the RNA binding protein TDP43 are found in the majority of individuals with ALS and FTD[6,7], implying the existence of a conserved pathogenic cascade responsible for neurodegeneration in both conditions. In support of this conclusion, many of the same mutations that cause familial ALS also result in familial FTD. First identified in 2011, the most prevalent mutation responsible for both familial ALS and FTD, accounting for up to 40% of inherited disease in some populations, consists of a hexanucleotide (GGGGCC) repeat expansion mutation in the first intron of chromosome 9 open reading frame 72, or *C9orf72*[4,5]. This thesis utilizes a model of ALS and FTD to investigate the molecular pathways leading to neurodegeneration in ALS and FTD with special focus on *C9orf72* and TDP43. In this introductory chapter, I review key overlapping features between ALS and FTD, the proposed mechanisms of *C9orf72*-related neurodegeneration, and aspects of TDP43

pathogenesis. The aim of this overview is to put into context the dissertation work on ALS and FTD disease pathogenesis described in subsequent chapters.

1.2 ALS and FTD clinical spectrum

ALS and FTD are thought to represent a continuous disease spectrum. Affecting around 2-3 per 100,000 individuals worldwide[8], ALS is the most common form of motor neuron disease caused by the degeneration of upper and lower motor neurons of the motor cortex and spinal cord[9,10]. This adult-onset neurodegenerative disease, with a peak age of onset between the ages of 40 and 70, rapidly leads to muscle weakness, paralysis, and ultimately death due to respiratory failure within two to five years from which symptoms are presented[9]. While the majority of ALS cases are sporadic, approximately 10% of ALS cases are familial and are usually inherited in an autosomal dominant pattern[11]. ALS patients may also present behavioral, personality, or languages abnormalities associated with FTD[2,3,12]. FTD is characterized by the loss of the frontal and anterior temporal lobes of the brain. FTD is the second most common form of dementia after Alzheimer's disease in people under the age of 65, with an incidence of 3-4 per 100,000 individuals[13]. Up to 40% of FTD cases are familial, and the majority of familial events are inherited in an autosomal-dominant manner[14]. Approximately 15% of FTD patients present clinical symptoms consistent with motor neuron dysfunction that is indistinguishable from ALS[11].

These rapidly progressive diseases have proven to be difficult to treat due to our incomplete understanding of the fundamental mechanisms leading to cell death. Riluzole, an anti-glutamatergic agent and the first drug approved by the FDA in 1995 to treat ALS, prolongs lifespan by a modest 2-3 months[15,16]. Since then, over 60 molecules have been

studied as a potential treatment for ALS, but the majority of drugs have failed to achieve efficacy in clinical trials[17]. In 2017, the FDA approved a second drug to treat ALS, edaravone, which slowed the progression of motor dysfunction over 6 months of treatment in patients with less disease severity[18,19]. Nonetheless, these drugs do not significantly extend the lifespan of affected individuals. One reason for this apparent disconnect may be the genetic heterogeneity of the population under study. With a better understanding of the genes associated with ALS and FTD and the molecular pathways responsible for neurodegeneration, more effective therapies can be developed.

1.3 The genetic overlap between ALS and FTD

The notion that ALS and FTD are fundamentally related and represent a continuous disease spectrum is supported further by evidence that mutations in several genes contribute to either familial ALS or FTD or both. Many of these disease-causing mutations impair similar molecular pathways (Table 1.1). Mutations in genes such as *PGRN*[20,21] or *MAPT*[22,23] are exclusively associated with FTD and *SOD1*[24] for ALS. However, mutations in genes such as *UBQLN2, VCP, FUS, TBK1, MATR3, TARDBP,* and *C9orf72* cause both ALS and FTD (Table 1.1). A hexanucleotide (GGGGCC) repeat expansion mutation in the first intron of chromosome 9 open reading frame 72, or *C9orf72,* is the most prevalent mutation responsible for both familial ALS and FTD, accounting up to 40% of inherited disease in some populations[4,5]. Mutations in the gene encoding TDP43, *TARDBP* (transactive response element DNA-binding protein 43), are rare in ALS and FTD, representing ~5% of affected individuals. However, pathologic cytoplasmic accumulations of TDP43 in affected brain regions and motor neurons are found in ~95% of ALS cases and over 50% of FTD

patients[6,25-27]. Over 40 mutations have been identified in TDP43 in ALS and FTD patients, and these mutations affect subcellular localization[28], aggregation[29], stability[30,31], and stress granule assembly[32,33]. Many of the genes associated with ALS, FTD, or both have been found to cause abnormalities in RNA metabolism[34] and impair protein homeostasis[35]. Genetic mutations associated with the RNA binding protein FUS and pathological cytoplasmic inclusions are observed in a small portion of patients with ALS and FTD[36,37]. Patients with mutations in other RNA binding proteins linked to ALS and FTD, such as TAF15 and EWSR1, also develop cytoplasmic inclusions in diseased brains[38,39]. Mutations in hnRNPA1 and hnRNPA2B1 RNA binding proteins result in TDP43 pathology[40]. Mutations in genes involved in protein homeostasis associated with ALS and FTD include *SQSTM1*[41], *VCP*[42], *OPTN*[43], *TBK1*[44], and *UBQLN2*[45], and these mutations result in TDP43 pathology (Table 1.1). The frequent presence of TDP43 pathology in ALS and FTD among different genes with disease-causing mutations strongly suggests a common pathogenic mechanism involving impaired RNA metabolism and protein homeostasis (Table 1.1).

1.4 *C9orf72*-related disease pathogenesis

Nucleotide repeat expansions contribute to several neurological diseases, including Huntington's disease, fragile X syndrome, and the spinocerebellar ataxias[46,47]. How exactly repeat expansions contribute to each disease state depends on the specific disease in question. In many conditions, the nucleotide repeat occurs within the open reading frame of a protein-coding gene, resulting in a repeating stretch of amino acids (usually glutamine) that renders the mutant protein prone to aggregation and neurotoxic. However, when the nucleotide repeat is located in non-coding regions of the genome, as with the *C9orf72* hexanucleotide repeat,

neurodegeneration may stem from one or more alternative mechanisms. In order to understand how *C9orf72* mutations cause neurodegeneration in ALS and FTD, it is important to understand the gene itself and features of the repeat expansion in *C9orf72*-mutation carriers.

1.4.1 The *C9orf72* gene, hexanucleotide repeat expansion, and transcript isoforms

In 2011, the most common cause of familial forms of both ALS and FTD was identified as a hexanucleotide (GGGGCC, or G_4C_2) expansion mutation within the first intron of the chromosome 9 open reading frame 72, or *C9orf72*[4,5]. This expansion mutation accounts for up to 40% of familial ALS, 29% of familial FTD, and 88% of both ALS and FTD cases[4,5]. The repeat expansion is most frequent among Caucasian populations in Europe and North America, particularly in Finland. In contrast, *C9orf72* mutations are rarely observed in Asian populations[4,5]. The mutation is inherited in an autosomal dominant manner with age-dependent penetrance[48,49]. However, it is possible for an affected individual to have an unaffected parent due to incomplete penetrance or a *de novo* variant, a phenomenon seen in nearly 10% of ALS patients and 5% of FTD cases[11]. Symptom onset occurs, approximately, between 30 to 80 years of age, with mean age onset in the late 50s[49]. Normal G_4C_2 repeat length observed in unaffected individuals is from 2 to 25 units; pathogenic alleles range from over 30 to hundreds or thousands of repeats[4,5,50]. Due to the high GC-content of the repeat expansion, it is difficult to accurately detect large repeat sizes (greater than 80 units), via techniques such as repeat primed polymerase chain reaction (PCR) or Southern Blot hybridization. Moreover, repeat instability of large expansions may lead to somatic mosaicism, making it difficult to accurately quantify repeat size. High variability in G_4C_2 repeat size exists between different regions of the brain and tissues[51]. Even within the

same individual, repeat length may also vary among different tissues[51-55]. Therefore, a correlation between repeat length and disease severity has not been convincingly identified among studies[51,53,55]. Taken together, the G_4C_2 expansion mutation has made it challenging to define a clear correlation between repeat length, disease onset, and disease severity.

The *C9orf72* gene consists of 11 exons and is transcribed into three major transcript variants encoding two protein isoforms. The transcript variants use alternative first exons, and depending on the transcript variant, the G_4C_2 expansion is either located in the promoter region or in the first intron of the gene (Fig. 1.1A). Transcript variants using exon 1a, which therefore include the hexanucleotide repeat, produce either a 222 or 481 amino acid protein isoform (short or long isoform, respectively). The third transcript variant using exon 1b generates the long protein isoform. Both fibroblasts derived from *C9orf72*-ALS patients and motor neurons derived from induced pluripotent stem cells (iPSCs) of *C9orf72*-ALS individuals suggest selective use of exon 1a compared to exon 1b[56].

Experiments using antibodies against either C9orf72 short or long protein isoforms indicate differences in subcellular localization, implying distinct cellular functions. By immunohistochemistry, the long isoform showed diffuse cytoplasmic staining in neurons[57] and is thought to be involved in autophagy[58-62]. In contrast, the short C9orf72 protein isoform is localized to the nuclear membrane in patient brain tissue. The short isoform is suggested to be involved in nucleocytoplasmic transport, as it has been reported to associate with nucleocytoplasmic transport proteins such as Ran-GTPase and Importin β[57]. Interestingly, loss of the short isoform nuclear membrane staining correlated with cytoplasmic TDP43 mislocalization in spinal motor neurons of *C9or72*-ALS cases[57]. Yet, it remains unclear the

precise biological function of each protein isoform, and exactly how these may be involved in disease.

Since the identification of the repeat expansion in *C9orf72*, there have been several studies investigating the clinical, pathological, and mechanistic features of *C9orf72* mutations in ALS and FTD. While the *C9orf72* field has rapidly gained insight into the potential mechanisms by which *C9orf72* mutations contribute to neurodegeneration, the primary driver of disease pathogenesis remains unclear. Three possible, non-mutually exclusive mechanisms have been proposed, including 1) loss of *C9orf72* function, 2) RNA-based gain of function toxicity, and 3) protein-based gain of function toxicity through non-canonical translation of the expanded repeat. These proposed mechanisms are illustrated in Figure 1.1 and are described in further detail in the space that follows.

1.4.2 Loss of *C9orf72* function

One proposed mechanism by which *C9orf72* may cause neurodegeneration is a loss of function toxicity due to decreased RNA and protein levels (Fig. 1.1B). Prior studies discovered decreased *C9orf72* mRNA and protein expression levels in both iPSCs and post-mortem brain tissue from *C9orf72*-mutation carriers[5,48,50,54,57,63-65]. The reduced expression of *C9orf72* suggests that haploinsufficiency could be responsible for disease. Since the hexanucleotide repeat expansion is G-rich, haploinsufficiency may stem from the formation of stable G-quadruplex structures that disrupt transcription[66,67]. Alternatively, the hexanucleotide repeat expansion may result in hypermethylation of the *C9orf72* locus. Upstream of the G_4C_2 repeat expansion lies a CpG island. Hypermethylation of this island is associated with expanded G_4C_2 repeats, resulting in decreased expression of *C9orf72* RNA levels in the blood, brain, and spinal cord[52,63,68]. Similar to other neurological disorders

involving repeat expansions located in noncoding regions, such as Friedrich ataxia[69] and fragile X syndrome[70,71], repeat-dependent hypermethylation results in gene expression silencing[52,72,73].

C9orf72 encodes an uncharacterized protein that is thought to function as a DENN (differentially expressed in normal and neoplastic cells) domain protein, regulating membrane trafficking in cooperation with Rab-GTPase switches that are involved in autophagy and endocytic transport. Prior studies demonstrated partial co-localization of C9orf72 with Rab proteins, such as Rab5, Rab7, and Rab11, implicated in autophagy and endocytic transport in primary neurons and *C9orf72*-ALS spinal cord motor neurons[74]. Furthermore, cell lines depleted of C9orf72 via siRNA resulted in autophagy dysregulation and inhibition of endocytosis[74]. Recent evidence suggests that the long C9orf72 protein isoform forms a complex with SMCR8 and WDR41, acting as a GDP/GTP exchange factor (GEF) for Rab proteins involved in vesicle trafficking and autophagy induction[60-62,74]. Investigations conducted by reducing expression of C9orf72 in cell lines and primary neurons provide evidence that C9orf72 regulates autophagy initiation by functioning as a Rab1a effector and interacting with ULK1 autophagy initiation complex[59,60,75]. In sum, these studies predict that the C9orf72 protein physiologically functions as a guanine exchange factor, though the precise role of C9orf72 in autophagy and endocytic transport leading to ALS and FTD pathogenesis remain unclear. Moreover, whether the hexanucleotide repeat expansion alters autophagy initiation or endocytic transport remains to be determined.

A number of studies have depleted *C9orf72* in different model systems to further understand the normal function of the protein. Knocking down the zebrafish orthologue of *C9orf72* with antisense morpholino oligonucleotides resulted in axonopathy and motor

defects; these phenotypes were rescued by expressing human *C9orf72* mRNA[64]. These findings were the first to provide *in vivo* evidence that loss of *C9orf72* function disrupts motor neuron function. In corroboration, deletion of the *C. elegans* orthologue of *C9orf72* resulted in motor neuron degeneration and locomotor deficits[76]. Notably, however, reduction of endogenous *C9orf72* RNA (30-40%) in mouse brain did not result in motor or behavioral defects[77], and neural-specific *C9orf72* knockout mice failed to produce any neuronal abnormalities[78]. Other *C9orf72* knockout mouse models demonstrated splenomegaly and lymphadenopathy phenotypes, suggesting immune system dysregulation; C9orf72-deficient spleens displayed elevated p62 and LC3 levels, indicating autophagy impairment[79-81].While reduction of C9orf72 protein in mouse models does not appear to induce neuronal toxicity, it does recapitulate pathological features of ALS and FTD. For example, knockdown of *C9orf72* resulted in the accumulation of p62 in primary neurons, and this phenotype was rescued upon expression of the long C9orf72 isoform[61]. Interestingly, cytoplasmic aggregates of phosphorylated TDP43 were observed in some cell types upon depletion of C9orf72[61]. However, the precise molecular mechanisms regarding C9orf72 and its role in membrane trafficking and autophagy and its contribution to ALS and FTD remain unclear. Taken together, *C9orf72* loss of function may not be driving disease pathogenesis since depletion of *C9orf72* is not sufficient to induce cell death; rather, a second stressor may be required for triggering neurodegeneration. Two gain of function mechanisms are discussed in the subsequent sections as potential mechanisms of disease pathogenesis.

1.4.3 RNA gain of function

The second proposed mechanism by which the G_4C_2 repeat expansion may cause disease is RNA-mediated toxicity and is associated with the formation of distinct RNA foci

composed of the expanded repeat-containing RNA predominantly found in the nucleus[5,77,82-84] and occasionally in the cytoplasm[56,65,77] (Fig 1.1C). Several neurological diseases caused by nucleotide repeat expansions such as myotonic dystrophy (DM), fragile X-associated tremor/ataxia syndrome (FXTAS) and spinocerebellar ataxia type 8 (SCA8), type 10 (SCA10), and type 31 (SCA31) demonstrate RNA-mediated toxicity and exhibit RNA foci[85,86]. Prior studies identified sense and antisense nuclear RNA foci in *C9orf72*-containing ALS/FTD patient tissue and iPSC-derived neurons[77,82,87-89], indicating that inappropriate sequestration of RNA binding proteins and splicing factors may be a causative factor in neurodegeneration. The existence of nuclear RNA foci observed in individuals carrying *C9orf72* mutations suggests abnormal alterations in RNA metabolism. Due to the G-rich content, G_4C_2 repeats can form highly stable DNA and RNA G-quadruplexes and promote DNA/RNA hybrids, or R-loops, *in vitro*[66,67,90,91]. Interestingly, a prior report suggests that the formation of G-quadruplex structures by G_4C_2 repeats is length-dependent, which may facilitate the formation of RNA foci and bind to proteins. Furthermore, Haeusler *et al.* suggests that secondary structures formed by G_4C_2 repeats bind to RNA binding proteins in a conformation-dependent manner. In sum, distinct secondary structures of the hexanucleotide repeat expansion could provide insight on the pathogenic mechanisms of neurodegeneration.

RNA-based toxicity is thought to arise from the sequestration of essential RNA binding proteins and splicing factors by sense and antisense repeat-containing transcripts[65,84,92], thereby inhibiting the normal functions of these proteins. A number of RNA binding proteins have been reported to co-localize with sense and anti-sense RNA foci or interact with G_4C_2 or C_4G_2 repeats in patient tissues and cell models, including ADARB2, FUS, hnRNPs, nucleolin, SRSF2, ALYREF, and Pur-α[56,65,67,84,88,92,93]. Using neuronal cell

lines expressing G_4C_2 repeats, sequestration of RNA binding proteins hnRNP-H and FUS, and RanGAP1, a nucleocytoplasmic trafficking protein, also interact directly with G_4C_2 repeats *in vitro*[94]. Mislocalization of RanGAP1 has been reported in flies overexpressing the repeat expansion and in cells from ALS patients harboring the *C9orf72* expansion mutation[94,95], indicating defects in nuclear import as a mechanism responsible for disease. In support of this observation, cells treated with ASOs targeting the repeat expansion rescued nucleocytoplasmic transport defects[94]. Taken together, changes in RNA processing and nucleolar stress caused by the *C9orf72* repeat expansion *in vitro* and *in vivo* suggest that G_4C_2 repeats may lead to disease phenotypes.

There is conflicting data on the extent to which RNA foci are responsible for toxicity. Overexpression studies using 'RNA only' based constructs in zebrafish suggest the repeat-containing RNA is sufficient to induce toxicity[96]. Conversely, transgenic flies expressing interrupted G_4C_2 repeats containing stop codons to prevent translation were not toxic[97]. Another transgenic fly model expressing G_4C_2 repeats embedded within an intron formed nuclear RNA foci, but the flies did not exhibit a neurodegenerative phenotype[98]. While RNA foci are detected in neurons and glial cells in *C9orf72* ALS and FTD cases, there seems to be no correlation between the burden of RNA foci with clinical and pathological features of disease; rather, cells with antisense RNA foci were reported to correlate with a later age at disease onset[87,99]. This would imply that RNA foci may be protective. Both sense and antisense RNA foci are not very abundant in cells, and one study suggests that these RNA foci are composed of a single molecule of repeat-containing RNA[100]. Therefore, the extent to which RNA foci impact disease remains debated.

1.4.4 Protein gain of function

The third proposed mechanism by which G_4C_2 repeats lead to neurodegeneration is by a gain of function in which expanded transcripts may undergo a unique mechanism of non-canonical translation, known as repeat-associated non-AUG (RAN) translation[101,102]. RAN translation has been shown to occur in other neurodegenerative diseases containing nucleotide repeat expansions, such as SCA8[103], myotonic dystrophy type 1 (DM1)[103], FXTAS[104], and Huntington's disease[105]. A number of studies have shown that *C9orf72*-linked ALS and FTD patient brains contain RAN translation products generated from both the sense and antisense strands of the hexanucleotide repeat in every reading frame[83,84,101,102]. This includes the following five different dipeptide repeat (DPR) proteins: glycine-alanine (GA), glycine-proline (GP), and glycine-arginine (GR) produced from the sense strand, and glycine-proline (GP), alanine-proline (AP), and proline-arginine (PR) derived from the antisense strand (Fig. 1.1D). Although GP peptides are produced from both sense and antisense strands, the GP product derived from the sense strand has a predicted unique C-terminal sequence, while the GP product derived from the antisense strand contains a stop codon immediately after the repeat[82].

With the added discovery of the different dipeptides generated from the hexanucleotide repeat expansion, the neuropathology of *C9orf72*-mutation is complex. In conjunction with neuronal and glial TDP43-positive inclusions, *C9orf72* ALS and FTD cases uniquely show p62-positive, TDP43-negative inclusions within neurons throughout different regions of the brain, particularly in the hippocampus, cerebral cortex, and cerebellum[106-108]. By immunohistochemistry, studies have demonstrated that these TDP43-negative inclusions contain each of the RAN translation products derived from the hexanucleotide repeat

expansion[82-84,101,102]. Immunohistochemical analyses of *C9orf72*-positive cases suggest differences in abundance among the DPRs with poly-GA-positive inclusions being the most abundant followed by poly-GP and poly-GR in the cortex and cerebellum[84,102,109]. Poly-PA and poly-PR-positive inclusions are rarely observed, suggesting that there is either more transcription of the sense transcript or RAN translation is more effective from the sense transcript[109].

Previous studies suggest that accumulation of DPRs within neurons may precede TDP43 pathology in *C9orf72*-mutation carriers[102,110,111]. Additionally, DPR accumulation does not closely predict neuron loss, unlike TDP43[110,112,113]. Abundant TDP43 deposition is observed in the most vulnerable brain regions related to disease, while DPR accumulation is present in unaffected areas of the brain. The disconnect may imply that either 1) RAN translation is not the primary driver of neurodegeneration, or 2) DPRs triggers pathways that subsequently results in the mislocalization and accumulation of TDP43, and eventually inducing cell death. Alternatively, it is possible that the *C9orf72* expanded RNA alone, or in combination with DPRs, contributes to neurodegeneration. Nonetheless, the lack of correlation between *C9orf72* and TDP43 pathology does not imply that the two are not linked through similar molecular pathways that are responsible for neurodegeneration. Rather, it is possible that these two features of disease share overlapping molecular mechanisms. Future studies are warranted to establish a link.

To understand the significance of DPRs in ALS and FTD pathogenesis, numerous studies have investigated the consequences of survival from DPR expression *in vitro* and *in vivo* and the mechanisms by which they cause toxicity. In some model systems, expression of the DPRs alone is sufficient to induce cell death. Several studies suggest that the arginine-

rich dipeptides, GR and PR, are the most toxic DPRs. *Drosophila*, zebrafish, and cell models expressing individual DPRs, in which alternative codons replace the G_4C_2-repeating sequence and thereby excluding any effects of RNA alone, demonstrate that GR and PR peptides are the most toxic[96,97,114,115]. Other fly models and studies involving cultured cells, including primary neurons, show similar toxicity upon GR and PR peptide expression[114,116-118]. These arginine-rich containing peptides localize to the nucleus, induce nucleolar stress, impair both pre-mRNA splicing and biogenesis of ribosomal RNA, and inhibit protein translation in cellular and neuronal models[114,116-118]. From a proteomics approach, poly-PR was found to interact with mRNA binding proteins, ribosomal proteins, translation initiation factors, and translation elongation factors, suggesting that arginine-rich DPRs may cause disease by inhibiting global protein translation[119].

Other cellular pathways in which GR or PR peptides may trigger neurodegeneration are evident in *Drosophila* and yeast screens that identified impairment of nucleocytoplasmic components upon expression of poly-GR or poly-PR[95,120,121]. Not only are the DPRs implicated in disruption of nucleocytoplasmic transport, but also the G_4C_2 repeat RNA may have a direct impact. Hexanucleotide repeat-containing RNA directly interacts with RanGAP and is mislocalized in fly models expressing G_4C_2 repeats, neurons derived from iPSCs of *C9orf72*-ALS patients, and *C9or72* ALS brain tissue[94]. Interestingly, nuclear import defects observed in fly models and iPSCs from *C9orf72*-ALS patients are rescued by antisense oligonucleotides (ASOs) targeting the G_4C_2 repeat-containing RNA[94].

Although many reports suggest that GR and PR dipeptides are the most toxic species, GA dipeptides have been shown to be toxic in mammalian cell lines and primary neurons. Expression of poly-GA in cell culture models results in the formation of predominantly

cytoplasmic aggregates[114,118,122-124]. These poly-GA aggregates inhibit the ubiquitin proteasome system (UPS), induce ER stress, and cause cell death[118,123]. Mass spectrometry analysis identified an enrichment of proteins involved in the UPS and the cargo adaptor Unc119 is sequestered by poly-GA inclusions in primary neurons[122]. In addition to the cell culture models, Zhang *et al.*[125] generated transgenic mice expressing abundant poly-GA pathology in the central nervous system (CNS). The expression of poly-GA in these mice led to sequestration and disruption of HR23 proteins involved in proteasomal degradation and proteins involved in nucleocytoplasmic transport. In sum, these studies suggest that each toxic DPR may uniquely contribute to neurodegeneration through different cellular pathways.

Of the studies that investigated PA or GP dipeptides, these RAN translation products have not resulted in toxicity in *Drosophila* and cell models[97,114,117,122]. Nonetheless, poly-GP is the only DPR produced from both sense and antisense G_4C_2 containing transcript. This dipeptide can be detected in the cerebral spinal fluid of *C9orf72*-ALS/FTD cases[126], making it a potentially important biomarker.

Whether one or more of these proposed mechanisms are primarily responsible for neurodegeneration has not been resolved. Determining the precise pathogenic mechanism is crucial for the development of effective therapies of ALS and FTD. In addition to investigating the mechanisms underlying *C9or72*-disease pathogenesis, it is essential to understand the molecular pathways affected by the accumulation of TDP43, a common feature in the brains and spinal cords of patients with ALS and FTD.

1.5 Features of TDP43 in ALS and FTD

For the majority of patients with ALS and FTD, including *C9orf72*-mutation carriers, the most striking pathologic characteristic is the cytoplasmic accumulation of TDP43[25]. TDP43 pathology is present in ~95% ALS cases and more than 50% of FTD cases[25-27]. In 2006, Neumann *et al*. identified TDP43 within neuronal and glial inclusions in the majority of ALS and FTD cases[6], suggesting that TDP43 is an important component in disease pathogenesis. Neuropathological features of TDP43 include nuclear exclusion, cytoplasmic inclusions, and/or aggregation. Other pathological features of TDP43 deposition include hyperphosphorylation, ubiquitination, and cleaved C-terminal fragments of TDP43 in the cytoplasm (~25 kDa and ~35kDa)[6,7,127]. Cytoplasmic TDP43 aggregates are detergent-insoluble and are mainly composed of the full-length protein or truncated C-terminal fragments of the protein. The mechanisms by which TDP43 accumulation results in disease pathogenesis remain unclear.

1.5.1 Structure and normal function of TDP43

A thorough understanding of the structure and function of TDP43 is critical for determining how TDP43 dysfunction leads to neurodegeneration (summarized in Fig. 1.2). Encoded by *TARDBP,* TDP43 was initially identified as a transcriptional repressor factor of HIV[128]. TDP43 is ubiquitously expressed and is necessary for early embryonic development[129-131]. TDP43 is a highly conserved 414 amino acid protein that shares homology to heterogeneous nuclear ribonucleoprotein (hnRNP) family members, which function in RNA metabolism. TDP43 is a RNA/DNA binding protein that harbors a nuclear localization signal (NLS) at the N-terminus, two RNA recognition motifs (RRM1 and RRM2), a proposed nuclear export signal (NES) within RRM2, and an intrinsically

disordered glycine-rich domain (GRD) at the C-terminal end (Fig. 1.2A). Monomers of TDP43 bind to each other to form homodimers *in vitro*[132-134], and the N-terminus of TDP43 has been shown to play a role in regulating homodimerization[134-136]. Furthermore, studies suggest that TDP43 dimerization is important for its normal function[134]. The GRD includes a glutamine/asparagine (Q/N) prion-like domain and is involved in protein-protein interactions, such as FUS, hnRNPA1 and hnRNPA2/B1[137-140]. A proteomics study in mammalian cell lines identified other TDP43 interacting proteins involved in RNA metabolism, such as splicing factors, and components involved in protein translation[140].

TDP43 is predominantly localized to the nucleus[6,141]. However, the presence of the NLS and NES enables TDP43 to actively shuttle between the nucleus and cytoplasm[141,142] with functions in RNA regulation such as control of transcription, splicing, transport, and translation[143-148]. In healthy neurons, a small portion of TDP43 is localized to the cytoplasm within RNA granules, functioning in the regulation of mRNA transport and local translation that are important for neuronal plasticity[149,150]. Under stress conditions, cytoplasmic TDP43 is recruited into stress granules[33,151,152]. Within these stress granules, TDP43 plays a role in trafficking and stabilizing mRNAs[153].

TDP43 binds thousands of RNA transcripts, particularly those containing UG/TG-rich sequences[143,144,154,155]. Both RRM1 and RRM2 are highly evolutionarily conserved domains[147]. Due to its aggregation propensity, currently there is limited structural data of TDP43; only structures of both RRMs bound to UG-rich RNA sequences have been determined[147,156,157]. RRM1 and RRM2 exhibit distinct binding characteristics. RRM1 is primarily involved in nucleic acid recognition and has a higher affinity for specific sequences, and while RRM2 alone binds little RNA, its presence enhances sequence

specificity of TDP43 binding[132,147,158,159]. TDP43 can recognize a minimum of 6 UG/TG repeats and binding affinity increases with the number of repeat elements[158]. Upon RNA binding, intramolecular interactions between RRM1 and RRM2, mediated by a salt bridge between Arg151 (located in RRM1) and Asp247 (located in RRM2) are necessary for maintaining TDP43's ability to recognize RNA[156].

The RNA binding activity of TDP43 is important for several RNA processes. *In vivo* studies using high-throughput RNA sequencing revealed that TDP43 preferentially binds to introns, 3' untranslated regions (UTRs), and non-coding RNAs[143,144,154]. RNA targets were enriched for Gene Ontology (GO) terms related to RNA metabolism, synaptic activity, and neuronal development[143,144,154]. TDP43 plays a role in alternative splicing by binding to UG-rich sequences near the 3' or 5' splice sites of exons. For example, TDP43 is involved in splicing the mRNA encoding the cystic fibrosis membrane conductor receptor (CFTR) by binding to the (TG)m repeat region near the 3'-splice site[158,160]. With the diverse set of functions for RNA regulation by TDP43, the loss of normal TDP43 function has the potential to disrupt many fundamental cellular processes and subsequently result in neurodegeneration.

1.5.2 Disease-associated *TARDBP* mutations

Over 40 mutations in *TARDBP* have been identified in both familial and sporadic ALS and FTD cases, with G298S, A315T, M337V, and A382T as the most frequent in patients[161]. TDP43 mutations have been found in ~3% of familial ALS and ~1.5% of sporadic ALS[162]. Majority of these mutations lie within the GRD. *TARDBP* mutations within the GRD reduce TDP43 solubility, enhance cytoplasmic TDP43 mislocalization, and affect TDP43 clearance[28,30,31,163,164]. Some mutations are located within the 3' or 5' UTRs[26,165], and these mutations may impair TDP43's ability to autoregulate. Only four disease-causing

mutations have been identified in the RNA recognition motifs: P112H, D169G, N259S, and K263E[166-169]. Very little is known about the effects of these mutations, but the P112H and D169G mutations do not appear to affect RNA binding[157,159]. Both D169G and K263E mutations result in an increase in thermal stability and resistance to aggregation *in vitro*[170]. Mutating K263 to alanine led to a small reduction on RNA binding, but this mutation had no effect on TDP43's ability to splice CFTR transcripts[156]. Overexpression of the N259S mutation in primary motor neurons resulted in cytoplasmic aggregates, but no studies have been conducted to determine its effect on binding[168]. Together, these data suggest that disease-associated TDP43 mutations located within the RRMs do not impair RNA binding. Despite the number of disease-associated *TARDBP* mutations, neurons of ALS and FTD patients show nuclear depletion and cytoplasmic accumulation of TDP43 even in the absence of a mutation.

How exactly TDP43 dysfunction leads to disease pathogenesis remains unclear. Because ALS and FTD patients exhibit a redistribution of nuclear TDP43 to cytoplasmic inclusions in neurons, both loss of function and gain of function mechanisms have been proposed. Given its essential role in RNA metabolism, loss of nuclear TDP43, accumulation of cytoplasmic TDP43, or both may have important consequences for RNA processing and subsequent neurodegeneration (Fig. 1.2B).

1.5.3 Loss of function mechanisms

Since nuclear exclusion of TDP43 is observed in ALS and FTD patients, neurodegeneration may stem from a TDP43 loss of function mechanism. The effect of the loss of TDP43 has been investigated in different model systems. Loss of *TDP-1* in *C. elegans*, the equivalent to human TDP43, results in subtle behavioral and locomotor

defects[171,172], but paradoxically displays a modest increase in lifespan[172,173]. Nevertheless, *TDP*-1 and human TDP43 have been demonstrated to be functionally conserved through RNA binding activity[147,172]. Knockdown of the *Drosophila* TDP43 homolog, *TBPH*, induces a neuromuscular phenotype and reduces lifespan[174]. Knockdown of TDP43 in zebrafish using antisense morpholino oligonucleotides results in shortened axons with abnormal branching and locomotor deficits [175,176]. Depletion of the TDP43 orthologs in zebrafish, *tardbp* and *tardbp1,* results in early lethality[177]. *TARDBP* null mice do not survive during embryonic development[129-131], and conditional *TARDBP* knockout embryonic stem cells fail to survive[178]. Together, these studies demonstrate that loss of TDP43 in most model systems results in phenotypes resembling neurodegeneration, and that TDP43 is essential for both development and survival.

At the molecular level, loss of TDP43 affects the expression and splicing of hundreds of different RNAs. TDP43 is a splicing repressor[143,144,179], and loss of TDP43 results in the inclusion of unannotated (cryptic) exons within target transcripts that lead to nonsense-mediated decay (NMD)[179-181]. In contrast, overexpressing TDP43 triggers exon skipping in most cases[182,183]. A study conducted in *Drosophila* either lacking *TBPH* or overexpressing TDP43 identified distinct differences in gene expression, despite both fly models presenting similar locomotor deficits. Knockdown of *TBPH* resulted in widespread gene upregulation, while flies overexpressing TDP43 led to the opposite effect[183]. Surprisingly, only 79 out of 1553 genes overlapped in both datasets, suggesting that different cellular pathways are affected based on TDP43 levels. Disease-associated *TARDBP* mutations are also capable of altering the expression and splicing of select RNAs through a loss of function mechanism[145].

1.5.4 Gain of function mechanisms

TDP43 regulates the expression and splicing of many RNA targets, including the expression and splicing of its own transcript by binding to an intron in the 3'UTR[137,144,146]. This allows for the autoregulation of *TARDBP* to maintain tight control of TDP43 protein levels via a negative feedback loop[146]. At high levels, TDP43 binds to its own transcript within the 3'UTR, triggers alternative splicing, destabilizes its mRNA, and reduces protein expression[143,144,146,184]. Degradation of its own mRNA occurs through NMD or exosome-mediated degradation[144,146]. Because TDP43 levels are important for cellular health, disruption of TDP43 autoregulation may play an important role in the cytoplasmic accumulation of TDP43. Indeed, increased mRNA and protein levels of TDP43 have been detected in ALS patients[185-188], emphasizing its significance to disease. Both overexpression and loss of TDP43 are toxic in cell and animals models[28,30,176,177,188-194]. Moreover, TDP43-related toxicity has been shown to be dose-dependent in primary neurons and in transgenic mice[28,30,190], suggesting that tight regulation of TDP43 levels is critical for cellular health.

Recent studies conducted in mice provide additional evidence for a gain of function mechanism. A *TARDBP* knock-in mouse model bearing a disease-associated mutation (Q331K) was created via CRISPR/Cas9[195]. These mice exhibited cognitive dysfunction and TDP43 autoregulation impairment, leading to an increase in TDP43 expression levels[195]. Moreover, changes in gene expression and splicing events were observed, including changes in other ALS-linked genes[195]. Another study using mice harboring a mutation within the endogenous *TARDBP* locus (M323K) observed novel splicing events associated with elevated TDP43 levels, suggesting a gain of splicing function[196].

It is well established that overexpression of WT TDP43 or disease-associated mutations in model systems, including primary neurons[28,30,176,189,197], rodents[190,193,198,199], *Drosophila*[194], yeast[200], and *C. elegans*[171,201], induce toxicity. In addition, overexpressing WT or mutant TDP43 recapitulates key features of sporadic ALS and FTD, including cytoplasmic mislocalization and aggregation of TDP43[28,141,190,200,201]. Previous studies support the hypothesis that cytoplasmic TDP43 is directly toxic. Disruption of the NLS in TDP43, thereby restricting TDP43 localization to the cytoplasm, enhances toxicity in primary neurons, flies, and transgenic mice[28,202,203]. In cell lines, the disruption of nuclear import leads to TDP43 cytoplasmic accumulation. Evidence in post-mortem FTD patient brains demonstrate reduced expression of nuclear import proteins[204]. Nucleocytoplasmic transport dysfunction has increasingly become a pathway of interest in ALS and FTD pathogenesis[94,95,120,121]. Mislocalization from the nucleus to the cytoplasm is not only observed with TDP43 but also with similar RNA binding proteins, such as FUS[37] and hnRNPA1[40], suggesting conserved mechanisms of disease.

The role of TDP43 aggregates in neurodegeneration is not fully understood. Two hypotheses have been proposed in which 1) cytoplasmic aggregates promote cell death by a gain of function mechanism, and 2) TDP43 aggregation sequesters functional TDP43 and other essential proteins involved in RNA metabolism by a loss of function mechanism. Evidence in yeast and cell lines supports the notion that TDP43 aggregates have a toxic role, and that aggregation is enhanced upon disease-associated mutations in the GRD[29,205-207]. Due to the low-complexity of the prion-like sequence within TDP43's C-terminal domain, this region is highly aggregation prone both *in vitro* and in cell culture and has the ability to induce cell death[29,205]. Prior studies demonstrated that the C-terminal domain of TDP43 is

capable of forming amyloid-like fibrils *in vitro*, subsequently recruiting other proteins and interfering with their normal function[29]. Evidence against TDP43 aggregation driving toxicity, however, suggests that TDP43 mislocalization and levels are more closely related to neurodegeneration[28,30,145].

1.5.5 The RNA binding properties of TDP43 and downstream toxicity

Previous studies have investigated the RNA binding properties of TDP43 in disease models and suggest that RNA binding by TDP43 mediates toxicity. In *Drosophila*, deletion of TDP43's RRM1 or substitution of Trp113 to Ala almost entirely eliminates RNA binding and prevents downstream toxicity from TDP43 accumulation[147,202]. Two Phe residues (F147 and F149) within RRM1 are required for nucleic acid recognition. Mutations of these residues to Leu completely abolish binding activity[146,158], and similar mutations rescue TDP43-dependent toxicity in yeast and *Drosophila*[191,192]. Moreover, toxicity was reduced in transgenic flies expressing a disease-associated TDP43 mutation and deletion of RRM1, suggesting that RNA binding is critical for toxicity[202]. Although these studies suggest that the RNA binding properties of TDP43 have important consequences on cellular toxicity, it is unclear the extent to which RNA binding impacts other features of TDP43, such as localization, stability, and RNA processing, and how these phenotypes lead to neurodegeneration.

1.6 Summary and goals of the dissertation

ALS and FTD are devastating neurodegenerative conditions that share key clinical, pathologic, and genetic characteristics. Since the identification of *C9orf72* as the most common genetic cause of ALS and FTD, numerous investigations sought to identify the

molecular mechanisms by which the repeat expansion mutation leads to neurodegeneration. Whether the repeat expansion leads to (1) loss of function, (2) RNA gain of function via sequestration of essential RNA binding proteins within RNA foci, or (3) protein gain of function via RAN translation, thereby contributing to neuronal loss still remains unanswered. TDP43 pathology is a characteristic feature in ALS/FTD, including *C9orf72*-linked disease, and changes in TDP43 localization and levels are strongly related to neuron loss in ALS and FTD. Despite growing interest in ALS and FTD and the rapid pace of genetic investigations, treatment options for these disorders remain extremely limited. This is a result, in part, from our incomplete understanding of the molecular mechanisms initiating and driving disease progression.

Previous investigations have demonstrated individually that *C9orf72* and TDP43 are strongly linked to both ALS and FTD; however, whether these common features of disease are connected remains unknown. The significance of the *C9orf72* RAN products remain unclear. Previous investigations suggest that TDP43's ability to bind RNA is essential for neurodegeneration upon TDP43 accumulation, but the mechanism by which TDP43 deposition leads to neuron loss in ALS and FTD remains unclear. The primary goal of my dissertation is to determine the molecular pathways leading to neurodegeneration in ALS and FTD. My central hypothesis is that *C9orf72* RAN peptides disrupt TDP43 metabolism, leading to neurodegeneration via TDP43-dependent RNA misprocessing. My dissertation research tests this hypothesis by 1) systematically evaluating RAN peptides for their effects on TDP43 metabolism and 2) by uncovering the molecular mechanisms responsible for neurodegeneration upon TDP43 accumulation. These aims are described in detail in Chapters 2-4. While it remains unclear how *C9orf72* mutations and TDP43 deposition are connected,

Chapter 4 sheds preliminary insight of a potential direct link between *C9orf72* and TDP43. Finally, Chapter 5 concludes the dissertation and addresses future directions.

References

1. Lomen-Hoerth, C. *et al.* Are amyotrophic lateral sclerosis patients cognitively normal? *Neurology* **60,** 1094–1097 (2003).
2. Geser, F. *et al.* Clinical and Pathological Continuum of Multisystem TDP-43 Proteinopathies. *Arch Neurol* **66,** (2009).
3. Geser, F., Lee, V. M. Y. & Trojanowski, J. Q. Amyotrophic lateral sclerosis and frontotemporal lobar degeneration: A spectrum of TDP-43 proteinopathies. *Neuropathology* **30,** 103–112 (2010).
4. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
5. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72,** 245–256 (2011).
6. Neumann, M. *et al.* Ubiquitinated TDP-43 in Frontotemporal Lobar Degenerationand Amyotrophic Lateral Sclerosis. *Science* **314,** 130–133 (2006).
7. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem and Biophy Res Commun* **351,** 602–611 (2006).
8. Marin, B. *et al.* Variation in worldwide incidence of amyotrophic lateral sclerosis: a meta-analysis. *Int. J. Epidemiol.* **78,** dyw061 (2016).
9. Kiernan, M. C. *et al.* Amyotrophic lateral sclerosis. *The Lancet* **377,** 942–955 (2011).
10. Goetz, C. G. Amyotrophic lateral sclerosis: early contributions of Jean-Martin Charcot. *Muscle & nerve* **23,** 336–343 (2000).
11. Renton, A. E., Chiò, A. & Traynor, B. J. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* **17,** 17–23 (2014).
12. Neary, D. *et al.* Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology* **51,** 1546–1554 (1998).
13. Onyike, C. U. & Diehl-Schmid, J. The epidemiology of frontotemporal dementia. *Int Rev Psychiatry* **25,** 130–137 (2013).
14. Rohrer, J. D. *et al.* The heritability and genetics of frontotemporal lobar degeneration. *Neurology* 1–6 (2009).
15. Bedlack, R. S. Amyotrophic lateral sclerosis: current practice and future treatments. *Current Opinion in Neurology* **23,** 524–529 (2010).
16. Miller, R. G., Mitchell, J. D., Lyon, M. & DH, M. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders* **4,** 191–206 (2009).
17. Petrov, D., Mansfield, C., Moussy, A. & Hermine, O. ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment? *Front. Aging Neurosci.* **9,** 610 (2017).
18. Dash, R. P., Babu, R. J. & Srinivas, N. R. Two Decades-Long Journey from Riluzole to Edaravone: Revisiting the Clinical Pharmacokinetics of the Only Two Amyotrophic Lateral Sclerosis Therapeutics. *Clinical Pharmacokinetics* 1–14 (2018). doi:10.1007/s40262-018-0655-4
19. Abe, K. *et al.* Confirmatory double-blind, parallel-group, placebo-controlled

study of efficacy and safety of edaravone (MCI-186) in amyotrophic lateral sclerosis patients. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* **15,** 610–617 (2014).
20. Baker, M. *et al.* Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature* **442,** 916–919 (2006).
21. Cruts, M. *et al.* Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **442,** 920–924 (2006).
22. Hutton, M. *et al.* Association of missense and 5'-splice-site mutations in *tau* with theinherited dementia FTDP-17. *Nature* **393,** 702–705 (1998).
23. Poorkaj, P. *et al.* Tau Is a Candidate Gene for Chromosome 17 Frontotemporal Dementia. *Ann Neurol.* **43,** 815–825 (1998).
24. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362,** 1–4 (1993).
25. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis. *Neuron* **79,** 416–438 (2013).
26. Buratti, E. *Functional Significance of TDP-43 Mutations in Disease*. *Advances in Genetics* **91,** 1–53 (Elsevier Ltd, 2015).
27. Neumann, M. Molecular Neuropathology of TDP-43 Proteinopathies. *IJMS* **10,** 232–246 (2009).
28. Barmada, S. J. *et al.* Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J. Neurosci* **30,** 639–649 (2010).
29. Guo, W. *et al.* An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struc. Mol. Biol.* **18,** 822–830 (2011).
30. Barmada, S. J. *et al.* Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. *Nat. Chem. Bio.* **10,** 677–685 (2014).
31. Ling, S.-C. *et al.* ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 13318–13323 (2010).
32. Liu-Yesucevitz, L. *et al.* Tar DNA Binding Protein-43 (TDP-43) Associates with Stress Granules: Analysis of Cultured Cells and Pathological Brain Tissue. *PLoS ONE* **5,** e13250 (2010).
33. McDonald, K. K. *et al.* TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. *Hum. Mol. Genet.* **20,** 1400–1410 (2011).
34. Kapeli, K., Martinez, F. J. & Yeo, G. W. Genetic mutations in RNA-binding proteins and their roles in ALS. *Hum. Mol. Genet.* **136,** 1193–1214 (2017).
35. Eykens, C. & Robberecht, W. The genetic basis of amyotrophic lateral sclerosis: recent breakthroughs. *AGG* 327–345 (2015). doi:10.2147/AGG.S57397
36. Kwiatkowski, T. J. *et al.* Mutations in the *FUS/TLS* Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. *Science* **323,** 1201–1205 (2009).
37. Vance, C. *et al.* Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. *Science* **323,** 1208–1210 (2009).

38. Couthouis, J. *et al.* Evaluating the role of the *FUS/TLS*-related gene *EWSR1* in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **21,** 2899–2911 (2012).
39. Couthouis, J. *et al.* A yeast functional screen predicts new candidate ALS disease genes. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 20881–20890 (2011).
40. Kim, H. J. *et al.* Mutations in prion-like domains inhnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495,** 467–473 (2013).
41. Fecto, F. *et al.* SQSTM1 Mutations in Familial Sporadic Amyotrophic Lateral Sclerosis. *Arch Neurol* **68,** 1440–1446 (2011).
42. Johnson, J. O. *et al.* Exome Sequencing Reveals VCP Mutations as a Cause of Familial ALS. *Neuron* **68,** 857–864 (2010).
43. Maruyama, H. *et al.* Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* **465,** 223–226 (2011).
44. Van Mossevelde, S. *et al.* Clinical features of *TBK1* carriers compared with *C9orf72, GRN* and non-mutation carriers in a Belgian cohort. *Brain* **139,** 452–467 (2016).
45. Deng, H.-X. *et al.* Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature* **477,** 211–215 (2011).
46. Almeida, B., Fernandes, S., Abreu, I. A. & Macedo-Ribeiro, S. Trinucleotide repeats: a structural perspective. *Front. Neurol.* 1–24 (2013). doi:10.3389/fneur.2013.00076/abstract
47. Orr, H. T. & Zoghbi, H. Y. Trinucleotide Repeat Disorders. *Annu. Rev. Neurosci.* **30,** 575–621 (2007).
48. Fratta, P. *et al.* Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathologica* **126,** 401–409 (2013).
49. Majounie, E. *et al.* Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *The Lancet Neurology* **11,** 323–330 (2012).
50. Gijselinck, I. *et al.* A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum:a gene identification study. *Lancet Neurol* **11,** 54–65 (2011).
51. van Blitterswijk, M. *et al.* Association between repeat sizes and clinical and pathological characteristics in carriers of C9orf72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *The Lancet Neurology* **12,** 978–988 (2013).
52. Gijselinck, I. *et al.* The *C9orf72* repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Molecular Psychiatry* **21,** 1112–1124 (2015).
53. Nordin, A. *et al.* Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. *Hum. Mol. Genet.* **24,** 3133–3142 (2015).
54. Waite, A. J. *et al.* Neurobiology of Aging. *NBA* **35,** 1779.e5–1779.e13 (2014).
55. Dols-Icardo, O. *et al.* Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. *Human Molecular Genetics* **23,** 749–754 (2014).
56. Sareen, D. *et al.* Targeting RNA Foci in iPSC-Derived Motor Neurons from ALS Patients with a C9ORF72 Repeat Expansion. *Science Translational Medicine* **5,**

208ra149–208ra149 (2013).
57. Xiao, S. *et al.* Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis. *Ann Neurol.* **78,** 568–583 (2015).
58. Xiao, S. *et al.* C9orf72 isoforms in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration. *Brain Research* 1–7 (2016). doi:10.1016/j.brainres.2016.04.062
59. Yang, M. *et al.* A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. *Sci Adv* **2,** e1601167–e1601167 (2016).
60. Sullivan, P. M. *et al.* The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. *Acta Neuropathol Commun* 1–16 (2016). doi:10.1186/s40478-016-0324-5
61. Sellier, C. *et al.* Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. *EMBO J* **35,** 1276–1297 (2016).
62. Amick, J., Roczniak-Ferguson, A. & Ferguson, S. M. C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. *Mol. Biol. Cell* **27,** 3040–3051 (2016).
63. Belzil, V. V. *et al.* Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathologica* **126,** 895–905 (2013).
64. Ciura, S. *et al.* Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. *Ann Neurol.* **74,** 180–187 (2013).
65. Donnelly, C. J. *et al.* RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. *Neuron* **80,** 415–428 (2013).
66. Fratta, P. *et al.* C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2,** 707 (2012).
67. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507,** 195–200 (2015).
68. Xi, Z. *et al.* Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *Am J Hum Genet* **92,** 981–989 (2013).
69. Castaldo, I. *et al.* DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. *J Med Genet* **45,** 808–812 (2008).
70. Oberlé, I. *et al.* Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252,** 1097–1102 (1991).
71. Sutcliffe, J. S. *et al.* DNA methylation represses FMR-1 transcription in fragile X syndrome. *Human Molecular Genetics* **1,** 397–400 (1992).
72. Russ, J. *et al.* Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. *Acta Neuropathologica* **129,** 39–52 (2014).
73. Liu, E. Y. *et al.* Aberrant RNA homeostasis in amyotrophic lateral sclerosis: potential for new therapeutic targets? *Neurodegenerative Disease Management* **128,** 417–437 (2014).
74. Farg, M. A. *et al.* C9ORF72, implicated in amyotrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum. Mol. Genet.* **23,** 3579–3595 (2014).

75. Webster, C. P. *et al.* The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *EMBO J* **35,** 1656–1676 (2016).
76. Therrien, M., Rouleau, G. A., Dion, P. A. & Parker, J. A. Deletion of C9ORF72 Results in Motor Neuron Degeneration and Stress Sensitivity in C. elegans. *PLoS ONE* **8,** e83450 (2013).
77. Lagier-Tourenne, C. *et al.* Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **110,** E4530–9 (2013).
78. Koppers, M. *et al.* C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Ann Neurol.* **78,** 426–438 (2015).
79. Atanasio, A. *et al.* C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. *Sci Rep* 1–14 (2016). doi:10.1038/srep23204
80. O'Rourke, J. G. *et al.* C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. *Neuron* **88,** 892–901 (2015).
81. Ji, Y. J., Ugolino, J., Brady, N. R., Hamacher-Brady, A. & Wang, J. Systemic deregulation of autophagy upon loss of ALS- and FTD-linked C9orf72. *Autophagy* **13,** 1254–1255 (2017).
82. Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc. Natl. Acad. Sci. U.S.A.* **110,** E4968–77 (2013).
83. Gendron, T. F. *et al.* Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathologica* **126,** 829–844 (2013).
84. Mori, K. *et al.* Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathologica* **126,** 881–893 (2013).
85. Echeverria, G. V. & Cooper, T. A. RNA-binding proteins in microsatellite expansion disorders: Mediators of RNA toxicity. *Brain Research* **1462,** 100–111 (2012).
86. Ueyama, M. & Nagai, Y. Repeat Expansion Disease Models. *Adv Exp Med Biol* **1076,** 63–78 (2018).
87. Mizielinska, S. *et al.* C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathologica* **126,** 845–857 (2013).
88. Cooper-Knock, J. *et al.* Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain* **137,** 2040–2051 (2014).
89. Cooper-Knock, J. *et al.* Antisense RNA foci in the motor neurons of. *Acta Neuropathologica* 1–13 (2015). doi:10.1007/s00401-015-1429-9
90. Reddy, K., Zamiri, B., Stanley, S. Y. R., Macgregor, R. B., Jr. & Pearson, C. E. The Disease-associated r(GGGGCC) nRepeat from the C9orf72Gene Forms Tract Length-dependent Uni- and Multimolecular RNA G-quadruplex Structures. *J. Biol. Chem.* **288,** 9860–9866 (2013).
91. Reddy, K. *et al.* Processing of double-R-loops in (CAG)·(CTG) and C9orf72 (GGGGCC)·(GGCCCC) repeats causes instability. *Nucleic Acids Research* **42,**

10473–10487 (2014).
92. Xu, Z. *et al.* Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proceedings of the National Academy of Sciences* **110,** 7778–7783 (2013).
93. Lee, Y.-B. *et al.* Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic. *Cell Rep* **5,** 1178–1186 (2013).
94. Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525,** 56–61 (2015).
95. Freibaum, B. D. *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525,** 129–133 (2015).
96. Swinnen, B. *et al.* A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta Neuropathologica* 1–17 (2018). doi:10.1007/s00401-017-1796-5
97. Mizielinska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345,** 1192–1194 (2014).
98. Tran, H. *et al.* Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* **87,** 1207–1214 (2015).
99. DeJesus-Hernandez, M. *et al.* In-depth clinico-pathological examination of RNA foci in a large cohort of. *Acta Neuropathologica* **134,** 255–269 (2017).
100. Liu, J. *et al.* c9orf72 Disease-Related Foci Are Each Composed of One Mutant Expanded Repeat RNA. *Cell Chemical Biology* 1–9 (2017). doi:10.1016/j.chembiol.2016.12.018
101. Ash, P. E. A. *et al.* Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron* 1–8 (2013). doi:10.1016/j.neuron.2013.02.004
102. Mori, K. *et al.* The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science* **339,** 1335–1338 (2013).
103. Zu, T. *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 260–265 (2011).
104. Todd, P. K. *et al.* CGG Repeat-Associated Translation Mediates Neurodegenerationin Fragile X Tremor Ataxia Syndrome. *Neuron* **78,** 440–455 (2013).
105. Bañez-Coronel, M. *et al.* RAN Translation in Huntington Disease. *Neuron* **88,** 667–677 (2015).
106. Al-Sarraj, S. *et al.* p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. *Acta Neuropathologica* **122,** 691–702 (2011).
107. Boxer, A. L. *et al.* Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. *J Neurol Neurosurg Psychiatry* **82,** 196–203 (2011).
108. Mann, D. M. *et al.* Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathol Commun* **1,**

68 (2013).
109. Mackenzie, I. R. A. *et al.* Quantitative analysis and clinico-pathological correlationsof different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathologica* **130,** 845–861 (2015).
110. Mackenzie, I. R. *et al.* Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathologica* **126,** 859–879 (2013).
111. Baborie, A. *et al.* Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in C9ORF72gene. *Neuropathol Appl Neurobiol* **41,** 601–612 (2015).
112. Davidson, Y. *et al.* Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. *Neuropathol Appl Neurobiol* **42,** 242–254 (2015).
113. Schludi, M. H. *et al.* Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathologica* 1–18 (2015). doi:10.1007/s00401-015-1450-z
114. Wen, X. *et al.* Antisense Proline-Arginine RAN Dipeptides Linkedto C9ORF72-ALS/FTD Form Toxic Nuclear Aggregates that Initiate In Vitro and In Vivo Neuronal Death. *Neuron* **84,** 1213–1225 (2014).
115. Yang, D. *et al.* FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathologica* **130,** 525–535 (2015).
116. Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345,** 1139–1145 (2014).
117. Tao, Z. *et al.* Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum. Mol. Genet.* **24,** 2426–2441 (2015).
118. Yamakawa, M. *et al.* Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum. Mol. Genet.* **24,** 1630–1645 (2014).
119. Kanekura, K. *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Human Molecular Genetics* **25,** 1803–1813 (2016).
120. Jovičić, A. *et al.* Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18,** 1226–1229 (2015).
121. Boeynaems, S. *et al.* Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci Rep* **6,** 20877 (2016).
122. May, S. *et al.* C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathologica* **128,** 485–503 (2014).
123. Zhang, Y.-J. *et al.* Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathologica* **128,** 505–524 (2014).
124. Nonaka, T. *et al.* C9ORF72 dipeptide repeat poly-GA inclusions promote

intracellular aggregation of phosphorylated TDP-43. *Hum. Mol. Genet.* **72,** 245 (2018).
125. Zhang, Y.-J. *et al.* C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci* 1–14 (2016). doi:10.1038/nn.4272
126. Su, Z. *et al.* Discovery of a Biomarker and LeadSmall Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. *Neuron* 1–41 (2014). doi:10.1016/j.neuron.2014.07.041
127. Hasegawa, M. *et al.* Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Ann Neurol.* **64,** 60–70 (2008).
128. Ou, S. H., Wu, F., Harrich, D., García-Martínez, L. F. & Gaynor, R. B. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J. Virol.* **69,** 3584–3596 (1995).
129. Kraemer, B. C. *et al.* Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta Neuropathologica* **119,** 409–419 (2010).
130. Sephton, C. F. *et al.* TDP-43 is a developmentally regulated protein essential for early embryonic development. *J. Biol. Chem.* **285,** 6826–6834 (2010).
131. Wu, L.-S. *et al.* TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. *genesis* **351,** n/a–n/a
132. Kuo, P.-H., Doudeva, L. G., Wang, Y.-T., Shen, C.-K. J. & Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37,** 1799–1808 (2009).
133. Shiina, Y., Arima, K., Tabunoki, H. & Satoh, J.-I. TDP-43 Dimerizes in Human Cells in Culture. *Cell Mol Neurobiol* **30,** 641–652 (2009).
134. Zhang, Y.-J. *et al.* The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum. Mol. Genet.* **22,** 3112–3122 (2013).
135. Chang, C.-K. *et al.* The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. *Biochem and Biophy Res Commun* **425,** 219–224 (2012).
136. Wang, Y.-T. *et al.* The Truncated C-terminal RNA Recognition Motif of TDP-43 Protein Plays a Key Role in Forming Proteinaceous Aggregates. *J. Biol. Chem.* **288,** 9049–9057 (2013).
137. Buratti, E. *et al.* TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. *J. Biol. Chem.* **280,** 37572–37584 (2005).
138. D'Ambrogio, A. *et al.* Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. *Nucleic Acids Res.* **37,** 4116–4126 (2009).
139. Deshaies, J.-E. *et al.* TDP-43 regulates the alternative splicing of hnRNP A1 to yield an aggregation-prone variant in amyotrophic lateral sclerosis. *Brain* **141,** 1320–1333 (2018).
140. Freibaum, B. D., Chitta, R. K., High, A. A. & Taylor, J. P. Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery. *J. Proteome Res.* **9,** 1104–1120 (2010).

141. Ayala, Y. M. *et al.* Structural determinants of the cellular localization and shuttling of TDP-43. *J. Cell Sci.* **121,** 3778–3785 (2008).
142. Winton, M. J. *et al.* Disturbance of Nuclear and Cytoplasmic TAR DNA-binding Protein (TDP-43) Induces Disease-like Redistribution, Sequestration, and Aggregate Formation. *J. Biol. Chem.* **283,** 13302–13309 (2008).
143. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* **14,** 452–458 (2011).
144. Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci* **14,** 459–468 (2011).
145. Arnold, E. S. *et al.* ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *Proc. Natl. Acad. Sci. U.S.A.* **110,** E736–45 (2013).
146. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J* **30,** 277–288 (2011).
147. Ayala, Y. M. *et al.* Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol.* **348,** 575–588 (2005).
148. Lagier-Tourenne, C. *et al.* Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat Neurosci* **15,** 1488–1497 (2012).
149. Wang, I.-F., Wu, L.-S., Chang, H.-Y. & Shen, C.-K. J. TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor. *J. Neurochem.* **105,** 797–806 (2008).
150. Liu-Yesucevitz, L. *et al.* ALS-Linked Mutations Enlarge TDP-43-Enriched Neuronal RNA Granules in the Dendritic Arbor. *Journal of Neuroscience* **34,** 4167–4174 (2014).
151. Colombrita, C. *et al.* TDP-43 is recruited to stress granules in conditions of oxidative insult. *J. Neurochem.* **111,** 1051–1061 (2009).
152. Dewey, C. M. *et al.* TDP-43 Is Directed to Stress Granules by Sorbitol, a Novel Physiological Osmotic and Oxidative Stressor. *Molecular and Cellular Biology* **31,** 1098–1108 (2011).
153. Volkening, K., Leystra-Lantz, C., Yang, W., Jaffee, H. & Strong, M. J. 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,** 168–182 (2009).
154. Sephton, C. F. *et al.* Identification of neuronal RNA targets of TDP-43-containing ribonucleoprotein complexes. *J. Biol. Chem.* **286,** 1204–1215 (2011).
155. Xiao, S. *et al.* RNA targets of TDP-43 identified by UV-CLIP are deregulated in ALS. *Mol. Cell. Neurosci.* **47,** 167–180 (2011).
156. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struc. Mol. Biol.* **20,** 1443–1449 (2013).
157. Kuo, P.-H., Chiang, C.-H., Wang, Y.-T., Doudeva, L. G. & Yuan, H. S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res.* **42,** 4712–

4722 (2014).
158. Buratti, E. & Baralle, F. E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276,** 36337–36343 (2001).
159. Furukawa, Y. *et al.* A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Sci Rep* **6,** 20576 (2016).
160. Buratti, E. *et al.* Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J* **20,** 1774–1784 (2001).
161. Corcia, P. *et al.* Phenotype and genotype analysis in amyotrophic lateral sclerosis with TARDBP gene mutations. *Neurology* **78,** 1519–1526 (2012).
162. Lattante, S., Rouleau, G. A. & Kabashi, E. TARDBP and FUS Mutations Associated with Amyotrophic Lateral Sclerosis: Summary and Update. *Human Mutation* **34,** 812–826 (2013).
163. Nishimura, A. L. *et al.* Allele-Specific Knockdown of ALS-Associated Mutant TDP-43 in Neural Stem Cells Derived from Induced Pluripotent Stem Cells. *PLoS ONE* **9,** e91269 (2014).
164. Watanabe, S., Kaneko, K. & Yamanaka, K. Accelerated disease onset with stabilized familial amyotrophic lateral sclerosis (ALS)-linked mutant TDP-43 proteins. *J. Biol. Chem.* **288,** 3641–3654 (2013).
165. Barmada, S. J. & Finkbeiner, S. Pathogenic TARDBP mutations in amyotrophic lateral sclerosis and frontotemporal dementia: disease-associated pathways. *Rev Neurosci* **21,** 251–272 (2010).
166. Moreno, F. A novel mutation P112H in the TARDBP gene associated with frontotemporal lobar degeneration without motor neuron disease and abundant neuritic amyloid plaques. 1–13 (2015). doi:10.1186/s40478-015-0190-6
167. Kabashi, E. *et al.* TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet* **40,** 572–574 (2008).
168. Maurel, C. *et al.* Mutation in the RRM2 domain of TDP-43 in Amyotrophic Lateral Sclerosis with rapid progression associated with ubiquitin positive aggregates in cultured motor neurons. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* **19,** 149–151 (2018).
169. Bhandare, V. V. & Ramaswamy, A. The proteinopathy of D169G and K263E mutants at the RNA Recognition Motif (RRM) domain of tar DNA binding protein (tdp43) causing neurological disorders: A computational study. *Journal of Biomolecular Structure and Dynamics* 1–0 (2017). doi:10.1080/07391102.2017.1310670
170. Austin, J. A. *et al.* Disease causing mutants of TDP-43 nucleic acid binding domains are resistant to aggregation and have increased stability and half-life. *Proc. Natl. Acad. Sci. U.S.A.* **111,** 4309–4314 (2014).
171. Ash, P. E. A. *et al.* Neurotoxic effects of TDP-43 overexpression in C. elegans. *Hum. Mol. Genet.* **19,** 3206–3218 (2010).
172. Zhang, T., Hwang, H.-Y., Hao, H., Talbot, C. & Wang, J. Caenorhabditis elegans RNA-processing protein TDP-1 regulates protein homeostasis and life span. *J. Biol. Chem.* **287,** 8371–8382 (2012).
173. Vaccaro, A. *et al.* TDP-1/TDP-43 Regulates Stress Signaling and Age-Dependent Proteotoxicity in Caenorhabditis elegans. *PLoS Genet* **8,** e1002806

(2012).
174. Feiguin, F. *et al.* Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior. *FEBS Letters* **583,** 1586–1592 (2009).
175. Laird, A. S. *et al.* Progranulin is Neurotrophic In Vivo and Protects against a Mutant TDP-43 Induced Axonopathy. *PLoS ONE* **5,** e13368 (2010).
176. Kabashi, E. *et al.* Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Hum. Mol. Genet.* **19,** 671–683 (2009).
177. Schmid, B. *et al.* Loss of ALS-associated TDP-43 in zebrafish causes muscle degeneration, vascular dysfunction, and reduced motor neuron axon outgrowth. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 4986–4991 (2013).
178. Chiang, P.-M. *et al.* Deletion of TDP-43 down-regulates Tbc1d1, a gene linked to obesity, and alters body fat metabolism. *Proceedings of the National Academy of Sciences* **107,** 16320–16324 (2010).
179. Ling, J. P., Pletnikova, O., Troncoso, J. C. & Wong, P. C. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* **349,** 650–655 (2015).
180. Humphrey, J., Emmett, W., Fratta, P., Isaacs, A. M. & Plagnol, V. Quantitative analysis of cryptic splicing associated with TDP-43 depletion. *BMC Med Genomics* **10,** 38 (2017).
181. Tan, Q. *et al.* Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. *Hum. Mol. Genet.* **48,** ddw337 (2016).
182. Bose, J. K., Wang, I.-F., Hung, L., Tarn, W.-Y. & Shen, C.-K. J. TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. *J. Biol. Chem.* **283,** 28852–28859 (2008).
183. Hazelett, D. J., Chang, J.-C., Lakeland, D. L. & Morton, D. B. 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* **2,** 789–802 (2012).
184. Bhardwaj, A., Myers, M. P., Buratti, E. & Baralle, F. E. Characterizing TDP-43 interaction with its RNA targets. *Nucleic Acids Res.* **41,** 5062–5074 (2013).
185. Kasai, T. *et al.* Increased TDP-43 protein in cerebrospinal fluid of patients with amyotrophic lateral sclerosis. *Acta Neuropathologica* **117,** 55–62 (2008).
186. Noto, Y.-I. *et al.* Elevated CSF TDP-43 levels in amyotrophic lateral sclerosis: Specificity, sensitivity, and a possible prognostic value. *Amyotrophic Lateral Sclerosis* **12,** 140–143 (2011).
187. Verstraete, E. *et al.* TDP-43 plasma levels are higher in amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis* **13,** 446–451 (2012).
188. Swarup, V. *et al.* Deregulation of TDP-43 in amyotrophic lateral sclerosis triggers nuclear factor κB–mediated pathogenic pathways. *J. Exp. Med.* **208,** 2429–2447 (2011).
189. Barmada, S. J. *et al.* Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 7821–7826 (2015).
190. Wils, H. *et al.* TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proc.*

Natl. Acad. Sci. U.S.A. **107,** 3858–3863 (2010).
191. Elden, A. C. *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466,** 1069–1075 (2010).
192. Voigt, A. *et al.* TDP-43-mediated neuron loss in vivo requires RNA-binding activity. *PLoS ONE* **5,** e12247 (2010).
193. Tatom, J. B. *et al.* Mimicking Aspects of Frontotemporal LobarDegeneration and Lou Gehrig's Diseasein Rats via TDP-43 Overexpression. *Molecular Therapy* **17,** 607–613 (2009).
194. Li, Y. *et al.* A Drosophila model for TDP-43 proteinopathy. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 3169–3174 (2010).
195. White, M. A. *et al.* TDP-43 gains function due to perturbed autoregulation in a Tardbp knock-in mouse model of ALS-FTD. *Nat Neurosci* 1–20 (2018). doi:10.1038/s41593-018-0113-5
196. Fratta, P. *et al.* Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis. *EMBO J* **37,** e98684 (2018).
197. Archbold, H. C. *et al.* TDP43 nuclear export and neurodegeneration in models of amyotrophic lateral sclerosis and frontotemporal dementia. *Sci Rep* 1–18 (2018). doi:10.1038/s41598-018-22858-w
198. Wegorzewska, I., Bell, S., Cairns, N. J., Miller, T. M. & Baloh, R. H. TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration. *Proceedings of the National Academy of Sciences* **106,** 18809–18814 (2009).
199. Xu, Y. F. *et al.* Wild-Type Human TDP-43 Expression Causes TDP-43 Phosphorylation, Mitochondrial Aggregation, Motor Deficits, and Early Mortality in Transgenic Mice. *J. Neurosci* **30,** 10851–10859 (2010).
200. Johnson, B. S., McCaffery, J. M., Lindquist, S. & Gitler, A. D. A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 6439–6444 (2008).
201. Liachko, N. F., Guthrie, C. R. & Kraemer, B. C. Phosphorylation Promotes Neurotoxicity in a Caenorhabditis elegans Model of TDP-43 Proteinopathy. *J. Neurosci* **30,** 16208–16219 (2010).
202. Ihara, R. *et al.* RNA binding mediates neurotoxicity in the transgenic Drosophila model of TDP-43 proteinopathy. *Hum. Mol. Genet.* **22,** 4474–4484 (2013).
203. Igaz, L. M. *et al.* Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. *J. Clin. Invest.* **121,** 726–738 (2011).
204. Nishimura, A. L. *et al.* Nuclear import impairment causes cytoplasmic trans-activation response DNA-binding protein accumulation and is associated with frontotemporal lobar degeneration. *Brain* **133,** 1763–1771 (2010).
205. TDP-43 Is Intrinsically Aggregation-prone, and Amyotrophic Lateral Sclerosis-linked Mutations Accelerate Aggregation and Increase Toxicity. *J. Biol. Chem.* **284,** 20329–20339 (2009).
206. Capitini, C. *et al.* TDP-43 Inclusion Bodies Formed in Bacteria Are Structurally Amorphous, Non-Amyloid and Inherently Toxic to Neuroblastoma Cells. *PLoS ONE* **9,** e86720 (2014).

207. Nonaka, T. *et al.* Prion-like Properties of Pathological TDP-43 Aggregates from Diseased Brains. *Cell Rep* **4,** 124–134 (2013).
208. Mackenzie, I. R. A. *et al.* The neuropathology of frontotemporal lobar degeneration caused by mutations in the progranulin gene. *Brain* **129,** 3081–3090 (2006).
209. Sreedharan, J. *et al.* TDP-43 Mutations in Familial and Sporadic Amyotrophic Lateral Sclerosis. *Science* **319,** 1665–1668 (2008).
210. Pesiridis, G. S., Lee, V. M. Y. & Trojanowski, J. Q. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **18,** R156–R162 (2009).
211. Johnson, J. O. *et al.* Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat Neurosci* **17,** 664–666 (2014).
212. Millecamps, S. *et al.* Genetic analysis of matrin 3 gene in French amyotrophic lateral sclerosis patients and frontotemporal lobar degeneration with amyotrophic lateral sclerosis patients. *NBA* **35,** 2882.e13–2882.e15 (2014).
213. Tada, M. *et al.* Matrin 3 Is a Component of Neuronal Cytoplasmic Inclusions of Motor Neurons in Sporadic Amyotrophic Lateral Sclerosis. *The American Journal of Pathology* **188,** 507–514 (2018).
214. Ticozzi, N. *et al.* Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. *Am. J. Med. Genet.* **156,** 285–290 (2011).
215. Neumann, M. *et al.* FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain* **134,** 2595–2609 (2011).
216. Seilhean, D. *et al.* Accumulation of TDP-43 and α-actin in an amyotrophic lateral sclerosis patient with the K17I ANG mutation. *Acta Neuropathologica* **118,** 561–573 (2009).
217. Mackenzie, I. R. *et al.* TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. 1–19 (2017). doi:10.1016/j.neuron.2017.07.025
218. Forman, M. S. *et al.* Novel ubiquitin neuropathology in frontotemporal dementia with valosin-containing protein gene mutations. *J. Neuropathol. Exp. Neurol.* **65,** 571–581 (2006).
219. Saracino, D. *et al.* Neurobiology of Aging. *NBA* 1–4 (2018). doi:10.1016/j.neurobiolaging.2018.06.037
220. Freischmidt, A. *et al.* Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nat Neurosci* **18,** 631–636 (2015).

Table 1.1 Genes associated with ALS, FTD, or both, that impair RNA and/or protein homeostasis.

Symbol	Gene Name	Clinical Manifestation	Pathology	Pathway implicated in disease	Refs.
PGRN	progranulin	FTD	TDP43, ubiquitin	Protein homeostasis	20,21,208
MAPT	microtubule-associated protein Tau	FTD	Tau filaments, ubiquitin	cytoskeleton dynamics	22,23
SOD1	Superoxide dismutase 1	ALS	SOD1/p62	Oxidative stress, Protein homeostasis	24
TDP43	TAR DNA-binding protein	ALS/FTD	TDP43/p62	RNA metabolism	6,7,167,209,210
FUS	Fused in sarcoma	ALS/FTD	FUS, ubiquitin, p62	RNA metabolism	36,37
hnRNPA1	Heterogeneous nuclear ribonucleoprotein A1	ALS, multisystem proteinopathy (MSP)	TDP43	RNA metabolism	40
hnRNPA2B1	Heterogeneous nuclear ribonucleoprotein A2B1	ALS, MSP	TDP43	RNA metabolism	40
MATR3	Matrin 3	ALS/FTD	MATR3	RNA metabolism	211-213
TAF15	TATA-box binding protein associated factor 15	ALS/FTD	TAF15	RNA metabolism	38,214,215
EWSR1	Ewing's sarcoma breakpoint region 1	ALS/FTD	EWSR1	RNA metabolism	38,215
ANG	Angiogenin	ALS	TDP43	RNA processing	216
C9orf72	Chromosome 9 open reading frame 72	ALS/FTD	TDP43/p62, DPR/p62	Reduced *C9orf72* expression [?] Toxic RNA [?] Toxic protein aggregation [?]	4,5

TIA1	TIA1 cytotoxic granule associated RNA binding protein	ALS/FTD	TDP43	RNA metabolism and stress granule dynamics	[217]
VCP	Valosin containing protein	ALS/FTD	TDP43, ubiquitin	Protein homeostasis	[42,218,219]
OPTN	Optineurin	ALS/FTD	TDP43/p62	Protein homeostasis	[43]
TBK1	TANK binding kinase 1	ALS/FTD	TDP43/p62	Protein homeostasis	[44,220]
UBQLN2	Ubiquilin 2	ALS/FTD	TDP43/p62, UBQLN2	Protein homeostasis	[45]
SQSTM1 (p62)	Sequestosome 1	ALS/FTD	TDP43/p62	Protein homeostasis	[41]

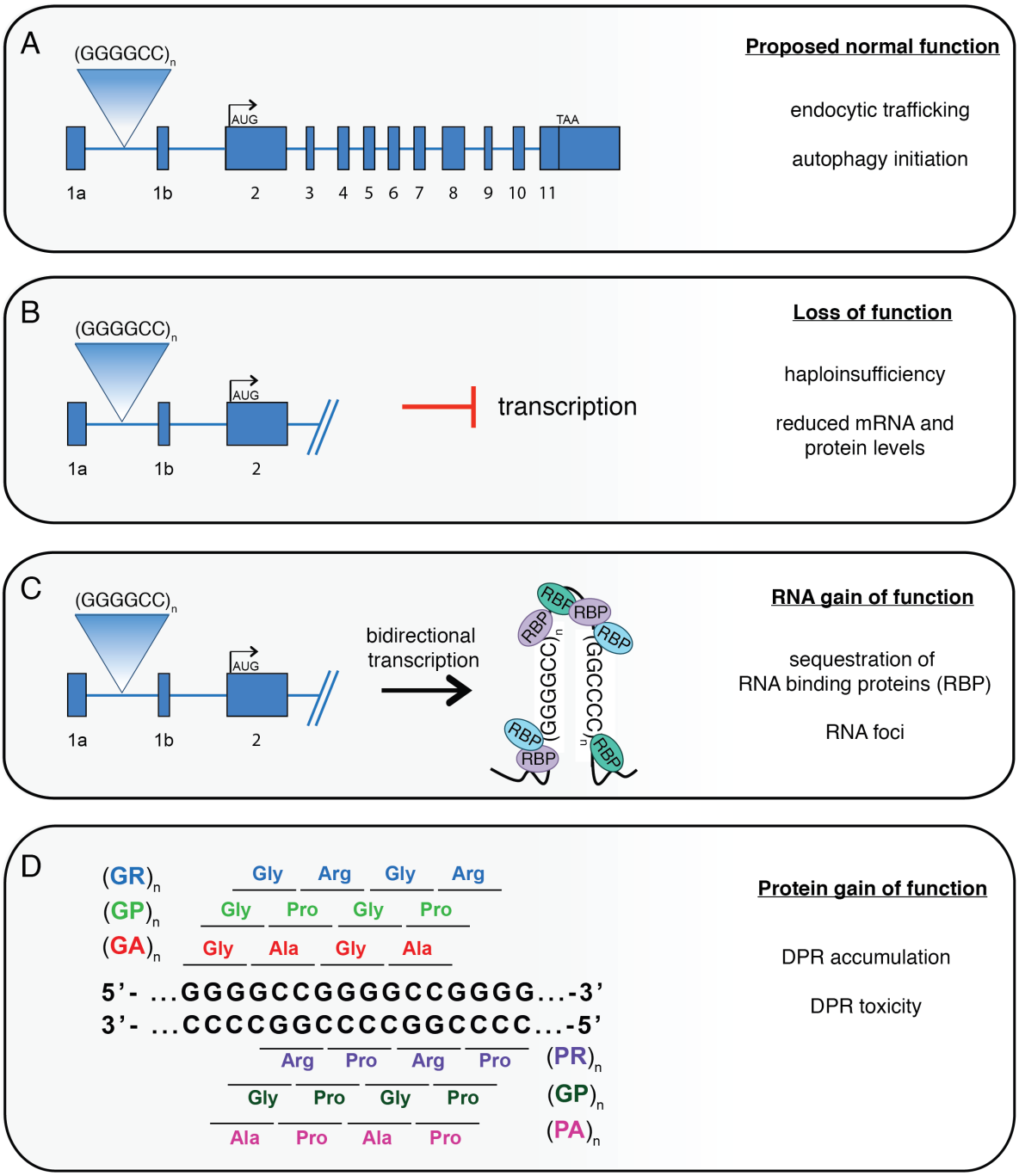

Figure 1.1 Potential pathogenic mechanisms for *C9orf72* mutations. (A) Schematic representation of the gene, *C9orf72*. *C9orf72* consists of 11 exons. Within the first intron is the hexanucleotide repeat (GGGGCC). Expanded repeats above 30 units are associated with ALS and FTD. It is thought that the normal function of the C9orf72 protein plays a role in endocytic trafficking and autophagy initiation. **(B)** The repeat expansion may disrupt transcription of *C9orf72*, resulting in reduced mRNA and protein levels as loss of function mechanism. **(C)** Sense and anti-sense repeat RNA may sequester essential RNA binding proteins within nuclear foci, resulting in an RNA gain of function mechanism.

(D) Mutant *C9orf72* transcripts may undergo repeat associated non-AUG (RAN) translation, generating five different dipeptide repeat proteins (DPRs): glycine-alanine (GA), glycine-proline (GP), glycine-arginine (GR), proline-alanine (PA), and proline-arginine (PR). By a protein gain of function mechanism, these DPRs accumulate within cells and are neurotoxic.

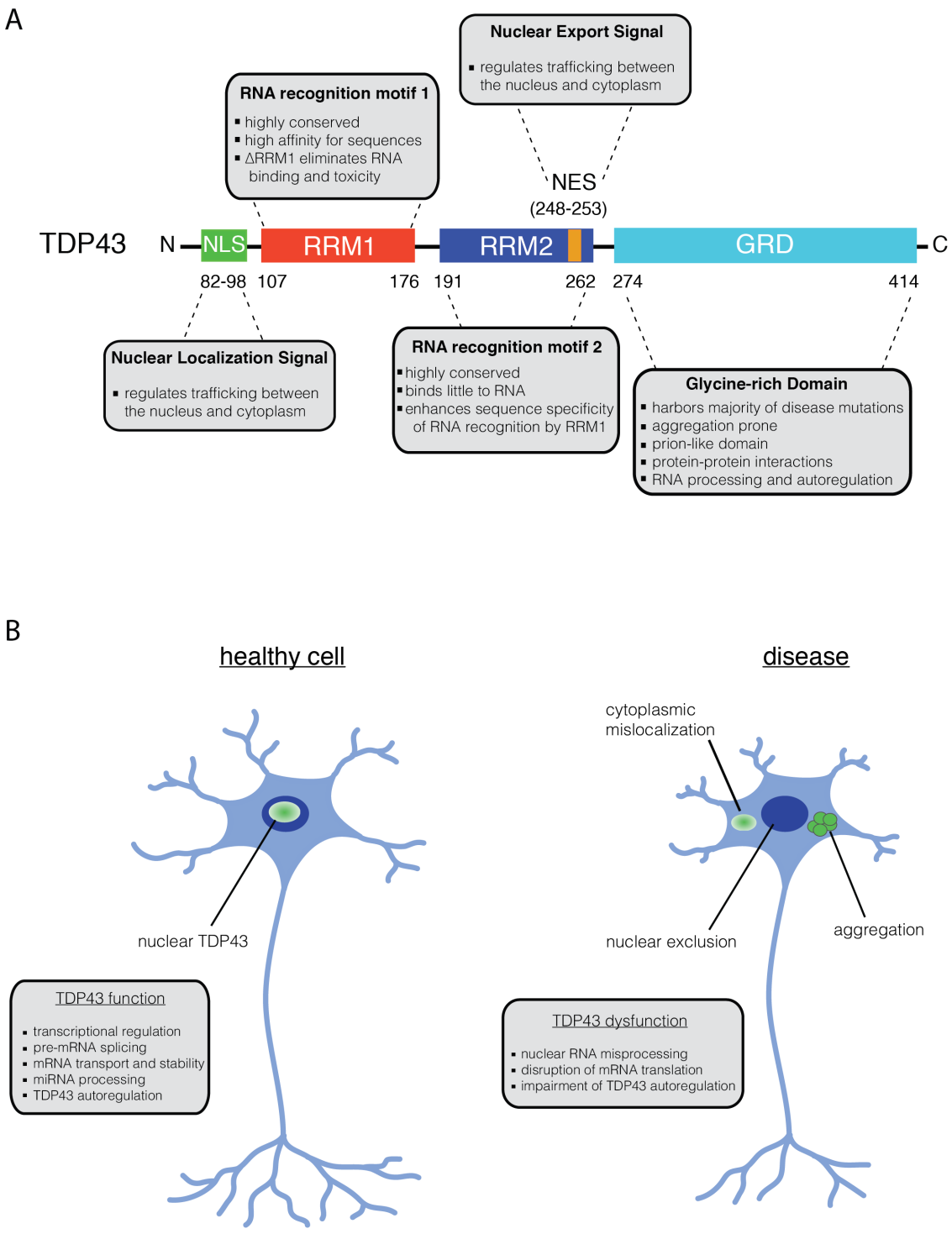

Figure 1.2 TDP43 structure and function. (A) TDP43 is a 414 amino acid protein that contains a nuclear localization signal (NLS), two RNA recognition motifs (RRM1 and RRM2), a nuclear export signal (NES) within RRM2, and a glycine-rich domain (GRD). The NLS and NES enable TDP43 to shuttle between the nucleus and cytoplasm. Although both RRM1 and RRM2 are involved in DNA/RNA binding, these motifs have distinct properties. RRM1 has high affinity for sequences and is primarily involved in

RNA recognition; RRM2 binds little to nucleic acids but enhances sequence specificity of DNA/RNA recognized by RRM1. The GRD is involved in protein-protein interactions, RNA processing, and autoregulation. The GRD is aggregation prone, and majority of the mutations associated with ALS and FTD are found in this region. **(B)** At steady state in a healthy cell (left), TDP43 is predominantly localized in the nucleus; however, it actively shuttles between the nucleus and cytoplasmic serving many functions in RNA metabolism. However, in a disease state, TDP43 is excluded from the nucleus, mislocalizes to the cytoplasm, and forms aggregates; these phenotypes result in the disruption of RNA and protein metabolism.

Chapter 2

Distinct *C9orf72*-aasociated dipeptide repeat structures correlate with neuronal toxicity*

2.1 Abstract

Hexanucleotide repeat expansions in C9orf72 are the most common inherited cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). The expansions elicit toxicity in part through repeat associated non-AUG (RAN) translation of the intronic (GGGGCC)n sequence into dipeptide repeat-containing proteins (DPRs). Little is known, however, about the structural characteristics and aggregation propensities of the dipeptide units comprising DPRs. To address this question, we synthesized dipeptide units corresponding to the three sense-strand RAN translation products, analyzed their structures by circular dichroism, electron microscopy and dye binding assays, and assessed their relative toxicity when applied to primary cortical neurons. Short, glycine-arginine (GR)3 dipeptides formed spherical aggregates and selectively reduced neuronal survival compared to glycine-alanine (GA)3 and glycine-proline (GP)3 dipeptides. Doubling peptide length had little effect on the structure of GR or GP peptides, but (GA)6 peptides formed β-sheet rich aggregates that bound thioflavin T and

* This chapter represents the following manuscript:

Flores BN[§], Dulchavsky ME[§], Krans A, Sawaya MR, Paulson HL, Todd PK, Barmada SJ, Ivanova MII. Distinct *C9orf72*-associated dipeptide repeat structures correlate with neuronal toxicity. 2016. *PLoS ONE* 11(10):e0165084. doi:10.1371/journal.pone.0165084.

Congo red yet lacked the typical fibrillar morphology of amyloids. Aging of (GA)6 dipeptides increased their β-sheet content and enhanced their toxicity when applied to neurons. We also observed that the relative toxicity of each tested dipeptide was proportional to peptide internalization. Our results demonstrate that different C9orf72-related dipeptides exhibit distinct structural properties that correlate with their relative toxicity.

2.2 Introduction

A GGGGCC repeat expansion in the first intron of C9orf72 is the most common known inherited cause of amyotrophic lateral sclerosis and frontotemporal dementia (C9ALS/FTD)[1,2] . This mutation explains up to 40% of all familial cases of ALS and FTD and 7% of all sporadic cases of these conditions. Most people have less than 28 repeats but in patients with expansions, hundreds to thousands of pathologic repeats are observed[3,4].

Three mechanisms have been proposed for neurodegeneration in association with (GGGGCC)n repeats in the C9orf72 locus: (a) inhibition of C9orf72 transcription and expression due to enhanced methylation and/or formation of DNA/RNA hybrids[1,5-7] (b) sequestration of essential RNA-binding proteins and other factors through the formation of insoluble RNA foci by sense (GGGGCC)n and antisense (GGCCCC)n repeat RNAs[1,8-11] and (c) repeat associated non-AUG (RAN) initiated translation of (GGGGCC)n and (GGCCCC)n repeat RNAs and the resulting accumulation of neurotoxic RAN dipeptide repeat proteins or DPRs[9,12-18]. Although more than one of these mechanisms may contribute to neuronal loss and disease features, we focus here on

the biochemical properties of the dipeptide units comprising DPRs as evidence increasingly supports a key role for DPRs in disease pathogenesis[8,16,17,19].

Both sense and antisense DPRs result from the translation of GGGGCC expansion repeats. All five of the potential DPRs (glycine-alanine GA-, glycine-proline GP-, glycine-arginine GR-,alanine-proline AP- and proline-arginine PR-) have been reported in disease brain[8,16], though the sense strand DPRs containing GP or GR appear to be most abundant[9,20]. Relatively little is known about the structure and morphology of DPR aggregates. Most of what is known is based on their morphological appearance in cellular models and patient tissues, except for the GA- containing peptide which was recently reported to have amyloid properties[14].

DPRs generated via AUG-initiated translation are toxic in yeast, cultured cells and Drosophila, and this toxicity is largely independent of (GGGGCC)n RNA repeat sequences[12,13,17,19,21-28].While several studies suggest that GR- and PR- repeat proteins are among the most toxic DPRs, GA-containing peptides are more common in individuals carrying C9orf72 expansions, and GA-containing peptides demonstrated selective toxicity in neuronal cell lines and primary neurons[9,14,17,19-30]. In contrast, no toxicity was observed with constructs selectively expressing GP- or PA- DPRs[17,23]. These features, together with the distinct composition and distribution of DPRs in C9ALS/FTD brains[8,16,20], suggest that each DPR possesses unique characteristics that can be best understood by utilizing different model systems.

Defining the structural characteristics of pathogenic proteins in neurodegenerative diseases has yielded important insights into the mechanisms of inclusion formation and disease pathogenesis[31-33]. Despite the prevalence of DPRs in C9ALS/FTD pathology and

their demonstrated toxicity in vitro, however, the structure of DPRs and how their structure relates to downstream toxicity remain unknown. Investigations of the minimal peptide segments that are the building blocks of amyloid fibrils have deepened our understanding of neurodegenerative disease pathogenesis and the aberrant behavior of disease-associated proteins[34-37]. Structural studies of these minimal units have the potential to uncover fundamental connections between protein structure and neurotoxicity, and accelerate the design of molecules that can prevent their formation.

Here we show that each of the three sense strand-derived dipeptide units underlying DPRs (GP-, GA-, and GR-), exhibits distinct structural properties that correlate with neurotoxicity when applied externally to primary rodent neurons. Using anti-DPR antibodies and fluorescence labeling, we also show that aggregates formed by the two toxic dipeptides, GA and GR, are internalized by neurons. These data suggest that short stretches of repeating GA- and GR dipeptides can adopt structural features with the potential to impact neuronal survival.

2.3 Results

In humans carrying pathogenic C9orf72 mutations, several RAN translation products differing in primary sequence and length may be produced simultaneously[16]. To assess the properties of different DPRs, we studied individual peptides with either three or six repeats of GA, GP or GR motifs. These three motifs correspond to the expected sense translation products of the mutant (GGGGCC)n repeat sequence.

Longer peptides, in particular GA-containing peptides were difficult to synthesize and displayed a high tendency to precipitate from solution. To assess how aging affects

the properties of the aggregates, we analyzed peptides immediately after solubilizing (fresh) or after incubation at 37°C for a week (aged). When dissolving peptides, we opted for a buffer that would be innocuous to neurons (phosphate-buffered saline) and sonicated the peptides to facilitate their solubilization rather than using organic solvents.

2.3.1 Glycine-proline peptides are resistant to aggregation and nontoxic

By TEM, we observed only sparse, round shaped aggregates in fresh and aged samples of peptides with 3 or 6 repeating GP motifs (Fig. 2.1A and 2.1E). The CD spectra of fresh and aged (GP)3 and (GP)6 were indistinguishable, resembling the spectra of other poly-proline peptides (Fig. 2.1B and 2.1F, Table 2.1)[43]. The FT-IR spectra of soluble and aged (GP)6 were also similar (Fig. S2.1). Additionally, the absorbance spectrum of CR, a marker of amyloid aggregates, did not change in the presence of (GP)3 or (GP)6 (Fig. 2.1C and 2.1G). Thus, under the conditions tested GP peptides formed only sparse aggregates and we did not observe any difference between fresh and aged GP- containing samples.

GP-containing peptides showed little toxicity in longitudinal assessments of neuronal survival. For these assays, we applied fresh or aged (GP)3 and (GP)6 peptides to rodent primary cortical neurons and tracked the survival of neurons using a fully-automated microscopy platform[40,41]. Neurons were transfected with a vector encoding mApple and imaged sequentially at 24-hour intervals. Individual cells were identified and tracked by custom-written image analysis algorithms and time of death was determined by changes in neuronal morphology or loss of fluorescence, sensitive measures of cell death in previous studies[40,41,44]. Using Cox proportional hazards analysis, these data were used to generate cumulative hazard plots and hazard ratios (HRs), representing relative

risks of death in comparison to a reference population (cells treated with buffer alone). Neither fresh nor aged GP samples were toxic when applied to primary neurons (HR 0.98–1.14, p>0.05; Fig. 2.1D and 2.1H). Thus, (GP)3 and (GP)6 rarely aggregate and are nontoxic in this model system, in accord with prior results[17,19,24].

2.3.2 Glycine-arginine dipeptides form spherical neurotoxic aggregates

Arginine-containing dipeptides demonstrated toxicity in several cell culture and animal models of C9ALS/FTD[19,21,23-26,29]. Fresh and aged (GR)3 peptides formed spherical aggregates by TEM, and aged (GR)3 samples appeared more homogeneous than the fresh samples (Fig. 2.2A). The CD spectra of fresh and aged (GR)3 were similar and contained a small negative band centered at 230 nm that is often associated with type I β-turns (Fig. 2.2B and 2.2F)[45]. Secondary structure predictions indicated that about 55% of the protein exhibits turn and/or random coil conformation and 41% is β-sheet (Table 2.1). Both fresh and aged peptides equally reduced CR absorbance and there was no apparent difference in their CR spectra (Fig. 2.2C). However, aged but not fresh (GR)3 significantly increased the risk of death in primary rodent cortical neurons (HR 1.36, p = 3.5x10-4) (Fig. 2.2D). The inability of (GP)3 and (GP)6 to elicit toxicity in similar assays (Fig. 2.1D and 2.1H) argues against a nonspecific reduction of neuronal survival caused by the quantities of (GR)3 used in these studies.

Similar to aged (GR)3, fresh and aged (GR)6 samples displayed spherical morphology by TEM (Fig. 2.2E). CD spectra for (GR)6 demonstrated that, like (GR)3, a large proportion of the peptide (56%) is turn or random coil and approximately 40% of fresh and aged (GR)6 samples were predicted to be in a β-sheet conformation (Fig. 2.2F and Table 2.1). As revealed by TEM, a large portion of the (GR)6 sample is present

within aggregates, which often reduces the measured CD signal. Accordingly, we utilized FT-IR, which has previously been used for secondary structure determination of protein aggregates[46]. With FT-IR, fresh and aged samples were nearly identical to one another, suggesting that there are no changes in their secondary structure (Fig. S2.1). (GR)6 also bound CR, with aged samples reducing the absorption maximum of CR to a lesser extent than fresh ones (Fig. 2.2G). Both fresh and aged (GR)6 increased the risk of death for rodent primary cortical neurons in longitudinal assays (HR 1.2, p = 4.8x10-3, and HR 1.5, p = 6.6x10-9, respectively) (Fig. 2.2H). In comparison to fresh (GR)6, aged peptide was significantly more toxic (HR 1.25, p 0.001). Thus, aged GR dipeptides, regardless of their length, formed aggregates with spherical morphology that elicited toxicity when applied to neurons.

2.3.3 (GA)6 forms atypical amyloid-like structures and is toxic to neurons

Fresh and aged (GA)3 were indistinguishable by TEM, both forming rare amorphous aggregates (Fig. 2.3A). CD spectra of fresh and aged (GA)3 were also similar (Fig. 2.3B). Neither fresh nor aged (GA)3 demonstrated detectable binding to CR (Fig. 2.3C), and neither sample reduced neuronal survival in longitudinal studies (HR 0.86–1.09, p>0.05; Fig 2.3D). Increasing the number of GA repeats from 3 to 6 had a striking effect on the structure of the peptide. By TEM, aged (GA)6 samples displayed large aggregates consisting of multilayered flat sheets (Fig. 2.3E), whereas fresh (GA)6 samples contained fibrillar aggregates in addition to the flat sheets. Moreover, secondary structure predictions indicated a 5% increase in β-sheet content for aged (GA)6 compared to fresh (GA)6 (Fig. 2.3F and (Table 2.1)[38]. Similarly, FT-IR spectra contained bands at 1622 cm-1 and 1698 cm-1, reminiscent of anti-parallel β-sheet (Fig. S2.1)[47]. As with CD

spectra, we noticed variations in FT-IR spectra of fresh versus aged (GA)6.

Both fresh and aged (GA)6 introduced a shift in the CR spectrum and an increase in overall CR absorbance, changes that are common for amyloid aggregates[39]. This shift in CR absorbance is highlighted by the residual absorbance of the difference spectra, calculated by subtracting the baseline CR spectra from the CR spectra collected in the presence of (GA)6 (Fig. 2.3G). Intriguingly, we observed a difference between the maximum residual absorbance produced by fresh (GA)6 (530 nm) and aged (GA)6 (515 nm). The shift demonstrated by fresh (GA)6 is similar to that produced by amyloid-β fibrils[39].

While both fresh and aged (GA)6 samples significantly reduced neuronal survival, the magnitude of the effect was much greater for aged (HR 2.88, p = 2x10-16) than for fresh (GA)6 samples (HR 1.29, p = 3.3x10-4; Fig. 3H). Thus, GA dipeptides displayed a strong correlation between β-sheet content and neuronal toxicity.

To further explore whether DPRs share features with amyloids, we assessed ThT binding for each of the peptides studied here. None of the shorter DPRs displayed detectable ThT binding (data not shown). Only the most toxic peptide, (GA)6, exhibited affinity for ThT (Fig. 2.4A), which is a common feature of amyloid aggregates. In addition, aged (GA)6 displayed yellow birefringence in the presence of CR (Fig. 2.4B), consistent with other amyloid-like aggregates[48].

Amyloid aggregates also exhibit a distinct x-ray diffraction pattern elicited by the unique β-sheet conformation typical of amyloid[49]. As shown in Fig. 2.4C, the x-ray fiber diffraction pattern of aged (GA)6 contains a strong, sharp 4.4 Å reflection and a broader, weaker 8.6 Å reflection. However, the strand-to strand spacing implied by the 4.4 Å

reflection of (GA)6 is smaller than the 4.8 Å typical of β-sheet structures formed by amyloids. Furthermore, the 8.6 Å reflection is different than the 3.9 Å - 5.3 Å spacing expected for β-sheets composed of glycine and alanine (Fig. 2.4C). Thus, the diffraction pattern for (GA)6 suggests a highly-ordered structure.

2.3.4 Externally applied arginine- and alanine-containing peptides are internalized by neurons

RAN translation products are usually detected within cytoplasmic inclusions of C9ALS/FTD brain tissue and are not typically observed extracellularly[16]. However, some amyloidogenic proteins, such as amyloid β1–42, bind lipid membranes and extracellular receptors to trigger toxic signal transduction cascades[50]. Recent evidence also suggests that some aggregation prone peptides and proteins may pass from one neuron to the next through endocytic and exosomal pathways[51,52].

We therefore sought to determine whether externally applied DPRs could be internalized by neurons in culture. For detection of GP, we used commercially available antibodies, but for GA and GR peptides we developed specific anti-DPR antibodies. Polyclonal rabbit antibodies were raised against synthetically-generated (GA)6 or (GR)6 peptide antigens. These antibodies were validated for specificity first by Western blot and ICC in cells transfected with ATG-V5-GA-GFP and ATG-V5- GR-GFP fusion proteins (Fig. S2.2). Both antibodies also recognized V5-reactive specific bands at different repeat sizes. Moreover, ICC using anti-DPR and –GFP antibodies showed significant co-localization of both antibodies with the epitope tag. To determine if these antibodies are capable of recognizing DPR aggregates in their native state, we performed IHC in C9ALS/FTD brain tissues and controls. Both GA and GR stained perinuclear aggregates

in patient brains but not in controls (Fig. S2.2), consistent with published results[8,9].

To determine if applied DPRs were internalized, primary neurons in culture were fixed 24h after peptide application, immunostained with antibodies recognizing each of the DPRs (GA, GP and GR) and imaged by automated microscopy (Fig. 2.5A; Fig. S2.3). Both nuclear and cytoplasmic inclusions were detected in several neurons treated with aged (GA)6 and (GR)6. In contrast, diffuse nuclear staining was noted in a small proportion of cells exposed to (GP)6, but the vast majority had no evidence of staining over background. Furthermore, DPR internalization was significantly more frequent for (GA)6 and (GR)6 than for (GP)6 (Fig. 2.5B), in accord with the toxicity of (GA)6 and (GR)6 in neurons. We did not detect intracellular DPRs in neurons treated with buffer alone.

To confirm these results, we also covalently coupled a fluorescent molecule, fluorescein-5-EX, succinimidyl ester (FITC), to aged (GA)6, (GP)6 and (GR)6. After applying the aged, labeled peptides to primary neurons in culture, the cells were fixed, immunostained with an antibody recognizing the pan-neuronal cytoplasmic marker MAP2, and imaged by laser scanning confocal microscopy (Fig. 2.5D–G). Intracellular inclusions were detected in neurons treated with FITC-(GA)6 and FITC-(GR)6, but not FITC-(GP)6 or in cells treated with FITC alone, in agreement with the lack of toxicity of GP dipeptides (Fig. 2.1H). Linear quantifications of the Z-sections shown in Fig. 2.5D–G demonstrated that the FITC-(GA)6 and FITC-(GR)6 peptides were indeed cytoplasmic, and not simply located along the extracellular face of the plasma membrane (Fig 2.5E–H). In addition, intracellular deposits of FITC-(GA)6 were significantly more frequent than deposits of FITC-(GP)6 or FITC-(GR)6 (Fig. 2.5C), consistent with the results

obtained using anti-DPR antibodies (Fig. 2.5B). DPR internalization occurred less frequently with FITC-labeled peptides than with unlabeled DPRs, and the distribution of internalized FITC-labeled peptides was also different, perhaps due to the size or charge of the FITC moiety. Nevertheless, these data suggest that (GA)6 and (GR)6 peptides are efficiently taken up by neurons, and that DPR internalization correlates with toxicity.

2.4 Discussion

Here we studied the structural characteristics of the three sense-strand derived DPRs that accumulate and aggregate in C9ALS/FTD. We demonstrate that different DPR peptides exhibit marked differences in their structural features and in their propensity to form aggregates. We also observed a correlation between these structures, their ability to enter cells, and their toxicity when applied exogenously to neurons. These findings suggest that specific structural conformations formed by GA- and GR- repeat proteins may be elemental to their toxicity and could serve as a target for future therapeutic strategies.

Our findings suggest that the length of the dipeptide repeat can influence aggregate structure, internalization and downstream neurotoxicity in a simple model system. Although RAN peptides produced in vivo from C9orf72 expansions may be significantly longer than those studied here, our data show that even short DPRs consisting of 3–6 repeat motifs formed amyloid-like structures and impaired neuronal survival, suggesting that these elements are sufficient for aggregation and neurotoxicity. Consistent with our results, Chang et al.[14] showed that externally-applied GA-containing peptides with fifteen repeats display features of amyloid, are taken up by cells, and are

toxic to transformed cell lines. Thus, the ability of relatively short peptides to cause toxicity suggests that these peptides could accelerate neurodegeneration if present at high local concentrations.

In individuals with C9orf72mutations, RAN proteins are likely produced in multiple reading frames within the same cell, thus the relevance of the exogenously applied individual dipeptide toxicity observed here to that in the human disease remains to be confirmed. However, RAN proteins are detectable in the cerebrospinal fluid of patients with C9orf72 mutations[53]. Given emerging evidence for prion-like spread of aggregation-prone proteins in other neurodegenerative disorders, these studies suggest that DPRs could enhance disease progression by catalyzing the spread of toxicity from one cell to another[14,51,52,54].

Of the peptides studied, (GA)6 and (GR)6 formed β-sheet structures, were taken up most effectively by neurons, and demonstrated toxicity in longitudinal assays, suggesting a potential relationship between structure, internalization, and cell death (Fig. 5). In contrast, GP-containing peptides formed amorphous aggregates, were not internalized by neurons, and were not toxic despite the relatively high concentrations used here. In prior investigations, intrinsically overexpressed GA DPRs form cytoplasmic aggregates, while overexpressed GR DPRs accumulate within nuclear foci[17,21,23,24,26]. In contrast, we detected externally applied, synthetic GR DPRs within the cytoplasm of neurons, and GA DPRs within the nucleus. The apparent discrepancy may be due to the method of delivery, since other investigations noted a similar cytoplasmic deposition of arginine-rich DPRs when applied externally to cells in culture[28]. While no previous study examined the subcellular localization following internalization of synthetic GA DPRs,

the nuclear localization of (GA)6 resembles the nuclear aggregation of other proteins containing polyalanine stretches[55].

Arginine-containing RAN translation products elicit toxicity in cellular and animal models of disease when exogenously applied or when intrinsically expressed under the control of a high-level promoter preceded by an AUG start codon[17,19,21-26,53]. Our results extend these observations, demonstrating that peptides containing three (GR) repeats form aggregates with distinct spherical morphology and cause neuronal death. Although prolonged incubation enhanced the toxicity of GR- containing peptides, the secondary structure and the morphology of the aggregates remained unchanged. However, we observed that fresh and aged (GR)6 have different affinities for CR, possibly reflecting subtle structural changes within the aggregates. Higher resolution studies will be required to reveal the packing of peptides within the aggregates that could account for the acquired toxicity.

Using several independent measures, we found that aged (GA)6 shares key properties with amyloid while remaining structurally distinct. Diffraction from amyloid fibrils typically yields a cross-β pattern, featuring two strong reflections at resolutions corresponding to strand-to-strand and sheet-to-sheet distances. If aged (GA)6 forms amyloid, we would predict a small sheet-to-sheet spacing consistent with the size of the glycine and alanine side chains (3.9 Å, and 5.3 Å, respectively).However, we did not observe either of these, nor did we observe a reflection at 4.8 Å, corresponding to the spacing between strands. Instead, the diffraction pattern shows strong reflections at 4.4, 3.6, and 8.6 Å (Fig. 2.4C). One possible explanation is that (GA)6 assumes a cross-β diffraction pattern, but the reflection corresponding to strand-to-strand spacing is

extinguished by the presence of a two-fold screw symmetry axis coincident with the fibril axis, as is common in crystals of amyloidogenic peptides with class 1 symmetry[36]. In this case, the 4.4 Å spacing would correspond to a set of Bragg planes, closely related to the strand-to-strand spacing but slightly offset in angle and spaced closer together, while the 8.6 Å spacing might arise from the distance between adjacent pairs of β-sheets. We also note that aged (GA)6 grows in large flat sheets, a morphology different from fibrillar aggregates formed by amyloids.

2.5 Conclusions

In summary, our work defines unique structural elements for individual DPRs and correlates the abundance of specific structures with their cellular internalization and relative toxicity. The most neurotoxic dipeptide in our study, (GA)6, exhibits a distinct structure that is β-sheet rich and shares features with amyloid. This link between structure and neurotoxicity suggests that strategies targeting such conformations may effectively reduce RAN peptide-induced neurodegeneration and slow disease progression.

2.6 Materials and methods

Aggregation assays

All peptides were purchased from GenScript. Prior all experiments, we dissolved (GA)3, (GP)3 and (GR)3 to 20 mM and (GA)6, (GP)6 and (GR)6 to 10mM. For transmission electron microscopy (TEM), Congo red (CR), and toxicity assays, peptides were solubilized in 25mM sodium phosphate pH 7.4, 0.1M NaCl. For circular dichroism(CD), peptides were solubilized in 25mM sodium phosphate pH 7.4. (GA)3,

(GP)3, (GP)6, (GR)3 and (GR)6 were soluble and their solutions remained clear after incubation. (GA)6 was not fully soluble at the concentrations used. Freshly dissolved solutions (termed “fresh”) were analyzed immediately after dissolving and were compared with samples (termed “aged”) that were instead incubated at 37°C with agitation in an EchoThermOrbitalmixer (level 9, Torrey Pines Scientific) for 6–8 days. For toxicity, CR and CD, samples were additionally diluted as described in the following sections.

Both sonicated and unsonicated samples were analyzed by CD, CR binding and TEM. Prior sonication, the peptides were additionally diluted to 4mM ((GA)3, (GP)3 and (GR)3) and 2mM ((GA)6, (GP)6 and (GR)6). The peptide samples were sonicated using continuous mode at level 10 (output power 7–8Watts) for 15 sec by immersing the micro-probe of 60 Sonic Dismembrator (Fisher Scientific). The probe was washed with water for 1 min between samples. Data from sonicated and unsonicated peptides were reproduced in three independent experiments, yielding similar results.

Thioflavin T (ThT) binding assay

ThT was added to each of the peptide solutions (20mM (GA)3, (GP)3 and (GR)3 and 10mM (GA)6, (GP)6 and (GR)6 dissolved in 25mM sodium phosphate pH 7.4, 0.1M NaCl) at a final concentration of 10μM, and 110 μl of this mixture was dispensed into each well of Falcon 96-well plate (black/clear, flat bottom, Corning, 353219). Plates were then incubated at 37°C in FLUOstar Omega (BMG Labtech Inc) and were shaken at 200rpmusingmeander corner well shaking mode. Fluorescence was measured with gain set at 90%, an excitation wavelength of 440 nm and emission wavelength of 490 nm. Four separate replicates were measured for each sample.

Transmission Electron Microscopy (TEM)

Negatively stained specimens for TEM were prepared by applying 5 μl of sample (20mM (GA)3, (GP)3, and (GR)3 and 10mM (GA)6, (GP)6, and (GR)6) to hydrophilic 400 mesh carbon coated Formvar support films mounted on copper grids (Ted Pella, Inc., 01702-F). The samples were allowed to adhere for 4 min, rinsed twice with distilled water, and stained for 60–90 sec with 5 μl of 1% uranyl acetate (Ted Pella, Inc.). All samples were imaged at an accelerating voltage of 60 kV in a CM-100 electron microscope (Philips, Inc.). At least four different grids from four independent experiments were examined for fresh and aged samples.

X-ray powder diffraction of (GA)6

Aged (GA)6 samples were centrifuged at 20,000 x g for 3 min. The pellet was washed twice with water, resuspended in 5 μl water, and placed between two fire-polished silanized glass capillaries (1 mm apart) and dried at room temperature. Diffraction patterns were collected using a Rigaku FRD rotating anode X ray generator and an RAXIS-4++ imaging plate detector. Samples were oscillated 1° over a 5 min exposure.

Circular Dichroism (CD)

CD experiments were performed with Jasco J-1500 spectrometer (JASCO Analytical Instruments). The samples were additionally diluted with 25mM sodium phosphate pH 7.5 to 0.6 mg/ml ((GA)3, (GR)3, (GA)6 and (GR)6) or 0.225mg/ml ((GP)3 and (GP)6) before measuring their spectra from 260 nm to 195 nm. Spectra were collected using scanning speed of 200nm/min, data interval of 0.5nmdata, response time of 1 sec and bandwidth of 1 nm. Four spectra were measured for each sample, and data

with a heat tension less than 500 were averaged. Secondary structure was predicted using programs described in[38].

Congo Red (CR) birefringence

For birefringence analysis, aged (GA)6 was pelleted by centrifugation at 20,000 x g for 1 min. 120μM CR (Sigma, C-6767) was added to the precipitate in 25mM sodium phosphate pH 7.4, 0.1M NaCl for 30 min, sedimented by centrifugation at 20 000 x g for 1 min, rinsed three times with water, resuspended in 10 μl of water, and dried on a glass slide to be examined by a light microscope (Zeiss, SteREO Discovery V8) equipped with light polarizers.

CR spectral shift

We followed protocols described by Klunk et al.[39] to determine which aggregates display a red shift in their spectra after binding to CR. A 7.5 μl sample of each peptide or buffer alone (25mM sodium phosphate pH 7.4, 0.1M NaCl) was added to 7.5 μl of 300 μM CR, 25 mM phosphate, pH 7.4, 0.1M NaCl. All specimens were incubated at 37°C for 30 min. Spectra were recorded immediately after incubation using a NanoDrop ND-100 Spectrophotometer (Thermo Scientific, Inc.). To account for contributions from light scatter, the spectra of untreated peptides were subtracted from the corresponding spectra collected from peptides treated with CR.

Fourier Transform Infrared Spectroscopy (FT-IR)

For FT-IR, (GA)6, (GP)6 and (GR)6 were dissolved to 10 mM in 25 mM sodium phosphate pH 7.5, 0.1M NaCl prepared with deuterated water. Immediately after solubilizing, fresh (GA)6, (GP)6 and (GR)6 were frozen in liquid nitrogen. Aged samples dissolved in deuterated buffer were incubated at 37°C with agitation in an EchoTherm

Orbital mixer (level 9, Torrey Pines Scientific) for 6–8 days, followed by flash freezing in liquid nitrogen. Both fresh and aged samples were lyophilized prior to FT-IR measurements. Spectra were collected at room temperature in a Perkin-Elmer FT-IR equipped with attenuated total reflectance (ATR). 128 scans were accumulated and averaged for each spectrum at a resolution of 4 cm-1. Spectra were corrected for absorption of buffer by subtracting the spectrum of buffer alone.

Primary neuron culture, transfection, and imaging

Primary mixed cortical neurons were dissected from embryonic day 20 rat pups and cultured at 0.6 x 106 cells/ml, as described previously[40,41]. Four days after plating, neurons were transfected with pGW1-mApple[40,41] using Lipofectamine 2000 (Invitrogen, 52887), and peptides (dissolved in 0.1M NaCl, 25 mM sodium phosphate pH 7.4), were applied at a final concentration of 1mM (diluted from 20mM for (GA)3, (GP)3 and (GR)3) or 0.5mM (diluted from 10mM for (GA)6, (GP)6 and (GR)6). “Fresh” peptides were applied to cells immediately after dissolving in solution whereas “aged” samples were applied after 6–8 days of agitation as described above. To avoid contamination, peptides were not subjected to sonication. Longitudinal imaging was initiated 24 hours post-transfection, and accomplished using an automated fluorescence microscopy platform as before[40,41]. Briefly, neurons were imaged using a Nikon TiE-B inverted microscope equipped with a high-numerical aperture 20X objective lens, a PerfectFocus3 system, an ASI2000 stage with rotary encoders in the x- and y-axes, a Lambda XL Xenon lamp (Sutter), and an Andor iXon 897 electron-multiplied charge coupled device (EMCCD) digital camera. All stage, shutter, and filter wheel movements were coordinated by code written in publicly-available software (μManager and ImageJ).

Image processing and survival analysis were accomplished by original code written in Python or the ImageJ macro language. All statistical analyses were performed in R using the survival analysis package, or Prism (GraphPad). Differences in survival between populations of neurons were determined using the log-rank test (for 2 conditions) or Cox proportional hazards analysis (for > 2 conditions).

Animal work

All vertebrate animal work was approved by the Committee on the Use and Care of Animals at the University of Michigan. All experiments were carefully planned so that we use as few animals as possible. Pregnant female wild-type, non-transgenic Long Evans rats (Rattus norvegicus) were housed singly in chambers equipped with environmental enrichment. They were fed ad libitum a full diet (30% protein, 13% fat, 57% carbohydrate; full information available at www.labdiet.com), and cared for by the Unit for Laboratory Animal Medicine (ULAM) at the University of Michigan. Veterinary specialists and technicians in ULAM are trained and approved in the care and long-term maintenance of rodent colonies, in accordance with the NIH-supported Guide for the Care and Use of Laboratory Animals. All rats were kept in routine housing for as little time as possible prior to euthanasia and dissection, minimizing any pain and/or discomfort. Pregnant dams were euthanized by CO2 inhalation at gestation day 20. For each animal, euthanasia was confirmed by bilateral pneumothorax. Euthanasia was fully consistent with the recommendations of the Guidelines on Euthanasia of the American Veterinary Medical Association and the University of Michigan Methods of Euthanasia by Species Guidelines. Following euthanasia, the fetuses were removed in a sterile manner from the uterus and decapitated. Primary cells from these fetuses were dissected

and cultured immediately afterwards, as described above.

Fluorescent labeling of peptides

A stock solution of Fluorescein-5-EX succinimidyl ester (FITC, Molecular Probes, F6130) was prepared by dissolving the reagent in DMSO to a final concentration of 20mM. (GA)6, (GP)6 and (GR)6 were dissolved in 100mM sodium bicarbonate pH 8.5 to 1.5mg/ml. FITC was added to (GA)6, (GP)6 and (GR)6 solution to a final concentration of 8.3mM, 7.0mM and 5.3mM, respectively. All reaction volumes were adjusted to 800 µl. The final concentration the peptides was 0.75mg/ml. To make the unlabeled (control) solution, 400 µl of 100 mM sodium bicarbonate pH 8.5 was mixed with 400 µl of 20 mM FITC stock. All solutions were incubated for 24 hours at 4°C on a nutator. The labeled peptides were dialyzed against 2 L of 25 mM sodium phosphate pH 7.4; 0.1M NaCl. The dialysis buffer was changed four times. Peptide labeling was confirmed with MALDI-TOF (Bruker AutoFlex Speed). After 3 days of dialysis, samples were incubated for a week at 37°C with shaking in EchoTherm Orbital mixer (level 9, Torrey Pines Scientific).

Detection of internalized FITC-labeled peptides

Following conjugation of each DPR to FITC, labeled DPRs were applied to primary rodent cortical neurons on DIV4 at a final concentration of 37.5 µg/ml. The cells were subjected to immunocytochemistry, as described below, using primary antibodies against MAP2 (dilution 1:500, Millipore, MAB3418, mouse monoclonal, clone AP20) and secondary anti-mouse Cy5-conjugated antibodies (dilution 1:250, Jackson Immunoresearch, 115-175-146, Goat whole IgG). The cells were then imaged using a Zeiss LSM510 laser scanning confocal microscope. For micrographs of neurons treated

with FITC-labeled peptides, 20 optical slices were taken for each cell with a step size of 5 nm between slices. Z-sections were constructed using LSM Browser software (Zeiss). A total of > 200 neurons were counted from each population to determine the frequency of internalization.

Generation and characterization of dipeptide repeat antibodies

Rabbit polyclonal antibodies were generated commercially by NeoScientific. Antisera were generated against synthetic (GA)6 or (GR)6 peptides and were affinity purified prior to use. Western blot validation was done on cell lysates from transfected COS-7 cells (purchased from ATCC) as previously described[42]. Briefly, COS-7 cells were lysed in RIPA (50mM Tris, pH8, 150mM NaCl, 0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate) buffer supplemented with Complete mini protease inhibitor cocktail tablets (Sigma/Roche, 11836153001) and passed through a 27-gauge needle to shear DNA. Equal amounts of protein were run on a 12% SDS polyacrylamide gel. After transfer to PVDF membrane (0.2μm, Bio-Rad, 162–0177), blots were incubated with the following antibodies:mouse anti-V5 (Abcam, ab27671, mouse monoclonal, IgG2a; recognizes a Pk epitope in P/V proteins of the paramyxovirus SV5; dilution 1:1000), rabbit anti-GA (dilution 1:100) or rabbit anti-GR (dilution 1:5000) or mouse anti-β-Actin (Sigma, A1978, mouse monocolonal, IgG1; recognizes an epitope located in the N-terminus of the β-isoform of actin; dilution 1:5000). At least three independent experiments were performed and scanned films were processed and quantified using ImageJ software. Peroxidase-AffiniPure Goat Anti-Rabbit IgG (H+L) (Jackson ImmunoResearch, 111-035-144; polyclonal; conjugated to Horseradish Peroxidase) and Peroxidase AffiniPureGoat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch, 115-035-

146; polyclonal; conjugated to Horseradish Peroxidase) were used as secondary antibodies. For immunofluorescence based validation, COS-7 cells were maintained 37°C in 5% CO2 incubators. Dulbecco's modified Eagle's medium (DMEM, Fisher, SH30022FS) supplemented with 10% fetal bovine serum (Fisher, MT35015CV), and 1% Pen-Strep were used as culture media. Cells were transfected using Lipofectamine LTX with Plus Reagent (Thermo Fisher, 15338100) using manufacturer's protocol. At 48 hours after transfection, cells were fixed with 4% paraformaldehyde for 15 minutes, washed, permeabilized with 0.1% triton X-100, blocked with 5% normal goat serum (NGS, Vector labs, S-1000) in 1X PBS containing 0.1% triton X-100, and incubated with mouse anti-GFP (Roche, 11814460001, mixture of clones 7.1 and 13.2 mouse monoclonal antibodies, IgG1k, which recognizes wild type and mutant forms of GFP; dilution 1:1000), rabbit anti-GA (dilution 1:100) or anti-GR (dilution 1:500) overnight. Slides were washed and probed with Goat anti-Mouse IgG (H+L) Secondary Antibody, Alexa Fluor 488 (Thermo Fisher Scientific,A-11001, goat polyclonal, immunogen: gamma immunoglobins heavy and light chains) and Alexa Fluor 555 labeled Goat-anti-Rabbit (Thermo Fisher Scientific,A-21428, goat polyclonal, immunogen: gamma immunoglobins heavy and light chains) antibodies and visualized with an inverted Olympus IX71 epifluorescence microscope with Slidebook software with identical fluorescent settings for each slide.

Immunohistochemical validation of the antibodies was done as in Todd et al [42]. Briefly, human cerebellum from control cases or cases with C9orf72 expansions confirmed by repeat primed PCR testing were deparaffinized and then processed through a basic antigen retrieval protocol to enhance detection. Sections on slides were then

permeabilized with 0.1% triton X- 100, blocked in 5% NGS and 0.1% BSA, and incubated in primary antibodies overnight at 4°C at the following concentrations: anti-GA (1:50), anti-GR (1:100). Following PBS washes, a Vectastain Elite ABC Kit (Rabbit IgG) (Vector Labs, PK-6101) was used according to manufacturer's protocols for DAB staining. Samples were counterstained with hematoxylin QS (Vector Labs, H-3404) to identify nuclei. Selected images are representative of staining from a least three different slides and two patients per group.

Detection of peptide internalization by immunocytochemistry

We added 0.5mM (final concentration) of fresh or aged (GA)6, (GP)6 and (GR)6 to primary neurons (cultured as described above) 4 days after dissection (DIV4). After 24 hours, the cells were rinsed twice in PBS (Life Technologies, 70011–044), then fixed in 4% paraformaldehyde for 10 min. Following 2 more rinses, neurons were permeabilized with 0.1% Triton X-100 in PBS for 20 min at room temperature, equilibrated with 10 mM glycine in PBS for 10 min at room temperature, then blocked in 0.1% Triton X-100, 3% BSA (Research Products International, 9048-46-8) and 0.2% goat serum PBS for 1 hour at room temperature. Primary antibodies against GA (1:100) and GR (1:500), both generated as described above, or GP (1:5000, polyclonal, Millipore ABN455, rabbit polyclonal), in addition to antibodies against MAP2 (1:500,Millipore MAB3418, mouse monoclonal, clone AP20), were added directly to the block and the samples incubated overnight at 4°C. All cells were rinsed twice quickly and 3 times for 10 min with PBS, then placed back in block solution containing the appropriate secondary antibodies (goat anti-mouse Cy5, Jackson ImmunoResearch, 115-175-146, whole IgG; and goat anti-rabbit Alexa Fluor 488, Life Technologies A-11008, whole IgG) at a dilution of 1:250.

The cells were rinsed twice quickly in PBS, and 3 times for 10 min each in PBS containing Hoescht dye (Invitrogen, 33342,) at 100 nM, then twice more in PBS before imaging by automated microscopy. A total of > 100 neurons were counted from each population to determine the frequency of internalization.

2.7 Acknowledgements

We thank Profs Samuel Krimm and Noemi Mirkin for discussions and interpretation of CD and FT-IR data.

2.8 Author contributions

Conceptualization: BNF MED PKT SJB MII.

Formal analysis: BNF MED AK MRS PKT SJB MII.

Funding acquisition: HLP PKT SJB MII.

Investigation: BNF MED AK MRS PKT SJB MII.

Resources: HLP PKT SJB MII.

Supervision: HLP PKT SJB MII.

Validation: BNF MED AK MRS HLP PKT SJB MII.

Visualization: BNF MED AK PKT SJB MII.

Writing—original draft: BNF MED PKT SJB MII.

Writing—review & editing: BNF MED AK MRS HLP PKT SJB MII.

References

1. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72,** 245–256 (2011).
2. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
3. van der Zee, J. *et al.* A Pan-European Study of the C9orf72Repeat Associated with FTLD: Geographic Prevalence, Genomic Instability, and Intermediate Repeats. *Human Mutation* **34,** 363–373 (2013).
4. Khan, B. K. *et al.* Atypical, slowly progressive behavioural variant frontotemporal dementia associated with C9ORF72hexanucleotide expansion. *J Neurol Neurosurg Psychiatry* **83,** 358–364 (2012).
5. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507,** 195–200 (2015).
6. Gijselinck, I. *et al.* A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum:a gene identification study. *The Lancet Neurology* **11,** 54–65 (2011).
7. Xu, Z. *et al.* Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proceedings of the National Academy of Sciences* **110,** 7778–7783 (2013).
8. Mori, K. *et al.* Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* **126,** 881–893 (2013).
9. Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proceedings of the National Academy of Sciences* **110,** E4968–77 (2013).
10. Gendron, T. F., Cosio, D. M. & Petrucelli, L. c9RAN translation: a potential therapeutic target for the treatment of amyotrophic lateral sclerosis and frontotemporal dementia. *Expert Opinion on Therapeutic Targets* **17,** 991–995 (2013).
11. Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525,** 56–61 (2015).
12. Freibaum, B. D. *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525,** 129–133 (2015).
13. Jovičić, A. *et al.* Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18,** 1226–1229 (2015).
14. Chang, Y.-J., Jeng, U.-S., Chiang, Y.-L., Hwang, I.-S. & Chen, Y.-R. The Glycine-Alanine Dipeptide Repeat from C9orf72 Hexanucleotide Expansions Forms Toxic Amyloids Possessing Cell-to-Cell Transmission Properties. *J. Biol. Chem.* **291,** 4903–4911 (2016).
15. Mori, K. *et al.* The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science* **339,** 1335–1338 (2013).
16. Ash, P. E. A. *et al.* Unconventional Translation of C9ORF72 GGGGCC Expansion

Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron* 1–8 (2014). doi:10.1016/j.neuron.2013.02.004
17. May, S. *et al.* C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* **128,** 485–503 (2014).
18. Gendron, T. F. *et al.* Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol* **126,** 829–844 (2013).
19. Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345,** 1139–1145 (2014).
20. Mackenzie, I. R. A. *et al.* Quantitative analysis and clinico-pathological correlationsof different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol* **130,** 845–861 (2015).
21. Mizielinska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345,** 1192–1194 (2014).
22. Zhang, Y.-J. *et al.* Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* **128,** 505–524 (2014).
23. Wen, X. *et al.* Antisense Proline-Arginine RAN Dipeptides Linkedto C9ORF72-ALS/FTD Form Toxic Nuclear Aggregates that Initiate In Vitro and In Vivo Neuronal Death. *Neuron* **84,** 1213–1225 (2014).
24. Yang, D. *et al.* FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathol* **130,** 525–535 (2015).
25. Tao, Z. *et al.* Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Human Molecular Genetics* **24,** 2426–2441 (2015).
26. Yamakawa, M. *et al.* Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Human Molecular Genetics* **24,** 1630–1645 (2014).
27. Tran, H. *et al.* Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* **87,** 1207–1214 (2015).
28. Kanekura, K. *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Human Molecular Genetics* **25,** 1803–1813 (2016).
29. Boeynaems, S. *et al.* Boeynaems et al Sci Reports 2016. *Science* **6,** 1–8 (2016).
30. Gendron, T. F. *et al.* Cerebellar c9RAN proteins associate with clinicaland neuropathological characteristics of C9ORF72 repeat expansion carriers. *Acta Neuropathol* **130,** 559–573 (2015).
31. Tiwari, M. K. & Kepp, K. P. Modeling the Aggregation Propensity and Toxicity of Amyloid-β Variants. *J. Alzheimers Dis.* **47,** 215–229 (2015).
32. Di Scala, C. *et al.* Mechanism of cholesterol-assisted oligomeric channel formation by a short Alzheimer β-amyloid peptide. *J. Neurochem.* **128,** 186–195 (2014).
33. Hung, L. W. *et al.* Amyloid- Peptide (A) Neurotoxicity Is Modulated by the Rate of Peptide Aggregation: A Dimers and Trimers Correlate with Neurotoxicity. *Journal of Neuroscience* **28,** 11950–11958 (2008).

34. Zanuy, D., Porat, Y., Gazit, E. & Nussinov, R. Peptide sequence and amyloid formation: molecular simulations and experimental study of a human islet amyloid polypeptide fragment and its analogs. *Structure* **12,** 439–455 (2004).
35. Saelices, L. *et al.* Uncovering the Mechanism of Aggregation of Human Transthyretin. *J. Biol. Chem.* **290,** 28932–28943 (2015).
36. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447,** 453–457 (2007).
37. Rodriguez, J. A. *et al.* Structure of the toxic core of α-synuclein from invisible crystals. *Nature* **525,** 486–490 (2015).
38. Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical Biochemistry* **287,** 252–260 (2000).
39. Klunk, W. E., Jacob, R. F. & Mason, R. P. Quantifying Amyloid -Peptide (A) Aggregation Using the Congo Red-A (CR–A) Spectrophotometric Assay. *Analytical Biochemistry* **266,** 66–76 (1999).
40. Barmada, S. J. *et al.* Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. *Nat. Chem. Bio.* **10,** 677–685 (2014).
41. Barmada, S. J. *et al.* Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 7821–7826 (2015).
42. Todd, P. K. *et al.* CGG Repeat-Associated Translation Mediates Neurodegenerationin Fragile X Tremor Ataxia Syndrome. *Neuron* **78,** 440–455 (2013).
43. Rucker, R. L. & Creamer, T. P. Polyproline II helical structure in protein unfolded states: Lysine peptides revisited. *Protein Science* **11,** 980–985 (2002).
44. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431,** 805–810 (2004).
45. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1751,** 119–139 (2005).
46. Jones, E. M. & Surewicz, W. K. Fibril Conformation as the Basis of Species- and Strain-Dependent Seeding Specificity of Mammalian Prion Amyloids. *Cell* **121,** 63–72 (2005).
47. Chirgadze, Y. N. & Nevskaya, N. A. Infrared spectra and resonance interaction of amide-I vibration of the paraellel-chain pleated sheets. *Biopolymers* **15,** 627–636 (1976).
48. Teng, P. K. & Eisenberg, D. Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Engineering Design and Selection* **22,** 531–536 (2009).
49. Serpell, L. C., Fraser, P. E. & Sunde, M. X-ray fiber diffraction of amyloid fibrils. *Methods Enzymol* 526–536 (1998).
50. Masters, C. L. & Selkoe, D. J. Biochemistry of Amyloid -Protein and Amyloid Deposits in Alzheimer Disease. *Cold Spring Harb Perspect Med* **2,** a006262–a006262 (2012).
51. Holmes, B. B. & Diamond, M. I. Prion-like Properties of Tau Protein: The

Importance of Extracellular Tau as a Therapeutic Target. *J. Biol. Chem.* **289,** 19855–19861 (2014).
52. Holmes, B. B. *et al.* Heparan sulfate proteoglycans mediate internalization and propagation of specific proteopathic seeds. *Proceedings of the National Academy of Sciences* **110,** E3138–47 (2013).
53. Su, Z. *et al.* Discovery of a Biomarker and LeadSmall Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. *Neuron* 1–41 (2014). doi:10.1016/j.neuron.2014.07.041
54. Luk, K. C. *et al.* Pathological alpha-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **338,** 949–953 (2012).
55. Tavanez, J. P. In vivo aggregation properties of the nuclear poly(A)-binding protein PABPN1. *RNA* **11,** 752–762 (2005).

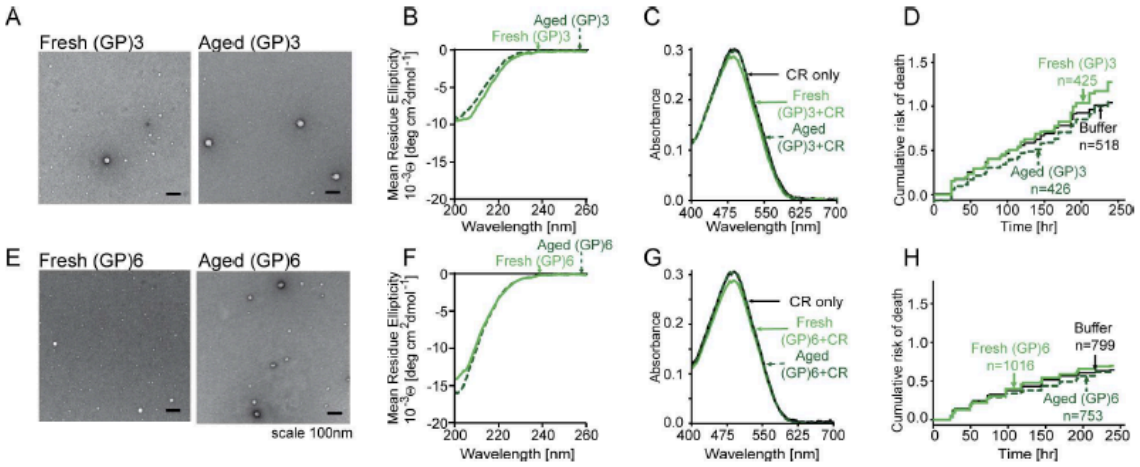

Figure 2.1 GP dipeptides form amorphous aggregates that are innocuous to neurons. (A, E) Fresh and aged (GP)3 and (GP)6 peptides formed low abundance amorphous aggregates by TEM. **(B, F)** CD spectra of fresh and aged (GP)3 and (GP)6 were indistinguishable from one another. **(C, G)** Neither (GP)3 nor (GP)6 peptides exhibited affinity for CR, and their absorbance spectra did not chance in the presence of CR. **(D, H)** Cumulative hazard plots for primary cortical neurons exposed to (GP)3 and (GP)6 peptides. Neither (GP)3 nor (GP)6 significantly impacted neuronal survival ($p > 0.05$) compared to buffer alone. n, number of neurons per condition. Survival data were pooled from 8 wells per condition, in each of 3 replicates.

Table 2.1 Secondary structure content of the peptides in this study calculated from their corresponding CD spectra[38].

Peptide		Helix [%]	Sheet [%]	Turn [%]	Unordered [%]	Range of data used for the predictions [nm]*
(GA)3	Fresh	4	44	21	31	195–260
	Aged	4	44	21	31	195–260
(GP)3	Fresh	7	32	24	37	190–260
	Aged	5	30	25	40	190–260
(GR)3	Fresh	4	41	21	34	200–260
	Aged	4	41	21	34	200–260
(GA)6	Fresh	2	50	19	28	190–260
	Aged	1	55	18	27	190–260
(GP)6	Fresh	7	29	25	40	190–260
	Aged	6	23	27	44	190–260
(GR)6	Fresh	4	40	21	35	200–260
	Aged	4	40	21	35	200–260

*The range specifies the wavelength range of the data which was used for the predictions.

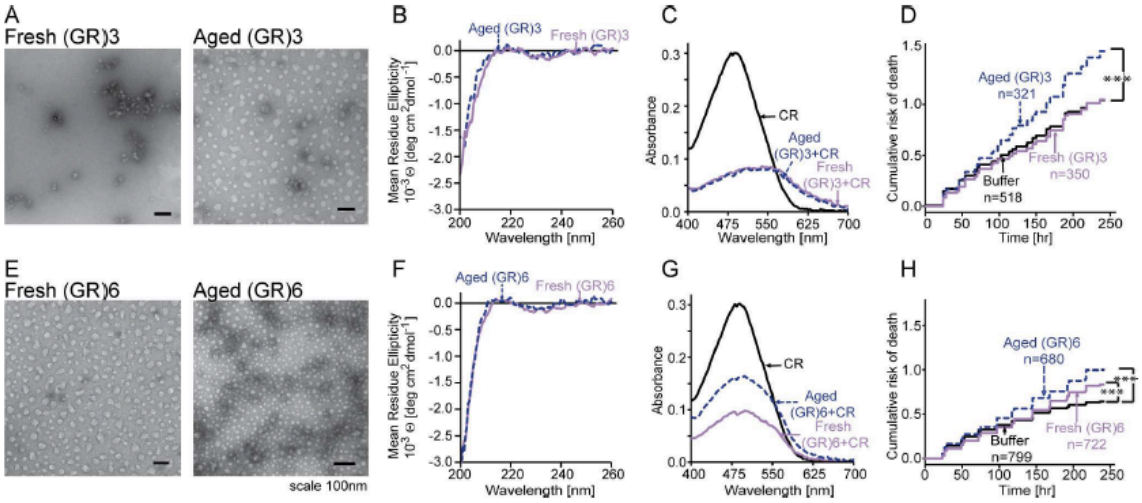

Figure 2.2 GR dipeptides form repeat length-independent neurotoxic aggregates. (A, E) The majority of aged (GR)3 and (GR)6 aggregates were spherical in shape, as assessed by TEM. **(B, F)** CD spectra of fresh and aged (GR)3 as well as fresh and aged (GR)6, were indistinguishable. **(C, G)** (GR)3 peptides produced a spectral shift from 490 nm to 550 nm and decreased the maximum of CR absorbance at 490 nm. CR spectra in the presence of fresh and aged (GR)3 were indistinguishable. In contrast, absorbance spectra of fresh and aged (GR)6 differed, with fresh peptide showing greater decrease in absorbance. **(D, H)** In longitudinal analyses of neuronal survival, aged but not fresh (GR)3 significantly increased the risk of death over control neurons exposed to buffer alone, whereas both fresh and aged (GR)6 were significantly toxic. n, number of neurons per condition. *** $p < 0.0003$, Cox proportional hazards analysis. Survival data were pooled from 8 wells per condition, in each of 3 replicates.

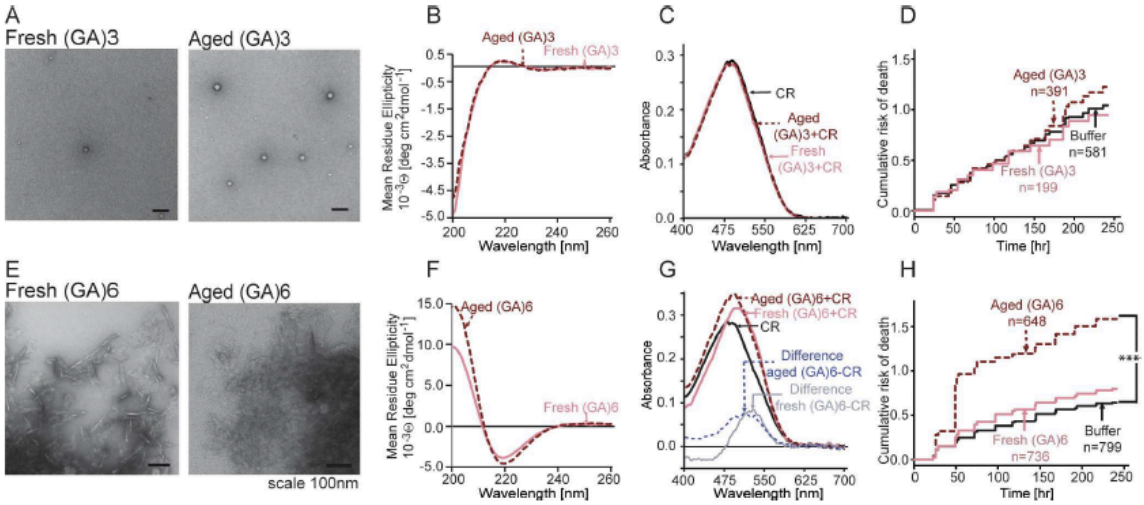

Figure 2.3 GA dipeptides form repeat length-dependent neurotoxic β-sheet rich aggregates. (A, E) By TEM, (GA)3 formed sparse amorphous aggregates. Fresh (GA)6 aggregates were heterogeneous, containing both fibrils and layered sheets. In contrast, aged (GA)6 formed large aggregates mostly composed of layered sheets. **(B, F)** CD spectra of fresh and aged (GA)3 were similar with 44% predicted β-sheet content. Aged (GA)6 had the largest β-sheet content among the peptides in this study (55%). **(C, G)** No change in the CR absorbance spectrum was noted for (GA)3, but fresh and aged (GA)6 displayed increased CR absorbance, as is commonly observed for amyloid-forming proteins. Fresh and aged samples introduced spectral shifts to 530 nm and 515 nm, respectively. **(D, H)** Aged (GA)6 peptides, but not (GA)3 peptides, were highly toxic to neurons compared to buffer only control. n, number of neurons per condition. *** $p < 0.0003$, Cox proportional hazard analysis. Survival data were pooled from 8 wells per condition, in each of 3 replicates.

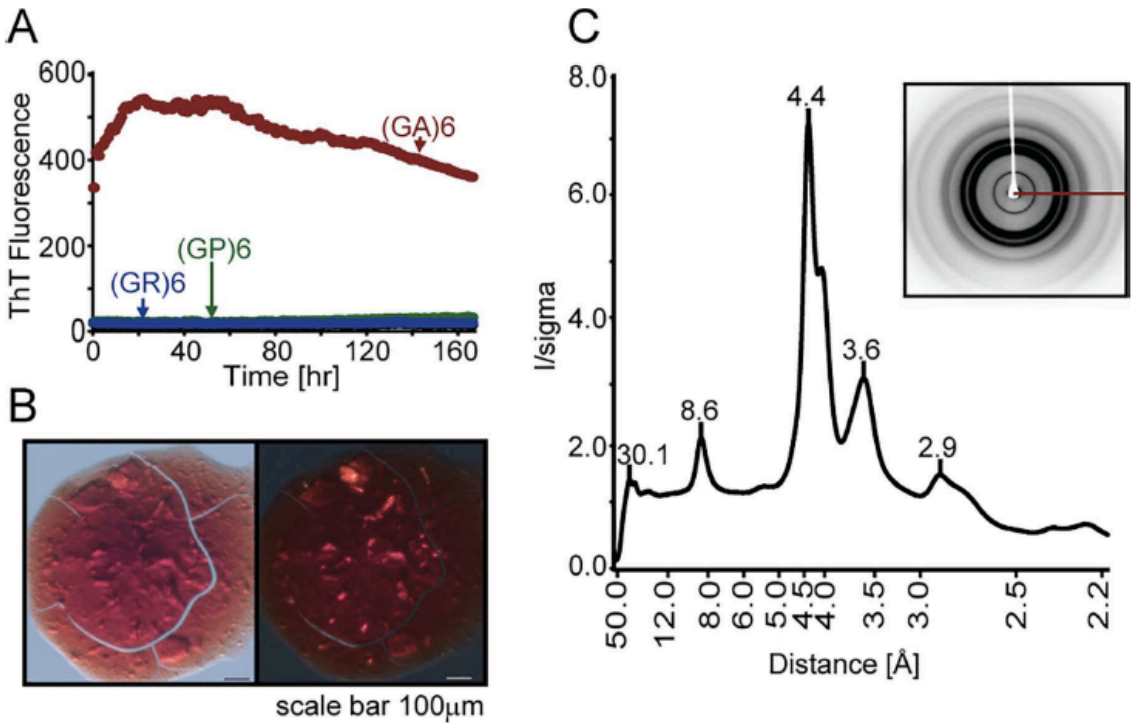

Figure 2.4 (GA)6 forms atypical amyloid-like aggregates. (A) ThT binding assays for all peptides, demonstrating significant binding only for (GA)6 (red line). **(B)** Aged (GA)6 in the presence of CR appeared red with unpolarized light (left) and displayed yellow birefringence with cross-polarized light (right). **(C)** X-ray diffraction pattern from (GA)6 peptides suggests that the aggregates assume a discrete structure.

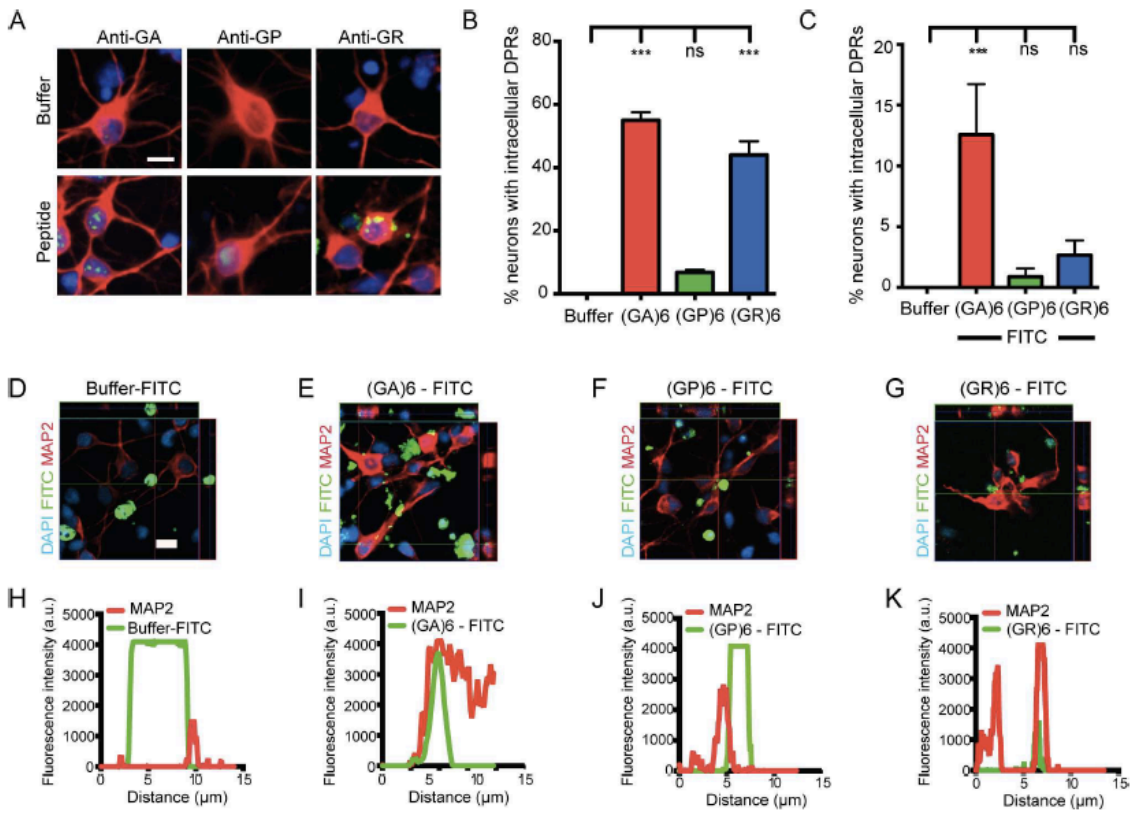

Figure 2.5 DPRs are internalized by neurons. (A) Fluorescence micrographs of neurons treated with the indicated peptide and stained with antibodies recognizing the pan-neuronal marker (MAP2, red), the indicated DPR (GA, GP, or GR; green), and a nuclear dye (4',6-diamidino-2-phenylindole (DAPI), blue). Scale bar, 20 µm. **(B)** Frequency of DPR internalization detected by immunocytochemistry. **(C)** Quantification of the data in panels **D-G**, showing internalization of FITC-labeled DPRs. **(D-G)** Confocal microscopy of neurons treated with the indicated peptide shows intracellular accumulations of FITC–(GA)6 and –(GR)6, but not –(GP)6 or buffer. Representations of the y-z axis appear above, and x-z plots to the right of each micrograph. Scale bar in **(D)**, 20 µm, applies to **D-G**. **(H-K)** Intensity plots of fluorescence across a line that intercepts the inclusions shown in D-G, demonstrating overlapping MAP2 and FITC signals for (GA)6 and (GR)6 but not (GP)6. For **(B)** and **(C)**, ***, $p < 0.05$ one-way ANOVA with Dunnett's test. n= 100-150 neurons per condition, pooled from 3 replicates and 2 independent experiments. ***, $p < 0.05$ one-way ANOVA with Dunnett's test.

Supplemental information

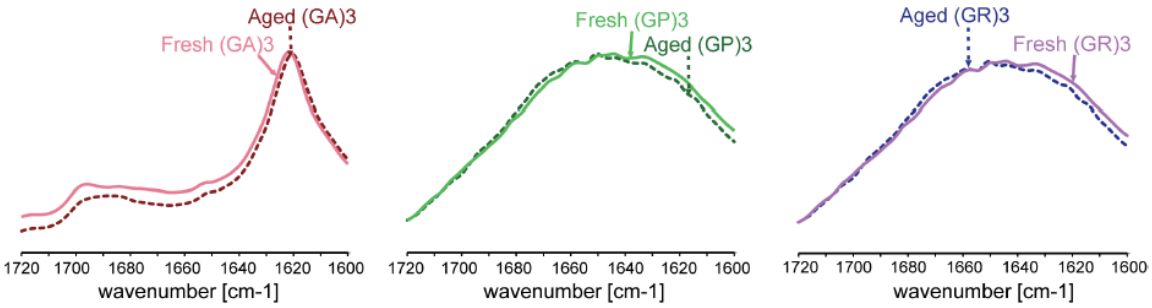

Figure S2.1 FT-IR spectra of dipeptide repeats. To assess for secondary structure, fresh and aged (GA)6, (GP)6 and (GR)6 peptides were subject to FT-IR. The two bands at 1622 cm^{-1} and 1698 cm^{-1} in (GA)6 spectra suggest that the secondary structure of these aggregates is mostly β-sheet.

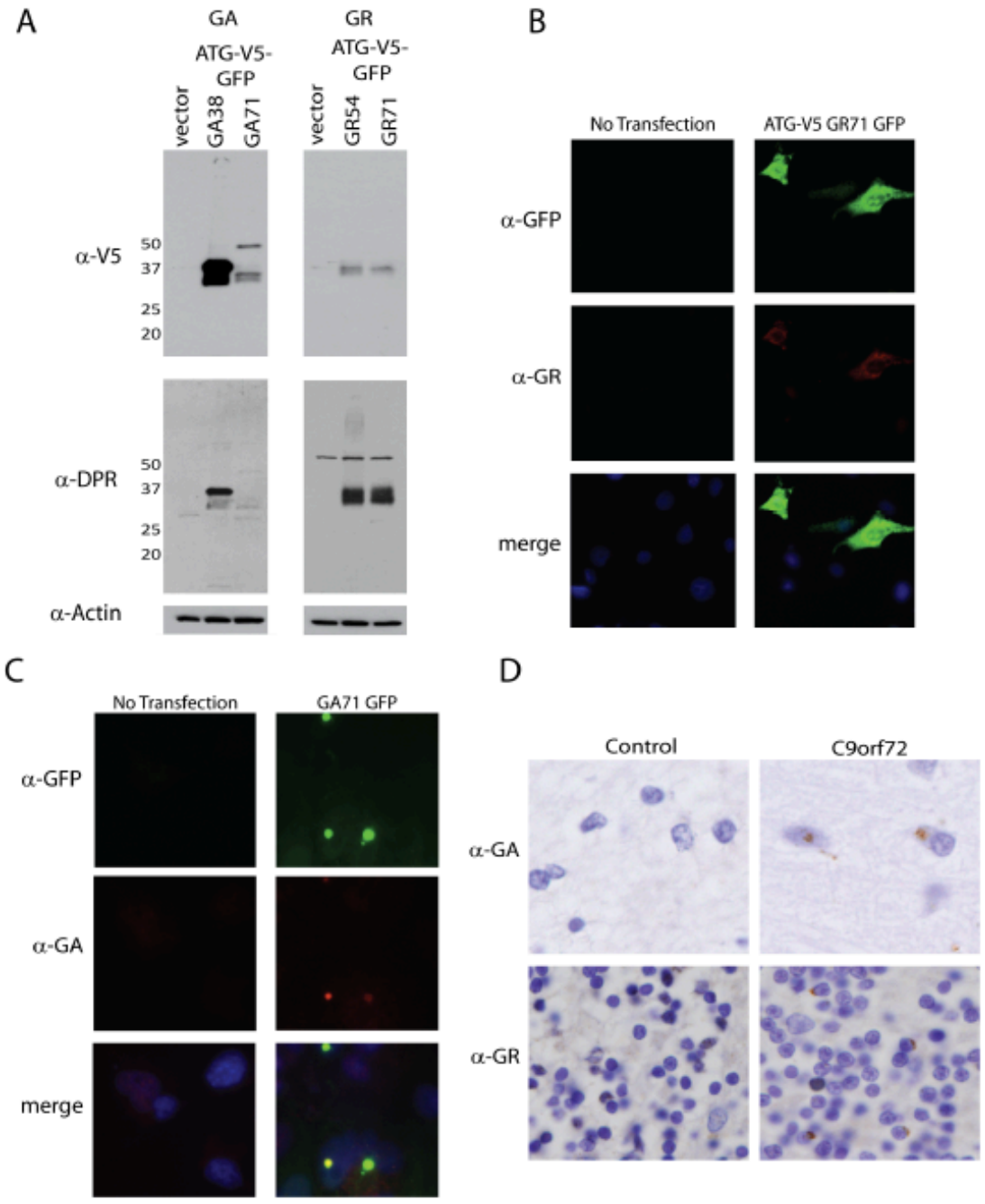

Figure S2.2. Characterization of new C9orf72 dipeptide repeat antibodies. Rabbit polyclonal antibodies were generated against the sense strand-derived DPRs GA and GR. **(A)** Western blot of lysates from COS-7 cells transfected with the indicated vectors. Blots were serially probed with the indicated DPR antibody, then V5 and then actin as a loading control. **(B, C)** Immunocytochemistry of COS-7 cells transfected with the indicated vectors and stained for GR **(B)** and GA **(C)** antibodies demonstrate specificity. **(D)** Immunohistochemistry of GA- and GR- protein aggregates in the cerebellum of a genetically confirmed *C9orf72*(+) ALS case. No significant staining was observed in control patient cerebellum.

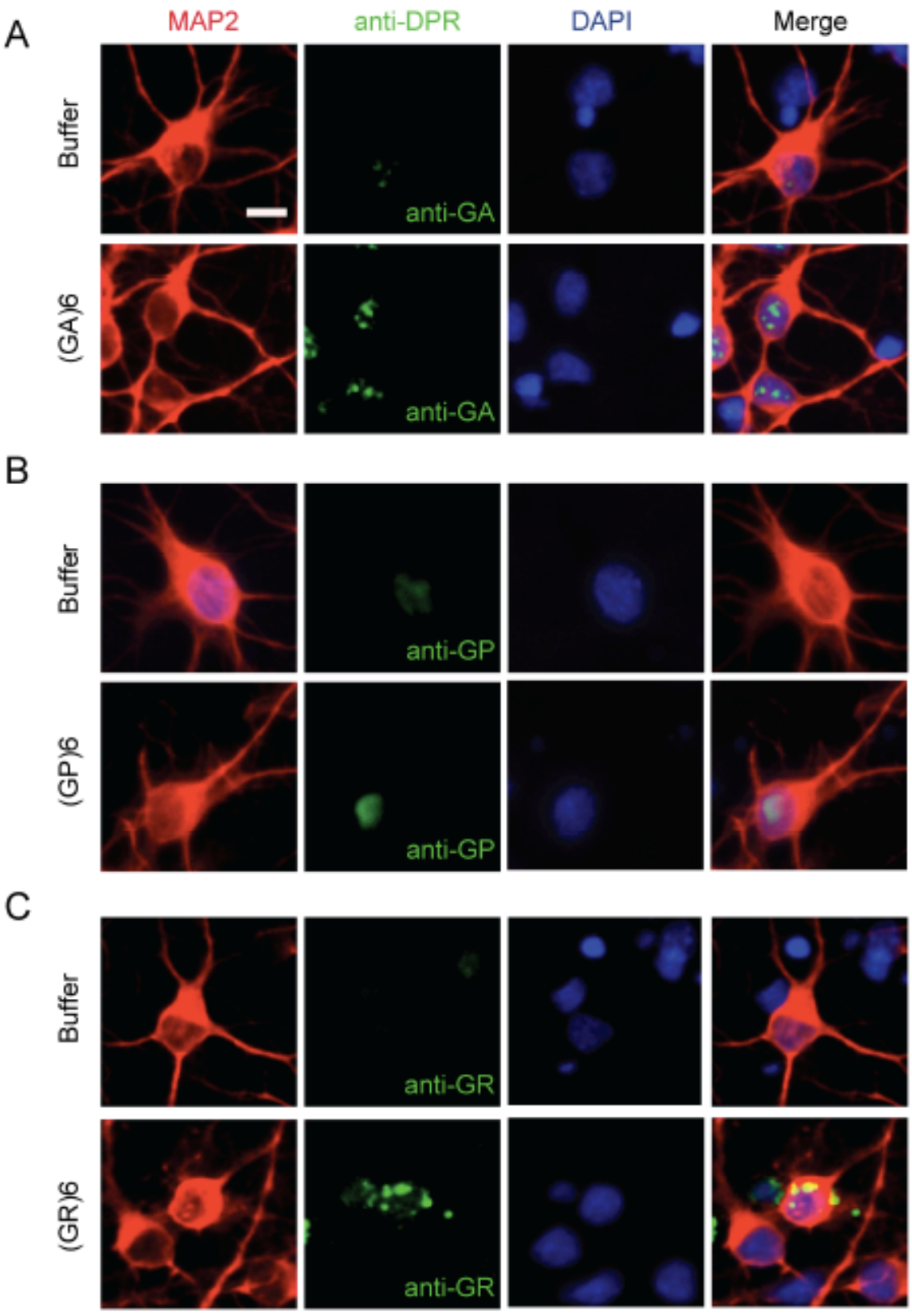
A
MAP2
anti-DPR
DAPI
Merge
Buffer
anti-GA
(GA)6
anti-GA
B
Buffer
anti-GP
(GP)6
anti-GP
C
Buffer
anti-GR
(GR)6
anti-GR

Figure S2.3 Detection of internalized DPRs by immunocytochemistry. 24 h after application of (GA)6, (GP)6 or (GR)6 peptides to live rodent primary cortical neurons, the cells were fixed and immunostained using antibodies against MAP2 (red) and each of the DPRs (green), and nuclei labeled with DAPI (blue). **(A)** Neurons treated with (GA)6 displayed nuclear foci. **(B)** Diffuse nuclear staining was noted in a small proportion of cells exposed to (GP)6. **(C)** Following application of (GR)6, DPRs were detected within neuronal cytoplasmic aggregates. See Figure 2.5B for quantification of DPR internalization. Scale bar, 20 μm.

Chapter 3

An intramolecular salt bridge linking TDP43's RNA recognition motifs dictates RNA binding and TDP43-dependent neurodegeneration*

3.1 Abstract

The majority of individuals with amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) exhibit neuronal cytoplasmic inclusions rich in the RNA binding protein TDP43. Even so, the relationship between TDP43's RNA binding properties and neurodegeneration remain obscure. Here we show that engineered mutations disrupting a salt bridge between TDP43's RNA recognition motifs interfere with nucleic acid binding and eliminate recognition of native TDP43 substrates. These same mutations dramatically destabilize TDP43, alter its subcellular localization and abrogate TDP43-dependent neurodegeneration. *C. elegans* harboring homologous TDP-1 mutations phenocopy knockout strains, confirming the necessity of the salt bridge residues for TDP43 function. Moreover, the accumulation of functional TDP43, but not RNA binding-deficient variants, disproportionately affects the abundance and splicing of

* This chapter represents the following manuscript:

Flores BN, Li X, Martinez J, Beg AA, Barmada SJ. An intramolecular salt bridge linking TDP43's RNA recognition motifs dictates RNA binding and TDP43-dependent neurodegeneration. *In revision.*

transcripts encoding oxidative phosphorylation and ribosome components. These studies demonstrate the significance of the salt bridge in sustaining TDP43 stability and RNA binding specificity, factors that are crucial for neurodegeneration arising from TDP43 deposition in ALS and FTD.

3.2 Introduction

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are distinct neurodegenerative disorders that share key pathologic and genetic features[1]. Mutations in several RNA binding proteins (RBPs) cause familial ALS and FTD, including TDP43, FUS, hnRNPA1 and hnRNPA2B1, MATR3, and TIA1[2]. Many of the most common mutations responsible for ALS and FTD, including the *C9orf72* hexanucleotide expansion[3,4], result in neuronal cytoplasmic inclusions rich in TDP43[5,6], a nuclear RBP involved in RNA processing, stability, and transport[7]. These observations strongly suggest that dysfunctional RNA metabolism is a convergent pathogenic mechanism responsible for neurodegeneration in ALS and FTD[8].

TDP43 contains two highly conserved RNA recognition motifs, RRM1 and RRM2, that exhibit distinct binding characteristics. RRM1 has a higher affinity for specific sequences, and while RRM2 alone binds little if any RNA, its presence enhances the sequence specificity of RNA recognition by RRM1[9-12]. In *Drosophila,* deletion of TDP43's RRM1 or substitution of Trp113 to alanine almost completely eliminates RNA binding and prevents downstream toxicity from TDP43 accumulation[10,13]. Likewise, mutation of two key phenylalanine residues (Phe147/149 to Leu) within RRM1 is sufficient to abrogate RNA binding by TDP43[9,14] and similar mutations rescue TDP43-

dependent toxicity in yeast and *Drosophila*[15,16]. RNA binding by TDP43 therefore appears to be critical for neurodegeneration in disease models. Even so, little is known about the significance of sequence-specific RNA binding by TDP43, and the RNA substrates that mediate TDP43-related neurodegeneration remain obscure.

Over 40 different pathogenic mutations within the gene encoding TDP43 (*TARDBP*) have been identified in families with ALS, FTD or both[17]. Disease-associated *TARDBP* mutations elicit gain-of-function toxicity by interfering with TDP43 autoregulation[18-20], enhancing cytoplasmic TDP43 mislocalization and deposition, and impacting TDP43 clearance[21-25]. Supporting the link between TDP43 turnover and neurodegeneration, toxicity is strongly and directly proportional to TDP43 abundance in individual cells, and stimulation of TDP43 turnover via autophagy extends neuronal survival and mitigates disease phenotypes in ALS and FTD models[23]. Additionally, the half-life of TDP43 within primary neurons is significantly longer than in fibroblasts or transformed cell lines[23,24,26], indicating preferential stabilization of the protein in neurons, and suggesting cell type-specific differences that could be important for the selective vulnerability of neurons in ALS and FTD. However, the factors responsible for maintaining TDP43 stability, and the contribution of TDP43 clearance to neurodegeneration, are unknown.

Upon RNA binding, intramolecular interactions between RRM1 and RRM2, mediated by a salt bridge between Arg151 (located in RRM1) and Asp247 (located in RRM2) are necessary for maintaining TDP43's ability to recognize RNA[27]. Here, we show that the RRM1-RRM2 salt bridge is pivotal not only for TDP43's RNA binding properties, but also its stability. TDP43 variants lacking the salt bridge are unable to bind

RNA substrates, degrade rapidly and ultimately incapable of triggering neurodegeneration, despite cytoplasmic mislocalization in many cases. Furthermore, although TDP43 overexpression engenders widespread changes in RNA abundance and splicing, most splicing events are unrelated to the RNA binding ability of TDP43. Among transcripts selectively affected by the accumulation of wild-type TDP43, the ribosomal and oxidative phosphorylation pathways are strongly enriched, emphasizing a potential role for TDP43 in regulating protein synthesis and energy production.

3.3 Results

3.3.1 The R151-D247 salt bridge is essential for nucleic acid binding by TDP43

To determine the significance of the TDP43 RRM1-RRM2 salt bridge, we mutated both residues participating in the intramolecular interaction[27]—Arg151 (R151) and Asp247 (D247)—to alanine (Fig. 3.1A, B). We purified recombinant TDP43(WT), TDP43(R151A), and TDP43(D247A) in *E. coli*, then tested the ability of each variant to bind nucleic acid *in vitro* via electromobility shift assays. Since TDP43(WT) primarily recognizes RNA and DNA sequences rich in UG and TG repeats[9,10], respectively, we first asked whether disruption of the RRM1-RRM2 salt bridge affects TDP43's affinity for repetitive UG elements. Mobility shifts were noted for all three variants (Fig. 3.1C), suggesting successful recognition of $(UG)_{12}$ oligonucleotides in each case, and indicating that the salt bridge is not necessary for RNA binding *per se*.

To determine if disruption of the RRM1-RRM2 salt bridge produced more subtle effects on RNA binding, we applied increasing concentrations of unlabeled $(UG)_{12}$ oligonucleotide to each preparation, and calculated a dissociation constant (K_d) and

cooperativity coefficient (Hill slope, h) for each TDP43 variant (Fig. 3.1D-I). The calculated K_d for TDP43(WT) in these assays was 41.1 ± 2.31 nM, similar to estimates from previous studies[9-12,28,29], and the Hill slope (h) was 1.44 ± 0.14, indicating that TDP43 binds RNA in a cooperative manner[11,30]. In these quantitative assays, both the R151A and D247A mutations significantly reduced the affinity of TDP43 for $(UG)_{12}$ (K_d= 150 ± 6.68 nM and 155.8 ± 8.59 nM, respectively) in comparison to TDP43(WT) (Fig. 3.1F-I). In contrast, the R151A and D247A mutations moderately enhanced binding cooperativity, increasing the Hill slope to 1.94 ± 0.14 and 1.84 ± 0.15, respectively.

TDP43 is capable of recognizing DNA as well as RNA[9,10]. Therefore, we conducted similar binding studies using $(TG)_{12}$ DNA oligonucleotides in place of $(UG)_{12}$, noting analogous mobility shifts for all three TDP43 variants (Fig. S3.1A). However, more quantitative analyses revealed unique effects of the R151A and D247A mutations on DNA binding by TDP43: while the dissociation constant for TDP43(R151A) (27.19 ± 1.9 nM) was not significantly different from that of TDP43(WT) (26.51 ± 2.3 nM) (Fig. S3.1B, C), this variant displayed marked reduction in the Hill slope, suggesting a loss of cooperative binding (Fig. S3.1D, E). TDP43(D247A) exhibited significantly weaker affinity for $(TG)_{12}$ (K_d= 47.88 ± 2.326 nM) in comparison to TDP43(WT) and TDP43(R151A), and an intermediate Hill slope of 1.387 ± 0.118 (Fig. S3.1F, G). Collectively, these data indicate that the salt bridge is required for maintaining both binding cooperativity and affinity of TDP43 for UG/TG-rich sequences, but the specific consequences of salt bridge disrupting-mutations are distinct for RNA and DNA.

We next asked whether disruption of the RRM1-RRM2 salt bridge affects the sequence specificity of nucleic acid recognition by TDP43. No significant mobility shifts

were detected upon incubation of TDP43(WT), TDP43(R151A) or TDP43(D247A) with $(GC)_{12}$ sequences, even at the highest protein concentrations (Fig. 3.1J). However, we observed clear shifts for both TDP43(R151A) and TDP43(D247A), but not TDP43(WT), in the presence of $(AT)_{12}$ oligomers (Fig. 3.1K), suggesting that the salt bridge mutants display aberrant sequence specificity *in vitro*. Notably, the $(AT)_{12}$-induced shift for TDP43(R151A) and TDP43(D247A) occurred at higher protein concentrations compared to those required for $(UG)_{12}$ or $(TG)_{12}$ sequences, implying a reduced affinity for $(AT)_{12}$ sequences. Indeed, in quantitative studies (Fig. S3.2) the K_d for both TDP43(R151A) (0.637 ± 0.052 μM) and TDP43(D247A) (1.524 ± 0.072 μM) for $(AT)_{12}$ oligomers was significantly higher than for $(UG)_{12}$ (Fig. 3.1) or $(TG)_{12}$ sequences (Fig. S3.1).

To examine the significance of the RRM1-RRM2 salt bridge for TDP43 RNA binding *in vivo*, we overexpressed EGFP-tagged TDP43(WT), TDP43(R151A), and TDP43(D247A) in HEK293T cells. As a control, we also overexpressed EGFP-tagged TDP43(F147L, F149L), a TDP43 variant harboring mutations of two critical phenylalanine residues within RRM1 required for nucleic acid binding[9]. Following transfection, each TDP43 variant was immunoprecipitated and TDP43-bound RNA assessed by qRT-PCR. As expected, two endogenous TDP43 substrates, *TARDBP* and *MALAT1*[14,30-33], were immunoprecipitated with TDP43(WT) but not TDP43(F147L, F149L) (Fig. 3.1L, M), consistent with the inability of the latter to recognize RNA[9]. We also noted striking reductions in the amount of *TARDBP* and *MALAT1* transcripts bound to TDP43(R151A) and TDP43(D247A). Together with the results of the above experiments, these data show that disruption of the salt bridge affects the affinity,

cooperativity and sequence specificity of nucleic acid recognition by TDP43, and effectively eliminates the recognition of at least a subset of native RNA targets *in vivo*.

3.3.2 The RRM1-RRM2 salt bridge is required for TDP43-mediated neurodegeneration

Since over 80% of ALS arises sporadically yet nonetheless displays TDP43 pathology[5,34,35], overexpression of TDP43(WT) may mimic aspects of sporadic disease pathogenesis. Supporting this notion, overexpression of TDP43(WT) in multiple model systems is sufficient to reproduce key features of disease, including cytoplasmic TDP43 mislocalization, the formation of ubiquitinated TDP43 aggregates, and neurodegeneration[15,21,23,36-39]. Given the effects of salt bridge-disrupting mutations on the nucleic acid binding properties of TDP43 (Fig. 3.1), and prior evidence suggesting that RNA binding by TDP43 is required for toxicity[13,15], we surmised that TDP43(R151A) and TDP43(D247A) would display reduced toxicity in comparison to TDP43(WT) when overexpressed. To test this hypothesis, we cultured primary mixed cortical neurons from rodents and transfected them with vectors encoding mApple, to visualize the cell body, and EGFP-tagged versions of TDP43(WT), TDP43(R151A) or TDP43(D247A). Neurons were also transfected with mApple and EGFP alone, as a negative control. Using automated microscopy[21,23], we imaged transfected neurons at regularly spaced 24 hr intervals, for a total of 10 days. Individual neurons were then identified and tracked by custom-written algorithms, and time of death determined by programs that detect abnormal cellular morphology (i.e. blebbing) or loss of mApple fluorescence (Fig. 3.2A), sensitive indicators of cell death established in prior studies[23,40,41]. Differences in survival between populations of neurons were measured relative to a reference group and

expressed as a hazard ratio (HR). As noted in previous investigations[21,23,41], overexpression of TDP43(WT) significantly increased the risk of death over EGFP alone (HR 5.04, $p < 2x10^{-16}$), and the F147L/F149L mutation effectively reduced TDP43-dependent toxicity. The HR for the comparison of TDP43(F147L, F149L) and TDP43(WT) was 0.27 ($p < 2x10^{-16}$), indicating an 80% reduction in toxicity (Fig. 3.2B) associated with the F147L/F149L double mutant. The R151A and D247A mutations similarly decreased the toxicity of TDP43 when overexpressed in neurons, by 75% and 68% respectively ($p < 2x10^{-16}$ for both comparisons). Given the reduced ability of TDP43(R151A) and TDP43(D247A) to bind nucleic acid (Fig. 3.1), these results imply that RNA binding activity is crucial for TDP43-mediated neurodegeneration.

Because R151 lies within RRM1 and D247 within RRM2, we next asked if deletion of RRM1 or RRM2 could recapitulate the effects of each mutation on TDP43-dependent toxicity. When expressed in primary mixed cortical neurons, both TDP43(ΔRRM1) and TDP43(ΔRRM2) were significantly less toxic than TDP43(WT) (Fig. 3.2C; HR 0.26 and HR 0.51, respectively, $p < 2x10^{-16}$ for both comparisons). In fact, neurons expressing TDP43(ΔRRM1) displayed a reduced risk of death in comparison to the negative control (EGFP alone; HR 0.7, $p < 2x10^{-16}$), suggesting a neuroprotective effect associated with this construct, and further implying that RRM1 is necessary for TDP43-related toxicity. Conversely, overexpression of TDP43(ΔRRM2) increased the risk of death by 42% in comparison to EGFP alone (HR 1.42, $p < 2x10^{-16}$). Together with prior data highlighting the ability of RRM1, but not RRM2, to bind RNA *in vitro*[9,11,12], these observations provide strong evidence linking RNA binding to toxicity from TDP43 overexpression.

Over 40 different mutations occurring within the *TARDBP* locus are associated with familial ALS and FTD[17], and overexpression of pathogenic TDP43 mutants is just as toxic or even more toxic than TDP43(WT) [21,23,38,42-44]. To determine if the RRM1-RRM2 salt bridge is required for mutant TDP43-related neurodegeneration, we transfected rodent primary cortical neurons with EGFP alone, or EGFP-tagged versions of the pathogenic TDP43(M337V) variant carrying the R151A, D247A or F147L/F149L mutations. Transfected neurons were imaged by automated microscopy, as before, and the risk of death within each population assessed by survival analysis. In doing so, we observed significant reductions in toxicity when the R151A, D247A, or F147L/F149L mutations were introduced on the pathogenic M337V background (Fig. 3.2D). The TDP43(D247A, M337V) double mutant displayed more toxicity than the other variants, suggesting that the D247A mutation prevents most but not all neurodegeneration associated with the M337V mutation. Nonetheless, these data demonstrate that the RRM1-RRM2 salt bridge and RNA binding are essential for toxicity caused by overexpression of WT as well as pathogenic mutant TDP43.

3.3.3 Mutations that disrupt the RRM1-RRM2 salt bridge enhance TDP43 turnover

The RRM1-RRM2 salt bridge helps maintain the tertiary structure of TDP43[27], which in turn may be important for protein folding and stabilization. To determine whether the RRM1-RRM2 salt bridge stabilizes TDP43, we turned to optical pulse labeling (OPL), a technique enabling measurement of protein half-life within living cells[23,45,46]. TDP43 variants were tagged with Dendra2, a green fluorescent protein that is irreversibly converted to a red fluorescent state upon illumination with low wavelength (405 nm) light[47] (Fig. 3.3A). Each pulse of 405 nm light creates a finite amount of

photoconverted, red-fluorescent TDP43-Dendra2. Over time, the red fluorescence decreases as the protein is degraded, providing a means of calculating TDP43-Dendra2 turnover within each cell[23].

Rodent primary cortical neurons were transfected with Dendra2-tagged versions of TDP43(WT) or TDP43 carrying the R151A, D247A, or F147L/F149L mutations, and imaged by automated microscopy after a brief pulse of 405 nm illumination. Single-cell red fluorescence intensities were tracked over time, and TDP43 half-life determined for neurons expressing each of the TDP43 variants (Fig. 3.3B, C). In this way, we observed striking destabilization of TDP43 carrying the R151A and D247A mutations—the median half-life dropped from 58.5 hr for TDP43(WT)-Dendra2, to 15.3 hr and 21.8 hr for Dendra2-tagged versions of TDP43(R151A) and TDP43(D247A), respectively ($p < 0.0001$ for both comparisons). Additionally, the F147L/F149L mutation similarly reduced TDP43 half-life, indicating that deficiencies in RNA binding, disruption of intramolecular interactions, or both are sufficient to destabilize TDP43.

Many misfolded proteins are quickly degraded by the ubiquitin-proteasome system (UPS) during or soon after translation[48]. To determine if the R151A, D247A or F147L/F149L mutations triggered UPS-dependent protein degradation, we treated neurons expressing each of the Dendra2-tagged TDP43 variants with 25 nM MG132, a reversible UPS inhibitor[49], then assessed protein half-life by automated microscopy and OPL (Fig. 3.3D). We again noted significant destabilization of TDP43 by the R151A, D247A and F147L/F149L mutations, but in each case, treatment with MG132 effectively prolonged protein half-life. At higher doses of MG132 we noted nearly complete inhibition of mutant TDP43 degradation (Fig. S3.3). These results imply that

manipulations that prevent TDP43 from binding RNA or interfere with the TDP43 RRM1-RRM2 salt bridge trigger protein misfolding and subsequent degradation by the UPS.

3.3.4 Both RNA binding and protein stability drive TDP43-dependent toxicity

Our data show that mutations that disrupt the RRM1-RRM2 salt bridge abrogate TDP43-mediated toxicity (Fig. 3.2) in conjunction with reduced nucleic acid binding (Fig. 3.1) and accelerated protein turnover (Fig. 3.3). To determine which of these effects is primarily responsible for mitigating neurodegeneration upon TDP43 overexpression, we pursued two complementary strategies. First, we asked if TDP43-related toxicity could be recapitulated by replacing the TDP43 RNA recognition motifs (RRM1 and RRM2) with those from a different RNA binding protein (Fig. 3.4A). Neurons overexpressing TDP43(PUM2)—a chimera consisting of the RNA binding domain of Pumilio 2 (PUM2, a murine Puf protein[50,51]) in place of RRM1 and RRM2—displayed significantly elevated risk of death in comparison to EGFP alone (Fig. 3.4B, C; HR 2.54, $p < 2x10^{-16}$). No significant difference in survival was observed between cells expressing TDP43(WT) or TDP43(PUM2), indicating that the PUM2 RNA binding domain was able to fully complement RRM1-RRM2 in reproducing TDP43-dependent toxicity in primary neurons. In contrast, only slight toxicity was observed upon overexpression of the PUM2 RNA binding domain alone in comparison to EGFP (HR 1.19, $p= 3.78 \times 10^{-5}$; Fig. 3.4B). We then mutated 3 conserved residues within helix 7 of the PUM2 RNA binding domain (mPUM2), effectively disrupting substrate recognition by PUM2[52]. These mutations significantly reduced the toxicity of TDP43(PUM2) (Fig. 3.4B), reinforcing the relationship between RNA binding activity and neurodegeneration. In addition, whereas

TDP43(PUM2) exhibited a diffuse distribution within the nucleus of transfected neurons, TDP43(mPUM2) formed multiple, spherical nuclear puncta (Fig. 3.4C). As discussed below, we observed similar puncta within the nuclei of neurons expressing TDP43(F147L, F149L). These results suggest that TDP43 forms nuclear puncta when it fails to bind to RNA, similar to the related RNA binding proteins FUS[53] and MATR3[54] (Malik *et al.*, in press).

Since the RNA sequence recognized by PUM2 (UGGANAUA)[50] shares some similarity with the UG-rich sequences bound by TDP43[32], we suspected that the recapitulation of TDP43-dependent toxicity by TDP43(PUM2) could be due to overlapping RNA substrates. To test this notion, we instead replaced the TDP43 RRM1-RRM2 domains with the RNA binding domain of the bacteriophage PP7 coat protein[55]. Unlike PUM2, PP7 recognizes unique stem-loop structures formed by repeating elements of 25 nucleotides[56,57]. While overexpression of TDP43(PP7) was significantly less toxic than TDP43(WT) (Fig. 3.4D; HR 0.73, $p < 2x10^{-16}$), it was still substantially more toxic than EGFP alone (HR 2.05, $p < 2x10^{-16}$). Taken together, these data show that TDP43-mediated neurodegeneration strongly depends on the ability of TDP43 to bind RNA, but the sequence specificity required for toxicity is less stringent.

We next asked if modulating TDP43 stability impacts neurodegeneration upon TDP43 overexpression. To enhance TDP43 clearance, we fused a destabilizing sequence (CL1)[58,59] to TDP43-Dendra2, and expressed the construct in primary neurons. Using automated microscopy and OPL, we noted a marked reduction in the half-life of TDP43(WT)-Dendra2-CL1 compared to TDP43(WT)-Dendra2 (Fig. 3.4E), indicating effective destabilization of the CL1 fusion protein. We then assessed neuronal lifetimes

in transfected neurons by automated microscopy and survival analysis. In these studies, the artificially destabilized TDP43(WT)-Dendra2-CL1 fusion was significantly less toxic than TDP43(WT)-Dendra2 (HR 0.64, $p < 2 \times 10^{-16}$; Fig. 3.4F), suggesting that rapid turnover of TDP43 limits its toxicity. We also sought to destabilize TDP43(WT) using a PEST sequence; like CL1, this domain triggers rapid protein clearance in many contexts[60]. Nevertheless, TDP43(WT)-PEST exhibited only a slight reduction in half-life compared to TDP43(WT), and a similarly modest enhancement of toxicity in longitudinal assessments of neuronal survival (Fig. S3.4). These data have two important implications: first, substantial destabilization of TDP43 is required to prevent neurodegeneration in overexpression models. Secondly, mutations that eliminate RNA binding or disrupt the RRM1-RRM2 salt bridge may mitigate toxicity not just by interfering with RNA recognition, but also by destabilizing TDP43.

3.3.5 Manipulation of the RRM1-RRM2 salt bridge modifies TDP43 stability and toxicity

Among the residues that are capable of forming non-covalent interactions with arginine, glutamate forms the most stable interaction, followed by aspartate and arginine[61,62]. We therefore generated the D247E mutation, which we predicted would form a stronger interaction with R151. We also swapped the residues participating in the salt bridge itself (R151D-D247R) (Fig. 3.4G). EGFP-tagged versions of each TDP43 variant were overexpressed in HEK293T cells, immunoprecipitated using anti-EGFP antibodies, and bound RNA substrates assessed by qRT-PCR. In contrast to TDP43(WT), TDP43 variants carrying the D247E and R151D-D247R mutations were incapable of pulling down *TARDBP* and *MALAT1* transcripts (Fig. 3.4H, I). Similar effects were

observed for TDP43 variants harboring R151D or D247R mutations individually (Fig. S3.5). These findings suggest that RNA recognition by TDP43 depends on unique properties of the R151-D247 interaction.

Since the RRM1-RRM2 salt bridge is necessary for maintaining TDP43 stability (Fig. 3.3), we expected that modification of the salt bridge would affect TDP43 turnover as well as its RNA binding properties. We therefore created Dendra2-tagged versions of TDP43 carrying the D247E, D247R, R151D and R151D-D247R mutations, expressed them in primary cortical neurons, and measured TDP43 half-life by automated microscopy and OPL (Fig. 3.4J, K). Compared to TDP43(WT), each of these variants demonstrated accelerated turnover, suggesting that targeted manipulation of the RRM1-RRM2 salt bridge, including changes meant to strengthen the salt bridge, leads to protein misfolding and subsequent degradation.

In light of these data, we predicted that modifications to the RRM1-RRM2 salt bridge would mitigate TDP43-dependent toxicity. We tested this idea by expressing EGFP-tagged versions of each variant in primary neurons, imaging the cells by automated microscopy and comparing the risk of death among populations by survival analysis (Fig. 3.4L). Overexpression of TDP43(D247E) and TDP43(R151D-D247R) was less toxic than TDP43(WT) (HR 0.6, $p < 2x10^{-16}$; and HR 0.21, $p < 2x10^{-16}$, respectively), consistent with our expectations. However, unlike TDP43(R151D-D247R)-expressing neurons, cells transfected with TDP43(D247E) exhibited a significantly greater risk of death compared to EGFP alone (HR 2.76, $p < 2x10^{-16}$). To determine if either the R151D or D247R mutations could account for the loss of toxicity upon TDP43(R151D-D247R) overexpression, we created TDP43 variants carrying each mutation singly. Neurons

expressing TDP43(R151D) showed no difference in survival compared to those transfected with EGFP alone, while overexpression of TDP43(D247R) significantly increased the risk of death in comparison to EGFP (HR 2.2, $p < 2x10^{-16}$). Thus, any changes to R151 completely prevented TDP43-mediated toxicity upon overexpression, indicating that this residue is critical for neurodegeneration upon TDP43 accumulation.

If R151 is indeed required for TDP43-related toxicity, then R151 mutations should outweigh other factors that govern TDP43 toxicity, including protein half-life. To test this, we compared the half-life of each TDP43 variant (as measured by OPL) with the risk of death in primary neurons overexpressing each variant. Among the R151 variants, TDP43(R151A) displayed the shortest half-life, followed by TDP43(R151D), and TDP43(R151D-D247R) (Fig. S3.6A). Despite their distinct half-lives, all R151 variants exhibited identical and minimal toxicity when overexpressed. Conversely, the toxicity of D247 variants varied proportionally with their measured half-life, with the least stable variant (D247A) demonstrating minimal toxicity, and the most stable variant (D247E) exhibiting the greatest toxicity (Fig. S3.6B). These results not only confirm the critical nature of the R151 residue for TDP43-related toxicity, but also indicate that half-life is an important predictor of neurodegeneration when this residue is intact.

3.3.6 Mutations affecting the RRM1-RRM2 salt bridge alter subcellular TDP43 localization

Given the prominent effects of salt bridge mutations on TDP43-mediated toxicity, the apparent relationship between cytoplasmic TDP43 mislocalization and neurodegeneration[13,21], and the physical proximity of the salt bridge to the putative TDP43 nuclear export signal (NES)[63,64], we asked whether the RRM1-RRM2 salt bridge

regulates subcellular TDP43 localization. Initial, qualitative inspection of transfected neurons demonstrated cytoplasmic TDP43 mislocalization in association with all mutations, with the exception of D247E and the F147L/F149L double mutant (Fig. S3.7A). We noted the formation of nuclear and cytoplasmic puncta in cells expressing TDP43(R151D), TDP43(R151D-D247R), and TDP43(D247A), as well as large, spherical nuclear puncta in neurons transfected with TDP43(F147L, F149L). Mirroring this phenomenon, select RNA binding-deficient TDP43 variants were enriched within the detergent-insoluble fraction of transfected HEK293T cells (Fig. S3.8). This insolubility suggests a phase transition similar to that observed for RNA binding-deficient MATR3 variants in neurons (Malik *et al.*, in press) and myoblasts[54].

To gauge TDP43 localization quantitatively and on the single-cell level, we measured the abundance of TDP43 separately within the nuclear and cytoplasmic compartments of transfected primary neurons, and calculated a nuclear-cytoplasmic ratio reflecting TDP43 distribution in individual cells (Fig. S3.7B, C). Consistent with our initial observations of TDP43 localization, we observed significant reductions in the nuclear-cytoplasmic ratio for all mutations affecting the RRM1-RRM2 salt bridge except D247E. Furthermore, the nuclear-cytoplasmic ratio was slightly but significantly higher in cells expressing TDP43(D247E) and TDP43(F147L, F149L). Since many variants were unable to recognize endogenous TDP43 targets by RNA IP (Figs. 3.1, 3.4) and displayed reduced toxicity in comparison to TDP43(WT) (Figs. 3.2, 3.4), these findings suggest that (a) nuclear TDP43 localization is independent of RNA binding *per se*, and (b) toxicity arising from cytoplasmic TDP43 mislocalization requires RNA binding.

3.3.7 A residue participating in the RRM1-2 salt bridge is required for native TDP43 function

All previous experiments were performed on a background of endogenous TDP43 expression, making it difficult to ascertain the functional significance of mutations affecting the RRM1-RRM2 salt bridge. To circumvent this limitation, we introduced the equivalent of the R151A mutation in the *C. elegans* TDP43 orthologue (*tdp-1*) by CRISPR/Cas9 genome engineering[65]. The RRM1 and RRM2 domains of TDP43 are highly conserved across species[10] (Fig. 3.5A), such that the R219A mutation in *C. elegans* TDP-1 is homologous to R151A in human TDP43. To target the *tdp-1* locus, a ribonucleoprotein (RNP) mixture containing a sgRNA specific for the *tdp-1* locus, recombinant *S. pyogenes* Cas9, and a single-stranded homology-directed repair template encoding the R219A mutation, was microinjected into the gonads of adult *C. elegans* hermaphrodites. Heterozygous F1 knock-in mutants were identified by PCR amplification and restriction enzyme digestion, and correct editing was confirmed by sequencing the entire *tdp-1* genomic open reading frame. Homozygous *tdp-1* (R219A) knock-in animals were subsequently identified in the F2 generation and heritable lines were established.

Because the R151A mutation in human TDP43 interferes with RNA binding (Fig. 3.1) and dramatically destabilizes the protein (Fig. 3.3), we suspected that the R219A mutation in *C. elegans* TDP-1 would functionally mirror the subtle behavioral and longevity phenotypes observed in TDP-1 knockout animals[66,67]. Supporting this notion, we observed a significant decrease in body length for *tdp-1* (R219A) worms compared to wild type N2 control worms; notably, these mutant worms were similar in length to *tdp-1*

(ok803) knockout animals (Fig. 3.5B, C). Locomotor metrics (track length, wavelength, amplitude) were similarly reduced in *tdp-1* (R219A) and *tdp-1 (ok803)* knockout worms compared to wild-type animals (Fig. 3.5D-F), while body bending in liquid media was reciprocally increased in *tdp-1* (R219A) and *tdp-1 (ok803)* knockout worms (Fig. 3.5G). Consistent with previous studies[67,68], we detected a modest but significant increase in the lifespan of *tdp-1 (ok803)* knockout animals in comparison to WT worms (Fig. 3.5H). *tdp-1* (R219A) animals also displayed an analogous lifespan extension, suggesting that the *tdp-1* (R219A) mutation closely phenocopies the TDP-1 knockout in *C. elegans*. Taken together, these data provide strong evidence that the R219 residue, homologous to R151A in human TDP43, is required for native TDP-1 function in *C. elegans*.

3.3.8 Overexpression of wild-type TDP43, but not salt bridge mutants, primarily affects ribosomal and mitochondrial transcripts

Given the consistent toxicity associated with WT TDP43, but not variants of TDP43 carrying salt bridge-disrupting mutations, we reasoned that transcriptomic comparison of cells expressing each of the variants would highlight the RNAs most closely associated with toxicity. We therefore overexpressed EGFP-tagged versions of TDP43(WT), TDP43(R151A), and TDP43(D247A) in HEK293T cells, purified total RNA, and performed high-throughput RNA sequencing to assess differences in transcript abundance and splicing associated with each variant. HEK293T cells transfected with EGFP alone and TDP43(F147L, F149L) served as negative controls for these experiments (Fig. 3.6A).

From an average of >26M reads, we mapped 17,620 unique transcripts in each of the conditions. In comparison to EGFP alone, we observed 1,026 differentially expressed

genes (DEGs) in cells overexpressing TDP43(WT) that were not detected upon overexpression of the other TDP43 variants (Fig. 3.6B). Conversely, only 21 and 24 DEGs were identified in cells transfected with TDP43(R151A) or TDP43(F147L, F149L), respectively, and an intermediate number (163) were noted in TDP43(D247A)-expressing cells (Fig. 3.6B). All TDP43 variants were overexpressed at equivalent levels, as determined by *TARDBP* transcript levels (Fig. 3.6C) and EGFP fluorescence (data not shown). Consistent with our previous data showing recognition of *MALAT1* transcripts by TDP43(WT) but not TDP43(R151A), TDP43(D247A) or TDP43(F147L, F149L) (Fig. 3.1), *MALAT1* was significantly downregulated only in TDP43(WT)-expressing cells (Fig. 3.6D).

To determine if transcripts preferentially affected by TDP43(WT) were concentrated within defined functional pathways, we classified transcripts according to gene ontology[69] (GO; Fig. 3.6E). In doing so, we noted profound enrichment for transcripts involved in the ribosomal and oxidative phosphorylation pathways (false discovery rate (FDR), 6.8×10^{-5} and 0.03, respectively; Fig. 3.6F, G). Based on the results of previous studies implicating RNA transport in ALS/FTD pathogenesis[70-73], we also focused on this pathway, noting a non-significant trend (FDR, 0.222) towards enrichment for transcripts that function in RNA transport (Fig. 3.6H). The effects of TDP43 variants on gene expression followed the same pattern we observed for toxicity in primary neurons (Fig. 3.2): in each case, the magnitude of change was greatest for TDP43(WT), intermediate for TDP43(D247A), and minimal for TDP43(R151A) and TDP43(F147L, F149L). Taken together, these data demonstrate a proportional

relationship between the degree of toxicity and differential gene expression in the ribosomal and oxidative phosphorylation pathways.

One of TDP43's primary functions is the suppression of unannotated (cryptic) exons during pre-mRNA splicing[32,74-76]. We therefore investigated differences in splicing generated by overexpression of EGFP-tagged versions of TDP43(WT), TDP43(R151A), TDP43(D247A) and TDP43(F147L, F149L). We calculated a percent splicing index (PSI) for all events[77], in the process identifying 9,339 unique splicing events that demonstrated a significant (FDR < 0.05) fold change of > 1.5 in cells overexpressing TDP43(WT)-EGFP compared to EGFP alone (Fig. 3.7A). These unique splicing events were not shared among cells expressing the other TDP43 variants. We then determined the PSI for each of these events upon expression of EGFP-tagged TDP43(F147L, F149L), TDP43(R151A), or TDP43(D247A) (Fig. 3.7B-D). Surprisingly, only 3,968 splicing events (42.5%) are unique to TDP43(WT), which alone among the tested variants demonstrates appreciable binding to native substrates (Fig. 3.1). These results therefore suggest that most splicing events attributable to TDP43 overexpression occur independently of direct RNA binding by TDP43, and have few consequences for cellular survival.

To isolate splicing changes tied specifically to TDP43-mediated toxicity, we concentrated our analysis on the splicing events detected selectively in cells expressing TDP43(WT) (Fig. 3.7E). By GO analysis, the ribosome pathway was significantly overrepresented among these transcripts (FDR 1.4×10^{-13}), consistent with the gene expression data for TDP43(WT) (Fig. 3.6). Transcripts selectively affected by TDP43(WT) were also highly enriched for the spliceosome, proteasome, and RNA

transport pathways (Table S3.1). To visualize genes most affected by TDP43(WT)-induced splicing abnormalities, we plotted the 100 splicing events with the largest magnitude change, including the 50 top repressed and included events (Fig. 3.7F). We noted several genes within this list that function within nucleocytoplasmic transport (KPNA2, TNPO1, NXF1) and ribosomal (RPS16, RPS20) pathways, in keeping with the results of GO analysis. In many cases, abnormal splicing was most significant for TDP43(WT), intermediate for TDP43(D247A), and minimal for TDP43(R151A) and TDP43(F147L, F149L) expression. Thus, as with transcriptional changes noted upon expression of TDP43 variants (Fig. 3.6), the phenotype observed for TDP43(WT) was most profound, followed by TDP43(D247A) and lastly TDP43(R151A) and TDP43(F147L, F149L).

We also examined the types of splicing abnormalities unique to TDP43(WT) overexpression (Fig. 3.7G). In agreement with the purported function of TDP43 as a splicing repressor[30,32,74], the largest single category of splicing changes elicited by TDP43(WT) overexpression featured abnormal intron retention (36%). The second most common change involved exon skipping—29% of changes demonstrated a single skipped exon, and another 5% exhibited skipping of 2 consecutive exons. The remaining 30% of changes featured alternative cassette splicing, including new 3' or 5' splice sites (13% and 10%, respectively), or alternative start sites, last exons or first exons (<3% each). Additionally, we noted several unannotated splicing events or junctions among those influenced selectively by TDP43(WT) (Fig. 3.7H, I), supporting a role for TDP43 in preventing cryptic exon inclusion[74-76].

3.4 Discussion

We took advantage of the intricate relationship between protein structure and function to probe the RNA binding properties of TDP43, and how these characteristics dictate its stability and propensity for toxicity in neuronal models of ALS and FTD. Engineered mutations that disrupt an intramolecular salt bridge between TDP43's RRMs reduce the affinity and sequence-specificity of RNA and DNA binding by TDP43. As a result of these mutations, TDP43 is incapable of recognizing its native substrates or stimulating neurodegeneration upon overexpression in neurons. We also leveraged the intrinsic contrast between TDP43(WT) and versions of TDP43 lacking an intact RRM1-RRM2 salt bridge to identify the direct transcriptomic and splicing changes most significantly associated with TDP43-dependent toxicity, thereby implicating dysfunction within the ribosomal and oxidative phosphorylation pathways in ALS and FTD characterized by TDP43 deposition.

The calculated Hill coefficient for TDP43(WT) ($h = 1.44 \pm 0.14$ for RNA, $h = 1.973 \pm 0.291$ for DNA) suggests that nucleic acid binding by TDP43 is highly cooperative, as could be seen with multiple allosteric binding sites or protein dimerization. The latter possibility is consistent with prior studies demonstrating TDP43 dimerization *in vitro* and *in vivo*[11,78-80]. Both the R151A or D247A mutations reduce the Hill coefficient for DNA binding, but subtly increase the Hill coefficient for RNA binding. Additionally, the R151A mutation reduced the affinity of TDP43 for RNA but not DNA substrates. These results imply that TDP43 binds RNA and DNA substrates through distinct mechanisms, mediated in part by the R151 and D247 residues. Notably, the D247A mutation within RRM2 affected both the affinity and cooperativity of

binding, suggesting that RRM2 plays a more substantial role in nucleic acid binding than previously appreciated[9,10,12]. Consistent with this observation, artificial mutations in the RRM2 domain of endogenous mouse TDP43 effectively impair RNA binding and result in a loss of native TDP43 function[19].

Unlike TDP43(WT), both TDP43(R151A) and TDP43(D247A) are capable of binding sequences other than $(UG)_{12}/(TG)_{12}$ *in vitro*, implying that the RRM1-RRM2 salt bridge is necessary for maintaining sequence specificity, as previously suggested[12,27]. These results are also consistent with a model whereby RRM1 provides a high-affinity nucleic acid binding surface, while RRM2 tunes the sequence specificity of binding. In this scenario, the R151-D247 salt bridge linking RRM1 and RRM2 is required for both affinity and specificity exhibited by full-length TDP43. Whether stochastic or regulated interruption of salt bridge *in vivo* may enable TDP43 to bind non-canonical sequences is unknown, and requires further investigation.

Consistent with an integral role for RNA binding in TDP43-mediated toxicity[13,15,16], the protective effect of the R151A and D247A mutations was comparable to that of mutations that eliminate RNA binding altogether (F147L, F149L), or deletion of RRM1 itself. Notably, the R151A mutation phenocopied the F147L/F149L double mutant more closely, while the D247A mutation elicited an intermediate phenotype. The D247A mutation leaves RRM1 fully intact, and since the majority of high affinity RNA binding is accomplished by RRM1[9,10,12], TDP43(D247A) may still recognize many conventional TDP43 targets. Supporting this notion, the number of DEGs and alternatively spliced transcripts in cells overexpressing TDP43(D247A) were more similar to TDP43(WT) than either TDP43(R151A) or TDP43(F147L, F149L) (Fig. 3.7).

In light of the dual effects of each of these engineered mutations on TDP43 RNA binding and stability, we pursued several approaches to determine which of these outcomes was most closely tied to TDP43-dependent toxicity. First, we substituted the RNA binding domains of PUM2 or PP7 for those of TDP43, and assessed neurodegeneration upon overexpression. Chimeric proteins incorporating the PUM2 RNA binding domains completely recapitulated TDP43-mediated toxicity, while those containing the PP7 RNA binding domain partially restored toxicity. Second, inactive versions of the PUM2 RNA binding domain were incapable of mirroring toxicity when fused to TDP43 and overexpressed in neurons. Third, attaching the CL1 degron to TDP43(WT) effectively destabilized the protein and significantly reduced its toxicity when overexpressed. These data indicate that both RNA binding and protein stability are important factors underlying TDP43-dependent toxicity.

Supporting this notion, we observed a distinct relation between the stability of D247 variants and TDP43-related toxicity in primary neurons, but a similar correlation was not observed for R151 variants. One possibility is that RNA recognition via RRM1, which requires the R151 residue[27], is crucial for downstream toxicity mediated by TDP43. In the setting of an intact RRM1, alterations to protein stability can further modulate the severity of neuron loss in TDP43-overexpressing cells. Since the half-life of TDP43 within neurons is substantially longer than in other cell types[23-25], the proportional relationship between TDP43 stability and toxicity may contribute to the enhanced susceptibility of neurons to TDP43-mediated cell death. Additional studies are warranted to determine if specific neuron subtypes display different TDP43 clearance rates, and

whether intrinsic differences in TDP43 turnover reflect selective neuronal vulnerability in ALS and FTD.

We observed substantial mislocalization of TDP43 variants carrying artificial mutations that interfered with RNA binding and the integrity of the RRM1-RRM2 salt bridge. These results suggest that the nucleocytoplasmic transport of TDP43 is closely linked to its ability to bind RNA[81]. Alternatively, mutations that disrupt the RRM1-RRM2 salt bridge may partially unfold the protein, making the putative NES within RRM2 more accessible. Arguing against this possibility, we and others were unable to detect TDP43 nuclear export driven by this putative NES or its predicted chaperone, exportin-1[81-83]. Despite the previously noted toxicity of cytoplasmic TDP43[13,21], all of the mislocalized TDP43 variants were less toxic than TDP43(WT), which maintained a primarily nuclear localization. Therefore, without the ability to recognize RNA, even cytoplasmic TDP43 is no longer toxic. A similar discrepancy was noted upon knockdown of the debranching enzyme DBR1; in this case, *DBR1* knockdown increased the abundance of unprocessed intron lariats that were recognized by TDP43, effectively sequestering or inactivating cytoplasmic TDP43[84].

Overexpressing TDP43(WT) recapitulates key aspects of sporadic ALS/FTD[15,21,23,36-39], and disease-associated *TARDBP* mutations elicit gain-of-function toxicity in animal models[18,19] that is mirrored by TDP43 overexpression. While useful for investigating the mechanisms of TDP43-dependent neurodegeneration, overexpression models complicate the study of loss-of-function mutations such as those affecting the TDP43 RRM1-RRM2 salt bridge. We therefore created *C. elegans* expressing endogenous TDP-1(R219A), a variant homologous to human TDP43(R151A). These

animals were almost identical to *tdp-1* knockout animals in morphology, behavior, and lifespan, indicating that the R219A mutation completely inactivates TDP-1. Unlike other TDP43 orthologs in *Drosophila* and mice[85-88], *C. elegans* TDP-1 is not required for survival, but it functions similarly to mammalian TDP43 in many aspects, including its ability to recognize UG-rich sequences with high affinity[10]. Our data therefore provide strong evidence that disrupting R151 and the RRM1-2 salt bridge is sufficient to eliminate TDP43 function.

TDP43 is a splicing repressor[30,32], and its depletion promotes the inclusion of cryptic exons within target transcripts and their subsequent destabilization by nonsense mediated RNA decay[74-76]. Conversely, TDP43 overexpression elicits exon skipping in most cases[19,89,90]. In accord with these findings, intron retention and exon skipping were among the most common splicing changes associated with TDP43 overexpression. Furthermore, we detected a profound enrichment of abnormal splicing events within ribosomal protein encoding transcripts selectively in cells overexpressing TDP43(WT), suggesting that functional TDP43 directly affects the splicing of these transcripts. Like many RNA binding proteins, including TDP43, ribosomal proteins bind their own transcripts and regulate their abundance by modulating mRNA splicing[91,92]. TDP43 may regulate ribosomal protein-encoding transcripts by repressing their splicing, much like ribosomal proteins themselves. We also detected a significant overrepresentation of the RNA transport and proteolysis pathways among differentially spliced transcripts in cells overexpressing TDP43(WT), but not TDP43 variants unable to bind RNA. These results reflect global deficits in RNA transport and proteolysis observed in ALS and FTD models[46,70-72,93], testifying to the conserved nature of these abnormalities and indicating a

potential niche for TDP43 in maintaining nucleocytoplasmic trafficking and protein homeostasis.

TDP43 recognizes thousands of RNAs bearing UG-rich sequences[30,32], complicating the identification of transcripts whose processing underlies TDP43-dependent neurodegeneration. To address this issue, we identified transcripts displaying significant changes in abundance or splicing in cells overexpressing TDP43(WT), but not in cells transfected with RNA binding-deficient TDP43 variants. In doing so, we noted that mRNAs encoding components of the ribosomal and oxidative phosphorylation pathways were highly enriched among downregulated transcripts selectively in TDP43(WT)-expressing cells. This pattern of enrichment is strikingly similar to that observed in human induced pluripotent stem cells (iPSCs) from individuals with pathogenic *C9orf72* mutations, one of the most common mutations responsible for ALS and FTD[3,4], and in iPSCs overexpressing TDP43 (**Tank and Figueroa-Romero *et al.*, in press). These observations imply that dysregulation of ribosomal and oxidative phosphorylation pathways are convergent, downstream events in ALS and FTD pathogenesis, inextricably linked with TDP43 deposition and RNA binding. Therapies that act on these pathways, therefore, are likely to be particularly effective in extending neuronal survival and preventing neurodegeneration in ALS and FTD.

3.5 Materials and methods

Plasmids

The plasmids pGW1-TDP43(WT)-EGFP, pGW1-TDP43(WT)-Dendra2, pGW1-TDP43(M337V)-EGFP, pGW1-EGFP, and pGW1-mApple were created as described

previously[21,23]. Primers used for site-directed mutagenesis (Table S3.2) or PCR amplification (Table S3.3) were ordered from Integrated DNA Technologies (IDT). TDP43 variants were generated from TDP43(WT)-EGFP, TDP43(M337V)-EGFP, or TDP43(WT)-Dendra2 by site-directed mutagenesis using the Pfu Ultra high-fidelity polymerase (Agilent Technologies) according to the manufacturer's protocol (Table S3.2).

To create TDP43(PUM2)-EGFP, the PUM2 RNA binding domain (courtesy of Dr. Aaron Goldstrohm) was excised from pFN21A-PUM2-R6SYE (a PUM2 variant in which the RNA recognition amino acids of the 6th repeat are mutated to SYE[50]) using XmaI and AgeI, and ligated into pGW1-EGFP cut with the same enzymes, generating pGW1-PUM2-EGFP. We then amplified amino acids 1-106 of TDP43(WT) from pGW1-TDP43(WT)-EGFP using PrimeStar GXL DNA polymerase (Takara) following the manufacturer's protocol (Table S3.3). The 318 bp product was digested with BstBI and XmaI restriction enzymes and ligated into pGW1-PUM2-EGFP that had been cut with the same enzymes. Next, amino acids 263-414 of TDP43(WT) from pGW1-TDP43(WT)-EGFP were amplified by PCR (Table S3.3). The resulting 453 bp fragment was digested with AgeI and ligated into pGW1-TDP43(AA1-106)-PUM2-EGFP that had also been cut with AgeI, generating pGW1-TDP43(PUM2)-EGFP. To create an RNA binding-deficient version of this construct (pGW1-TDP43(mPUM2)-EGFP), 3 critical residues within helix 7 of the PUM2 RNA binding domain[52] were mutated by site-directed mutagenesis (Table S3.2).

To make pGW1-TDP43(PP7)-EGFP, the TDP43(PUM2)-EGFP plasmid described above was cut with XmaI and AgeI to remove both PUM2 (1044 bp fragment)

and the C-terminal end (AA263-414) of TDP43. A 850 bp fragment encoding the RNA binding domain of PP7 (plasmid pHR-PP7-3xmCherry[94], courtesy of Dr. Nils Walter) was amplified by PCR (Table S3.3) and ligated into TDP43(AA1-106)-EGFP-pGW1 cut with XmaI and AgeI. The vector was then digested with AgeI to re-insert the C-terminal end of TDP43 to generate the final product, pGW1-TDP43(PP7)-EGFP.

For pGW1-TDP43(ΔRRM1-2)-EGFP, the pGW1-TDP43(AA1-106)-PUM2-EGFP plasmid described above was digested with XmaI and AgeI to remove the PUM2 fragment. The C-terminal end of TDP43 (AA263-414) was then amplified by PCR using the primers described above and inserted into the cut vector. To generate TDP43(ΔRRM1)-EGFP or TDP43(ΔRRM2)-EGFP, we first amplified the RRM2 and RRM1 domains from pGW1-TDP43-EGFP by PCR (Table S3.3). The TDP43(ΔRRM1-2)-EGFP construct was then cut with XmaI followed by insertion of the PCR products encoding the RRM2 or RRM1 domains, respectively, cut with the same enzyme.

To generate pGW1-TDP43(WT)-Dendra2-CL1, TDP43(WT) was PCR amplified from pGW1-TDP43(WT)-EGFP (Table S3.3) and ligated into pGW1-empty vector using NheI and AgeI restriction sites. The following Dendra2-CL1 sequence with flanking AgeI and SalI restriction sites was ordered as a G-block from IDT (the CL1 sequence is underlined): 5' – ACC GGT TGC CAC CAT GAA CAC CCC GGG AAT TAA CCT GAT CAA GGA GGA CAT GCG CGT GAA GGT GCA CAT GGA GGG CAA CGT GAA CGG CCA CGC CTT CGT GAT CGA GGG CGA GGG CAA GGG CAA GCC CTA CGA GGG CAC CCA GAC CGC CAA CCT GAC CGT GAA GGA GGG CGC CCC CCT GCC CTT CAG CTA CGA CAT CCT GAC CAC CGC CGT GCA CTA CGG CAA CCG GGT GTT CAC CAA GTA CCC CGA GGA CAT CCC CGA CTA

CTT CAA GCA GAG CTT CCC CGA GGG CTA CAG CTG GGA GCG CAC CAT GAC CTT CGA GGA CAA GGG CAT CTG CAC CAT CCG CAG CGA CAT CAG CCT GGA GGG CGA CTG CTT CTT CCA GAA CGT GCG CTT CAA GGG CAC CAA CTT CCC CCC CAA CGG CCC CGT GAT GCA GAA GAA GAC CCT GAA GTG GGA GCC CAG CAC CGA GAA GCT GCA CGT GCG CGA CGG CCT GCT GGT GGG CAA CAT CAA CAT GGC CCT GCT GCT GGA GGG CGG CGG CCA CTA CCT GTG CGA CTT CAA GAC CAC CTA CAA GGC CAA GAA GGT GGT GCA GCT GCC CGA CGC CCA CTT CGT GGA CCA CCG CAT CGA GAT CCT GGG CAA CGA CAG CGA CTA CAA CAA GGT GAA GCT GTA CGA GCA CGC CGT GGC CCG CTA CAG CCC CCT GCC CAG CCA GGT GTG GGC ATG CAA GAA CTG GTT CTC TAG CCT GTC CCA CTT CGT CAT CCA TCT GTA AGT CGA C– 3'. The Dendra2-CL1 fragment was digested with AgeI and SalI and inserted into pGW1-TDP43(WT) cut with the same enzymes to make TDP43(WT)-Dendra2-CL1.

To create pGW1-TDP43(WT)-PEST-Dendra2, the following PEST sequence with flanking AgeI restriction sites was ordered as a G-Block from IDT: 5' – ACC GGT TGC CAC CAT GCA TGG CTT CCC TCC AGA GGT GGA GGA GCA AGA TGA TGG CAC TCT CCC CAT GAG CTG CGC TCA AGA GAG TGG CAT GGA TCG TCA CCC CGC TGC TTG CGC CTC GGC TCG CAT CAA CGT GAA ACC GGT – 3'. The PEST sequence was digested with AgeI and inserted into TDP43(WT)-Dendra2 cut with the same enzyme to generate TDP43(WT)-PEST-Dendra2.

Purification of recombinant TDP43

TDP43(WT), TDP43(R151A) or TDP43(D247A) was expressed in BL21 DE3 *E. coli* cells from the plasmid pE-6xHis-SUMO-TDP43 (a gift from Dr. James Shorter).

Induction was carried out with 1 mM isopropyl-β-D-1-thiogalactopyranoside and cells were grown at 15°C for 16 hr. Cell pellets were resuspended in lysis buffer (50 mM HEPES, 2% TritonX-100, 300 mM NaCl, 5% glycerol, 50 mM imidazole, 2 mM BME, EDTA-free protease inhibitor cocktail, 5 μM pepstatin, and 20 mg/mL lysozyme) and incubated on ice for 30 minutes. Following sonication on ice, cell lysates were centrifuged for 20 min at 11,000 x g at 4°C. Recombinant protein was purified by binding to Ni-NTA resin (Qiagen), rinsed with 25 mL of wash buffer 4 times (50 mM HEPES, 2 % TritonX-100, 300 mM NaCl, 5% glycerol, 50 mM imidazole, and 2 mM BME), and released with 2 mL of elution buffer (50 mM HEPES, 500 mM NaCl, 300 mM imidazole, 5% glycerol, and 5 mM DTT) at room temperature (RT), collecting five 2 mL fractions. Protein was dialyzed twice for 1 hr in 1 L of final buffer (50 mM HEPES and 500 mM NaCl), and dialyzed once in 1 L of final buffer overnight at 4°C.

Electromobility shift assays

Binding assays were performed with purified full-length recombinant TDP43 protein and either ssRNA labeled probes tagged with a 5' 800nm infrared (IR) moiety or ssDNA labeled probes tagged with a 5' 700nm IR moiety (IDT). Binding reactions were performed in binding buffer (12.5 mM HEPES, pH 7.8, 50 mM KCl, 2.5 mM MgCl2, 0.5 mM TCEP, 25 μg/mL BSA, 0.01% NP-40) with 50% glycerol, 1μg/μl poly-dIdC, 100 pM of labeled probe, and recombinant protein (concentrations indicated in figure legends). Reactions were incubated on ice for 5 min followed by 25 min at RT. Electrophoresis of 6% acrylamide gels were performed at 100 V. For competition assays, labeled ssRNA or ssDNA probes were incubated prior to the addition of unlabeled ssRNA or ssDNA oligos, respectively, for another 30 min at RT. Images were acquired

using the LI-COR Odyssey platform. Intensities for each of the IR-labeled probe/protein complexes were determined using Image Studio 2.0 and plotted in GraphPad Prism. The dissociation constant (K_d) and hill coefficient (h) were determined in GraphPad Prism using the nonlinear least squares regression fit equation.

RNA immunoprecipitation

HEK293T cells were transfected with WT and pGW1-TDP43-EGFP variants using Lipofectamine 2000 (ThermoFisher) following the manufacturer's protocol. Forty-eight hours post-transfection, cell pellets were resuspended in lysis buffer (50 mM HEPES, pH 7.5, 50 mM KCl, 0.5% sodium deoxycholate, 0.1% SDS, 0.5% NP-40, 0.5 mM DTT, and EDTA-free protease inhibitor cocktail) and incubated on ice for 30 min. Following sonication on ice, cell lysates were centrifuged at 13,000 x g for 10 min at 4°C. 400 µg protein was added to Dynabeads Protein G magnetic beads (Invitrogen) and 1 µg/µl of anti-GFP (Rabbit, Immunology Consultants Laboratory) and incubated overnight at 4°C. Protein/antibody/bead complexes were washed with 50 mM HEPES pH 7.5, 300 mM KCl, 0.5% NP-40, 0.5 mM DTT, and EDTA-free protease inhibitor cocktail.

RT-PCR and quantitative RT-PCR

Total RNA was extracted using RNeasy Mini Kit (Qiagen) following the manufacturer's protocol. To synthesize cDNA, 1 µg of total RNA was used in a 20 µl reaction volume with the Bio-Rad iScript cDNA synthesis kit according the manufacturer's protocol. The reactions were incubated at 25°C for 5 min, 42°C for 30 min, and 85°C for 5 min. PCR was carried out in a 25 µl reaction volume with 0.5 µl of cDNA, 200 nM primers, and GoTaq Master Mix (Promega). For quantitative RT-PCR

(qRT-PCR), reactions were carried out using Step One Plus Realtime PCR system (Applied Biosystems). Reactions were carried out using SYBR Green Master Mix (Applied Biosystems), with 200 nM primers, and 0.5 µl cDNA, according to the following parameters: 4 min at 95°C, then 30 cycles at 95°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec. Relative gene expression was calculated using the ΔΔCt method. Values obtained from TDP43 variants were scaled to those from TDP43(WT) and plotted in GraphPad Prism. The following primers were used for qRT-PCR amplification: *TARDBP*, 5'- GTG GCT CTA ATT CTG GTG CAG -3' and 5'- CAC AAC CCC ACT GTC TAC ATT -3', and *MALAT1,* 5'- GAC GGA GGTT GAG ATG AAG C -3' and 5'- ATT CGG GGC TCT GTA GTC CT -3'.

Differential solubility

HEK293T cells were transfected with WT and pGW1-TDP43-EGFP variants using Lipofectamine 2000 (Thermofisher) according to the manufacturer's protocol. Cells were washed and collected in cold PBS 48 hr after transfection. Cells were centrifuged at 7,000 x g for 5 min at 4°C and resuspended in RIPA buffer with protease inhibitors. Following lysis on ice for 15 min, cells were centrifuged at 21,000 x g for 15 min at 4°C. The supernatant was removed and saved as the RIPA-soluble fraction. The RIPA-insoluble pellet was then washed in 1 mL RIPA, and contents were resuspended in urea buffer (7 M urea, 2 M thiourea, 4% CHAPS, 30 mM Tris, pH 8.5). Samples were centrifuged at 21,000 x g for 15 min at 4°C, and the supernatant was saved as the urea-soluble fraction.

For SDS-PAGE, stock sample buffer (10% SDS, 20% glycerol, 0.0025% bromophenol blue, 100 mM EDTA, 1 M DTT, 20 mM Tris, pH 8) was diluted in all

samples. For RIPA-soluble fractions, 10 ug of sample were boiled for 10 min. For urea-soluble fractions, equal volumes of sample for each condition were mixed 1:1 with water, and these samples were not boiled. Samples were then loaded onto 4-15% gradient gels (Bio-Rad). After transfer to 0.2 μm PVDF membrane (Bio-Rad), blots were incubated in 3% BSA in 0.2% Tween-20 with the following antibodies: mouse anti-GAPDH (Millipore, MAB374; dilution 1:1000) and rabbit anti-TDP43 (Cell Signaling; dilution 1:1000). Blots were then washed in 0.2% TBST and incubated at RT for 1 hr with AlexaFluor goat anti-mouse 594 (ThermoFisher) and goat anti-rabbit 488 (ThermoFisher), both diluted 1:10,000 in 3% milk in 0.2% TBST. Blots were washed in 0.2% TBST before imaging with the LI-COR Odyssey platform.

Primary neuron transfections

Primary mixed cortical neurons were dissected from embryonic day 19-20 Long-Evans rat pups and cultured at 0.6 x 10^6 cells/mL in 96 well cell culture plates (TPP), as previously described[23,41]. At *in vitro* day (DIV) 3-4, neurons were transfected with 0.2 μg DNA and 0.5 μL Lipofectamine 2000 (ThermoFisher) per well, per the manufacturer's protocol, with the exception that cells were incubated with Lipofectamine/DNA complexes for only 20 min at 37°C before rinsing. For live-cell nuclear staining, cells were rinsed with media containing Hoechst 33258 Dye (Invitrogen) at 1:5000. Following transfection, cells were maintained in NEUMO photostable medium containing SOS supplement (Cell Guidance Systems). For experiments that involved MG132 (Millipore), drug was added to a final concentration of either 25 nM or 500 nM 18 hr post-transfection; cells were imaged 6 h following drug addition.

Longitudinal fluorescence microscopy and image analysis

Automated longitudinal fluorescence microscopy began 24 hr post-transfection for 10 days, as previously described[21,40,41,45]. Briefly, images were acquired by an inverted Nikon Ti microscope equipped with a 20x objective lens, a PerfectFocus system, a Lambda XL Xenon lamp (Sutter) with 5 mm liquid light guide (Sutter), and either an Andor iXon3 897 EMCCD camera or Andor Zyla4.2 (+) sCMOS camera. All stage, shutter, and filter wheel movements were carried out by custom code written in publicly available software (μManager, ImageJ).

Image processing, survival analyses, and fluorescent intensity measurements were accomplished using scripts written in Python or the ImageJ macro language. Neurons were identified by morphology, size, and fluorescence intensity. Cell death was determined for each neuron by rounding of the soma, degenerating processes, and loss of fluorescence.

***C. elegans* tdp-1(R219A) CRISPR/Cas9 mutant generation**

For all experiments, the Bristol N2 strain was used as the wild type control and was the parental strain for all transgenic lines created in this study. N2 animals were modified by CRISPR/Cas9 gene editing using direct injection of CRISPR/Cas9 ribonucleoproteins as previously described[65]. In brief, single-stranded oligonucleotide homology directed repair templates (ssODN-HDR) containing 40-50 bp 5' and 3' homology arms flanking the 20-mer sgRNA target site were designed containing the edit of interest, a unique in-frame restriction site (SacI), and conservative nucleotide changes to prevent sgRNA:Cas9 cleavage. The ssODN-HDR templates were ordered as Ultramers from Integrated DNA Technologies (IDT), and recombinant *S. pyogenes* Cas9

nuclease was purchased from IDT as well. A synthetic single guide RNA (5'- UUU GUC AGA AUG UCA UCA GU-3') was purchased from Synthego, Inc. The R219A ssODN-HDR repair template contained the following sequence: 5'- TCT TCT TCA GAT AAA ACG AAA GTC GGA TGG AAA CTC AAA AGG ATT TGG ATT CGT TGC GAT GAG CTC TGT AGG TGA ACA AAA TAA AGT ATT AGC TAT CCC ACA GCA CAT GAT TGA TGG TCG -3'. Correctly edited animals were back crossed six times to N2 and the entire *tdp-1* genomic DNA open reading frame was analyzed by PCR and Sanger sequencing.

Behavioral and longevity assays in *C. elegans*

Worms were grown at 20°C on NGM plates seeded with OP50 *Escherichia coli*. For all behavioral assays, synchronized L4 larval animals were obtained using sodium hypochlorite. For all assays, the scorer was blind to the genotype. For locomotion assays, 10-15 seven-day adult worms were picked and transferred to 35mm NGM plates seeded with OP50. Worms were filmed for one-minute using custom built video-acquisition hardware, and locomotor metrics were analyzed using WormLab software (MBF Biosciences). For locomotion assays on NGM plates, at least 127 animals/genotype were quantified in 3 replicate trials. For thrash assays, 10-15 animals were placed in 1 mL of M9 media at room temperature. After a one-minute acclimation period, a one-minute video was recorded and analyzed using the thrash assay function in WormLab software (MBF Biosciences). At least 96 animals/genotype were quantified in 3 replicate trials. Longevity assays were conducted at 20°C to minimize temperature fluctuations and ensure constant conditions. For each strain at least 20 L4 larvae were transferred to NGM plates containing OP50 *E. coli* lawns. A total of at least 5 plates were

used per strain for a total of at least 100 worms per strain. Longevity assays were replicated in 3 independent trials. Animals were transferred every 2-3 days as needed to a fresh plate. The viability of the worms was assessed daily, and animals that failed to respond to stimulation by touch were considered dead. In respect to the timeline, day 0 reflected the day L4 larvae were transferred to NGM plates containing OP50 *E. coli* lawns.

RNA-sequencing, gene expression and splicing analysis

HEK293T cells were transfected with pGW1-EGFP, pGW1-TDP43(WT)-EGFP, pGW1-TDP43(R151A)-EGFP, pGW1-TDP43(D247A)-EGFP, or pGW1-TDP43(F147L, F149L)-EGFP using Lipofectamine 2000 (ThermoFisher), following the manufacturer's protocol. Total RNA was extracted using RNeasy Mini Kit (Qiagen) following the manufacturer's protocol. Purified RNA was submitted to the University of Michigan Sequencing Core for library generation (Illumina) and sequenced using the Illumina Hi-Seq 4000 platform (stranded mRNA, paired end 50 cycle, 2 lanes). FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) (version v0.11.3) was used for quality control. The Tuxedo Suite software package was used for alignment, differential expression analysis, and post-analysis diagnostics[95-97]. Briefly, reads were aligned to the reference mRNA transcriptome (hg19)[98] using TopHat (version 2.0.13) and Bowtie2 (version 2.2.1). Default parameter settings for alignment were used, with the exception of "b2-very-sensitive." To ensure high quality data was used for expression quantification and differential expression analysis, FastQC was used post-alignment. For expression quantitation, normalization, and differential expression analysis, Cufflinks/CuffDiff (version 2.2.1) was used, and hg19.fa was set as the reference genome sequence. The

following parameter settings was used: “multi-read-correct” to adjust expression calculations for reads that map in more than one locus, and “compatible-hits-norm” and “upper-quartile-norm” form normalization of expression value. Diagnostic plots were generated using the CummeRbund R package. Custom scripts were used to format and annotate the differential expression data from CuffDiff. Genes and transcripts that are differentially expressed were identified based on the following criteria: test status = “OK”, FDR ≤ 0.05, and fold change $\geq \pm 1.5$. To model the biological relevance of differentially expressed genes, Advaita Bio’s iPathwayGuide (http://www.advaitabio.com/ipathwayguide) was used, and EGFP was set as the control.

For analysis of splicing changes, trimmed reads were aligned to hg38 genome reference using HISAT2[99] with Gencode gene annotation (version 27). Discordant or mixed mapping between mates in a pair was disabled using the “--no-mixed --no-discordant” parameters. Properly mapped alignments with quality > 20 were selected and analyzed using the bioconductor packages SGSeq[77] and Limma[100]. Specifically, *analyzeFeatures()* was run to predict transcript features from the alignments; *annotate()* was run to annotate features with known transcripts from library “TxDb.Hsapiens.UCSC.hg38.knownGene”; *analyzeVariants()* was run to identify splice events, obtain counts for each splice variant and compute estimates of relative splice usage; *plotFeatures() and plotVariants()* were used to plot splice graphs and heatmaps of expression levels for selected genes; and *diffSplice()* from Limma was run to test differential splice variant usage for the four comparisons: TDP43(WT)-EGFP vs EGFP, TDP43(F147L, F149L)-EGFP vs EGFP, TDP43(R151A)-EGFP vs EGFP, and TDP43(D247A)-EGFP vs EGFP.

Statistical analyses

Statistical analyses were performed in either R or GraphPad Prism. For survival analyses, the publically available R survival package was used to determine differences among populations through Cox proportional hazards analysis. Half-life was determined for individual cells by using a custom-written R script that fits log-transformed TRITC intensities for each cell, normalized to the intensity measured immediately following photoconversion, to a linear equation[23]. Statistical differences among half-lives, nuclear/cytoplasmic ratios, and locomotor metrics were determined by one-way ANOVA with either Dunnett's multiple comparisons test or Tukey's post hoc test in GraphPad Prism. For qRT-PCR analyses, differences among groups were determined using the Kruskal-Wallis test.

3.6 Acknowledgements

We thank all members of the Barmada lab for helpful discussions and technical advice, Roberto Miguez for technical assistance; Drs. Robert Fuller, Henry Paulson, and Peter Todd for their suggestions and mentoring; Drs. Yuna Ayala and James Shorter for assistance in purifying recombinant protein; Dr. Jiou Wang for the kind donation of *C. elegans* strain *ok803*; and Dr. Aaron Haeusler for advice on EMSAs. We also thank Ahmed Malik for assistance in manuscript preparation. This work utilized core services provided by the Bioinformatics Core at the University of Michigan. We are indebted to Richard McEachin and Weisheng Wu for helpful discussions and technical support.

3.7 Author contributions

B.N.F and S.J.B designed the study, planned experiments and interpreted the data. BNF performed all transfections, survival analyses, half-life determinations, site-directed mutagenesis and cloning, EMSA experiments, and sample preparation for RNA-IP experiments and RNA-sequencing. X.L isolated primary neurons, assisted with the RNA-IP experiments, and performed qRT-PCR. J.L.M and A.A.B designed, performed and analyzed the *C. elegans* studies. B.N.F and S.J.B wrote and edited the manuscript with input from all authors.

Funding

This work was supported by National Institutes of Health (NIH), General Medical Sciences (GM) T-32-GM00731.5 (BNF), National Institute for Neurological Disorders and Stroke (NINDS) R01-NS097542 (SJB) and R01-NS094678 (AAB), National Institute for Aging (NIA) P30 AG053760 (SJB), the University of Michigan Protein Folding Disease Initiative (SJB), and the Fred A. and Barbara M. Erb Family Foundation (SJB).

References

1. Robberecht, W. & Philips, T. The changing scene of amyotrophic lateral sclerosis. *Nat. Rev. Neurosci.* **14,** 248–264 (2013).
2. Therrien, M., Dion, P. A. & Rouleau, G. A. ALS: Recent Developments from Genetics Studies. *Curr Neurol Neurosci Rep* **16,** 59 (2016).
3. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72,** 245–256 (2011).
4. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
5. Neumann, M. *et al.* Ubiquitinated TDP-43 in Frontotemporal Lobar Degenerationand Amyotrophic Lateral Sclerosis. *Science* **314,** 130–133 (2006).
6. Arai, T. *et al.* TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Biochem and Biophy Res Commun* **351,** 602–611 (2006).
7. Ratti, A. & Buratti, E. Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. *J. Neurochem.* **138 Suppl 1,** 95–111 (2016).
8. Barmada, S. J. Linking RNA Dysfunction and Neurodegeneration in Amyotrophic Lateral Sclerosis. *Neurotherapeutics* **12,** 340–351 (2015).
9. Buratti, E. & Baralle, F. E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276,** 36337–36343 (2001).
10. Ayala, Y. M. *et al.* Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol.* **348,** 575–588 (2005).
11. Kuo, P.-H., Doudeva, L. G., Wang, Y.-T., Shen, C.-K. J. & Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37,** 1799–1808 (2009).
12. Furukawa, Y. *et al.* A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Sci Rep* **6,** 20576 (2016).
13. Ihara, R. *et al.* RNA binding mediates neurotoxicity in the transgenic Drosophila model of TDP-43 proteinopathy. *Hum. Mol. Genet.* **22,** 4474–4484 (2013).
14. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J* **30,** 277–288 (2011).
15. Elden, A. C. *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466,** 1069–1075 (2010).
16. Voigt, A. *et al.* TDP-43-mediated neuron loss in vivo requires RNA-binding activity. *PLoS ONE* **5,** e12247 (2010).
17. Barmada, S. J. & Finkbeiner, S. Pathogenic TARDBP mutations in amyotrophic lateral sclerosis and frontotemporal dementia: disease-associated pathways. *Rev Neurosci* **21,** 251–272 (2010).
18. White, M. A. *et al.* TDP-43 gains function due to perturbed autoregulation in a Tardbp knock-in mouse model of ALS-FTD. *Nat Neurosci* 1–20 (2018). doi:10.1038/s41593-018-0113-5

19. Fratta, P. *et al.* Mice with endogenous TDP-43 mutations exhibit gain of splicing function and characteristics of amyotrophic lateral sclerosis. *EMBO J* **37,** e98684 (2018).
20. Koyama, A. *et al.* Increased cytoplasmic *TARDBP* mRNA in affected spinal motor neurons in ALS caused by abnormal autoregulation of TDP-43. *Nucleic Acids Res.* **44,** 5820–5836 (2016).
21. Barmada, S. J. *et al.* Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J. Neurosci* **30,** 639–649 (2010).
22. Nishimura, A. L. *et al.* Allele-Specific Knockdown of ALS-Associated Mutant TDP-43 in Neural Stem Cells Derived from Induced Pluripotent Stem Cells. *PLoS ONE* **9,** e91269 (2014).
23. Barmada, S. J. *et al.* Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. *Nat. Chem. Bio.* **10,** 677–685 (2014).
24. Watanabe, S., Kaneko, K. & Yamanaka, K. Accelerated disease onset with stabilized familial amyotrophic lateral sclerosis (ALS)-linked mutant TDP-43 proteins. *J. Biol. Chem.* **288,** 3641–3654 (2013).
25. Ling, S.-C. *et al.* ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 13318–13323 (2010).
26. Ling, S.-C., Polymenidou, M. & Cleveland, D. W. Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis. *Neuron* **79,** 416–438 (2013).
27. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struc. Mol. Biol.* **20,** 1443–1449 (2013).
28. Kuo, P.-H., Chiang, C.-H., Wang, Y.-T., Doudeva, L. G. & Yuan, H. S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res.* **42,** 4712–4722 (2014).
29. Mackness, B. C., Tran, M. T., McClain, S. P., Matthews, C. R. & Zitzewitz, J. A. Folding of the RNA Recognition Motif (RRM) Domains of the Amyotrophic Lateral Sclerosis (ALS)-linked Protein TDP-43 Reveals an Intermediate State. *J. Biol. Chem.* **289,** 8264–8276 (2014).
30. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* **14,** 452–458 (2011).
31. Bhardwaj, A., Myers, M. P., Buratti, E. & Baralle, F. E. Characterizing TDP-43 interaction with its RNA targets. *Nucleic Acids Res.* **41,** 5062–5074 (2013).
32. Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci* **14,** 459–468 (2011).
33. Guo, W. *et al.* An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat. Struc. Mol. Biol.* **18,** 822–830 (2011).
34. Scialò, C. *et al.* Clinical epidemiology of amyotrophic lateral sclerosis in Liguria, Italy: An update of LIGALS register. *Amyotrophic Lateral Scler. and Frontotemporal Degener.* **17,** 535–542 (2016).

35. Traxinger, K., Kelly, C., Johnson, B. A., Lyles, R. H. & Glass, J. D. Prognosis and epidemiology of amyotrophic lateral sclerosis: Analysis of a clinic population, 1997-2011. *Neurology: Clinical Practice* **3,** 313–320 (2013).
36. Wils, H. *et al.* TDP-43 transgenic mice develop spastic paralysis and neuronal inclusions characteristic of ALS and frontotemporal lobar degeneration. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 3858–3863 (2010).
37. Johnson, B. S., McCaffery, J. M., Lindquist, S. & Gitler, A. D. A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 6439–6444 (2008).
38. Liachko, N. F., Guthrie, C. R. & Kraemer, B. C. Phosphorylation Promotes Neurotoxicity in a Caenorhabditis elegans Model of TDP-43 Proteinopathy. *J. Neurosci* **30,** 16208–16219 (2010).
39. Wang, D. B., Gitcho, M. A., Kraemer, B. C. & Klein, R. L. Genetic strategies to study TDP-43 in rodents and to develop preclinical therapeutics for amyotrophic lateral sclerosis. *Euro. J. Neurosci.* **34,** 1179–1188 (2011).
40. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431,** 805–810 (2004).
41. Barmada, S. J. *et al.* Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 7821–7826 (2015).
42. Janssens, J. *et al.* Overexpression of ALS-Associated p.M337V Human TDP-43 in Mice Worsens Disease Features Compared to Wild-type Human TDP-43 Mice. *Mol Neurobiol* **48,** 22–35 (2013).
43. Arnold, E. S. *et al.* ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43. *Proc. Natl. Acad. Sci. U.S.A.* **110,** E736–45 (2013).
44. Kabashi, E. *et al.* Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. *Hum. Mol. Genet.* **19,** 671–683 (2009).
45. Tsvetkov, A. S. *et al.* Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration. *Nat. Chem. Bio.* **9,** 586–592 (2013).
46. Gupta, R. *et al.* The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome. *eNeuro* **4,** ENEURO.0249–16.2017 (2017).
47. Chudakov, D. M., Lukyanov, S. & Lukyanov, K. A. Using photoactivatable fluorescent protein Dendra2 to track protein movement. *BioTechniques* **42,** 553–555– 557 passim (2007).
48. Wang, F., Durfee, L. A. & Huibregtse, J. M. A Cotranslational Ubiquitination Pathway for Quality Control of Misfolded Proteins. *Mol. Cell* **50,** 368–378 (2013).
49. Kisselev, A. F. & Goldberg, A. L. Proteasome inhibitors: from research tools to drug candidates. *Chem. Biol.* **8,** 739–758 (2001).
50. Van Etten, J. *et al.* Human Pumilio proteins recruit multiple deadenylases to efficiently repress messenger RNAs. *J. Biol. Chem.* **287,** 36370–36383 (2012).
51. White, E. K., Moore-Jarrett, T. & Ruley, H. E. PUM2, a novel murine puf

protein, and its consensus RNA-binding site. *RNA* **7,** 1855–1866 (2001).
52. Wang, X., McLachlan, J., Zamore, P. D. & Hall, T. M. T. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **110,** 501–512 (2002).
53. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* eaar7366 (2018). doi:10.1007/978-0-387-46312-4
54. Iradi, M. C. G. *et al.* Characterization of gene regulation and protein interaction networks for Matrin 3 encoding mutations linked to amyotrophic lateral sclerosis and myopathy. *Sci Rep* 1–15 (2018). doi:10.1038/s41598-018-21371-4
55. Lim, F. & Peabody, D. S. RNA recognition site of PP7 coat protein. *Nucleic Acids Res.* **30,** 4138–4144 (2002).
56. Chao, J. A., Patskovsky, Y., Almo, S. C. & Singer, R. H. Structural basis for the coevolution of a viral RNA–protein complex. *Nat Struc & Mol Biol* **15,** 103–105 (2007).
57. Bin Wu, Chao, J. A. & Singer, R. H. Fluorescence Fluctuation Spectroscopy Enables Quantitative Imaging ofSingle mRNAs in Living Cells. *Biophysj* **102,** 2936–2944 (2012).
58. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the Ubiquitin-Proteasome System by Protein Aggregation. *Science* **292,** 1552–1555 (2001).
59. Gilon, T., Chomsky, O. & Kulka, R. G. Degradation signals for ubiquitin system proteolysis in *Saccharomyces cerevisiae*. *EMBO J* **17,** 2759–2766 (1998).
60. Li, B. *et al.* Hax-1 is rapidly degraded by the proteasome dependent on its PEST sequence. *BMC Cell Biol.* **13,** 1–10 (2012).
61. White, A. D. *et al.* Free energy of solvated salt bridges: a simulation and experimental study. *J Phys Chem B* **117,** 7254–7259 (2013).
62. Magalhaes, A., Maigret, B., Hoflack, J., Gomes, J. N. & Scheraga, H. A. Contribution of unusual arginine-arginine short-range interactions to stabilization and recognition in proteins. *J. Protein Chem.* **13,** 195–215 (1994).
63. Winton, M. J. *et al.* Disturbance of Nuclear and Cytoplasmic TAR DNA-binding Protein (TDP-43) Induces Disease-like Redistribution, Sequestration, and Aggregate Formation. *J. Biol. Chem.* **283,** 13302–13309 (2008).
64. Ayala, Y. M. *et al.* Structural determinants of the cellular localization and shuttling of TDP-43. *J. Cell Sci.* **121,** 3778–3785 (2008).
65. Prior, H., Jawad, A. K., MacConnachie, L. & Beg, A. A. Highly Efficient, Rapid and Co-CRISPR-Independent Genome Editing in Caenorhabditis elegans. *G3* **7,** 3693–3698 (2017).
66. Ash, P. E. A. *et al.* Neurotoxic effects of TDP-43 overexpression in C. elegans. *Hum. Mol. Genet.* **19,** 3206–3218 (2010).
67. Zhang, T., Hwang, H.-Y., Hao, H., Talbot, C. & Wang, J. Caenorhabditis elegans RNA-processing protein TDP-1 regulates protein homeostasis and life span. *J. Biol. Chem.* **287,** 8371–8382 (2012).
68. Vaccaro, A. *et al.* TDP-1/TDP-43 Regulates Stress Signaling and Age-Dependent Proteotoxicity in Caenorhabditis elegans. *PLoS Genet* **8,** e1002806 (2012).
69. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45,**

D362–D368 (2017).
70. Freibaum, B. D. *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525,** 129–133 (2015).
71. Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525,** 56–61 (2015).
72. Jovičić, A. *et al.* Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18,** 1226–1229 (2015).
73. Boeynaems, S. *et al. Drosophila* screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci Rep* **6,** 1–8 (2016).
74. Ling, J. P., Pletnikova, O., Troncoso, J. C. & Wong, P. C. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* **349,** 650–655 (2015).
75. Humphrey, J., Emmett, W., Fratta, P., Isaacs, A. M. & Plagnol, V. Quantitative analysis of cryptic splicing associated with TDP-43 depletion. *BMC Med Genomics* **10,** 38 (2017).
76. Tan, Q. *et al.* Extensive cryptic splicing upon loss of RBM17 and TDP43 in neurodegeneration models. *Hum. Mol. Genet.* **48,** ddw337 (2016).
77. Goldstein, L. D. *et al.* Prediction and Quantification of Splice Events from RNA-Seq Data. *PLoS ONE* **11,** e0156132 (2016).
78. Shiina, Y., Arima, K., Tabunoki, H. & Satoh, J.-I. TDP-43 Dimerizes in Human Cells in Culture. *Cell Mol Neurobiol* **30,** 641–652 (2009).
79. Sun, Y., Arslan, P. E., Won, A., Yip, C. M. & Chakrabartty, A. Binding of TDP-43 to the 3′UTR of Its Cognate mRNA Enhances Its Solubility. *Biochemistry* **53,** 5885–5894 (2014).
80. Zhang, Y.-J. *et al.* The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Hum. Mol. Genet.* **22,** 3112–3122 (2013).
81. Ederle, H. *et al.* Nuclear egress of TDP-43 and FUS occurs independently of Exportin-1/CRM1. *Sci Rep* **8,** 1–18 (2018).
82. Archbold, H. C. *et al.* TDP43 nuclear export and neurodegeneration in models of amyotrophic lateral sclerosis and frontotemporal dementia. *Sci Rep* 1–18 (2018). doi:10.1038/s41598-018-22858-w
83. Pinarbasi, E. S. *et al.* Active nuclear import and passive nuclear export are the primary determinants of TDP-43 localization. *Sci Rep* **8,** 1–16 (2018).
84. Armakola, M. *et al.* Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nature Genet* **44,** 1302–1309 (2012).
85. Diaper, D. C. *et al.* Loss and gain of Drosophila TDP-43 impair synaptic efficacy and motor control leading to age-related neurodegeneration by loss-of-function phenotypes. *Hum. Mol. Genet.* **22,** 1539–1557 (2013).
86. Feiguin, F. *et al.* Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior. *FEBS Letters* **583,** 1586–1592 (2009).
87. Sephton, C. F. *et al.* TDP-43 is a developmentally regulated protein essential for early embryonic development. *J. Biol. Chem.* **285,** 6826–6834 (2010).
88. Kraemer, B. C. *et al.* Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. *Acta Neuropathol* **119,** 409–419 (2010).

89. Bose, J. K., Wang, I.-F., Hung, L., Tarn, W.-Y. & Shen, C.-K. J. TDP-43 overexpression enhances exon 7 inclusion during the survival of motor neuron pre-mRNA splicing. *J. Biol. Chem.* **283,** 28852–28859 (2008).
90. Hazelett, D. J., Chang, J.-C., Lakeland, D. L. & Morton, D. B. 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* **2,** 789–802 (2012).
91. Burgos, H. L., O'Connor, K., Sanchez-Vazquez, P. & Gourse, R. L. Roles of Transcriptional and Translational Control Mechanisms in Regulation of Ribosomal Protein Synthesis in *Escherichia coli*. *J. Bacteriol.* **199,** e00407–17 (2017).
92. Nomura, M., Yates, J. L., Dean, D. & Post, L. E. Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein MRNA. *Proc. Natl. Acad. Sci. U.S.A.* **77,** 7084–7088 (1980).
93. Guo, Q. *et al.* In Situ Structure of Neuronal C9orf72 Poly-GA Aggregates Reveals Proteasome Recruitment. *Cell* **172,** 696–705.e12 (2018).
94. Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165,** 976–989 (2017).
95. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10,** R25 (2009).
96. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
97. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with rNA-seq. *Nature Biotechnology* **31,** 46–53 (2012).
98. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Research* **12,** 996–1006 (2002).
99. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12,** 357–360 (2015).
100. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47–e47 (2015).

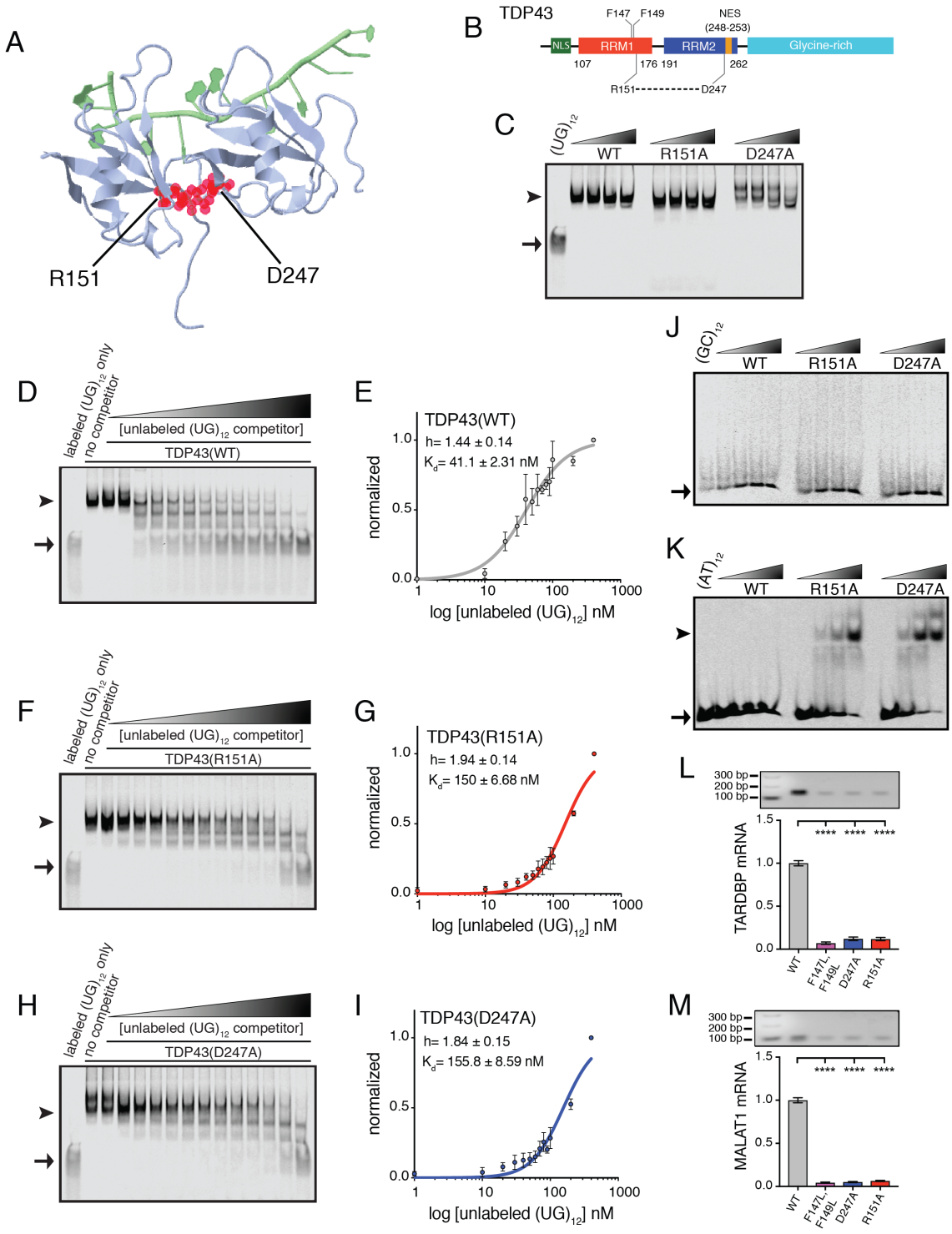

Figure 3.1 Disruption of TDP43's salt bridge impairs nucleic acid recognition. **(A)** Structure of RRM1 and RRM2 from TDP43 (blue) in complex with UG-repeat RNA (in green), highlighting the salt bridge between Arg151 and Asp247 (in red). The structure was generated in *FirstGlance in Jmol,* PDB identification code 4bs2, deposited by Lukavsky *et al*. (2013)[27]. **(B)** Schematic representation of TDP43, depicting the RRM1-

RRM2 salt bridge and F147/F149, two residues necessary for nucleic acid recognition[9]. **(C)** Electromobility shift assay (EMSA) of recombinant TDP43(WT) and mutants at increasing protein concentrations (12 fmol to 4 pmol) mixed with labeled $(UG)_{12}$ oligonucleotides (100 pM). **(D, F, H)** Unlabeled $(UG)_{12}$ oligonucleotides were added at increasing concentrations (1-400 nM) with 0.5 pmol protein and 100 pM of labeled $(UG)_{12}$ sequences. Three independent replicates for each protein were quantified in **E, G, I,** to generate dissociation constants (K_d) and Hill slopes for each variant. **(J)** EMSA of TDP43 variants at increasing protein concentrations (0.5-16 pmol) with 100 pM of labeled $(GC)_{12}$ oligonucleotides. Among the conditions tested, no shift was observed for WT or mutant TDP43 variants. **(K)** EMSA of TDP43 variants at increasing protein concentrations (0.5-16 pmol) with 100 pM of labeled $(AT)_{12}$ sequences, demonstrating shifts for both TDP43(R151A) and TDP43(D247A), but not TDP43(WT). **(L, M)** RNA immunoprecipitation of EGFP-tagged versions of TDP43(WT), TDP43(F147L, F149L), TDP43(R151A), or TDP43(D247A) expressed in HEK293T cells. The abundance of bound *TARDBP* **(L)** or *MALAT1* **(M)** transcripts were measured by qRT-PCR. For **C, D, F, H, J,** and **K,** arrowheads indicate protein-DNA complexes, while arrows point to unbound oligonucleotides. For **E, G,** and **I**, data were pooled from 3 independent replicates, the K_d and Hill coefficient were determined by the nonlinear least squares regression fit equation, and the plots show mean ± SD. For **L** and **M**, data were pooled from 3 independent replicates, and plots show mean ± SEM.

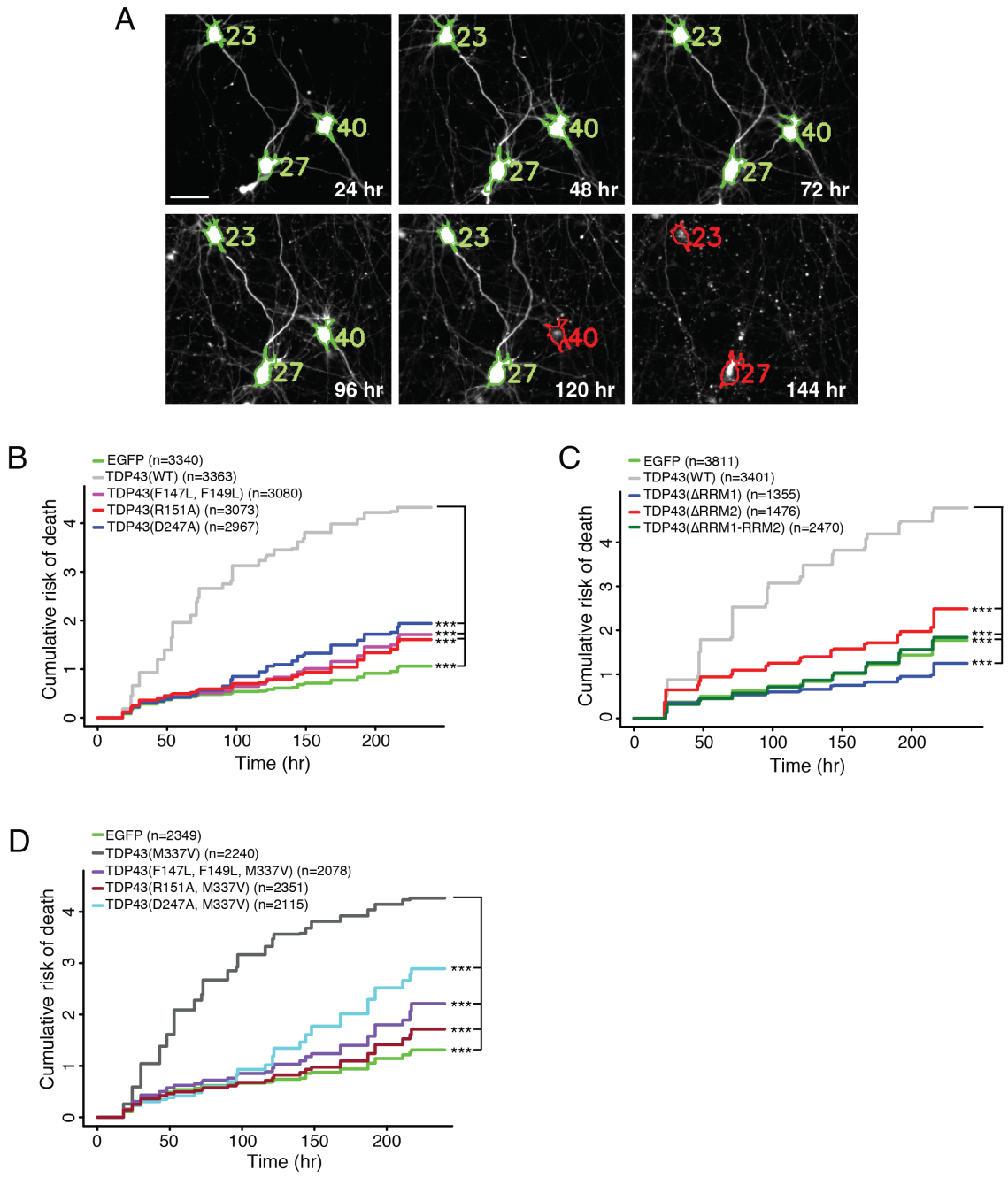

Figure 3.2 Disruption of the RRM1-RRM2 salt bridge abrogates TDP43-mediated neurotoxicity. (A) Automated fluorescence microscopy. Primary rodent cortical neurons were transfected with mApple, a survival and morphology marker, and TDP43 variants fused to EGFP. Individual neurons were programmatically identified by mApple fluorescence (green outlines) and given unique identifiers (green numbers). Cell death (red outline) was determined by loss of fluorescence, rounding of the soma, or elimination of neuritic processes. Scale bar: 20 μm. **(B)** Overexpressing TDP43(WT)-EGFP significantly increases the cumulative risk of death compared to neurons

expressing EGFP (HR 5.04, $p < 2x10^{-16}$). Mutations that disrupt the salt bridge (R151A or D247A) and mutations that eliminate RNA binding (F147L, F149L), significantly reduce the risk of death compared to neurons expressing TDP43(WT) (HR 0.25, 0.32, and 0.27, respectively, $p < 2x10^{-16}$ for all comparisons). **(C)** Deletion of RRM2 or both RRM1-RRM2 significantly improve survival compared to TDP43(WT) (HR 0.51 and 0.31, respectively, $p < 2x10^{-16}$, for both comparisons). Deletion of RRM1 reduces the risk of death compared to TDP43(WT) (HR 0.26, $p < 2x10^{-16}$) and EGFP alone (HR 0.7, $p < 2x10^{-16}$). **(D)** Expression of TDP43(M337V)-EGFP significantly increases the risk of death (HR 4.26, $p < 2x10^{-16}$) compared to EGFP. The F147L/F149L, R151A, and D247A mutations significantly reduce the toxicity of TDP43(M337V)-EGFP (HR 0.34, 0.26, and 0.39, respectively, $p < 2x10^{16}$ for all comparisons). n, number of neurons. *** $p < 2x10^{-16}$, Cox proportional hazards. Survival analyses were pooled from 3 independent experiments, with 8 wells per condition for each replicate.

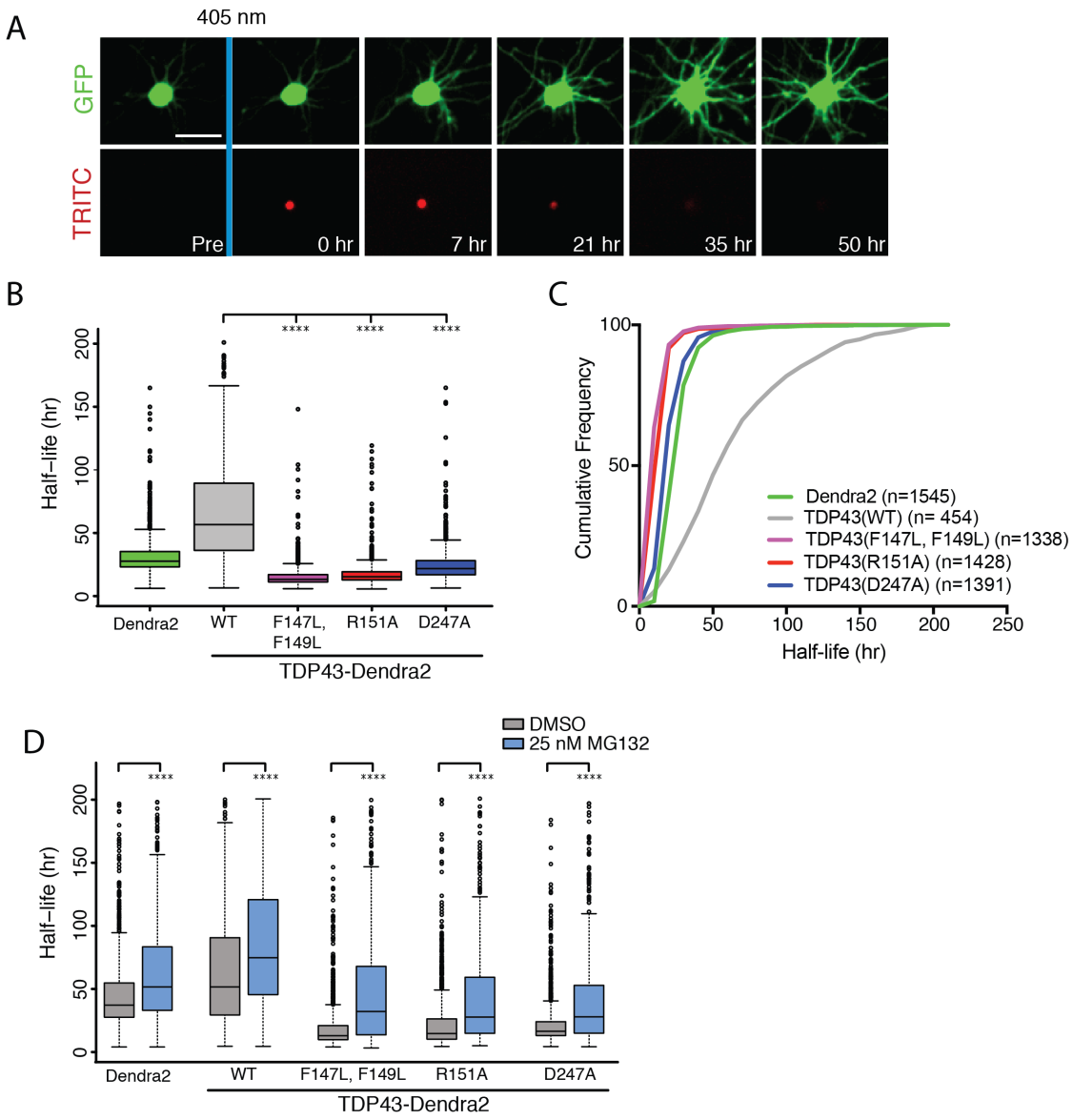

Figure 3.3 Mutations that interfere with the RRM1-RRM2 salt bridge or RNA binding destabilize TDP43. (A) Optical pulse labeling (OPL) of primary cortical neurons using automated fluorescence microscopy. Neurons were transfected with EGFP and TDP43-Dendra2 variants, photoconverted with a pulse of 405 nm of light, and imaged over time. The time-dependent loss of photoconverted TDP43-Dendra2 (detected in the TRITC channel) for each cell was used to calculate protein half-life. Scale bar: 50 µm. Pre, before photoconversion. **(B)** Box plot of protein half-life. The median TDP43(WT)-Dendra2 half-life (58.5 hr) was significantly greater than that of TDP43(R151A)-Dendra2 (15.3 hr), TDP43(D247A)-Dendra2 (21.8 hr), and TDP43(F147L, F149L)-Dendra2 (13.2 hr). **(C)** Cumulative frequency plot of the protein half-life data from **B**. **** $p < 0.01$, one-way ANOVA with Dunnett's test**.** n, number of neurons**. (D)** 6 hr prior to OPL, cells were treated with either DMSO or 25 nM MG132. Treatment of MG132 significantly extended the half-life of Dendra2 and TDP43-Dendra2 variants compared to DMSO controls. **** $p < 0.01$, one-way ANOVA with Tukey's test. For **B** data pooled from 4 biological replicates. For **D** data accumulated from 300-1400 cells, over 3 biological replicates.

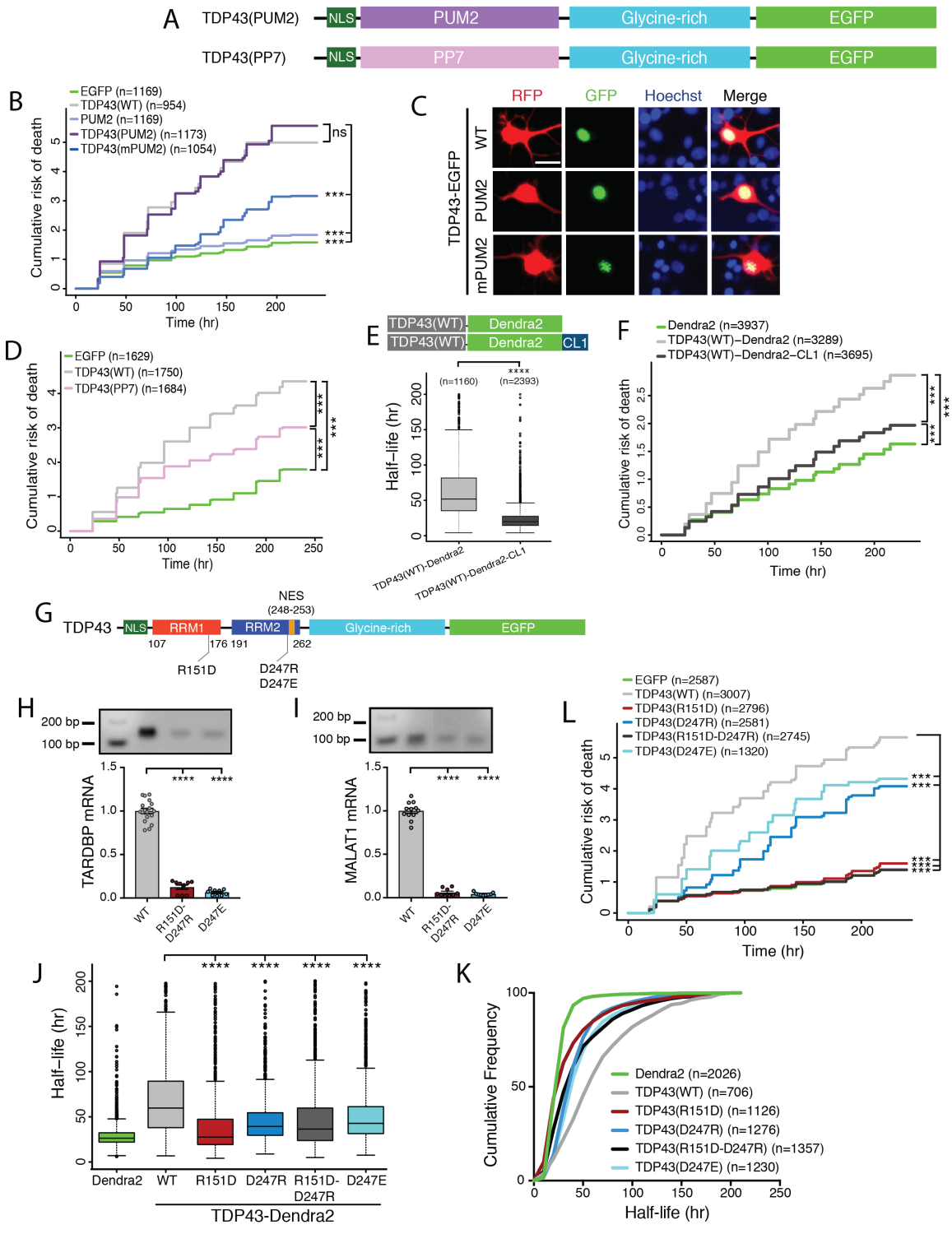

Figure 3.4 RNA binding and protein stability are essential for TDP43-dependent neurodegeneration. (A) Schematic representation of EGFP-tagged TDP43 constructs in which the region comprised of RRM1-RRM2 is replaced with RNA binding domains (RBDs) from PUM2 or PP7. **(B)** TDP43(PUM2)-EGFP expression significantly increased the cumulative risk of death compared to EGFP alone (HR 2.54, $p < 2x10^{-16}$).

No significant difference was observed among neurons expressing TDP43(WT)-EGFP or TDP43(PUM2)-EGFP. Expression of the PUM2 RBD marginally elevated the risk of death compared to EGFP alone (HR 1.19, $p= 3.78 \times 10^{-5}$). RNA binding-deficient versions of the PUM2 RBD (mPUM2) significantly reduced the risk of death compared to TDP43(PUM2)-EGFP (HR 0.45, $p < 2x10^{-16}$). **(C)** Fluorescent images of representative neurons overexpressing mApple (RFP) and EGFP-tagged TDP43(WT), TDP43(PUM2), or TDP43(mPUM2), demonstrating a diffuse distribution within the nucleus for all TDP43 variants, but TDP43(mPUM2) exhibited multiple puncta within the nucleus. Scale bar: 20 μm. **(D)** Expression of TDP43(PP7) reduced the risk of death in comparison to TDP43(WT) (HR 0.73, $p < 2x10^{-16}$) but remained significantly toxic compared to EGFP alone (HR 2.05, $p < 2x10^{-16}$). **(E)** Schematic representation of Dendra2-tagged TDP43(WT) and TDP43(WT)-CL. By OPL, the median TDP43-Dendra2 half-life dropped from 51.9 h to 19.8 h with addition of the destabilizing CL degron. **f**, Both TDP43(WT)-Dendra2 and TDP43(WT)-Dendra2-CL significantly elevated the risk of death compared to Dendra2 alone (HR 1.82 and 1.18, $p < 2x10^{-16}$ and $p= 1.23 \times 10^{-10}$, respectively). Expression of TDP43(WT)-Dendra2-CL significantly decreased the risk of death compared to TDP43(WT)-Dendra2 (HR 0.64, $p < 2x10^{-16}$). **(G)** Diagram of EGFP-tagged TDP43 carrying mutations tested in **H-I**. **H-I**, HEK293T cells were transfected with EGFP-tagged versions of TDP43(WT), TDP43(R151D-D247R) or TDP43(D247E) prior to immunoprecipitation with anti-GFP antibodies. *TARDBP* **(H)** and *MALAT1* **(I)** transcripts were amplified from total RNA by qRT-PCR. For each, a representative image of the PCR products assessed by PAGE (top) and quantification of band intensity normalized to TDP43(WT) (bottom) are shown. **(J)** A significant decrease in protein half-life was observed for all TDP43 mutants compared to TDP43(WT), as determined by OPL. **(K)** Cumulative frequency plot of data shown in **J**. **(L)** Longitudinal fluorescence microscopy of neurons overexpressing EGFP-tagged TDP43 variants or EGFP alone. TDP43(WT)-EGFP expression was significantly toxic to cells compared to EGFP alone (HR 4.42, $p < 2x10^{-16}$). A significant reduction in toxicity was observed for all TDP43 mutants compared to TDP43(WT)-EGFP (TDP43(R151D)-EGFP, HR 0.23; TDP43(D247R)-EGFP, HR 0.48; TDP43(R151D-D247R)-EGFP, HR 0.21; TDP43(D247E)-EGFP, HR 0.6; $p < 2x10^{-16}$ for all comparisons). Cells expressing either TDP43(D247R)-EGFP or TDP43(D247E)-EGFP demonstrated a significantly elevated risk of death compared to those expressing EGFP alone (HR 2.2 and 2.76, respectively, $p < 2x10^{-16}$ for both comparisons). Data were pooled from 3 independent replicates for **B, D,** and **L** and 4 replicates for **F**, *** $p < 2x10^{-16}$, Cox proportional hazards**.** For **E**, data were pooled from 4 independent experiments, **** $p < 0.0001$, unpaired t test. For **H** and **I**, data were pooled from 3 independent experiments, **** $p < 0.05$, Kruskal-Wallis with Dunn's test. For **J**, data were pooled from 4 replicates, **** $p < 0.05$, one-way ANOVA with Dunnett's test. For **B**, **D-F**, **K**, and **L**, n, number of neurons.

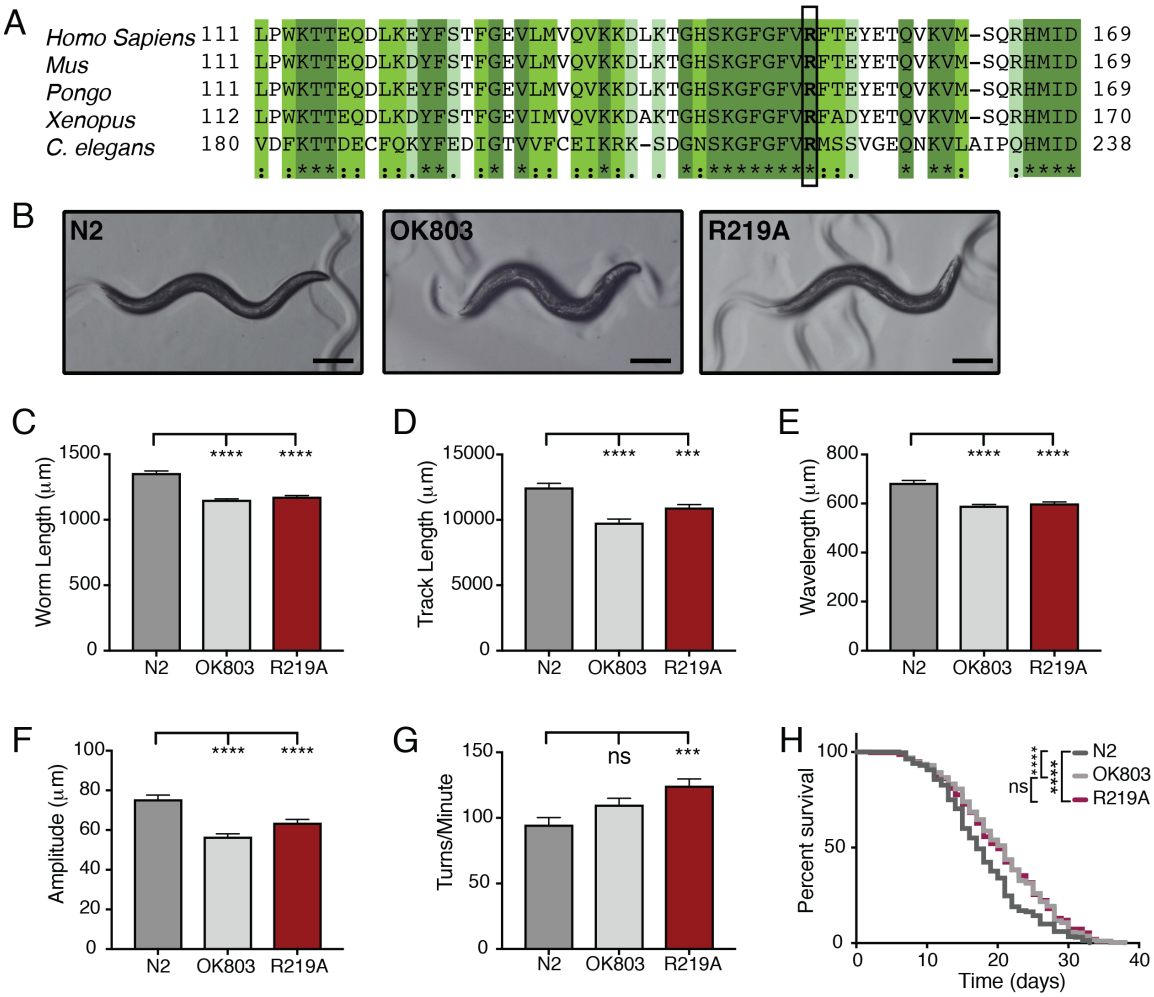

Figure 3.5 The *tdp-1*(R219A) knock-in phenocopies TDP-1 null mutations in *C. elegans*. (A) Interspecies sequence conservation of TDP43's RRM1 domain. Highly conserved residues are highlighted in green. Conservation of R151 (in bold) is highlighted. **(B)** Representative images of 7 d old wild-type (N2), *tdp-1 (ok803)* knockout, and mutant *tdp-1* (R219A) animals. Scale bar: 200 mm. **(C-F)** Behavioral assays of 7 d old animals measuring worm length, track length, wavelength, and amplitude on NGM plates. Both *ok803* and R219A animals displayed significantly reduced locomotor metrics compared to N2 animals on NGM plates. **(G)** Thrash assay measuring the number of body bends per minute in liquid media. R219A, but not *ok803,* animals displayed a significant increase in the number of body bends per minute compared to N2. **(H)**, An increase in survival was observed for both *ok803* and R219A animals compared to N2, but no significant difference was observed between *ok803* and R219A worms. **(C-F)** data were pooled from 3 independent experiments, totaling n≥127 animals/genotype; **** $p < 0.0001$, one-way ANOVA, Dunnett's post-hoc. **(G)**, data were pooled from 3 independent experiments, totaling n≥96 animals/genotype; **** $p < 0.001$, one-way ANOVA, Dunnett's post-hoc. **(H)**, data were pooled from 3 replicates, totaling 418 animals/genotype; **** $p < 0.0001$, log-rank test. Plots in **C-G** show mean ± SEM.

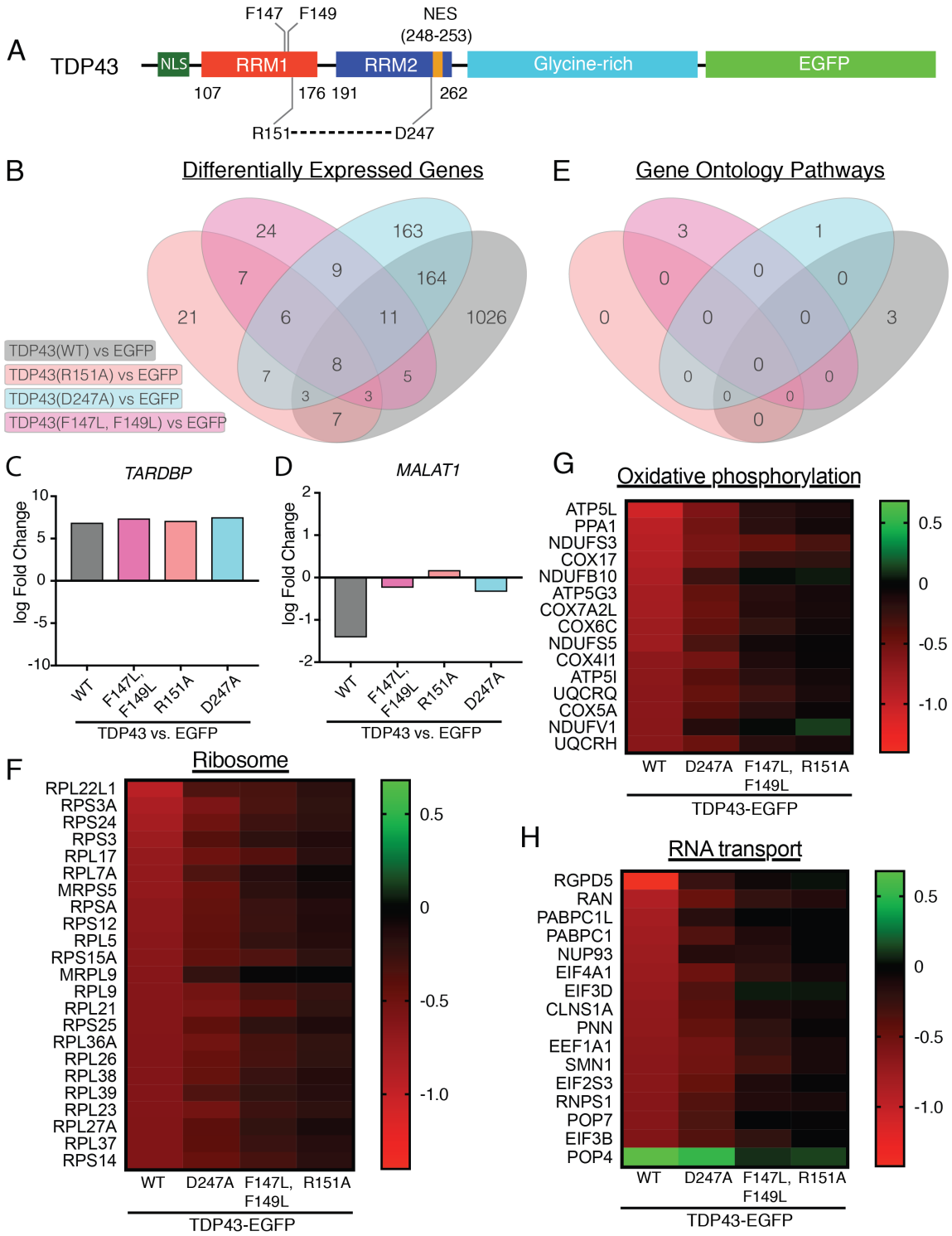

Figure 3.6 TDP43(WT) overexpression selectively affects the ribosomal and oxidative phosphorylation pathways. (A) Diagram of EGFP-tagged TDP43 variants used for subsequent experiments. HEK293T cells were transfected with EGFP-tagged versions of TDP43(WT), TDP43(R151A), TDP43(D247A), TDP43(F147L, F149L), or EGFP alone, and subject to high-throughput RNA-sequencing. **(B)** Venn diagram

depicting the number of differentially expressed genes (DEGs) for each condition, in comparison to EGFP alone. **(C)** *TARDBP* transcript levels are similar among all TDP43 variants, demonstrating equivalent overexpression in each case. **(D)** Levels of *MALAT1*, an established TDP43 target transcript, were significantly downregulated only in TDP43(WT)-expressing cells. **(E)** Venn diagram showing significant enrichment for functional pathways according to gene ontology (GO) analysis. Among downregulated mRNAs, transcripts involved in the ribosomal **(F)** and oxidative phosphorylation **(G)** pathways (false discovery rate (FDR) = 6.8×10^{-5} and 0.03, respectively) were highly enriched, with a trend towards enrichment for transcripts related to RNA transport (**H**, FDR = 0.222).

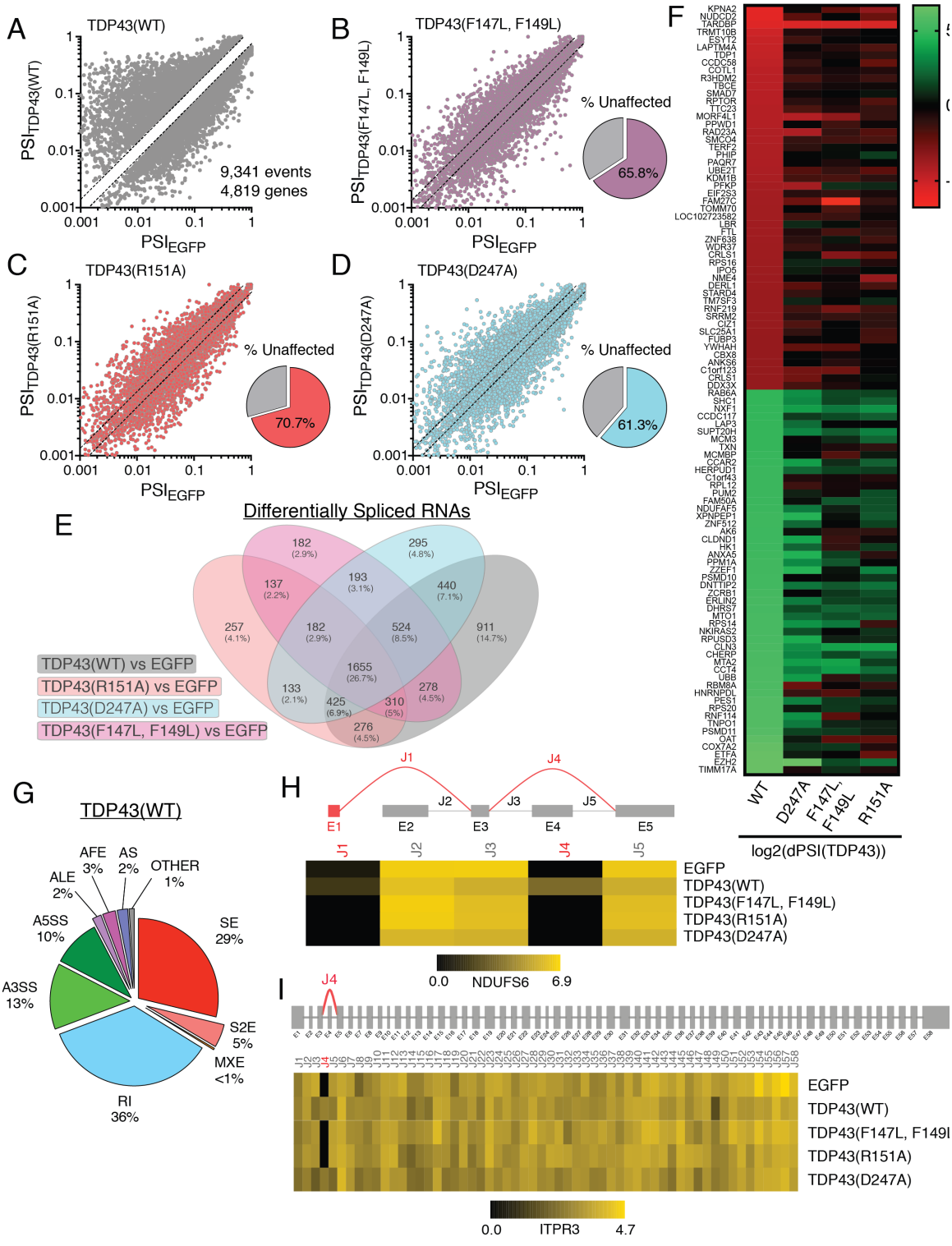

Figure 3.7 TDP43(WT) overexpression promotes intron retention and exon skipping. (A) 9,339 differential splicing events involving 4,819 genes were noted in TDP43(WT)-EGFP overexpressing cells. **(B-D)** Of these splicing events, 34.2% were also found upon overexpression of TDP43(F147L, F149L)-EGFP, 29.3% with

TDP43(R151A)-EGFP, and 38.7% with TDP43(D247A)-EGFP. **(E)** 3,968 transcripts (42.5%) were differentially and selectively spliced by TDP43(WT)-EGFP. **(F)** Heat map of the 50 top repressed and 50 top included splicing events. **(G)** Categorization of abnormal splicing events noted selectively with TDP43(WT) overexpression. RI, intron retention. A3SS, alternative 3' splice sites. A5SS, alternative 5' splice sites. ALE, alternative last exon. AFE, alternative first exon. AS, alternative start. SE, skipping of one exon. S2E, skipping of 2 exons. MXE, mutually exclusive exons. **(H)** Representative examples of abnormal splicing in *NDUFS6* **(H)** and *ITPR3* **(I)** transcripts, with schematic diagrams of exons (E) and splice junctions (J) above heat maps displaying the read density for each junction. Unannotated exons/junctions are shown in red.

Supplemental information

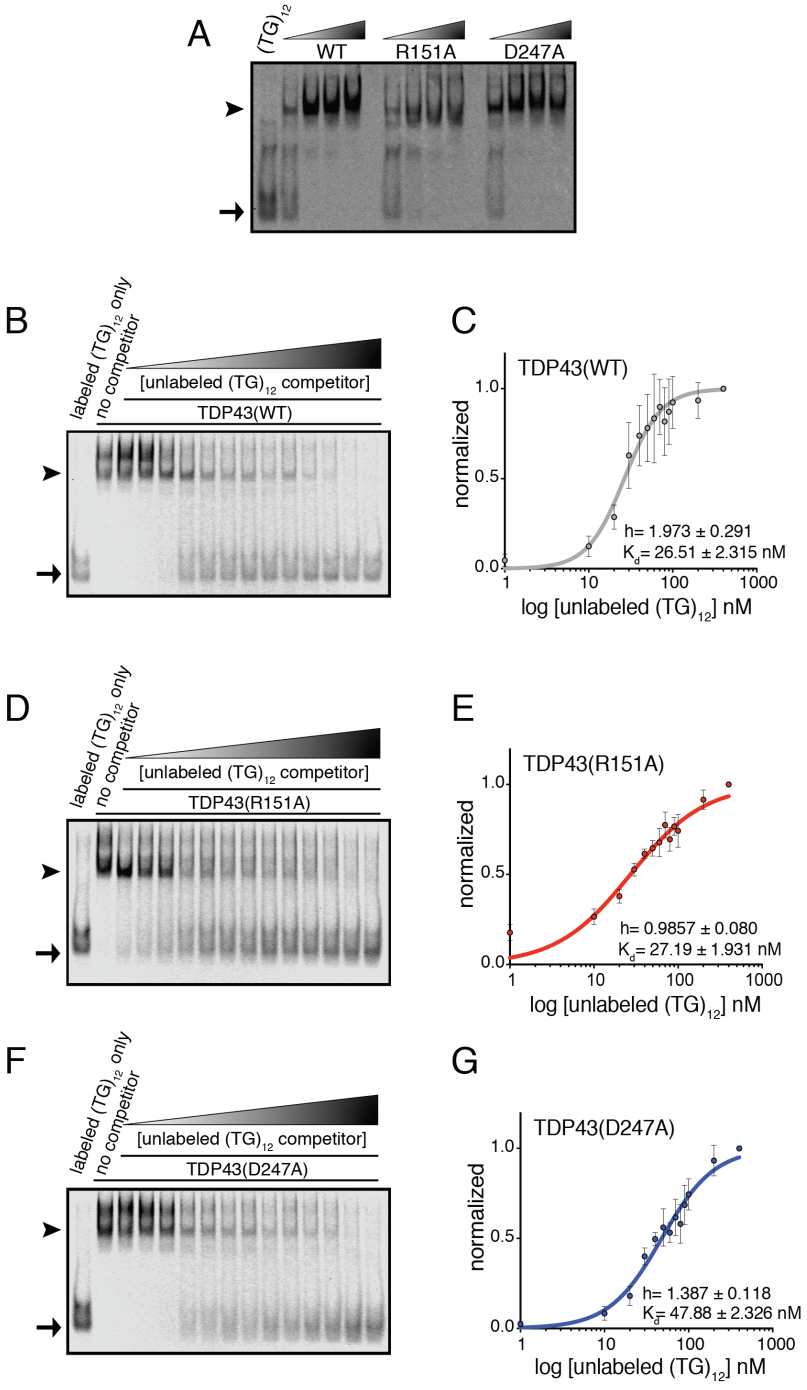

Figure S3.1 Salt bridge disrupting mutations of TDP43 reduce the cooperativity and affinity for $(TG)_{12}$ sequences. (A) EMSA of recombinant TDP43 variants at increasing protein concentrations (12 fmol to 4 pmol) incubated with labeled $(TG)_{12}$ oligonucleotides (100 pM). **(B, D, F)** Unlabeled $(TG)_{12}$ oligomers were added at increasing concentrations (1-400 nM) with 0.5 pmol protein and 100 pM of labeled

$(TG)_{12}$ sequences. Three independent replicates for each protein were quantified in **C, E, G** to calculate dissociation constants (K_d) and Hill slopes. For **A**, **B, D, F,** arrowheads mark protein-DNA complexes, while arrows indicate free DNA oligomers. For **C, E,** and **G**, K_d and Hill coefficient were determined by the nonlinear least squares regression fit equation, and each plot displays mean ± SD.

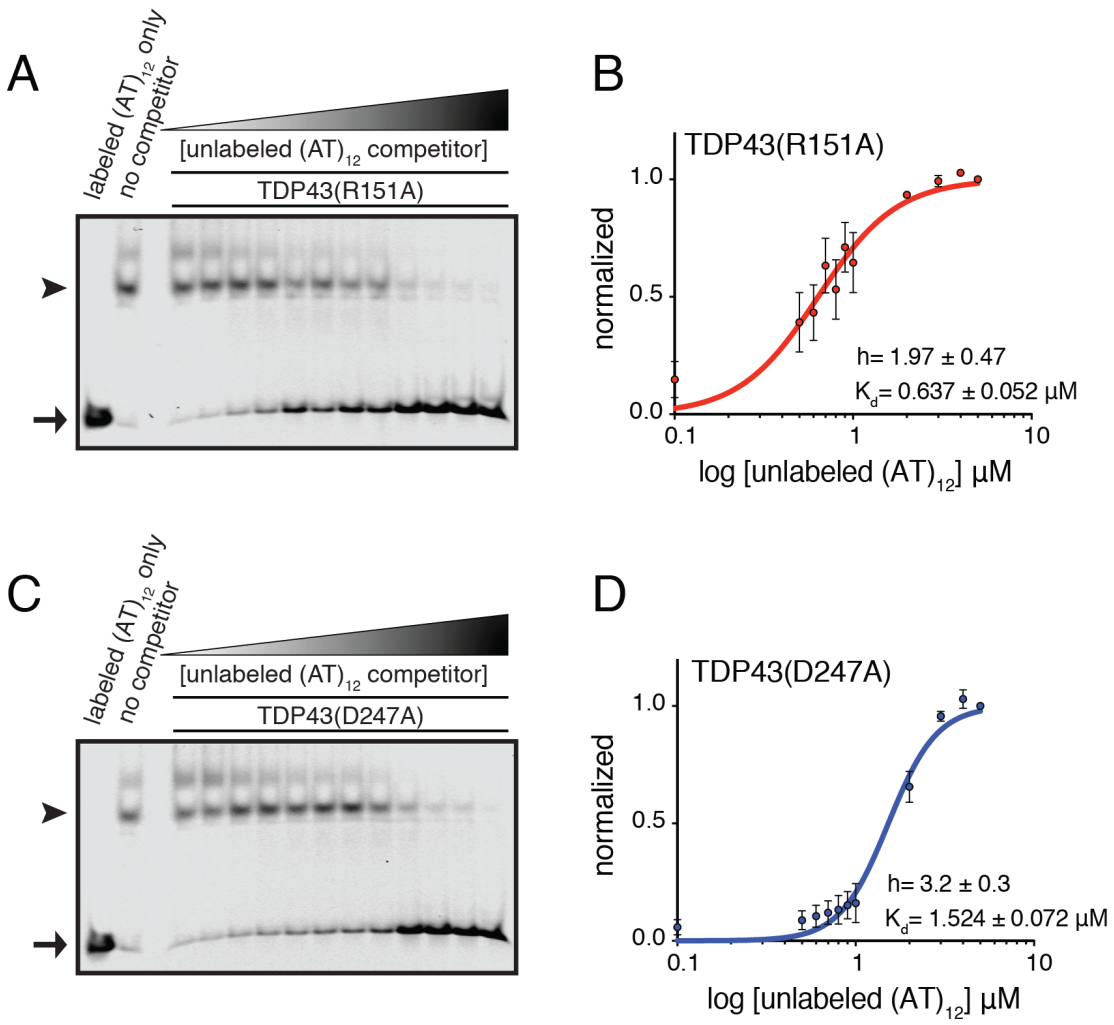

Figure S3.2 TDP43(R151A) and TDP43(D247A) recognize $(AT)_{12}$ oligonucleotides *in vitro*. (A, C) Competition EMSA of recombinant TDP43(R151A) or TDP43(D247A) (16 pmol) incubated with unlabeled $(AT)_{12}$ sequences at increasing concentrations (0.01-5 µM) and 100 pM of labeled $(AT)_{12}$. Dissociation constants (K_d) and Hill slopes were quantified from three independent replicates in **B** and **D**. For **A** and **C**, arrowheads point to protein-DNA complexes, and arrows show unbound DNA. For **B** and **D,** the nonlinear least squares regression fit equation was used to determine both K_d and Hill coefficient; each plot displays mean ± SEM.

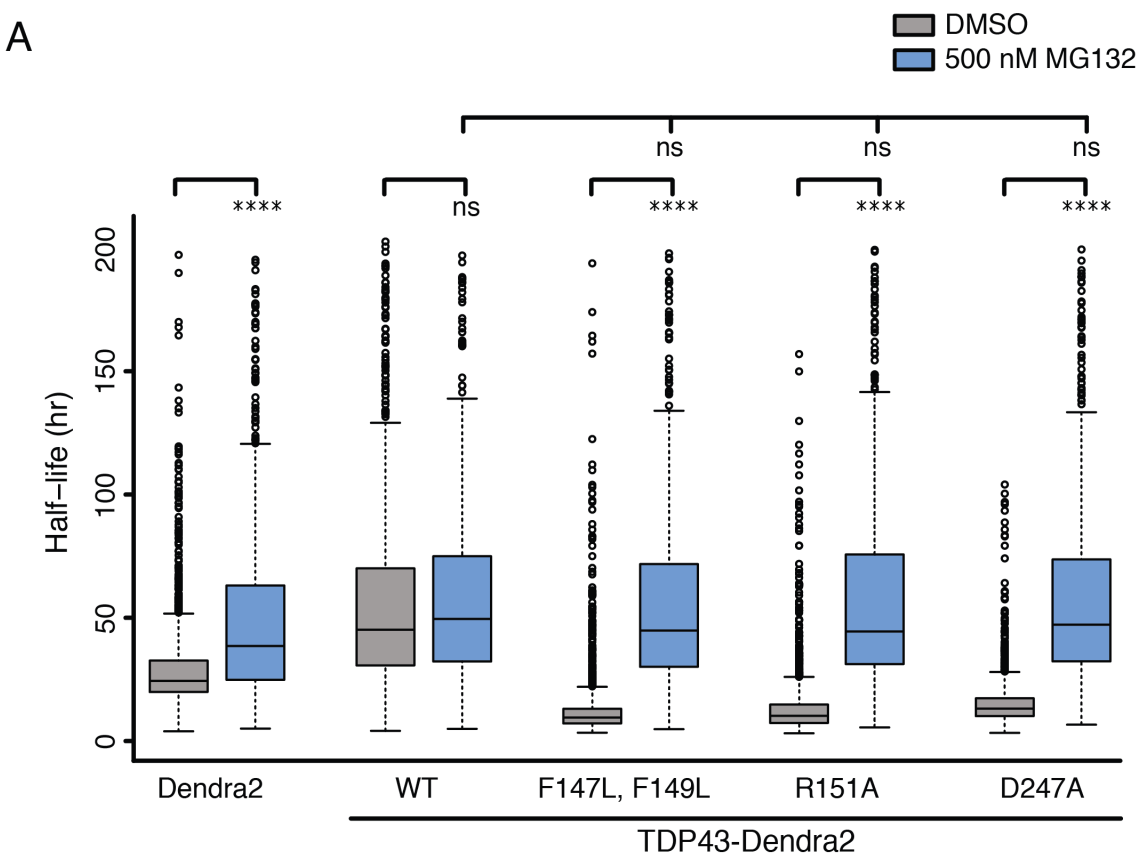

Figure S3.3 High doses of MG132 completely stabilize RNA binding-deficient TDP43 variants. (A) Box plot of protein half-life. Neurons were treated with either DMSO or 500 nM MG132 6 hr prior to OPL. The half-life of Dendra2 and TDP43-Dendra2 mutants significantly increased compared to DMSO controls. **** $p < 0.01$ one-way ANOVA with Tukey's test. Data represents from 300-1700 cells, over 3 independent replicates. Plots show median (black line), interquartile range (IQR, thick box), and 3 x IQR (whiskers).

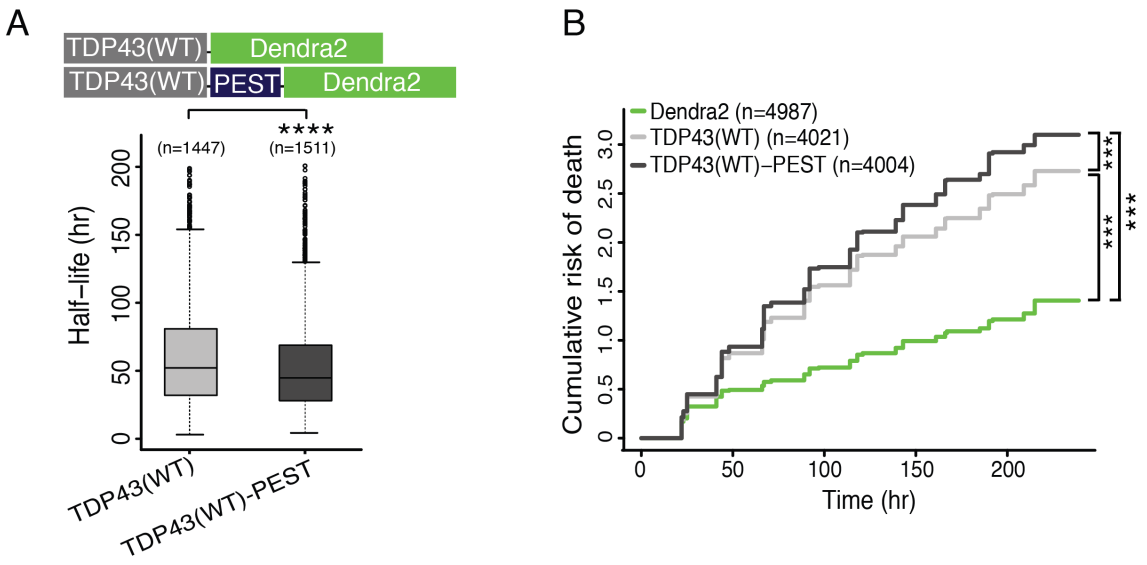

Figure S3.4 Fusion of a PEST sequence modestly affects TDP43(WT) stability and toxicity. (A) Schematic representation of Dendra2-tagged TDP43(WT) and TDP43(WT)-PEST. By OPL, the median half-life of TDP43-Dendra2 dropped from 52.2 hr to 44.8 hr in the presence of the destabilizing PEST sequence. **(B)** Neurons expressing TDP43(WT)-Dendra2 or TDP43(WT)-PEST-Dendra2 demonstrated an elevated risk of death in comparison to cells expressing Dendra2 alone (HR 1.93 and 2.14, respectively; $p < 2x10^{-16}$ for both, Cox proportional hazards). TDP43(WT)-PEST-Dendra2 was modestly more toxic than TDP43(WT)-Dendra2 (with TDP43(WT)-Dendra2 as a reference, HR 1.11, $p= 6.37x10^{-6}$). For **A**, data were pooled from 3 biological replicates, **** $p < 0.0001$, unpaired t test. Plot shows median (black line), interquartile range (IQR, thick box), and 3 x IQR (whiskers). For **B**, data were pooled from 3 independent replicates, *** $p < 2x10^{-16}$, Cox proportional hazards. For **A** and **B**, n represents the number of neurons.

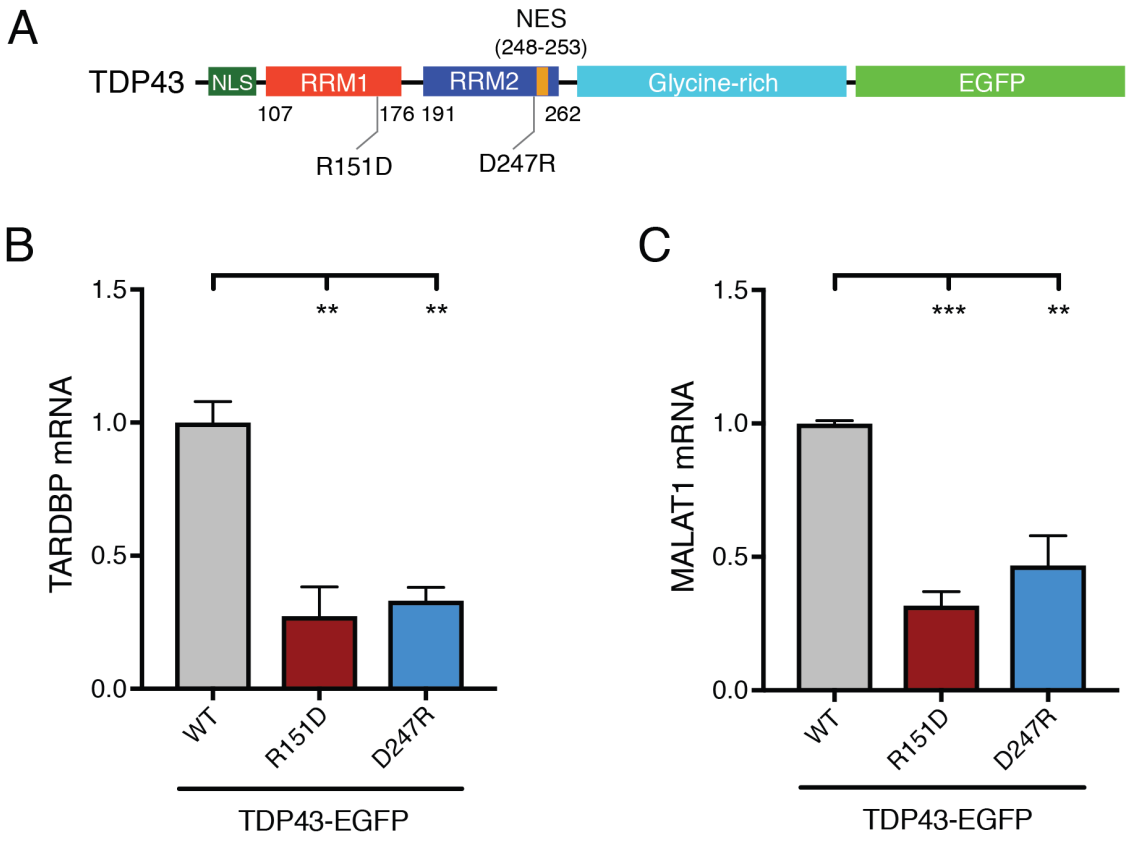

Figure S3.5 Mutations disrupting the RRM1-RRM2 salt bridge weaken TDP43's ability to recognize native substrates. (A) Schematic representation of EGFP-tagged TDP43, illustrating mutations tested in **B** and **C**. HEK293T cells were transfected with TDP43(WT), TDP43(R151D), or TDP43(D247R) and immunoprecipitated with antibodies against GFP. Bound *TARDBP* **(B)** or *MALAT1* **(C)** transcripts were amplified from total RNA by qRT-PCR. For **B** and **C**, data were pooled from 3 independent replicates. Plots show mean ± SEM, ** $p < 0.05$, Kruskal-Wallis with Dunn's test.

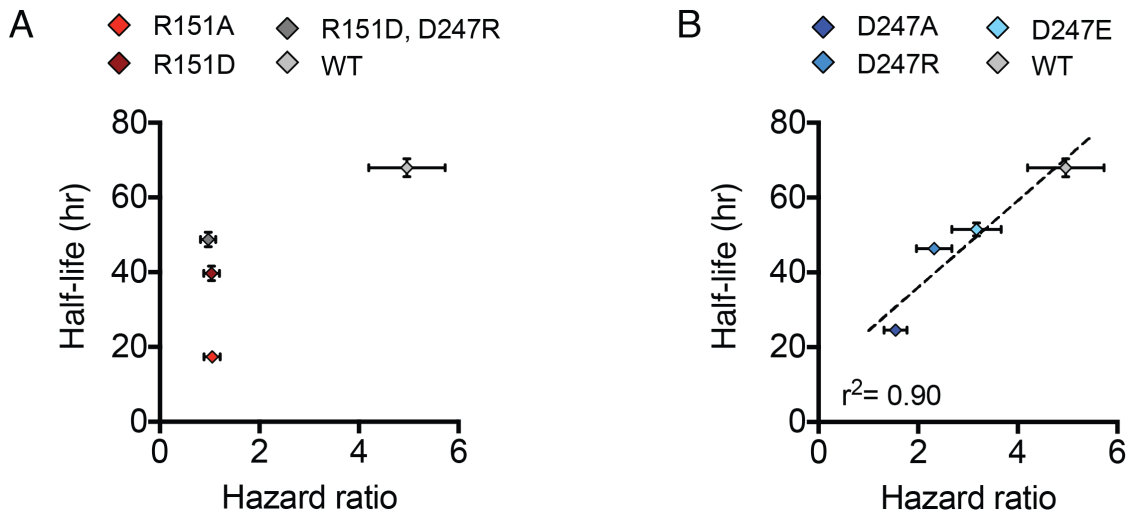

Figure S3.6 Arg151 mutations interrupt the proportional relationship between TDP43 stability and toxicity. While no relationship was observed between half-life and toxicity for R151 variants **(A)**, D247 variants displayed a positive linear relationship between half-life and toxicity **(B)**. Nonlinear least square regression equation was used to fit a line and generate the r^2 value for the relationship between half-life and toxicity. For **A** and **B**, error bars represent 95% confidence intervals.

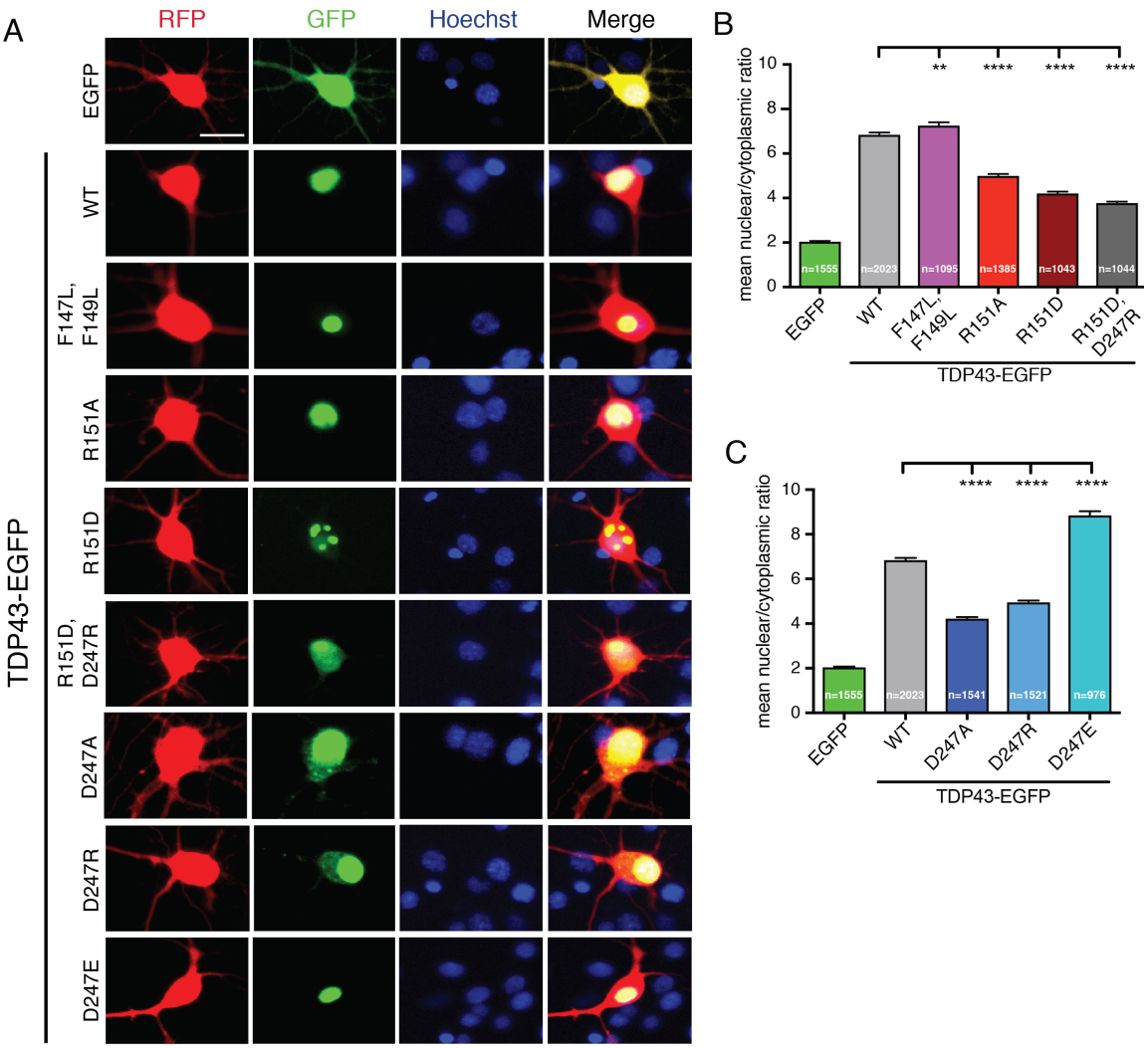

Figure S3.7 Manipulating the RRM1-RRM2 salt bridge affects subcellular TDP43 localization. (A) Fluorescent microscopy of primary neurons expressing mApple (RFP) and EGFP-tagged TDP43 variants. Nuclei were stained with Hoechst 33258. Scale bar: 20 µm. **(B** and **C)** Subcellular TDP43-EGFP localization was quantified by measuring the fluorescence intensity of TDP43 within the nuclear and cytoplasmic compartments separately for each neuron. The nuclear-cytoplasmic ratio for TDP43(WT) was 6.9 ± 0.087 SEM. **(B)** All variants of R151 significantly decreased the TDP43-EGFP nuclear-cytoplasmic ratio. **(C)** Mutating D247 to either alanine or arginine significantly decreased TDP43-EGFP nuclear-cytoplasmic ratio, while mutating D247 to glutamate significantly increased TDP43-EGFP nuclear localization. Data in **B** and **C** represent at least 3 independent experiments, ****$p < 0.0001$, one-way ANOVA with Dunnett's post-hoc. Plots in **B, C** show mean ± SEM; n, number of neurons.

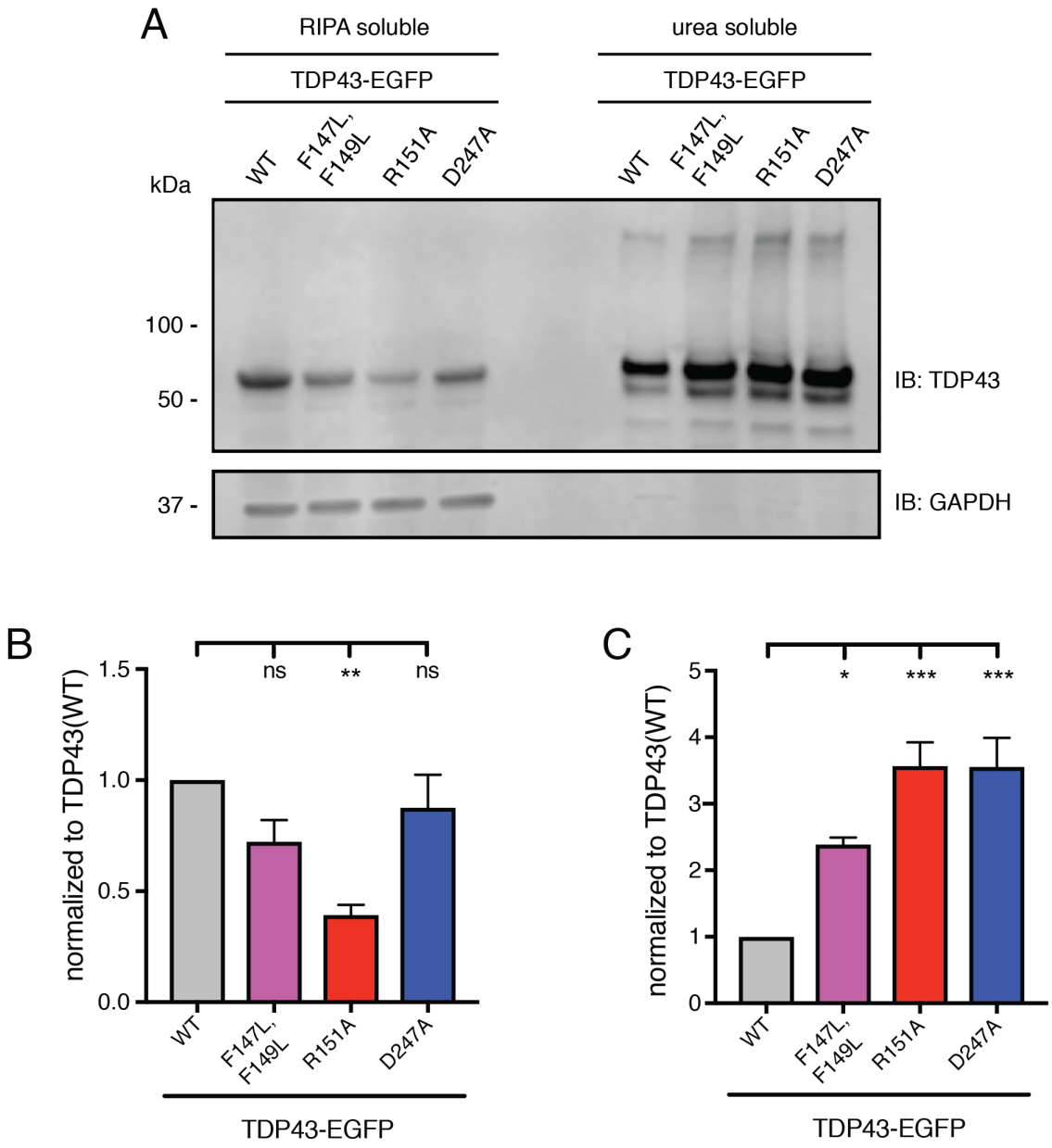

Figure S3.8 Mutations that interfere with RNA binding also reduce solubility. A, HEK293T cells were transfected with TDP43 variants, sonicated in RIPA buffer or extracted with urea buffer as indicated, and immunoblotted with anti-TDP43 antibodies. **(B**, **C)** Each TDP43 mutant was enriched within the urea-soluble fraction, compared to TDP43(WT). Data represent 3 independent replicates, ***p< 0.0001, one-way ANOVA with Dunnett's post-hoc. Plots in **B**, **C** show mean ± SEM.

Table S3.1 Pathways in which spliced transcripts were most affected from TDP43(WT) overexpression.

KEGG Term	p	FE	Bonf.	Benj.	FDR	Genes
Ribosome	$6.6x10^{-17}$	3.0	$3.1x10^{-14}$	$3.1x10^{-14}$	$1.4x10^{-13}$	RPL18, RPL17, RPL19, RPL14, RPL13, RPS27L, RPS2, RPS3, RPLP0, RPLP1, RPL10, RPL11, RPL12, MRPL32, MRPL33, MRPL2, MRPL1, MRPL4, MRPL3, MRPS5, MRPL9, MRPS2, RPS18, RPS16, MRPS18A, RPS13, UBA52, MRPS17, MRPS16, MRPS11, MRPS10, RPS15A, RPL36, RPL37, MRPL10, RPS27, RPL30, RPS28, MRPL15, RPL32, RPL6, RPL31, RPL34, RPL9, RPL8, RPL3, RPL5, RPL7A, RPL10A, RPS24, RPSA, RPL26, RPL27, MRPS21, RPS6, RPS8, RPL28, RPS7, RPL23, RPL18A, RPL21, RPL37A
Spliceosome	$5.5x10^{-10}$	2.5	$1.5x10^{-7}$	$7.7x10^{-8}$	$7.2x10^{-7}$	HNRNPA1L2, NCBP1, CHERP, SRSF10, TRA2B, U2AF2, U2SURP, TRA2A, CWC15, SNU13, SF3B6, XAB2, SF3B4, SF3B2, HNRNPA3, HNRNPM, SF3B1, TCERG1, PLRG1, USP39, DHX15, MAGOHB, HNRNPC, ACIN1, RBM25, PRPF40A, RBM22, SNRPA1, ALYREF, SF3A2, DDX5, HNRNPA1, SF3A1, RBMX, HNRNPU, PRPF6, SRSF2, EIF4A3, PPIE, SRSF5, PPIH, SRSF6, SLU7, SNRNP40, SNRPC, PUF60, PRPF38B, THOC1, PRPF38A, SNRPG
Proteasome	$6.1x10^{-8}$	3.4	$1.7x10^{-5}$	$5.7x10^{-6}$	$8.1x10^{-5}$	PSMA7, PSMB5, PSMA1, PSMB4, PSMD14, PSMB7, PSMD13, PSMC5, PSMC4, PSMD11, PSMA5, PSMC3, PSME2, PSMA4, PSMC2, PSMA3, POMP, PSMD2, PSMD3, PSME3, PSME4, PSMD7, PSMD8
Cell cycle	$1.5x10^{-7}$	2.3	$4.3x10^{-5}$	$1.1x10^{-5}$	$2.0x10^{-4}$	E2F3, FZR1, E2F4, PKMYT1, TTK, CHEK1, PTTG1, CDC16,ZBTB17, CDC45, ORC4, MYC, ORC1, CUL1, BUB3, TFDP1, STAG1, CDC7, CDK1, ANAPC2, ANAPC5, RBL1, ANAPC4, TP53, SKP2, MCM2, SKP1, MCM3, MCM4, CDC27, WEE1, CDC25A, MCM6, CDC25B, CCNB1, HDAC1, PLK1, PCNA, BUB1B, MDM2, ANAPC7, SMC1A, GADD45B
RNA transport	$8.6x10^{-7}$	2.0	$2.4x10^{-4}$	$4.9x10^{-5}$	$1.1x10^{-3}$	XPO1, NCBP1, RANGAP1, PNN, PRMT5, MAGOHB, NUP37, DDX20, ACIN1, TPR, TGS1, KPNB1, EIF2B4, EIF2B5, CLNS1A, NUP88, RAN, EIF2S3, NUP85, UBE2I, TACC3, EIF4G1, EIF4A3, EIF4G2, EIF4G3, AAAS, EIF4A2, EIF4A1, THOC6, NUP107, THOC1, NUP98, ELAC2, STRAP, PABPC4, NDC1, SUMO2, NUP214, EIF3B, EIF3G, EIF3E, NUP50, XPOT, GEMIN2, ALYREF, RNPS1, NXF1, CASC3, FXR1, EIF4B, SEC13, EIF4E2
Base excision repair	$6.6x10^{-6}$	3.4	$1.9x10^{-3}$	$3.1x10^{-4}$	$8.8x10^{-3}$	APEX2, LIG1, NEIL3, POLE, NEIL1, MBD4, XRCC1, SMUG1, POLD4, MPG, MUTYH, POLE2, POLD1, POLD2, TDG, PCNA, PARP2
DNA replication	$2.6x10^{-5}$	3.1	$7.4x10^{-3}$	$1.1x10^{-3}$	$3.5x10^{-2}$	LIG1, POLE, RNASEH1, POLA2, MCM2, MCM3, MCM4, RNASEH2C, MCM6, RFC5, POLD4, RFC1, POLE2, POLD1, PRIM2, POLD2,PCNA

Table S3.2 Primer sets for PCR amplification.

Point Mutations	Amino Acid(s)	Primers	Sequences
F147L/F149L	147,149	Forward	5’ – GGT CAT TCA AAG GGG CTT GGC CTT GTT CGT TTT ACG G – 3’
		Reverse	5’ – CCG TAA AAC GAA CAA GGC CAA GCC CCT TTG AAT GAC C – 3’
R151A	151	Forward	5’ – GGT TTG GCT TTG TTG CTT TTA CGG AAT ATG – 3’
		Reverse	5’ – CAT ATT CCG TAA AAG CAA CAA AGC CAA ACC – 3’
R151D	151	Forward	5’ – GGG GTT TGG CTT TGT TGA TTT TAC GGA ATA TG – 3’
		Reverse	5’ – CAT ATT CCG TAA AAT CAA CAA AGC CAA ACC CC – 3’

D247A	247	Forward	5’ – CTC TTT GTG GAG AGG CCT TGA TCA TTA AAG G – 3’
		Reverse	5’ – CCT TTA ATG ATC AAG GCC TCT CCA CAA AGA G – 3’
D247R	247	Forward	5’ – CTC TTT GTG GAG AGC GCT TGA TCA TTA AAG – 3’
		Reverse	5’ – CTT TAA TGA TCA AGC GCT CTC CAC AAA GAG – 3’
D247E	247	Forward	5’ – GTC TCT TTG TGG AGA GGA GTT GAT CAT TAA AGG AAT C – 3’
		Reverse	5’ – GAT TCC TTT AAT GAT CAA CTC CTC TCC ACA AAG AGA C – 3’
mPUM2	255, 256, 259	Forward	5’ – CAC AAA TTT GCC GCC GCT GTA GTA GCA AAG TGT GTT ACT C – 3’
		Reverse	5’ – GAG TAA CAC ACT TTG CTA CTA CAG CGG CGG CAA ATT TGT G – 3’

Table S3.3 Primer sets for site-directed mutagenesis.

PCR Product	Amino Acid(s)	Primers	Sequences
PP7	1-283	Forward	5’ – TCC AAA ACC ATC GTT CTT TCG G – 3’
		Reverse	5’ – TCC TCC TCC GCT TCC TCC ACT A – 3’
TDP43 N-terminus	1-106	Forward	5’ – ATG TCT GAA TAT ATT CGG GTA ACC G – 3’
		Reverse	5’ – TAA ATC GGA TGT TTT CTG GAC T – 3’
TDP43 C-terminus	263-414	Forward	5’ – AAG CAC AAT AGC AAT AGA CAG TTA G– 3’
		Reverse	5’ – TCC CCA GCC AGA AGA CTT AGA ATC C – 3’
TDP43(RRM1)	107-176	Forward	5’ – ATA GTG TTG GGT CTC CCA T– 3’
		Reverse	5’ – GCT TCT CAA AGG CTC ATC TT– 3’
TDP43(RRM2)	191-262	Forward	5’- CTT CCT AAT TCT AAG CAA AGC CAA G -3’
		Reverse	5’– AGG TTC GGC ATT GGA TAT ATG AAC GC – 3
TDP43(WT)	1-414	Forward	5’– GCT AGC GCC ACC ATG TCT GAA TAT ATT– 3’
		Reverse	5’– ACC GGT CCC AAA CCT CTA CCG TCC CA– 3’

Chapter 4

C9orf72 mutations enhance TDP43 cytoplasmic mislocalization and neuronal toxicity

4.1 Abstract

The majority of individuals with ALS and FTD exhibit neuronal cytoplasmic inclusions rich in the RNA binding protein, TDP43, and the most common mutation responsible for familial forms of both disorders consists of a hexanucleotide (G_4C_2) repeat expansion mutation in the first intron of chromosome 9 open reading frame 72, or *C9orf72*. Family members with this mutation may develop ALS, FTD, or both. TDP43 pathology is a characteristic feature of *C9orf72*-linked disease, and changes in TDP43 localization and levels are strongly predictive of neuron loss in ALS/FTD. Even so, the link between *C9orf72* expansions, TDP43 deposition, and neurodegeneration remains unclear. TDP43 binds thousands of RNA transcripts, particularly UG-rich sequences, and TDP43-dependent toxicity is closely tied to its ability to recognize RNA. Intramolecular interactions between TDP43's RNA binding domains, mediated by a salt bridge, are necessary for maintaining sequence specificity. Here, we show that TDP43 harboring mutations that disrupt the salt bridge recognize G_4C_2 oligonucleotides *in vitro*. Mutations that alter the salt bridge also enhance TDP43 mislocalization, but abrogate TDP43-dependent toxicity. In ongoing work, we are testing if TDP43 directly binds the *C9orf72* repeat expansion upon disruption of the salt bridge. These studies provide a potential

connection between *C9orf72* mutations and TDP43 pathology, and highlight previously uncharacterized properties of TDP43 that may serve as targets for therapy development.

4.2 Introduction

A hexanucleotide (GGGGCC) repeat expansion located in the first intron of *C9orf72* is the most common genetic cause of both ALS and FTD[1,2]. Unaffected individuals possess less than 30 repeats; expansions greater than 30 and up to 4000 units are linked to ALS and FTD[1-3]. Because the repeat expansion is located in a non-coding region, neurodegeneration may stem from one or more overlapping mechanisms. Briefly, first the expansion may disrupt *C9orf72* transcription, resulting in loss of function[4-10]. *C9orf72* knockout animals fail to show symptoms of ALS/FTD[11-13], however, arguing against this possibility. Second, repeat RNA may sequester essential RNA binding proteins within nuclear foci[14-16]. Although, these foci are typical of *C9orf72*-related ALS/FTD, they do not appear to affect cellular health. Third, mutant *C9orf72* transcripts undergo translation through repeat associated non-AUG (RAN) translation, generating six dipeptide repeat proteins (DPRs)[17-21]. These peptides can be detected in *C9orf72* expansion carriers, but are concentrated in brain regions that appear to be unaffected in ALS/FTD (i.e. cerebellum)[22,23]. Conversely, changes in TDP43 deposition and levels are strongly predictive of neuron loss in ALS/FTD[24]. Even so, the connection between mutant *C9orf72* transcripts, TDP43 deposition, and neurodegeneration remains fundamentally unclear.

TDP43 pathology is a characteristic feature in ALS and FTD, including *C9orf72*-linked disease[1]. TDP43 binds thousands of transcripts, particularly UG-rich sequences,

and TDP43-dependent toxicity is tightly tied to its ability to recognize RNA[25,26]. Intramolecular interactions between 2 RNA recognition motifs (RRM1 and RRM2)—mediated by a salt bridge between Arg151 (located in RRM1) and Asp247 (located in RRM2)—are necessary for maintaining specificity for UG sequences[27]. In Chapter 3, we demonstrated that engineered mutations disrupting TDP43's salt bridge reduce the affinity and sequence specificity of nucleic acid binding by TDP43. Moreover, these mutations dramatically abrogate TDP43-dependent toxicity. Our work demonstrates the significance of the salt bridge in sustaining RNA binding specificity, and TDP43's ability to bind to RNA is an important driver for TDP43-dependent toxicity. Here, we hypothesize that disruption of the salt bridge enables sequestration of TDP43 by mutant *C9orf72* transcripts, eventually leading to nuclear TDP43 clearance, cytoplasmic deposition of TDP43, and subsequent cell death. Consistent with this notion, here we show that salt-bridge disrupting TDP43 variants recognize a single G_4C_2 stretch of DNA oligonucleotides *in vitro*. Co-expression of the *C9orf72* repeat expansion and TDP43 variants enhances cytoplasmic TDP43 mislocalization and toxicity in primary neurons. Additional work is warranted to thoroughly evaluate the potential interaction between G_4C_2 repeats and mutations disrupting TDP43's RRM1-RRM2 salt bridge. Below I will discuss our preliminary work thus far.

4.3 Results

4.3.1 *C9orf72* hexanucleotide repeat expansions are selectively toxic to neurons

We first determined whether overexpression of the G_4C_2 repeat expansion is toxic to primary neurons. To do this, we generated EGFP-tagged constructs containing G_4C_2

repeats of pathogenic length (71, 66, or 69 units) (Fig. 4.1A). Repeats were placed in front of EGFP in each of the three reading frames relative to EGFP. Repeats in the +0, or native, reading frame correspond to poly-GA, and repeats in the +1 or +2 reading frames correspond to poly-GP or poly-GR, respectively. Although EGFP contains a start and stop codon, no AUG codon is present above the repeat (Fig. 4.1A). Rodent primary cortical neurons were transiently transfected with vectors encoding mApple, to visualize the cell body, and the EGFP-tagged G_4C_2 repeats. As a negative control, neurons were transfected with mApple and EGFP alone. By automated fluorescence microscopy[28,29], transfected neurons were imaged longitudinally at 24-hour intervals for 10 days. Using custom-written algorithms, individual neurons were identified and tracked to note the last time the cell was seen alive. Death was determined by cellular morphology or loss of fluorescence, sensitive measures of cell death in previous studies[28,30,31]. Using Cox proportional hazards analysis, differences in survival between populations of neurons were measured relative to a reference group and given a hazard ratio (HR). We observed differences in toxicity between each of the repeat-containing constructs compared to EGFP alone. Repeats in the +0 reading frame were not significantly toxic to neurons (HR 1.07, p= 0.21) (Fig. 4.1B). However, repeats in the +1 and +2 reading frames increased the cumulative risk of death compared to EGFP alone (HR 1.32 and 1.16, p= $2.45x10^{-7}$ and 0.008, respectively) (Fig. 4.1B). Interestingly, repeats in the GP reading frame relative to EGFP were the most toxic. Among the DPRs, multiple studies have shown that the arginine-rich dipeptides, GR and PR, are the most toxic DPRs[32-36], and GA peptides has some selective toxicity in neuronal cell lines and primary neurons[37-39]. However, GP peptides do not elicit toxicity. While we cannot discriminate between G_4C_2 RNA versus

individual DPRs with these constructs, our results suggest that there may be sequence context differences that cause toxicity; though, this remains to be determined. For subsequent overexpression studies, we used (G_4C_2)66-EGFP because it was the most toxic to neurons.

Next, we sought to determine whether our repeat-containing constructs undergo RAN translation to produce DPRs. To test this, primary neurons overexpressing mApple and (G_4C_2)66-EGFP were fixed 48h after transfection, immunostained with antibodies detecting each DPR (GA, GP, and GR) and imaged by automated microscopy (Fig. 4.1C). Both nuclear and cytoplasmic diffuse staining was observed with anti-GA antibodies in cells expressing the repeat expansion, but no signal was detected in EGFP alone expressing cells (Fig. 4.1C). Using antibodies against GP dipeptides, nuclear staining was detected in neurons expressing EGFP alone, suggesting non-specific interactions (Fig. 4.1C). However, neurons expressing the repeats presented a greater signal above background in the nucleus compared to EGFP alone, suggesting GP production via RAN translation. Neurons expressing (G_4C_2)66 resulted in a diffuse nuclear stain with antibodies against GR dipeptides (Fig. 4.1C). We observed similar patterns for each DPR in cells expressing (G_4C_2)69-EGFP or (G_4C_2)71-EGFP plasmids (data not shown). Taken together, these findings demonstrate that overexpression G_4C_2 repeats, of pathogenic length, are toxic to neurons, and these repeats are RAN translated, resulting in detectable DPRs.

4.3.2 Localization of endogenous TDP43 is unaffected by G_4C_2 repeats or DPRs

TDP43 mislocalization and aggregation are characteristic features of *C9orf72*-related disease[1], and changes in TDP43 localization and levels are strongly related to

neuron loss in ALS and FTD[28,29], but the connection between *C9orf72* mutations and TDP43 metabolism remains unclear. Therefore, we asked whether *C9orf72* RAN translation products disrupt TDP43 trafficking. To determine the subcellular localization of endogenous TDP43 in the presence of *C9orf72* RAN translation products, either fresh or aged synthetic (GA)3, (GP)3, or (GR)3 dipeptides were applied to primary neurons. Cells were also treated with buffer (buffer in which peptides were solubilized) alone as a control. Twenty-four hours after peptide application, cells were immunostained with antibodies against MAP2, a pan-neuronal cytoplasmic marker, and TDP43 to detect endogenous TDP43. To quantitatively assess TDP43 localization, the abundance of TDP43 was measured separately within the nuclear and cytoplasmic compartments of individual neurons, and a nuclear-cytoplasmic ratio was calculated to reflect TDP43 distribution in each neuron. We observed no significant change of endogenous TDP43 localization when treated with dipeptides compared to buffer control (Fig. 4.2A). It is possible that (GN)3 dipeptides are not efficiently internalized by neurons, and thus, incapable of altering TDP43 localization effectively. Alternatively, perhaps mislocalization of TDP43 by the DPRs is dependent on TDP43 levels. Therefore, we next overexpressed EGFP-tagged TDP43(WT) and mApple, a cytoplasmic marker, and treated cells with either fresh or aged synthetic (GA)3, (GP)3, or (GR)3 dipeptides and assessed the localization of exogenous TDP43. Here, we detected a significant reduction in the nuclear-cytoplasmic ratio of TDP43 among cells treated with fresh or aged (GA)3 dipeptides, but not (GP)3 or (GR)3 dipeptides. This suggests that short GA dipeptides are sufficient to induce mislocalization of exogenous TDP43 to the cytoplasm.

4.3.3 Disruption of TDP43's RRM1-RRM2 salt bridge enables TDP43 to bind to G_4C_2 oligonucleotides *in vitro*

In Chapter 3 we engineered mutations that disrupt an intramolecular salt bridge between RRM1-RRM2 of TDP43 to investigate the RNA binding properties of TDP43, and how these characteristics dictate its vulnerability for toxicity in primary neurons. TDP43 variants bearing mutations that disrupt the salt bridge exhibit a reduction in affinity and sequence-specificity of nucleic acid recognition by TDP43 (Fig. 3.1). Moreover, these TDP43 variants impair binding to its native substrates (Fig. 3.1). From these observations, we next sought to determine whether TDP43 is capable of binding to G_4C_2 oligonucleotides *in vitro*. Using purified recombinant TDP43(WT), TDP43(R151A), and TDP43(D247A), we tested the ability of each variant to bind G_4C_2 DNA via electromobility shift assays (EMSAs). We first tested the ability of each TDP43 variant to recognize oligonucleotides of a single G_4C_2 unit, $(G_4C_2)1$. At increasing protein concentrations, we observed clear shifts for both TDP43(R151A) and TDP43(D247A), but not TDP43(WT), in the presence of $(G_4C_2)1$ oligonucleotides (Fig. 4.3A), suggesting that the salt bridge mutants display aberrant sequence recognition *in vitro*.

We also asked whether the salt bridge-disrupting mutations are capable of recognizing G_4C_2 expanded repeats. Next we tested the ability of each TDP43 variant to recognize six G_4C_2 units, $(G_4C_2)6$. Unlike the $(G_4C_2)1$ oligonucleotides, we detected a faint shift for both TDP43(R151A) and TDP43(D247A), but not TDP43(WT), induced by $(G_4C_2)6$ oligomers (Fig 4.3B). Notably, we observed a strong signal in the wells of the gel for all conditions, including the negative control ($(G_4C_2)6$ only). This observation suggests the possibility of $(G_4C_2)6$ sequences forming secondary structures, hindering its

ability to migrate completely through the acrylamide gel. Supporting this notion, previous investigations demonstrated that short $(G_4C_2)_n$ RNA and DNA (2-8 units) are capable of forming G-quadruplex structures *in vitro*[16,40-42], and the mobility of these oligomers is decreased in the presences of potassium ions, suggesting that the formation of G-quadruplexes is K^+ dependent[16,40]. Our electromobility shift assays were performed in presence of KCl, suggesting that our $(G_4C_2)6$ DNA oligomers are likely forming secondary structures and impairing its mobility. Therefore, it is possible that TDP43 variants are capable of binding to $(G_4C_2)6$ oligomers *in vitro,* and these effects may be observed under conditions less favorable for secondary structure formation by G_4C_2 repeats. Further investigation is required to assess the interaction between salt bridge-disrupting TDP43 variants and expanded G_4C_2 repeats.

4.3.4 Neuronal toxicity and cytoplasmic mislocalization of TDP43 are enhanced upon co-expression of $(G_4C_2)66$ and TDP43 variants.

Because we identified a potential interaction between G_4C_2 expanded repeats and TDP43 salt bridge-disrupting variants *in vitro*, we sought to determine the effects on neuronal survival upon overexpression of the G_4C_2 repeat expansion and TDP43 variants. Primary neurons were co-transfected with $(G_4C_2)66$-EGFP or EGFP alone and mApple-tagged versions of TDP43(WT, R151A, or D247A) or mApple alone. As a negative control, neurons were co-transfected with EGFP and mApple. By longitudinal fluorescence microscopy, transfected cells were assessed for survival. Expression of $(G_4C_2)66$ alone modestly increased the risk of death compared to EGFP-expressing cells (Fig. 4.4A, B; HR 1.25, p= 0.0003). As noted in previous investigations[28-30], overexpression of TDP43(WT) significantly increased the risk of death compared to

mApple alone (HR 2.58, $p < 2x10^{-16}$). However, co-expression of TDP43(WT) and $(G_4C_2)66$ dramatically enhanced toxicity based on expression of either alone, suggesting a synergistic effect from the accumulation of TDP43 (Fig. 4.4A, B; HR 4.2, $p < 2x10^{-16}$). Similar to our prior studies investigating neuronal survival upon overexpression of TDP43(R151A) or TDP43(D247A) (Fig. 3.2B), both variants significantly abrogated toxicity compared to TDP43(WT)-expressing cells (Fig. 4.4A, B; HR 0.35 and 0.66, respectively, $p < 2x10^{-16}$ for both comparisons). Cells expressing both TDP43(R151A) and $(G_4C_2)66$ displayed a significant increase in the cumulative risk of death compared to cells expressing TDP43(R151A) and EGFP (HR 1.39, $p= 1.24x10^{-7}$) (Fig 4.4A). In addition, cells expressing both TDP43(D247A) and $(G_4C_2)66$ significantly enhanced toxicity compared to cells expressing TDP43(D247A) and EGFP (Fig. 4.4B; HR 2.68, $p < 2x10^{-16}$). Among the salt bridge-disrupting TDP43 mutants, the effect of the repeat expansion was additive for TDP43(R151A) but synergistic for TDP43(D247A), suggesting differences in the interaction between mutant TDP43 and expanded repeats. In Chapter 3, we observed differences in toxicity upon expression of TDP43(R151A) and TDP43(D247A), where the latter TDP43 variant was more toxic. The differences in toxicity may be due to the particular residue that is altered: any manipulation to R151 resulted in minimal toxicity, whereas manipulations to D247 varied proportionally, depending on the strength of the interaction with R151. Therefore, R151 seems to have a pivotal role in TDP43-related toxicity. Taken together, these data suggest a potential interaction between TDP43 and G_4C_2 repeats with consequences for neuronal survival. Additional studies are necessary to confirm and extend these findings.

Our studies involving synthetic RAN peptides suggested that TDP43 localization is mostly unaffected by exogenous peptide application, with the exception of cells overexpressing TDP43(WT) in the presence of (GA)3 peptides (Fig. 4.2). While the application of synthetic peptides to neurons allows for the assessment on the effect of each unique DPR, it does not accurately reflect disease since all RAN proteins are likely generated within the same cell. Moreover, it excludes any effects that G_4C_2 expanded RNA may have within the cell. It is likely that either transcribed G_4C_2 expanded RNA or RAN translation, or both, contribute to disease pathogenesis. Therefore, we asked whether expression of the G_4C_2 repeat expansion affects TDP43 localization. Primary neurons were co-transfected with $(G_4C_2)66$-EGFP or EGFP alone and mApple-tagged versions of TDP43(WT, R151A, or D247A) or mApple alone. The subcellular localization of TDP43 was determined by fluorescence microscopy, measuring mApple fluorescence in the nucleus and cytoplasm separately to generate a nuclear-cytoplasmic ratio. Similar to our previous study (Fig. S3.7), both TDP43(R151A) and TDP43(D247A) reduce the nuclear-cytoplasmic ratio compared to TDP43(WT)-expressing cells (Fig. 4.4C). However, cytoplasmic mislocalization of TDP43 was further enhanced upon overexpression of $(G_4C_2)66$ among all TDP43 variants (Fig. 4.4C). These data suggest that expression of *C9orf72* RNA or peptides produced by RAN translation, or both, affect TDP43 localization.

4.4 Discussion and future directions

Here, we started to investigate whether there is a connection between *C9orf72* and TDP43 and disease pathogenesis. Very little is known as to whether these two common

features of disease converge to similar molecular pathways. Our preliminary data suggest that salt bridge-disrupting mutations (R151A or D247A) are capable of recognizing G_4C_2 DNA sequences *in vitro*, unlike TDP43(WT) (Fig. 4.3). These results are consistent with our previous observations that both TDP43(R151A) and TDP43(D247A) recognize sequences other than $(UG)_{12}/(TG)_{12}$ *in vitro* (Figs. 3.1 and S3.2). Furthermore, these observations confirm that the RRM1-RRM2 salt bridge is critical for maintaining sequence specificity of TDP43 for TG or UG-rich sequences, as previously identified[27,43]. While our EMSAs revealed a clear shift, indicating binding, of TDP43(R151A) and TDP43(D247A) with a single stretch of G_4C_2 DNA, it was not as obvious with multiple units of G_4C_2 DNA oligomers (Fig. 4.3). One possible explanation for the unclear results with $(G_4C_2)6$ DNA oligonucleotides is that the repeats have the propensity to form secondary structures, impairing its ability to migrate through the acrylamide gel. Supporting this notion, prior studies demonstrated the formation of highly stable G-quadruplexes of short $(G_4C_2)_n$ DNA or RNA sequences[16,40-42], and these secondary structures migrate slowly through acrylamide gels[16,40]. Furthermore, G-quadruplexes are stabilized by potassium or sodium ions but destabilized in the presence of lithium ions[44,45]; our assays were performed in the presence of potassium ions (Fig. 4.3). In the presence of lithium, we did not detect binding between $(G_4C_2)6$ DNA and TDP43 variants (data not shown). This observation may suggest that the secondary structure of G_4C_2 repeats could be important for nucleic acid recognition by TDP43. Nonetheless, further investigations are required in order to draw firm conclusions.

As part of our investigation to test the interaction between G_4C_2 repeats and TDP43, we started a collaboration with Dr. Christopher Pearson's group from the

Hospital for Sick Children to test if TDP43 variants directly interact with $(G_4C_2)_n$ or $(C_4G_2)_n$ RNA or DNA oligonucleotides via EMSAs. Preliminary evidence suggests that TDP43(WT) binds G-rich DNA and RNA and C-rich RNA *in vitro*, and that TDP43's structure changes in the presence of G-rich RNA and DNA, determined by circular dichroism (CD) spectra (data not shown). Additionally, we are interested in determining the CD spectra of TDP43(R151A) and TDP43(D247A) in the presence of G-rich RNA and DNA. Another important step to take is to determine whether an interaction between TDP43 variants and G_4C_2 repeats occurs in cells. One prior study suggests that TDP43 recognizes and transports G-quadruplex-forming, including $(G_4C_2)_n$, RNAs in cells[46]. To test whether an interaction occurs between TDP43 variants and G_4C_2 repeats, we are interested in performing pull-down assays upon co-expression of TDP43 variants and G_4C_2 expanded repeats in HEK293 cells.

We observed a synergistic effect in neurons upon overexpression of TDP43(WT) or TDP43(D247A) in the presence of expanded G_4C_2 repeats (Fig. 4.4B). Notably, the synergistic effect observed in TDP43(D247A) and $(G_4C_2)66$ co-expressing cells did not fully mimic the effect seen in TDP43(WT) and $((G_4C_2)66$ co-expressing cells (Fig. 4.4B). Additionally, the effects on survival observed in neurons co-expressing TDP43(R151A) and $(G_4C_2)66$ was minimal (Fig. 4.4A). The differences in TDP43-dependent toxicity we observed in the presence of $(G_4C_2)66$ followed a similar trend based on our previous studies, in which cells expressing TDP43(WT) was the most toxic, intermediate for TDP43(D247A), and minimal for TDP43(R151A) (Fig. 3.2B). Taken together, these data imply a potential difference in $(G_4C_2)_n$ RNA substrate recognition by TDP43 variants, if such a direct interaction is occurring in neurons. However, to confirm this notion,

additional studies are necessary. Mutating two key Phe residues (F147L/F149L) within RRM1 is sufficient to eliminate RNA binding by TDP43[47,48], and similar mutations rescue TDP43-dependent toxicity in yeast and *Drosophila*[49,50]. To determine whether G_4C_2 repeats mediate TDP43-related toxicity upon binding, we are interested in assessing neuronal survival upon co-expressing TDP43(F147L/F149L) and $(G_4C_2)66$. If G_4C_2 RNA mediates TDP43-related toxicity through a direct interaction, then we expect to not observe neuronal toxicity in these conditions.

We hypothesize that disruption of the salt bridge enables sequestration of TDP43 by mutant *C9orf72* transcripts, eventually leading to nuclear TDP43 clearance, cytoplasmic deposition of TDP43, and subsequent cell death. Consistent with this notion, our preliminary data suggest that expression of expanded G_4C_2 RNA enhances cytoplasmic mislocalization of not only the salt bridge-disrupting TDP43 mutants but also TDP43(WT) (Fig. 4.4C). To confirm the relationship between G_4C_2 expanded RNA and TDP43 mislocalization, the next step is to assess the localization of TDP43(F147L/F149L), which forms large spherical nuclear puncta (Fig. S3.7), upon expression of $(G_4C_2)66$. If TDP43 and G_4C_2 expanded RNA interact directly, then no change in the subcellular distribution should occur, as the F147L/F149L double mutant abrogates TDP43's ability to bind RNA[47]. To determine whether *C9orf72* repeat expansions affect the accumulation of TDP43, we can also assess the solubility and stability of each TDP43 variant in the presence of G_4C_2 repeats. If mutant *C9orf72* RNA interacts with TDP43 variants, then we would expect a decrease in the solubility of TDP43, since the repeat RNA forms nuclear foci and cytoplasmic granules. To assess for protein stability of each TDP43 variant in the presence of *C9orf72* mutations, we can co-

express Dendra2-tagged TDP43 variants and EGFP-tagged G_4C_2 repeats in primary neurons, and measure protein half-life by optical pulse labeling and automated microscopy. If *C9orf72* mutations disrupt TDP43 turnover, then the half-life for each TDP43 variant would increase, resulting in the accumulation of TDP43.

While our (G_4C_2)66-EGFP reporter is capable of producing RAN dipeptides (Fig. 4.1C), it is still unclear whether the G_4C_2 RNA or RAN translation products are driving neuronal toxicity and mislocalization of TDP43. With the exception of synthetic (GA)3 dipeptides upon overexpression of TDP43(WT), we did not detect significant differences of TDP43 localization (Fig. 4.2). Whether these synthetic dipeptides affect the localization of TDP43(R151A) or TDP43(D247A) are unknown. Expressing constructs that use alternative codons for expression of individual DPRs, resulting in the exclusion of G_4C_2 RNA, could offer insight as to whether G_4C_2 expanded RNA or the DPRs, contribute to neuronal toxicity and TDP43 mislocalization. Furthermore, overexpressing the G_4C_2 repeat expansion containing a stop codon in front of the repeat, thereby inhibiting RAN translation, could also shed light as to whether production of the DPRs contribute to TDP43-related toxicity. These additional studies are warranted to determine their respective contributions of TDP43-dependent toxicity.

Based on our preliminary observations, we are interested in pursuing a thorough investigation to determine whether *C9orf72* mutant transcripts and TDP43 deposition converge upon similar molecular pathways that contribute to neurodegeneration. The studies described here have the potential to uncover and highlight critical mechanisms responsible for neurodegeneration in *C9orf72*-related ALS/FTD that may serve as targets for therapy development.

4.5 Materials and methods

Plasmids

The plasmids pGW1-EGFP, pGW1-mApple, and pGW1-TDP43(WT)-mApple were generated as described previously[28,29]. Primers used for PCR amplification or site-directed mutagenesis (SDM) were ordered from Integrated DNA Technologies (IDT). Mutant TDP43 was created from pGW1-TDP43(WT)-mApple by site-directed mutagenesis using the Pfu Ultra high-fidelity polymerase (Agilent Technologies) according to the manufacturer's protocol.

To produce EGFP-2A-mApple, EGFP-2A was PCR amplified using PrimeStar GXL DNA polymerase (Takara) from pGW1-EGFP with the following primers: 5' – ATA TAA GCT TGC CAC CAT GGT GAG CAA GGG CGA GGA GCT – 3' and 5'– AAG CTT TAG GGC CGG GAT TCT CCT CCA CGT CAC CTG CTT GTT TGA GTA GTG AGA AGT TTG TTG CTC CAG ATC CC TTG TAC AGC TCG TCC ATG C – 3'. Note that the reverse primer contains the 2A peptide sequence (5'– GGA TCT GGA GCA ACA AAC TTC TCA CTA CTC AAA CAA GCA GGT GAC GTG GAG GAG AAT CCC GGC CCT. The 783 bp product was digested with HindIII and ligated into pGW1-mApple. To generate pGW1-EGFP-2A-TDP43(WT)-mApple, EGFP-2A was PCR amplified as described above; the PCR product was cut with HindIII and ligated into pGW1-TDP43(WT)-mApple. The following primers were used to generate pGW1-EGFP-2A-TDP43(R151A)-mApple via SDM: 5' – GGT TTG GCT TTG TTG CTT TTA CGG AAT ATG – 3' and 5' – CAT ATT CCG TAA AAG CAA CAA AGC CAA ACC – 3'. The following primers were used to generate pGW1-EGFP-2A-TDP43(D247A)-

mApple by SDM: 5' – CTC TTT GTG GAG AGG CCT TGA TCA TTA AAG G – 3' and 5' – CCT TTA ATG ATC AAG GCC TCT CCA CAA AGA G – 3'.

To generate each construct containing the hexanucleotide repeat expansion of *C9orf72* (G_4C_2) tagged with EGFP in pGW1, the $(G_4C_2)_n$-EGFP fragment was subcloned from pcDNA3.1-$(G_4C_2)_n$-EGFP (provided by Dr. Peter Todd). Both 71 and 66 repeats and its EGFP tag were cut with the restriction enzymes ApaI and NheI and inserted into the pGW1 vector. (G_4C_2)69-EGFP was digested with EcoRI and NheI and ligated into pGW1. Note that a single C to A mutation at repeat 13 (GGGCA) is present within the repeat sequence.

Electrophoretic mobility shift assays

Purification of full-length recombinant TDP43(WT), TDP43(R151A), and TDP43(D247A) are described in Chapter 3. Binding assays were performed as described in Chapter 3. Briefly, binding reactions were carried out in binding buffer (12.5 mM HEPES, pH 7.8, 50 mM KCl, 2.5 mM MgCl2, 0.5 mM TCEP, 25 μg/mL BSA, 0.01% NP-40) with 50% glycerol, 1μg/μl poly-dIdC, recombinant TDP43 variants and ssDNA labeled probes tagged with a 5' 700nm infrared (IR) moiety (purchased from IDT). Concentrations of protein and ssDNA are indicated in figure legends. Each reaction was incubated on ice for 5 min followed by 25 min at room temperature (RT). 6% acrylamide gels were performed at 100 V. The LI-COR Odyssey platform was used to image the gels.

Primary neuron culture and transfection

Mixed primary cortical neurons were harvested from embryonic day 19-20 Long-Evans rat pups and cultured at 0.6 x 10^6 cells/mL in 96 well cell culture plates (TPP), as

previously described[28,30]. On day *in vitro* 3 or 4, neurons were transfected with 0.2 μg DNA and 0.5 μL Lipofectamine 2000 (ThermoFisher) per well. Cells were rinsed with media containing Hoescht 33258 Dye (1:5000, Invitrogen) to label nuclei. For experiments that involved synthetic peptides, "fresh" peptides were added directly to cells immediately after dissolving in solution (0.1 M NaCl, 25 mM sodium phosphate pH 7.4) while "aged" peptides were added following 6-8 days of agitation as previously described[51]. All peptides were purchased from GenScript and applied to cells at a final concentration of 1mM.

Longitudinal fluorescence microscopy

Automated longitudinal fluorescence microscopy started 24 hours post-transfection, as previously described[29-31,52]. Briefly, imaging was accomplished by an inverted Nikon Ti microscope with a 20x objective lens, a Lambda XL Xenon lamp (Sutter) with 5 mm liquid light guide (Sutter), a PerfectFocus system, and either an Andor iXon3 897 EMCCD camera or Andor Zyla4.2 (+) sCMOS camera. Custom code was written in publically available software (μManager, ImageJ) to control all stage, shutter, and filter wheel movements.

Post-imaging processing, analyses for survival, and fluorescent intensity measurements were achieved with custom scripts written in Python or ImageJ macro language. Morphology, size, and fluorescence intensity were used to label neurons. Loss of fluorescence, degenerating processes, and rounding of the soma were used as indicators to identify individual neurons as dead.

Immunocytochemistry

Forty-eight hours after transfection, primary neurons were rinsed twice in PBS (Life Technologies) and fixed in 4% paraformaldehyde for 10 min at RT. Following two rinses in PBS, neurons were permeabilized with 0.1% Triton-X-100 (Bio-Rad Laboratories) for 20 min at RT. Following equilibration with 10 mM glycine in PBS for 10 min at RT, samples were blocked in 0.1% Triton-X-100, 3% BSA (Research Products International) and 0.2% goat serum in PBS for 1 hour at RT. Primary antibodies against poly-GA (1:100) and poly-GR (1:500), both previously described[51], or poly-GP (1:5000, rabbit polyclonal, Millipore) were added to the samples in block and incubated overnight at 4°C. Antibodies against TDP43 (1:1000, rabbit polyclonal, Proteintech) were added to samples to detect endogenous TDP43. Samples were rinsed 3 times for 5 min with PBS. Goat anti-mouse Cy5 secondary antibody (1:250, Jackson ImmunoResearch) was then placed in block for 1 hr at RT. Samples were then washed 5 times for 5 min with PBS, then twice more containing Hoescht 33258 Dye (Invitrogen) at 1:5000. Cells were rinsed twice for 5 min with PBS before imaging. Samples were assessed for localization by automated custom-written algorithms.

For assessing localization of endogenous TDP43 in the presence of synthetic DPRs by immunocytochemistry, primary neurons were treated with fresh or aged dipeptides on day *in vitro* 4. Twenty-four hours after peptide application, cells were fixed and immunostained with an antibody against the pan-neuronal cytoplasmic marker, MAP2 (1:500, mouse monoclonal, Millipore) and an antibody against TDP43 (1:1000, rabbit polyclonal, Proteintech). Samples were imaged by fluorescence microscopy and assessed manually using ImageJ.

Statistical analysis

Statistical analyses were accomplished in either GraphPad Prism or R. For survival analyses, differences among cell populations were determined via Cox proportional hazards analysis in the publically available R survival package. Statistical differences among nuclear/cytoplasmic ratios were determined by one-way ANOVA with Tukey's post hoc test in GraphPad Prism.

4.6 Acknowledgements

I would like to thank members of the Barmada lab for helpful comments and technical advice. I thank Dr. Peter Todd for providing reagents.

References

1. Buratti, E. & Baralle, F. E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276,** 36337–36343 (2001).
2. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **72,** 245–256 (2011).
3. Ayala, Y. M. *et al.* TDP-43 regulates its mRNA levels through a negative feedback loop. *EMBO J* **30,** 277–288 (2011).
4. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
5. Tollervey, J. R. *et al.* Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nat Neurosci* **14,** 452–458 (2011).
6. Polymenidou, M. *et al.* Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat Neurosci* **14,** 459–468 (2011).
7. Gijselinck, I. *et al.* A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum:a gene identification study. *Lancet Neurol* **11,** 54–65 (2011).
8. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struc. Mol. Biol.* **20,** 1443–1449 (2013).
9. Xi, Z. *et al.* Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *Am J Hum Genet* **92,** 981–989 (2013).
10. Ciura, S. *et al.* Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. *Ann Neurol.* **74,** 180–187 (2013).
11. Farg, M. A. *et al.* C9ORF72, implicated in amyotrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum. Mol. Genet.* **23,** 3579–3595 (2014).
12. Sellier, C. *et al.* Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. *EMBO J* **35,** 1276–1297 (2016).
13. Webster, C. P. *et al.* The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *EMBO J* **35,** 1656–1676 (2016).
14. Sullivan, P. M. *et al.* The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. *Acta Neuropathol Commun* 1–16 (2016). doi:10.1186/s40478-016-0324-5
15. Yang, M. *et al.* A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. *Sci Adv* **2,** e1601167–e1601167 (2016).
16. Barmada, S. J. *et al.* Autophagy induction enhances TDP43 turnover and survival in neuronal ALS models. *Nat. Chem. Bio.* **10,** 677–685 (2014).
17. Barmada, S. J. *et al.* Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 7821–7826 (2015).
18. O'Rourke, J. G. *et al.* C9orf72 is required for proper macrophage and microglial function in mice. *Science* **351,** 1324–1329 (2016).
19. Barmada, S. J. *et al.* Cytoplasmic mislocalization of TDP-43 is toxic to neurons

and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J. Neurosci* **30,** 639–649 (2010).
20. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R. & Finkbeiner, S. Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431,** 805–810 (2004).
21. Cooper-Knock, J. *et al.* Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain* **137,** 2040–2051 (2014).
22. Donnelly, C. J. *et al.* RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. *Neuron* **80,** 415–428 (2013).
23. Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507,** 195–200 (2015).
24. Yamakawa, M. *et al.* Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum. Mol. Genet.* **24,** 1630–1645 (2014).
25. Ash, P. E. A. *et al.* Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to c9FTD/ALS. *Neuron* 1–8 (2013). doi:10.1016/j.neuron.2013.02.004
26. Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345,** 1139–1145 (2014).
27. Gendron, T. F. *et al.* Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathologica* **126,** 829–844 (2013).
28. Wen, X. *et al.* Antisense Proline-Arginine RAN Dipeptides Linkedto C9ORF72-ALS/FTD Form Toxic Nuclear Aggregates that Initiate In Vitro and In Vivo Neuronal Death. *Neuron* **84,** 1213–1225 (2014).
29. Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc. Natl. Acad. Sci. U.S.A.* **110,** E4968–77 (2013).
30. Tao, Z. *et al.* Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum. Mol. Genet.* **24,** 2426–2441 (2015).
31. Mori, K. *et al.* Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathologica* **126,** 881–893 (2013).
32. Mizielinska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345,** 1192–1194 (2014).
33. Mori, K. *et al.* The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science* **339,** 1335–1338 (2013).
34. Chang, Y.-J., Jeng, U.-S., Chiang, Y.-L., Hwang, I.-S. & Chen, Y.-R. The Glycine-Alanine Dipeptide Repeat from C9orf72 Hexanucleotide Expansions Forms Toxic Amyloids Possessing Cell-to-Cell Transmission Properties. *J. Biol. Chem.* **291,** 4903–4911 (2016).
35. Al-Sarraj, S. *et al.* p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. *Acta Neuropathologica* **122,** 691–702 (2011).
36. May, S. *et al.* C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins

cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathologica* **128,** 485–503 (2014).
37. Mann, D. M. *et al.* Dipeptide repeat proteins are present in the p62 positive inclusions in patients with frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathol Commun* **1,** 68 (2013).
38. Zhang, Y.-J. *et al.* Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathologica* **128,** 505–524 (2014).
39. Mackenzie, I. R. A. *et al.* Quantitative analysis and clinico-pathological correlationsof different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathologica* **130,** 845–861 (2015).
40. Reddy, K., Zamiri, B., Stanley, S. Y. R., Macgregor, R. B., Jr. & Pearson, C. E. The Disease-associated r(GGGGCC) nRepeat from the C9orf72Gene Forms Tract Length-dependent Uni- and Multimolecular RNA G-quadruplex Structures. *J. Biol. Chem.* **288,** 9860–9866 (2013).
41. Fratta, P. *et al.* C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2,** 707 (2012).
42. Zamiri, B., Mirceta, M., Bomsztyk, K., Macgregor, R. B., Jr. & Pearson, C. E. Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)8•(GGCCCC)8 repeat: effect of CpG methylation. *Nucleic Acids Res.* **43,** 10055–10064 (2015).
43. Furukawa, Y. *et al.* A molecular mechanism realizing sequence-specific recognition of nucleic acids by TDP-43. *Sci Rep* **6,** 20576 (2016).
44. Campbell, N. H. & Neidle, S. Interplay between Metal Ions and Nucleic Acids. *Met Ions Life Sci* **10,** 119–134 (2012).
45. Neidle, S. The structures of quadruplex nucleic acids and their drug complexes. *Curr Opin Struct Biol* **19,** 239–250 (2009).
46. Ishiguro, A., Kimura, N., Watanabe, Y., Watanabe, S. & Ishihama, A. TDP-43 binds and transports G-quadruplex-containing mRNAs into neurites for local translation. *Genes Cells* **21,** 466–481 (2016).
47. Elden, A. C. *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466,** 1069–1075 (2010).
48. Voigt, A. *et al.* TDP-43-mediated neuron loss in vivo requires RNA-binding activity. *PLoS ONE* **5,** e12247 (2010).
49. Atanasio, A. *et al.* C9orf72 ablation causes immune dysregulation characterized by leukocyte expansion, autoantibody production, and glomerulonephropathy in mice. *Sci Rep* 1–14 (2016). doi:10.1038/srep23204
50. Koppers, M. *et al.* C9orf72 ablation in mice does not cause motor neuron degeneration or motor deficits. *Ann Neurol.* **78,** 426–438 (2015).
51. Flores, B. N. *et al.* Distinct C9orf72-Associated Dipeptide Repeat Structures Correlate with Neuronal Toxicity. *PLoS ONE* **11,** e0165084 (2016).
52. Tsvetkov, A. S. *et al.* Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration. *Nat. Chem. Bio.* **9,** 586–592 (2013).

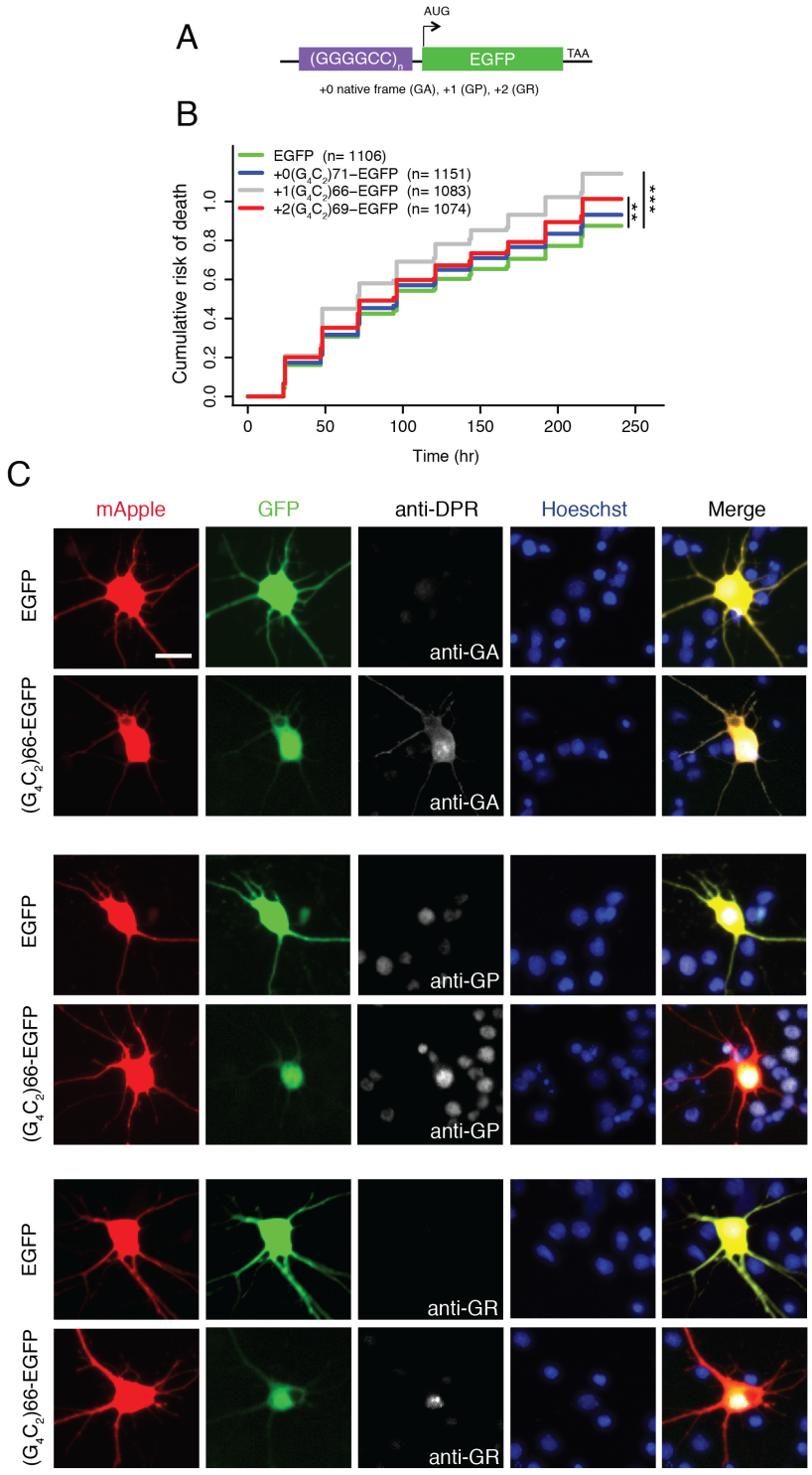

Figure 4.1 Overexpressing *C9orf72* expanded repeats are toxic to neurons. (A) Diagram of EGFP-tagged expanded repeats. Constructs contained 71, 66, or 69 repeats. Repeats were placed in front of EGFP in separate reading frames relative to EGFP. **(B)** Overexpressing *C9orf72* expanded repeats in the GP or GR reading frames significantly increased the cumulative risk of death compared to neurons expressing EGFP (HR 1.32,

1.16, respectively). n, number of neurons. ** $p < 0.008$, *** $p < 2.45x10^{-7}$, Cox proportional hazards. Survival analyses were pooled from 3 independent experiments, with 8 wells per condition for each replicate. **(C)** Fluorescence microscopy of primary cortical neurons overexpressing mApple (RFP) and either EGFP or $(G_4C_2)66$-EGFP and stained for each dipeptide repeat protein (GA, GP, or GR). Both GA and GR are detected in $(G_4C_2)66$-EGFP expressing cells, but not in EGFP expressing cells. GP is detected in both cells expressing $(G_4C_2)66$-EGFP or EGFP alone. However, signal above background is observed in $(G_4C_2)66$-EGFP expressing cells. Nuclei were stained with Hoechst 33258. Scale bar: 20 μm.

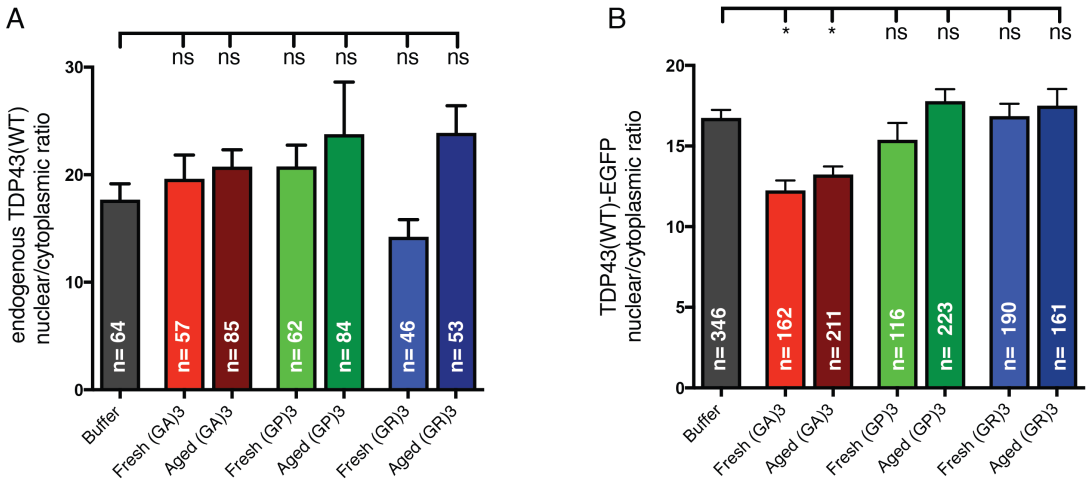

Figure 4.2 C9*orf72* RAN translation products have little effect on TDP43(WT) localization. (A) Neither fresh nor aged (GA)3, (GP)3, or (GR)3 peptides[51] altered endogenous TDP43 localization when applied to primary neurons, compared to buffer control. **(B)** Applying either fresh or aged (GA)3 peptides, but not (GP)3 or (GR)3, significantly reduced the nuclear-cytoplasmic ratio of TDP43(WT)-EGFP expressing neurons. n, number of neurons. ns, not significant. Data in **A** represents one experiment. Data in **B** represent 3 independent experiments, *$p < 0.001$, one-way ANOVA with Dunnett's post-hoc. Plots in **A** and **B** show mean ± SEM; n, number of neurons.

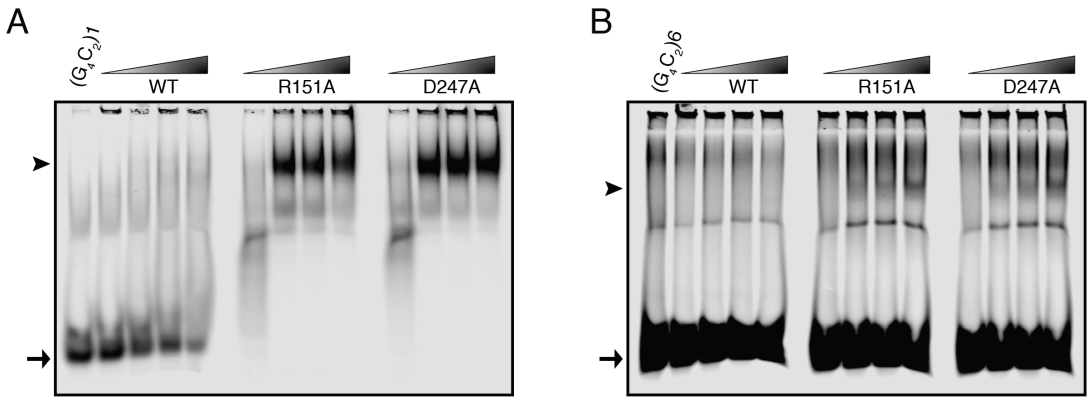

Figure 4.3 TDP43 binds to G_4C_2 oligonucleotide upon disruption of its R151-D247 salt bridge. (A-B) Electromobility shift assay (EMSA) of recombinant TDP43(WT) and salt bridge disrupting mutants, TDP43(R151A) or TDP43(D247A), at increasing concentrations (0.5 to 16 pmol) were mixed with either 5 nM of labeled $(G_4C_2)1$ oligonucleotides in **A** or $(G_4C_2)6$ in **B**. Among the conditions tested, a clear shift was observed for mutant TDP43 variants in **A**, but not **B**. For **A-B,** arrowheads indicate protein-DNA complexes, while arrows point to unbound oligonucleotides. Data in **A** represents 3 replicates and 1 replicate in **B**.

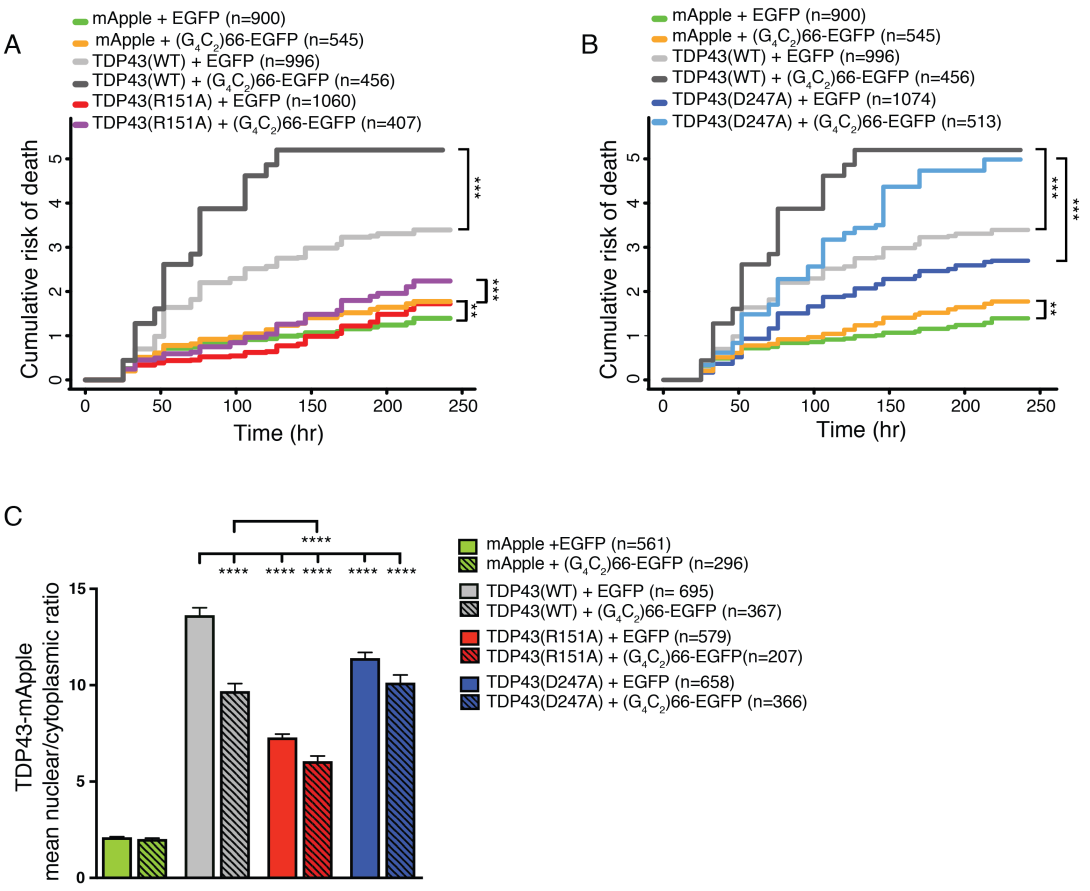

Figure 4.4 Co-expression of $(G_4C_2)66$ and TDP43 enhances neuronal toxicity and TDP43 mislocalization. (A-B) Primary cortical neurons were co-transfected with $(G_4C_2)66$-EGFP or mApple-tagged TDP43 variants, TDP43(R151A) or TDP43(D247A), constructs. Overexpressing mApple-tagged TDP43(WT) significantly increased the cumulative risk of death compared to neurons expressing mApple (HR 2.58, $p < 2x10^{-16}$). Mutations that disrupt the salt bridge (R151A in **A** or D247A in **B**) significantly reduce the risk of death compared to neurons expressing TDP43(WT) (HR 0.35 and 0.66, respectively, $p < 2x10^{-16}$, for both comparisons). A significant increase in the cumulative risk of death was observed in $(G_4C_2)66$-EGFP expressing cells compared to EGFP alone (HR 1.25, p= 0.0003). Co-expressing TDP43(WT) and the expanded repeats enhanced toxicity compared to mApple control (HR 4.2, $p < 2x10^{-16}$). **(A)** In the presence of the expanded repeats, the cumulative risk of death increased from HR 0.96 ($p < 2x10^{-16}$) to 1.39 ($p= 1.24x10^{-7}$) in cells expressing TDP43(R151A), compared to mApple and EGFP control. **(B)** The cumulative risk of death increased in neurons co-expressing TDP43(D247A) and $(G_4C_2)66$-EGFP from HR 1.74 to 2.68 ($p < 2x10^{-16}$ for both comparisons) compared to mApple and EFFP control. **(C)** Co-expression of mApple-tagged TDP43(WT) or salt bridge disrupting mutants with $(G_4C_2)66$-EGFP enhance TDP43 cytoplasmic mislocalization in neurons. For **A-B**, data were pooled from 2 independent experiments. For **C,** data were pooled from 2 replicates, **** $p < 0.0001$ one-way ANOVA with Tukey's test. For **A-C**, n, number of neurons.

Chapter 5

Discussion and future directions

The research described in this dissertation provides novel insight into the molecular mechanisms contributing to neurodegeneration in models of ALS and FTD. In Chapter 2, in collaboration with Dr. Ivanova, we synthesized short polymers corresponding to the three *C9orf72* RAN products derived from the sense strand, analyzed their structures by electron microscopy and assessed their relative toxicity when applied to rodent primary cortical neurons. We observed unique structural features for each dipeptide that correlates with their cellular internalization and relative toxicity. In Chapter 3, we demonstrated that the RNA binding properties of TDP43, mediated by the formation of a salt bridge between TDP43's RNA recognition motifs, are critical for maintaining TDP43's RNA binding properties, stability, localization, and toxicity in primary neurons. In addition, we identified select transcripts encoding oxidative phosphorylation and ribosome components affected by the accumulation of functional TDP43. In Chapter 4, I shed light into a potential connection between *C9orf72* mutations and TDP43 deposition that may uncover novel disease pathways eliciting neurodegeneration. Despite the significant contribution of the main findings of this thesis to the ALS and FTD field, many questions remain for future study. In this section I outline important steps to take moving forward from this work.

5.1 *C9orf72*: DPR vs. RNA-mediated toxicity

Work in this thesis has helped cement the toxic properties of different DPRs in a neuronal system. We systematically examined each of the sense-derived *C9orf72* RAN products to determine whether these are sufficient for neuronal toxicity by applying synthetic dipeptides to primary neurons and tracking their survival. Here, we showed that short peptides composed of 3 or 6 repeat unit lengths are capable of inducing toxicity when applied externally to neurons, and were the first to show that RAN peptides adopt abnormal and toxic conformations even at such short lengths. In various model systems, prior studies have used AUG-initiated constructs to drive DPR expression from engineered cDNA, including some that are independent of repeat-containing RNA[1-10]. However, it is still unknown how many repeating units constitute DPR pathological aggregates in *C9orf72*-mutation carriers. The fact that we observed differences in the structure and toxicity of the peptides between two short repeating-units emphasizes the importance of future studies investigating different DPR lengths.

The extent to which DPRs might spread through the disease brain is an intriguing avenue of exploration. We observed a relationship between peptide structures, their ability to enter cells, and their toxicity to neurons. Specifically, we observed a strong correlation between β-sheet content and toxicity for GA dipeptides, and cells internalized these peptides most frequently. Our experiments imply that RAN peptides may enable the spread or progression of disease at the cellular level. Similar to our studies, a previous investigation found that (GA)15 synthetic peptides formed amyloid-like fibrils, displayed cell-to-cell transmission properties, and were toxic to cells[11]. These data suggest that therapies capable of blocking RAN peptide internalization may prevent or slow disease

progression. Therefore, an important first step towards investigating this avenue is to determine the mechanism by which the RAN peptides are internalized by neurons. One study suggests that most DPRs are capable of spreading from one cell to another via exosomes upon overexpression of codon-optimized DPR constructs in cell culture[12]. Another study demonstrated the spread of poly-GA aggregates in neurons and antibodies targeting poly-GA inhibited cell-to-cell transmission[13]. Given that GA dipeptides appear to be the most abundant in *C9orf72*-mutation carriers[14], and one of the most toxic species in cells[3,6,9,15], determining non-cell autonomous mechanisms could reveal important information. To further highlight its relevance to physiologic disease conditions, a previous study demonstrated that the spread of DPRs from one cell to another is present in iPSC-derived motor neurons from *C9orf72*-mutation carriers[12]. Cell-to-cell transmission of toxic proteins, such as α-synuclein, and β-amyloid, have been reported for other neurodegenerative diseases[16,17]. Taken together, determining the factors and precise pathways facilitating peptide release and cellular uptake could be useful in preventing disease progression for *C9orf72*-ALS/FTD.

While many different model systems have been generated to study *C9orf72* mutations, no model has been reported to clearly delineate the relative contribution of DPR versus RNA-mediated toxicity. A number of mouse models have been generated to investigate the effects of the G_4C_2 repeat expansion. Generally, these models display key features of *C9orf72*-related disease, including RNA foci and the accumulation of DPRs, but present variable neurodegenerative phenotypes[18]. Notably, a model using AAV to express $(G_4C_2)_{66}$ was the only mouse model to show robust TDP43 pathology[19]. With this particular mouse model, it would be interesting to parse the relative contributions of

repeat-containing RNA and DPRs towards disease pathogenesis by either generating constructs that contain stop codons to interrupt the repeat expansion or by inserting a stop codon before the repeat in order to block RAN translation. It would also be of therapeutic interest to test antisense oligonucleotides (ASOs) targeting the repeat expansion in the mouse model expressing $(G_4C_2)_{66}$ to see if the neurodegenerative phenotypes could be rescued. Whether it is the repeat-containing RNA, DPRs, or both driving toxicity, targeting the repeat expansion with ASOs should mitigate the effects of both toxic species.

5.2 The role of RNA binding and stability in TDP43-dependent neurodegeneration

In Chapter 3, we thoroughly investigated the consequences of TDP43's RRM1-RRM2 salt bridge on TDP43 localization, stability, RNA processing, and toxicity. While prior studies have looked at TDP43's RNA binding properties[20-22], these did not systematically connect RNA binding to TDP43's biophysical properties and downstream toxicity. We determined that both RNA binding and protein stability are important drivers of TDP43-dependent toxicity. We identified the precise RNAs recognized by TDP43 that are responsible for neurodegeneration upon TDP43 accumulation. By comparing the transcriptome and differentially spliced RNAs in human cell lines overexpressing TDP43(WT) or RNA binding-deficient mutants, we found that the RNAs preferentially affected by TDP43(WT) were concentrated within the ribosomal and oxidative phosphorylation pathways. These results are consistent with separate findings from our laboratory showing that TDP43 destabilizes transcripts encoding ribosomal

components and enzymes of the oxidative phosphorylation pathway in iPSCs from *C9orf72*-mutation carriers and in iPSCs overexpressing TDP43[23]. The splicing data we obtained from overexpressing TDP43(WT) suggests that the accumulation of functional TDP43 primarily affects splicing events for transcripts enriched for the spliceosome, proteasome, and RNA transport pathways. Together, these data highlight abnormalities in RNA metabolism, nucleocytoplasmic trafficking, and protein homeostasis in ALS and FTD pathogenesis. Future studies are warranted to investigate the precise mechanisms through which TDP43 accumulation leads to altered gene expression and alternative splicing of transcripts enriched in these pathways. Additional studies should aim to determine how TDP43 acts to destabilize ribosome protein-encoding and oxidative phosphorylation transcripts, and to investigate pathways that can prevent or counteract such destabilization. Whether disease-associated mutations in *TARDBP* enhance the altered gene expression and splicing events identified from cells overexpressing TDP43(WT) remains to be determined.

Although there are limited structural data for TDP43[20,24,25], and given that we demonstrate the importance of the salt bridge for maintaining TDP43's biophysical properties and downstream toxicity, it may be useful to determine the structure of TDP43 upon disruption of the salt bridge. Disruption of the salt bridge impairs RNA binding and destabilizes the protein. Our data suggests that RNA binding may influence the folding of TDP43. Therefore, determining the specific structural differences of TDP43(R151A) and TDP43(D247A) compared to TDP43(WT) via NMR or crystallography could provide insight as to how disruption of the salt bridge leads to changes in RNA binding and protein stability. However, determining the structure of TDP43 poses a challenge. A

complete structure of TDP43 has not been solved, most likely due to its aggregation propensity. The only structural data of TDP43 in the literature comprises RRM1 and RRM2 in complex with UG-rich sequences[20,24,25]. Since the R151A and D247A mutations impair RNA binding, these mutations may add more difficulty to obtain structural information.

Our laboratory could further investigate the molecular interactions between TDP43's RRMs using Förster resonance energy transfer (FRET) by fusing RRM1 and RRM2 to the fluorescent proteins mClover and mRuby2. These fluorescent proteins provide a measurable signal only when they are located within short distances (< 6 nm) from one another[26]. Therefore, this will allow us to measure intramolecular distances and TDP43's binding partners in living cells. Using automated microscopy will also enable us to correlate FRET intensity with neuronal survival. However, a drawback to this approach is that the fluorescent proteins are considerably larger in size compared to the RRMs, which may alter the native folding of TDP43 and also possibly limit the spatial resolution of FRET.

A substantial amount of recent evidence reveals the importance of RBPs involved in familial ALS and FTD, including TDP43, FUS, hnRNPA1 and hnRNPA2B1, MATR3, and TIA1[27]. The intrinsically disordered regions (IDRs) within these RBPs mediate the assembly of ribonucleoprotein granules and other membraneless organelles through a reversible process known as liquid-liquid phase separation (LLPS) *in vitro* and in cells. Disease-associated mutations are frequently located within the IDRs and enhance LLPS, which may be precursory to forming cytoplasmic aggregates that resemble pathological inclusions observed in ALS and FTD post-mortem brains[28-33]. A recent study provides

evidence suggesting that the amount of RNA regulates the phase transition of such RBPs[34]. Low amounts of RNA promotes LLPS, while high amounts of RNA inhibits the formation of liquid-like droplets[34]. Given that we observed differences in the subcellular localization and solubility of RNA binding-deficient TDP43 variants, it would be interesting to determine whether salt bridge-disrupting mutations of TDP43 impact its liquid-like properties *in vitro* and in cells. We hypothesize that salt bridge-disrupting mutations enhance LLPS by interfering with the ability of TDP43 to bind RNA. Additional studies should also determine exactly which RNAs regulate phase separation since we identified specific transcripts linked to the accumulation of TDP43(WT). Phase transitions also impact protein-protein interactions, and these effects have been observed for RNA binding deficient-MATR3 variants[35]. Future studies should investigate whether RNA binding-deficient TDP43 variants also affect protein-protein interactions. Another approach towards investigating the liquid-like properties of the RNA binding-deficient TDP43 variants is to perform fluorescence recovery after photobleaching (FRAP) upon overexpression in primary neurons. Together, these future studies provide an interesting avenue for investigating the exact properties mediating phase transitions.

Finally, future studies should determine whether TDP43's native salt bridge could be disrupted *in vitro* as a therapeutic approach. Small molecule drug screens could be developed to identify compounds that interrupt the salt bridge. Our data suggest that compounds that accomplish this should eliminate TDP43-dependent toxicity in ALS and FTD. The identification of candidate compounds could further be extended to successfully developed disease models in our laboratory involving human neurons derived from iPSCs, donated by individuals with ALS and FTD.

5.3 Defining a link between *C9orf72* mutations and TDP43-related toxicity

To date, it is unknown whether the most common genetic mutation causing ALS and FTD, *C9orf72*, and the pathological hallmark of ALS and FTD, TDP43, are directly linked. Some studies suggest that DPR accumulation precedes TDP43 pathology[36,37], and TDP43 accumulation more closely predicts neuron loss[14,37,38]. Some studies failed to observe an effect on TDP43 localization upon expression of individual DPRs in cell lines[3,6], while others suggest that poly-GA aggregates induce partial mislocalization of TDP43[39] or trigger the formation of TDP43 aggregates[15]. Nevertheless, strong evidence defining a link between *C9orf72* mutations and TDP43 is still lacking.

In Chapter 4, we present preliminary evidence of a possible direct connection between G_4C_2 repeats and TDP43. *In vitro*, TDP43(R151A) and TDP43(D247A), but not TDP43(WT), are capable of recognizing at least one unit of G_4C_2 oligonucleotides. As a first step, future experiments are warranted to confirm whether TDP43(R151A) and TDP43(D247A) are capable of recognizing longer G_4C_2 repeats *in vitro*. Gel shift assays should be optimized for determining whether TDP43 variants can recognize longer G_4C_2 repeats. Because the secondary structure of the repeat expansion poses a great challenge to successfully demonstrate an interaction with the TDP43 variants, an *in vivo* approach is another important step to pursue in the future. Determining an interaction can be accomplished by pull-down assays from cells co-expressing TDP43 variants and long G_4C_2 repeats.

We observed a synergistic effect on neuronal toxicity upon co-expression of *C9orf72* mutations and TDP43(WT) or TDP43(D247A), but not TDP43(R151A). The

enhanced toxicity may imply that G_4C_2 repeats and TDP43 are inducing toxicity through a conserved mechanism. We also noted an increase of TDP43 mislocalization in neurons co-expressing G_4C_2 repeats and TDP43 variants. Given that prior studies suggest *C9orf72* disrupts nucleocytoplasmic transport[1,2,40,41], future studies seeking to establish a mechanism by which *C9orf72* mutations disrupt TDP43 localization are warranted. Specifically determining whether the repeat-expanded RNA or DPRs are responsible for the enhanced TDP43 mislocalization would add additional value to these studies. If the repeat-containing RNA plays a more direct role on toxicity and TDP43 mislocalization, future studies should test co-expression of the RNA binding-deficient TDP43 double mutant (F147L/F149L) and the repeat expansion. If the G_4C_2 RNA is primarily responsible, then no change in neuronal toxicity or TDP43 localization would be expected. Using constructs containing interrupted codons or inserting of a stop codon in front of the repeat could also discriminate between the role of *C9orf72* RNA and DPRs in driving toxicity.

TDP43 pathology is correlated with neurodegeneration in *C9orf72*-mutation carriers, and *C9orf72* RAN translation products exhibit toxicity in model systems, but it is unclear whether *C9orf72* RAN induces neurodegeneration through a TDP43-dependent mechanism. To answer this question, one could evaluate the toxicity of each RAN peptide in primary neurons deficient in TDP43 by transfecting cells with (a) shRNA against TDP43, and (b) plasmids containing G_4C_2 repeats, presumably producing RAN translation products. Alternatively, one could express *in vitro* transcribed G_4C_2 repeat RNA in TDP43-deficient cells to determine whether toxicity depends on the G_4C_2 RNA or RAN products. If TDP43 is necessary for neurodegeneration due to *C9orf72*

mutations, then elimination of endogenous TDP43 should mitigate toxicity when mutant *C9orf72* is expressed in neurons, provided that TDP43 knock-down is not substantially toxic itself.

Prior studies have looked at changes in the transcriptome caused by *C9orf72* mutations either from brain tissue or motor neurons derived from *C9orf72*-ALS/FTD patients[42,43]. These studies observed an enrichment of transcripts involved in RNA processing and oxidative phosphorylation. Additionally, more aberrant splicing events were observed in *C9orf72*-ALS cases compared to sporadic ALS, suggesting dysfunction of RBPs. Zhang *et al.* analyzed the transcriptome of mice expressing poly-GR, and transcripts encoding ribosomal components were significantly altered[44]. With the transcriptome analyses we conducted upon TDP43 overexpression, these studies together suggest dysregulation of ribosomal and oxidative phosphorylation pathways by *C9orf72* and TDP43 are convergent downstream events in ALS and FTD. Therefore, future studies seeking to establish a precise connection between *C9orf72* and TDP43, with special focus on ribosomal and oxidative phosphorylation components, will be instrumental to understanding the disease and developing novel therapies. The next step towards this end is to understand specifically how either the hexanucleotide repeat-containing RNA or the DPRs and TDP43 leads to gene expression changes and abnormal splicing events. Future studies using the *C9orf72* mouse model generated by Chew *et al.* could be a valuable tool to investigate the transcriptomic changes that could lead to TDP43 dysfunction in the context of *C9orf72*, given that these mice exhibit abundant TDP43 pathology.

Given that TDP43 and other related RBPs undergo LLPS *in vitro* and in cells with implications of RNA mediating phase transitions, it would be interesting to investigate whether *C9orf72* mutations affect the biophysical behavior of TDP43. Recent reports suggest that both the hexanucleotide repeat expansion[45,46] and arginine-rich dipeptides[47-49] play a role in LLPS *in vitro* and in cells. Phase separation driven by expanded G_4C_2 RNA is length and structure-dependent[45], while arginine-rich dipeptides interact directly with low-complexity sequences in proteins, undergo LLPS, and disrupt the dynamics of membraneless organelles such as stress granules[47-49]. Although these reports observe phase transitions in the presence of the G_4C_2 expanded RNA or DPRs, it is unknown whether the DPRs or repeat-containing RNA interact with TDP43 directly in such phase transitions. Future studies investigating the molecular mechanisms mediating LLPS in the context of the *C9orf72* and TDP43 could provide insight on a potential connection.

5.4 Concluding remarks

ALS and FTD are debilitating neurodegenerative conditions resulting in the progressive loss of motor and cognitive function for thousands of people each year. Despite the truly remarkable progress made in the genetics and biology underlying ALS and FTD, effective therapies remain extremely limited. The studies described here have underscored important mechanisms responsible for neurodegeneration in ALS and FTD that may serve as targets for therapy development. This dissertation investigated *C9orf72* mutations, abnormal TDP43 deposition, and their respective contributions to neuronal toxicity.

Despite the significant advancements that have been made since the identification of the *C9orf72* mutation as the most common genetic cause of ALS and FTD, it remains crucial, yet challenging, to understand exactly how the G_4C_2 repeats cause neurodegeneration. Although three mechanisms have been proposed by which *C9orf72* mutations may cause disease, it remains to be determined which mechanism is the primary driver of pathogenicity and if there is a cascade of events leading to cell death. Determining the relative contributions of each of these potential mechanisms will certainly influence therapeutic strategies.

Despite our incomplete understanding of the key mechanisms underlying TDP43-dependent neurodegeneration, this dissertation highlights critical features of TDP43's RNA binding properties and the neurodegeneration that arises from its accumulation in ALS and FTD. The fact that we detected a significant portion of transcripts encoding components of the ribosomal and oxidative phosphorylation pathways affected by the accumulation of TDP43, and transcripts involved in these pathways were also destabilized in a separate study using patient-derived neurons underscores the importance of RNA homeostasis in ALS and FTD. Determining the precise mechanisms leading to RNA dysregulation will absolutely aid in the development of therapeutic strategies to delay or prevent neurodegeneration in ALS and FTD.

The direct connection, if any, between *C9orf72* mutations and TDP43 is yet to be determined. While there are events such as nucleocytoplasmic transport and RNA misprocessing that are affected by *C9orf72* mutations and TDP43, determining a direct interaction will be instrumental for subsequent studies. Efforts towards defining a link will certainly contribute to our understanding of the molecular pathways responsible for

neurodegeneration. Together, the proposed experiments outlined in this thesis will drive further investigation and promise an interesting study that will be useful for identifying critical targets for effective therapies.

References

1. Freibaum, B. D. *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525,** 129–133 (2015).
2. Jovičić, A. *et al.* Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18,** 1226–1229 (2015).
3. May, S. *et al.* C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathologica* **128,** 485–503 (2014).
4. Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345,** 1139–1145 (2014).
5. Mizielinska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345,** 1192–1194 (2014).
6. Zhang, Y.-J. *et al.* Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathologica* **128,** 505–524 (2014).
7. Yang, D. *et al.* FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathologica* **130,** 525–535 (2015).
8. Tao, Z. *et al.* Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum. Mol. Genet.* **24,** 2426–2441 (2015).
9. Yamakawa, M. *et al.* Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum. Mol. Genet.* **24,** 1630–1645 (2014).
10. Kanekura, K. *et al.* Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Human Molecular Genetics* **25,** 1803–1813 (2016).
11. Chang, Y.-J., Jeng, U.-S., Chiang, Y.-L., Hwang, I.-S. & Chen, Y.-R. The Glycine-Alanine Dipeptide Repeat from C9orf72 Hexanucleotide Expansions Forms Toxic Amyloids Possessing Cell-to-Cell Transmission Properties. *J. Biol. Chem.* **291,** 4903–4911 (2016).
12. Westergard, T. *et al.* Cell-to-Cell Transmission of Dipeptide Repeat Proteins Linked to C9orf72-ALS/FTD. *Cell Rep* **17,** 645–652 (2016).
13. Zhou, Q. *et al.* Antibodies inhibit transmission and aggregation of C9orf72poly-GA dipeptide repeat proteins. *EMBO Mol Med.* **9,** 687–702 (2017).
14. Mackenzie, I. R. A. *et al.* Quantitative analysis and clinico-pathological correlationsof different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathologica* **130,** 845–861 (2015).
15. Nonaka, T. *et al.* C9ORF72 dipeptide repeat poly-GA inclusions promote intracellular aggregation of phosphorylated TDP-43. *Hum. Mol. Genet.* **72,** 245 (2018).
16. Gallegos, S., Pacheco, C., Peters, C., Opazo, C. M. & Aguayo, L. G. Features of alpha-synuclein that could explain the progression and irreversibility of Parkinson's disease. *Front. Neurosci.* **9,** 59 (2015).
17. Nath, S. *et al.* Spreading of Neurodegenerative Pathology via Neuron-to-Neuron Transmission of -Amyloid. *J of Neurosci* **32,** 8767–8777 (2012).

18. Batra, R. & Lee, C. W. Mouse Models of C9orf72 Hexanucleotide Repeat Expansion in Amyotrophic Lateral Sclerosis/ Frontotemporal Dementia. *Front. Cell. Neurosci.* **11,** 385 (2017).
19. Chew, J. *et al.* C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* **348,** 1151–1154 (2013).
20. Lukavsky, P. J. *et al.* Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struc. Mol. Biol.* **20,** 1443–1449 (2013).
21. Buratti, E. & Baralle, F. E. Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.* **276,** 36337–36343 (2001).
22. Ayala, Y. M. *et al.* Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. *J. Mol. Biol.* **348,** 575–588 (2005).
23. Tank, E. M. *et al.* Abnormal RNA stability in amyotrophic lateral sclerosis. *Nat Commun* 1–16 (2018). doi:10.1038/s41467-018-05049-z
24. Kuo, P.-H., Doudeva, L. G., Wang, Y.-T., Shen, C.-K. J. & Yuan, H. S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **37,** 1799–1808 (2009).
25. Kuo, P.-H., Chiang, C.-H., Wang, Y.-T., Doudeva, L. G. & Yuan, H. S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res.* **42,** 4712–4722 (2014).
26. Lam, A. J. *et al.* Improving FRET dynamic range with bright green and red fluorescent proteins. *Nat Methods* **9,** 1005–1012 (2012).
27. Ito, D., Hatano, M. & Suzuki, N. RNA binding proteins and the pathological cascade in ALS/FTD neurodegeneration. *Science Translational Medicine* **9,** (2017).
28. Gopal, P. P., Nirschl, J. J., Klinman, E. & Holzbaur, E. L. F. Amyotrophic lateral sclerosis-linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. *Proceedings of the National Academy of Sciences* **114,** E2466–E2475 (2017).
29. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162,** 1066–1077 (2015).
30. Ryan, V. H. *et al.* Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Molecular Cell* **69,** 465–479.e7 (2018).
31. Mackenzie, I. R. *et al.* TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. 1–19 (2017). doi:10.1016/j.neuron.2017.07.025
32. Molliex, A. *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell* **163,** 123–133 (2015).
33. Malik, A. M. *et al.* Matrin 3-dependent neurotoxicity is modified by nucleic acid binding and nucleocytoplasmic localization. *Elife* **7,** (2018).
34. Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* eaar7366 (2018). doi:10.1007/978-0-387-46312-4
35. Iradi, M. C. G. *et al.* Characterization of gene regulation and protein interaction

networks for Matrin 3 encoding mutations linked to amyotrophic lateral sclerosis and myopathy. *Sci Rep* 1–15 (2018). doi:10.1038/s41598-018-21371-4
36. Baborie, A. *et al.* Accumulation of dipeptide repeat proteins predates that of TDP-43 in frontotemporal lobar degeneration associated with hexanucleotide repeat expansions in C9ORF72gene. *Neuropathol Appl Neurobiol* **41,** 601–612 (2015).
37. Mackenzie, I. R. *et al.* Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathologica* **126,** 859–879 (2013).
38. Davidson, Y. S. *et al.* Brain distribution of dipeptide repeat proteins in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9ORF72. *Acta Neuropathol Commun* **2,** 70 (2014).
39. Khosravi, B. *et al.* Cytoplasmic poly-GA aggregates impair nuclear import of TDP-43 in C9orf72ALS/FTLD. *Hum. Mol. Genet.* ddw432 (2017). doi:10.1093/hmg/ddw432
40. Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525,** 56–61 (2015).
41. Boeynaems, S. *et al.* Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci Rep* **6,** 20877 (2016).
42. Prudencio, M. *et al.* Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci* 1–10 (2015). doi:10.1038/nn.4065
43. Cooper-Knock, J. *et al.* C9ORF72 GGGGCC Expanded Repeats Produce Splicing Dysregulation which Correlates with Disease Severity in Amyotrophic Lateral Sclerosis. *PLoS ONE* **10,** e0127376 (2015).
44. Zhang, Y.-J. *et al.* Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nature Medicine* 1–13 (2018). doi:10.1038/s41591-018-0071-1
45. Fay, M. M., Anderson, P. J. & Ivanov, P. ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions In Vitro and in Cells. *Cell Rep* **21,** 3573–3584 (2017).
46. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546,** 243–247 (2017).
47. Boeynaems, S. *et al.* Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Molecular Cell* **65,** 1044–1055.e5 (2017).
48. Lin, Y. *et al.* Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **167,** 789–796.e12 (2016).
49. Lee, K.-H. *et al.* C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **167,** 774–778.e17 (2016).