Investigating the Role of Matrin-3 in Amyotrophic Lateral Sclerosis and Distal Myopathy

by

Nandini Ramesh

BTech, Anna University, India, 2014

Submitted to the Graduate Faculty of

the Department of Human Genetics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Nandini Ramesh

It was defended on

March 31, 2020

and approved by

Committee Members: Zsolt Urban, PhD, Associate Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Committee Member: Quasar S. Padiath, PhD, Associate Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Committee Member: Christopher J. Donnelly, PhD, Assistant Professor, Department of Neurobiology, School of Medicine, University of Pittsburgh

Committee Member: Sokol V. Todi, PhD, Associate Professor, Department of Pharmacology and Neurology, School of Medicine, Wayne State University

Dissertation Advisor: Udai B. Pandey PhD, Associate Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Udai B. Pandey, PhD

Investigating the Role of Matrin-3 in Amyotrophic Lateral Sclerosis and Distal Myopathy

Nandini Ramesh, PhD

University of Pittsburgh, 2020

Matrin-3 (MATR3) is a ubiquitously expressed DNA/RNA-binding protein involved in regulating various aspects of RNA metabolism. MATR3-positive cytoplasmic inclusions have been identified in post-mortem brain tissues of amyotrophic lateral sclerosis (ALS) patients with G4C2 hexanucleotide repeat expansions in *C9orf72* gene. Furthermore, pathogenic mutations in *MATR3* have been discovered in patients with ALS and distal myopathy, suggesting multisystem proteinopathy involving degeneration of motor neurons and muscles. However, the underlying mechanisms of MATR3-mediated neuromuscular degeneration are still unknown.

Here, I developed *Drosophila* models expressing MATR3-WT and ALS/myopathy linked mutations that recapitulate key features of human disease including motor dysfunction and muscular atrophy. Using site-directed mutagenesis to investigate structure-function correlations, I found that MATR3 exerts toxicity through its RNA-binding domains. Through a genetic screen, I found that knocking down rump, *Drosophila* homolog of human hnRNPM, suppressed mutant MATR3 toxicity. In mammalian cells, MATR3 and hnRNPM colocalized together and showed RNA-mediated physical interaction, suggesting a role for hnRNPM in mediating mutant MATR3 toxicity. Additionally, I found that MATR3 genetically and functionally interacts with G4C2 repeat expansion in iPSC-derived motor neurons and *Drosophila* models of C9orf72-ALS through RNA-dependent mechanism(s).

Neuromuscular diseases impose a huge burden on public health with no effective therapies available. The work presented in this dissertation highlights the role of MATR3-mediated neuromuscular degeneration in ALS and myopathy, particularly through its RNA-binding functions; and identifies strong genetic modifiers that are potential candidates for future therapeutic intervention

Table of Contents

List of Tables

List of Figures

Preface

First and foremost, I would like to sincerely thank my supervisor and mentor, Dr. Udai Pandey, for his endless support and patience in guiding me through my graduate career. I am extremely grateful for his knowledge, thoughtful motivation and encouragement of independent thinking that has aided my professional and personal growth. He has taught me a great deal about scientific research and guided me in the clear and impactful presentation of my thesis work. I could not have asked for a better advisor and I am privileged to have had the opportunity to work with him.

My sincere thanks also to my committee members for their time and valuable suggestions that were immensely helpful in improving my dissertation research. I would also like to extend my thanks to my graduate faculty advisor, Dr. Candace Kammerer, for her continued support.

My PhD training and research would not be possible without the collaboration with current and former members of the Pandey lab. My sincere thanks to Sukhleen, Ian, Eric, Gavin, John, Tyler and Darilang for your thoughtful discussions and giving me the opportunity to learn from you. Above all, thank you for your friendship and making each day enjoyable to work in the lab. I would like to thank all the collaborators who helped with this work – Dr. Evangelos Kiskinis, Dr. Sami Barmada, Dr. Dhivyaa Rajasundaram, Dr. Julia Kofler, Dr. Amanda Gleixner, Jake Mann, Elizabeth Daley and Ahmed Malik.

I am deeply grateful for the love and support of my family and friends. Thank you to my parents, V. Ramesh and Vijayalakshmi Ramesh, and my brother, Rahul, for their unrelenting support and encouragement to follow my passion. Special thanks to my friend Nivedhitha for

enriching my life with cat videos. Finally, thank you to Srinidhi for the laughs, unwavering support and constant motivation throughout my PhD.

1.0 Chapter 1: A current perspective on RNA-binding protein Matrin-3 and its role in neuromuscular degeneration

1.1 Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal adult-onset neurodegenerative disorder that was first described over 150 years ago by the renowned French clinician and neurobiologist, Jean-Martin Charcot (1825-93) [1, 2]. Charcot described lesions in the lateral portion and anterior horn of the spinal cord, both resulting in progressive degeneration. Initially known as Charcot's Sclerosis, it is now more commonly known as Lou Gehrig's disease, named after the famous American baseball player who brought overwhelming attention to the disease following his diagnosis.

ALS is characterized by the progressive degeneration of upper motor neurons in the motor cortex, and lower motor neurons in the brainstem and spinal cord [3]. This leads to progressive weakness and atrophy of skeletal muscles, causing paralysis and resulting in death due to diaphragmatic failure within 2-4 years of disease onset. However, there is considerable heterogeneity in the disease prognosis, with 10-20% of cases surviving over 5 years, and 5-10% of cases surviving over 10 years [4]. The average age of disease onset is currently 58-60 years [5], though some genetic forms of ALS present with an early-onset and aggressive form [6]. Epidemiologically, ALS has an incidence of approximately 1-2.6 cases per 100,000 persons annually, and a prevalence of approximately 6 cases per 100,000 [5]. Although it is relatively rare, ALS is the most common adult-onset motor neuron disease. Moreover, with an ageing population, the number of ALS cases is projected to increase by 69% over the next 25 years [7]. However, the

disease etiology is still unknown, and there is an urgent need to develop therapeutics. Currently, there are two FDA-approved therapies available: riluzole and edaravone. However, these drugs have modest effects on patient outcome, extending their lifespan by a just few months in a subset of patients [8].

1.2 Genetics of ALS

ALS is predominantly considered to be a "sporadic" disease, in that, >90% of ALS patients do not have a family history of the disease [9]. Approximately 5-10% of patients have familial ALS caused by mutations in roughly 30 genes identified so far, inherited mostly in an autosomal dominant inheritance pattern [10]. Genome-Wide Association studies (GWAS) comparing genomes of ALS-cases and controls have identified common variants that are associated with disease susceptibility [10]. Taking together high-risk monogenic variants and low-risk common variants, so far >110 genetic factors have been associated to ALS (ALSoD: ALS online genetics database) (11).

The clinical and pathological features of familial and sporadic forms of ALS are indistinguishable. At a cellular level, they are both characterized by loss of cell bodies on upper and lower motor neurons, gliosis and axonal retraction [12]. Importantly, the surviving neurons in post-mortem brains from either sporadic or familial ALS patients show ubiquitinated inclusions which are positive for the ALS-associated protein TDP-43 in ~95% of the patients [12]. Furthermore, a subset of sporadic ALS patients harbors mutations in the same genes as familial ALS patients. This postulates that ALS pathogenesis is due to multiple overlapping mechanisms, and that the proteins encoded by the ALS-linked genes are key molecular players in these

processes. The causative genes in ALS are involved in a combination of proposed disease mechanisms, including nucleocytoplasmic transport defects [*C9orf72, TARDBP*] [13, 14]; dysregulation of RNA metabolism [*TARDBP, FUS, EWSR1, MATR3, C9orf72, TAF15, ATXN2, TIA1, HNRNPA1, SETX, ANG*] [15]; failure of protein quality control [*C9orf72, OPTN, SQSTM1, VCP, UBQLN2*] [16, 17]; defects in DNA damage repair [*FUS, TAF15, SETX, EWSR1 NEK1, C9orf72, MATR3*] [18–21]; oxidative stress and mitochondrial dysfunction [*SOD1, CHCHD10, TARDBP*] [22, 23]; vesicle transport defects [*OPTN, VAPB, CHMPB2, UNC13A*] [24], and axonal transport defects [*TARDBP, TUBA4A, PFN1*] [25, 26]. Among these, overwhelming research has thus far focused on dysregulation of RNA-binding proteins and RNA metabolism.

1.3 RNA-binding proteins in ALS pathogenesis

RNA-binding proteins are involved in post-transcriptional mRNA processing through binding with specific RNA targets. Post-transcriptional regulation includes polyadenylation, splicing, localization, transport, decay and turnover, and translation. The central nervous system relies heavily on post-transcriptional RNA-processing to maintain morphological and functional diversity [27]. In fact, the brain has the highest level of alternative splicing, thus contributing to the proteomic diversity in the central nervous system (CNS) [27]. Furthermore, mRNA transport along axons mediated by RNA-binding proteins is required for local translation and maintaining local protein homeostasis [28]. Thus, dysfunction of RNA-binding proteins has been implicated in multiple neurological disorders, including ALS/FTD, Fragile-X-associated tremor/ataxia syndrome, Huntington's disease and Spinal Muscular Atrophy [29].

A majority of ALS-causing genes encode RNA-binding proteins (RBPs), including fused in sarcoma (*FUS*), TAR-binding protein (*TARDBP*), Matrin-3 (*MATR3*), TATA-box binding protein, associated factor 15 (*TAF15*), heterogenous nuclear ribonucleoprotein A1 (*HNRNPA1*) and A2B1 (*HNRNAPA2B1*), and Ewing's sarcoma breaking point region 1 (*EWSR1*). In addition, one of the key pathological hallmarks of ALS is the presence of neuronal and glial intracellular inclusions of RNA-binding proteins, notably TDP-43 [30], FUS [6], EWSR1 [31], TAF15 [32] and MATR3 [33]. Furthermore, the most common genetic cause of ALS is a hexanucleotide (GGGGCC or G4C2) repeat expansion in the first intron of *C9orf72* gene [34], commonly termed as C9orf72-ALS. One of the key pathologic features found in C9orf72-ALS patient tissues is aggregates of the repeat-containing RNA into "RNA foci" that sequester RNA-binding proteins [34], further expanding the scope of RNA-binding proteins involved in disease pathogenesis. Taken together, these findings point to dysregulated RNA metabolism as a key mechanism underlying neurodegeneration. Understanding how mutations in RNA-binding proteins affect their respective functions is important to answer overarching questions about RNA dysregulation in ALS.

1.4 Mutations in *MATR3* cause ALS

Matrin-3 (MATR3) is a ubiquitously expressed DNA/RNA binding protein that was initially discovered as a component of the nuclear matrix [35]. While nuclear matrix proteins are important for maintaining the structural integrity of the nucleus, MATR3 also plays a role in multiple aspects of RNA metabolism (outlined in section 1.7.3). The protein is 847 amino acids in

length and comprised mostly of intrinsically disordered regions (IDRs), with the exception of 4 distinct functional domains, including two RNA-recognition motifs (RRM1 and RRM2) and two zinc-finger domains (ZF1 and ZF2) **(Fig 1).**

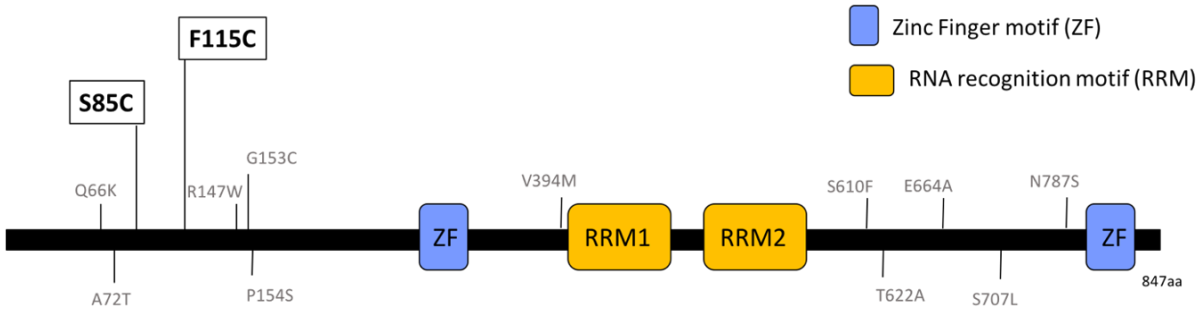

Figure 1: MATR3 protein schematic

Schematic representation of MATR3 protein showing functional domains (RRM1/2 and ZF1/2) and ALS-associated mutations. The F115C and S85C mutations are characterized further in Chapter 2 of this dissertation

ALS-associated mutations in *MATR3* were first described in 3 independent cohorts of European ancestry, each carrying unique mutations in MATR3: Phe115Cys (F115C) and Thr622Ala (T622A) mutations in familial ALS patients with autosomal dominant inheritance pattern; and Pro154Ser (P154S) mutation in a sporadic ALS patient [36]. However, the pathogenicity of T622A mutation was not fully established in this study, and was later found to be a non-disease associated polymorphism in the Genome Aggregation Database (gnomAD) [37], suggesting that the mutation is likely not causal. Prior to this study, a Ser85Cys (S85C) mutation in *MATR3* was discovered to segregate in an autosomal dominant manner in a large multigenerational family, who exhibited distal and asymmetrical myopathy, as well as vocal cord and pharyngeal weakness (VCPDM) [38]. Interestingly the neurophysiological and muscle biopsy-based observations from these patients indicated a neurogenic (arising from the nervous system) or a myopathic (arising from the muscles) pattern. They exhibited signs of brisk knee and upper limb reflexes, tongue fasciculations, brunt jaw-jerk and a “split-hand” pattern, all indicative of

upper motor neuron lesions. In light of other mutations in *MATR3* causally linked to ALS, the lead neurologist re-evaluated the S85C-carrying patients and discovered that they developed a progressive respiratory failure, that resulted in their death 15 years-post onset [36], and subsequently re-classified the S85C-associated disorder to "slow progressive ALS with distal myopathy". Further expanding the phenotypic spectrum of MATR3-associated ALS, 2 out of 5 of the patients with F115C mutation displayed cognitive impairments several years after the initial presentation of the disease, and also had a family history of dementia, indicating that this mutation could be resulting in an ALS/FTD disease phenotype [36].

Following the initial discovery, other mutations in *MATR3* have been discovered in Taiwanese [39], Chinese [40], French [41] and Italian [42] cohorts of both familial and sporadic ALS patients. In addition to phenotypic heterogeneity, another characteristic feature of MATR3-associated ALS is a variation in disease course, with a majority of patients exhibiting relatively prolonged survival, ranging from 7 to 15 years [36, 39, 42]. This alludes to mutations in MATR3 typically causing a slow, progressive form of the disease. *In silico* analysis of severity of the mutations using prediction algorithms such as PolyPhen2 reveal that some of variants might be benign **(Table 1)**. Nevertheless, pooling population metrics from all mutations, the frequency of MATR3-associated ALS is currently estimated at approximately 0.5% of all ALS cases [41]

Table 1: PolyPhen2 and SIFT scores and corresponding prediction of pathogenicity of MATR3 mutations

	SIFT Prediction	PolyPhen2 Prediction	Allele Frequency (gnomAD)
Q66K	Score =0.00 AFFECT PROTEIN FUNCTION	Score = 0.924 POSSIBLY DAMAGING	Not found
A72T	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.759 POSSIBLY DAMAGING	Not found
S85C	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.999 PROBABLY DAMAGING	Not found
F115C	Score = 0.01 AFFECT PROTEIN FUNCTION	Score = 0.999 PROBABLY DAMAGING	1.41e-5
R147W	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.997 PROBABLY DAMAGING	Not found
G153S	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.816 POSSIBLY DAMAGING	Not found
P154S	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.001 BENIGN	Not found
V394M	Score = 0.02 AFFECT PROTEIN FUNCTION	Score = 0.001 BENIGN	3.98e-6

Table 1 Continued

S610F	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.995 PROBABLY DAMAGING	Not found
E664A	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.993 PROBABLY DAMAGING	3.04e-4
S707L	Score = 0.00 AFFECT PROTEIN FUNCTION	Score = 0.20 BENIGN	7.98e-6
N787S	Score = 0.03 AFFECT PROTEIN FUNCTION	Score = 0.383 BENIGN	3.75e-4

1.5 MATR3-related multisystem proteinopathy

While the S85C mutation in MATR3 was initially reclassified as leading to a combined ALS/myopathy disorder, there have also been independent reports causally linking the variant to distal myopathy without neurogenic symptoms [43–45]. Distal myopathies are a group of muscular dystrophies where the distal muscles of the upper and lower limbs are selectively affected [46]. Distal myopathy associated with MATR3-S85C mutation is characterized by adult onset muscle weakness in the feet and hands, which then slowly progresses to the proximal limb muscles. It is also combined with vocal and/or swallowing difficulties and thus diagnosed as vocal cord and pharyngeal distal myopathy (VCPDM) [38]. This pleiotropy adds *MATR3* to the list of genes, including *VCP* [47], *SQSTM1* [48], *HNRNPA1* [49] and *HNRNPA2B1* [49], that are associated

with multisystem proteinopathy, an inherited neurodegenerative disorder that affects the muscle, bone and/or the nervous system [50].

Apart from adult-onset degenerative disorders, MATR3 has also been linked to developmental delays and congenital heart defects [51]. Specifically, the Developmental Genome Anatomy Project, which analyses chromosomes of individuals with major congenital abnormalities [52], described one individual with a disruption in *MATR3* gene. Phenotypic characterization revealed global developmental delay with a marked delay in speech, and was clinically diagnosed with Noonan syndrome [53] along with multiple cardiac defects. This disruption resulted in an overall increase in protein levels of MATR3 in the patient lymphoblasts compared to control [51]. While the ALS and myopathy-linked mutations are missense mutations, the congenital defect is due to a disruption of the gene, and the mechanisms may not be necessarily similar (gain-of-function vs loss-of-function). It is also possible that MATR3 has different roles in development and disease pathogenesis.

1.6 Pathological features of MATR3-ALS/VCPDM

Morphological analyses of muscle biopsy samples from VCPDM patients show dystrophic features including partially rimmed vacuoles in both mildly and severely degenerating muscles [43–45], and proliferation of fat [44, 45] indicative of end-stage myopathy. On the ultrastructural level, observations from muscle biopsies in German VCPDM patients also showed irregularly shaped nuclei in the skeletal muscle fibers and in the satellite cells, characterized by indentations and segmentations. These features mimic the morphological changes observed in nuclear envelopathies in humans [54–56], and in senescent cells [57, 58].

In brains, immunohistochemical analysis of healthy control subjects showed a granular staining pattern for MATR3 in the nuclei of motor neurons and surrounding glia [36]. In tissues from patients with F115C mutation, MATR3 immunoreactivity was intense within the nucleus accompanied by diffuse cytoplasmic staining in some neurons [36]. In myonuclei, on the other hand, MATR3 appears to show varying degrees of immunoreactivity in patient tissues (26,27), in that, some myonuclei exhibited intense immunoreactivity, while some exhibited loss of immunoreactivity of MATR3 [44]. TDP-43 [36] and phosphorylated TDP-43 [44] aggregates were detected in the myofibers of some patients [36], with variable observations of co-aggregation with MATR3 [36]. Interestingly, sarcoplasmic p62 aggregates accompanied by granular ubiquitin in the sarcoplasm of these fibers have also been shown in a fraction of myofibers [44]. This is characteristic of inclusions found in many neurodegenerative disorders, including ALS, and suggest deficits in protein degradation mechanisms.

Mutations in other ALS-associated RNA-binding proteins, such as TDP-43 [59] and FUS [6], have been widely shown to lead to neuronal and/or cytoplasmic pathological inclusions. While MATR3-positive cytoplasmic aggregates were described in the myofibers of a subset of VCPDM [44], it remains to be seen if this is a common feature across the board. It is likely that mutations in MATR3 could possibly be causing a gain-of-function within the nucleus, and thus leading to disease pathogenesis.

1.7 Functions of MATR3

MATR3 protein was initially discovered along with other nuclear matrins that compose the nuclear matrix, a salt-resistant and detergent resistant network that is distributed throughout the

nucleus [35]. The nuclear matrix plays a crucial role in maintaining the skeletal framework of the nucleus through interactions with the chromosome as well as the nuclear lamina [35]. The functional domains in MATR3, which includes RNA-recognition motifs (RRM1/2), C2H2-type zinc-finger domains (ZF1/ZF2) and a highly acidic carboxy-terminal that binds to histones, allow the protein to participate in multiple other cellular functions. Broadly, the role of MATR3 in the cell can be summarized by its interaction with DNA and/or DNA-binding proteins to regulate chromosomal and genomic integrity, association with the nuclear lamina to maintain nuclear framework, and post-transcriptional mRNA regulation that is mediated by RNA-binding

1.7.1 MATR3 in chromatin regulation and DNA-damage response

The chromatin is tethered to the nuclear matrix through DNA elements called scaffold/matrix attachment regions (S/MAR), which associate with matrix-binding proteins to anchor the chromatin into loop domains [60]. This interplay between S/MAR and matrix proteins determines the structure of the chromatin loop domains, which, along with chromatin modifications regulates gene expression through modulating active and silent chromatin states [61]. MATR3 has been demonstrated to bind to these S/MAR regions, particularly in the euchromatin, suggesting that it may play a direct role in euchromatin organization and transcription [60]. Additionally, MATR3 may affect chromatin structure indirectly through post-transcriptional regulation of genes that encode chromatin remodeling factors. Indeed, functional annotation of transcripts that are either differentially regulated by MATR3 or directly-bound by MATR3 revealed significant enrichment in chromatin regulation and remodeling [62–64].

MATR3 has also been implicated in maintaining genomic integrity through aiding DNA-damage response [19]. In response to genomic stress and DNA damage, MATR3 is phosphorylated

through the ATM/ATR kinase cascade which stalls the cells at S-phase until the damage is repaired. While MATR3 is not directly recruited to the double-stand break, it interacts with PSF and NONO, both of which are recruited to the site of the damage. Additionally, knocking down MATR3 altered the retention time of these proteins at the site of damage [19], suggesting that MATR3 may be involved in dynamic recruitment and release of other DNA damage response proteins at the site of damage. Consistent with this idea, many proteins in the MATR3 protein interactome (outlined below) are DNA-damage response factors.

1.7.2 The role of MATR3 in maintaining nuclear architecture

The nuclear lamina, which is primarily composed of Lamin proteins, is a dense network of filaments at the nuclear periphery important for maintaining the structural integrity of the nucleus [65]. The Lamins interact with many binding partners to perform this role. An investigation of the nuclear lamina proteome in C2C12, mouse myoblast cell lines, further differentiated into myotubes identified MATR3 as a key interactor of the nuclear membrane protein Lamin A [66]. MATR3 localized to the nuclear matrix and the nuclear membrane, where it overlapped with Lamin A signal. Further, localization of MATR3 to the nuclear membrane appeared to be specific to myoblasts that have already entered myogenic differentiation, suggesting that MATR3 localization to the nuclear membrane and colocalization with Lamin A are possibly important for myogenic differentiation [66]. Interestingly, this was the first report to indicate MATR3 localization at the nuclear periphery, highlighting the cell-type and developmental-stage specificity of MATR3 functions.

More than 300 mutations in *LMNA* have been found to lead to various skeletal and cardiac myopathies [67]. In patient-derived fibroblasts harboring *LMNA-Δ303* truncation that causes

cardiomyopathy and muscular dystrophy (29-31), there is decreased spatial and physical interaction between the MATR3 and Lamin A resulting in increased distance between the nuclear membrane and the nuclear matrix [66]. MATR3 is possibly an important mediator in the interaction between the nuclear matrix and lamina to maintain nuclear integrity. Particularly in differentiated muscle cells, reduced interaction between the nuclear matrix and lamina could make these cells especially susceptible to mechanical forces that compromise the nuclear architecture, and consequent loss of nuclear membrane integrity. Congruently, mutations in MATR3 could also possibly disrupt interactions between the nuclear lamina and matrix, and thus contribute to disease pathogenesis. This is especially highlighted by nuclear membrane invaginations observed in skeletal muscle cells of a subset of patients with MATR3-associated distal myopathy [43].

1.7.3 Post-transcriptional gene regulation by MATR3

MATR3 plays an important role in RNA metabolism mediated by direct interaction with mRNA through its RRM2 domain [68]. Depletion of MATR3 led to down-regulation of several transcripts, which, in some cases, was due to decreased half-life [68], suggesting that MATR3 is involved in maintaining mRNA stability. It is likely that MATR3 does not perform this function alone, as several other proteins implicated to have a role in maintaining mRNA stability are binding partners of MATR3 [69–71]. While more compelling evidence has emerged supporting the role of MATR3 in alternative splicing, the two post-transcriptional events are not necessarily mutually exclusive. In fact, an independent report on changes to the transcriptome upon MATR3 knockdown noted that most of the transcripts that were downregulated, also showed alternative splicing events that enriched isoforms with premature stop codons [63]. This suggests that MATR3

could be indirectly impacting mRNA decay/stability through degradation of these isoforms by nonsense-mediated decay (NMD).

A prominent tool that has aided in deciphering RNA targets of RNA-binding proteins and their cognate binding motifs has been cross-linking immunoprecipitation (CLIP) combined with high-throughput sequencing [72]. Utilizing PAR-CLIP (photoactivatable-ribonucleoside-enhanced CLIP) in human neuroblastoma-derived SH-SY5Y cells, MATR3 was shown to preferentially bind to pyrimidine-rich sequences within intronic regions of pre-mRNAs [64]. Analysis of alternative splicing events corresponding to surrounding exons revealed that MATR3 acts predominantly as a splicing repressor. Further single-nucleotide resolution of MATR3-bound transcriptome using iCLIP (individual-nucleotide resolution CLIP) also supported MATR3's affinity to pyrimidine-rich motifs, and optimally to AUUCU-motifs in introns flanking repressed exons [63]. Annotation of the splicing map of MATR3 suggests that MATR3 initially binds to a specific site, which is followed by recruitment of more MATR3 around the site through self-association [60], leading to repression of target exons [63]. Analysis of functional terms associated with regulated cassette exons revealed enrichment of genes encoding chromatin/chromatin-binding proteins and cytoskeletal proteins [63]. Thus, MATR3 could also potentially be affecting mRNA levels and/or stability through its alternative splicing regulation of chromatin-related proteins.

A curious feature of RNA-binding proteins (RBPs), including MATR3, is that they are ubiquitously expressed in many different cell types, and yet, the central nervous system is especially susceptible to mutations in RBPs. Indeed, the central nervous system shows high levels of alternative splicing [73], and aberrant splicing has been observed in bio-samples from ALS patients [74, 75]. Thus, determining cell/tissue-type specific transcriptomic targets and alternative

splicing events regulated by MATR3 will be important for interpreting the basis of disease pathogenesis. MATR3 has been reported to be a part of a large assembly of splicing regulators (LASR) in the mouse brain, which includes RBFOX, hnRNPM, hnRNPH, hnRNPC, hnRNPUL2, NF110, NF45 and DDX5 [76]. The major component of this complex, RBFOX1, regulates alternative splicing of important transcripts that are pertinent to neurodevelopment and synaptic functions (76). This suggests a potentially analogous role for MATR3 in regulating alternative splicing in neurons through the LASR complex.

While mutations in RNA-binding proteins often cause neurological diseases [77–79], they are also linked to the pathogenesis of muscular dystrophies such as myotonic dystrophy [80], and oculopharyngeal muscular dystrophy (OPMD) (81). OPMD, in particular, is an autosomal dominant disease caused by a mutation in the *PABPN1* gene which encodes the poly(A) RNA binding protein nuclear 1, also known as polyadenylate-binding protein 2 [81]. In mouse skeletal muscles, MATR3 was found to be a key interactor of PABPN1, and also found to bind to several essential myogenic transcripts including *Myog*, *Myod* and *Pitx2* [82]. Consequently, knocking down MATR3 affected the proliferation and differentiation of primary myoblasts to myotubes [82]. Further elucidation of MATR3-mediated RNA regulation in specific cell types would be imperative for understanding susceptibility of neurons and muscles to mutations in *MATR3*.

1.8 Defining MATR3 protein interactome

There is modest evidence that RNA targets are shared between the different RNA-binding proteins, suggesting that there could be coordinated roles for regulation by a complex of RNA-binding proteins. Indeed, the combinatorial nature of splicing strongly indicates that groups of

regulatory proteins can act together [83–86]. Thus, considerable focus has been on elucidating the protein-protein interactions of RNA-binding proteins, including MATR3. Three different groups have utilized high throughput immunoprecipitation coupled to mass spectrometry (IP-MS) to build the MATR3 protein interactome, with prominent overlap in protein-protein interactions (PPIs) identified between the different groups [69–71]. A majority of the interactors include heterogenous ribonucleoproteins (hnRNPs) that have ubiquitous roles in RNA processing [87]. Indeed, functional annotation of protein interactors that were identified in two or more IP-MS studies revealed enrichment of proteins involved in various aspects of RNA metabolism, including mRNA stability, RNA splicing, miRNA-mediated gene silencing and RNA catabolic processes **(Fig 2)**.

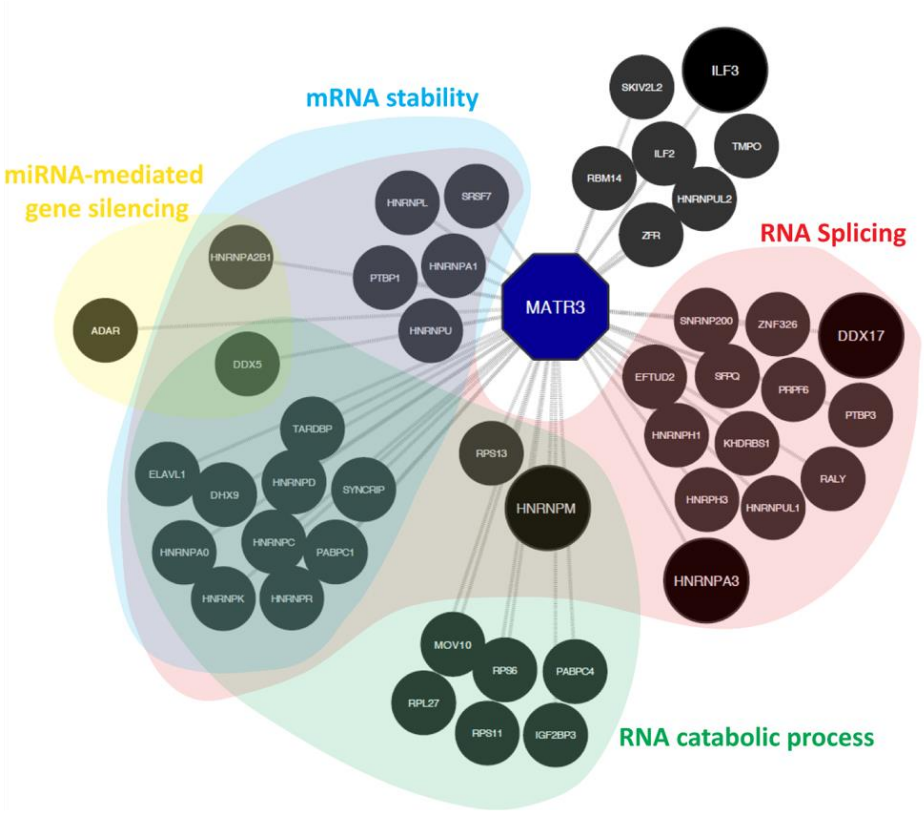

Figure 2: MATR3 protein-protein interactome and enriched biological pathways

Network representation of protein interactors of MATR3 identified in more than one mass spectrometry-based study. Node size represents number of studies that shows interaction with MATR3. The interactome is classified based on annotated biological processes.

The interactome of RBPs gives an insight into the normal cellular functions of the protein. Moreover, determining how disease-causing mutations alter the protein's interactome will help in shedding light on novel associations that contribute to the disease mechanism. Boehringer et al. [71] discovered that mutations in MATR3 changed its interactions with about 60% of its targets in a mouse motor neuron-like cell lines (NSC-34). A few of the proteins that showed higher affinity for mutant MATR3, including ALY, DDX39B and SARNP, were members of the Transcription-Export (TREX) complex involved in mediating mRNA export. These aberrant interactions postulate that mutations in MATR3 could be causing defects in nuclear mRNA export. In support

of this hypothesis, it was shown that mutations in MATR3 did in fact impede global mRNA export, suggesting that increased associations with TREX complex proteins could be the underlining mechanism [71].

In contrast, Iradi et al. [70] did not observe any changes in the strength of normal protein interactors or uncover any new interactions for mutant MATR3. The difference could be attributed, at least partially, to differences in cell systems used, C2C12 myoblasts versus NSC-34 motor neuron like cells between the two studies [70, 71]. However, Iradi et al. was the first to characterize changes to MATR3 interactome when the known functional domains are deleted. They observed that the most dramatic change in the interactome occurred when the RNA-binding domain, RRM2, was deleted, with over half of the novel interactors consisting of proteins that have low-complexity domains [70]. Interestingly, they also found that MATR3-ΔRRM2 formed spherical structures in cells, reminiscent of liquid-like droplet structures formed by TDP-43 and FUS *in vitro* [88–90]. The authors hypothesized that this phenotype is a consequence of increased association of MATR3- ΔRRM2 with other proteins that undergo phase separation to form liquid-like droplets.

Both studies were imperative in underlining the importance of MATR3's role in RNA metabolism, and also suggest that mutations in MATR3 could cause aberrant protein-protein interactions that perturbs normal RNA processing. However, it is still unclear if RNA is the mediator of these interactions. While Iradi et al. partially addressed this question through elucidation of MATR3-ΔRRM1/2 PPI, more studies delineating RNA-dependent PPIs from RNA-independent PPIs will help us in further understanding the role of MATR3 in RNA metabolism and how the mutations in MATR3 perturb this function.

1.9 Current models of MATR3-ALS

1.9.1 Lessons from cellular models of MATR3-ALS

Since the discovery of ALS-causing mutations in MATR3, several groups have developed *in vitro* and *in vivo* models to decipher how these mutations lead to disease pathogenesis. The first *in vitro* characterization of MATR3 mutants focused on assessment of the subcellular localization of the wildtype (WT) and mutant proteins in various cell lines [91]. Overexpression of MATR3 fusion protein (MATR3-YFP) revealed that both wildtype and mutant proteins confined largely to the nucleus in a punctate pattern. Diffuse cytoplasmic staining was also observed in a small percentage of cells, however in wildtype and mutants alike, indicating that it was not a mutation-specific phenomenon [91].

Other ALS-associated RBPs, particularly TDP-43 and FUS, are recruited into cytoplasmic stress granules when cells are stressed [92]. Furthermore, the mutations in these genes result in increased localization of the encoded proteins in cytoplasmic stress granules and are hypothesized to be primers to toxic aggregation [93]. Despite being structurally and functionally similar to FUS and TDP-43, MATR3 is not recruited to stress granules in cells when stressed. However, formation of stress granules following sodium arsenite treatment is markedly reduced in S85C patient-derived fibroblasts compared to control fibroblasts [94]. Furthermore, persistence of stress for 24 hours reduced the viability of cells to a higher degree in the S85C fibroblasts [94]. Thus, while MATR3 is not directly involved in stress granule formation, it may have an indirect role in regulating stress response pathways, that is disrupted by disease-causing mutations.

In neurons, it has been shown that excessive activation of NMDA receptors leads to increased phosphorylation and subsequent degradation of MATR3 [95]. Excessive activation of

NMDA is neurotoxic, suggesting that MATR3 could be one of the intermediates to neuronal death. The relationship between MATR3 levels and neurotoxicity has been explored in rat cortical neurons [96], where longitudinal fluorescence microcopy (LFM) [97], a powerful high-content imaging system that allows the analysis of survival at a single-cell level, was employed to model MATR3-mediated neurotoxicity. It was shown that either knocking down endogenous rat MATR3 or overexpressing human MATR3 induced neurotoxicity, with knockdown having a more drastic effect of neuronal survival compared to overexpression [96]. Furthermore, overexpression of MATR3 harboring ALS-linked mutations resulted in a mild increase in neurotoxicity compared to WT. Deletion of the zinc-finger (ZF2) domain mitigated MATR3 overexpression-associated neurotoxicity. While deletion of either of the RRM domains had no effect on the toxicity, RRM2 deletion elicited phase separation of MATR3 into spherical liquid-like droplets, consistent with the previously reported observation C2C12 cells [70]. Interestingly, although overall toxicity remains unchanged when the RRM2 domain is deleted, neurons with diffuse MATR3-ΔRRM2 exhibited reduced survival compared to neurons with MATR3- ΔRRM2 localized into spherical droplets. This suggests that formation of phase-separated droplets by MATR3-ΔRRM2 in this context could be neuroprotective. Additionally, overexpression of MATR3 with a deletion in putative nuclear localization signal, that caused retention of exogenous MATR3 in the cytoplasm, resulted in reduced neurotoxicity [96]. This further supports the hypothesis that the when overexpressed, MATR3 possibly leads to a toxic gain-of-function in the nucleus.

1.9.2 Lessons from animal models

While cellular models have demonstrated acute toxicity caused by MATR3 overexpression, animal models so far have revealed tissue- and mutation-specific susceptibilities

to MATR3 overexpression. Utilizing a mouse prion promoter (MoPrP) that is active in the spinal cord and muscles, transgenic MATR3 WT and F115C mutant mouse models showed robust expression in skeletal muscles, but not in spinal cords [98]. Both the WT and F115C transgenic mice showed initial muscle pathology at 2 months, which progressed noticeably by 10 months to include frequent, large vacuoles, rounded fibers, and internalized nuclei [98]. Accompanying the myopathic changes, the F115C mice also developed an overt motor phenotype, which included paresis and paralysis. The WT mice also developed muscle pathology, albeit at a much slower rate, requiring more than 20 months of aging before reaching the same severity observed in younger F115C mice. Concurrently, F115C, and later WT, transgenic mice also developed gross muscle atrophy, reminiscent to that observed in humans with MATR3-associated distal myopathy [38, 43, 44]. Although F115C mutation-carrying patients exhibit a more profound ALS/FTD phenotype rather than myopathy, these transgenic mice did not develop any neurogenic phenotypes. However, this could potentially be explained by the relatively low transgene expression levels observed in the mouse spinal cord.

1.10 Summary

Amyotrophic lateral sclerosis (ALS) is complex multigenic disease that is rising in prevalence, and there is an urgent need to develop effective therapies. In recent years, mutations in many genes, particularly those that encode RNA-binding proteins, have been discovered in ALS patients. Given the large overlap in molecular pathways perturbed between familial and sporadic ALS, elucidation of pathogenic mutations-mediated disease mechanisms is a key to identifying potential therapeutic targets. Mutations in *MATR3* were initially identified in patients with distal

myopathy with vocal and pharyngeal weakness (VCPDM), who then went on to develop neurologic symptoms and respiratory failure, leading to re-classification of disease to slow progressive ALS with distal myopathy. Furthermore, the disease spectrum also extends to frontotemporal dementia, highlighting the complex pleiotropy of MATR3-associated ALS. While *in vitro* studies so far have been important for determining cellular and neuronal susceptibility to changes in MATR3 levels, *in vivo* models are required to understand multi-system susceptibility to disease-causing mutations. Particularly, small model organisms, such as *Drosophila*, are useful for dissecting the underlying genetic factors that mediate neurodegeneration in disease models.

For over a century, *Drosophila* has been used as a versatile model system to study many biological processes and their molecular mechanisms. It offers many attractive features, including short lifespan, high number of offspring, low cost of maintenance, and relatively easy genetic tractability. A comparison of human and *Drosophila* genome revealed that over 70% of human disease-causing genes have functional homologs in *Drosophila* [99]. Furthermore, the fly nervous system, albeit simple compared to the human brain, consists of multiple neuron types that share similar cell biology to human neurons. In the past two decades, *Drosophila* has been reliably used to model neurodegenerative and neuromuscular disorders, such as Alzheimer's, Parkinson's, ALS, Huntington's disease and Muscular Dystrophies [100–104].

Here, I have developed the first *Drosophila* models of MATR3 wildtype and ALS/myopathy-associated mutations, F115C and S85C. The goal of this dissertation research is to utilize *Drosophila* and cellular model systems, combined with genetic, molecular and biochemical tools, to elucidate mechanisms of MATR3-mediated neuromuscular degeneration. Chapter 2 of this dissertation builds on the MATR3 *Drosophila* models to examine behavioral deficits driven by tissue-specific expression of MATR3 and elucidate the underlying mechanisms

in both fly and cellular models. In addition to disease-causing mutations, MATR3 also plays a more ubiquitous role in ALS pathogenesis, as evidenced by cytoplasmic neuronal inclusions in C9orf72-ALS/FTD and in sporadic ALS patient tissues [33, 36]. Chapter 3 of this dissertation describes a novel role for MATR3 in C9orf72 GGGGCC hexanucleotide repeat expansion-mediated ALS pathogenesis across patient-derived neurons and *Drosophila* models.

2.0 Chapter 2: RNA-recognition motif of Matrin-3 mediates neurodegeneration through interaction with hnRNPM

Disclosure: A revised manuscript based on this chapter is under review at Molecular Neurodegeneration

Ramesh N, Kour S, Anderson EN, Rajasundaram D, Pandey UB. "RNA-recognition motif in Matrin-3 mediates neurodegeneration through interaction with hnRNPM"

2.1 Abstract

Amyotrophic lateral sclerosis (ALS) is an adult-onset, fatal neurodegenerative disease characterized by progressive loss of upper and lower motor neurons. While pathogenic mutations in the DNA/RNA-binding protein Matrin-3 (MATR3) are linked to ALS, the molecular mechanisms underlying MATR3-mediated neuromuscular degeneration remain unclear. We generated *Drosophila* lines with transgenic insertion of human MATR3 wildtype, disease-associated variants F115C and S85C, and deletion variants in functional domains, ΔRRM1, ΔRRM2, ΔZNF1 and ΔZNF2. We utilized genetic, behavioral and biochemical tools for comprehensive characterization of our models *in vivo* and *in vitro*. Additionally, we employed *in silico* approaches to find transcriptomic targets of MATR3 and hnRNPM from publicly available eCLIP datasets. We found that targeted expression of MATR3 in *Drosophila* muscles or motor neurons shorten lifespan and produces progressive motor defects, muscle degeneration and atrophy. Strikingly, deletion of its RNA-recognition motif (RRM2) mitigates MATR3 toxicity.

We identified rump, the *Drosophila* homolog of human RNA-binding protein hnRNPM, as a modifier of mutant MATR3 toxicity *in vivo*. Interestingly, hnRNPM physically and functionally interacts with MATR3 in an RNA-dependent manner in mammalian cells. Furthermore, common RNA targets of MATR3 and hnRNPM converge in biological processes important for neuronal health and survival. We propose a model of MATR3-mediated neuromuscular degeneration governed by its RNA-binding domains and modulated by interaction with splicing factor hnRNPM.

2.2 Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disorder that leads to progressive loss of upper and lower motor neurons [105]. ALS pathogenesis is increasingly linked to mutations in genes encoding RNA-binding proteins [106–108]. Indeed, in post-mortem brain and spinal cord tissues, many such RNA-binding proteins are sequestered into pathological aggregates in the cytoplasm [109]. Preclinical models implicate dysregulation of RNA metabolism as an underlying mechanism linking RNA-binding proteins to ALS pathogenesis [15, 110–112].

Mutations in Matrin-3 (*MATR3*), a DNA/RNA-binding nuclear matrix protein, were initially discovered in North American patients with familial ALS [36]. Additional ALS-causing mutations in *MATR3* were later identified in other cohorts; currently 13 nonsynonymous point mutations are implicated in familial and sporadic ALS [39, 40, 42, 113]. One such mutation, p.S85C, is also linked to inherited distal myopathy with vocal and pharyngeal weakness (VCPDM) followed by neurogenic changes [36]. Interestingly, the same mutation is also found in patients with only myopathic symptoms [38, 43, 44]. Thus, *MATR3* is among the family of genes including

VCP, *HNRNPA1*, *HNRNPA2B1* and *SQSTM1* that cause "multisystem proteinopathy" associated with either one or a combination of ALS/frontotemporal dementia (FTD), VCPDM and Paget's disease of bone [50]. Additionally, MATR3-positive inclusions have been discovered in C9orf72-mediated- [36] and sporadic ALS [33], underscoring the link between MATR3 and neurodegenerative pathology.

As an essential nuclear matrix protein, MATR3 maintains the fibrogranular network and has functions in various aspects of RNA metabolism including alternative splicing [63, 64, 82], maintaining mRNA stability [68], transcription [68, 114] and mRNA export [71, 115]. Additionally, MATR3 interacts with SFPQ and NONO to mediate the DNA-damage response [19]. The protein is composed mostly of intrinsically disordered regions (IDRs), except for 4 functional domains: two tandem RNA-recognition motifs and two C2H2-type zinc finger domains (Fig 3A). Mutations associated with ALS do not fall within any known functional domain, however most pathogenic mutations are clustered in the protein N-terminus (Fig 3A). The MATR3 N-terminus region forms droplet-like structures in cells [116], indicating that this region is required for mediating physiological liquid-liquid phase separation of MATR3. However, the physiological role of MATR3 phase separation and its relevance to disease pathogenesis is yet to determined.

In vitro and *in vivo* studies indicate that MATR3 dosage is key to neuronal survival. In NSC-34 motor neuron-like cells, ectopic expression of wild type MATR3 and its ALS-associated mutations impedes nuclear mRNA export, including export of *FUS* and *TDP-43* mRNA [71]. Either overexpression or knockdown of MATR3 in rat cortical neurons is sufficient to negatively impact their survival [96]. In mice, homozygous knockout of *Matr3* is embryonic lethal [51]. On the other hand, transgenic overexpression of the MATR3 ALS-linked F115C mutant leads to development of severe myopathic degeneration and paralysis in an age-dependent manner [98].

The crucial remaining question is how mutations in MATR3 lead to disease pathogenesis of ALS/FTD and distal myopathy.

To address this, we developed a transgenic *Drosophila* model expressing either the human wild type or mutant (F115C and S85C) MATR3. Our model exhibits significant neuromuscular degeneration, enabling us to dissect mechanisms of MATR3-ALS and MATR3-myopathy. We show that MATR3-mediated toxicity *in vivo* is regulated by its RNA-binding domains, suggesting that mutations in MATR3 lead to disease pathogenesis through its RNA-binding functions. Furthermore, we show that MATR3 genetically and physically interacts with the splicing factor hnRNPM, which is required for mutant MATR3 to exert toxicity in *Drosophila*.

2.3 Results

2.3.1 Ubiquitous expression of *MATR3* in *Drosophila* leads to neuromuscular junction defects followed by motor dysfunction

To investigate the role of MATR3 in an *in vivo* model system, we generated transgenic *Drosophila* lines with site-specific integration of either wild type (WT) human *MATR3* or *MATR3* with ALS-associated mutations Ser85Cys (S85C) and Phe115Cys (F115C) with an N-terminal FLAG tag. We confirmed similar transgene expression levels between groups before further analysis **(Fig 3B, C)**. We then utilized the well-established UAS/GAL4 ectopic expression system to systematically assess the effect of MATR3 expression in target tissues. Constitutive ubiquitous expression of WT or mutant MATR3 in *Drosophila* caused developmental lethality, leading to complete loss of egg-to-adult viability **(Fig 3D)**. In contrast, neither the non-transgenic control

(standard w^{1118} strain crossed with Tub-Gal4 driver) nor ubiquitous expression of an irrelevant transgene, *EGFP*, caused any lethality, confirming that the phenotype was exclusively due to MATR3 expression **(Fig 3D).**

To assess our model for specific motor neuron defects, we utilized the GeneSwitch (GS) expression system induced by a mild dose of RU486 for controlled conditional expression of MATR3 through development to larval stages **(Fig 3E)**. We labeled the third-instar larval neuromuscular junction (NMJ), a well-established model for investigating neurodegeneration, for the presynaptic marker HRP to assess synaptic bouton morphology. Expression of MATR3 WT and F115C, but not S85C, reduced the number of synaptic boutons (normalized to muscle area) in the NMJ compared to controls **(Fig 3F, G)**, with F115C showing a statistically significant loss of synaptic boutons **(Fig 3G)**. To determine if ubiquitous expression of MATR3 results in motor dysfunction, we used the inducible expression system to bypass expression in developmental stages and instead conditionally express MATR3 in adults, induced by RU486 **(Fig 3E)**. Consistent with NMJ defects in larvae, conditional expression of MATR3 in adults reduced their locomotion ability compared to controls, with F115C mutant flies showing statistically significant motor defects **(Fig 3I)**. Additionally, these flies had a significantly shortened lifespan compared to controls **(Fig 3H)**. We observed a strong nuclear WT and mutant MATR3 signal with a granular expression pattern **(Fig 4)** in the *Drosophila* ventral nerve chord (VNC), suggesting that pathogenic mutations do not perturb the subcellular distribution of MATR3 protein in our model.

We next performed a pulse-chase experiment to analyze MATR3 degradation kinetics *in vivo*. We briefly induced MATR3 expression using RU486, then chased protein turnover using immunoblotting. Since MATR3 associates with both detergent (NP40)-soluble and insoluble fractions, we analyzed protein stability in both fractions. NP40-soluble wild type and F115C

MATR3 exhibited similar degradation kinetics **(Fig 3J),** with the F115C mutant protein exhibiting a slightly longer half-life ($t_{1/2}$ = 44 hours) compared to WT ($t_{1/2}$ = 42 hours). The NP40-soluble S85C mutant, on the other hand, accumulated for the first 12 hours then degraded, for an overall increased half-life of the protein ($t_{1/2}$ = 51 hours) **(Fig 3J)**. However, NP40-insoluble MATR3 exhibited different kinetics: insoluble F115C and S85C mutants accumulated before degradation and took longer to degrade compared to the WT protein **(Fig 3K)**. This suggests that pathogenic mutations in MATR3 result in decreased protein turn-over, leading to more stable insoluble forms of the protein that might disrupt its nuclear functions.

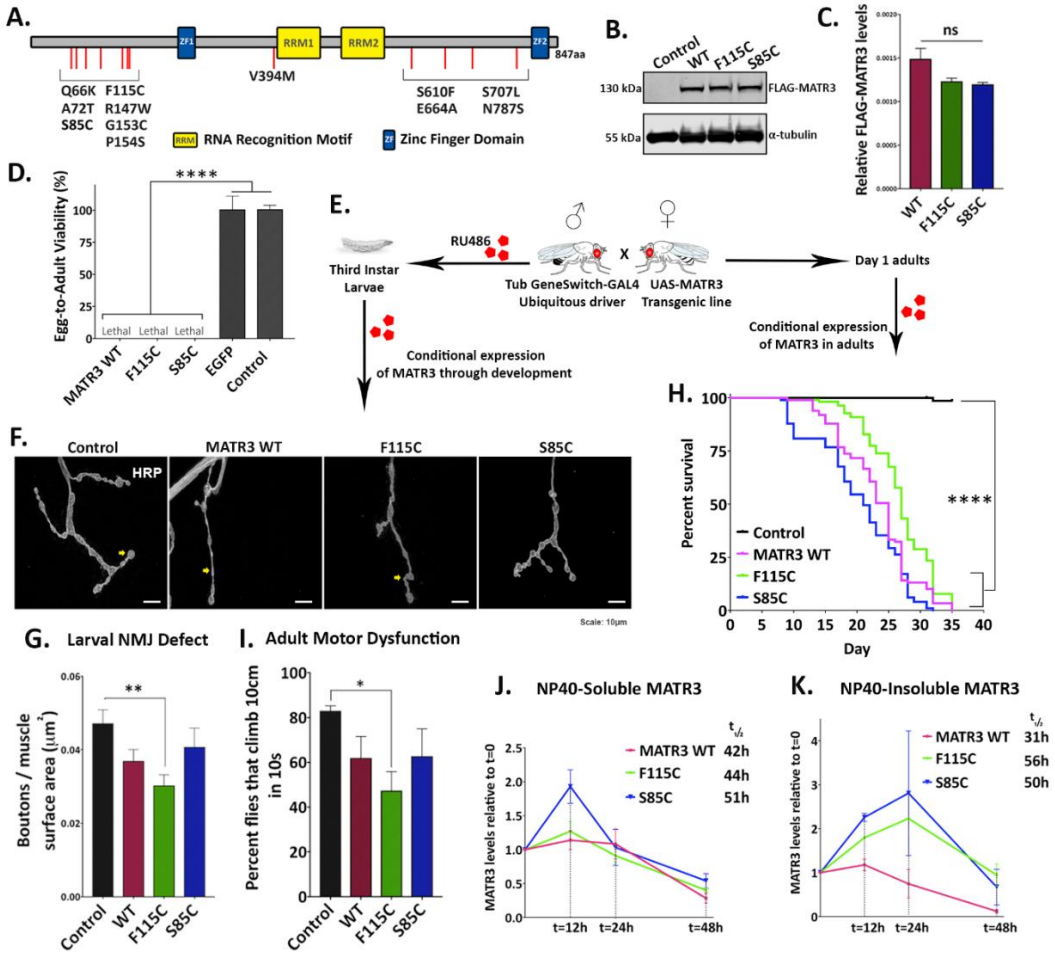

Figure 3: Ubiquitous expression of MATR3 is toxic in Drosophila

(A) Schematic diagram of MATR3 protein domain architecture, consisting of two tandem RNA-recognition motifs (RRM1 and RRM2) and two Zinc Finger motifs (ZF1 and ZF2). ALS-associated mutations in MATR3 are spread across intrinsically disordered regions of the protein. (B) Immunoblot showing transgenic expression of UAS-FLAG-MATR3 wildtype and mutants – F115C and S85C (C) Quantification of replicate western blots to confirm equal transgenic expression levels. α-tubulin is used as loading control (n=3; *One-way ANOVA*) (D) Constitutive ubiquitous expression of MATR3 wildtype and mutants, driven by Tub-Gal4, is toxic in *Drosophila* leading to complete lethality. Quantification of egg-to-adult viability showed that control (driver-alone) and transgenic UAS-EGFP expression had no effect on viability (n=3, *One-way ANOVA*) (E) Schematic of conditional expression of MATR3 during development and in adults using the inducible driver, TubGS-Gal4, that is activated by RU486. (F) Representative IF images of third-instar larval neuromuscular junction (NMJ) immunostained for presynaptic marker, HRP. Yellow arrows point to the synaptic boutons. (G) Quantification of boutons, normalized to surface area, showed reduced number of boutons in MATR3-expressing larvae (n=8; *One-way ANOVA*) (H) Kaplan-Meier survival curve of adults ubiquitously expressing MATR3 showing statistically-significant reduction in longevity of MATR3-expressing flies compared to driver-alone control flies (n=80; *Log-Rank Mantel-Cox test*) (I) Quantification of motor dysfunction in day-12 adults expressing MATR3, induced on day-1 (n=30, *One-way ANOVA*). (J) Quantification of NP40-soluble and (K) NP40-insoluble MATR3 degradation *in vivo* by a pulse-chase assay at t=0, 12, 24 and 48 hours following inhibition of transgene production (n=3 per group). NP40-soluble and insoluble MATR3 mutants have a higher half-life compared to MATR3 WT. Error bars indicate S.E.M. *p<0.05, **p<0.01, ****p<0.0001

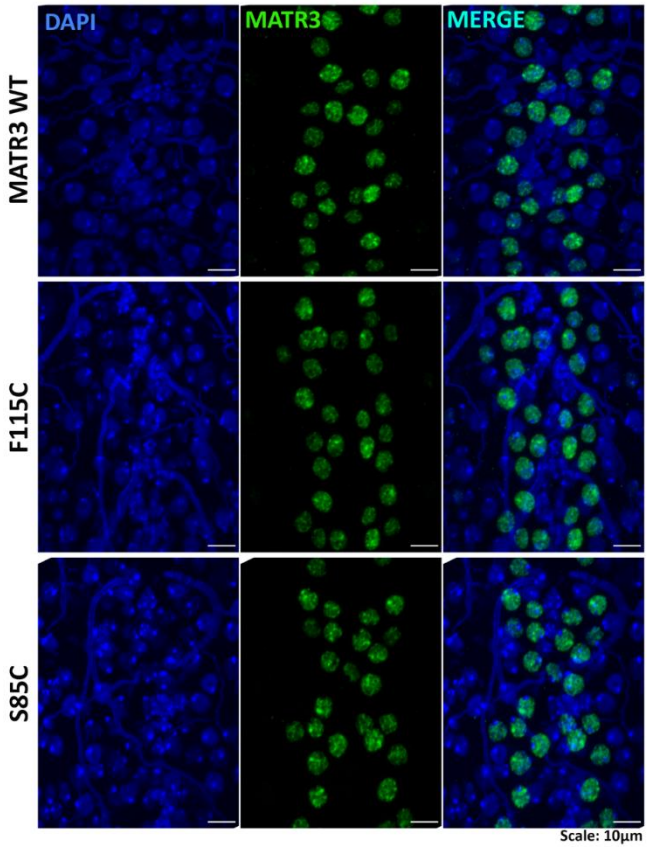

Figure 4: MATR3 localizes in nucleus in Drosophila model

Larval ventral nerve chord (VNC) immunostained for FLAG in larvae expressing FLAG-MATR3 WT and mutants, F115C and S85C, in motor neurons. FLAG-MATR3 WT and mutants localize to the nucleus (Dapi) in VNC cells

2.3.2 Muscles exhibit higher vulnerability to the ALS/myopathy-linked MATR3 S85C mutation

To dissect the tissue-specific roles of MATR3, we expressed MATR3 in disease-relevant tissues, including motor neurons and muscles. While not as potent as in the ubiquitous expression paradigm, expression of MATR3 in either motor neurons **(Fig 5A, Fig 6A)** or muscles **(Fig 5B, Fig 6B)** significantly shortened the lifespan compared to controls, indicating that tissue-specific MATR3 expression results in neuronal and muscular degeneration, respectively. Furthermore,

expression of MATR3 in *Drosophila* eyes did not cause external or internal eye degeneration, even when aged (**Fig 7).** This finding suggests that MATR3-mediated toxicity is not systemic, and that motor neurons and muscles may be specifically susceptible to MATR3 expression.

To examine whether reduced longevity is supported by concurrent motor dysfunction, we tested the locomotion abilities of flies expressing MATR3 in muscles. At day 2, MATR3-expressing flies do not exhibit any obvious motor deficit **(Fig 5C, D).** However, at day 15 of the adult lifespan, while MATR3 flies retained their climbing ability, the speed at which they climbed was significantly impaired compared to controls **(Fig 5D).** Interestingly, the motor dysfunction in S85C mutant flies was exacerbated in an age-dependent manner; by day 30, approximately 25% entirely lost their ability to climb **(Fig 5C).** Among the flies that did climb, their climbing velocity remained significantly lower compared to that of controls **(Fig 5D).** This suggests that while loss of lifespan induced by tissue-specific MATR3 expression occurs at a later age, climbing defects start to manifest as early as day-15, suggesting a slow, progressive deterioration.

To investigate if these locomotion defects are caused by morphological defects in the muscles, we performed hematoxylin and eosin (H&E) staining on cross-sections of the dorsal longitudinal muscles (DLMs) in the thorax. DLMs in MATR3-expressing flies appeared smaller compared to those of controls, most likely due to muscle atrophy in these animals **(Fig 5E, F)**. Complementary to motor dysfunction, the atrophy phenotype was particularly severe in S85C mutant flies, which exhibited enlarged gaps between muscle fibers, indicating muscle degeneration **(Fig 5E)**. F115C mutant flies showed similar degeneration, but to a lesser degree than the S85C mutant flies **(Fig 5E)**. This finding demonstrates that muscle-targeted expression of MATR3 leads to adult-onset degeneration that is progressive in the pathogenic S85C mutant.

We next utilized this model to identify biochemical alterations caused by MATR3 mutations that lead to acquisition of toxic properties, particularly solubility of MATR3 in NP40-containing buffers. In *Drosophila* muscles, MATR3 WT and F115C mutant proteins were distributed almost evenly between the soluble and insoluble fractions **(Fig 5G, H).** However, the S85C mutation drastically reduced the solubility of the protein, with most of the protein accumulating in the insoluble fraction **(Fig 5G, H).** Multiple disease-associated proteins can misfold into toxic conformations that render them insoluble [109, 117, 118]; our finding suggests that the S85C mutation might lead to disease pathogenesis through similar mechanisms.

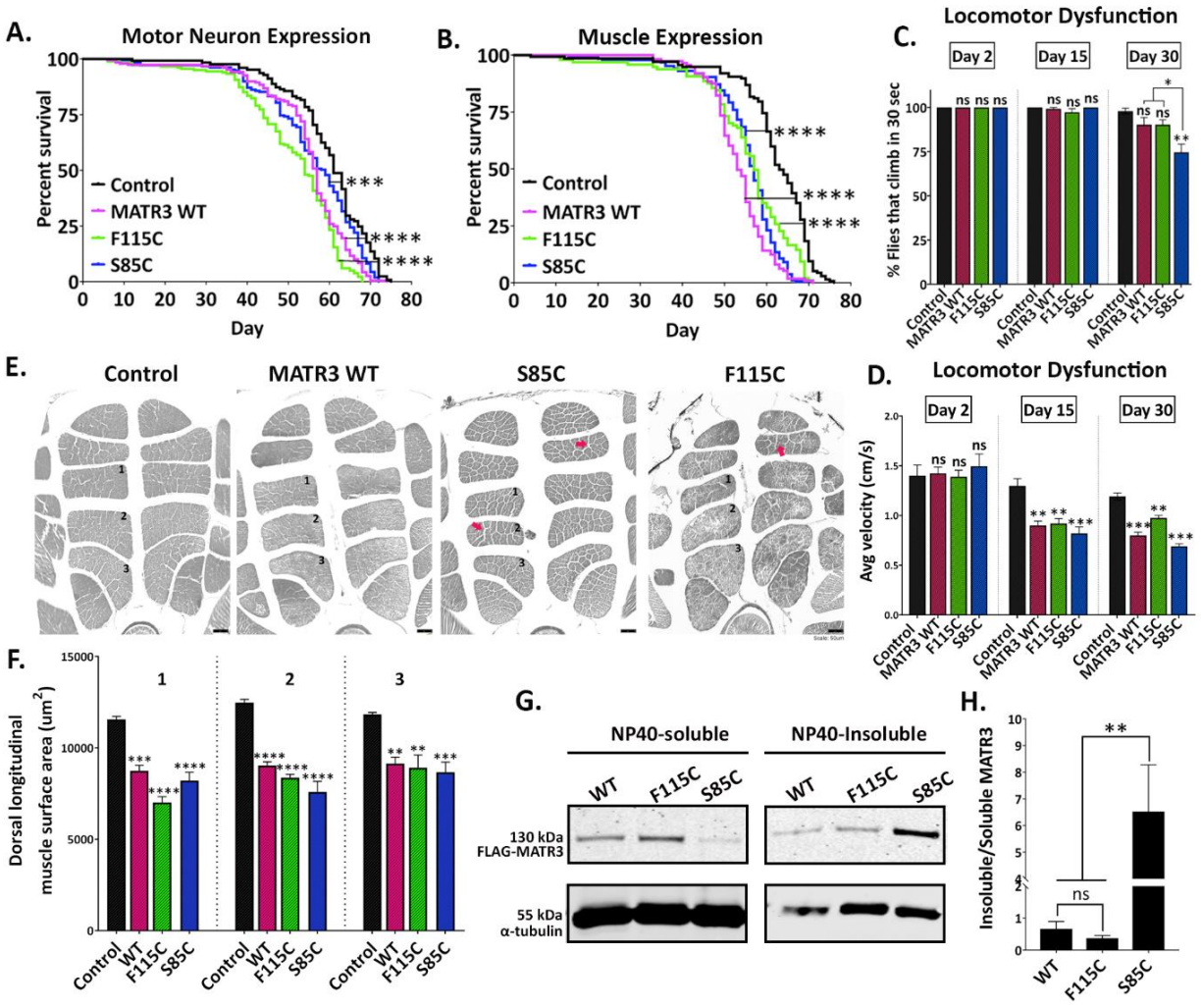

Figure 5: Tissue-specific expression of MATR3 leads to progressive degeneration in motor neurons and muscles

(A) Kaplan-Meier survival curve in flies constitutively expressing MATR3 in motor neurons, driven by OK371-Gal4, and **(B)** in muscles, driven by MHC-Gal4. Expression of MATR3 WT and mutants in either tissue moderately, but significantly, reduced longevity of the flies compared to driver-alone control (n=100; *Log-Rank Mantel-Cox test*) **(C)** Quantification of the percent flies that can climb and **(D)** average velocity at which they climb at day-2, day-15 and day-30 of adult lifespan (n=50; *One-way ANOVA*) **(E)** H&E stained indirect flight muscles in transverse thoracic cross-sections in 30-day old flies expressing MATR3 in muscles. Red arrows indicate enlarged gaps between muscle fibers in the S85C and F115C-expressing flies indicating muscle degeneration. **(F)** Quantification of the cross-sectional area of dorsal-longitudinal muscle (DLM) segments 1, 2 and 3. DLMs of MATR3-expressing flies showed lower cross-sectional area compared to driver-alone control, with F115C and S85C flies exhibiting more pronounced atrophy (n=4 hemi-segments per group; *One-way ANOVA*). **(G)** Immunoblot showing the NP40-soluble and NP40-insoluble fractions of MATR3 in *Drosophila* thorax lysates from flies expressing MATR3 in muscles. **(H)** Quantification of insoluble/soluble MATR3 protein fraction. S85C mutant protein is significantly more enriched in the insoluble fraction (n=4-6 per group; *One-way ANOVA*). α-tubulin is used as loading control. Error bars indicate S.E.M. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

A.

Motor neuron expression	Days			
	Control	MATR3 WT	F115C	S85C
25% Death	56	52	48	44
50% Death	61	57	59	54
75% Death	67	61	65	60

B.

Muscle expression	Days			
	Control	MATR3 WT	F115C	S85C
25% Death	58	49	65	50
50% Death	62	51	57	57
75% Death	68	56	60	61

Figure 6: MATR3 expression in Drosophila motor neurons and muscles reduces longevity of flies

(A) Tabular representation of number of days it takes for 25%, 50% and 75% death in flies expressing MATR3 WT and mutants in motor neurons and **(B)** muscles

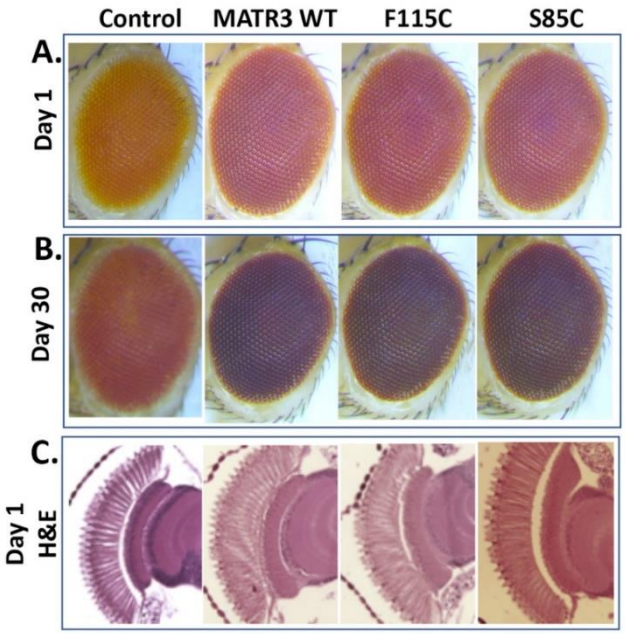

Figure 7: MATR3 expression in Drosophila eyes does not cause external or internal degeneration

(**A**) External eye phenotypes of flies expressing MATR3 in the eye, driven by GMR-gal4, at day 1 post-eclosion and **(B)** day 30 post-eclosion. Expression of either WT or mutant MATR3 did not cause any external eye degeneration at early or later time points. **(C)** H&E stained photoreceptors in cross-sections of *Drosophila* eyes. Expression of MATR3 did not result in any internal degeneration.

2.3.3 RNA-binding domains of MATR3 drive toxicity in vivo

We sought to determine which functional domain(s) of MATR3 is responsible for mediating toxicity. We generated transgenic MATR3 lines with deletion mutations in each of the four functional domains: ΔRRM1, ΔRRM2, ΔZNF1 and ΔZNF2 **(Fig 8A)**. When ubiquitously expressed, deletion of the RRM2 domain completely rescued developmental lethality, yielding adults expressing MATR3 **(Fig 8B)**. RRM1 deletion also partially rescued developmental toxicity **(Fig 8B)**. However, deletion of either of the zinc-finger domains (ZNF1 or ZNF2) did not suppress toxicity **(Fig 8B)**. To determine if RRM1/2 deletion similarly modulates the NMJ defect, we induced conditional MATR3 expression with RU486 and labeled the third instar larval NMJs for HRP. Importantly, RRM2 deletion was sufficient to strongly rescue the NMJ defects caused by MATR3 **(Fig 8C),** restoring the number of synaptic boutons to near-control levels **(Fig 8D)**. We then moved to an adult expression paradigm to evaluate if RRM2 deletion retains its rescue ability. In adults, deletion of RRM1 and RRM2, independently, significantly extended the lifespans of flies expressing MATR3 **(Fig 8E)**. However, unlike in developmental toxicity, ΔRRM2 adults retain some toxicity, with a shorter lifespan compared to controls **(Fig 8E)**. Taken together, adult longevity analyses suggest that MATR3 toxicity may be mediated by both RRM1 and RRM2 RNA-binding domains. ZNF1/2 deletion, on the other hand, exacerbated toxicity in the adults by significantly shortening their lifespan **(Fig 8F)**.

These observations prompted us to examine the role of RRM1 and RRM2 domains in mediating toxicities associated with disease-causing mutations. To test this, we generated mutant transgenic lines including F115C-ΔRRM1, F115C-ΔRRM2, S85C-ΔRRM1 and S85C-ΔRRM2. RRM2 deletion was equally successful in rescuing developmental toxicity in flies expressing pathogenic mutant MATR3 **(Fig 9A)**. Interestingly, ΔRRM1, while mildly protective on its own,

did not have any rescue effect on mutant MATR3 developmental toxicity **(Fig 9B)**. A possible explanation for this could be that F115C and S85C mutant MATR3 exert higher developmental toxicity compared to WT MATR3 *in vivo*, and thus, any mild protective effect of RRM1 deletion in developmental stages is negated, or perhaps irreversible. However, in adults, both RRM1 and RRM2 deletion significantly extended the lifespan of MATR3 mutant flies **(Fig 9C, D)**.

We then assessed the effect of RRM1 and RRM2 deletion on total levels of MATR3 WT and mutants. We observed that, especially in the WT background, the total levels of ΔRRM1 and ΔRRM2 were significantly higher compared to full-length MATR3 **(Fig 10)**, indicating that mitigated toxicities in ΔRRM1 and ΔRRM2-expressing flies is not a consequence of reduced protein expression. We asked if removal of either RRM1 or RRM2 domains has any effect on modulating solubility of the MATR3 protein. Analysis of insoluble and soluble fractions showed that RRM2 deletion significantly increased MATR3 solubility **(Fig 9E, F, G, H).** Deletion of the RRM1 domain, however, had a milder impact on increasing MATR3 solubility. Interestingly, while RRM2 deletion decreased the insolubility of both WT and mutant MATR3, the relative decrease in insolubility was higher in the WT background compared to the mutant background **(Fig 9F, G, H).** This correlated with our longevity analyses, where RRM2 deletion in the WT background extended the lifespan of MATR3 flies more than in the mutant background **(Fig 8E, 9C, D).** Overall, our results highlight an imperative role of the MATR3 RNA-binding domains, particularly the RRM2 domain, in mediating MATR3 WT and mutant toxicity. Thus, pathogenic mutations in MATR3 could result in a toxic gain-of-function in motor neurons and muscles, resulting from aberrant regulation of RNA targets that bind to MATR3.

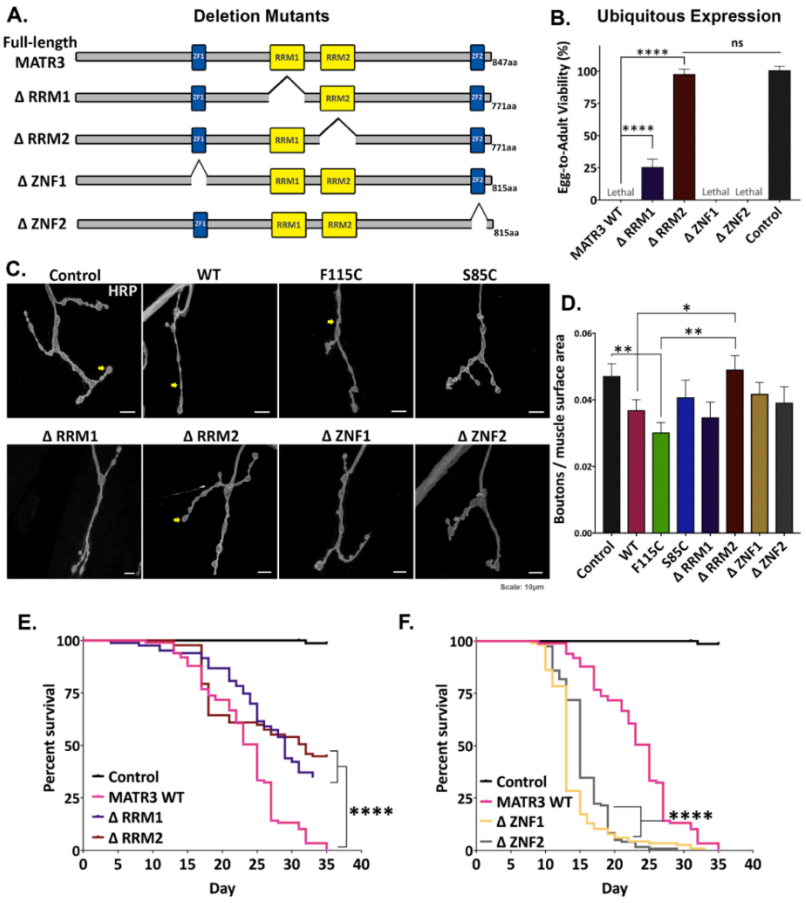

Figure 8: MATR3 toxicity is mediated through its RNA-binding domains

(A) Schematic diagram of MATR3 protein domain architecture in deletion mutant-transgenic flies, where each of the known functional domains are deleted **(B)** Quantification of egg-to-adult viability in flies ubiquitously expressing MATR3 deletion mutants, driven by Tub-Gal4 driver. Constitutive ubiquitous expression of MATR3 deletion mutants showed partial rescue by ΔRRM1 and complete rescue by ΔRRM2 (n=3, *One-way ANOVA*) **(C)** Representative immunofluorescence images of third-instar larval neuromuscular junction (NMJ) immunostained for presynaptic marker, HRP. Yellow arrows point to the synaptic boutons. **(D)** Quantification of number of synaptic boutons, normalized to surface area, showed that ΔRRM2 restored number of synaptic boutons back to near-control levels *(*n=8, *One-way ANOVA)* **(E)** Kaplan-Meier survival curve of adults ubiquitously expressing MATR3 deletion mutants under the conditional driver. Both ΔRRM1 and ΔRRM2 significantly extended MATR3 lifespan, while **(F)** ΔZNF1 and ΔZNF2 further reduced MATR3 lifespan *(*n=100, *Log-rank Mantel-Cox test)* Error bars indicate S.E.M. *p<0.05, **p<0.01, ****p<0.0001

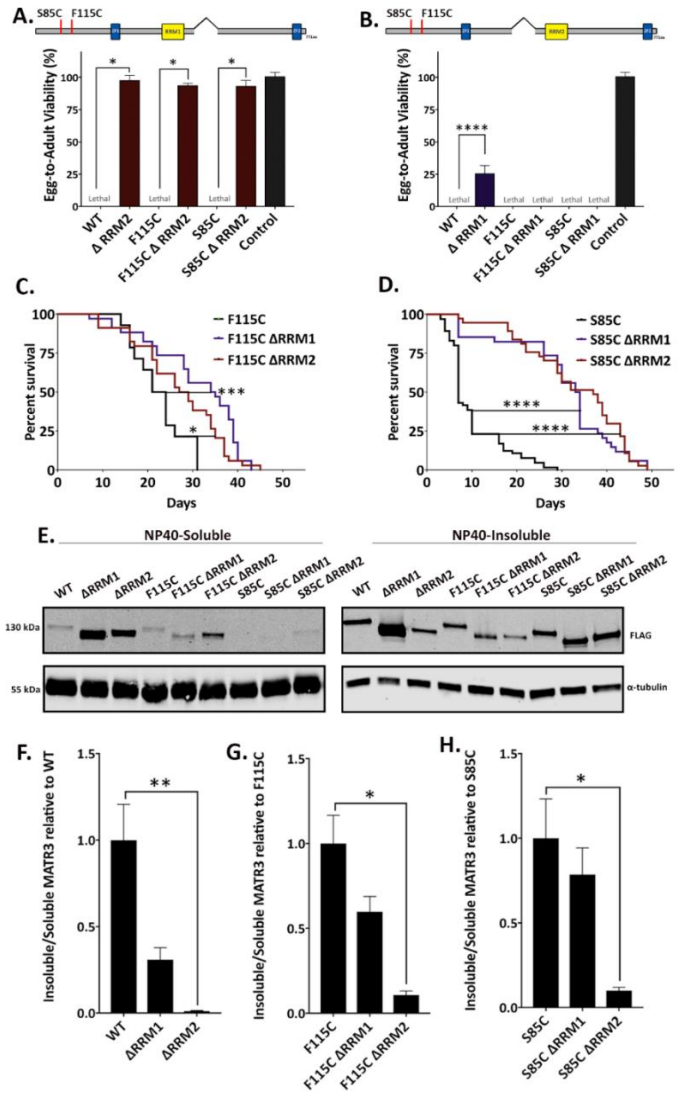

Figure 9: RRM2 deletion recues mutant MATR3 toxicity in vivo

(A) Quantification of egg-to-adult viability in flies expressing double mutants: F115C ΔRRM2 and S85C ΔRRM2. Constitutive ubiquitous expression of ΔRRM2 in the mutant background completely rescued mutant MATR3-mediated developmental toxicity (n=3, *One-way ANOVA*) **(B)** Quantification of egg-to-adult viability in flies expressing double mutants: F115C ΔRRM1 and S85C ΔRRM1. Constitutive ubiquitous expression of ΔRRM1 in the mutant background is not sufficient to rescue mutant MATR3-mediated developmental toxicity **(C)** Kaplan-Meier survival curve of adults ubiquitously expressing MATR3 deletion mutants in F115C background and **(D)** S85C background under conditional driver (TubGS-Gal4). ΔRRM1 and ΔRRM2 significantly increased longevity of adults expressing the MATR3 mutants *(*n=50, *Log-rank Mantel-Cox test)*. **(E)** Immunoblot of NP40-soluble and NP40-insoluble fractions of MATR3 from lysates of flies expressing full-length WT and mutant MATR3, and corresponding deletion mutants. α-tubulin is used as loading control. **(F)** Quantification of the insoluble/soluble fractions of full-length MATR3 and deletion mutants, ΔRRM1 and ΔRRM2, in the background of WT, **(G)** F115C and **(H)** S85C. Deletion of RRM2 significantly increased the solubility of MATR3 WT, F115C and S85C (n=4 per group; *One-way ANOVA)* Error bars indicate S.E.M. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

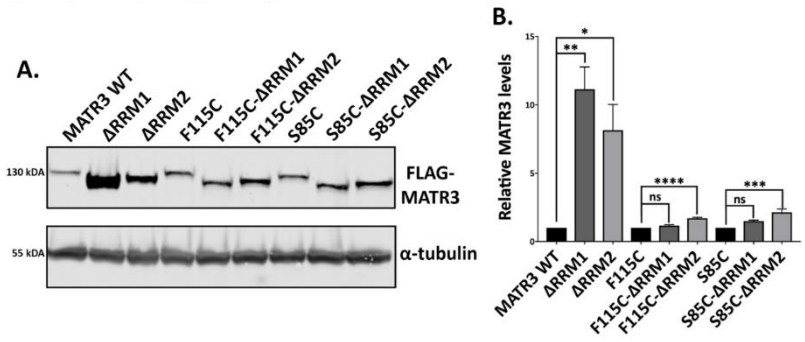

Figure 10: Total MATR3 protein expression levels in flies expressing ΔRRM1 and ΔRRM2 variants

(A) Immunoblot showing MATR3 protein levels in flies ubiquitously expressing MATR3 WT, F115C and S85C and corresponding RRM1 and RRM2 deletions. **(B)** Quantification of replicate western blots shows increased levels on ΔRRM1 and ΔRRM2-MATR3 compared to full-length MATR3. (n=4; *One-way ANOVA*). Error bars indicate S.E.M. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001

2.3.4 Rump, a homolog of hnRNPM, is a strong modifier of MATR3 toxicity in vivo

To identify modifiers of MATR3 toxicity, we focused on proteins that have been identified as high-confidence interactors of MATR3 in two or more high-throughput studies [69–71]. Most protein-protein interactors were RNA-binding proteins involved in multiple aspects of RNA metabolism, primarily splicing. We screened the interactome for proteins previously implicated in neurodegenerative diseases, including hnRNP family proteins. We obtained publicly available RNAi lines for the selected candidate genes and screened for those that do not cause any intrinsic toxicity with ubiquitous expression **(Appendix Table 1)**. To perform the screen, we combined MATR3 WT and mutants with each candidate RNAi line, crossed them with the ubiquitous driver and looked for viable adults in the progeny **(Fig 12A).**

We identified Rumpelstiltskin (*rump*), the *Drosophila* homolog of human *HNRNPM*, as a strong modifier of MATR3 toxicity **(Fig 12A, B) (Appendix Table 1).** hnRNPM is an RNA-binding protein that binds to pre-mRNA and splicing regulator complexes to regulate alternative

splicing [119, 120]. Furthermore, hnRNPM is one of the strongest interactors of MATR3, identified through multiple high-throughput protein-protein interaction studies [69–71, 76]. Knocking down *rump* **(Fig 11A, B)** significantly rescued S85C toxicity, as evident from rescue of egg-to-adult viability in S85C flies also expressing rump RNAi **(Fig 12B)**. Interestingly, rump was unique in suppressing MATR3 toxicity, suggesting a potentially prominent role for hnRNPM in mediating MATR3 toxicity *in vivo* **(Appendix Table 1).** To validate this interaction, we knocked down *rump* in adults conditionally expressing MATR3 and assessed longevity. rump knockdown (KD) significantly increased the lifespans of flies expressing F115C- or S85C-mutant MATR3 **(Fig 12C, D, F).** rump KD on its own did not exert intrinsic toxicity during development **(Fig 12B)**, however it did promote decreased longevity in adults compared to controls **(Fig 11C).** Interestingly, rump KD did not have any obvious effect on WT lifespan **(Fig 12E, F)**, indicating a role for hnRNPM in mediating toxicity caused by pathogenic mutations in MATR3. This evidence further highlights the disease relevance of a MATR3 and hnRNPM genetic interaction *in vivo.* Furthermore, rump KD decreased the insolubility of MATR3 S85C in muscles **(Fig 12G, H)**, suggesting a potential mechanism for alleviating MATR3 toxicity *in vivo*.

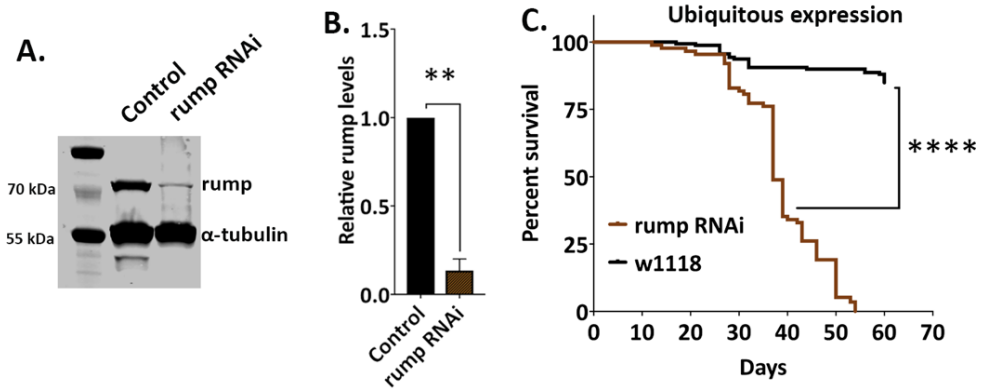

Figure 11: Knockdown of rump in Drosophila reduces adult longevity

(A) Immunoblot showing reduced rump protein levels in rump RNAi line. **(B)** Quantification of replicate western blots to confirm reduced rump levels in rump RNAi flies, driven by Tub-Gal4, compared to driver-alone control (n=3; *Kruskall-Wallis test*). **(C)** Kaplan-Meier survival curve of adults ubiquitously expressing rump RNAi. Knockdown of

rump conditionally in adults reduced longevity of flies compared to driver-alone control *(*n=50, *Log-rank Mantel-Cox test)* Error bars indicate S.E.M. **p<0.01, ****p<0.0001

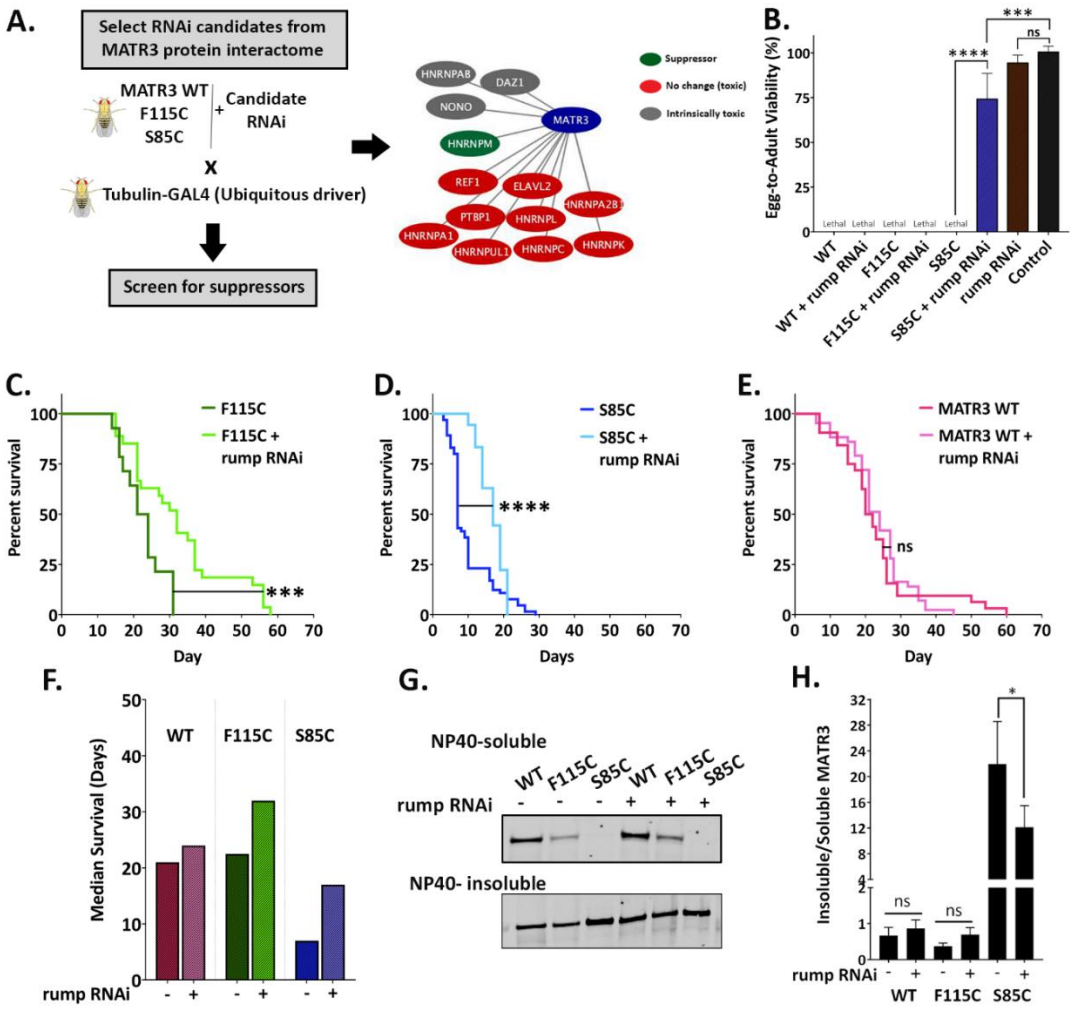

Figure 12: Rumpelstiltskin, homolog of human hnRNPM, is a strong modifier of MATR3 toxicity

(A) Schematic of candidate RNAi screen in MATR3 *Drosophila* model. Diagrammatic representation of MATR3 protein interactors that were screened for suppression of MATR3-mediated toxicity. hnRNPM homolog in *Drosophila*, Rumpelstiltskin (rump), is a unique suppressor of MATR3 toxicity. (B) Quantification of egg-to-adult viability in flies expressing MATR3 with and without rump RNAi. Constitutive ubiquitous expression of S85C with rump knockdown rescues S85C-mediated developmental toxicity (n=3, *One-way ANOVA*). (C) Kaplan-Meier survival curve of adults ubiquitously expressing MATR3 S85C, (D) MATR3 F115C and (E) MATR3 WT, with and without rump knockdown. Knocking down rump significantly improved longevity of F115C and S85C-expressing flies, and had no obvious effect on WT longevity (n=50, *Log-rank Mantel-Cox* test). (F) Median survival of MATR3-expressing flies with and without rump KD showed increased median survival upon rump KD in F115C- and S85C-expressing flies (G) Immunoblot of NP40-soluble and NP40-insoluble fractions of MATR3 in *Drosophila* thorax expressing MATR3, with or without rump RNAi, in the muscles. α-tubulin is used as loading control. (H) Quantification of the insoluble/soluble MATR3 protein fractions. Rump KD decreased insolubility of MATR3 S85C (n=6 per group; *One-way ANOVA*). Error bars indicate S.E.M. ***p<0.001, ****p<0.0001

2.3.5 hnRNPM genetically and physically interacts with MATR3 via its RRM2 domain in mammalian cells

To identify functional interactions between MATR3 and hnRNPM, we turned to mammalian cell systems. Mouse myoblast C2C12 cells are particularly susceptible to MATR3 overexpression, as ectopic expression of MATR3 results in cytoplasmic mislocalization of MATR3 and accumulation into cytoplasmic granules in a subset of cells **(Fig 13A, B).** We also observed that overexpression of the S85C mutation led to significantly increased accumulation of MATR3 into cytoplasmic granules compared to WT overexpression **(Fig 13B)**. Moreover, we observed clear colocalization between MATR3 and endogenous hnRNPM in the cytoplasmic granules, suggesting that MATR3 overexpression concurrently leads to mislocalization of hnRNPM and sequestration into cytoplasmic granules **(Fig 13A).** Interestingly, overexpression of F115C- and S85C-mutant MATR3 caused a higher degree of sequestration of hnRNPM into the cytoplasmic granules compared to WT **(Fig 13C).** This correlates with selective suppression of MATR3-mutant toxicity by hnRNPM KD in our *Drosophila* models, supporting the hypothesis that hnRNPM plays a role in disease pathogenesis caused by mutations in MATR3.

While hnRNPM was shown to interact with wildtype MATR3 through mass spectrometric studies [69–71], we sought to assess the physical interaction between hnRNPM and mutant MATR3. Co-immunoprecipitation revealed a physical interaction between hnRNPM and WT MATR3 as well as F115C and S85C mutants **(Fig 13D).** Interestingly, deletion of the RRM2 domain interrupted this interaction, suggesting that the RRM2 domain is required for mediating interaction between MATR3 and hnRNPM **(Fig 13D).** To investigate if this interaction is RNA-dependent, we treated the immunoprecipitated lysate with RNaseA to degrade the RNA. Treatment with RNaseA decreased the interaction between MATR3 and hnRNPM in mammalian cells **(Fig**

13E), suggesting that these proteins interact, at least partially, through binding to shared RNA targets.

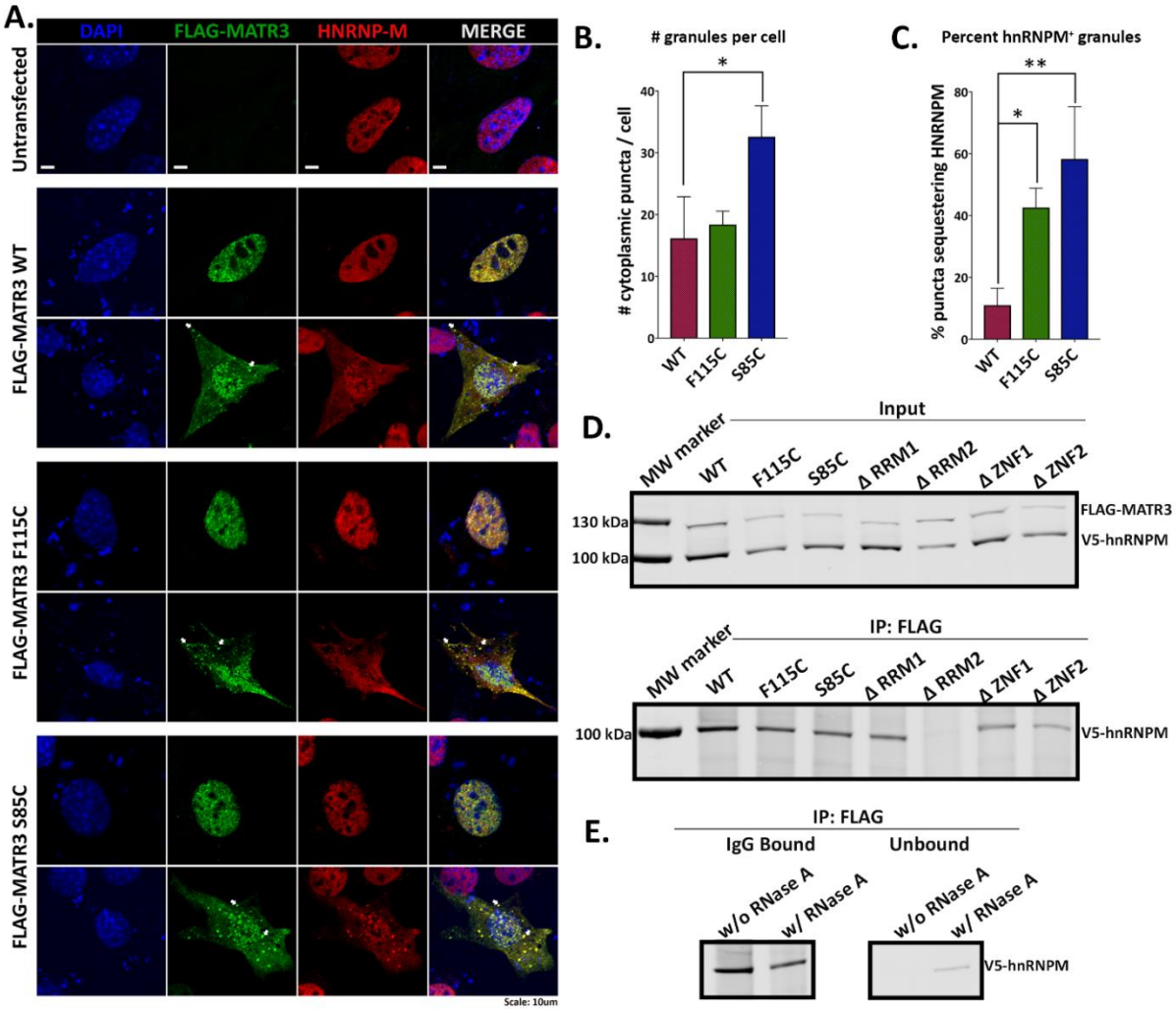

Figure 13: hnRNPM physically interacts with MATR3 in mammalian cells

(A) C2C12 cells ectopically expressing WT and mutant FLAG-MATR3 showed colocalization of FLAG-MATR3 (green) and endogenous hnRNPM (red) in the nucleus and cytoplasm. White arrows point to sequestration of endogenous hnRNPM with FLAG-MATR3 in cytoplasmic granules. **(B)** Quantification of number of cytoplasmic granules per cell showed that cell expressing F115C and S85C mutant MATR3 had increased formation of cytoplasmic granules (n=6 per group; *One-way ANOVA*). **(C)** Quantification of percentage cytoplasmic granules that are positive for hnRNPM. Cells ectopically expressing mutant MATR3 showed increased sequestration of hnRNPM in cytoplasmic granules compared to WT (n=6 per group; *One-way ANOVA*) **(D)** Immunoblot of co-immunoprecipitation of FLAG-MATR3 from HEK293T cells expressing FLAG-MATR3 and V5-hnRNPM. Immunoprecipitation with FLAG antibody showed V5-hnRNPM pulled down with FLAG-MATR3 WT, F115C and S85C. RRM2 deletion disrupts interaction between FLAG-MATR3 and V5-hnRNPM. **(E)** Immunoblot of lysates immunoprecipitated with FLAG antibody and treated with RNase A. Treatment with RNase A decreases interaction between FLAG-MATR3 WT and V5-hnRNPM (left). Elution of V5-hnRNPM in the unbound fraction after RNaseA treatment (right). Error bars indicate S.E.M. *p<0.05, **p<0.01

2.3.6 MATR3 and hnRNPM share common transcriptomic targets

We hypothesized that MATR3 and hnRNPM functionally interact to regulate metabolism of shared RNA targets and that dysregulation of these transcripts leads to disease pathogenesis. We employed an *in-silico* approach to further elucidate this functional interaction. We compared published eCLIP datasets [121] for MATR3 and hnRNPM from two different cell types, K562 (lymphocytes) and HepG2 (hepatocytes), and mined transcripts bound by both proteins **(Appendix Table 2).** MATR3 and hnRNPM share appreciable overlap in the transcripts that they bind **(Fig 14A, B) (Appendix Table 2).** In K562 cells ~46% of MATR3-bound transcripts are also bound by hnRNPM **(Fig 14A)**; in HepG2 cells, ~26% of MATR3-bound transcripts are also bound by hnRNPM **(Fig 14B).** Gene ontology analysis of the common targets **(Appendix Table 3)** revealed that the top 20 most enriched unique biological processes shared between the two cell types included significant processes such as neurogenesis, proteasomal protein ubiquitination, histone modification & chromosome organization **(Fig 14C,D).**

Furthermore, gene ontology analysis on the basis of disease associations indicated that the shared targets are enriched in genes associated with neurodevelopmental disorders **(Appendix Table 3).** These results indicate that MATR3 and hnRNPM both bind to, and thus may co-regulate, transcripts that have important functions in nervous system development and maintaining cellular/neuronal health. Thus, it is likely that aberrant regulation of these transcripts, and consequently these processes, caused by mutations in MATR3 result in disease pathogenesis.

We sought to further validate this hypothesis in our *in vivo* model. We rationally selected candidate targets from the MATR3-hnRNPM shared transcriptome, focusing on candidates that have previously been shown by other studies to be regulated by MATR3 in mammalian cells [19, 63, 121], and additionally exhibit disease-relevance in ALS and other neurodegenerative disorders.

Assessment of levels of candidate transcripts, including Dystrophin (*Dys*), Ataxin-1 (*Atx-1*), Adenylate kinase 3 (*Adk3*) and Sialyltransferase (*SiaT*), revealed significantly increased mRNA levels in mutant MATR3-expressing flies compared to control **(Fig 15).** Interestingly, we observed that levels of these transcripts were altered in flies expressing either one or both MATR3 mutations, and unchanged in flies expressing wildtype MATR3, further supporting a role for the physical and genetic interaction between MATR3 and hnRNPM in the disease pathogenesis of mutant MATR3.

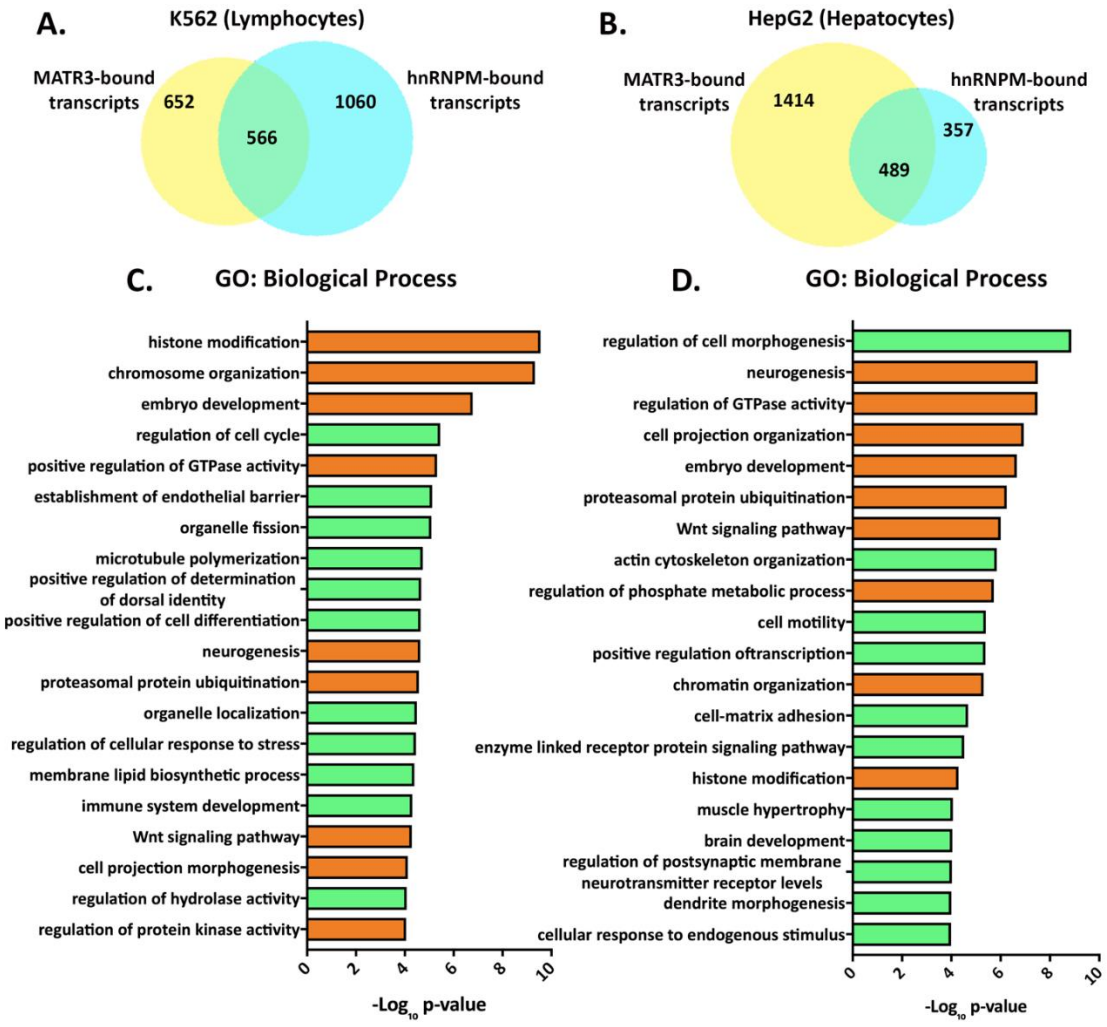

Figure 14: MATR3 and hnRNPM share transcriptomic targets

(A) Venn diagram representing transcriptomic targets from *in silico* analysis of MATR3 eCLIP and hnRNPM eCLIP (ENCODE) in K562 cells (lymphocytes) and **(B)** HepG2 cells (hepatocytes). MATR3 and hnRNPM share

transcriptomic targets in both cell types. **(C)** Top 20 unique GO:Biological Process terms that are enriched in gene ontology assessment of shared transcriptomic targets from K562 cells and **(D)** HepG2 cells. Green bars indicate biological processes unique to the cell type. Orange bars indicate biological processes commonly enriched in both cell types.

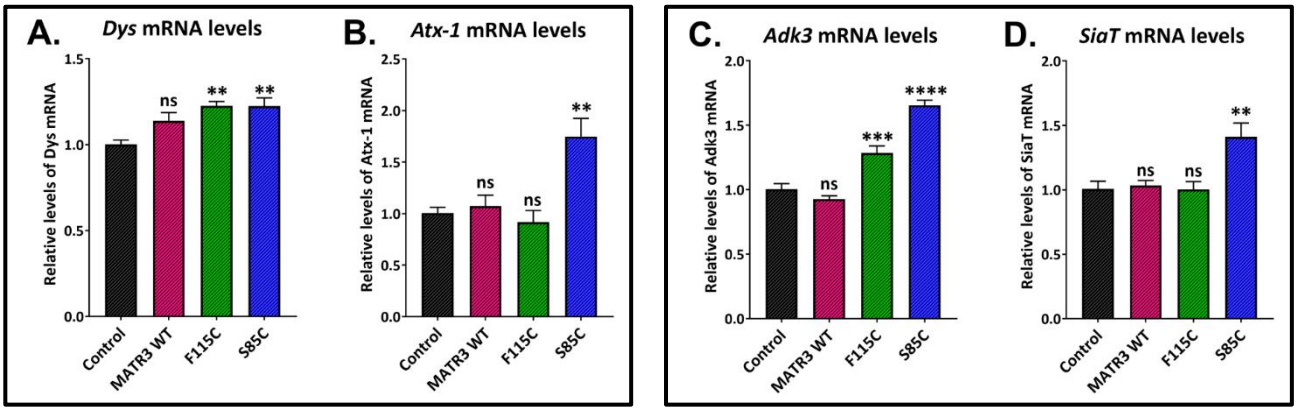

Figure 15: Levels of candidate targets from MATR3-hnRNPM shared transcriptome altered in Drosophila model

Quantitative graph showing fold change difference in mRNA levels of **(A)** Dystrophin, *Dys*, **(B)** Ataxin-1, *Atx-1*, **(C)** Adenylate kinase 3, *Adk3*, and **(D)** Sialyltransferase, *SiaT*, in flies ubiquitously expressing WT and mutant MATR3. mRNA levels of candidate targets are significantly higher in both F115C- and S85C-expressing flies (*Dys* and *Adk3*) or only S85C-expressing flies (*Atx-1* and *SiaT*) compared to driver-alone control (n=5 per group; *One-way ANOVA*). Error bars indicate S.E.M. **p<0.01, ***p<0.001, ****p<0.0001

2.4 Discussion

We comprehensively characterized MATR3-mediated toxicity in a novel transgenic *Drosophila* model. *Drosophila* is a versatile model system that has been used for over a century to study the molecular mechanisms of key biological functions given their genetic and overall experimental tractability [122]. In the context of ALS, *Drosophila* has been extensively used to model both genetic (mutations in ALS-associated genes) and environmental insults (traumatic brain injury) that lead to disease pathogenesis [100, 123–125].

Although similar in structure and function to *Drosophila* orthologs of many RNA-binding proteins, there is no known *Drosophila* ortholog for MATR3. Thus, our transgenic model allowed

us to investigate expression of wild type and mutant MATR3 *in vivo* without interference from endogenous protein. Ubiquitous expression of MATR3 in *Drosophila* negatively impacted viability during development and in adults, supported by underlying neuromuscular junction defects and motor dysfunction. Tissue-specific expression revealed higher susceptibility to MATR3 expression in muscles and motor neurons. Particularly in muscles, MATR3 expression led to early development of motor deficits that persisted with aging and eventually decreased longevity. This age-dependent decline in motor function was also reported in a transgenic mouse model expressing the F115C mutant [98]. Another striking similarity between the two models is development of myopathic histopathological changes underlying the motor deficits. In *Drosophila*, expression of MATR3 in muscles results in atrophy of the indirect flight muscles, exacerbated by the F115C and S85C mutations.

A recurring theme in the behavioral assessment of our transgenic MATR3 model is that expression of the wild type protein has comparable toxicity to the mutants. These results support evidence from rat cortical neurons where ectopic expression of either wild type or ALS-associated mutant MATR3 reduces neuron survival [96]. Similarly, in mice, transgenic expression of either wild type or F115C mutant MATR3 results in myopathic changes in pathology, and F115C expression leads to more severe paralysis [98]. Mechanistically, overexpression of wild type and mutant MATR3 impedes nuclear mRNA export in NSC-34 cells, suggesting that MATR3 function is similarly impacted in either condition. Overall, our model and others indicate a potential gain-of-function in MATR3 caused by pathogenic mutations. While MATR3 immunoreactivity appears to be higher in a few ALS and myopathy patient tissues [36, 38], further analysis of patient tissues, and/or discovery of a duplication mutation in MATR3 resulting in similar phenotypes, is required to validate the gain-of-function models.

The sub-cellular localization of MATR3 is unaffected by the mutations in our model. However, the biochemical properties of MATR3 are significantly changed by the mutations. MATR3 protein is typically distributed between detergent-soluble and insoluble fractions. In muscles, the S85C mutation drastically decreases the protein solubility. In addition to solubility, assessment of protein stability *in vivo* demonstrates a higher half-life, particularly of insoluble mutant MATR3 compared to wild type. This suggests that MATR3 mutations result in more stable and insoluble species that might be disrupting protein function. Interestingly, both mutations result in cysteine residues. Cysteine residues are capable of forming stable disulfide bridges, which might explain the increased stability and insolubility of MATR3 mutants compared to wild type. TDP-43, another protein that forms detergent-insoluble species [126], can form more insoluble species in response to oxidative stress. This occurs because of oxidization of cysteine residues, which results in crosslinking of TDP-43 through disulfide bridges [127]. Comprehensive characterization of the biochemical and biophysical properties of wild type and mutant MATR3, particularly in response to stress, may help elucidate disease pathogenesis.

Since none of the pathogenic mutations in MATR3 lie within the known functional domains of the protein, the mutations could be acting in *cis* with the functional domains to negatively impact MATR3 function. By generating and characterizing transgenic lines with deletion mutations in each functional domain of the protein, we showed that MATR3 toxicity *in vivo* is governed by its RNA-binding domains. Indeed, RRM1 and RRM2 deletion strongly mitigated MATR3 toxicity, which was protective not only in the wild type background, but also to some extent in the mutant background. This diminished rescue ability of RRM1/2 deletion in the mutant background may be explained by the higher potency of MATR3 mutation-mediated toxicity in our models. Nevertheless, this is compelling evidence that mutations in MATR3 could

lead to disease pathogenesis through its RNA-binding domains. The RNA-recognition motifs in MATR3 share sequence similarity to the RRM domain found in PTBP1, hnRNPL and hnRNPI [128]. While both RRM1 and RRM2 domains can mediate regulation of splicing and maintenance of mRNA stability, RNA-binding ability itself has only been shown for the RRM2 domain [68]. Previous studies have shown that RNA-binding ability is essential for causing the toxicities associated with FUS [129] and TDP-43 [130] and mutating the RRM domain abolishes the toxic effects.

A previous study in rat cortical neurons did not observe any change in toxicity exerted by MATR3 when the RRM domains are deleted [96]. The difference could be attributed to the different model systems used - cultured rat cortical neurons versus a whole-animal *Drosophila* model system. A main difference between our models is the absence of endogenous MATR3 in *Drosophila*, whereas, rat cortical neurons do have an endogenous MATR3. Thus, it is likely that presence of endogenous MATR3 is sufficient to mask the effect of RRM deletion in rat cortical neurons. Contrastingly, in flies, overexpression of MATR3 WT and mutants, in the absence of interference from endogenous protein, leads to aberrant binding and dysregulation of RNA which is inhibited by removing RNA-binding ability. Additionally, MATR3 toxicity was exacerbated by deletion of either zinc-finger domain (ZF1 and ZF2) in our model. The functions of these domains in MATR3 are less understood compared to the RRM domains. Both zinc-finger domains are required for mediating MATR3 interaction with DNA [131]. Thus, one possibility is that the zinc-finger domains interact with DNA and regulate transcription and chromosomal arrangements that are imperative for viability in our model. Another possible explanation could be that deletion of the zinc-finger domains re-distributes the protein and enhances its RNA-binding affinity, and concurrently its RNA-binding toxicity. In fact, it has been shown that removal of zinc-finger (ZF1)

domain increased the splicing repressor activity of MATR3 [63]. Thus, deletion of zinc-finger domains could be enhancing the toxic gain-of function activity of MATR3, mediated by its RNA-binding and RNA dysregulation.

RNA-binding proteins form dynamic and complex networks to perform their functions in an age- and tissue-dependent manner [132]. The multisystem disease etiology of MATR3 strongly indicates that there are other gene/protein modifiers that act alongside MATR3 in mediating disease pathogenesis. One of the strengths of the *Drosophila* model is the ability to perform screens to identify genetic interactors. Unsurprisingly, the MATR3 protein interactome reveals interactions with other RNA-binding proteins involved primarily in RNA splicing [69–71]. Through our candidate screen to elucidate genetic interactions between MATR3 and its interactome, we discovered *Drosophila* Rumpelstiltskin (rump), the homolog of human hnRNPM, as a strong modifier of MATR3 toxicity. Interestingly, our work is not the first to discover a potential role for rump in ALS. Investigation of hnRNPs that modify TDP-43 toxicity in a *Drosophila* model also identified rump as a strong modifier [133], suggesting that MATR3 and TDP-43 could have converging pathological mechanisms. Mass spectrometry-based studies have consistently found hnRNPM to be one of the top interactors of MATR3 [69–71]. A study comparing the MATR3 protein interactome in deletion mutants found that interaction with hnRNPM is diminished when the MATR3 RRM2 domain is deleted [70]. Consistent with this finding, using co-immunoprecipitation we showed that wild type and mutant MATR3 interacts with hnRNPM, and that this interaction is mediated by the RRM2 domain.

hnRNPM is a splicing factor that associates with pre-spliceosome and mature spliceosome complexes to regulate alternative splicing [119, 120]. A recent study showed that MATR3 and hnRNPM interact in a large assembly of splicing regulators (LASR) complex in conjunction with

the brain-specific RBFOX splicing protein [76]. Accordingly, we showed that MATR3 and hnRNPM interaction *in vitro* occurs in an RNA-dependent manner. Furthermore, there was a significant overlap in transcripts co-regulated by these proteins and the biological processes that they regulate. Transcripts that are co-regulated by MATR3 and hnRNPM converge primarily on developmental processes, particularly those imperative for nervous system development including neurogenesis and neuronal differentiation. Considering the importance of tight regulation of alternative splicing during development, it is not surprising that MATR3 and hnRNPM co-regulate developmentally important transcripts. It remains to be determined if any of these processes are perturbed in age-related neuromuscular degeneration caused by MATR3 mutations.

After neurodevelopment, chromatin remodeling and histone modification were the most prominent biological processes in our analysis. In post-mitotic neurons, these processes might be actively involved in regulating the cellular response to aging and stress. In fact, TDP-43 was recently implicated in sequestering the chromatin remodeling complex, Chd1, in *Drosophila* and disrupting its function in activating stress response genes [134]. An attractive candidate transcript shared between MATR3 and hnRNPM is HDAC4, an enzyme that modifies chromatin through histone deacetylation (Appendix Table 2). In the SOD1 model of ALS, HDAC4 plays a protective role in the neuromuscular junction and muscle innervation [135]. Other biological processes with strong evidentiary basis in neurodegenerative disease biology include the ubiquitin-proteasome system [136] and Wnt signaling pathway [137, 138]. It could be inferred that these biological processes are perturbed by dysregulation of MATR3-bound transcripts due to the disease-causing mutations. However, we only compared targets bound by wild-type MATR3. It is plausible that mutant MATR3 does not bind to the same targets and thus may regulate a different set of transcripts, resulting in perturbations in other biological processes not uncovered in our analysis.

Further CLIP-seq approaches probing for differences in MATR3 wild type and mutant transcriptomic binding, particularly in neuronal cells, will help in deciphering the role of RNA dysregulation in MATR3-ALS.

2.5 Materials and Methods

Plasmids

FLAG-MATR3 WT, ΔRRM1, ΔRRM2, ΔZNF1 and ΔZNF2 in pCMV Tag2B vector were a gift from Yossi Shiloh (Addgene plasmid #32880, 32881. 32882, 32883, 32884). pCMV-Tag2B FLAG-MATR3 S85C and pCMV-Tag2B FLAG-MATR3 F115C plasmids were generated by acquiring IDT gene fragments (gBlock) corresponding to S85C and F115C sequences and sub-cloning it into pCMV-Tag2B MATR3 WT plasmid between ScaI-EcoRI restriction sites. For generating *Drosophila* lines, FLAG-MATR3 constructs were cloned into pUASTaTTB vector between NotI-XhoI restriction sites. All sequences were verified by Sanger sequencing. The gBlock fragment and primer sequences are listed in Supplementary Table S2.4 (APPENDIX A). pT7-V5-SBP-C1-HshnRNPM was a gift from Elisa Izaurralde (Addgene plasmid # 64924).

Generation and maintenance of *Drosophila* lines

UAS-MATR3 lines were generated by site-specific insertion of the transgene in a w^{1118} background by BestGene Inc. The detailed list of other lines used in this study and their respective sources are outlined in **Supplementary Table S2.5 (APPENDIX 1).** All *Drosophila* stocks were cultured on standard dextrose media on a 12-hour light/dark cycle.

Egg-to-adult viability assay

UAS-MATR3 lines were crossed with ubiquitous driver, Tubulin-Gal4/TM3, on standard media at 28°C. The crosses were set up in triplicate. The expected progeny from the cross is 50% of flies carrying the TM3 balancer, i.e. flies not expressing MATR3, and 50% of flies not carrying TM3 balancer, i.e. flies that are expressing MATR3. Egg-to-adult viability was measured as a percentage of observed # flies/expected # flies.

Larval neuromuscular junction (NMJ) analysis

UAS-MATR3 lines were crossed with inducible driver Tubulin-GS-Gal4, at 28°C on standard media mixed with 10μM final concentration RU486 (Cayman Chemicals) for inducing transgene expression. Third-instar wandering larvae were picked for NMJ analysis. For immunofluorescence, the larvae were rinsed in PBS (Lonza 17-512F) and dissected along the dorsal midline to expose the NMJs and performed as outlined in "Immunofluorescence" section of Methods. Confocal images were acquired in Z-stacks using Nikon A1 microscope at 60x (oil) magnification. NMJs innervating Muscle 4 at hemi-segments A3-A4 were used for analyzing synaptic bouton quantities. For each NMJ, number of synaptic boutons per NMJ was normalized to surface area of innervating muscle. ImageJ (NIH) was used for quantification of bouton number and NMJ area.

Adult motor dysfunction

For conditional ubiquitous expression, UAS-MATR3 lines were crossed with inducible driver, TubGS-Gal4. Day 1 adults from the F1 progeny were collected every 24 hours and moved to standard media mixed with 20mM RU486. For constitutive expression in specific tissues, UAS-

MATR3 lines were crossed with either muscle-specific driver, MHC-Gal4, or motor neuron-specific driver, OK371-Gal4. The F1 adults were cultured at 28°C.

Locomotion was assessed using the RING assay, as previously described[139] with a few modifications. Briefly, flies were transferred, without anesthetization, into plastic vials and placed in the RING apparatus. The vials were tapped down against the bench and the climbing was recorded on video. Quantifications were performed manually by a third party in a blinded manner.

Drosophila sectioning and H&E staining

For assessment of muscle morphology, UAS-MATR3 flies were crossed with muscle-specific MHC-Gal4 driver. The F1 progeny adults were aged for 30 days when the thorax was dissected out and fixed in Davidson's fixative, modified (Electron Microscopy Sciences 64133-10). For assessment of retinal morphology, UAS-MATR3 flies were crossed with eye-specific GMR-Gal4 driver. The heads were resected from F1 adults and fixed using Davidson's fixative. Sectioning and H&E staining was done at Excalibur Pathologies. Light microscopy images were acquired in using Leica DM5500 at 10X magnification and quantification of muscle surface area was performed using ImageJ (NIH).

Pulse-chase assay

UAS-MATR3 flies were crossed with inducible ubiquitous driver, TubGS-Gal4. F1 adults were starved overnight (placed in vials with Kimwipe soaked in water) and then placed on standard media mixed with 20mM RU486 for 12 hours to induce transgene expression. The flies were then transferred non-RU486 media to stop transgene production and start the "chase" experiment. Protein degradation was chased by flash-freezing flies at 0, 12, 24 and 48 hours after transferring

flies to non-RU486 media and start of chase, followed by preparation of fly lysates and SDS-PAGE as described below. For determining half-life, time taken for protein levels to reach 0.5 relative to starting protein amount (t=0) was interpolated from the curve. If relative protein levels did not reach 0.5 at 48 hours, the graph was extrapolated to determine half-life.

Preparation of fly lysates and SDS-page

Drosophila tissue was flash frozen on dry ice and crushed using pestles. For preparation of standard lysate, crushed tissue was incubated in RIPA buffer: 150 mM NaCl, 1% NP40, 0.1% SDS, 1% sodium deoxycholate, 50 mM NaF, 2 mM EDTA, 1 mM DTT, 2.5 mM Na orthovanadate, 1x protease inhibitor cocktail (Roche 11836170001), pH 7.4. The samples were sonicated in an ultrasonic bath and centrifuged down at 12000xg for 10 min. The supernatant was boiled in 1X NuPage™ LDS-Sample buffer (Invitrogen NP0007) at 95°C for 5 minutes.

For soluble-insoluble fractionation, flash frozen tissue was crushed and resuspended in NP40 lysis buffer: 0.5% NP40, 10mM Tris HCl pH 7.8, 10mM EDTA, 150mM NaCl, 2.5mM Na orthovanadate, 1x protease inhibitor cocktail (Roche 11836170001). The lysate was sonicated in an ultrasonic bath and centrifuged at 21000xg for 25 min. The supernatant (soluble fraction) was collected and boiled in 1X NuPage™ LDS-Sample buffer (Invitrogen, NP0007) at 95°C for 5 minutes. The pellet was washed by resuspending in Washing buffer: 50mM Tris HCl pH 7.4, 150mM NaCl, and centrifuged at 21000xg for 5 min. The pellet was then resuspended in resolubilization buffer: 50mM Tris HCl pH 6.8, 5% SDS, 10% glycerol followed by sonication and centrifugation at 12000xg for 10min. The supernatant (insoluble fraction) was collected and boiled in 1X NuPage™ LDS-Sample buffer (Invitrogen, NP0007) at 95°C for 5 minutes.

SDS-PAGE was performed using 3-8% NuPage™ Tris-Acetate gels (Invitrogen) and the separated proteins were transferred onto nitrocellulose membranes using the iBlot2 system (Life Technologies 13120134)

Co-immunoprecipitation

HEK293T cells were cultured in 10-cm dishes and co-transfected with FLAG-MATR3 and V5-HNRNPM as described below. 24 hours post-transfection, cells were trypsinized and pelleted down. Nuclear lysates were extracted from the cell pellets using the NE-PER kit according manufacturer's instructions (Thermo Scientific 78833) and incubated with 8µl/ml FLAG antibody (Sigma F1804) overnight at 4°C. The lysates were incubated with Dynabeads™ Protein G superparamagnetic beads (Invitrogen 10004D) for 4 hours at 4°C. For RNase A treatment, the samples were treated with 1mg/ml RNaseA at this stage for 1 hour at 4°C. The beads were washed and purified using MagnaRack™ (Invitrogen CS15000). The immunoprecipitated samples were resuspended in 1X NuPage™ LDS-Sample buffer (Invitrogen NP0007) and boiled at 95°C for 5 minutes, followed by SDS-PAGE and Immunoblotting.

Cell culture and Transfections

HEK293T cells (ATCC® CRL-3216™) and C2C12 cells (ATCC® CRL-1772™) were cultured in Advanced DMEM supplemented with 10% FBS and 1% Glutamax and grown at 37°C and 5% CO_2. HEK293T cells were transiently transfected using Lipofectamine 3000 (Invitrogen L3000001) and used 24 hours after transfection. C2C12 cells were transiently transfected using Turbofect (Invitrogen) and fixed for immunofluorescence at 48 hours post-transfection.

Immunoblotting

Nitrocellulose membranes were blocked in blocking buffer: 5% milk (BLOT-QuickBlocker™ EMD Millipore WB57) in TBST followed by overnight incubation with primary antibody at 4°C. Membranes were washed with TBST and incubated with secondary antibody for 1 hour at room temperature, followed by washed with TBST. The membranes were imaged on Odyssey® CLx (LI-COR Biosciences) and quantification of bands was performed using Image Studio™ (LI-COR Biosciences).

Primary and secondary antibodies were prepared in blocking buffer.

Primary antibodies: Mouse anti-FLAG (1:1000; Sigma F1804); Rabbit anti-MATR3 (1:4000; Abcam ab151714); Mouse anti-V5 (1:1000; Invitrogen R960-25); Mouse anti- α tubulin (1:8000; Sigma T5168)

Secondary antibodies: Goat anti-mouse Dylight 680 (1:10000; LI-COR 925-68070); Goat anti-rabbit Dylight 680 (1:10000; Invitrogen 35568); Goat anti-mouse Dylight 800 (1:10000; Invitrogen SA5-10176); Goat anti-rabbit Dylight 800 (1:10000; Invitrogen SA5-35571)

Immunofluorescence

Dissected *Drosophila* tissues or C2C12 cells grown on coverslips were rinsed in PBS (Lonza 17-512F) and fixed in 4% paraformaldehyde (Sigma P6148) for 20 minutes at room temperature. Following fixation, the samples were washed four times (x10 min) in PBS and blocked with blocking buffer: 5% normal goat serum (NGS; Abcam AB7681) in PBS with 0.1% TritonX-100 (PBST). The samples were incubated with primary antibody overnight at 4°C, washed four times (x10 min) with 0.1% PBST, incubated with secondary antibody for 2 hours at

room temperature followed by 0.1% PBST washes. Samples were mounted onto slides using either ProLong® Gold Antifade mounting reagent (Invitrogen P36930) or Fluoroshield (Sigma F6057).

Primary and secondary antibodies were prepared in blocking buffer.

Primary antibodies: Cy3-conjugated anti-HRP (1:100, Jackson ImmunoResearch 123-165-021); Rabbit anti-FLAG (1:500, Sigma F7425); Mouse anti-hnRNPM 2A6 (1:100; Santa Cruz sc-20001); Rabbit anti-MATR3 (1:500; Abcam ab151714)

Secondary antibodies: Goat anti-rabbit Alexa Fluor 488 (1:1000; Invitrogen A11008); Goat anti-mouse Alexa Fluor 546 (1:1000; Invitrogen A11030)

Quantitative Real Time-PCR

RNA was extracted from the heads of flies ubiquitously expressing MATR3 using PureLink RNA Mini Kit (Invitrogen 12183018A0) following the manufacturer's instructions. cDNA synthesis was performed using the iScript Select cDNA Synthesis Kit (BioRad; 170-8897) and was subsequently run using the Bio-Rad iQ™Supermix (170-8862) on a 96-well plate (Applied Biosystems, #4306737) on Applied Biosystems StepOnePlus Real-Time PCR system. Primer pairs and probes were designed for each target of interest and housekeeping control α-tubulin using Integrated DNA Technologies PrimeTime qPCR Assay (www.idtdna.com). The comparative Ct method was used for analyzing the fold change differences as previously described [140] and statistical analysis was performed on Prism Graphpad.

Primer pairs and probe:

Dys

Primer 1: 5'-GCAGCACCGACTTATCGAA-3'

Primer 2: 5’-GCTGTCGTTCCGAGTACAAT-3’

Probe: 5’-/56-FAM/AACGAGCCG/ZEN/AGTGGTACTGGAG/3IABkFQ/-3’

Atx-1

Primer 1: 5’-CAGCCCTGTCCGTAAACAA-3’

Primer 2: 5’-GTGACACTCACCTTCAGCTAC-3’

Probe: 5’-/56-FAM/ACACACCAC/ZEN/GCCAAGATGGACTT/3IABkFQ/-3’

Adk3

Primer 1: 5’-GCTCTAGTTCTTCAACCTCTCC-3’

Primer 2: 5’-ACGAGAATTGCTACCGAGATAC -3’

Probe: 5’-/56-FAM/AAATTTCGG/ZEN/GTGTCCGCCGAC/3IABkFQ/-3’

SiaT

Primer 1: 5’-TGGGCGCATGATTGAATCTC -3’

Primer 2: 5’-TCCAGTGCGGGATCCTT -3’

Probe: 5’-/56-FAM/CGCTTCATT/ZEN/GACACGCACGACATT /3IABkFQ/-3’

eCLIP data analysis

The previously published[121] datasets for MATR3 eCLIP and hnRNPM eCLIP and corresponding input controls from HepG2 and K562 cell lines were acquired from the ENCODE portal. The pipeline from the standard operating procedure (SOP) published on the ENCODE website was followed[121]. In brief, FASTQ files were adapter-trimmed, mapped, and PCR duplicates removed. The uniquely mapped reads termed as usable reads were used for downstream analysis. As a second processing pipeline, CLIP BAM files were normalized over the input and

fold-change enrichment within enriched peak regions was estimated (*p*-value for enrichment was calculated by Yates' Chi-Square test or Fisher Exact Test). Enriched peaks were then annotated to GRCh38 version of the genome using HOMER[141] and filtered based on significance (p-value < 0.05) and fold-change (≥ 4) for further analysis.

Gene Ontology enrichment

List of unique annotated transcripts bound by MATR3 and hnRNPM at 3'-UTR, 5'-UTR, intron and exons were obtained from eCLIP analysis and compared for shared targets (Supplementary Table S2.2 – APPENDIX 1). Gene ontology analysis was performed using ToppGene Suite (ToppFun). A list of enriched "GO:Biological Processes" and "GO:Diseases" were generated by ToppFun. The top-20 unique biological processes were plotted (Fig 2.7C, D) and the corresponding gene list for each biological process was listed. **(Appendix Table 4).**

Statistical analysis

All statistical analyses were performed on Graphpad Prism6.

3.0 Chapter 3: RNA-dependent suppression of C9orf72 neurodegeneration by Matrin-3

Disclosure: A manuscript based on this chapter is in preparation

Ramesh N, Daley E, Mann J, Gleixner A, Kour S, Mawrie D, Anderson EN, Kofler J, Donnelly CJ, Kiskinis E, Pandey UB. "RNA-dependent suppression of C9orf72 neurodegeneration by Matrin-3"

3.1 Abstract

The most common genetic cause of Amyotrophic lateral sclerosis (ALS) is GGGGCC (G4C2) hexanucleotide repeat expansions in first intron of *C9orf72* gene. The repeat RNA mediates toxicity potentially through aggregation into intranuclear RNA foci that sequester key RNA-binding proteins (RBPs), and non-ATG mediated translation into toxic dipeptide protein repeats (DPRs). However, the role of RBP sequestration in the mechanisms underlying RNA-mediated toxicity are still under investigation. Here we show that ALS-associated RNA-binding protein, Matrin-3 (MATR3), colocalizes with G4C2 RNA foci in patient tissues and neurons with *C9orf72* mutation. Additionally, MATR3 subcellular distribution and levels are perturbed in patient-derived motor neurons. Interestingly, ectopic expression of MATR3 strongly mitigates G4C2-mediated neurodegeneration *in vivo.* This suppression is abrogated by deletion of RNA-binding domain of MATR3. Finally, we show that overexpression of MATR3-WT in HEK293T cells co-expressing G4C2 repeats reduced production of RAN-translated DPRs. Overall, we describe a novel role for MATR3 in G4C2 repeat RNA-mediated neurodegeneration.

3.2 Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder that is characterized by degeneration of upper and lower motor neurons, leading to progressive atrophy and weakness in skeletal muscles, eventually resulting in death due to respiratory failure [142, 143]. An expansion of hexanucleotide GGGGCC (G4C2) repeat within the first intron of C9orf72 gene is the most common cause of familial ALS as well as frontotemporal dementia (FTD), commonly termed as C9-ALS/FTD [34, 144]. While unaffected individuals carry less than 10 G4C2 repeats (3,4), C9-ALS/FTD patients carry anywhere between 24 to 1600 repeats [34].

Extensive investigation into the mechanisms underlying C9-ALS has identified three potentially pathogenic mechanisms for the G4C2 hexanucleotide repeat expansion (G4C2-HRE) [146–148] – (i) Loss-of-function of endogenous C9orf72 protein that could, in turn, affect endosomal trafficking and autophagy pathways [149, 150]; (ii) Gain-of-function RNA-toxicity that arises from transcription of G4C2 repeats in sense and antisense direction [151, 152]; and, (iii) Gain-of-function protein toxicity caused by repeat associated non-ATG (RAN) translation of G4C2-HRE RNA to produce toxic dipeptide repeat products (DPRs) [152–155]. Both gain-of-function models are supported by observations of pathologic aggregation of G4C2-HRE RNA into intranuclear RNA foci and cytoplasmic inclusions of the DPRs in patient neurons [153, 154, 156, 157]. While these mechanisms have been proposed alternately and have been independently investigated, they are not necessarily mutually exclusive.

The RNA-toxicity hypothesis is also supported by intronic repeat expansions in other neurological disorders, such as myotonic dystrophy and fragile X tremor ataxia syndrome, also characterized by mutant RNA aggregates or RNA foci [158, 159]. The accumulation of expanded repeat-containing RNA transcripts, including G4C2 RNA foci, sequesters essential RNA-binding

proteins (RBPs) [151, 160–162], thus leading to a potential loss-of-function of these proteins and, consequently, dysregulation of RNA metabolism [151, 152]. Thus, identification of these RBPs might be critical for understanding the mechanism of C9orf72-mediated neurodegeneration, and possibly serve as therapeutic targets. Two independent studies have so far identified the ALS-associated RNA-binding protein Matrin-3 (MATR3) as an interactor of G4C2 repeat RNA *in vitro* [163, 164]. However, the disease-relevance of these observations is yet unclear. MATR3 is an essential nuclear matrix protein that, in addition to contributing to maintaining the fibrogranular nuclear matrix network [60, 66], is also involved in regulating post-transcriptional RNA processing, including alternative splicing, mRNA stability and mRNA export [63, 64, 68, 128]. Interestingly, pathogenic mutations in MATR3 have been discovered in a subset of familial and sporadic ALS cases [36, 39, 41, 42]. Furthermore, MATR3-positive cytoplasmic inclusions were found in post-mortem brain tissue of a C9-ALS and sporadic ALS patients [33, 36].

Here, we demonstrate that MATR3 is an important component of C9orf72-associated disease pathology. We found colocalization between MATR3 and pathogenic G4C2 foci in C9-ALS patient-derived neurons and in post-mortem brain tissues. Additionally, we found that MATR3 sub-cellular localization and levels are altered in C9-ALS patient neurons. Further, we show that ectopic expression of MATR3 strongly suppressed G4C2-HRE toxicity in *Drosophila*, mediated by its RNA-binding domain. Collectively, our data presents evidence for a previously unknown role for MATR3 in G4C2 repeat RNA-mediated toxicity.

3.3 Results

3.3.1 Pathologic G4C2 RNA foci sequester MATR3 protein in C9-ALS patient neurons

Proteomic studies have suggested physical interaction between G4C2 RNA and MATR3 [163, 164], however, the functional significance of this interaction in unclear. This led us to investigate relationship between MATR3 and G4C2 foci localization in neurons. To that end, we utilized C9-ALS patient-derived iPSCs that were further differentiated into motor neurons, hereby referred to as C9 iPSC-MN. Using fluorescent in situ hybridization (FISH) coupled to immunocytochemistry, we observed colocalization between G4C2 RNA foci and MATR3 in three independent C9-ALS patient-derived iPSC-MNs **(Fig 16A).**

Quantitative analysis revealed that about 50% of neurons exhibited colocalization between G4C2 RNA foci and MATR3 **(Fig 16B),** within which 35% of foci colocalized with MATR3 (**Fig 16C)**. To further validate this finding in end-stage disease condition, we analyzed post-mortem brain tissue from C9-ALS patients which showed colocalization between G4C2 RNA and MATR3 in motor cortex sections of patient brains **(Fig 16D, E).**

To confirm that colocalization is due to physical interaction between MATR3 and G4C2 RNA, we performed a biotinylated-G4C2 RNA pull down assay with nuclear lysates from HEK293T cells transiently expressing MATR3 **(Fig 16F)**. Consistent with previous reports, MATR3 wildtype successfully pulled down with biotinylated $(G4C2)_{10}$, indicating that there is physical interaction between G4C2 RNA and MATR3 **(Fig 16G)**. Interestingly, the interaction was diminished by RNA-binding deficient variant, MATR3-ΔRRM2, that has a truncation in the primary RNA-binding domain of the protein **(Fig 16G)**. This suggests that MATR3 binds to G4C2

RNA through its RNA-binding domain and is subsequently sequestered into pathogenic G4C2 foci found in patient-derived tissues and iPSC-MNs.

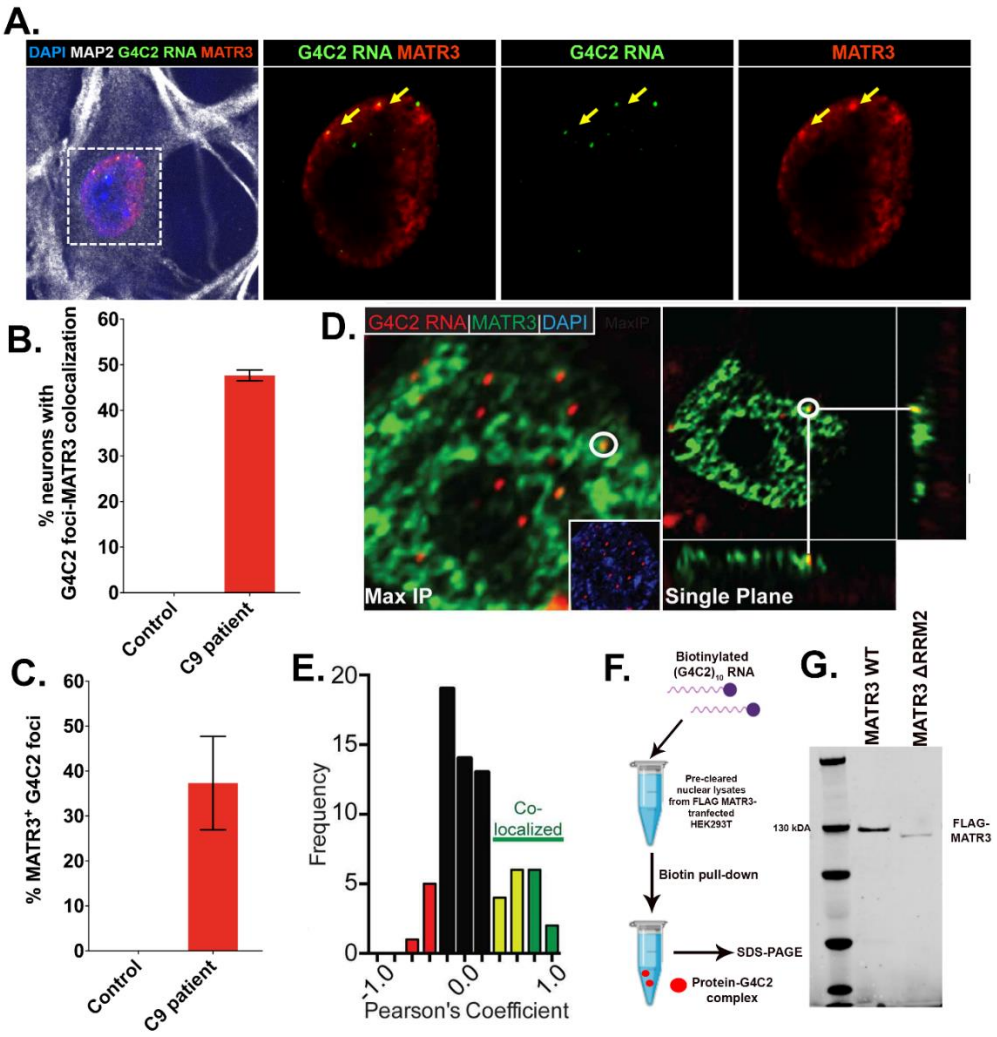

Figure 16: MATR3 colocalizes with pathogenic G4C2 RNA foci in C9-ALS patient neurons

(A) Representative single-plane image of colocalization (yellow arrows) between G4C2 RNA foci (green) and MATR3 protein (red) in C9-ALS patient-derived iPSCs differentiated into motor neurons (iPSC-MN), indicated by MAP2 (grey). **(B)** Quantification of percentage neurons that exhibit co-localization between MATR3 and G4C2 RNA foci reveals close to half the neurons showing colocalization (n = 3 independent C9-ALS and control iPSC-MN; *Student t-test*). **(C)** Quantification of percentage G4C2 foci per cell that colocalize with MATR3 puncta indicates that, on average, ~35% of G4C2 foci are also positive for MATR3 colocalization (n = 3 independent C9-ALS and control iPSC-MN; *Student t-test* for comparison). **(D)** Representative images of immunohistochemical analysis of MATR3 signal (green) in motor cortex neurons from control (top panel) and C9orf72 (bottom panel) patient tissue. NeuN staining (white) demarcates neuronal cells **(E)** RNA FISH analysis performed to examine co-localization between G4C2 RNA foci (red) and MATR3 (green) in nuclei of C9ORF72 patient tissue cells. Maximum intensity projection (left) and single plane (right) representative images are shown. Moderate (yellow) to strong (green) Pearson's coefficients indicate co-localization between RNA foci and MATR3 signal. N = 70 RNA foci. **(F)** Diagrammatic representation of $G4C2_{10}$ pull-down assay. Nuclear lysates from HEK293T cells transiently transfected with either FLAG-MATR3-WT or FLAG-MATR3-ΔRRM2 were incubated with biotinylated G4C210 RNA. The protein-G4C2

RNA complex was pulled down with streptavidin and the separated by SDS-PAGE. **(G)** Immunoblot of Biotin-G4C2$_{10}$ pull-down fraction probed for FLAG-MATR3 shows interaction between MATR3 and G4C2 RNA. The interaction is moderately diminished between MATR3-ΔRRM2 and G4C2 RNA. Error bars indicate S.E.M.

3.3.2 Increased cytoplasmic MATR3 in C9-ALS patient post-mortem brain tissues

G4C2 hexanucleotide repeat expansion (HRE) is widely reported to lead to disruption of nucleocytoplasmic transport resulting in the mislocalization of nuclear proteins, specifically proteins that have a nuclear localization signal (NLS) [165, 166]. It has been postulated that MATR3 consists an NLS that allows it to shuttle between nucleus and cytoplasm [96, 167]. Thus, we used immunohistochemistry to determine if MATR3 subcellular localization is affected in C9-ALS patient post-mortem brain tissue. In control non-ALS cases, MATR3 was predominantly nuclear in entorhinal cortex sections, with sparse cytoplasmic signal detected in a small percentage of cells **(Fig 17A).** On the other hand, examination of the entorhinal cortex sections from C9-ALS patient post-mortem brains revealed a significantly high percentage of cells with cytoplasmic immunoreactivity for MATR3 compared to that in control subjects **(Fig 17A, B).** Additionally, among cells that showed cytoplasmic staining for MATR3, we also observed a more intense cytoplasmic immunoreactivity for MATR3 in patient tissue when compared to control **(Fig 17A).**

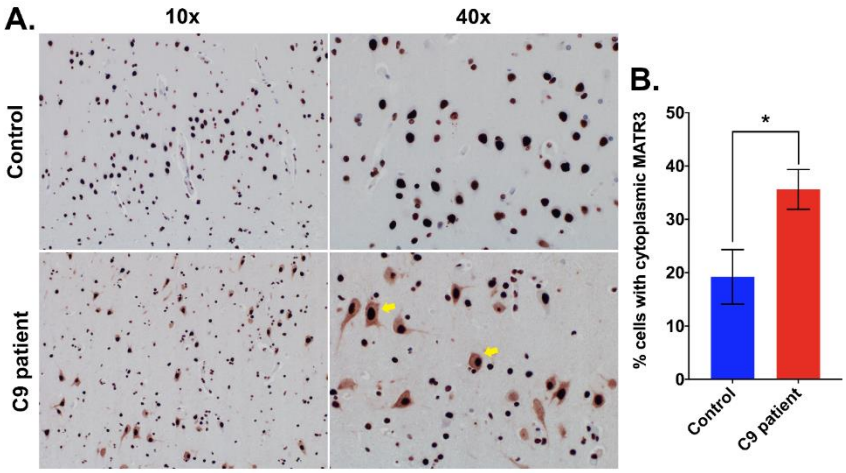

Figure 17: Increased cytoplasmic MATR3 in C9-ALS patient post-mortem brain tissue

(A) Representative IHC for MATR3 in entorhinal cortex from control and C9-ALS patient post-mortem brain. MATR3 is predominantly nuclear in control tissue. C9-ALS patient tissue shows increased cytoplasmic staining of MATR3 (yellow arrows). **(B)** Quantification of percentage of cells with cytoplasmic MATR3 shows increased number of cells with cytoplasmic MATR3 staining in C9-ALS compared to control (control, n=3; C9-ALS, n=6; *Student t-test* for comparison). Error bars indicate S.E.M. * p-value<0.05

3.3.3 MATR3 levels and subcellular localization are altered in C9-ALS patient-derived iPSC-neurons

To further investigate MATR3 distribution, we performed immunocytochemistry in C9-ALS iPSC-MNs. We observed a marked decrease in nuclear MATR3 immunoreactivity in the C9orf72 iPSC-MNs compared to that in control iPSC-MNs **(Fig 18A)**. Quantitative analysis from three independent control and C9-ALS iPSC-MNs showed a consistent decrease in MATR3 immunoreactivity in patient neurons compared that in control **(Fig 18B)**. Additionally, we observed that while MATR3 has a predominantly nuclear signal in control iPSC-MNs, C9-ALS iPSC-MNs exhibited cytoplasmic MATR3 signal distributed in a fine granular pattern **(Fig 18A).** Quantitative analysis showed that, despite some patient-to-patient variability, C9-ALS iPSC-MNs had a significantly higher number of MATR3-positive cytoplasmic granules per cell compared to control iPSC-MNs **(Fig 18C).**

Complementary to protein immunoreactivity levels, we also observed a concurrent decrease in *MATR3* mRNA levels in C9-ALS iPSC-MNs compared to controls (avg. fold change = 0.47 for C9 #1 and 0.43 for C9 #2) **(Fig 18D).** We next investigated if reversing the C9orf72 mutation directly impacts MATR3 levels. To this end, we analyzed MATR3 protein levels in two independent iPSC-MNs derived from C9-ALS patients, and their respective isogenic controls iPSC-MNs in which C9orf72 mutation had been corrected. Western blot analysis revealed that MATR3 protein levels were lower in both patient-derived iPSC-MNs compared to their respective

isogenic controls, indicating that correcting the mutation was sufficient to increase MATR3 protein levels **(Fig 19A, B).**

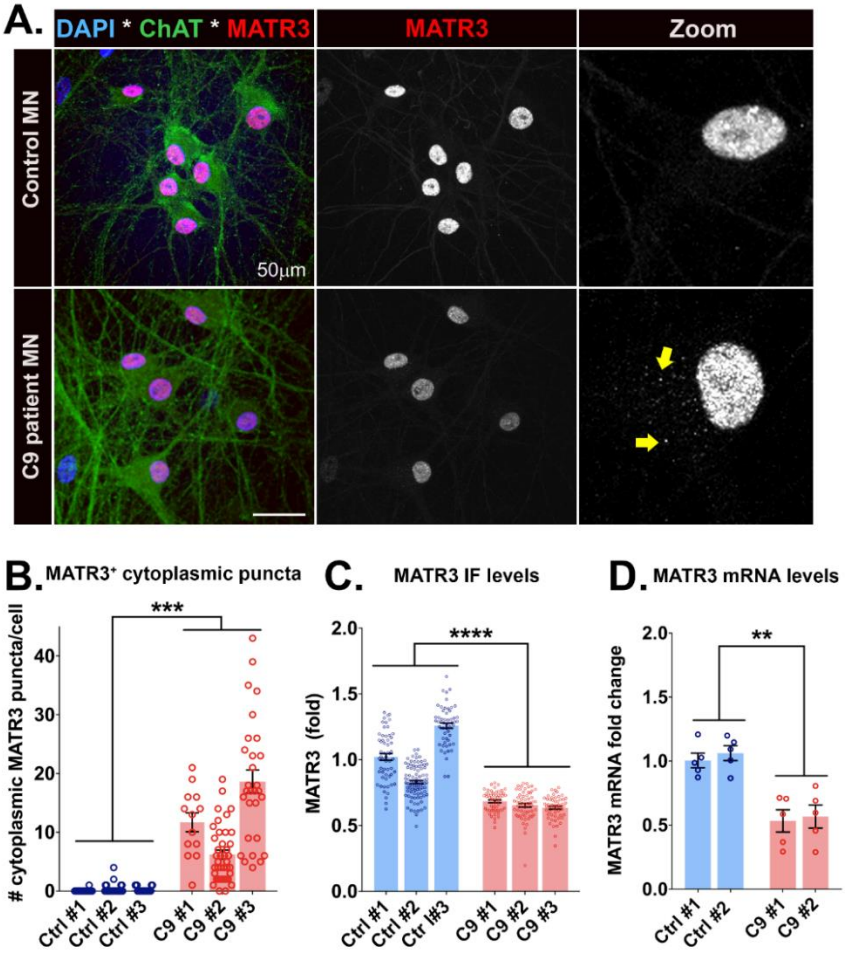

Figure 18: MATR3 levels and sub-cellular localization are altered in C9-ALS patient-derived iPSC-neurons

(A) Representative images of control and C9-ALS patient-derived iPSCs that have been differentiated to neurons ($ChAT^+$, green) and immunostained for endogenous MATR3 (red/gray). MATR3 localizes in cytoplasmic granules in C9-ALS iPSC-MNs (yellow arrows) **(B)** Quantification of number of cytoplasmic MATR3 puncta per cell indicates a significantly high cytoplasmic MATR3 granules per cell in C9-ALS iPSC neurons (Ctrl #1, n=18; Ctrl #2, n=50; Ctrl #3, n=31; C9 #1, n=13; C9 #2, n=40; C9 #3; n=29; *One-way ANOVA* for comparison). **(C)** Quantification of endogenous MATR3 IF levels in $ChAT^+$ neurons indicates lower MATR3 levels in C9 patient-derived neurons (Ctrl #1, n=57; Ctrl #2, n=91; Ctrl #3, n=55; C9 #1, n=55; C9 #2, n=60; C9 #3; n=57; *One-way ANOVA* for comparison). **(D)** Quantification of *MATR3* mRNA fold change compared to control (Ctrl #1) showing lowered mRNA levels in C9-ALS iPSC-MNs (n=5 per group; *One-way ANOVA* for comparison). Error bars indicate S.E.M. ** p-value<0.01; *** p-value<0.001

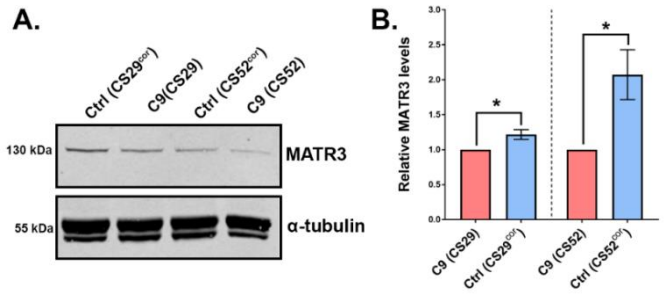

Figure 19: MATR3 protein levels increased in isogenic control iPSC-MNs harboring correction of C9orf72 mutation

(A) Immunoblot showing levels of endogenous MATR3 protein in C9-ALS iPSC-MNs and respective isogenic control harboring correction of C9orf72 mutation. α-tubulin is used as loading control **(B)** Quantification of replicate blots showing significantly increased levels of MATR3 control groups compared to respective isogenic C9-ALS groups. Error bars indicate S.E.M. * p-value<0.05

3.3.4 MATR3 is a strong suppressor of C9orf72 G4C2 HRE-mediated neurodegeneration *in vivo* via its RNA-binding domain

Our results so far indicated that G4C2-HRE in C9orf72 not only physically associates with MATR3, but also significantly impacts levels and sub-cellular localization of MATR3. So, we sought to determine whether MATR3 and G4C2 HRE genetically interact *in vivo* in *Drosophila*. We utilized previously published *Drosophila* models of G4C2-HRE toxicity, particularly transgenic models expressing 3 (3R), 30 (30R), 36 (36R) and 58 repeats (58R) of G4C2 respectively [166, 168, 169]. When expressed in the eye, these models exhibit varying degrees of eye degeneration characterized by depigmentation, ommatidial fusion and/or loss, bristle disorganization, and in the case of 36R, also accompanied by necrotic patches **(Fig 21A).** We found that ectopic expression of human transgenic MATR3 wildtype in these fly eyes **(Fig 20)** significantly ameliorated repeat-induced eye degeneration, including amelioration of necrotic patches caused by 36R **(Fig 21B).** Phenotypic quantification of eye degeneration showed

statistically significant suppression of toxicity across all G4C2-HRE models upon expression of MATR3 **(Fig 21C).** Additionally, we also observed that while G4C2 30R-mediated eye degeneration is exacerbated with aging to 30-days, expression of MATR3 in these flies continues to suppress the degeneration, even at later ages **(Fig 22A)**. Using RT-qPCR, we determined that the mRNA levels of G4C2 are not significantly changed upon expression of MATR3 **(Fig 21D)**, indicating that the suppression in eye phenotype is most likely not due to reduced levels of G4C2 transcripts.

To determine if ectopic expression of MATR3 in G4C2-HRE flies can suppress any motor deficit, we assessed locomotion ability in adult flies conditionally expressing G4C2-30R in adult neurons and aged for 30 days. Expression of G4C2-30R alone results in severe motor degeneration in these animals, with only ~10% flies able to climb at all **(Fig 21G).** Ectopic expression of MATR3 in these flies significantly improved the motor function **(Fig 21G).** Concurrently, conditional expression of 30R in neurons markedly the reduced survival ability of these flies, which was also rescued by MATR3 expression **(Fig 21E, F).**

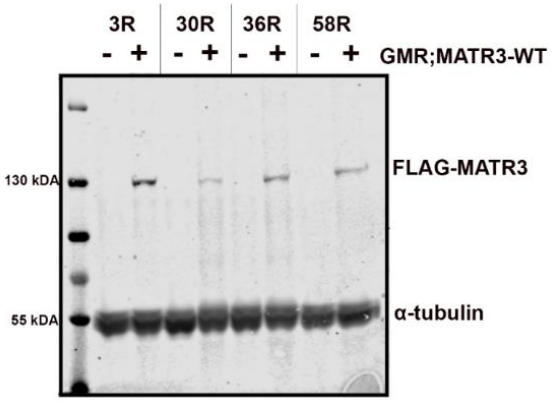

Figure 20: Ectopic MATR3 expression in G4C2-HRE Drosophila models

Immunoblot probed for MATR3 showing ectopic expression of FLAG-MATR3 in fly eyes in the G4C2 hexanucleotide repeat lines – 3R, 30R, 36R and 58R. No signal for MATR3 detected in flies not carrying the transgene. α-tubulin is used as loading control

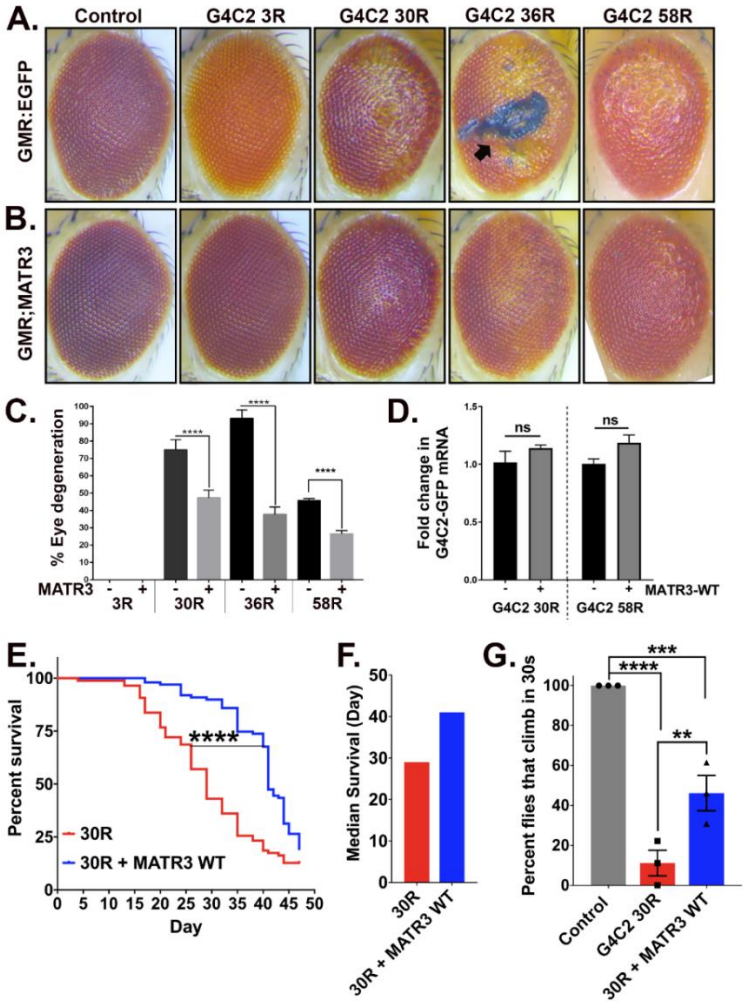

Figure 21: MATR3 is a strong modifier of C9orf72 G4C2 hexanucleotide repeat expansion-mediated neurodegeneration in vivo

(A) Representative eye images of G4C2 hexanucleotide repeat expansion mediated eye degeneration in transgenic flies expressing G4C2 3 repeats, 30 repeats, 36 repeats and 58 repeats in the eyes driven by GMR-gal4 driver. Flies expressing G4C2 3R have comparable eye phenotype to control. Flies expressing G4C2 30R, 36R and 58R repeat exhibit signs of degeneration including ommatidial fusion, bristle disorganization, depigmentation, and necrotic patches (arrow) due to 36R. These flies are expressing the G4C2 repeat expansion in the background of UAS-EGFP (dummy transgene) to account for GAL4 dilution. **(B)** Transgenic expression of human MATR3 wildtype in the G4C2 flies significantly ameliorates the degenerative phenotypes and restores ommatidial structure and pigmentation. **(C)** Quantification of eye phenotypes showing statistically significant rescue in eye degeneration in G4C2 repeat expansion flies upon expression of MATR3 (n~50 per group; *Kruskall-Wallis test*) **(D)** Quantification of mRNA levels of G4C2-GFP in heads from flies expressing G4C2 30R and 58R, with and without MATR3 WT, in fly eyes reveals no change in G4C2-GFP mRNA levels upon MATR3 expression (n=5 per group; *One-way ANOVA*) **(E)** Kaplan-Meier survival curve of flies conditionally expressing G4C2 30R in adult neurons, driven by ElavGS-GAL4 driver, shows significant reduction in overall longevity and **(F)** median survival. Expression of MATR3 wildtype significantly extends the lifespan **(E)**, and increases median survival of G4C2 30R flies **(F)** (n=100 per group; *Log-Rank Mantel Cox test* for comparison). **(G)** Neuronal expression of G4C2 30R, driven by ElavGS-GAL4 driver, causes profound

motor dysfunction. Quantification of percentage of flies that can climb in 30s indicates severe locomotion defects in G4C2 30R-expressing flies. Expression of MATR3 wildtype partially rescues the motor defects (n=3 per group; *One-way ANOVA* for comparison). Error bars indicate S.E.M. **p-value<0.01; ****p-value<0.001

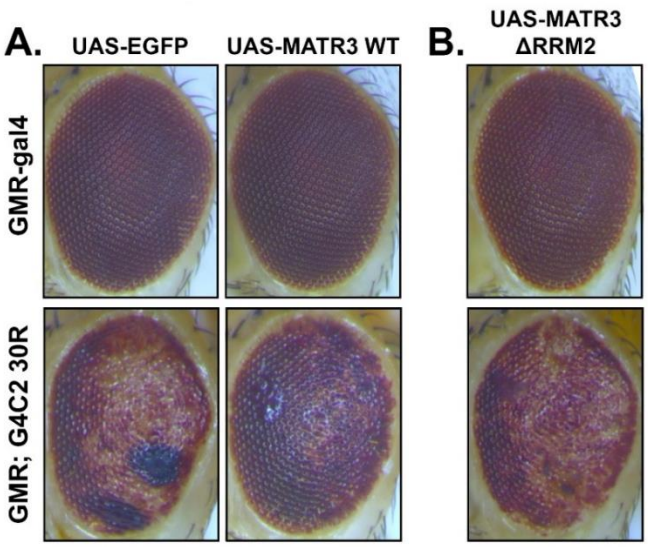

Figure 22: MATR3-mediated suppression of G4C2 toxicity persists at later age

(A) Representative eye images of G4C2-30R flies aged for 30 days and co-expressing UAS-EGFP, UAS-MATR3-WT, or **(B)** UAS-MATR3-ΔRRM2. G4C2-30R co-expressed with UAS-EGFP was used as control for basal G4C2-30R toxicity. Aging for 30 days exacerbates eye degeneration in flies expressing G4C2-30R showed by increase in de-pigmentation, ommatidial fusion and development of new necrotic patches. Ectopic expression of MATR3-WT strongly suppresses eye degeneration in G4C2-30R flies at day-30. Ectopic expression of MATR3-ΔRRM2 also suppresses eye degeneration in G4C2-30R flies at 30-days, however to a much lesser extent than control. Aging does not result in any eye degeneration in flies expressing just EGFP or MATR3.

We next sought to investigate which potential function of MATR3 is mediating suppression of G4C2-HRE toxicity. MATR3 has four known functional domains: two zinc finger motifs (ZF1 and ZF2), known to have DNA-binding and putative RNA-binding properties [114, 128], and two tandem RNA recognition motifs (RRM1 and RRM2), with RRM2 domain demonstrated to exhibit more predominant RNA-binding activity **(Fig 23A)** [68]. To map the genetic interaction between G4C2 and MATR3 to the functional domains of the protein, we generated transgenic lines with each of the four functional domains deleted: ΔRRM1, ΔRRM2, ΔZF1 and ΔZF2. MATR3 with truncation in either zinc-finger domains (ΔZF1 and ΔZF2) still retained its suppressive effect on G4C2-30R eye degeneration **(Fig 23B, C)**. On the other hand,

deletion of RRM domains, particularly the RRM2 domain, was not as efficient as full-length MATR3 in suppressing G4C2-30R toxicity in the eyes **(Fig 23B, C).** Expression of MATR3 deletion mutants alone in *Drosophila* eyes did not cause any external eye degeneration **(Fig 24),** confirming that the phenotypes observed in flies co-expressing G4C2-30R and MATR3-ΔRRM2 is not due to cumulative toxicities. Consistent with the eye phenotypes, in flies conditionally expressing G4C2-30R in adult neurons, ectopic expression of MATR3-ΔRRM2 was also not successful in increasing their longevity **(Fig 23D).** This suggests that the RNA-binding domain of MATR3 is required to suppress G4C2 HRE toxicity *in vivo*.

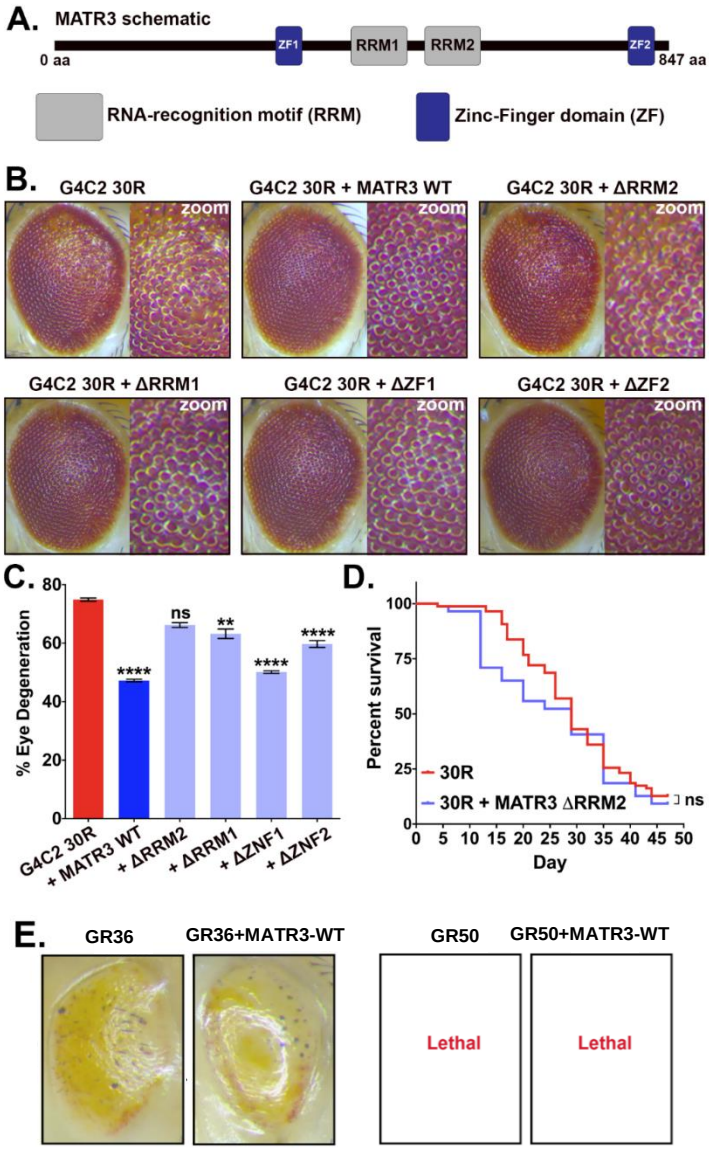

Figure 23: RRM2 domain of MATR3 required to mediate G4C2 HRE toxicity in vivo

(A) Schematic of 847 amino acid (aa)-long MATR3 protein and its functional domains: two RNA-recognition motifs (RRM1/2) and two Zinc-finger domains (ZF1/2). **(B)** Representative images of flies co-expressing G4C2 30R with MATR3 wildtype (full-length) or deletion mutants (ΔRRM1, ΔRRM2, ΔZF1, ΔZF2) in the eyes. Zoom panels emphasize on the degree of ommatidial disorganization and depigmentation. MATR3 ΔRRM2 suppresses G4C2 30R toxicity to a lesser degree compared to MATR3 wildtype. Flies co-expressing G4C2 30R and ΔRRM2 mutant retain some of the degenerative phenotypes including de-pigmentation and ommatidial fusion. **(C)** Quantification of eye degeneration reveals no difference in G4C2 30R + MATR3 ΔRRM2 compared to G4C2 30R alone (n~50 per group; *Kruskall-Wallis test*). **(D)** Kaplan-Meier survival curve showing neuronal expression of MATR3 ΔRRM2 mutant in G4C2 30R flies is not successful at modifying its lifespan (n=100 per group; *Log-Rank Mantel Cox test* for comparison). **(E)** Expression of MATR3 wildtype in codon-optimized dipeptide repeat lines that do not contain G4C2 repeat sequence (GR36 and GR50) does not suppress their respective phenotypes in the fly eyes. between G4C2 RNA and MATR3 WT. This interaction is diminished when RNA-binding domain (RRM2) is deleted. Error bars indicate S.E.M. ****p-value<0.001

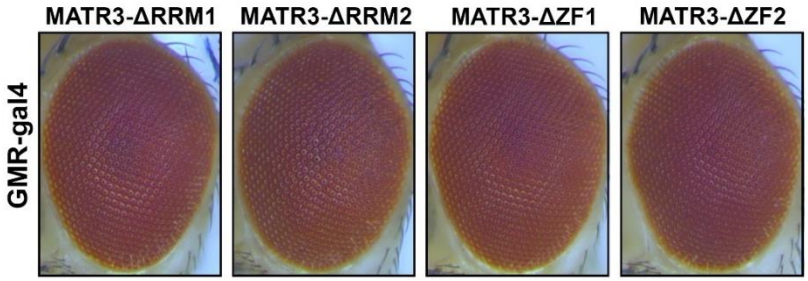

Figure 24: Deletion of different functional domains of MATR3 does not cause any external eye degeneration

Representative eye images of flies expressing MATR3 deletion variants: ΔRRM1, ΔRRM2, ΔZF1 and ΔZF2. There is no external eye degeneration caused by ectopic expression of MATR3 deletion variants

3.3.5 Ectopic expression of MATR3 does not modulate codon optimized DPR toxicity in vivo

We hypothesized that since the RNA-binding domain was essential for MATR3-mediated suppression of G4C2-HRE toxicity *in vivo*, that vice versa MATR3 suppresses toxicity possibly through G4C2 RNA-dependent mechanisms. To test this *in vivo*, we utilized codon-optimized dipeptide repeat (DPR) fly lines that produce the end-product toxic peptides, however, bypassing the production of transcripts that carry the G4C2 repeat RNA [169, 170]. The codon-optimized expression of GR dipeptide (GR36 and GR50) in fly eyes results in severe degenerative phenotype, and in fact, results in lethality in GR50-expressing flies **(Fig 23E).** Ectopic expression of MATR3 in these flies did not alleviate the degenerative phenotype **(Fig 23E),** thus confirming the reverse paradigm that when G4C2 HRE-containing transcript is unavailable, MATR3 is unable to suppress toxicity *in vivo*.

3.3.6 Overexpression of MATR3 WT suppresses RAN translation in mammalian cells

G4C2-HRE transcript potentially induces neurodegeneration through formation of toxic nuclear RNA foci and/or through RAN-translation into toxic DPRs that form neuronal inclusions [146, 171]. To investigate which of these pathogenic mechanisms is potentially altered by MATR3 expression, we used a reporter G4C2 construct, consisting of ~60 repeats (termed $G4C2_{60}$), with a fluorescent Dendra2 reporter in GR-frame which is only detected if the transcript undergoes RAN translation. We co-transfected HEK293T cells with the $G4C2_{60}$-Dendra2 reporter plasmid along with either FLAG or FLAG-MATR3 WT plasmid. Using RNA-FISH combined with ICC, we confirmed that co-expression of the FLAG and $G4C2_{60}$-Dendra2 reporter in HEK293T cells results in formation of G4C2 RNA foci in the nucleus. Furthermore, a subset of these cells (~59%) produced GR-Dendra2 RAN-translated product **(Fig 25A).** We observed that co-expression of exogenous FLAG-MATR3-WT with $G4C2_{60}$-Dendra2 did not affect either transcription **(Fig 25B)** or formation of G4C2 RNA foci **(Fig 25C)** in HEK293T cells. However, we found a significant decrease (by ~58%) in the percentage of G4C2-expressing cells that produced GR-Dendra2 RAN product **(Fig 25A, D).** Complementary to this, in flies expressing G4C2-58R, we observed reduced levels of poly-GR and poly-GP RAN-translated products upon MATR3 expression **(Fig 26).**

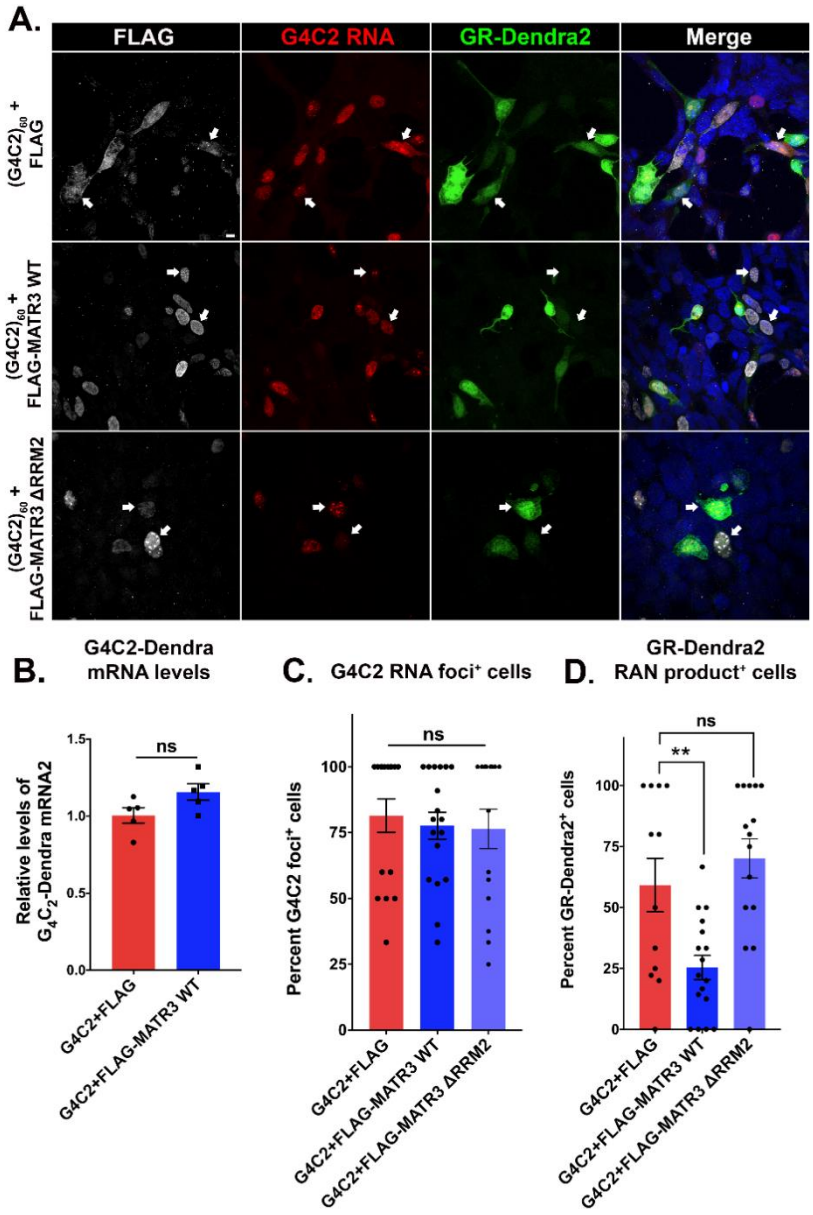

Figure 25: MATR3 overexpression mitigates RAN translation in mammalian cells

(A) Representative of HEK293T cells co-transfected with G4C2$_{60}$-Dendra2 reporter plasmid and either FLAG, FLAG-MATR3 WT or FLAG-MATR3-ΔRRM2 plasmids (gray). G4C2$_{60}$ expression results in formation of G4C2 RNA foci (red). Further, the G4C2$_{60}$ also undergoes RAN translation in a subset of cells to produce dipeptide repeats (DPRs), indicated by Dendra2 (green) signal that is in frame with polyGR DPR (white arrows). **(B)** Quantification of mRNA levels of G4C2$_{60}$-Dendra2 reveals no fold change difference between G4C2$_{60}$-Dendra2+FLAG and G4C2$_{60}$-Dendra2+FLAG-MATR3 WT. **(C)** Quantification of percent cells that form G4C2 RNA foci and **(D)** percent cells that produce GR-Dendra2 (green) RAN product. There is no change in percent cells that form G4C2 foci across the groups **(C)**. Overexpression of FLAG-MATR3 WT significantly reduces percent cells that produce GR-Dendra2. Overexpression of FLAG-MATR3-ΔRRM2 does not change percent cells that produce GR-Dendra2 **(D).** (G4C260-Dendra2 + FLAG, n=16 ; G4C260-Dendra2 + FLAG-MATR3-WT, n=18 ; G4C260-Dendra2 + FLAG-MATR3-ΔRRM2, n=15 ; *One-way ANOVA used for comparison*) Error bars indicate S.E.M. **p-value<0.01; ****p-value<0.001

Given that RRM2-deficient MATR3 was not able to suppress G4C2 toxicity *in vivo*, we investigated if RRM2 deletion results in a similar outcome *in vitro*. We thus co-expressed $G4C2_{60}$-Dendra2 with FLAG-MATR3-ΔRRM2 in HEK293T cells and assessed its impact on G4C2 foci formation and RAN translation. We found that co-expression of MATR3-ΔRRM2 did not alter the percentage of G4C2-expressing cells that produced GR-Dendra2 (Fig 25A, D), nor did it have any effect on G4C2 RNA foci formation (Fig 25C). Taken together, our data suggest that overexpression of MATR3 could be suppressing C9orf72-mediated neurodegeneration through post-transcriptional RNA-mediated mechanisms that ultimately suppress formation of toxic dipeptides.

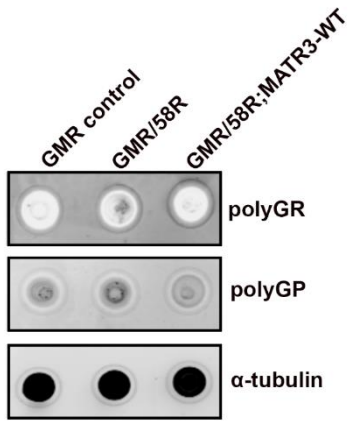

Figure 26: DPR levels reduced by MATR3 expression in G4C2-58R flies

Representative dot blot probed for polyGR and polyGP DPRs in flies expressing G4C2-58R with or without MATR3-WT. Ectopic expression of MATR3 reduces the amount of RAN-translated DPRs in G4C2-58R expressing flies. α-tubulin is used as loading control

3.4 Discussion

GGGGCC hexanucleotide repeat expansion in first intron of C9orf72 gene leads to neurodegeneration through a combination of loss-of-function of endogenous C9orf72 protein, and gain-of-function of toxic repeat-associated RNA and protein aggregates [172]. The gain-of-function RNA toxicity, that arises from bidirectional transcription of the repeats, can manifest in multiple ways. G4C2 and C4G2 repeat RNA form pathologic RNA aggregates, or RNA foci, within the brain and spinal cord of C9-ALS/FTD patients [151, 152, 173–176]. These RNA foci directly interact with and sequester essential RNA-binding proteins and thereby compromise their normal function [162], which culminates in defective RNA metabolism [75].

Colocalization and G4C2-pull down studies in C9-ALS brain tissues as well as iPSC-derived neurons have identified multiple RNA-binding proteins (RBPs) that interact with the G4C2 repeat RNA, including hnRNP-family proteins (hnRNPA1, A2/B1, A3, K, H1, H3), ALYREF, ADARB2, RanGAP1 and MATR3 [147]. Meta-analysis of all proteins that make up the G4C2-RBP interactome has identified two independent reports that determined MATR3 to be one of the interactors of G4C2 RNA [147, 163, 164]. Consistent with these reports, we also found that MATR3 binds to G4C2 RNA, mediated partially by its RRM2 RNA-binding domain. However, the disease relevance of this interaction was previously unknown. In this study we identify for the first time colocalization between G4C2 RNA foci and MATR3 in C9-ALS patient-derived brain tissue and iPSC-neurons, thus linking MATR3 to C9-ALS disease pathogenesis. RBPs have been shown to have differential binding affinities to G4C2 RNA depending on its secondary structure. For example, hnRNPH binds to both G-quadruplex and hairpin conformations of G4C2 [163], and on the other hand, SRSF1, nucleolin and RanGAP1 exhibit higher binding

affinity to G-quadruplex [163, 165]. Elucidation of differential binding affinities of MATR3 to G4C2-secondary structures may help further understand this interaction.

Our data also indicates that sub-cellular localization of MATR3 is perturbed in C9-ALS patient tissues and iPSC-neurons. MATR3 is predominantly nuclear in control subjects, whereas, in C9-ALS cases, MATR3 also exhibits significant cytoplasmic staining. Disruption in MATR3 subcellular localization has previously been reported in other ALS cases, including in C9-ALS and sporadic-ALS patient brain tissues [33, 36]. In some sporadic ALS cases, MATR3 also form cytoplasmic skein-like inclusions in patient neurons, reminiscent of TDP-43 cytoplasmic inclusions [33]. However, while we did not observe any pathologic MATR3 inclusions in the post-mortem brain tissues analyzed in this study, we did observe that cytoplasmic MATR3 in C9-ALS patient-derived iPSC-MNs distributed in a fine granular pattern. However, the nature of the cytoplasmic granules and its functional significance is yet to be determined.

In addition to perturbed sub-cellular localization, the levels of MATR3 are significantly reduced in C9-ALS neurons; interestingly at the transcriptomic level. Like many other RBPs, MATR3 is known to autoregulate its own levels [63, 68, 95]. Thus, reduced MATR3 levels in C9-ALS neurons could be a direct consequence of sequestration of MATR3 into G4C2 RNA foci, leading to disruption in its autoregulation. This raises the possibility that MATR3 sequestration in G4C2 foci could be resulting in disruption of its other roles. Initially discovered as a component of the nuclear matrix, MATR3 is best known for its role in RNA metabolism, including in alternative splicing, mRNA stability and mRNA export [63, 64, 68, 71, 82, 115]. In fact, knocking down MATR3 has been reported to lead to transcriptome-wide changes in mRNA levels as well as alternative splicing events in mammalian cells [63, 68]. In C9-ALS. widespread alterations of alternative splicing and alternative polyadenylation site usage has been identified in cerebellar

tissues from patients [75]. Thus, depleted levels of MATR3 in C9-ALS patient neurons might be one of the mechanistic links to wide-scale splicing defects that contributes to neurodegeneration.

A compelling evidence that links MATR3 to G4C2 pathogenesis is the genetic interaction between MATR3 and G4C2 HRE-mediated toxicity *in vivo*. Herein we identified MATR3 as a strong suppressor of G4C2-HRE toxicity *in vivo*. Furthermore, our data here is consistent with a model of RNA-mediated suppression of G4C2 toxicity. Specifically, we found that deletion of RNA-binding domain of MATR3 (MATR3-ΔRRM2) nullifies MATR3-mediated suppression of G4C2 toxicity *in vivo*. Likewise, MATR3 was unable to suppress toxicity in *Drosophila* models that form dipeptide repeat proteins (polyGR) through translation of codon-optimized transcripts. However, this does not preclude MATR3 altering DPR gain-of-function toxicity through G4C2 RNA-dependent post-transcriptional mechanisms. Consistent with this hypothesis, we show using a reporter for RAN translation that co-expression of MATR3 with G4C2-HRE in HEK293T (human embryonic kidney 293T) cells suppresses DPR production, that is dependent on the presence of the RRM2 RNA-binding domain in MATR3. Congruently we also show that ectopic expression of MATR3 in *Drosophila* expressing G4C2-58R repeats reduces levels of RAN-translated DPRs.

Taken together, we present a possible role for MATR3 in the pathomechanism(s) of C9orf72 G4C2-HRE mediated neurodegeneration. While MATR3-mediated suppression occurs through RNA-dependent mechanisms, it is unclear if MATR3 overexpression has any impact on RNA foci formation. A caveat of G4C2 overexpression models used here is that while endogenous mechanisms govern DPR production by RAN translation, G4C2 transcription and consequent RNA foci formation are partly regulated by exogenous overexpression-dependent mechanisms. Thus, MATR3 overexpression in disease-relevant systems, such as patient-derived fibroblasts and

iPSC-MNs, might prove more reliable for determining its impact on G4C2 transcription and aggregation into RNA foci. Future investigations into the mechanisms of neuroprotection conferred by MATR3 overexpression or restoring normal levels of MATR3 are needed for evaluating the therapeutic potential of targeting MATR3 in C9-ALS.

3.5 Materials and Methods

Plasmids

FLAG-tagged MATR3 WT, ΔRRM1, ΔRRM2, ΔZNF1 and ΔZNF2 in pCMV-Tag2B mammalian expression vector were a gift from Yossi Shiloh (Addgene plasmid #32880, 32881. 32882, 32883, 32884). For generating *Drosophila* lines, FLAG-MATR3 constructs were cloned into pUASTaTTB vector between NotI-XhoI restriction sites. All sequences were verified by Sanger sequencing. pcDNA3.0-IRES2X-G4C2$_{60}$-Dendra2 construct was a generous gift from Dr. Aaron Haeusler at Thomas Jefferson University.

Immunohistochemistry

Formalin-fixed, paraffin-embedded human entorhinal cortex sections of ALS/FTD subjects were retrieved from the Neurodegenerative Brain Bank at the University of Pittsburgh, following protocols approved by the University of Pittsburgh Committee for Oversight of Research and Clinical Training Involving Decedents (CORID).

Sections were deparaffinized in xylene and rehydrated through graded ethanol (100%, 95%, 70%, and 50% to water). Antigen retrieval was performed with citrate buffer, pH 6.0, for 15 min at 95°C, followed with 3% hydrogen peroxide treatment for 10 min at room temperature (RT).

Sections were then blocked for 1 hour at RT with 10% serum diluted in PBS with 0.3% Triton X-100. Sections were incubated in primary antibody (rabbit anti-MATR3, 1:200; Abcam 151714) diluted in blocking solution at 4°C overnight. The next day, horseradish peroxidase (HRP)-linked secondary antibody and ABC reagent (Vector Laboratories, #PK-6100 or #PK-6102) were prepared according to the manufacturer instructions. After 1 hr of incubation with secondary antibody, slides were immersed in ABC solution for 1 hr at room temperature and visualized by DAB (Vector Laboratories , #SK-4100). After counterstaining with hematoxylin (Dako, S3309), slides were dehydrated in a series of ethanols, cleared in xylene, and mounted with permanent mounting medium (Vector Laboratories, #H-5000).

Patient tissue RNA FISH and immunohistochemistry

Human control and ALS sections (formalin-fixed, paraffin-embedded) sections were then permeabilized with 0.4% Triton X-100 for 15 minutes at room temperature following two 1X PBS washes and subsequently equilibrated in 2X SSC for 10 minutes at room temperature. Sections were then incubated in 50% formamide solution (in 1X SSC) for 10 minutes at 85°C while 5' TYE 563-labeled LNA probes (5TYE563/CCCCGGCCCCGGCCC, Exiqon, Batch # 620253) were incubated at 95°C for 5 minutes. Probe hybridization was then performed by first adding 200µL probe solution (300µg/mL salmon sperm, 300µg/mL E.Coli tRNA, 1.5ng/uL probe in 100% formamide) and 200uL hybridization buffer (20% dextran sulfate in PBS, 4mg/mL BSA, 2X SSC, 2mM RVC, 0.1X PBS) per section and incubation for 1 hour at 66°C. Sections were then washed first in 50% formamide (in 2X SSC) at 80°C for 15 minutes prior to three 5-minute washes in 2X SSC at room temperature, one 5-minute wash in Tris-glycine solution (200mM Tris pH 7.4,

7.5mg/mL glycine) at room temperature, and one five minute wash in TBS-50 (50mM Tris pH 7.4, 150mM NaCl) at room temperature.

Following UV crosslinking, sections were then blocked overnight at 4°C (0.1% Triton X-100, 2% heat shock BSA in TBS-50). The next day, sections were incubated with primary antibodies (rabbit anti-Matrin-3, Abcam: ab84422, 1:500; mouse anti-NeuN, Abcam: ab104224, 1:500) in IF buffer (0.5% Protease-free BSA Fraction V (Roche), 0.5% Heat-shocked BSA Fraction V (Roche), 0.1% Triton X-100 in TBS-50) overnight at 4°C. Sections were then exposed to four 5-minute washes in IF buffer at room temperature prior to incubation with secondary antibodies (AlexaFluor 488/AlexaFlour 647, 1:800) for 1 hour at room temperature. Four 5-minute washes in IF buffer, two 5-minute washes in TBS-50, two five-minute washes in 1X PBS + 2mM MgCl2, and one 5-minute wash in 1X PBS were then performed prior to mounting slides with Prolong Antifade Mounting Media with DAPI (Invitrogen).

Cell culture and Transfections

HEK293T cells (ATCC® CRL-3216™) were cultured in Advanced DMEM supplemented with 10% FBS and 1% Glutamax and grown at 37°C and 5% CO2. HEK293T cells were transiently transfected using Lipofectamine 3000 (Invitrogen L3000001) following manufacturer instructions.

Induced pluripotent stem cell (iPSC) culture and motor neuron differentiation

iPSC differentiation into motor neurons was performed as previously described (55,56). Briefly, iPSCs were dissociated with Accutase and plated at a density of 100,000 cells/cm2 in mTESR1 media substituted with and 10μM ROCK inhibitor (DNSK International, Y-27632). Next day, media was replaced with N2B27 medium (50% DMEM:F12, 50% Neurobasal, supplemented

with NEAA, Glutamax, N2 and B27; Gibco) supplemented with SB431542 (10µM, DNSK International), LDN-193189 (100nM, DNSK International), Retinoic Acid (1 µM RA, Sigma) and Smoothened-Agonist (1µM SAG, DNSK International) and fed daily with same media for 6 days. On day 7, media was switched to N2B27 supplemented with 1µM RA, 1µM SAG, 5µM DAPT (DNSK International) and 4µM SU5402 (DNSK International) and fed daily with same media for 7 days. Cell were then dissociated using TrypLE Express (Gibco) supplemented with DNase I (Worthington) and were plated onto pre-coated Matrigel-coated surfaces (BD Biosciences) and cultured in Neurobasal medium supplemented with NEAA, Glutamax, N2, B27, Ascorbic acid (0.2 µg/ml) and BDNF, CNTF and GDNF (10 ng/mL, R&D systems). For imaging analysis, motor neurons (MNs) were seeded on pre-plated mouse glia cells.

Immunocytochemistry

Motor neurons, matured for 50 days, and transiently transfected HEK293T cells, at 72-hours post transfection, were used for immunocytochemistry. Cells were washed and fixed with 4% paraformaldehyde and blocked for 1h in Phosphate-buffered saline (PBS) containing 0.2% Triton X-100 and 10% normal donkey serum for iPSC-MNs and 5% normal goat serum for HEK293Ts respectively (Jackson ImmunoResearch). Samples were then incubated overnight at 4ºC with primary antibodies. Next day, cells were washed with PBS with 0.1% Triton X-100. Samples were then incubated with the appropriate secondary antibodies (diluted in blocking solution) conjugated to Alexa488, Alexa555 or Alexa647 fluorophores (1:500 to 1:1000 Molecular Probes) for 1 hr at RT. Cell nuclei were labeled by DNA staining using DAPI or Hoechst (Life Technologies).

Primary antibodies: Chicken anti-MAP2 (1:5000, Abcam ab5392); Goat anti-ChAT (1:500); Rabbit anti-MATR3 (1:200; Abcam 151714); Mouse Anti-digoxin (1:200; Jackson Immunoresearch 200-002-156); Rabbit anti-FLAG (1:500, Sigma F7425)

Quantification of nuclear MATR3 immunosignal

Nuclear region was determined using DAPI channel. Mean MATR3 signal within the nucleus was normalized to mean DAPI signal within each respective nucleus. Fold change was determined by averaging the values from all control neurons and normalizing to this average.

Drosophila assays

Generation and maintenance of *Drosophila* lines

FLAG-tagged UAS-MATR3 lines were generated by site-specific insertion of the transgene in a w1118 background by BestGene Inc. The detailed list of other lines used in this study and their respective sources are outlined in **Appendix Table 5**. All *Drosophila* stocks were cultured on standard dextrose media on a 12-hour light/dark cycle.

Eye degeneration quantification

Expression of transgene in *Drosophila* eyes was driven using the GMR-gal4 driver. 1-day old F1 adults (or 30-day old adults for aging experiments) were collected and external images of the eye were using Leica M205C microscope equipped with a Leica DFC450 camera. Percent external eye degeneration was quantified using a previously published scoring system (57).

Adult survival and motor dysfunction

For conditional pan-neuronal expression, G4C2-30R and UAS-MATR3 lines were crossed with inducible driver, ElavGS-Gal4. Day 1 adults from the F1 progeny were collected every 24 hours and moved to standard media mixed with 20mM RU486 to induce transgene expression in adults. The adults were cultured at 28°C and death was recorded every other day. Motor function was assessed on day-30 of their lifespan. Locomotion was assessed using the RING assay, as previously described (58) with a few modifications. Briefly, flies were transferred, without anesthetization, into plastic vials and placed in the RING apparatus. The vials were tapped down against the bench and the climbing was recorded on video. Quantifications were performed manually by a third party in a blinded manner.

Biotinylated G4C2 RNA pull-down

$G4C2_{10}$ oligomer biotinylated at the 5' end was synthesized by Integrated DNA Technologies Inc. (IDT) and RNA pull-down was performed as previously described [164]. Nuclear lysates from HEK293T cells were extracted using the NE-PER nuclear cytoplasmic extraction kit (ThermoScientific 78833) following manufacturer instructions. 0.5mg of lysates made upto 500µl with in RNA structure buffer (10 mM Tris pH 7, 0.1 M KCl, 10 mM $MgCl_2$) were pre-cleared with 50µl of pre-washed Dynabeads Streptavidin Magnetic Beads (Thermofisher 11205D) at 4°C overnight on a rotary shaker. 4µg of biotinylated RNA was heated to 90 °C for 2 min in RNA structure buffer (10 mM Tris pH 7, 0.1 M KCl, 10 mM MgCl2) and incubated at for 20 min at RT. Biotin-$G4C2_{10}$ RNA oligomer was added to the pre-cleared nuclear extract and incubated at 4°C for 4 hours. 50µl of pre-washed Dynabeads Streptavidin Magnetic Beads was added to each reaction and incubated at 4°C for 1 hour. After incubation with the nuclear extracts,

beads were washed 5 times in wash buffer (10mM Tris-Cl, pH 7.5; 1mM EDTA, 2M NaCl). Co-precipitated proteins were eluted from bound biotinylated RNAs in wash buffer and boiled in 1X NuPage LDS-Sample buffer (Invitrogen NP0007) at 95°C for 5 minutes. Co-precipitated proteins were detected by Western Blot. All buffers were freshly added with 0.5mM DTT, 0.1 U/ml RNase inhibitor, 0.1 µg/µl tRNA and 5mM EDTA.

Preparation of lysates for SDS-PAGE and Dot blot

Drosophila tissue was first flash frozen on dry ice and crushed using pestles. For preparation of standard lysate, crushed tissue or cells were incubated in RIPA buffer: 150 mM NaCl, 1% NP40, 0.1% SDS, 1% sodium deoxycholate, 50 mM NaF, 2 mM EDTA, 1 mM DTT, 2.5 mM Na orthovanadate, 1x protease inhibitor cocktail (Roche 11836170001), pH 7.4. The samples were sonicated in an ultrasonic bath and centrifuged down at 12000xg for 10 min. The supernatant was boiled in 1X NuPage LDS-Sample buffer (Invitrogen NP0007) at 95°C for 5 minutes.

For dot blot analysis, lysates were prepared from flies expressing transgenes in the eye, under GMR-gal4 driver. 2 µl of the boiled lysates were blotted onto dry nitrocellulose membrane. The blot was left to dry for 15 minutes at room temperature. SDS-PAGE was performed using 3-8% NuPage™ Tris-Acetate gels (Invitrogen) and the separated proteins were transferred onto nitrocellulose membranes using the iBlot2 system (Life Technologies 13120134).

Nitrocellulose membranes were blocked in blocking buffer: 5% milk (BLOT-QuickBlocker™ EMD Millipore WB57) in TBST followed by overnight incubation with primary antibody at 4°C. Membranes were washed with TBST and incubated with secondary antibody for 1 hour at room temperature, followed by washed with TBST. The membranes were imaged on

Odyssey® CLx (LI-COR Biosciences) and quantification of bands was performed using Image Studio™ (LI-COR Biosciences).

Primary antibodies: Mouse anti-FLAG (1:1000; Sigma F1804); Rabbit anti-MATR3 (1:4000; Abcam ab151714); Mouse anti-α tubulin (1:8000; Sigma T5168); Rabbit anti-GR (1:4000; Millipore MABN778); Rabbit anti-GP (1:4000; Proteintech 24494-1-AP)

RNA preparation and RT-quantitative PCR (qPCR)

RNA was isolated using RNeasy Mini Kit (Qiagen) according to manufacturer instructions. cDNA synthesis was performed using the iScript Select cDNA Synthesis Kit (BioRad; 170-8897) and was subsequently run using the Bio-Rad iQ™Supermix (170-8862) on a 96-well plate (Applied Biosystems, #4306737) on Applied Biosystems StepOnePlus Real-Time PCR system. Primer pairs and probes were designed for each target of interest and housekeeping control α-tubulin (for *Drosophila*) and GAPDH (for human) using Integrated DNA Technologies PrimeTime qPCR Assay (www.idtdna.com). The comparative Ct method was used for analyzing the fold change differences. The list of primer-probe sequences are listed in **Appendix Table 6**

Statistical analysis

All statistical analyses were performed on GraphPad Prism6.

4.0 Chapter 4: Conclusions and future directions

4.1 Summary of dissertation research

4.1.1 *Drosophila* model of MATR3-ALS/myopathy

The work in this dissertation provides evidence for the functional impact of pathogenic mutations in MATR3 in disease the pathogenesis of amyotrophic lateral sclerosis and distal myopathy. MATR3 is a ubiquitously expressed DNA/RNA-binding protein that has many roles in the cell, including in RNA metabolism, DNA damage repair, transcription and maintenance of nuclear integrity [128]. In the context of ALS, MATR3 came to prominence when autosomal-dominant mutations were identified in European-ancestry cohorts of familial and sporadic ALS [36]. Since then, multiple reports of disease-causing mutations in MATR3 have emerged in Italian, Japanese, French and Taiwanese cohorts, adding up to 13 nonsynonymous ALS-linked variants identified in *MATR3* so far . Interestingly, one of these variants was previously identified in patients with autosomal dominant distal myopathy with vocal and pharyngeal weakness (commonly termed as VCPDM) [38]. Further re-evaluation of these patients revealed development of neurogenic symptoms, reminiscent of that in ALS, and death due to respiratory symptoms 15-years from disease onset. Thus, the disease was re-classified as slow-progressive ALS with distal myopathy. The S85C mutation has also separately been identified in patients with VCPDM, without development of ALS phenotypes [43–45]. This indicates that mutations in MATR3 leads to multisystem proteinopathy, involving degeneration of motor neurons and muscles. However, the underlying mechanism of MATR3-mediated neuromuscular degeneration is still unknown.

In vitro models indicate that either overexpression of wildtype and mutant MATR3 or knockdown of endogenous MATR3 is neurotoxic [96]. Transgenic mouse models overexpressing MATR3 WT and F115C mutant developed overt myopathic phenotypes and paralysis, which manifested earlier in F115C transgenic mice compared to WT [98]. This highlights the importance of *in vivo* modeling of MATR3 for further elucidation of age and tissue susceptibilities of MATR3 expression. *Drosophila* models have been widely used to uncover novel mechanisms of neuromuscular degeneration. Availability of various tools for spatiotemporal control of transgenic expression combined with genetic tractability allows for identification of genetic and small molecule modifiers of toxicity, providing key insight into molecular pathways of the disease and potential therapeutic targets [101, 102].

I generated transgenic *Drosophila* models expressing MATR3 WT and ALS/myopathy linked mutants, F115C and S85C, that recapitulate key features of neuromuscular degeneration and behavioral defects. Ubiquitous expression of both WT and mutant MATR3 is toxic, resulting in development lethality as well as a drastic loss of lifespan and motor dysfunction when conditionally expressed in adults. Targeted expression of MATR3 in disease-relevant tissue systems, including motor neurons and muscles, reveals adult-onset degeneration characterized by reduced lifespan, motor dysfunction, and muscular atrophy. Interestingly, adults that express the S85C mutant in the muscles develop a more profound age-related motor deficit, suggesting differential tissue susceptibilities to MATR3 expression. To make structure-function correlations in my models, I generated transgenic *Drosophila* lines expressing MATR3 WT and mutants, with truncations in each known functional domain of the protein. Deletion of RRM domains, particularly the RRM2 domain, rescued both WT and mutant MATR3 toxicity, suggesting that the RNA-binding domain of MATR3 might be involved in mediating disease pathogenesis.

I utilized the *Drosophila* models to perform a candidate-RNAi screen, based on high-confidence protein-interactors of MATR3. I found that knockdown of *Drosophila* rump, the homolog of human hnRNPM, significantly mitigated mutant MATR3 toxicity. Further translating to mammalian cells, I found RNA-mediated physical and functional interaction between MATR3 in hnRNPM in human embryonic kidney 293T (HEK293T) and mouse myoblast derived C2C12 cells, respectively. We hypothesized that MATR3 and hnRNPM might synergistically mediate disease pathogenesis through dysregulation of specific RNA targets. Functional annotation of transcripts that are commonly bound by MATR3 and hnRNPM in two different cell types (hepatocyte-derived HepG2 cells and lymphoblasts-derived K562 cells) reveals an enrichment of biological processes important for neuronal survival, notably neurogenesis and ubiquitin-proteasome degradation.

4.1.2 Role of MATR3 in C9orf72-ALS

MATR3 has also been linked to ALS through detection of MATR3-positive neuronal cytoplasmic inclusions in post-mortem brain tissues of C9orf72-ALS and sporadic ALS [33, 36]. C9orf72-ALS is the most common genetic form of ALS caused by expansion of GGGGCC (G4C2) hexanucleotide repeats located in the first intron of *C9orf72* gene [34, 168]. Three potential mechanisms have been proposed for G4C2 hexanucleotide repeat expansion mediated neurodegeneration: 1. Loss-of-function of endogenous C9orf72 protein that cause defects in endosomal transport; 2. Gain-of-function through intranuclear aggregation of G4C2 repeats into RNA foci that sequester key RNA-binding proteins, and; 3. Repeat-associated non-ATG mediated translation (RAN translation) of G4C2 repeats into dipeptide repeat products (DPRs) that form

toxic neuronal inclusions [148]. Investigation of RNA-binding proteins that interact with G4C2 RNA identified MATR3 independently in two studies [147].

To examine the role of MATR3 in disease context of C9orf72-ALS, I utilized post-mortem brain tissue from ALS patients with C9orf72 mutations, as well as C9orf72-ALS patient-derived iPSCs that were differentiated into motor neurons (referred to as iPSC-MNs). I found that MATR3 colocalizes strongly with pathogenic G4C2 foci in patient brains and in iPSC-MNs. This colocalization is mediated by direct interactions between G4C2 RNA and MATR3, via the RRM2 RNA-binding domain of MATR3. The subcellular localization of MATR3 is also perturbed in C9orf72-ALS patient tissues, with increased cytoplasmic MATR3 immunoreactivity in C9orf72-ALS brains. This phenotype is recapitulated in C9-ALS iPSC-MNs, which also showed significant decrease in levels of MATR3.

Utilizing the transgenic MATR3 *Drosophila* models, I translated the findings *in vivo* to show genetic interaction between G4C2-HRE and MATR3. Ectopic expression of MATR3 in *Drosophila* models of G4C2-HRE strongly suppressed neurodegeneration mediated by the repeats. Furthermore, the MATR3-mediated suppression of G4C2 toxicity occurs through RNA-mediated mechanism(s) and, along the same line, is dependent on the RNA-binding domain of MATR3. Interestingly, I found that overexpression of MATR3 was significantly mitigating the production of RAN-translated DPR products through its RNA-binding domain, suggesting that MATR3 may play a direct role in G4C2 RNA-mediated toxicity that ultimately impacts toxic DPR synthesis.

4.2 Conclusions

Overall, the work presented in this dissertation sheds light on the broad role of MATR3 in mediating disease pathogenesis of ALS and distal myopathy, particularly through its RNA-binding functions. My work here is the first to show that MATR3 toxicity is mediated by its RNA-binding domains *in vivo*. Furthermore, the work here also uncovers a novel role for MATR3 in C9orf72-ALS through G4C2 RNA-dependent mechanisms.

One of the main findings from the *Drosophila* model of MATR3-ALS/myopathy is that mutant toxicity is alleviated when rumpelstiltiskin, the *Drosophila* homolog of human hnRNPM, is knocked down. This suggests that hnRNPM is possibly involved in mediated disease pathogenesis in patients with mutations in MATR3. Interestingly, rump knock down has also been shown to suppress TDP-43 toxicity in *Drosophila* models [133], suggesting that there might be converging mechanisms between MATR3 and TDP-43-mediated neurodegeneration. Cells consist of multiple RNA-binding proteins that work synergistically to regulate RNA processing. Thus, it is plausible that cells maintain a strict equilibrium of RNA-binding proteins, and that mutations in any of these RNA-binding proteins that affects the localization and biochemical properties of the same protein and/or interacting proteins could drive neuromuscular degeneration in ALS and myopathy. Particularly in the context of MATR3, we found that when overexpressed in mouse myoblast cells, mutant MATR3 forms cytoplasmic granules that recruit endogenous hnRNPM, thus impacting its sub-cellular localization. However, the cellular implications of this phenotype is not yet clear. Since MATR3 and hnRNPM physically interact in mammalian cells [69–71], that is RNA-dependent [70], we explored the hypothesis that MATR3 and hnRNPM might be functionally interacting to co-regulate common transcripts, and that mutations in MATR3 could be leading to dysregulation of these target transcripts. We identified a significant number of

transcripts that are bound by both MATR3 and hnRNPM, that play roles in a wide range of biological processes, particularly those related to neurodevelopment. Further investigation into how mutations in MATR3 affect interaction with these transcripts and, consequently, impact processing of these transcripts, is imperative to understand how mutations in MATR3 result in RNA dysregulation, and related to that, elucidate how functional interaction with hnRNPM contributes to it.

In addition to neuromuscular degeneration caused by pathogenic mutations in MATR3, the work in this dissertation also sheds light on previously unknown role of MATR3 in C9orf72-ALS. While it was known that MATR3 physically interacts with G4C2 RNA *in vitro*, it was not known if it had any functional implication. We found that in patient post-mortem brain tissue as well patient iPSC-derived motor neurons, MATR3 is sequestered into intranuclear G4C2 RNA foci. Furthermore, there is also evidence for higher cytoplasmic localization of MATR3 in patient neurons. This raises the question whether sequestration of MATR3 plays a causal role in cytoplasmic localization of MATR3. Defects in nucleocytoplasmic transport resulting in cytoplasmic mislocalization and subsequent aggregation of several proteins, particularly those that contain a nuclear localization signal (NLS), is an emerging pathomechanism in ALS and other neurodegenerative diseases [14, 165, 166]. Since MATR3 also contains a putative NLS [96] coupled with evidence for cytoplasmic mislocalization of MATR3 also observed in non-C9 ALS cases [33], it is plausible that sequestration of MATR3 into G4C2 foci and cytoplasmic localization of MATR3 are mutually exclusive phenomena.

We also discovered that RNA-binding domain of MATR3 is required for interaction with G4C2 RNA. Curiously, MATR3 has so far been shown to have affinity to bind to AU-rich sequence within mRNA [64]. It remains to be determined if MATR3 interaction with G4C2 repeats

is disease-specific, and in the same context, whether MATR3 interacts with other disease-associated GC-repeats. Furthermore, it is unclear if MATR3 exhibits higher binding affinity to different lengths of G4C2 repeats, or to known secondary structures formed by G4C2, namely G-quadruplex and hairpin loop. Further elucidation of the nature of interaction between MATR3 and G4C3 RNA would help determine if blocking this interaction attenuates neurodegeneration in C9-ALS.

In addition to physical interaction, we also observed genetic interaction between MATR3 and G4C2 repeat expansion. Overexpression of MATR3 was protective in *Drosophila* models of G4C2 repeat toxicity. Interestingly, *Drosophila* does not have a known ortholog for MATR3. This raises the possibility that overexpression of MATR3 could be suppressing G4C2 toxicity through functionally compensating for loss of secondary protein, or through interaction with a secondary protein. For instance, TDP-43 was found to suppress CGG repeat induced toxicity in *Drosophila* models of FXTAS, particularly through interaction with *Drosophila* homologs of hnRNPA2/B1, proteins that also binds to CGG repeats [179]. Thus, MATR3 could be suppressing G4C2 neurodegeneration similarly through interaction with other proteins that play a role in mediating G4C2 toxicity. Alternately, overexpression of MATR3 could have a direct impact on pathological mechanisms of G4C2 repeat-induced toxicity: specifically formation of toxic intranuclear RNA foci and/or RAN translation of G4C2 RNA into toxic dipeptide repeats (DPRs). Using a reporter for RAN translation, we found that overexpression of MATR3 in HEK293T cells reduces the percentage of cells that undergo RAN translation despite forming RNA foci. This suggests that MATR3 could have a more direct, intrinsic role in RNA-mediated toxicity of G4C2 repeat expansions in C9orf72-ALS.

4.3 Future directions

4.3.1 Investigating MATR3 mutation-dependent disease pathogenesis

The work described in Chapter 2 of this dissertation shows a potential role for hnRNPM in mediating mutant-MATR3 toxicity *in vivo*. Intriguingly, knocking down the *Drosophila* homolog of hnRNPM, rump, did not affect MATR3 WT toxicity *in vivo*. There was no obvious difference in the binding affinities of MATR3 WT and mutants with hnRNPM in mammalian cells, suggesting that the differential effect of hnRNPM knockdown is likely not due to differential binding. *In silico* analysis of transcripts bound by MATR3 and hnRNPM indicated a significant overlap between the two proteins. Interestingly, analysis of candidate targets levels from the shared transcriptome revealed differential regulation of these transcripts only in MATR3 mutants (F115C and S85C)-expressing flies. Thus, one hypothesis is that MATR3 WT and mutants might be causing RNA dysregulation through different mechanisms. The transgenic *Drosophila* model can be utilized for testing this hypothesis. Specifically, RNA-seq in flies expressing MATR3 WT and mutants, with and without rump RNAi, would identify differentially expressed transcripts between WT and mutants, and simultaneously identify transcripts that are altered by rump knockdown in these models.

Furthermore, an interesting finding from the transgenic MATR3 model is the differential tissue susceptibility to MATR3 expression. Targeted expression of MATR3 in *Drosophila* muscles and motor neurons led to toxic phenotypes. On the other hand, targeted expression of MATR3 in *Drosophila* eyes did not result in any toxicity, even upon aging. Interestingly, the two mutations, F115C and S85C, also seem to exhibit mildly differential toxicities depending on the tissue system. For example, flies expressing S85C in the muscles had a more profound age-dependent motor

dysfunction at day-30 of their lifespan compared to those expressing WT and F115C. Additionally, motor neuron-specific or pan-neuronal expression of MATR3 revealed slightly higher toxicity exerted by F115C mutant. These observations are especially interesting considering differential clinical disease progression associated with each mutant, with F115C carriers developing classic ALS/FTD phenotype, and S85C carriers developing distal myopathy, with or without ALS symptoms. Further examination of MATR3 toxicities in motor neurons, muscles, and other systems including glia, might help shed further light on tissue susceptibilities.

Additionally, the differential toxicities in *Drosophila* eyes versus muscles/motor neurons can also be exploited to address questions about tissue-specific molecular mechanisms that govern MATR3 toxicities. Unbiased genome-wide screens to identify modifiers of MATR3 toxicity in different tissue systems could identify molecular pathways that may be contributing to toxicity in neurons and muscles, while also identifying potentially protective pathways in the eye tissue. Another way would be to utilize high-throughput methods, including RNA-seq and CLIP, to determine transcripts that are differentially regulated and/or differentially bound by MATR3 in various tissues. Thus, a comparison of transcriptomic changes upon MATR3 expression in different tissues could shed light on RNA dysregulation events that drive toxicity in neurons and muscles.

4.3.2 Examining the mechanisms of MATR3-mediated suppression of C9orf72-ALS

The work described in Chapter 3 this dissertation sheds light on a potentially novel role for MATR3 in C9orf72 GGGGCC hexanucleotide repeat expansion (G4C2-HRE)-mediated ALS pathogenesis. The critical question that remains to be answered is if MATR3 plays any direct causal role in C9orf72-ALS. The sequestration of MATR3 in G4C2 RNA foci and concurrent

perturbations to levels and subcellular localization suggests a potential loss-of-function of MATR3 in C9orf72-ALS. Knockdown of MATR3 is known to be neurotoxic, however the precise mechanisms underlying neurotoxicity is largely unknown [51, 96]. Given the role of MATR3 in mRNA regulation, reduced levels of MATR3 might directly lead to dysregulation of MATR3-regulated transcripts, resulting in neurodegeneration. Investigation into MATR3 targets that are misregulated in C9orf72-ALS could shed light on specific alterations that drive C9orf72 G4C2-HRE toxicity. We have analyzed published eCLIP datasets to identify transcriptomic targets of MATR3, which could be a good starting point to test this hypothesis.

Interestingly, overexpression of MATR3 along with G4C2-HRE *in vivo* and *in vitro* revealed a direct genetic interaction between the two in an RNA-dependent manner. Moreover, MATR3 overexpression appears to alleviate toxicity through mitigating RAN translation into toxic DPRs. This raises the possibility that MATR3 might have a direct role in either transcriptional and/or post-transcriptional regulation of C9orf72 G4C2-HRE containing transcripts, including splicing and export. Future research bridging the knowledge gap between MATR3 binding to G4C2 transcripts and MATR3 overexpression ultimately suppressing RAN translation is required to understand the precise mode of suppression of C9orf72-ALS neurodegeneration by MATR3 overexpression.

Appendix A : Supplementary material for Chapter 1

Appendix Table 1: List of candidate genes and their respective Drosophila homologs used for RNAi screen

Human Gene	***Drosophila* Homolog**	**RNAi viable with Tub-Gal4?**	**Screen phenotype**	
hnRNPM	rump	Y	WT	no modification
			F115C	no modification
			S85C	Suppression
hnRNPA2B1	hrb98de	Y	WT	no modification
			F115C	no modification
			S85C	no modification
hnRNPA1	hrb87f	Y	WT	no modification
			F115C	no modification
			S85C	no modification
hnRNPC	cg42458	Y	WT	no modification
			F115C	no modification
			S85C	no modification
hnRNPUL1	cg30122	Y	WT	no modification
			F115C	no modification
			S85C	no modification
PTBP1	heph	Y	WT	no modification
			F115C	no modification
			S85C	no modification
ELAVL2	fne	Y	WT	no modification
			F115C	no modification
			S85C	no modification

Appendix Table 1 continued

hnRNPL	sm	Y	WT	no modification
			F115C	no modification
			S85C	no modification
hnRNPK	hrb57a	Y	WT	no modification
			F115C	no modification
			S85C	no modification
REF1	alyref	Y	WT	no modification
			F115C	no modification
			S85C	no modification
DAZ1	rb97d	N	NA	
hnRNPAB	sqd	N	NA	
NONO	nona	N	NA	

Appendix Table 2: List of transcripts bound by MATR3 and hnRNPM in K562 and HepG2 cells

MATR3_HNRNPM_K562					
ABCB7	CR1L	GTPBP4	NSD2	RPS6KC1	UBA6
ABL1	CRAMP1	GUCA2A	NSL1	RRAGC	UBAC1
ACER3	CREB3L2	GYPC	NSMCE2	RREB1	UBE2E3
ACSL3	CRIPAK	HADH	NUP214	RSRC1	UBE2H
ACVR1	CRY1	HBP1	NUP35	RTN4	UBE2L3
ADAM17	CTNNB1	HDLBP	NUP54	RUNX2	UBE3A
ADAMTS14	CTNND1	HERC2	NUTM2B	RYBP	UBP1
ADGRB3	CUL1	HGC6.3	OCA2	SAMSN1	UCK2
ADGRF2	CUL3	HIVEP1	OLA1	SART3	UGCG

Appendix Table 2 continued

ADIPOR2	CUX1	HPCAL1	OR51B6	SBF1	UGGT2
ADNP2	CYHR1	IGHMBP2	OR52D1	SCARB1	URI1
AFDN	CYP51A1	IL12RB2	ORC4	SCCPDH	USP36
AGAP3	DBNL	IMMP1L	OSBPL3	SCD5	USP48
AGFG1	DCTN4	IMMP2L	OSBPL8	SCGB3A1	USP6NL
AIG1	DEK	IMPG2	OSBPL9	SCLT1	VAPA
AKAP11	DENND1A	INTS1	OTUD5	SCMH1	VAPB
AKAP9	DGCR8	INVS	PAFAH1B1	SCML2	VKORC1L1
ALDH1A2	DHX35	ITGB1	PAK2	SEC63	VPS13B
ALG13	DIDO1	JADE1	PCCB	SERPINB6	VPS26C
AMBRA1	DIP2B	JAK2	PCM1	SETD2	VWA8
ANKLE2	DLG1	JMJD1C	PCMTD1	SETD5	VWDE
ANKRD11	DLG5	JPT2	PCSK9	SGMS1	WAC
ANKRD12	DMXL1	KANSL1	PDCD10	SH2B3	WASL
ANKRD17	DMXL2	KCTD3	PDE3B	SH3GL1	WDR20
ANKRD28	DNAH14	KDM6A	PDE8A	SH3GL3	WDR27
ANO10	DNAJC6	KIDINS220	PDK1	SHH	WDR4
ANTXR2	DRC1	KIF13A	PDS5A	SHPRH	WDR60

Appendix Table 2 continued

ANXA11	DROSHA	KIF2A	PHF14	SIL1	WDSUB1
AP3M1	DTNBP1	KLHDC4	PHF20	SKI	WNT11
ARFGAP2	DTYMK	KMT2A	PHKB	SKP1	WNT16
ARHGAP12	DUS4L	KMT2C	PHTF1	SLAIN2	XPO7
ARHGAP17	DYNC1H1	KMT2E	PI4KA	SLC25A24	XPR1
ARHGAP21	DYRK1A	KRIT1	PICALM	SLC25A37	XRCC3
ARHGEF12	E2F3	LARP4B	PIGG	SLC25A40	XRRA1
ARHGEF18	ECHDC1	LARS2	PIGV	SLC38A1	XXYLT1
ARID1A	EDEM1	LEPR	PIKFYVE	SLC39A11	YEATS2
ARL15	EHMT1	LMBR1	PIP4K2A	SLC7A11	YIPF6
ARL8B	EIF3E	LOC100287036	PLA2G12B	SLIT1	YTHDC2
ARVCF	EIF4G1	LRBA	PLCL1	SMAP2	ZAP70
ASXL1	ELF2	LRP12	PLEKHB2	SMCHD1	ZC2HC1B
ATF7IP2	ELOVL6	LRPPRC	PLXNC1	SMG7	ZCCHC7
ATL2	EPC2	LRRC17	POFUT2	SMURF1	ZDHHC14
ATP6V1B2	EPS15	LRRN3	POLE	SMYD3	ZDHHC7
ATP7B	ESYT2	LRTM2	POM121	SND1	ZFAND3
AUH	EVI5	LSS	POM121C	SNX13	ZFPM2

Appendix Table 2 continued

B3GALT1	EXO1	LUC7L2	PPARD	SNX14	ZMIZ1
B3GALT2	EZH2	LYPLA1	PPIG	SNX9	ZNF121
B3GNTL1	FAF1	MAGI3	PPIL4	SPATA5	ZNF124
BCL6	FAM120AOS	MAN2A1	PPP1R12C	SPATA9	ZNF131
BEND2	FAM126A	MANBA	PPP1R14C	SPPL2B	ZNF280B
BIRC2	FAM13A	MAP3K4	PPP2R1B	SPRN	ZNF292
BMS1	FAM168B	MAP3K7	PPP2R2A	SPRY3	ZNF451
BRD7	FAM178B	MAP4K3	PPP2R5C	SQLE	ZNF496
BRF1	FAM193A	MAPK1	PPP2R5E	SS18L1	ZNF518A
C16orf72	FAM8A1	MAPK6	PPP3R1	SSBP3	ZNF609
C1GALT1	FANCI	MAPK8	PPP4R1	ST3GAL2	ZNF644
C1orf146	FANCL	MAPKAPK2	PPP6R3	ST8SIA6	ZNF680
C22orf34	FBXL4	MARK2	PRAM1	STAM	ZNF746
C2orf88	FBXO11	MBOAT2	PRKAR1B	STK17A	ZNF750
C3orf35	FBXO8	MECP2	PRKAR2B	STK3	ZNF777
CA8	FDFT1	MED10	PRKCB	STRBP	ZNF780A
CAMKK2	FGF13	MED13	PROS1	STT3B	ZNF782

Appendix Table 2 continued

CAMSAP1	FGF7	MED13L	PRRC2C	STXBP2	ZNRF1
CAMTA1	FGFR1OP	MED14OS	PSAP	SUCO	ZRANB3
CCDC77	FILIP1L	MED15	PSMA1	SYNJ2	ZSWIM7
CCDC8	FLNC	MEMO1	PSMD13	TAL1	ZZZ3
CCNY	FLRT1	METTL9	PTK2	TARBP1	
CD46	FNDC3B	MFSD10	PTPN18	TBC1D5	
CD55	FNIP1	MFSD14A	PTTG1IP	TBL1XR1	
CDC14A	FNTA	MIER1	PURG	TCEA1	
CDC42BPA	FOXJ3	MLLT10	PXDNL	TCERG1	
CDC42SE2	FOXK2	MOB1B	QKI	TCF12	
CDC45	FOXN2	MON2	R3HCC1L	TCF4	
CDCA2	GAB2	MPC2	RAB7A	TENT2	
CDK13	GAB3	MPP1	RAB8B	TENT4A	
CDK6	GABRE	MRPL19	RABGAP1L	TFG	
CDKAL1	GALK2	MRPS5	RAD51B	THEM4	
CDYL	GALNT11	MRPS6	RANBP9	THOC2	
CEBPD	GALNT2	MSI2	RAPGEF6	TJP1	

Appendix Table 2 continued

CELF1	GALNT5	MTG1	RB1CC1	TLK1	
CENPE	GATAD2A	MTX2	RBBP5	TMCO3	
CEP192	GDI2	MYO1D	RBM20	TMEM135	
CEP350	GDNF	NADK	RBM33	TMEM161B	
CHD9	GINS2	NBAS	RBM41	TMEM164	
CHSY1	GJD4	NCAPG2	RCOR1	TMEM245	
CLCC1	GLDN	NCKAP1	RFX7	TMEM38B	
CLDN5	GLMN	NCOA1	RIC1	TNKS	
CLIC2	GLRX3	NDEL1	RICTOR	TNPO3	
CLTCL1	GNAQ	NDUFA10	RNF114	TNRC6A	
CMC1	GNB1	NDUFC1	RNF146	TOP2B	
CMC2	GPBP1	NDUFS4	RNF217	TP53BP2	
CMIP	GPR18	NEK1	RNF24	TPGS2	
CNOT2	GPR22	NEK7	RNF32	TRA2B	
CNOT7	GPR52	NELFA	RNF4	TRAPPC10	
CNTF	GPR88	NET1	ROCK1	TRIM24	
COA1	GPR89A	NFATC3	ROCK2	TRIM44	
COL4A3BP	GRB10	NIPBL	ROR2	TRIM58	
CORO1C	GRB2	NOL10	RPAP2	TTC28	
COX6C	GSG1L	NPLOC4	RPAP3	TWSG1	

Appendix Table 2 continued

CPQ	GTPBP10	NR3C1	RPRD1B	U2AF2	

MATR3_HNRNPM_HEPG2				
ABCD3	DHX35	KDM6A	PRPF40A	TBC1D22A
ACACA	DIDO1	KIAA0391	PRPSAP1	TBC1D4
ACSF3	DIP2B	KIF16B	PRR4	TBC1D5
ACSL3	DLC1	LAMC1	PSMA6	TBC1D8
ACTN1	DLG1	LARP4B	PSMD14	TBL1XR1
ACTN4	DNAH14	LCOR	PSMG4	TCF12
ACVR2A	DNAJC3	LEMD3	PTCHD4	TDRP
ADD1	DOCK1	LHFPL2	PTK2	TEAD1
ADIPOR2	DOCK5	LHFPL5	PTPN18	TENT4A
ADK	DRC1	LOC100130880	PTPRA	TET3
AFDN	DRD5	LOC100287036	PTPRF	TGFBR3
AGAP3	DUS4L	LPIN2	PTPRG	THAP2
AGFG1	DUSP16	LPP	PTPRK	TLK1
AGPS	DYRK1A	LRBA	PTTG1IP	TMCO3
AIG1	E2F3	LRMDA	PXDNL	TMEM131
AK4	EDA	LRP5	QKI	TMEM135
AKAP1	EHBP1	LRRC1	RAB28	TMEM178A
AKAP13	EHMT1	MAD1L1	RAB3GAP1	TMEM181
AKT2	EIPR1	MAFA	RANBP9	TMEM185B
ALG13	EP300	MAGI1	RAP1GAP	TMEM248
AMBRA1	EPB41L2	MAGI3	RB1CC1	TNFRSF21
AMFR	EPC2	MAN2A1	RBFOX2	TNKS
ANKLE2	EPN1	MAP4K3	RBM26	TNRC18
ANKRD11	EPS8	MAP4K4	RBM33	TNRC6A
ANKRD12	ESYT2	MAPK6	RELN	TNRC6B
ANKRD17	EVI5	MATR3	RERE	TNS3
ANKRD28	EXOC6B	MED13	RFFL	TOMM70
AP1AR	EXT1	MEMO1	RGL4	TOP2B

Appendix Table 2 continued

AP2A2	FAF1	MERTK	RNASET2	TOX3
APLP2	FAM107B	MGAT4A	RNF213	TRA2B
ARHGAP17	FAM168B	MGAT5	RNF32	TRAP1
ARHGAP18	FAM193A	MIA2	ROCK2	TRIM24
ARHGAP5	FANCL	MKLN1	RPRD1B	TRIM71
ARHGEF12	FARP2	MLLT10	RPS4Y1	TRIO
ARID1A	FAT1	MMADHC	RPTOR	TSNARE1
ARSB	FBH1	MORC2	RRBP1	TTC3
ASAP2	FBXL17	MPP6	RREB1	U2AF2
ASB1	FBXL4	MROH8	RUNX2	UBE2H
ASPH	FBXO33	MSI2	RYBP	UBE2K
ASXL1	FBXO7	MTG1	SAMD4A	UBP1
ASXL2	FBXW11	MTSS1	SARDH	UBR3
ATF7IP2	FBXW4	MTX2	SBF2	UCK2
ATL2	FCHSD2	N4BP1	SCAF11	UGP2
ATP7B	FGFR1OP	NAPB	SCFD1	UHRF2
ATP9B	FLNB	NCKAP1	SDC2	URI1
ATRN	FMNL2	NCOA6	SEC14L1	USP10
ATXN1	FNDC3A	NDEL1	SESTD1	USP3
B3GALT2	FNDC3B	NDUFA10	SETD2	USP31
B3GNTL1	FOXJ3	NDUFAF2	SETD4	USP36
BANP	FOXN3	NDUFB4	SETD5	UTRN
BCAR1	FOXO1	NDUFB9	SH3BP4	VAPB
BCAR3	FOXO3	NEDD4L	SHANK2	VWA8
BCL6	FOXP1	NFATC3	SHPRH	WAC
BRD1	FRAS1	NHSL1	SIAH1	WASF1
BRF1	G2E3	NINL	SIK3	WDR27
C14orf119	GAB2	NKD1	SKI	WDR35
C1GALT1	GALK2	NKX2-4	SLC20A2	WDR4
C3	GALNT11	NOL10	SLC22A23	WWOX
C7orf50	GALNT2	NOS1AP	SLC25A13	WWTR1
C9orf129	GATAD2B	NPAS2	SLC25A37	XRN2

Appendix Table 2 continued

CAMSAP1	GBX2	NPSR1	SLC26A5	ZAP70
CAMTA1	GDI2	NRIP1	SLC29A4	ZBED4
CCNY	GDNF	NSD2	SLC2A9	ZCCHC14
CDC7	GHSR	NSL1	SLC38A5	ZCCHC2
CDH2	GLMN	NUBPL	SLC39A14	ZCCHC7
CDK13	GMDS	NUTM2B	SLC39A9	ZDHHC20
CDK5RAP2	GNA12	OLA1	SLC5A3	ZDHHC21
CDK6	GNA14	OMG	SLMAP	ZFAND3
CDYL	GNAQ	ORC4	SMAD3	ZFHX3
CELF1	GPC4	OSMR	SMARCA1	ZHX2
CFDP1	GPC6	PAK2	SMCHD1	ZMYND11
CFI	GPHN	PAPSS1	SMYD3	ZNF124
CHD4	GPR18	PARD3	SND1	ZNF292
CHD6	GPR22	PCM1	SNTB1	ZNF44
CHD7	GPR52	PCNX1	SNX13	ZNF496
CHDH	GRB2	PCSK6	SNX25	ZNF518A
CHST15	GSG1L	PCSK9	SNX29	ZNF609
CLASP1	HDAC4	PDE3B	SNX9	ZNF638
CLK3	HDLBP	PDE8A	SPATA5	ZNF644
CMC2	HGC6.3	PDIA5	SPATS2L	ZNF680
CNOT3	HIPK2	PDLIM1	SPDYA	ZNF750
CNOT7	HIVEP1	PHF21A	SPRN	ZNF780A
CNTF	HLF	PHLPP1	SPTBN1	ZNF780B
COA1	HNMT	PHYHIPL	SRPK2	ZNF782
COP1	HS2ST1	PICALM	SS18L1	ZNF860
COQ8A	HSD17B4	PLXNC1	ST3GAL6	ZNRF1
COX11	ILF3	POLE4	STARD13	ZNRF3
CRCP	IMMP2L	POM121	STK24	ZRANB1
CREB3L2	INSM2	PPIL4	STK39	ZZZ3
CRIM1	IQGAP2	PPP1R9A	STT3B	
CTNNA1	ITFG1	PPP2R2A	SUCLG2	
CTNNB1	ITGA2	PPP2R5E	SYNJ2	

Appendix Table 2 continued

CTNND1	ITPR2	PPP4R2	TAOK2	
CTTN	JMJD1C	PPP6R2	TARBP1	
CUL3	KAAG1	PPP6R3	TAS2R14	
CUX1	KANK1	PQLC1	TAS2R30	
CYP26B1	KANSL1	PQLC3	TAS2R31	
CYP4F3	KCMF1	PRIM2	TAS2R4	
D2HGDH	KCTD3	PRKAR1B	TAS2R50	
DGCR8	KDM4C	PRKCZ	TBC1D14	

Appendix Table 3: Top 20 unique GO:Biological Process terms for MATR3 and hnRNPM shared transcriptomic targets in K562 and HepG2 cells

K562 GO: Biological processes		
ID	Name	p-value
GO:0016570	histone modification	2.80E-10
GO:0051276	chromosome organization	4.72E-10
GO:0009790	embryo development	1.65E-07
GO:0051726	regulation of cell cycle	3.54E-06
GO:0043547	positive regulation of GTPase activity	4.79E-06
GO:0061028	establishment of endothelial barrier	7.62E-06
GO:0048285	organelle fission	8.12E-06
GO:0046785	microtubule polymerization	1.83E-05
GO:2000017	positive regulation of determination of dorsal identity	2.12E-05
GO:0045597	positive regulation of cell differentiation	2.25E-05

Appendix Table 3 continued

GO:0022008	neurogenesis	2.31E-05
GO:0043161	proteasomal protein ubiquitination	2.64E-05
GO:0051640	organelle localization	3.20E-05
GO:0080135	regulation of cellular response to stress	3.49E-05
GO:0046467	membrane lipid biosynthetic process	4.06E-05
GO:0002520	immune system development	4.97E-05
GO:0016055	Wnt signaling pathway	5.16E-05
GO:0048858	cell projection morphogenesis	7.52E-05
GO:0051336	regulation of hydrolase activity	8.32E-05
GO:0045859	regulation of protein kinase activity	8.85E-05

HepG2 GO: Biological processes		
ID	Name	p-value
GO:0022604	regulation of cell morphogenesis	1.36E-09
GO:0022008	neurogenesis	3.12E-08
GO:0043087	regulation of GTPase activity	3.19E-08
GO:0030030	cell projection organization	1.15E-07
GO:0031589	cell-substrate adhesion	1.66E-07
GO:0009790	embryo development	2.18E-07
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	5.61E-07
GO:0016055	Wnt signaling pathway	9.94E-07

Appendix Table 3 continued

GO:0030036	actin cytoskeleton organization	1.43E-06
GO:0019220	regulation of phosphate metabolic process	1.90E-06
GO:0048870	cell motility	3.92E-06
GO:0006325	chromatin organization	4.87E-06
GO:0045893	positive regulation of transcription, DNA-templated	9.47E-06
GO:0007167	enzyme linked receptor protein signaling pathway	2.96E-05
GO:0016570	histone modification	5.08E-05
GO:0014896	muscle hypertrophy	8.54E-05
GO:0007420	brain development	9.03E-05
GO:0099072	regulation of postsynaptic membrane neurotransmitter receptor levels	9.38E-05
GO:0048813	dendrite morphogenesis	9.90E-05
GO:0071495	cellular response to endogenous stimulus	1.03E-04

Appendix Table 4: List of primers used for pCMVTag2B FLAG-MATR3 variants sequence verification

Primers for verification by Sanger Sequencing	Sequence
Primer for ΔRRM1 and ΔRRM1	ATCCATTCATGTTGCAGCAGTC
Primer for ΔZnF1	GATGAACTGAGTCGTTATCCAGAGG
Primer for ΔZnF2	TTCTGAGAACGCTGATGATCC
Primer for S85C and F115C	GGCGGATCCTCCAAGTCATT

Appendix Table 5: List of Drosophila lines used in Chapter 2

Line	Source
Transgenic lines	
UAS-MATR3 WT	This manuscript (Bestgene Inc.)
UAS-MATR3 F115C	This manuscript (Bestgene Inc.)
UAS-MATR3 S85C	This manuscript (Bestgene Inc.)
UAS-MATR3 ΔRRM1	This manuscript (Bestgene Inc.)
UAS-MATR ΔRRM2	This manuscript (Bestgene Inc.)
UAS-MATR3 ΔZNF1	This manuscript (Bestgene Inc.)
UAS-MATR3 ΔZNF2	This manuscript (Bestgene Inc.)
UAS-MATR3 F115C ΔRRM1	This manuscript (Bestgene Inc.)

Appendix Table 5 continued

UAS-MATR3 F115C ΔRRM2	This manuscript (Bestgene Inc.)
UAS-MATR3 S85C ΔRRM1	This manuscript (Bestgene Inc.)
UAS-MATR3 S85C ΔRRM2	This manuscript (Bestgene Inc.)
Driver lines	
GMR-Gal4	Bloomington Stock Center #1104
MHC-Gal4	Gift from Dr. ChunLai Wu (Zhu et al. 2015)
Tub-Gal4	Bloomington Stock Center #5138
TubGS-Gal4	Gift from Dr. Scott Pletcher
OK371-Gal4	Bloomington Stock Center #26160
RNAi lines	
rump	Vienna Drosophila Stock Center 10001KK
hrb98de	Bloomington Stock Center #31303
hrb87f	Vienna Drosophila Stock Center 100732KK
cg42458	Bloomington Stock Center #42506
cg30122	Vienna Drosophila Stock Center 106984KK
heph	Vienna Drosophila Stock Center 110749KK

Appendix Table 5 continued

fne	Vienna Drosophila Stock Center 101508KK
sm	Vienna Drosophila Stock Center 108351KK
hrb57a	Vienna Drosophila Stock Center 105271KK
alyref	Vienna Drosophila Stock Center 104471KK
rb97d	Vienna Drosophila Stock Center 102159KK
sqd	Vienna Drosophila Stock Center 32395GD
nona	Vienna Drosophila Stock Center 100723KK

Appendix B : Supplementary material for Chapter 3

Appendix Table 6: List of Drosophila lines used in Chapter 3

Line	Source
Transgenic lines	
UAS-MATR3 WT	This manuscript
UAS-MATR3 F115C	This manuscript
UAS-MATR3 S85C	This manuscript
UAS-MATR3 ΔRRM1	This manuscript
UAS-MATR ΔRRM2	This manuscript
UAS-MATR3 ΔZNF1	This manuscript
UAS-MATR3 ΔZNF2	This manuscript
Driver lines	
GMR-Gal4	Bloomington Stock Center #1104
OK371-Gal4	Bloomington Stock Center #26160
ElavGS-Gal4	Gift from Dr. Haig Keshishian
G4C2-HRE lines	
UAS-G4C2-3R	Gift from Dr. Peng Jin
UAS-G4C2-30R	Gift from Dr. Peng Jin
UAS-G4C2-58R	Gift from Dr. Paul Taylore

Appendix Table 6 continued

UAS-G4C2-36R	Gift from Dr. Adrian Isaacs
UAS-GR36	Gift from Dr. Adrian Isaacs
UAS-GR50	Generated by Pandey lab

Appendix Table 7: RT-qPCR primer and probe sequences

Dros α-tubulin	
Primer 1	ACCAGCCTGACCAACATG
Primer 2	CCTCGAAATCGTAGCTCTACAC
Probe	/56FAM/TCACACGCG/ZEN/ACAAGGAAAATTCACAGA/3IABkFQ/
MATR3	
Primer 1	CTTCTTCTGTCTGCGTTCTTCT
Primer 2	TACTGTAAGCTGTGTTCACTCTT
Probe	/56FAM/ACTCATTGC/ZEN/AGCAGCCTTCCTCA/3IABkFQ/
GAPDH	
Primer 1	GTGGAGTCATACTGGAACATGTAG
Primer 2	AATGGTGAAGGTCGGTGTG
Probe	/56-FAM/TGCAAATGG/ZEN/CAGCCCTGGTG/3IABkFQ/
Dendra2	
Primer 1	ACTTCAAGCAGAGCTTCCC
Primer 2	ACGTTCTGGAAGAAGCAGTC
Probe	/56-FAM/CAGATGCCC/ZEN/TTGTCCTCGAAGGTC/3IABkFQ/
GFP	

Appendix Table 7 continued

Primer 1	GAACCGCATCGAGCTGAA
Primer 2	TGCTTGTCGGCCATGATATAG
Probe	/56-FAM/ATCGACTTC/ZEN/AAGGAGGACGGCAAC/3IABkFQ/

Appendix C : Abbreviations

ALS - Amyotrophic lateral sclerosis

ANG - Angiogenin

ATXN2 - Ataxin-2

C9orf72 - Chromosome 9 open reading frame 72

CHCHD10 - Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10

CHMP2B - Charged Multivesicular Body Protein 2B

CHO - Chinese hamster ovary

CLIP - Cross-linking immunoprecipitation

CNS - Central nervous system

DLM - Dorsal-longitudinal muscle

EWSR1 - EWS RNA Binding Protein 1

FISH - Fluorescence in situ hybridization

FTD - Frontotemporal dementia

FUS - Fused In Sarcoma

GO - Gene ontology

GWAS - Genome-wide association study

H&E - Hematoxylin and eosin

HEK293T - Human embryonic kidney 293T

hnRNP - Heterogenous ribonucleoprotein

HRE - Hexanucleotide repeat expansion

IDR - Intrinsically disordered region

iPSC - Induced pluripotent stem cell

IP-MS - Immunoprecipitation coupled to mass spectrometry

KD - Knockdown

LASR - Large assembly of splicing regulators

MATR3 - Matrin-3

MN - Motor neuron

N2A - Neuro-2A

NEK1 - NIMA Related Kinase 1

NCT - Nucleocytoplasmic transport

NLS - Nuclear localization signal

NMJ - Neuromuscular junction

OPMD - Oculopharyngeal muscular dystrophy

OPTN - Optineurin

PFN1 - Profilin 1

PPI - Protein-protein interactome

RRM - RNA recognition motif

RAN - Repeat-associated non-ATG (translation)

RBP - RNA-binding protein

RRM - RNA-recognition motif

RT-qPCR - Real-time quantitative polymerase chain reaction

S/MAR - Scaffold/matrix attachment regions

SETX - Senataxin

SQSTM1- Sequestosome

TAF15 - TATA-Box Binding Protein Associated Factor 15

TARDBP - Transactive Response DNA Binding Protein 43 kDa

TUBA4A - Tubulin Alpha 4A

UBQLN2 - Ubiquilin 2

VAPB - Vesicle-Associated Membrane Protein-Associated Protein B

VCPDM - Vocal cord and pharyngeal distal myopathy

VCP - Valosin Containing Protein

VNC - Ventral nerve cord

ZNF/ZF - Zinc Finger

Bibliography

[1] Charcot JM, Joffory A. Deux cas d'atrophie musculaire progressive avec lesions de la substance grise et des faisceaux antero-lateraux de la moelle epiniere

[2] Goetz CG. Amyotrophic lateral sclerosis: early contributions of Jean-Martin Charcot. *Muscle Nerve* 2000; 23: 336–343.

[3] Rowland LP, Shneider NA. Amyotrophic lateral sclerosis. *N Engl J Med* 2001; 344: 1688–1700.

[4] Chio A, Logroscino G, Hardiman O, et al. Prognostic factors in ALS: A critical review. *Amyotroph Lateral Scler* 2009; 10: 310–323.

[5] Talbott EO, Malek AM, Lacomis D. The epidemiology of amyotrophic lateral sclerosis. *Handb Clin Neurol* 2016; 138: 225–238.

[6] Bäumer D, Hilton D, Paine SML, et al. Juvenile ALS with basophilic inclusions is a FUS proteinopathy with FUS mutations. *Neurology* 2010; 75: 611–618.

[7] Arthur KC, Calvo A, Price TR, et al. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. *Nat Commun* 2016; 7: 12408.

[8] Sawada H. Considerations for pharmacotherapy use in patients with amyotrophic lateral sclerosis: the earlier it starts, the better the results. *Expert Opin Pharmacother* 2019; 20: 1671–1674.

[9] Volk AE, Weishaupt JH, Andersen PM, et al. Current knowledge and recent insights into the genetic basis of amyotrophic lateral sclerosis. *Med Genet* 2018; 30: 252–258.

[10] Al-Chalabi A, van den Berg LH, Veldink J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. *Nat Rev Neurol* 2017; 13: 96–104.

[11] Abel O, Powell JF, Andersen PM, et al. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum Mutat* 2012; 33: 1345–1351.

[12] Saberi S, Stauffer JE, Schulte DJ, et al. Neuropathology of Amyotrophic Lateral Sclerosis and Its Variants. *Neurol Clin* 2015; 33: 855–876.

[13] Hutten S, Dormann D. Nucleocytoplasmic transport defects in neurodegeneration - Cause or consequence? *Semin Cell Dev Biol*. Epub ahead of print 5 June 2019. DOI: 10.1016/j.semcdb.2019.05.020.

[14] Kim HJ, Taylor JP. Lost in transportation: nucleocytoplasmic transport defects in ALS and other neurodegenerative diseases. *Neuron* 2017; 96: 285–297.

[15] Butti Z, Patten SA. RNA Dysregulation in Amyotrophic Lateral Sclerosis. *Front Genet* 2018; 9: 712.

[16] Ramesh N, Pandey UB. Autophagy Dysregulation in ALS: When Protein Aggregates Get Out of Hand. *Front Mol Neurosci*; 10. Epub ahead of print 22 August 2017. DOI: 10.3389/fnmol.2017.00263.

[17] Medinas DB, Valenzuela V, Hetz C. Proteostasis disturbance in amyotrophic lateral sclerosis. *Hum Mol Genet* 2017; 26: R91–R104.

[18] Konopka A, Atkin JD. The Emerging Role of DNA Damage in the Pathogenesis of the C9orf72 Repeat Expansion in Amyotrophic Lateral Sclerosis. *Int J Mol Sci*; 19. Epub ahead of print 12 October 2018. DOI: 10.3390/ijms19103137.

[19] Salton M, Lerenthal Y, Wang S-Y, et al. Involvement of Matrin 3 and SFPQ/NONO in the DNA damage response. *Cell Cycle* 2010; 9: 1568–1576.

[20] Kim BW, Jeong YE, Wong M, et al. DNA damage accumulates and responses are engaged in human ALS brain and spinal motor neurons and DNA repair is activatable in iPSC-derived motor neurons with SOD1 mutations. *Acta Neuropathologica Communications* 2020; 8: 7.

[21] Higelin J, Catanese A, Semelink-Sedlacek LL, et al. NEK1 loss-of-function mutation induces DNA damage accumulation in ALS patient-derived motoneurons. *Stem Cell Research* 2018; 30: 150–162.

[22] Martin LJ. Mitochondrial pathobiology in ALS. *J Bioenerg Biomembr* 2011; 43: 569–579.

[23] Smith EF, Shaw PJ, De Vos KJ. The role of mitochondria in amyotrophic lateral sclerosis. *Neuroscience Letters* 2019; 710: 132933.

[24] Burk K, Pasterkamp RJ. Disrupted neuronal trafficking in amyotrophic lateral sclerosis. *Acta Neuropathol* 2019; 137: 859–877.

[25] Van Damme P, Robberecht W, Van Den Bosch L. Modelling amyotrophic lateral sclerosis: progress and possibilities. *Dis Model Mech* 2017; 10: 537–549.

[26] De Vos KJ, Hafezparast M. Neurobiology of axonal transport defects in motor neuron diseases: Opportunities for translational research? *Neurobiol Dis* 2017; 105: 283–299.

[27] Li Q, Lee J-A, Black DL. Neuronal regulation of alternative pre-mRNA splicing. *Nat Rev Neurosci* 2007; 8: 819–831.

[28] Glock C, Heumüller M, Schuman EM. mRNA transport & local translation in neurons. *Curr Opin Neurobiol* 2017; 45: 169–177.

[29] Lukong KE, Chang K, Khandjian EW, et al. RNA-binding proteins in human genetic disease. *Trends Genet* 2008; 24: 416–425.

[30] Neumann M, Sampathu DM, Kwong LK, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 2006; 314: 130–133.

[31] Couthouis J, Hart MP, Erion R, et al. Evaluating the role of the FUS/TLS-related gene EWSR1 in amyotrophic lateral sclerosis. *Hum Mol Genet* 2012; 21: 2899–2911.

[32] Neumann M, Bentmann E, Dormann D, et al. FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain* 2011; 134: 2595–2609.

[33] Tada M, Doi H, Koyano S, et al. Matrin 3 Is a Component of Neuronal Cytoplasmic Inclusions of Motor Neurons in Sporadic Amyotrophic Lateral Sclerosis. *Am J Pathol* 2018; 188: 507–514.

[34] DeJesus-Hernandez M, Mackenzie IR, Boeve BF, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 2011; 72: 245–256.

[35] Nakayasu H, Berezney R. Nuclear matrins: identification of the major nuclear matrix proteins. *Proc Natl Acad Sci U S A* 1991; 88: 10312–10316.

[36] Johnson JO, Pioro EP, Boehringer A, et al. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat Neurosci* 2014; 17: 664–666.

[37] Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536: 285–291.

[38] Senderek J, Garvey SM, Krieger M, et al. Autosomal-Dominant Distal Myopathy Associated with a Recurrent Missense Mutation in the Gene Encoding the Nuclear Matrix Protein, Matrin 3. *Am J Hum Genet* 2009; 84: 511–518.

[39] Lin K-P, Tsai P-C, Liao Y-C, et al. Mutational analysis of MATR3 in Taiwanese patients with amyotrophic lateral sclerosis. *Neurobiol Aging* 2015; 36: 2005.e1–4.

[40] Xu L, Li J, Tang L, et al. MATR3 mutation analysis in a Chinese cohort with sporadic amyotrophic lateral sclerosis. *Neurobiol Aging* 2016; 38: 218.e3-218.e4.

[41] Millecamps S, De Septenville A, Teyssou E, et al. Genetic analysis of matrin 3 gene in French amyotrophic lateral sclerosis patients and frontotemporal lobar degeneration with amyotrophic lateral sclerosis patients. *Neurobiol Aging* 2014; 35: 2882.e13-2882.e15.

[42] Marangi G, Lattante S, Doronzio PN, et al. Matrin 3 variants are frequent in Italian ALS patients. *Neurobiol Aging* 2017; 49: 218.e1-218.e7.

[43] Müller TJ, Kraya T, Stoltenburg-Didinger G, et al. Phenotype of matrin-3-related distal myopathy in 16 German patients. *Ann Neurol* 2014; 76: 669–680.

[44] Yamashita S, Mori A, Nishida Y, et al. Clinicopathological features of the first Asian family having vocal cord and pharyngeal weakness with distal myopathy due to a MATR3 mutation. *Neuropathology and Applied Neurobiology* 2015; 41: 391–398.

[45] Feit H, Silbergleit A, Schneider LB, et al. Vocal cord and pharyngeal weakness with autosomal dominant distal myopathy: clinical description and gene localization to 5q31. *Am J Hum Genet* 1998; 63: 1732–1742.

[46] Dimachkie MM, Barohn RJ. DISTAL MYOPATHIES. *Neurol Clin* 2014; 32: 817–842.

[47] Watts GDJ, Wymer J, Kovach MJ, et al. Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nat Genet* 2004; 36: 377–381.

[48] Bucelli RC, Arhzaouy K, Pestronk A, et al. SQSTM1 splice site mutation in distal myopathy with rimmed vacuoles. *Neurology* 2015; 85: 665–674.

[49] Kim HJ, Kim NC, Wang Y-D, et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 2013; 495: 467–473.

[50] Taylor JP. Multisystem proteinopathy: intersecting genetics in muscle, bone, and brain degeneration. *Neurology* 2015; 85: 658–660.

[51] Quintero-Rivera F, Xi QJ, Keppler-Noreuil KM, et al. MATR3 disruption in human and mouse associated with bicuspid aortic valve, aortic coarctation and patent ductus arteriosus. *Hum Mol Genet* 2015; 24: 2375–2389.

[52] Higgins AW, Alkuraya FS, Bosco AF, et al. Characterization of apparently balanced chromosomal rearrangements from the developmental genome anatomy project. *Am J Hum Genet* 2008; 82: 712–722.

[53] Roberts AE, Allanson JE, Tartaglia M, et al. Noonan syndrome. *Lancet* 2013; 381: 333–342.

[54] Ozawa R, Hayashi YK, Ogawa M, et al. Emerin-lacking mice show minimal motor and cardiac dysfunctions with nuclear-associated vacuoles. *Am J Pathol* 2006; 168: 907–917.

[55] Park Y-E, Hayashi YK, Goto K, et al. Nuclear changes in skeletal muscle extend to satellite cells in autosomal dominant Emery-Dreifuss muscular dystrophy/limb-girdle muscular dystrophy 1B. *Neuromuscul Disord* 2009; 19: 29–36.

[56] Fidziańska A, Glinka Z. Nuclear architecture remodelling in envelopathies. *Folia Neuropathol* 2007; 45: 47–55.

[57] Tomonaga M. Histochemical and ultrastructural changes in senile human skeletal muscle. *J Am Geriatr Soc* 1977; 25: 125–131.

[58] Righolt CH, van 't Hoff MLR, Vermolen BJ, et al. Robust nuclear lamina-based cell classification of aging and senescent cells. *Aging (Albany NY)* 2011; 3: 1192–1201.

[59] Guo W, Chen Y, Zhou X, et al. An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. *Nat Struct Mol Biol* 2011; 18: 822–830.

[60] Zeitz MJ, Malyavantham KS, Seifert B, et al. Matrin 3: chromosomal distribution and protein interactions. *J Cell Biochem* 2009; 108: 125–133.

[61] Wilson RHC, Coverley D. Relationship between DNA replication and the nuclear matrix. *Genes Cells* 2013; 18: 17–31.

[62] Tank EM, Figueroa-Romero C, Hinder LM, et al. Abnormal RNA stability in amyotrophic lateral sclerosis. *Nat Commun* 2018; 9: 1–16.

[63] Coelho MB, Attig J, Bellora N, et al. Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. *EMBO J* 2015; 34: 653–668.

[64] Uemura Y, Oshima T, Yamamoto M, et al. Matrin3 binds directly to intronic pyrimidine-rich sequences and controls alternative splicing. *Genes Cells* 2017; 22: 785–798.

[65] de Leeuw R, Gruenbaum Y, Medalia O. Nuclear Lamins: Thin Filaments with Major Functions. *Trends Cell Biol* 2018; 28: 34–45.

[66] Depreux FF, Puckelwartz MJ, Augustynowicz A, et al. Disruption of the lamin A and matrin-3 interaction by myopathic LMNA mutations. *Hum Mol Genet* 2015; 24: 4284–4295.

[67] Bertrand AT, Chikhaoui K, Yaou RB, et al. Clinical and genetic heterogeneity in laminopathies. *Biochem Soc Trans* 2011; 39: 1687–1692.

[68] Salton M, Elkon R, Borodina T, et al. Matrin 3 binds and stabilizes mRNA. *PLoS ONE* 2011; 6: e23882.

[69] Chi B, O'Connell JD, Yamazaki T, et al. Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins. *Sci Rep* 2018; 8: 8755.

[70] Iradi MCG, Triplett JC, Thomas JD, et al. Characterization of gene regulation and protein interaction networks for Matrin 3 encoding mutations linked to amyotrophic lateral sclerosis and myopathy. *Sci Rep* 2018; 8: 4049.

[71] Boehringer A, Garcia-Mansfield K, Singh G, et al. ALS Associated Mutations in Matrin 3 Alter Protein-Protein Interactions and Impede mRNA Nuclear Export. *Sci Rep* 2017; 7: 14529.

[72] Lee FCY, Ule J. Advances in CLIP Technologies for Studies of Protein-RNA Interactions. *Molecular Cell* 2018; 69: 354–369.

[73] Grosso AR, Gomes AQ, Barbosa-Morais NL, et al. Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res* 2008; 36: 4823–4832.

[74] Ling JP, Pletnikova O, Troncoso JC, et al. TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* 2015; 349: 650–655.

[75] Prudencio M, Belzil VV, Batra R, et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci* 2015; 18: 1175–1182.

[76] Damianov A, Ying Y, Lin C-H, et al. Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell* 2016; 165: 606–619.

[77] Verkerk AJ, Pieretti M, Sutcliffe JS, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 1991; 65: 905–914.

[78] Pak C, Garshasbi M, Kahrizi K, et al. Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. *Proc Natl Acad Sci USA* 2011; 108: 12390–12395.

[79] Ling S-C, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* 2013; 79: 416–438.

[80] Liquori CL, Ricker K, Moseley ML, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 2001; 293: 864–867.

[81] Brais B, Bouchard J-P, Xie Y-G, et al. Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nature Genetics* 1998; 18: 164–167.

[82] Banerjee A, Vest KE, Pavlath GK, et al. Nuclear poly(A) binding protein 1 (PABPN1) and Matrin3 interact in muscle cells and regulate RNA processing. *Nucleic Acids Res* 2017; 45: 10706–10725.

[83] Barash Y, Calarco JA, Gao W, et al. Deciphering the splicing code. *Nature* 2010; 465: 53–59.

[84] Campbell ZT, Bhimsaria D, Valley CT, et al. Cooperativity in RNA-Protein Interactions: Global Analysis of RNA Binding Specificity. *Cell Rep* 2012; 1: 570–581.

[85] Zhang C, Lee K-Y, Swanson MS, et al. Prediction of clustered RNA-binding protein motif sites in the mammalian genome. *Nucleic Acids Res* 2013; 41: 6793–6807.

[86] Cereda M, Pozzoli U, Rot G, et al. RNAmotifs: prediction of multivalent RNA motifs that control alternative splicing. *Genome Biol* 2014; 15: R20.

[87] Geuens T, Bouhy D, Timmerman V. The hnRNP family: insights into their role in health and disease. *Hum Genet* 2016; 135: 851–867.

[88] Molliex A, Temirov J, Lee J, et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 2015; 163: 123–133.

[89] Murakami T, Qamar S, Lin JQ, et al. ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* 2015; 88: 678–690.

[90] Patel A, Lee HO, Jawerth L, et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* 2015; 162: 1066–1077.

[91] Gallego-Iradi MC, Clare AM, Brown HH, et al. Subcellular Localization of Matrin 3 Containing Mutations Associated with ALS and Distal Myopathy. *PLOS ONE* 2015; 10: e0142144.

[92] Aulas A, Vande Velde C. Alterations in stress granule dynamics driven by TDP-43 and FUS: a link to pathological inclusions in ALS? *Front Cell Neurosci*; 9. Epub ahead of print 23 October 2015. DOI: 10.3389/fncel.2015.00423.

[93] Dobra I, Pankivskyi S, Samsonova A, et al. Relation Between Stress Granules and Cytoplasmic Protein Aggregates Linked to Neurodegenerative Diseases. *Curr Neurol Neurosci Rep* 2018; 18: 107.

[94] Mensch A, Meinhardt B, Bley N, et al. The p.S85C-mutation in MATR3 impairs stress granule formation in Matrin-3 myopathy. *Exp Neurol* 2018; 306: 222–231.

[95] Giordano G, Sánchez-Pérez AM, Montoliu C, et al. Activation of NMDA receptors induces protein kinase A-mediated phosphorylation and degradation of matrin 3. Blocking these effects prevents NMDA-induced neuronal death. *J Neurochem* 2005; 94: 808–818.

[96] Malik AM, Miguez RA, Li X, et al. Matrin 3-dependent neurotoxicity is modified by nucleic acid binding and nucleocytoplasmic localization. *eLife* 2018; 7: e35977.

[97] Barmada SJ, Skibinski G, Korb E, et al. Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J Neurosci* 2010; 30: 639–649.

[98] Moloney C, Rayaprolu S, Howard J, et al. Analysis of spinal and muscle pathology in transgenic mice overexpressing wild-type and ALS-linked mutant MATR3. *acta neuropathol commun* 2018; 6: 137.

[99] Pandey UB, Nichols CD. Human disease models in Drosophila melanogaster and the role of the fly in therapeutic drug discovery. *Pharmacol Rev* 2011; 63: 411–436.

[100] Casci I, Pandey UB. A fruitful endeavor: modeling ALS in the fruit fly. *Brain Res* 2015; 1607: 47–74.

[101] Lu B, Vogel H. Drosophila models of neurodegenerative diseases. *Annu Rev Pathol* 2009; 4: 315–342.

[102] McGurk L, Berson A, Bonini NM. Drosophila as an In Vivo Model for Human Neurodegenerative Disease. *Genetics* 2015; 201: 377–402.

[103] Pantoja M, Ruohola-Baker H. Drosophila as a starting point for developing therapeutics for the rare disease Duchenne Muscular Dystrophy. *Rare Dis*; 1. Epub ahead of print 10 May 2013. DOI: 10.4161/rdis.24995.

[104] Plantié E, Migocka-Patrzałek M, Daczewska M, et al. Model Organisms in the Fight against Muscular Dystrophy: Lessons from Drosophila and Zebrafish. *Molecules* 2015; 20: 6237–6253.

[105] Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. *N Engl J Med* 2017; 377: 162–172.

[106] Corcia P, Couratier P, Blasco H, et al. Genetics of amyotrophic lateral sclerosis. *Revue Neurologique* 2017; 173: 254–262.

[107] Kapeli K, Martinez FJ, Yeo GW. Genetic mutations in RNA-binding proteins and their roles in ALS. *Hum Genet* 2017; 136: 1193–1214.

[108] Zhao M, Kim JR, van Bruggen R, et al. RNA-Binding Proteins in Amyotrophic Lateral Sclerosis. *Mol Cells* 2018; 41: 818–829.

[109] Blokhuis AM, Groen EJN, Koppers M, et al. Protein aggregation in amyotrophic lateral sclerosis. *Acta Neuropathol* 2013; 125: 777–794.

[110] Droppelmann CA, Campos-Melo D, Ishtiaq M, et al. RNA metabolism in ALS: when normal processes become pathological. *Amyotroph Lateral Scler Frontotemporal Degener* 2014; 15: 321–336.

[111] Liu EY, Cali CP, Lee EB. RNA metabolism in neurodegenerative disease. *Dis Model Mech* 2017; 10: 509–518.

[112] Donnelly CJ, Grima JC, Sattler R. Aberrant RNA homeostasis in amyotrophic lateral sclerosis: potential for new therapeutic targets? *Neurodegener Dis Manag* 2014; 4: 417–437.

[113] Leblond CS, Gan-Or Z, Spiegelman D, et al. Replication study of MATR3 in familial and sporadic amyotrophic lateral sclerosis. *Neurobiol Aging* 2016; 37: 209.e17-209.e21.

[114] Hibino Y, Ohzeki H, Sugano N, et al. Transcription modulation by a rat nuclear scaffold protein, P130, and a rat highly repetitive DNA component or various types of animal and plant matrix or scaffold attachment regions. *Biochem Biophys Res Commun* 2000; 279: 282–287.

[115] Kula A, Gharu L, Marcello A. HIV-1 pre-mRNA commitment to Rev mediated export through PSF and Matrin 3. *Virology* 2013; 435: 329–340.

[116] Gallego-Iradi MC, Strunk H, Crown AM, et al. N-terminal sequences in matrin 3 mediate phase separation into droplet-like structures that recruit TDP43 variants lacking RNA binding elements. *Lab Invest* 2019; 99: 1030–1040.

[117] Miguel L, Avequin T, Delarue M, et al. Accumulation of insoluble forms of FUS protein correlates with toxicity in Drosophila. *Neurobiol Aging* 2012; 33: 1008.e1–15.

[118] Rotunno MS, Bosco DA. An emerging role for misfolded wild-type SOD1 in sporadic ALS pathogenesis. *Front Cell Neurosci* 2013; 7: 253.

[119] Passacantilli I, Frisone P, De Paola E, et al. hnRNPM guides an alternative splicing program in response to inhibition of the PI3K/AKT/mTOR pathway in Ewing sarcoma cells. *Nucleic Acids Res* 2017; 45: 12270–12284.

[120] Harvey SE, Xu Y, Lin X, et al. Coregulation of alternative splicing by hnRNPM and ESRP1 during EMT. *RNA* 2018; 24: 1326–1338.

[121] Van Nostrand EL, Pratt GA, Shishkin AA, et al. Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods* 2016; 13: 508–514.

[122] Bier E. Drosophila, the golden bug, emerges as a tool for human genetics. *Nat Rev Genet* 2005; 6: 9–23.

[123] Olesnicky EC, Wright EG. Drosophila as a Model for Assessing the Function of RNA-Binding Proteins during Neurogenesis and Neurological Disease. *J Dev Biol*; 6. Epub ahead of print 18 August 2018. DOI: 10.3390/jdb6030021.

[124] Walters R, Manion J, Neely GG. Dissecting Motor Neuron Disease With Drosophila melanogaster. *Front Neurosci*; 13. Epub ahead of print 2019. DOI: 10.3389/fnins.2019.00331.

[125] Anderson EN, Gochenaur L, Singh A, et al. Traumatic injury induces stress granule formation and enhances motor dysfunctions in ALS/FTD models. *Hum Mol Genet* 2018; 27: 1366–1381.

[126] Berning BA, Walker AK. The Pathobiology of TDP-43 C-Terminal Fragments in ALS and FTLD. *Front Neurosci*; 13. Epub ahead of print 2019. DOI: 10.3389/fnins.2019.00335.

[127] Cohen TJ, Hwang AW, Unger T, et al. Redox signalling directly regulates TDP-43 via cysteine oxidation and disulphide cross-linking. *The EMBO Journal* 2012; 31: 1241–1252.

[128] Coelho MB, Attig J, Ule J, et al. Matrin3: connecting gene expression with the nuclear matrix. *Wiley Interdiscip Rev RNA* 2016; 7: 303–315.

[129] Daigle JG, Lanson NA, Smith RB, et al. RNA-binding ability of FUS regulates neurodegeneration, cytoplasmic mislocalization and incorporation into stress granules associated with FUS carrying ALS-linked mutations. *Hum Mol Genet* 2013; 22: 1193–1205.

[130] Voigt A, Herholz D, Fiesel FC, et al. TDP-43-mediated neuron loss in vivo requires RNA-binding activity. *PLoS ONE* 2010; 5: e12247.

[131] Hibino Y, Usui T, Morita Y, et al. Molecular properties and intracellular localization of rat liver nuclear scaffold protein P130. *Biochim Biophys Acta* 2006; 1759: 195–207.

[132] Dassi E. Handshakes and Fights: The Regulatory Interplay of RNA-Binding Proteins. *Front Mol Biosci*; 4. Epub ahead of print 29 September 2017. DOI: 10.3389/fmolb.2017.00067.

[133] Appocher C, Mohagheghi F, Cappelli S, et al. Major hnRNP proteins act as general TDP-43 functional modifiers both in Drosophila and human neuronal cells. *Nucleic Acids Res* 2017; 45: 8026–8045.

[134] Berson A, Sartoris A, Nativio R, et al. TDP-43 Promotes Neurodegeneration by Impairing Chromatin Remodeling. *Curr Biol* 2017; 27: 3579-3590.e6.

[135] Pigna E, Simonazzi E, Sanna K, et al. Histone deacetylase 4 protects from denervation and skeletal muscle atrophy in a murine model of amyotrophic lateral sclerosis. *EBioMedicine* 2019; 40: 717–732.

[136] Atkin G, Paulson H. Ubiquitin pathways in neurodegenerative disease. *Front Mol Neurosci*; 7. Epub ahead of print 8 July 2014. DOI: 10.3389/fnmol.2014.00063.

[137] Chen Y, Guan Y, Zhang Z, et al. Wnt signaling pathway is involved in the pathogenesis of amyotrophic lateral sclerosis in adult transgenic mice. *Neurol Res* 2012; 34: 390–399.

[138] González-Fernández C, Mancuso R, del Valle J, et al. Wnt Signaling Alteration in the Spinal Cord of Amyotrophic Lateral Sclerosis Transgenic Mice: Special Focus on Frizzled-5 Cellular Expression Pattern. *PLoS One*; 11. Epub ahead of print 18 May 2016. DOI: 10.1371/journal.pone.0155867.

[139] Gargano JW, Martin I, Bhandari P, et al. Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in Drosophila. *Exp Gerontol* 2005; 40: 386–395.

[140] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001; 25: 402–408.

[141] Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010; 38: 576–589.

[142] Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. *http://dx.doi.org/10.1056/NEJMra1603471*. Epub ahead of print 12 July 2017. DOI: 10.1056/NEJMra1603471.

[143] Al-Chalabi A, Hardiman O, Kiernan MC, et al. Amyotrophic lateral sclerosis: moving towards a new classification system. *Lancet Neurol* 2016; 15: 1182–1194.

[144] Renton AE, Majounie E, Waite A, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011; 72: 257–268.

[145] Smith BN, Newhouse S, Shatunov A, et al. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur J Hum Genet* 2013; 21: 102–108.

[146] Kumar V, Hasan GM, Hassan MdI. Unraveling the Role of RNA Mediated Toxicity of C9orf72 Repeats in C9-FTD/ALS. *Front Neurosci*; 11. Epub ahead of print 15 December 2017. DOI: 10.3389/fnins.2017.00711.

[147] Haeusler AR, Donnelly CJ, Rothstein JD. The expanding biology of the C9orf72 nucleotide repeat expansion in neurodegenerative disease. *Nat Rev Neurosci* 2016; 17: 383–395.

[148] Balendra R, Isaacs AM. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol* 2018; 14: 544–558.

[149] Farg MA, Sundaramoorthy V, Sultana JM, et al. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol Genet* 2014; 23: 3579–3595.

[150] Shi Y, Lin S, Staats KA, et al. Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nature Medicine* 2018; 24: 313–325.

[151] Lee Y-B, Chen H-J, Peres JN, et al. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep* 2013; 5: 1178–1186.

[152] Zu T, Liu Y, Bañez-Coronel M, et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci USA* 2013; 110: E4968-4977.

[153] Mori K, Weng S-M, Arzberger T, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* 2013; 339: 1335–1338.

[154] Mackenzie IR, Arzberger T, Kremmer E, et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* 2013; 126: 859–879.

[155] Mori K, Arzberger T, Grässer FA, et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* 2013; 126: 881–893.

[156] Ash PEA, Bieniek KF, Gendron TF, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 2013; 77: 639–646.

[157] Gendron TF, van Blitterswijk M, Bieniek KF, et al. Cerebellar c9RAN proteins associate with clinical and neuropathological characteristics of C9ORF72 repeat expansion carriers. *Acta Neuropathol* 2015; 130: 559–573.

[158] Mankodi A, Lin X, Blaxall BC, et al. Nuclear RNA foci in the heart in myotonic dystrophy. *Circ Res* 2005; 97: 1152–1155.

[159] Tassone F, Iwahashi C, Hagerman PJ. FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol* 2004; 1: 103–105.

[160] Swinnen B, Robberecht W, Van Den Bosch L. RNA toxicity in non-coding repeat expansion disorders. *EMBO J* 2020; 39: e101112.

[161] Wojciechowska M, Krzyzosiak WJ. Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum Mol Genet* 2011; 20: 3811–3821.

[162] Miller JW, Urbinati CR, Teng-Umnuay P, et al. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J* 2000; 19: 4439–4448.

[163] Haeusler AR, Donnelly CJ, Periz G, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 2014; 507: 195–200.

[164] Mori K, Lammich S, Mackenzie IRA, et al. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol* 2013; 125: 413–423.

[165] Zhang K, Donnelly CJ, Haeusler AR, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* 2015; 525: 56–61.

[166] Freibaum BD, Lu Y, Lopez-Gonzalez R, et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* 2015; 525: 129–133.

[167] Hisada-Ishii S, Ebihara M, Kobayashi N, et al. Bipartite nuclear localization signal of matrin 3 is essential for vertebrate cells. *Biochem Biophys Res Commun* 2007; 354: 72–76.

[168] Xu Z, Poidevin M, Li X, et al. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc Natl Acad Sci USA* 2013; 110: 7778–7783.

[169] Mizielinska S, Grönke S, Niccoli T, et al. C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* 2014; 345: 1192–1194.

[170] Wen X, Tan W, Westergard T, et al. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron* 2014; 84: 1213–1225.

[171] Tran H, Almeida S, Moore J, et al. Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* 2015; 87: 1207–1214.

[172] Loureiro JR, Oliveira CL, Silveira I. Unstable repeat expansions in neurodegenerative diseases: nucleocytoplasmic transport emerges on the scene. *Neurobiol Aging* 2016; 39: 174–183.

[173] Donnelly CJ, Zhang P-W, Pham JT, et al. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* 2013; 80: 415–428.

[174] Lagier-Tourenne C, Baughn M, Rigo F, et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc Natl Acad Sci USA* 2013; 110: E4530-4539.

[175] Mizielinska S, Lashley T, Norona FE, et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol* 2013; 126: 845–857.

[176] Cooper-Knock J, Higginbottom A, Stopford MJ, et al. Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathol* 2015; 130: 63–75.

[177] Pandey UB, Nie Z, Batlevi Y, et al. HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. *Nature* 2007; 447: 859–863.

[178] Nichols CD, Becnel J, Pandey UB. Methods to assay Drosophila behavior. *J Vis Exp*. Epub ahead of print 7 March 2012. DOI: 10.3791/3795.

[179] He F, Krans A, Freibaum BD, et al. TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. *Hum Mol Genet* 2014; 23: 5036–5051.