Identification of Novel Modifiers of RAN Translation at *FMR1* and *C9ORF72*

by

Alexander Linsalata

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Cellular and Molecular Biology)
in the University of Michigan
2020

Doctoral Committee:

Associate Professor Peter K. Todd, Chair
Associate Professor David Turner
Professor Nils Walter
Associate Professor Bing Ye

Alexander Edward Linsalata
linsalex@med.umich.edu
ORCID iD: 0000-0001-9961-4181

Acknowledgements

I write these acknowledgements with the benefit of almost exactly one year spent in medical school and away from my doctoral research, as well as 2 weeks of all the simultaneous comforts and miseries of sheltering in place due to COVID-19. I like to think this has given me some added perspective on my experience in graduate school.

First, to Peter Todd—thank you for your patience, your guidance, and your generosity. I am as stubborn as they come (about as stubborn as I am frequently wrong), and you still allowed me the freedom to conduct, design, and prioritize experiments as I thought best. Our weekly discussions nudged (occasionally dragged) me in the direction I needed to go. I would not be the scientist I am today without the tangible and intangible resources you have invested in my growth since my first nervous day in lab. I think of you as family.

To the members of the Todd lab, past and present—you are an outrageously intelligent, talented bunch, and I still cannot believe I was fortunate enough to share in your creativity, insight, mettle, and community. There's not one piece of data I've generated that does not have your collective fingerprints on it in some way. In particular, thank you to Caitlin Rodriguez, Katelyn Green, and Shannon Wright for your collective wisdom and understanding as we all moved through PhD life together; to Mike Kearse, my first mentor in Peter's Lab, for your infinite patience and desire to teach; and to Yi-Ju (Lulu) Tseng for your endless enthusiasm, tireless labor, and readiness to tackle our genome-wide screens head-on.

To the administration at the Cell and Molecular Biology Graduate Program and the Medical Scientist Training Program (MSTP)—thank you for smoothing out all the wrinkles that have popped up over the past 7 years. I'm very confident that if it were not for your dedication to the students here, I literally would have starved many winters ago.

To my co-fellows of the MSTP—I feel like the luckiest idiot in the world to have joined this program and met each of you. Graduate school can be an oddly atomizing

experience, but we've stuck together and stuck through. I love you all like siblings, and I can't wait to see the brilliant directions your lives will take.

To the members of Funkytown Monkeypants, new and old, near and far—thank you for every single adventure. Thank you for every single pizza, taco, margarita, Ale 8, bourbon, belay, soft-catch, goat, jackfruit, debris fire, and near-death experience. Thank you for bringing me back to life when I needed it. It pains me to know that so many will soon leave Michigan, and that I have spent so much of this past year studying for medical school instead of in your company. Please know I'm already looking forward to a reunion.

To my mother, Marilyn Linsalata—you're the one who taught me by your own example how to think, how to be creative, how to be resourceful, and how to work hard. It's from you that I learned how to be a scientist without even realizing it, even before I knew I wanted to be one. Every day that we talk, and every day we don't, I become a bit more aware of how much everything I am and everything I have comes from you. Thank you for starting me on this path.

To my father, Nicholas Linsalata—it still baffles me that I cannot express to you, right now, how much I am covered in your fingerprints. But you always knew. It's from you I get my sense of humor, my sense of adventure, my patience (or, what little I have). You're in everything I do, big or small. And though you could not be there at my defense, that feeling, that knowledge, is always with me. Thank you, Dad.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

3xF— 3x FLAG
ALS— amyotrophic lateral sclerosis
AuC— area under the curve
Bel— belle
BDSC— Bloomington Drosophila Stock Center
CCDS— consensus coding sequence
CHX— cycloheximide
CNS— central nervous system
DMEM— Dulbecco's Modified Eagle Medium
DPR— dipeptide repeat
eIF— eukaryotic initiation factor
FF— firefly luciferase
FMRpolyA— *FMR1* polyalanine RAN product
FMRpolyG— *FMR1* polyglycine RAN product
FTD— frontotemporal dementia
FXPOI— Fragile X-associated premature ovarian insufficiency
FXTAS— Fragile X-associated tremor/ataxia syndrome
GA— glycine-alanine *C9ORF72* RAN product
GAP— GTPase-activating protein
GO— gene ontology
GP— glycine-proline *C9ORF72* RAN product
GR— glycine-arginine *C9ORF72* RAN product
HCR— hybridization chain reaction
ICC— immunocytochemistry
IHC— immunohistochemistry
iPSC— induced pluripotent stem cell
IRES— internal ribosome entry site
ISR— integrated stress response
LCD— low-complexity domain
LNA— locked nucleic acids
m^6A— N6-methyladenosine
m^7G— methyl-7-guanosine
MBNL— muscleblind
NL— nanoluciferase

NLS— nuclear localization signal
NPC— nuclear pore complex
NRE— nucleotide-repeat expansion
O/E— overexpression
ORF— open reading frame
P bodies— processing bodies
PA— proline-alanine *C9ORF72* RAN product
PABP— polyA-binding protein
PIC— pre-initiation complex
PR— proline-arginine *C9ORF72* RAN product
RAN— repeat-associated, non-AUG
RIP— RNA immunoprecipitation
RIPA— radioimmunoprecipitation assay
RNA-Seq— RNA sequencing
RNP— ribonucleoprotein
SCA— spinocerebellar ataxia
shRNA— small hairpin RNA
siNL— anti-nanoluciferase siRNA
siNT— non-targeting siRNA
siRNA— small interfering RNA
sgRNA— single guide RNA
TPM— transcripts per million
$tRNA_i^{Met}$— methionine-conjugated initiator tRNA
UAS— upstream activating sequence
uORF— upstream open reading frame
UTR— untranslated region

Abstract

Microsatellites are tracts of short (2-6 nucleotides) DNA units repeated in tandem. These occur throughout eukaryotic genomes and are prone to inter-generational expansion and contraction. Microsatellite expansions, also known as nucleotide-repeat expansions (NREs), underlie a diverse set of clinical disorders, many of them neurodegenerative. Historically, researchers hypothesized that NREs could elicit toxicity through a variety of pathogenic mechanisms, including protein-mediated gain- and loss-of-function, RNA-mediated gain- and loss-of-function, induction of mitochondrial stress, and DNA damage. More recently, investigators recognized that transcribed NREs can trigger initiation of translation and protein synthesis in the absence of an AUG codon, violating a widely-held assumption of protein synthesis. This process, known as repeat-associated, non-AUG (RAN) translation, yields aberrant protein products that are detected in patient tissue. Expression of RAN products is sufficient and in certain contexts necessary for those NREs outside of annotated open-reading frames (ORFs) to elicit toxicity in animal and cellular models. Despite its significance to multiple neurodegenerative disorders, the mechanism by which RAN translation occurs remains enigmatic. The goal of this thesis is to identify modulators of RAN translation at two different NREs—CGG repeats in *FMR1*, associated with Fragile X-associated tremor/ataxia syndrome (FXTAS), and G_4C_2 repeats in *C9ORF72*, associated with C9 amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)—as a means of defining the mechanics of RAN translation and identifying potential targets for therapeutic intervention.

To identify proteins and pathways that regulate RAN translation, I conducted two screens. First, I performed a candidate-based screen of eukaryotic initiation factors (eIFs), RNA helicases, and known translational regulators in a *Drosophila melanogaster* model of FXTAS. This screen identified and subsequent experiments validated five

factors as modifiers of RAN translation: the RNA helicase belle (bel)/DDX3X, eIF4B, eIF4H, eIF1, and eIF5. Knockdown of *bel*/*DDX3X* suppressed the expression of *FMR1*- and *C9ORF72*-associated RAN products in flies, heterologous human cells, primary rodent neurons, and patient-derived induced pluripotent stem cells (iPSCs), while also mitigating *FMR1* $(CGG)_n$-elicited toxicity in flies and rodent neurons. Identification of these specific factors is suggestive of the roles that RNA secondary structure and start-codon fidelity play in RAN translation.

Second, I performed a high-throughput, genome-wide, small-interfering RNA (siRNA) screen in HEK293 cells that expressed a reporter for RAN translation of the G_4C_2 NRE in *C9ORF72*, with additional screening for modifiers of *FMR1*-based RAN translation in a subset of the genome. This identified 561 and 75 high-confidence, RAN translation-selective candidate modifiers of *C9ORF72* and *FMR1* RAN translation, respectively. In subsequent experiments depletion of two candidates, the RNA-binding protein hnRNPA_2/B_1 and the proteasomal subunit PSMB5, significantly and selectively suppressed the expression of multiple *FMR1*- and *C9ORF72*-associated RAN products tested in reporter assays. These studies provide insight into the shared factors and pathways required for RAN translation at two independent NREs, while providing a mechanistic framework for developing new therapeutic strategies for FXTAS, C9 ALS/FTD, and other NRE-associated neurodegenerative disorders.

Chapter 1
Introduction[1]

Abstract

Nucleotide-repeat expansions (NREs) underlie a heterogeneous group of neurodegenerative and neuromuscular disorders for which there are currently no effective therapies. Recently, it was discovered that NREs in mRNA can support translation initiation in the absence of an AUG start codon across a wide variety of sequence contexts, and that the products of these atypical translation initiation events contribute to neuronal toxicity. The mechanism of this process, known as repeat-associated, non-AUG (RAN) translation, remains unclear. RAN translation at two loci at least—CGG repeats in *FMR1* [(associated with Fragile X-associated tremor/ataxia syndrome (FXTAS)] and G_4C_2 repeats in *C9ORF72* [(associated with amyotrophic lateral sclerosis (ALS)/frontotemporal dementia (FTD)]—generates homopolymeric and dipeptide-repeat (DPR) proteins that elicit toxicity through a variety of mechanisms. Additional research into which protein factors [particularly eukaryotic initiation factors (eIFs) and other known regulators of translation] mediate or modulate RAN translation would prove advantageous, both in understanding this odd biological phenomenon and identifying potential targets for therapeutic intervention.

[1] Parts of this chapter were originally published as Green KM, Linsalata AE, Todd PK. RAN translation—What makes it run? *Brain Research* **1647**, 30-42 (2016).

Nucleotide-repeat expansions (NREs) underlie a heterogeneous group of primarily neurological diseases that, in aggregate, impact a significant number of patients[1,2]. They elicit toxicity through a variety of mechanisms delineated over the past 25 years. When located within protein-coding open reading frames (ORFs), NREs lead to cellular toxicity via the synthesis of homopolymeric (frequently polyglutamine) proteins[3-6]. Toxicity results from both disruption of the native functions of the proteins in which the repeats reside (loss-of-function), as well as perturbations in cellular proteostasis independent of protein context (gain-of-function). By contrast, NREs located outside of known protein-coding ORFs can elicit changes in the expression of the gene in which they reside, leading to reduced or enhanced expression at the transcript and protein level[7]. Non-coding NREs can also elicit toxicity as RNA by binding to and sequestering specific RNA-binding proteins via presentation of a repetitive sequence or structural motif[8-10], or more novelly, forming phase-separated gels[11].

The discovery of repeat-associated, non-AUG (RAN) translation blurred the line that defines which NREs elicit toxicity via protein gain-of-function/loss-of-function and which act through RNA-mediated gain-of-function[12,13]. This non-canonical process of translation initiation, which occurs in the absence of an AUG initiation codon and in multiple reading frames, enables elongation through NREs in mRNA, producing multiple homopolymeric or polypeptide repeat-containing proteins. Originally described in association with CAG NREs causative for spinocerebellar ataxia type 8 (SCA8), this process also occurs in association with expansions of CAG, CUG, G_4C_2, G_2C_4, CCUG, CAGG, UGGAA, CGG, and CCG[14-26]. The recognition of these initiation events has led to a flurry of activity within the scientific and medical communities, with a significant

body of work now demonstrating 1) the presence of RAN-translated peptides in patient and animal-model tissue across a wide spectrum of neurodegenerative disorders, and 2) a link between these peptides and cellular toxicity. Despite this interest, the mechanism or mechanisms by which RAN translation initiates—and the ways in which it departs from general, AUG-initiated translation—remains unclear.

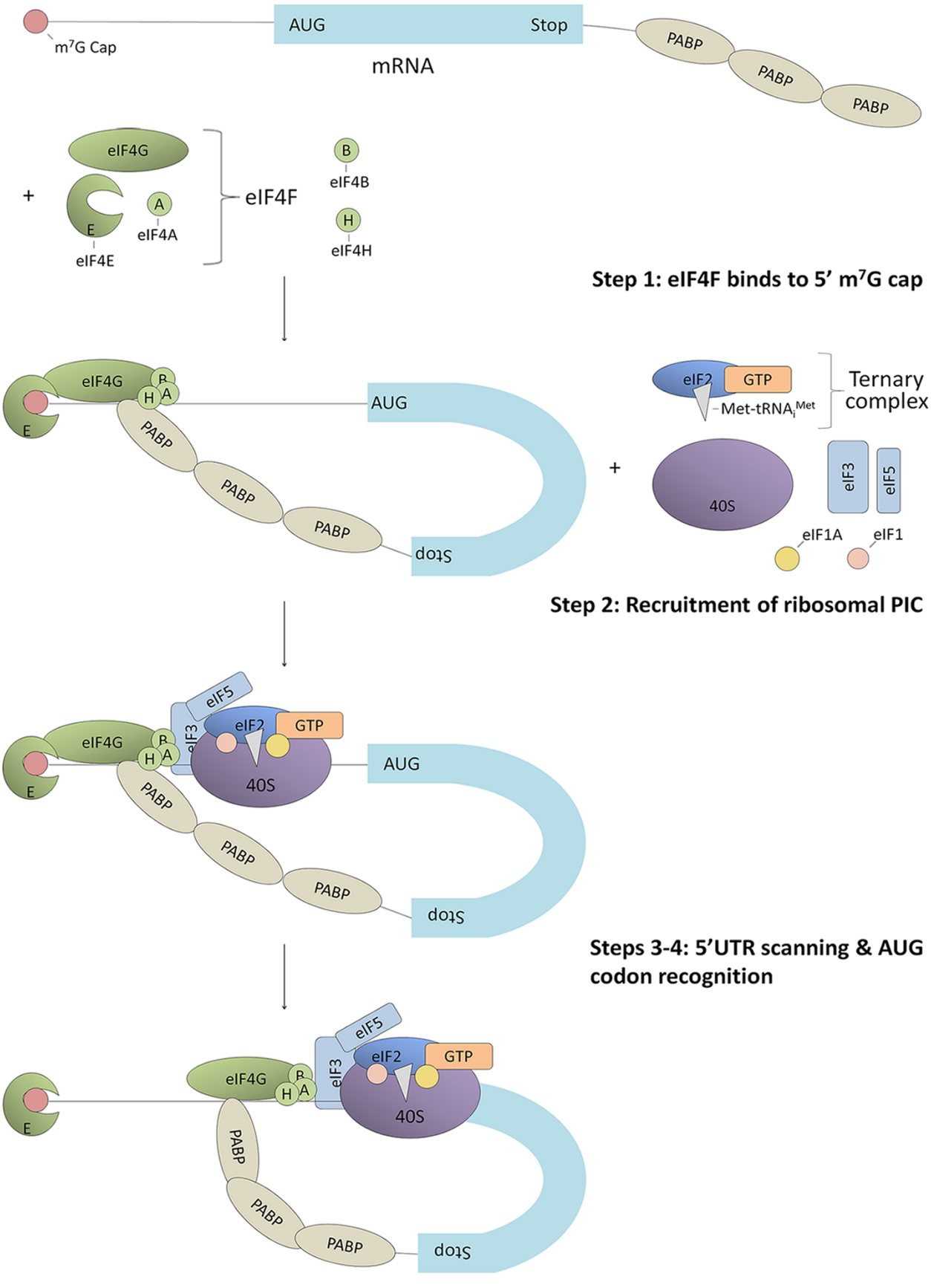
Translation Initiation, Classically

Translation initiation is the step-wise assembly of elongation-competent 80S ribosomes at start codons of mRNA. It is a highly complex process, entailing the concerted activity of at least nine eukaryotic initiation factors (eIFs)[27]. In most cases, initiation begins with recognition of the 5' methyl-7-guanosine (m^7G) cap on mRNA by the eIF4F complex (Figure 1.1, Step 1)[28-30]. The eIF4F complex is composed of eIF4E (the direct cap-binding subunit), eIF4G (a scaffolding subunit), and eIF4A (a DEAD-box RNA helicase). eIF4G also recognizes the polyA-binding protein (PABP)[31,32], which in turn binds to the 3' polyA tail on mRNAs. This is thought to result in circularization of the mRNA and greater initiation efficiency[33,34].

The eIF4F complex, still bound to the m^7G cap, is joined by the 43S pre-initiation

---

Figure 1.1: Canonical Scanning Model of Translation Initiation, Part I.
Step 1: The eIF4F complex, composed of eukaryotic initiation factors (eIFs) eIF4E, eIF4G, and eIF4A, binds to the 5' methyl-7-guanosine (m^7G) cap with eIF4B and eIF4H. PolyA-binding protein (PABP) associates with eIF4G to circularize the mRNA.
Step 2: The eIF4F complex recruits the 43S pre-initiation complex (PIC), composed of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, and the ternary complex, itself consisting of the initiator methionine-tRNA ($tRNA_i^{Met}$), eIF2, and GTP.
Step 3-4: The PIC and components of the eIF4F complex scan through the 5' untranslated region (UTR) in the 5' to 3' direction until encountering an AUG start codon in a good Kozak context.


m7G Cap
AUG
Stop
mRNA
PABP
PABP
PABP
eIF4G
eIF4E
eIF4A
eIF4F
eIF4B
eIF4H
Step 1: eIF4F binds to 5' m7G cap
AUG
Stop
eIF2
GTP
Met-tRNAiMet
Ternary complex
40S
eIF3
eIF5
eIF1A
eIF1
Step 2: Recruitment of ribosomal PIC
Steps 3-4: 5'UTR scanning & AUG codon recognition

complex (PIC), composed of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, and the ternary complex [in turn composed of methionine-conjugated tRNA ($tRNA_i^{Met}$) and eIF2-GTP; Figure 1.1, Step 2]. This joining of the 43S PIC to the eIF4F complex is mediated by an eIF4G-eIF3 interaction[35]. Successful translation of most eukaryotic mRNAs is thought to require the RNA helicase activity of eIF4A in order to resolve RNA-RNA secondary structures adjacent to the m^7G cap and prepare a "landing pad" for the 43S PIC[36].

According to the scanning model of translation initiation[37,38], the 43S PIC and components of the eIF4F complex scan through the 5' untranslated region (UTR) in the 5' to 3' direction (Figure 1.1, Step 3). This stage is also known to require eIF4A in order to resolve weaker internal secondary structures[39,40], though additional helicases such as DHX29 and DHX33[41-43] assist in melting stronger structures. eIF4A is also joined by two co-stimulatory factors, eIF4B and eIF4H, both of which are required to achieve maximal eIF4A-mRNA binding, helicase catalytic efficiency, and translational efficiency[44-46].

A related RNA helicase, the DEAD-box protein DDX3X [Ded1 in yeast, belle (bel) in *Drosophila melanogaster*], has received particular attention for its function in 43S PIC scanning. DDX3X/Ded1 exhibits efficient, interdependent ATPase and RNA helicase activities *in vitro*[47-51] and acts by binding to single-stranded RNA and destabilizing local RNA-RNA helices[52-55]. Like most DEAD-box helicases, DDX3X/Ded1 exhibits no specificity or preference for RNA sequences. *In vivo*, DDX3X/Ded1 directly interacts with eIF4E and PABP[56,57], eIF4G[58], and the 40S ribosomal subunit itself[59]. It has been shown to be a component of 43S PICs[60,61] with functions distinct from and complementary to those of eIF4A and its cofactors[46,50,62-64]. On both individual-transcript

and translatome-wide levels, DDX3X/Ded1/bel has been demonstrated to be required for ribosomal scanning and translation initiation of transcripts bearing long, GC-rich (and putatively secondary-/tertiary-structured) 5' UTRs[46,59,61-68]. DDX3X/Ded1 has also been identified as a component of stress granules and processing (P)-bodies—membrane-less organelles with roles in mRNA triage, mRNA degradation, and translational suppression—and can, under certain circumstances, inhibit translation[56,58,69-73]. Thus, DDX3X/Ded1/bel is a multifunctional RNA helicase positioned at the nexus of multiple translational regulatory pathways, with demonstrated functions and binding partners consistent with roles in both facilitation and inhibition of mRNA translation.

Having unwound any secondary structures, the 43S PIC scans until encountering an AUG codon in good Kozak context, (A/G)NNAUGG (Figure 1.1, Step 4)[74,75]. At this point, base-pairing between the AUG codon and CAU anti-codon loop on $tRNA_i^{Met}$ results in the ejection of eIF1, a factor which, along with eIF1A, increases the stringency of AUG start-codon selection[76-82]. eIF2 hydrolyzes its bound GTP with the assistance of eIF5, its associated GTPase-activating protein (GAP; Figure 1.2, Step 5). Regulating the abundance and availability of eIF5 is one mechanism through which cells can "tune" start-codon fidelity in response to stimuli[83-86]. At this point, the 40S ribosome is committed to its selection of start codon, and forms a tighter interaction with the substrate mRNA, collectively known as the 48S PIC. In the final stages of initiation, the 40S subunit is joined by the 60S ribosomal subunit (Figure 1.2, Step 6), the majority of remaining eIFs are ejected, eIF5B hydrolyzes its bound GTP (Figure 1.2, Step 7), and translation elongation begins with formation of the first peptide bond[87].

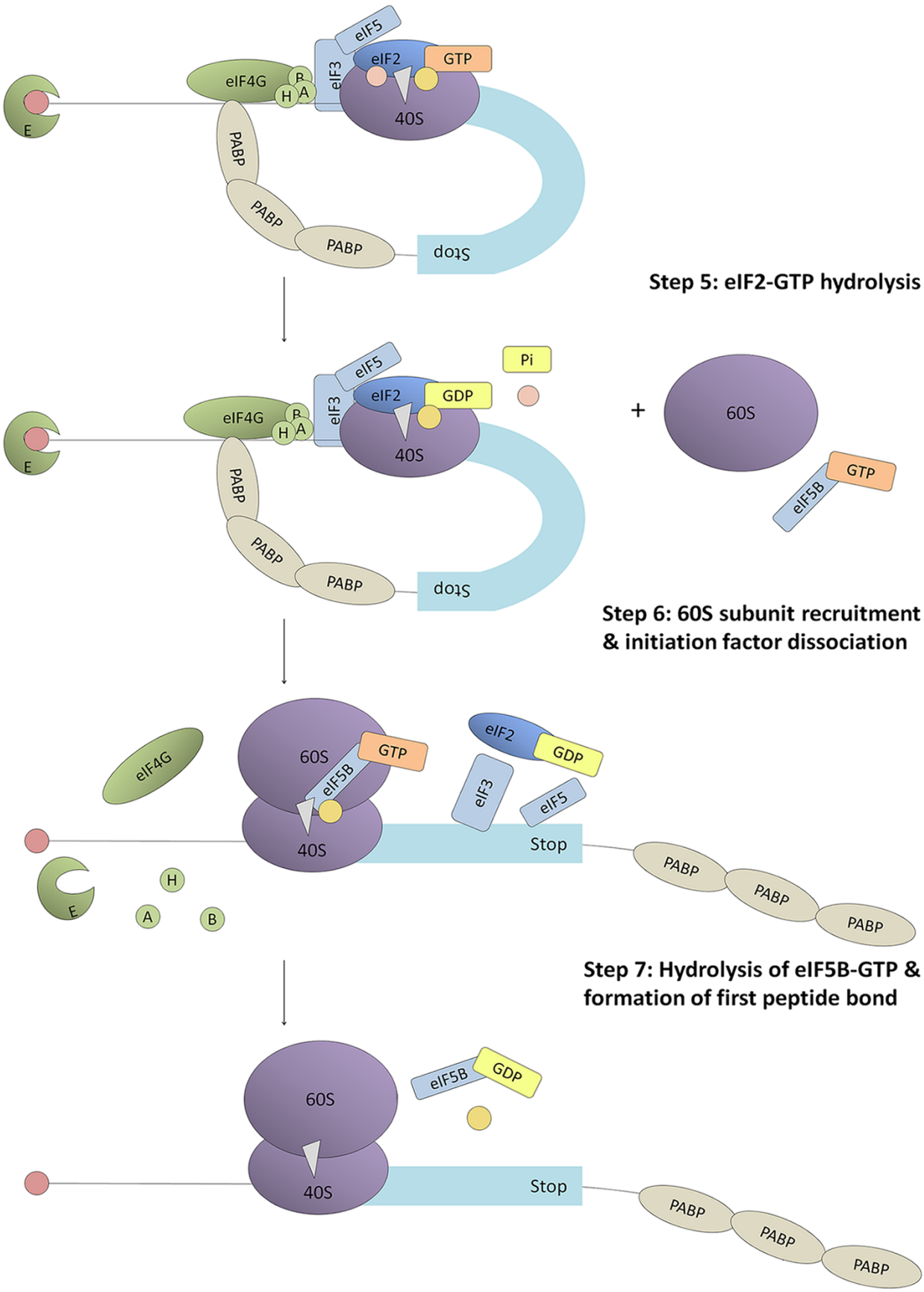

eIF5
eIF3
eIF2
GTP
eIF4G
R
H
A
40S
E
PABP
PABP
PABP
Stop
Step 5: eIF2-GTP hydrolysis
Pi
GDP
+
60S
GTP
eIF5B
Step 6: 60S subunit recruitment
& initiation factor dissociation
eIF4G
60S
GTP
eIF5B
eIF2
GDP
eIF3
eIF5
40S
Stop
H
A
B
Step 7: Hydrolysis of eIF5B-GTP &
formation of first peptide bond

Figure 1.2: Canonical Scanning Model of Translation Initiation, Part II.
Step 5: eIF1 dissociates from the PIC, and eIF2 hydrolyzes GTP with the assistance of eIF5. This commits the 40S ribosome to translation initiation at the present AUG codon.
Step 6: eIF5B-GTP promotes the recruitment of the 60S subunit and displacement of most remaining eIFs.
Step 7: eIF5B hydrolyzes its bound GTP and dissociates with eIF1A, establishing the elongation-competent 80S ribosome.

The Role of Secondary Structure in Translation Initiation

Secondary and tertiary structures in 5' UTRs impact translation initiation both positively and negatively, depending on the structure's location. When placed upstream of an AUG start codon, they inhibit initiation, either by blocking the eIF4E-m^7G interaction when located adjacent to the cap or by impeding 5'-to-3' translocation of the 43S PIC when located internally in the 5' UTR[75,88-91]. In contrast, secondary structures downstream of start codons facilitate initiation at imperfect start codons (i.e., those with poor Kozak context and even non-AUG codons[90,92]). Initiation at non-AUG codons occurs at reduced efficiency relative to AUG codons in *in vitro* translation systems[93,94], but the presence of secondary structures downstream markedly increases initiation efficiency[90-92,95].

Recent advances in ribosomal footprinting methodologies suggest that these *in vitro* findings may reflect a common but heretofore unrecognized set of initiation events *in vivo.* Ribosome profiling combines the traditional aspects of an RNase-protection assay with next-generation sequencing to identify the positions of initiating and elongating ribosomes on mRNAs on a transcriptome-wide scale[96]. This technique has generated evidence for thousands of unpredicted translation initiation events, many of which occur at non-AUG codons[96-99]. This is especially true for upstream open reading

frames (uORFs): short ORFs upstream of canonical, annotated ORFs in the same mRNA transcript. Many uORFs appear to play regulatory roles in translation that are dependent on metabolic conditions and cell-cycle stage[100]. These findings are also supported by a variety of studies utilizing mass spectroscopy to confirm the presence of uORF-encoded peptides *in vitro* and *in vivo*, many of which possess functional roles[101-107]. Thus, initiation at non-AUG codons occurs *in vivo*, is regulated in part by mRNA-specific *cis* factors, might be regulated in part by *trans*-acting eIFs as well, and appears to play important regulatory roles in protein synthesis.

RAN Translation: an Unexpected Finding at NREs

Above, mRNA secondary structure encodes "instructions" for how a given transcript is to be translated. Laura Ranum and colleagues' discovery of RAN translation introduced a novel mode of translation to this mechanistic multitude[14]. Expansions of protein-coding CAG repeats in *ATXN8* cause the neurodegenerative disorder SCA8. Unexpectedly, Zu *et al*.[14] demonstrated that mutation of the only AUG codon upstream of expanded $(CAG)_n$ does not abrogate protein synthesis. In fact, translation initiated in multiple reading frames, generating homopolymeric proteins with glutamine, serine, or alanine repeats depending on the reading frame. RAN translation products from all three reading frames accumulated in cells transfected with expanded $(CAG)_n$ reporters, occasionally even within the same transfected cell. Antisense *ATXN8* transcripts bearing expanded CUG repeats also supported RAN translation. Antibodies generated against the predicted polyalanine product of *ATXN8* sense mRNA recognized

a protein in the cerebellums of SCA8 human patients and mouse models. Finally, a similar approach provided *in vivo* evidence of a polyglutamine RAN product from *DMPK* antisense RNAs bearing expanded CAG repeats, causative for myotonic dystrophy type I.

In silico modeling predicts and *in vitro* analysis suggests that consecutive GC-rich repeats form hairpin structures[108-110] sufficiently stable to impede 43S PICs[75,88-91,111]. In parallel, Zu *et al*.[14] demonstrated that RAN translation of *ATXN8* mRNAs is dependent on the GC content and length of these repeats—factors that directly correlate with the stability of secondary structure. These results support the hypothesis that secondary-structured elements within mRNA are critical for RAN translation to initiate. Kozak's work[90,92,95] could explain how: downstream secondary structures enhance initiation at upstream non-AUG codons. Specifically, the increase in non-AUG initiation is maximal when a hairpin falls 14 nucleotides downstream of the initiator codon. Based on the known size of ribosomes, this orientation would place the start codon within the P site of the 40S ribosome, opposite the anti-codon loop of $tRNA_i^{Met}$ [92]. These observations have led to our current working model of RAN translation, in which secondary structures formed by GC-rich NREs cause 43S PICs to stall, increasing initiation at optimally positioned non-AUG codons. The link between 43S PIC stalling and non-AUG initiation could be explained by several possible mechanisms: increased dwell time of the 40S subunit over an imperfect codon-anticodon match, congestion of mRNA-bound 43S PICs upstream of the NRE[90,92,95], dissociation of key eIFs that help determine AUG start codon fidelity (e.g., eIF1[112,113]), or even alternative ribosomal conformations[114,115].

Fragile X-Associated Tremor/Ataxia Syndrome and RNA-Mediated Neurodegeneration

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset neurodegenerative disorder caused by a CGG NRE in the 5' UTR of *FMR1*. In unaffected individuals, repeats number less than 45. Individuals with FXTAS carry between 55 and 200 repeats, known as the "premutation" range[116-119]. Premutation NREs result in enhanced transcription of *FMR1* mRNA[120-123]. In contrast, expansion to >200 repeats triggers transcriptional silencing of the *FMR1* locus, leading to loss of *FMR1* mRNA and the Fragile X protein, FMRP. Transcriptional silencing manifests in Fragile X syndrome, a clinically distinct neurodevelopmental disorder characterized by intellectual disability and features of autism-spectrum disorder[124,125]. Approximately 40% of male premutation carriers develop FXTAS (approximately 1:3,000 of the total population), with increased penetrance at older ages and larger repeat sizes[126-128]. FXTAS is characterized clinically by action tremors, ataxia, parkinsonism, and cognitive decline, and pathologically by both neuronal and non-neuronal ubiquitinated inclusions throughout the cerebral cortex, brainstem, and cerebellum[129]. Premutation carrier women are also at increased risk of Fragile X-associated premature ovarian insufficiency (FXPOI)[130,131].

Before the discovery of RAN translation, the dominant hypothesis of FXTAS pathogenesis explained NRE-elicited toxicity by invoking an RNA gain-of-function mechanism[132], in which expanded $(CGG)_n$ RNAs sequester essential RNA-binding proteins[10]. This model is not without supporting evidence. CGG-expanded *FMR1*

mRNAs are more abundant than non-expanded *FMR1*[120-123]. Expanded $(CGG)_n$ in heterologous mRNAs is sufficient to drive toxicity in Purkinje neurons[133], which hints at an RNA-mediated gain-of-function mechanism. In addition, RNA pathology is observed—in at least one patient, $(CGG)_{113}$ *FMR1* mRNAs formed nuclear foci[134], which can be recapitulated in cultured COS7 cells[135]. $(CGG)_n$ RNAs could perturb the functions of clinically relevant RNA-binding proteins, since ubiquitinated inclusions purified from FXTAS-patient frontal cortex were found to contain both muscleblind (MBNL)[136]—an RNA splicing regulator whose sequestration by expanded CUG repeats is a key event in myotonic dystrophy pathogenesis[8,9]—and hnRNPA_2/B_1—an RNA-binding protein which binds to the *FMR1* 5' UTR[137-139] and mutations in which are associated with other neurodegenenrative disorders[140]. In addition, overexpression of hnRNPA_2/B_1 ameliorates $(CGG)_n$-and NRE-elicited toxicity[22,139,141], implying that its sequestration might have contributed to $(CGG)_n$-elicited toxicity. Expanded $(CGG)_n$ RNAs also form complexes with the splicing factor Sam68, and Sam68 disruption recapitulates some of the splicing changes observed in FXTAS brains[135]. Expanded $(CGG)_n$ RNAs form complexes with DGCR8 and DROSHA[142], disrupt miRNA maturation[142], and elicit a change in the distrubution of mRNAs critical for proteostasis[143]. Finally, $(CGG)_n$-elicted DNA damage and mitochondrial stress have also been considered as possible pathogenic contributers[144].

FXTAS and RAN Translation

It was in this context that our lab demonstrated that the CGG-expanded *FMR1* 5'

UTR supports RAN translation[19]. Initiation within the 5' UTR occurs in at least two reading frames in the absence of an AUG start codon: the GGC (+1) frame yields a polyglycine product (FMRpolyG), and the GCG (+2) frame yields a polyalanine product (FMRpolyA; Figure 1.3A). FMRpolyG accumulates in ubiquitinated inclusions in patient tissue and cellular and animal disease models, is necessary to elicit toxicity in *Drosophila* and mouse models of disease[145], induces proteasome perturbations in *Drosophila* and HeLa cells[19,146,147], and might be responsible for laminopathy in transgenic mice and FXTAS patient induced pluripotent stem cell (iPSC)-derived neurons[145]. Antisense *FMR1* transcripts bearing expanded $(CCG)_n$ repeats also undergo RAN translation in transfected cells and FXTAS patient tissue[21]. In an inducible mouse model of FXTAS that expresses the *FMR1* 5'UTR with 90 CGG repeats, turning off transgene expression reversed the formation of neuronal FMRpolyG-positive inclusions and repeat-elicited behavioral deficits[148]. Finally, FMRpolyG-positive inclusions are observed in ovarian stromal cells in FXTAS mouse models and a FXPOI patient, suggesting FMRpolyG expression could be linked to other Fragile X-related clinical phenotypes[149].

In experiments using transfectable reporters for FMRpolyG and FMRpolyA in cultured human cells, we detected synthesis of FMRpolyG constructs bearing 0-100 CGG repeats, with expression strongly correlating with repeat number[19,150]. Using a combination of point mutagenesis and stop-codon insertion, we also identified the start codons used for FMRpolyG synthesis: an ACG and a GUG 5' to the NRE. In contrast, synthesis of FMRpolyA reporters was more stringently repeat-dependent. Moreover, inserting a stop codon immediately 5' to the NRE did not preclude FMRpolyA

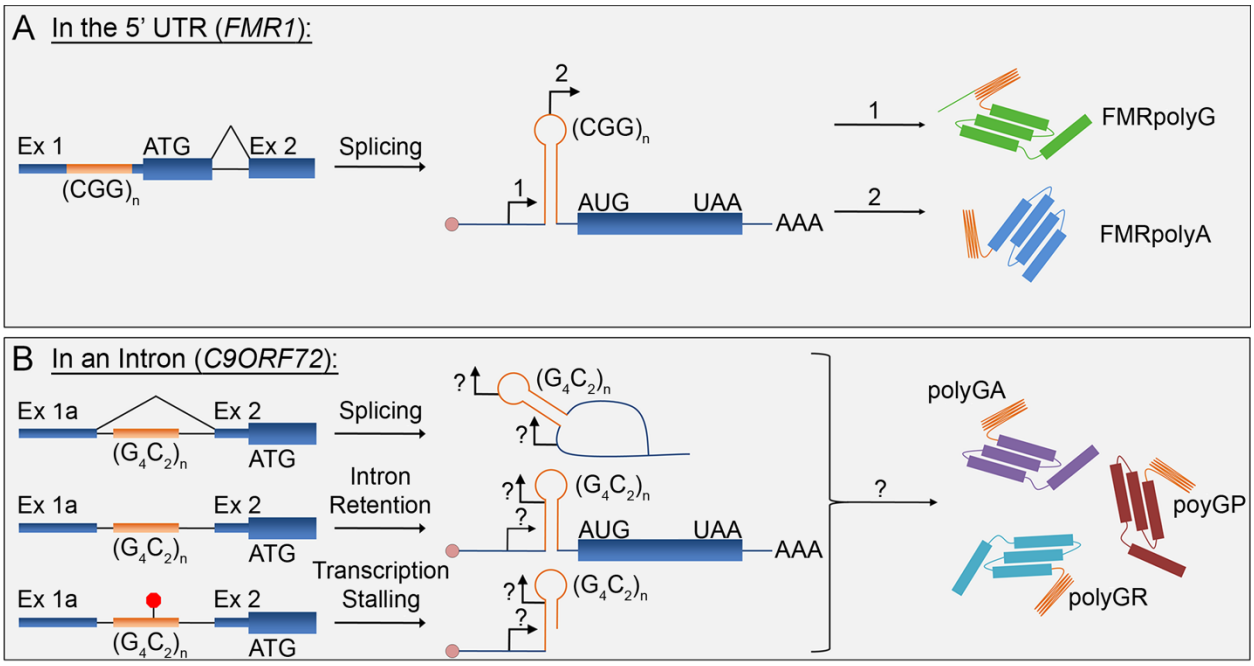


Figure 1.3: Production of RAN-Translated Proteins Across Different Sequence Contexts.
A When located in the 5′ UTR, as in *FMR1*, expanded GC-rich repeats trigger initiation of repeat-associated, non-AUG (RAN) translation upstream of the canonical AUG start codon, leading to the production of FMRpolyG (1) and FMRpolyA (2).
B When located in an intron, as in *C9ORF72*, it is unclear what RNA species is the substrate for RAN translation: a spliced lariat, an aberrantly spliced transcript in which the intron is retained, or a 3′ truncated RNA resulting from stalled transcription. The translated RNA species yields GA, GR, and GP dipeptide repeat (DPR)-containing RAN-translation products. In both panels, generation of RAN products from antisense transcripts has been excluded for clarity.

---

expression, suggesting that RAN translation can initiate within the NRE in the GCG frame. These results raise the intriguing possibility that RAN translation of the same sequence can differ mechanistically in different reading frames. Additional experiments demonstrated that FMRpolyG and FMRpolyA synthesis is dependent on the initial eIF4E-m^{7}G interaction and the helicase activity of eIF4A[150], all of which suggest that the initial stages of RAN translation closely resemble canonical AUG-initiated translation. These results have been foundational for our current model of RAN translation.

RAN Translation at G_4C_2/G_2C_4 Repeats in C9 Amyotrophic Lateral Sclerosis and Frontotemporal Dementia

The *C9ORF72* G_4C_2/G_2C_4 NRE was identified by two groups in 2011 as the most common known cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)[151,152]. ALS is the most frequently occurring form of motor neuron disease, affecting approximately 2-4/100,000 individuals[153], and is characterized by progressive paralysis typically leading to death within two to three years after onset. FTD is the second most common form of presenile dementia and affects approximately 20/100,000 individuals between the ages of 45 and 65[154]. Although ALS and FTD each manifest with a unique set of symptoms and pathology, they are believed to constitute two ends of a single disease spectrum. Approximately 50% of ALS patients develop FTD-like cognitive and behavioral impairment[155,156], while up to 50% of FTD patients develop motor dysfunction[157]. Additionally, TDP-43-positive inclusions are present within the neurons and glia of a majority of ALS patients, as well as in the most common familial variants of FTD[158].

The G_4C_2/G_2C_4 NRE, located within the first intron of transcript isoforms 1 and 3 and the promoter region of isoform 2, is expanded from 2-25 repeats in healthy individuals to >1,000 repeats in C9-associated ALS/FTD patients[151,152,159]. Both sense and antisense strands of *C9ORF72* are transcribed in mutation carriers, resulting in the production of G_4C_2/G_2C_4-repeat containing RNAs[17,18,26]. These expanded repeats are both predicted to form highly stable RNA secondary structures, with the sense RNA repeat generating G-quadruplexes and hairpins *in vitro*[160-163] and the antisense RNA

repeat shown to assume A form-like double helices *in vitro*[116].

Dipeptide Repeat Proteins in Patients and Model Systems

In addition to TDP-43-positive inclusions within both neurons and glia[158], neuronal TDP-43-negative inclusions that co-stain for ubiquitin and ubiquitin-binding proteins are found throughout the central nervous system (CNS) of C9 ALS/FTD patients[164,165]. Immunohistochemical (IHC) analysis by multiple laboratories indicates that RAN translation-derived proteins constitute these TDP43-negative inclusions[15-18,26]. A total of six different dipeptide repeat (DPR) proteins are generated from G_4C_2/G_2C_4 transcripts (Figure 1.3B). Specifically, glycine-alanine (GA) and glycine-arginine (GR) DPRs are generated from the sense strand, proline-alanine (PA) and proline-arginine (PR) arise from the antisense strand, and two glycine-proline (GP)-containing proteins arise from RAN translation of both strands.

DPRs form cytoplasmic and intranuclear inclusions throughout neurons in the CNS. The distribution of DPRs throughout the brain is highly variable, however, with the highest burden occurring in the hippocampus, cerebellum, neocortex, and thalamus[15,166,167]. Although limited by potential differences in antibody affinities, IHC studies also suggest that the different DPRs are not present in equal abundance. In several brain regions assessed with multiple independently-generated antibodies for each DPR, GA appears to be the most abundant, followed by GP and GR, while the DPRs derived exclusively from antisense transcripts (PA and PR) appear to be least abundant[16,17,168]. Expression of $(G_4C_2)_n$ reporters in *in vitro* expression systems,

transfected HEK293 cells, and transfected rodent neurons recapitulates this differential abundance[169,170].

Although their distribution throughout the brain raises questions about their exact role in disease, it is clear from studies *in vitro* and *in vivo* that DPR expression in isolation can induce neurodegeneration. From yeast[171], to *Drosophila*[172-176], cultured cells[18,167,177,178], and primary mammalian neurons[167,175,179], DPR expression leads to cell death and/or reduced survival. In many of these studies, DPR expression was sufficient to trigger toxicity, as demonstrated by the use of alternative codons in place of a G_4C_2 NRE that allow for DPR production in the absence of the potentially toxic repeat-containing RNA species[167,171,173,175-179]. Furthermore, transgenic flies expressing G_4C_2 repeats of various lengths with stop codons in all three reading frames formed RNA foci, but only flies containing pure repeats produced GR and GP DPRs and exhibited significant cell death[173].

Of the myriad possible pathogenic effects of DPR expression, disruption of nuclear-cytoplasmic transport received the most attention initially. Multiple components of the nuclear pore complex (NPC), m^7G cap-binding proteins, and factors that mediate and regulate nuclear import/export interact genetically with G_4C_2 repeats in *Drosophila* and yeast[171,172,180]. This interaction is mediated at least in part by the GA and PR RAN products[171,180,181], though G_4C_2-containing RNAs might also contribute[182]. G_4C_2-expressing flies *in vivo* and S2 cells *in vitro*, and GA_{50}-expressing murine cortical neurons *in vivo*, exhibit nuclear envelope abnormalities[172], mislocalization of key NPC components[181,182], and defects in nuclear-cytoplasmic transport[171,172,180,182], much of which is also observed in C9 ALS/FTD patient tissue and iPSC-derived neurons[172,182].

The hypothesis that the G_4C_2 NRE drives toxicity by impairing nuclear-cytoplasmic transport is particularly intriguing because it conceptually links the NRE to the mislocalization of TDP43, a pathognomonic hallmark of ALS.

Further investigation has focused on the propensity of DPRs to phase-separate[183-187] and their inclusion in membraneless organelles[11]. Several ALS/FTD-associated proteins, including FUS, TIA1, $hnRNPA_1$, and $hnRNPA_2/B_1$, undergo liquid-liquid phase separation, a process in which proteins containing low-complexity domains (LCDs), sometimes along with RNA[188], form "droplets" *in vitro* and "granules" *in vivo*[189-193]. Under normal, physiological conditions these membraneless organelles—nucleoli, nuclear speckles, Cajal bodies, P-bodies, stress granules, and others—are thought to play various roles in RNA metabolism. However, disease-associated mutations in LCD-containing proteins frequently alter the dynamics of these structures, either accelerating their formation, inhibiting the diffusion of molecules within them, or preventing their disassembly[189-193]. In rapid succession, multiple groups demonstrated that arginine-rich DPRs (GR and PR) interact with LCD-containing proteins and alter the dynamics of liquid-liquid droplets *in vitro* and *in vivo*[180,194-196], leading to toxicity and cell death by disrupting nucleoli, inhibiting synthesis of rRNA, impairing pre-mRNA splicing, and suppressing mRNA translation[175,177,178,194,195,197]. It is worth noting that $(G_4C_2)_n$-containing RNAs can also promote phase-separation of LCD-containing proteins[198]. Nevertheless, neutralizing the numerous possible toxic effects of DPRs must remain a high priority for any therapeutic intervention for C9 ALS/FTD.

The Mechanism of RAN Translation at G_4C_2/G_2C_4 Repeats

Despite compelling evidence that RAN translation and the so-generated DPRs are involved in pathogenesis, little is known about the mechanism by which expanded G_4C_2/G_2C_4 repeats trigger RAN translation. RAN translation at G_4C_2/G_2C_4 repeats is repeat length-dependent with more robust DPR production occurring with longer repeats[17,18,163,169,170], consistent with observations at CAG and CGG repeats[14,19,150]. Interestingly, the apparent minimum repeat-length requirements for RAN translation, determined using heterologous *in vitro* G_4C_2 expression, in the sense reading frames correlate indirectly with respective DPR abundance in patient tissue[168]. While these differences in length requirements may reflect artifacts of detection based on antibody avidity, differences in DPR solubility, or differential rates of protein degradation, they could also indicate an inherent discrepancy in RAN translational efficiency across reading frames. For example, different RNA secondary structures might favor initiation in certain frames at shorter repeats, while an increase in structural variability or frame-shifting at larger repeat sizes becomes more prominent[170].

Beyond these initial insights lie a series of unanswered questions regarding the mechanism of RAN translation at G_4C_2/G_2C_4 repeats. First, it remains unknown exactly what *C9ORF72* RNA species actually undergo RAN translation *in vivo*. The G_4C_2 NRE is located within the first intron of *C9ORF72*. Therefore, in patients, RAN-translated G_4C_2 repeats could conceivably derive from a retained intron, a spliced intron lariat, or within aberrant disease-specific transcripts generated by transcriptional stalling (Figure 1.3B). There is evidence for generation of such aberrant transcripts, at least *in vitro*[161]. The ratio of exon1a-intron1 (unspliced or abortive) RNA to exon1-exon2 (mature, spliced) RNA, however, is not altered in C9 ALS/FTD iPSC-derived neurons and patient

brain tissue relative to controls, arguing against significant production of truncated transcripts or increased intron retention[174]. However, a recent study suggests that intron retention occurs with some frequency in both control and C9 patient cells[199], and this may only have pathological consequences when the expanded repeat is present. Therefore, the lack of increased retention does not rule out the possibility that such transcripts undergo RAN translation in C9 patients.

In our and others' hands, m^7G-capped reporters bearing expanded G_4C_2 repeats in their 5' leader sequences were translated more efficiently than A-capped reporters, indicating that the G_4C_2 NRE does not support internal ribosome entry and suggesting that spliced G_4C_2-containing lariats are not as efficiently translated[169,170]. In earlier *Drosophila* experiments, insertion of the NRE into an efficiently spliced intron dramatically reduced both RAN translation and NRE-elicited toxicity compared to repeats placed into a 5' leader[174]. However, DPR synthesis from intronic G_4C_2 repeats was still sufficient to elicit toxicity when *Drosophila* are grown at elevated temperatures, indicating that an intronic context is able to support RAN translation under certain conditions. Whether the limited amount of DPRs observed was produced from a spliced or retained intronic repeat in this experiment is unclear. If a lariat is the RNA species translated, then some mechanism must exist for it to escape normal nuclear degradation mechanisms, translocate to the cytoplasm, and engage with translational machinery.

RAN Translation and The Integrated Stress Response

Initiation at non-AUG codons is a defining feature of RAN translation. In canonical translation, start-codon fidelity is regulated by several eIFs—including eIF1, eIF1A, and eIF5—as well as multiple upstream pathways, each responsive to a particular insult to cellular homeostasis[36,76-81,85,100,200,201]. In brief, viral infection and oxidative, starvation, and proteostatic stress converge to simultaneously inhibit AUG-initiated translation and promote non-AUG-initiated translation via the integrated stress response (ISR), culminating in phosphorylation of eIF2α and subsequent limited bioavailability of ternary complexes[201]. Such stressors also lead to a selective increase in RAN translation of CGG-expanded *FMR1* and G_4C_2-expanded *C9ORF72* reporters in a phospho-eIF2α-dependent manner[169,202-204]. These results indicate that RAN translation is subject to some of the same regulatory mechanisms as AUG-initiated translation, though the outcomes of that regulation can differ.

Hypotheses and Addressable Questions

RAN translation represents a new and provocative mechanism by which protein translation can occur in the setting of NREs to produce a novel set of toxic proteins. The novelty of RAN translation may well prove to be its greatest value, both in revealing interesting biology and in providing a particularly good target for therapy development. **If RAN translation drives neurodegeneration, as current data supports, then factors that are selectively required for RAN translation, but not canonical transition, offer a major opportunity for therapeutic intervention.**

Nevertheless, it is important to recognize that RAN translation might be one sliver

of a greater scheme regulating protein synthesis[205,206]. Initiation in the absence of an AUG start codon is not as uncommon as once thought; data from ribosome profiling datasets suggest significant non- and near-AUG initiated translation throughout the transcriptome[207]. Thus, RAN translation may reflect physiological but non-canonical initiation processes that, in the setting of NREs, produce toxic proteins but also have valuable, adaptive, or necessary functions in other settings. Defining these normal functions and their roles in neuronal biology will be critical if RAN translation is to serve as a therapeutic target.

Finally, are the mechanisms underlying RAN translation the same or different across repeat types, reading frames, and sequence contexts? Discrepancies do suggest that different mechanisms are in play. Thus far, data on *FMR1*- and *C9ORF72*-based RAN translation are most consistent with the scanning model and use of a near-AUG codon for initiation just 5' to the repeat[19,150,169,170], but this would seem unable to explain RAN initiation within an ORF, as apparently occurs in Huntington disease[20]. And though reporters bearing expanded G_4C_2 repeats are translated most efficiently in a m^7G-dependent manner[169,170], this does not preclude a mechanism in which spliced G_4C_2-containing lariats are translated[202]. In this fashion, RAN translation may be analogous to the diversity of viral internal ribosome entry site (IRES) elements, which display significant variance in both eIF requirements and sequence to achieve the same goal of bypassing cap-dependent ribosomal loading. Only after careful identification of the key factors required for RAN translation can this delineation be made across different NREs and disease contexts.

I hypothesize that RAN translation is modulated by many of the same eIFs

that regulate canonical translation. To test this hypothesis, I have taken advantage of several models and techniques. A *Drosophila* model of FXTAS, which expresses an EGFP reporter 3' to the human *FMR1* 5' UTR bearing 90 CGG repeats, permits ready assessment of NRE-elicited phenotype, as well as quantification of FMRpolyG abundance *in vivo*. HeLa and HEK293 cells, in conjunction with transfectable reporters for canonical and *FMR1*-/*C9ORF72*-based RAN translation across repeat sizes and reading frames, enable rapid (high-throughput, even) quantification of RAN-product abundance as well as more thorough, mechanistic investigation of putative modulators of RAN translation. Automated longitudinal microscopy of primary rodent neurons expressing CGG-expanded *FMR1* provides a sensitive measure of cellular toxicity in mammalian neurons. Finally, quantification of RAN-product abundance in C9 ALS/FTD patient-derived iPSCs provides information on the physiological relevance of putative modulators.

Chapter 2

DDX3X and Specific Initiation Factors Modulate *FMR1* RAN Translation[2]

Abstract

A CGG trinucleotide repeat expansion in the 5' UTR of *FMR1* causes the neurodegenerative disorder Fragile X-associated tremor/ataxia syndrome (FXTAS). This repeat supports a non-canonical mode of protein synthesis known as repeat-associated, non-AUG (RAN) translation. The mechanism underlying RAN translation at CGG repeats remains unclear. To identify modifiers of RAN translation and potential therapeutic targets, we performed a candidate-based screen of eukaryotic initiation factors and RNA helicases using both a *Drosophila melanogaster* model of FXTAS and cell-based assays. We identified multiple modifiers of expanded $(CGG)_n$-elicited toxicity and *FMR1* RAN translation. These include the DEAD-box RNA helicase *belle*/*DDX3X*, the helicase accessory factors *EIF4B/4H*, and the start-codon selectivity factors *EIF1* and *EIF5.* Disrupting *belle*/*DDX3X* selectively inhibited *FMR1* RAN translation in *Drosophila in vivo* and cultured human cells, and mitigated repeat-induced toxicity in *Drosophila* and primary rodent neurons. These findings implicate RNA secondary structure and start-codon fidelity as critical elements regulating *FMR1* RAN translation and identify potential targets for treating repeat-associated neurodegeneration.

[2] The majority of this chapter has been published as: Linsalata, AE, He F, Malik AM, Glineburg MR, Green KM, Natla S, Flores BN, Krans A, Archbold HC, Fedak SJ, Barmada SJ, Todd PK. DDX3X and Specific Initiation Factors Modulate *FMR1* Repeat-Associated Non-AUG (RAN) Translation. *EMBO Reports* **20**, e47498 (2019).

Introduction

Over thirty different NREs cause neurodegeneration in humans[1]. NREs within consensus coding sequences (CCDS) cause disease predominantly via protein-based gain-of-function mechanisms that depend on the intrinsic toxicity of homopolymeric peptides or dysfunction of the proteins in which they reside[3-6]. Alternatively, NREs can elicit toxicity via mRNA-based mechanisms where expanded repeats sequester essential RNA-binding proteins, leading to transcriptome dysregulation (e.g., myotonic dystrophy I and II) and gelation of RNA-protein complexes into RNA containing foci[8-11]. More recently, NREs were found to support translational initiation in the absence of an AUG start codon through a process known as RAN translation[14]. Proteins generated through RAN translation accumulate in patient tissues[14,15,17-19] and are toxic in animal and cellular models of disease[14,19,145,149,173]. Since its discovery, RAN translation has been implicated in several NRE-associated neurodegenerative disorders[25], including FXTAS, C9 ALS/FTD, and Huntington's disease[20].

FXTAS is an adult-onset neurodegenerative disorder caused by a CGG NRE in the 5' UTR of *FMR1* from approximately 30 repeats to 55-200 repeats[208]. The NRE is transcribed into mRNA, which can bind to and sequester specific RNA-binding proteins [135-139,142]. In addition, expanded CGG repeats are translated via RAN translation into toxic proteins, which accumulate in ubiquitinated aggregates in tissue of both FXTAS patients and animal models of FXTAS[19,145,149]. Synthesis of RAN products is necessary for CGG repeats to elicit toxicity in overexpression systems, including *Drosophila melanogaster,* cultured human cells, and transgenic mice[19,145,209]. FXTAS shares its

causative locus with the neurodevelopmental disorder Fragile X syndrome, but it is clinically and mechanistically distinct: Fragile X syndrome results from larger (>200) CGG NREs that transcriptionally silence the Fragile X locus, resulting in loss of *FMR1* mRNA, no expression of expanded CGG repeats as RNA, and absence of FMRP[210].

The mechanism of RAN translation, and how it differs from that of canonical translation, remains unclear. Early reports demonstrated that, at least under some circumstances, RAN translation's initial stages resemble canonical translation[150,169,170]: the 43S PIC—composed of the 40S ribosomal subunit, $tRNA_i^{Met}$, and a number of essential eIFs—binds to the 5' m^7G cap on mRNA and scans through the 5' UTR[27]. In canonical translation, the 43S PIC scans until it encounters an AUG start codon, which triggers a cascade of structural rearrangements that ends with binding of the 60S ribosomal subunit and initiation of translation. In RAN translation, initiation occurs at non-AUG codons, either upstream of or within the NRE[14,150,169,170,203]. At CGG repeats and possibly G_4C_2 repeats, this failure of codon fidelity is thought to result from impairment of 43S PIC scanning by stable RNA secondary structures formed by the expanded repeats, since such structures facilitate initiation at non-AUG sites[75,95,111,211]. At other repeats and cellular contexts, RAN translation may utilize cap-independent initiation mechanisms and/or initiator tRNAs other than $tRNA_i^{Met}$ [14,202,212]. Which mechanisms occur in the context of each human disease is unclear and could vary based on gene context, repeat content, and cell type[25,213,214].

Discerning how RAN translation initiates, and how that process diverges from canonical translation, might reveal new therapeutic strategies for FXTAS, C9 ALS/FTD, and other NRE-associated disorders. Two features distinguish RAN translation from

canonical translation: the presence of highly stable RNA secondary structures composed of NREs[109,110,160] and the use of non-AUG start codons. 43S PIC scanning is known to require several RNA helicases in order to resolve mRNA structure within 5' UTRs, including Ded1/bel/DDX3X[46,50,58,64,66-68], eIF4A and its cofactors eIF4B and eIF4H[40,44-46], and DHX29[41,215]. In addition, start codon fidelity in yeast is regulated by a series of eIFs—including eIF1, eIF1A, eIF2, and eIF5—and upstream signaling pathways[36,76,78-81,200,201].

With these features in mind, we conducted a candidate-based screen of eIFs, RNA helicases, and other RNA-binding proteins to identify regulators of RAN translation, using both cell-based assays and a *Drosophila melanogaster* model of FXTAS[216]. In FXTAS, RAN translation from the sense strand of *FMR1* yields at least two RAN products from the CGG repeat: a polyglycine peptide (FMRpolyG), reflecting initiation in the GGC (+1, relative to the reading frame of the downstream CCDS product, FMRP) frame, and a polyalanine peptide (FMRpolyA), reflecting initiation in the GCG (+2) frame[19,150]. Antisense *FMR1* transcripts also support RAN translation, yielding three distinct RAN products[21]. In *Drosophila*, we previously demonstrated that CGG repeat-elicited toxicity is largely dependent on RAN translation[19,147]. Our screen identified multiple factors that, when disrupted, suppressed CGG NRE-elicited toxicity in *Drosophila* and were necessary for *FMR1* RAN translation. Disruption of one in particular, the DEAD-box RNA helicase *DDX3X* (*bel* in *Drosophila*), selectively inhibited *FMR1* RAN translation in human cell-based systems and suppressed repeat-induced neurodegeneration in rodent neurons. Our findings implicate RNA secondary structure and start-codon fidelity in *FMR1* RAN translation and suggest specific targets for future

therapeutic development.

Results

A Screen for Modifiers of CGG-Repeat Toxicity in *Drosophila Melanogaster*

To identify regulators of *FMR1* RAN translation, we conducted a candidate-based screen using a *Drosophila melanogaster* model of FXTAS (Figure 2.1)[216]. This model carries an EGFP transgene 3' to an upstream activation sequence (UAS) and the 5' UTR of human *FMR1* with 90 CGG repeats, with EGFP in the GGC (+1, FMRpolyG) reading frame [$(CGG)_{90}$-EGFP)]. Expression in the eye via a GMR-GAL4 driver manifests in a significant rough-eye phenotype observable at eclosion, with ubiquitin-positive aggregates of the RAN product FMRpolyG-EGFP accumulating in retinal neurons[19]. For the screen, females expressing $(CGG)_{90}$-EGFP under a GMR-GAL4 driver were crossed to males carrying germline mutations in, UAS-driven transgenes of, and UAS-driven small hairpin RNA (shRNA) constructs targeting 10 canonical eIFs, 4 RNA helicases, a ribosomal protein associated with non-canonical translation initiation[217,218], and an RNA-binding protein implicated in ALS/FTD[219] (Table A.1). We selected these candidates in a hypothesis-driven fashion based on their known functions in non-canonical translation initiation and regulation of start-codon fidelity, as well their potential to modulate GC-rich secondary structures in RNA. By design, this candidate list was non-exhaustive. Because previous work has demonstrated the importance of eIF4E-m^7G binding and 43S PIC scanning to RAN translation[150,169,170] and because eIF4E and eIF4G are necessary for canonical translation, we did not

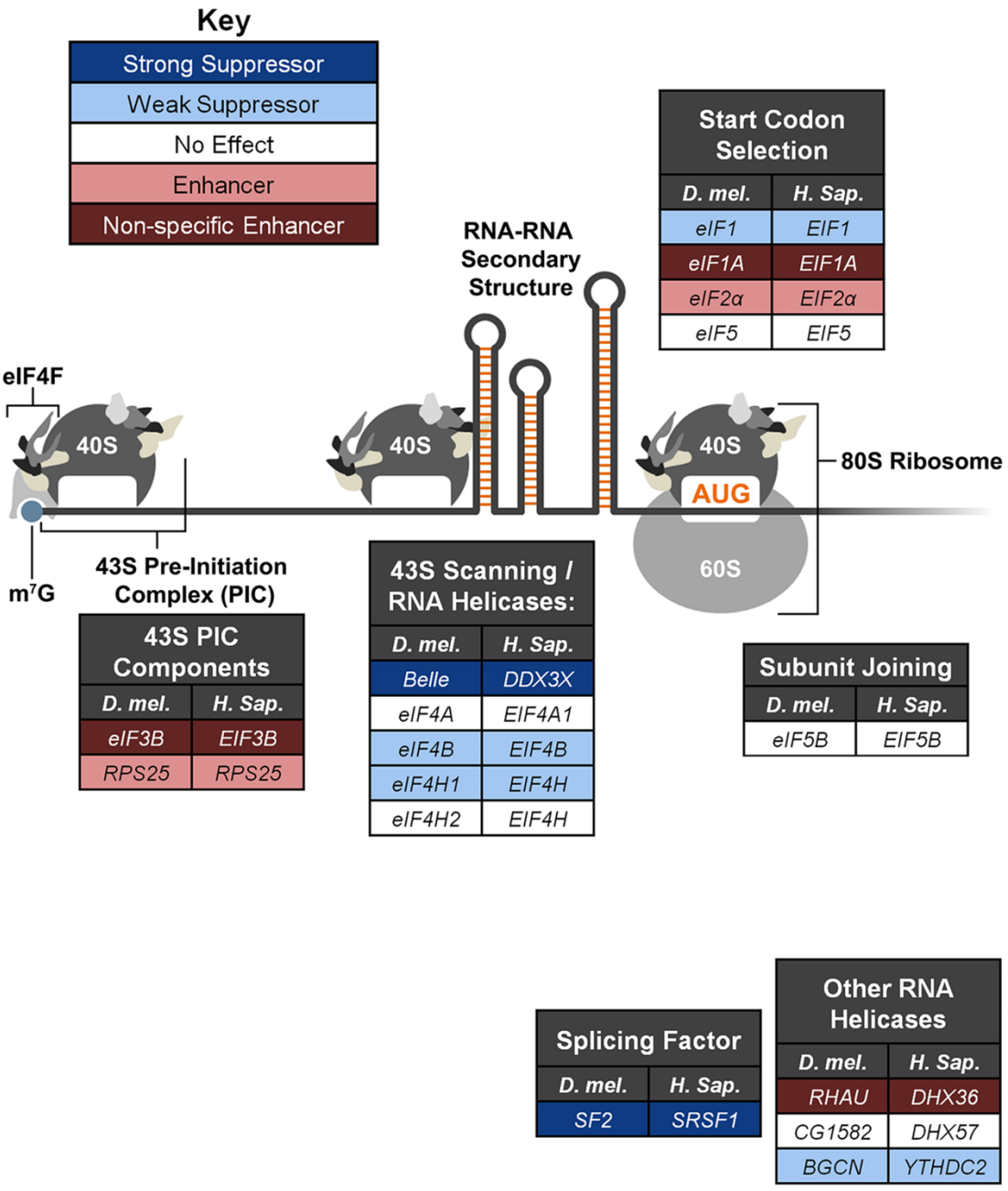


Figure 2.1: A Candidate-Based Screen Identifies Modifiers of Repeat-Elicited Toxicity in *Drosophila Melanogaster*.
Candidate modifiers are categorized here based on their known functions in gene expression. Fly genes are listed in the left columns, while their human homologs are listed in the right columns. Disruption of genes highlighted in dark blue strongly suppressed $(CGG)_{90}$-EGFP toxicity. Disruption of genes highlighted in light blue weakly suppressed $(CGG)_{90}$-EGFP toxicity. Disruption of genes highlighted in light red enhanced $(CGG)_{90}$-EGFP toxicity selectively. Disruption of genes highlighted in dark red enhanced the toxicity of both $(CGG)_{90}$-EGFP and GMR-GAL4 (these were toxic independent of the repeat.) All other genes are displayed in white. The m^7G cap, CGG nucleotide-repeat expansion (NRE), eIF4F complex, 43S PIC, and ribosomal subunits are indicated.

evaluate them in this screen.

We examined the rough-eye phenotype of F_1 flies at eclosion to identify suppressors and enhancers of $(CGG)_{90}$-driven toxicity. Of the 57 candidate lines tested, 21 acted as suppressors of toxicity, and 17 as enhancers. All lines were subsequently crossed to flies carrying GMR-GAL4 alone to control for toxic effects independent of $(CGG)_{90}$-based toxicity (Table A.1). Six of the 17 toxicity-enhancing lines had no effect in the absence of $(CGG)_{90}$-EGFP, suggesting that this enhancement is specific to the presence of expanded CGG repeats. We selected 3 suppressors for further analysis—*bel*/*DDX3X*, *eIF4B*, and *eIF4H1*—based on their individual functions in translation initiation.

Bel/DDX3X Selectively Modulates *FMR1* RAN Translation in *Drosophila*

In our candidate-based *Drosophila* screen, *bel* disruption by multiple genetic means suppressed CGG_{100}-elicited toxicity. Four shRNAs against *bel* and five heterozygous loss-of-function *bel* mutants significantly suppressed the rough-eye phenotype in $(CGG)_{90}$-EGFP-coexpressing flies (Figure 2.2A, B; Figure A.1A, B). *Bel* mutants tested included nonsense mutations and P[lacW] and P[PZ] element insertions[220-223]. The *bel* shRNA lines generally suppressed $(CGG)_{90}$-EGFP toxicity more effectively than the heterozygous, loss-of-function *bel* mutants, potentially because the shRNAs had stronger effects on the abundance of functional bel protein. None of the *bel* shRNAs had phenotypic effects in flies expressing an AUG-initiated EGFP transgene under a GMR-GAL4 driver (Figure A.1C, D). Modulation of $(CGG)_{90}$-driven toxicity was not limited to the eye, since three *bel* shRNAs increased the lifespan

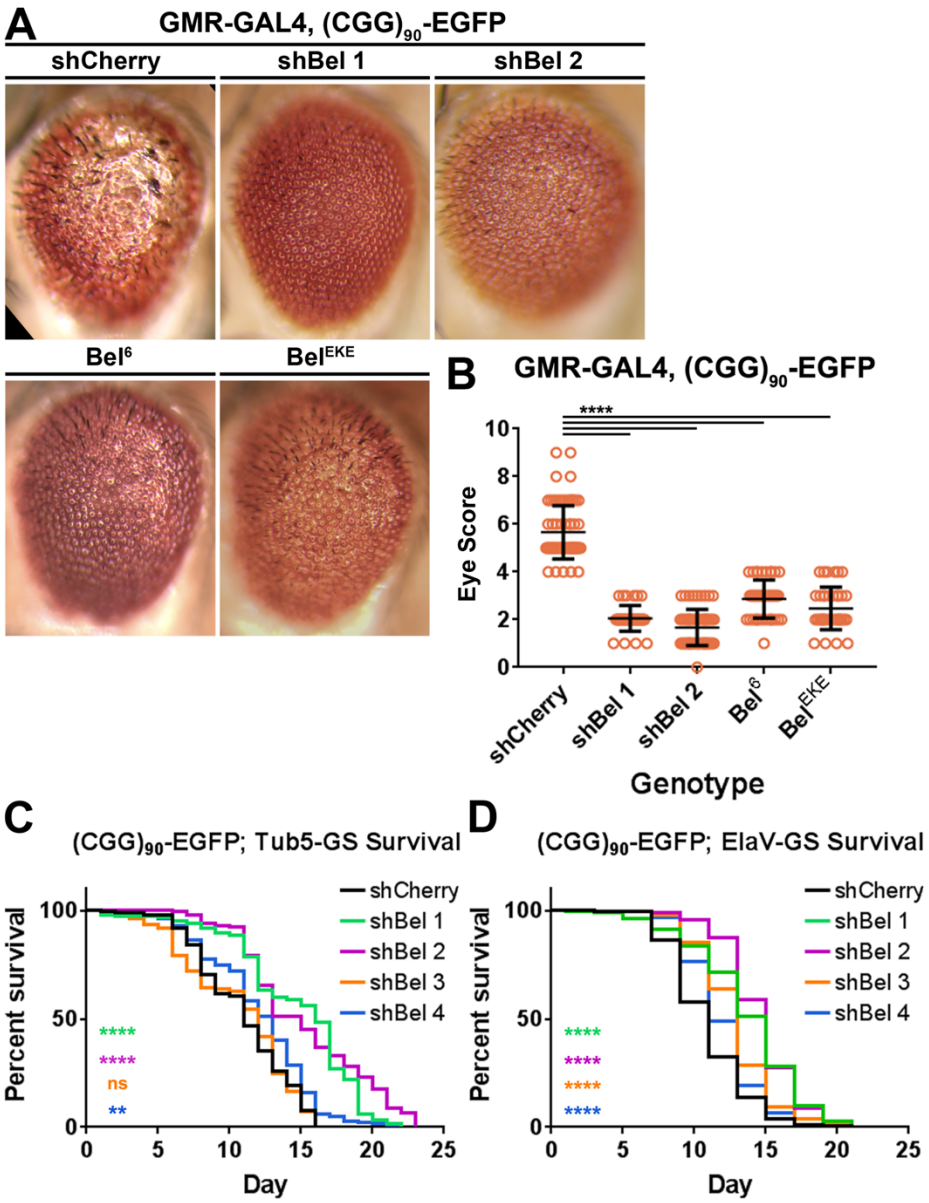


Figure 2.2: Knockdown of *Belle* (*Bel*) Mitigates Repeat-Elicited Toxicity in *Drosophila*.
A Representative photographs of fly eyes expressing $(CGG)_{90}$-EGFP under a GMR-GAL4 driver along with various *bel* disruptions.
B Quantitation of GMR-GAL4, $(CGG)_{90}$-EGFP eye phenotypes with *bel* disruptions (Mann-Whitney U test with Bonferonni corrections for multiple comparisons; *n*=35-77/genotype).
C, D Longevity assays of $(CGG)_{90}$-EGFP; Tub5-GS (C; Log-rank Mantel-Cox test with Bonferroni corrections for multiple comparisons; *n*=110-219/genotype) and $(CGG)_{90}$-EGFP; ElaV-GS (D; *n*=147-299/genotype) flies with *bel* knockdown.

Data Information: For all panels, ns=non-significant, ** P≤0.01, **** P≤0.0001 for the specified statistical test. All data in all panels are presented as mean ± SD (compiled from ≥3 replicates).

of adult flies expressing $(CGG)_{90}$-EGFP ubiquitously post-eclosion under an inducible Tub5 Geneswitch driver[224] (Figure 2.2C). Similarly, four *bel* shRNAs increased lifespan when $(CGG)_{90}$-EGFP was expressed pan-neuronally in adult flies under an inducible Geneswitch ElaV driver (Figure 2.2D).

Bel and its homologs in yeast (*Ded1*) and humans (*DDX3X*) are important for translation of specific mRNAs, particularly those with long or structured 5' UTRs[46,50,58,59,64,66-68]. Given the role that secondary structure is hypothesized to play in the initiation of RAN translation, we asked whether knockdown of *bel* suppressed the $(CGG)_{90}$-EGFP phenotype by suppressing RAN translation. Knockdown of *bel* by two independent shRNAs reduced the expression of the RAN product FMRpolyG-EGFP in $(CGG)_{90}$-expressing flies (Figure 2.3A, B), supporting the hypothesis that suppression of $(CGG)_{90}$ toxicity is driven by inhibition of RAN translation. In contrast, *bel* knockdown had no effect on the abundance of $(CGG)_{90}$-EGFP transcripts (Figure 2.3C). Finally, knockdown of *bel* had no effect on expression of an AUG-initiated EGFP reporter lacking the *FMR1* 5' UTR (Figure 2.3D, E), suggesting that the decrease in FMRpolyG-EGFP we observed reflects a selective effect on RAN translation rather than a global decline in translation.

<u>*Bel/DDX3X* Selectively Modulates *FMR1* RAN Translation in Human Cells</u>

We next asked whether *DDX3X*, the human homolog of *bel*, might play a similar role in facilitating RAN translation of CGG repeats in human cells. We previously generated transfectable luciferase-based reporters consisting of a 3xFLAG-tagged nanoluciferase (NL-3xF) downstream of the 5'UTR of human *FMR1*, with multiple repeat

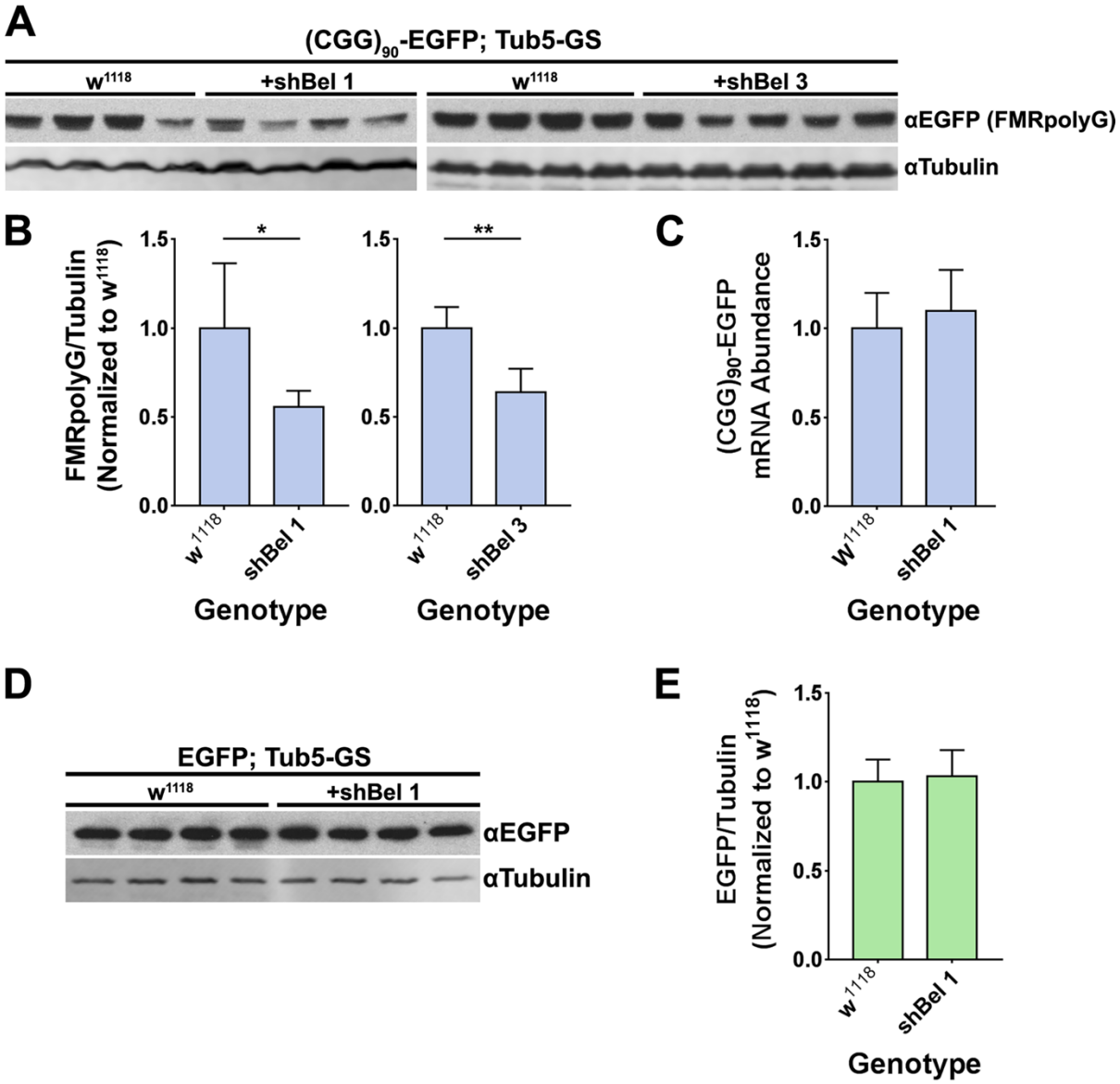


Figure 2.3: Knockdown of *Bel* Specifically Inhibits RAN Translation in *Drosophila*.

A Western blots of the FMRpolyG-EGFP RAN product from $(CGG)_{90}$-EGFP; Tub5-GS fly tissue with and without *bel* knockdown by two independent small hairpin RNAs (shRNAs).

B Quantitation of FMRpolyG-EGFP band density, normalized to β tubulin band density, from blots in A (Student's t test; *n*=4-5/genotype).

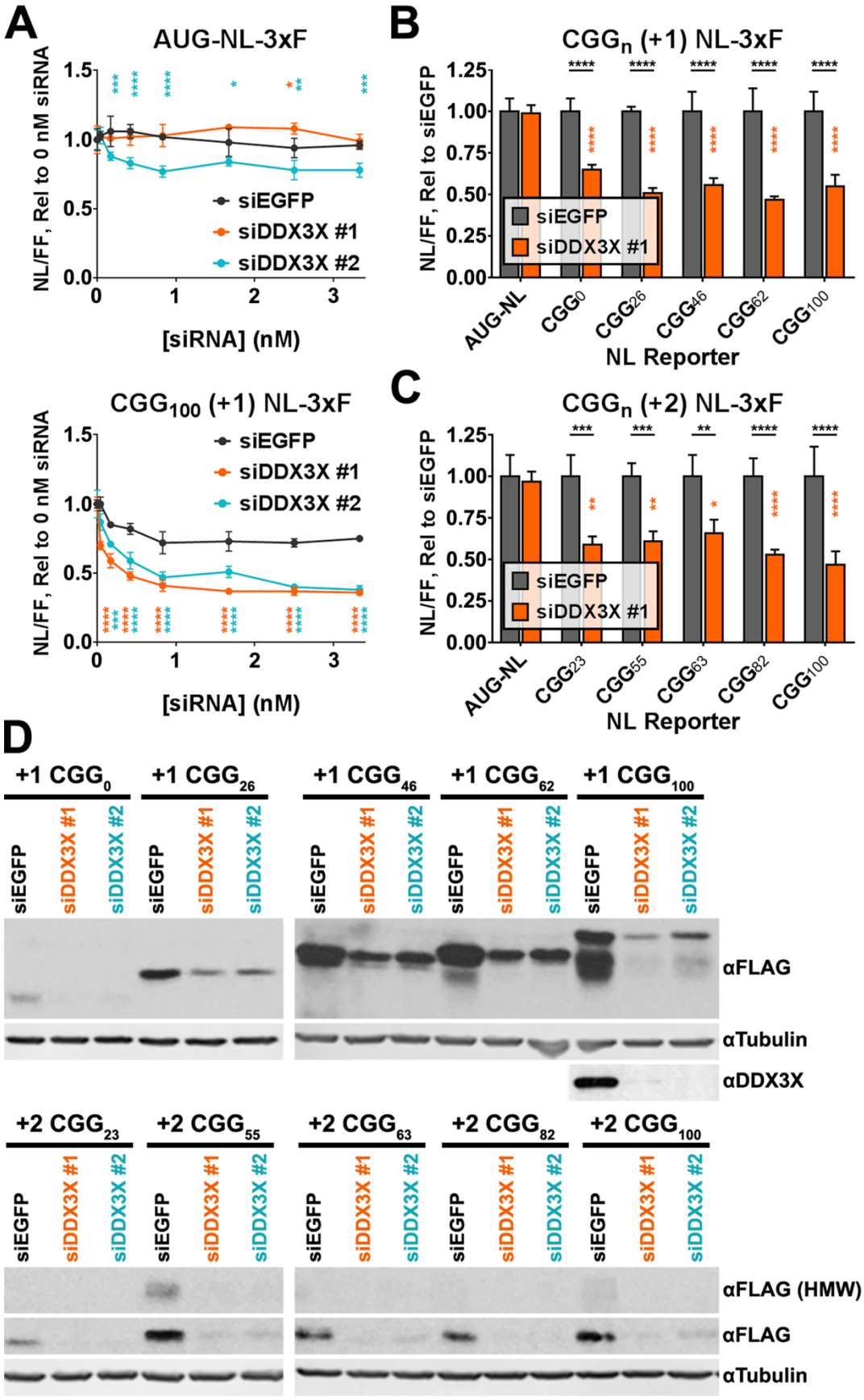
C Abundance by qRT-PCR of $(CGG)_{90}$-EGFP mRNA normalized to *RPL32* mRNA from $(CGG)_{90}$-EGFP; Tub5-GS fly tissue, following *bel* knockdown (*n*=8/genotype).

D Western blot of AUG-driven EGFP from EGFP; Tub5-GS fly tissue with and without *bel* knockdown.

E Quantitation of EGFP band density, normalized to β tubulin band density, from blot in D *(n*=4/genotype).

Data Information: For all panels, * P≤0.05, ** P≤0.01 for the specified statistical test. All data in all panels are presented as mean + SD (compiled from ≥3 replicates).

sizes (0-100 repeats) and with the NL-3xF in both the GGC (+1, FMRpolyG) and GCG (+2, FMRpolyA) reading frames (Figure A.2)[150]. These reporters enable quantitative and qualitative detection of RAN-product expression by luminescence assays and western blotting, respectively. Knockdown of *DDX3X* by five independent small interfering RNAs (siRNAs) reduced the expression of plasmid-based +1 $(CGG)_{100}$ NL-3xF reporters in a dose-dependent manner (Figure 2.4A; Figure A.3). As in our *Drosophila* experiments, these siRNAs had minimal effect on the expression of an AUG-initiated NL-3xF reporter (AUG-NL-3xF). To further test whether *DDX3X* knockdown inhibits protein synthesis across mRNAs, we tested the effects of two *DDX3X* siRNAs on NL-3xF reporters bearing the short, minimally-structured 5' UTRs of β actin (*ACTB*) and β globin (*HBB*). siDDX3X #1 had no effect on Actin-NL-3xF and increased the expression of Globin-NL-3xF but decreased the expression of +1 $(CGG)_{100}$ NL-3xF (Figure A.4A). Though siDDX3X #2 decreased both Actin- and Globin-NL-3xF, it did so significantly less than it inhibited +1 $(CGG)_{100}$ NL-3xF. Finally, to assess the effects of *DDX3X* knockdown on global protein synthesis of endogenous mRNAs, we performed polysome fractionation on HeLa cells transfected with siDDX3X or siEGFP (Figure A.5A, B). Consistent with our NL-3xF reporter data, knockdown of *DDX3X* (Figure A.5C) did not result in a reproducible shift in the relative monosome and polysome fractions. This indicates that *DDX3X* knockdown does not lead to global inhibition of mRNA translation, again highlighting the role of DDX3X in the expression of select genes. This finding is consistent with previous work demonstrating that the expression of only a specific subset of mRNAs—those with long and/or secondary-structured 5' UTRs—are reduced following *DDX3X*/*Ded1* disruption[46,59,64,72].


A
AUG-NL-3xF
NL/FF, Rel to 0 nM siRNA
[siRNA] (nM)
siEGFP
siDDX3X #1
siDDX3X #2
CGG100 (+1) NL-3xF
B
CGGn (+1) NL-3xF
NL Reporter
AUG-NL
CGG0
CGG26
CGG46
CGG62
CGG100
C
CGGn (+2) NL-3xF
CGG23
CGG55
CGG63
CGG82
D
+1 CGG0
+1 CGG26
+1 CGG46
+1 CGG62
+1 CGG100
+2 CGG23
+2 CGG55
+2 CGG63
+2 CGG82
+2 CGG100
αFLAG
αTubulin
αDDX3X
αFLAG (HMW)

We next asked whether DDX3X regulates RAN translation of *FMR1* transcripts in other reading frames. Expression of the +2 $(CGG)_{100}$ NL-3xF RAN product ($FMRpolyA_{100}$), which likely derives from initiation within the NRE[150], was reduced to a similar degree as the $FMRpolyG_{100}$ product in luminescence assays (Figure 2.4B, C; Figure A.4B, C). We also observed these effects following detection of the $FMRpolyG_n$ and $FMRpolyA_n$ RAN products by western blotting (Figure 2.4D). These data indicate that the function of DDX3X in promoting RAN translation is not limited to a single reading frame.

DDX3X and its homologs in yeast and *Drosophila* function in several aspects of RNA metabolism. We therefore asked whether DDX3X functions in RAN translation directly or whether its effects might instead be mediated by modulating RNA stability or transcription. Knockdown of *DDX3X* reduced expression of +1 and +2 $(CGG)_{100}$ NL-3xF, regardless of whether the reporters were transfected as plasmids or *in-vitro* transcribed

---

Figure 2.4: Knockdown of *DDX3X* Inhibits RAN Translation in Cultured Human Cells.

A Dose-response curves plotting the effects of two independent anti-*DDX3X* small interfering RNAs (siRNAs) on the expression of AUG-NL-3xF (top) and $(CGG)_{100}$ +1 NL-3xF (bottom) reporters. Plasmid-based reporters were transfected into HeLa cells 24 hours after knockdown, and reporter expression was quantified by luminescence. Nanoluciferase-3xFLAG (NL-3xF) luminescence has been normalized to luminescence from firefly luciferase (FF), which was co-transfected, in order to control for transfection variability. Asterisks refer to comparisons between anti-*DDX3X* siRNAs and siRNAs against EGFP (siEGFP; two-way ANOVA with Dunnett's multiple comparisons test; *n*=12/condition).

B, C $(CGG)_n$ +1 and $(CGG)_n$ +2 NL-3xF expression (normalized to FF) with and without *DDX3X* knockdown across a range of CGG repeat sizes. Black asterisks refer to comparisons between siDDX3X- and siEGFP-treated cells; orange asterisks refer to comparisons between siDDX3X-treated cells expressing AUG-NL-3xF and those expressing a different reporter (two-way ANOVA with Tukey's multiple comparisons test; *n*=17-30/condition).

D Western blots of FMRpolyG-NL-3xF and FMRpolyA-NL-3xF products with and without *DDX3X* knockdown across a range of repeat sizes.

Data Information: For all panels, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001 for the specified statistical test. All panels depict data as means ± SD (compiled from ≥3 replicates).

RNAs (Figure 2.5A; Figure A.4D, E), suggesting that DDX3X acts post-transcriptionally. We next asked whether *DDX3X* knockdown impacted the abundance of reporter mRNAs following plasmid transfection. We observed that the mRNA abundance of both AUG-NL-3xF and $(CGG)_{100}$-NL-3xF in the +1 and +2 frames was not consistently affected after transfection with siDDX3X #1 and #2 (Figure 2.5B). In order to isolate translational effects *per se* from other, concurrent effects on gene expression, we conducted *in vitro* translation assays using transcribed reporter RNAs and cytoplasmic extracts generated from cells depleted of DDX3X by two independent siRNAs. DDX3X-depleted extracts yielded reduced translation of +1 $(CGG)_{100}$ NL-3xF, while having either no effect on or increasing the synthesis of an AUG-NL-3xF reporter (Figure 2.5C). This effect was consistent across independently prepared *in vitro* translation extracts replicates (4 extracts per siRNA, two anti-*DDX3X* siRNAs; Figure A.6A, B). These *in vitro* experiments point specifically to a direct translational function of DDX3X, while leaving open the possibility that changes in mRNA stability and abundance may further impact the expression of RAN products.

We next asked how DDX3X regulates *FMR1* RAN translation. To determine whether DDX3X can directly interact with CGG reporter mRNAs, we performed photo-crosslinking RNA immunoprecipitation (RIP) assays to probe for an interaction between DDX3X and our *FMR1* $(CGG)_{100}$ reporters in cultured cells. As we anticipated, significantly more *HSPA1A* mRNA (translation of which requires DDX3X[68]) was co-purified using antibodies against DDX3X than using isotype control IgG (Figure 2.5D). Similarly, significantly more +1 $(CGG)_{100}$ NL-3xF mRNA was co-purified using antibodies against DDX3X than antibodies against EGFP or isotype control IgG (Figure

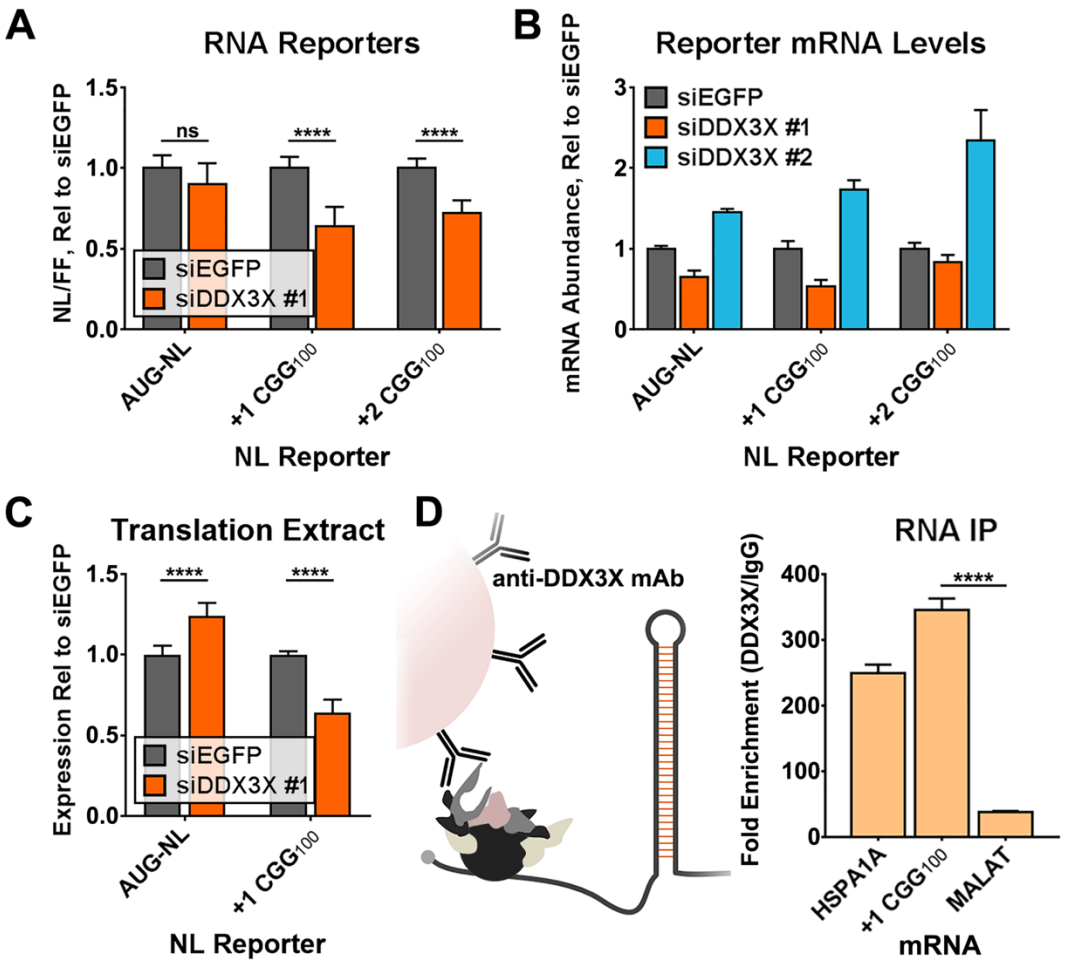


Figure 2.5: DDX3X Facilitates Expression of RAN Products at the Level of Translation.
A Expression of *in vitro* transcribed AUG-, +1 $(CGG)_{100}$-, and +2 $(CGG)_{100}$-NL-3xF RNAs following *DDX3X* knockdown in HeLa cells, expressed as NL luminescence normalized to FF luminescence (Student's t test with Bonferroni corrections for multiple comparisons; *n*=21/condition).
B Abundance of reporter mRNAs following *DDX3X* knockdown and plasmid-reporter transfection, determined by qRT-PCR (*n*=7/condition). This panel depicts data as means + SEM.
C Expression of AUG-NL-3xF and +1 $(CGG)_{100}$ NL-3xF in *in vitro* translation extracts, collected from HeLa cells treated with siRNAs against EGFP or *DDX3X* (two-way ANOVA with Tukey's multiple comparisons test; *n*=4/condition). Experiments with independent, replicate lysates are presented in Figure A.6A, B.
D Enrichment of *HSPA1A*, +1 $(CGG)_{100}$ NL-3xF, and *MALAT* RNA following anti-DDX3X RNA immunoprecipitation (RIP), relative to incubation with isotype control IgG (Student's t test, *n*=3). Data from the additional replicate is presented in Figure A.7A.

Data Information: For all panels, ns=non-significant, **** P≤0.0001 for the specified statistical test. All panels depict data as means + SD, unless indicated otherwise (compiled from ≥3 replicates).

2.5D; Figure A.7A). +1 $(CGG)_{100}$ NL-3xF mRNA was also enriched by DDX3X RIP in comparison to endogenous *MALAT* RNA, indicating a transcript-selective interaction between DDX3X and +1 $(CGG)_{100}$ NL-3xF mRNA. To determine whether that interaction is an artifact of the NL-3xF tag on +1 $(CGG)_{100}$ NL-3xF mRNA, we repeated the experiment using a 3'-truncated "tagless" construct in which the NL-3xF tag had been deleted (leaving the *FMR1* 5' UTR and minimal vector sequence intact). In two independent replicates, tagless $(CGG)_{100}$ mRNA also co-precipitated with DDX3X, near or above the levels of *HSPA1A* mRNA as a positive control (Figure A.7B). Finally, to determine whether that interaction depends on the expanded CGG repeats in the tagless $(CGG)_{100}$ construct, we repeated this experiment using a modified tagless construct bearing 0 CGG repeats [tagless $(CGG)_0$]. $(CGG)_0$ mRNAs co-precipitated with DDX3X at levels comparable to $(CGG)_{100}$ and *HSPA1A* mRNA, indicating that expanded CGG repeats are unnecessary for this interaction. This is not surprising, however, as the 5'UTR of *FMR1* is highly GC-rich (76%) even excluding the CGG NRE, and previously published work has demonstrated that DDX3X's yeast homolog Ded1 preferentially binds 5' to secondary structures within 5' UTRs[59].

To determine which features of the *FMR1* 5'UTR enable DDX3X to modulate RAN translation, we first varied the size of the CGG NRE in our reporters. Using plasmid-based reporters, we observed no significant effect of CGG repeat size on the impact of *DDX3X* knockdown in either the GGC (+1) frame or the GCG (+2) frame (Figure 2.4B-D; Figure A.4B, C). All exhibited decreased expression as measured by luminescence assays and western blotting. We observed similar results when we transfected reporters as *in vitro*-transcribed RNA (Figure A.4D, E). These results

indicate that expanded CGG repeats are unnecessary for DDX3X to modulate RAN translation initiating within the *FMR1* 5' UTR.

One of the unique features of RAN translation is its use of non-AUG codons for initiation. We therefore asked whether DDX3X facilitated RAN translation by regulating start-codon selection. We first tested the effect of *DDX3X* knockdown on a modified +1 $(CGG)_{100}$ reporter, in which the major near-AUG codon (ACG) utilized for GGC (+1)-frame RAN translation had been removed and replaced with a nearby AUG (Figure A.8)[150]. This change enhanced basal expression of the construct, but *DDX3X* knockdown still impaired AUG-$(CGG)_{100}$ (+1) NL-3xF expression (Figure 2.6A). Similarly, insertion of an AUG codon in a strong Kozak context 5' to the NRE in the GCG (+2) frame enhanced basal expression, but expression of this AUG-$(CGG)_{100}$ (+2) NL-3xF reporter remained *DDX3X*-dependent. In a complementary experiment, we evaluated whether *DDX3X* knockdown affected translation of NL-3xF reporters that initiate with the near-cognate codons CUG, GUG, UUG, or ACG absent any NRE or *FMR1* sequence. As expected, expression of CUG-, GUG-, UUG-, and ACG-NL-3xF reporter plasmids was lower than expression of AUG-NL-3xF[95,169]. When these plasmids were transfected into cells along with siRNAs, *DDX3X* knockdown reduced their expression compared to AUG-NL-3xF but significantly less than +1 $(CGG)_{100}$ NL-3xF (Figure 2.6B). In contrast, expression of *in vitro*-transcribed near-AUG-initiated NL-3xF reporters were unaffected by DDX3X depletion in *in vitro* translation assays (Figure 2.6C; Figure A.6C). In total, our results indicate that altered start-codon fidelity is unlikely to be the sole factor mediating the effects *of DDX3X* knockdown on *FMR1* RAN translation.

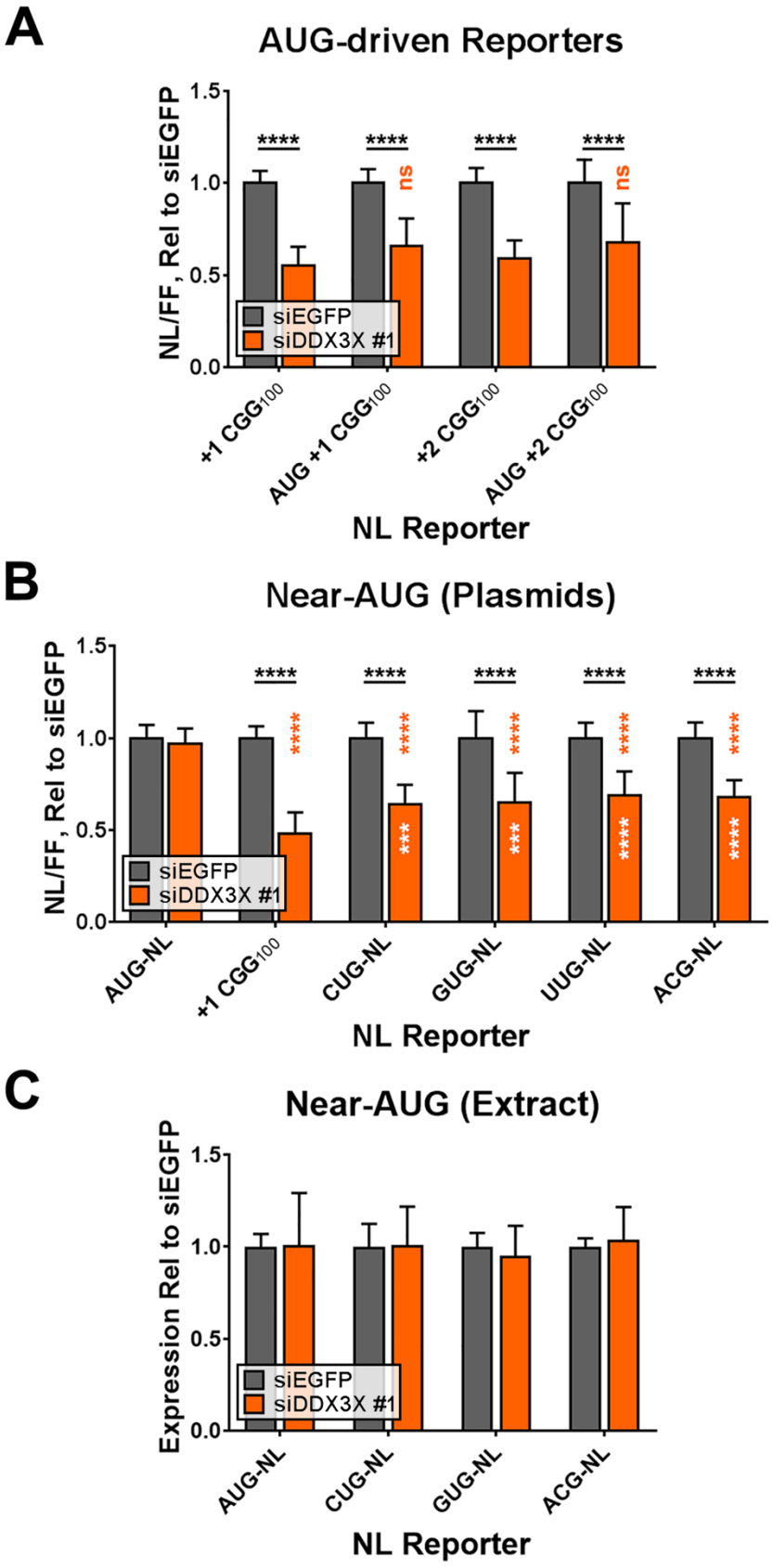

A
AUG-driven Reporters
NL/FF, Rel to siEGFP
1.5
1.0
0.5
0.0
****
****
****
****
ns
ns
siEGFP
siDDX3X #1
+1 CGG100
AUG +1 CGG100
+2 CGG100
AUG +2 CGG100
NL Reporter
B
Near-AUG (Plasmids)
NL/FF, Rel to siEGFP
1.5
1.0
0.5
0.0
****
****
****
****
****
****
****
****
****
***
***
****
****
siEGFP
siDDX3X #1
AUG-NL
+1 CGG100
CUG-NL
GUG-NL
UUG-NL
ACG-NL
NL Reporter
C
Near-AUG (Extract)
Expression Rel to siEGFP
1.5
1.0
0.5
0.0
siEGFP
siDDX3X #1
AUG-NL
CUG-NL
GUG-NL
ACG-NL
NL Reporter

Figure 2.6: In Cultured Human Cells, Near-AUG Start Codons Are Sufficient but Not Necessary to Confer Dependency on DDX3X.

A Expression of +1 and +2 $(CGG)_{100}$ NL-3xF plasmid reporters with and without an AUG inserted 5' to the CGG repeat, with and without *DDX3X* knockdown. Black asterisks refer to comparisons between siDDX3X- and siEGFP-treated cells; orange asterisks refer to comparisons between siDDX3X-treated cells expressing either +1 or +2 $(CGG)_{100}$ NL-3xF and those expressing the respective AUG-driven variant (two-way ANOVA with Tukey's multiple comparisons test; *n*=11-12/condition).

B Expression of NL-3xF plasmids with initiator AUG codons mutated to near-AUG codons, with and without *DDX3X* knockdown (two-way ANOVA with Dunnett's multiple comparisons test; *n*=18-24/condition). Black asterisks refer to comparisons between siEGFP-treated and siDDX3X-treated cells; orange asterisks refer to comparisons between siDDX3X-treated cells expressing AUG-NL-3xF and those expressing a different reporter; white asterisks refer to comparisons between siDDX3X-treated cells expressing +1 $(CGG)_{100}$ NL-3xF and those expressing a different reporter.

C Expression of *in vitro* transcribed near-AUG reporter RNAs in *in vitro* translation extracts, collected from HeLa cells treated with siRNAs against EGFP or *DDX3X* (*n*=4/group). Results from independent, replicate lysates are presented in Figure A.6C.

Data Information: For all panels, ns=non-significant, *** P≤0.001, **** P≤0.0001 for the specified statistical test. All panels depict data as means + SD, unless indicated otherwise (compiled from ≥3 replicates).

---

eIF4B and eIF4H Modulate RAN Translation at CGG Repeats

eIF4B and eIF4H are co-stimulatory factors for the RNA helicase eIF4A, and like bel/DDX3X, they are required for translation of mRNAs with long or structured 5' UTRs[40,44-46]. Previous work demonstrated that eIF4A is specifically required for RAN translation at both CGG repeats and at G_4C_2 repeats associated with ALS/FTD[150,169,170]. We were therefore interested to see whether these co-stimulatory factors might play a similar, specific function in RAN translation like DDX3X. In our initial *Drosophila* screen, three shRNAs against *eIF4H1*, one shRNA against *eIF4H2*, and one shRNA against *eIF4B* suppressed the rough-eye phenotype induced by $(CGG)_{90}$-EGFP (Figure 2.7A, B; Figure A.9A, B). In addition, *eIF4B* overexpression exacerbated this phenotype, while having no effect in the absence of $(CGG)_{90}$-EGFP (Figure A.9C, D). As with *bel* shRNAs, these effects were not limited to the eye, as *eIF4B* shRNA increased and

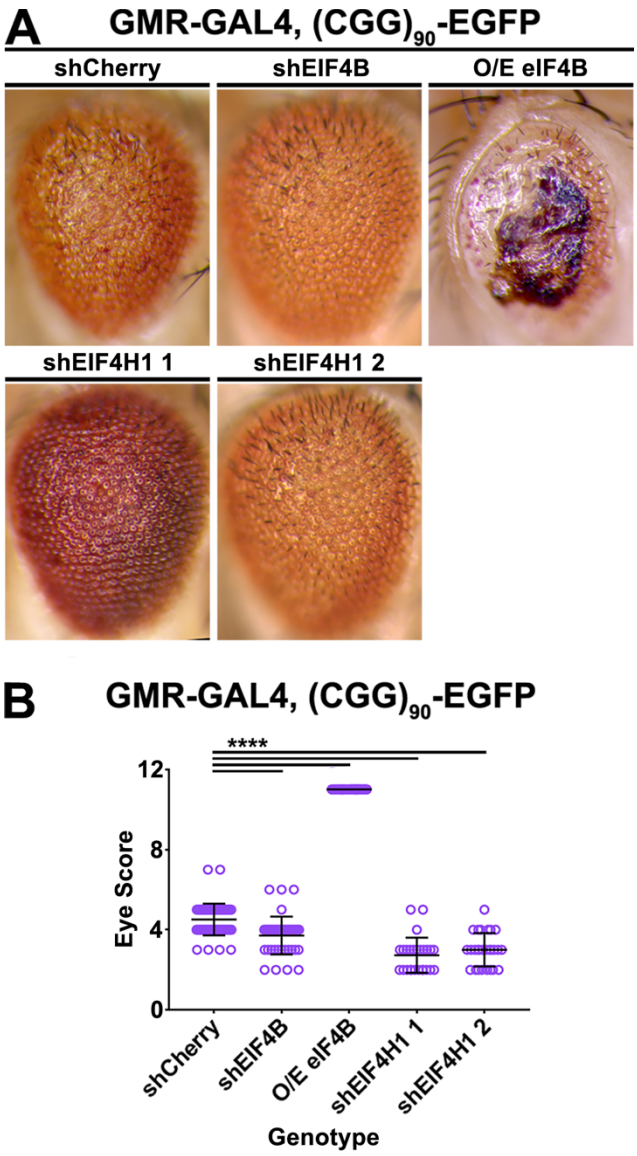


Figure 2.7: *eIF4B* and *eIF4H* Modulate NRE-Elicited Toxicity in *Drosophila*.
A Representative photographs of GMR-GAL4, $(CGG)_{90}$-EGFP fly eyes expressing genetic disruptions of *eIF4B* and *eIF4H1*.
B Quantitation of GMR-GAL4, $(CGG)_{90}$-EGFP eye phenotypes with *eIF4B* and *H1* disruptions (Mann-Whitney U test with Bonferroni corrections for multiple comparisons; *n*=26-55/genotype).

Data Information: **** P≤0.0001 for the specified statistical test. All panels present data as means ± SD (compiled from ≥3 replicates).

---

overexpression decreased the lifespan of flies expressing $(CGG)_{90}$-EGFP under an inducible Tub5 driver (Figure A.9E). These experiments suggest that, like bel/DDX3X, eIF4B and eIF4H are capable of modulating RAN translation.

In contrast to bel/DDX3X, the impact of modulating eIF4B or eIF4H expression was not specific to RAN translation. In cultured HeLa cells, knockdown of either *EIF4B* or *EIF4H* (Figure A.9F) similarly decreased expression of +1 CGG_{100} NL-3xF and a co-transfected AUG-driven firefly luciferase (AUG-FF) reporter (Figure 2.8A, B), suggesting that we cannot separate the role of eIF4B/H in RAN translation from their functions in translation generally. In support of this interpretation, overexpression of *EIF4B* alone, *EIF4H* alone, or *EIF4B* and *EIF4H* together, significantly increased the expression of AUG-NL-3xF (Figure 2.8C). These data suggest that eIF4B and eIF4H both regulate RAN translation but do so in a manner that is not specific to transcripts that are RAN translated.

<u>eIF1 and eIF5 Modulate RAN Translation in Human Cells Via Start-Codon Selectivity</u>

By definition, RAN translation follows a failure in start-codon fidelity. Therefore, we predict that eIFs that enhance start-codon fidelity suppress RAN translation, while eIFs that reduce start-codon fidelity enhance RAN translation. We and others have pursued this concept in the context of the integrated stress response[169,170,202]: cellular stressors trigger phosphorylation of eIF2α, which suppresses global protein translation by reducing ternary complex (eIF2α-GTP-$tRNA_i^{Met}$) recycling and availability. Cellular stressors and mutations in multiple eIF2 subunits have both been shown to enhance initiation at non-AUG codons in yeast and mammalian cells[225,226]. Consistent with this, we and others observed that exogenous stressors or eIF2α phosphorylation selectively enhance RAN translation at both CGG repeats and G_4C_2 repeats associated with ALS[169,170,202] in a near-AUG codon-dependent manner.

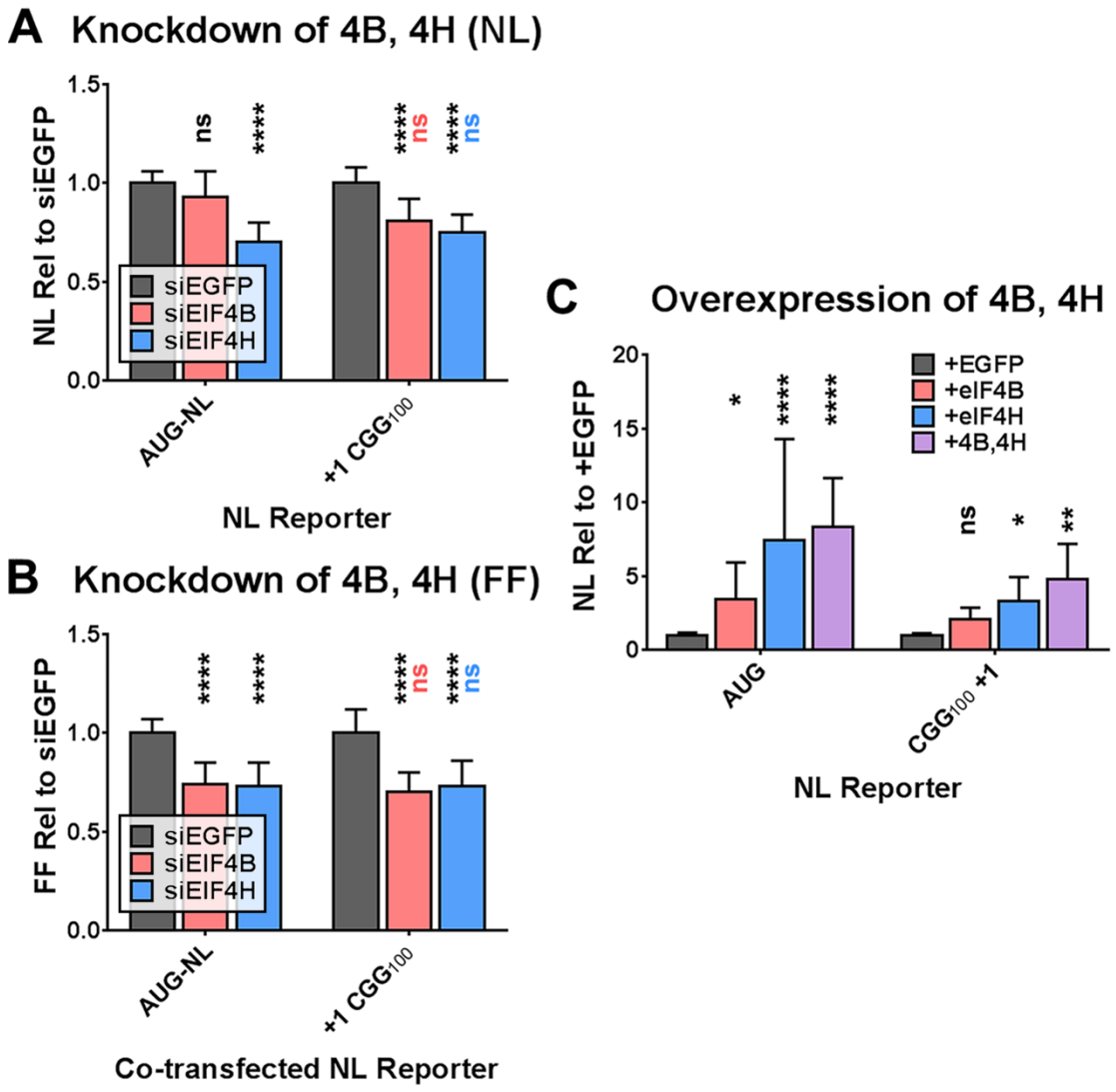


Figure 2.8: In Cultured Human Cells, *EIF4B* and *EIF4H* Modulate Both RAN and Canonical Translation.

A, B Expression of plasmid-based AUG-NL-3xF and +1 $(CGG)_{100}$ NL-3xF reporters (A) or the co-transfected AUG-FF reporters (B) following knockdown of *EIF4B* or *EIF4H*. Black asterisks refer to comparisons between siEGFP- and siEIF4B/H-treated cells; pink and blue asterisks refer to comparisons between siEIF4B- (pink) or siEIF4H- (blue) treated cells expressing AUG-NL-3xF and those expressing +1 $(CGG)_{100}$ NL-3xF (two-way ANOVA with Tukey's multiple comparisons test, *n*=9/condition).

C Expression of plasmid-based AUG-NL-3xF and $(CGG)_{100}$ +1 NL-3xF reporters with and without overexpression of *EIF4B*, *EIF4H*, or both (two-way ANOVA with Dunnett's multiple comparisons test; *n*=20/condition). Asterisks refer to comparisons between cells expressing the same reporter and either EGFP, *EIF4B, EIF4H,* or *EIF4B* and *EIF4H*.

Data Information: For all panels, ns=non-significant, * P≤0.05, ** P≤0.01, **** P≤0.0001 for the specified statistical test. All panels present data as means + SD (compiled from ≥3 replicates).

Several other eIFs are known to modulate start-codon fidelity[27,206]. eIF1 maintains the fidelity of scanning 43S PICs for AUG start codons by antagonizing the structural reconfigurations that follow AUG recognition[36,76,79-81]. We therefore asked whether *EIF1* overexpression impacts RAN translation. Overexpression of *EIF1* in cultured human cells decreased the expression of $(CGG)_{100}$ NL-3xF in the +1 and +2 frames (Figure 2.9A). Notably, inserting an AUG codon upstream of the CGG NRE in the +1 frame abolished this effect, indicating that the identity of the start codon is essential for modulation by eIF1. This demonstrates that manipulation of the molecular machinery that determines start-codon fidelity can modulate RAN translation at CGG repeats in human cells.

We next asked whether siRNA-mediated *EIF1* knockdown would modulate RAN translation, and we observed that *EIF1* knockdown resulted in significant inhibition of all transfected reporters (Figure A.10A). This finding is consistent with the known scanning-promoting functions of eIF1 during general translation initiation[36,80] and potentially explains the toxicity we observed with some *EIF1*-disrupting *Drosophila* lines in the absence of $(CGG)_{90}$-EGFP (Table A.1), as well as the reduced $(CGG)_{90}$-elicited toxicity we observed with *EIF1*-disrupting lines that were not, on their own, toxic.

We next asked whether enhancing expression of *EIF5* would affect RAN translation. eIF5 is an eIF2γ-specific GAP[227,228]. Once a 43S PIC recognizes a start codon, eIF1 dissociates from the PIC, and eIF5 promotes the hydrolysis of eIF2γ-bound ATP[229-232], a critical step in the dissociation of other eIFs that must precede ribosomal subunit joining. Furthermore, Loughran *et al*.[85] demonstrated that higher eIF5 abundance leads to increased initiation at non-AUG codons. Consistent with these

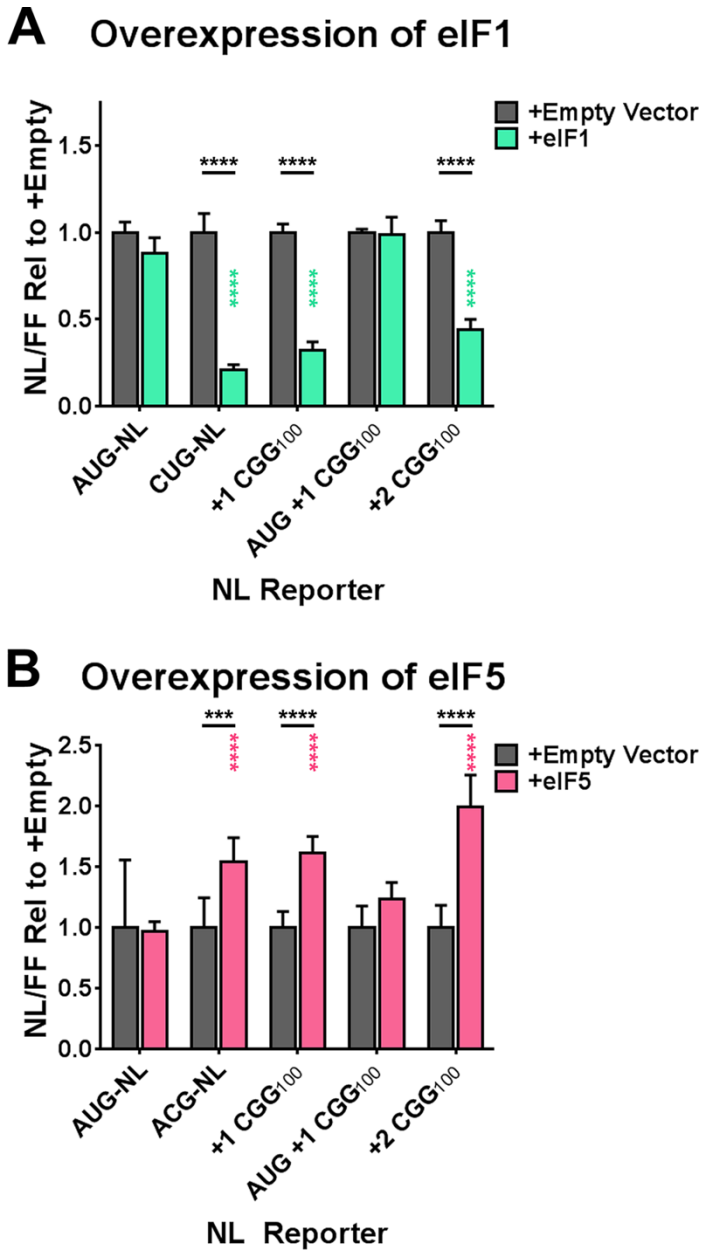


Figure 2.9: *EIF1* and *EIF5* Modulate RAN Translation by Determining AUG Start-Codon Specificity.

A Expression of plasmid-based NL-3xF reporters in HEK293 cells with and without overexpression of *EIF1* (two-way ANOVA with Sidak's multiple comparisons test; *n*=9-12/condition). Black asterisks refer to comparisons between empty vector-transfected and *EIF1*-transfected cells; green asterisks refer to comparisons between *EIF1*-transfected cells expressing AUG-NL-3xF and those expressing a different reporter.

B Expression of plasmid-based NL-3xF reporters in HEK293 cells with and without overexpression of *EIF5* (two-way ANOVA with Sidak's multiple comparisons test; *n*=9-12/condition). Black asterisks refer to comparisons between empty vector-transfected and *EIF5*-transfected cells; pink asterisks refer to comparisons between *EIF5*-transfected cells expressing AUG-NL-3xF and those expressing a different reporter.

Data Information: For all panels, *** P≤0.001, **** P≤0.0001 for the specified statistical test. All bars represent mean + SEM (compiled from ≥3 replicates).

---

results, we observed that *EIF5* overexpression in cultured human cells led to higher expression of ACG-initiated NL-3xF and +1 and +2 $(CGG)_{100}$ NL-3xF reporter plasmids, but not AUG-NL-3xF or AUG-initiated +1 $(CGG)_{100}$ NL-3xF reporters (Figure 2.9B). Moreover, we confirmed by Western blot that transfection of *EIF1* and *EIF5* cDNA-containing plasmids resulted in a higher level of eIF1 and eIF5 expression (Figure A.10B). In total, these experiments demonstrate that manipulation of factors that influence start-codon fidelity can up- or down-regulate RAN translation at CGG repeats.

Knockdown of *DDX3X* Suppresses Toxicity in $(CGG)_{100}$-Expressing Primary Neurons

The toxicity of the CGG NRE in *FMR1* is driven at least in part by the products of RAN translation[19,145,149]. Having demonstrated that DDX3X regulates the abundance of RAN products in mammalian cells, we asked whether knockdown of *DDX3X* can mitigate the toxicity of $(CGG)_{100}$ repeats in mammalian neurons. We transfected primary rat neurons with locked nucleic acids (LNAs) against *DDX3X* or non-targeting controls, along with plasmids containing the human *FMR1* 5' UTR with 100 CGG repeats upstream of an EGFP reporter in the +1 frame [+1 $(CGG)_{100}$ EGFP], along with an AUG-driven mApple construct to identify transfected cells. Over the following 10 days, we used automated longitudinal fluorescence microscopy to track the survival of transfected cells (Figure 2.10A)[233-235]. Expression of +1 $(CGG)_{100}$ EGFP markedly reduced survival compared to neurons expressing either EGFP (Figure A.11) or an EGFP reporter in which the AUG start codon has been replaced with a GGG codon (GGG-EGFP). Knockdown of *DDX3X* (Figure 2.10B) by two independent LNAs

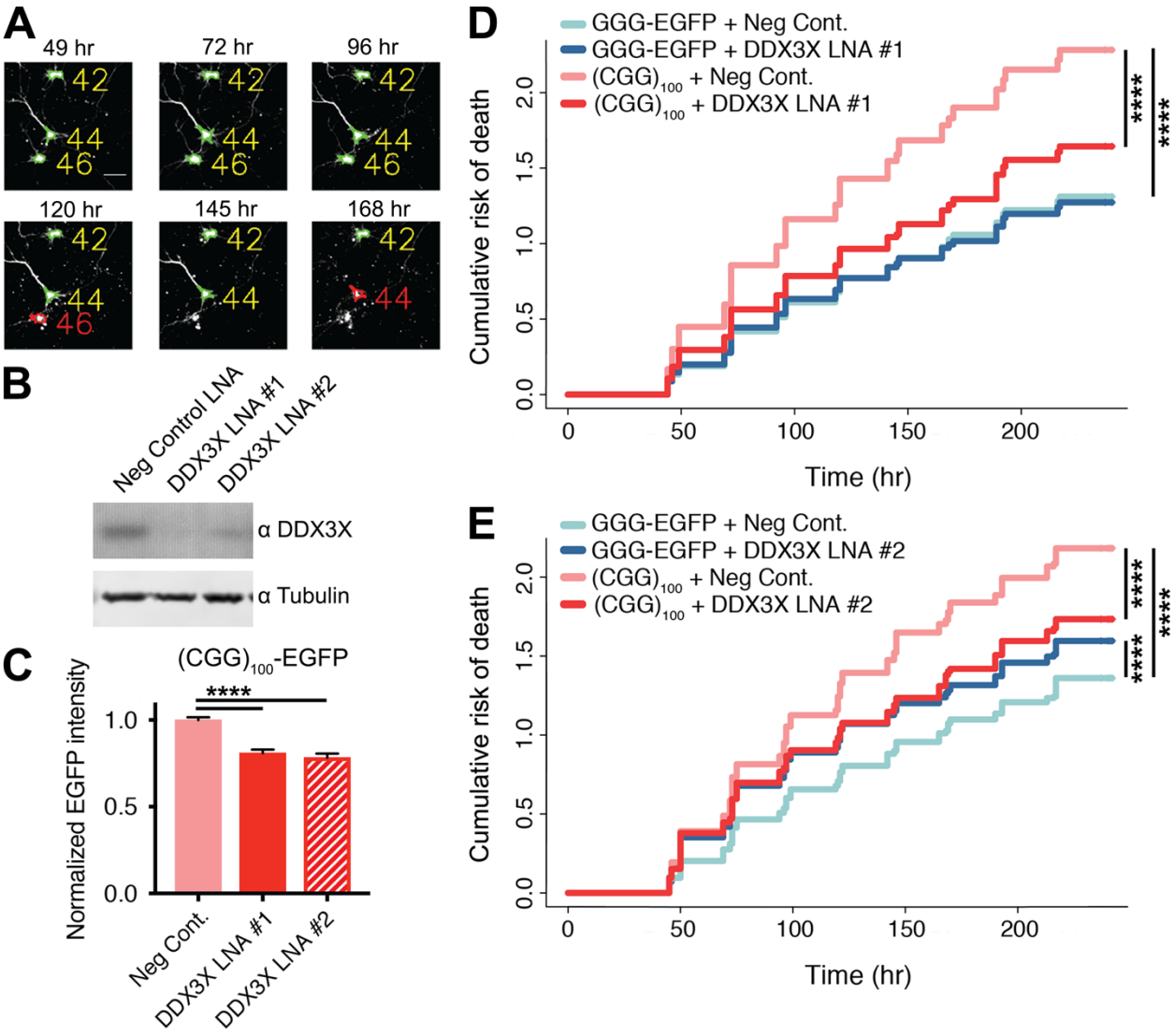


Figure 2.10: Knockdown of *DDX3X* Mitigates $(CGG)_{100}$ Toxicity in Primary Rodent Neurons.
A Sample micrographs collected by automated longitudinal fluorescence microscopy, demonstrating the automated determination of cell death in cultured rat neurons. Cells outlined in green are alive; those outlined in red are dead or dying.
B Anti-DDX3X western blot of B35 cells transfected with either of two independent anti-*DDX3X* locked nucleic acids (LNAs) or a control LNA.
C EGFP fluorescence in primary rat neurons transfected with +1 $(CGG)_{100}$-EGFP and either anti-*DDX3X* LNAs or a negative control LNA (one-way ANOVA with Tukey's multiple comparisons test; *n*=2408-5689 cells/condition). Graph depicts data normalized first within the replicate.
D, E Cumulative risk of death by automated longitudinal microscopy in rat neurons transfected with either GGG-EGFP control or +1 $(CGG)_{100}$-EGFP, as well as non-targeting LNAs, anti-*DDX3X* LNA #1 (D), or anti-*DDX3X* LNA #2 (E; Cox proportional hazard analysis; *n*=2408-3676 cells/condition).

Data Information: For all panels, **** P≤0.0001 for the specified statistical test (compiled from ≥3 replicates).

reduced the expression of FMRpolyG$_{100}$-EGFP in these neurons (Figure 2.10C) and significantly improved the survival of (CGG)$_{100}$-expressing neurons relative to transfection of non-targeting control LNAs (Figure 2.10D, E). We observed some neurotoxicity with one of the LNAs targeted DDX3X (#2) but not the other, potentially indicative of adverse off-target effects from LNA #2. These findings suggest that, like in *Drosophila*, knockdown of *DDX3X* suppresses CGG NRE-elicited toxicity by inhibiting RAN translation.

Knockdown of *DDX3X* Suppresses RAN Translation of G_4C_2 Repeats in *C9ORF72*

The mechanism of RAN translation might differ from NRE to NRE and from sequence context to sequence context[214]. We and others have previously demonstrated that initiation of RAN translation of *FMR1* and *C9ORF72* mRNAs is dependent on the sequence upstream of the NRE[150,169,170,203], indicating that *cis* features of mRNA can and do impact this process. Therefore, we asked whether DDX3X modulates RAN translation of other mRNAs, specifically, expanded G_4C_2 repeats in intron 1 of *C9ORF72*. In HeLa cells, knockdown of *DDX3X* by 5 independent siRNAs resulted in a dose-dependent suppression of two different transfected, plasmid-based G_4C_2 reporters, both of which report on the GA RAN product: one bearing 70 G_4C_2 repeats in the native intronic sequence upstream of NL-3xF in the +0 frame (Int 1 GA$_{70}$ NL-3xF; Figure 2.11A), and one bearing 71 G_4C_2 repeats embedded in non-native sequence upstream of the same reporter (GA$_{71}$ NL-3xF; Figure 2.11B). Unlike the +1 and +2 (CGG)$_n$ NL-3xF reporters for FMRpolyG and FMRpolyA expression, the effects of *DDX3X* knockdown on Int 1 GA$_n$ NL-3xF were repeat-dependent: GA$_3$ and GA$_{35}$

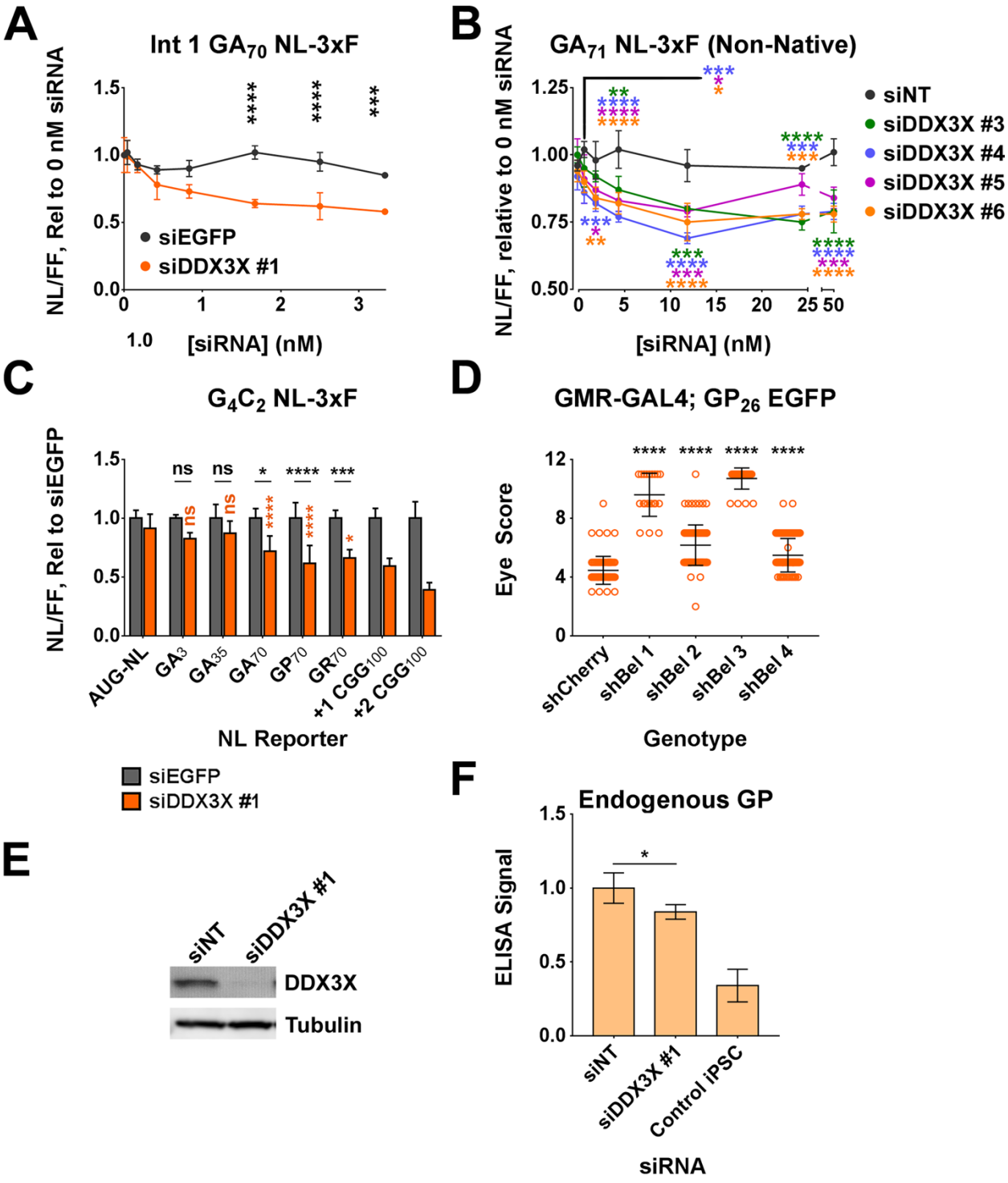


Figure 2.11: Knockdown of *DDX3X* Inhibits RAN Translation of Expanded G_4C_2 Repeats in *C9ORF72*.

A, B Dose-response curves plotting the effects of siDDX3X, siEGFP, or non-targeting siRNAs (siNT) on the expression of *C9ORF72* Intron 1 GA_{70} NL-3xF (A) and non-native GA_{71} NL-3xF (B) reporters [two-way ANOVA with Sidak's multiple comparisons test (A) or Dunnett's multiple comparisons test (B); *n*=3/condition]. Asterisks refer to comparisons between siRNAs at the

indicated concentration.
C Expression of *C9ORF72*- and *FMR1*-based RAN reporters with and without *DDX3X* knockdown across a range of repeat sizes (two-way ANOVA with Tukey's multiple comparisons test; *n*=3-13/condition compiled from ≥3 replicates). Black asterisks refer to comparisons between siDDX3X- and siEGFP-treated cells; orange asterisks refer to comparisons between siDDX3X-treated cells expressing AUG-NL-3xF and those expressing a different reporter.
D Quantitation of GMR-GAL4; $(GP)_{28}$ EGFP rough-eye phenotypes with *bel* disruptions (Kruskal-Wallis test; *n*=35-77/genotype).
E Anti-DDX3X western blot of C9 ALS iPSC lysates transfected with siNT or siDDX3X.
F Abundance of endogenous GP RAN product by ELISA in C9 amyotrophic lateral sclerosis (ALS) patient-derived and control induced pluripotent stem cell (iPSC) lines (Student's t test, *n*=3-5/group, compiled from ≥3 replicates).

Data Information: For all panels, * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001 for the specified statistical test. All panels depict data as mean ± SD.

---

reporters in the native intronic context were not significantly affected by *DDX3X* knockdown (Figure 2.11C). However, RAN translation of expanded $(G_4C_2)_{70}$ in all 3 sense reading frames of *C9ORF72*, including the GP and GR DPRs, were inhibited by *DDX3X* knockdown.

To determine whether *bel/DDX3X* knockdown might reverse $(G_4C_2)_n$-elicited toxicity, we crossed flies expressing shRNAs against *bel* to mates expressing 28 G_4C_2 repeats [$(G_4C_2)_{28}$ EGFP] under a GMR-GAL4 driver. Basally, expression of $(G_4C_2)_{28}$ EGFP in the *Drosophila* eye manifests in a marked rough-eye phenotype similar to $(CGG)_{100}$ EGFP (Todd lab; manuscript in preparation). To our surprise, *bel* knockdown significantly potentiated this phenotype (Figure 2.11D). This is the opposite of what we had expected based on our findings in HeLa cells, as well as the opposite of what we had observed in flies expressing $(CGG)_{90}$-EGFP following *bel* knockdown. The reason for this discrepancy is unclear but might entail the synthesis of aggregate-prone proteins out-of-frame from the EGFP ORF. To resolve this discrepancy, we turned to a C9 ALS/FTD patient-derived iPSC line, in which we can quantify the abundance of the GP

DPR using an ELISA[163]. In C9 iPSCs, *DDX3X* knockdown significantly reduced the abundance of the GP DPR (Figure 2.11E, F). This final experiment is a critical one, establishing that the effects of *DDX3X* knockdown are consistent between our reporter-based overexpression system (HeLa cells) and this endogenous-RAN product context.

Discussion

We performed a screen of eIFs and RNA helicases to identify modifiers of *FMR1*-associated RAN translation and toxicity in *Drosophila* and human cells, and found both selective (*bel*/*DDX3X*, *EIF1*, and *EIF5*) and non-selective (*EIF4B* and *EIF4H*) modifiers. Manipulation of these genes both reduced the expression of the RAN products FMRpolyG and FMRpolyA and mitigated CGG repeat-associated toxicity. This work both extends our understanding of RAN translation mechanisms and identifies potential therapeutic targets for FXTAS and other RAN translation-associated disorders.

RAN translation occurs in association with expanded, GC-rich, secondary structure-forming repeats that promote initiation in the absence of an AUG codon[14,19,150]. We originally hypothesized that disruption of RNA helicases, which resolve RNA-RNA secondary structures, would enhance RAN translation at CGG repeats. Our results demonstrated the opposite: knockdown of *bel*/*DDX3X*, *EIF4B*, or *EIF4H* inhibited *FMR1* RAN translation. How can we account for this discrepancy? *DDX3X* and its yeast homolog *Ded1* are DEAD-box RNA helicases that are required for resolution of RNA-RNA structures in long, GC-rich 5’ UTRs of particular genes[46,50,58,59,61,64-68]. The *FMR1* 5’ UTR is GC-rich (76%) independent of the CGG repeats and is predicted to form

highly-stable secondary structures[236-238] capable of stalling scanning 43S PICs[75,111]. Our results, in which DDX3X is required for initiation within the *FMR1* 5' UTR but not for expression of AUG-NL-3xF, are most consistent with a model in which DDX3X interacts with and resolves RNA-RNA secondary structures within the *FMR1* 5' UTR 5' to and within the CGG repeat, which allows access of scanning PICs to sites of initiation. According to this model, without DDX3X, scanning 43S PICs are unable to access the initiation sites for RAN translation, leading to a selective decrease in their use. In contrast, *EIF4B* and *EIF4H* are stimulatory factors for the DEAD-box RNA helicase eIF4A[40,44,45,239,240], which is critical for 43S PIC-mRNA binding and PIC scanning[46]. Our data, in which eIF4B and eIF4H regulate not only RAN translation but also general translation, is best explained by a model in which eIF4B and eIF4H facilitate 43S PIC attachment and basal scanning in the initial stages of RAN translation in a similar fashion to that for canonical translation. This model is consistent with previous observations that *FMR1* RAN translation resembles canonical translation during these early stages of initiation[150,169,170].

A key feature that distinguishes RAN translation from canonical translation is its use of non-AUG codons for initiation[14,150,169,170]. The specificity for AUG start codons is simultaneously central to the integrity of eukaryotic and prokaryotic proteomes and an essential point of regulation for determining which, when, and how much protein is synthesized from a given mRNA transcript[98,206,241-243]. RNA secondary structure-forming elements (including GC-rich NREs) are predicted to slow or stall PICs during scanning, enhancing the dwell time of codon-anticodon interactions at non-AUG codons and increasing the frequency of initiation events at non-AUG sites[95]. RAN translation might

initiate in a similar manner. If correct, this model predicts that increasing the abundance of eIFs that boost start-codon fidelity would inhibit RAN translation, while increasing abundance of eIFs that reduce start-codon fidelity would enhance RAN translation. Multiple factors and regulatory pathways converge to govern start codon selection, including *EIF1* and *EIF5*. eIF1 (originally identified as Sui1) is known to increase the specificity of scanning 43S PICs for AUG start codons[76,77], while higher availability and abundance of eIF5 (originally Sui5) have been shown to decrease start-codon fidelity[83,85,86].

Consistent with our prediction, we observed that *EIF1* overexpression decreased RAN translation in a start codon-dependent manner, while overexpression of *EIF5* did the opposite. However, as factors such as *EIF1* have multiple roles in translational initiation, suppression of their expression led to global decreases in translation that impacted both *FMR1* RAN translation and canonical translation, which elicited intrinsic toxicity in *Drosophila*. These results align with previous work demonstrating that RAN translation at both CGG and G_4C_2 NREs and across multiple reading frames is induced by activation of the ISR in a start codon-dependent manner via a mechanism that impinges on the AUG selectivity of the 43S PIC[169,202,203]. This line of research supports a model in which RAN translation represents a failure in start codon fidelity, suggesting that factors that regulate start-codon fidelity also regulate this pathologic process.

A central goal of understanding how RAN translation occurs, and how it is distinct from canonical translation, is to identify potential therapeutic targets. Prevention of *FMR1* RAN translation is sufficient to suppress toxicity/neurodegeneration in human cells *in vitro* and *Drosophila* and mice *in vivo*[19,145,149]. Here, we demonstrate that

disruption of *bel*/*DDX3X* not only inhibits *FMR1* RAN translation selectively *in vitro* and *in vivo*, but significantly mitigates repeat-induced toxicity across model systems. We suggest that targeting factors critical for resolving RNA secondary structures and/or enhancing start codon fidelity could represent viable therapeutic strategies for FXTAS and related neurodegenerative disorders. *DDX3X* in particular is currently the target of multiple lines of pharmacological research aimed at treating various cancers and viruses[244-246]. By targeting a proximal event in the pathophysiology of RAN translation-associated neurodegenerative disorders, this strategy has the potential to be more efficacious than targeting the toxic effects of each RAN product.

Materials and Methods

Drosophila Lines

All fly lines used here and their sources are listed in Table A.2.

Antibodies

For western blotting, all primary antibodies were used at 1:1,000 in 5% non-fat dairy milk (wt/vol), 0.02% NaN_3 (wt/vol), 0.1% Tween-20 (vol/vol) in TBS. Monoclonal mouse anti-EGFP antibody was acquired from Sigma (clones 7.1 and 13.1, catalog #11814460001). Monoclonal mouse anti-β tubulin antibody, developed by Michael Klymkowsky, was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH, and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. Monoclonal mouse anti-FLAG antibody was acquired from Sigma (clone M2, catalog #F1804). Polyclonal rabbit anti-DDX3X antibody (catalog

#2635), anti-eIF4B antibody (catalog #3592), anti-eIF4H antibody (catalog #2444), anti-eIF1 antibody (catalog #12496), and anti-eIF5 antibody (catalog #13894) were acquired from Cell Signaling Technology. For information on the anti-DDX3X and isotype control antibodies used for RIP, see the relevant section below.

Drosophila Phenotyping

All flies were raised and crossed at 25°C on SY10 food unless otherwise stated. For the screen, virgin female flies carrying the UAS-*FMR1* $(CGG)_{90}$-EGFP reporter[216] and a GMR-GAL4 driver [Bloomington Drosophila Stock Center (BDSC) 8605] were crossed to males carrying UAS-driven shRNA constructs against, UAS-driven transgenes of, or germline mutations in candidate genes. Rough-eye phenotypes in F_1 progeny were photographed and scored at 0-1 day post-eclosion according to a rubric adapted from Pandey *et al*.[247]. One point was given for each of the following morphological aberrations: supernumerary inter-ommatidial bristles, abnormal orientation of inter-ommatidial bristles, disorganization of the ommatidial array, ommatidial fusion, and total loss of the ommatidial array over 10% of the eye surface. Two points were given for each of the following: the presence of necrotic lesions, collapse of the eye's convex surface, and shrinkage of the eye's surface area by 25%. Individual flies could therefore score between 0 and 11, with higher scores indicating a more severe phenotype. Eye images were captured using a Leica M125 stereomicroscope and a Leica DFC425 digital camera.

For longevity assays, flies carrying $(CGG)_{90}$-EGFP and either a Tub5-GAL4 GeneSwitch or ElaV-GAL4 GeneSwitch driver (both RU486-inducible)[224] were placed

on SY10 food supplemented with 200 μM RU486 and flipped onto fresh RU486-supplemented SY10 every 24 (Tub5) or 48 (ElaV) hours, and kept at 29°C until expiration. Deaths were counted and dead flies removed every 24 or 48 hours.

Western Blotting and qRT-PCR of *Drosophila* Tissue

For western blotting and qRT-PCR of fly tissue, <2 days post-eclosion flies carrying $(CGG)_{90}$-EGFP and the Tub5-GAL4 GeneSwitch driver were placed on 200 μM RU486-supplemented SY10 for 72 hours, with fresh RU486-supplemented food provided every 24 hours, at 29°C. For western blotting, flies were homogenized at 4°C in ice-cold radioimmunoprecipitation assay (RIPA) buffer [50 mM Tris (pH 8.0), 150 mM NaCl, 0.1% SDS (wt/vol), 0.5% sodium deoxycholate (wt/vol), 1% IGEPAL CA-630 (vol/vol), and cOmplete-Mini protease inhibitor (Roche)], then briefly centrifuged at 12K RPM for 2 minutes at 4°C to pellet cuticle and wing debris. The supernatant was removed and the chromatin sheared by 10 strokes through a 28.5G syringe. Lysates were subsequently mixed with 6x reducing Laemmli buffer, separated by SDS-PAGE, and transferred to PVDF membranes (Biorad).

For qRT-PCR assays, flies were homogenized in TRIzol (Thermo Fisher Scientific) and total RNA was subsequently extracted. Ten μg of RNA per sample were twice incubated with 2 U of TURBO DNase (Thermo Fisher Scientific) in reaction volumes of 50 μl at 37°C for 30 mins, per manufacturer's instructions, then recovered using RNA Clean and Concentrator-25 kits (Zymo Research). Five hundred ng of DNase-treated RNA per sample were used to generate cDNAs using a mixture of oligo(dT) and random hexamer primers (iScript cDNA Synthesis Kit, Biorad). cDNA

abundance was measured using iQ SYBR Green Supermix (Biorad), appropriate primers at 300 nM (Table A.3), and an iQ5 qPCR system (Biorad).

Plasmids

All reporter constructs used here were generated and described by Kearse *et al*.[150] and/or Green *et al*.[169]. In brief, all nanoluciferase reporters were developed by site-directed mutagenesis or digestion/ligation from pcDNA3.1(+)/AUG-NL-3xF. All pcDNA3.1(+)/*FMR1* $(CGG)_n$ NL-3xF constructs bear the full human 5' UTR of *FMR1* upstream of the CGG repeats. In addition, in all *FMR1* $(CGG)_n$ constructs used here, the initiator ATG of NL-3xF has been mutated to GGG to abolish initiation at this site. The FF luciferase construct pGL4.13 was acquired from Promega. For *in vitro* transcription of FF luciferase RNA, the FF luciferase construct was digested and ligated into a pcDNA3.1(+) vector using 5' HindIII and 3' XbaI restriction sites (Rapid DNA Dephos and Ligation Kit, Sigma). All plasmids used for transfection and *in vitro* transcription were prepared from *E. coli* cultures using ZymoPURE Plasmid Midiprep Kits (Zymo Research).

pCMV6-XL5/EIF4B and pCMV5-XL5/EIF4H, which drive expression of human *EIF4B* and *EIF4H*, respectively, were acquired from OriGene. pcDNA3.1(+)/EIF1 and pcDNA3.1D/EIF5-V5-His, which drive expression of human *EIF1* and *EIF5*, respectively, were acquired from J. Schofield.

Cell Culture and Transfection

HeLa (CCL-2, ATCC), HEK293 (CRL-1573, ATCC), HEK293T (CRL-3216,

ATCC), and B35 (CRL-2754, ATCC) cells were cultured and passaged at 37°C, 5% CO_2, with HeLa cells in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% non-essential amino acids, and HEK293, HEK293T, and B35 cells in DMEM supplemented with 10% FBS.

For luciferase assays, HeLa cells were plated in 96-well plates at 1.0×10^4 cells/well in 100 µl media and reverse transfected with Stealth siRNAs against human *DDX3X* (DDX3XHSS102712 and DDX3XHSS176054, Thermo Fisher Scientific) or EGFP at 1.67 nM, unless otherwise noted, using Lipofectamine RNAiMAX (Thermo Fisher Scientific). Alternatively, they were transfected with ON-TARGETplus SMART*pool* siRNAs against human *EIF4B* (L-020179-00, Dharmacon), *EIF4H* (L-013054-00), or *EIF1* (L-015804-02) or a non-targeting pool (D-001810-10). In brief, siRNA and RNAiMAX were diluted in Opti-MEM, combined, incubated for 10 mins at room temperature, then mixed with cells. For subsequent plasmid transfection, 24 hours after plating cells were transfected with 25 ng/well pcDNA3.1(+)/NL-3xF plasmid and 25 ng/well pGL4.13 as transfection control using jetPRIME (Polyplus). The transfection media was removed and replaced 4 hours post-transfection. For RNA reporter transfection, 24 hours after plating cells were transfected with 25 ng/well *in vitro*-transcribed nanoluciferase RNA and 25 ng/well firefly luciferase RNA using TransIT mRNA (Mirus Bio). Luciferase assays were performed 24 hours after plasmid transfection, as described by Kearse *et al*.[150] and Green *et al*.[169].

For overexpression experiments, cells were plated in 96-well plates at 1.0×10^4 cells/well (HeLa) or 2.0×10^4 cells/well (HEK293T) in 100 µl media. Twenty-four hours after plating, HeLa cells were transfected with 5 ng/well pcDNA3.1(+)/NL-3xF, 5 ng/well

pGL4.13, and 40 ng/well pCMV6-XL5/EIF4B, pCMV5-XL5/EIF4H, pEGFP N1, or a combination thereof, using jetPRIME (Polyplus). HEK293T cells were transfected with 25 ng/well pcDNA3.1(+)/NL-3xF, 25 ng/well pGL4.13, and 250 ng/well pcDNA3.1(+)/EIF1, empty pcDNA3.1(+), pcDNA3.1D/V5-His-EIF5, or empty pcDNA3.1D/V5-His using FuGene HD (Promega). Luciferase assays were performed as above.

For western blotting experiments, HeLa cells were plated in 12-well plates at 1.5×10^5 cells/well in 1 ml media and reverse transfected, as above, with Stealth siRNAs against *DDX3X* or EGFP at 1.67 nM using Lipofectamine RNAiMAX. Alternatively, they were transfected with ON-TARGETplus SMART*pool* siRNAs against *EIF4B*, *EIF4H*, or a non-targeting pool (siNT) at 15 nM. Twenty-four hours after plating cells were transfected with 500 ng/well pcDNA3.1(+)/NL-3xF using jetPRIME (Polyplus). The transfection media was removed and replaced 4 hours post-transfection. Twenty-four hours after plasmid transfection, cells were lysed on-plate in RIPA buffer. The lysate was homogenized by 10 strokes through a 28.5 G syringe (without centrifugation), mixed with 6x reducing Laemmli buffer, heated at 90°C for 10 mins, resolved by SDS-PAGE, and transferred to a PVDF membrane before incubation in primary antibody.

For qRT-PCR experiments, HeLa cells were plated in 6-well plates at 2.5×10^5 cells/well in 2.5 ml media and reverse transfected with Stealth siRNAs against DDX3X or EGFP at 1.67 nM using Lipofectamine RNAiMAX. Twenty-four hours after plating cells were transfected with 625 ng/well pcDNA3.1(+)/NL-3xF and 625 ng/well pGL4.13 using jetPRIME (Polyplus). The transfection media was removed and replaced 4 hours post-transfection. Twenty-four hours after plasmid transfection, cells were lysed and

total cellular RNA collected using Quick-RNA MiniPrep Kit (Zymo Research). Five µg of RNA per sample were incubated twice with 2 U of TURBO DNase (Thermo Fisher Scientific) for 30 minutes at 37°C to remove contaminating genomic and plasmid DNA, then recovered using the RNA Clean and Concentrator-25 Kit (Zymo Research). cDNA was generated from 500 ng of DNase-treated RNA per sample and a mixture of oligo(dT) and random hexamer primers (iScript cDNA Synthesis Kit, Biorad). cDNA abundance was measured using iQ SYBR Green Supermix (Biorad), an iQ5 qPCR system (Biorad), and the appropriate primers at 300 nM. cDNA abundance was quantified using a modified ΔΔ C_t method recommended by the manufacturer and have been presented as normalized to spiked-in *in-vitro* transcribed RNAs to account for differences in RT efficiency.

For confirmation of anti-*DDX3X* LNA efficacy, B35 cells were plated in 12-well plates at 2.0×10^5 cells/well in 1 ml media and reverse transfected with anti-*DDX3X* Silencer Select LNAs (s165214 and s165216, Thermo Fisher Scientific) or a non-targeting control (4390843, Thermo Fisher Scientific) at 40 nM using Lipofectamine RNAiMAX. Forty-eight hours after plating, cells were lysed as above for western-blot analysis.

Cross-linking and RNA Immunoprecipitation

HEK293 cells were plated in poly-L-lysine-coated 10-cm plates at 3.0×10^6 cells/plate. Forty-eight hours after plating cells were transfected with 5 µg of pcDNA3.1+/CGG_{100} (+1) NL-3xF and 5 µg of either pGL4.13 or pEGFP N1 using Viafect (Promega). Twenty-four hours post-transfection the media was aspirated and replaced

with fresh media supplemented with 6-thioguanisine[248] at 100 μM and allowed to incubate for 12 hours.

Cells were rinsed 3x in PBS (pH 7.4), the PBS was aspirated, and the cells were irradiated uncovered with 0.6 J/cm^2 of 365 nm UV light using a Stratalinker 2400 (Stratagene). Cells were then harvested using trypsin (0.25%)-EDTA and rubber policemen, collected by centrifugation, rinsed 2x in ice-cold PBS, flash-frozen in a dry ice/EtOH bath, and stored at -80°C. For processing, cell pellets were lysed in NP-40 lysis buffer [50 mM Tris (pH 7.5), 150 mM NaCl, 1 mM DTT, 0.5% IGEPAL CA-630 (wt/vol)] supplemented with cOmplete Mini EDTA-free protease inhibitors (Roche), PhosSTOP phosphatase inhibitors (Roche), 1 U/μl recombinant RNAsin (Promega), and 200 U/ml SuperaseIN (Thermo Fisher Scientific) for 25 mins on ice, incubated with 42 U/ml RQ1 DNase (Promega) at 37°C for 10 mins, then centrifuged at 10K *g*, 4°C for 10 mins.

Protein G Dynabeads (Thermo Fisher Scientific) were prepared and incubated with mouse anti-DDX3X (clone 2253C5a, Santa Cruz sc-81247), mouse anti-EGFP (clones 7.1/13.1, Sigma 11814460001), or mouse isotype control IgG (Thermo Fisher Scientific 10400C) as per the manufacturer's instructions. Antibody-conjugated beads were rinsed 3x in NP-40 lysis buffer and incubated with the cleared lysate for 16 hours with inversion at 4°C. Only those cells transfected with pEGFP N1 were subjected to anti-EGFP RIP. The lysate was removed, and the beads were washed 3x with NP-40 lysis buffer and 3x with 5x PBS (pH 7.4) supplemented with 0.5% IGEPAL CA-630. RNA was eluted by incubation with 2 mg/ml Proteinase K at 55°C for 1 hour in 50 mM Tris-HCl (pH 7.0), 75 mM NaCl, 6 mM EDTA, and 2% SDS (wt/vol), extracted using

TRIzol (Thermo Fisher Scientific) and GlycoBlue co-precipitant (Thermo Fisher Scientific), treated twice with 40 U/ml TURBO DNase at 37°C for 30 mins, and purified using the RNA Clean and Concentrator-5 Kit (Zymo Research). cDNAs were generated using the iScript cDNA Synthesis Kit, as described above, with each reaction spiked with equal amounts of *in vitro*-transcribed AUG-FF or EGFP RNA (depending on which plasmid had not been co-transfected) as a reverse-transcription control. qPCR was performed as described above, with cDNA abundance normalized to spiked-in FF or EGFP cDNA abundance.

In vitro Transcription and Translation Reactions

pcDNA3.1(+)/NL-3xF and pcDNA3.1(+)/FF were linearized by PspOMI and XbaI restriction enzymes (NEB), respectively, and recovered using DNA Clean and Concentrator-25 kits (Zymo Research). m^7G-capped and poly-adenylated RNAs were transcribed *in vitro* from these plasmids using HiScribe T7 ARCA mRNA kit (with tailing; NEB) as per the manufacturer's instructions and recovered using RNA Clean and Concentrator-25 kits (Zymo Research). The integrity and size of all transcribed RNAs were confirmed by denaturing agarose gel electrophoresis with formaldehyde/formamide.

For preparation of translation-competent extracts, HeLa cells were plated in 14.5-cm dishes at 8×10^6 cells/plate. Twenty-four hours later they were forward transfected with Stealth siRNAs against *DDX3X* or EGFP at 1.67 nM using RNAiMAX, as described above (adapted from Rakotondrafara & Heintze[249]). The transfection media was removed 5 hours post-transfection and replaced with fresh media. Two days post-

transfection, cells were harvested using trypsin (0.25%)-EDTA, centrifuged, and rinsed 3x with PBS (pH 7.4). Cells were allowed to swell on ice in a volume of hypotonic lysis buffer [10 mM HEPES-KOH (pH 7.6), 10 mM KOAc, 0.5 mM Mg_2OAc, 5 mM DTT, supplemented with cOmplete Mini, EDTA-free protease inhibitor] equal to the cell pellet volume for 30 mins. Cells were mechanically disrupted at 4°C using 20 strokes in a 27G syringe, then allowed to incubate on ice for an additional 20 minutes. Lysis was confirmed visually in >95% of cells by trypan blue inclusion. The lysate was centrifuged at 10K *g* for 10 mins at 4°C. The supernatant was then collected, diluted in lysis buffer to 8.0 µg/µl using a modified Bradford protein quantification assay (Biorad), flash frozen in liquid N_2, and stored at -80°C.

For *in vitro*-translation reactions, lysates were brought to final concentrations of 20 mM HEPES-KOH (pH 7.6), 44 mM KOAc, 2.2 mM Mg_2AOc, 2 mM DTT, 20 mM creatine phosphate (Roche), 0.1 µg/µl creatine kinase (Roche), 0.1 mM spermidine, and on average 0.1 mM of each amino acid (with relative amounts approximating those in eukaryotes[250]). To this, *in vitro*-transcribed RNAs were added to 4 nM in a final volume of 10 µl per reaction. After incubation at 30°C for 30 min, 25 µl room-temperature Glo Lysis buffer (Promega) was added to halt the reaction and allowed to incubate for 5 mins at room-temperature. To 25 µl of this mixture was added 25 µl of NanoGlo substrate freshly diluted in NanoGlo buffer (Promega). This mixture was allowed to incubate in opaque 96-well plates on a rocking shaker in the dark for 5 minutes before the luminescence detection and quantification using a GloMax microplate luminometer (Promega).

Polysome Fractionation

HeLa cells were seeded in four to eight 10-cm dishes per condition. Twenty-four hours after plating, cells were transfected with siRNAs against *DDX3X* or *EGFP* at 1.6 nM using RNAiMAX, as above, with the media exchanged at 5 hours post-transfection. When cells reached 70-90% confluency, 24-36 hours post-knockdown, they were treated with 100 µg/mL cycloheximide (CHX) for 5 minutes at 37° C. Cells were then transferred to ice and washed with 2.5 mL ice-cold PBS containing 100 µg/mL CHX, collected by scraping in 2.5 mL cold PBS+CHX, and pelleted at 1200 RPM and 4° C for 5 minutes. PBS was aspirated and pellets re-suspended in polysome-profiling lysis buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 15 mM $MgCl_2$, 8% (vol/vol) glycerol, 20 U/mL SUPERase, 80 U/mL Murine RNase Inhibitor, 0.1 mg/mL heparin, 100 µg/mL CHX, 1 mM DTT, 1x EDTA-free protease inhibitor cocktail, 20 U/mL Turbo DNase, 1% Triton X-100)[251]. Lysates were passed through a 20G needle 10x and incubated on ice for 5 minutes. Cellular debris was pelleted at 14,000 g and 4° C for 5 minutes, and supernatant transferred to a fresh tube. Total lysate RNA was estimated by NanoDrop. Lysates were flash frozen in liquid N_2 and stored at -80°C until fractionation.

Sucrose gradients were prepared by successively freezing equal volumes of 50%, 36.7%, 23.3%, and 10% sucrose (wt/vol) in 12-mL Seton tubes. Sucrose-gradient buffer consisted of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 15 mM $MgCl_2$, 10 U/mL SUPERase, 20 U/mL Murine RNase Inhibitor, 100 µg/mL CHX, and 1 mM DTT[251]. Prior to use, gradients were allowed to thaw and linearize overnight at 4° C[252]. For fractionation, approximately 90 (trial 1 with four 10-cm dishes), 220, and 250 µg (trials 2 and 3, respectively, with eight 10-cm dishes each) total RNA was applied to the top of

the sucrose gradient. Gradients were spun at 35,000 RPM and 4° C for 3 hours using a Beckman Coulter Optima L-90K ultracentrifuge and SW 41 Ti swinging-bucket rotor.

Gradients were fractionated with Brandel's Gradient Fractionation System, measuring absorbance at 254 nm. The detector was base-lined with 60% sucrose chase solution, and its sensitivity set to 0.5 for trial 1 and 1.0 for trials 2 and 3. For fractionation, 60% sucrose was pumped at a rate of 1.5 mL/min. Brandel's PeakChart software was used to collect data, overlay profiles, and calculate the area under the curve for monosome and polysome fractions.

Primary Neuronal Cultures and Automated Fluorescence Microscopy

Embryonic day (E) 19-20 Long-Evans rat (*Rattus norvegicus*) cortices were harvested and the neurons dissociated and plated in 96-well plates at 6.0×10^6 cells/mL as previously described[253]. On *in vitro* day (DIV) 4, neurons were transfected using Lipofectamine 2000 (Thermo Fisher Scientific)[254] with 100 ng/well pGW/$(CGG)_{100}$ +1 EGFP or pGW/GGG-EGFP, 50 ng/well pGW/mApple, and LNAs to a final concentration of 40 nM. Following transfection, neurons were maintained in NEUMO photostable media (Cell Guidance Systems) for the duration of the experiment.

Neurons were imaged using a Nikon Eclipse Ti inverted microscope with PerfectFocus3 and Nikon Plan Fluor 20X objective lens[234]. Cells were illuminated with a Lambda XL Xenon lamp (Sutter Instrument) and detected using an Andor iXon3 897 EMCCD or Andor Zyla4.2 (+) sCMOS camera. Stage, filter, and shutter movements were controlled with scripts written in Beanshell for use in μManager. Separate ImageJ/Fiji macros and Python scripts were employed for automated identification of

transfected neurons and the drawing of regions of interest around each neuron[235]. Cell death was indicated by rounding of the cell body, deterioration of neuronal processes, and loss of mApple fluorescence intensity.

iPSC Culturing, Transfection, and GP ELISA

iPSCs derived from human fibroblasts that were harvested from control patients and C9 ALS patients were grown using standard techniques in Essential-8 media (Invitrogen). The day before transfection, cells were split using 50 mM EDTA with vigorous trituration in Essential-8 media containing RockI (Y27632, Fisher) on vitronectin (Fisher)-coated plates. The next day, media was replaced with mTESR (StemCell). For siRNA transfection, 72 pmols of each siRNA were transfected using 4.8 µl Lipofectamine RNAiMax reagent (Invitrogen) per well of a 6-well dish. The transfection mixture was added to cells in 1.6 ml of media and removed 20-24 hours later. Cells were harvested for protein isolation 48 hours post-transfection and lysed in 50 mM Tris-HCl (pH 7.4), 300 mM NaCl, 5 mM EDTA, 1.0% (vol/vol) Triton X-100, 2% (wt/vol) SDS in the presence of phosphatase and protease inhibitors. Protein concentration was quantified using BCA Protein Assay Kit (Thermo Fisher Scientific) and the lysates analyzed by ELISA as previously described[163].

Data Analysis

Band intensity on western blots was quantified using ImageJ (NIH; anti-EGFP) or Odyssey Image Studio (LI-COR; anti-tubulin) software. Primary neuron survival analysis and determination of hazard ratios through Cox proportional hazard analysis was

conducted using the publicly available survival package in R. All other data were analyzed using GraphPad Prism 7.00.

Acknowledgements

We thank Scott Pletcher's lab at the University of Michigan for providing Tub5-GS and ElaV-GS lines and all fly food. We thank Peng Jin's lab at Emory University for providing the $(CGG)_{90}$-EGFP line. We thank Wu-Min Deng's lab at Florida State University for providing mutant *bel* lines. We thank Ting Xie's lab at the University of Kansas for providing the UAS-eIF4A line. We thank J. Schofield for pcDNA3.1+/EIF1 and pcDNA3.1D/EIF5-V5-His plasmids. We thank David Turner for technical assistance with photo-crosslinking and RIP experiments. We thank Michael Kearse and everyone in the Todd lab for many thoughtful conversations and their collective wisdom.

.

Chapter 3

A High-Throughput, Genome-Wide Screen Identifies Modifiers of *C9ORF72*- and *FMR1*-Associated RAN Translation[3]

Abstract

The G_4C_2 nucleotide-repeat expansion (NRE) in *C9ORF72* is the most frequent cause of familial and sporadic amyotrophic lateral sclerosis (ALS)/frontotemporal dementia (FTD). One mechanism through which expanded G_4C_2 RNA repeats elicit toxicity is by inducing aberrant protein synthesis in the absence of an AUG codon, leading to the production of toxic dipeptide repeat (DPR)-containing proteins. This non-canonical mode of translation, known as repeat-associated, non-AUG (RAN) translation, is observed in the context of several NRE-associated neurodegenerative disorders. In an effort to elucidate the mechanism of RAN translation and identify novel therapeutic targets, we conducted a high-throughput, genome-wide siRNA screen for genes that modulated expression of the glycine-alanine (GA) DPR RAN product using luciferase-based reporters. Following a series of progressively discriminating screens, we identified 552 genes as selective modifiers of GA expression, the majority of them suppressors. 190 of these genes also modified the expression of RAN products synthesized from expanded CGG repeats in the 5' UTR of *FMR1*, another disease-associated, RAN-competent gene. Among these were the RNA-binding protein *HNRNPA2B1* and the proteasomal subunit *PSMB5*. In follow-up experiments using independent siRNAs, we demonstrated that knockdown of *HNRNPA2B1* or *PSMB5* modulated the expression of both *C9ORF72* and *FMR1* RAN products across reading frames, frequently in a repeat-dependent manner. Additional work will be required to narrow and prioritize this list of RAN modifiers, including elimination of putative off-target effects.

[3] Tseng Y, Santoro N, and Glineburg MR contributed significantly to the work presented in this chapter. YT assisted with screen data collection, NS with high-throughput procedure design, and MRG with HCR experiments.

Introduction

ALS and FTD are progressive, fatal neurodegenerative disorders for which there are no effective disease-modifying therapies. Median survival after symptom onset is ~30 months for ALS[255,256] and ~8 years for FTD[257-260]. Despite differences in their initial clinical presentation, diagnostic criteria, and prognoses, ALS and FTD are linked by their shared disease-causing mutations, histopathological hallmarks, and some clinical features[155-157]. Major hypotheses regarding the etiology and pathogenesis of ALS/FTD vary widely[261] and include persistent enteroviral infection[262]; mitochondrial dysfunction[263]; disruption of protein homeostasis[264,265]; DNA structural aberrations, double-stranded breaks, and epigenetic dysregulation[161,266-269]; and perturbations of RNA metabolism and ribonucleoprotein (RNP) phase separation[270,271]. In the last decade, extensive research has focused for multiple reasons on one ALS/FTD-associated mutation, the NRE of G_4C_2 repeats in the gene *C9ORF72*. It is the most frequent genetic cause of sporadic and familial ALS/FTD[151,152,272]. It can be modeled with relative ease in a variety of *in vitro* and *in vivo* systems. And, critically, expanded G_4C_2 DNA and/or RNA elements can be plausibly linked to several of the previously hypothesized causes of ALS/FTD.

Genetic screens have proved effective tools to generate novel hypotheses regarding the pathophysiology and treatment of NRE-associated neurodegeneration[273-278]. Much of the initial study of *C9ORF72*-associated ALS/FTD pathogenesis was launched using genetic screens in a variety of model systems[172,180,182,279,280]. These were designed to identify genetic modifiers of G_4C_2 NRE-elicited toxicity, often broadly.

By design, identified modifiers could genetically interact with any of several possible molecular sources of G_4C_2 toxicity: G_4C_2 DNA, G_4C_2 and G_2C_4 RNA, or the aberrant DPR-containing protein products of G_4C_2 RAN translation[15,17,161,173,194,281,282]. In contrast, we are interested specifically in how RAN translation of G_4C_2-expanded *C9ORF72* initiates, because 1) RAN translation appears to violate basic assumptions regarding protein synthesis and start-codon fidelity, and 2) a mechanistic understanding of RAN translation might generate novel therapeutic strategies for C9 ALS/FTD and other RAN translation-associated neurodegenerative disorders[25].

Identifying the proteins and pathways that modulate RAN translation will be key to understanding its mechanism and how it departs from canonical, AUG-initiated translation. To do this, we and others have previously used either hypothesis-driven approaches[23,150,169,202,283] or limited, candidate-based screens (see Chapter 2). To interrogate comprehensively the mechanism of RAN translation, as well as identify potential therapeutic targets, we conducted a high-throughput, genome-wide screen for modifiers of *C9ORF72*-based RAN translation and an abbreviated screen for modifiers of *FMR1*-based RAN translation in cultured human cells.

Results

A Non-Biased, Genome-Wide, High-Throughput Screen Identifies Modifiers of *C9ORF72* RAN Translation

We plated HEK293 cells and transfected them with a genome-wide library of siRNAs (the human siGENOME SMARTpool library, Dharmacon), then co-transfected

two plasmids (Figure 3.1A): 1) a FLAG-tagged nanoluciferase reporter (NL-3xF) downstream of the human *C9ORF72* intron 1 bearing 70 G_4C_2 repeats (GA_{70} NL-3xF), which is translated to produce a fusion of the GA_{70} RAN product and NL-3xF reporter; and 2) an AUG-initiated FF luciferase reporter for general translation and cell viability. The GA_{70} construct was selected from the available DPR reporters because of its high signal and wide dynamic range[169] and because the GA product is the most abundant DPR protein in C9 ALS/FTD patient tissue[16,17,168]. Dual transfection with subsequent independent luminescence assays enabled the differentiation of modifiers specific to RAN translation (NL-3xF) from modifiers of general translation and cell viability (FF). Non-targeting siRNAs (siNT; Dharmacon) and siRNAs targeting the ORF of NL (siNL) served as negative and positive controls, respectively (Figure 3.1B). In addition, siRNAs against *PLK1* (siPLK1) served as positive controls for loss of cell viability, as *PLK1* knockdown leads to mitotic arrest. For the purpose of identifying hits, our chief screening statistic was the GA_{70} NL-3xF luminescence normalized to AUG-FF luminescence (NL/FF). Across the entire *C9ORF72* primary screen the average Z′ score, calculated from NL/FF values of siNT- and siNL-transfected cells, was 0.681,

---

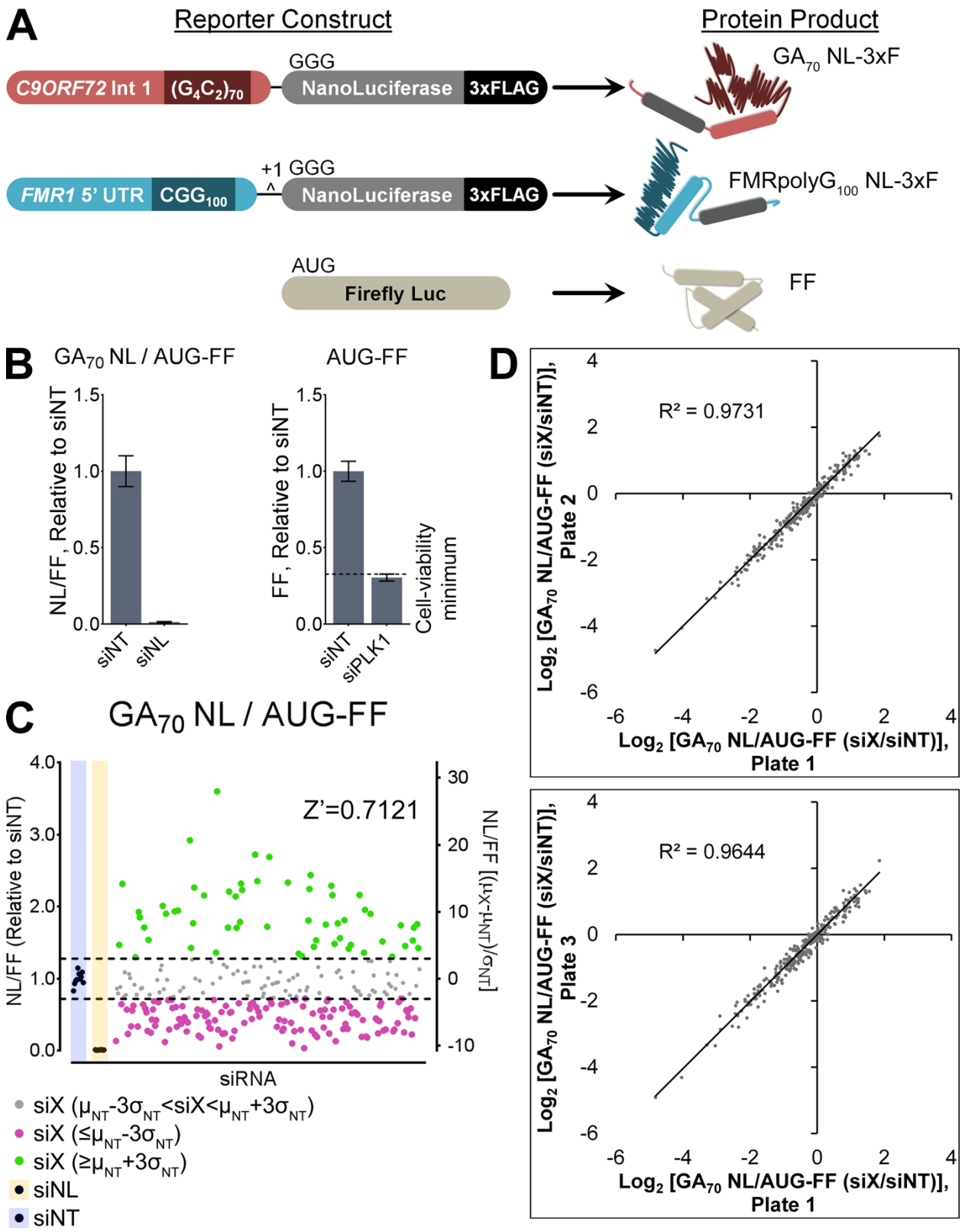
Figure 3.1: Design of High-Throughput siRNA Screens for Modifiers of *C9ORF72* and *FMR1* RAN Translation.

A Schematics of the NL- and FF-based reporter constructs used for *C9ORF72*-based RAN, *FMR1*-based RAN, and AUG-initiated translation.

B Normalized expression of GA_{70} NL-3xF following transfection of siNT and nanoluciferase siRNA (siNL; negative and positive controls respectively; left panel), and relative expression of AUG-FF following transfection of siNT and siPLK1 (right panel). The FF luminescence value representing minimally accepted cell viability [mean + 1 standard deviation (SD)] is drawn.

C NL/FF ratios gathered from a representative assay plate, presented normalized to the mean of siNT-transfected cells (left y axis) and as the number of SDs from the mean of siNT-transfected cells (right y axis). siNL-transfected cells are included as a positive control. The lines demarcating +3 and -3 SD from the mean siNT are drawn.

D Correlations between Log_2 NL/FF (siX/siNT) ratios by gene across triplicates.


A
Reporter Construct
Protein Product
C9ORF72 Int 1
(G4C2)70
GGG
NanoLuciferase
3xFLAG
GA70 NL-3xF
FMR1 5' UTR
CGG100
+1 GGG
FMRpolyG100 NL-3xF
AUG
Firefly Luc
FF
B
GA70 NL / AUG-FF
AUG-FF
NL/FF, Relative to siNT
FF, Relative to siNT
siNT
siNL
siPLK1
Cell-viability minimum
C
GA70 NL / AUG-FF
Z'=0.7121
NL/FF (Relative to siNT)
NL/FF [(xX-xNT)/σNT]
siRNA
siX (μNT-3σNT<siX<μNT+3σNT)
siX (≤μNT-3σNT)
siX (≥μNT+3σNT)
siNL
siNT
D
R² = 0.9731
Log2 [GA70 NL/AUG-FF (siX/siNT)], Plate 2
Log2 [GA70 NL/AUG-FF (siX/siNT)], Plate 1
R² = 0.9644
Log2 [GA70 NL/AUG-FF (siX/siNT)], Plate 3
Log2 [GA70 NL/AUG-FF (siX/siNT)], Plate 1

indicating high robustness and replicability (Figure 3.1C). Each siRNA was transfected in triplicate, with a high correlation between replicates (Figure 3.1D).

We determined primary hits with the algorithm detailed (Figure 3.2). We first excluded from further analysis any genes whose knockdown resulted in reduced cell viability, i.e., genes with FF signal < $\mu_{PLK1}+\sigma_{PLK1}$ (Figure 3.1B, right panel). Second, we counted as primary hits any genes which met the following criteria. Genes were included as suppressors of *C9ORF72* RAN translation if 1) knockdown reduced NL/FF values to less than the mean NL/FF of siNT-treated cells minus 3 standard deviations ($NL/FF_{siX} \leq \mu_{NT}-3\sigma_{NT}$) in at least 2 of 3 replicates; 2) knockdown reduced the mean NL value to below 50% of the mean NL value of siNT-transfected cells ($NL_{siX}/NL_{siNT} \leq 0.50$);

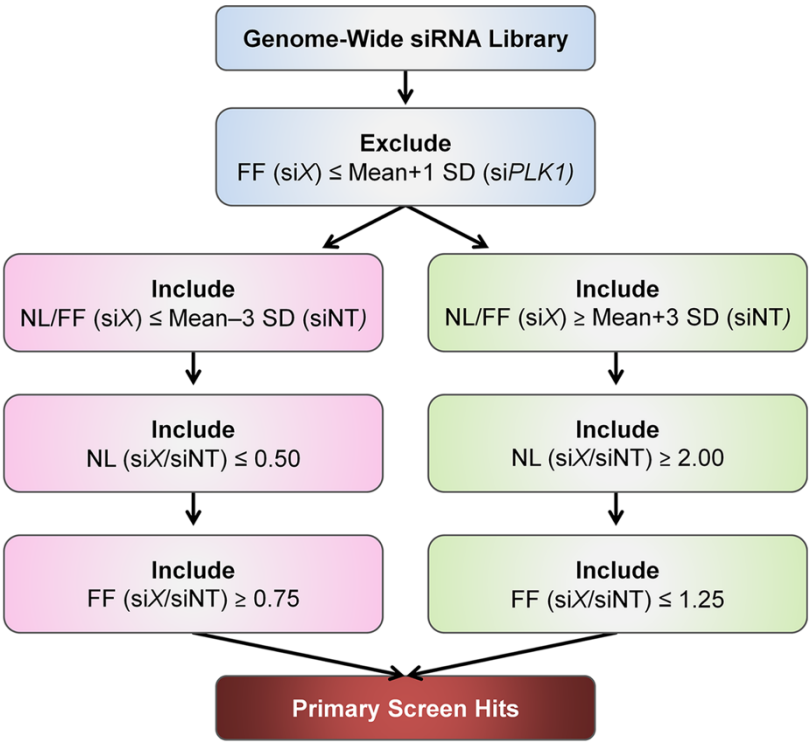


Figure 3.2: Schematic of the Algorithm by Which RAN-Specific Suppressors and Enhancers of GA_{70} NL-3xF Were Determined During the Primary Screen.

and 3) knockdown *did not* reduce the mean FF value to below 75% of the mean FF value of siNT-transfected cells ($FF_{siX}/FF_{siNT} \geq 0.75$). In contrast, genes were included as enhancers of *C9ORF72* RAN translation if 1) $NL/FF_{siX} \geq \mu_{NT}+3\sigma_{NT}$ in at least 2 of 3 replicates; 2) $NL_{siX}/NL_{siNT} \geq 2.0$; and 3) $FF_{siX}/FF_{siNT} \leq 1.25$.

Of 18,110 genes analyzed this algorithm identified 1,624 (8.97%) suppressors and 119 (6.57%) enhancers of *C9ORF72* RAN translation (Figure 3.3A; Figure B.1; Table B.1; Table B.2; Table B.3; Table B.4). To increase our confidence that this set represents replicable, RAN-specific modifiers of translation, we performed both a validation screen using the same GA_{70} NL-3xF reporter as well as a counter screen using an AUG-initiated NL-3xF reporter with a short, unstructured 5' UTR (AUG-NL-3xF), which we have previously used as a reporter for general translation[150,169].

In the *C9ORF72* validation screen, we transfected cells with GA_{70} NL-3xF and AUG-FF reporter plasmids, as well as siRNAs targeting each primary hit. Hit criteria were slightly less stringent: NL/FF-suppressing siRNAs were included if $NL_{siX}/NL_{siNT} \leq 0.75$, and NL/FF-enhancing siRNAs were included if $NL_{siX}/NL_{siNT} \geq 1.5$. All other criteria remained identical. Primary hits validated at a high rate (74.99% total), with 1,228 (6.8% of the whole genome) of suppressors and 79 (0.44% of the whole genome) of enhancers replicating as hits (Figure B.2, top panel). Primary and validation GA_{70} NL values for each gene were significantly correlated (Figure B.3A), demonstrating the high replicability of our assay. In addition, we find it noteworthy that this validation screen chiefly served to eliminate those primary hits that affected AUG-FF expression. siRNAs that failed to validate exhibited NL/FF ratios closer to that of siNT-transfected cells in the validation screen (Figure B.3A), an effect driven by changes in AUG-FF (Figure B.3B,

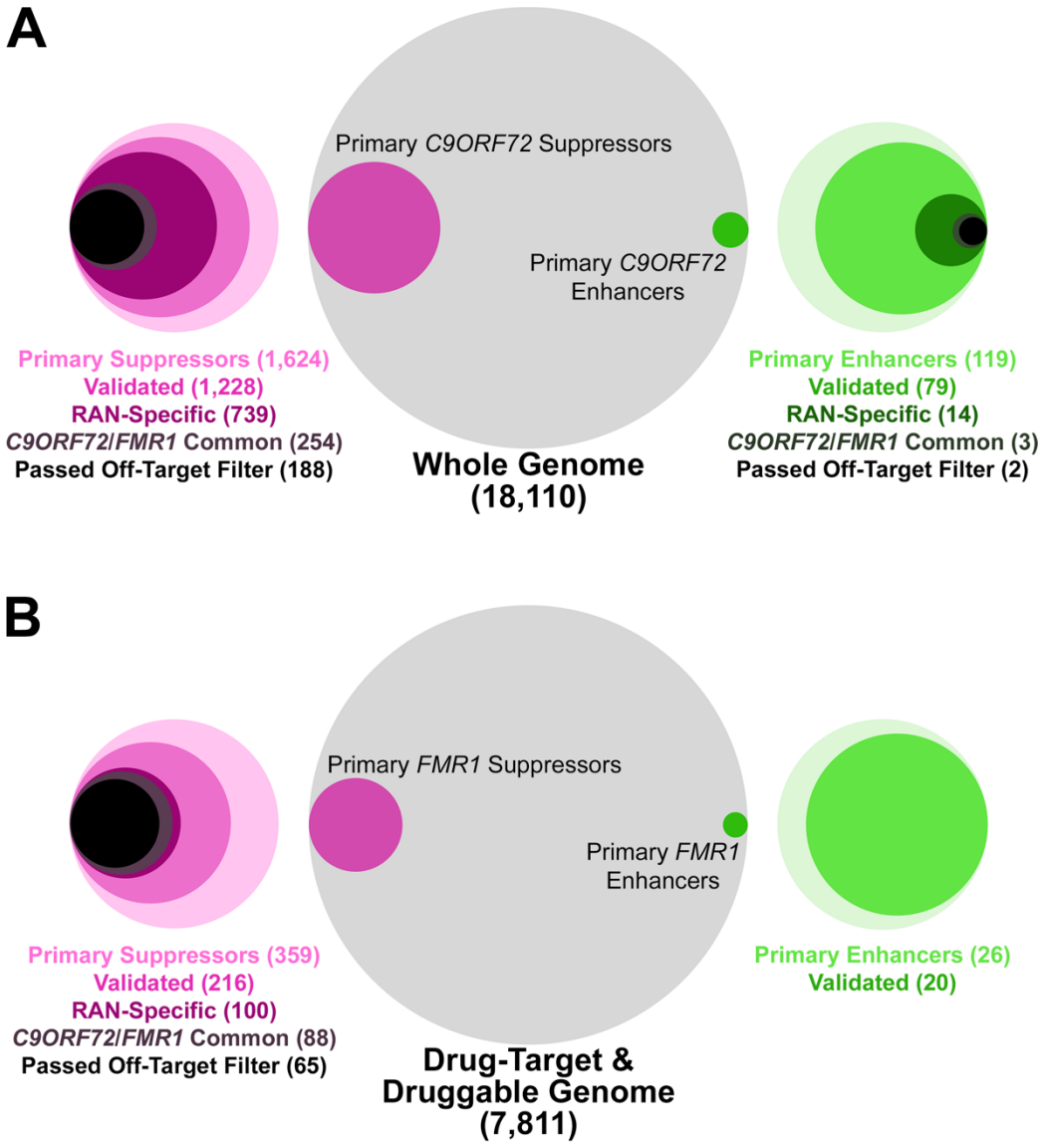


Figure 3.3: Serial Screens Identify RAN-Specific Modifiers of *C9ORF72* and *FMR1* RAN Translation.

A, B To-scale illustration of the numbers of genes identified as selective suppressors and enhancers of GA_{70} NL-3xF (A) and +1 $(CGG)_{100}$ NL-3xF (B) following the primary and subsequent validation, AUG-counter, and cross-reporter validation screens, as well as the off-target filter.

C).

In the AUG counter screen, we transfected cells with AUG-NL-3xF and AUG-FF plasmids, as well as siRNAs against all primary hits. *C9ORF72* suppressors that also

suppressed AUG-NL-3xF ($NL_{siX}/NL_{siNT} \leq 0.80$) and *C9ORF72 e*nhancers that also enhanced AUG-NL-3xF ($NL_{siX}/NL_{siNT} \geq 1.2$) were eliminated, leaving 739 (4.1% of the whole genome) suppressors and 14 (0.08% of the whole genome) enhancers (Figure B.2, middle panel). This attrition indicates that despite filtering our primary screen data using AUG-FF, numerous modifiers of general gene expression had escaped elimination. Finally, we tested each validated, RAN-specific *C9ORF72* modifier with a reporter for RAN translation at expanded CGG repeats in the *FMR1* 5' UTR: +1 $(CGG)_{100}$ NL-3xF. GA_{70} NL values and +1 $(CGG)_{100}$ NL values were significantly correlated (Figure B.4A, left panel), even after we eliminated siRNAs that modified AUG-NL-3xF expression (Figure B.4B). Through cross-validation with the *FMR1* RAN reporter, we identified 254 RAN translation-specific suppressors of both *C9ORF72* and *FMR1* and 3 enhancers. In brief, this set comprised genes which, following knockdown, reliably modified the expression of two different RAN products of two different NREs in two different sequence contexts.

In reviewing the outcomes of the *C9ORF72* validation and AUG counter screens, we noted that the frequency of validated *C9ORF72* modifiers that also validated for *FMR1* would have been higher had non-specific modulators not been eliminated from the datasets by the AUG counter screen. The strongest suppressors of both *C9ORF72* and *FMR1* reporters also suppressed AUG-NL-3xF, resulting in their elimination (Figure B.4A, left panel), and the correlation between relative GA_{70} NL-3xF expression and relative AUG-NL-3xF expression among all *C9ORF72*-validated siRNAs, including those the counter screen would eliminate, was significant (Figure B.4A, right panel). Our screen design nevertheless permitted identification of siRNAs that modified both

C9ORF72 and *FMR1* reporters without affecting AUG-NL-3xF; there remained a significant correlation between GA_{70} NL and +1 CGG_{100} NL expression for each siRNA, even after elimination of siRNAs that impacted AUG-NL expression (Figure B.4B). This perspective supports our decision to construct and conduct these screens in the manner we did.

Analogous Screens Identify Modifiers of *FMR1* RAN Translation

In parallel, we conducted a series of screens to identify specific modifiers of *FMR1* RAN translation. For these we replaced the GA_{70} NL-3xF reporter of the *C9ORF72* screens with +1 $(CGG)_{100}$ NL-3xF, translation of which yields the polyglycine RAN product FMRpolyG^{150}. In the primary *FMR1* screen we screened only the Drug Target and Druggable Genome subsets of the siGENOME siRNA library (Dharmacon). These enrich for identified and characterized targets of pharmacologic agents as well as kinases, phosphatases, proteases, ion channels, and others which are potentially amenable to pharmacological inhibition. *FMR1* screens followed the same high-throughput format, hit-detection algorithm, and validation/counter-screen design described above for *C9ORF72* screens (Figure 3.2).

Of 7,811 siRNAs tested, we identified 359 (4.60%) primary suppressors and 26 (0.33%) primary enhancers of *FMR1* RAN translation (Figure 3.3B; Figure B.5; Table B.5; Table B.6; Table B.7). In the subsequent validation screen 216 (2.77% of the primary siRNA library) suppressors and 20 (0.26%) enhancers validated (Figure B.6). Of these, 100 (1.28% of the siRNA library) suppressors and 0 enhancers remained following the AUG counter screen. Finally, 88 (1.13%) so-validated *FMR1*-suppressing

siRNAs cross-validated against the GA_{70} NL-3xF reporter. While this number is lower than the number of validated, AUG-cleared *C9ORF72* modifiers, it actually represents a larger fraction of the primary +1 $(CGG)_{100}$ NL-3xF suppressors (88/359, 24.51%) than does the fraction of primary GA_{70} suppressors that validated, cleared the AUG counter screen, and cross-validated against +1 $(CGG)_{100}$ NL-3xF (286/1,624, 17.61%). As in the *C9ORF72*-modifier set, primary and validation +1 $(CGG)_{100}$ NL values were significantly correlated (data not shown).

There is Significant Overlap Between *C9ORF72* and *FMR1* RAN-Translation Modifiers

The frequency of siRNAs that dual-validated for *C9ORF72* and *FMR1* hint that there is significant overlap in the biological pathways that regulate RAN translation at these loci. Comparison of the primary *C9ORF72* and *FMR1* modifiers revealed that 198 of the *C9ORF72* (198/647, 28.6%) and *FMR1* hits (198/359, 51.4%) in the Drug-Target and Druggable-Genome library either suppressed or enhanced both RAN translation reporters (Figure 3.4A, B). Overlap was reduced following validation and AUG counter screens [to 58 of the *C9ORF72* (58/380, 15.1%) and *FMR1* hits (58/100, 58%)], but remained significant (Figure 3.4C, D). With analysis restrained to the Drug-Target/Druggable-Genome *C9ORF72* primary hits, siRNAs that likewise modified *FMR1* validated against GA_{70} NL-3xF and cleared the AUG counter screen with significantly less frequency than siRNAs that did not modify *FMR1* (Figure 3.4E). In contrast, among primary *FMR1* modifiers, siRNAs that likewise modified *C9ORF72* validated against +1 $(CGG)_{100}$ NL-3xF with significantly more frequency than siRNAs that did not modify *C9ORF72*. The results of this analysis were the same when we considered only

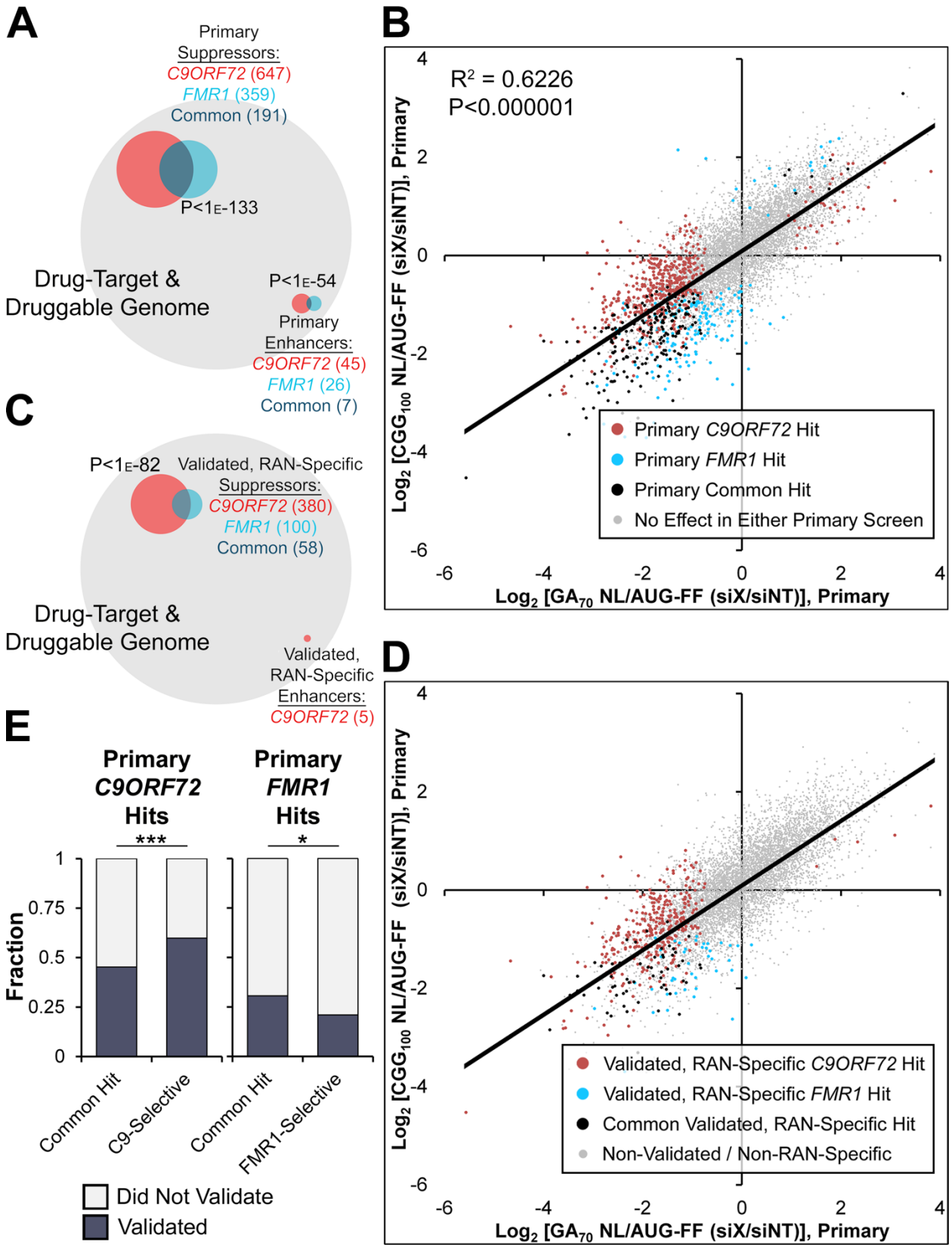

A
Primary
Suppressors:
C9ORF72 (647)
FMR1 (359)
Common (191)
P<1E-133
Drug-Target &
Druggable Genome
P<1E-54
Primary
Enhancers:
C9ORF72 (45)
FMR1 (26)
Common (7)
B
R2 = 0.6226
P<0.000001
Log2 [CGG100 NL/AUG-FF (siX/siNT)], Primary
Log2 [GA70 NL/AUG-FF (siX/siNT)], Primary
Primary C9ORF72 Hit
Primary FMR1 Hit
Primary Common Hit
No Effect in Either Primary Screen
C
P<1E-82
Validated, RAN-Specific
Suppressors:
C9ORF72 (380)
FMR1 (100)
Common (58)
Drug-Target &
Druggable Genome
Validated,
RAN-Specific
Enhancers:
C9ORF72 (5)
D
Validated, RAN-Specific C9ORF72 Hit
Validated, RAN-Specific FMR1 Hit
Common Validated, RAN-Specific Hit
Non-Validated / Non-RAN-Specific
E
Primary
C9ORF72
Hits
Primary
FMR1
Hits
***
*
Fraction
Common Hit
C9-Selective
Common Hit
FMR1-Selective
Did Not Validate
Validated

Figure 3.4: Identification of a Subset of Genes that Modifies Both *C9ORF72* and *FMR1* RAN Translation.

A To-scale illustration of the number of siRNAs identified in the primary screens for *C9ORF72* and *FMR1* modifiers, including common hits (χ^2 test).

B Scatterplot of Log_2 NL/FF ratios by siRNA across the *C9ORF72* vs. *FMR1* primary screens of the Drug-Target/Druggable Genome siRNA library. siRNAs have been categorized based on identification as a hit in either/both primary screens.

C To-scale illustration of the number of siRNAs identified in the primary screens for *C9ORF72* and *FMR1* modifiers that validated and were RAN-specific, including common hits (χ^2 test).

D Scatterplot of Log_2 NL/FF ratios by siRNA across the *C9ORF72* vs. *FMR1* primary screens of the Drug-Target/Druggable Genome siRNA libraries, categorized based on identification as a hit following primary, validation, and AUG counter screens.

E Percentage of primary *C9ORF72* (left panel) and *FMR1* (right panel) hits that validated and cleared the AUG counter screen, based on whether siRNAs modified both *C9ORF72* and *FMR1* or either alone (χ^2 test; * P≤0.05, *** P≤0.001).

suppressors of *C9ORF72* and *FMR1* (data not shown). This suggests that our *C9ORF72* primary screen was more selective for RAN-specific modifiers than was our *FMR1* primary screen.

Eliminating Genes with Low Expression in HEK293 Cells Reduces False Positives Due to Off-Target Effects of siRNAs

Given the high expected frequency of false positives in all siRNA screens due to off-target effects[284-286], we conducted a pilot experiment to estimate this frequency in our sets of candidate RAN modifiers. Of 156 independent-library (Dharmacon ON-TARGET*plus*) siRNA pools targeting our siGENOME-validated modifiers of *C9ORF72* and/or *FMR1*, only 100 (64%) cleared an AUG counter screen, of which only 30 (19%) had the same effects as siGENOME siRNAs on GA_{70} NL-3xF or +1 $(CGG)_{100}$ NL-3xF (Table 3.1). The 156 genes so tested included 41 dual-primary and dual-validated RAN modifiers, which validated at approximately the same frequency (8/41, 20%) as the other modifiers tested. We can offer three potential explanations for this low validation

Candidate Set	Number of ON-TARGET*plus* Candidates*						
	Submitted	Passed AUG	Validated Against RAN		Passed AUG and Validated Against RAN		
			C9ORF72	*FMR1*	*C9ORF72*	*FMR1*	Both
C9ORF72 Primary Suppressors	143	91 (64%)	40 (28%)	26 (18%)	28 (20%)	13 (9%)	12 (8%)
+ *FMR1* Validation	71	47 (66%)	20 (28%)	14 (20%)	15 (21%)	8 (11%)	8 (11%)
FMR1 Primary Suppressors	49	31 (63%)	11 (22%)	10 (20%)	8 (16%)	6 (12%)	5 (10%)
+ *C9ORF72* Validation	46	30 (65%)	11 (24%)	10 (22%)	8 (17%)	6 (13%)	5 (11%)
C9ORF72/*FMR1* Dual Primary Suppressors**	41	26 (63%)	11 (27%)	9 (22%)	8 (20%)	5 (12%)	5 (12%)
C9ORF72 Primary Enhancers	5	4 (80%)	0	0	0	0	0
Total Candidates	156	100 (64%)	40 (26%)	27 (17%)	28 (18%)	14 (9%)	12 (8%)

Table 3.1: Validation of RAN-Translation Modifiers Using the ON-TARGET*plus* siRNA Library.
Data Information: * All percentages reflect fractions of the indicated candidate set that were submitted to ON-TARGET*plus* testing. ** These candidates were independently identified during both primary screens and subsequently validated against both GA_{70} and $FMRpolyG_{100}$ reporters using siGENOME siRNAs.

frequency: 1) off-target effects in the original, siGENOME library generated false positives against *C9ORF72*-/*FMR1*-RAN expression in the primary and validation screens; 2) ON-TARGET*plus* siRNAs are less efficacious, generating false negatives against *C9ORF72*/*FMR1* reporters in this pilot experiment; or 3) off-target effects or greater efficacy of knockdown in the second, ON-TARGET*plus* library impacted AUG-NL-3xF expression, leading to erroneous elimination of true positives. We favor the first explanation, since more candidates failed to validate against the *C9ORF72*-/*FMR1*-RAN reporters than were eliminated in the ON-TARGET*plus* AUG-counter screen, and we (admittedly anecdotally) have not observed failure of siRNA-mediated knockdown with the ON-TARGET*plus* library in other experiments.

We therefore sought to reduce the frequency of false positives by identifying and eliminating those genes that are unexpressed or minimally expressed in HEK293 cells—if gene *X* is not expressed in HEK293 cells, any effects of anti-*X* siRNAs likely reflect off-target interaction(s)[287]. We therefore selected 4 publicly available RNA sequencing (RNA-Seq) libraries collected from HEK293 cells with minimal genetic or pharmacological manipulation—essentially, control sets for their respective experiments[288-291]. Each set exhibited a multimodal distribution of mRNA expression across genes [transcripts per million (TPM)], with one mode near 0 TPM, corresponding to no or minimal expression. Within each set, we identified the TPM value at the inter-modal nadir and set that value as a threshold denoting sufficient expression. This

RNA-Seq Library	Number of Total Transcripts Detected	Transcripts Below Expression Threshold (% of Total)
GSE103719	23,388	7,867 (33.6%)
GSE113550	23,615	12,398 (52.5%)
GSE79133	17,675	6,711 (38.0%)
GSE52447	23,308	7,266 (31.2%)

Table 3.2: RNA Sequencing (RNA-Seq) Libraries Used to Construct the Off-Target Filter.

method eliminated 31-53% of genes of each individual RNA-Seq set (Table 3.2). We then compared the four RNA-Seq libraries (Table B.8), and those genes which were beneath threshold in ≥3 libraries were eliminated from our sets of *C9ORF72* (163/739 removed, 23%) and *FMR1* RAN-translation modifiers (25/100 removed, 25%), as delineated in Table 3.3. As crude proof of concept, application of this filter to the list of 156 candidates tested using ON-TARGET*plus* siRNAs resulted in preferential filtration of candidates that failed to validate: whereas 33/126 (26%) of those candidates that failed ON-TARGET*plus* validation were eliminated by this off-target filter, significantly fewer—only 3/30 (10%)—of candidates that successfully validated were eliminated (Table 3.3). Our pilot results imply that this filter reduced the number of false positives due to off-target siRNA-mediated knockdown, thus further refining our list of modifiers of *C9ORF72*, *FMR1*, or both (Table B.9; Table B.10; Table B.11; Table B.12; Table B.13).

Gene Ontology Analysis Reveals Biological Processes That Modulate RAN Product Abundance

Table 3.3: Off-Target Filtration of Candidate Modifiers by Modifier Class.
Data Information: * All submitted genes in this column passed the ON-TARGET*plus* AUG-counter screen and validated against either *C9ORF72* or *FMR1*. ** These candidates were independently identified during both primary screens and subsequently validated against both GA_{70} and FMRpolyG_{100} reporters. ^ P=0.0437 (χ^2 test).

Candidate Set	Number of Gene Candidates		Number of Genes Validated by On-TARGET*plus* siRNAs*		Number of Genes That Failed On-TARGET*plus* Validation	
	Submitted	Passed	Submitted	Passed	Submitted	Passed
C9ORF72 Primary Suppressors	739	544 (74%)	29	26 (90%)		
+ *FMR1* Validation	254	188 (74%)	15	14 (93%)		
FMR1 Primary Suppressors	100	75 (75%)	9	8 (89%)		
+ *C9ORF72* Validation	88	65 (74%)	9	8 (89%)		
C9ORF72/*FMR1* Dual Primary Suppressors**	58	42 (72%)	8	7 (88%)		
C9ORF72 Primary Enhancers	14	8 (57%)				
+ FMR1 Validation	3	2 (66%)				
Total Candidates	795	585 (74%)^	30	27 (90%)^	126	93 (74%)

Next, we sought to identify biological pathways and molecular functions common to modifiers of RAN translation of *C9ORF72*, *FMR1*, or both using gene ontology (GO) analysis. The 585 genes remaining after off-target filtration were analyzed using Advaita Bio's iPathwayGuide in collaboration with the Bioinformatics Core at the University of Michigan. This software analysis tool implements the Impact Analysis approach that takes into consideration the direction and type of all signals on a pathway, as well as the position, role, and type of every gene[292-295]. This analysis identified biological processes and molecular functions enriched among suppressors of *C9ORF72* (Figure B.7A, B) and *FMR1* (Figure B.8A, B) RAN translation. Among *C9ORF72* RAN suppressors, GO terms of notable interest included proteasome-mediated protein degradation ("SCF-dependent proteasomal ubiquitin-dependent protein catabolic process," P=4.60E-5; and "Proteasome-activating ATPase activity," P=1.26E-2), NFκB/TNF signaling ("NIK/NF-κB signaling," P=1.50E-4; and "TNF-mediated signaling pathway," P=5.80E-4), "mRNA methyltransferase" (P=1.73E-2), "C-terminal protein methylation" (P=8.9E.4), "Golgi-to-lysosome transport" (P=1.32E-3), and "Positive regulation of canonical Wnt signaling pathway" (P=3.60E-4). In parallel, GO terms of interest among *FMR1* RAN suppressors included "IRES-dependent viral translational initiation" (P=8.5E-4) and "Chaperone-binding" (P=7.50E-3). Each of these terms has been linked to either neurodegeneration in general, autosomal dominantly inherited variants of ALS[296], non-canonical modes of mRNA translation initiation, or functions of known modifiers of RAN translation (see Discussion).

To our surprise, despite the significant overlap between our sets of *C9ORF72* and *FMR1* RAN translation modifiers (Figure 3.4C, D), the 10 most statistically

significant GO terms for each set were typically not enriched in the other. Among biological processes only “IRES-dependent viral translational initiation” and “Vagina development” were common to both sets; among molecular functions only “ATP-binding” and “Protein serine/threonine kinase activity” were common to both sets. This distinction might be attributed to “true” biological deviation in mechanism, differences in NL-3xF reporter design (the *FMR1* reporter had a PEST tag to optimize its linear range, whereas the *C9ORF72* reporter did not), or the difference in genes screened.

Depletion of hnRNPA$_2$/B$_1$ Suppresses *C9ORF72* and *FMR1* RAN Reporters

hnRNPA$_2$/B$_1$ is an RNA-binding protein that, in humans and mice, regulates pre-mRNA splicing and polyadenylation[190,297,298]. In addition, hnRNPA$_2$/B$_1$ is linked to neurological disease: the D290V mutation is associated with multisystem proteinopathy (related to ALS/FTD)[140], hnRNPA$_2$/B$_1$ binds expanded CGG repeats in *FMR1*[137-139] and recombinant GR$_{50}$ and PR$_{50}$[195], and overexpression of human *HNRNPA2B1* or its *Drosophila* orthologs suppresses (CGG)$_n$ toxicity in flies[141]. Therefore, we were intrigued to find that knockdown of *HNRNPA2B1* suppressed the expression of both *C9ORF72* and *FMR1* RAN products in our screens. In follow-up luciferase assays and western blots, *HNRNPA2B1* knockdown using individual and pooled siRNAs significantly and selectively reduced the expression of +1 (CGG)$_{100}$ NL-3xF and GA$_{70}$ NL-3xF in HEK293 cells (Figure 3.5A, B). In contrast, AUG-NL-3xF was not suppressed by *HNRNPA2B1* knockdown. In the inverse experiment, overexpression of *HNRNPA2B1* in HeLa cells had the opposite effects, selectively enhancing the expression of +1 (CGG)$_{100}$ NL-3xF and GA$_{70}$ NL-3xF (Figure 3.5C). In addition, *HNRNPA2B1* knockdown reduced the

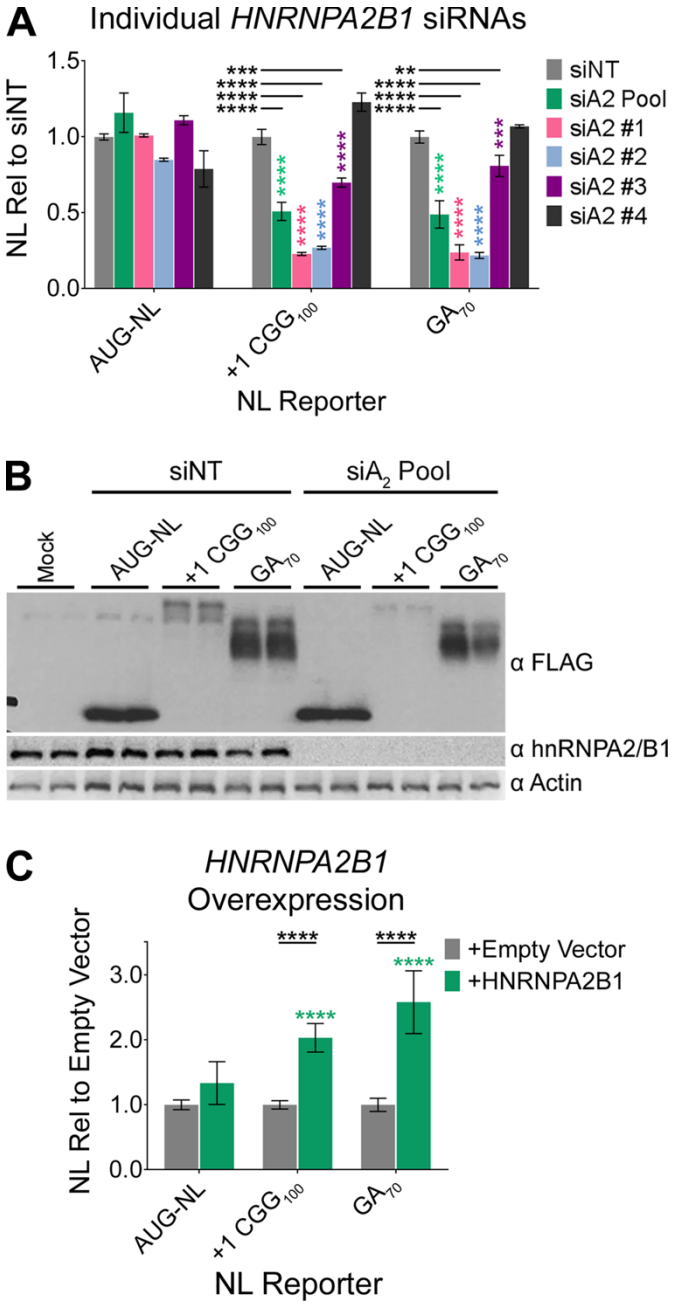


Figure 3.5: Knockdown of *HNRNPA2B1* Selectively Inhibits RAN Translation of G_4C_2-Expanded *C9ORF72* and CGG-Expanded *FMR1*, Part 1.

A, B Expression of transfected, plasmid-based NL-3xF reporters by luminescence assay (A) and western blot (B) in HEK293 cells following transfection of siNT or pooled and individual siRNAs against *HNRNPA2B1* (two-way ANOVA with Tukey's multiple comparisons test; *n*=6/condition). Black asterisks indicate comparisons between siNT- and siHNRNPA2B1-transfected cells expressing the same reporter; colored asterisks indicate comparisons between siHNRNPA2B1-transfected cells expressing AUG-NL-3xF and those expressing either GA_{70} NL-3xF or +1 $(CGG)_{100}$ NL-3xF.

C Expression of transfected, plasmid-based NL-3xF reporters by luminescence assay in HeLa cells following overexpression of *HNRNPA2B1* or empty vector (two-way ANOVA with Tukey's

multiple comparisons test; *n*=9/condition). Black asterisks indicate comparisons between empty vector- and *HNRNPA2B1*-transfected cells expressing the same reporter; green asterisks indicate comparisons between *HNRNPA2B1*-transfected cells expressing AUG-NL-3xF and those expressing either GA_{70} NL-3xF or +1 $(CGG)_{100}$ NL-3xF.

Data Information: All panels depict data as mean ± SD (compiled from ≥2 replicates). ** P≤0.01, *** P≤0.001, **** P≤0.0001.

---

expression of the *C9ORF72* glycine-proline (GP_{70} NL-3xF) and glycine-arginine (GR_{70} NL-3xF) DPR RAN products, as well as the *FMR1* alanine RAN product [+2 $(CGG)_{100}$ NL-3xF] (Figure 3.6A), demonstrating that the effects of *HNRNPA2B1* knockdown are not limited to single reading frames and suggesting that our use of single-reading-frame RAN reporter constructs does not hinder extrapolation of our results to other reading frames. Finally, the effects of *HNRNPA2B1* knockdown on +1 $(CGG)_n$ NL-3xF and GA_n NL-3xF reporters were repeat-dependent, whereas the effects of knockdown on +2 $(CGG)_n$ NL-3xF reporters were not (Figure 3.6B-D). The reason for this discrepancy is unclear, but it might represent differences in the site of translation initiation relative to the NRE[19,150]. In total, our results suggest that hnRNPA_2/B_1 is a specific regulator of RAN-product expression.

We next asked whether hnRNPA_2/B_1 acts peri-transcriptionally or peri-translationally, since hnRNPA_2/B_1 has described functions in both pre-mRNA processing[190] and translational regulation[299-302]. We first replaced our plasmid-based reporters with *in vitro*-transcribed RNA reporters for RAN translation. To our surprise, *HNRNPA2B1* knockdown significantly increased the expression of both GA_{70} NL-3xF and +1 $(CGG)_{100}$ NL-3xF RNA reporters (Figure 3.6E) without impacting the expression of AUG-FF RNA reporters. We next asked whether *HNRNPA2B1* knockdown impacted the abundance of mRNAs transcribed from GA_{70} NL-3xF or +1 $(CGG)_{100}$ NL-3xF

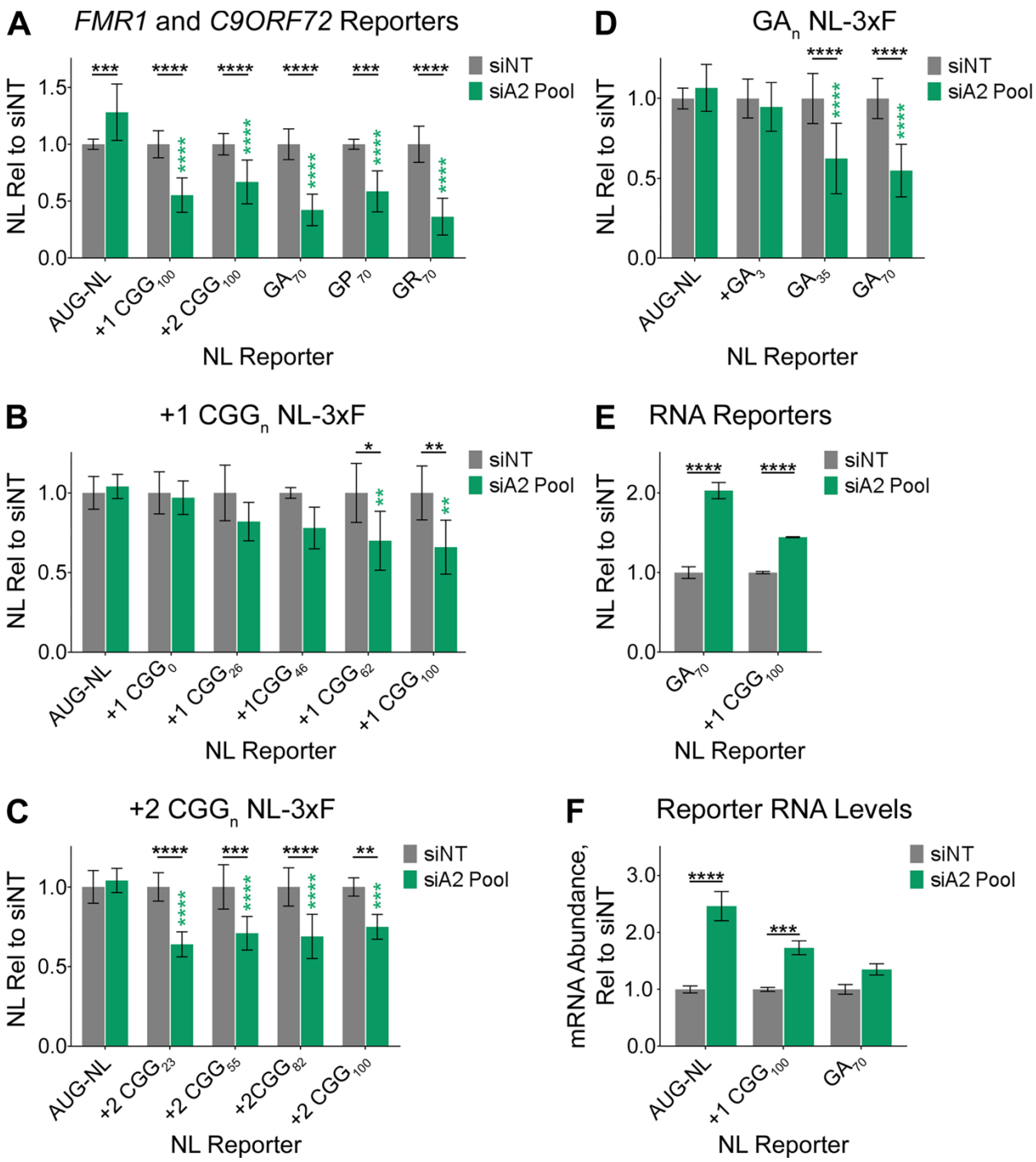


Figure 3.6: Knockdown of *HNRNPA2B1* Selectively Inhibits RAN Translation of G_4C_2-Expanded *C9ORF72* and CGG-Expanded *FMR1*, Part 2.

A Expression of transfected, plasmid-based AUG-NL-3xF and multiple RAN reporters with varied sequence contexts and reading frames. Following transfection of siNT or pooled siRNAs against *HNRNPA2B1*, expression was quantified by luminescence assay in HEK293 cells (two-way ANOVA with Tukey's multiple comparisons test; *n*=6/condition). Black asterisks indicate comparisons between siNT- and siHNRNPA2B1-transfected cells expressing the same reporter;

green asterisks indicate comparisons between siHNRNPA2B1-transfected cells expressing AUG-NL-3xF and those expressing another NL-3xF reporter.
B-D Expression of transfected, plasmid based +1 $(CGG)_n$ NL-3xF (B), +2 $(CGG)_n$ NL-3xF (C), and $(GA)_n$ NL-3xF (D) reporters in HEK293 cells with a range of repeat sizes following *HNRNPA2B1* knockdown (two-way ANOVA with Tukey's multiple comparisons test; *n*=3-9/condition).
E Expression of transfected, RNA-based NL-3xF reporters by luminescence assay in HEK293 cells following *HNRNPA1B1* knockdown (two-way ANOVA with Sidak's multiple comparisons test; *n*=5-6/condition).
F Abundance of NL-3xF reporter mRNAs by qRT-PCR following *HNRNPA2B1* knockdown (two-way ANOVA with Sidak's multiple comparisons test; *n*=5-6/condition).

Data Information: All panels depict data as mean ± SD (compiled from ≥2 replicates). * P≤0.05, ** P≤0.01, *** P≤0.001, **** P≤0.0001.

---

plasmids. Again bucking our expectations, qRT-PCR revealed that *HNRNPA2B1* knockdown significantly increased the abundance of AUG-NL-3xF and +1 $(CGG)_{100}$ NL-3xF mRNAs and slightly increased (non-significantly) GA_{70} NL-3xF mRNA abundance (Figure 3.6F). This increase in mRNA abundance, coupled with the lower abundance of plasmid-encoded RAN products following *HNRNPA2B1* knockdown, implied that hnRNPA_2/B_1 depletion was preventing RAN-competent mRNAs (when transcribed in the nucleus) from engaging the translational machinery. To assess whether *HNRNPA2B1* knockdown alters the cellular distribution of RAN-competent mRNAs, we performed hybridization chain reaction (HCR) to fluorescently label and localize +1 $(CGG)_{100}$ NL-3xF mRNAs in HEK293 cells, with and without *HNRNPA2B1* knockdown. This revealed that +1 $(CGG)_{100}$ NL-3xF mRNAs cluster throughout both the nucleus and cytoplasm (Figure 3.7). Following *HNRNPA2B1* knockdown, +1 $(CGG)_{100}$ NL-3xF mRNA signal increased in both cellular compartments, consistent with the increased abundance of these mRNAs we observed by qRT-PCR. But we observed no gross difference in the distribution of these mRNAs following knockdown. Therefore, the relationship between human hnRNPA_2/B_1 and repeat-expanded mRNAs, and the mechanism by which its

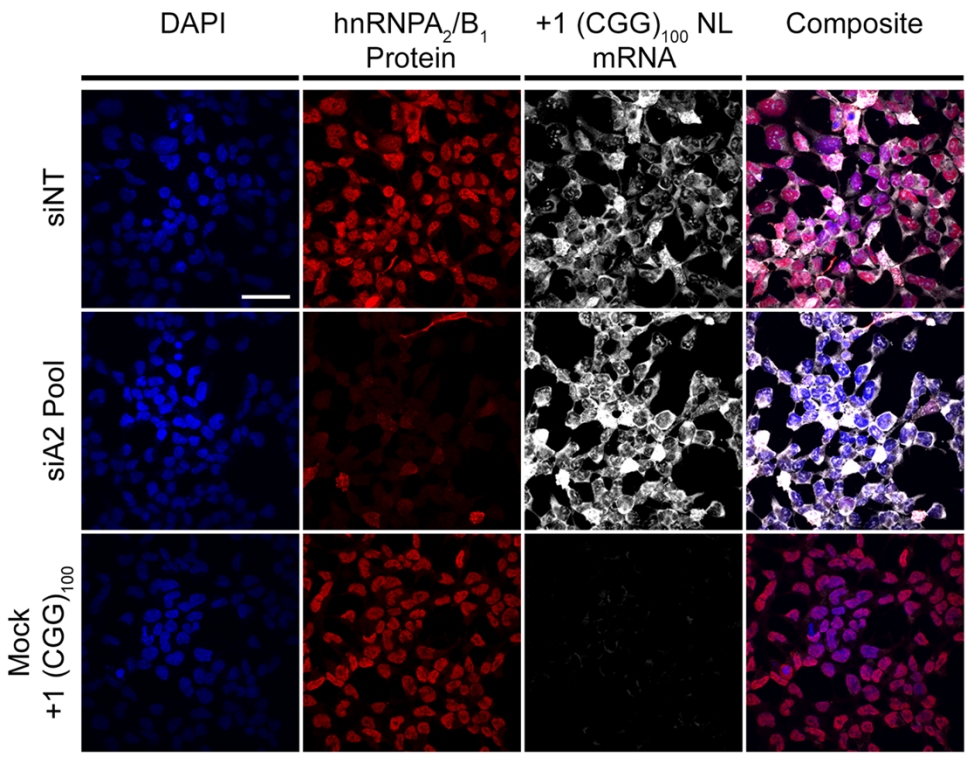


Figure 3.7: Knockdown of *HNRNPA2B1* Does Not Change the Subcellular Localization of +1 $(CGG)_{100}$ NL-3xF mRNA.
Hybridization chain reaction (HCR) against +1 $(CGG)_{100}$ NL-3xF mRNA and immunocytochemistry (ICC) against hnRNPA_2B_1 protein following knockdown of *HNRNPA2B1* in HEK293 cells. Scale bar = 50 µm.

---

depletion suppresses RAN translation of these reporters, remains unclear.

hnRNPA_2/B_1 and its *Drosophila* Homologs Modulate NRE-elicited toxicity, but Not Via RAN Translation

In parallel, we investigated the effects of modulating expression of *HNRNPA2B1* and its homologs in *Drosophila* on NRE-induced toxicity. Sofola *et al*.[139] demonstrated that overexpression of human *HNRNPA2B1* or its fly homologs *HRB87F* and *HRB98DE* mitigated the toxicity of expanded CGG repeats in *FMR1*, while knockdown of *HRB87F* or *HRB98DE* potentiated toxicity. We replicated these results using flies that expressed +1 $(CGG)_{90}$-EGFP under a GMR-GAL4 driver (Figure 3.8A). Neither overexpression of

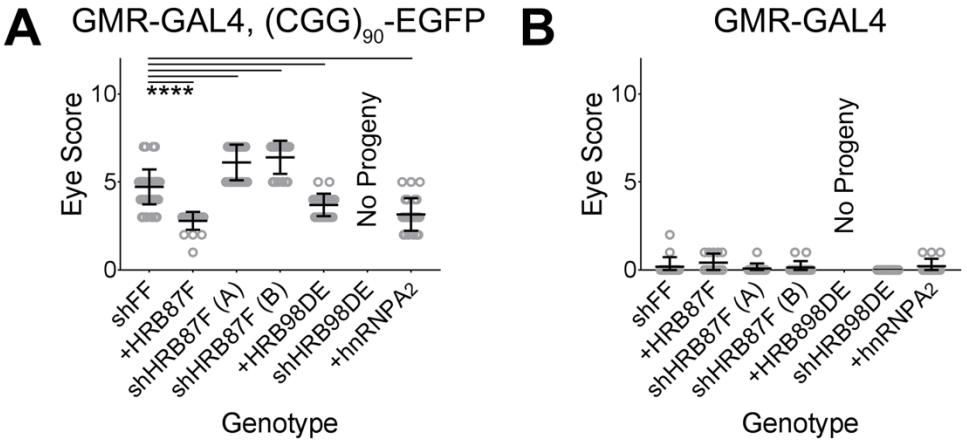


Figure 3.8: Modulation of *HNRNPA2B1* and its Homologs Bidirectionally Modifies $(CGG)_{90}$ Toxicity in *Drosophila*.

A Quantification of rough-eye phenotypes in flies expressing +1 $(CGG)_{90}$-EGFP under the GMR-GAL4 driver, along with various disruptions of *HNRNPA2B1*'s fly homologs (one-way ANOVA with Dunnett's multiple comparisons test; *n*=20-47/genotype).

B Quantification of rough-eye phenotypes in flies expressing various disruptions of *HNRNPA2B1*'s fly homologs under the GMR-GAL4 driver alone.

Data Information: Data are presented as mean ± SD (compiled from ≥3 replicates). **** P≤0.0001.

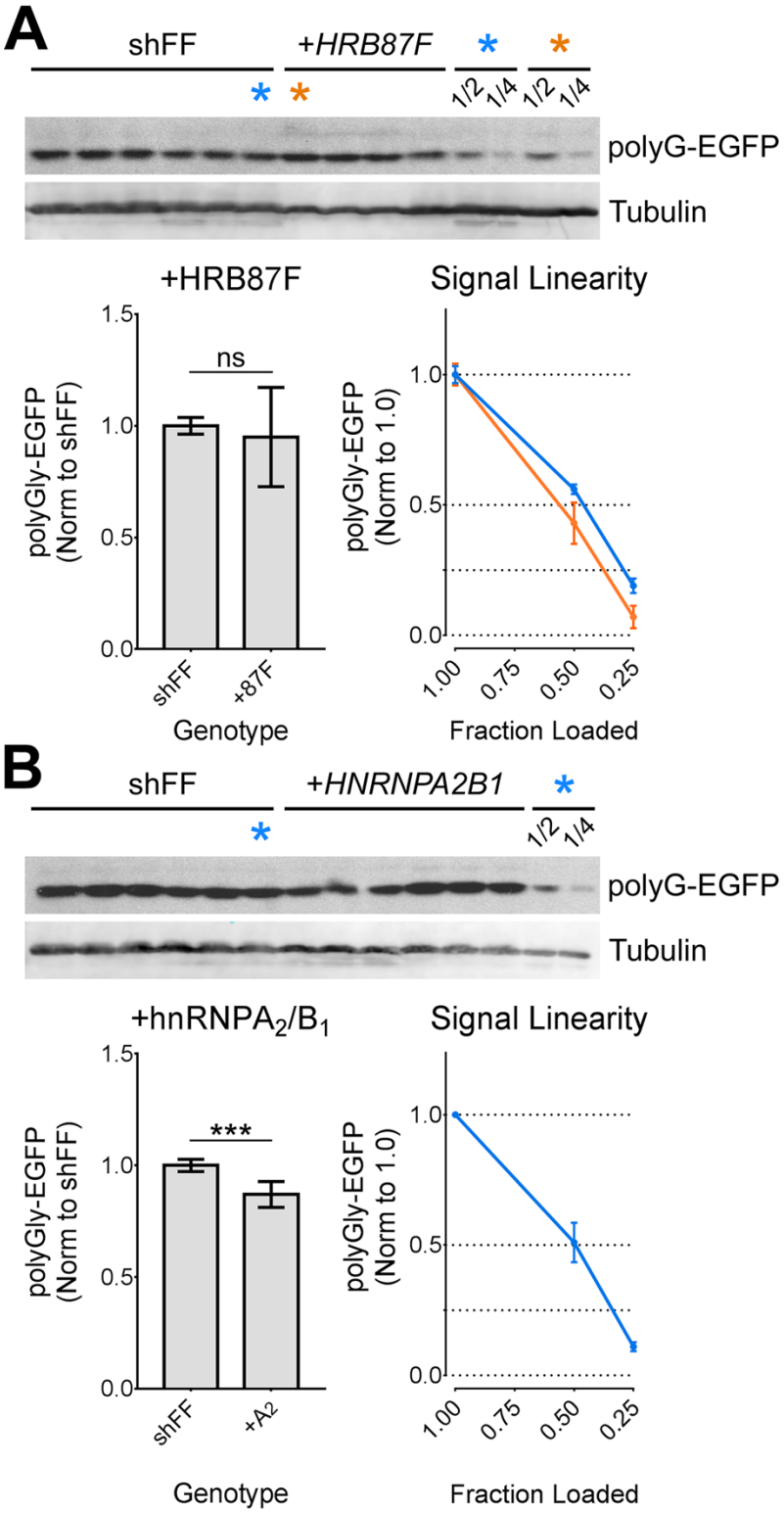
HNRNPA2B1 or *HRB87F* nor knockdown of *HRB87F* or *HRB98DE* affected eye phenotype in flies carrying the GMR-GAL4 driver alone (Figure 3.8B). Given the possibility that *HNRNPA2B1* modulates expression of RAN products, we next asked whether these genetic disruptions modulated expression of the FMRpolyG-EGFP RAN product in +1 $(CGG)_{90}$-EGFP-expressing flies. Western blotting revealed, however, that neither overexpression nor knockdown of *HNRNPA2B1*, *HRB87F*, and *HRB98DE* had consistent effects on FMRpolyG-EGFP (Figure 3.9A, B; Figure 3.10A, B). These data

---

Figure 3.9: Modulation of *HNRNPA2B1* and Its Homologs Has Inconsistent Effects on FMRpolyG Expression in *Drosophila*, Part 1.

A-B Western blots of the FMRpolyG-EGFP RAN product and tubulin in flies expressing +1 $(CGG)_{90}$-EGFP and either shRNAs against firefly luciferase (shFF) or UAS-driven *HRB87F* (A) or UAS-driven human *HNRNPA2B1* (B) under the inducible Tub5-GS driver, with quantification. Colored asterisks above westerns indicate which samples were loaded at ½ and ¼ volume to determine the linear range suitable for quantification (Student's t test; *n*=4-6 bands/genotype).

Data Information: Data are presented as mean ± SD. ns=non-significant, *** P≤0.001.


A
shFF
+HRB87F
1/2
1/4
1/2
1/4
polyG-EGFP
Tubulin
+HRB87F
ns
polyGly-EGFP (Norm to shFF)
1.5
1.0
0.5
0.0
shFF
+87F
Genotype
Signal Linearity
polyGly-EGFP (Norm to 1.0)
1.0
0.5
0.0
1.00
0.75
0.50
0.25
Fraction Loaded
B
shFF
+HNRNPA2B1
1/2
1/4
polyG-EGFP
Tubulin
+hnRNPA2/B1
***
polyGly-EGFP (Norm to shFF)
1.5
1.0
0.5
0.0
shFF
+A2
Genotype
Signal Linearity
polyGly-EGFP (Norm to 1.0)
1.0
0.5
0.0
1.00
0.75
0.50
0.25
Fraction Loaded

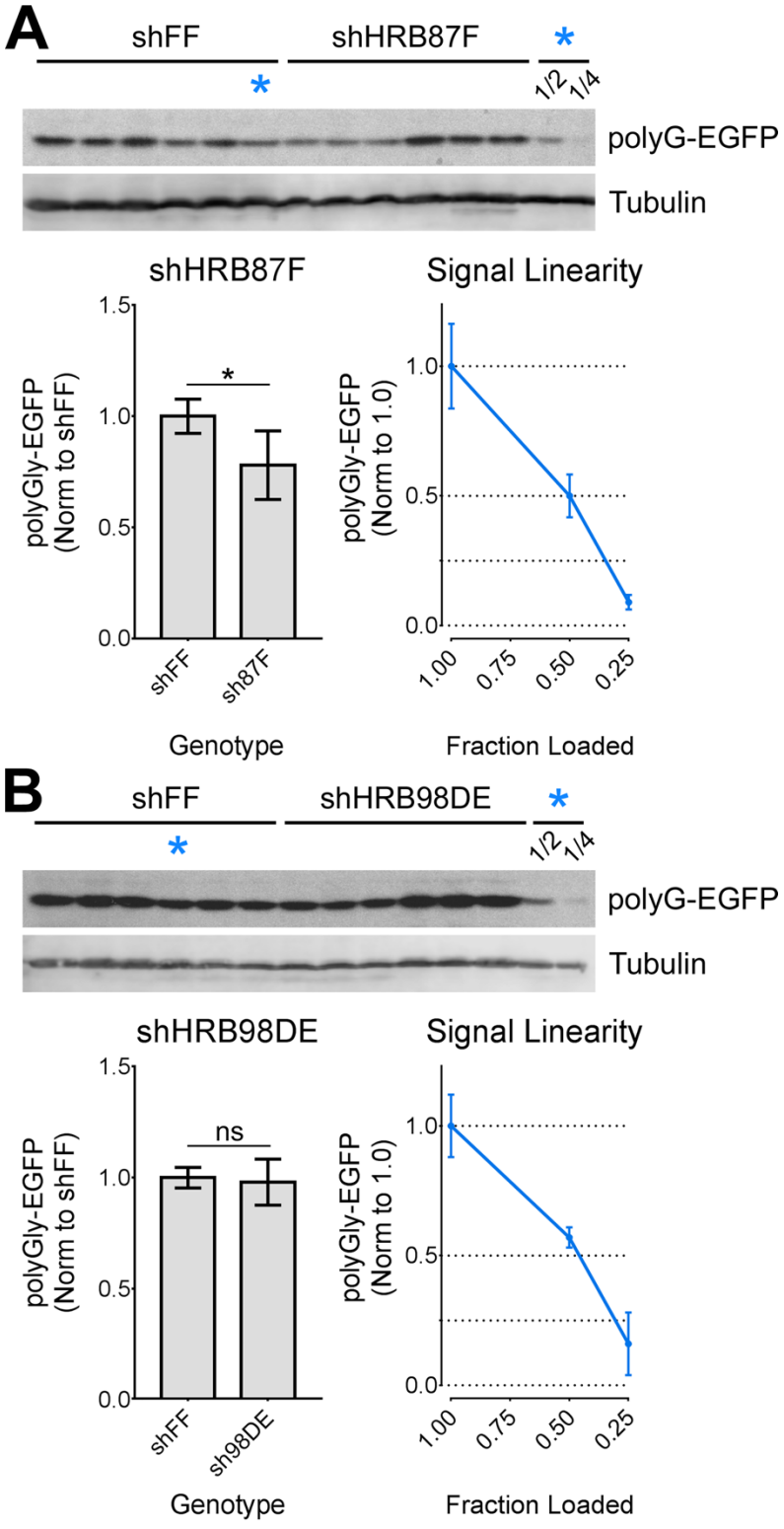

A
shFF
shHRB87F
*
*
1/2
1/4
polyG-EGFP
Tubulin
shHRB87F
Signal Linearity
polyGly-EGFP (Norm to shFF)
1.5
1.0
0.5
0.0
*
shFF
sh87F
Genotype
polyGly-EGFP (Norm to 1.0)
1.0
0.5
0.0
1.00
0.75
0.50
0.25
Fraction Loaded
B
shFF
shHRB98DE
*
*
1/2
1/4
polyG-EGFP
Tubulin
shHRB98DE
Signal Linearity
polyGly-EGFP (Norm to shFF)
1.5
1.0
0.5
0.0
ns
shFF
sh98DE
Genotype
polyGly-EGFP (Norm to 1.0)
1.0
0.5
0.0
1.00
0.75
0.50
0.25
Fraction Loaded

Figure 3.10: Modulation of *HNRNPA2B1* and Its Homologs Has Inconsistent Effects on FMRpolyG Expression in *Drosophila*, Part 2.
A-B Western blots of the FMRpolyG-EGFP RAN product and tubulin in flies expressing +1 $(CGG)_{90}$ EGFP and either shFF or shRNAs against *HRB87F* (A) or against *HRB98DE* (B) under the inducible Tub5-GS driver, with quantification. Colored asterisks above westerns indicate which samples were loaded at ½ and ¼ volume to determine the linear range suitable for quantification (Student's t test; *n*=6 bands/genotype).

Data Information: Data are presented as mean ± SD. ns=non-significant, * P≤0.05.

suggest that *HNRNPA2B1* and its fly homologs modulate the toxicity of CGG repeats via a different mechanism than RAN translation.

In an attempt to resolve the discrepancies between our *Drosophila* and HEK293 findings, we utilized an ELISA to detect the endogenous GP RAN product in a C9 ALS patient-derived iPSC line[163] following *HNRNPA2B1* knockdown (Figure 3.11A). As expected, GP was more abundant in C9 iPSCs compared to a control line. However, we could detect no difference in GP abundance following $hnRNPA_2/B_1$ depletion (Figure 3.11B).

Depletion of PSMB5 Suppresses *C9ORF72* and *FMR1* RAN Reporters

Our screens identified multiple subunits of the proteasome, knockdown of which suppressed expression of *C9ORF72* reporters (Figure B.9, Table B.3, Table B.4, Table B.7). Among these was the proteolytic subunit PSMB5. In parallel, other investigators have discovered that a polymorphism in *PSMB5* that decreases its expression was protective against FXTAS in individuals carrying *FMR1* CGG repeats in the premutation range (P. Jin, personal communication). To confirm the effects of *PSMB5* knockdown we observed in our screen, we first tested an independent pool of anti-*PSMB5* siRNAs (ON-TARGET*plus*, Dharmacon) on multiple reporters for canonical and RAN translation.

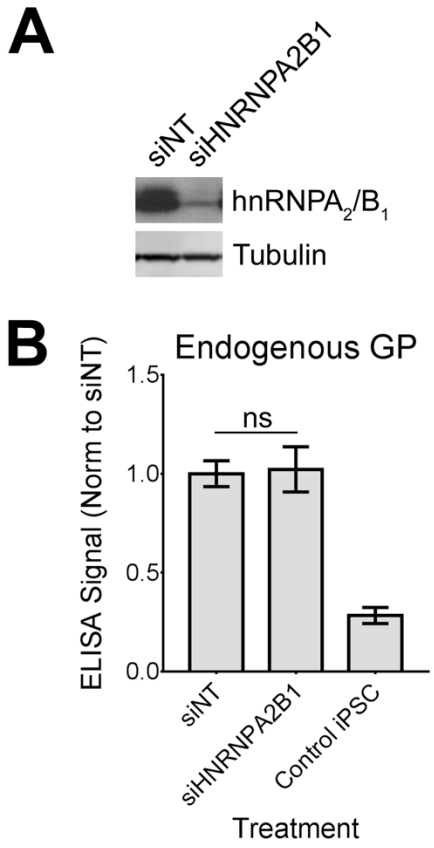


Figure 3.11: Knockdown of *HNRNPA2B1* in C9 iPSCs Does Not Affect GP Expression.
A Anti-hnRNPA$_2$/B$_1$ and anti-tubulin western blots of lysates of iPSCs derived from a C9 ALS patient. iPSCs were transfected with siNT or siRNAs against *HNRNPA2B1*.
B Abundance of the GP RAN product by ELISA in C9 ALS iPSCs or controls, following transfection of siNT or siRNAs against *HNRNPA2B1* (Student's t test; *n*=5/condition). Abundance of the GP RAN product in a control iPSC line is shown for comparison.

Data Information: Data are presented as mean ± SD (compiled from 3 replicates). ns=non-significant.

In our initial studies with this independent siRNA pool, its transfection led to widespread loss of cell viability and a global reduction in reporter expression (Figure 3.12A). To reconcile these results, we tested the effects of the individual ON-TARGET*plus* siRNAs on AUG-NL-3xF, +1 $(CGG)_{100}$ NL-3xF, and GA_{70} NL-3xF (Figure 3.12B) while monitoring *PSMB5* knockdown by western blot (Figure 3.12C). Transfection of siPSMB5 #1 had no effect on any reporter and did not reduce *PSMB5* expression; #2 reproduced

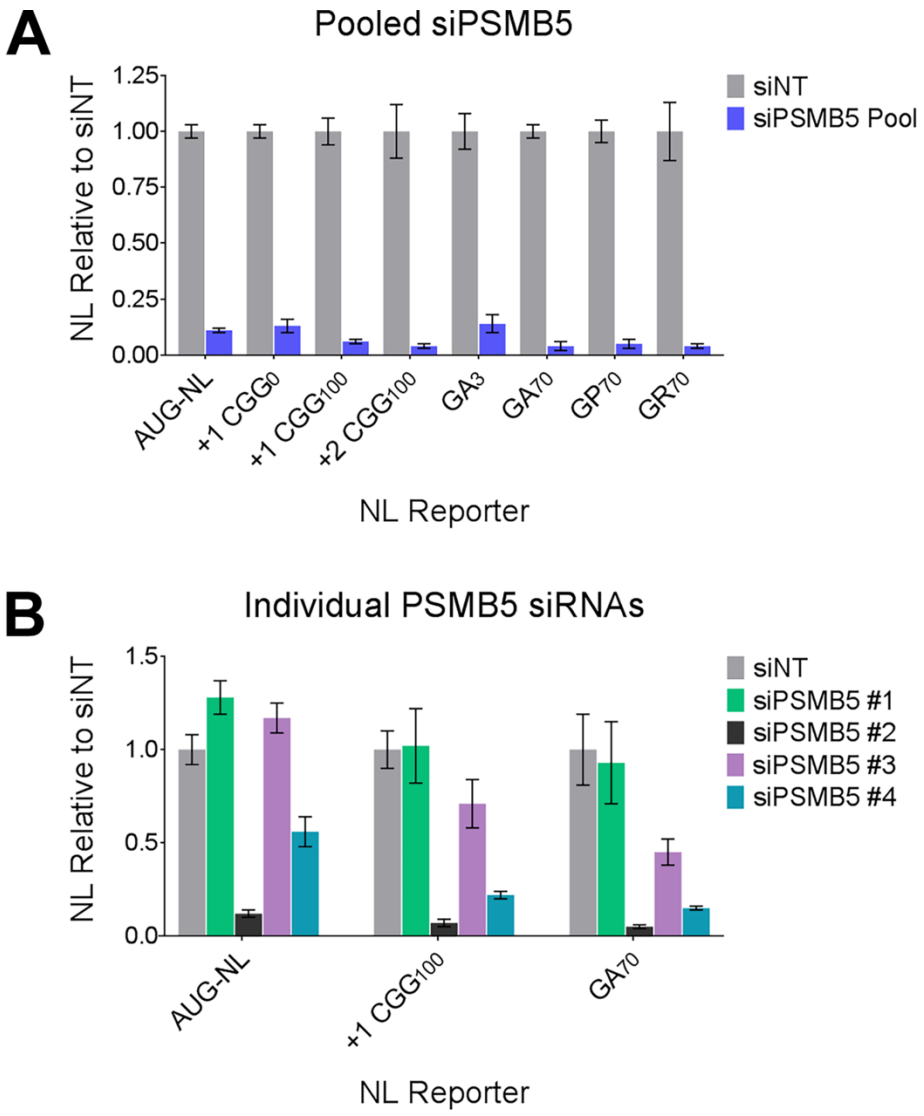

A
Pooled siPSMB5
NL Relative to siNT
1.25
1.00
0.75
0.50
0.25
0.00
siNT
siPSMB5 Pool
AUG-NL
+1 CGG0
+1 CGG100
+2 CGG100
GA3
GA70
GP70
GR70
NL Reporter
B
Individual PSMB5 siRNAs
NL Relative to siNT
1.5
1.0
0.5
0.0
siNT
siPSMB5 #1
siPSMB5 #2
siPSMB5 #3
siPSMB5 #4
AUG-NL
+1 CGG100
GA70
NL Reporter
C
+1 (CGG)100 NL-3xF
GA70 NL-3xF
siNT
siB5 Pool
siB5 #1
siB5 #2
siB5 #3
siB5 #4
FLAG
PSMB5 (HMW)
PSMB5 (LMW)
Tubulin

Figure 3.12: *PSMB5* Knockdown with Pooled and Individual siRNAs Have Differential Effects on Expression of *C9ORF72* RAN Products, *FMR1* RAN Products, and PSMB5.
A, B Expression of transfected, plasmid-based NL-3xF reporters by luminescence assay in HEK293 cells following transfection of siNT and pooled (A) or individual (B) siRNAs against *PSMB5* (siPSMB5).
C Anti-FLAG, anti-PSMB5, and anti-tubulin western blot of lysates from HEK293 cells transfected with siRNAs against *PSMB5* and either +1 $(CGG)_{100}$ or GA_{70} NL-3xF.

Data Information: In all panels data are presented as mean ± SD (compiled from ≥2 replicates).

---

the loss of cell viability and global reduction in reporter expression, but also did not reduce *PSMB5* expression; #3 specifically inhibited expression of +1 $(CGG)_{100}$ NL-3xF and GA_{70} NL-3xF while slightly reducing *PSMB5* expression; and #4, which exhibited the strongest *PSMB5* knockdown, manifested in a global reduction in reporter expression but with more pronounced effects on +1 $(CGG)_{100}$ NL-3xF and GA_{70} NL-3xF. The effects of siPSMB5 #3 and #4, therefore, are most consistent with our screen results. The RAN-selective nature of their effects are best illustrated following normalization of +1 $(CGG)_{100}$ NL-3xF and GA_{70} NL-3xF expression to either total protein loaded (western blot; Figure 3.12C) or expression of a co-transfected AUG-FF reporter (luminescence assay; Figure 3.13A): their effects persit.

We next tested whether the effects of siPSMB5 #3 and #4 on *C9ORF72* and *FMR1* RAN translation were frame- or repeat-dependent. Transfection with siPSMB5 #3 reduced the expression of expanded-repeat *FMR1* [+1 and +2 $(CGG)_{100}$ NL-3xF] and *C9ORF72* constructs (GA_{70}, GP_{70}, and GR_{70} NL-3xF), but had no effect on AUG-NL-3xF, +1 CGG_0 NL-3xF, or GA_3 NL-3xF (Figure 3.13B). In addition, the effects of siPSMB5 #4 were significantly stronger on +1 and +2 $(CGG)_{100}$ NL-3xF and GA_{70}, GP_{70}, and GR_{70} NL-3xF than AUG-NL-3xF, but AUG-NL-3xF, +1 CGG_0 NL-3xF, and GA_3 NL-3xF were all inhibited to a comparable degree (Figure 3.13C). These results indicate

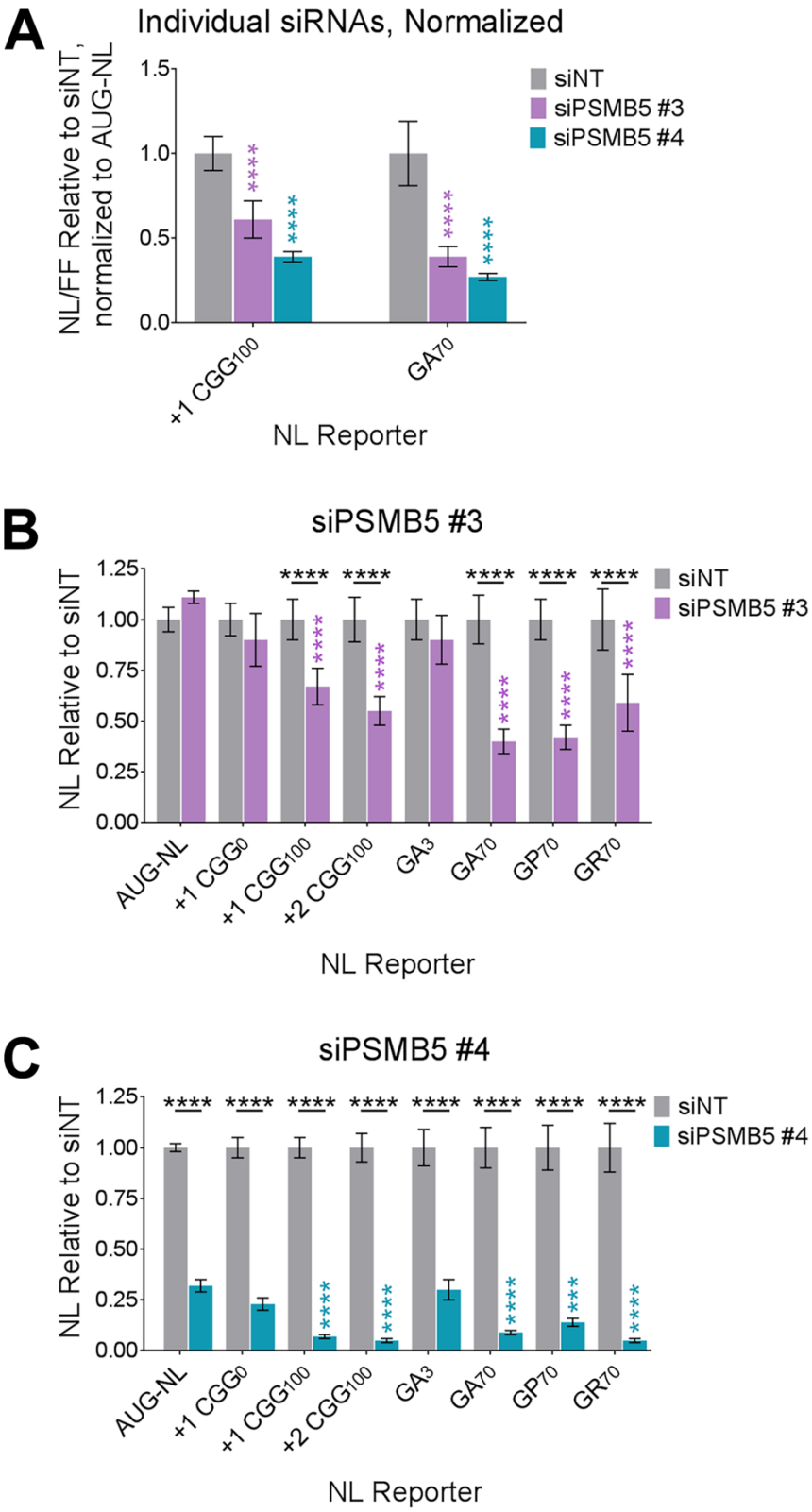

A
Individual siRNAs, Normalized
NL/FF Relative to siNT, normalized to AUG-NL
1.5
1.0
0.5
0.0
siNT
siPSMB5 #3
siPSMB5 #4
****
+1 CGG100
GA70
NL Reporter
B
siPSMB5 #3
NL Relative to siNT
1.25
1.00
0.75
0.50
0.25
0.00
AUG-NL
+1 CGG0
+1 CGG100
+2 CGG100
GA3
GA70
GP70
GR70
NL Reporter
C
siPSMB5 #4
NL Relative to siNT
NL Reporter

Figure 3.13: Knockdown of *PSMB5* Selectively Suppresses Expression of *C9ORF72* and *FMR1* RAN Products at Expanded Repeat Sizes.

A Expression of +1 $(CGG)_{100}$ or GA_{70} NL-3xF, normalized to a co-transfected AUG-FF reporter, in HEK293 cells transfected with siNT or siPSMB5. Colored asterisks indicate comparisons between siPSMB5-transfected cells expressing AUG-NL-3xF and those expressing a different NL-3xF reporter.

B, C Expression of NL-3xF reporters in HEK293 cells following transfection with siNT or siPSMB5 (two-way ANOVA with Sidak's multiple comparisons test; *n*=6/condition). Black asterisks indicate comparisons between siNT- and siPSB5-transfected cells expressing the same reporter.

Data Information: In all panels data are presented as mean ± SD (compiled from ≥2 replicates). **** P≤0.0001.

---

that depletion of PSMB5 results in a repeat-dependent inhibition of RAN-product expression regardless of frame, repeat, or sequence context.

Previous studies have suggested that FMRpolyG is degraded by the proteasome in *Drosophila* and HeLa cells[19,147]. If so, one would expect that depletion of PSMB5, a key proteasome component, would lead to higher, not lower, abundance of RAN products. We observed the opposite. To investigate this discrepancy, we tested whether pharmacological inhibition of the proteasome with the PSMB5-targeting proteasome inhibitor bortezomib[303,304] has the same effect as siRNAs on RAN product abundance. In HEK293 cells transfected with +1 $(CGG)_{100}$ NL-3xF reporter plasmids 24 hours prior, treatment with 100 nM bortezomib resulted in higher reporter expression relative to a vehicle control. We did not observe the same effect on an AUG-NL-3xF reporter (Figure 3.14A, B). These results are consistent with the field's general consensus, suggesting that FMRpolyG is degraded by the proteasome. We next asked whether proteasome impairment might affect the rate of FMRpolyG synthesis in addition to its degradation. To test this, we treated cells with 100 nM bortezomib, transfected plasmid-based reporters, then monitored reporter expression at multiple time points afterwards. The rate of neither AUG-NL-3xF nor +1 $(CGG)_{100}$ NL-3xF synthesis varied significantly

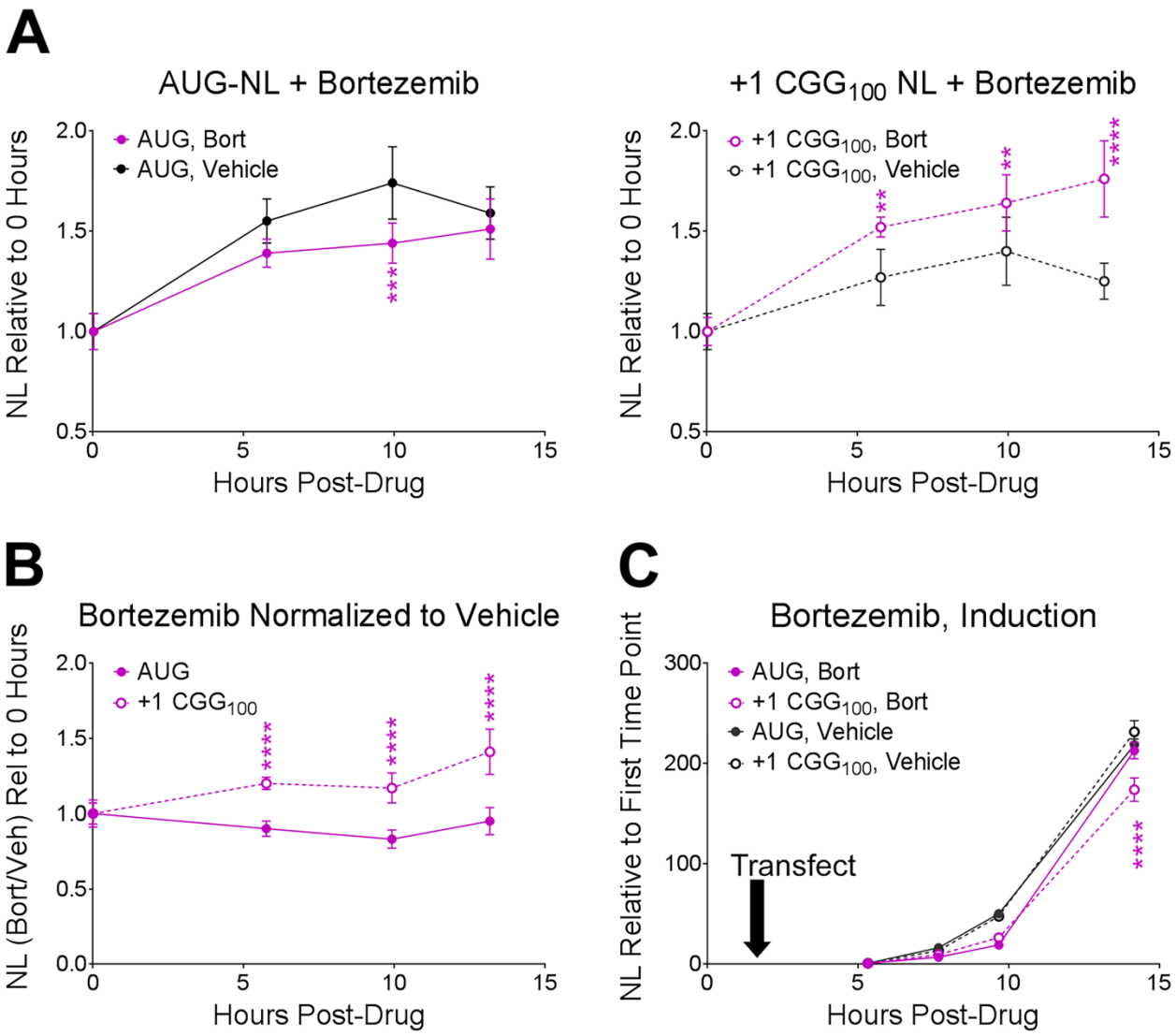


Figure 3.14: Inhibition of PSMB5 by Bortezomib Increases the Expression of FMRpolyG Reporters.

A Expression of AUG-NL-3xF (left) or +1 $(CGG)_{100}$ NL-3xF (right) in the presence of 100 nM bortezomib or vehicle. Bortezomib was added 24 hours post-transfection (two-way ANOVA with Sidak's multiple comparisons test; *n*=6/condition). Asterisks indicate comparisons between bortezomib- and vehicle-treated cells.

B Expression of AUG-NL-3xF and +1 $(CGG)_{100}$ NL-3xF with bortezomib normalized to vehicle. Asterisks indicate comparisons between cells expressing AUG-NL-3xF and those expressing +1 $(CGG)_{100}$ NL-3xF (two-way ANOVA with Sidak's multiple comparisons test; *n*=6/condition).

C Expression of AUG-NL-3xF and +1 $(CGG)_{100}$ NL-3xF with bortezomib or vehicle, with bortezomib added 2 hr before reporter transfection.

Data Information: In all panels data are presented as mean ± SD (compiled from ≥2 replicates). ** P≤0.01, *** P≤0.001, **** P≤0.0001.

following bortezomib administration over most of the experiment (Figure 3.14C). These results support the conclusion that knockdown of *PSMB5* exerts its effects through some pathway other than general proteasomal inhibition.

Discussion

We conducted a high-throughput, non-biased screen of the full human genome to identify modifiers of RAN translation of G_4C_2-expanded *C9ORF72*. The five DPRs synthesized via *C9ORF72* RAN translation are hypothesized to elicit toxicity through a variety of interactions with endogenous, physiological proteins and RNAs[305,306]. Previous screens for modifiers of $(G_4C_2)_n$-elicited toxicity focused, by design, on these interactions[171,172,180,181,195,280]. In contrast, this screen targeted modifiers of DPR abundance directly. Furthermore, we have validated each candidate modifier using both *C9ORF72* and *FMR1* reporters, allowing us to delineate a set of candidates that modulate RAN translation across NREs and local sequence context. We sought to better understand the mechanism of RAN translation through recognition of the genetic factors and pathways that can inflect it. This knowledge might be applied to devise new, effective therapeutic interventions that target RAN translation, a proximal pathogenic event shared by the five DPRs.

We so identified 585 RAN-selective modifiers of the GA DPR, the most abundant RAN product of G_4C_2-expanded *C9ORF72*[16,17,168-170], and/or FMRpolyG, a RAN product of expanded $(CGG)_n$ in *FMR1*. We have already launched studies of two promising candidates, the RNA-binding protein hnRNPA_2/B_1 and the catalytic proteasomal subunit

PSMB5. In addition, by using GO analysis we have identified multiple biological pathways enriched among *C9ORF72* and *FMR1* RAN suppressors, hinting at future lines of research into the mechanism of RAN translation.

hnRNPA$_2$/B$_1$ is an RNA-binding protein that interacts with the *FMR1* 5' UTR, likely through the UTR's CGG repeats[137-139]. Overexpressing *HNRNPA2B1* or its *Drosophila* homologs suppresses the toxicity of expanded CGG repeats, and the suppressing effects of both *TARDBP* and *CUGBP1* overexpression on CGG repeats are mediated by hnRNPA$_2$/B$_1$[139,141]. Mutations in the LCD of *HNRNPA2B1*, though rare, have been linked to ALS/FTD[140]. Finally, there is some evidence to suggest that hnRNPA$_2$/B$_1$ can tune the translation of mRNAs bearing expanded CGG repeats[307]. Thus, there are multiple lines of evidence that situate hnRNPA$_2$/B$_1$ at the intersection of RNA metabolism and neurodegeneration. We demonstrated here that, in HEK293 cells, *HNRNPA2B1* knockdown suppressed the expression of *C9ORF72* and *FMR1* RAN products regardless of frame and (generally) in a repeat size-dependent manner. However, in *Drosophila*, altering the expression of *HNRNPA2B1* or its homologs did not consistently affect the abundance of the FMRpolyG-EGFP RAN product, despite bi-directionally modulating the toxicity of (CGG)$_{90}$ repeats. The cause of this discrepancy is unclear. Our *Drosophila* data not only contradict our findings in HEK293 cells, but also imply that changes in RAN translation do not underlie the phenotypic effects of *HNRNPA2B1* modulation that we and others have observed[139]. A major concern here is the possibility that the effect of *HNRNPA2B1* knockdown in HEK293 cells is an artifact of our transfected, plasmid-based reporters. This concern is compounded by our observation that knockdown of *HNRNPA2B1* had no effect on the abundance of the GP

RAN product in expanded G_4C_2-expressing iPSCs. At least, this discrepancy should give pause to future investigators utilizing these tools—it points to the need for validation in model systems expressing RAN products in an endogenous context.

Nevertheless, we were intrigued that, in HEK293 cells, *HNRNPA2B1* knockdown also increased the abundance of +1 $(CGG)_{100}$ reporter mRNAs, despite decreasing the abundance of proteins synthesized from these transcripts. This implies that hnRNPA_2/B_1 might act at two distinct stages of gene expression—one impacting mRNA abundance, one impacting mRNA translation—in opposing directions. Alternatively, an overabundance of +1 $(CGG)_{100}$ reporter mRNAs might overwhelm some peri-translational rate-limiting step in RAN-product expression—more mRNAs leading directly to less protein. Based on our HCR results, nuclear export does not appear to be this limiting step. Given our concerns with heterologous expression in HEK293 cells, the effects of *HNRNPA2B1* knockdown must first be validated in an alternate system, preferably one with physiological *C9ORF72* expression. Still, we find it tempting to speculate that changes in the abundance or processing of $(CGG)_n$ or $(G_4C_2)_n$-containing mRNAs might underlie the phenotypic rescue we observed in *Drosophila*.

PSMB5 is a catalytic subunit of the proteasome, possessing “chymotrypsin-like” activity[308]. Previous work has demonstrated that *C9ORF72* and *FMR1* RAN products interact with and are regulated by the proteasome[19,147,167,178,179,309,310]. We were therefore surprised to find that knockdown of *PSMB5* and other proteasomal subunits suppressed GA_{70} NL-3xF expression in our screen. The mechanism of this effect is unclear, but it might involve substitution of PSMB5 with PSMB8, a closely related component of the immunoproteasome with higher catalytic efficiency[311-316]. This would

explain why *PSMB5* knockdown suppressed GA_{70} and +1 $(CGG)_{100}$ NL-3xF while bortezomib, a PSMB5-binding proteasome inhibitor, enhanced +1 $(CGG)_{100}$ NL-3xF expression: bortezomib inhibits both PSMB5 and PSMB8[304,317-320].

The effects of *PSMB5* knockdown on AUG-, +1 $(CGG)_{100}$-, and GA_{70} NL-3xF expression were siRNA-dependent: though several siRNAs suppressed *C9ORF72* and *FMR1* RAN reporters, two (#2 and #4) also had effects on AUG-NL-3xF (though the effects of siPSMB5 #4 were significantly stronger on +1 $(CGG)_{100}$ and GA_{70} NL-3xF). We had not observed any effects of *PSMB5* knockdown on AUG-NL-3xF in our screen, in which we used a different siRNA library. The effects of siPSMB5 #4 on AUG-NL-3xF might be explained by its stronger depletion of PSMB5 relative to the others. In contrast, siPSMB5 #2 exhibited the strongest effect on global reporter expression by luciferase assay, most likely as a side-effect of loss of cell viability, while having no effect on the abundance of mature or immature PSMB5 levels. This exemplifies a common issue with siRNA screens: how to eliminate off-target effects of siRNAs without introducing new ones[284-286]. In the ideal scenario, investigators can validate hits using a second, independent siRNA library. Acquiring additional siRNAs on a large scale can prove financially prohibitive, however, depending on institutional resources and the number of hits to be validated. Furthermore, as happened here, these new siRNAs can introduce their own unanticipated effects, whether based on off-target recognition or differences in the efficiency of knockdown.

Besides *HNRNPA2B1* and *PSMB5*, 583 genes—candidate modifiers of either *C9ORF72*, *FMR1*, or both—remain. We based our dual *C9ORF72* and *FMR1* screen design on the conjecture that RAN translation might differ mechanistically based on the

sequences of NREs and surrounding contexts[169], as NRE-adjacent sequences often harbor stop codons and near-AUG initiation sites[150,169,170], and RNA-binding proteins with high consensus-sequence specificity (MBNL, for example[23]) might affect RAN translation in a transcript-specific manner. As our screens identified 372 modifiers specific to either *C9ORF72* (362) or *FMR1* (10), this work provides some support for this hypothesis. The disparity between the number of *C9ORF72*-specific and *FMR1*-specific modifiers is intriguing, however. It can be attributed somewhat to the size of the siRNA libraries screened: 1,052/1,742 (60.4%) of *C9ORF72* primary modifiers were not tested in the *FMR1* primary screen because they were not included in the Drug-Target/Druggable siRNA library. However, assuming 1) equal hit frequency between *C9ORF72* and *FMR1* primary screens and 2) similar distribution of hits between Drug-Target/Druggable and whole-genome libraries, we would have expected to identify ~60% fewer *FMR1*-specific than *C9ORF72*-specific modifiers, not 97% (352/362) fewer. One of these assumptions must be wrong. We can offer no explanation why *FMR1* modifiers are less likely to be receptors, kinases, phosphatases, proteases, or ion channels than *C9ORF72* modifiers are. But slight variations in the structures of *C9ORF72* and *FMR1* reporter could generate different hit frequency. In order to maintain expression of the *FMR1* reporter within the linear range of the luminescence assay, we tagged this reporter with a C-terminal PEST tag, which decreases reporter abundance by destabilizing the protein product and rapidly targeting it for degradation. This PEST tag might explain why knockdown of multiple proteasomal subunits suppressed the expression of GA_{70} NL-3xF but had no effect on our +1 CGG_{100} NL-3xF-PEST reporters, for example (despite reducing the expression of non-PEST-tagged +1

CGG_{100} NL-3xF reporters in subsequent experiments). If this interpretation is correct and multiple "*C9ORF72*-specific" siRNAs failed to affect +1 CGG_{100} NL-3xF-PEST in our screen because of differences in reporter-protein stability, those genes whose knockdown did impact both *C9ORF72* and *FMR1* reporter expression might be more likely to modulate RAN-product synthesis than their degradation. These are therefore prime candidates for follow-up—in brief, a narrow, translation-specific set came at the cost of a comprehensive list of *FMR1* modifiers.

GO analysis revealed significant enrichment of GO terms with high *a priori* relevance to RAN translation and RAN-product metabolism, including "IRES-dependent viral translational initiation" (*C9ORF72* and *FMR1*), "positive regulation of canonical Wnt signaling pathway," proteasome-mediated protein catabolism, mRNA/DNA methyltransferase activity, "transcription factor-binding," "C-terminal protein methylation," and "Golgi-to-lysosome transport" (all *C9ORF72* alone), as well as the unanticipated GO term "NIK/NFκB signaling" (*C9ORF72* alone). Besides these, "chaperone-binding" was significantly enriched within the *FMR1* set, but the corresponding genes—*BIRC5*, *BAX*, and *LRP2*—have been more thoroughly described in contexts other than unfolded proteins and proteotoxic stress[321,322]. The question of whether RAN translation initiates via an IRES-like mechanism has been interrogated, discussed, and debated on its empiric merits and mechanistic implications elsewhere[150,169,170,202,214,323-325]; suffice it to say that there is evidence from heterologous expression systems both in favor of and against IRES-mediated initiation for both *C9ORF72* and *FMR1*. (Notable, however, are the two genes associated with this GO term: PTB preferentially binds expanded $(G_4C_2)_n$ RNA in *vitro*[281], undergoes phase

separation[187], interacts with GR_{50} and PR_{50} DPRs in HEK293 cells, and modulates PR_{50}-elicited toxicity in *Drosophila*[195]; and eIF3F, though a canonical eIF, appears to also modulate RAN translation of *SCA8*[283].) In addition, DDX3X is a Wnt-dependent regulator of casein kinase 1[326], besides its more extensively explored functions in translation initiation.

The ubiquitin-proteasome system is a central determinant of cellular toxicity secondary to either misfolded or intrinsically disordered proteins, both with regard to neurodegeneration in general[327-330] and to RAN-associated disease in particular (see discussion of PSMB5 above). That we identified several proteasomal subunits and proteasome regulators as RAN modifiers is therefore intriguing. But in addition to its role in general proteome turnover, the proteasome promotes NFκB signaling by degrading IκBα, thus disinhibiting NFκB proteins from entering the nucleus and regulating gene expression[331]. Multiple ligands, receptors, signal-transduction components, and regulators of the NFκB/TNF pathway were also identified by GO analysis of our *C9ORF72* RAN modifiers, including TNF receptor superfamily 9, RhoA, MAP3K14, IκB kinase β, Fbxw11, ectodysplasin A, and ectodysplasin A receptor. The mechanism by which inhibition of NFκB activity might suppress RAN translation is unclear, but there is evidence in support of mutual up-regulation of NFκB activity and the ISR[332-334]. Though ours is not the first screen of *C9ORF72* RAN modifiers to identify multiple subunits of the proteasome[171,280], it is the first (at least among those published) to potentially relate proteasomal function to NFκB/TNF signaling in the context of RAN translation.

mRNA methyltransferases and demethylases (such as the METTL3-METTL14-WTAP complex and ALKBH5, respectively) regulate translation efficiency, the

translational response to cellular stress, phase separation of mRNAs, and the requirement of various eIFs in translation[335-339] through introduction and removal of N^6-methyladenosine (m^6A) and other mRNA modifications. Here, knockdown of the mRNA methyltransferase *METTL16* suppressed *C9ORF72* RAN translation. Little is known of METTL16 other than it is predominantly cytoplasmic, mediates m^6A modification at a UACAGAGAA consensus site (found in neither genomic nor transcriptomic *C9ORF72* sequences), and regulates transcript-selective intron retention[340-343]. In addition, knockdown of *METTL7A*—a presumed nucleic-acid methyltransferase (based on homology) that is known to regulate replication of Hepatitis C virus (an RNA virus, translation of which also requires DDX3X[60]) on the surface of lipid-rich, membraneless organelles[344]—also suppressed *C9ORF72*. Oddly, knockdown of several RNA demethylases likewise suppressed RAN product expression. *ALKBH5* knockdown suppressed both *C9ORF2* and *FMR1* RAN translation, while knockdown of *FTO* and *ALKBH2* (by either siGENOME or ON-TARGET*plus* siRNAs) selectively suppressed *C9ORF72* RAN translation. Fto is a highly CNS-expressed protein that mediates mRNA demethylation and thereby regulates mRNA stability and abundance in a transcript-specific manner[338,340,345-348]. ALKBH2 might regulate RAN translation in similar fashion, though evidence to date suggests that it exhibits preferential demethylation of DNA substrates[349]. Our screens also identified as a *C9ORF72* modifier *ADAR1*, an RNA-editing enzyme that mediates A-to-I deamination and thereby regulates RNA-RNA secondary structure and mRNA stability[350,351]. Alternatively, methylation and/or demethylation of DNA NREs or flanking regions by these enzymes might be the critical regulatory step: multiple lines of evidence implicate dysregulation of DNA methylation in

C9 ALS/FTD[268,269], including the observation that patients with C9 ALS and those with non-C9 ALS exhibit similar patterns of DNA methylation in post-mortem brain[352]. We here and others previously[180,195] have implicated XRCC proteins in C9 ALS/FTD (and here, *FMR1* RAN translation); these proteins are key mediators of repair of DNA double-stranded breaks induced by, among other stressors, alkylation[353].

Boeynaems *et al*.[180] screened for modifiers of eye toxicity in *Drosophila* expressing an AUG-initiated, codon-optimized PR_{25} transgene and demonstrated that depletion of 4 different arginine (R) methyltransferases enhanced PR_{25} toxicity: *FBXO11* (either *FBXO11* or *FBXO10* in human), *Art1*/*PRMT1*, *Art6*, and *Art7*/*PRMT7*. They went on to demonstrate that PRMT1 colocalized with both PR and GR in heterologous DPR-overexpressing cultured cells and GR was methylated in C9 ALS/FTD post-mortem brain. That our screen—interrogating non-AUG initiation of the native-sequence GA DPR and FMRpolyG in a totally orthologous experimental system—identified the same category of genetic modifiers is tremendous: *PRMT2* (*C9ORF72* and *FMR1*), *PRMT7* (*FMR1*), *FBXO11* (*C9ORF72* and *FMR1*), as well as related leucine and cysteine methyltransferases *LCMT1* and *ICMT* (both *C9ORF72*). Knockdown of these suppressed RAN translation. We also point out that *PRMT1* knockdown strongly reduced expression of both *C9ORF72* and *FMR1* RAN products, but narrowly failed to meet inclusion criteria during the primary screens. How might these methyltransferases regulate RAN translation? The ALS/FTD-associated protein FUS might be an informative example. The subcellular localization of multiple predominantly nuclear RNA-binding proteins (FUS, EWS, TAF15) is dependent on the methylation of R residues adjacent to their nuclear localization signals (NLSs), which is mediated by,

among other enzymes, *PRMT1*[354,355]—R-dimethylated FUS is predominantly cytoplasmic, and *PRMT1* inhibition appears to promote cytoplasmic-to-nuclear transport of WT FUS and ALS-associated FUS mutants[355,356]. But in addition, non-methylated WT FUS exhibits liquid-liquid phase separation at lower concentrations than R-dimethylated FUS *in vitro*, reduced intra-droplet diffusion, increased incorporation into stress granules in cultured heterologous cells, and increased resistance to proteasomal degradation in *Drosophila*[357,358]. This literature implies that R methyltransferases exert neuroprotective effects in the setting of FUS mutation by mitigating FUS phase separation and the subsequent dysregulation of membraneless organelles (stress granules, P bodies, nucleoli, and others), consistent with the observed potentiation of PR_{50}- and mutant FUS-elicited toxicity by *PRMT1* knockdown[180,356]. One might speculate that depletion of R methyltransferases could therefore inhibit RAN translation of non-R products by leading to the triage and sequestration of RAN-competent mRNAs (though it remains unclear how to reconcile that hypothesis with the facilitation of RAN translation following ISR activation[169,202]) or by disrupting the activity of PRMT1 targets such as Rap55/Scd6, which binds eIF4G, recruits PICs to stress granules, and mediates general translational repression in the setting of ISR activation and in an R dimethylation-dependent manner[359-362]. In our hands, Rap55 depletion, like that of several PRMTs, selectively suppressed *C9ORF72* RAN translation. Alternatively, perhaps R-dimethylated RAN products exhibit different molecular characteristics than FUS. Overall, GO analysis serves two purposes: it is a catalyst for new hypotheses as to how RAN translation is regulated, as well as a map, specifying which genes of interest are most ripe for follow-up mechanistic experiments when testing those hypotheses.

Besides *HNRNPA2B1* and *PSMB5*, 211 promising candidate genes remain: replicable, RAN-selective modifiers with some effect on both *C9ORF72* and *FMR1* RAN translation, with minimization of off-target false positives. With this smaller, higher-confidence set, determining therapeutic potential and the phenotypic effects of gene disruption (using *Drosophila* disease models), physiological relevance (using patient-derived iPSCs), and possibly on-target validity (using individual siRNAs rather than siRNA pools) is more feasible. Our experience validating *HNRNPA2B1* and *PSMB5* has been informative—these candidates hint at some of the stimulating biology our screen reveals, while highlighting which experiments must be performed in order to validate other candidates with confidence. We eagerly anticipate our next steps.

Materials and Methods

Plasmids

pcDNA3.1(+)/*C9ORF72* GA_{70} NL-3xF and pGL4.13 (AUG-FF) have been described previously[169]. pcDNA3.1(+)/AUG-NL-3xF-PEST was generated from pcDNA3.1(+)/AUG-NL-3xF[150] using Q5 Site-Directed Mutagenesis (New England BioLabs) and the primers detailed in Table B.14. pcDNA3.1(+)/*FMR1* +1 $(CGG)_{100}$ NL-3xF-PEST was generated from pcDNA3.1(+)/*FMR1* +1 $(CGG)_{100}$ NL-3xF[150] using the same method. All plasmids used throughout the screen were purified from *E. coli* cultures using the Maxiprep Plasmid DNA Purification Kit (Qiagen).

C9ORF72 and *FMR1* High-Throughput Screens

HEK293 cells (ATCC, CRL-1573) were maintained and passaged at 37° C, 5% CO_2 in DMEM supplemented with 10% (vol/vol) FBS. Use of the Human siGENOME SMART*pool* siRNA Library (Dharmacon) with full-genome coverage was acquired through the Center for Chemical Genomics at the University of Michigan, Ann Arbor. siRNAs were reconstituted in 1x siRNA Buffer (Dharmacon), sealed, and stored under RNase-free conditions in 384-well stock plates (Thermo Fisher Scientific, AB-0781) at -20° C.

siGENOME siRNAs were moved to 4° C and allowed to thaw for 24 hours. Using a Biomek FXP (Beckman-Coulter), siRNAs were transferred to white 384-well assay plates (Greiner Bio-One, cat. # 781080) in triplicate, diluted to 13.17 nM in OptiMEM Reduced Serum Medium (Thermo Fisher Scientific) containing 450 nl/well of Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher Scientific; final volume of 10.4 µl/well), briefly centrifuged, and incubated at room temperature for 30 minutes. siNT, siNL, and siPLK1 were added to each assay plate and handled identically. Using a MultiDrop Dispenser (Thermo Fisher Scientific), dissociated HEK293 cells in DMEM were added at 6,000 cells (20 µl)/well, with final concentrations of 10% FBS, 100 U/ml penicillin, 100 µg/ml streptomycin (Thermo Fisher Scientific), and 39.5 nM siRNA.

Forty-eight hours later, cells were transfected with 15 ng/well firefly luciferase plasmid (pGL4.13) and 15 ng/well either pcDNA3.1(+)/*C9ORF72* GA_{70} NL-3xF plasmid or pcDNA3.1(+)/*FMR1* +1 CGG_{100} NL-3xF-PEST plasmid using Viafect (Promega). Twenty-four hours after plasmid transfection, excess media was aspirated using a 405 Select Microplate Washer (BioTek) to leave 10 µl media/well. The Nano-Glo Dual Luciferase Reporter Assay System (Promega) and Envision 2104 Multilabel Reader

(Perkin-Elmer) were used per manufacturer's instructions to quantify expression of GA_{70} NL-3xF (or +1 CGG_{100} NL-3xF-PEST) and AUG-FF reporters independently.

Low-Throughput Luminescence Assays

All cell plating, siRNA transfections, plasmid transfections, and luminescence assays in low-throughput format were performed as described previously (Chapter 2).

Western Blotting

Western blotting was performed as described previously (Chapter 2). All antibodies used for blotting were diluted in 5% non-fat dairy milk (wt/vol), 0.02% NaN_3 (wt/vol), 0.1% Tween-20 (vol/vol) in TBS. Mouse anti-hnRNPA_2/B_1 antibody (Dp3B3; Santa Cruz, sc32316) was used at 1:1,000 dilution. Monoclonal mouse anti-EGFP antibody was acquired from Sigma (clones 7.1 and 13.1, catalog #11814460001). Monoclonal mouse anti-β tubulin antibody, developed by Michael Klymkowsky, was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH, and maintained at The University of Iowa, Department of Biology, Iowa City, IA 52242. Monoclonal mouse anti-FLAG antibody was acquired from Sigma (clone M2, catalog #F1804). Rabbit anti-β actin antibody (N21; Santa Cruz, sc130656) was used at 1:10,000.

qRT-PCR Experiments

All qRT-PCR experiments were performed as described previously (Chapter 2).

Drosophila melanogaster Phenotyping and Western Blotting

All fly lines used in this study and their sources are listed (Table B.15). Eye phenotypes were quantified and *Drosophila* tissue prepared for western blotting as previously described (Chapter 2). For western blot quantification, prepared lysates were loaded onto two duplicate gels and transferred, blotted, developed, and quantified identically using ImageJ (NIH). The mean band density was used for subsequent statistical analysis.

Hybridization Chain Reaction (HCR)

HCR version II was done following Molecular Instruments "Mammalian cells on slide" protocol[363] with the following modifications. HEK293 cells were transfected in 8-well slides and fixed for 10 minutes in 4% PFA, washed 3x in PBS, and permeabilized in 70% ethanol overnight at 4° C. Cells were washed 3x in PBS, treated with DNase (10 U/well Turbo DNase; Invitrogen) in 1X Turbo DNase buffer at 37° C for 30 minutes. Cells were washed 3x in PBS and prehybridized in 50% formamide hybridization buffer (Molecular Instruments) for 30 minutes at 45° C, followed by hybridization with 1 pmol $(CCG)_8$ probe/well overnight at 45° C. Cells were washed 4x with probe wash buffer (Molecular Instruments), 2x with 5X SSCT, and incubated for 30 minutes at room temperature in amplification buffer (Molecular Instruments). Cells were incubated with 3.75 pmols of snap-cooled Alexa Fluor 647-labeled B1H1 and B1H2 hairpin probes (Molecular Instruments) per well in amplification buffer at room temperature overnight. Cells were washed 5x with 5X SSCT and mounted with ProLong Gold Antifade Mount with DAPI. Fifty transfected cells/condition at minimum were imaged at 40x with oil on a

confocal microscope. Representative images are shown.

iPSC Culturing, Transfection, and GP ELISA

All iPSC and ELISA experiments were performed as described previously (Chapter 2).

Acknowledgements

We thank everyone at the Center for Chemical Genomics at the University of Michigan, Ann Arbor for many hours of technical assistance and even more hours of thoughtful high-throughput design. We thank Juan Botas at the Baylor College of Medicine for the *HNRNPA2B1*-, *HRB87F*-, and *HRB98DE*-overexpressing fly lines. We thank the Liebermann lab at the University of Michigan for the bortezomib. We apologize to future generations for the tonnage of plastic consumed by this and other high-throughput screens.

Chapter 4
Conclusions and Future Directions

Summary and Significance

Candidate-Based and Non-Biased Screens Identified Modifiers of RAN Translation

The objective of this thesis work was to identify proteins and regulatory pathways that modulate RAN translation. This objective is significant for two reasons. First, RAN-translation products are sufficient and in some contexts necessary to elicit toxicity in overexpression models of FXTAS and C9 ALS/FTD[15,19,173], suggesting they are directly involved in pathogenesis. Understanding how RAN translation occurs and how it might be blocked, therefore, carries significant clinical implications for neurodegenerative disorders associated with RAN translation[25]. Second, by initiating at non-AUG codons, RAN translation violates a fundamental assumption of protein synthesis that underlies how we understand and define eukaryotic proteomes. Though it has been known for decades that PICs can and do initiate at non-AUG codons[92,95], it is only recently that investigators have begun to appreciate the biological significance of these events[100,206,207]. Therefore, RAN translation represents a new, important, and clinically meaningful mechanism by which mRNAs are translated and protein synthesis regulated that likely has implications beyond a relatively rare set of neurodegenerative disorders[205].

To meet this objective, we conducted two screens for modifiers of RAN translation: one a candidate-based, phenotype-modifier screen in a *Drosophila melanogaster* model of FXTAS, and the other a genome-wide, high-throughput screen for modifiers of *C9ORF72* RAN-product expression in HEK293 cells. Our *Drosophila* screen closely resembles previous modifier screens that have been used to identify genetic suppressors and enhancers of toxicity elicited by other NREs[273-278], using a degenerative, "rough-eye" phenotype as a visible proxy for FMRpolyG expression. In contrast, our HEK293 screen provided >100,000 direct measurements of RAN-product (GA_{70}- and FMRpolyG$_{100}$-NL) and canonical-translation (AUG-FF) reporters, enabling us to identify selective modifiers of RAN translation (explaining why *EIF4B* and *EIF4H* were identified as hits in *Drosophila* but not HEK293 cells, for example). This high-throughput screen is among the first of its kind to directly identify genes that specifically impact RAN-product expression, rather than those that modulate RAN product-elicited toxicity[171,180,279,280]. In combination, these screens identified three proteins—bel/DDX3X, hnRNPA_2/B_1, and PSMB5—as novel modifiers of RAN-product expression across gene contexts and reading frame, along with others (eIF4B, eIF4H, eIF1, and eIF5) that influence RAN translation through mechanisms that also impact translation globally.

Bel/DDX3X Facilitates RAN Translation, and its Disruption Mitigates NRE Toxicity

We demonstrated that knockdown of *bel*/*DDX3X* both reduces the expression of FMRpolyG and FMRpolyA in *Drosophila*, cultured HeLa cells, and primary rodent neurons, as well as mitigates the toxicity of expanded CGG repeats in *Drosophila* and cultured neurons. We then demonstrated that DDX3X acts co-translationally, consistent

with its described interactions with eIFs and function in promoting 43S PIC scanning[58-61,64-68,364]. Finally, we demonstrated that *DDX3X* knockdown reduced the expression of multiple *C9ORF72* RAN products, including GA_{70} reporters in HeLa cells and endogenous GP in patient-derived iPSCs. This line of research is the first in-depth, mechanistic investigation of a RAN-modulating factor, drawing on experiments performed in *Drosophila in vivo*, HeLa cells, cultured rodent neurons, and patient-derived iPSCs. Though other researchers have previously identified factors such as MBNL and eIF3F that impact the abundance of RAN products[23,283], they did not demonstrate co-translational activity of these factors nor modification of NRE-elicited toxicity *in vivo* following their manipulation. Besides building a strong, multi-tiered argument that DDX3X modulates RAN translation, this course of experiments delineates a strategy to validate candidate modifiers of RAN translation on the basis of efficacy, mechanism, reduction in toxicity, and physiological relevance.

Highly Coincident Sets of Genes Modify *C9ORF72* and *FMR1* RAN Translation

Our high-throughput screens tested the 7,811 genes of Dharmacon's Drug-Target and Druggable Genome libraries for modifiers of GA_{70} and FMRpolyG$_{100}$ independently. In addition, we tested those siRNAs that modified GA_{70} from across the human genome for congruent effects on FMRpolyG$_{100}$. Both schemes revealed significant, but not complete, overlap between modifiers of *C9ORF72* and *FMR1* RAN translation. Moreover, knockdown of at least one modifier (*PSMB5*) inhibited expression of all tested *C9ORF72* and *FMR1* RAN products despite meeting hit criteria only for *C9ORF72*, implying that our design actually underestimates *C9ORF72*/*FMR1*-modifier

congruency. This work is the first of its kind to systematically compare modifiers of RAN translation of different NREs within different genetic contexts. In light of the significant congruency of RAN modifiers observed, it is both reasonable to conclude that RAN translation exhibits some mechanistic commonalities despite sequence disparities, as well as tempting to speculate that these commonalities are amenable to therapeutic targeting with a common agent.

Outstanding Questions

How Does DDX3X Modulate RAN Translation?

This thesis work establishes that bel/DDX3X modulates RAN translation. It does not directly address how it does so. DDX3X and its homologs in yeast and *Drosophila* are RNA helicases[51,53,55,65,66] involved in multiple aspects of mRNA metabolism[58,61,68,73,365-367]. On the basis of our RIP assay results, we cannot conclude that the RAN-specific effects of *bel*/*DDX3X* knockdown reflect transcript-specific binding. (Our results are consistent with previous work, in which Ded1 did not demonstrate sequence-specific binding across or within 5' UTRs[59].) And because the effects of *DDX3X* knockdown are not dependent on the size of the CGG repeat (following transfection of plasmid reporters, at least), but rather on the 5' UTR of *FMR1* itself, we cannot conclude that DDX3X's activity is mediated by molecular interaction with the NRE.

Because knockdown of *DDX3X* suppressed RAN translation of +1 $(CGG)_0$ and $(CGG)_{100}$ NL-3xF comparably, we hypothesize that DDX3X interacts with the *FMR1* 5'

UTR itself. The *FMR1* 5' UTR is highly GC-rich (80%, excluding the repeat), and is therefore predicted to form stable secondary structures. (Intron 1 of *C9ORF72* is likewise highly GC-rich.) Such structures are capable of impeding scanning 43S PICs[91,111] and preventing them from accessing downstream initiation sites, leading to a decrease in protein synthesis. Bel/DDX3X might act to resolve these structures, facilitating 43S PIC scanning and permitting them access to the downstream initiation sites for RAN translation[19,150]. Testing this hypothesis will require *in vivo* probing of the structure of the *FMR1* 5' UTR in the presence and absence of DDX3X. Such techniques, including variations on DMS-Seq and SHAPE-Seq, exist and have been used successfully[368-372]. These tools would enable investigators to understand the structure (or structures) of the *FMR1* 5' UTR on a molecule-by-molecule basis. Based on our hypothesis, we would predict that *DDX3X* knockdown would result in a greater degree of secondary-structure stability 5' to the start sites for RAN translation.

How Does PSMB5 Modulate the Expression of RAN Products?

Depletion of a conspicuous number of proteasomal subunits, including the catalytic subunits PSMB1 and PSMB5[308], suppressed the expression of the GA *C9ORF72* RAN product. Despite not suppressing *FMR1* RAN translation in our high-throughput screen, follow-up experiments demonstrated that knockdown of *PSMB5* did, in fact, comparably suppress the RAN products of *C9ORF72* and *FMR1* across reading frames and in a repeat-dependent manner. This was unexpected, as the FMRpolyG RAN product is a substrate of the proteasome[147], and pharmacological inhibition of PSMB5 resulted in an increase in *FMR1* +1 $(CGG)_{100}$ NL-3xF reporter expression, both

overall and relative to an AUG-initiated reporter. Clearly, knockdown of *PSMB5* does not have the same effect as pharmacological proteasome inhibition. How can we account for this difference?

Three non-constitutive proteasomal subunits—PSMB8, PSMB9, and PSMB10—are induced by interferon γ and, in response to interferon γ, substitute for the catalytic subunits PSMB5, PSMB1, and PSMB2, respectively[373,374]. This interferon γ-induced, re-constituted proteasome, known as the immunoproteasome[375], exhibits accelerated protein degradation, potentially to facilitate antigen presentation in anticipation of viral infection[311-313,315,316]. Intriguingly, researchers have also observed neuronal induction of immunoproteasomal subunits in patient tissue and animal models of Huntington's disease, Parkinson's disease, and ALS[376-378]. Could knockdown of *PSMB5* induce incorporation of the more catalytically-active PSMB8, leading to a higher rate of RAN product degradation? This could explain the divergent results of *PSMB5* knockdown versus bortezomib administration, since bortezomib targets both constitutive and immuno-proteasomes. This hypothesis is readily testable using traditional methods of PSMB8 detection (western blotting and qRT-PCR) and genetic and pharmacological inhibitors[379,380] of the immunoproteasome in cultured HEK293 cells, patient-derived iPSCs, and primary rodent neurons.

Do RAN-Translation Modifiers That Act Via Different Mechanisms Converge to Suppress NRE-Elicited Toxicity?

The work presented here makes a strong argument that disruption of *bel/DDX3X* mitigates NRE-elicited toxicity by inhibiting RAN translation *per se*. In parallel, others

have demonstrated that knockdown of multiple proteasomal subunits mitigates toxicity elicited by exogenous, recombinant PR_{20} and GR_{20}[280], reflecting effective modulation of toxicity *after* these DPRs have been synthesized. This prompts the question: might the combination of therapeutic interventions that impact different events in RAN products' lifespans lead to an additive reduction in NRE-elicited toxicity? In that regard, further understanding of the relation between hnRNPA_2/B_1 and NRE-elicited toxicity could prove useful. Though we demonstrated that disrupting *HNRNPA2B1*'s homologs results in bi-directional modulation of NRE-elicited toxicity in *Drosophila*, we could not point conclusively to its mechanism of action: overexpression and knockdown of *HNRNPA2B1* or its fly homologs did not yield consistent effects on FMRpolyG expression. It is possible that hnRNPA_2/B_1 exerts its protective effects via modulation of RNA-mediated toxicity, invoking the previously dominant RNA-mediated gain-of-function hypothesis of FXTAS pathogenesis, in which expanded $(CGG)_n$ sequester essential RNA-binding proteins[10,132]. Indeed, the C9 ALS/FTD field has long recognized that G_4C_2-containing RNAs likely have their own downstream toxicities[381]. Further comparison of known modifiers of RAN translation (from ours and others' genome-wide screens[325,382]) to known modifiers of NRE-elicited toxicity in model systems[171,172,180,182,279,280,383-388] and genotype-phenotype correlations in patients could shed more light on the respective contributions of RAN products and NRE-containing RNAs to NRE-associated neurodegeneration, in turn highlighting potential combination therapeutic strategies.

How to Minimize False Positives Due to Off-Target siRNA Interactions?

Our high-throughput screens identified an ample set of putative RAN translation-specific modifiers: 544 *C9ORF72* RAN-suppressing and 8 RAN-enhancing siRNA pools (of which 188 suppressors and 2 enhancers had analogous effects on *FMR1* RAN translation), as well as 75 independently identified *FMR1*-suppressing siRNA pools (of which 65 had analogous effects on *C9ORF72*). Based on previous investigations of off-target siRNA interactions[284-286] and our own pilot data, this set is not without false positives due to off-target effects, however. These present a significant but not insurmountable challenge. Of course, those 30 genes already validated using the alternative, ON-TARGET*plus* siRNA library have already cleared this hurdle and await further study. Using the ON-TARGET*plus* library does effectively eliminate off-target candidates, but it does so with significant financial burden relative to the number of candidates that validate. In addition, it risks attrition of true positives due to off-target effects of ON-TARGET*plus* siRNAs. (*PSMB5,* a promising candidate, is a key example.) For the 465 remaining candidates not yet tested with ON-TARGET*plus* siRNAs, we favor a different approach: increasing the stringency of the off-target filter generated from HEK293 RNA-Seq sets. That this filter preferentially eliminated ON-TARGET*plus* non-validating candidates supports its continued use. It is high-yield, low-cost, and adjustable—we can exploit the ON-TARGET*plus* pilot data to generate a receiver-operator characteristic curve and "tune" the exclusion criteria to maximize elimination of false positives and minimize attrition of true positives. One could even consider testing additional ON-TARGET*plus* siRNAs to increase statistical confidence in this method. Alternatively, one could consider constructing a custom single-guide RNA (sgRNA) library of modifiers and leveraging CRISPR-mediated disruption to further eliminate off-

target effects. Advantages of this method include an orthogonal disruption strategy, as well as the availability of computational tools to identify and minimize off-target effects with non-pooled disruption reagents[280,287].

How to Determine Which Candidate Modifiers are Most Promising?

Prioritizing candidates based first on dual activity against GA and FMRpolyG and second on the results of GO analysis is perhaps our best option to identify the modifiers most ripe for further study. With mechanistic investigation of the proteasome already launched with our experiments on *PSMB5,* elucidating the potential roles of NFκB/TNF signaling, RNA methyltransferases/demethylases, and R methyltransferases would likely have the highest yield. (In fact, it is through modulation of NFκB that proteasomal perturbation might impact RAN translation[331].) NFκB/TNF signaling and pharmacological inhibitors thereof have been extensively studied in the context of immunology, rheumatological disease, and multiple cancers[389]. Its implication in RAN translation, if verified, should immediately spark novel therapeutic hypotheses with small-molecule and recombinant therapies borrowed from these fields (with caution[390,391]). With regard to covalent nucleic-acid modifications, modulating the enzymes that mediate such changes has the potential to mitigate neurotoxicity via effects on *C9ORF72* transcription[392], mRNA translation *per se*[335-339], or more global dysregulation of gene expression[268,269].

Once identified, how best to proceed with investigation of likely high-yield RAN modifiers? Further study must prioritize reversal of NRE-elicited toxicity (in *Drosophila*, potentially rodent neurons), validation of RAN-product expression effects in endogenous

systems (in patient-derived iPSCs, quantified by ELISA or CRISPR-inserted tagging), mechanistic novelty, and therapeutic potential. In that regard, we stand to benefit from the number of $(G_4C_2)_n$- and DPR-elicited phenotype-modifier screens published in the last 5 years[171,172,180,182,279,280,383-388]. As a first step, our candidate genes of particular interest should be cross-referenced with published lists of *C9ORF72* modifiers. Modifiers for which suppression of RAN translation in our hands correlates with reversal of NRE- or DPR-elicited toxicity (as for *bel*/*DDX3X*, *eIF4B*, *eIF4H*, and possibly *PSMB5*) might avoid some of the mechanistic confusion we encountered with *HNRNPA2B1*. In addition, if consistent phenotypic effects have been confirmed elsewhere previously, it is prudent to proceed to other high-priority tests rather than essentially repeat these experiments. If such effects have not been confirmed, doing so using *Drosophila* disease models is the fastest, most cost-effective method (provided the cross scheme is not onerous) to arrive at a set of genes with high technical confidence, cross-model validity, and demonstrated therapeutic potential. Alternatively, if a time- and labor-intensive series of crosses is required to generate the experimental genotype, testing should emphasize early validation of RAN-product expression effects in endogenous systems (iPSCs). At that point, mechanistic studies should proceed. Here again, cross-referencing with lists of previously identified DPR-binding proteins, $(G_4C_2)_n$-binding proteins, or components of membraneless organelles[187,195,196,281,388,393] might enable prioritization of modifiers, as well as imply whether coprecipitation or colocalization experiments are appropriate next steps. So contextualizing our screens' modifiers within these previous screens' results will enable us to understand our own data better. And faster.

Appendix A
Supplement to Chapter 2

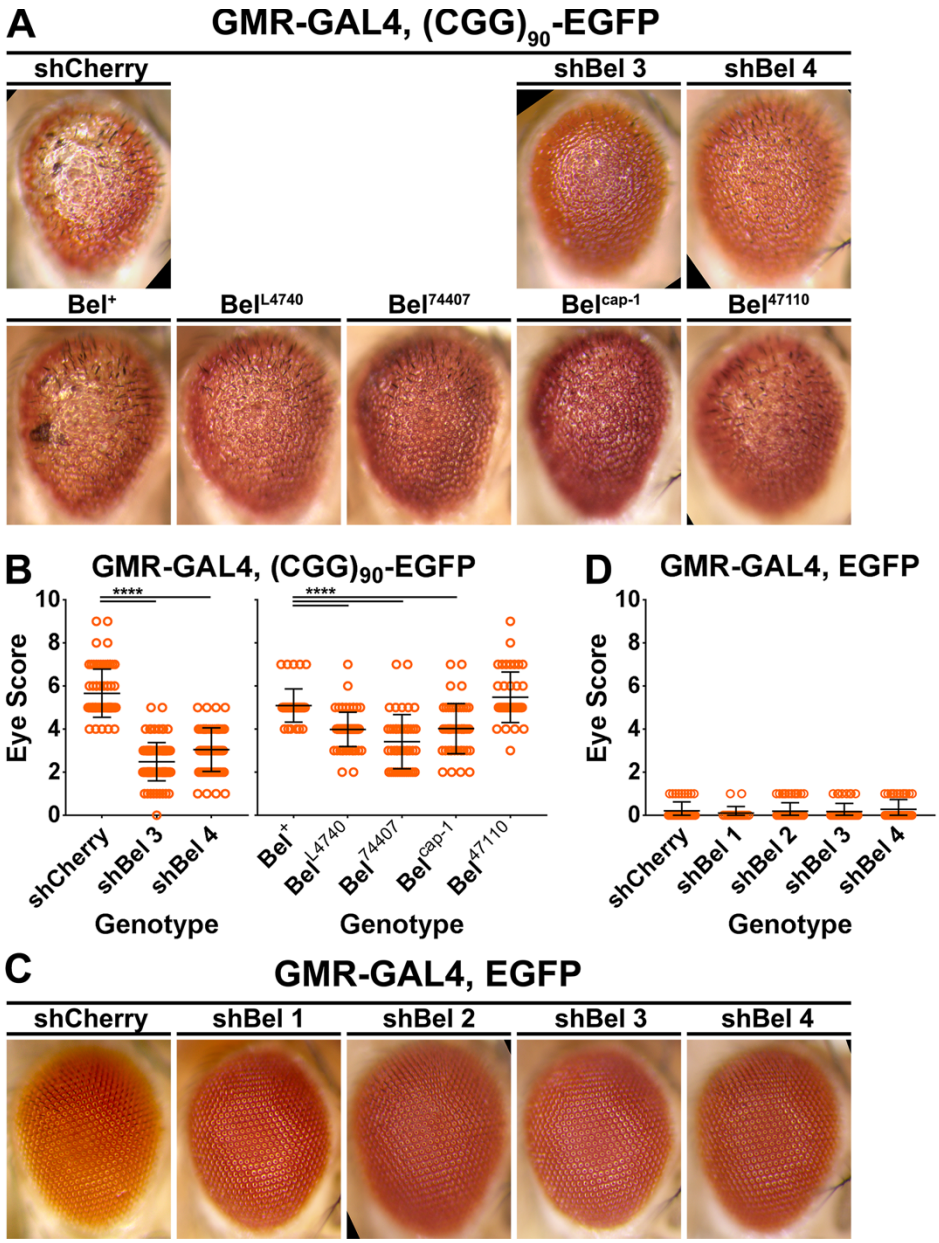


Figure A.1: Supplement to Figure 2.2.

A Representative photographs of fly eyes expressing GMR-GAL4, $(CGG)_{90}$-EGFP with additional *bel* disruptions.

B Quantitation of GMR-GAL4, $(CGG)_{90}$-EGFP eye phenotypes with additional *bel* disruptions (Mann-Whitney U test with Bonferonni corrections for multiple comparisons; *n*=42-98/genotype).

C Representative photographs of fly eyes expressing an AUG-initiated EGFP along with *bel* shRNAs under a GMR-GAL4 driver.

D Quantitation of GMR-GAL4, EGFP eye phenotypes with *bel* shRNAs (*n*=20-64/genotype).

Data Information: For all panels, **** P≤0.0001 for the specified statistical test (compiled from ≥3 replicates).

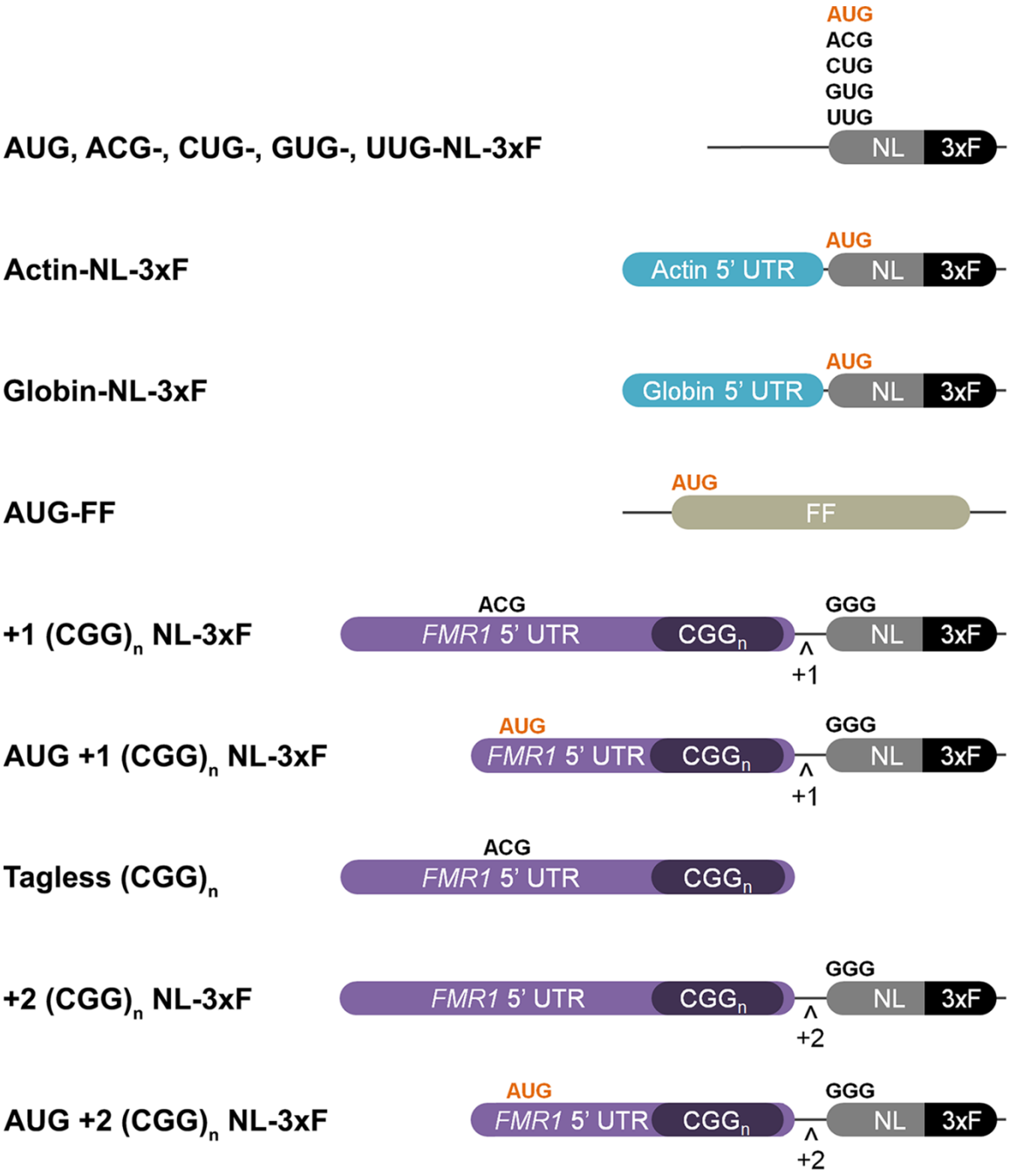


Figure A.2: NL-3xF and FF Reporter Constructs Used in This Study.
The tag (e.g., NL-3xF, FF, tagless), 5' leader sequence (e.g., β Actin or β Globin 5' UTR), start codon (e.g., AUG, ACG, GGG), and tag reading frame relative to the CGG repeat (e.g., +1, +2) are indicated.

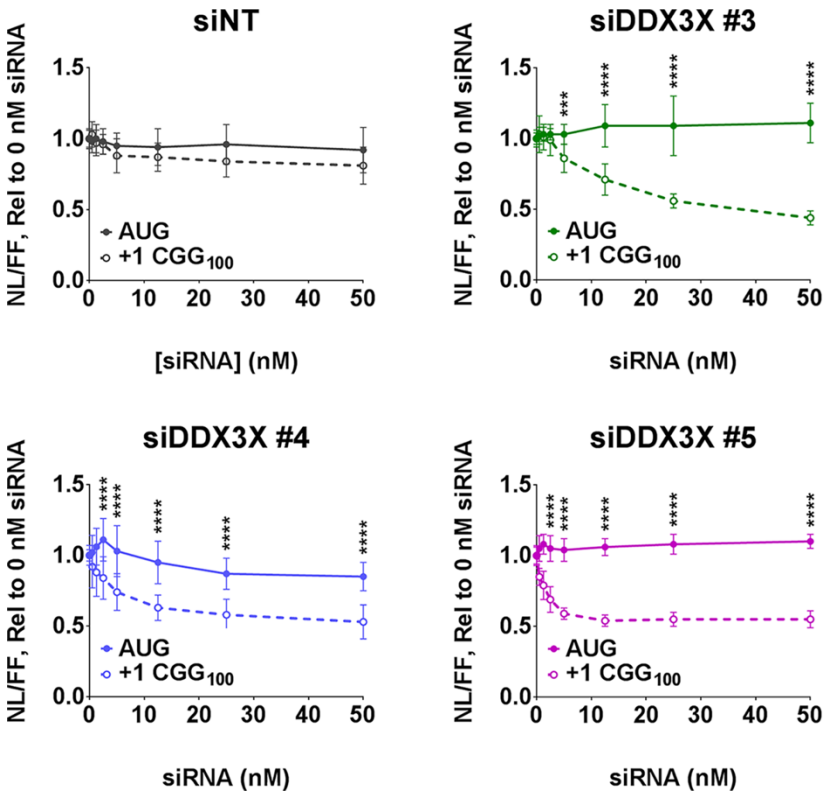


Figure A.3: Supplement to Figure 2.4.
Asterisks indicate comparisons between relative AUG-NL-3xF and +1 $(CGG)_{100}$ NL-3xF expression (two-way ANOVA with Sidak's multiple comparisons test; *n*=9-21/condition).

Data Information: *** P≤0.001, **** P≤0.0001 for the specified statistical test at the indicated siRNA concentration. All data presented as means ± SD (compiled from ≥3 replicates).

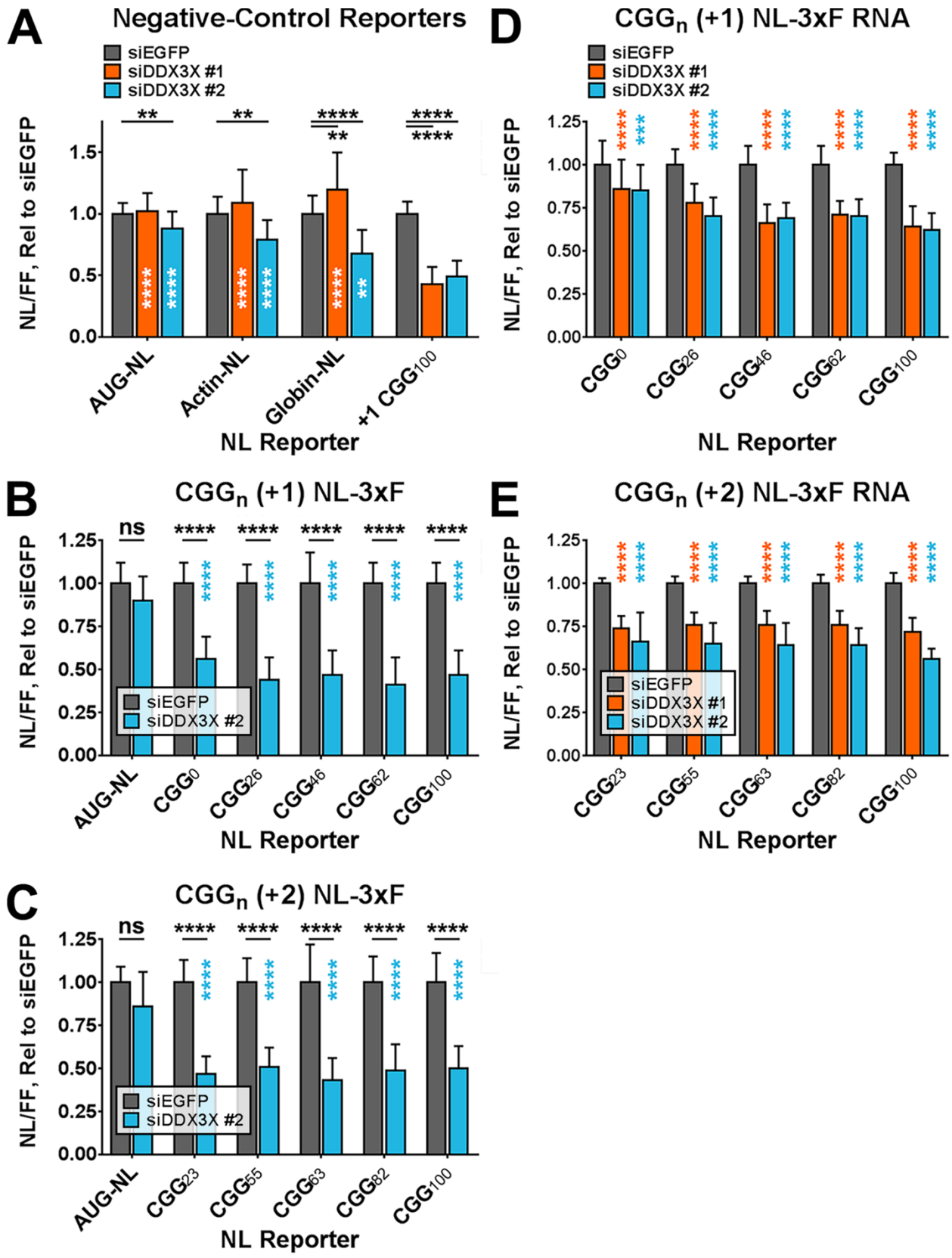

A
Negative-Control Reporters
siEGFP
siDDX3X #1
siDDX3X #2
NL/FF, Rel to siEGFP
AUG-NL
Actin-NL
Globin-NL
+1 CGG100
NL Reporter
B
CGGn (+1) NL-3xF
ns
siEGFP
siDDX3X #2
AUG-NL
CGG0
CGG26
CGG46
CGG62
CGG100
NL Reporter
C
CGGn (+2) NL-3xF
ns
siEGFP
siDDX3X #2
AUG-NL
CGG23
CGG55
CGG63
CGG82
CGG100
NL Reporter
D
CGGn (+1) NL-3xF RNA
siEGFP
siDDX3X #1
siDDX3X #2
CGG0
CGG26
CGG46
CGG62
CGG100
NL Reporter
E
CGGn (+2) NL-3xF RNA
siEGFP
siDDX3X #1
siDDX3X #2
CGG23
CGG55
CGG63
CGG82
CGG100
NL Reporter

Figure A.4: Supplement to Figures 2.4 and 2.5.

A Expression in HeLa cells of transfected AUG-NL-3xF and +1 $(CGG)_{100}$ NL-3xF reporter plasmids, with and without *DDX3X* knockdown, compared to the expression of AUG-initiated NL-3xF reporters bearing the short, minimally-structured 5' UTRs of human β actin and β globin (two-way ANOVA with Tukey's multiple comparisons test; *n*=15-39/condition). Black asterisks refer to comparisons between siEGFP-treated and siDDX3X-treated cells; white asterisks refer to comparisons between siDDX3X-treated cells expressing +1 $(CGG)_{100}$ NL-3xF and those expressing a different reporter.

B, C $(CGG)_n$ +1 and $(CGG)_n$ +2 NL-3xF expression with and without *DDX3X* knockdown across a range of CGG repeat sizes (two-way ANOVA with Tukey's multiple comparisons test; *n*=12-18/condition). Black asterisks refer to comparisons between siDDX3X- and siEGFP-treated cells; blue asterisks refer to comparisons between siDDX3X-treated cells expressing AUG-NL-3xF and those expressing a different reporter.

D, E Expression of *in vitro*-transcribed $(CGG)_n$ +1 and $(CGG)_n$ +2 NL-3xF reporter RNAs with and without *DDX3X* knockdown across a range of CGG repeat sizes (two-way ANOVA with Tukey's multiple comparisons test; *n*=12-24/condition). Orange and blue asterisks refer to comparisons between siDDX3X- and siEGFP-treated cells.

Data Information: For all panels, ns=non-significant, ** P≤0.01, *** P≤0.001, **** P≤0.0001 for the specified statistical test. All panels depict data as means + SD (compiled from ≥3 replicates).

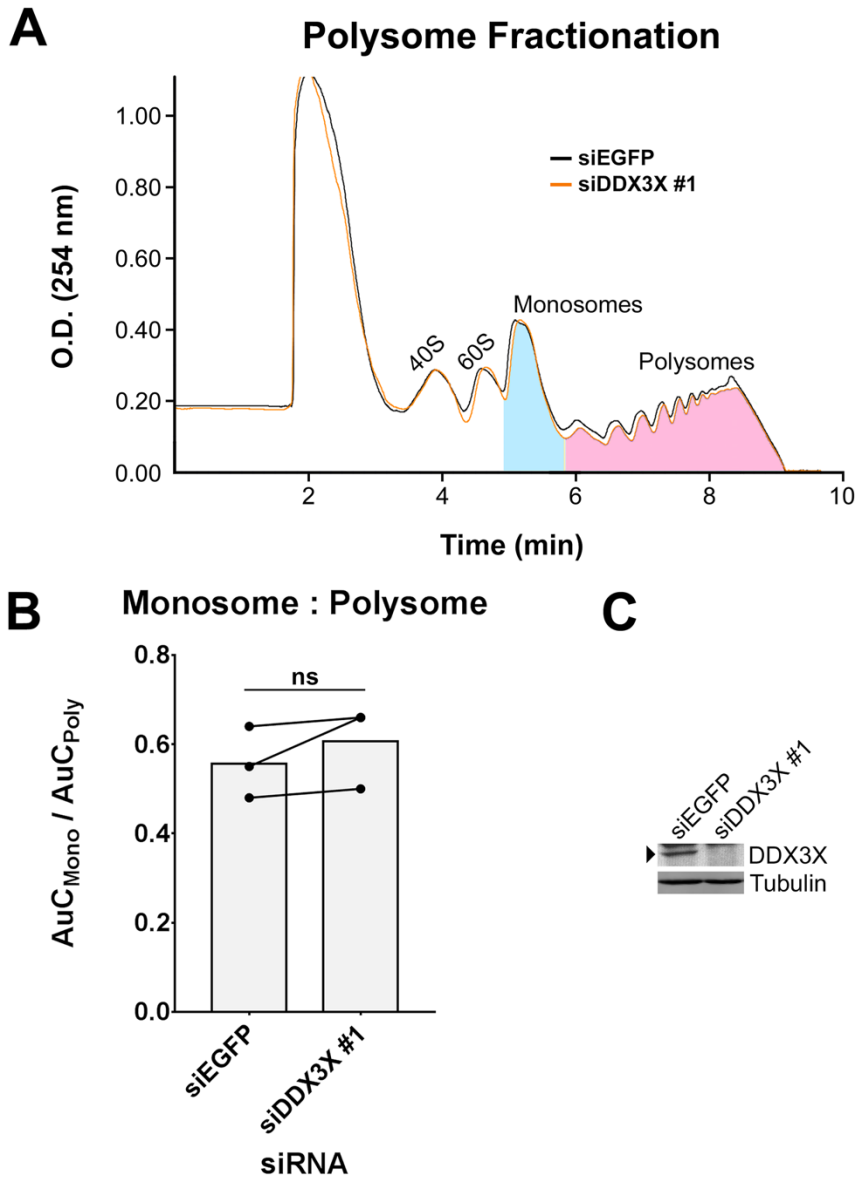


Figure A.5: Knockdown of *DDX3X* Does Not Inhibit Global Translation.

A Representative polysome-fractionation profiles of HeLa cell lysates transfected with siDDX3X #1 or siEGFP. The areas-under-the-curve (AuC) for monosomes and polysomes are shaded blue and pink, respectively.

B Mean ratios of the AuCs of monosomes to polysomes across three replicates (Student's paired t-test, *n*=3/condition). ns=non-significant.

C Anti-DDX3X western blot of HeLa lysates processed for polysome fractionation. The DDX3X band is indicated.

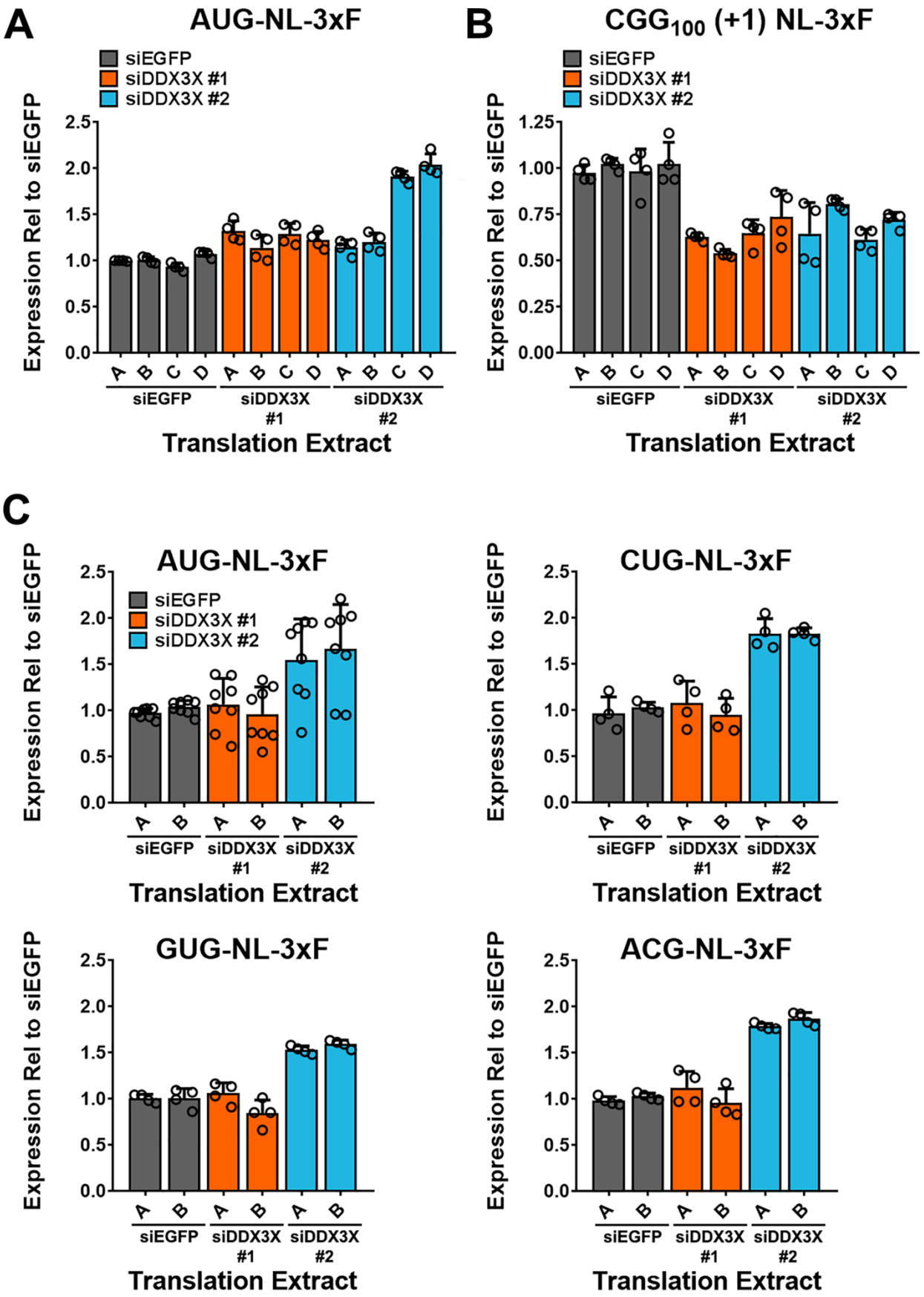

A
AUG-NL-3xF
siEGFP
siDDX3X #1
siDDX3X #2
Expression Rel to siEGFP
2.5
2.0
1.5
1.0
0.5
0.0
A B C D A B C D A B C D
siEGFP
siDDX3X #1
siDDX3X #2
Translation Extract
B
CGG100 (+1) NL-3xF
siEGFP
siDDX3X #1
siDDX3X #2
Expression Rel to siEGFP
1.25
1.00
0.75
0.50
0.25
0.00
A B C D A B C D A B C D
siEGFP
siDDX3X #1
siDDX3X #2
Translation Extract
C
AUG-NL-3xF
siEGFP
siDDX3X #1
siDDX3X #2
Expression Rel to siEGFP
2.5
2.0
1.5
1.0
0.5
0.0
A B A B A B
siEGFP
siDDX3X #1
siDDX3X #2
Translation Extract
CUG-NL-3xF
Expression Rel to siEGFP
2.5
2.0
1.5
1.0
0.5
0.0
A B A B A B
siEGFP
siDDX3X #1
siDDX3X #2
Translation Extract
GUG-NL-3xF
Expression Rel to siEGFP
2.5
2.0
1.5
1.0
0.5
0.0
A B A B A B
siEGFP
siDDX3X #1
siDDX3X #2
Translation Extract
ACG-NL-3xF
Expression Rel to siEGFP
2.5
2.0
1.5
1.0
0.5
0.0
A B A B A B
siEGFP
siDDX3X #1
siDDX3X #2
Translation Extract

Figure A.6: Reporter Expression in *In Vitro* Translation Assays is Consistent Across Independently Prepared, Replicate Extracts.

A, B Expression of *in vitro*-transcribed AUG-NL-3xF (A) and +1 $(CGG)_{100}$ NL-3xF RNAs (B) in *in vitro* translation assays, collected from HeLa cells treated with siRNAs against EGFP or *DDX3X*. Four replicate lysates ("A-D") were generated per siRNA (*n*=4/lysate).

C Expression of *in vitro*-transcribed near-AUG reporter RNAs in *in vitro* translation extracts, collected from HeLa cells treated with siRNAs against EGFP or *DDX3X*. Reporter RNAs were tested in duplicate lysates (2 lysates per siRNA; n=*4*/lysate).

Data Information: All panels depict pooled data (means + SD) gathered across two replicates.

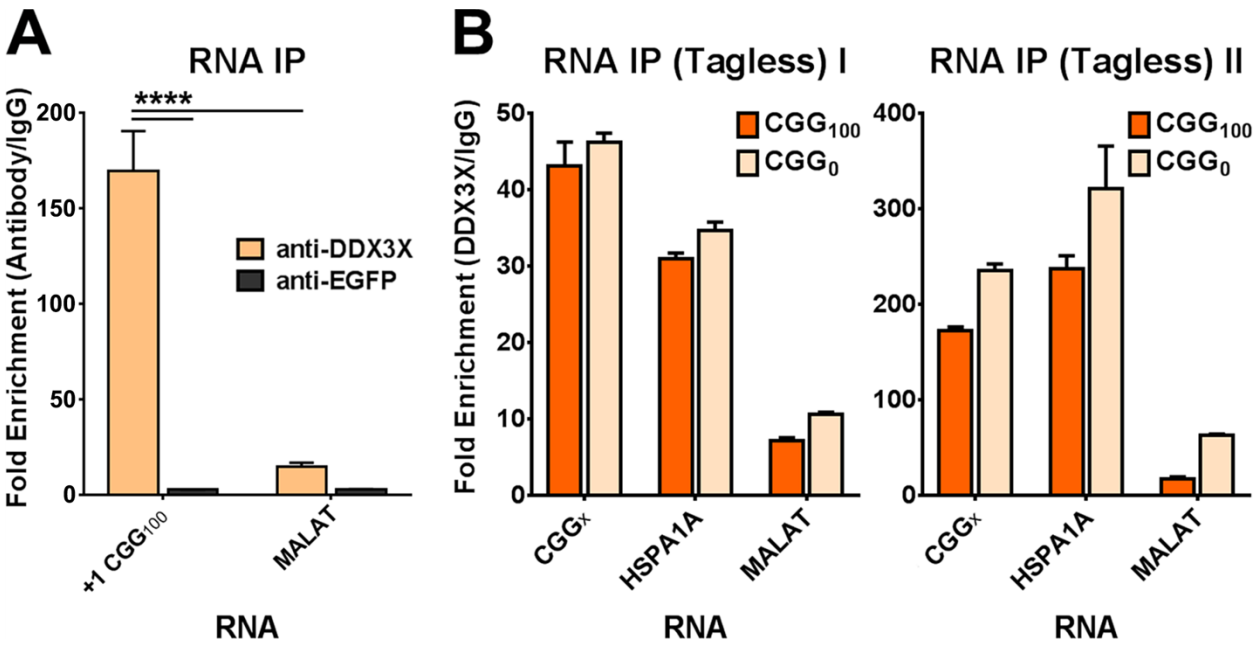


Figure A.7: Anti-DDX3X RIP Co-Precipitates +1 $(CGG)_{100}$ NL-3xF RNA Independently of the NL-3xF Tag or NRE Size.

A Enrichment of +1 $(CGG)_{100}$ NL-3xF and *MALAT* RNA following anti-DDX3X and anti-EGFP RIP, relative to incubation with isotype control IgG (Student's t test, *n*=3). In addition, +1 $(CGG)_{100}$ NL-3xF mRNA is not enriched following anti-EGFP RIP from cells expressing EGFP (Student's t test, *n*=3). This experiment is a replicate of that presented in Figure 2.5D.

B Enrichment of $(CGG)_{100}$, $(CGG)_0$, and *HSPA1A*, but not *MALAT* RNA, following anti-DDX3X RIP in two independent replicates. In contrast with panel A and Figure 2.5D, the $(CGG)_{100}$ and $(CGG)_0$ constructs here do not have a NL-3xF tag.

Data Information: **** P≤0.0001. All panels depict technical triplicates (means + SD) within the given experiment.

+1 (CGG)$_n$ NL-3xF

TAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTA
AGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTTA
CAGATCTGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGGCCG
CCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGGCGGCGGCGGT
G**ACG**GAGGCGCCGCTGCCAGGGGGCGTGCGGCAGCG(**CGG**)$_n$CTGG
GCCTCGAGGATATCAAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCGGT
ACTGTTGGTAAAGCCACCGGGGTCTTCACACTC...

AUG +1 (CGG)$_{100}$ NL-3xF

TAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTA
AGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTTA
CACC**ATG**GCGCCGCTGCCAGGGGGCGTGCGGCAGCG(**CGG**)$_{100}$CTGGG
CCTCGAGGATATCAAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCGG
TACTGTTGGTAAAGCCACCGGGGTCTTCACACTC...

+2 (CGG)$_n$ NL-3xF

TAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTA
AGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTTA
ACAGATCTGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGGCCG
CCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGGCGGCGGCGGT
GACGGAGGCGCCGCTGCCAGGGGGCGTGCGGCAGCG(**CGG**)$_n$CTGGG
CCATCAAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCGGTACTGTTG
GTAAAGCCACCGGGGTCTTCACACTC...

AUG +2 (CGG)$_{100}$ NL-3xF

TAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTAAACTTA
AGCTTGGTACCGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTT
ACACC**ATG**GGGCGCCGCTGCCAGGGGGCGTGCGGCAGCG(**CGG**)$_{100}$CTG
GGCCTCGAGATTCGATCGTCGTGATATCAAGATCTGGCCTCGGCGG
CCAAGCTTGGCAATCCGGTACTGTTGGTAAAGCCACCGGGGTCTTC
ACACTC...

T7 Promotor Site
FMR1 5' UTR
NL Open Reading Frame

AUG Start Codon
Non-AUG Start Codon
CGG NRE

Figure A.8: Sequence Maps of the +1 and +2 (CGG)$_{100}$ NL-3xF Reporter Constructs, With and Without an AUG Inserted 5' to the CGG NRE.

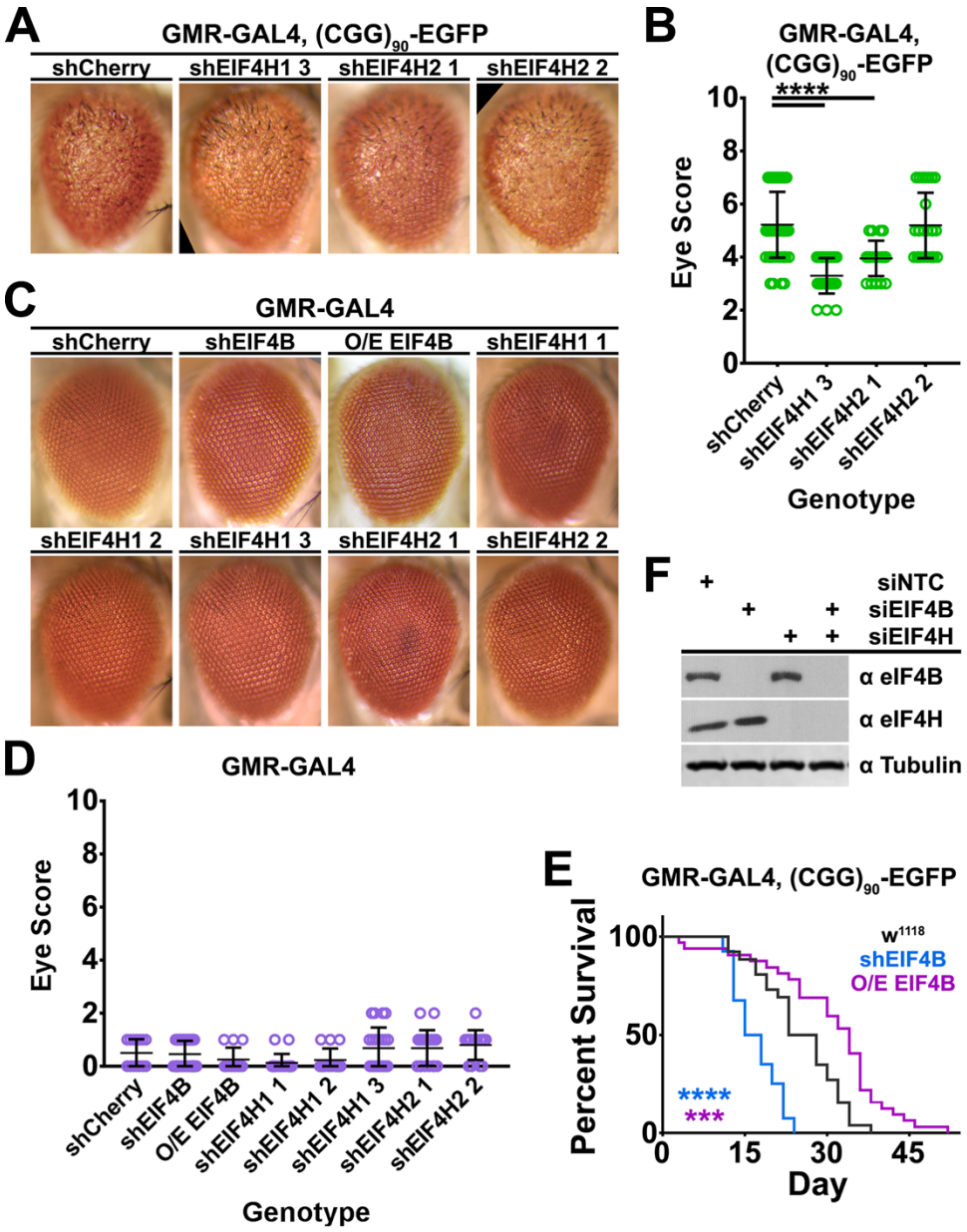


Figure A.9: Supplement to Figures 2.7 and 2.8.

A Representative photographs of fly eyes expressing GMR-GAL4, $(CGG)_{90}$-EGFP with additional *eIF4H1* and *eIF4H2* shRNAs.

B Quantitation of GMR-GAL4, $(CGG)_{90}$-EGFP eye phenotypes with additional *eIF4H1* and *eIF4H2* shRNAs (Mann-Whitney U test with Bonferonni corrections for multiple comparisons; *n*=21-55/genotype).

C Representative photographs of fly eyes expressing *eIF4H1* and *eIF4H2* shRNAs alone under a GMR-GAL4 driver.

D Quantitation of eye phenotypes of *eIF4B*, *eIF4H1*, and *eIF4H2* shRNAs expressed under a GMR-GAL4 driver (*n*=12-33/genotype).

E Longevity assays of $(CGG)_{90}$-EGFP; Tub5-GS flies with knockdown and overexpression of *eIF4B* (Log-rank Mantel-Cox test with Bonferroni corrections for multiple comparisons; *n*=26-32/genotype).

F Anti-eIF4B and anti-eIF4H western blot of HeLa cells transfected with siRNAs against *EIF4B* or *EIF4H*.

Data Information: For all panels, ns=non-significant, *** P≤0.001, **** P≤0.0001 for the specified statistical test. All panels depict data as means ± SD (compiled from ≥3 replicates).

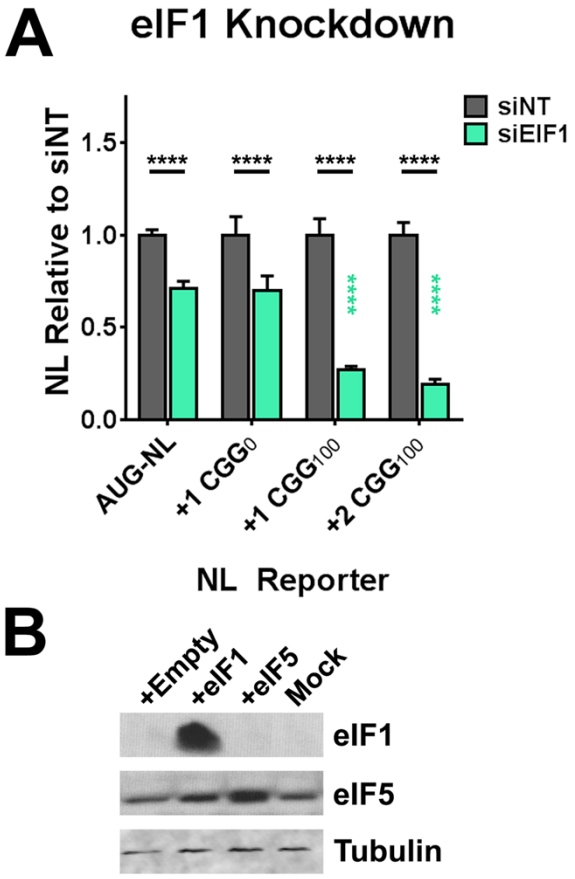


Figure A.10: Supplement to Figure 2.9.

A Expression of plasmid-based NL-3xF reporters in HEK293 cells with and without knockdown of *EIF1* (two-way ANOVA with Sidak's multiple comparisons test; *n*=12/condition). Black asterisks refer to comparisons between siNT-transfected and siEIF1-transfected cells; green asterisks refer to comparisons between siEIF1-transfected cells expressing AUG-NL-3xF and those expressing a different reporter. Graph depicts data as means + SD (compiled from ≥3 replicates). **** P≤0.0001.

B Anti-eIF1 and -eIF5 western blot demonstrating overexpression of eIF1 and eIF5 after transfection of the respective plasmid constructs.

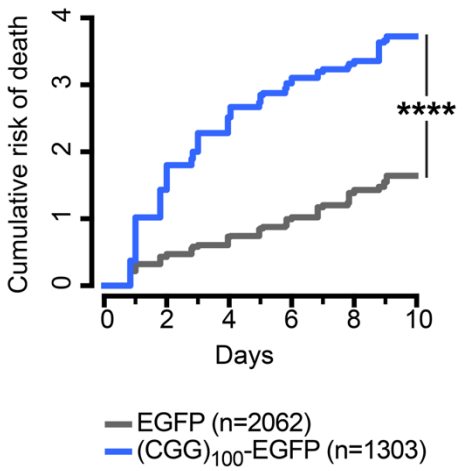


Figure A.11: Supplement to Figure 2.10.
Transfection of +1 $(CGG)_{100}$ EGFP plasmid-based reporters increased the cumulative risk of death in primary rodent neurons, relative to transfection of EGFP reporters (Cox proportional hazard analysis; *n*=1303-2062 cells/condition). **** P<0.0001. Data compiled over 3 replicates.

		GMR-GAL4, $(CGG)_{90}$-EGFP	GMR-GAL4
Control	shCherry		
Belle	shBelle 1		
	shBelle 2		
	shBelle 3		
	shBelle 4		
	$Belle^{6}$		
	$Belle^{EKE}$		
	$Belle^{L4740}$		
	$Belle^{74407}$		
	$Belle^{cap-1}$		
	$Belle^{47110}$		
eIF4B, eIF4H	shEIF4B		
	UAS-EIF4B		
	shEIF4H1 1		
	shEIF4H1 2		
	shEIF4H1 3		
	shEIF4H2 1		
	shEIF4H2 2		
eIF1	shEIF1 1		
	shEIF1 2		
	$EIF1^{EY02210}$		
	shEIF1 3		
	shEIF1 4		
	shEIF1 5		
eIF1A	shEIF1A 1		
	shEIF1A 2		
	shEIF1A 3		
	shEIF1A 4		
	$EIF1A^{645}$		
	$EIF1A^{2232}$		
	$EIF1A^{c04533}$		
	$EIF1A^{EP935}$		
	UAS-EIF1A		

		GMR-GAL4, $(CGG)_{90}$-EGFP	GMR-GAL4
eIF2α	shEIF2α 1		
	shEIF2α 2		
	shEIF2α 3		
	$EIF2α^{815-29}$		
	UAS-EIF2α1		
eIF3	$EIF3B^{EY14430}$		
eIF4A	shEIF4A		
	UAS-EIF4A		
	$EIF4A^{1013}$		
eIF5, eIF5B	shEIF5 1		
	shEIF5 2		
	UAS-EIF5		
	$EIF5B^{09143}$		
	$EIF5B^{EY01401}$		
RNA Helicases	shRHAU		
	UAS-RHAU		
	shDHX57		
	shBGCN 1		
	shBGCN 2		
	shBGCN 3		
	$BGFCN^{KG08129}$		
Ribosomal Proteins	shRPS25 1		
	shRPS25 2		
RNA-Binding Proteins	shSF2 1		
	shSF2 2		

Suppressor	
Enhancer	
No Effect	

Table A.1: Summary of the Candidate-Based Screen for Modifiers of $(CGG)_{90}$-Elicited Toxicity. Candidate modifier genes for RAN translation were crossed to flies carrying either GMR-GAL4,

$(CGG)_{90}$-EGFP or GMR-GAL4 alone and their eye phenotypes were assessed. Blue indicates suppression of the rough-eye phenotype. Red indicates exacerbation or induction of a rough-eye phenotype. Each candidate modifier was screened against both GMR-GAL4, $(CGG)_{90}$-EGFP flies and GMR-GAL4 alone across a minimum of two independent crosses with at least 25 flies evaluated per cross.

Fly Stock	Source	Catalog #
shCherry	BDSC	35785
shBel 1	BDSC	35185
shBel 2	BDSC	35302
shBel 3	VDRC	6299
shBel 4	BDSC	28049
w^{1118}	BDSC	5905
Bel^{6}	BDSC	4024
Bel^{EKE}	Deng Lab	-
Bel^{L7470}	Deng Lab	-
Bel^{74407}	Deng Lab	-
Bel^{cap-1}	BDSC	1178
Bel^{47110}	Deng Lab	-
shEIF4B	BDSC	31364
UAS-EIF4B	Todd Lab	-
shEIF4H1 1	VDRC	100817
shEIF4H1 2	VDRC	34301
shEIF4H1 3	VDRC	48119
shEIF4H2 1	VDRC	102825
shEIF4H2 2	VDRC	32192
shEIF1 1	VDRC	29216
shEIF1 2	BDSC	55232
$EIF1^{EY02210}$	BDSC	15406
shEIF1 3	BDSC	57174
shEIF1 4	VDRC	29215
shEIF1 5	VDRC	105763
shEIF1A 1	VDRC	100611
shEIF1A 2	BDSC	29316
shEIF1A 3	BDSC	31185
shEIF1A 4	VDRC	26022
$EIF1A^{645}$	BDSC	23925
$EIF1A^{2232}$	BDSC	23941
$EIF1A^{c04533}$	BDSC	11495
$EIF1A^{EP935}$	BDSC	17203
UAS-EIF1A	FlyORF	F000848

Fly Stock	Source	Catalog #
shEIF2α 1	VDRC	7799
shEIF2α 2	VDRC	104562
shEIF2α 3	VDRC	7798
$EIF2α^{815-29}$	BDSC	4926
UAS-EIF2α	FlyORF	F000983
$EIF3B^{EY14430}$	BDSC	20931
shEIF4A	BDSC	33970
UAS-EIF4A	Xie Lab	-
$EIF4A^{1013}$	BDSC	8647
shEIF5 1	BDSC	34841
shEIF5 2	VDRC	29070
UAS-EIF5	BDSC	22132
$EIF5B^{09143}$	BDSC	11735
$EIF5B^{EY01401}$	BDSC	19641
shRHAU	VDRC	44984
St r	BDSC	20040
shDHX57	BDSC	55373
shBGCN 1	VDRC	108334
shBGCN 2	VDRC	25590
shBGCN 3	BDSC	36636
$BGCN^{KG08129}$	BDSC	14687
shRPS25 1	VDRC	101342
shRPS25 2	VDRC	52602
shSF2 1	BDSC	29522
shSF2 2	BDSC	32367
GMR-GAL4	BDSC	8605
Tub5-GS	Pletcher Lab	-
ElaV-GS	Pletcher Lab	-
CGG_{90}-EGFP	Jin Lab	-
EGFP	BDSC	6874

Table A.2: Fly Stocks Used in This Study and Their Sources.

Primer Name	Primer Sequence (5’ to 3’)
EGFP (Forward)	TCTTCTTCAAGGACGACGGCAACTAC
EGFP (Reverse)	GTACTCCAGCTTGTGCCCCAGGATGT
Belle (Forward)	CAGTAGCTTGTGGAACGTAAGAAGTTT
Belle (Reverse)	TTACTCATATTATCCTCCAATCAGTTGC
RPL32 Dmel (Forward)	GTTGTGCACCAGGAACTTCTTGAATCCG
RPL32 Dmel (Reverse)	CTTCCAGCTTCAAGATGACCATCCGC
Nanoluciferase (Forward)	GGTGGTGTACCCTGTGGATG
Nanoluciferase (Reverse)	AACCCCGTCGATTACCAGTG
Firefly luciferase (Forward)	GCAGTACCGGATTGCCCAAG
Firefly luciferase (Reverse)	GTCGGGGATGATCTGGTTGC
MALAT (Forward)	TGGTGATGAAGGTAGCAGGC
MALAT (Reverse)	GGCATGCTGGTCTAGGATCC

Table A.3: Primers Used in This Study for qRT-PCR.

Appendix B
Supplement to Chapter 3

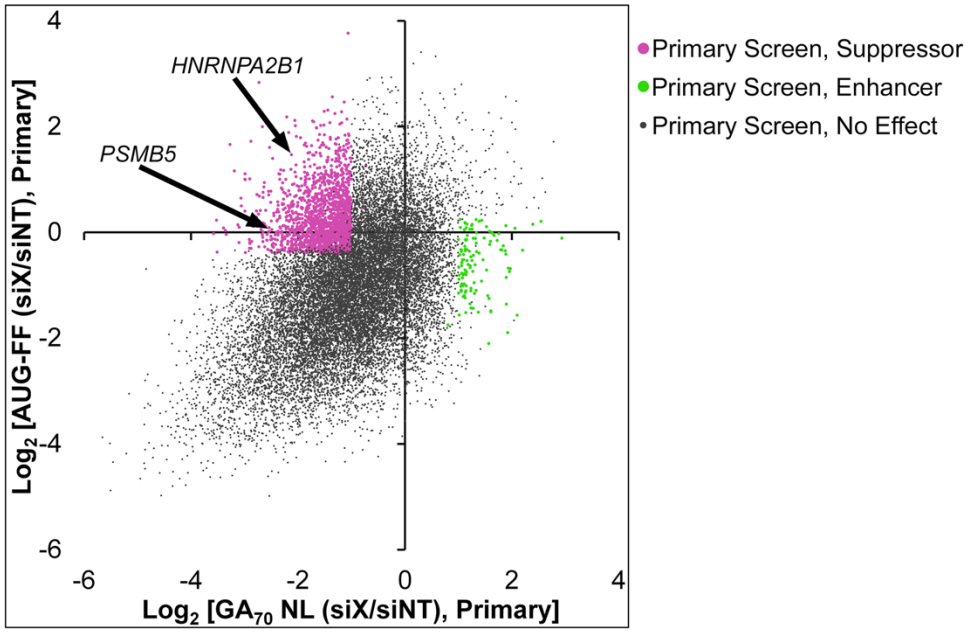
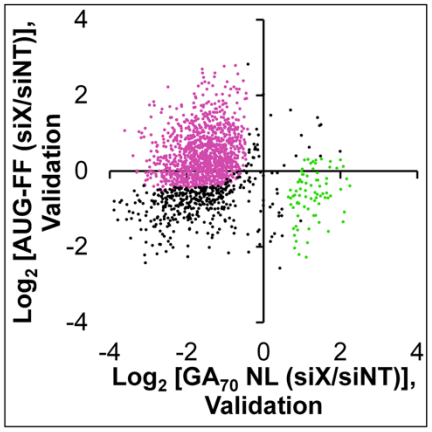


Figure B.1: Primary Screen Identifies Candidate Suppressors and Enhancers of *C9ORF72* RAN Translation.

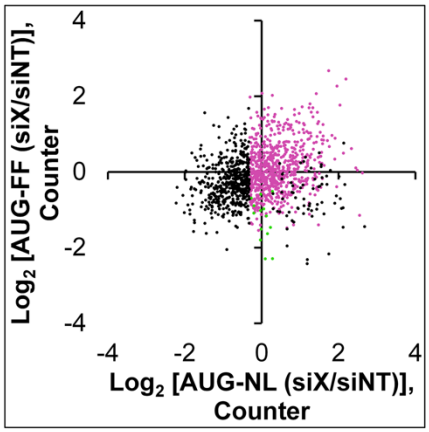
Scatterplot of Log_2 GA_{70} NL vs. Log_2 AUG-FF values, each relative to siNT-transfected cells, gathered across the entire *C9ORF72* primary screen. siRNAs are colored based on whether they suppressed, enhanced, or had no effect on GA_{70} NL-3xF. Data points corresponding to *HNRNPA2* and *PSMB5* are indicated.

Primary Hits



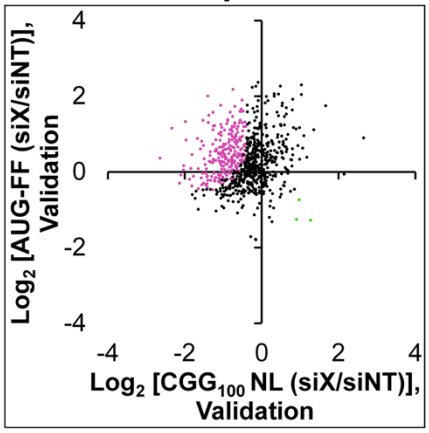
- C9-Validated, Suppressor
- C9-Validated, Enhancer
- Did Not C9-Validate

Primary Hits, C9-Validated



- C9-Validated, RAN-Specific Suppressor
- C9-Validated, RAN-Specific Enhancer
- Not RAN-Specific

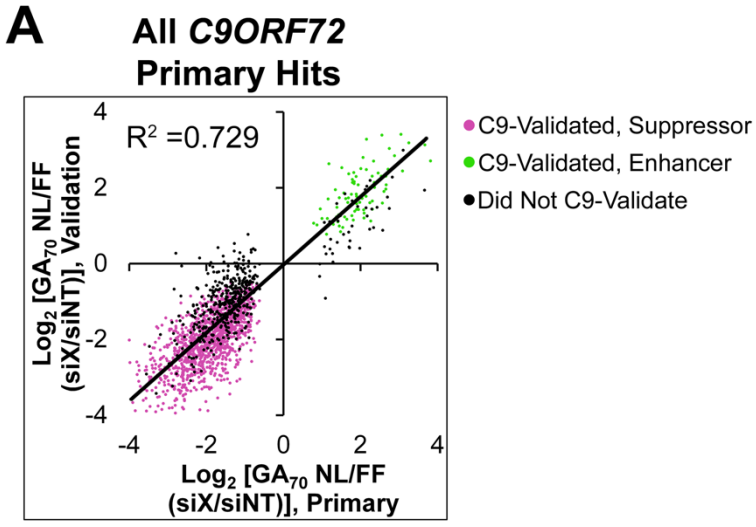
Primary Hits, C9-Validated, RAN-Specific



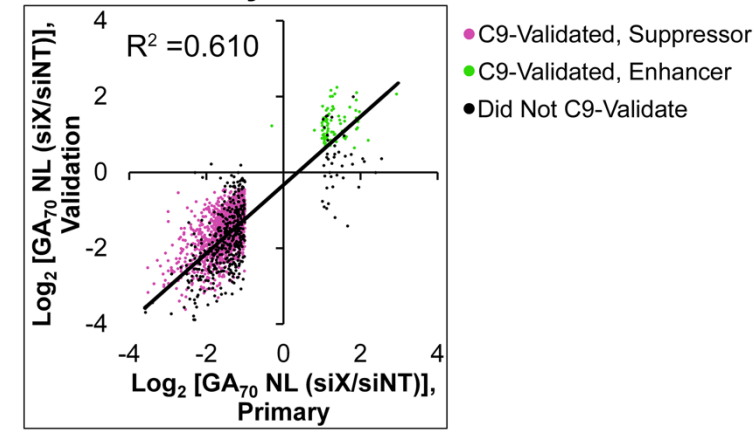
- C9-Validated, RAN-Specific, FMR1-Validated Suppressor
- C9-Validated, RAN-Specific, FMR1-Validated Enhancer
- Did Not FMR1-Validate

Figure B.2: Serial Validation and Refinement of RAN-Specific *C9ORF72* Candidate Modifiers. Scatterplot of Log_2 GA_{70} NL vs. Log_2 AUG-FF values gathered during the *C9ORF72*-validation screen (top panel), Log_2 AUG NL vs. Log_2 AUG-FF values gathered during the AUG counter screen (middle panel), and Log_2 CGG_{100} NL vs. Log_2 AUG-FF values gathered during the *FMR1* validation screen (bottom panel). Data points are colored based on whether the corresponding siRNAs progressed to the subsequent screen.

A


All C9ORF72 Primary Hits
R² =0.729
Log2 [GA70 NL/FF (siX/siNT)], Validation
Log2 [GA70 NL/FF (siX/siNT)], Primary
C9-Validated, Suppressor
C9-Validated, Enhancer
Did Not C9-Validate

B


All C9ORF72 Primary Hits
R² =0.610
Log2 [GA70 NL (siX/siNT)], Validation
Log2 [GA70 NL (siX/siNT)], Primary
C9-Validated, Suppressor
C9-Validated, Enhancer
Did Not C9-Validate

C

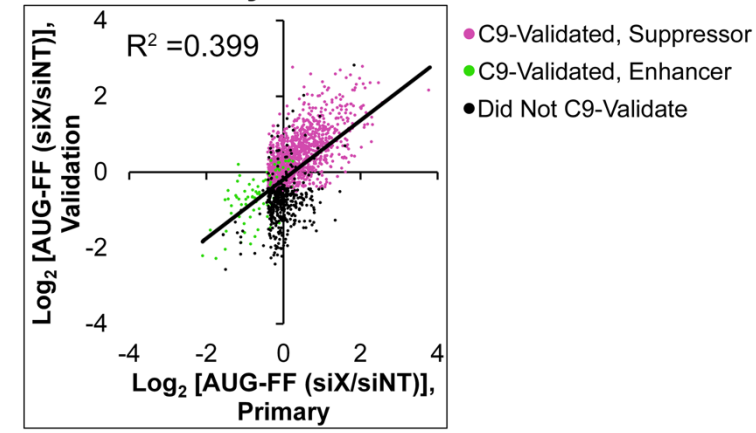

All C9ORF72 Primary Hits
R² =0.399
Log2 [AUG-FF (siX/siNT)], Validation
Log2 [AUG-FF (siX/siNT)], Primary
C9-Validated, Suppressor
C9-Validated, Enhancer
Did Not C9-Validate

Figure B.3: The Failure of Some Candidate *C9ORF72* Modifiers to Validate Was Driven by Inconsistencies in Effects on AUG-FF Expression, not GA_{70} NL-3xF Expression.

A Scatterplot of Log_2 GA_{70} NL/FF values in the primary vs. the *C9ORF72* validation screen. All primary hits are graphed. Each is colored depending on whether it validated against the *C9ORF72* reporter (Pearson's correlation; *n*=1,734; P<1_E-6).

B Scatterplot of Log_2 GA_{70} NL values in the primary vs. the *C9ORF72* validation screen. All primary hits are graphed. Each is colored depending on whether it validated against the *C9ORF72* reporter (Pearson's correlation; *n*=1,734; P<1_E-6).

C Scatterplot of Log_2 AUG-FF values in the primary vs. the *C9ORF72* validation screen. All primary hits are graphed. Each is colored depending on whether it validated against the *C9ORF72* reporter (Pearson's correlation; *n*=1,734; P<1_E-6).

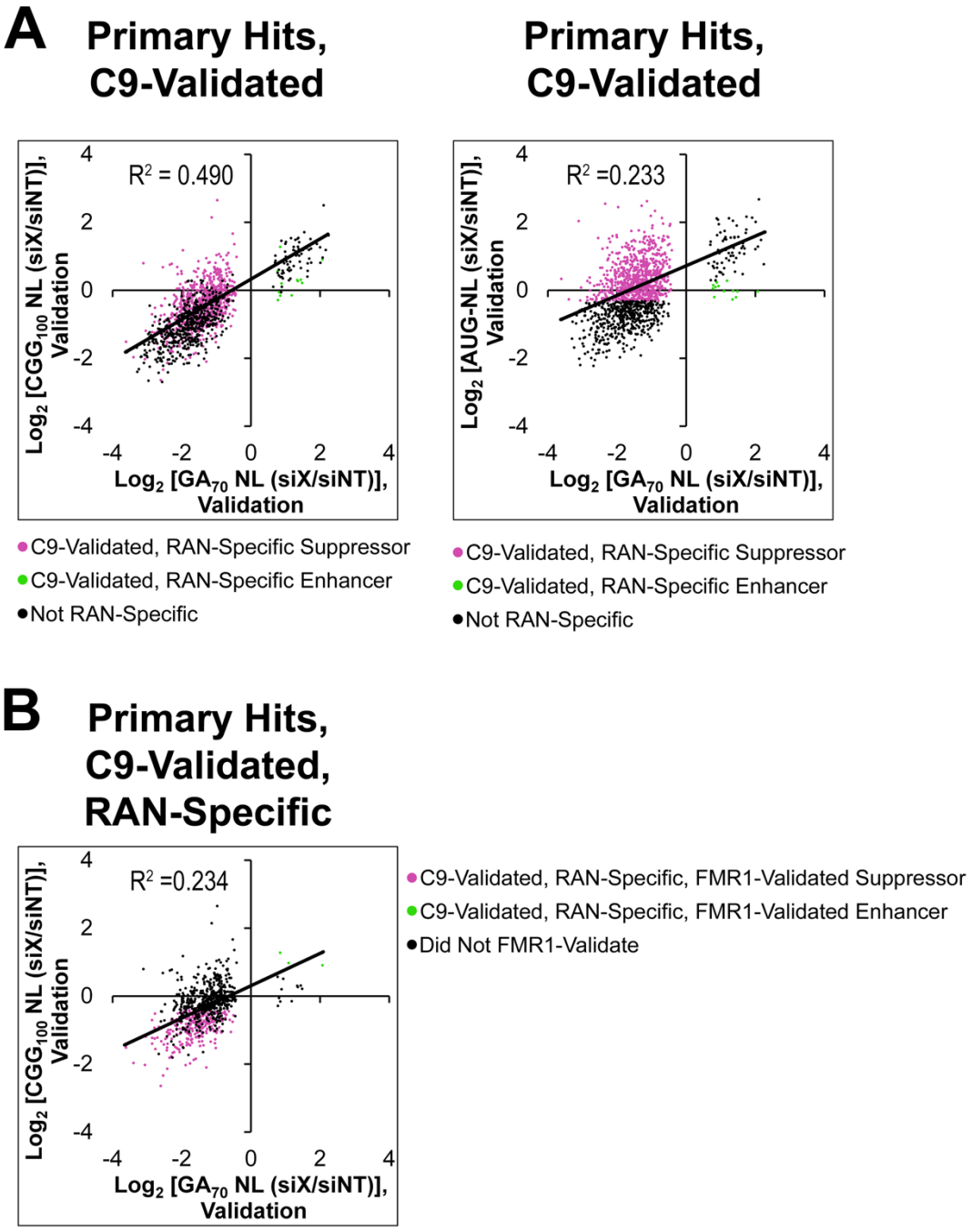


Figure B.4: The AUG Counter Screen Eliminates Several Hits That Modified Both *C9ORF72* and *FMR1* Reporters.

A Scatterplots of validation Log_2 GA_{70} NL values vs. validation Log_2 CGG_{100} NL values (left panel; Pearson's correlation; *n*=1,307; P<1E-6) and validation Log_2 GA_{70} NL data vs. Log_2 AUG-NL data (right panel; Pearson's correlation; *n*=1,307; P<1E-6). Each point represents a primary hit that successfully validated against the *C9ORF72* reporter. These have been categorized and colored based on whether they were eliminated during the AUG counter screen.

B Scatterplot of validation Log_2 GA_{70} NL data vs. validation Log_2 CGG_{100} NL values, plotting only *C9ORF72*-validated and RAN-specific siRNAs. siRNAs have been categorized ad colored

based on whether they validated against the *FMR1* reporter (Pearson's correlation; n=753; P<1E-6).

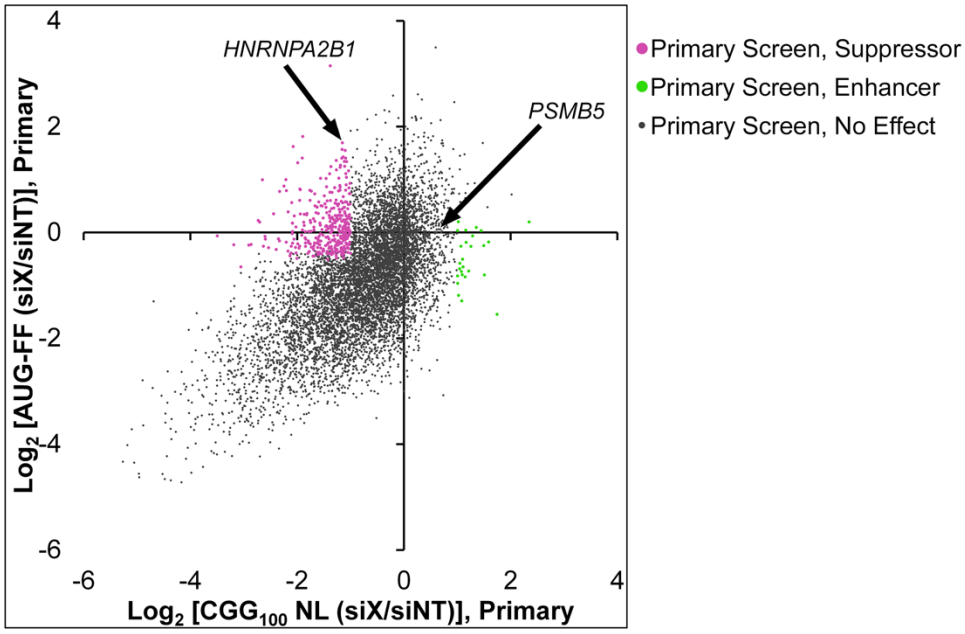
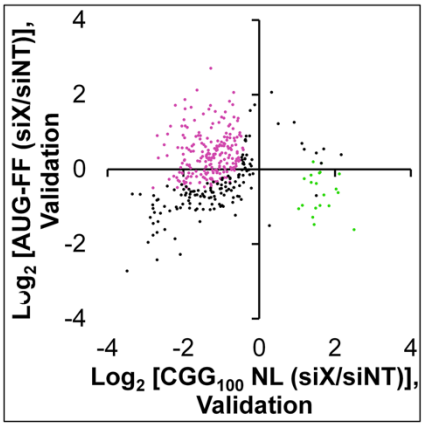


Figure B.5: An Independent Primary Screen Identifies Candidate Suppressors and Enhancers of *FMR1* RAN Translation.

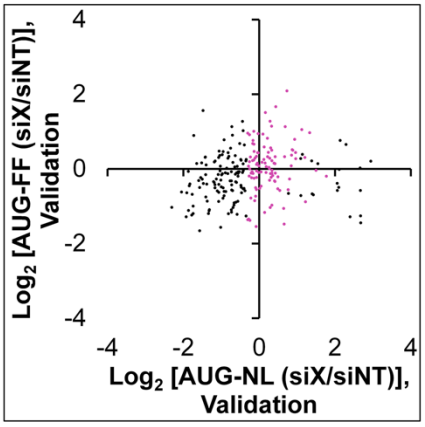
Scatterplot of Log_2 CGG_{100} NL vs. Log_2 AUG-FF values, each relative to siNT-transfected cells, gathered across the *FMR1* primary screen of Drug-Target/Druggable Genome siRNA library. siRNAs are colored based on whether they suppressed, enhanced, or had no effect on +1 CGG_{100} NL-3xF. Data points corresponding to *HNRNPA2B1* and *PSMB5* are indicated.

Primary Hits



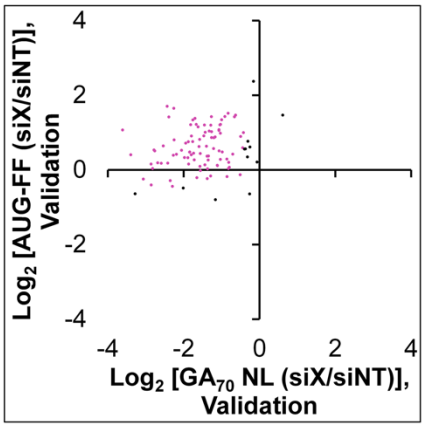
- FMR1-Validated, Suppressor
- FMR1-Validated, Enhancer
- Did not FMR1-Validate

Primary Hits, FMR1-Validated



- FMR1-Validated, RAN-Specific, Suppressor
- FMR1-Validated, RAN-Specific, Enhancer
- Not RAN-Specific

Primary Hits, FMR1-Validated, RAN-Specific



- FMR1-Validated, RAN-Specific, C9-Validated, Suppressor
- FMR1-Validated, RAN-Specific, C9-Validated, Enhancer
- Did not C9-Validate

Figure B.6: Serial Validation and Refinement of RAN-Specific *FMR1* Candidate Modifiers. Scatterplot of Log_2 CGG_{100} NL vs. Log_2 AUG-FF values gathered during the validation screen (top panel), Log_2 AUG NL vs. Log_2 AUG-FF values gathered during the AUG counter screen (middle panel), and Log_2 GA_{70} NL vs. Log_2 AUG-FF values gathered during the *C9ORF72* validation screen (bottom panel). Data points are colored based on whether the corresponding siRNAs progressed to the subsequent screen.

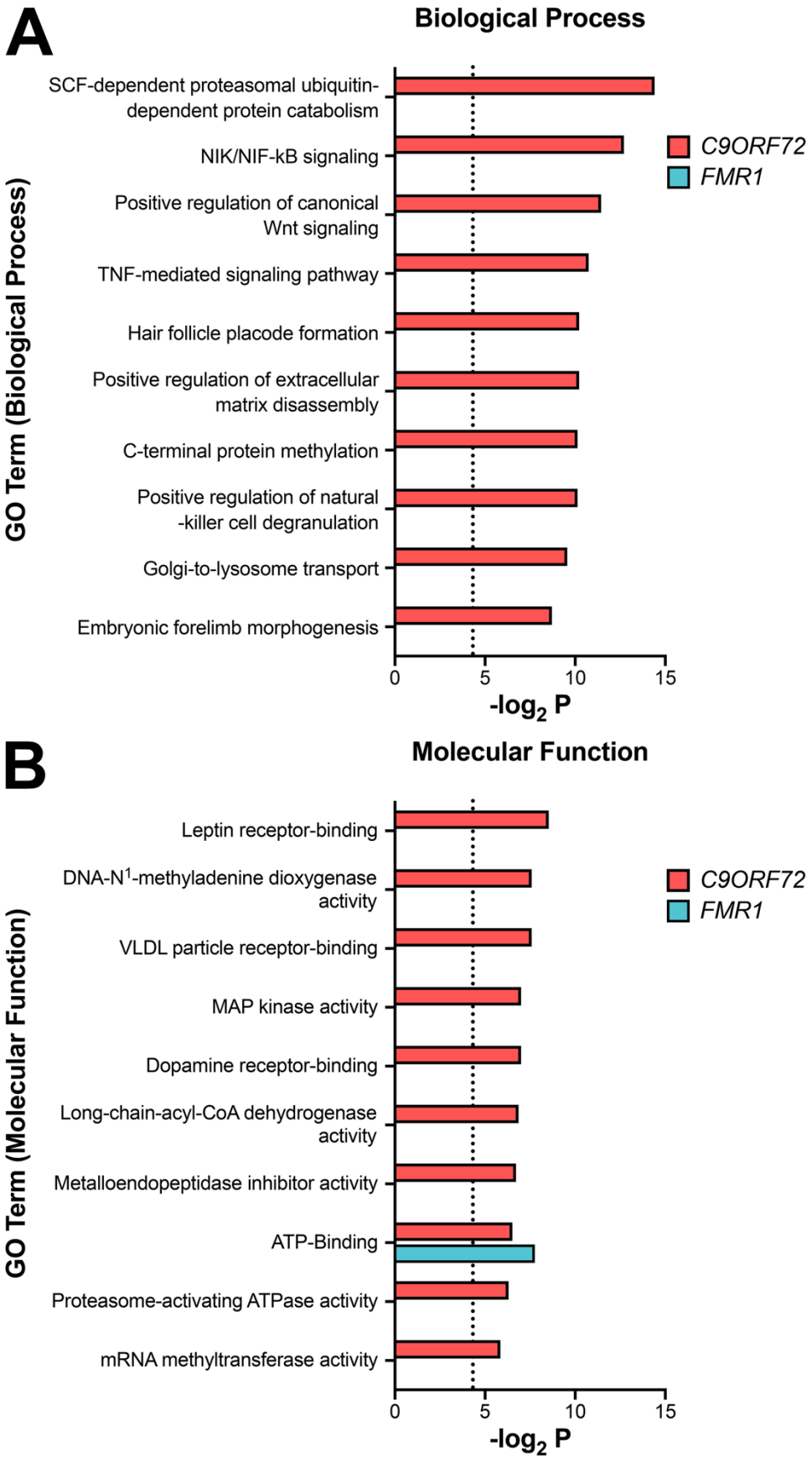

A
Biological Process
SCF-dependent proteasomal ubiquitin-dependent protein catabolism
NIK/NIF-kB signaling
Positive regulation of canonical Wnt signaling
TNF-mediated signaling pathway
Hair follicle placode formation
Positive regulation of extracellular matrix disassembly
C-terminal protein methylation
Positive regulation of natural -killer cell degranulation
Golgi-to-lysosome transport
Embryonic forelimb morphogenesis
GO Term (Biological Process)
C9ORF72
FMR1
0
5
10
15
-log2 P
B
Molecular Function
Leptin receptor-binding
DNA-N1-methyladenine dioxygenase activity
VLDL particle receptor-binding
MAP kinase activity
Dopamine receptor-binding
Long-chain-acyl-CoA dehydrogenase activity
Metalloendopeptidase inhibitor activity
ATP-Binding
Proteasome-activating ATPase activity
mRNA methyltransferase activity
GO Term (Molecular Function)
C9ORF72
FMR1
0
5
10
15
-log2 P

Figure B.7: Gene Ontology (GO) Analysis of Validated, RAN-Specific Modifiers of *C9ORF72*.
A, B Top 10 GO terms (biological process, A; or molecular function, B) by descending statistical significance among validated, RAN-specific *C9ORF72* suppressors that passed the off-target filter (red). Representation and statistical significance of these terms among validated, RAN-specific, filtered *FMR1* suppressors, if significant, is illustrated by teal bars. P=0.05 is demarcated by the dashed vertical line. Some biological process terms have been excluded due to complete overlap of their corresponding genes to those of the term “SCF-dependent proteasomal ubiquitin-dependent protein catabolism;” these include “hematopoietic stem cell differentiation,” “ubiquitin-protein ligase activity involved in mitotic cell cycle,” “regulation of G2/M transition of mitotic cell cycle,” and “antigen processing/presentation via MHC class I.”

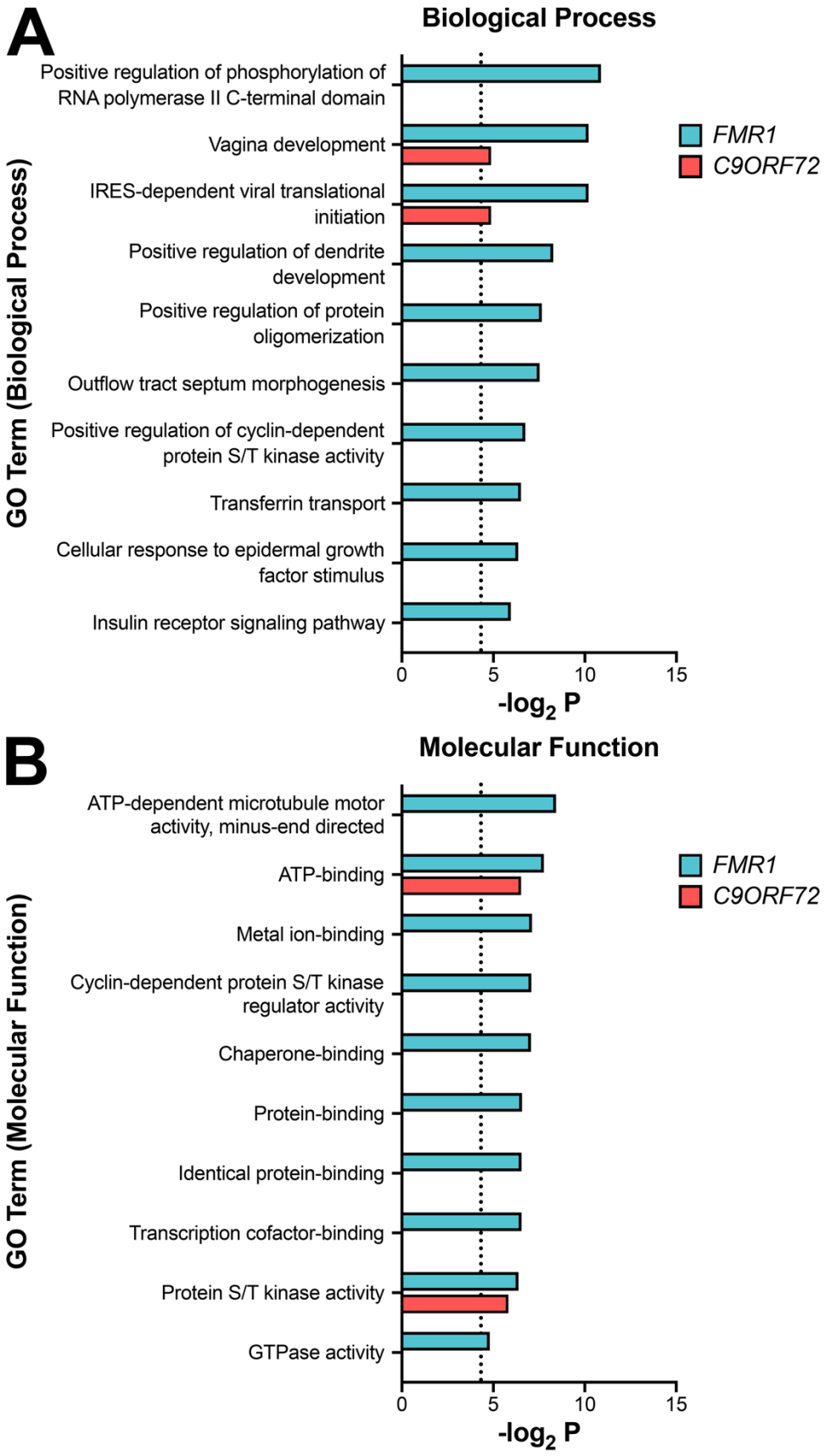

A
Biological Process
Positive regulation of phosphorylation of RNA polymerase II C-terminal domain
Vagina development
IRES-dependent viral translational initiation
Positive regulation of dendrite development
Positive regulation of protein oligomerization
Outflow tract septum morphogenesis
Positive regulation of cyclin-dependent protein S/T kinase activity
Transferrin transport
Cellular response to epidermal growth factor stimulus
Insulin receptor signaling pathway
GO Term (Biological Process)
FMR1
C9ORF72
0
5
10
15
-log2 P
B
Molecular Function
ATP-dependent microtubule motor activity, minus-end directed
ATP-binding
Metal ion-binding
Cyclin-dependent protein S/T kinase regulator activity
Chaperone-binding
Protein-binding
Identical protein-binding
Transcription cofactor-binding
Protein S/T kinase activity
GTPase activity
GO Term (Molecular Function)
FMR1
C9ORF72
0
5
10
15
-log2 P

Figure B.8: GO Analysis of Validated, RAN-Specific Modifiers of *FMR1*.
A, B Top 10 GO terms (biological process, A; or molecular function, B) by descending statistical significance among validated, RAN-specific *FMR1* suppressors that passed the off-target filter (teal). Representation and statistical significance of these terms among validated, RAN-specific, filtered *C9ORF72* suppressors, if significant, is illustrated by red bars. P=0.05 is demarcated by the dashed vertical line.

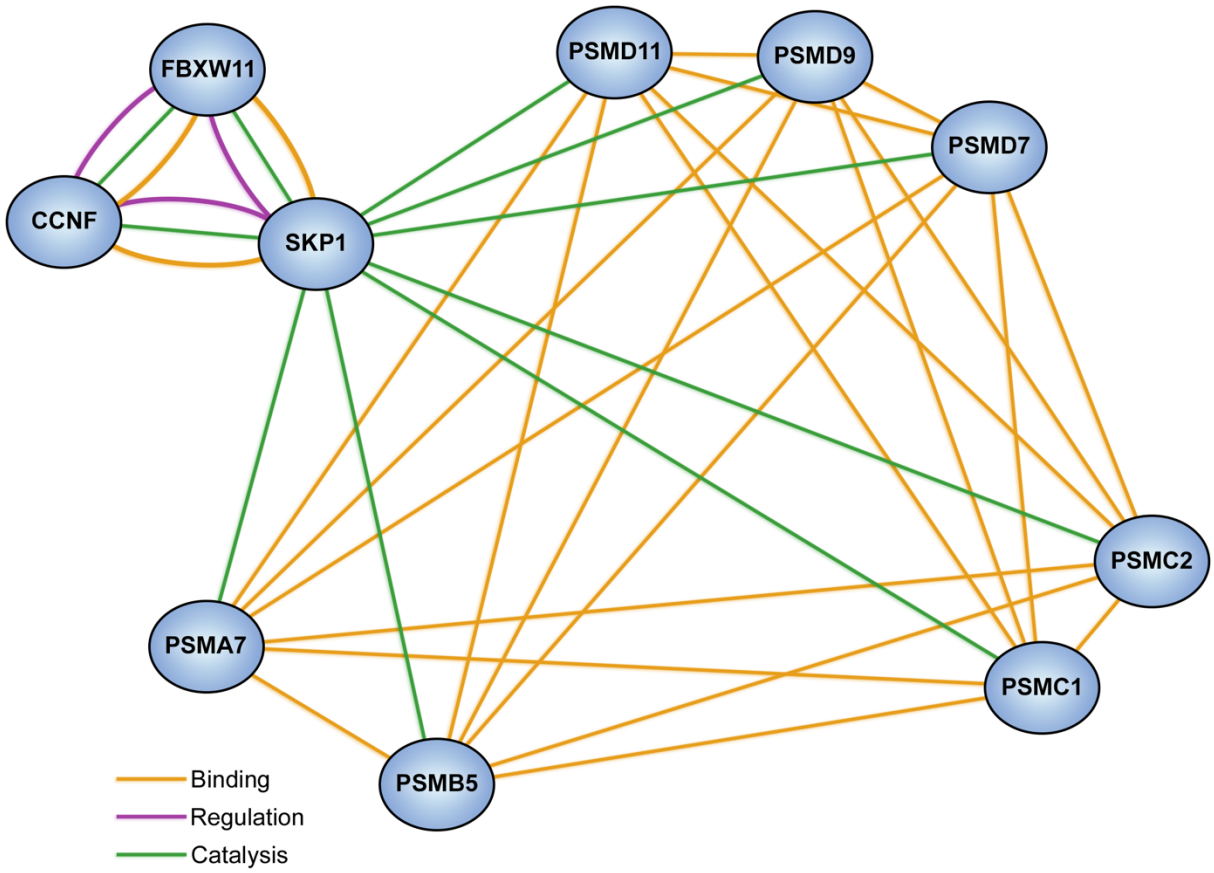


Figure B.9: The Relationship Between Multiple Modifiers of *C9ORF72* RAN Translation, All Related to the Proteasome.

The candidates depicted were identified and validated against the GA_{70} NL-3xF and AUG-NL-3xF reporters, were not eliminated by the off-target filter, and are associated with the GO term “SCF-dependent proteasomal ubiquitin-dependent protein catabolic process.”

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data	
GENE ID	**GENE SYMBOL**	NL	FF	NL	FF	NL	FF	NL	FF
10987	COPS5	1.168	0.970						
8562	DENR	0.596	1.097						
54505	DHX29	0.320	0.414						
10209	EIF1	0.204	0.294						
1964	EIF1AX	0.880	0.660						
9086	EIF1AY	1.010	0.445						
10289	EIF1B	1.020	1.643						
83939	EIF2A	0.762	0.422						
1967	EIF2B1	0.195	0.314						
8892	EIF2B2	0.688	0.501						
8891	EIF2B3	0.360	0.523						
8890	EIF2B4	0.784	0.332						
8893	EIF2B5	0.309	0.242						
1939	EIF2D	0.550	2.244						
1965	EIF2S1	0.242	0.251						
8894	EIF2S2	0.084	0.071						
1968	EIF2S3	0.253	0.221						
8661	EIF3A	0.125	0.114						
8662	EIF3B	0.324	0.139						
8663	EIF3C	0.186	0.280						
8664	EIF3D	1.883	1.979						
3646	EIF3E	1.119	0.501						
8665	EIF3F	0.433	0.907	0.477	2.081	0.906	2.006	0.492	1.319
8666	EIF3G	0.830	0.154						
8667	EIF3H	0.535	0.392						
8668	EIF3I	0.220	0.119						
8669	EIF3J	0.482	0.265						
27335	EIF3K	1.025	1.019						
51386	EIF3L	0.770	0.450						
10480	EIF3M	0.396	0.403						
1973	EIF4A1	0.381	0.598						
1974	EIF4A2	0.968	1.002						
9775	EIF4A3	0.543	0.102						
1975	EIF4B	0.441	0.434						
1977	EIF4E	0.223	0.237						
9470	EIF4E2	0.197	0.205						
317649	EIF4E3	0.436	0.618						
8637	EIF4EBP3	0.249	0.132						

Table B.1: Modulation of GA_{70} NL-3xF by Knockdown of eIFs, Part I.
siRNAs from the whole-genome library that target genes annotated with the GO terms “Translation initiation factor activity,” “Translation pre-initiation complex,” and “eIF4F complex” are listed above. Pink shading indicates that siRNA validated as a suppressor in the indicated screen; green shading indicates that siRNA validated as an enhancer in the indicated screen.

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data	
GENE ID	**GENE SYMBOL**	NL	FF	NL	FF	NL	FF	NL	FF
1981	EIF4G1	0.092	0.062						
1982	EIF4G2	0.474	0.094						
8672	EIF4G3	0.175	0.341						
7458	EIF4H	0.842	0.360						
1983	EIF5	0.683	0.587						
9669	EIF5B	0.480	1.918	0.426	4.054	0.933	2.777	0.618	2.406
3692	EIF6	1.530	2.525						
27086	FOXP1	0.595	0.738						
28985	MCTS1	0.414	0.514						
84285	MGC11102	1.789	0.912						
4528	MTIF2	1.018	1.713						
219402	MTIF3	1.669	1.154						
51633	OTUD6B	1.131	0.494						

Table B.2: Modulation of GA_{70} NL-3xF by Knockdown of eIFs, Part II.

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data	
GENE ID	**GENE SYMBOL**	NL	FF	NL	FF	NL	FF	NL	FF
11047	ADRM1	0.281	0.769	0.364	1.058	0.409	0.803	0.450	1.178
23392	ECPAS	0.524	0.364						
80227	PAAF1	0.268	0.523						
51371	POMP	0.292	2.374	0.116	1.870	4.097	3.396	1.736	3.109
5682	PSMA1	0.286	0.473						
5683	PSMA2	0.052	0.093						
5684	PSMA3	0.420	0.178						
5685	PSMA4	0.164	0.436						
5686	PSMA5	0.239	0.260						
5687	PSMA6	0.134	0.318						
5688	PSMA7	0.263	1.840	0.509	2.410	3.751	1.853	6.281	1.864
143471	PSMA8	0.915	0.350						
5689	PSMB1	0.348	1.048	0.701	1.006	5.095	0.966	1.805	1.046
5690	PSMB2	0.329	0.291						
5691	PSMB3	0.059	0.252						
5692	PSMB4	0.600	1.132						
5693	PSMB5	0.198	1.073	0.171	0.852	2.902	0.752	1.603	0.977
5694	PSMB6	0.164	0.239						
5695	PSMB7	0.623	1.913						
5696	PSMB8	0.148	0.354						
5698	PSMB9	1.026	1.406						
5699	PSMB10	0.428	0.713						
5700	PSMC1	0.470	1.641	0.455	0.924	6.149	0.982	4.412	0.959
5701	PSMC2	0.361	0.914	0.197	0.957	5.839	0.452	1.684	0.702
5702	PSMC3	0.067	0.208						
5704	PSMC4	0.084	0.462						
5705	PSMC5	0.563	1.375						
5706	PSMC6	0.027	0.055						

Table B.3: Modulation of GA_{70} NL-3xF by Components of the Proteasome, Part I.
siRNAs from the whole-genome library that target genes annotated with the GO terms “Proteasome core complex,” “Proteasome regulatory particle,” “Proteasome-activating ATPase activity,” “Proteasome accessory complex,” and “Proteasome assembly” are listed above. Pink shading indicates that siRNA validated as a suppressor in the indicated screen; green shading indicates that siRNA validated as an enhancer in the indicated screen.

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data	
GENE ID	**GENE SYMBOL**	NL	FF	NL	FF	NL	FF	NL	FF
5707	PSMD1	0.328	0.344						
5708	PSMD2	0.080	0.251						
5709	PSMD3	0.258	0.555						
5710	PSMD4	0.202	0.884	0.097	0.695	0.549	0.777	0.199	0.629
5711	PSMD5	0.721	1.299						
9861	PSMD6	0.202	0.225						
5713	PSMD7	0.362	1.139	0.409	1.265	5.615	1.032	1.923	1.618
5714	PSMD8	0.386	1.351	0.485	0.566	2.751	0.467	3.035	0.827
5715	PSMD9	0.372	1.467	0.367	1.570	1.939	1.092	1.112	1.080
5716	PSMD10	1.058	0.446						
5717	PSMD11	0.394	1.014	0.627	1.322	1.247	0.889	1.740	1.528
5718	PSMD12	0.219	0.328						
5719	PSMD13	0.108	0.197						
10213	PSMD14	0.262	0.875	0.167	0.303	3.000	0.235	1.204	0.352
5720	PSME1	0.854	0.344						
5721	PSME2	0.121	0.229						
10197	PSME3	0.314	0.341						
23198	PSME4	0.386	2.142	0.280	1.697	0.593	1.203	0.511	1.402
9491	PSMF1	0.486	0.249						
8624	PSMG1	1.487	0.762	0.350	0.649	0.602	0.447	0.523	0.952
56984	PSMG2	0.533	2.186						
84262	PSMG3	1.205	0.387						
5886	RAD23A	0.178	0.286						
5887	RAD23B	1.324	0.561						
7979	SHFM1	0.351	0.869						

Table B.4: Modulation of GA_{70} NL-3xF by Components of the Proteasome, Part II.

		FMR1 Primary Data		*FMR1* Validation Data		AUG Counter Data		*C9ORF72* Validation Data	
GENE ID	**GENE SYMBOL**	NL	FF	NL	FF	NL	FF	NL	FF
10987	COPS5	1.576	0.935						
54505	DHX29	0.287	0.390						
10209	EIF1	0.416	0.317						
1964	EIF1AX	0.816	0.455						
9086	EIF1AY	1.030	0.504						
10289	EIF1B	0.651	1.952						
83939	EIF2A	0.639	0.477						
1967	EIF2B1	0.318	0.504						
8892	EIF2B2	0.701	0.488						
8891	EIF2B3	0.247	0.462						
8890	EIF2B4	0.850	0.338						
8893	EIF2B5	0.432	0.373						
1965	EIF2S1	0.409	0.247						
8894	EIF2S2	0.119	0.062						
1968	EIF2S3	0.305	0.192						
8661	EIF3A	0.069	0.089						
8662	EIF3B	0.121	0.091						
8663	EIF3C	0.105	0.186						
8664	EIF3D	0.578	0.698						
3646	EIF3E	0.430	0.353						
8665	EIF3F	0.390	0.750	0.492	1.319	0.906	2.006	0.477	2.081
8666	EIF3G	0.249	0.183						
8667	EIF3H	0.295	0.239						
8668	EIF3I	0.138	0.114						
8669	EIF3J	0.880	0.238						
27335	EIF3K	0.557	0.999						
1973	EIF4A1	0.403	0.464						
1974	EIF4A2	0.977	0.899						
9775	EIF4A3	0.194	0.095						
1975	EIF4B	0.291	0.280						
1977	EIF4E	0.272	0.268						
9470	EIF4E2	0.234	0.330						
317649	EIF4E3	0.563	0.684						
1981	EIF4G1	0.095	0.092						
1982	EIF4G2	0.316	0.189						
8672	EIF4G3	0.314	0.433						
7458	EIF4H	1.166	0.336						

Table B.5: Modulation of +1 CGG_{100} NL-3xF by Knockdown of eIFs, Part I.
siRNAs from the Drug-Target and Druggable Genome libraries that target genes annotated with the GO terms “Translation initiation factor activity,” “Translation pre-initiation complex,” and “eIF4F complex” are listed above. Pink shading indicates that siRNA validated as a suppressor in the indicated screen; green shading indicates that siRNA validated as an enhancer in the indicated screen.

		FMR1 Primary Data		*FMR1* Validation Data		AUG Counter Data		*C9ORF72* Validation Data	
GENE ID	**GENE SYMBOL**	NL	FF	NL	FF	NL	FF	NL	FF
1983	EIF5	1.013	0.667						
9669	EIF5B	0.653	1.961						
3692	EIF6	1.726	1.971						
27086	FOXP1	0.767	0.870						
4528	MTIF2	1.036	1.428						
219402	MTIF3	2.026	1.151	2.999	0.934	4.313	0.991	1.810	0.959

Table B.6: Modulation of +1 CGG_{100} NL-3xF by Knockdown of eIFs, Part II.

		FMR1 Primary Data		*FMR1* Validation Data		AUG Counter Data		*C9ORF72* Validation Data	
GENE ID	**GENE SYMBOL**	NL	FF	NL	FF	NL	FF	NL	FF
80227	PAAF1	0.446	0.309						
51371	POMP	1.604	2.564						
5682	PSMA1	2.012	0.514	3.038	0.511	6.381	0.418	0.565	0.503
5683	PSMA2	0.163	0.150						
5684	PSMA3	1.257	0.167						
5685	PSMA4	0.882	0.549						
5686	PSMA5	0.644	0.278						
5687	PSMA6	0.303	0.324						
5688	PSMA7	2.294	1.886						
143471	PSMA8	0.832	0.258						
5689	PSMB1	0.852	1.282						
5690	PSMB2	0.954	0.319						
5691	PSMB3	0.231	0.233						
5692	PSMB4	0.926	1.245						
5693	PSMB5	1.683	1.054						
5694	PSMB6	0.926	0.241						
5695	PSMB7	4.059	1.640						
5696	PSMB8	0.277	0.289						
5698	PSMB9	1.369	0.953						
5699	PSMB10	0.850	0.568						
5700	PSMC1	0.872	1.645						
5701	PSMC2	0.517	0.985						
5702	PSMC3	0.051	0.124						
5704	PSMC4	0.484	0.506						
5705	PSMC5	5.080	1.149	4.481	1.312	4.404	0.381	0.396	1.244
5706	PSMC6	0.070	0.127						
5713	PSMD7	0.661	1.462						
10213	PSMD14	0.459	0.967	1.204	0.352	3.000	0.235	0.167	0.303
5721	PSME2	0.315	0.372						
5886	RAD23A	0.448	0.417						
5887	RAD23B	0.896	0.488						
7979	SHFM1	0.462	1.389	0.523	0.952	0.602	0.447	0.350	0.649

Table B.7: Modulation of +1 CGG_{100} NL-3xF by Components of the Proteasome.
siRNAs from the Drug-Target and Druggable Genome libraries that target genes annotated with the GO terms “Proteasome core complex,” “Proteasome regulatory particle,” “Proteasome-activating ATPase activity,” “Proteasome accessory complex,” and “Proteasome assembly” are listed above. Pink shading indicates that siRNA validated as a suppressor in the indicated screen; green shading indicates that siRNA validated as an enhancer in the indicated screen.

RNA-Seq Library	Number Below Threshold In That Set Alone (%)	Number of Transcripts Below Threshold in Common (% of Library in Left Column)			
		GSE103719	GSE113550	GSE79133	GSE52447
GSE103719	1,237 (15.7%)	*	6,118 (77.8%)	768 (9.8%)	1,580 (20.1%)
GSE113550	3,303 (26.6%)	6,118 (49.3%)	*	2,629 (21.2%)	2635 (21.3%)
GSE79133	3,821 (56.9%)	768 (11.4%)	2,629 (39.2%)	*	761 (11.3%)
GSE52447	1,882 (25.9%)	1,580 (21.7%)	2,635 (36.3%)	761 (10.5%)	*

Table B.8: Comparison of the Number of Genes Beneath Expression Threshold Across RNA-Seq Libraries.

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data		
GENE ID	GENE SYMBOL	NL	FF	NL	FF	NL	FF	NL	FF	
7539	ZFP37	0.154	1.136	0.095	1.325	0.832	0.874	0.255	1.795	**
56953	NT5M	0.172	1.012	0.082	2.095	1.261	1.164	0.350	1.626	
22862	FNDC3	0.087	1.173	0.175	1.128	1.121	1.140	0.372	0.914	
3782	KCNN3	0.150	0.828	0.138	0.756	1.207	0.809	0.426	0.981	
342933	LOC342933	0.109	2.239	0.180	2.393	0.989	2.655	0.418	2.319	**
403278	OR5K4	0.136	0.811	0.165	1.579	0.839	0.965	0.732	1.189	**
23291	FBXW11	0.159	1.132	0.145	1.449	1.164	1.253	0.420	1.238	
30844	EHD4	0.159	0.887	0.164	1.267	0.806	0.838	0.160	1.285	
340069	LOC340069	0.182	1.285	0.158	0.825	0.911	0.570	0.307	0.704	**
79092	CARD14	0.162	0.829	0.180	1.190	1.207	0.925	0.546	1.025	
8050	PDHX	0.146	1.323	0.207	1.311	1.846	1.414	0.581	1.301	
905	CCNT2	0.234	1.175	0.119	0.842	0.811	0.389	0.245	1.119	
1545	CYP1B1	0.239	1.032	0.121	1.205	1.058	0.866	0.437	0.904	**
5803	PTPRZ1	0.180	2.632	0.185	3.254	0.850	1.245	0.428	2.340	**
5693	PSMB5	0.198	1.073	0.171	0.852	2.902	0.752	1.603	0.977	
33	ACADL	0.189	1.002	0.186	0.964	1.136	0.744	0.350	0.814	
134860	TAAR9	0.157	3.998	0.228	3.262	3.025	1.927	1.605	1.984	**
84071	ARMC2	0.192	1.352	0.200	1.351	0.897	0.805	0.502	1.402	
10129	13CDNA73	0.136	3.298	0.264	2.124	0.961	2.356	0.740	2.490	
1359	CPA3	0.255	1.265	0.147	1.422	0.835	1.884	0.340	1.757	**
4717	NDUFC1	0.263	1.360	0.144	0.909	1.298	0.957	0.523	1.306	
51371	C13ORF12	0.292	2.374	0.116	1.870	4.097	3.396	1.736	3.109	
51012	C20ORF45	0.164	1.335	0.244	1.083	0.820	0.638	0.369	1.059	**
259289	TAS2R43	0.242	1.296	0.168	1.313	1.117	0.600	0.728	0.707	**
26267	FBXO10	0.221	1.726	0.191	2.673	0.981	1.924	0.501	1.662	
57038	RARSL	0.259	1.574	0.158	2.917	1.094	2.324	0.533	2.424	
54904	WHSC1L1	0.252	0.852	0.166	1.129	0.841	0.998	0.398	1.198	
4534	MTM1	0.193	0.965	0.225	0.915	1.246	0.517	0.540	0.758	
4281	MID1	0.237	1.443	0.193	1.470	1.075	1.013	0.564	2.165	
80714	PBX4	0.245	1.619	0.186	1.870	0.833	0.717	0.480	2.074	
51318	MRPL35	0.234	1.278	0.209	1.866	1.233	1.283	0.618	1.460	**
23510	KCTD2	0.153	1.331	0.296	3.814	2.438	1.923	0.921	3.313	
1382	CRABP2	0.265	1.376	0.185	1.279	0.967	1.112	0.331	1.524	
643923	LOC643923	0.237	1.129	0.214	0.955	0.953	0.655	0.438	0.488	
6462	SHBG	0.205	1.914	0.247	2.191	2.123	1.308	0.800	1.275	
286046	C8ORF21	0.164	1.086	0.290	1.222	1.232	0.957	0.488	0.661	
6605	SMARCE1	0.230	1.424	0.227	2.188	2.166	0.789	0.838	1.858	
4000	LMNA	0.157	0.969	0.305	1.953	1.958	0.869	0.866	1.634	
8618	CADPS	0.194	1.014	0.270	1.253	1.195	1.028	0.482	1.125	
84926	FLJ14800	0.197	2.005	0.272	1.613	0.835	1.325	0.566	1.620	

Table B.9: Most Efficacious *C9ORF72*-Validated, RAN-Specific Suppressors.
siRNAs are ranked by their mean $NL_{siX/siNT}$ value between the primary and validation *C9ORF72* screens in ascending order. siRNAs shaded pink in the final column also validated as suppressors of *FMR1*.

Data Information: ** Gene failed to meet expression minimum in off-target filter.

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data		
GENE ID	GENE SYMBOL	NL	FF	NL	FF	NL	FF	NL	FF	
221091	LOC221091	7.649	0.926	4.189	0.394	0.976	0.287	1.874	0.419	**
25802	LMOD1	3.702	0.751	2.711	0.549	0.867	0.470	1.238	0.517	**
10189	THOC4	4.591	0.791	1.796	0.363	1.067	0.204	2.419	0.413	
60509	FLJ21839	3.815	0.626	2.528	0.675	0.829	0.617	1.231	1.002	
57496	MRTF-B	3.783	0.268	2.262	0.346	1.216	0.204	0.901	0.288	**
221150	C13ORF3	2.700	0.528	2.793	0.633	0.977	0.494	1.160	0.534	
10422	UBADC1	2.880	0.610	2.128	0.700	0.981	0.666	1.965	0.600	
55794	DDX28	3.042	0.357	1.849	0.340	1.171	0.362	1.149	0.435	
26508	HEYL	2.137	0.748	2.558	0.444	0.988	0.351	1.203	0.401	**
10554	AGPAT1	2.352	0.465	1.798	0.276	1.214	0.698	0.896	0.290	**
7023	TFAP4	2.329	0.631	1.716	0.349	1.108	0.323	0.820	0.305	
22947	DUX4	2.134	0.572	1.911	0.639	1.089	0.446	1.425	0.664	**
26511	CHIC2	2.087	0.674	1.752	0.644	0.871	0.538	0.961	0.563	
80071	FLJ13215	2.126	0.431	1.706	0.631	1.037	0.508	1.483	0.781	

Table B.10: Most Efficacious *C9ORF72*-Validated, RAN-Specific Enhancers.
siRNAs are ranked by their mean $NL_{siX/siNT}$ value between the primary and validation *C9ORF72* screens in descending order. siRNAs shaded green in in the final column also validated as enhancers of *FMR1*.

Data Information: ** Gene failed to meet expression minimum in off-target filter.

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data		
GENE ID	GENE SYMBOL	NL	FF	NL	FF	NL	FF	NL	FF	
30844	EHD4	0.159	0.887	0.164	1.267	0.806	0.838	0.160	1.285	
7539	ZFP37	0.154	1.136	0.095	1.325	0.832	0.874	0.255	1.795	**
905	CCNT2	0.234	1.175	0.119	0.842	0.811	0.389	0.245	1.119	
56953	NT5M	0.172	1.012	0.082	2.095	1.261	1.164	0.350	1.626	
22862	FNDC3	0.087	1.173	0.175	1.128	1.121	1.140	0.372	0.914	
7363	UGT2B4	0.325	2.133	0.172	1.562	0.908	1.584	0.198	2.229	**
342933	LOC342933	0.109	2.239	0.180	2.393	0.989	2.655	0.418	2.319	**
3782	KCNN3	0.150	0.828	0.138	0.756	1.207	0.809	0.426	0.981	
7572	ZNF24	0.307	1.303	0.187	1.132	2.331	0.544	0.227	0.815	
23291	FBXW11	0.159	1.132	0.145	1.449	1.164	1.253	0.420	1.238	
33	ACADL	0.189	1.002	0.186	0.964	1.136	0.744	0.350	0.814	
1359	CPA3	0.255	1.265	0.147	1.422	0.835	1.884	0.340	1.757	**
51012	C20ORF45	0.164	1.335	0.244	1.083	0.820	0.638	0.369	1.059	**
1382	CRABP2	0.265	1.376	0.185	1.279	0.967	1.112	0.331	1.524	
5803	PTPRZ1	0.180	2.632	0.185	3.254	0.850	1.245	0.428	2.340	**
4036	LRP2	0.317	1.120	0.195	0.819	0.825	0.394	0.284	0.773	
1545	CYP1B1	0.239	1.032	0.121	1.205	1.058	0.866	0.437	0.904	**
112	ADCY6	0.285	2.829	0.209	3.135	1.069	1.909	0.315	1.841	
389058	SP5	0.278	1.092	0.215	0.929	0.803	0.779	0.319	0.736	
54904	WHSC1L1	0.252	0.852	0.166	1.129	0.841	0.998	0.398	1.198	
2926	GRSF1	0.319	0.924	0.188	0.886	0.853	0.537	0.378	0.857	
79092	CARD14	0.162	0.829	0.180	1.190	1.207	0.925	0.546	1.025	
84071	ARMC2	0.192	1.352	0.200	1.351	0.897	0.805	0.502	1.402	
84620	ST6GAL2	0.402	0.931	0.209	1.793	0.992	0.756	0.285	1.374	**
23612	PHLDA3	0.329	1.277	0.227	1.005	1.099	0.993	0.345	0.977	**
390197	OR4D10	0.390	5.905	0.138	3.563	0.812	3.943	0.379	4.013	**
78990	OTUB2	0.246	1.214	0.295	0.882	1.043	0.554	0.366	0.792	
56987	BBX	0.281	2.177	0.252	0.833	1.049	0.733	0.376	0.753	
80714	PBX4	0.245	1.619	0.186	1.870	0.833	0.717	0.480	2.074	
4762	NEUROG1	0.337	2.166	0.238	0.877	0.899	0.537	0.336	0.895	**
26267	FBXO10	0.221	1.726	0.191	2.673	0.981	1.924	0.501	1.662	
161176	C14ORF49	0.358	2.236	0.240	1.490	1.483	1.080	0.319	1.062	
391112	OR6Y1	0.340	1.161	0.153	1.144	0.830	0.828	0.428	0.868	**
113540	CMTM1	0.231	0.914	0.308	2.051	1.046	1.738	0.386	2.819	
11095	ADAMTS8	0.231	1.405	0.284	2.311	1.332	1.102	0.415	1.661	
4717	NDUFC1	0.263	1.360	0.144	0.909	1.298	0.957	0.523	1.306	
8570	KHSRP	0.437	1.095	0.203	0.961	0.990	0.457	0.292	0.925	
8050	PDHX	0.146	1.323	0.207	1.311	1.846	1.414	0.581	1.301	
4670	HNRPM	0.242	4.046	0.285	3.204	1.806	3.140	0.416	2.977	
84916	CIRH1A	0.244	1.303	0.284	1.382	1.470	1.039	0.415	1.199	

Table B.11: Most Efficacious *C9ORF72*-Validated, RAN-Specific, *FMR1*-Validated Suppressors.
siRNAs are ranked by their mean $NL_{siX/siNT}$ value between the *C9ORF72* primary, *C9ORF72* validation, and *FMR1* validation screens in ascending order.

Data Information: ** Gene failed to meet expression minimum in off-target filter.

		C9ORF72 Primary Data		*C9ORF72* Validation Data		AUG Counter Data		*FMR1* Validation Data		
GENE ID	GENE SYMBOL	NL	FF	NL	FF	NL	FF	NL	FF	
221091	LOC221091	7.649	0.926	4.189	0.394	0.976	0.287	1.874	0.419	**
10189	THOC4	4.591	0.791	1.796	0.363	1.067	0.204	2.419	0.413	
10422	UBADC1	2.880	0.610	2.128	0.700	0.981	0.666	1.965	0.600	

Table B.12: Most Efficacious *C9ORF72*-Validated, RAN-Specific, *FMR1*-Validated Enhancers. siRNAs are ranked by their mean $NL_{siX/siNT}$ value between the *C9ORF72* primary, *C9ORF72* validation, and *FMR1* validation screens in descending order.

Data Information: ** Gene failed to meet expression minimum in off-target filter.

		FMR1 Primary Data		*FMR1* Validation Data		AUG Counter Data		*C9ORF72* Validation Data		
GENE ID	GENE SYMBOL	NL	FF	NL	FF	NL	FF	NL	FF	
84868	HAVCR2	0.164	0.951	0.233	0.945	0.859	0.582	0.409	0.892	
7572	ZNF24	0.260	1.273	0.227	0.815	2.331	0.544	0.187	1.132	
4437	MSH3	0.267	2.650	0.253	2.513	1.088	1.189	0.295	2.529	
10859	LILRB1	0.238	3.080	0.288	3.168	1.252	2.210	0.485	2.304	**
1382	CRABP2	0.241	1.242	0.287	1.395	0.967	1.112	0.156	1.142	
7363	UGT2B4	0.381	2.159	0.198	2.229	0.908	1.584	0.172	1.562	**
9630	GNA14	0.231	0.835	0.355	1.273	0.951	0.963	0.219	1.335	
905	CCNT2	0.341	1.421	0.245	1.119	0.811	0.389	0.119	0.842	
5516	PPP2CB	0.236	0.870	0.366	0.975	0.844	0.384	0.386	1.018	
581	BAX	0.338	0.984	0.272	0.874	1.458	0.670	0.318	0.866	
80347	COASY	0.256	0.814	0.376	0.929	1.574	0.358	0.287	1.325	
54496	PRMT7	0.346	1.413	0.300	1.402	0.862	0.924	0.368	1.097	
84675	TRIM55	0.311	0.839	0.338	1.160	1.052	1.129	0.339	0.988	
4036	LRP2	0.365	1.094	0.284	0.773	0.825	0.394	0.195	0.819	
56953	NT5M	0.309	1.198	0.350	1.626	1.261	1.164	0.082	2.095	
862	CBFA2T1	0.384	0.882	0.279	1.064	0.822	1.051	0.331	1.542	
892	CCNC	0.475	2.532	0.192	0.831	1.258	1.416	0.103	0.642	
7539	ZFP37	0.436	1.742	0.255	1.795	0.832	0.874	0.095	1.325	**
1359	CPA3	0.367	1.296	0.340	1.757	0.835	1.884	0.147	1.422	**
9152	SLC6A5	0.380	0.824	0.328	1.872	1.265	0.974	0.613	1.340	**
152	ADRA2C	0.289	0.896	0.424	0.827	0.937	0.602	0.446	0.574	
79705	LRRK1	0.277	0.932	0.438	1.236	1.135	1.044	0.379	1.552	
8618	CADPS	0.235	1.013	0.482	1.125	1.195	1.028	0.270	1.253	
23012	STK38L	0.284	1.282	0.441	1.327	1.588	1.254	0.335	1.757	
2846	GPR23	0.345	1.039	0.394	1.350	1.406	0.864	0.240	1.441	**
7168	TPM1	0.287	0.761	0.460	1.541	1.223	0.488	0.253	1.255	
643	BLR1	0.351	1.209	0.409	0.842	0.870	0.847	0.833	0.640	**
54904	WHSC1L1	0.364	0.938	0.398	1.198	0.841	0.998	0.166	1.129	
55507	GPRC5D	0.323	1.397	0.443	1.474	0.952	1.086	0.300	1.362	
3801	KIFC3	0.319	1.295	0.461	0.938	1.105	0.445	0.566	1.026	
23291	FBXW11	0.369	1.080	0.420	1.238	1.164	1.253	0.145	1.449	
3782	KCNN3	0.365	1.025	0.426	0.981	1.207	0.809	0.138	0.756	
22894	KIAA1008	0.455	0.962	0.337	0.785	1.828	1.426	0.203	0.736	
9550	ATP6V1G1	0.449	0.924	0.345	0.909	0.835	1.144	0.141	1.025	
11095	ADAMTS8	0.385	1.520	0.415	1.661	1.332	1.102	0.284	2.311	
112	ADCY6	0.492	1.663	0.315	1.841	1.069	1.909	0.209	3.135	
51506	UFC1	0.365	1.031	0.464	0.960	0.885	1.166	0.274	1.116	
26353	HSPB8	0.358	1.457	0.480	2.390	1.233	1.477	0.500	2.534	
379	ARL4D	0.424	1.421	0.414	1.187	0.995	0.935	0.488	1.975	
3416	IDE	0.349	0.785	0.506	1.360	1.339	1.252	0.801	1.273	

Table B.13: Most Efficacious *FMR1*-Validated, RAN-Specific Suppressors.
siRNAs are ranked by their mean $NL_{siX/siNT}$ value between the primary and validation *FMR1* screens in ascending order. siRNAs shaded pink in the final column also validated as suppressors of *C9ORF72*.

Data Information: ** Gene failed to meet expression minimum in off-target filter.

Primer Name	Primer Sequence
F_Q5_nLuc-3xF P st1	AAGCCGCCGGTACATTGCCTATGTCCTGCGCACATAAAAG GGTCAAGACAATCC
R_Q5_nLuc-3xF P st1	GCTCTTCAACCTCAGGCGGAAAGCCGTGAGAATTCTTGTC ATCGTCATCCTTG
F_Q5_nLuc-3xF P st2	GCTTGTGCTTCAGCTCGCATCAACGTCTAAGGCCGCGACT CTAGAG
R_Q5_nLuc-3xF P st2	GGCTGGGTGCCGGTCCATACCGCTTTCTTGTGCGCAGGA CTAGGC

Table B.14: Primers Used in This Study.

Fly Stock	Source	Catalog #
CGG_{90}-EGFP	Jin Lab	-
GMR-GAL4	BDSC	8605
Tub5-GS	Pletcher Lab	-
shFirefly Luciferase	BDSC	35788
shHRB87F (A)	BDSC	31244
shHRB87F (B)	BDSC	31472
shHRB98DE	BDSC	32351
UAS-HNRNPA2B1	Botas Lab	-
UAS-HRB87F	Botas Lab	-
UAS-HRB98DE	Botas Lab	-

Table B.15: Fly Lines Used in This Study and Their Sources.

References

1 Nelson, D. L., Orr, H. T. & Warren, S. T. The unstable repeats--three evolving faces of neurological disease. *Neuron* **77**, 825-843 (2013).

2 Mason, A. R., Ziemann, A. & Finkbeiner, S. Targeting the low-hanging fruit of neurodegeneration. *Neurology* **83**, 1470-1473 (2014).

3 La Spada, A. R., Paulson, H. L. & Fischbeck, K. H. Trinucleotide repeat expansion in neurological disease. *Ann Neurol* **36**, 814-822 (1994).

4 Paulson, H. L. *et al.* Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron* **19**, 333-344 (1997).

5 Warrick, J. M. *et al.* Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in Drosophila. *Cell* **93**, 939-949 (1998).

6 Paulson, H. L., Shakkottai, V. G., Clark, H. B. & Orr, H. T. Polyglutamine spinocerebellar ataxias - from genes to potential treatments. *Nature Reviews. Neuroscience* **18**, 613-626 (2017).

7 He, F. & Todd, P. K. Epigenetics in nucleotide repeat expansion disorders. *Seminars in Neurology* **31**, 470-483 (2011).

8 Miller, J. W. *et al.* Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J* **19**, 4439-4448 (2000).

9 Mankodi, A. *et al.* Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum Mol Genet* **10**, 2165-2170 (2001).

10 Ranum, L. P. & Cooper, T. A. RNA-mediated neuromuscular disorders. *Annu Rev Neurosci* **29**, 259-277 (2006).

11 Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247 (2017).

12 Cleary, J. D. & Ranum, L. P. Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Current Opinion in Genetics & Development* **26**, 6-15 (2014).

13 Kearse, M. G. & Todd, P. K. Repeat-associated non-AUG translation and its impact in neurodegenerative disease. *Neurotherapeutics : the Journal of the American Society for Experimental NeuroTherapeutics* **11**, 721-731 (2014).

14 Zu, T. *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *PNAS* **108**, 260-265 (2011).

15 Ash, P. E. *et al.* Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646 (2013).

16 Mori, K. *et al.* Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* **126**, 881-893 (2013).

17 Mori, K. *et al.* The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335-1338 (2013).

18 Zu, T. *et al.* RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *PNAS* **110**, E4968-4977 (2013).

19 Todd, P. K. *et al.* CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440-455 (2013).

20 Banez-Coronel, M. *et al.* RAN Translation in Huntington Disease. *Neuron* **88**, 667-677 (2015).

21 Krans, A., Kearse, M. G. & Todd, P. K. Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Annals of Neurology* **80**, 871-881 (2016).

22 Ishiguro, T. *et al.* Regulatory Role of RNA Chaperone TDP-43 for RNA Misfolding and Repeat-Associated Translation in SCA31. *Neuron* **94**, 108-124 e107 (2017).

23 Zu, T. *et al.* RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. *Neuron* **95**, 1292-1305 e1295 (2017).

24 Soragni, E. *et al.* Repeat-Associated Non-ATG (RAN) Translation in Fuchs' Endothelial Corneal Dystrophy. *Investigative Ophthalmology & Visual Science* **59**, 1888-1896 (2018).

25 Zu, T., Pattamatta, A. & Ranum, L. P. W. Repeat-Associated Non-ATG Translation in Neurological Diseases. *Cold Spring Harbor Perspectives in Biology* (2018).

26 Gendron, T. F. *et al.* Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol* **126**, 829-844 (2013).

27 Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews. Molecular Cell Biology* **11**, 113-127 (2010).

28 Sonenberg, N., Morgan, M. A., Merrick, W. C. & Shatkin, A. J. A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *PNAS* **75**, 4843-4847 (1978).

29 Sonenberg, N., Rupprecht, K. M., Hecht, S. M. & Shatkin, A. J. Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *PNAS* **76**, 4345-4349 (1979).

30 Grifo, J. A., Tahara, S. M., Morgan, M. A., Shatkin, A. J. & Merrick, W. C. New initiation factor activity required for globin mRNA translation. *The Journal of Biological Chemistry* **258**, 5804-5810 (1983).

31 Imataka, H., Gradi, A. & Sonenberg, N. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J* **17**, 7480-7489 (1998).

32 Kessler, S. H. & Sachs, A. B. RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. *Molecular and Cellular Biology* **18**, 51-57 (1998).

33 Borman, A. M., Michel, Y. M. & Kean, K. M. Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. *Nucleic Acids Research* **28**, 4068-4075 (2000).

34 Kahvejian, A., Svitkin, Y. V., Sukarieh, R., M'Boutchou, M. N. & Sonenberg, N. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes & Development* **19**, 104-113 (2005).

35 LeFebvre, A. K. *et al.* Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *The Journal of Biological Chemistry* **281**, 22917-22932 (2006).

36 Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes & Development* **16**, 2906-2922 (2002).

37 Kozak, M. & Shatkin, A. J. Migration of 40 S ribosomal subunits on messenger RNA in the presence of edeine. *The Journal of Biological Chemistry* **253**, 6568-6577 (1978).

38 Kozak, M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109-1123 (1978).

39 Jackson, R. J. The ATP requirement for initiation of eukaryotic translation varies according to the mRNA species. *European Journal of Biochemistry* **200**, 285-294 (1991).

40 Svitkin, Y. V. *et al.* The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA* **7**, 382-394 (2001).

41 Dhote, V., Sweeney, T. R., Kim, N., Hellen, C. U. & Pestova, T. V. Roles of individual domains in the function of DHX29, an essential factor required for translation of structured mammalian mRNAs. *PNAS* **109**, E3150-3159 (2012).

42 Parsyan, A. *et al.* The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. *PNAS* **106**, 22217-22222 (2009).

43 Zhang, Y., You, J., Wang, X. & Weber, J. The DHX33 RNA Helicase Promotes mRNA Translation Initiation. *Molecular and Cellular Biology* **35**, 2918-2931 (2015).

44 Rogers, G. W., Jr., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *The Journal of Biological Chemistry* **276**, 30914-30922 (2001).

45 Rogers, G. W., Jr., Richter, N. J. & Merrick, W. C. Biochemical and kinetic characterization of the RNA helicase activity of eukaryotic initiation factor 4A. *The Journal of Biological Chemistry* **274**, 12236-12244 (1999).

46 Sen, N. D., Zhou, F., Harris, M. S., Ingolia, N. T. & Hinnebusch, A. G. eIF4B stimulates translation of long mRNAs with structured 5' UTRs and low closed-loop potential but weak dependence on eIF4G. *PNAS* **113**, 10464-10472 (2016).

47 Iost, I., Dreyfus, M. & Linder, P. Ded1p, a DEAD-box protein required for translation initiation in Saccharomyces cerevisiae, is an RNA helicase. *The Journal of Biological Chemistry* **274**, 17677-17683 (1999).

48 Liu, F., Putnam, A. & Jankowsky, E. ATP hydrolysis is required for DEAD-box protein recycling but not for duplex unwinding. *PNAS* **105**, 20209-20214 (2008).

49 Liu, F., Putnam, A. A. & Jankowsky, E. DEAD-box helicases form nucleotide-dependent, long-lived complexes with RNA. *Biochemistry* **53**, 423-433 (2014).

50 Marsden, S., Nardelli, M., Linder, P. & McCarthy, J. E. Unwinding single RNA molecules using helicases involved in eukaryotic translation initiation. *Journal of Molecular Biology* **361**, 327-335 (2006).

51 Yang, Q. & Jankowsky, E. ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1. *Biochemistry* **44**, 13591-13601 (2005).

52 Chen, Y. *et al.* DEAD-box proteins can completely separate an RNA duplex using a single ATP. *PNAS* **105**, 20203-20208 (2008).

53 Sharma, D., Putnam, A. A. & Jankowsky, E. Biochemical Differences and Similarities between the DEAD-Box Helicase Orthologs DDX3X and Ded1p. *J Mol Biol* **429**, 3730-3742 (2017).

54 Yang, Q., Del Campo, M., Lambowitz, A. M. & Jankowsky, E. DEAD-box proteins unwind duplexes by local strand separation. *Molecular Cell* **28**, 253-263 (2007).

55 Yang, Q. & Jankowsky, E. The DEAD-box protein Ded1 unwinds RNA duplexes by a mode distinct from translocating helicases. *Nat Struct Mol Biol* **13**, 981-986 (2006).

56 Shih, J. W., Tsai, T. Y., Chao, C. H. & Wu Lee, Y. H. Candidate tumor suppressor DDX3 RNA helicase specifically represses cap-dependent translation by acting as an eIF4E inhibitory protein. *Oncogene* **27**, 700-714 (2008).

57 Shih, J. W. *et al.* Critical roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response. *The Biochemical Journal* **441**, 119-129 (2012).

58 Hilliker, A., Gao, Z., Jankowsky, E. & Parker, R. The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. *Molecular Cell* **43**, 962-972 (2011).

59 Guenther, U. P. *et al.* The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature* **559**, 130-134 (2018).

60 Geissler, R., Golbik, R. P. & Behrens, S. E. The DEAD-box helicase DDX3 supports the assembly of functional 80S ribosomes. *Nucleic Acids Res* **40**, 4998-5011 (2012).

61 Lai, M. C., Chang, W. C., Shieh, S. Y. & Tarn, W. Y. DDX3 regulates cell growth through translational control of cyclin E1. *Molecular and Cellular Biology* **30**, 5444-5453 (2010).

62 Gao, Z. *et al.* Coupling between the DEAD-box RNA helicases Ded1p and eIF4A. *eLife* **5** (2016).

63 Gupta, N., Lorsch, J. R. & Hinnebusch, A. G. Yeast Ded1 promotes 48S translation pre-initiation complex assembly in an mRNA-specific and eIF4F-dependent manner. *eLife* **7** (2018).

64 Sen, N. D., Zhou, F., Ingolia, N. T. & Hinnebusch, A. G. Genome-wide analysis of translational efficiency reveals distinct but overlapping functions of yeast DEAD-box RNA helicases Ded1 and eIF4A. *Genome Research* **25**, 1196-1205 (2015).

65 Berthelot, K., Muldoon, M., Rajkowitsch, L., Hughes, J. & McCarthy, J. E. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol Microbiol* **51**, 987-1001 (2004).

66 Chuang, R. Y., Weaver, P. L., Liu, Z. & Chang, T. H. Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science* **275**, 1468-1471 (1997).

67 Ihry, R. J., Sapiro, A. L. & Bashirullah, A. Translational control by the DEAD Box RNA helicase belle regulates ecdysone-triggered transcriptional cascades. *PLoS Genetics* **8** (2012).

68 Soto-Rifo, R. *et al.* DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. *EMBO J* **31**, 3745-3756 (2012).

69 Beckham, C. *et al.* The DEAD-box RNA helicase Ded1p affects and accumulates in Saccharomyces cerevisiae P-bodies. *Molecular Biology of the Cell* **19**, 984-993 (2008).

70 Gotze, M. *et al.* Translational repression of the Drosophila nanos mRNA involves the RNA helicase Belle and RNA coating by Me31B and Trailer hitch. *RNA* **23**, 1552-1568 (2017).

71 Oh, S. *et al.* Medulloblastoma-associated DDX3 variant selectively alters the translational response to stress. *Oncotarget* **7**, 28169-28182 (2016).

72 Valentin-Vega, Y. A. *et al.* Cancer-associated DDX3X mutations drive stress granule assembly and impair global translation. *Scientific Reports* **6**, 25996 (2016).

73 Yarunin, A., Harris, R. E., Ashe, M. P. & Ashe, H. L. Patterning of the Drosophila oocyte by a sequential translation repression program involving the d4EHP and Belle translational repressors. *RNA Biol* **8**, 904-912 (2011).

74 Kozak, M. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature* **308**, 241-246 (1984).

75 Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *PNAS* **83**, 2850-2854 (1986).

76 Castilho-Valavicius, B., Yoon, H. & Donahue, T. F. Genetic characterization of the Saccharomyces cerevisiae translational initiation suppressors sui1, sui2 and SUI3 and their effects on HIS4 expression. *Genetics* **124**, 483-495 (1990).

77 Ivanov, I. P., Loughran, G., Sachs, M. S. & Atkins, J. F. Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *PNAS* **107**, 18056-18060 (2010).

78 Maag, D., Algire, M. A. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 5 and 1A within the ribosomal pre-initiation complex plays a role in start site selection. *J Mol Biol* **356**, 724-737 (2006).

79 Passmore, L. A. *et al.* The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Molecular Cell* **26**, 41-50 (2007).

80 Pestova, T. V., Borukhov, S. I. & Hellen, C. U. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394**, 854-859 (1998).

81 Pisarev, A. V. *et al.* Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes & Development* **20**, 624-636 (2006).

82 Unbehaun, A., Borukhov, S. I., Hellen, C. U. & Pestova, T. V. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes & Development* **18**, 3078-3093 (2004).

83 Kozel, C. *et al.* Overexpression of eIF5 or its protein mimic 5MP perturbs eIF2 function and induces ATF4 translation through delayed re-initiation. *Nucleic Acids Research* **44**, 8704-8713 (2016).

84 Loughran, G., Firth, A. E., Atkins, J. F. & Ivanov, I. P. Translational autoregulation of BZW1 and BZW2 expression by modulating the stringency of start codon selection. *PloS One* **13**, e0192648 (2018).

85 Loughran, G., Sachs, M. S., Atkins, J. F. & Ivanov, I. P. Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Research* **40**, 2898-2906 (2012).

86 Tang, L. *et al.* Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Research* **45**, 11941-11953 (2017).

87 Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harbor Perspectives in Biology* **4**, a013706 (2012).

88 Kozak, M. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell* **19**, 79-90 (1980).

89 Kozak, M. Leader length and secondary structure modulate mRNA function under conditions of stress. *Molecular and Cellular Biology* **8**, 2737-2744 (1988).

90 Kozak, M. Features in the 5' non-coding sequences of rabbit alpha and beta-globin mRNAs that affect translational efficiency. *Journal of Molecular Biology* **235**, 95-110 (1994).

91 Sagliocco, F. A. *et al.* The influence of 5'-secondary structures upon ribosome binding to mRNA during translation in yeast. *J Biol Chem* **268**, 26522-26530 (1993).

92 Kozak, M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *PNAS* **87**, 8301-8305 (1990).

93 Peabody, D. S. Translation initiation at an ACG triplet in mammalian cells. *The Journal of Biological Chemistry* **262**, 11847-11851 (1987).

94 Peabody, D. S. Translation initiation at non-AUG triplets in mammalian cells. *The Journal of Biological Chemistry* **264**, 5031-5035 (1989).

95 Kozak, M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Molecular and Cellular Biology* **9**, 5073-5080 (1989).

96 Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223 (2009).

97 Gao, X. *et al.* Quantitative profiling of initiating ribosomes in vivo. *Nature Methods* **12**, 147-153 (2015).

98 Ingolia, N. T. *et al.* Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Reports* **8**, 1365-1379 (2014).

99 Lee, S. *et al.* Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *PNAS* **109**, E2424-2432 (2012).

100 Brar, G. A. *et al.* High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552-557 (2012).

101 Anderson, D. M. *et al.* A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595-606 (2015).

102 Chanut-Delalande, H. *et al.* Pri peptides are mediators of ecdysone for the temporal control of development. *Nature Cell Biology* **16**, 1035-1044 (2014).

103 Magny, E. G. *et al.* Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116-1120 (2013).

104 Menschaert, G. *et al.* Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence

of alternative translation products and near-cognate translation initiation events. *Molecular & Cellular Proteomics: MCP* **12**, 1780-1790 (2013).

105 Pauli, A. *et al.* Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science* **343**, 1248636 (2014).

106 Slavoff, S. A. *et al.* Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nature Chemical Biology* **9**, 59-64 (2013).

107 Vanderperre, B. *et al.* Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PloS One* **8**, e70698 (2013).

108 Kiliszek, A., Kierzek, R., Krzyzosiak, W. J. & Rypniewski, W. Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Research* **39**, 7308-7315 (2011).

109 Sobczak, K. *et al.* Structural diversity of triplet repeat RNAs. *The Journal of Biological Chemistry* **285**, 12755-12764 (2010).

110 Zumwalt, M., Ludwig, A., Hagerman, P. J. & Dieckmann, T. Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene. *RNA Biology* **4**, 93-100 (2007).

111 Kozak, M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Molecular and Cellular Biology* **9**, 5134-5142 (1989).

112 Cheung, Y. N. *et al.* Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo. *Genes & Development* **21**, 1217-1230 (2007).

113 Martin-Marcos, P. *et al.* Enhanced eIF1 binding to the 40S ribosome impedes conformational rearrangements of the preinitiation complex and elevates initiation accuracy. *RNA* **20**, 150-167 (2014).

114 Fernandez, I. S., Bai, X. C., Murshudov, G., Scheres, S. H. & Ramakrishnan, V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell* **157**, 823-831 (2014).

115 Muhs, M. *et al.* Cryo-EM of ribosomal 80S complexes with termination factors reveals the translocated cricket paralysis virus IRES. *Molecular Cell* **57**, 422-432 (2015).

116 Dodd, D. W., Tomchick, D. R., Corey, D. R. & Gagnon, K. T. Pathogenic C9ORF72 Antisense Repeat RNA Forms a Double Helix with Tandem C:C Mismatches. *Biochemistry* **55**, 1283-1286 (2016).

117 Hagerman, R. J. *et al.* Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57**, 127-130 (2001).

118 Pembrey, M. E., Winter, R. M. & Davies, K. E. A premutation that generates a defect at crossing over explains the inheritance of fragile X mental retardation. *Am J Med Genet* **21**, 709-717 (1985).

119 Sherman, S. L. *et al.* Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet* **69**, 289-299 (1985).

120 Tassone, F. *et al.* Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA* **13**, 555-562 (2007).

121 Tassone, F., Hagerman, R. J., Chamberlain, W. D. & Hagerman, P. J. Transcription of the FMR1 gene in individuals with fragile X syndrome. *Am J Med Genet* **97**, 195-203 (2000).

122 Tassone, F. *et al.* Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *Am J Med Genet* **94**, 232-236 (2000).

123 Tassone, F. *et al.* Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am J Hum Genet* **66**, 6-15 (2000).

124 Pieretti, M. *et al.* Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817-822 (1991).

125 Verkerk, A. J. *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* **65**, 905-914 (1991).

126 Dombrowski, C. *et al.* Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum Mol Genet* **11**, 371-378 (2002).

127 Jacquemont, S. *et al.* Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA* **291**, 460-469 (2004).

128 Rousseau, F., Rouillard, P., Morel, M. L., Khandjian, E. W. & Morgan, K. Prevalence of carriers of premutation-size alleles of the FMRI gene--and implications for the population genetics of the fragile X syndrome. *Am J Hum Genet* **57**, 1006-1018 (1995).

129 Leehey, M. A. & Hagerman, P. J. Fragile X-associated tremor/ataxia syndrome. *Handb Clin Neurol* **103**, 373-386 (2012).

130 Allingham-Hawkins, D. J. *et al.* Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study--preliminary data. *Am J Med Genet* **83**, 322-325 (1999).

131 Cronister, A. *et al.* Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. *Am J Med Genet* **38**, 269-274 (1991).

132 Garcia-Arocena, D. & Hagerman, P. J. Advances in understanding the molecular basis of FXTAS. *Hum Mol Genet* **19**, R83-89 (2010).

133 Hashem, V. *et al.* Ectopic expression of CGG containing mRNA is neurotoxic in mammals. *Hum Mol Genet* **18**, 2443-2451 (2009).

134 Tassone, F. *et al.* Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome. *J Med Genet* **41**, e43 (2004).

135 Sellier, C. *et al.* Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J* **29**, 1248-1261 (2010).

136 Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256-271 (2006).

137 Jin, P. *et al.* Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. *Neuron* **55**, 556-564 (2007).

138 Muslimov, I. A., Patel, M. V., Rose, A. & Tiedge, H. Spatial code recognition in neuronal RNA targeting: role of RNA-hnRNP A2 interactions. *J Cell Biol* **194**, 441-457 (2011).

139 Sofola, O. A. *et al.* RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. *Neuron* **55**, 565-571 (2007).

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

141 He, F., Krans, A., Freibaum, B. D., Taylor, J. P. & Todd, P. K. TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. *Hum Mol Genet* **23**, 5036-5051 (2014).

142 Sellier, C. *et al.* Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep* **3**, 869-880 (2013).

143 Qurashi, A., Li, W., Zhou, J. Y., Peng, J. & Jin, P. Nuclear accumulation of stress response mRNAs contributes to the neurodegeneration caused by Fragile X premutation rCGG repeats. *PLoS Genet* **7**, e1002102 (2011).

144 Hagerman, R. J. & Hagerman, P. Fragile X-associated tremor/ataxia syndrome - features, mechanisms and management. *Nat Rev Neurol* **12**, 403-412 (2016).

145 Sellier, C. *et al.* Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron* **93**, 331-347 (2017).

146 Buijsen, R. A. *et al.* FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun* **2**, 162 (2014).

147 Oh, S. Y. *et al.* RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. *Human Molecular Genetics* **24**, 4317-4326 (2015).

148 Hukema, R. K. *et al.* Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. *Hum Mol Genet* **24**, 4948-4957 (2015).

149 Buijsen, R. A. *et al.* Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. *Hum Reprod* **31**, 158-168 (2016).

150 Kearse, M. G. *et al.* CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**, 314-322 (2016).

151 DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256 (2011).

152 Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268 (2011).

153 Johnston, C. A. *et al.* Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. *J Neurol* **253**, 1642-1643 (2006).

154 Onyike, C. U. & Diehl-Schmid, J. The epidemiology of frontotemporal dementia. *Int Rev Psychiatry* **25**, 130-137 (2013).

155 Lomen-Hoerth, C. *et al.* Are amyotrophic lateral sclerosis patients cognitively normal? *Neurology* **60**, 1094-1097 (2003).

156 Ringholz, G. M. *et al.* Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology* **65**, 586-590 (2005).

157 Lomen-Hoerth, C., Anderson, T. & Miller, B. The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. *Neurology* **59**, 1077-1079 (2002).

158 Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130-133 (2006).

159 Gijselinck, I. *et al.* A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol* **11**, 54-65 (2012).

160 Fratta, P. *et al.* C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2**, 1016 (2012).

161 Haeusler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195-200 (2014).

162 Reddy, K., Zamiri, B., Stanley, S. Y., Macgregor, R. B., Jr. & Pearson, C. E. The disease-associated r(GGGGCC)n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J Biol Chem* **288**, 9860-9866 (2013).

163 Su, Z. *et al.* Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**, 1043-1050 (2014).

164 Al-Sarraj, S. *et al.* p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS. *Acta Neuropathol* **122**, 691-702 (2011).

165 Boxer, A. L. *et al.* Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. *J Neurol Neurosurg Psychiatry* **82**, 196-203 (2011).

166 Schludi, M. H. *et al.* Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathol* **130**, 537-555 (2015).

167 Zhang, Y. J. *et al.* Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* **128**, 505-524 (2014).

168 Mackenzie, I. R. *et al.* Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol* **130**, 845-861 (2015).

169 Green, K. M. *et al.* RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications* **8**, 2005 (2017).

170 Tabet, R. *et al.* CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nature Communications* **9**, 152 (2018).

171 Jovicic, A. *et al.* Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18**, 1226-1229 (2015).

172 Freibaum, B. D. *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129-133 (2015).

173 Mizielinska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192-1194 (2014).

174 Tran, H. *et al.* Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* **87**, 1207-1214 (2015).

175 Wen, X. *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* **84**, 1213-1225 (2014).

176 Yang, D. *et al.* FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathol* **130**, 525-535 (2015).

177 Tao, Z. *et al.* Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet* **24**, 2426-2441 (2015).

178 Yamakawa, M. *et al.* Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum Mol Genet* **24**, 1630-1645 (2015).

179 May, S. *et al.* C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* **128**, 485-503 (2014).

180 Boeynaems, S. *et al.* Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Sci Rep* **6**, 20877 (2016).

181 Zhang, Y. J. *et al.* C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci* **19**, 668-677 (2016).

182 Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56-61 (2015).

183 Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729-1732 (2009).

184 Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *PNAS* **108**, 4334-4339 (2011).

185 Hubstenberger, A., Noble, S. L., Cameron, C. & Evans, T. C. Translation repressors, an RNA helicase, and developmental cues control RNP phase transitions during early development. *Dev Cell* **27**, 161-173 (2013).

186 Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767 (2012).

187 Lin, Y., Protter, D. S., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell* **60**, 208-219 (2015).

188 Han, T. W. *et al.* Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768-779 (2012).

189 Mackenzie, I. R. *et al.* TIA1 Mutations in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Promote Phase Separation and Alter Stress Granule Dynamics. *Neuron* **95**, 808-816 e809 (2017).

190 Martinez, F. J. *et al.* Protein-RNA Networks Regulated by Normal and ALS-Associated Mutant HNRNPA2B1 in the Nervous System. *Neuron* **92**, 780-795 (2016).

191 Molliex, A. *et al.* Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123-133 (2015).

192 Murakami, T. *et al.* ALS/FTD Mutation-Induced Phase Transition of FUS Liquid Droplets and Reversible Hydrogels into Irreversible Hydrogels Impairs RNP Granule Function. *Neuron* **88**, 678-690 (2015).

193 Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066-1077 (2015).

194 Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139-1145 (2014).

195 Lee, K. H. *et al.* C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **167**, 774-788 e717 (2016).

196 Lin, Y. *et al.* Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell* **167**, 789-802 e712 (2016).

197 Yin, S. *et al.* Evidence that C9ORF72 Dipeptide Repeat Proteins Associate with U2 snRNP to Cause Mis-splicing in ALS/FTD Patients. *Cell Rep* **19**, 2244-2256 (2017).

198 Fay, M. M., Anderson, P. J. & Ivanov, P. ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions In Vitro and in Cells. *Cell Rep* **21**, 3573-3584 (2017).

199 Niblock, M. *et al.* Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. *Acta Neuropathol Commun* **4**, 18 (2016).

200 Singh, C. R. *et al.* Eukaryotic translation initiation factor 5 is critical for integrity of the scanning preinitiation complex and accurate control of GCN4 translation. *Molecular and Cellular Biology* **25**, 5480-5491 (2005).

201 Wek, R. C. Role of eIF2alpha Kinases in Translational Control and Adaptation to Cellular Stress. *Cold Spring Harbor Perspectives in Biology* **10** (2018).

202 Cheng, W. *et al.* C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation. *Nat Commun* **9**, 51 (2018).

203 Sonobe, Y. *et al.* Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol Dis* **116**, 155-165 (2018).

204 Westergard, T. *et al.* Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med* **11** (2019).

205 Rodriguez, C. M. *et al.* A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat Neurosci* **23**, 386-397 (2020).

206 Kearse, M. G. & Wilusz, J. E. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes & Development* **31**, 1717-1731 (2017).

207 Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802 (2011).

208 Hagerman, P. J. & Hagerman, R. J. Fragile X-associated tremor/ataxia syndrome. *Ann N Y Acad Sci* **1338**, 58-70 (2015).

209 Glineburg, M. R., Todd, P. K., Charlet-Berguerand, N. & Sellier, C. Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. *Brain Res* **1693**, 43-54 (2018).

210 Hagerman, R. J. *et al.* Fragile X syndrome. *Nat Rev Dis Primers* **3**, 17065 (2017).

211 Leppek, K., Das, R. & Barna, M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews. Molecular Cell Biology* **19**, 158-174 (2018).

212 Jan, E. *et al.* Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. *Cold Spring Harbor Symposia on Quantitative Biology* **66**, 285-292 (2001).

213 Gao, F. B., Richter, J. D. & Cleveland, D. W. Rethinking Unconventional Translation in Neurodegeneration. *Cell* **171**, 994-1000 (2017).

214 Green, K. M., Linsalata, A. E. & Todd, P. K. RAN translation-What makes it run? *Brain Res* (2016).

215 Pisareva, V. P., Pisarev, A. V., Komar, A. A., Hellen, C. U. & Pestova, T. V. Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. *Cell* **135**, 1237-1250 (2008).

216 Jin, P. *et al.* RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron* **39**, 739-747 (2003).

217 Carvajal, F. *et al.* Structural domains within the HIV-1 mRNA and the ribosomal protein S25 influence cap-independent translation initiation. *The FEBS Journal* **283**, 2508-2527 (2016).

218 Landry, D. M., Hertz, M. I. & Thompson, S. R. RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes & Development* **23**, 2753-2764 (2009).

219 Hautbergue, G. M. *et al.* SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nature Communications* **8**, 16063 (2017).

220 Castrillon, D. H. *et al.* Toward a molecular genetic analysis of spermatogenesis in Drosophila melanogaster: characterization of male-sterile mutants generated by single P element mutagenesis. *Genetics* **135**, 489-505 (1993).

221 Johnstone, O. *et al.* Belle is a Drosophila DEAD-box protein required for viability and in the germ line. *Developmental Biology* **277**, 92-101 (2005).

222 Poulton, J. S. *et al.* The microRNA pathway regulates the temporal pattern of Notch signaling in Drosophila follicle cells. *Development* **138**, 1737-1745 (2011).

223 Spradling, A. C. *et al.* The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* **153**, 135-177 (1999).

224 Osterwalder, T., Yoon, K. S., White, B. H. & Keshishian, H. A conditional tissue-specific transgene expression system using inducible GAL4. *PNAS* **98**, 12596-12601 (2001).

225 Hinnebusch, A. G., Ivanov, I. P. & Sonenberg, N. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413-1416 (2016).

226 Young, S. K. & Wek, R. C. Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. *The Journal of Biological Chemistry* **291**, 16927-16935 (2016).

227 Chakrabarti, A. & Maitra, U. Function of eukaryotic initiation factor 5 in the formation of an 80 S ribosomal polypeptide chain initiation complex. *The Journal of Biological Chemistry* **266**, 14039-14045 (1991).

228 Das, S., Maiti, T., Das, K. & Maitra, U. Specific interaction of eukaryotic translation initiation factor 5 (eIF5) with the beta-subunit of eIF2. *The Journal of Biological Chemistry* **272**, 31712-31718 (1997).

229 Nanda, J. S., Saini, A. K., Munoz, A. M., Hinnebusch, A. G. & Lorsch, J. R. Coordinated movements of eukaryotic translation initiation factors eIF1, eIF1A, and eIF5 trigger phosphate release from eIF2 in response to start codon recognition by the ribosomal preinitiation complex. *The Journal of Biological Chemistry* **288**, 5316-5329 (2013).

230 Nanda, J. S. *et al.* eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation. *Journal of Molecular Biology* **394**, 268-285 (2009).

231 Saini, A. K. *et al.* Eukaryotic translation initiation factor eIF5 promotes the accuracy of start codon recognition by regulating Pi release and conformational transitions of the preinitiation complex. *Nucleic Acids Research* **42**, 9623-9640, doi:10.1093/nar/gku653 (2014).

232 Pisareva, V. P. & Pisarev, A. V. eIF5 and eIF5B together stimulate 48S initiation complex formation during ribosomal scanning. *Nucleic Acids Research* **42**, 12052-12069 (2014).

233 Archbold, H. C. *et al.* TDP43 nuclear export and neurodegeneration in models of amyotrophic lateral sclerosis and frontotemporal dementia. *Scientific Reports* **8**, 4606 (2018).

234 Barmada, S. J. *et al.* Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *PNAS* **112**, 7821-7826 (2015).

235 Malik, A. M. *et al.* Matrin 3-dependent neurotoxicity is modified by nucleic acid binding and nucleocytoplasmic localization. *eLife* **7** (2018).

236 Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Research* **36**, W70-74 (2008).

237 Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology : AMB* **6**, 26 (2011).

238 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415 (2003).

239 Grifo, J. A. *et al.* Characterization of eukaryotic initiation factor 4A, a protein involved in ATP-dependent binding of globin mRNA. *The Journal of Biological Chemistry* **257**, 5246-5252 (1982).

240 Ray, B. K. *et al.* ATP-dependent unwinding of messenger RNA structure by eukaryotic initiation factors. *The Journal of Biological Chemistry* **260**, 7651-7658 (1985).

241 Hann, S. R. Regulation and function of non-AUG-initiated proto-oncogenes. *Biochimie* **76**, 880-886 (1994).

242 Meijer, H. A. & Thomas, A. A. Control of eukaryotic protein synthesis by upstream open reading frames in the 5'-untranslated region of an mRNA. *The Biochemical Journal* **367**, 1-11 (2002).

243 Touriol, C. *et al.* Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the Cell* **95**, 169-178 (2003).

244 Bol, G. M. *et al.* Expression of the RNA helicase DDX3 and the hypoxia response in breast cancer. *PloS One* **8**, e63548 (2013).

245 Brai, A. *et al.* Human DDX3 protein is a valuable target to develop broad spectrum antiviral agents. *PNAS* **113**, 5388-5393 (2016).

246 Samal, S. K., Routray, S., Veeramachaneni, G. K., Dash, R. & Botlagunta, M. Ketorolac salt is a newly discovered DDX3 inhibitor to treat oral cancer. *Scientific Reports* **5**, 9982 (2015).

247 Pandey, U. B. *et al.* HDAC6 rescues neurodegeneration and provides an essential link between autophagy and the UPS. *Nature* **447**, 859-863 (2007).

248 Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).

249 Rakotondrafara, A. M. & Hentze, M. W. An efficient factor-depleted mammalian in vitro translation system. *Nature Protocols* **6**, 563-571 (2011).

250 Caschera, F. & Noireaux, V. Preparation of amino acid mixtures for cell-free expression systems. *BioTechniques* **58**, 40-43 (2015).

251 Simsek, D. *et al.* The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell* **169**, 1051-1065 e1018 (2017).

252 Luthe, D. S. A simple technique for the preparation and storage of sucrose gradients. *Anal Biochem* **135**, 230-232 (1983).

253 Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M. E. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* **95**, 55-66 (1998).

254 Barmada, S. J. *et al.* Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience* **30**, 639-649 (2010).

255 del Aguila, M. A., Longstreth, W. T., Jr., McGuire, V., Koepsell, T. D. & van Belle, G. Prognosis in amyotrophic lateral sclerosis: a population-based study. *Neurology* **60**, 813-819 (2003).

256 Mateen, F. J., Carone, M. & Sorenson, E. J. Patients who survive 5 years or more with ALS in Olmsted County, 1925-2004. *J Neurol Neurosurg Psychiatry* **81**, 1144-1146 (2010).

257 Coyle-Gilchrist, I. T. *et al.* Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. *Neurology* **86**, 1736-1743 (2016).

258 Garcin, B. *et al.* Determinants of survival in behavioral variant frontotemporal dementia. *Neurology* **73**, 1656-1661 (2009).

259 Hodges, J. R., Davies, R., Xuereb, J., Kril, J. & Halliday, G. Survival in frontotemporal dementia. *Neurology* **61**, 349-354 (2003).

260 Roberson, E. D. *et al.* Frontotemporal dementia progresses to death faster than Alzheimer disease. *Neurology* **65**, 719-725 (2005).

261 Taylor, J. P., Brown, R. H., Jr. & Cleveland, D. W. Decoding ALS: from genes to mechanism. *Nature* **539**, 197-206 (2016).

262 Xue, Y. C. *et al.* Enteroviral Infection Leads to Transactive Response DNA-Binding Protein 43 Pathology in Vivo. *Am J Pathol* **188**, 2853-2862 (2018).

263 Smith, E. F., Shaw, P. J. & De Vos, K. J. The role of mitochondria in amyotrophic lateral sclerosis. *Neurosci Lett* **710**, 132933 (2019).

264 Medinas, D. B., Valenzuela, V. & Hetz, C. Proteostasis disturbance in amyotrophic lateral sclerosis. *Hum Mol Genet* **26**, R91-R104 (2017).

265 Ramesh, N. & Pandey, U. B. Autophagy Dysregulation in ALS: When Protein Aggregates Get Out of Hand. *Front Mol Neurosci* **10**, 263 (2017).

266 Armon, C. From Snow to Hill to ALS: An epidemiological odyssey in search of ALS causation. *J Neurol Sci* **391**, 134-140 (2018).

267 Konopka, A. & Atkin, J. D. The Emerging Role of DNA Damage in the Pathogenesis of the C9orf72 Repeat Expansion in Amyotrophic Lateral Sclerosis. *Int J Mol Sci* **19** (2018).

268 Belzil, V. V., Katzman, R. B. & Petrucelli, L. ALS and FTD: an epigenetic perspective. *Acta Neuropathol* **132**, 487-502 (2016).

269 Wang, J., Haeusler, A. R. & Simko, E. A. Emerging role of RNA*DNA hybrids in C9orf72-linked neurodegeneration. *Cell Cycle* **14**, 526-532 (2015).

270 Alberti, S., Mateju, D., Mediani, L. & Carra, S. Granulostasis: Protein Quality Control of RNP Granules. *Front Mol Neurosci* **10**, 84 (2017).

271 Ito, D., Hatano, M. & Suzuki, N. RNA binding proteins and the pathological cascade in ALS/FTD neurodegeneration. *Sci Transl Med* **9** (2017).

272 Majounie, E. *et al.* Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol* **11**, 323-330 (2012).

273 Bilen, J. & Bonini, N. M. Genome-wide screen for modifiers of ataxin-3 neurodegeneration in Drosophila. *PLoS Genet* **3**, 1950-1964 (2007).

274 Elden, A. C. *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466**, 1069-1075 (2010).

275 Kaltenbach, L. S. *et al.* Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genet* **3**, e82 (2007).

276 Kazemi-Esfarjani, P. & Benzer, S. Genetic suppression of polyglutamine toxicity in Drosophila. *Science* **287**, 1837-1840 (2000).

277 Lessing, D. & Bonini, N. M. Polyglutamine genes interact to modulate the severity and progression of neurodegeneration in Drosophila. *PLoS Biol* **6**, e29 (2008).

278 Li, L. B., Yu, Z., Teng, X. & Bonini, N. M. RNA toxicity is a component of ataxin-3 degeneration in Drosophila. *Nature* **453**, 1107-1111 (2008).

279 Chai, N. & Gitler, A. D. Yeast screen for modifiers of C9orf72 poly(glycine-arginine) dipeptide repeat toxicity. *FEMS Yeast Res* **18** (2018).

280 Kramer, N. J. *et al.* CRISPR-Cas9 screens in human cells and primary neurons identify modifiers of C9ORF72 dipeptide-repeat-protein toxicity. *Nat Genet* **50**, 603-612 (2018).

281 Donnelly, C. J. *et al.* RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* **80**, 415-428 (2013).

282 Lopez-Gonzalez, R. *et al.* Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. *Neuron* **92**, 383-391 (2016).

283 Ayhan, F. *et al.* SCA8 RAN polySer protein preferentially accumulates in white matter regions and is regulated by eIF3F. *EMBO J* **37** (2018).

284 Buehler, E. *et al.* siRNA off-target effects in genome-wide screens identify signaling pathway members. *Sci Rep* **2**, 428 (2012).

285 Marine, S., Bahl, A., Ferrer, M. & Buehler, E. Common seed analysis to identify off-target effects in siRNA screens. *J Biomol Screen* **17**, 370-378 (2012).

286 Schultz, N. *et al.* Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-beta pathway and reveal microRNA regulation of TGFBR2. *Silence* **2**, 3 (2011).

287 Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol* **34**, 634-636 (2016).

288 Arguel, M. J. *et al.* A cost effective 5 selective single cell transcriptome profiling approach with improved UMI design. *Nucleic Acids Res* **45**, e48 (2017).

289 Lahvic, J. L. *et al.* Specific oxylipins enhance vertebrate hematopoiesis via the receptor GPR132. *PNAS* **115**, 9252-9257 (2018).

290 Ricci, E. P. *et al.* Staufen1 senses overall transcript secondary structure to regulate translation. *Nat Struct Mol Biol* **21**, 26-35 (2014).

291 Woo, Y. M. *et al.* TED-Seq Identifies the Dynamics of Poly(A) Length during ER Stress. *Cell Rep* **24**, 3630-3641 e3637 (2018).

292 Ahsan, S. & Draghici, S. Identifying Significantly Impacted Pathways and Putative Mechanisms with iPathwayGuide. *Curr Protoc Bioinformatics* **57**, 7 15 11-17 15 30 (2017).

293 Donato, M. *et al.* Analysis and correction of crosstalk effects in pathway analysis. *Genome Res* **23**, 1885-1893 (2013).

294 Draghici, S. *et al.* A systems biology approach for pathway level analysis. *Genome Res* **17**, 1537-1545 (2007).

295 Tarca, A. L. *et al.* A novel signaling pathway impact analysis. *Bioinformatics* **25** (2009).

296 Deng, H. X. *et al.* Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. *Nature* **477**, 211-215 (2011).

297 Golan-Gerstl, R. *et al.* Splicing factor hnRNP A2/B1 regulates tumor suppressor gene splicing and is an oncogenic driver in glioblastoma. *Cancer Res* **71**, 4464-4472 (2011).

298 Moran-Jones, K., Grindlay, J., Jones, M., Smith, R. & Norman, J. C. hnRNP A2 regulates alternative mRNA splicing of TP53INP2 to control invasive cell migration. *Cancer Res* **69**, 9219-9227 (2009).

299 Hung, C. Y. *et al.* Nm23-H1-stabilized hnRNPA2/B1 promotes internal ribosomal entry site (IRES)-mediated translation of Sp1 in the lung cancer progression. *Sci Rep* **7**, 9166 (2017).

300 Kosturko, L. D. *et al.* Heterogeneous nuclear ribonucleoprotein (hnRNP) E1 binds to hnRNP A2 and inhibits translation of A2 response element mRNAs. *Mol Biol Cell* **17**, 3521-3533 (2006).

301 White, R. *et al.* Heterogeneous nuclear ribonucleoprotein (hnRNP) F is a novel component of oligodendroglial RNA transport granules contributing to regulation of myelin basic protein (MBP) synthesis. *J Biol Chem* **287**, 1742-1754 (2012).

302 White, R. *et al.* Activation of oligodendroglial Fyn kinase enhances translation of mRNAs transported in hnRNP A2-dependent RNA granules. *J Cell Biol* **181**, 579-586 (2008).

303 Adams, J. *et al.* Proteasome inhibitors: a novel class of potent and effective antitumor agents. *Cancer Res* **59**, 2615-2622 (1999).

304 Berkers, C. R. *et al.* Activity probe for in vivo profiling of the specificity of proteasome inhibitor bortezomib. *Nat Methods* **2**, 357-362 (2005).

305 Freibaum, B. D. & Taylor, J. P. The Role of Dipeptide Repeats in C9ORF72-Related ALS-FTD. *Front Mol Neurosci* **10**, 35 (2017).

306 Yuva-Aydemir, Y., Almeida, S. & Gao, F. B. Insights into C9ORF72-Related ALS/FTD from Drosophila and iPSC Models. *Trends Neurosci* **41**, 457-469 (2018).

307 Khateb, S., Weisman-Shomer, P., Hershco-Shani, I., Ludwig, A. L. & Fry, M. The tetraplex (CGG)n destabilizing proteins hnRNP A2 and CBF-A enhance the in vivo translation of fragile X premutation mRNA. *Nucleic Acids Res* **35**, 5775-5788 (2007).

308 Voges, D., Zwickl, P. & Baumeister, W. The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem* **68**, 1015-1068 (1999).

309 Gupta, R. *et al.* The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome. *eNeuro* **4** (2017).

310 Cristofani, R. *et al.* The small heat shock protein B8 (HSPB8) efficiently removes aggregating species of dipeptides produced in C9ORF72-related neurodegenerative diseases. *Cell Stress Chaperones* **23**, 1-12 (2018).

311 Boes, B. *et al.* Interferon gamma stimulation modulates the proteolytic activity and cleavage site preference of 20S mouse proteasomes. *J Exp Med* **179**, 901-909 (1994).

312 Driscoll, J., Brown, M. G., Finley, D. & Monaco, J. J. MHC-linked LMP gene products specifically alter peptidase activities of the proteasome. *Nature* **365**, 262-264 (1993).

313 Gaczynska, M., Rock, K. L., Spies, T. & Goldberg, A. L. Peptidase activities of proteasomes are differentially regulated by the major histocompatibility complex-encoded genes for LMP2 and LMP7. *PNAS* **91**, 9213-9217 (1994).

314 Groettrup, M. *et al.* The interferon-gamma-inducible 11 S regulator (PA28) and the LMP2/LMP7 subunits govern the peptide production by the 20 S proteasome in vitro. *J Biol Chem* **270**, 23808-23815 (1995).

315 Mishto, M. *et al.* Proteasome isoforms exhibit only quantitative differences in cleavage and epitope generation. *Eur J Immunol* **44**, 3508-3521 (2014).

316 Raule, M., Cerruti, F. & Cascio, P. Enhanced rate of degradation of basic proteins by 26S immunoproteasomes. *Biochim Biophys Acta* **1843**, 1942-1947 (2014).

317 Altun, M. *et al.* Effects of PS-341 on the activity and composition of proteasomes in multiple myeloma cells. *Cancer Res* **65**, 7896-7901 (2005).

318 Blackburn, C. *et al.* Characterization of a new series of non-covalent proteasome inhibitors with exquisite potency and selectivity for the 20S beta5-subunit. *Biochem J* **430**, 461-476, doi:10.1042/BJ20100383 (2010).

319 Demo, S. D. *et al.* Antitumor activity of PR-171, a novel irreversible inhibitor of the proteasome. *Cancer Res* **67**, 6383-6391 (2007).

320 Muchamuel, T. *et al.* A selective inhibitor of the immunoproteasome subunit LMP7 blocks cytokine production and attenuates progression of experimental arthritis. *Nat Med* **15**, 781-787 (2009).

321 Christ, A., Herzog, K. & Willnow, T. E. LRP2, an auxiliary receptor that controls sonic hedgehog signaling in development and disease. *Dev Dyn* **245**, 569-579 (2016).

322 Kelly, R. J., Lopez-Chavez, A., Citrin, D., Janik, J. E. & Morris, J. C. Impacting tumor cell-fate by targeting the inhibitor of apoptosis protein survivin. *Mol Cancer* **10**, 35 (2011).

323 Choi, J. H. *et al.* hnRNP Q Regulates Internal Ribosome Entry Site-Mediated fmr1 Translation in Neurons. *Mol Cell Biol* **39** (2019).

324 Ludwig, A. L., Hershey, J. W. & Hagerman, P. J. Initiation of translation of the FMR1 mRNA Occurs predominantly through 5'-end-dependent ribosomal scanning. *J Mol Biol* **407**, 21-34 (2011).

325 Yamada, S. B. *et al.* RPS25 is required for efficient RAN translation of C9orf72 and other neurodegenerative disease-associated nucleotide repeats. *Nat Neurosci* **22**, 1383-1388 (2019).

326 Cruciat, C. M. *et al.* RNA helicase DDX3 is a regulatory subunit of casein kinase 1 in Wnt-beta-catenin signaling. *Science* **339**, 1436-1441 (2013).

327 Bustamante, H. A. *et al.* Interplay Between the Autophagy-Lysosomal Pathway and the Ubiquitin-Proteasome System: A Target for Therapeutic Development in Alzheimer's Disease. *Front Cell Neurosci* **12**, 126 (2018).

328 Ciechanover, A. & Kwon, Y. T. Protein Quality Control by Molecular Chaperones in Neurodegeneration. *Front Neurosci* **11**, 185 (2017).

329 Maurel, C. *et al.* Causative Genes in Amyotrophic Lateral Sclerosis and Protein Degradation Pathways: a Link to Neurodegeneration. *Mol Neurobiol* **55**, 6480-6499 (2018).

330 Opoku-Nsiah, K. A. & Gestwicki, J. E. Aim for the core: suitability of the ubiquitin-independent 20S proteasome as a drug target in neurodegeneration. *Transl Res* **198**, 48-57 (2018).

331 O'Dea, E. & Hoffmann, A. NF-kappaB signaling. *Wiley Interdiscip Rev Syst Biol Med* **1**, 107-115 (2009).

332 Tan, J. *et al.* Tumor Necrosis Factor-alpha Attenuates the Osteogenic Differentiation Capacity of Periodontal Ligament Stem Cells by Activating PERK Signaling. *J Periodontol* **87**, e159-171 (2016).

333 Qiao, Q. *et al.* Endoplasmic reticulum stress pathway PERK-eIF2alpha confers radioresistance in oropharyngeal carcinoma by activating NF-kappaB. *Cancer Sci* **108**, 1421-1431 (2017).

334 Denis, R. G. *et al.* TNF-alpha transiently induces endoplasmic reticulum stress and an incomplete unfolded protein response in the hypothalamus. *Neuroscience* **170**, 1035-1044 (2010).

335 Coots, R. A. *et al.* m(6)A Facilitates eIF4F-Independent mRNA Translation. *Mol Cell* **68**, 504-514 e507 (2017).

336 Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m(6)A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol Cell* **62**, 335-345 (2016).

337 Meyer, K. D. *et al.* 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell* **163**, 999-1010 (2015).

338 Zhou, J. *et al.* Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591-594 (2015).

339 Ries, R. J. *et al.* m(6)A enhances the phase separation potential of mRNA. *Nature* **571**, 424-428 (2019).

340 Koh, C. W. Q., Goh, Y. T. & Goh, W. S. S. Atlas of quantitative single-base-resolution N(6)-methyl-adenine methylomes. *Nat Commun* **10**, 5636 (2019).

341 Nance, D. J. *et al.* Characterization of METTL16 as a cytoplasmic RNA binding protein. *PLoS One* **15**, e0227647 (2020).

342 Pendleton, K. E. *et al.* The U6 snRNA m(6)A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169**, 824-835 e814 (2017).

343 Warda, A. S. *et al.* Human METTL16 is a N(6)-methyladenosine (m(6)A) methyltransferase that targets pre-mRNAs and various non-coding RNAs. *EMBO Rep* **18**, 2004-2014 (2017).

344 Park, E. M., Lim, Y. S., Ahn, B. Y. & Hwang, S. B. AAM-B Interacts with Nonstructural 4B and Regulates Hepatitis C Virus Propagation. *PLoS One* **10**, e0132839 (2015).

345 Berulava, T. *et al.* FTO levels affect RNA modification and the transcriptome. *Eur J Hum Genet* **21**, 317-323 (2013).

346 Fu, Y. *et al.* FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat Commun* **4**, 1798 (2013).

347 Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol* **7**, 885-887 (2011).

348 Mauer, J. *et al.* Reversible methylation of m(6)Am in the 5' cap controls mRNA stability. *Nature* **541**, 371-375 (2017).

349 Falnes, P. O., Bjoras, M., Aas, P. A., Sundheim, O. & Seeberg, E. Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res* **32**, 3456-3461 (2004).

350 Solomon, O. *et al.* RNA editing by ADAR1 leads to context-dependent transcriptome-wide changes in RNA secondary structure. *Nat Commun* **8**, 1440 (2017).

351 Wang, I. X. *et al.* ADAR regulates RNA editing, transcript stability, and gene expression. *Cell Rep* **5**, 849-860 (2013).

352 Ebbert, M. T. W. *et al.* Conserved DNA methylation combined with differential frontal cortex and cerebellar expression distinguishes C9orf72-associated and sporadic ALS, and implicates SERPINA1 in disease. *Acta Neuropathol* **134**, 715-728 (2017).

353 Caldecott, K. W. XRCC1 protein; Form and function. *DNA Repair (Amst)* **81**, 102664 (2019).

354 Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. *Mol Cell* **33**, 1-13 (2009).

355 Dormann, D. *et al.* Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. *EMBO J* **31**, 4258-4275 (2012).

356 Scaramuzzino, C. *et al.* Protein arginine methyltransferase 1 and 8 interact with FUS to modify its sub-cellular distribution and toxicity in vitro and in vivo. *PLoS One* **8**, e61576 (2013).

357 Hofweber, M. *et al.* Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell* **173**, 706-719 e713 (2018).

358 Lo Piccolo, L., Mochizuki, H. & Nagai, Y. The lncRNA hsromega regulates arginine dimethylation of human FUS to cause its proteasomal degradation in Drosophila. *J Cell Sci* **132** (2019).

359 Matsumoto, K. *et al.* PRMT1 is required for RAP55 to localize to processing bodies. *RNA Biol* **9**, 610-623 (2012).

360 Poornima, G., Shah, S., Vignesh, V., Parker, R. & Rajyaguru, P. I. Arginine methylation promotes translation repression activity of eIF4G-binding protein, Scd6. *Nucleic Acids Res* **44**, 9358-9368 (2016).

361 Yang, W. H., Yu, J. H., Gulick, T., Bloch, K. D. & Bloch, D. B. RNA-associated protein 55 (RAP55) localizes to mRNA processing bodies and stress granules. *RNA* **12**, 547-554 (2006).

362 Rajyaguru, P., She, M. & Parker, R. Scd6 targets eIF4G to repress translation: RGG motif proteins as a class of eIF4G-binding proteins. *Mol Cell* **45**, 244-254 (2012).

363 Choi, H. M. *et al.* Mapping a multiplexed zoo of mRNA expression. *Development* **143**, 3632-3637 (2016).

364 Lee, C. S. *et al.* Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. *Nucleic Acids Res* **36**, 4708-4718 (2008).

365 Lo, P. K. *et al.* RNA helicase Belle/DDX3 regulates transgene expression in Drosophila. *Dev Biol* **412**, 57-70 (2016).

366 Pek, J. W. & Kai, T. DEAD-box RNA helicase Belle/DDX3 and the RNA interference pathway promote mitotic chromosome segregation. *PNAS* **108**, 12007-12012 (2011).

367 Tsai, T. Y. *et al.* RNA helicase DDX3 maintains lipid homeostasis through upregulation of the microsomal triglyceride transfer protein by interacting with HNF4 and SHP. *Sci Rep* **7**, 41452 (2017).

368 Homan, P. J. *et al.* Single-molecule correlated chemical probing of RNA. *Proc Natl Acad Sci U S A* **111**, 13858-13863 (2014).

369 Krokhotin, A., Mustoe, A. M., Weeks, K. M. & Dokholyan, N. V. Direct identification of base-paired RNA nucleotides by correlated chemical probing. *RNA* **23**, 6-13 (2017).

370 Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA-Protein Interactions in Living Cells with SHAPE. *Biochemistry* **54**, 6867-6875 (2015).

371 Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* **10**, 1643-1669 (2015).

372 Talkish, J., May, G., Lin, Y., Woolford, J. L., Jr. & McManus, C. J. Mod-seq: high-throughput sequencing for chemical probing of RNA structure. *RNA* **20**, 713-720 (2014).

373 Aki, M. *et al.* Interferon-gamma induces different subunit organizations and functional diversity of proteasomes. *J Biochem* **115**, 257-269 (1994).

374 Hisamatsu, H. *et al.* Newly identified pair of proteasomal subunits regulated reciprocally by interferon gamma. *J Exp Med* **183**, 1807-1816 (1996).

375 Kimura, H., Caturegli, P., Takahashi, M. & Suzuki, K. New Insights into the Function of the Immunoproteasome in Immune and Nonimmune Cells. *J Immunol Res* **2015**, 541984 (2015).

376 Diaz-Hernandez, M. *et al.* Neuronal induction of the immunoproteasome in Huntington's disease. *J Neurosci* **23**, 11653-11661 (2003).

377 Nardo, G. *et al.* Transcriptomic indices of fast and slow disease progression in two mouse models of amyotrophic lateral sclerosis. *Brain* **136**, 3305-3332 (2013).

378 Ugras, S. *et al.* Induction of the Immunoproteasome Subunit Lmp7 Links Proteostasis and Immunity in alpha-Synuclein Aggregation Disorders. *EBioMedicine* **31**, 307-319 (2018).

379 Ettari, R. *et al.* Immunoproteasome-selective and non-selective inhibitors: A promising approach for the treatment of multiple myeloma. *Pharmacol Ther* **182**, 176-192 (2018).

380 Ogorevc, E., Schiffrer, E. S., Sosic, I. & Gobec, S. A patent review of immunoproteasome inhibitors. *Expert Opin Ther Pat* **28**, 517-540 (2018).

381 Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat Rev Neurol* **14**, 544-558 (2018).

382 Cheng, W. *et al.* CRISPR-Cas9 Screens Identify the RNA Helicase DDX3X as a Repressor of C9ORF72 (GGGGCC)n Repeat-Associated Non-AUG Translation. *Neuron* **104**, 885-898 e888 (2019).

383 Chai, N. *et al.* Genome-wide synthetic lethal CRISPR screen identifies FIS1 as a genetic interactor of ALS-linked C9ORF72. *Brain Res* **1728**, 146601 (2020).

384 Goodman, L. D. *et al.* Toxic expanded GGGGCC repeat transcription is mediated by the PAF1 complex in C9orf72-associated FTD. *Nat Neurosci* **22**, 863-874 (2019).

385 Goodman, L. D. *et al.* eIF4B and eIF4H mediate GR production from expanded G4C2 in a Drosophila model for C9orf72-associated ALS. *Acta Neuropathol Commun* **7**, 62 (2019).

386 Lopez-Gonzalez, R. *et al.* Partial inhibition of the overactivated Ku80-dependent DNA repair pathway rescues neurodegeneration in C9ORF72-ALS/FTD. *Proc Natl Acad Sci U S A* **116**, 9628-9633 (2019).

387 Yuva-Aydemir, Y., Almeida, S., Krishnan, G., Gendron, T. F. & Gao, F. B. Transcription elongation factor AFF2/FMR2 regulates expression of expanded GGGGCC repeat-containing C9ORF72 allele in ALS/FTD. *Nat Commun* **10**, 5466 (2019).

388 Moens, T. G. *et al.* C9orf72 arginine-rich dipeptide proteins interact with ribosomal proteins in vivo to induce a toxic translational arrest that is rescued by eIF1A. *Acta Neuropathol* **137**, 487-500 (2019).

389 Gilmore, T. D. & Herscovitch, M. Inhibitors of NF-kappaB signaling: 785 and counting. *Oncogene* **25**, 6887-6899 (2006).

390 Borjesson, A., Grundmark, B., Olaisson, H. & Waldenlind, L. Is there a link between amyotrophic lateral sclerosis and treatment with TNF-alpha inhibitors? *Ups J Med Sci* **118**, 199-200 (2013).

391 Petitpain, N. *et al.* Is TNF inhibitor exposure a risk factor for amyotrophic lateral sclerosis? *Fundam Clin Pharmacol* **33**, 689-694 (2019).

392 Bauer, P. O. Methylation of C9orf72 expansion reduces RNA foci formation and dipeptide-repeat proteins expression in cells. *Neurosci Lett* **612**, 204-209 (2016).

393 Berson, A. *et al.* Drosophila Ref1/ALYREF regulates transcription and toxicity associated with ALS/FTD disease etiologies. *Acta Neuropathol Commun* **7**, 65 (2019).