Repeat-Associated non-AUG Translation Initiation at Expanded GGGGCC Repeats in C9orf72-Associated Amyotrophic Lateral Sclerosis and Frontotemporal Dementia

by

Katelyn M. Green

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

Doctoral Committee:

Associate Professor Peter K. Todd, Chair
Associate Professor Sami J. Barmada
Professor Miriam Meisler
Assistant Professor Jayakrishnan Nandakumar

Katelyn M. Green

rkatelyn@umich.edu

ORCID: 0000-0001-5562-9920

Dedication

To my best friend and husband, Dan Green.

Acknowledgements

Over the last five and a half years, the guidance, assistance, and support of more people than I can name made this thesis possible. However, the contributions of those listed below have been so great that I could not consider this written dissertation complete without their formal acknowledgement.

- First and foremost, I must acknowledge my mentor and principle investigator, Dr. Peter Todd, for being the biggest advocate of my career, and always being available to offer guidance and insights, but also allowing me space to grow as an independent scientist.
- Dr. Mike Kearse, for sharing with me his passion and seemingly endless knowledge of RNA biology and translation.
- Amy Krans, for always having an answer to my countless questions.
- Udit Sheth, for being a motivated and hardworking undergraduate assistant, and helping advance our studies with small molecule RAN translation inhibitors and the eIF2A KO mice.
- All other Todd Lab members, past and present, for the scientific discussions, troubleshooting advice, and many laughs. As Peter always says, science is a team sport, and I consider myself lucky to have had each of you on my team.

- The Barmada Lab, for the scientific camaraderie, and especially Drs. Sami Barmada, Elizabeth Tank, and Brittany Flores, for assistance with studies involving primary rodent neurons and patient-derived cell lines.
- The Paulson Lab for creating a pleasant shared workspace, and especially Dr. Hayley McLoughlin, Rob Komlo, Sveta Fischer, and Dr. Maria Do Carmo Costa, for training in mouse injections and behavioral assays.
- The Ivanova Lab for assistance and training in circular dichroism.
- My thesis committee members, Drs. Sami Barmada, Miriam Meisler, and JK Nandakumar, for their advice, critiques, and scientific inspiration throughout my thesis work.
- The Cellular and Molecular Biology Graduate Program, especially Dr. Bob Fuller and Pat Ocelnik, for academic guidance and assistance.
- And, most importantly, my family for supporting me throughout this long pursuit, celebrating my successes, and being a source of comfort or perspective during my setbacks.

During my dissertation work, I was supported by the following grants.

- NIH/NINDS NRSA F31NS100302 – June 2017 to May 2020
- NIH Cellular and Molecular Biology Training Grant T32-GM007315 – July 2015 to May 2017
- Rackham Graduate Student Research Grant – May 2017 & February 2018
- Rackham Conference Travel Grant - October 2015, November 2016, September 2017 & July 2019

Table of Contents

List of Figures

List of Tables

Abstract

A GGGGCC repeat expansion in C9orf72 causes amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Through a process termed repeat-associated non-AUG (RAN) translation, the expanded repeat is translated in all three readings frame, in the absence of an AUG start codon, to produce dipeptide repeat-containing proteins (DPRs). DPRs accumulate in patient neurons and several are toxic in diverse disease models. However, the non-canonical translation initiation mechanism through which DPRs are synthesized is poorly understood.

I developed luciferase-based GGGGCC RAN translation-specific reporters to quantitatively assess RAN translation levels in all three reading frames. Both *in vitro* in a rabbit reticulocyte lysate (RRL) and in cultured cells and neurons, I found that RAN translation of 70 GGGGCC repeats occurs most readily on mRNAs containing a functional 5′ mRNA cap, and through eIF4A-mediated ribosomal scanning. Additionally, GGGGCC RAN translation occurs most efficiently in the glycine-alanine (GA) reading frame, followed by the glycine-proline (GP) and glycine-arginine (GR) frames. This is due in part to use of a CUG initiation codon located upstream of the repeat in the GA frame, as mutating this codon greatly decreases polyGA production. Interestingly, unlike global translation, C9RAN translation is not inhibited by cellular stress. Instead, eIF2α phosphorylation during stress enhances C9RAN translation in a manner that depends upon its non-AUG initiation.

These findings identify aspects of RAN translation that are mechanistically distinct from canonical translation, suggesting the possibility of developing RAN translation-selective inhibitors that do not impair global translation. To assess this possibility, I adapted our RRL *in vitro* RAN translation assay to perform a high-throughput screen of 3253 small molecules. From this screen, I identified five small molecules that more significantly inhibit RAN translation than AUG-initiated translation across multiple reading frames of the GGGGCC repeat, as well as the CGG repeat causative of fragile X-associated tremor/ataxia syndrome. I further showed that while three of these inhibitors directly interact with repeat RNAs, two do not, suggesting multiple mechanisms by which RAN translation can be selectively targeted.

Additionally, as a potential mechanism by which C9RAN translation evades downregulation following eIF2α phosphorylation, I investigated the role of eIF2 alternatives in supporting C9RAN translation. eIF2A, eIF2D, and DENR/MCTS1 can function in place of eIF2 and deliver the initiator methionine tRNA to the pre-initiation ribosome. In HEK293 cells, DENR/MCTS1 knockdown more significantly reduces RAN translation than AUG-initiated translation, as does deletion of eIF2A in *in vitro* translation lysates. Furthermore, knockdown of each factor modestly reduces some repeat-associated toxicity in model organisms.

Lastly, while C9RAN translation reporters are a powerful tool, it is critical to understand how RAN translation occurs on endogenous repeat containing RNAs in patient cells. To determine the C9orf72 sequence found upstream of the normally intronic GGGGCC repeat, I performed repeat-primed 5′ RNA ligation mediated rapid amplification of cDNA ends on actively translated RNA isolated from control and

C9ALS/FTD patient-derived iNeurons following polysome profiling. This identified novel C9orf72 5′ ends that position the expanded GGGGCC repeat in a non-intronic context, providing evidence for one pathway through which the repeat becomes available to ribosomes.

Together, these studies support a cap- and scanning-dependent model of C9RAN translation, that uses a CUG start codon, occurs on 5′ truncated C9orf72 transcripts, is enhanced by cellular stress, and is selectively inhibited with small molecules or genetic manipulations, and provide groundwork for future RAN translation-targeting therapeutic development.

Chapter 1: Introduction

This chapter is adapted from the following publication:

Green KM*, Linsalata AE*, Todd PK. RAN translation – what makes it run? *Brain Research.* 2016 Sep 1. 1647: 30-42. **co-first authorship.*

1.1 Statement of others' contribution to work presented in this chapter

Alexander Linsalata was co-first author on the manuscript from which this chapter is derived. As such, he contributed approximately half the text, and generated **Fig. 1-2** and **Fig. 1-3**. However, as co-first author, I was involved in all decisions related to the content and structure of the entire manuscript, including each figure.

1.2 RAN Translation Overview

Nucleotide-repeat expansions underlie a heterogeneous group of primarily neurological diseases that impact a large number of patients (1). Repeats can cause problems through a variety of mechanisms delineated over the past 25 years. Expansion of trinucleotide repeats within protein-coding open reading frames (ORFs) cause a gain-of-function toxicity downstream of the production of polyglutamine or (less frequently) polyalanine proteins (2). This toxicity results from both alterations in the native functions of the protein in which the repeat resides as well as toxicity independent of protein context, related to perturbations in neuronal proteostasis. Repeat expansions 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 (3). Such non-coding repeats can also elicit toxicity as RNA by binding to and sequestering specific RNA-binding proteins via presentation of a repetitive motif (4).

The discovery of repeat-associated non-AUG (RAN)-initiated translation blurs the lines that define which repeats elicit toxicity via protein gain-of-function and which act through RNA repeat-elicited gain-of-function (5,6). This non-canonical translational initiation process enables elongation through a repeat strand in the absence of an AUG initiation codon and in multiple reading frames, producing multiple homopolymeric or dipeptide repeat-containing proteins. Originally described in association with CAG-repeat expansions causative for spinocerebellar ataxia type 8 (SCA8), this process also occurs in association with expansions of CAG, CUG, GGGGCC, GGCCCC, CGG, CCUG, and CAGG repeats (7-17). Repeats can drive RAN translation in a surprising variety of RNA contexts, including within 5' untranslated regions (UTRs), protein-coding ORFs, or introns and "non-coding" RNAs. The identification of this novel translational initiation event has led to a flurry of activity within the research community, with a significant body of work now demonstrating 1) the presence of RAN-translated peptides across a wide spectrum of neurodegenerative disorders and 2) an association between these short repeat-containing proteins and neuronal toxicity. Despite this interest, the mechanism or mechanisms by which RAN translation occurs remains largely unknown. The focus of this review will be what we know thus far about RAN translation initiation, with particular attention paid to how RAN translation compares to canonical translation and other forms of initiation described over the past 40 years. We hope that revisiting these foundational

experiments will shed light on this new disease-relevant process.

1.3 Translation Initiation

Translation initiation is the stepwise 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) (18). A comprehensive account of the roles of each eIF and each stage of initiation is beyond the scope of this review, but several steps are worth highlighting. In most cases, initiation begins with the recognition of the 5′ methyl-7-guanosine (m^7G) cap on mRNA by the eIF4F complex (**Fig.1-1, Step 1**) (19-21). The eIF4F complex is composed of eIF4E (the direct cap-binding subunit), eIF4G (a scaffolding subunit), eIF4A (a DEAD-box RNA helicase), with eIF4B and eIF4H serving as additional helicase stimulatory factors. eIF4G recognizes the poly-adenosine-binding protein (PABP) (22,23), 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 (24,25).

The eIF4F complex, still bound to the m^7G cap, is joined by the 43S pre-initiation complex (PIC), composed of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, and the ternary complex [composed of methionine-conjugated tRNA ($tRNA^{Met}$), and eIF2-GTP; **Fig.1-1, Step 2**]. This joining of the 43S PIC to the eIF4F complex is mediated by an eIF4G-eIF3 interaction (26). 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 (27). In the ribosomal scanning model of translation initiation (28), the 43S PIC and components

of the eIF4F complex scan through the 5′ UTR in the 5′ to 3′ direction (**Fig. 1-1, Step 3**). This stage is also known to require eIF4A in order to resolve weaker internal secondary structures, though additional helicases assist in melting stronger structures (29-33). The 43S PIC scans until encountering an AUG codon in a good Kozak context, (A/G)NNAUGG (**Fig. 1-1, Step 4**) (34,35). At this point, base-pairing between the AUG codon and CAU anti-codon loop on $tRNA^{Met}$ results in the ejection of eIF1, a factor which, along with eIF1A, increases the stringency of AUG start-codon selection (27,36-39). eIF2 hydrolyzes its bound GTP with the assistance of eIF5, the associated GTPase-activating protein (GAP; **Fig.1-1, Step 5**). 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 (**Fig. 1-1, Step 6**), the majority of remaining eIFs are ejected, eIF5B hydrolyzes its bound GTP (**Fig. 1-1, Step 7**) and translation elongation begins with formation of the first peptide bond (40).

1.4 The role of secondary structure in translational initiation

RNA secondary and tertiary structures contribute significantly to the dynamics and regulation of translation initiation. Structures in the 5′ UTR impact translation initiation both positively and negatively, depending on the structure's location. When placed upstream of an AUG start codon, highly structured regions are known to 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 (41-44). In contrast, secondary structures downstream of start codons facilitate initiation at

imperfect start codons with poor Kozak context and even non-AUG codons (44,45). Initiation at non-AUG codons occurs at reduced efficiency relative to AUG codons in *in vitro* translation systems (46,47), but the presence of secondary structures downstream markedly increases initiation efficiency (44,45,48).

Recent advances in ribosomal footprinting methodologies suggest that these *in vitro* findings may reflect common but heretofore unrecognized 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 level (49). This technique has found evidence for thousands of unpredicted translation initiation events, many of which occur at non-AUG codons (50-52). This is especially true for upstream open reading frames (uORFs), which are short ORFs upstream of canonical, annotated ORFs in the same mRNA transcript. Many of these uORFs appear to play regulatory roles in translation that are dependent on metabolic conditions and cell-cycle stage (53). These findings are now supported by a variety of studies utilizing mass spectroscopy to confirm the presence of these uORF-coded peptides, many of which possess functional roles (54-60). Thus, initiation at non-AUG codons occurs *in vivo*, may be regulated in part by *trans*-acting eIFs as well as mRNA-specific *cis* factors, and appears to play important regulatory roles in global protein translation.

While the majority of eukaryotic mRNAs are likely translated via the canonical mechanism described above, multiple atypical mechanisms exist, and like canonical initiation, atypical modes of initiation are modulated by mRNA secondary structure. Multiple viral and cellular RNAs are translated via an internal ribosome entry site (IRES)-

mediated pathway (**Fig. 1-2, B**) (61). IRESes are complex RNA structures that directly recruit ribosomal subunits and eIFs to internal sites within RNA transcripts. These are highly heterogeneous structurally and functionally, but the common, central feature is that translation initiation bypasses the eIF4E-m^7G interaction. Therefore, IRES-based translation is said to be m^7G cap-independent, allowing for initiation within bicistronic or circular RNA elements. In addition, IRESs are heterogeneous in which eIFs are and aren't recruited or required, with some forms requiring most eIFs while others require only the 40S ribosomal subunit and an alanine-conjugated tRNA (**Fig. 1-2, C**) (62-64). The latter mechanism, employed by the cricket paralysis virus (CrPV), intriguingly utilizes CCU for an initiation codon and codes for a protein with an N-terminal alanine. Though viral IRESs are the most thoroughly characterized, multiple eukaryotic mRNAs are also thought to contain IRESs, including c-Myc, p53, Bcl-2, and others (65). Translation of these proteins is maintained under various stress conditions in which canonical translation is inhibited, implying a unique mechanism that escapes this general inhibition.

In a second example of non-canonical modes of initiation, *histone 4* mRNA appears to be translated through a ribosomal tethering mechanism (**Fig. 1-2, D**) (66,67). The 5′ UTR of *histone 4* is shorter than most 5′ UTRs (the mouse homolog is 9 nucleotides, at the short extreme). It is efficiently translated, however, despite translation initiation generally requiring a 5′ UTR of at least 20 nucleotides (68,69). Translation of histone 4 mRNA begins with the recruitment of the eIF4F complex to a structural element within the ORF. eIF4F then binds to the m^7G cap, which is buried within a nearby structural element. The 43S PIC is subsequently recruited to eIF4F, and is then transferred directly to the AUG start codon (70). Thus, translation of *histone 4* is cap-

dependent, but deviates from canonical translation in that the ribosome is initially recruited internally. And, as with IRES-mediated translation, translation of *histone 4* is mediated through interactions between eIFs and mRNA structural elements.

Ribosomal shunting represents yet another alternative mode of translation initiation (**Fig. 1-2, E**). As with IRES-mediated translation, ribosomal shunting was originally described in association with viral RNAs (71-73). In ribosomal shunting, the eIF4F complex and 43S PIC still bind to the m^7G cap, but the 43S PIC is able to "jump" past sections of the 5′ UTR, rather than translocating linearly and performing a base-by-base inspection. Generally, this is thought to require a direct interaction between an mRNA structural element and the 18S rRNA (66). As with IRESes, certain eukaryotic mRNAs are also thought to be translated through a shunting-related mechanism, including HSP70 (72), BACE1 (74), and cIAP2 (75).

1.5 RAN translation: an unexpected finding at nucleotide repeats

In each of the above examples, 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 (7). Expansions of protein-coding CAG repeats in the gene Ataxin 8 (*ATXN8*) lead to the neurodegenerative disorder SCA8. Unexpectedly, mutation of the only AUG codon upstream of expanded CAG repeats did not abrogate protein translation (7). Nevertheless, translation initiated in multiple reading frames, generating homopolymeric proteins with glutamine, serine, or alanine repeats, depending on the reading frame. RAN initiation on ATXN8 transcripts depended on the stability of secondary structures formed

from the expanded CAG repeats, as decreasing the number of CAG repeats or their GC content reduced RAN translation (7). RAN translation products from all three reading frames accumulated in cells transfected with expanded CAG reporters, occasionally even within the same transfected cell. Antisense ATXN8 transcripts bearing expanded CUG repeats also supported RAN translation (7). Antibodies generated against the predicted polyalanine product of the ATXN8 sense transcript recognized a protein in the cerebellums of SCA8 human patients and mouse models. A similar approach provided *in vivo* evidence of a polyglutamine RAN product from DMPK antisense transcripts bearing expanded CAG repeats, associated with myotonic dystrophy type 1 (DM1) (7).

Since this initial observation, several groups have demonstrated that RAN translation occurs at a wide variety of different repeat expansions (7-17). This chapter and dissertation focuses on RAN translation in two distinct repeat-expansion disorders that occur in different sequence contexts: CGG repeats in the 5′ UTR of the fragile X mental retardation 1 (*FMR1*) gene, as occurs in fragile X-associated tremor/ataxia syndrome (FXTAS), and intronic GGGGCC repeats in *C9orf72*-associated amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Each example offers unique insights into how RAN translation occurs, but also presents unique challenges in our effort to understand this process mechanistically.

1.6 RAN translation of CGG repeats in FXTAS

FXTAS is a late-onset neurodegenerative disorder caused by the expansion of CGG repeats 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 (76-78). Premutation repeat expansions result in enhanced transcription of FMR1 mRNA bearing these repeats (79,80). In contrast, expansions to greater than 200 repeats trigger transcriptional silencing of the *FMR1* locus, leading to loss of FMR1 mRNA and the Fragile X protein, FMRP. Transcriptional silencing manifests as Fragile X syndrome, a clinically distinct neurodevelopmental disorder characterized by intellectual disability and autistic features (81,82). Approximately 40% of male premutation carriers develop FXTAS (approximately 1:3000 of the total population), with increased penetrance at older ages and larger repeat sizes (83-85). 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 (86,87). Premutation carrier women are also at increased risk of premature ovarian insufficiency (FXPOI) (88,89).

Our lab demonstrated that the CGG-expanded *FMR1* 5′ UTR supports RAN translation initiation (12). 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). FMRpolyG accumulates in ubiquitinated inclusions in patient tissue and cellular and animal disease models, is necessary to elicit toxicity in *Drosophila* models of disease, and induces proteasome perturbations in *Drosophila* and HeLa cells (12,90,91). In an inducible mouse model of FXTAS that expresses the *FMR1* 5′ UTR with 90 CGG repeats, turning off transgene expression reverses the formation of neuronal FMRpolyG-positive inclusions and repeat-elicited behavioral deficits (92). Finally, FMRpolyG-positive inclusions have been found in ovarian stromal cells in FXTAS mouse models and a FXPOI

patient, suggesting FMRpolyG expression is linked to other Fragile X-related clinical phenotypes (17).

In an effort to investigate RAN translation of CGG-expanded *FMR1* mechanistically, our lab developed several transfectable reporters for expression of FMRpolyG and FMRpolyA. First, we observed production of an FMRpolyG-green fluorescent protein (GFP) fusion reporter construct bearing 30, 50, or 88 CGG repeats, suggesting that RAN translation can occur in the absence of pathological expansions (12). In contrast, FMRpolyA was expressed from reporter constructs bearing 88 repeats but not 30 repeats (12). Second, insertion of a stop codon immediately upstream of the CGG repeats precluded expression of FMRpolyG reporters, indicating that RAN translation of FMRpolyG initiates upstream of the repeats (12). Further mutational analysis revealed that initiation in this frame can occur at multiple upstream near-AUG codons in the human 5′ UTR (12). In contrast, insertion of a stop codon did not prevent expression of FMRpolyA (12). This suggests that RAN translation can initiate in the GCG frame within the repeats. These results raise the intriguing possibility that RAN translation of the same sequence can differ mechanistically in different reading frames.

One important question is whether RAN translation of CGG-expanded *FMR1* is cap-dependent or utilizes an IRES-like cap-independent mechanism (**Fig. 1-2, A-C**). There is some evidence to suggest that the *FMR1* 5′ UTR possesses IRES activity. Insertion of the *FMR1* 5′ UTR between two ORFs in a plasmid-based bicistronic reporter doesn't eliminate translation of the second ORF (93,94). Although some expression of the second ORF may be due to the presence of a cryptic promoter element within the *FMR1* 5′ UTR, this same finding was observed when *in vitro* transcribed bicistronic RNA

was transfected into cells (94). Further experiments indicated that translation from monocistronic reporters with the 5′ UTR of *FMR1* were less cap-dependent than reporters bearing the 5′ UTR of β-globin, as would be predicted if an IRES was utilized (94). This putative IRES activity was partially dependent on the CGG repeat and required both the approximately 100 nucleotides upstream of the repeats in *FMR1* 5′ UTR, as well as the region containing the repeat itself. However, these studies were done before a conceptualization of RAN translation was in place, and it is unclear whether these initiation events occurred at the AUG of the reporter ORF or within the FMR1 5′ UTR itself. Also, these studies were performed with a very short repeat (9 CGGs). Follow-up work suggested that a significant fraction of FMRP is translated through a cap-dependent process, highlighting the significance of the eIF4E-m^7G interaction both in cells and *in vitro* translation systems (95,96).

More recently, Kearse *et al*. (2016) demonstrated that RAN translation of CGG-expanded *FMR1* reporters is cap-dependent (97). Either when transfected into HeLa cells or incubated in an *in vitro* translation system, *in vitro* transcribed, m^7G-capped RAN-translation reporter RNAs are efficiently translated (97). In contrast, when m^7G is substituted with A-cap, which is not recognized by eIF4E, expression of these RAN reporters decreased (97). Furthermore, addition of excess free m^7G cap to sequester eIF4E, also reduced the expression of RAN translation reporters. In contrast, neither of these manipulations affected expression of an IRES-bearing control reporter (97). These experiments suggest that RAN translation of *FMR1* is a cap-dependent process, resembling the first stage of canonical translation.

In the next stage of canonical initiation, the 43S PIC scans through the 5′ UTR,

performing a base-by-base inspection for an AUG codon (**Fig. 1-1, Step 3**). If the scanning model were to hold for RAN translation of CGG-expanded *FMR1*, then the 43S PIC would need to scan through the CGG repeats. *In silico* modeling predicts and *in vitro* analysis suggests that consecutive CGG repeats form a stable hairpin structure (98,99), presenting a significant impedance to scanning 43S PICs. However, per the ribosomal shunting mechanisms (**Fig. 1-2, E**), it is possible that 43S PICs could bypass this hairpin, as in this model, 43S PICs do not scan linearly through the 5′ UTR. Increasing the length of CGG repeats above 30 repeats reduces the expression of a downstream reporter (95,96), suggesting that scanning is in fact linear (or, potentially, larger CGG repeats interfere with an alternative mechanism). In addition, Ludwig *et al*. (2011) observed initiation at a near-AUG codon in an artificial hairpin inserted in the 5′ UTR in place of the CGG repeats, which suggests that this section of the UTR is scanned rather than "skipped," as could be the case in a tethering-like mechanism (96). More directly, to determine whether the scanning model holds true for RAN translation of CGG-expanded *FMR1*, Kearse *et al.* (2016) treated cells with hippuristanol, an inhibitor of eIF4A (the RNA helicase that promotes 43S PIC scanning) (97,100). When administered in cells or supplemented in an *in vitro* translation system, hippuristanol reduced expression of RAN translation reporters, but had no effect on IRES-mediated translation (97). Again, this would suggest that the initial stages of RAN translation resemble canonical initiation. Though these experiments suggest that mechanisms of non-linear scanning are unlikely, further, more rigorous experimentation is required to show whether in RAN translation each codon is inspected, likely through the introduction of AUG codons throughout the 5′ UTR, not just near the hairpin.

If the scanning model of translation applies to *FMR1*, the 43S PIC will encounter a stable hairpin formed by the CGG repeats before encountering the AUG codon of FMRP. In their initial characterization of RAN translation of CAG repeats in *SCA8*, Zu *et al*. (2011) demonstrated that RAN translation in transfected cells is dependent on the GC content and length of these repeats—factors that would increase the stability of secondary structure (7). Together, these lines of evidence suggest that RAN translation initiation entails elements of secondary structure within the mRNA. In this case, Kozak's early experiments might help explain why: downstream secondary structure can enhance initiation at upstream non-AUG codons (44,45,48). Specifically, the increase in non-AUG codon usage is maximal when a hairpin falls 14 nucleotides downstream of the AUG 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 $tRNA^{Met}$ (45). These findings have led to the hypothesis that secondary structure causes scanning 43S PICs to stall, increasing initiation at optimally positioned non-AUG codons. RAN translation on CGG-expanded FMR1 mRNA to generate FMRpolyG may utilize a similar mechanism (12). However, further work is needed to test this model and its application both to other *FMR1* reading frames as well as other repeats and sequence contexts.

If downstream mRNA secondary structures are necessary for RAN translation initiation, then it is critical to define exactly how such structures promote initiation at non-AUG codons. For example, stalling of scanning ribosomes is predicted to lead to both the congestion of 43S PICs on mRNAs upstream of the repeat and an increase in the dwell time of the 40S subunit over an imperfect codon-anticodon match (44,45,48). Both of these events could presumably enhance the rate of enzymatic catalysis and 48S complex

formation at multiple upstream non-AUG codons. Alternatively, stalling could favor the dissociation of key eIFs that help determine AUG start codon fidelity (eIF1 and eIF1A) or even alternative ribosomal conformations as may occur with IRES-mediated translation (101,102).

In consideration of the existing literature, our current working model of RAN translation at CGG-expanded FMR1 is as follows (**Fig. 1-3, A**): the eIF4F complex and 43S PIC bind to the m^7G cap on FMR1 mRNA. This complex then scans downstream through the 5′ UTR until encountering secondary structure formed either by CGG repeats or the surrounding, intrinsic sequence of the 5′ UTR. Ribosomal stalling results in aberrant translation initiation at non-AUG codons either upstream of or within the repeat in the +1 and +2 frames, resulting in the production of FMRpolyG and FMRpolyA. Whether initiation also occurs in the CGG (+0, polyarginine) frame remains unknown, although there is no evidence for an N-terminal polyarginine extension on FMRP or incorporation of FMRP into ubiquitinated inclusions in FXTAS patients (103). However, as discussed below, there is good reason to suspect that the characteristics of RAN translation of CGG-expanded FMR1 may not be completely applicable to RAN translation of other repeats (capable of forming different secondary structures) or other sequence contexts.

1.7 RAN translation at GGGGCC and GGCCCC repeats in C9 ALS/FTD

The *C9orf72* GGGGCC/GGCCCC hexanucleotide repeat expansions was identified by two groups in 2011 as the most common known cause of ALS and FTD (104,105). ALS is the most frequently occurring form of motor neuron disease, affecting approximately 2-4/100,000 individuals (106), and is characterized by progressive

paralysis typically leading to death within two to three years of 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 (107,108). FTD presents heterogeneously and is divided into three clinical syndromes; behavioral variant, semantic dementia, and progressive nonfluent aphasia (109). The *C9orf72* hexanucleotide repeat expansions is most frequently associated with the behavioral variant (104,105,110), characterized by changes in personality and conduct. 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 (111,112); while up to 50% of FTD patients develop motor dysfunction (113). 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 variant of FTD (FTLD-TDP) (114).

In C9ALS/FTD, the GGGGCC repeat, 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 upwards of more than a thousand repeats in C9ALS/FTD patients (104,105,110). Both the sense and antisense strands of *C9orf72* are transcribed in mutation carriers, resulting in the production of GGGGCC and GGCCCC-repeat containing RNAs (9,11,13). Both expanded repeat sequences are predicted to form highly stable RNA secondary structures, with the sense RNA repeat generating a G-quadruplex and hairpin *in vitro* (115-118) and the antisense RNA repeat shown to assume an A-form-like double helix *in vitro* (119).

1.8 DPR pathology in C9 ALS/FTD patients

In addition to TDP-43-positive inclusions within both neurons and glia (114), neuronal TDP-43-negative inclusions that co-stain for ubiquitin and ubiquitin-binding proteins are uniquely found throughout the CNS of C9-associated ALS and FTD patients (120,121). Immunohistochemical (IHC) analysis by multiple laboratories indicates that RAN-translation-derived proteins constitute these TDP-43-negative inclusions (8-11,13). A total of six different dipeptide repeat proteins (DPRs) are generated from the GGGGCC and GGCCCC transcripts (**Fig. 1-3, B & D**). 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 (8-11,13).

DPRs form both neuronal cytoplasmic and intranuclear inclusions (NCIs and NIIs) throughout the CNS. However, the distribution of DPRs throughout the brain is highly variable, with the highest burden occurring in the hippocampus, cerebellum, neocortex, and thalamus (8,122,123). Additionally, 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 antibodies for each DPR, polyGA appears to be most abundant, followed by polyGP and polyGR, while the DPRs derived exclusively from antisense transcripts (polyPA and polyPR) appear to be least abundant (10,11,124).

Despite the different CNS regions that exhibit marked neurodegeneration in ALS and FTD—the motor cortex and spinal cord in ALS, and the frontal and temporal lobes in FTD—quantitative IHC studies show that DPR abundance within the frontal cortex and

lower motor neurons is not significantly different between C9 patients with pure ALS or FTD (125). Several additional studies have similarly shown that DPR burden is not well-correlated with degeneration (123,124,126). This is in contrast to TDP-43-positive inclusions, which are most abundant in the most severely affected brain regions (124-126). This lack of correlation may suggest that RAN translation is not the driving force in disease pathogenesis. Alternatively, it may be that DPR inclusion formation is neuroprotective while the soluble DPRs oligomers drive toxicity, as has been proposed in several other neurodegenerative proteinopathies (127,128). Alternatively, Edbauer and Haas (2015) propose an “amyloid-like” mechanism of toxicity for the DPRs, in which accumulation of DPRs initiates a cascade of events that leads to TDP-43 mis-localization and aggregation in selectively vulnerable neurons (129). Work is still needed to distinguish between these possibilities, perhaps using the AAV $GGGGCC_{66}$ mouse model in which neuronal loss and TDP-43 pathology is detectable (130).

1.9 Toxicity of DPRs in model systems

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 (131), to *Drosophila* (132-136), cultured cells (13,122,137,138) and primary mammalian neurons (122,133,139), DPR expression leads to cell death and/or reduced survival. Significantly, in many of these systems, DPR expression is sufficient to trigger toxicity, as demonstrated by the use of alternative codons in place of GGGGCC that allow for DPR production in the absence of the potentially toxic repeat-containing RNA species (122,131-134,137-139). Furthermore,

transgenic flies expressing various length GGGGCC repeats with stop codons in all three reading frames form RNA foci, but only flies containing pure repeats produce polyGR and polyGP proteins and undergo significant cell death (134).

Multiple studies suggest that arginine-containing DPRs are the most toxic RAN species. Both polyGR and polyPR form intranuclear aggregates that disrupt nucleoli when overexpressed in model systems (133,137,138,140). However, nucleolar DPRs are not detected in patient brain tissue (123,125), and when co-expressed with polyGA, polyGR proteins are recruited into cytoplasmic polyGA inclusions (132), suggesting that nucleolar stress may not be a significant driver of toxicity in patients. Alternatively, Wen *et al.* (2014) identified nucleolar polyPR-positive inclusions in spinal cord tissue from a C9ALS patient, and suggest that the high toxicity of nucleolar polyPR results in increased vulnerability of neurons containing these species (133). PolyGR proteins can also mediate toxicity through impairment of the Notch pathway (132), and polyPR and polyGR inhibit nucleocytoplasmic transport in flies and yeast (131,135). Importantly, GGGGCC repeats directly interact with RanGAP, a regulator of nucleocytoplasmic transport (141), suggesting that the arginine-containing DPRs and the repeat-containing RNA both contribute to this mode of toxicity.

Although comparatively less toxic than polyGR or polyPR when each is expressed in isolation (131-135,138), non-arginine-containing DPR proteins also appear important to neurodegeneration in model systems. Adult flies expressing exclusively GA-100 DPRs within neurons have significantly reduced survival (134), and expression of polyGA in primary mammalian neurons causes increased toxicity through impairment of the

ubiquitin-proteasome system (122), induction of ER stress (122), and sequestration of Unc119, a trafficking protein with a GAGASA binding motif (139).

1.10 Mechanism of RAN translation at GGGGCC/GGCCCC repeats

Despite compelling evidence that RAN translation and the resulting DPRs are involved in disease pathogenesis, little is known about the mechanism by which the expanded GGGGCC and GGCCCC repeats trigger DPR production. When placed in a 5′ leader context, RAN translation at GGGGCC/GGCCCC repeats is repeat-length dependent, with more robust DPR production occurring with longer repeats (11,13,118), consistent with observations at CAG and CGG repeats (7,12). The repeat-length requirement for initiation also appears to be different for different reading frames and in different sequence contexts. When GGGGCC and GGCCCC repeats are placed downstream of synthetic sequence, all DPRs are detected by immunocytochemistry (ICC) with as few as 30 or 40 repeats, respectively (13). When GGGGCC repeats are instead placed downstream of 113 nucleotides from intron1, in a partially native context (**Fig. 1-3, B**), production of polyGA occurs similarly with as little as 38 repeats (11). However, within this sequence context, polyGP detection required 66 repeats in one report and 145 in another, while polyGR was not detected within cells expressing constructs containing up to 145 repeats (11,118). Interestingly, the apparent order of the minimum repeat-length-requirement for RAN in each sense reading frame, from constructs containing the GGGGCC repeat downstream of native sequence, correlates with sense DPR abundance in patient tissue (125). Antisense DPRs also showed different length requirements for detection when placed downstream of native sequence; while polyGP and polyPR are

detected in cells expressing 66 GGCCCC repeats downstream of 99 native nucleotides, polyPA is not (9). While these apparent differences in length requirements may reflect artifacts of detection based on antibody avidity or differences in DPR solubility, it 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, with an increase in promiscuity or frameshifting at larger repeat sizes becomes more prominent.

Beyond these initial insights, however, are a series of unanswered questions related to the mechanism of RAN translation at GGGGCC and GGCCCC repeats. First, it remains unknown exactly what RNA species undergo RAN translation in C9 repeat expansion patients. The GGGGCC repeat expansion is located within the first intron of *C9orf72*. Therefore, in patients, RAN-translated GGGGCC repeats could conceivably derive from a retained intron, a spliced intron in a lariat, or within aberrant disease-specific transcripts generated by transcriptional stalling (**Fig. 1-3, B**). There is some evidence for generation of such aberrant transcripts, at least *in vitro* (117). The ratio of exon1a-intron1 (unspliced or abortive) RNA to exon1-exon2 (mature, spliced) RNA (**Fig. 1-3, B**), however, is not altered in C9 iPSC-derived neurons and patient brain tissue relative to controls, arguing against significant production of truncated transcripts or increased intron retention (136). However, a recent study suggests that intron retention occurs with some frequency in both control and C9 patient cells (142), 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 *Drosophila,* placement of the repeat into an efficiently spliced intron dramatically reduces both RAN translation and its relative toxicity compared to repeats placed into a 5′ leader sequence, suggesting that spliced lariats containing GGGGCC repeats may not be efficiently utilized to produce DPRs (136). However, DPR production from the intronic repeat becomes sufficient to elicit toxicity when *Drosophila* are grown at elevated temperatures, indicating that an intronic context is able to support pathological RAN translation under certain conditions (136). Whether the limited amount of DPRs observed was produced from a spliced or retained intronic repeat is unclear. However, if an intron lariat is the transcript subtype utilized, then some mechanism must exist for it to bypass normal degradation mechanisms, exit to the cytoplasm, and become engaged with translational machinery.

Each of these target transcript possibilities has significant implications for what translational initiation factors would be required and what translational mode would be preferentially utilized. For instance, if an intronic lariat RNA is the substrate of GGGGCC/GGCCCC RAN translation, then this almost by definition rules out a role for cap-dependent, canonical processes and strongly favors mechanisms more in line with internal ribosome entry. Similarly, where initiation occurs in each reading frame and what *trans* factors are required will likely be highly dependent on the RNA species being studied. Therefore, more direct studies are needed to address these questions. Doing so will help formulate a clearer picture of how RAN translation occurs and will likely provide new potential targets for therapeutically inhibiting this pathological process.

1.11 Summary

RAN translation represents a new and provocative mechanism by which protein translation can occur in the setting of nucleotide repeat expansions to produce a novel set of toxic proteins. This chapter has tried to take the limited mechanistic data generated prior to this dissertation work and place it into the context of known canonical and non-canonical translation initiation processes. These different modes of translation provide a framework for which questions are of greatest importance. By determining the cap-dependency, the requirement for linear and continuous 5′-to-3′ scanning, and the N-terminal amino acid used during RAN translation, we will be able to take advantage of previous work on processes with similar biology. Such an approach can also narrow down and prioritize which of a myriad of potential *trans* factors should be studied and guide strategies for interventions that might selectively preclude RAN translation.

One important question going forward is whether the mechanisms underlying RAN translation are the same or different across repeat types, reading frames and sequence contexts. Some discrepancies suggest that different mechanisms may be in play. For example, data on RAN translation at CGG repeats thus far is most consistent with a scanning mechanism and use of a near-AUG codon for initiation just 5′ to the repeat (12), but this would seem to be quite unlikely as a mechanism to explain initiation within an open reading frame, as occurs in Huntington disease (14). In this fashion RAN translation may be analogous to the situation with viral IRES RNA elements, which display a significant variance in both sequence and initiation factor requirements to achieve the same goal of bypassing cap-dependent ribosomal loading. We may need to begin thinking of these processes as *RAN Translations* rather than as a single entity. However,

only after careful identification of the key factors required for RAN translation can this delineation really be made across different repeat and disease contexts.

Upon these same lines, it is important to recognize that aspects of what are currently observed in RAN translation may be part of a larger recognition that translation initiation in the absence of an AUG start codon may not be as uncommon as once thought. Data from ribosome profiling datasets suggest significant non- and near-AUG initiated translation throughout the transcriptome (50). Thus, RAN translation may reflect aberrancy of normal non-canonical initiation processes that produces toxic proteins but which otherwise has valuable 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.

Lastly, it is worth noting that 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 this process proves important to neurodegeneration, as current data would support, then identification of factors that are selectively critical for RAN translation, but not canonical transition, offer a real opportunity going forward.

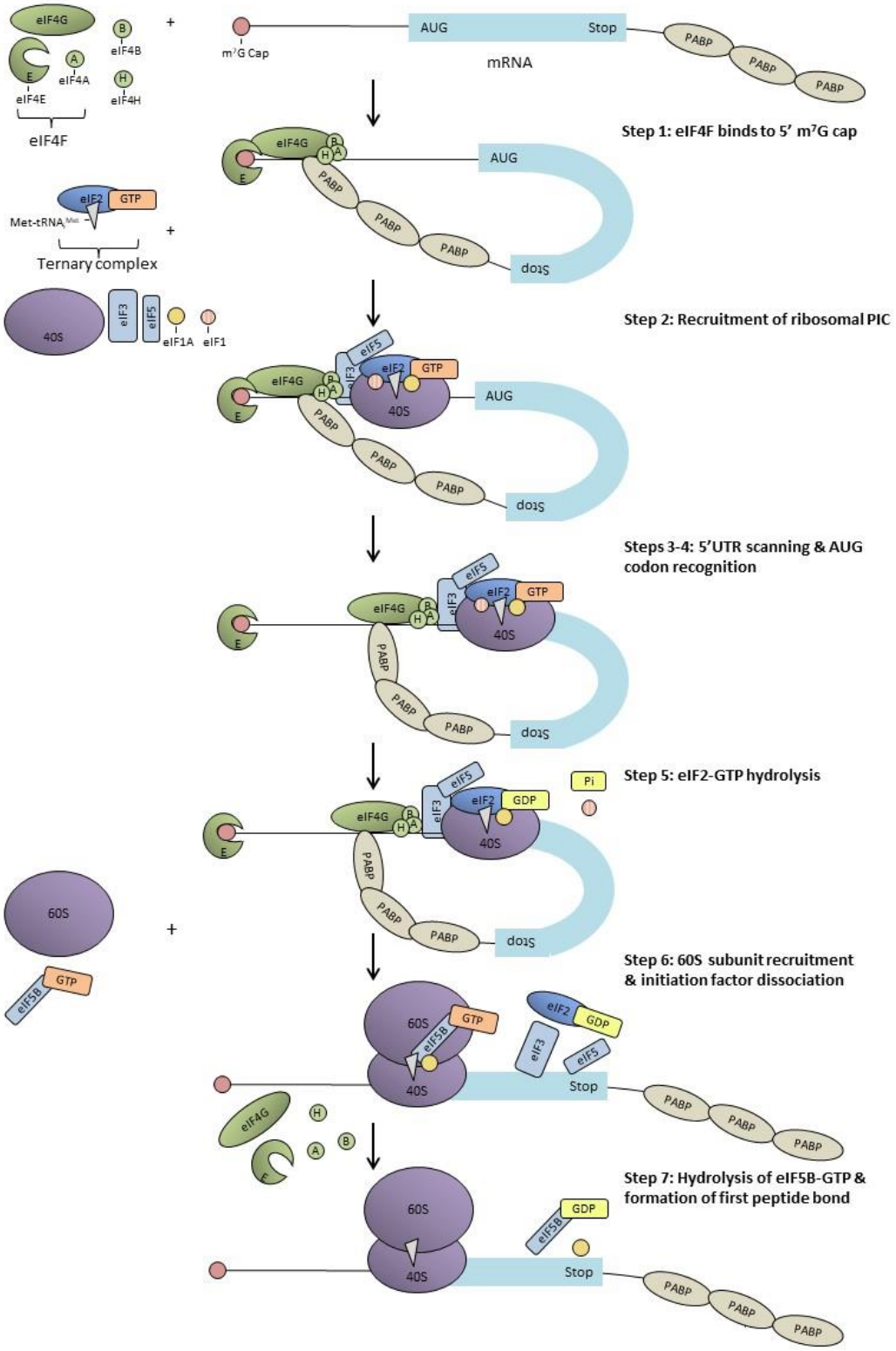
To this end, the following chapters of this dissertation focus on understanding the mechanism of C9RAN translation at GGGGCC repeats. As such, advances to our understanding of C9RAN translation over the last five will be discussed throughout. These advances, including those made through this dissertation work and those made within the broader field, will then be summarized in detail within the final discussion chapter.

1.12 Chapter-specific acknowledgements

We thank members of the Todd lab for fruitful suggestions and discussions. KMG and AEL were supported by NIH T-32-GM007315. PKT was supported by VAMC BLRD 1I21BX001841 and 1I01BX001689 and NIH R01NS086810.

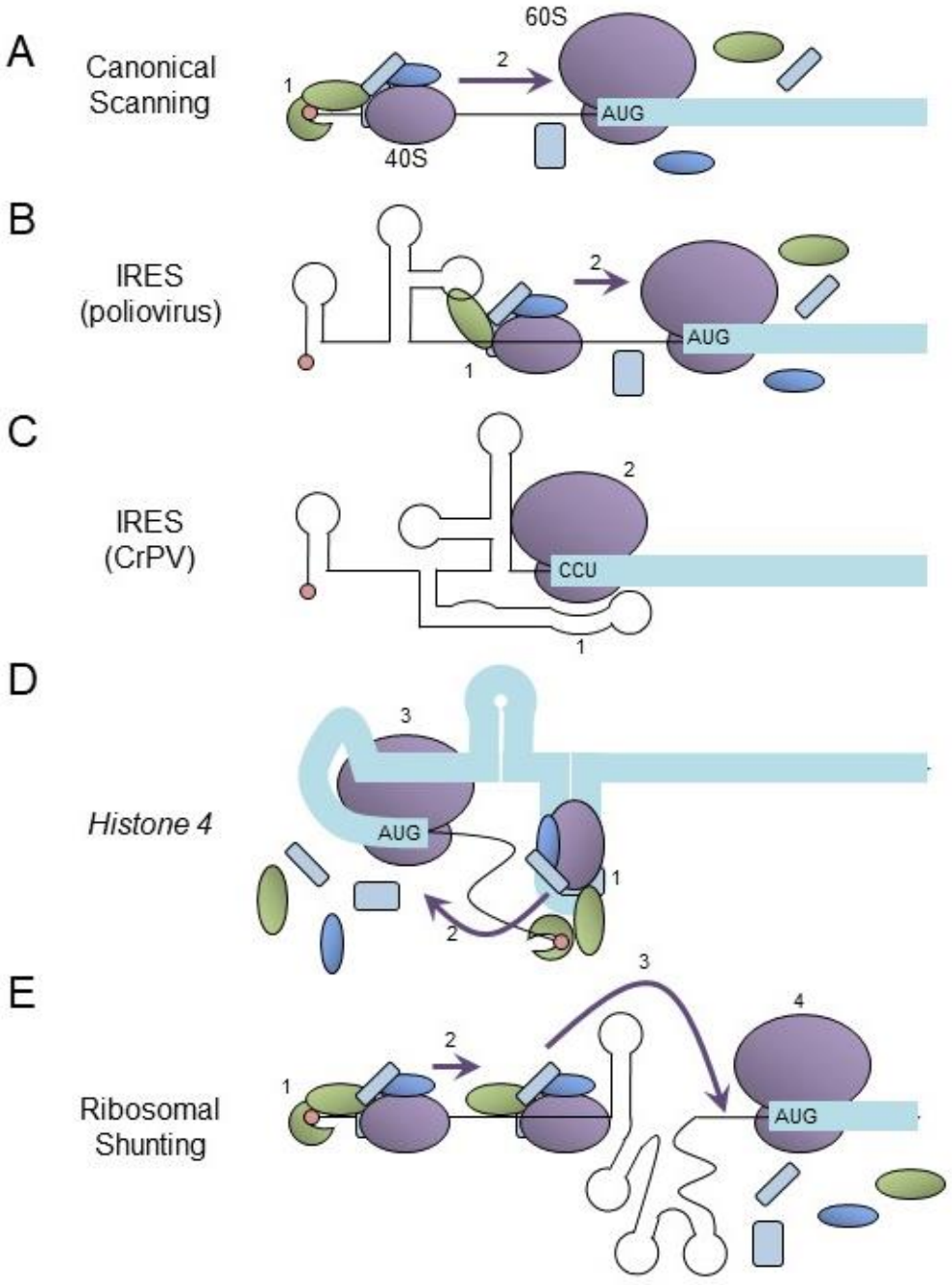
1.13 Figures

Figure 1-1. Canonical scanning model of translation initiation



Step 1 - The eIF4F complex, composed of eIF4E, eIF4G, eIF4A, binds to the 5′ m^7G cap with eIF4B and/or eIF4H. PABP associates with eIF4G to circularize the mRNA. *Step 2* - The eIF4F complex recruits the 43S PIC, composed of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, and the ternary complex, consisting of the initiator methionine-tRNA, eIF2, and GTP. *Step 3-4* - The PIC and components of the eIF4F complex scan through the 5′ UTR in the 5′ to 3′ direction until encountering an AUG start codon in a good Kozak context. *Step 5* – eIF1 dissociates from the PIC, and eIF2 hydrolyzes GTP with the assistance of eIF5, committing the 40S ribosome to translation initiation at the present AUG codon. *Step 6* - eIF5B-GTP promotes association of the 60S subunit and displacement of most eIFs. *Step 7* – eIF5B hydrolyzes its bound GTP and dissociates with eIF1A, establishing the elongation-competent 80S ribosome. eIF = eukaryotic initiation factor, m^7G = 7 methylguanosine, PABP = poly-adenosine binding protein, PIC = preinitiation complex.

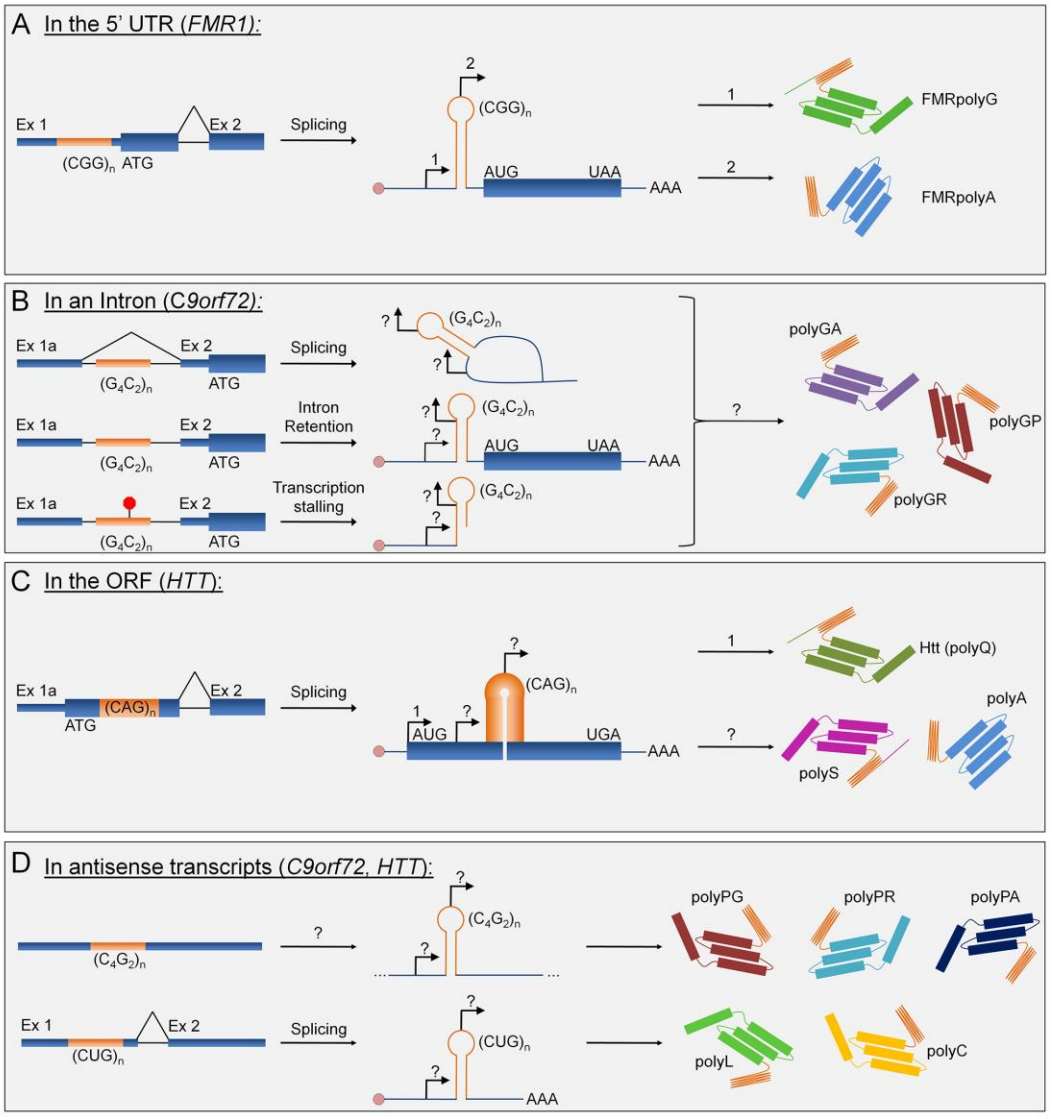
Figure 1-2. Canonical and alternative mechanisms of translation initiation



A) In canonical translation initiation, the 43S PIC with a full complement of eIFs binds to the 5′ m^7G cap (Step 1), then linearly scans through the 5′ UTR until reaching an AUG start codon in good Kozak context (Step 2). B) In contrast, the poliovirus IRES uses

complex RNA secondary structure to recruit nearly all eIFs to an internal site within the transcript, bypassing the eIF4E-m^7G interaction for recruiting the 43S PIC (Step 1), which then scans 5′ to 3′ until encountering the AUG start codon (Step 2). C) In contrast, the CrPV IRES requires and recruits only the 40S and 60S ribosomal subunits and an alanine-conjugated tRNA (Step 1), initiating translation at a CCU codon in the absence of any eIFs (Step 2). D) Translation of *Histone 4* mRNA begins with tethering of eIF4F (Step 1) and the 43S PIC (Step 2) to internal secondary structures. The 40S subunit is then transferred to the AUG start codon upstream (Step 3). E) In ribosomal shunting, eIF4F and the 43S PIC are recruited to the 5′ m^7G cap (Step 1), and scan (Step 2). However, secondary structure within the 5′ UTR binds to the 18S rRNA and facilitates jumping across elements that would otherwise preclude efficient scanning (Step 3). The 43S PIC then scans until reaching a start codon (Step 4).

Figure 1-3. Production of RAN proteins across repeats



A) When located in the 5′ UTR, as in *FMR1*, expanded GC-rich repeats trigger initiation of RAN translation upstream of the canonical AUG start codon, leading to the production of FMRpolyG and FMRpolyA. 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 relevant RNA species produces polyGA, polyGR, and polyGP RAN-translation products. C) When located within an ORF, as in *HTT*, canonical

translation still initiates at the AUG codon upstream of the repeats. However, expression of polyserine (polyS) and polyalanine (polyA) proteins occurs by a combination of RAN-translation through the repeats and frameshifting out of the native (polyglutamine; polyQ) frame. D) Repeats located in antisense transcripts, as in *C9orf72* and *HTT*, are also substrates for RAN translation, further expanding the number of dipeptide or homopolymeric RAN proteins.

1.14 References

1. Mason, A. R., Ziemann, A., and Finkbeiner, S. (2014) Targeting the low-hanging fruit of neurodegeneration. *Neurology* **83**, 1470-1473
2. Orr, H. T., and Zoghbi, H. Y. (2007) Trinucleotide repeat disorders. *Annu Rev Neurosci* **30**, 575-621
3. He, F., and Todd, P. K. (2011) Epigenetics in nucleotide repeat expansion disorders. *Semin Neurol* **31**, 470-483
4. Mohan, A., Goodwin, M., and Swanson, M. S. (2014) RNA-protein interactions in unstable microsatellite diseases. *Brain Res* **1584**, 3-14
5. Cleary, J. D., and Ranum, L. P. (2014) Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Curr Opin Genet Dev* **26**, 6-15
6. Kearse, M. G., and Todd, P. K. (2014) Repeat-associated non-AUG translation and its impact in neurodegenerative disease. *Neurotherapeutics* **11**, 721-731
7. Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., Nan, Z., Forster, C.,

Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* **108**, 260-265

8. Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646

9. Gendron, T. F., Bieniek, K. F., Zhang, Y. J., Jansen-West, K., Ash, P. E., Caulfield, T., Daughrity, L., Dunmore, J. H., Castanedes-Casey, M., Chew, J., Cosio, D. M., van Blitterswijk, M., Lee, W. C., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013) 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

10. Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C., and Edbauer, D. (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335-1338

11. Mori, K., Arzberger, T., Grässer, F. A., Gijselinck, I., May, S., Rentzsch, K., Weng, S. M., Schludi, M. H., van der Zee, J., Cruts, M., Van Broeckhoven, C., Kremmer, E., Kretzschmar, H. A., Haass, C., and Edbauer, D. (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are

translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* **126**, 881-893

12. Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. C., Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., and Paulson, H. L. (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440-455
13. Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., Ostrow, L. W., Rothstein, J. D., Troncoso, J. C., and Ranum, L. P. (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-4977
14. Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., Pletnikova, O., Borchelt, D. R., Ross, C. A., Margolis, R. L., Yachnis, A. T., Troncoso, J. C., and Ranum, L. P. (2015) RAN Translation in Huntington Disease. *Neuron* **88**, 667-677
15. Zu, T., Cleary, J. D., Liu, Y., Bañez-Coronel, M., Bubenik, J. L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H. B., Yachnis, A. T., Swanson, M. S., and Ranum, L. P. W. (2017) RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. *Neuron* **95**, 1292-1305.e1295
16. Soragni, E., Petrosyan, L., Rinkoski, T. A., Wieben, E. D., Baratz, K. H., Fautsch, M. P., and Gottesfeld, J. M. (2018) Repeat-Associated Non-ATG (RAN)

Translation in Fuchs' Endothelial Corneal Dystrophy. *Invest Ophthalmol Vis Sci* **59**, 1888-1896

17. Buijsen, R. A., Visser, J. A., Kramer, P., Severijnen, E. A., Gearing, M., Charlet-Berguerand, N., Sherman, S. L., Berman, R. F., Willemsen, R., and Hukema, R. K. (2016) 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
18. Jackson, R. J., Hellen, C. U., and Pestova, T. V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-127
19. Sonenberg, N., Morgan, M. A., Merrick, W. C., and Shatkin, A. J. (1978) A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc Natl Acad Sci U S A* **75**, 4843-4847
20. Sonenberg, N., Rupprecht, K. M., Hecht, S. M., and Shatkin, A. J. (1979) Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proc Natl Acad Sci U S A* **76**, 4345-4349
21. Grifo, J. A., Tahara, S. M., Morgan, M. A., Shatkin, A. J., and Merrick, W. C. (1983) New initiation factor activity required for globin mRNA translation. *J Biol Chem* **258**, 5804-5810
22. Imataka, H., Gradi, A., and Sonenberg, N. (1998) 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

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

24. Borman, A. M., Michel, Y. M., and Kean, K. M. (2000) 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 Res* **28**, 4068-4075

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

26. LeFebvre, A. K., Korneeva, N. L., Trutschl, M., Cvek, U., Duzan, R. D., Bradley, C. A., Hershey, J. W., and Rhoads, R. E. (2006) Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *J Biol Chem* **281**, 22917-22932

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

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

29. Jackson, R. J. (1991) The ATP requirement for initiation of eukaryotic translation varies according to the mRNA species. *Eur J Biochem* **200**, 285-294

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

31. Svitkin, Y. V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G. J., and Sonenberg, N. (2001) 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

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

33. Zhang, Y., You, J., Wang, X., and Weber, J. (2015) The DHX33 RNA Helicase Promotes mRNA Translation Initiation. *Mol Cell Biol* **35**, 2918-2931

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

35. Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292

36. Yoon, H. J., and Donahue, T. F. (1992) The suil suppressor locus in Saccharomyces cerevisiae encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. *Mol Cell Biol* **12**, 248-260

37. Unbehaun, A., Borukhov, S. I., Hellen, C. U., and Pestova, T. V. (2004) 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 Dev* **18**, 3078-3093

38. Maag, D., Fekete, C. A., Gryczynski, Z., and Lorsch, J. R. (2005) A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol Cell* **17**, 265-275

39. Passmore, L. A., Schmeing, T. M., Maag, D., Applefield, D. J., Acker, M. G., Algire, M. A., Lorsch, J. R., and Ramakrishnan, V. (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell* **26**, 41-50

40. Dever, T. E., and Green, R. (2012) The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol* **4**, a013706

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

42. Kozak, M. (1986) Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci U S A* **83**, 2850-2854

43. Kozak, M. (1988) Leader length and secondary structure modulate mRNA function under conditions of stress. *Mol Cell Biol* **8**, 2737-2744

44. Kozak, M. (1994) Features in the 5' non-coding sequences of rabbit alpha and beta-globin mRNAs that affect translational efficiency. *J Mol Biol* **235**, 95-110

45. Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A* **87**, 8301-8305

46. Peabody, D. S. (1987) Translation initiation at an ACG triplet in mammalian cells. *J Biol Chem* **262**, 11847-11851

47. Peabody, D. S. (1989) Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* **264**, 5031-5035

48. Kozak, M. (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* **9**, 5073-5080

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

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

51. Lee, S., Liu, B., Huang, S. X., Shen, B., and Qian, S. B. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A* **109**, E2424-2432

52. Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., and Qian, S. B. (2015) Quantitative profiling of initiating ribosomes in vivo. *Nat Methods* **12**, 147-153

53. Brar, G. A., Yassour, M., Friedman, N., Regev, A., Ingolia, N. T., and Weissman, J. S. (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**, 552-557

54. Magny, E. G., Pueyo, J. I., Pearl, F. M., Cespedes, M. A., Niven, J. E., Bishop, S. A., and Couso, J. P. (2013) Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science* **341**, 1116-1120

55. Menschaert, G., Van Criekinge, W., Notelaers, T., Koch, A., Crappé, J., Gevaert, K., and Van Damme, P. (2013) 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. *Mol Cell Proteomics* **12**, 1780-1790

56. Slavoff, S. A., Mitchell, A. J., Schwaid, A. G., Cabili, M. N., Ma, J., Levin, J. Z., Karger, A. D., Budnik, B. A., Rinn, J. L., and Saghatelian, A. (2013) Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* **9**, 59-64

57. Vanderperre, B., Lucier, J. F., Bissonnette, C., Motard, J., Tremblay, G., Vanderperre, S., Wisztorski, M., Salzet, M., Boisvert, F. M., and Roucou, X. (2013) Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS One* **8**, e70698

58. Chanut-Delalande, H., Hashimoto, Y., Pelissier-Monier, A., Spokony, R., Dib, A., Kondo, T., Bohère, J., Niimi, K., Latapie, Y., Inagaki, S., Dubois, L., Valenti, P., Polesello, C., Kobayashi, S., Moussian, B., White, K. P., Plaza, S., Kageyama, Y., and Payre, F. (2014) Pri peptides are mediators of ecdysone for the temporal control of development. *Nat Cell Biol* **16**, 1035-1044

59. Pauli, A., Norris, M. L., Valen, E., Chew, G. L., Gagnon, J. A., Zimmerman, S., Mitchell, A., Ma, J., Dubrulle, J., Reyon, D., Tsai, S. Q., Joung, J. K., Saghatelian, A., and Schier, A. F. (2014) Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science* **343**, 1248636

60. Anderson, D. M., Anderson, K. M., Chang, C. L., Makarewich, C. A., Nelson, B. R., McAnally, J. R., Kasaragod, P., Shelton, J. M., Liou, J., Bassel-Duby, R., and Olson, E. N. (2015) A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell* **160**, 595-606

61. Lozano, G., and Martínez-Salas, E. (2015) Structural insights into viral IRES-dependent translation mechanisms. *Curr Opin Virol* **12**, 113-120

62. Wilson, J. E., Pestova, T. V., Hellen, C. U., and Sarnow, P. (2000) Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**, 511-520

63. Wilson, J. E., Powell, M. J., Hoover, S. E., and Sarnow, P. (2000) Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol Cell Biol* **20**, 4990-4999

64. Jan, E., Thompson, S. R., Wilson, J. E., Pestova, T. V., Hellen, C. U., and Sarnow, P. (2001) Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. *Cold Spring Harb Symp Quant Biol* **66**, 285-292

65. Komar, A. A., and Hatzoglou, M. (2005) Internal ribosome entry sites in cellular mRNAs: mystery of their existence. *J Biol Chem* **280**, 23425-23428

66. Chappell, S. A., Dresios, J., Edelman, G. M., and Mauro, V. P. (2006) Ribosomal shunting mediated by a translational enhancer element that base pairs to 18S rRNA. *Proc Natl Acad Sci U S A* **103**, 9488-9493

67. Chappell, S. A., Edelman, G. M., and Mauro, V. P. (2006) Ribosomal tethering and clustering as mechanisms for translation initiation. *Proc Natl Acad Sci U S A* **103**, 18077-18082

68. Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**, 8125-8148

69. Kozak, M. (1991) A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expr* **1**, 111-115

70. Martin, F., Barends, S., Jaeger, S., Schaeffer, L., Prongidi-Fix, L., and Eriani, G. (2011) Cap-assisted internal initiation of translation of histone H4. *Mol Cell* **41**, 197-209

71. Fütterer, J., Kiss-László, Z., and Hohn, T. (1993) Nonlinear ribosome migration on cauliflower mosaic virus 35S RNA. *Cell* **73**, 789-802

72. Yueh, A., and Schneider, R. J. (2000) Translation by ribosome shunting on adenovirus and hsp70 mRNAs facilitated by complementarity to 18S rRNA. *Genes Dev* **14**, 414-421

73. Firth, A. E., and Brierley, I. (2012) Non-canonical translation in RNA viruses. *J Gen Virol* **93**, 1385-1409

74. Rogers, G. W., Edelman, G. M., and Mauro, V. P. (2004) Differential utilization of upstream AUGs in the beta-secretase mRNA suggests that a shunting mechanism regulates translation. *Proc Natl Acad Sci U S A* **101**, 2794-2799

75. Sherrill, K. W., and Lloyd, R. E. (2008) Translation of cIAP2 mRNA is mediated exclusively by a stress-modulated ribosome shunt. *Mol Cell Biol* **28**, 2011-2022

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

77. Dorn, M. B., Mazzocco, M. M., and Hagerman, R. J. (1994) Behavioral and psychiatric disorders in adult male carriers of fragile X. *J Am Acad Child Adolesc Psychiatry* **33**, 256-264

78. Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P. J. (2001) Intention tremor,

parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57**, 127-130

79. Tassone, F., Beilina, A., Carosi, C., Albertosi, S., Bagni, C., Li, L., Glover, K., Bentley, D., and Hagerman, P. J. (2007) Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA* **13**, 555-562

80. Tassone, F., Hagerman, R. J., Taylor, A. K., Gane, L. W., Godfrey, T. E., and Hagerman, P. J. (2000) 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

81. Pieretti, M., Zhang, F. P., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., and Nelson, D. L. (1991) Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**, 817-822

82. Verkerk, A. J., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., and Zhang, F. P. (1991) 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

83. Rousseau, F., Rouillard, P., Morel, M. L., Khandjian, E. W., and Morgan, K. (1995) 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

84. Dombrowski, C., Lévesque, S., Morel, M. L., Rouillard, P., Morgan, K., and Rousseau, F. (2002) 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

85. Jacquemont, S., Hagerman, R. J., Leehey, M. A., Hall, D. A., Levine, R. A., Brunberg, J. A., Zhang, L., Jardini, T., Gane, L. W., Harris, S. W., Herman, K., Grigsby, J., Greco, C. M., Berry-Kravis, E., Tassone, F., and Hagerman, P. J. (2004) Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA* **291**, 460-469

86. Leehey, M. A. (2009) Fragile X-associated tremor/ataxia syndrome: clinical phenotype, diagnosis, and treatment. *J Investig Med* **57**, 830-836

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

88. Cronister, A., Schreiner, R., Wittenberger, M., Amiri, K., Harris, K., and Hagerman, R. J. (1991) Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. *Am J Med Genet* **38**, 269-274

89. Allingham-Hawkins, D. J., Babul-Hirji, R., Chitayat, D., Holden, J. J., Yang, K. T., Lee, C., Hudson, R., Gorwill, H., Nolin, S. L., Glicksman, A., Jenkins, E. C., Brown, W. T., Howard-Peebles, P. N., Becchi, C., Cummings, E., Fallon, L., Seitz, S., Black, S. H., Vianna-Morgante, A. M., Costa, S. S., Otto, P. A., Mingroni-Netto, R. C., Murray, A., Webb, J., and Vieri, F. (1999) 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

90. Buijsen, R. A., Sellier, C., Severijnen, L. A., Oulad-Abdelghani, M., Verhagen, R. F., Berman, R. F., Charlet-Berguerand, N., Willemsen, R., and Hukema, R. K. (2014) 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

91. Oh, S. Y., He, F., Krans, A., Frazer, M., Taylor, J. P., Paulson, H. L., and Todd, P. K. (2015) RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. *Hum Mol Genet* **24**, 4317-4326

92. Hukema, R. K., Buijsen, R. A., Schonewille, M., Raske, C., Severijnen, L. A., Nieuwenhuizen-Bakker, I., Verhagen, R. F., van Dessel, L., Maas, A., Charlet-Berguerand, N., De Zeeuw, C. I., Hagerman, P. J., Berman, R. F., and Willemsen, R. (2015) Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. *Hum Mol Genet* **24**, 4948-4957

93. Chiang, P. W., Carpenter, L. E., and Hagerman, P. J. (2001) The 5'-untranslated region of the FMR1 message facilitates translation by internal ribosome entry. *J Biol Chem* **276**, 37916-37921

94. Dobson, T., Kube, E., Timmerman, S., and Krushel, L. A. (2008) Identifying intrinsic and extrinsic determinants that regulate internal initiation of translation mediated by the FMR1 5' leader. *BMC Mol Biol* **9**, 89

95. Chen, L. S., Tassone, F., Sahota, P., and Hagerman, P. J. (2003) The (CGG)n repeat element within the 5' untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of a downstream reporter. *Hum Mol Genet* **12**, 3067-3074

96. Ludwig, A. L., Hershey, J. W., and Hagerman, P. J. (2011) Initiation of translation of the FMR1 mRNA Occurs predominantly through 5'-end-dependent ribosomal scanning. *J Mol Biol* **407**, 21-34

97. Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., and Todd, P. K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**, 314-322

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

99. Sobczak, K., Michlewski, G., de Mezer, M., Kierzek, E., Krol, J., Olejniczak, M., Kierzek, R., and Krzyzosiak, W. J. (2010) Structural diversity of triplet repeat RNAs. *J Biol Chem* **285**, 12755-12764

100. Bordeleau, M. E., Mori, A., Oberer, M., Lindqvist, L., Chard, L. S., Higa, T., Belsham, G. J., Wagner, G., Tanaka, J., and Pelletier, J. (2006) Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. *Nat Chem Biol* **2**, 213-220

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

102. Muhs, M., Hilal, T., Mielke, T., Skabkin, M. A., Sanbonmatsu, K. Y., Pestova, T. V., and Spahn, C. M. (2015) Cryo-EM of ribosomal 80S complexes with

termination factors reveals the translocated cricket paralysis virus IRES. *Mol Cell* **57**, 422-432

103. Iwahashi, C. K., Yasui, D. H., An, H. J., Greco, C. M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C. B., Hagerman, R. J., and Hagerman, P. J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256-271

104. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256

105. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V. M., Kaivorinne, A. L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E.,

Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C., Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J., Traynor, B. J., and Consortium, I. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268

106. Johnston, C. A., Stanton, B. R., Turner, M. R., Gray, R., Blunt, A. H., Butt, D., Ampong, M. A., Shaw, C. E., Leigh, P. N., and Al-Chalabi, A. (2006) Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. *J Neurol* **253**, 1642-1643

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

108. Luukkainen, L., Bloigu, R., Moilanen, V., and Remes, A. M. (2015) Epidemiology of Frontotemporal Lobar Degeneration in Northern Finland. *Dement Geriatr Cogn Dis Extra* **5**, 435-441

109. Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P. H., Albert, M., Boone, K., Miller, B. L., Cummings, J., and Benson, D. F. (1998) Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* **51**, 1546-1554

110. Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S., Engelborghs, S., Sieben, A., De Jonghe, P., Vandenberghe, R., Santens, P., De Bleecker, J., Maes, G., Bäumer, V., Dillen, L., Joris, G., Cuijt, I., Corsmit, E., Elinck, E., Van Dongen, J., Vermeulen, S., Van den Broeck, M., Vaerenberg, C.,

Mattheijssens, M., Peeters, K., Robberecht, W., Cras, P., Martin, J. J., De Deyn, P. P., Cruts, M., and Van Broeckhoven, C. (2012) 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

111. Lomen-Hoerth, C., Murphy, J., Langmore, S., Kramer, J. H., Olney, R. K., and Miller, B. (2003) Are amyotrophic lateral sclerosis patients cognitively normal? *Neurology* **60**, 1094-1097

112. Ringholz, G. M., Appel, S. H., Bradshaw, M., Cooke, N. A., Mosnik, D. M., and Schulz, P. E. (2005) Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology* **65**, 586-590

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

114. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130-133

115. Fratta, P., Mizielinska, S., Nicoll, A. J., Zloh, M., Fisher, E. M., Parkinson, G., and Isaacs, A. M. (2012) C9orf72 hexanucleotide repeat associated with amyotrophic

lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* **2**, 1016

116. Reddy, K., Zamiri, B., Stanley, S. Y., Macgregor, R. B., and Pearson, C. E. (2013) 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

117. Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., and Wang, J. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195-200

118. Su, Z., Zhang, Y., Gendron, T. F., Bauer, P. O., Chew, J., Yang, W. Y., Fostvedt, E., Jansen-West, K., Belzil, V. V., Desaro, P., Johnston, A., Overstreet, K., Oh, S. Y., Todd, P. K., Berry, J. D., Cudkowicz, M. E., Boeve, B. F., Dickson, D., Floeter, M. K., Traynor, B. J., Morelli, C., Ratti, A., Silani, V., Rademakers, R., Brown, R. H., Rothstein, J. D., Boylan, K. B., Petrucelli, L., and Disney, M. D. (2014) Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**, 1043-1050

119. Dodd, D. W., Tomchick, D. R., Corey, D. R., and Gagnon, K. T. (2016) Pathogenic C9ORF72 Antisense Repeat RNA Forms a Double Helix with Tandem C:C Mismatches. *Biochemistry*

120. Al-Sarraj, S., King, A., Troakes, C., Smith, B., Maekawa, S., Bodi, I., Rogelj, B., Al-Chalabi, A., Hortobágyi, T., and Shaw, C. E. (2011) 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

121. Boxer, A. L., Mackenzie, I. R., Boeve, B. F., Baker, M., Seeley, W. W., Crook, R., Feldman, H., Hsiung, G. Y., Rutherford, N., Laluz, V., Whitwell, J., Foti, D., McDade, E., Molano, J., Karydas, A., Wojtas, A., Goldman, J., Mirsky, J., Sengdy, P., Dearmond, S., Miller, B. L., and Rademakers, R. (2011) Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. *J Neurol Neurosurg Psychiatry* **82**, 196-203

122. Zhang, Y. J., Jansen-West, K., Xu, Y. F., Gendron, T. F., Bieniek, K. F., Lin, W. L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., Chew, J., Belzil, V. V., Prudencio, M., Stankowski, J. N., Castanedes-Casey, M., Whitelaw, E., Ash, P. E., DeTure, M., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2014) Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* **128**, 505-524

123. Schludi, M. H., May, S., Grässer, F. A., Rentzsch, K., Kremmer, E., Küpper, C., Klopstock, T., Arzberger, T., Edbauer, D., Degeneration, G. C. f. F. L., and Alliance, B. B. B. (2015) Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathol* **130**, 537-555

124. Mackenzie, I. R., Arzberger, T., Kremmer, E., Troost, D., Lorenzl, S., Mori, K., Weng, S. M., Haass, C., Kretzschmar, H. A., Edbauer, D., and Neumann, M. (2013) Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* **126**, 859-879

125. Mackenzie, I. R., Frick, P., Grässer, F. A., Gendron, T. F., Petrucelli, L., Cashman, N. R., Edbauer, D., Kremmer, E., Prudlo, J., Troost, D., and Neumann, M. (2015) Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol* **130**, 845-861

126. Davidson, Y., Robinson, A. C., Liu, X., Wu, D., Troakes, C., Rollinson, S., Masuda-Suzukake, M., Suzuki, G., Nonaka, T., Shi, J., Tian, J., Hamdalla, H., Ealing, J., Richardson, A., Jones, M., Pickering-Brown, S., Snowden, J. S., Hasegawa, M., and Mann, D. M. (2015) Neurodegeneration in Frontotemporal Lobar Degeneration and Motor Neurone Disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. *Neuropathol Appl Neurobiol*

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

128. Haass, C., and Selkoe, D. J. (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat Rev Mol Cell Biol* **8**, 101-112

129. Edbauer, D., and Haass, C. (2015) An amyloid-like cascade hypothesis for C9orf72 ALS/FTD. *Curr Opin Neurobiol* **36**, 99-106

130. Chew, J., Gendron, T. F., Prudencio, M., Sasaguri, H., Zhang, Y. J., Castanedes-Casey, M., Lee, C. W., Jansen-West, K., Kurti, A., Murray, M. E., Bieniek, K. F., Bauer, P. O., Whitelaw, E. C., Rousseau, L., Stankowski, J. N., Stetler, C.,

Daughrity, L. M., Perkerson, E. A., Desaro, P., Johnston, A., Overstreet, K., Edbauer, D., Rademakers, R., Boylan, K. B., Dickson, D. W., Fryer, J. D., and Petrucelli, L. (2015) Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* **348**, 1151-1154

131. Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., Paul, J. W., Sun, S., Herdy, J. R., Bieri, G., Kramer, N. J., Gage, F. H., Van Den Bosch, L., Robberecht, W., and Gitler, A. D. (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18**, 1226-1229

132. Yang, D., Abdallah, A., Li, Z., Lu, Y., Almeida, S., and Gao, F. B. (2015) FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathol* **130**, 525-535

133. Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., Lin, S., Shneider, N. A., Monaghan, J., Pandey, U. B., Pasinelli, P., Ichida, J. K., and Trotti, D. (2014) 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

134. Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., Moens, T., Norona, F. E., Woollacott, I. O., Pietrzyk, J., Cleverley, K., Nicoll, A. J., Pickering-Brown, S., Dols, J., Cabecinha, M., Hendrich, O., Fratta, P., Fisher, E. M., Partridge, L., and Isaacs, A. M. (2014) C9orf72 repeat expansions cause

neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192-1194

135. Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., Kim, N. C., Almeida, S., Lee, K. H., Badders, N., Valentine, M., Miller, B. L., Wong, P. C., Petrucelli, L., Kim, H. J., Gao, F. B., and Taylor, J. P. (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129-133

136. Tran, H., Almeida, S., Moore, J., Gendron, T. F., Chalasani, U., Lu, Y., Du, X., Nickerson, J. A., Petrucelli, L., Weng, Z., and Gao, F. B. (2015) Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* **87**, 1207-1214

137. Tao, Z., Wang, H., Xia, Q., Li, K., Jiang, X., Xu, G., Wang, G., and Ying, Z. (2015) Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet* **24**, 2426-2441

138. Yamakawa, M., Ito, D., Honda, T., Kubo, K., Noda, M., Nakajima, K., and Suzuki, N. (2015) Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum Mol Genet* **24**, 1630-1645

139. May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grässer, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., Mann, M., Meissner, F., and Edbauer, D. (2014) C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* **128**, 485-503

140. Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S. L. (2014) Poly-dipeptides encoded by the

C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139-1145

141. Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., Daley, E. L., Miller, S. J., Cunningham, K. M., Vidensky, S., Gupta, S., Thomas, M. A., Hong, I., Chiu, S. L., Huganir, R. L., Ostrow, L. W., Matunis, M. J., Wang, J., Sattler, R., Lloyd, T. E., and Rothstein, J. D. (2015) The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56-61

142. Niblock, M., Smith, B. N., Lee, Y. B., Sardone, V., Topp, S., Troakes, C., Al-Sarraj, S., Leblond, C. S., Dion, P. A., Rouleau, G. A., Shaw, C. E., and Gallo, J. M. (2016) Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. *Acta Neuropathol Commun* **4**, 18

Chapter 2: RAN Translation at C9orf72-Associated Repeat Expansions is Selectively Enhanced by the Integrated Stress Response

This chapter is published as:

Green KM, Glineburg MR, Kearse MG, Flores BN, Linsalata AE, Fedak SJ, Goldstrohm AC, Barmada SJ, Todd PK. RAN translation at *C9orf72*-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications.* 2017 Dec 8. 8(1): 2005.

2.1 Statement of others' contributions to data presented in this chapter

M. Rebecca Glineburg and Mike Kearse performed the majority of experiments usings the CGG reporters. M. Rebecca Glineburg, Mike Kearse, and Stephen Fedak performed the experiments using the near-AUG reporters. Brittany Flores and Sami Barmada performed all experiments with primary rodent neurons and analyzed the results, except for the experiments in Fig. 2-1, E. Alex Linsalata generated the HeLa translation lysate, and performed the experiments using it. I generated the C9RAN translation reporters and performed all experiments assessing their expression. Peter Todd and I wrote the manuscript, with extensive feedback from M. Rebecca Glineburg and Mike Kearse. I also genertated all figures.

2.2 Introduction

Nucleotide repeat expansions cause multiple neurodegenerative disorders (1). Recently, an unconventional form of translation initiation known as repeat-associated non-AUG (RAN) translation has emerged as a novel mechanism by which repeat expansions cause toxicity (2,3). RAN translation occurs in the absence of an AUG start codon, in multiple reading frames, through an expanded repeat to produce homopolymeric or dipeptide repeat-containing proteins (DPRs). This non-canonical initiation event occurs in multiple disorders, including at CAG and CUG repeats in spinocerebellar ataxia type 8 (SCA8) and Huntington's disease, and at CGG and CCG repeats in fragile X-associated tremor/ataxia syndrome (FXTAS) (2,4-6).

A G_4C_2 repeat expansion located in the first intron of *C9orf72* is the most common known inherited cause of both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (7,8). In C9ALS/FTD, this repeat is often expanded from <25 units to upwards of several hundred, although disease occurs with as few as 70 repeats (7-9). Despite its intronic localization, RAN translation occurs at this locus (C9RAN) at both sense strand-derived G_4C_2 repeats and antisense strand-derived C_4G_2 repeat transcripts to generate six different DPRs (10-12). These DPRs accumulate in p62 and ubiquitin positive aggregates in C9ALS/FTD neurons, consistent with pathology observed in many repeat expansion disorders (10,11).

DPRs are both necessary and sufficient to induce neurodegeneration in simple model systems (13-15). DPRs elicit toxicity through a number of mechanisms, including altered ribosomal biogenesis, impaired nucleocytoplasmic transport, shifts in RNA metabolism, protein sequestration, and impaired protein quality control pathways

(13,16-21). The charged DPRs, glycine-arginine and proline-arginine, in particular accumulate in membrane-less organelles, including RNA granules, and are associated with suppressed global protein synthesis and altered granule dynamics (17,22-24). However, most of these findings originated from studies that relied upon DPR production not through RAN translation, but through AUG-initiated translation of a synthetic non-repetitive RNA sequence. As such, while the relative toxicity of different DPR species in isolation is established, their relative stoichiometry and translation kinetics remain unclear.

Despite a potentially central role in multiple neurodegenerative disorders, our understanding of the mechanism(s) of RAN translation is incomplete. Canonical eukaryotic translation initiation follows a scanning mechanism, where the 5′ m^7G-cap recruits the cap-binding complex eIF4F (composed of the cap-binding protein eIF4E, eIF4G, and the DEAD box helicase eIF4A) to the 5′ end of the mRNA (25,26). In parallel, the multi-subunit GTPase initiation factor eIF2 binds to the initiator methionine tRNA ($tRNA_i^{Met}$) in its GTP-bound state to generate the ternary complex, which then assembles with the 40S ribosomal subunit and other initiation factors to form the 43S pre-initiation complex (PIC). The PIC associates with eIF4F at the mRNA 5′ m^7G-cap. This complex scans along the mRNA in a 5′ to 3′ direction in a process promoted by eIF4A, until it encounters an AUG start codon in an appropriate Kozak sequence context in the P-site (25,26). eIF5 then promotes hydrolysis of GTP to GDP on eIF2, and eIF2-GDP and P_i is released, allowing for recruitment of the 60S subunit and decoding of the second codon in the A-site (25-27).

While the scanning model of translation initiation applies to many transcripts under basal conditions, a variety of alternative initiation mechanisms exist that bypass these requirements. Internal ribosomal entry sites (IRESs) are often complex RNA structures that promote translation initiation independent of the 5′ cap, specific initiation factors and, in certain cases (*e.g.* the cricket paralysis virus [CrPV] IRES), bypass the need for any initiation factors or an AUG codon (26,28). In addition, cells actively regulate translation initiation after exposure to a variety of perturbations in cellular homeostasis through the integrated stress response (ISR, reviewed in 25,29,30). ER stress, viral infection, amino acid starvation and other triggers stimulate ISR kinase cascades that converge to phosphorylate the regulatory initiation factor eIF2α at serine 51. This phosphorylation event suppresses global protein synthesis by inhibiting eIF2B, the GEF that exchanges GDP for GTP on eIF2, thus preventing eIF2 rebinding to $tRNA_i^{Met}$ and forming additional ternary complexes. However, a subset of mRNAs escapes this suppression through use of upstream open reading frames (uORFs), IRES elements, and/or non-AUG initiation codons and retain expression under stress conditions (25,30-35).

How canonical and non-canonical translation initiation rules intersect with RAN translational requirements is not yet known. If RAN translation contributes meaningfully to pathogenicity in repeat expansion disorders, then identification of specific factors that selectively favor RAN translation may reveal novel targets for therapeutic development across a range of neurological disorders. Moreover, by identifying what cellular conditions influence RAN translation, we can gain insights into critical disease mechanisms underlying C9ALS/FTD and other neurodegenerative diseases. To these

ends, we established a series of C9RAN translation-specific reporters and investigated the mechanisms mediating RAN translation at G_4C_2 repeat expansions using both *in vitro* and cell-based assays. C9RAN translation utilizes a cap-, eIF4E-, and eIF4A-dependent scanning mechanism to initiate translation predominantly at a CUG codon just upstream of the repeat. RAN translation at both CGG and G_4C_2 repeats is selectively enhanced by ISR activation and eIF2α phosphorylation. These same disease-causing repeats independently impair global protein synthesis and activate stress granule formation, creating a potential feed-forward loop that drives a toxic cascade towards neurodegeneration.

2.3 Results

G_4C_2 RAN translation levels differ across reading frames

To determine how RAN translation occurs at G_4C_2 repeats, we designed a series of reporters containing the first *C9orf72* intron through the G_4C_2 repeat, for use in *in vitro* and cell-based assays (**Fig. 2-1, A**). This sequence was inserted upstream of a modified NanoLuciferase (NLuc) reporter with its AUG start codon mutated to GGG (36). A carboxy-terminal 3xFLAG-tag was included for western blot detection and a precision protease (PSP) cleavage site was introduced between the repeat and reporter sequences to allow for efficient release of NLuc from the DPR.

Consistent with published results (36), mutating NLuc's AUG (AUG-NLuc) to GGG (GGG-NLuc) resulted in a >1,000 fold reduction in luciferase activity in rabbit reticulocyte lysate (RRL) *in vitro* translation assays and loss of the major immunoreactive protein detected by western blot (**Fig. 2-1, B & C**). When *C9orf72*

intron 1 containing 70 G_4C_2 repeats was inserted upstream of this reporter in the glycine-alanine (GA) reading frame (**Fig. 2-1, A**), there was an approximately 300-fold recovery of luciferase signal and the appearance of a higher molecular weight species by western blot (**Fig. 2-1, B & C**). Consistent with initiation upstream or within the expanded repeat, the observed molecular weight of GA-NLuc fusion protein increased proportionally with repeat length (**Fig. 2-1, B**). Similar results were seen when the reporters were expressed in HEK293 cells and in a distinct *in vitro* system generated from HeLa cell lysates (**Fig. 2-2, A-C**) (37).

As C9RAN also occurs in the glycine-proline (GP) and glycine-arginine (GR) reading frames, we generated additional reporters for production of these DPRs by inserting one or two nucleotides, respectively, between the repeat and NLuc sequences (**Fig. 2-1, A**). When expressed in RRL, HeLa cell lysate, HEK293 cells, and primary rat hippocampal neurons, C9RAN reporter expression was significantly lower in the GP and GR frames, relative to the GA frame, but still above the GGG-NLuc control (**Fig. 2-1, C-E** and **Fig. 2-2, B & C**). This difference in NLuc expression between reading frames was likely not a result of differences in protein stability, as the stability of GA, GP, and GR-NLuc fusion proteins were similar in HEK293 cells and not more stable than the AUG-NLuc control (**Fig. 2-2, D**). Additionally, to control for the possibility that each DPR differentially affects NLuc function, we compared luciferase activity of each C9RAN reporter expressed in RRL before and after cleaving at the engineered PSP site (**Fig.2-1, A** and **Fig. 2-2, E & F**) (36). A 15% increase in NLuc activity was observed for the GR-NLuc reporter upon cleavage (**Fig. 2-1, E**), but this small affect cannot account for the nearly 140 fold difference in expression between the GR and GA frames (**Fig. 2-1,**

C). Importantly, this difference in expression level between the three reading frames is consistent with differences in DPR abundance measured in C9ALS/FTD autopsy brain samples by immunohistochemistry (38). Thus, our C9RAN reporters are specific to each reading frame, exhibit consistent patterns across four systems, and recapitulate the expression pattern seen in disease tissue.

To determine whether differential elongation rates contributed to the observed difference in RAN translation levels across the three sense reading frames, AUG-driven reporters for each reading frame were generated. These reporters contained an AUG start codon in optimal Kozak sequence context immediately upstream of the 70 G_4C_2 repeats and lacked the UAG stop codon that natively occurs in the GP reading frame immediately upstream the repeat (**Fig. 2-3, A**). When expressed in RRL and HEK293 cells, NLuc levels from the GP and GA reporters were no longer significantly different, while GR-NLuc production remained lower than both (**Fig. 2-2, G & H**). This suggests that the ribosome can synthesize poly-GA and poly-GP products with similar efficiency, but that differences in initiation rates impede poly-GP RAN translation. In contrast, these data indicate that lower synthesis rates of the GR DPR may be caused by differences in both elongation and initiation rates.

RAN translation at G_4C_2 repeats is cap- and eIF4A-dependent

We next examined the requirement of the 5′ m^7G-cap for C9RAN translation by transcribing C9RAN NLuc reporters with either the canonical m^7G-cap or an A-cap analog that cannot recruit the cap-binding initiating factor eIF4E, but protects the mRNA from degradation (**Fig. 2-3, A**)**.** As a control for cap-independent initiation, 5′ m^7G- or A-

capped mRNAs with the CrPV IRES placed upstream of NLuc were also generated (36). In RRL and HEK293 cells, A-capped C9RAN reporter mRNAs had dramatically decreased expression in all reading frames compared to m^7G-capped mRNAs, whereas translation from the CrPV IRES was unaffected (**Fig. 2-3, B** and **Fig. 2-4, A**). Similarly, addition of free m^7G-cap to the RRL translation reaction *in trans*, to competitively inhibit eIF4E binding to reporter mRNAs, significantly reduced C9RAN in all three readings frames without affecting CrPV expression levels (**Fig. 2-3, C**). Together, these data indicate that RAN translation from these reporters proceeds through a cap- and eIF4E-dependent mechanism, and that *C9orf72* intron1 with 70 G_4C_2 repeats does not act as an IRES.

We next assessed whether C9RAN translation requires ribosomal scanning after recruitment of the PIC to the 5′ m^7G-cap. PIC scanning is dependent upon the RNA helicase, eIF4A, which is specifically inhibited by hippuristanol (39). Addition of hippuristanol to RRL reactions dramatically inhibited translation of the control AUG-NLuc reporter, whereas expression of the CrPV IRES reporter was unaffected (**Fig. 2-3, D**), consistent with previous reports (36,39). Expression of all three C9RAN reporters was significantly reduced (over 900-fold) with eIF4A inhibition by hippuristanol (**Fig. 2-3, D**). Thus, our C9RAN reporters exhibit a strong dependence on eIF4A, similar to what we have previously shown for RAN translation of expanded CGG repeats within the 5′ UTR of *FMR1* in FXTAS (36). These results are consistent with a scanning model of initiation.

C9RAN translation uses a near-cognate codon for initiation

RAN translation at CGG repeats initiates upstream of the repeat at near-cognate codons (codons that differ from AUG by a single nucleotide) in some reading frames (36,40). To determine if a similar mechanism occurs in C9RAN, the sequence upstream of the repeat in all three reading frames was examined. In the GA frame, there are two near-cognate codons: a CUG at position -24 and an AGG at position -15 relative to the first nucleotide of the repeat (**Fig. 2-3, A**). The CUG codon is in a strong Kozak sequence context, while the AGG codon is not. Mutating the AGG codon to AAA alone had little effect on C9RAN translation in RRL or HEK293 cells for any reading frame (**Fig. 2-3, E** and **Fig. 2-4, B**). In contrast, mutating the CUG codon to CCC either in the presence or absence of the AGG codon led to a marked reduction in C9RAN in the GA reading frame, in RRL, HEK293 cells, and HeLa cell lysate (**Fig. 2-3, E** and **Fig. 2-4, B & C**), suggesting that this near-cognate codon is utilized for the majority of RAN translation initiation in the GA frame. Surprisingly, despite being located in the GA frame, mutating the CUG to CCC also suppressed RAN translation in the GR reading frame in RRL (**Fig. 2-3, E**), and enhanced RAN translation in the GP reading frame in both RRL and HeLa cell lysate (**Fig. 2-3, E** and **Fig. 2-4, C**). However, these correlative and anti-correlative effects were not observed in transfected HEK293 cells, where the CUG to CCC mutation did not significantly alter translation in the GR reading frame and decreased translation in the GP reading frame (**Fig. 2-4, B**).

In a reciprocal experiment, we converted this same CUG codon to AUG. This mutation significantly decreased expression in the GP frame in RRL and HEK293 cells (**Fig. 2-4, D & F**). In contrast, the CUG to AUG mutation significantly increased expression in the GR frame in RRL, although this effect was not observed in HEK293

cells (**Fig. 2-4, D & F**). As expected, placing an AUG start codon above the repeat in the GA reading frame greatly enhanced production of GA-NLuc in both systems (**Fig. 2-4, E & G**). Consequently, inhibiting or enhancing translation in the GA frame by modifying start codon usage alters expression in the GP and GR frames. Interestingly, the interplay between translation in the GA and GR frames differs between HEK293 cells and RRL, suggesting functional differences between these assay systems.

Together, these data support a model for C9RAN initiation, in which the PIC is recruited to the mRNA's 5′ cap via interaction with eIF4E and utilizes the eIF4A helicase to scan in the 3′ direction. Initiation at a CUG codon upstream of the repeat is important for translation in the GA reading frame. However, if the ribosome fails to initiate at this CUG codon, it may continue scanning into the repeat, where it could initiate in the GP frame in the absence of any near-cognate codon.

Cellular stress selectively enhances RAN translation

Cellular stressors such as viral infection, misfolded proteins and amino acid starvation, can activate the ISR through one of four kinases (interferon-induced double-stranded RNA-dependent eIF2α kinase [PKR], endoplasmic reticulum [ER]-resident kinase [PERK], general control non-derepressible 2 [GCN2], or heme-regulated inhibitor kinase [HRI]), that all phosphorylate eIF2α at serine 51 (**Fig. 2-5, A**) (29,33). As both start codon stringency and initiation kinetics are modulated in response to eIF2α phosphorylation following ISR activation (30,31,33,41), we hypothesized that RAN translation might be refractory to ISR-activation. To test this, cells transfected with C9RAN reporters were exposed to the ER calcium pump inhibitor, thapsigargin (TG), to

cause ER stress and activate the ISR through PERK (**Fig. 2-5, A**). As expected, ER stress induction by TG led to PERK phosphorylation, BiP upregulation, and increased eIF2α phosphorylation (**Fig. 2-5, B**), as well as global translation repression (**Fig. 2-6A**) (42). Consistent with this, expression of both the AUG-NLuc and co-transfected firefly luciferase (FLuc), which serves as an independent internal control, decreased when cells were stressed with TG (**Fig. 2-5, C**). This effect was less pronounced for AUG-NLuc than FLuc, as expected due to its heightened stability (**Fig. 2-2, D**). Therefore, when destabilized with a PEST tag, AUG-NLuc expression was more greatly decreased by TG treatment (**Fig. 2-6, B**). In contrast, expression of all C9RAN translation reporters was significantly increased during ER stress with TG treatment as shown by both luciferase activity and western blot (**Fig. 2-5, B & C**).

To determine if this enhancement was unique to C9RAN, we also interrogated the effect of ISR induction on CGG RAN translation. RAN translation at CGG repeats occurs predominantly in two reading frames (**Fig. 2-5, D**) (5). Initiation in the +1 (GGC, glycine) reading frame occurs mainly at either an ACG or GUG codon just upstream of the repeat to generate FMRpolyG, a protein that accumulates in intranuclear inclusions in FXTAS (5,36,40). In contrast, RAN translation in the +2 (GCG, alanine) reading frame is less robust and likely initiates in the repeat sequence itself to produce FMRpolyA (5,36). TG-induced ER stress significantly enhanced luciferase activity and immunodetection of both +1 or +2 CGG RAN translation reporters, but not the internal FLuc control, compared to vehicle treatment in HEK293T cells (**Fig. 2-5, D to F**). Similarly, stress-stimulated global translation attenuation with tunicamycin (TM), which blocks N-linked glycosylation in the Golgi to cause ER stress, or sodium arsenite (SA),

which causes oxidative stress and activation of the HRI kinase (**Fig. 2-5, A** and **Fig. 2-6, A**), also either spared or enhanced CGG RAN translation while significantly inhibiting AUG-initiated translation (**Fig. 2-6, C & D**). Thus, activation of ISR pathways enhances RAN translation across at least two repeats and five separate reading frames.

To assess the effect of ISR activation of C9RAN in neurons, we utilized automated fluorescent microscopy of primary rat cortical neurons co-transfected with GFP or (G_4C_2)x66-GFP (a reporter containing 66 G_4C_2 repeats in the GP reading frame just upstream of GFP) and mApple, a red fluorescent marker used to facilitate longitudinal tracking. Single-cell fluorescence intensity for both reporters was measured for three days after TG treatment at varying doses. As observed with AUG-initiated luciferase reporters in HEK cells, ER stress induction reduced signal from AUG-initiated mApple and AUG-initiated GFP (**Fig. 2-5, G**)**.** However, expression of (G_4C_2)x66-GFP reporter increased in a dose-dependent manner with TG (**Fig. 2-5, G**). Consequently, cellular stress induction in multiple cell types, including neurons, selectively enhances the production of neurotoxic RAN proteins involved in two distinct neurodegenerative diseases.

eIF2α phosphorylation selectively enhances RAN translation

We next explored additional approaches to more directly evaluate the role of eIF2α phosphorylation in RAN translation. Salubrinol (Sal003) selectively inhibits PP1, the major phosphatase that acts on eIF2α (**Fig. 2-5, A**); treatment with Sal003 thus increases cellular levels of phosphorylated eIF2α (**Fig. 2-8, A**) (43). Addition of Sal003 to transfected cells had only modest inhibitory effects on production of both canonically-

translated AUG-NLuc and FLuc reporters (**Fig. 2-7, A & B**). In contrast, treatment with Sal003 significantly enhanced RAN translation from both C9RAN and CGG RAN reporters, by both luciferase activity and western blot (**Fig. 2-7, A to C**). Furthermore, co-transfecting cells with NLuc reporters and a phosphomimetic form of eIF2α (S51D) suppressed translation of both AUG-NLuc and control FLuc reporters, relative to co-transfection with WT eIF2α, while selectively enhancing RAN translation from C9 and CGG RAN reporters in all assessed reading frames (**Fig. 2-7, D & E**). These data show that eIF2α phosphorylation is sufficient to enhance RAN translation.

To determine if eIF2α phosphorylation is also necessary for stress-induced RAN translation, MEFs homozygous for the WT (S51 S/S) or a non-phosphorylatable eIF2α (S51 A/A) (42), were co-transfected with CGG or C9RAN NLuc reporters and a FLuc internal control, and treated with TM or TG. While both TM and TG treatments increased CGG and C9RAN NLuc expression relative to FLuc in WT MEFs, this enhancement was lost in the S51 A/A MEFs (**Fig. 2-7, F and Fig. 2-8, B**). Thus, eIF2α phosphorylation following induction of the ISR is both necessary and sufficient to selectively enhance RAN translation under conditions that simultaneously suppress global canonical translation initiation.

Stress-induced RAN translation requires a non-AUG codon

Cellular stress and ISR activation can favor initiation at near-cognate codons (30,32,33,35). To determine if the initiation codon is important in ISR-mediated activation of RAN translation, we inserted an AUG start codon upstream of the repeat in C9RAN and CGG RAN reporters (36) to drive canonical translation of the expanded

repeat (**Fig. 2-9, A & B**). Unlike RAN translation, translation of the repeats from a canonical AUG start codon did not show enhancement in response to treatment with TG or Sal003, but behaved similarly to AUG-NLuc (**Fig. 2-9, A & B and Fig. 2-10, A**). Next, to determine if a near-cognate codon alone was sufficient to allow initiation in the setting of cellular stress, we created a set of reporters with the AUG codon of NLuc mutated to one of the near-cognate codons utilized for C9 and CGG RAN translation (**Fig. 2-9, C**). Mutation to ACG, CUG, or GUG significantly impaired translation under basal conditions compared to AUG (**Fig. 2-10, B**) (44). However, initiation at near-cognate codons was enhanced in response to treatment with TG and Sal003 (**Fig. 2-9, C & D**), and was relatively spared when compared to AUG-NLuc when co-transfected with eIF2α S51D (**Fig. 2-10, C**) or treated with SA and TM (**Fig. 2-10, D & E**). Therefore, initiation at a non-AUG start codon is necessary to promote RAN translation in response to stress.

Repeats trigger stress granules & inhibit global translation

The ISR is activated in multiple neurodegenerative disorders (45). C9RAN DPRs can suppress global translation, and overexpression of GA DPR proteins elicits ER stress in neurons (23,24,46,47,48). We therefore evaluated whether and how repeat containing constructs impact global protein synthesis and activate the ISR. A common phenomenon during cellular stress is the formation of stress granules, membrane-less structures composed of mRNAs, stalled translation pre-initiation complexes, and multiple RNA binding proteins (*e.g.* FMRP and G3BP) (49). When either G_4C_2 or CGG repeat-containing reporters were overexpressed in HEK293 cells, they elicited cytoplasmic FMRP and G3BP-positive stress granule formation (**Fig. 2-11, A** and **Fig.**

2-12, A) (48). Concomitantly, overexpression of the +1 CGG RAN reporter inhibited global translation in HEK293 cells, relative to AUG-NLuc, as measured by reduced puromycin incorporation through the surface sensing of translation (SUnSET) assay (**Fig. 2-11, B**) (50). These results are consistent with the previously reported translational suppression stimulated by G_4C_2 repeats (48).

To determine if similar effects were observed in cells directly impacted in the disease state, we assessed whether CGG repeat overexpression affected translation in primary rat cortical neurons tracked by automated fluorescence microscopy. Neurons were co-transfected with mApple and either a reporter containing 100 CGG repeats upstream of GFP (CGGx100-GFP) or GFP alone. mApple expression was then measured over a ten-day time course of imaging to determine if expanded CGG repeats caused translation attenuation. mApple fluorescence intensity remained stable in neurons co-transfected with GFP alone, but decreased by nearly 50% in neurons expressing CGGx100-GFP (**Fig. 2-11, C**). This was independent of cytotoxicity elicited by the repeat expansions (data not shown).

Stress granules induced by ISR activation are dependent on eIF2α phosphorylation (42). To investigate whether G_4C_2 and CGG repeat-induced stress granules require phosphorylation of eIF2α, reporters were transfected into eIF2α S51 S/S and S51 A/A MEFs (42). In agreement with previous studies (42), S/S MEFs formed robust FMRP and G3BP1-positive stress granules in response to TG, while A/A MEFs did not (**Fig. 2-12, B**). Similarly, both G_4C_2 and CGG repeat reporters readily induced stress granule formation in S/S MEFs, but exhibited an approximately 10-fold decrease in stress granule formation in eIF2α A/A MEFs (**Fig. 2-11, D** and **Fig. 2-12, C**).

Together, these data suggest that expression of both G_4C_2 and CGG repeat expansions impairs global translation and stimulate the formation of phosphorylated-eIF2α-dependent stress granules.

2.4 Discussion

RAN translation from repeat expansions contributes significantly to the pathology of multiple neurodegenerative disorders, including C9ALS/FTD (2,5,13-15,18,23,40,47). Here we find that C9RAN translation initiates through a 5′ m^7G-cap-, eIF4E-, and eIF4A-dependent mechanism. RAN translation starts either at a near-cognate CUG start codon just upstream of the repeat or potentially within the repeat itself, depending on the reading frame. RAN translation at both G_4C_2 and CGG repeats is enhanced by activation of ISR pathways which normally suppress global translation. This effect is independent of the stress stimuli applied, as it can be recapitulated directly by altering eIF2α phosphorylation, but is dependent upon the initiation codon (*i.e.* AUG versus a near-cognate codon). Moreover, overexpression of either CGG or G_4C_2 repeats impairs global translation and induces stress granules in an eIF2α phosphorylation-dependent manner. Thus, repeat expansions can trigger a feed-forward loop that drives RAN translation while impairing global translation and altering RNA metabolism (**Fig. 2-13**). In the context of data implicating RAN translation products in the pathogenesis of both FXTAS (5,40) and C9ALS/FTD (13-15,20,23,47), our findings support a model whereby an inefficient translation mechanism such as RAN translation might meaningfully contribute to neuronal dysfunction and death in disease.

The G_4C_2 repeat containing transcripts studied here are capped and polyadenylated linear mRNAs, but the repeat normally resides within an intron in *C9orf72.* The exact RNA species that undergoes RAN translation in C9ALS/FTD has not been determined empirically. G_4C_2 repeat expansions can trigger intron retention, altered transcription initiation, as well as premature transcription termination, all of which could generate repeat-containing linear mRNAs subject to both 5′ m^7G-capping and polyadenylation (9,51,52). Moreover, recent data suggest that mRNAs containing the repeat within a retained intron are trafficked to the cytoplasm more efficiently due to interactions with the RNA binding protein and nuclear export adapter SRSF1 (53). Given the strong cap-dependence we observe with our C9RAN reporters, our data argue that such events, even if rare, could significantly enhance RAN translation efficiency.

The requirements for C9RAN translation initiation closely mirror those previously described for CGG repeats in FXTAS, including initiation at near-cognate codons located 5′-proximal to the repeat in the most robustly-translated reading frames (GA for C9RAN and +1 FMRpolyG reading frame for CGG RAN) (36). However, here we show that altering initiation levels in one reading frame, by modifying the initiating CUG codon in the GA frame, also alters translation levels in the other two reading frames. In *in vitro* systems, removing the CUG codon enhances production in the GP reading frame but reduces translation in the GR reading frame. This increase in the GP frame suggests that it competes for initiation with the GA reading frame, and, based on the scanning model of translation initiation (25), that the CUG codon is located upstream of the GP initiation site. A UAG stop codon is positioned immediately upstream of the repeat in the

GP frame, suggesting GP initiation occurs within the repeat sequence itself. However, alternative possibilities, such as stop codon read-through or frame-shifting upstream of the stop codon remain to be explored. In contrast, the impaired production in the GR reading frame may be consistent with either a +2 or -1 nucleotide frameshift from the GA to the GR reading frame (54). This is intriguing given evidence linking G-quadruplex structures to -1 ribosomal frameshifting (55). Such translational frameshifts can occur at other nucleotide repeats (54), although frameshifts are not the dominant cause of RAN translation across different frames at other repeats (2,5). Given that each repeat's surrounding sequence context and RNA structure may confer different constraints on RAN initiation, generalizing these findings to RAN initiation at C_4G_2, CCG, CAG and CUG repeats may not be possible without further studies.

Interestingly, the mRNA reporters used in these studies can generate a GA product from only 3 or 35 G_4C_2 repeats, suggesting that the CUG codon, in good Kozak context, does not require an expanded repeat for use by the initiating ribosome. This is consistent with a report finding sparse, neuronal DPR inclusions in a cognitively normal 84 year old woman harboring 30 *C9orf72* G_4C_2 repeats (56). The absence of DPR accumulation in individuals with normal repeat sizes (<25) may indicate that these smaller species are rapidly cleared by cells, or that proper splicing and degradation of the intronic sequence containing the G_4C_2 repeat precludes its translation.

Our results indicate that the non-AUG initiation utilized by RAN translation is critical for its enhancement under stress conditions. We also observe that initiation at near cognate codons in the absence of any repeat sequence can be enhanced by some forms of ISR activation, suggesting that such codons may be sufficient for stress-

induced initiation. However, our RAN reporters consistently demonstrate more robust ISR activation than the near-cognate codon reporters lacking a repeat. This suggests that repeats enhance initiation during stress, possibly by creating a blockade for scanning 40S ribosomes that increases near-cognate codon initiation (44,57). There may thus be functional overlap between RAN translation and the translational mechanism used by single-stranded alphaviruses, where a hairpin structure within the coding region just 3′ to the start codon maintains active translation in the setting of eIF2α phosphorylation (58). Additionally, for at least two stress-enhanced RAN events (GP for G_4C_2 repeats and +2 FMRpolyA for CGG repeats) initiation likely occurs within the repeat in the absence of any near cognate codon (36). Thus, alternative modes of initiation may depend on the repeat structure to bypass canonical translational control mechanisms and respond to cellular stress pathways.

Recent data from ribosome profiling studies suggest that initiation at near cognate codons may be much more common than previously appreciated (59,60), and our findings delineate a specific role for near cognate codons in RAN translation. We also observe a role for eIF2 in RAN translation initiation, but the relationship between eIF2α-phosphorylation and start codon fidelity is complicated. ISR activation reduces mature GTP-eIF2-$tRNA_i^{Met}$ ternary complex availability (25). This can trigger leaky scanning, where uORFs in a poor Kozak sequence context are bypassed for initiation or re-initiation to allow for enhanced translation from the main ORF (30,31,34,61,62), suggesting that ISR activation enhances start codon fidelity (62). In contrast, and in agreement with our own findings, translation initiation at near-cognate codons can be enhanced by ISR activation (31,33,35). For example, translation of uORFs that initiate

from UUG and CUG start codons in the transcript encoding the ER stress chaperone protein BIP are maintained under stress conditions by utilizing non-canonical initiation factors (33). Several initiation factors have been implicated in ISR-resistant translation, including eIF2D, eIF2A, and eIF5B (33,58,63-65). Such factors can promote $tRNA_i^{Met}$ recruitment to the ribosome, as well as allow for initiation with elongator tRNAs, such as leucine-tRNA at CUG codons, when functional eIF2 is limited (33,58,63-66). Empirically identifying the specific tRNA and initiation factors required for RAN initiation at G_4C_2 and CGG repeats will be important moving forward.

In sum, our findings create a framework for better understanding how C9RAN translation occurs mechanistically, while also providing a potential explanation for how such an inefficient form of protein translational initiation can contribute to neurodegeneration. By identifying a central cellular pathway (eIF2α phosphorylation) as a trigger for selective enhancement of RAN translation, we are now well-positioned to explore how both exogenous and endogenous cellular stressors, including repetitive RNAs and RAN translation products themselves, can contribute to neurodegeneration. When coupled with the inherent toxicity of RAN-derived proteins and their resistance to degradation, this mechanism creates a mutually reinforcing system that feeds forward to enhance RAN translation and its toxic downstream consequences (**Fig. 2-13**) (22,24,48). Interventions which selectively intercede in this feed-forward loop are thus promising targets for future therapeutic development.

2.5 Experimental Procedures

Antibodies

The following antibodies were used for western blots as specified; 1:1,000 FLAG-M2 (mouse, Sigma F1804), 1:1,000 GAPDH 65C (mouse, Santa Cruz sc32233), 1:1,000 tubulin (mouse, DSHB 12G10), 1:1,000 PERK (rabbit, CST 3192S), 1:1,000 phospho-eIF2α (rabbit, Thermo MA5-15133), 1:1,000 BiP (rabbit, CST 3177S), 1:5,000 puromycin 12D10 (mouse, Millipore MABE434), in 5% non-fat dry milk (NFDM). LI-COR IRDye 680RD goat-anti-mouse secondary antibody (96-68070) was used 1:10,000 in 5% NFDM for GAPDH and tubulin loading controls. HRP-conjugated goat-anti-mouse (115-035-146) or goat-anti-rabbit (111-035-144) antibodies from Jackson ImmunoResearch Laboratories were used at 1:10,000 in 5% NFDM for all other western blots.

The following antibodies were used for immunocytochemistry as specified; 1:500 FLAG (rabbit, Cell Signaling #2368), 1:200 G3BP (mouse, BD Transduction Laboratories 23/G3BP), 1:200 FMRP (mouse, Covance 6B8), and 1:200 FMRP (rabbit, abcam17722) in 5% normal goat serum (NGS – HEK293 cells) or 2% bovine serum albumin (BSA – MEFs). Secondary goat-anti-mouse Alexa Fluor 488 (A-11029) and goat-anti-rabbit Alexa Fluor 555 (A-21428) from Life Technologies were applied at 1:500.

Plasmids

For C9RAN reporters, the 5′ end of *C9rof72* intron1 was PCR-amplified from human fibroblast DNA and inserted upstream of previously published GGG-NL-3xF in pcDNA3.1(+) via NheI (36). Native intronic near-cognate codons were mutated using Q5 Site-Directed Mutagenesis (SDM) Kit (NEB). An AUG start codon was then added to the

intronic sequence through a similar strategy. Annealed primers containing PSP cleavage sequence were ligated into an engineered AgeI site upstream of GGG-NL-3xF sequence. Q5 SDM was used to add one or two nucleotides immediately 5′ to the PSP site, to generate reporters for all three sense reading frames, and remove 3′ AgeI site resulting from PSP insertion. 70 G_4C_2 repeats were transferred from a published construct (67) immediately 3′ to the intronic sequence and 5′ to the PSP site, via engineered EagI and AscI sites. Repeat sequence contains a single C to A mutation resulting in an imperfect GGGGCA at repeat 13.

Near-cognate NLuc reporters were constructed by mutating the start codon of pcDNA3.1(+)/AUG-NLuc-3xF using the Q5 SDM (36). pcDNA3.1(+)/*ATF4* 5' leader-NLuc-3xF was constructed by subcloning a synthetic insert (Integrated DNA Technologies) into pcDNA3.1(+) via SacI/XbaI. This reporter was designed as previously published for a *ATF4* 5' leader-FLuc reporter (33) which harbors the complete 5' leader of the human *ATF4* including the annotated AUG start of ATF4 and the complete overlapping inhibitory uORF.

See **Tables 2-1 and 2-2** for primer and C9RAN NLuc reporter sequences.

RNA synthesis

RNAs were *in vitro* transcribed from linearized plasmids (36). pcDNA3.1(+) reporter plasmids were linearized with PspOMI; pCRII FLuc reporter with HindIII-HF. Linearized DNA was *in vitro* transcribed using HiScribe T7 High Yield RNA Synthesis Kit (NEB), with 3'-O-Me-m^7GpppG anti-reverse cap analog (ARCA) or ApppG cap (NEB)

added at eight times the concentration of GTP, for a capping efficiency of ~90%. 10 µL T7 reactions were carried out at 37ºC for 2 hours. Reactions were then treated with 2 U RNase-free DNaseI (NEB) for 15 minutes at 37ºC to remove DNA template, and then poly-adenylated with 5 U *E. coli* Poly-A Polymerase, 10X buffer, and 10 mM ATP (NEB) for 1 hour at 37ºC. Synthesized mRNAs were clean and concentrated with RNA Clean and Concentrator-25 Kit from Zymo Research. The size and quality of all synthesized mRNAs were verified on a denaturing formaldehyde RNA gel.

Rabbit reticulocyte lysate in vitro translation

mRNAs were *in vitro* translated with Flexi Rabbit Reticulocyte Lysate System from Promega, that is supplemented with calf liver tRNA (36). Reactions for luminescence assays were programmed with 3 nM mRNA and contained 30% RRL, 10 mM amino-acid mix minus methionine, 10 mM amino acid mix minus leucine, 0.5 mM MgOAc, 100 mM KCl, and 0.8 U/uL Murine RNAse Inhibitor (NEB), and incubated at 30ºC for 30 minutes before termination by incubation at 4ºC. Reactions were then diluted 1:7 in Glo Lysis Buffer (Promega), and incubated 1:1 for 5 minutes in the dark in opaque 96 well plates with NanoGlo Substrate freshly diluted 1:50 in NanoGlo Buffer (Promega). Luminescence was measured on a GloMax 96 Microplate Luminometer.

For comparison of translation levels between m^7G- and A-capped reporters, seven-molar excess of m^7G-capped and polyadenylated FLuc mRNA was added to reactions as this has been shown to better recapitulate the endogenous cap and poly(A) synergy (68). For eIF4E competition assays, 250 µM free ARCA (m^7G-cap) or A-cap was added to reaction mixture. For eIF4A inhibition, RRL mix was pre-incubated with 4

μM hippuristanol (a kind gift from Jerry Pelletier, McGill University), prior to addition of NLuc reporters and seven-molar excess FLuc mRNA.

Reactions for western blot assays were performed as above, except 50 ng mRNA was used. 10 μL reactions were mixed with 40 μL sample buffer and heated at 70^{o}C for 15 minutes, and 20 μL was run on a 12% polyacrylamide gel.

For precision protease (PSP) site cleavage, 4 μL RRL reaction was mixed either with 17.78 μM cycloheximide, 4 μL RNase free water, and either 2 U PSP (GE Health Sciences) or vehicle, and incubated for 30 minutes at 30^{o}C, prior to processing for luminescence or western blot analysis.

Cell lines, transfection, drug treatments, and analysis

HEK293, HEK293T, and HeLa cells were purchased from American Type Culture Collection (ATCC). WT and A/A MEFs were received from Randal Kaufman (Sanford Burnham Prebys Medical Discovery Institute).

For C9RAN luminescence assays, HEK293 cells were seeded in 96-well plates at 2x10^{4} cells/well and transfected 24 hours later at ~80% confluency. RNA transfections were performed with *Trans*IT-mRNA Transfection Kit from Mirus, per manufacturer's recommended protocol, with 90 ng reporter mRNA and 200 ng pGL4.13 FLuc control DNA added to each well in triplicate. C9RAN DNA transfections were performed in triplicate with FuGene HD at a 3:1 ratio to DNA, except where specified below, with 50 ng NLuc reporter DNA and 50 ng pGL4.13 FLuc reporter added per well. Cells were then lysed 24 hours post transfection with 60 μL Glo Lysis Buffer for 5 minutes at room temperature. 25 μL of lysate was mixed with NanoGlo Substrate

prepared as for RRL reactions, and 25 µL of ONE-Glo Luciferase Assay System (Promega), for 5 minutes in the dark, in opaque 96-well plates. Luminescence measurements were obtained as with RRL reactions. P-values were calculated using Student's t test.

For C9RAN reporter luminescence analysis following ISR activation, HEK293 cells were seeded and transfected as above for 19 hours, followed by 5 hour drug treatment. For CGG RAN reporter luminescence analysis, HEK293T cells were seeded at $1.5x10^4$ cells/96 well for 36 hours, then transfected with 4:1 Viafect:DNA for 1 hour, followed by 5 hour drug treatment. MEFs were seeded at $1x10^4$ cells/well for 24 hours, then transfected 2:1 with jetPRIME:DNA for 1 hour, followed by a 5 hour drug treatment. All cell types were lysed and luciferase activity measured as above. Drugs used: TG (Thermo), Sal003 (Sigma), TM (Sigma), and SA (Sigma).

For assessment of RAN translation in the presence of overexpressed eIF2α-S51D, HEK293 and 293T cells were grown as above. NLuc reporters were co-transfected 1:1 with pGL4.13 FLuc, and 1:10 with an effector plasmid (eIF2α WT or S51D) for 24 hours. Cells were lysed and luciferase activity was measured as above.

For C9RAN western blots, HEK293 cells were seeded in 12-well plates at $2x10^5$ cells/well and transfected 24 hours later at ~80% confluency with 500 ng NLuc reporter DNAs and 4:1 FuGene HD. Cells were lysed in 300 µL RIPA buffer with protease inhibitor 24 hours post transfection, for 30 minutes at 4^oC. Lysates were homogenized by passing through a 28G syringe, mixed with 6X sample buffer, and stored at -20^oC.

For western analysis of RAN reporters following stress induction, HEK293T cells were seeded at $7.5x10^4$ cells/well in 24-well plates. 24 hours post seeding, cells were

transfected with 250 ng NLuc reporter plasmids and 250 ng pGL4.13 FLuc plasmids with 3:1 FuGene HD for 18 hours, followed by 5 hour drug treatment. Cells were lysed as with C9RAN western blot transfections. Each genotype was run in triplicate on a 12% SDS-PAGE along with a standard curve for quantification of protein expression. Band intensities were measured using ImageJ, and quantified by extrapolating off the standard curve, and normalizing to alpha-tubulin.

HeLa cell lysate in vitro translation

In vitro translation extracts were prepared from cultured HeLa cells (37) (ATCC) maintained at 37°C/5% CO_2 in DMEM supplemented with 10% FBS and 1% non-essential amino acids. To prepare extracts, adherent cells were trypsinized, centrifuged, and washed in PBS. Cell pellets were resuspended in RNAse-free hypotonic buffer containing 10 mM HEPES-KOH (pH 7.6), 10 mM potassium acetate, 0.5 mM magnesium acetate, 5 mM DTT, and EDTA-free protease inhibitor cocktail (Roche). Cell pellets were incubated on ice for 20 minutes, mechanically disrupted by a 27G syringe, incubated for another 20 minutes on ice, and centrifuged at 10k *g* for 10 minutes at 4°C. The supernatant was removed, then brought to 4 μg/μl in additional hypotonic buffer. For *in vitro* translation reactions, lysates were supplemented to final concentrations of 20 mM HEPES-KOH (pH 7.6), 44 mM potassium acetate, 2.2 mM magnesium acetate, 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 (69). To this, *in vitro* transcribed reporter RNAs were

added to 4 nM. After incubation at 30°C for 30 minutes, luciferase assays were carried out as with RRL reactions.

Primary rat hippocampal neuron transfection

Rat hippocampi were harvested from postnatal day 0-2 pups, dissociated with L-cysteine-activated papain, and 60,000 neurons were plated per well on poly-D-lysine coated coverslips in neuronal growth media (NGM). Neurons were allowed to mature for 13 days *in vitro*, with half NGM media changes, supplemented with glial and cortical enriched media, every 2-3 days. On DIV13, neurons were transfected with 5 µg DNA and 10 µL Lipo2000 per well. 48 hours post transfection, neurons were lysed in 300 µL Glo Lysis Buffer for 5 minutes at room temperature. 80 µL of lysed cells were incubated with 80 µL freshly prepared NanoGlo Substrate in NanoGlo Buffer or ONE-Glo, and luminescence measured as with other assays.

Protein stability analysis

24 hours post RAN plasmids transfection, performed as above, HEK293 cells were treated with 10 µg/mL puromycin for 0, 6, and 24 hours. After each timepoint, cells were lysed in 60 µL Glo Lysis Buffer for 5 minutes at room temperature and stored at -20ºC. After all time points were collected, NLuc and FLuc activities was measured simultaneously.

Automated fluorescence microscopy imaging of primary neurons

Rat cortical primary neurons were harvested from E20 pups and cultured at 0.6 x 10^6 cells/mL *in vitro*. On DIV4, neurons were co-transfected with 0.1 µg pGW1-GFP, pGW1-(G_4C_2)x66-GFP, or pGW1-FMRP-(CGG)x100-GFP DNA and 0.1 µg pGW1-mApple with 2:1 Lipo2000 (Invitrogen, 52887). Beginning one day post transfection, neurons were reiteratively imaged with automated fluorescent microscopy for four to ten days (70,71). Image processing and fluorescent intensity measurements for GFP and mApple (n>30 neurons) were obtained for each timepoint using custom code written in Python or the ImageJ macro language. To assess the effects on ISR activation on RAN translation in neurons, cells with treated with 0.5, 1, or 2 µM TG following the first timepoint.

Monitoring translation by puromycin incorporation

Translation levels were assessed using the surface sensing of translation (SUnSET) method[50]. HEK293 cells were seeded at 1x10^5 cells/well in 24-well plates, and transfected 24 hours later with 250 ng CGG RAN reporters and 4:1 FuGene HD. 24 hours after transfection, cells were incubated with fresh media containing 10 µg/mL puromycin for 10 minutes at room temperature. Cells were then placed on ice and washed with ice-cold PBS, prior to lysis in 150 µL RIPA buffer containing protease inhibitor.

Stress granule analysis

HEK293 cells were seeded at 1x10^5 cells/well in 4-well chamber slides 24 hours prior to FuGene HD transfection of 250 ng DNA reporters. 24 hours post transfection,

cells were fixed with 4% paraformaldehyde (PFA) in PBS-MC for 15 minutes at room temperature, permeabilized with 0.1% triton-X in PBS-MC for 5 minutes at room temperature, blocked with 5% NGS, and incubated overnight with primary antibodies in 5% NGS at 4^{o}C in a humidity chamber. The following morning, cells were incubated with Alexa-Fluor secondary antibodies for 1 hour at room temperature in the dark. Coverslips were then applied to slides with ProLong Gold Antifade Mountant with DAPI. 3-5 fields per condition were imaged at 20x1.6 magnification with Olympus IX71 fluorescent microscope and Slidebook 5.5 software.

WT and A/A mutant MEFs were seeded at 1x10^{5} cells/well for 24 hours, then transfected with 500 ng NLuc reporters and 2:1 jetPRIME for 24 hours. Cells were fixed and permeabilized as above, blocked with 2% BSA for 20 minutes at room temperature, and incubated overnight with primary antibodies in 2% BSA at 4^{o}C. Secondary antibodies were applied the following morning for 1 hour at room temperature, in the dark. 10-20 fields per condition were taken at 20x1.6 magnification, as above.

For stress granule analysis, signals for each channel were normalized prior to quantification. For HEK239 cells, >450 cells were counted for each genotype (>70 transfected cells/genotype). For MEFs, >370 cells were counted for each genotype (>40 transfected cells/genotype). Quantification was performed using ImageJ analysis. P values were calculated using Fisher's exact test.

Statistical methods

Statistical analysis was performed using GraphPad Prism7. For comparison of NLuc reporter luciferase activity, one-way ANOVAs were performed to confirm statistical

difference between control and experimental groups. Post-hoc Student's t tests were then performed with Bonferroni correction for multiple comparisons and Welch's correction for unequal variance. Fisher's exact tests were used for immunocytochemistry experiments, to determine if there was a statistical difference between the proportion of control or RAN transfected cells that contained stress granules.

Data Availability

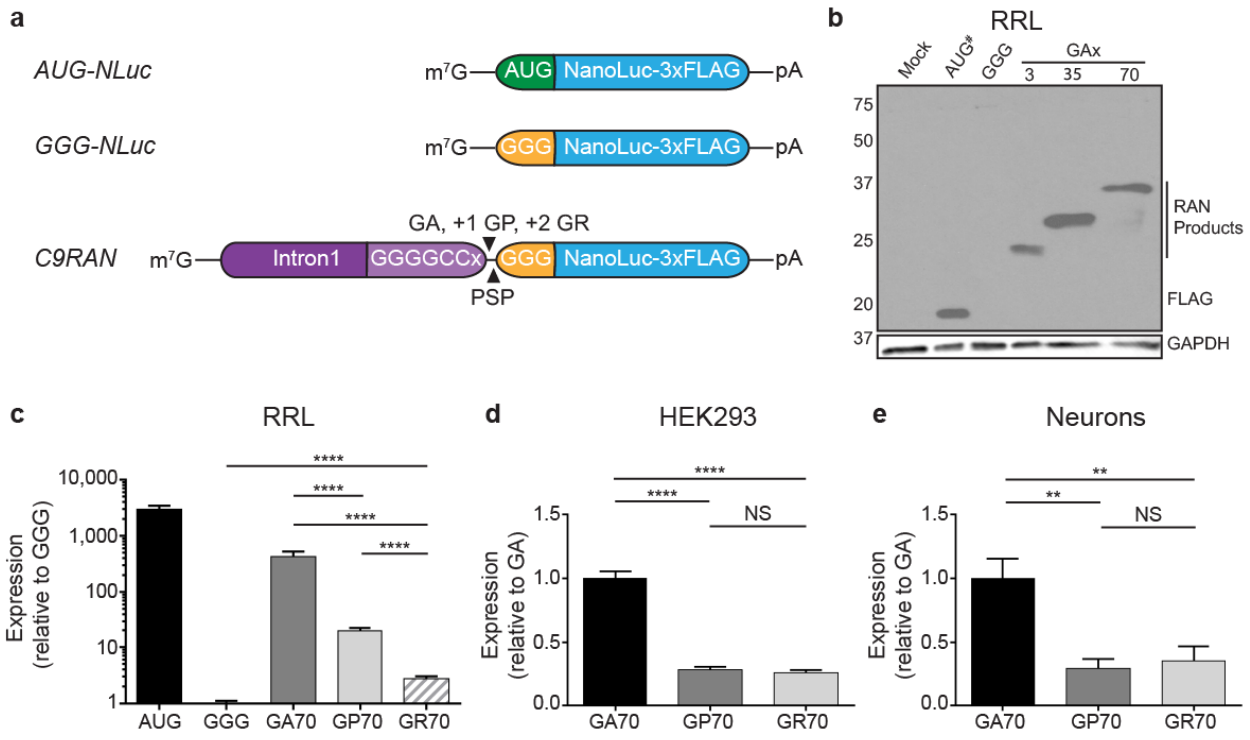
All relevant data are available from the authors upon request.

2.6 Chapter-specific acknowledgements

The authors would like to thank members of the Todd lab for input on this manuscript. They also thank Jerry Pelletier (McGill University) for kindly providing hippuristanol, and Randal Kaufman (Sanford Burnham Prebys Medical Discovery Institute) for supplying them with WT and eIF2α S51 A/A MEFs. This work was funded by grants from the VA BLRD (1I21BX001841 and 1I01BX003231), the NIH (R01NS099280 and R01NS086810), and the Michigan Alzheimer's Disease Center and Protein Folding Disease Initiative to PKT. KMG, BNF, and AEL were supported by NIH T32GM007315. KMG was further supported by NIH F31NS100302 and AEL by NIH F30NS098571. MRG was supported by NIH T32NS007222-35S1. MGK was supported by the NIH F32NS089124. SJB and BNF were supported by the NIH (R01-NS097542, 1P30AG053760-01) and Ann Arbor Active Against ALS.

2.7 Figures

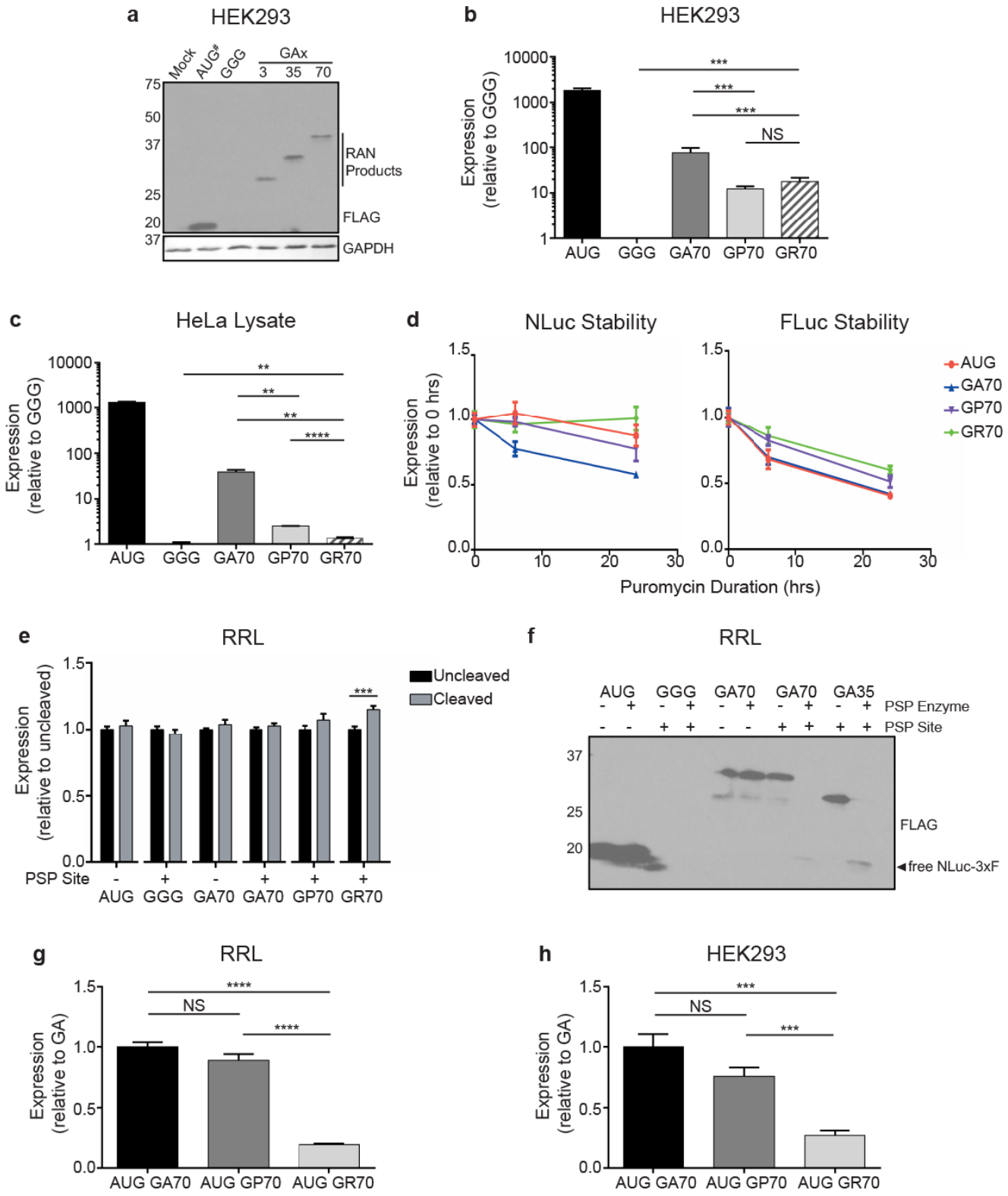
Figure 2-1 C9RAN translation-specific reporters reveal differential expression across reading frames



(a) Schematic of reporters used in this study. C9RAN translation reporters were designed by placing the *C9orf72* intron 1 sequence, including 70 G_4C_2 repeats, upstream of a start codon mutant NanoLucifearse (NLuc) and a C-terminal 3xFLAG-tag, in separate reading frames relative to the repeat. (b) Anti-FLAG western blot of control and C9RAN translation reporters expressed in rabbit reticulocyte lysate (RRL). GAPDH is used as a loading control. To prevent over-exposure, the AUG-NLuc control reaction was diluted 1:5 in sample buffer (indicated by $^{\#}$). (c-e) Relative expression from C9RAN NLuc reporters (c) normalized to GGG-NLuc in RRL (n=15), or normalized to GA-NLuc in (d) HEK293 cells (n=18), and (e) primary rat hippocampal neurons (n=9). GA, glycine-alanine. GP, glycine-proline. GR, glycine-arginine. PSP, precision protease

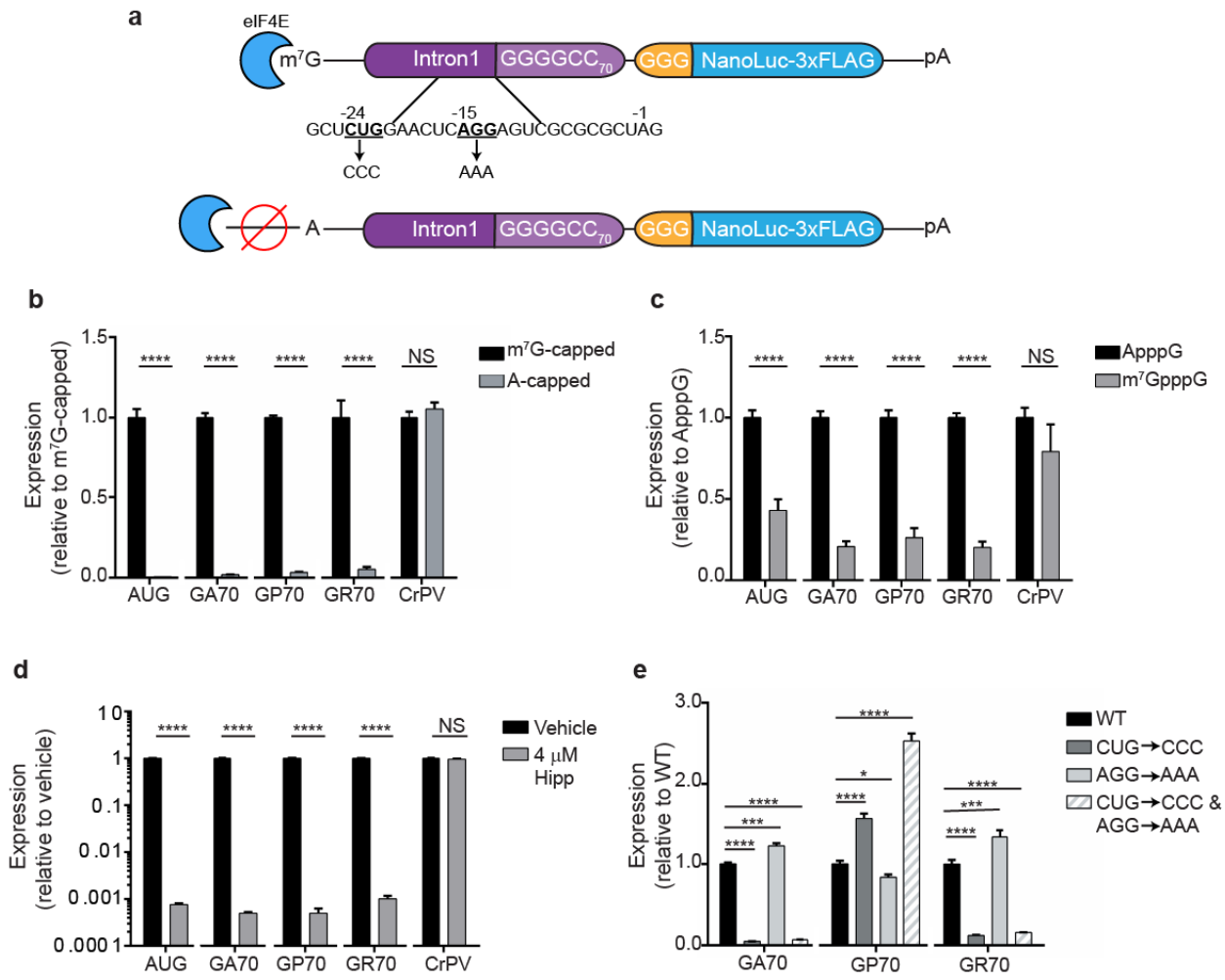
cleavage site. Graph in (c) represents mean ± SD. Graphs in (d) and (e) represent mean ± SEM. Two-tailed Student's t test with Bonferroni and Welch's correction, $**p < 0.01$; $****p<0.0001$.

Figure 2-2 C9RAN translation reporter system allows for quantitative assessment of RAN translation in all three sense reading frames



(a) Anti-FLAG western blot analysis of control and GA C9RAN translation reporters in HEK293 cells. #One-tenth AUG reporter was transfected into cells to prevent over-exposure. (b) Expression of control and C9RAN reporters in mRNA transfected HEK293 cells, n=6. (c) Representative expression of control and C9RAN reporter mRNAs in HeLa cell lysate, n=3. (d) The stability of control and C9RAN reporter proteins was assessed in transfected HEK293 cells by treating cells with 10 µg/mL puromycin and measuring reporter activity at 0, 6, and 24 hours later, n=9 (0 and 6 hours), n=12 (24 hours). (e)The hindrance of the DPR fusion on NLuc activity was assessed in RRL by incubating completed reactions with PSP enzyme to cleave DPRs from NLuc, n = 6-9. (f) Anti-FLAG Western blot to confirm PSP cleavage of C9RAN fusion proteins in RRL. (g) Expression from AUG-driven reporters for each sense reading frame, relative to AUG-GA70, in RRL, n=6 and (h) HEK293 cells, n=6. RRL, rabbit reticulocyte lysate. Graphs in (b) and (c) represent mean ± SD. Remaining graphs show mean ± SEM. Two-tailed Student's t test with Welch's correction, $**p < 0.01$; $***p < 0.001$; ****p<0.0001.

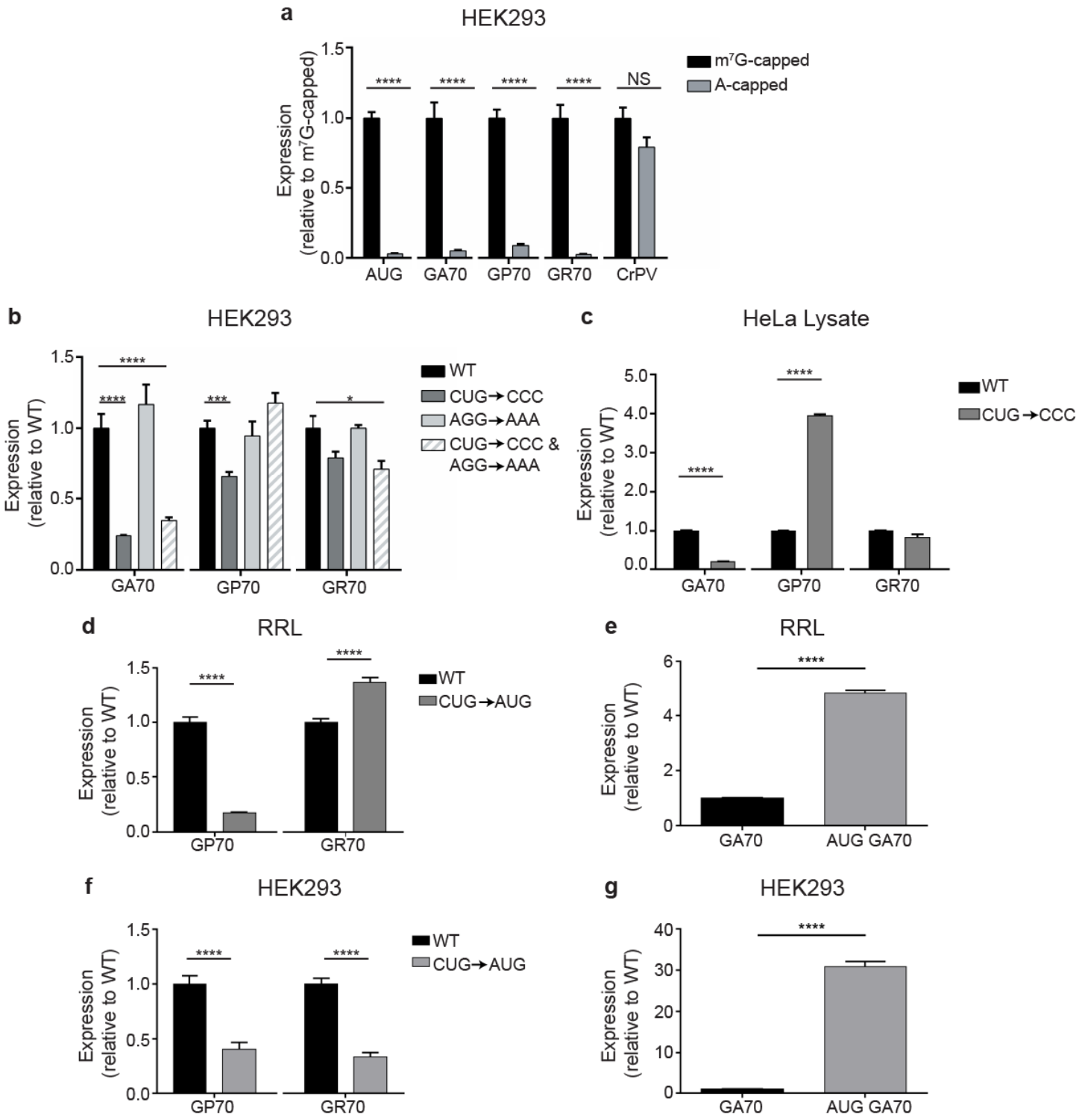
Figure 2-3 C9RAN is cap- and eIF4A-dependent and can initiate at a near-cognate start codon



(a) Schematic of 5′-cap C9RAN reporter mRNAs and near-cognate start codon mutations. (b) Expression of m^{7}G-capped and A-capped control and C9RAN reporters in RRL, n=6. (c) Expression of control and C9RAN reporters in RRL when excess free m^{7}G (250 μM) or equimolar A-cap was added to inhibit eIF4E *in trans*, n=6. (d) Expression of reporter mRNAs when eIF4A, the canonical helicase required for ribosome scanning during initiation, is inhibited with 4 μM hippuristanol (Hipp), n=6. (e) Mutational analysis of near-cognate start codons upstream of the repeat in the GA

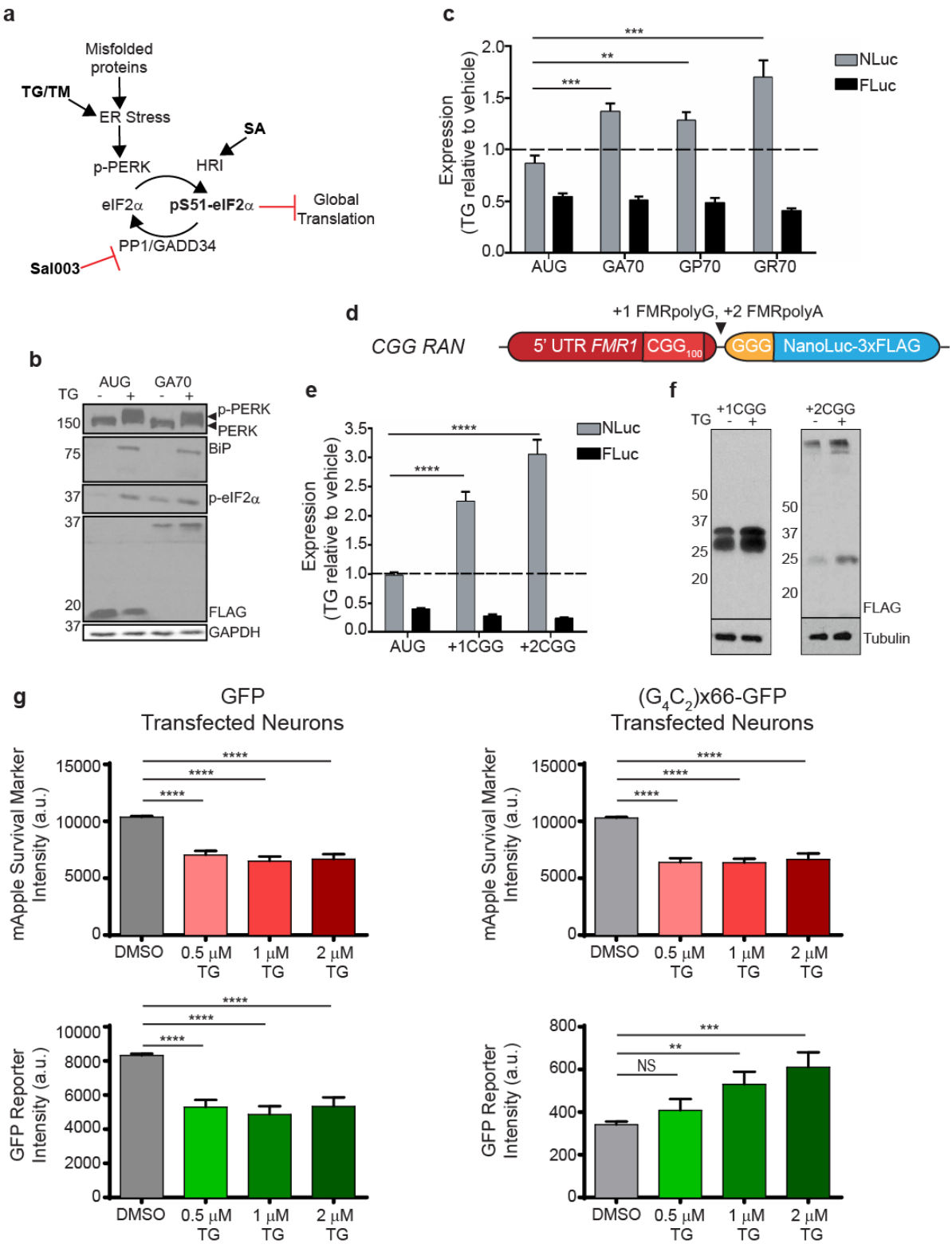
frame as depicted in (a), n=6. GA, glycine-alanine. GP, glycine-proline. GR, glycine-arginine. CrPV, cricket paralysis virus. Graphs represent mean ± SEM. Two-tailed Student's t test with Bonferroni and Welch's correction, $*p < 0.05$; $***p < 0.001$; $****p<0.0001$.

Figure 2-4 C9RAN translation in HEK293 cells is cap-dependent and can initiate at a near-cognate start codon



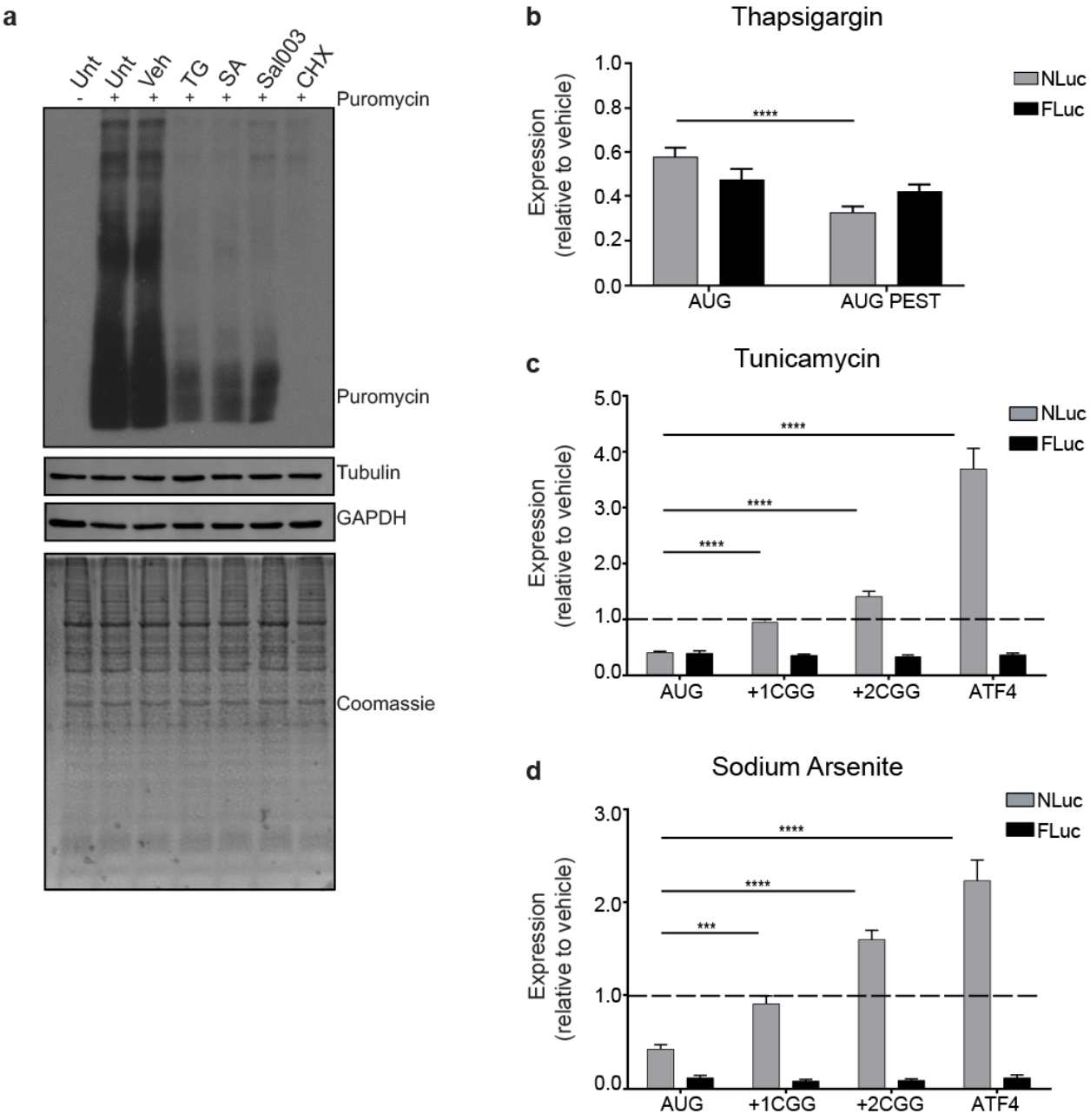
(a) Expression of m^7G-capped and A-capped control and C9RAN mRNA reporters in HEK293 cells, n=9. (b) Expression of all three sense C9RAN mRNAs in HEK293 cells following mutation of near-cognate codons in GA frame, n=6. (c) Expression of sense C9RAN mRNAs in HeLa cell lysate with or without CUG codon mutated to CCC, n=11. (d) Expression of GP and GR-NLuc reporters in RRL from constructs with CUG codon mutated to AUG, relative to WT sequence in RRL, n=6. (e) Insertion of an AUG codon upstream of the repeat in the GA frame enhances GA-NLuc expression in RRL, n=6. (f) Mutating CUG codon to AUG decreases expression of GP and GR-NLuc reporters in HEK293 cells, n=6. (g) Insertion of an AUG codon upstream of the repeat in the GA frame enhances GA-NLuc expression in HEK293 cells, n=24. Graphs represent mean ± SEM. Two-tailed Student's t test with Welch's correction, $*p < 0.05$; $***p < 0.001$; $****p<0.0001$.

Figure 2-5 RAN translation is selectively activated by the integrated stress response



(a) Schematic of the integrated stress response pathway. PERK, endoplasmic reticulum ER-resident kinase. HRI, heme-regulated inhibitor kinase. SA, sodium arsenite. TG, thapsigargin. TM, tunicamycin. (b) Western blot analysis of the ER stress pathway and C9RAN reporter levels in HEK293 cells after treatment with 2 μM TG. GAPDH was used as a loading control. (c) Expression of control and C9RAN NLuc reporters and co-transfected FLuc in HEK293 cells treated with 2 μM TG, n=9. (d) Schematic of the previously published[36] +1 and +2 CGG RAN translation NLuc reporters. (e-f) Expression of control and CGG RAN translation reporters and co-transfected FLuc in HEK293Tcells treated with 2 μM TG analyzed by (e) luciferase activity, n=9, and (f) anti-FLAG western blot. Tubulin was used as a loading control. (g) Fluorescence intensity of mApple and co-transfected GFP (left) or (G_4C_2)x66 -GFP (right) in primary rat cortical neurons, imaged with automated fluorescent microscopy three days after treatment with 0.5, 1, or 2 μM TG, n>30. Graphs represent mean ± SEM. Two-tailed Student's t test with Bonferroni and Welch's correction, $**p < 0.01$; $***p < 0.001$; $****p<0.0001$.

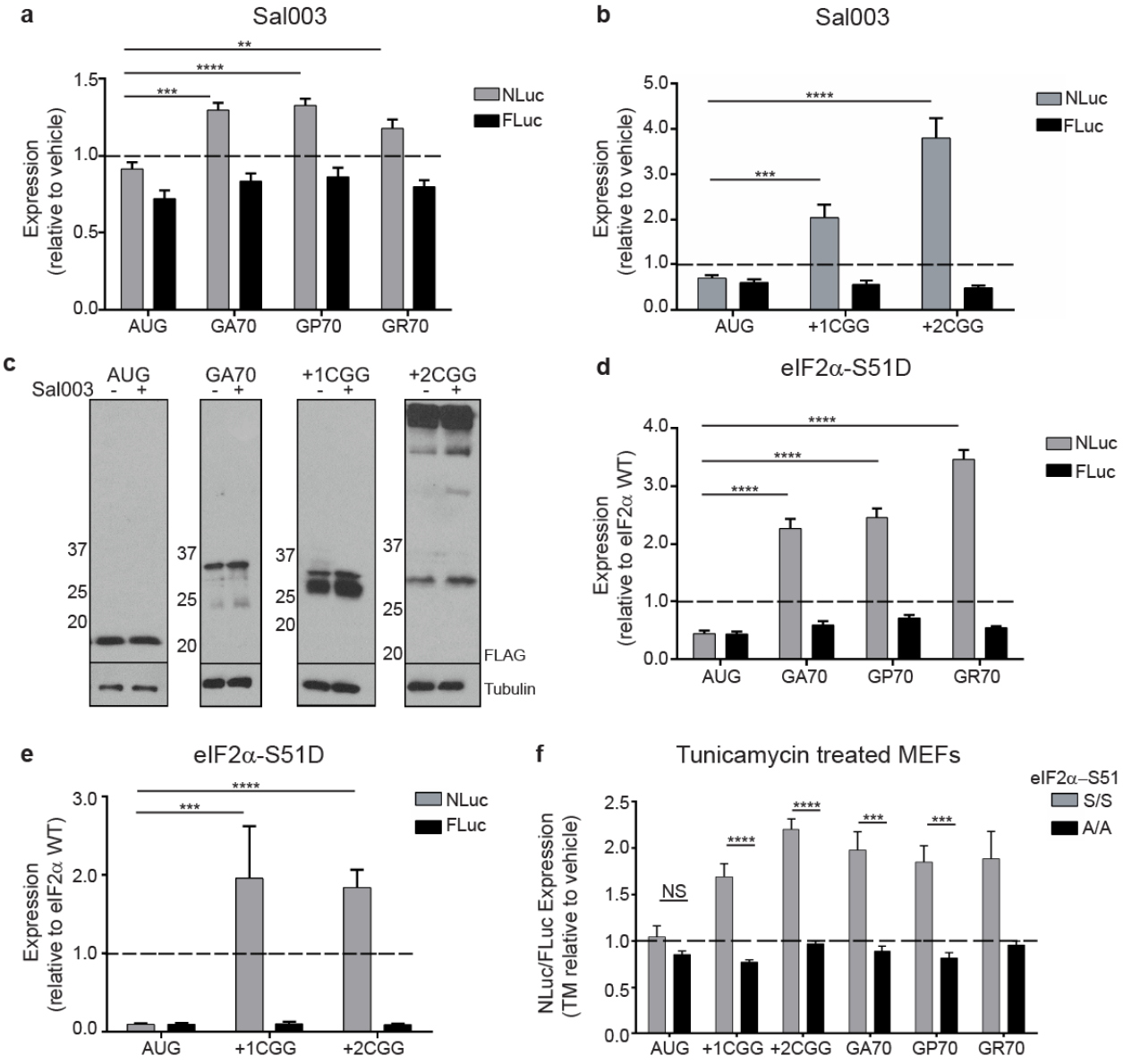
Figure 2-6 CGG RAN translation in multiple reading frames is refractory to translation attenuation during ER and oxidative stress



(a) Anti-puromycin western blot of cells treated with various cell stress inducers and 10 μg/mL puromycin to monitor global translation activity. Tubulin, GAPDH, and Coomassie stain were used as loading controls. Unt, untreated. Veh, vehicle. TG, 2μM

thapsigargin. SA, 10μM sodium arsenite. Sal003, 20μM Salubrinol. CHX, 100μg/mL cycloheximide. (b) Destabilization of AUG-NLuc reporter with PEST tag results in greater decrease in AUG-NLuc expression with TG treatment. (c-d) Expression of control and CGG RAN NLuc reporters in transfected HEK293T cells when treated with (c) 2.5 μg/mL Tunicamycin, n=9 and (d) 20 μM sodium arsenite, n=9, for 5 hours. FLuc was co-expressed as an internal control with each NLuc reporter. AUG and ATF4 reporters serve as reporters that are attenuated and stimulated, respectively, during activation of the ISR. Graphs represent mean ± SEM. Two-tailed Student's t test with Welch's correction, ∗∗∗$p < 0.001$; ∗∗∗∗$p<0.0001$.

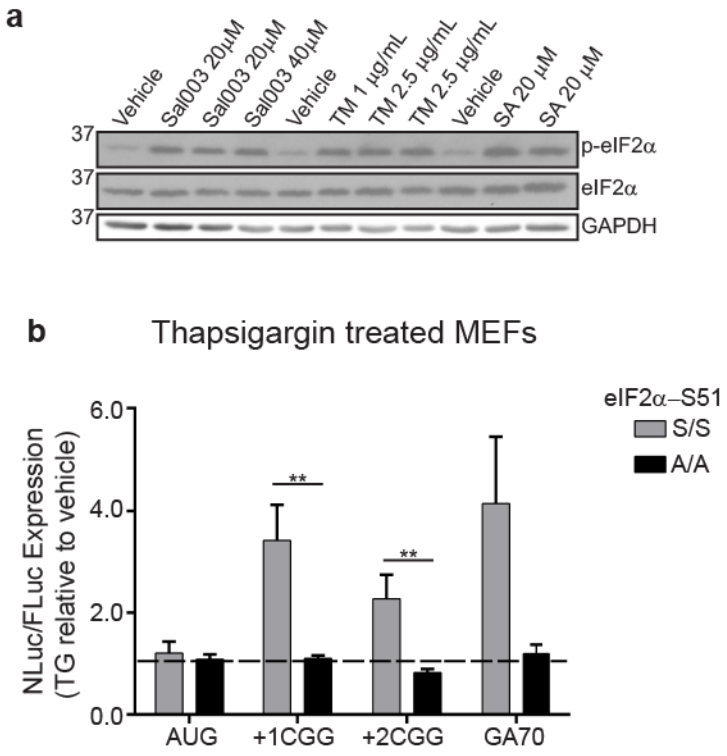
Figure 2-7 RAN translation is resistant to eIF2α phosphorylation



(a) Expression of control and C9RAN NLuc reporters and co-transfected FLuc in HEK293 cells treated with 40 μM Sal003, n=6. (b) Expression of control and CGG RAN NLuc reporters and co-transfected FLuc in HEK293T cells treated with 20μM Sal003, n=9. (c) Western blot analysis of control, GA70 RAN and CGG RAN NLuc reporters in HEK293T cells treated with 20 μM Sal003. (d) Expression of control and C9RAN NLuc reporters and co-transfected FLuc in HEK293 cells transfected with either WT or S51D

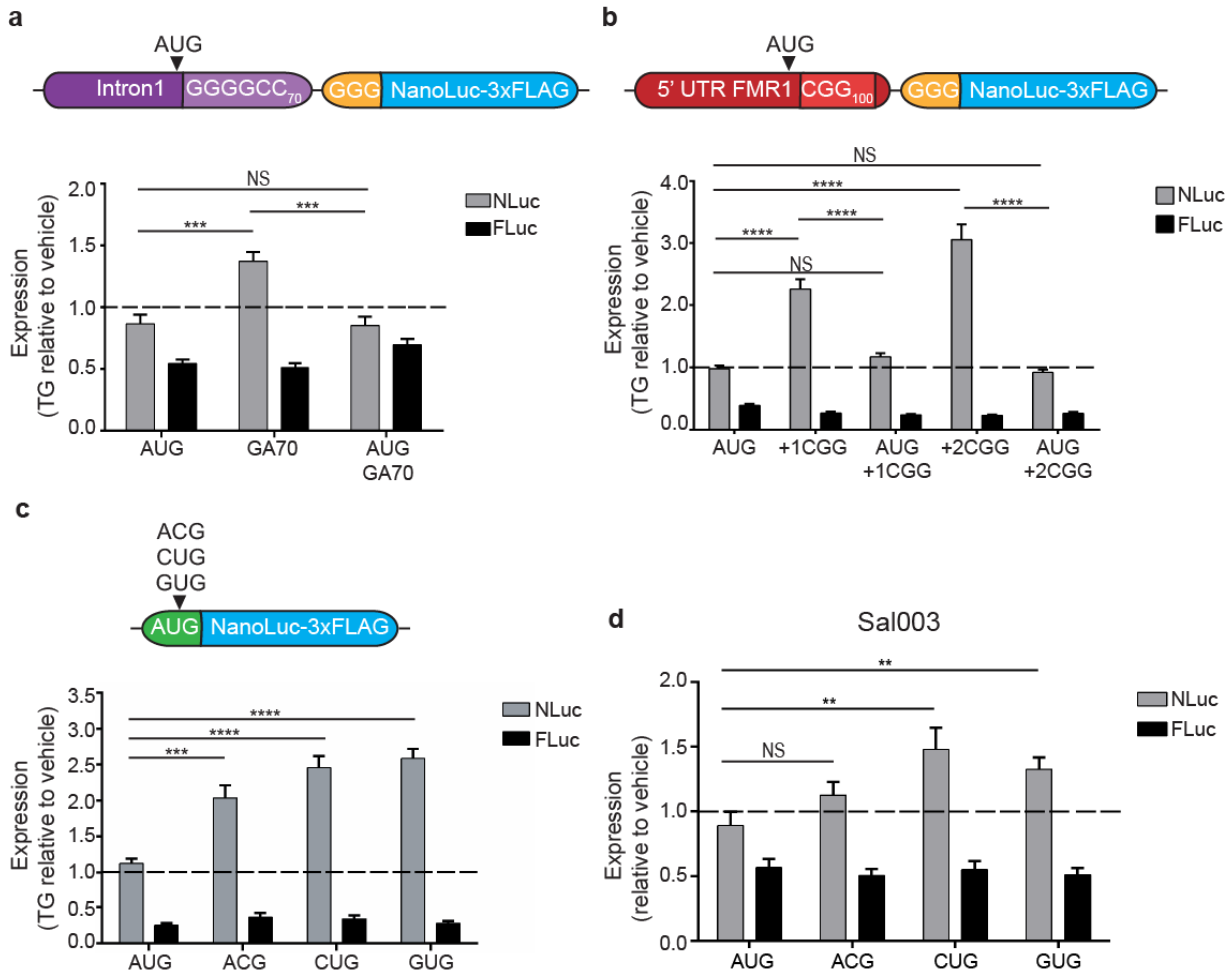
(phosphomimetic) eIF2α, n=6. (e) Expression of control and CGG RAN NLuc reporters and co-transfected FLuc in HEK293T cells transfected with either WT or S51D (phosphomimetic) eIF2α, n=9-15. (f) Expression of control, CGG, and C9RAN NLuc reporters normalized to co-transfected FLuc in WT eIF2α-S51 (S/S) and non-phosphorylatable homozygous eIF2α-S51 A/A mutant mouse embryonic fibroblasts (MEFs) following treatment with 1 μg/mL tunicamycin (TM), n=6-9. Graphs represent mean ± SEM. Two-tailed Student's t test with Bonferroni and Welch's correction, $**p < 0.01$; $***p < 0.001$; $****p<0.0001$.

Figure 2-8 Thapsigargin-induced enhancement of C9 and CGG RAN translation requires phosphorylated eIF2α



(a) Western blot showing increased phosphorylation of eIF2α in HEK293 cells following treatment with Sal003 (20 or 40 μM), TM (1 or 2.5 μg/mL), and SA (20μM). (b) Expression of control, CGG, and C9RAN NLuc reporters normalized to co-transfected FLuc in WT and eIF2α-S51A/A homozygous mutant MEFs following treatment with 1 μM thapsigargin (TG), n=6-9. Graph represent mean ± SEM. Two-tailed Student's t test with Bonferroni and Welch's correction, $**p < 0.01$

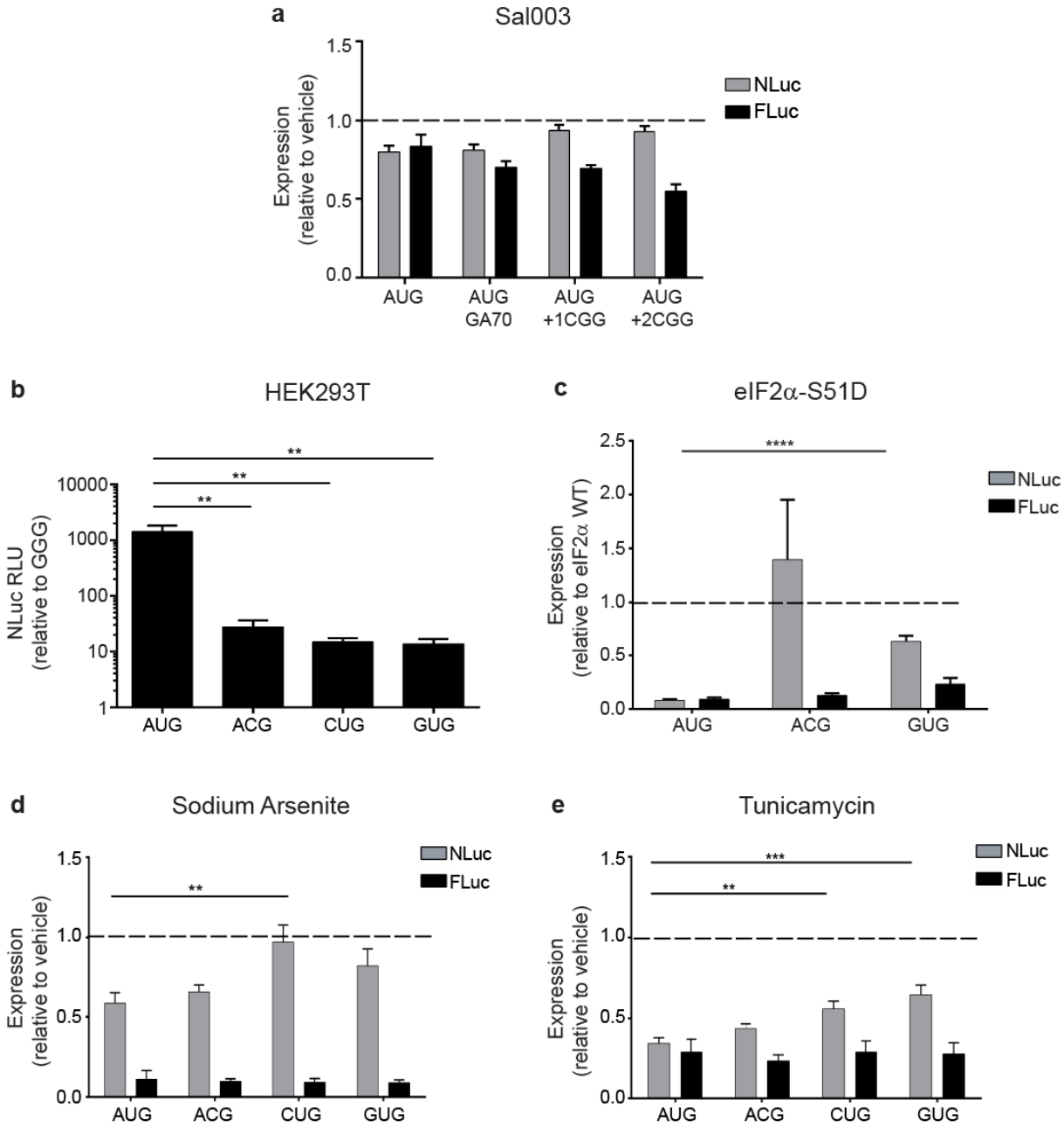
Figure 2-9 Near-cognate codons are sufficient to allow for stress-induced translation



(a) *Top:* Schematic of C9RAN reporter with AUG codon inserted upstream of repeat. *Bottom:* Expression of control, C9RAN NLuc, and AUG-driven C9RAN reporters and co-transfected FLuc in HEK293 cells treated with 2 μM TG, n=6-9. (b) *Top:* Schematic of CGG RAN reporter with a near-cognate codon mutated to AUG. *Bottom:* Expression of control, CGG RAN NLuc, and AUG-driven CGG RAN reporters and co-transfected FLuc in HEK293T cells treated with 2 μM TG, n=9. (c) *Top:* Schematic showing location of near-cognate codon substitutions made in AUG-NLuc. *Bottom:* Expression of control and near-cognate codon NLuc reporters and co-transfected FLuc in HEK293T cells

treated with 2 μM TG, n=9. (d) Expression of control and near-cognate codon NLuc reporters and co-transfected FLuc in HEK293T cells treated with 20 μM Sal003 for 5 hours, n=12. Graphs represent mean ± SEM. Two-tailed Student's t test with Bonferroni and Welch's correction, $**p < 0.01$; $***p < 0.001$; $****p<0.0001$.

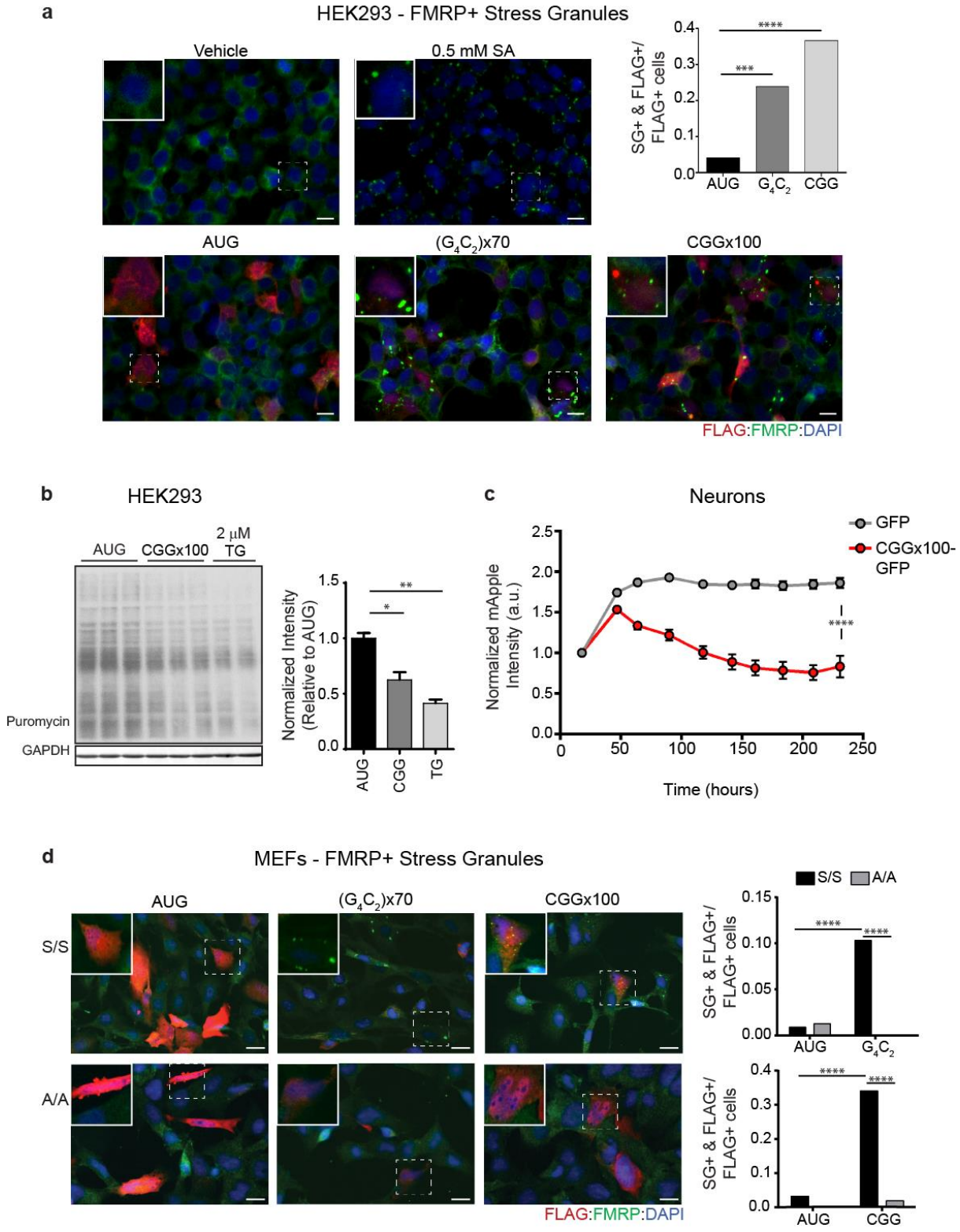
Figure 2-10 Initiation at near-cognate codons is refractory to multiple stress stimuli



(a) Expression of the control AUG-NLuc and AUG-initiated reporters harboring 100 CGG or 70 G_4C_2 repeats in multiple reading frames in HEK293 cells when treated with 20uM Sal003, for 5 hours, n=6-12. (b) Expression of NLuc reporters with varying start

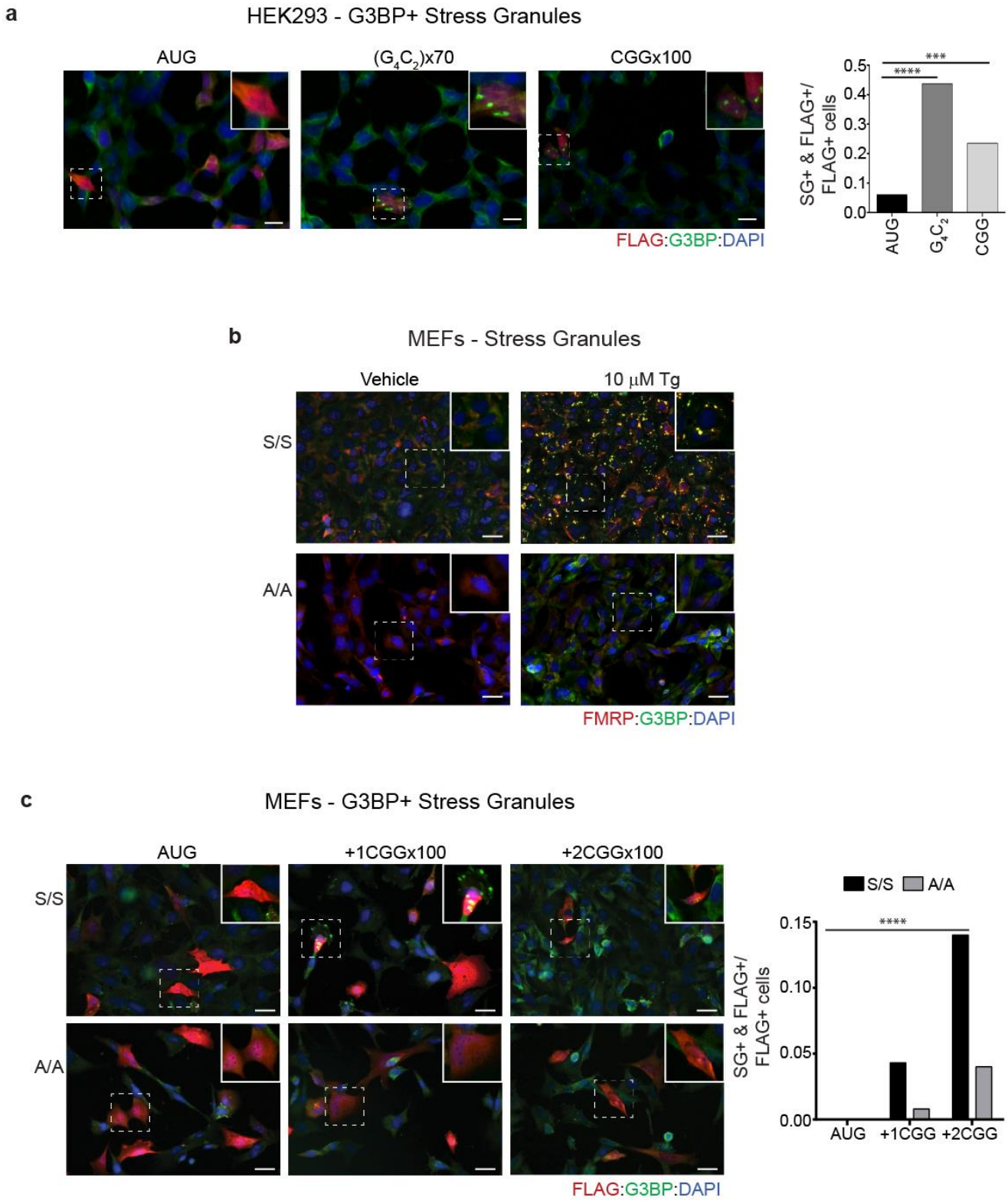
codon mutations in HEK293T cells relative to the negative control GGG-NLuc, n=9. (c-d) Response of the AUG-NLuc and near cognate-NLuc reporters, co-transfected with the internal FLuc control, in HEK293T cells (c) co-transfected with either WT or S51D (phosphomimetic) eIF2α, n=12-15, or treated with (d) 2.5 μg/ml tunicamycin, n=9, or (e) 20 μM sodium arsenite, n=9, for 5 hours. Graphs represent mean ± SEM. Two-tailed Student's t test with Welch's correction, $**p < 0.01$; $***p < 0.001$, $****p<0.0001$.

Figure 2-11 CGG and G_4C_2 repeat expansions induce phosphorylated-eIF2α dependent stress granules



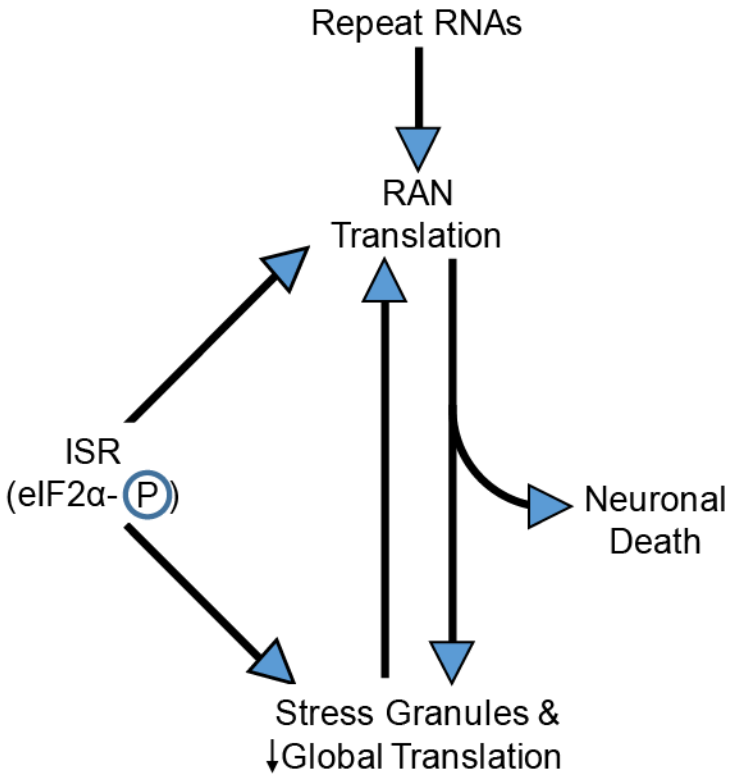
(a) *Top left:* Immunofluorescent images of HEK239 cells treated with vehicle or 0.5M SA. *Bottom:* Immunofluorescent images of HEK239 cells expressing control, (G_4C_2)x70, or CGGx100 reporters, scale bar=100 µm. *Top right:* Quantification of the proportion of FLAG-positive cells with FMRP-positive stress granules (SGs) for each genotype, n>70. (b) Western blot and quantification of puromycin incorporation in cells transfected with control AUG-NLuc or CGGx100 reporter, or treated with 2 µM TG as a positive control. GAPDH is used as a loading control. Graph represents mean ± SEM. (c) mApple fluorescent intensity in primary rat cortical neurons co-transfected with GFP or CGGx100-GFP, longitudinally imaged with automated fluorescent microscopy for ten days following transfection, n>68. Graph represents mean ± 95% confidence interval. (d) *Left:* Immunofluorescent images of WT eIF2α- S51 S/S and eIF2α- S51 A/A MEFs expressing control, (G_4C_2)x70, or CGGx100 reporters, scale bar=100 µm. *Right:* Quantification of the proportion of FLAG-positive cells with FMRP-positive SGs for each genotype, n>40. FLAG marks reporter expressing cells, FMRP mark SGs. For (a) and (d), Fisher's exact test, ∗∗∗$p < 0.001$; ∗∗∗∗$p < 0.0001$. For (b) and (c), two-tailed Student's t test with Bonferroni and Welch's correction, ∗$p < 0.05$; ∗∗$p < 0.01$; ∗∗∗∗$p<0.0001$.

Figure 2-12 CGG and G4C2 repeats induce G3BP-positive stress granules in a phosphorylated-eIF2α dependent manner



(a) *Left*: Immunofluorescent images of HEK293 cells expressing control, $(G_4C_2)_{70}$, or $(CGG)_{100}$ reporters, scale bar=100 µm. *Right:* Quantification of the proportion of FLAG-positive cells with G3BP-positive stress granules (SGs) for each genotype, n>45. (b) Immunofluorescent images of WT and eIF2α- S51A/A MEFs treated with vehicle or 10 µM TG for 3 hours, scale bar=100 µm. (c) *Left:* Immunofluorescent images of WT and eIF2α-S51A/A MEFs expressing control, $+1(CGG)_{100}$, or $+2(CGG)_{100}$ RAN reporters, scale bar=100 µm. *Right:* Quantification of the proportion of FLAG-positive cells with G3BP-positive SGs for each genotype, n>40. FLAG marks reporter expressing cells, G3BP mark SGs. Fisher's exact test, ***p < 0.001.; ****p<0.0001.

Figure 2-13 Working model for how a feed-forward loop activates RAN translation and cellular stress pathways



Repeat expansions trigger RAN translation. RAN proteins or the repeat RNAs themselves then elicit stress granules and suppress global protein synthesis in a phosphorylated-eIF2α-dependent fashion. Activation of the integrated stress response (ISR) and phosphorylation of eIF2α, either by the repeat RNAs or RAN proteins directly or through exogenous cellular stress can further trigger stress granule formation and suppress global translation while selectively enhancing RAN translation. This creates a feed-forward loop that can contribute to neuronal dysfunction and death.

2.8 Tables

Table 2-1. Primers for reporter generation

Reporter	Forward Primer Sequence	Reverse Primer Sequence
PSP-GGG-NL-3xFLAG	CCGGTCTCGAGGTCCTCTTCCAGGGACCCA	CCGGTGGGTCCCTGGAAGAGGACCTCGAGA
GA frame PSP site	CCAGGGACCCGATGGGGTCTTCAC	AAGAGGACCTCGAGACCG
GP frame PSP site	CCTCGAGGTCCTCTTCCAGGGACCCGATGG	ACCGGTGGGCGCGCCCGG
GR frame PSP site	CCCTCGAGGTCCTCTTCCAGGACCCGATGG	
Intron1 CTG-CCC	GTAGCAAGCTCCCGAACTCAGGAGTCGC	AGGCTGCGGTTGTTTCCC
Intron1 AGG-AAA	TCTGGAACCAAAAGTCGCGCGC	GCTTGCTACAGGCTGCGG
Intron1 CTG-CCC and AGG-AAA	CTAAAAGTCGCGCGCTAGCGGCC	TTCGGGAGCTTGCTACAGGCTGCGTTG
Intron CTG-ATG	GTAGCAAGCTATGGAACTCAGG	AGGCTGCGGTTGTTTCCC
Intron1 AUG	ATGAGTCGCGCGCTATCTA	CCTGAGTTCCAGAGCTTG
AUG-GA	CTAGCTAACTAACACCATGGC	GGCCGCCATGGTGTTAGTTAG
AUG-GP	CTAGCTAACTAACACCATGGGGC	GGCCGCCCCATGGTGTTAGTTAG
AUG-GR	CTAGCTAACTAACACCATGGGC	GGCCGCCCATGGTGTTAGTTAG
CUG-NLuc	ACCCTGGTCTTCACACTCGAAGATTTC	GGCTTATTTACCAACAGTACCGGATTG
GUG-NLuc	ACCGTGGTCTTCACACTCGAAGATTTC	
ACG-NLuc	ACCACGGTCTTCACACTCGAAGATTTC	

Table 2-2. C9RAN construct sequences

Construct	Sequence
Nhe1-Intron1-GA70-PSP-GGG-NLuc-3xFLAG-PspOMI	GCTAGCGTGTGTGTTTTTGTTTTTCCCACCCTCTCTCCCCACTA CTTGCTCTCACAGTACTCGCTGAGGGTGAACAAGAAAAGACCT GATAAAGATTAACCAGAAGAAAACAAGGAGGGAAACAACCGCA GCCTGTAGCAAGCTCTGGAACTCAGGAGTCGCGCGCTAGCGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CAGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGTCGTGG AAGGGTGGGCGCGCCcACCGGTCTCGAGGTCCTCTTCCAGGG ACCCGATGGGGTCTTCACACTCGAAGATTTCGTTGGGGACTGG CGACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACAG GGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTAA CTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGAA GATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGGC GACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTACC CTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCAC ACTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTTC GGACGGCCcTATGAAGGCATCGCCGTGTTCGACGGCAAAAAG ATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATCG ACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAG TAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCA TTCTGGCGGACTACAAAGACCATGACGGTGATTATAAAGATCA TGACATCGATTACAAGGATGACGATGACAAGTAAGGCCGCGAC TCGAGAGGGCCC
Nhe1-Intron1-GP70-PSP-GGG-NLuc-3xFLAG-PspOMI	GCTAGCGTGTGTGTTTTTGTTTTTCCCACCCTCTCTCCCCACTA CTTGCTCTCACAGTACTCGCTGAGGGTGAACAAGAAAAGACCT GATAAAGATTAACCAGAAGAAAACAAGGAGGGAAACAACCGCA GCCTGTAGCAAGCTCTGGAACTCAGGAGTCGCGCGCTAGCGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CAGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG

	CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGTCGTGG AAGGGTGGGCGCGCCcACCGGTcCTCGAGGTCCTCTTCCAGG GACCCGATGGGGTCTTCACACTCGAAGATTTCGTTGGGGACTG GCGACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACA GGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTA ACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGA AGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGG CGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTAC CCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCA CACTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTT CGGACGGCCcTATGAAGGCATCGCCGTGTTCGACGGCAAAAA GATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATC GACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGA GTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGC ATTCTGGCGGACTACAAAGACCATGACGGTGATTATAAAGATC ATGACATCGATTACAAGGATGACGATGACAAGTAAGGCCGCGA CTCGAGAGGGCCC
Nhe1-Intron1-GR70-PSP-GGG-NLuc-3xFLAG-PspOMI	GCTAGCGTGTGTGTTTTTGTTTTTCCCACCCTCTCTCCCCACTA CTTGCTCTCACAGTACTCGCTGAGGGTGAACAAGAAAAGACCT GATAAAGATTAACCAGAAGAAAACAAGGAGGGAAACAACCGCA GCCTGTAGCAAGCTCTGGAACTCAGGAGTCGCGCGCTAGCGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CAGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGTCGTGG AAGGGTGGGCGCGCCcACCGGTccCTCGAGGTCCTCTTCCAGG GACCCGATGGGGTCTTCACACTCGAAGATTTCGTTGGGGACTG GCGACAGACAGCCGGCTACAACCTGGACCAAGTCCTTGAACA GGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGGGTGTCCGTA ACTCCGATCCAAAGGATTGTCCTGAGCGGTGAAAATGGGCTGA AGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGG CGACCAAATGGGCCAGATCGAAAAAATTTTTAAGGTGGTGTAC CCTGTGGATGATCATCACTTTAAGGTGATCCTGCACTATGGCA CACTGGTAATCGACGGGGTTACGCCGAACATGATCGACTATTT CGGACGGCCcTATGAAGGCATCGCCGTGTTCGACGGCAAAAA GATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATC GACGAGCGCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGA

	GTAACCATCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGC ATTCTGGCGGACTACAAAGACCATGACGGTGATTATAAAGATC ATGACATCGATTACAAGGATGACGATGACAAGTAAGGCCGCGA CTCGAGAGGGCCC

2.9 References

1. Nelson, D. L., Orr, H. T., and Warren, S. T. (2013) The unstable repeats--three evolving faces of neurological disease. *Neuron* **77**, 825-843
2. Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., Nan, Z., Forster, C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* **108**, 260-265
3. Kearse, M. G., and Todd, P. K. (2014) Repeat-associated non-AUG translation and its impact in neurodegenerative disease. *Neurotherapeutics* **11**, 721-731
4. Krans, A., Kearse, M. G., and Todd, P. K. (2016) Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Ann Neurol* **80**, 871-881
5. Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. C., Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., and Paulson, H. L. (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440-455

6. Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., Pletnikova, O., Borchelt, D. R., Ross, C. A., Margolis, R. L., Yachnis, A. T., Troncoso, J. C., and Ranum, L. P. (2015) RAN Translation in Huntington Disease. *Neuron* **88**, 667-677

7. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256

8. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V. M., Kaivorinne, A. L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C.,

Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J., Traynor, B. J., and Consortium, I. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268

9. Sareen, D., O'Rourke, J. G., Meera, P., Muhammad, A. K., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., Gendron, T., Petrucelli, L., Baughn, M., Ravits, J., Harms, M. B., Rigo, F., Bennett, C. F., Otis, T. S., Svendsen, C. N., and Baloh, R. H. (2013) Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med* **5**, 208ra149

10. Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646

11. Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C., and Edbauer, D. (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335-1338

12. Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., Ostrow, L. W., Rothstein, J. D., Troncoso, J. C., and Ranum, L. P. (2013) RAN proteins and RNA foci from

antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-4977

13. Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., Moens, T., Norona, F. E., Woollacott, I. O., Pietrzyk, J., Cleverley, K., Nicoll, A. J., Pickering-Brown, S., Dols, J., Cabecinha, M., Hendrich, O., Fratta, P., Fisher, E. M., Partridge, L., and Isaacs, A. M. (2014) C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192-1194

14. Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., Lin, S., Shneider, N. A., Monaghan, J., Pandey, U. B., Pasinelli, P., Ichida, J. K., and Trotti, D. (2014) 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

15. Tran, H., Almeida, S., Moore, J., Gendron, T. F., Chalasani, U., Lu, Y., Du, X., Nickerson, J. A., Petrucelli, L., Weng, Z., and Gao, F. B. (2015) Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS. *Neuron* **87**, 1207-1214

16. Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., Daley, E. L., Miller, S. J., Cunningham, K. M., Vidensky, S., Gupta, S., Thomas, M. A., Hong, I., Chiu, S. L., Huganir, R. L., Ostrow, L. W., Matunis, M. J., Wang, J., Sattler, R., Lloyd, T. E., and Rothstein, J. D. (2015) The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56-61

17. Yamakawa, M., Ito, D., Honda, T., Kubo, K., Noda, M., Nakajima, K., and Suzuki, N. (2015) Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum Mol Genet* **24**, 1630-1645

18. Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., Paul, J. W., Sun, S., Herdy, J. R., Bieri, G., Kramer, N. J., Gage, F. H., Van Den Bosch, L., Robberecht, W., and Gitler, A. D. (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18**, 1226-1229

19. Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., Kim, N. C., Almeida, S., Lee, K. H., Badders, N., Valentine, M., Miller, B. L., Wong, P. C., Petrucelli, L., Kim, H. J., Gao, F. B., and Taylor, J. P. (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129-133

20. May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grässer, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., Mann, M., Meissner, F., and Edbauer, D. (2014) C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* **128**, 485-503

21. Zhang, Y. J., Gendron, T. F., Grima, J. C., Sasaguri, H., Jansen-West, K., Xu, Y. F., Katzman, R. B., Gass, J., Murray, M. E., Shinohara, M., Lin, W. L., Garrett, A., Stankowski, J. N., Daughrity, L., Tong, J., Perkerson, E. A., Yue, M., Chew, J., Castanedes-Casey, M., Kurti, A., Wang, Z. S., Liesinger, A. M., Baker, J. D., Jiang, J., Lagier-Tourenne, C., Edbauer, D., Cleveland, D. W., Rademakers, R., Boylan, K. B., Bu, G., Link, C. D., Dickey, C. A., Rothstein, J. D., Dickson, D. W.,

Fryer, J. D., and Petrucelli, L. (2016) C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci* **19**, 668-677

22. Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., Guharoy, M., De Decker, M., Jaspers, T., Ryan, V. H., Janke, A. M., Baatsen, P., Vercruysse, T., Kolaitis, R. M., Daelemans, D., Taylor, J. P., Kedersha, N., Anderson, P., Impens, F., Sobott, F., Schymkowitz, J., Rousseau, F., Fawzi, N. L., Robberecht, W., Van Damme, P., Tompa, P., and Van Den Bosch, L. (2017) Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics. *Mol Cell* **65**, 1044-1055.e1045
23. Lee, K. H., Zhang, P., Kim, H. J., Mitrea, D. M., Sarkar, M., Freibaum, B. D., Cika, J., Coughlin, M., Messing, J., Molliex, A., Maxwell, B. A., Kim, N. C., Temirov, J., Moore, J., Kolaitis, R. M., Shaw, T. I., Bai, B., Peng, J., Kriwacki, R. W., and Taylor, J. P. (2016) C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **167**, 774-788.e717
24. Kanekura, K., Yagi, T., Cammack, A. J., Mahadevan, J., Kuroda, M., Harms, M. B., Miller, T. M., and Urano, F. (2016) Poly-dipeptides encoded by the C9ORF72 repeats block global protein translation. *Hum Mol Genet* **25**, 1803-1813
25. Jackson, R. J., Hellen, C. U., and Pestova, T. V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113-127
26. Sonenberg, N., and Hinnebusch, A. G. (2009) Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136**, 731-745

27. Das, S., Ghosh, R., and Maitra, U. (2001) Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein. *J Biol Chem* **276**, 6720-6726

28. Wilson, J. E., Pestova, T. V., Hellen, C. U., and Sarnow, P. (2000) Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**, 511-520

29. Walter, P., and Ron, D. (2011) The unfolded protein response: from stress pathway to homeostatic regulation. *Science* **334**, 1081-1086

30. Young, S. K., and Wek, R. C. (2016) Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. *J Biol Chem* **291**, 16927-16935

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

32. Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., Levorse, J., Dill, B. D., Schramek, D., Molina, H., Weissman, J. S., and Fuchs, E. (2017) Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**, 494-499

33. Starck, S. R., Tsai, J. C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., Martins-Green, M., Shastri, N., and Walter, P. (2016) Translation from the 5' untranslated region shapes the integrated stress response. *Science* **351**, aad3867

34. Young, S. K., Baird, T. D., and Wek, R. C. (2016) Translation Regulation of the Glutamyl-prolyl-tRNA Synthetase Gene EPRS through Bypass of Upstream Open Reading Frames with Noncanonical Initiation Codons. *J Biol Chem* **291**, 10824-10835

35. Schwab, S. R., Shugart, J. A., Horng, T., Malarkannan, S., and Shastri, N. (2004) Unanticipated antigens: translation initiation at CUG with leucine. *PLoS Biol* **2**, e366

36. Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., and Todd, P. K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**, 314-322

37. Rakotondrafara, A. M., and Hentze, M. W. (2011) An efficient factor-depleted mammalian in vitro translation system. *Nat Protoc* **6**, 563-571

38. Mackenzie, I. R., Frick, P., Grässer, F. A., Gendron, T. F., Petrucelli, L., Cashman, N. R., Edbauer, D., Kremmer, E., Prudlo, J., Troost, D., and Neumann, M. (2015) Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol* **130**, 845-861

39. Bordeleau, M. E., Mori, A., Oberer, M., Lindqvist, L., Chard, L. S., Higa, T., Belsham, G. J., Wagner, G., Tanaka, J., and Pelletier, J. (2006) Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. *Nat Chem Biol* **2**, 213-220

40. Sellier, C., Buijsen, R. A., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziane, H., Vincent, A., Champy, M. F., Sorg, T., Pavlovic, G., Wattenhofer-Donze, M., Birling, M. C., Oulad-Abdelghani, M., Eberling, P., Ruffenach, F., Joint, M., Anheim, M., Martinez-Cerdeno, V., Tassone, F., Willemsen, R., Hukema, R. K., Viville, S., Martinat, C., Todd, P. K., and Charlet-

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

41. Andreev, D. E., O'Connor, P. B., Fahey, C., Kenny, E. M., Terenin, I. M., Dmitriev, S. E., Cormican, P., Morris, D. W., Shatsky, I. N., and Baranov, P. V. (2015) Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife* **4**, e03971
42. McEwen, E., Kedersha, N., Song, B., Scheuner, D., Gilks, N., Han, A., Chen, J. J., Anderson, P., and Kaufman, R. J. (2005) Heme-regulated inhibitor kinase-mediated phosphorylation of eukaryotic translation initiation factor 2 inhibits translation, induces stress granule formation, and mediates survival upon arsenite exposure. *J Biol Chem* **280**, 16925-16933
43. Boyce, M., Bryant, K. F., Jousse, C., Long, K., Harding, H. P., Scheuner, D., Kaufman, R. J., Ma, D., Coen, D. M., Ron, D., and Yuan, J. (2005) A selective inhibitor of eIF2alpha dephosphorylation protects cells from ER stress. *Science* **307**, 935-939
44. Peabody, D. S. (1989) Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* **264**, 5031-5035
45. Smith, H. L., and Mallucci, G. R. (2016) The unfolded protein response: mechanisms and therapy of neurodegeneration. *Brain* **139**, 2113-2121
46. Amick, J., and Ferguson, S. M. (2017) C9orf72: At the intersection of lysosome cell biology and neurodegenerative disease. *Traffic* **18**, 267-276

47. Zhang, Y. J., Jansen-West, K., Xu, Y. F., Gendron, T. F., Bieniek, K. F., Lin, W. L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., Chew, J., Belzil, V. V., Prudencio, M., Stankowski, J. N., Castanedes-Casey, M., Whitelaw, E., Ash, P. E., DeTure, M., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2014) Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* **128**, 505-524

48. Rossi, S., Serrano, A., Gerbino, V., Giorgi, A., Di Francesco, L., Nencini, M., Bozzo, F., Schininà, M. E., Bagni, C., Cestra, G., Carrì, M. T., Achsel, T., and Cozzolino, M. (2015) Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS. *J Cell Sci* **128**, 1787-1799

49. Buchan, J. R., and Parker, R. (2009) Eukaryotic stress granules: the ins and outs of translation. *Mol Cell* **36**, 932-941

50. Schmidt, E. K., Clavarino, G., Ceppi, M., and Pierre, P. (2009) SUnSET, a nonradioactive method to monitor protein synthesis. *Nat Methods* **6**, 275-277

51. Niblock, M., Smith, B. N., Lee, Y. B., Sardone, V., Topp, S., Troakes, C., Al-Sarraj, S., Leblond, C. S., Dion, P. A., Rouleau, G. A., Shaw, C. E., and Gallo, J. M. (2016) Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. *Acta Neuropathol Commun* **4**, 18

52. Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., and

Wang, J. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195-200

53. Hautbergue, G. M., Castelli, L. M., Ferraiuolo, L., Sanchez-Martinez, A., Cooper-Knock, J., Higginbottom, A., Lin, Y. H., Bauer, C. S., Dodd, J. E., Myszczynska, M. A., Alam, S. M., Garneret, P., Chandran, J. S., Karyka, E., Stopford, M. J., Smith, E. F., Kirby, J., Meyer, K., Kaspar, B. K., Isaacs, A. M., El-Khamisy, S. F., De Vos, K. J., Ning, K., Azzouz, M., Whitworth, A. J., and Shaw, P. J. (2017) SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nat Commun* **8**, 16063

54. Wojciechowska, M., Olejniczak, M., Galka-Marciniak, P., Jazurek, M., and Krzyzosiak, W. J. (2014) RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. *Nucleic Acids Res* **42**, 11849-11864

55. Yu, C. H., Teulade-Fichou, M. P., and Olsthoorn, R. C. (2014) Stimulation of ribosomal frameshifting by RNA G-quadruplex structures. *Nucleic Acids Res* **42**, 1887-1892

56. Gami, P., Murray, C., Schottlaender, L., Bettencourt, C., De Pablo Fernandez, E., Mudanohwo, E., Mizielinska, S., Polke, J. M., Holton, J. L., Isaacs, A. M., Houlden, H., Revesz, T., and Lashley, T. (2015) A 30-unit hexanucleotide repeat expansion in C9orf72 induces pathological lesions with dipeptide-repeat proteins and RNA foci, but not TDP-43 inclusions and clinical disease. *Acta Neuropathol* **130**, 599-601

57. Kozak, M. (1989) Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* **9**, 5073-5080

58. Ventoso, I., Sanz, M. A., Molina, S., Berlanga, J. J., Carrasco, L., and Esteban, M. (2006) Translational resistance of late alphavirus mRNA to eIF2alpha phosphorylation: a strategy to overcome the antiviral effect of protein kinase PKR. *Genes Dev* **20**, 87-100

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

60. Gao, X., Wan, J., Liu, B., Ma, M., Shen, B., and Qian, S. B. (2015) Quantitative profiling of initiating ribosomes in vivo. *Nat Methods* **12**, 147-153

61. Vattem, K. M., and Wek, R. C. (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci U S A* **101**, 11269-11274

62. Palam, L. R., Baird, T. D., and Wek, R. C. (2011) Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation. *J Biol Chem* **286**, 10939-10949

63. Pestova, T. V., de Breyne, S., Pisarev, A. V., Abaeva, I. S., and Hellen, C. U. (2008) eIF2-dependent and eIF2-independent modes of initiation on the CSFV IRES: a common role of domain II. *EMBO J* **27**, 1060-1072

64. Dmitriev, S. E., Terenin, I. M., Andreev, D. E., Ivanov, P. A., Dunaevsky, J. E., Merrick, W. C., and Shatsky, I. N. (2010) GTP-independent tRNA delivery to the

ribosomal P-site by a novel eukaryotic translation factor. *J Biol Chem* **285**, 26779-26787

65. Skabkin, M. A., Skabkina, O. V., Dhote, V., Komar, A. A., Hellen, C. U., and Pestova, T. V. (2010) Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev* **24**, 1787-1801
66. Starck, S. R., Jiang, V., Pavon-Eternod, M., Prasad, S., McCarthy, B., Pan, T., and Shastri, N. (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719-1723
67. Lee, Y. B., Chen, H. J., Peres, J. N., Gomez-Deza, J., Attig, J., Stalekar, M., Troakes, C., Nishimura, A. L., Scotter, E. L., Vance, C., Adachi, Y., Sardone, V., Miller, J. W., Smith, B. N., Gallo, J. M., Ule, J., Hirth, F., Rogelj, B., Houart, C., and Shaw, C. E. (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep* **5**, 1178-1186
68. Soto Rifo, R., Ricci, E. P., Décimo, D., Moncorgé, O., and Ohlmann, T. (2007) Back to basics: the untreated rabbit reticulocyte lysate as a competitive system to recapitulate cap/poly(A) synergy and the selective advantage of IRES-driven translation. *Nucleic Acids Res* **35**, e121
69. Caschera, F., and Noireaux, V. (2015) Preparation of amino acid mixtures for cell-free expression systems. *Biotechniques* **58**, 40-43
70. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R., and Finkbeiner, S. (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805-810

71. Barmada, S. J., Ju, S., Arjun, A., Batarse, A., Archbold, H. C., Peisach, D., Li, X., Zhang, Y., Tank, E. M., Qiu, H., Huang, E. J., Ringe, D., Petsko, G. A., and Finkbeiner, S. (2015) Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc Natl Acad Sci U S A* **112**, 7821-7826

Chapter 3: High-Throughput Screening to Identify Small Molecule Inhibitors of RAN Translation

This chapter is published as:

Green KM, Sheth U, Flores BN, Wright SE, Sutter A, Kearse MG, Barmada S, Ivanova M, Todd PK. High-throughput screening yields several small-molecule inhibitors of repeat-associated non-AUG translation. *Journal of Biological Chemistry.* 2019 Dec 6. 294(49):18624-18638.

3.1 Statement of others' contribution to work presented in this chapter

Mike Kearse assisted me with the initial design of the screen. Udit Sheth assisted me in validation experiments with the 22 small molecules purchased based upon screen results. Alexandra Sutter and Magdalena Ivanova assisted me in performing circular dichroism experiments and analyzing results. Brittany Flores and Sami Barmada perform experiments in primary rodent neurons. Shannon Wright collected data with the near-AUG reporters. I performed all other experiments, generated all figures, and wrote the manuscript from which this chapter is derived, with Peter Todd's oversight.

3.2 Introduction

Repeat-associated non-AUG (RAN) translation has recently emerged as a common pathogenic mechanism among nucleotide repeat expansion diseases. Through this process, the ribosome initiates translation upstream of or within stable secondary structures formed by repetitive RNA sequences, in the absence of an AUG start codon (1). This leads to production of RAN peptides; often aggregate-prone proteins with repetitive amino acid sequences resulting from translation of the repetitive RNA element.

To date, RAN translation is known to occur in at least nine different human diseases: fragile X-associated tremor ataxia syndrome (FXTAS) (2,3), amyotrophic lateral sclerosis (ALS) (4-7), frontotemporal dementia (FTD) (4-7), spinocerebellar ataxia type 8 (1), myotonic dystrophy types 1 and 2 (1,8), Huntington's disease (9), fragile X-associated primary ovarian insufficiency syndrome (10), and Fuchs' endothelial corneal dystrophy (11). These diseases are caused by multiple different microsatellite repeat sequences located within various regions (e.g., coding sequence, UTRs, intron) of different genes.

FXTAS is a late onset neurodegenerative disease characterized by difficulty walking, loss of fine motor skills, and progressive cognitive and behavioral changes (12), that affects approximately 1 in 10,000 men over the age of 50 (13). FXTAS results from the expansion of CGG repeats within the 5′ leader of *FMR1*, from 30 or fewer CGGs, to 55-200 (14). RAN translation of this repeat in the +1 (GGC) reading frame generates a polyglycine protein, FMRPolyG, that is found within inclusions in patient neurons (2,15,16). When overexpressed from a CGG repeat in *Drosophila* or mice, FMRPolyG is neurotoxic, causing motor deficits, neurodegeneration and reduced lifespans (2,15).

ALS and FTD are also neurodegenerative diseases. ALS is the most common form of motor neuron disease, and FTD the second most common form of early-onset dementia. The most common genetic cause of both ALS and FTD is 70 or more GGGGCC repeats within the first intron of *C9orf72* (17-19). RAN translation in all three reading frames of the expanded GGGGCC repeat produces three different dipeptide repeat proteins (DPRs); glycine-alanine (GA: GGG-GCC), glycine-proline (GP: GGG-CCG), and glycine-arginine (GR: GGC-CGG). All three DPRs are detected in patient neurons (4,5,7), and the GR product has been repeatedly shown to be highly toxic across model systems (20-22), with more moderate toxicity from the GA product also reported (20,23-25).

Despite differences in repeat sequence and context, our lab and others have shown that RAN translation from reporter constructs with FXTAS-associated CGG repeats (CGG RAN) and C9ALS/FTD-associated GGGGCC repeats (C9RAN) share common mechanistic features (26-29). RAN translation in all three reading frames across both repeats occurs most efficiently when the ribosome is able to bind at the 5′ cap and scan in a 5′ to 3′ direction (26-29). Additionally, for both repeats, RAN translation in the reading frame that generates the most abundant RAN peptide in patient tissue (FMRPolyG and GA) (2,30), predominantly utilizes a near-AUG codon upstream of the repeat for initiation (26-28,31). However, evidence also suggests that initiation can occur, at lower levels, within the repeat sequence itself and in a cap-independent manner (1,29,31). Intriguingly, activation of the integrated stress response and phosphorylation of the initiation factor eIF2α, enhances RAN translation of both the CGG and GGGGCC repeats in model systems (27,29,31,32).

To better define the mechanism of RAN translation and identify potential inhibitors of this process, we developed an *in vitro*, reporter-based small molecule screen for bioactive compounds that selectively and dose-dependently inhibit RAN translation at CGG repeats. From this screen, we identified five novel CGG RAN translation inhibitors, and found that each compound also inhibits RAN translation at C9ALS/FTD-associated GGGGCC repeats, in multiple sense-strand reading frames. Using circular dichroism and native gel analysis, we show that three of these compounds bind the repeats and altering their secondary structure in a manner similar to previously developed small molecule inhibitors (33-35). These studies establish that RAN translation from multiple repeat expansions, and across multiple reading frames, can be selectively targeted by small molecule inhibitors, providing insights into the mechanisms by which RAN translation occurs and establishing a framework for future therapeutic development in nucleotide repeat expansion disorders.

3.3 Results

Primary screen of 3253 bioactive compounds for inhibitors of CGG RAN translation

To identify small molecule inhibitors of RAN translation of FXTAS-associated CGG repeats, we adapted a previously developed *in vitro* RAN translation assay (26) to a 384-well format (**Fig. 3-1, A**). This assay utilized a +1CGGx100 RAN translation nanoluciferase (NLuc) reporter mRNA, which was added to wells containing rabbit reticulocyte lysate (RRL) treated with one of 3253 bioactive compounds at 20 µM (**Fig. 3-1, A**). After a 30-minute incubation at 30ºC, luminescence was measured as a readout for production of the neurotoxic polyglycine RAN protein (FMRpolyG) in each well (**Fig.**

3-1, A). We confirmed that under these conditions, the reporter mRNA produced robust RAN-specific luminescent signal within the dynamic linear range of detection (**Fig. 3-1, B & C** and **Fig. 3-2, A**).

The 3253 compounds screened with this assay came from the Pilot LOPAC, Pilot Prestwick, Pilot NCC-focused and Navigator Pathways small molecule libraries. These libraries contain pharmacologically active compounds, small molecules previously tested in clinical trials, and FDA-approved drugs, all in DMSO. As a positive control for translation inhibition, the last two columns of each plate were treated with 30 nM cycloheximide, and as an internal negative control, the first two columns were treated with DMSO (**Fig. 3-1, A**). The average Z-factor and coefficient of variability (CV) value, calculated per plate, was 0.79 and 6.93%, respectively (**Table 3-1**), indicating that the screen was of high quality. This allowed us to identify 289 "hits," which we defined as compounds that reduced +1CGG RAN translation by greater than 20% or 3 standard deviations relative to vehicle-treated controls (**Fig. 3-1, D, Table 2,** and Supplementary Table 1). These hits were structurally and mechanistically diverse. While 14 known ribosomal inhibitors and 17 DNA/RNA intercalators were identified, the majority of hits represented other classes of compounds, acting through less obvious modes of translational inhibition (**Fig. 3-1, E**).

Concentration response and counter screen

From 289 primary hits, we selected 110 inhibitors to advance to a secondary concentration response curve analysis (**Fig. 3-3, A**). Compounds were selected primarily based on their percent inhibition and how many standard deviations their effect differed from vehicle-only controls. Preference was given to compounds that were FDA-approved

or that were hits multiple times when present in multiple libraries. Compounds that were flagged for concerns about toxicity, active in greater than 90% of luciferase assays previously performed at our screening facility or known translation inhibitors of a mechanism already represented among selected hits, were excluded unless of particular interest. Additionally, a standard deviation cut-off of 2.5 was applied to compounds added to wells at the edge of the plate, where signal was more variable.

Selected compounds were added to the RRL at 8 concentrations from 50 to 8 µM, in duplicate. To obtain a 50 µM dose, the volume of compound added to each well increased the final percentage of DMSO in the reaction mixture from 1% in the primary screen, to 2%. We confirmed that at this level, DMSO did not significantly inhibit +1CGG RAN translation on its own (**Fig. 3-4, A**). Of the 110 compounds selected, 77 exhibited dose-dependent inhibition of +1CGG RAN translation, while 33 did not (**Fig. 3-3, A** and Supplementary Table 2). Among the 33 compounds that failed to validate, only three reduced signal in the primary screen by both more than 20% and 4 standard deviations relative to controls (Supplementary Table 2).

To eliminate compounds that non-specifically inhibited all translation, with no specificity for RAN translation, the concentration response curve assay was simultaneously performed with an NLuc reporter mRNA lacking any repeat element and translated using a canonical AUG start codon (AUG-NLuc) (**Fig. 3-4, B & C**).

Of the 77 validated hits, 54 also inhibited translation of the AUG-initiated canonical reporter in a dose-dependent manner. Consequently, 23 compounds exclusively and dose-dependently inhibited +1CGG RAN translation, which were called “RAN specific inhibitors” (**Fig. 3-3, B** and Supplementary Table 2). None of these compounds were of

high potency; the IC50 of the most potent, amlexanox, was 26.3 μM (Supplementary Table 2). An additional 30 compounds were more potent inhibitors of +1CGG RAN translation than canonical translation (**Fig. 3-3, C to F** and Supplementary Table 2). These were called “RAN selective inhibitors.” As a group, the RAN selective inhibitors were more potent than the RAN specific inhibitors, with IC50s ranging from 0.72 to 81.3 μM (Supplementary Table 2).

Of the remaining 24 compounds that reduced both +1CGG RAN and AUG-NLuc translation, 16 had similar activity with both, while 8 were more potent inhibitors of canonical translation, including cycloheximide, consistent with a recent report (36) (Supplementary Table 2).

Five bioactive compounds selectively inhibit CGG RAN translation in multiple reading frames

Based on results from the concentration response and counter screen, 22 compounds were selected for independent validation, including 13 RAN specific and nine RAN selective inhibitors (**Fig. 3-3, A** and **Table 3-3**). Each compound was added at 4-5 doses to a 10 μL RRL *in vitro* translation assay using the same reporter mRNAs as in the primary and counter screens. Of these compounds, only one RAN specific and four RAN selective inhibitors significantly and dose-dependently inhibited +1CGG RAN translation, while leaving translation from the canonical AUG-NLuc reporter relatively spared (**Table 3-3**). These compounds included cephalothin, a discontinued beta-lactam antibiotic; BIX01294, a histone lysine methyltransferase inhibitor; anthralin, an FDA-approved drug used in topical treatments for psoriasis; propidium iodide, a fluorescent nucleic acid

intercalating agent; and CP-31398, a p53 stabilizer (**Fig. 3-5, A to F**). As a group, these compounds have diverse biological targets and structures. However, interestingly, BIX01294 and CP-31398 share a similar functional group (**Fig. 3-5, A**).

The remaining 17 compounds either had no effect on RAN translation or also inhibited canonical translation to a similar degree during independent validation (**Fig. 3-6, A to D; Table 3-3**). To assess the possibility that results observed in the primary screen for these 17 compounds came from a by-product produced through extended storage durations, we incubated a subset at 37^{o}C for 1 week. However, this did not substantially affect their inhibitory properties (data not shown).

RAN translation of the expanded CGG repeat also occurs in the +2 reading frame (GCG), generating a polyalanine protein (FMRPolyA) (2). However, unlike initiation in the +1 reading frame, initiation in the +2 frame does not utilize a near-AUG codon, and likely occurs within the repeat sequence itself (26). We were therefore interested to see if the five inhibitors of +1CGG RAN translation also reduced +2CGG RAN translation. Cephalothin, BIX01294, anthralin, and CP-31398 all significantly inhibited translation from a +2CGGx100 RAN translation reporter mRNA (26), relative to the canonical AUG-NLuc control (**Fig. 3-5, B to D and F**). However, at the doses tested, propidium iodide did not (**Fig. 3-5, E**). Consequently, RAN translation resulting in the synthesis of distinct polypeptides, from initiation at different codons in different reading frames, can be inhibited simultaneously by the same small molecule, although this effect cannot be generalized to all inhibitors.

We also assessed the inhibitory activity of two additional small molecules, protoporphyrin IX (PPIX) and TMPyP4. PPIX is a natural compound that consists of

porphyrin ring. It is known to bind g-quadruplex structures and increases levels of FMRP in fragile-X patient cells (37). TMPyP4 is structurally similar to PPIX, also consisting of a porphyrin ring, and known to bind CGG and GGGGCC repeat RNA (38,39). When tested, both compounds selectively inhibited CGG RAN translation, relative to the AUG-NLuc control (**Fig. 3-5, G and H**). However, unlike PPIX, TMPyP4 only had relative selectivity for CGG RAN translation in the +1, but not +2, reading frame for the doses tested, much like what we observed with propidium iodide (**Fig. 3-5, H**).

Inhibitors of CGG RAN translation also inhibit C9RAN translation

Recent work suggests mechanistic similarity between CGG RAN translation and C9ALS/FTD-associated GGGGCC RAN translation (C9RAN translation) (27-29). We therefore tested if these small molecule inhibitors of CGG RAN translation also impacted C9RAN translation. To accomplish this, we repeated the RRL *in vitro* translation assays using previously developed C9RAN translation-specific NLuc reporter mRNAs that contained 70 GGGGCC repeats (27). All five compounds identified in the CGG RAN translation screen, as well as PPIX, selectively and dose-dependently decreased C9RAN translation in multiple reading frames (GA, GP, and GR), relative to the AUG-NLuc control (**Fig. 3-7, A to C; Fig. 3-8, A to C**). We further confirmed the decrease in luminescence from the CGG and C9RAN reporter mRNAs was due to a decrease in RAN polypeptide synthesis by western blot. BIX01294, anthralin, and PPIX all decreased poly-GA and FMRPolyG synthesis in RRL, in a dose-dependent manner, while leaving canonical NLuc synthesis relatively spared (**Fig. 3-7, D to F)**.

RAN translation inhibitors have varying effects on AUG-initiated translation of expanded repeats

To determine if these small molecules inhibit RAN translation by targeting its non-AUG initiation, we repeated the *in vitro* translation reactions using reporters with AUG start codons inserted upstream of the expanded CGG repeats in either the +1 or +2 reading frames (**Fig. 3-9, A** and **Fig. 3-10, A**). Anthralin, TMPyP4, and PPIX all inhibited AUG +1CGGx100 and AUG +2CGGx100 translation to a similar or greater extent than +1CGG and +2CGG RAN translation (**Fig. 3-9, B to D**). Consequently, the selectivity of these compounds for translation of expanded CGG repeats, relative to AUG-NLuc translation, was maintained regardless of whether repeat translation initiated with an AUG or non-AUG codon (**Fig. 3-9, B to D**). Conversely, BIX01294 and CP-31398 inhibition was markedly decreased when translation of the expanded repeats initiated with an AUG start codon (**Fig. 3-9, E & F**), indicating that their activity is dependent upon a non-AUG initiation event. Therefore, these five small molecules fall within two distinct groups that produce similar inhibitory effects by targeting different aspects of RAN translation.

To assess whether the inhibitory activity of BIX01294 and CP-31398 on non-AUG initiated-translation was specific to mRNAs containing a repeat, we utilized two additional NLuc reporters; one that initiates at an ACG and the other at a CUG start codon (ACG-NLuc and CUG-NLuc), both lacking a repetitive element (**Fig. 3-10, B**). We specifically chose these two non-AUG start codons as they were previously shown to be used for FMRpolyG and poly-GA synthesis, respectively (26-28). Interestingly, relative to the canonical AUG-NLuc control, BIX01294 and CP-31398 more significantly inhibited translation from both the ACG-NLuc and CUG-NLuc reporters (**Fig. 3-10, C & D**).

Therefore, a repeat is not required for translation inhibition by either compound. This supports the idea that non-AUG-initiated translation can be sufficiently distinct from canonical AUG-initiated translation to allow for selective targeting (40).

In contrast, anthralin, PPIX and TMPyP4 all failed to inhibit translation of ACG-NLuc and CUG-NLuc reporter mRNAs, relative to the AUG-NLuc control (**Fig. 3-10, E to G**), indicating that their inhibitory effects are repeat-dependent. Additionally, as expression of the near-AUG initiated reporters is significantly lower than the AUG-initiated control (**Fig. 3-10, B**), this suggests the relatively low expression levels of the repeat-containing reporters does not alone account for their selective inhibition by these compounds.

BIX01294 interacts with repeat RNAs

To test if the RAN translation inhibitors directly interact with the repeat RNAs, as a possible mechanism for their inhibition, we utilized circular dichroism (CD). First, we attained the CD spectrum of CGGx16 repeat RNA folded in the presence of 100 mM KCl (the concentration present in the RRL reactions) by heating to 95ºC and returning to room temperature. UUUx16 RNA, folded under the same conditions, was used as a control lacking both sequence and structural similarity. We then incubated each RNA with 25 and 50 µM TMPyP4, as a positive control for CGG repeat RNA binding. Consistent with previous findings (38,41), TMPyP4 caused a dose-dependent shift in the CGGx16 RNA CD spectrum, while a higher dose of TMPyP4 was required to alter UUUx16's CD spectrum (**Fig. 3-11, A & B**).

We next incubated the RNAs with increasing amounts of BIX01294 and anthralin. At both 50 and 100 μM, BIX01294 shifted CGGx16's CD spectrum, but had little effect on UUUx16's (**Fig. 3-11, C & D**). Conversely, incubation with anthralin had no effect on the CD spectra of either CGGx16 or UUUx16, even at four times a dose of BIX01294 that caused a clear shift (50 vs. 200 μM, **Fig. 3-11, E & F**). Higher doses of anthralin were not tested by CD, due to interference by the increasing concentration of DMSO with the absorbance spectra.

To determine if these compounds interact similarly with the *C9ORF72*-associated GGGGCC repeat, we performed CD using a GGGGCCx8 repeat RNA. This RNA was folded at room temperature following heating to 95°C in the presence of either 100 mM KCl, to promote the formation of a g-quadruplex structure, or in the presence of 100 mM NaCl, to promote hairpin formation (33,39). Regardless of the cation used, TMPyP4 and BIX01294 shifted the GGGGCCx8 RNA CD spectra, while anthralin did not (**Fig. 3-12**), mirroring their interactions with CGGx16.

As a secondary measure of compound interaction with CGG repeat RNA, we utilized native gel electrophoresis to assay for an electrophoretic shift indicative of changes in the RNA structure due to compound binding. For this, we used 5′ IRdye 800CW-labeled CGGx16 and UUUx16 RNAs folded in the presence of 100 mM KCl. 50 nM of each RNA was incubated with increasing concentrations of each inhibitor and run on a 12% native polyacrylamide gel. The labeled RNAs were then visualized at 800nm.

TMPyP4 was again used as a positive control for direct binding to the CGG repeat RNA (38). Increasing concentration of TMPyP4 led to a dose-dependent decrease in CGGx16's band intensity, despite equal RNA loading (**Fig. 3-13, A**). This differs from a

previous report using a radio-labeled CGG RNA with extensive sequence 5′ to the repeat (41), and may be the result of TMPyP4 binding interfering with detection of the RNA's IRdye. A similar result was obtained when TMPyP4 was incubated with the UUUx16 RNA (**Fig. 3-13, A**).

BIX01294, propidium iodide, and CP-31398 all had a similar effect on the CGGx16 RNA as TMPyP4, causing a dose-dependent decrease in its band intensity (**Fig. 3-13, B to D**). Additionally, they showed a preferential effect on the CGGx16 RNA relative to the control UUUx16 RNA, as lower doses of each compound were required to reduce the intensity of the CGGx16 band, relative to those needed to reduce intensity of the UUUx16 band (**Fig. 3-13, B to D**). However, only incubation with increasing concentrations of CP-31398 resulted in a clear upshift indicative of altered RNA secondary structure (**Fig. 3-13, D**).

PPIX also modestly, but preferentially decreased the CGGx16 RNA band intensity in a dose-dependent manner (**Fig. 3-13, E**), while anthralin and cephalothin showed little evidence of interacting with either CGGx16 or UUUx16 RNA, even at excessively high doses (**Fig. 3-13, F & G**).

Therefore, by CD and/or native gel analysis, BIX01294 and CP-31398 exhibit preferential interaction with CG-rich repeats RNAs. Native gel results suggest a possible interaction between propidium iodide and CGGx16 RNA, while anthralin and cephalothin show no evidence of interaction. This may indicate that these sets of compounds utilize different mechanisms for inhibiting CGG RAN translation.

BIX01294 inhibits CGG RAN in cultured cells, but is toxic

We next asked whether these small molecule inhibitors of RAN translation were active in cultured cells. To do so, we applied BIX01294 to HEK293 cells transfected with plasmids expressing either AUG-NLuc, +1CGGx100-NLuc, or GA70-NLuc reporters. As an internal control for transfection efficiency and overall cellular health, all wells were co-transfected with a firefly luciferase reporter. Twenty-four hours post treatment, luciferase activity was measured as a readout for translation of each reporter.

Relative to vehicle-treated controls, 25 μM BIX01294 significantly reduced +1CGG RAN translation, without affecting AUG-NLuc expression (**Fig. 3-14**). However, as assessed by changes in cellular morphology (data not shown), and a reduction in firefly luciferase signal (**Fig. 3-15, A**), BIX01294 treatment at this dose was moderately toxic to cells. At doses above 25 μM, BIX01294 caused cellular death, with complete cellular detachment from plates (data not shown) and loss of firefly luciferase expression (**Fig. 3-15, A**). In contrast, 25 μM BIX01294 treatment unexpectedly increased poly-GA expression (**Fig. 3-15, B**), perhaps indirectly through changes in gene expression related to its primary biological activity as a histone lysine methyltransferase inhibitor.

We also applied BIX01294, anthralin, propidium iodide, and PPIX to primary rodent neurons at 1 μM, and tracked the survival of these neurons over the course of ten days using automated longitudinal fluorescence microscopy. Even at a dose well below the small molecules' *in vitro* IC50s (**Fig. 3-5**), treatment by each increased the neurons' cumulative risk of death (**Fig. 3-15, C**). Consequently, none of these small molecules are likely to represent good therapeutic candidates.

3.4 Discussion

In recent years, an emerging body of research has shown RAN translation to be a common factor in numerous repeat-expansion diseases. Although much work has established that RAN peptides are sufficient to cause neurotoxicity in a broad range of disease models, our understanding of the mechanisms of RAN translation, as well as how to selectively target it, is still incomplete.

Here we used an *in vitro* RAN translation reporter assay to perform a screen of bioactive compounds and identified five novel small molecule inhibitors of this non-canonical translation event. Although none of these compounds are bioavailable, by interrogating their activity, they have revealed new insights into the mechanism of RAN translation.

We performed the initial screens using only a +1CGGx100 RAN reporter mRNA. However, each compound identified in the screen had similar activity against GGGGCCx70 RAN translation reporter mRNAs. Therefore, despite being different repeats, located within different sequence contexts, that produce different RAN peptide products, this finding supports mechanistic overlap between these translation events, and suggests that future strategies to modify RAN translation in one disease could be more broadly applicable to the growing class of diseases associated with this event.

Previous studies to develop inhibitors of RAN translation have specifically focused on identifying small molecules that bind to the repeat RNAs (33-35,42). For instance, compound 1a was identified based on its ability to bind to loops produced by the unpaired G-G nucleotides in the hairpin stem structures formed by both CGG and GGGGCC repeats (35). Subsequent studies established that 1a, and structurally similar compounds,

reduce RAN translation of both RNAs in cultured cells, including *C9ORF72*-patient-derived iNeurons (33,42). Additionally, a recent screen for molecules that bind G-quadruplexes successfully identified two such molecules that reduced C9RAN translation in GGGGCCx36-expressing *Drosophila* and *C9ORF72*-patient-derived motor neurons (34).

Consequently, small molecule binding to repeat RNAs has proven to be a successful strategy at inhibiting RAN translation. This is consistent with the findings we report here, as three of the small molecule inhibitors we identified in the screen, BIX01294, CP-31398, and propidium iodide, show evidence of interacting with CGG and/or GGGGCC repeat RNAs by circular dichroism and/or native gel analysis. Future work to determine if the functional group shared by BIX01294 and CP-31398 is important for their interaction with the repeat RNAs and/or selective inhibition of RAN translation, could provide insight into strategies for optimizing this effect. However, these compounds also interacted with the control UUUx16 RNA at higher doses. Therefore, we have not excluded the possibility that interactions of these compounds with other RNAs or proteins in our translation system, such as ribosomal RNAs, contribute to their relatively selective inhibition of RAN translation.

Unlike previous studies, by using a luciferase-based RAN translation reporter assay to directly measure RAN translation, we were able to unbiasedly screen a diverse library of bioactive compounds, and potentially identify small molecules that inhibit RAN translation through novel mechanisms not related to repeat RNA binding. In this context, our finding that neither anthralin nor cephalothin directly interact with the CGG and/or GGGGCC repeat RNAs as measured by circular dichroism or native gel analysis, is quite

intriguing. This suggests that these compounds inhibit RAN translation through novel means, potentially by interacting with or inhibiting non-RNA-based factors that are selectively required for RAN translation.

Our findings also suggest that relatively selective inhibition of RAN translation can occur by targeting different aspects of the translation process. Relative to +1 and +2CGGx100 RAN translation, anthralin, TMPyP4 and PPIX inhibited AUG-initiated repeat-translation to a similar or greater extent. This indicates that these compounds do not target the non-AUG initiation event of CGG RAN translation but may instead impair translation elongation through the repetitive RNA elements. Alternatively, BIX01294 and CP-31398 do specifically target the non-AUG initiation event; they were less effective at inhibiting AUG-initiated repeat-translation compared to RAN translation, and selectively inhibited translation from CUG-NLuc and ACG-NLuc reporters, relative to an AUG-NLuc control.

In conclusion, this study serves as a proof-of-principle that small, bioactive compounds can selectively inhibit RAN translation across multiple disease-causing repeat expansion mutations. It has also revealed new lead compounds that could be used in future work to identify novel, targetable binding pockets in the repeat RNAs, as well as new factors involved in RAN translation.

3.5 Experimental Methods

Reporter RNA in vitro transcription

pcDNA3.1(+) NLuc reporter plasmids (26,27,36) were linearized with PspOMI. Capped and polyadenylated mRNAs were synthesized as previously described, using

HiScribe T7 High Yield RNA Synthesis Kit (NEB, Catalog No. E2040S), 3′-O-Me-m^7GpppG anti-reverse cap analog (ARCA) (NEB, Catalog No. S1411S), RNAse-free DNaseI (NEB, Catalog No. M0303S), and *E. coli* Poly-A Polymerase (NEB, Catalog No. M0276S) (26). Following synthesis, mRNAs were purified with RNA Clean and Concentrator-25 Kit from Zymo Research (Catalog No. R1017). The integrity and size of all *in vitro* transcribed mRNAs was verified on a denaturing formaldehyde RNA gel.

Primary and secondary screens

Rabbit reticulocyte lysate (RRL) *in vitro* translation reactions were performed in white, flat bottom, polypropylene 384-well plates from Greiner (Catalog No. 784075). RRL reaction mixture was prepared using Promega's Flexi System (Catalog No. L4540), and consisted of 30% RRL, 10 µM minus leucine amino acid mix, 10 µM minus methionine amino acid mix, 0.5 mM magnesium acetate, 100 mM potassium chloride, 4U murine RNAse inhibitor (NEB, Catalog No. M0314L), and 0.05% Tween-20. The addition of Tween-20 was necessary to minimize adhesion of the reaction mixture to the multidrop tubing and even dispensing between wells, and did not alter the levels of translation from the +1 CGG RAN or AUG-NLuc reporter RNAs (**Fig. 3-2, B**). 4 µL of RRL mixture was dispensed per well using Thermo Labsystems multidrop plate dispenser, at medium speed. 50 nL of DMSO was then multidropped into columns 1 and 2, and 50 nL cycloheximide was added to columns 23 and 24, for a final concentration of 3 µM. To the remaining wells, 50 nL of 3253 compounds, at 2 mM, from the Pilot LOPAC, Pilot Prestwick, Pilot NCC Focused, and Navigators Pathways Libraries were pintooled using a Sciclone ALH 300 advanced liquid handling system, for a final concentration of 20 µM.

1 µL (3 fmol +1 CGG RAN, 1 fmol AUG-NLuc) *in vitro* transcribed reporter RNA was then multidropped into each well.

Plates were incubated for 30 minutes at 30ºC. Translation reactions were then terminated by addition of 50 nL 10 µM cycloheximide to each well, using a multidrop dispenser. 10 µL of room temperature Glo Lysis Buffer (Promega, Catalog No. E2661) was dispensed into each well with the multidrop, at low speed, which was essential for maintaining the stability of NLuc luminescence. To this, 5 µL of room temperature NanoGlo Substrate diluted 1:50 in NanoGlo Buffer (Promega, Catalog No. N1120) was dispensed via multidrop at low speed. Plates were then covered, mixed gently, and incubated at room temperature for 5 minutes. Luminescence was measured with a Perkin Elmer EnVision plate reader.

The concentration response curve and counter screen were performed using all the same reagents, materials, and equipment as in the primary screen. The only difference was that selected compounds were added to 384-well plates at 8 doses from 50-8 µM, in duplicate, using TTP Labtech Mosquito X1 Hit-Picking Liquid Handler, prior to addition of RRL.

After measuring luminescence, pIC50s were calculated through the University of Michigan's Center for Chemical Genomics' MScreen software, using a 4 parameter logistic equation (43). Initial minimum limits for curve fits were informed by the minimum value of each plate, determined by the cycloheximide-treated internal positive controls. Regression analysis then derived a curve fit, with a calculated minimum and maximum limit, for each compound. IC50s greater than the maximum dose in the assay (50 µM), were extrapolated from the curve fit

Independent validation of hits in RRL system

The following compounds were purchased through Sigma Marketsite: cephalothin sodium (Sigma-Aldrich; C4520), rufloxacin HCl (Vitas-M Lab., Ltd.; STK711124), rofecoxib (Sigma-Aldrich; MFCD00935806), alfuzosin HCl (Sigma-Aldrich; A0232), olanzapine (Sigma-Aldrich; O1141), protoporphyrin IX (Chem-Impex Int, Inc; 21661), pefloxacine mesylate (Selleck Chemicals, S1855), reserpine (Sigma-Aldrich; MFCD00005091), isoxicam (Sigma-Aldrich; MFCD000079374), phenazopyridine HCl (Sigma-Aldrich; MFCD00035347), parbendazole (Sigma-Aldrich; MFCD018910864), balsalazide disodium salt dihydrate (Key Organics/BIONET; KS-5216), efavirenz (Sigma-Aldrich; SML0536), oxiconazole nitrate (Key Organics/BIONET; KS-5288), sulfamethazine sodium salt (Sigma;S5637-25G), kenpaullone (Adooq Bioscience; A11220), propidium iodide (Sigma; P4170), BIX01294 HCl hydrate (Cayman Chemical; 13124), anthralin (Key Organics/BIONET; KS-5183), CP-31398 dihydrochloride hydrate (Tocris Bioscience; 3023), X80 (Sigma; X3629), amlexanox (Sigma; 68302-57-8).

Compounds were all dissolved to 10-100 mM stocks in DMSO and stored in single-use aliquots at -20 ºC. For *in vitro* translation assays, compounds were diluted in DMSO and added to RRL for a final DMSO concentration of 1% total reaction volume. mRNAs were *in vitro* translated with Flexi Rabbit Reticulocyte Lysate System from Promega, as performed for the primary screen, with a few slight modifications. First, 0.05% Tween-20 was excluded from reactions, as regular pipetting eliminated the need to prevent RRL adhesion to multidrop dispenser tubing. Additionally, the reaction volume was increased to 10 µL, with RNA concentration unchanged (0.3 nM). Incubations were performed in

polypropylene PCR tubes, reactions were terminated on ice, and then transferred to black 96-well plates. Samples were diluted 1:7 in Glo Lysis Buffer (Promega) prior to a 5 minute in the dark with NanoGlo Substrate freshly diluted 1:50 in NanoGlo Buffer (Promega). Luminescence was measured on a GloMax 96 Microplate Luminometer.

Reactions for western blot assays were performed as above, except 50 ng mRNA was used, as previously described (27). Membranes were probed with the following antibodies: 1:1,000 FLAG-M2 (mouse, Sigma F1804), 1:1,000 GAPDH 6C5 (mouse, Santa Cruz sc32233).

Circular dichroism

CGGx16, UUUx16, and GGGGCCx8 RNAs were purchased from IDT, with RNAse-free HPLC purification. CGGx16 and UUUx16 RNAs were diluted to 5 µM in 300 µL RNA folding buffer consisting of 100 mM KCl, 10 mM Tris-HCl, and 0.1 mM EDTA (39). GGGGCCx8 RNA was diluted to 2.5 µM in 250 µL RNA folding buffer, consisting of either 100 mM KCl or 100 mM NaCl, with 10 mM Tris-HCl, and 0.1 mM EDTA. Diluted RNAs were heated to 95ºC for 1 minute and cooled to room temperature.

CD was performed using a Jasco J-1500 Circular Dichroism Spectrophotometer. Sample were scanned from 320-220 nm at 25ºC (39). Spectra were collected with scanning speed of 20nm/min (39), data interval of 0.1nm, response time of 1 sec and 1nm bandwidth. Spectra are the average of six accumulations, and only data with a heat tension less than 500 is shown. Increasing volumes of compounds were progressively added to the same RNA sample, to obtain increasing doses. The resulting decrease in RNA concentration due to increasing volume was accounted for in the molar ellipticity

calculations. Spectra collected on the same day are compared to the same untreated RNA spectra, for each RNA and cation tested. All spectra were corrected for the background, by subtracting the buffer spectra using Jasco software, and baselined by setting the first value of the RNA-only spectra to zero.

Native gel analysis

5′ IRDye-800CW-modified CGGx16 and UUUx16 RNAs were purchased from IDT, with RNAse-free HPLC purification. RNAs was diluted to 50 nM in buffer containing 100 mM KCl, 10 mM Tris-HCl, and 0.1 mM EDTA, heated to 95ºC for 1 minute, and cooled to room temperature.

Folded RNAs were then incubated with the indicated compounds and at the indicated doses for 30 minutes at room temperature. 4 µL of Orange-G loading Dye (40% glycerol, 0.1% Orange-G, 1x TBE, and sterile MQ H_2O) was added to each reaction, and entire reactions were loaded into 12% native polyacrylamide gels (12% acrylamide:bis acrylamide 29:1, 1X TBE, 0.1% APS, TEMED, and sterile MQ H_2O), with 4% stacks. Prior to loading, gels were pre-ran in 1x TBE at 100V for approximately 45 minutes. Samples were run on ice at 70V for approximately 95 minutes, and entire gels within glass cartridges were imaged with LI-COR Odyssey CLx Imaging System.

Compound activity in HEK293 cells

HEK293 cells (ATCC CRL-1573) were maintained at 37ºC and 5% CO_2 in 10 cm dishes containing 10 mL DMEM + high glucose (GE Healthcare Bio-Science, SH30022FS) supplemented with 9.09% fetal bovine serum (50 mL added to 500 mL

DMEM; Bio-Techne, S11150). 100 µL cells were plated at $2.2x10^5$ cells/mL in 96-well plates (Fisher, FB012931). Approximately 24 hours post plating, at 50-60% confluency, cells were transfected with pcDNA3.1+ NLuc reporter plasmids and control pGL4.13 Firefly Luciferase plasmid (26,27). 50 ng of each plasmid was transfected per well, using FuGene HD (Promega, E2312) at a 3:1 ratio of reagent to total DNA, according to manufacturer's recommended protocol. Approximately 24 hours post transfection, at approximately 90% confluency, cells were treated with small molecules at indicated doses. 24 hours post treatment, cells were lysed in 60 µL Glo Lysis Buffer per well, for 5 minutes at room temperature with rocking. In opaque, black 96-well plates (Grenier Bio-One, 655076), 25 µL lysate was then separately incubated for 5 minutes in the dark at room temperature with 25 µL NanoGlo Substrate freshly diluted 1:50 in NanoGlo Buffer or 25 µL ONE-Glo Buffer (Promega, E6110). NLuc and Firefly luminescence were measured on a GloMax 96 Microplate Luminometer.

Primary neuron survival experiments

Primary cortical neurons from embryonic day 19-20 rats were harvested and cultured at 0.6×10^6 cells/mL in 96 well cell culture plates, as previously described (44,45). Neurons were cultured in NEUMO photostable medium containing SOS supplement (Cell Guidance Systems) at 37ºC in 5% CO_2. Primary neurons were co-transfected with 0.1 µg of the survival marker, pGW1-mApple, and 0.1 µg of pGW1-EGFP on DIV4 using Lipofectamine 2000 (Invitrogen). Twenty-four hours post-transfection, compounds or DMSO were added to neurons, immediately following the first imaging run. Images were taken for 10 consecutive days with an automated fluorescent microscopy platform, as

previously described (44-47). Image processing and survival analysis were acquired for each neuron at each timepoint using custom code written in Python or the ImageJ macro language. For survival analyses, differences among populations through Cox proportional hazards analysis was determined with the publicly available R survival package.

3.6 Chapter-specific acknowledgements

We thank Nicholas Santoro, Tom McQuade, Renju Jacob, and Martha Larsen at the Center for Chemical Genomics in the Life Sciences Institute at University of Michigan for their help with the compound library screen, data mining, and MScreen software use. We also thank Veronica Towianski for her technical assistance during the independent validation of compound hits, Luke Parks and the University of Michigan's Biophysics Department for their invaluable training and assistance with the circular dichroism experiments, Sarah Cox for assistance with the circular dichroism software and ChemDraw, Fang He and all members of the Todd Lab for the feedback on this project and manuscript.

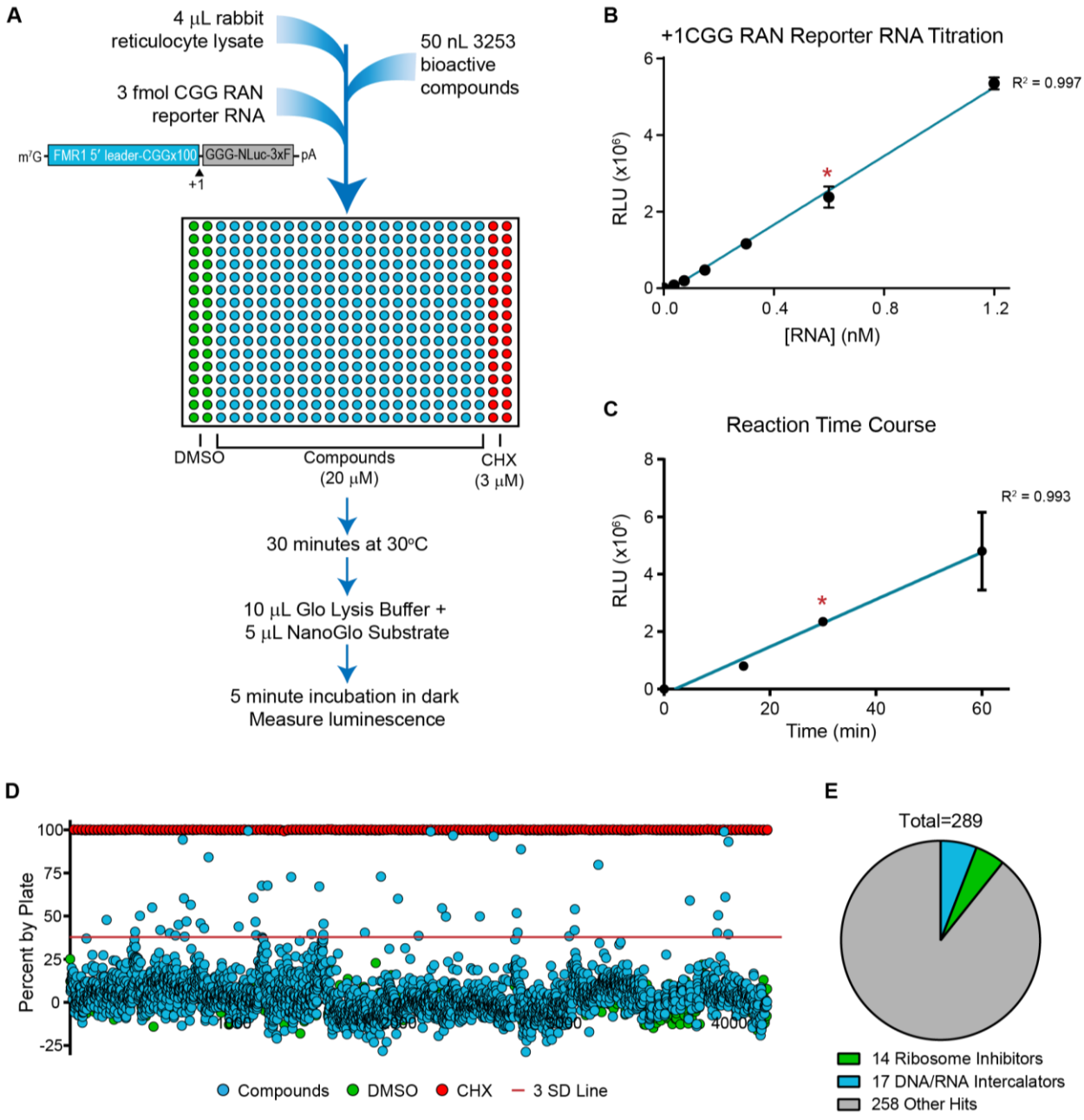
This work was funded by grants from the VA BLRD (1I21BX001841 and 1I01BX003231), the NIH (R01NS099280 and R01NS086810), and the Michigan Alzheimer's Disease Center and Protein Folding Disease Initiative to PKT, MII and AS. MII was additionally supported by the NIH R01NS096785-10. KMG was supported by NIH T32GM007315 and NIH F31NS100302. MGK was supported by NIH F32NS089124.

3.7 Conflict of interest statement

PKT serves as a consultant with Denali Therapeutics, and he, KMG, and MGK licensed technology through the University of Michigan to Denali Therapeutics that is based on the work presented here.

3.8 Figures

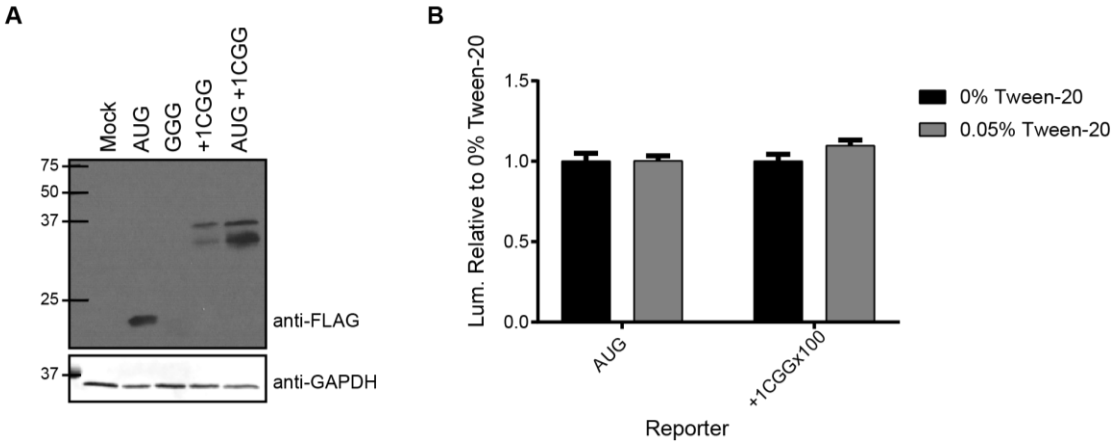
Figure 3-1. Screen of 3253 bioactive compounds for inhibitors of RAN translation



(A) Schematic of primary screen design. In each plate, DMSO (columns 1 and 2) served as an internal negative control and 3 µM cycloheximide (CHX, columns 23 and

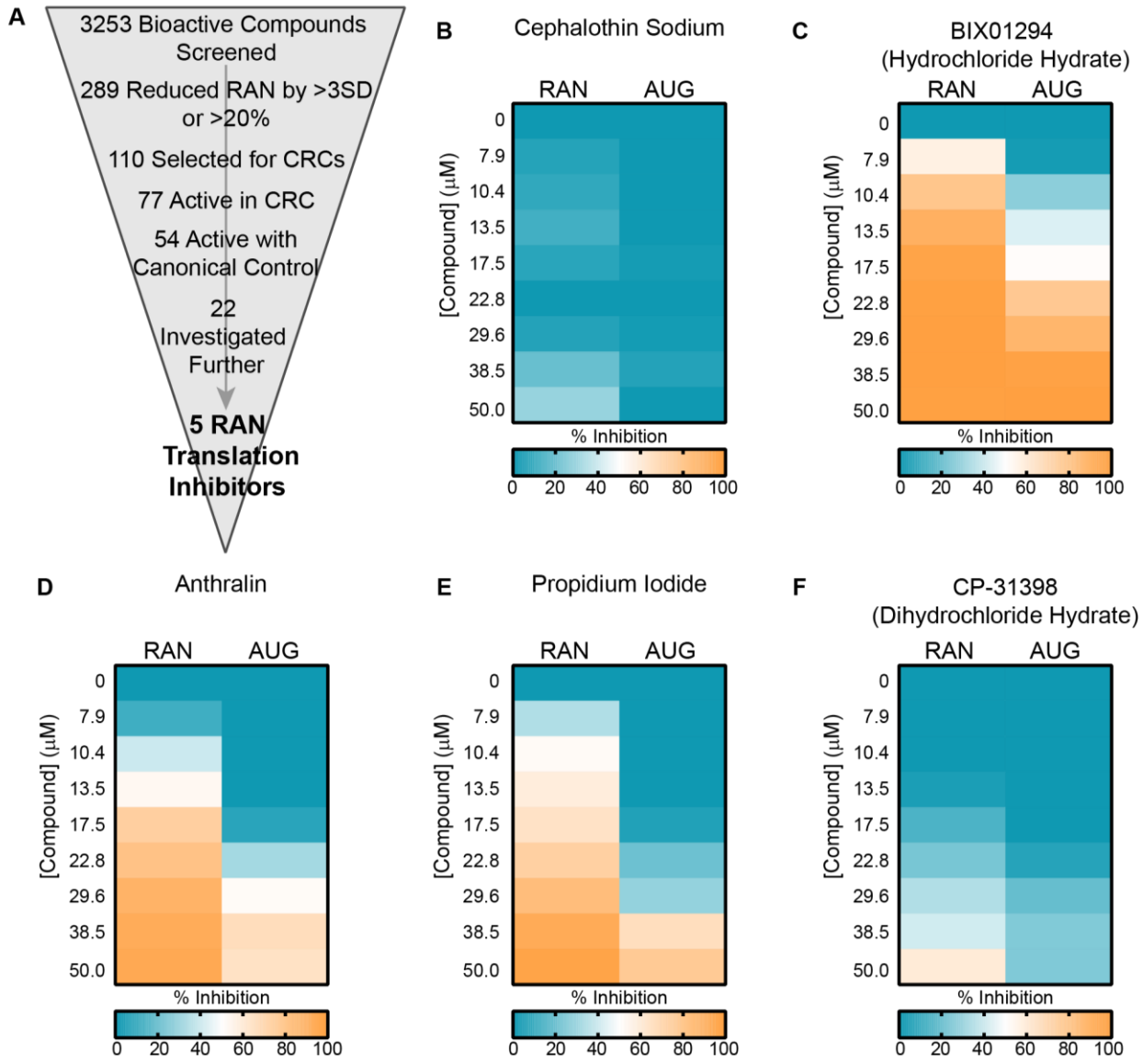
24) as an internal positive control for translation inhibition. Compounds screened were from the Pilot LOPAC, Pilot Prestwick, Pilot NCC-Focused, and Navigator Pathways libraries. 3xF = 3xFLAG tag (B) Linear relationship between +1CGG RAN reporter mRNA concentration and luminescence for *in vitro* translation assay under conditions used for primary screen. (C) Linear relationship between reaction time and luminescence for *in vitro* translation assay with 3fmol +1CGG nanoluciferase (NLuc) RAN reporter mRNA under conditions used for primary screen. For (B) and (C), teal lines represent linear regression fit, with red asterisk indicating conditions used in screen. Each point represents the mean of n=3, and error bars represent +/- standard deviation. (D) Percent change in NLuc signal for all 3,253 compounds relative to the DMSO vehicle controls. Red line approximately represents 3 standard deviations from vehicle controls. (E) Pie chart of screen hits, indicating the proportion of hits that are known ribosome inhibitors or DNA/RNA intercalators.

Figure 3-2: Primary screen controls



(A) When expressed in rabbit reticulocyte lysate (RRL), the +1CGGx100 NLuc reporter mRNA generates a RAN-specific FMRPolyG product that is detectable by western blot against its c-terminal FLAG tag. GAPDH was used as a loading control. To prevent over-exposure, one-half the amount of the AUG-NLuc reaction was loaded relative to the other samples. (B) Addition of 0.05% Tween-20 to RRL did not substantially alter total luminescence from AUG-NLuc and +1CGGx100 NLuc reactions, relative to 0% Tween-20. Graph represents mean of n=3 +/- standard deviation.

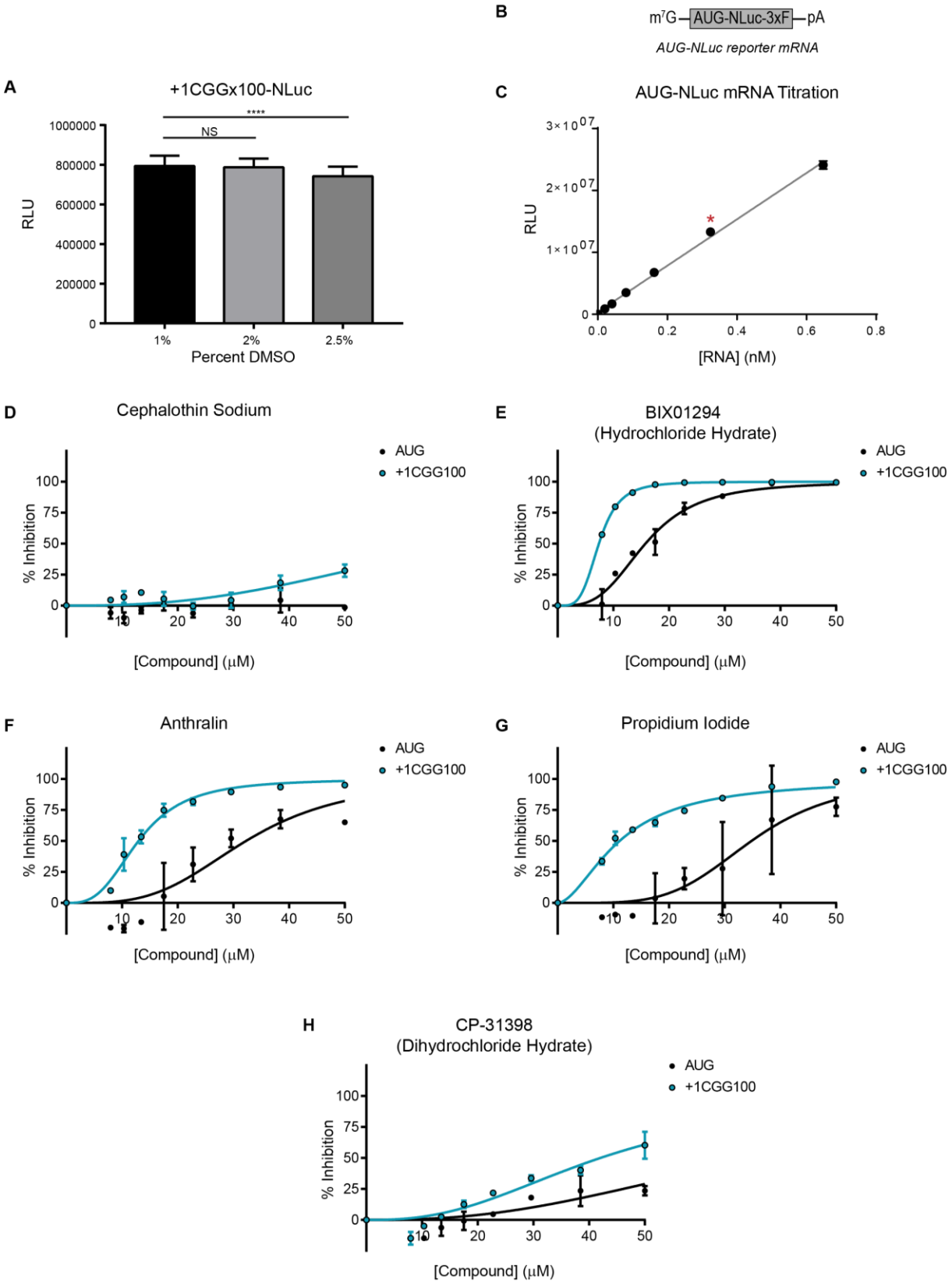
Figure 3-3. Concentration response curves and counter screen



(A) Schematic of screen work flow, with number of hits at each step indicated. SD = standard deviation, CRC = concentration response curve. (B) Representative heat map of a “RAN specific inhibitor,” cephalothin, showing percent inhibition of +1CGG RAN translation and canonical AUG translation at indicated concentrations. (C-F) Representative heat maps of “RAN selective inhibitors,” showing percent inhibition of +1CGG RAN translation and canonical AUG translation at indicated concentrations. For

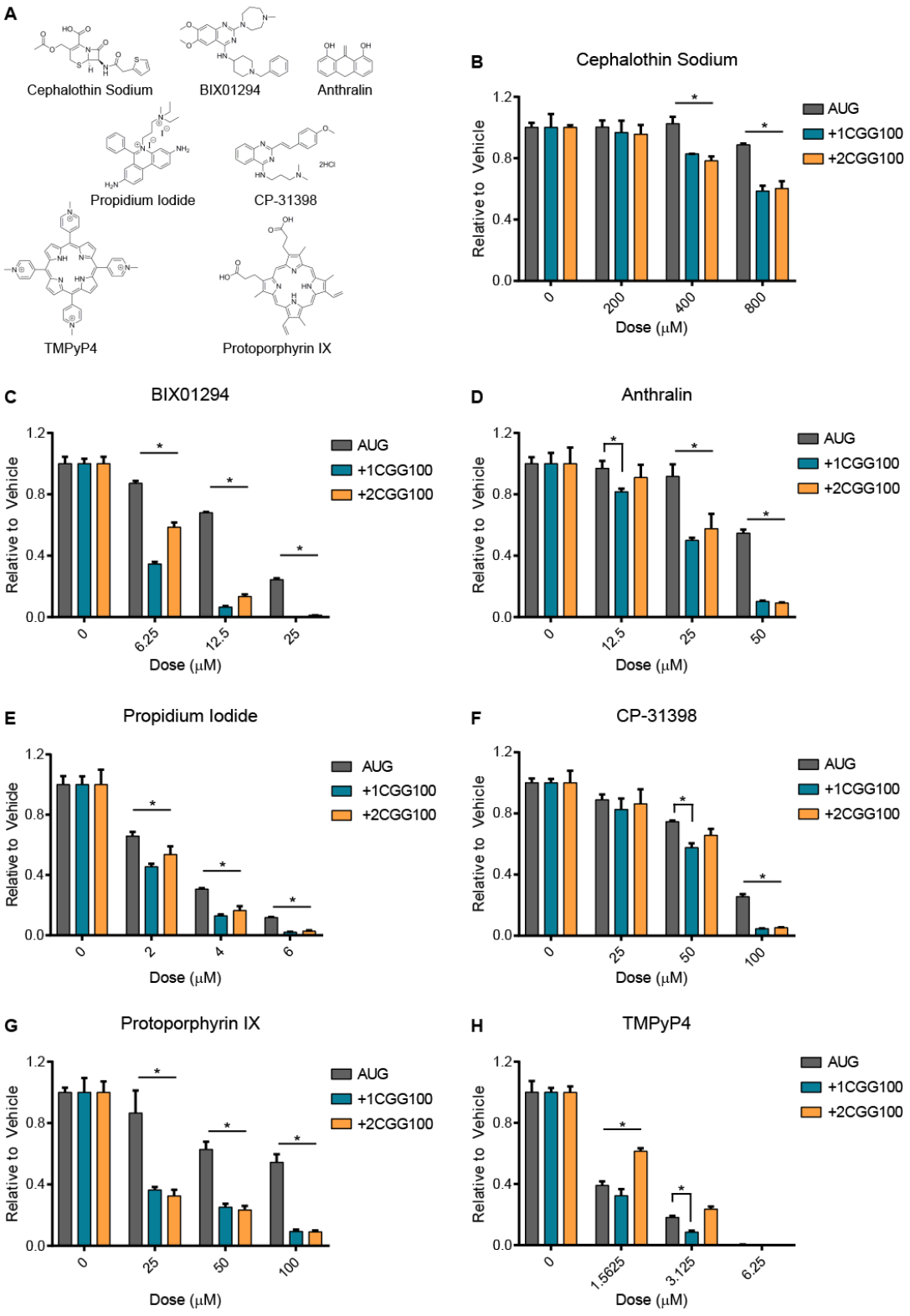
(B-F), heat map values are the average of independent duplicates. For visual clarity, effects of <0% inhibition (increases) were set to zero.

Figure 3-4: Concentration response curve and counter screen controls



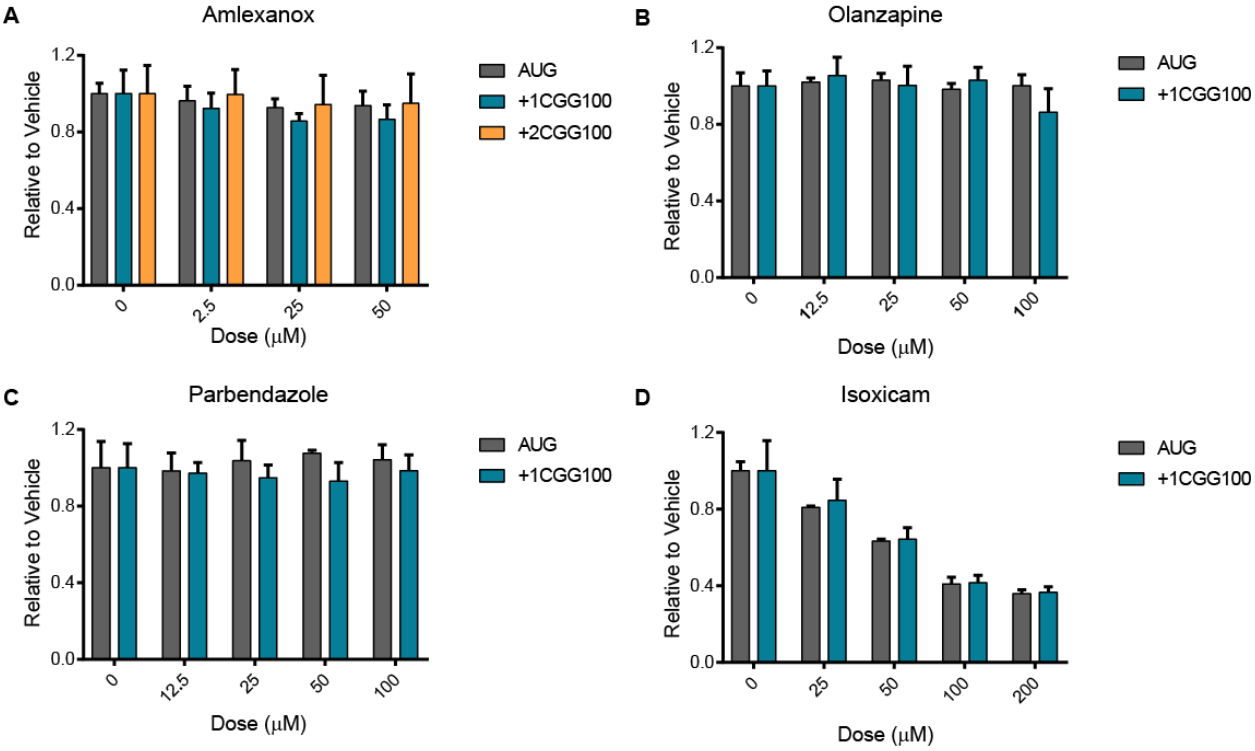
(A) 2% DMSO did not significantly reduce +1CGG RAN translation levels in rabbit reticulocyte lysate in vitro translation assay, relative to 1% DMSO. Graph represents mean of n=32, and error bars represent +/- standard deviation. **** p > 0.0001, one-way ANOVA with Dunnett's multiple comparison test. (B) Schematic of AUG-initiated nanoluciferase (AUG-NLuc) reporter mRNA used during screen and followup experiments. 3xF = 3x-FLAG tag. (C) 0.3 nM (1.5 fmol) AUG-NLuc reporter mRNA was added per well (red asterisks) in the counter screen. This concentration produced a robust signal within the linear range of detection. Each point represents the mean of n=14, and error bars represent +/- standard deviation. Data are fit with a linear regression curve; R2=0.9972. (D-E) Concentration response curves, showing percent inhibition, relative to DMSO, of AUG-NLuc and +1CGGx100-NLuc reporter mRNAs for each dose of compound, corresponding to heat maps in Fig. 2B-F. Each point is the average of n=2 +/- standard deviation. Data are fitted with four-parameter variable slope inhibition curves, with bottom and top constraints of 0% (DMSO treatment) and 100% inhibition, respectively.

Figure 3-5: Small molecules inhibit CGG RAN translation in multiple reading frames



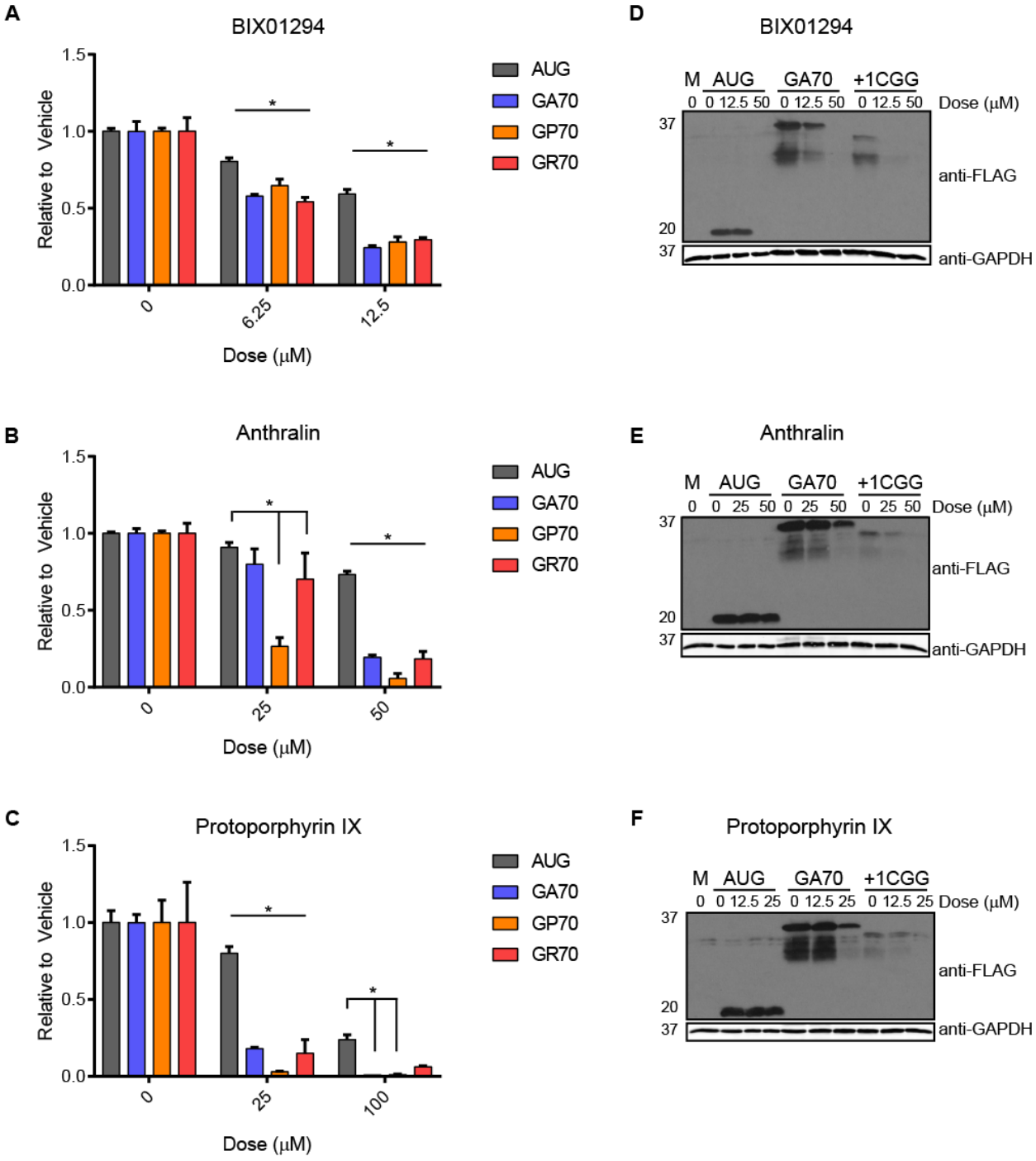
(A) Structures of compounds used in panels B-H. (B-F) Compounds identified from screen as RAN translation inhibitors were independently re-assessed for their activity in RRL *in vitro* translation assays. Indicated compounds were added at increasing doses to 30% RRL reactions with AUG, +1CGGx100, or +2CGGx100-NLuc reporter mRNAs and luminescence measured relative to vehicle (DMSO) treatment. (G-H) Candidates compounds protoporphyrin IX and TMPyP4 were added at increasing concentration to RRL reactions with AUG, +1CGGx100, or +2CGGx100-NLuc reporter mRNAs. Luminescence was measured relative to vehicle (DMSO) treatment. All graphs represent n=3 +/- standard deviation. *$p < 0.05$, 2-way ANOVA with Dunnett's multiple comparison test.

Figure 3-6: Representative compounds that failed independent validation for selective +1CGG RAN translation inhibition



In independent in-house trials, (A) amlexanox, (B) olanzapine, (C) parbendazole, failed to show selective inhibition of CGG RAN translation, relative to the AUG-NLuc control, and had no effect on expression of either reporter in RRL, at indicated doses. (D) Isoxicam also failed to replicate selective inhibition of +1CGG RAN translation during independent validation and inhibited +1CGG RAN and AUG-NLuc translation to a similar extent at the indicated doses. Graph represents mean of n=3 +/- standard deviation.

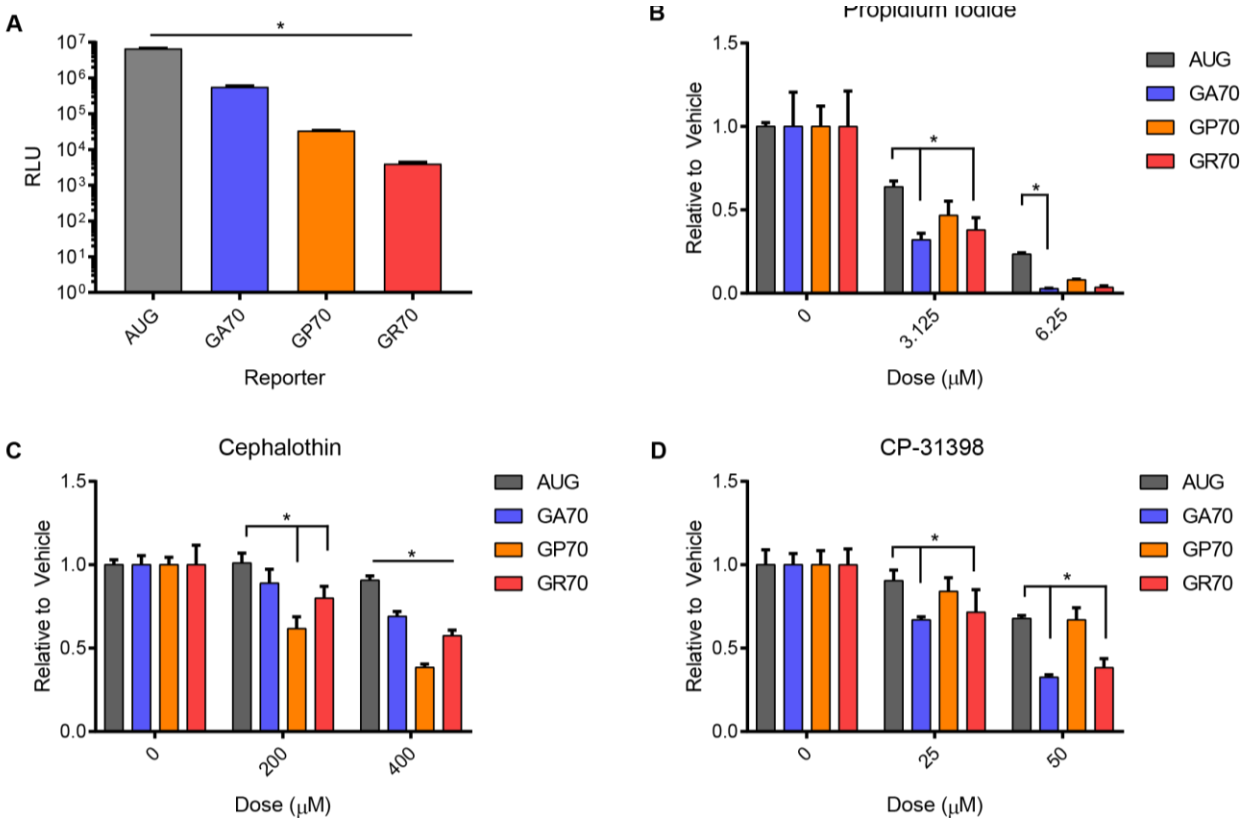
Figure 3-7: Small molecules inhibit C9RAN translation in all three reading frames



(A) BIX01294, (B) anthralin, and (C) protoporphyrin IX were added at increasing concentrations to 30% RRL *in vitro* translation reactions with AUG-NLuc or GGGGCCx70 reporter mRNAs for all three reading frames, and luminescence was

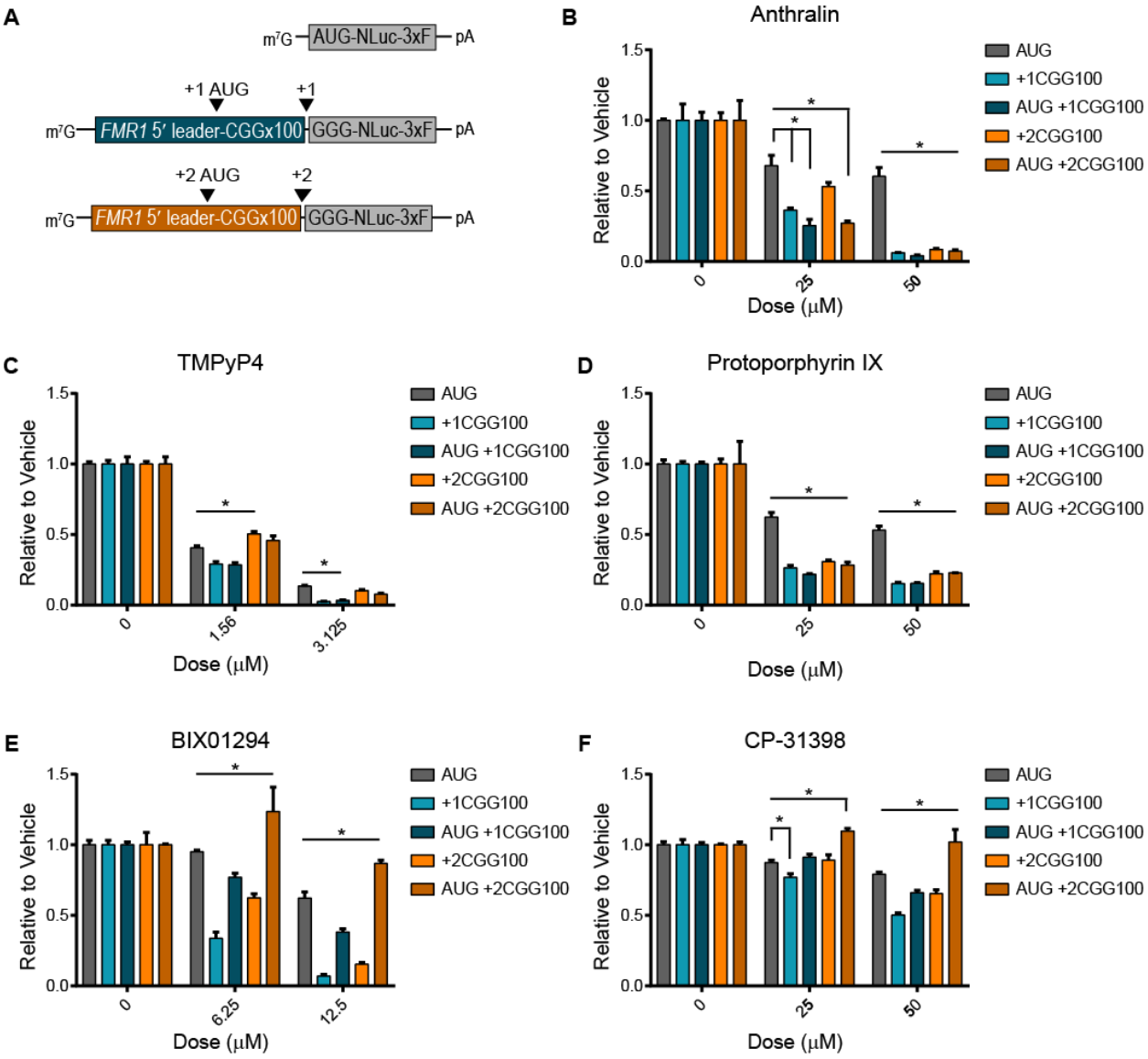
measured relative to vehicle (DMSO) control. (D-F) Selective inhibition of RAN translation of GGGGCCx70 and CGGx100 NLuc reporter mRNAs, relative to AUG-NLuc control, was confirmed by western blot against a c-terminal FLAG tag for the glycine-alanine (GA) and FMRPolyG (+1CGG) products, following incubation with BIX01294, anthralin, or protoporphyrin IX at indicated doses. GAPDH was used a loading control. For (A-C), all graphs represent mean of n=3 +/- standard deviation. *$p < 0.05$, 2-way ANOVA with Dunnett's multiple comparison test. M=mock, GA=glycine-alanine, GP=glycine-proline, GR=glycine-arginine.

Figure 3-8:Additional small molecule inhibitors of C9RAN translation



(A) Representative graph comparing expression levels of each C9RAN translation reporter mRNA to the AUG-NLuc control. RLU values are from experiment represented in Fig. 4A. (B) Propidium iodide, (C) cephalothin, and (D) CP-31398 selectively and dose-dependently inhibit C9RAN translation in multiple reading frames, relative to AUG-NLuc, when incubated in a 30% RRL lysate at indicated concentrations. All graphs represent mean of n=3 +/- standard deviation. For A, *p < 0.0001, one-way ANOVA with Dunnett's multiple comparison test. For B-D, values are relative to DMSO controls; *p < 0.05, 2-way ANOVA with Dunnett's multiple comparison test. RLU= relative light units, GA=glycine-alanine, GP=glycine-proline, GR=glycine-arginine.

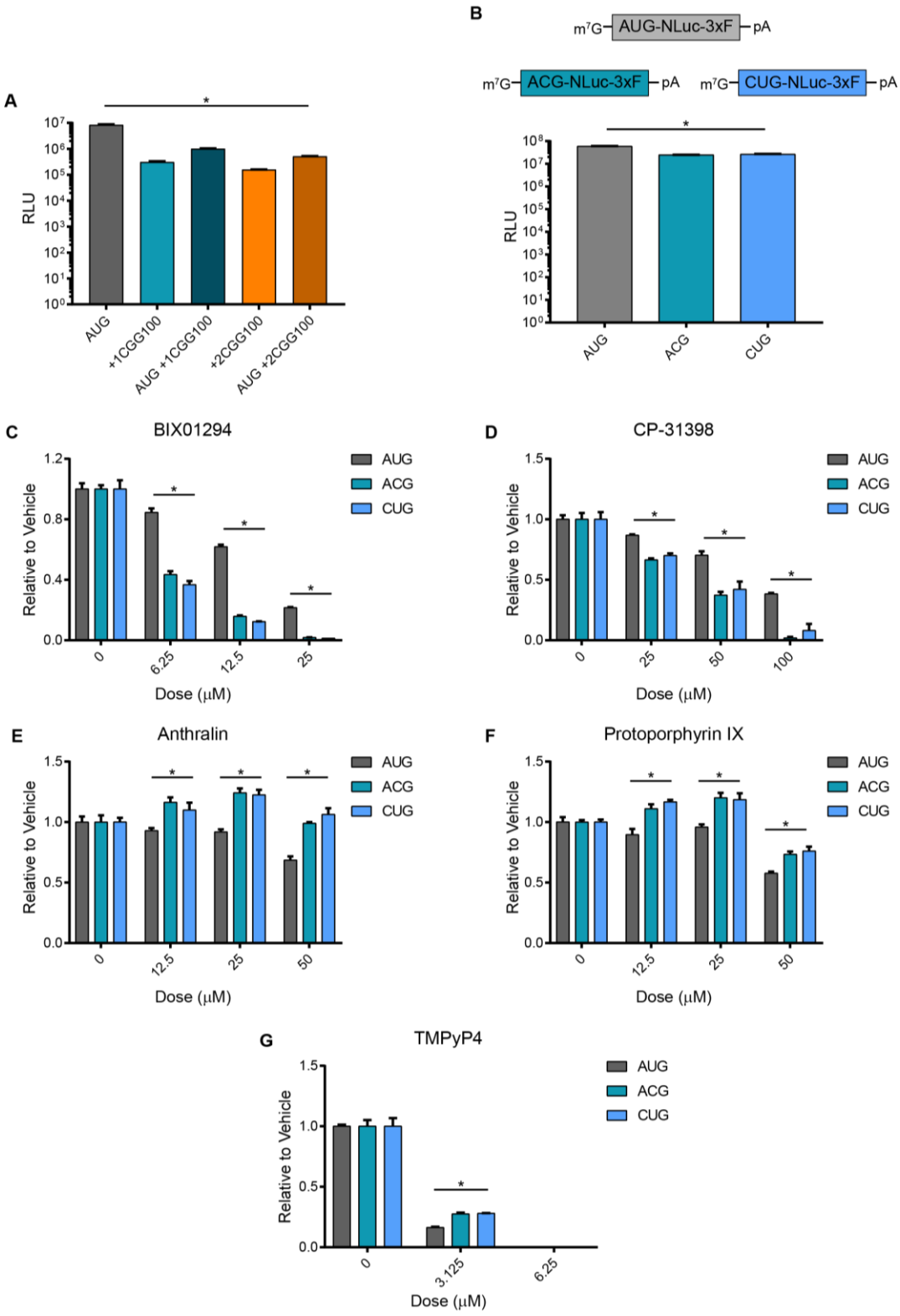
Figure 3-9. RAN translation inhibitors have varying effects on AUG-initiated repeat translation



(A) Schematic of nanoluciferase (NLuc) reporter mRNAs used in 30% RRL *in vitro* translation assays, illustrating the position of the AUG start codons inserted upstream of the repeats, to drive translation in the indicated reading frames. The effect of (B) anthralin, (C) TMPyP4, (D) protoporphyrin IX, (E) BIX01294, and (F) CP-31398 on +1 and +2CGGx100 RAN translation, compared to AUG-initiated CGGx100 translation,

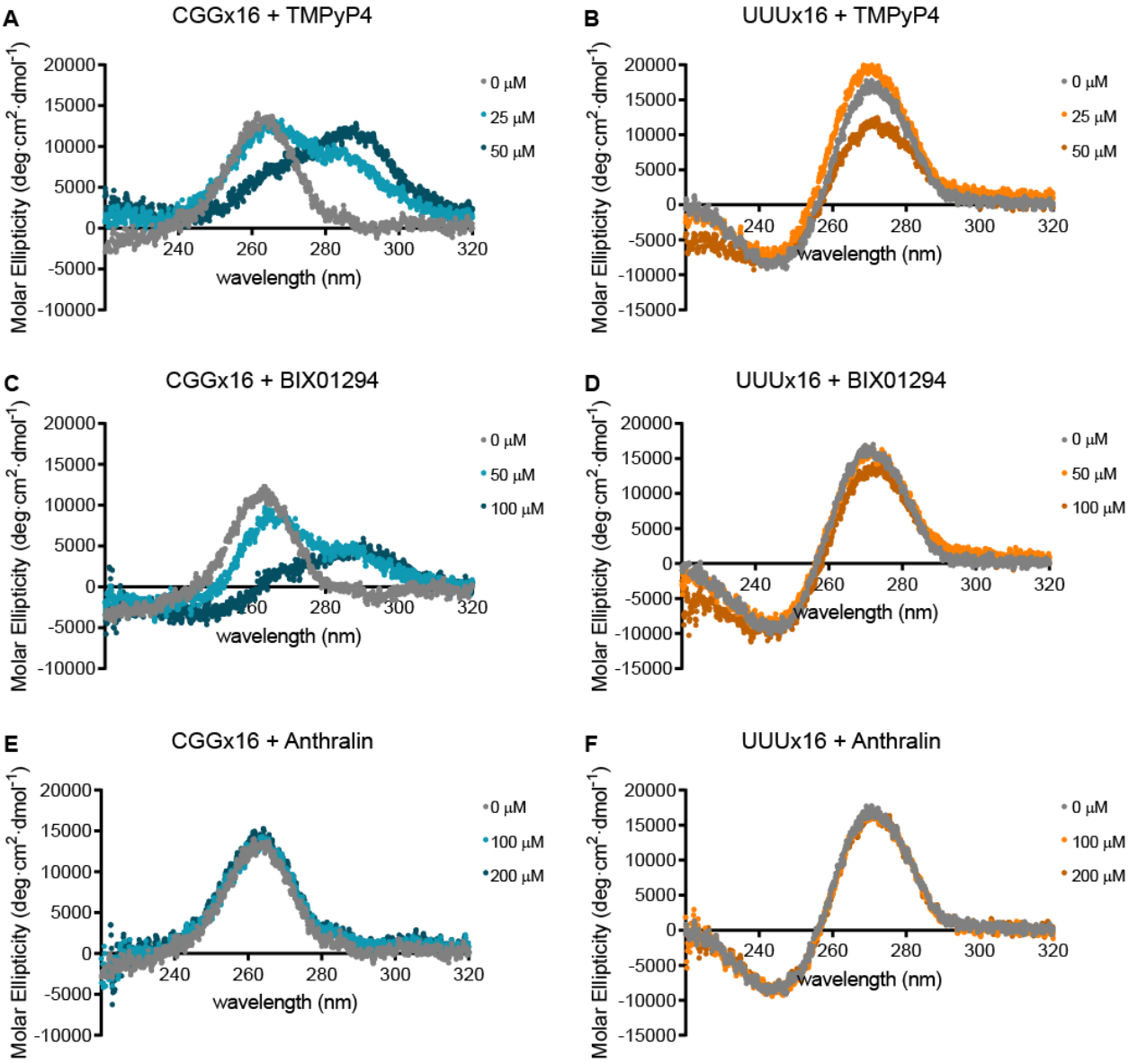
measured in the presence of increasing concentrations of each small molecule. For (B-F), values are relative to DMSO controls, and graphs represent the mean of n=3 +/- standard deviation. *$p < 0.05$, 2-way ANOVA with Dunnett's multiple comparison test. 3xF = 3xFLAG tag.

Figure 3-10: Small molecule inhibitors have varying effects on translation from near-AUG reporter RNAs



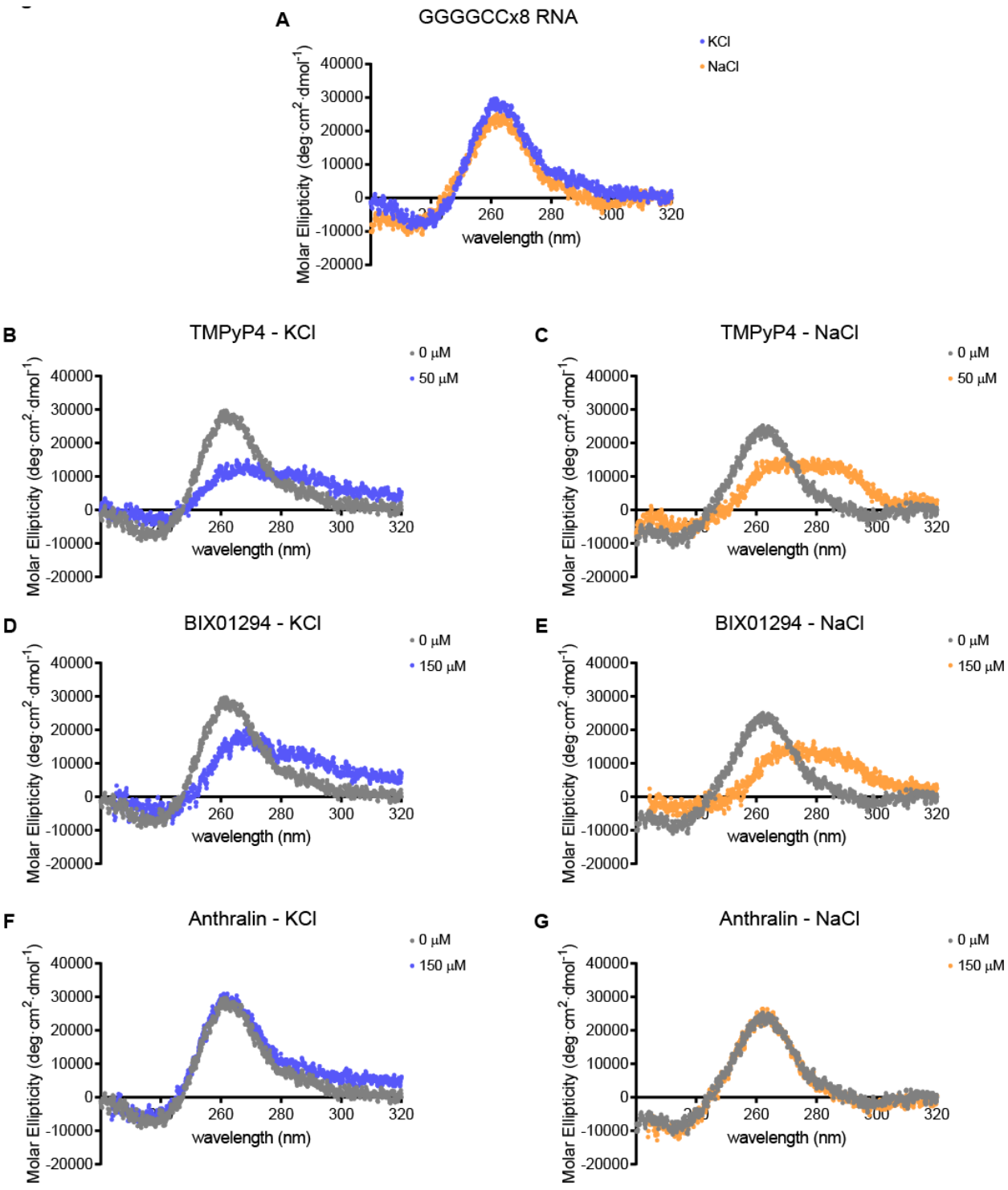
(A) Representative graph comparing expression levels of each CGG reporter mRNA to the AUG-NLuc control. RLU values are from experiment represented in Fig. 5F. (B) Schematic of near-AUG NLuc reporter mRNAs, and representative graph comparing their expression levels to the AUG-NLuc control. RLU values are from experiment represented in Supplementary Fig. 5C. The effect of (C) BIX01294, (D) CP-31398, (E) anthralin, (F) protoporphyrin IX, and (G) TMPyP4 on translation from ACG and CUG-NLuc synthesis, compared to AUG-NLuc control, in 30% RRL at indicated concentrations. All graphs represent the mean of n=3 +/- standard deviation. For A and B, *$p < 0.0001$, one-way ANOVA with Dunnett's multiple comparison test. For C-G, values are relative to DMSO controls; *$p < 0.001$, 2-way ANOVA with Dunnett's multiple comparison test. RLU= relative light units, 3xF = 3xFLAG tag

Figure 3-11. Circular dichroism analysis for small molecule interactions with CGG repeat RNA



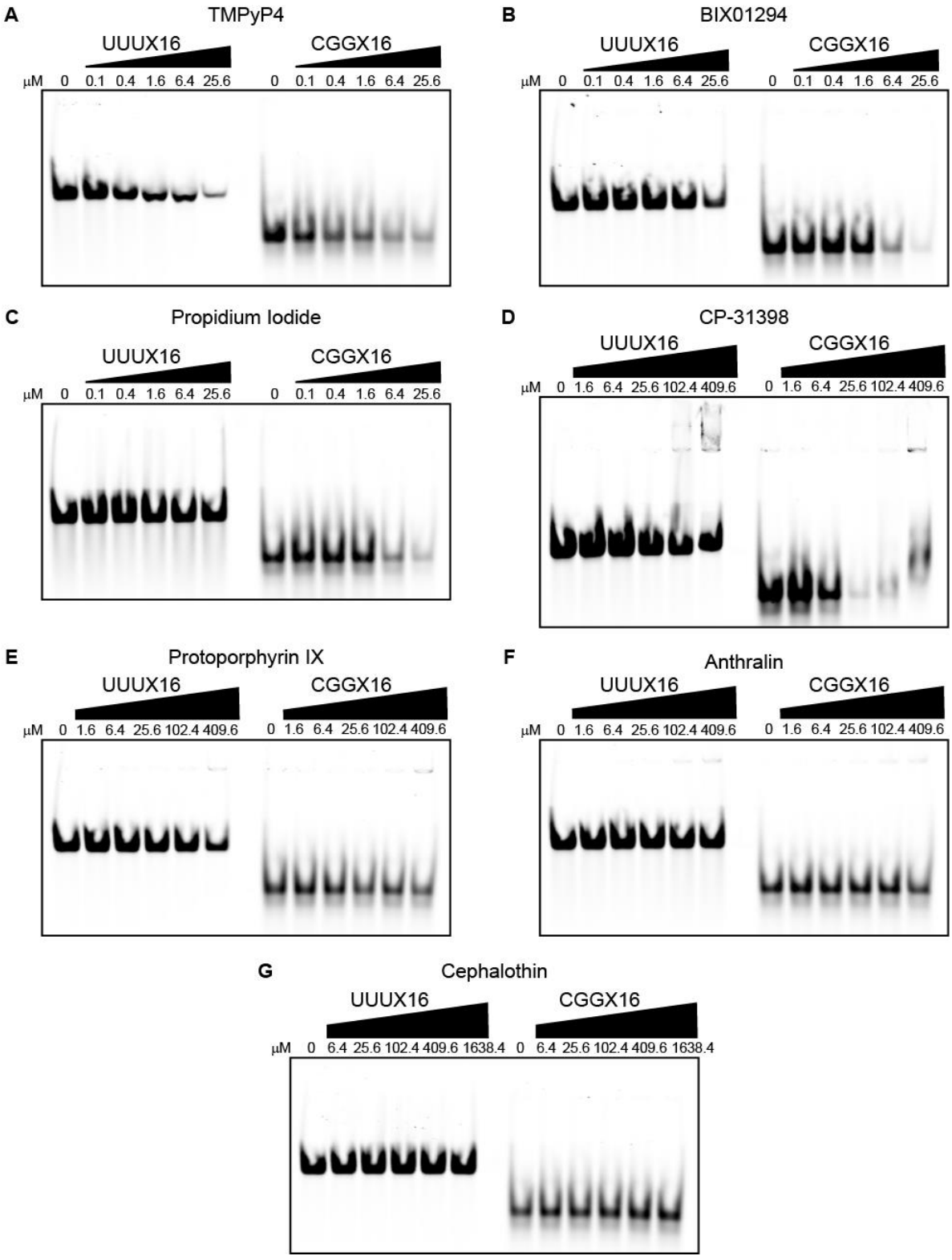
Molar ellipticity of 5 µM CGGx16 or UUUx16 RNA, measured between 320-220 nm with circular dichroism (CD), following incubation with TMPyP4 (A-B), BIX01294 (C-D), or anthralin (E-F) at the indicated concentrations. All CD spectra were acquired at 25ºC, with 300 µL RNA samples folded in the presence of 100 mM KCl, and represent the average of six accumulations collected at 20 nm/minute.

Figure 3-12: Interaction of small molecule RAN translation inhibitors with GGGGCC repeat RNA by circular dichroism



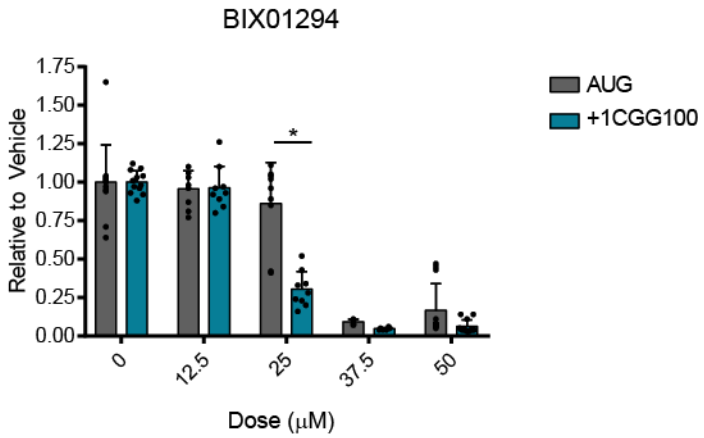
(A) Comparison of the molar ellipticity of 2.5 µM GGGGCCx8 RNA, measured between 320-220 nm with circular dichroism (CD), following folding in the presence of 100 mM KCl (blue) or 100 mM NaCl (orange). (B-G) CD spectra of GGGGCCx8 RNA folded in the presence of either 100 mM KCl (blue) or 100 mM NaCl (orange), following incubation with the indicated concentrations of TMPyP4 (B-C), BIX01294 (D-E), or anthralin (F-G). All CD spectra were acquired at 25ºC, with 250 µL RNA samples, and represent the average of six accumulations collected at 20 nm/minute. The untreated RNA spectra in (A) are used for comparison to the RNA+compound spectra in (B-G).

Figure 3-13. Native gel analysis for small molecule interactions with CGG repeat RNA



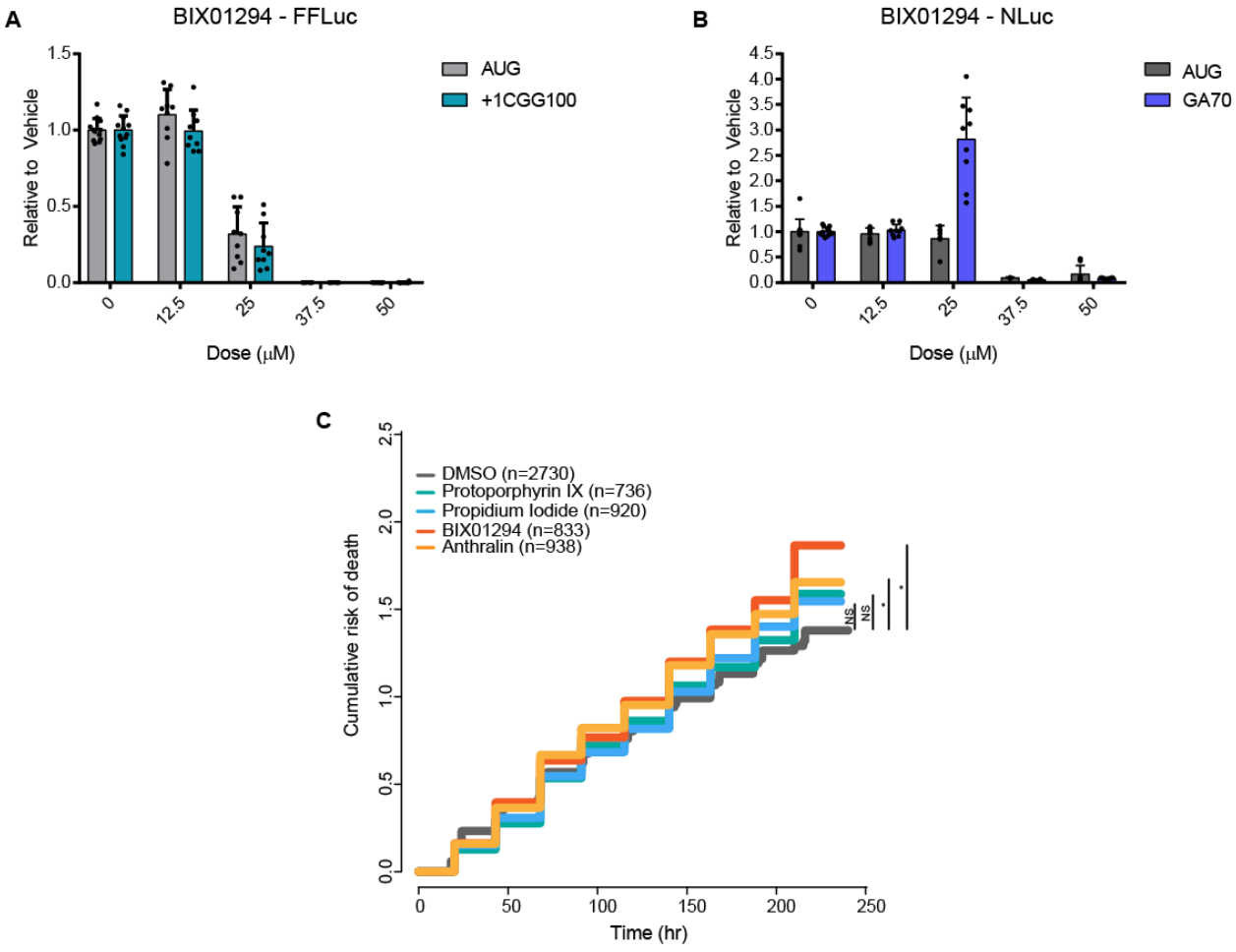
(A-G) Native gel analysis of 5′ IRDye 800CW labeled-CGGx16 or UUUx16 RNA following incubation with the indicated small molecules, at the indicated concentrations, for 30 minutes at room temperature. RNA was imaged at 800nm. TMPyP4 was used as a positive control for CGG repeat-RNA binding, with increasing doses causing a decrease in CGGx16 and UUUx16 RNA band intensity at 800nm (A).

Figure 3-14. Effect of BIX01294 on +1CGG RAN translation in HEK293 cells



NLuc levels from HEK293 cells transfected with AUG-NLuc or +1CGGx100-NLuc expressing plasmids and treated with BIX01294 at the indicated doses for 24 hours. Luminescence values are expressed relative to DMSO treated cells (0 µM). Graphs represent mean +/- standard deviation, with each n represesnted by a dot. n=12 for 0 and 50 µM doses, n =9 for 12.5 and 25 µM doses, and n=6 for 37.5 µM, from 2-4 independent experiments with n=3 each. *p < 0.0001, 2-way ANOVA with Sidak's multiple comparison test.

Figure 3-15. Small molecule inhibitors are toxic in cultured cells



(A) Firefly luciferase (FFLuc) levels from HEK293 cells co-transfected with either AUG-NLuc or +1CGGx100-NLuc reporter plasmids and treated with BIX01294 at indicated doses for 24 hours. (B) Nanoluciferase (NLuc) expression of AUG-NLuc and poly-GA70-NLuc expressing plasmids in HEK293 cells following treatment with BIX01294 for 24 hours. AUG-NLuc values are the same as those reported in Fig. 8, as experiments with all three reporters were performed simultaneously in the same plates. For A and B, luminescence values are expressed relative to DMSO treated cells (0 μM). Graphs represent mean +/- standard deviation, with each n represesnted by a dot. n=12 for 0

and 50 μM doses, n =9 for 12.5 and 25 μM doses, and n=6 for 37.5 μM, from 2-4 independent experiments with n=3 each. (C) Small molecule inhibitors BIX01294, anthralin, propidium iodide, and protoporphyrin IX were added at 1 μM to primary cortical rat neurons transfected with fluorescent markers. Neuron survival was tracked over the course of 10 days, and the cumulative risk of death calculated for each condition. *p < 0.05, cox proportional hazards analysis. GA=glycine-alanine.

3.9 Tables

Table 3-1. Primary screen statistics

Z′ Scores per Plate	
Average	0.79
Minimum	0.70
Maximum	0.88
Coefficients of Variance per Plate	
Average	6.93%
Minimum	3.88%
Maximum	9.82%

Table 3-2. Summary of primary screen hits

Signal Reduction Range	**Number of Compounds**
20-39%	154
40-59%	25
60-79%	9
80-99%	8
>3 SD Below Mean	211
Total (>20% or >3 SD)	274

Table 3-3. Activity of compounds in independent validation

Compound Name	ID Number	Active against +1CGGx100?	Active against AUG?
Kenpaullone	CCG-35778	No	No
BIX01294 (hydrochloride hydrate)	**CCG-208677**	**Yes**	**Less**
Propidium Iodide	**CCG-220792**	**Yes**	**Less**
Amlexanox	CCG-100953	No	No
Anthralin	**CCG-38920**	**Yes**	**Less**
Rosiglitazone Maleate	CCG-100943	No	No
X80	CCG-222138	Yes	Yes
Balsalazide Disodium	CCG-213078	No	No
CP-31398 dihydrochloride hydrate	**CGG-222578**	**Yes**	**Less**
Phenazopyridine hydrochloride	CGG-39935	No	No
Rufloxacin hyrdochloride	CGG-100908	No	No
Reserpine	CGG-204168	No	No
Efavirenz	CGG-101011	No	No
Alfuzosin hydrochloride	CGG-213373	No	No
Cephalothin Sodium	**CCG-38923**	**Yes**	**Less**
Rofecoxib	CGG-40253	No	No
Isoxicam	CCG-40307	Yes	Yes
Pefloxacine Mesylate	CCG-39994	No	No
Parbendazole	CCG-221110	No	No
Sulfamethazine sodium salt	CCG-220775	No	No
Olanzapine	CCG-100922	No	No
Oxiconazole Nitrate	CCG-40020	No	No

Table 3-4. NLuc reporter mRNA sequences

Reporter Name	DNA sequence to 3′ PspOMI cut site, excluding addition of poly-A tail
+1CGG100-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGTAC CGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTTAAC AGATCTGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGG CCGCCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGGC GGCGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTGCGG CAGCGCGGCGGCGGCGGCGGCGGCGGCGGAGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC

	GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCTGGG CCTCGAGATTCGATCGTCGTGATATCAAGATCTGGCCTCGGC GGCCAAGCTTGGCAATCCGGTACTGTTGGTAAAGCCTCCCTC GAGGTCCTCTTCCAGGGACCCGGGGTCTTCACACTCGAAGAT TTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGA CCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAAT CTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGC GGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCG TATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAA ATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGG TGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGC CGAACATGATCGACTATTTCGGACGGCCGTATGAAGGCATCG CCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGT GGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCG ACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCG GCTGGCGGCTGTGCGAACGCATTCTGGCGGACTACAAAGAC CATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATG ACGATGACAAGTAAGGCCGCGACTCTAGAG
AUG-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGCAA TCCGGTACTGTTGGTAAAGCCACCATGGTCTTCACACTCGAA GATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTG GACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAG AATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTG AGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATC CCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGA AAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTT AAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTT ACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGC ATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACC CTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAAC CCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTG ACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGACTACAA AGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG GATGACGATGACAAGTAAGGCCGCGACTCTAGAG
+2CGG100- NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGTAC CGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTTAAC AGATCTGCTCAGCTCCGTTTCGGTTTCACTTCCGGTGGAGGG CCGCCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGGC GGCGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTGCGG CAGCGCGGCGGCGGCGGCGGCGGCGGCGGAGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC

	GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGGGCCTCG AGATTCGATCGTCGTGATATCAAGATCTGGCCTCGGCGGCCA AGCTTGGCAATCCGGTACTGTTGGTAAAGCCTCCCTCGAGGT CCTCTTCCAGGGACCCGGGGTCTTCACACTCGAAGATTTCGT TGGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAG TCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCG GGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTG AAAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGA AGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTT TAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATC CTGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAAC ATGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTG TTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAAC GGCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGC TCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGG CGGCTGTGCGAACGCATTCTGGCGGACTACAAAGACCATGAC GGTGATTATAAAGATCATGACATCGATTACAAGGATGACGATG ACAAGTAAGGCCGCGACTCTAGAG
Intron-GA70-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTGTGTGTTTTTGTTTTTCCC ACCCTCTCTCCCCACTACTTGCTCTCACAGTACTCGCTGAGG GTGAACAAGAAAAGACCTGATAAAGATTAACCAGAAGAAAACA AGGAGGGAAACAACCGCAGCCTGTAGCAAGCTCTGGAACTCA GGAGTCGCGCGCTAGCGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCAGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGTCGTGGAAGGGTGGGCGCGCCCACCGG TCTCGAGGTCCTCTTCCAGGGACCCGATGGGGTCTTCACACT CGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAA CCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTT TCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGT CCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCAT CATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGAT CGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCAC TTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGG GTTACGCCGAACATGATCGACTATTTCGGACGGCCCTATGAA GGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGG ACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATC

	AACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGA GTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGACTA CAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTAC AAGGATGACGATGACAAGTAAGGCCGCGACTCGAGAG
Intron-GP70-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTGTGTGTTTTTGTTTTTCCC ACCCTCTCTCCCCACTACTTGCTCTCACAGTACTCGCTGAGG GTGAACAAGAAAAGACCTGATAAAGATTAACCAGAAGAAAACA AGGAGGGAAACAACCGCAGCCTGTAGCAAGCTCTGGAACTCA GGAGTCGCGCGCTAGCGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCAGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGTCGTGGAAGGGTGGGCGCGCCCACCGG TCCTCGAGGTCCTCTTCCAGGGACCCGATGGGGTCTTCACAC TCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACA ACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGT TTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTG TCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCA TCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGA TCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCA CTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGG GGTTACGCCGAACATGATCGACTATTTCGGACGGCCCTATGA AGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGG GACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGAT CAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGG AGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGACT ACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTA CAAGGATGACGATGACAAGTAAGGCCGCGACTCGAGAG
Intron-GR70-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTGTGTGTTTTTGTTTTTCCC ACCCTCTCTCCCCACTACTTGCTCTCACAGTACTCGCTGAGG GTGAACAAGAAAAGACCTGATAAAGATTAACCAGAAGAAAACA AGGAGGGAAACAACCGCAGCCTGTAGCAAGCTCTGGAACTCA GGAGTCGCGCGCTAGCGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCAGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG

	GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGGCCGGGGCCGGTCGTGGAAGGGTGGGCGCGCCCACCGG TCCCTCGAGGTCCTCTTCCAGGGACCCGATGGGGTCTTCACA CTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTAC AACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTG TTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATT GTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTC ATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAG ATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATC ACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACG GGGTTACGCCGAACATGATCGACTATTTCGGACGGCCCTATG AAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAG GGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGA TCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACG GAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGAC TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATT ACAAGGATGACGATGACAAGTAAGGCCGCGACTCGAGAG
AUG +1CGG100- NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGTAC CGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTTACA CCATGGCGCCGCTGCCAGGGGGCGTGCGGCAGCGCGGCGG CGGCGGCGGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGG CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG CGGCGGCGGCGGCGGCGGCGGCGGCTGGGCCTCGAGGATA TCAAGATCTGGCCTCGGCGGCCAAGCTTGGCAATCCGGTACT GTTGGTAAAGCCACCGGGGTCTTCACACTCGAAGATTTCGTT GGGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGT CCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGG GGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGA AAATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAA GGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTT AAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCC TGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACA TGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGT TCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACG GCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCT CCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGC GGCTGTGCGAACGCATTCTGGCGGACTACAAAGACCATGACG

	GTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGA CAAGTAAGGCCGCGACTCTAGAG
AUG +2CGG100- NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGTAC CGAGCTCGGATCCACTAGTCCAGTGTGGTGGAATTCGTTACA CCATGGGGCGCCGCTGCCAGGGGGCGTGCGGCAGCGCGGC GGCGGCGGCGGCGGCGGCGGAGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCGGCGGCGGCGGCGGCGGCGGCTGGGCCTCGAGAT TCGATCGTCGTGATATCAAGATCTGGCCTCGGCGGCCAAGCT TGGCAATCCGGTACTGTTGGTAAAGCCACCGGGGTCTTCACA CTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTAC AACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTG TTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATT GTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTC ATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAG ATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATC ACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACG GGGTTACGCCGAACATGATCGACTATTTCGGACGGCCGTATG AAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAG GGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGA TCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACG GAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGAC TACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATT ACAAGGATGACGATGACAAGTAAGGCCGCGACTCTAGAG
AUG Intron- GA70-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTGTGTGTTTTTGTTTTTCCC ACCCTCTCTCCCCACTACTTGCTCTCACAGTACTCGCTGAGG GTGAACAAGAAAAGACCTGATAAAGATTAACCAGAAGAAAACA AGGAGGGAAACAACCGCAGCCTGTAGCAAGCTCTGGAACTCA GGATGAGTCGCGCGCTATCTAGAGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCAGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGTCTAGAAGCTTGGCAATCCGGTACTG TTGGTAAAGCCACCGGGGTCTTCACACTCGAAGATTTCGTTG

	GGGACTGGCGACAGACAGCCGGCTACAACCTGGACCAAGTC CTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAATCTCGGG GTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGCGGTGAA AATGGGCTGAAGATCGACATCCATGTCATCATCCCGTATGAA GGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAAATTTTT AAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCC TGCACTATGGCACACTGGTAATCGACGGGGTTACGCCGAACA TGATCGACTATTTCGGACGGCCGTATGAAGGCATCGCCGTGT TCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACG GCAACAAAATTATCGACGAGCGCCTGATCAACCCCGACGGCT CCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCGGCTGGC GGCTGTGCGAACGCATTCTGGCGGACTACAAAGACCATGACG GTGATTATAAAGATCATGACATCGATTACAAGGATGACGATGA CAAGTAAGGCCGCGACTCGAGAG
CUG-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGCAA TCCGGTACTGTTGGTAAAGCCACCCTGGTCTTCACACTCGAA GATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTG GACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAG AATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTG AGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATC CCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGA AAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTT AAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTT ACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGC ATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACC CTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAAC CCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTG ACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGACTACAA AGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG GATGACGATGACAAGTAAGGCCGCGACTCTAGAG
ACG-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGCAA TCCGGTACTGTTGGTAAAGCCACCACGGTCTTCACACTCGAA GATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTG GACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAG AATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTG AGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATC CCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGA AAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTT AAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGGGTT ACGCCGAACATGATCGACTATTTCGGACGGCCGTATGAAGGC ATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACC CTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATCAAC CCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTG ACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGACTACAA AGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAG GATGACGATGACAAGTAAGGCCGCGACTCTAGAG

3.10 References

1. Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., Nan, Z., Forster, C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* **108**, 260-265
2. Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. C., Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., and Paulson, H. L. (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440-455
3. Krans, A., Kearse, M. G., and Todd, P. K. (2016) Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Ann Neurol* **80**, 871-881
4. Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646
5. Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C., and

Edbauer, D. (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335-1338

6. Mori, K., Arzberger, T., Grässer, F. A., Gijselinck, I., May, S., Rentzsch, K., Weng, S. M., Schludi, M. H., van der Zee, J., Cruts, M., Van Broeckhoven, C., Kremmer, E., Kretzschmar, H. A., Haass, C., and Edbauer, D. (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* **126**, 881-893
7. Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., Ostrow, L. W., Rothstein, J. D., Troncoso, J. C., and Ranum, L. P. (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-4977
8. Zu, T., Cleary, J. D., Liu, Y., Bañez-Coronel, M., Bubenik, J. L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H. B., Yachnis, A. T., Swanson, M. S., and Ranum, L. P. W. (2017) RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. *Neuron* **95**, 1292-1305.e1295
9. Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., Pletnikova, O., Borchelt, D. R., Ross, C. A., Margolis, R. L., Yachnis, A. T., Troncoso, J. C., and Ranum, L. P. (2015) RAN Translation in Huntington Disease. *Neuron* **88**, 667-677
10. Buijsen, R. A., Visser, J. A., Kramer, P., Severijnen, E. A., Gearing, M., Charlet-Berguerand, N., Sherman, S. L., Berman, R. F., Willemsen, R., and Hukema, R.

K. (2016) 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

11. Soragni, E., Petrosyan, L., Rinkoski, T. A., Wieben, E. D., Baratz, K. H., Fautsch, M. P., and Gottesfeld, J. M. (2018) Repeat-Associated Non-ATG (RAN) Translation in Fuchs' Endothelial Corneal Dystrophy. *Invest Ophthalmol Vis Sci* **59**, 1888-1896
12. Jacquemont, S., Hagerman, R. J., Leehey, M., Grigsby, J., Zhang, L., Brunberg, J. A., Greco, C., Des Portes, V., Jardini, T., Levine, R., Berry-Kravis, E., Brown, W. T., Schaeffer, S., Kissel, J., Tassone, F., and Hagerman, P. J. (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am J Hum Genet* **72**, 869-878
13. Jacquemont, S., Hagerman, R. J., Leehey, M. A., Hall, D. A., Levine, R. A., Brunberg, J. A., Zhang, L., Jardini, T., Gane, L. W., Harris, S. W., Herman, K., Grigsby, J., Greco, C. M., Berry-Kravis, E., Tassone, F., and Hagerman, P. J. (2004) Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA* **291**, 460-469
14. Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P. J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* **57**, 127-130
15. Sellier, C., Buijsen, R. A., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziane, H., Vincent, A., Champy, M. F., Sorg, T., Pavlovic, G.,

Wattenhofer-Donze, M., Birling, M. C., Oulad-Abdelghani, M., Eberling, P., Ruffenach, F., Joint, M., Anheim, M., Martinez-Cerdeno, V., Tassone, F., Willemsen, R., Hukema, R. K., Viville, S., Martinat, C., Todd, P. K., and Charlet-Berguerand, N. (2017) Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron* **93**, 331-347

16. Buijsen, R. A., Sellier, C., Severijnen, L. A., Oulad-Abdelghani, M., Verhagen, R. F., Berman, R. F., Charlet-Berguerand, N., Willemsen, R., and Hukema, R. K. (2014) 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

17. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256

18. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O.,

Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V. M., Kaivorinne, A. L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C., Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J., Traynor, B. J., and Consortium, I. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268

19. Sareen, D., O'Rourke, J. G., Meera, P., Muhammad, A. K., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., Gendron, T., Petrucelli, L., Baughn, M., Ravits, J., Harms, M. B., Rigo, F., Bennett, C. F., Otis, T. S., Svendsen, C. N., and Baloh, R. H. (2013) Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med* **5**, 208ra149

20. Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., Moens, T., Norona, F. E., Woollacott, I. O., Pietrzyk, J., Cleverley, K., Nicoll, A. J., Pickering-Brown, S., Dols, J., Cabecinha, M., Hendrich, O., Fratta, P., Fisher, E. M., Partridge, L., and Isaacs, A. M. (2014) C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192-1194

21. Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., Lin, S., Shneider, N. A., Monaghan, J., Pandey, U. B., Pasinelli, P., Ichida, J. K., and Trotti, D. (2014) 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

22. Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., and McKnight, S. L. (2014) Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139-1145

23. Zhang, Y. J., Jansen-West, K., Xu, Y. F., Gendron, T. F., Bieniek, K. F., Lin, W. L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., Chew, J., Belzil, V. V., Prudencio, M., Stankowski, J. N., Castanedes-Casey, M., Whitelaw, E., Ash, P. E., DeTure, M., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2014) Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* **128**, 505-524

24. May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grässer, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., Mann, M., Meissner, F., and Edbauer, D. (2014) C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* **128**, 485-503

25. Flores, B. N., Dulchavsky, M. E., Krans, A., Sawaya, M. R., Paulson, H. L., Todd, P. K., Barmada, S. J., and Ivanova, M. I. (2016) Distinct C9orf72-Associated

Dipeptide Repeat Structures Correlate with Neuronal Toxicity. *PLoS One* **11**, e0165084

26. Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., and Todd, P. K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**, 314-322
27. Green, K. M., Glineburg, M. R., Kearse, M. G., Flores, B. N., Linsalata, A. E., Fedak, S. J., Goldstrohm, A. C., Barmada, S. J., and Todd, P. K. (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications* **8**, 2005
28. Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C. Z., Lin, C. C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., Désaubry, L., Gendron, T., Petrucelli, L., Martin, F., and Lagier-Tourenne, C. (2018) CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* **9**, 152
29. Cheng, W., Wang, S., Mestre, A. A., Fu, C., Makarem, A., Xian, F., Hayes, L. R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., Cleveland, D. W., and Sun, S. (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat Commun* **9**, 51
30. Mackenzie, I. R., Arzberger, T., Kremmer, E., Troost, D., Lorenzl, S., Mori, K., Weng, S. M., Haass, C., Kretzschmar, H. A., Edbauer, D., and Neumann, M. (2013) Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* **126**, 859-879

31. Sonobe, Y., Ghadge, G., Masaki, K., Sendoel, A., Fuchs, E., and Roos, R. P. (2018) Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol Dis* **116**, 155-165

32. Westergard, T., McAvoy, K., Russell, K., Wen, X., Pang, Y., Morris, B., Pasinelli, P., Trotti, D., and Haeusler, A. (2019) Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med* **11**

33. Su, Z., Zhang, Y., Gendron, T. F., Bauer, P. O., Chew, J., Yang, W. Y., Fostvedt, E., Jansen-West, K., Belzil, V. V., Desaro, P., Johnston, A., Overstreet, K., Oh, S. Y., Todd, P. K., Berry, J. D., Cudkowicz, M. E., Boeve, B. F., Dickson, D., Floeter, M. K., Traynor, B. J., Morelli, C., Ratti, A., Silani, V., Rademakers, R., Brown, R. H., Rothstein, J. D., Boylan, K. B., Petrucelli, L., and Disney, M. D. (2014) Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**, 1043-1050

34. Simone, R., Balendra, R., Moens, T. G., Preza, E., Wilson, K. M., Heslegrave, A., Woodling, N. S., Niccoli, T., Gilbert-Jaramillo, J., Abdelkarim, S., Clayton, E. L., Clarke, M., Konrad, M. T., Nicoll, A. J., Mitchell, J. S., Calvo, A., Chio, A., Houlden, H., Polke, J. M., Ismail, M. A., Stephens, C. E., Vo, T., Farahat, A. A., Wilson, W. D., Boykin, D. W., Zetterberg, H., Partridge, L., Wray, S., Parkinson, G., Neidle, S., Patani, R., Fratta, P., and Isaacs, A. M. (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol Med* **10**, 22-31

35. Disney, M. D., Liu, B., Yang, W. Y., Sellier, C., Tran, T., Charlet-Berguerand, N., and Childs-Disney, J. L. (2012) A small molecule that targets r(CGG)(exp) and

improves defects in fragile X-associated tremor ataxia syndrome. *ACS Chem Biol* **7**, 1711-1718

36. Kearse, M. G., Goldman, D. H., Choi, J., Nwaezeapu, C., Liang, D., Green, K. M., Goldstrohm, A. C., Todd, P. K., Green, R., and Wilusz, J. E. (2019) Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev* **33**, 871-885

37. Kumari, D., Swaroop, M., Southall, N., Huang, W., Zheng, W., and Usdin, K. (2015) High-Throughput Screening to Identify Compounds That Increase Fragile X Mental Retardation Protein Expression in Neural Stem Cells Differentiated From Fragile X Syndrome Patient-Derived Induced Pluripotent Stem Cells. *Stem Cells Transl Med* **4**, 800-808

38. Weisman-Shomer, P., Cohen, E., Hershco, I., Khateb, S., Wolfovitz-Barchad, O., Hurley, L. H., and Fry, M. (2003) The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res* **31**, 3963-3970

39. Zamiri, B., Reddy, K., Macgregor, R. B., and Pearson, C. E. (2014) TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J Biol Chem* **289**, 4653-4659

40. Starck, S. R., Jiang, V., Pavon-Eternod, M., Prasad, S., McCarthy, B., Pan, T., and Shastri, N. (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719-1723

41. Ofer, N., Weisman-Shomer, P., Shklover, J., and Fry, M. (2009) The quadruplex r(CGG)n destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res* **37**, 2712-2722

42. Yang, W. Y., Wilson, H. D., Velagapudi, S. P., and Disney, M. D. (2015) Inhibition of Non-ATG Translational Events in Cells via Covalent Small Molecules Targeting RNA. *J Am Chem Soc* **137**, 5336-5345

43. Sebaugh, J. L. (2011) Guidelines for accurate EC50/IC50 estimation. *Pharm Stat* **10**, 128-134

44. Barmada, S. J., Skibinski, G., Korb, E., Rao, E. J., Wu, J. Y., and Finkbeiner, S. (2010) Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *J Neurosci* **30**, 639-649

45. Barmada, S. J., Ju, S., Arjun, A., Batarse, A., Archbold, H. C., Peisach, D., Li, X., Zhang, Y., Tank, E. M., Qiu, H., Huang, E. J., Ringe, D., Petsko, G. A., and Finkbeiner, S. (2015) Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. *Proc Natl Acad Sci U S A* **112**, 7821-7826

46. Flores, B. N., Li, X., Malik, A. M., Martinez, J., Beg, A. A., and Barmada, S. J. (2019) An Intramolecular Salt Bridge Linking TDP43 RNA Binding, Protein Stability, and TDP43-Dependent Neurodegeneration. *Cell Rep* **27**, 1133-1150.e1138

47. Arrasate, M., Mitra, S., Schweitzer, E. S., Segal, M. R., and Finkbeiner, S. (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* **431**, 805-810

Chapter 4 : The role of eIF2A, eIF2D, and DENR/MCTS1 in Repeat-Associated Non-AUG Translation Initiation

4.1 Statement of others' contribution to data presented in this chapter

Fang He and Amy Krans generated the GGGGCCx28 *Drosophila* lines. Indranil Malik assisted in tracking fly survival in a blinded manner. Amy Krans harvested the brains of all mice used in this chapter, and Udit Sheth performed sectioning and staining of the mouse brain tissue in a blinded manner. I performed all other experiments, analyzed all data, generated all figures, and wrote the chapter, except the method section on generating the repeat flies, which was written by Fang He.

4.2 Introduction

While the majority of eukaryotic translation initiation events utilize a canonical AUG start codon, advances in transcriptome-wide ribosome footprinting have revealed a greater prevalence of non-AUG translation initiation than previously appreciated (1,2). Cellular proteins synthesized from non-AUG start codons can have essential functions that are often associated with cellular stress responses (3). For instance, eukaryotic initiation factor 4G2 (eIF4G2; also known as DAP5), plays an important role in regulating cap-independent translation, and initiates at a GUG codon (4,5). Furthermore, upstream open reading frames (uORFs) are enriched for non-AUG start

codon use (1,2). In the case of the endoplasmic reticulum (ER) chaperone protein BIP, a pair of near-AUG initiated uORFs in its 5′ leader are believed to be important for promoting its continues synthesis during ER stress (6).

Non-AUG initiation can be subject to different regulatory mechanisms than canonical AUG-initiated translation. This is particularly true during the integrated stress response (ISR). The ISR reduces global cellular translation by inducing phosphorylation of the essential eukaryotic initiation factor eIF2 at serine 51 of its alpha subunit. Phosphorylation prevents eIF2 from exchanging GDP for GTP, which subsequently prevents eIF2 from binding and delivering the initiator methionine tRNA (Met-tRNA$_i^{Met}$) to the 40S ribosome. While this strategy is effective at dramatically decreasing most AUG-initiated translation, some non-AUG initiation events are spared or upregulated during the ISR.

Multiple mechanisms have been proposed for how specific non-AUG initiation events evade downregulation during ISR activation. One model depends on the ability of specific non-AUG start codons to receive initiator tRNAs independent of eIF2. Factors such as eIF2A, eIF2D, and DENR/MCTS1 are not subject to regulation by the ISR, and can deliver initiator tRNAs to the pre-initiation ribosome (7,8). While these factors can all bind and deliver the Met-tRNA$_i^{Met}$, each has also been shown to deliver non-methionyl tRNAs cognate to the near-AUG codons used for initiation (7-9).

While non-AUG initiation is important in normal cellular processes, its mis-regulation has also been implicated in human disease (3,10,11). In particular, repeat expansion mutations associated with several neurodegenerative and neuromuscular diseases are known to undergo a process called repeat-associated non-AUG (RAN)

translation (11-13). RAN translation initiates upstream of or within the expanded repeats, and results in translation through the repetitive RNA sequence. This produces RAN peptides that contain large, repetitive amino acid sequences that are often aggregation-prone and neurotoxic in model systems.

By its very nature, RAN translation exclusively utilizes non-AUG start codons for synthesis of these toxic proteins. Recently, our lab and others showed that RAN translation of expanded GGGGCC repeats associated with C9orf72 amyotrophic lateral sclerosis and frontotemporal dementia (C9ALS/FTD), as well as of expanded CGG repeats associated with fragile X-associated tremor/ataxia syndrome (FXTAS), behaves like some other non-AUG translation events and is increased following ISR activation, while global translation is simultaneously downregulated (14-17). Here, we investigate the ability of C9 and CGG RAN translation to utilize the eIF2-alternatives, eIF2A, eIF2D, and DENR/MCTS1, during translation initiation. We find evidence that eIF2A and DENR/MCTS1 can promote C9 and CGG RAN translation under specific conditions and knocking down each factor modestly improves repeat-mediated toxicity in *in vivo* disease models. However, these factors do not appear responsible for the increase in RAN translation that is observed following ISR activation.

4.3 Results

C9 and CGG RAN translation is reduced by small molecules that inhibit non-methionine and methionine-mediated translation initiation

Previous work from our group and others has established that near-AUG start codons are used for RAN translation initiation of both the GGGGCC and CGG repeats

in the reading frames that produce the most abundant RAN peptides in patient tissues (14,16,18,19). Specifically, a CUG codon is important for production of the poly-glycine-alanine (GA) peptide from the GGGGCC repeat, while ACG or GUG codons are utilized for the majority of poly-glycine (FMRpolyG) synthesis from the CGG repeat (14,16,18,19). It is possible that these near-AUG initiation events could utilize either the canonical methionine tRNA or their cognate tRNAs, i.e. leucine for the CUG codon, threonine for the ACG codon, and valine for the GUG codon.

To assess the contributions of methionine and non-methionine initiation during C9 and CGG RAN translation, we took advantage of small molecules previously shown to selectively inhibit translation depending upon the tRNA/amino acid used during initiation. Acriflavine, a nucleic acid intercalator (20), has been shown to reduce protein synthesis that initiates with leucine more significantly than protein synthesis initiating with the canonical methionine (9). To confirm this specificity, we utilized a canonical nanoluciferase (NLuc) reporter that initiates translation at an AUG codon, and a CrPV-NLuc reporter, in which the cricket paralysis virus internal ribosome entry site (IRES) is inserted upstream of NLuc control (**Fig 4-1, A)**. The CrPV IRES initiates translation at an alanyl GCU codon independent of the Met-tRNA$_i^{Met}$. Consistent with previous results, in a rabbit reticulocyte lysate (RRL) *in vitro* translation system, 50 μM acriflavine more greatly inhibited expression of the CrPV-NLuc reporter than the AUG-NLuc control (**Fig 4-1, B & C**). Like the CrPV-NLuc reporter, RAN translation from GGGGCCx70 reporter mRNAs for all three reading frames, as well as CGGx100 reporter mRNAs for the +1 (FMRpolyG), but not +2 (FMRpolyA), reading frames, was more significantly inhibited by

acriflavine treatment than AUG-NLuc (**Fig 4-1, B & C**), suggesting a potential for some RAN translation initiation events to utilize non-methionyl tRNAs.

We next expressed these reporters in the presence of NSC119893, a small molecule that inhibits formation of the eIF2·GTP·Met-tRNA$_i^{Met}$ ternary complex (21). As expected, the addition of 10 μM NSC119893 to RRL reactions greatly inhibited AUG-NLuc expression, but did not affect expression of an A-capped CrPV-NLuc reporter that initiates translation entirely through a non-methionyl IRES-mediated mechanism (**Fig 4-1, D & E**). Interestingly, expression of all C9 and CGG RAN translation reporters in RRL was inhibited to an even greater extent than AUG-NLuc by NSC119893 treatment (**Fig 4-1, D & E**). The activity of these compounds together suggests that C9 and CGG RAN translation are capable of utilizing a non-methionyl translation initiation process, but the majority of RAN translation initiation of these repeats, under basal conditions, requires the canonical eIF2·GTP·Met-tRNA$_i^{Met}$ ternary complex.

Loss of eIF2A reduces C9 and CGG RAN translation in in vitro translation assays

Starck et al. previously showed that acriflavine-sensitive translation can use eIF2A in place of the canonical eIF2 during translation initiation (9). eIF2A is a monomeric protein, that, like the trimeric eIF2, can deliver the Met-tRNA$_i^{Met}$ to the 40S ribosome to promote translation initiation (8). However, *in vitro*, eIF2A is outcompeted by eIF2 in directing protein synthesis, and is considered an alternative, non-essential initiation factor as its deletion is well-tolerated in yeast and mice (22-24). eIF2A can also bind and supply non-methionyl tRNAs for initiation, and is not subjected to inhibition during ISR activation (6,8).

To determine the role of eIF2A in C9 and CGG RAN translation, we generated *in vitro* translation lysates from wildtype (WT) and CRISPR-mediated eIF2A knockout (KO) HAP1 cells (25,26) (**Fig. 4-2, A**). Relative to WT lysates consisting of the same total protein concentration, basal translation levels in eIF2A KO lysates was significantly reduced (**Fig. 4-2, B**). However, after adjusting for this basal difference, RAN translation of the GGGGCCx70 repeat reporter mRNA in the GA reading frame, as well as of the +1 CGGx100 repeat reporter mRNA, was significantly reduced in the absence of eIF2A compared to expression of the canonical AUG-NLuc reporter mRNA (**Fig. 4-2, C**). This finding is consistent with a previous report showing that loss of eIF2A reduces polyGA production in multiple cells lines and chick embryos (16).

However, in HEK293 cells, siRNA-mediated knockdown (KD) of eIF2A had little effect on C9 and CGG RAN translation from transfected reporter plasmids (**Fig. 4-2, D to F and Fig. 4-4, A**). While it modestly decreased polyGA expression 10-20% relative to cells treated with a non-targeting siRNA (**Fig. 4-2, D and Fig. 4-4, A)**, no inhibitory effect was observed for the polyGP and polyGR frames, nor for the +1 and +2 reading frames of the CGG repeat (**Fig. 4-2, E & F**). Furthermore, no inhibition was observed when the mRNA reporters like those used in the *in vitro* lysate experiments were directly transfected into cells with eIF2A KD (**Fig. 4-2, G**).

Previous studies showing inhibition of C9RAN translation in the polyGA frame utilized bi-cistronic reporters in which the GGGGCC repeat was placed in the second cistron (16). In this sequence context, cap-independent RAN translation of the GGGGCC repeat may contribute a greater proportion of the luminescent signal than the signal generated from our functionally capped, monocistronic reporter mRNAs, and cap-

independent translation initiation may be more sensitive to eIF2A levels. To test this, we next transfected eIF2A KD HEK293 cells with A-capped reporter RNAs (14,18). Unlike the functional m^7G, the A-cap cannot interact with the cap-binding factors to support cap-dependent translation initiation. As previously reported (14,18), relative to functionally capped RAN reporters, the A-capped reporters supported significantly lower levels of translation (**Fig. 4-2, H**). However, the remaining fraction of translation observed, which likely occurs through a cap-independent mechanism, also was not inhibited be eIF2A KD (**Fig. 4-2, I**).

The inconsistency between these two systems may indicate that our siRNA is unable to reduce eIF2A levels to the extent necessary for its clear inhibition of RAN translation, and thus an effect is only seen when eIF2A expression is completely, or nearly completely, lost. Alternatively, the interplay between eIF2A's role in regulating translation during cellular stress (6,27,28), the cellular stress resulting from transient transfections, and the increase in RAN translation during cellular stress conditions (14-17), may obscure the role eIF2A plays in regulating RAN translation under the non-stressed conditions in *in vitro* translation lysates.

The effect of eIF2D and DENR KD on RAN translation

Besides eIF2 and eIF2A, additional translation factors can support translation initiation by delivering the Met-$tRNA_i^{Met}$ to the 40S ribosome, including the monomeric eIF2D and the obligate dimeric DENR/MCTS1 (7). Additionally, under specific conditions in *in vitro* re-constitution assays, eIF2D and DENR/MCTS1 also promote

translation initiation at CUG and GUG codons through delivery of the cognate leucyl or valinyl tRNAs, respectively (7).

To determine if either eIF2D and/or DENR/MCTS1 are involved in C9 and CGG RAN translation, we expressed RAN reporters in HEK293 cells following KD of either protein/complex. Reduced levels of eIF2D did not inhibit C9 or CGG RAN translation in any reading frame tested (**Fig. 4-3, A to C**). In fact, it significantly increased expression of the polyGR product from the GGGGCC repeat (**Fig. 4-3, B**).

In contrast, reduction of DENR/MCTS1 levels, through use of a DENR-specific siRNA, significantly reduced NLuc expression of C9 and CGG RAN reporters across all reading frames assayed, without significantly altering AUG-NLuc expression (**Fig. 4-3, D to F**). Reduction of polyGA synthesis, along with unaltered AUG-NLuc expression, following DENR KD was confirmed by Western blot using an antibody against the c-terminal FLAG tag present on each reporter (**Fig. 4-3, G**).

Interestingly, the effect of DENR KD on the GGGGCCx70 and CGGx100 reporter expression was dependent on their non-AUG initiation, as reporters with an AUG start codon inserted upstream of either repeat, to drive canonical translation in the GA or +1 FMRpolyG reading frames, were not significantly inhibited (**Fig. 4-3, E & F**). Furthermore, this effect was specific to repeat-associated non-AUG translation, as opposed to no-repeat non-AUG translation, as DENR KD, like eIF2A and eIF2D KD, did not reduce expression of CUG-NLuc or ACG-NLuc reporters (**Fig. 4-4, A to D**). However, it is important to note that, although DENR KD had no effect on expression of the highly stable AUG-NLuc protein, it did significantly reduce expression of the less stable, co-transfected AUG-firefly luciferase (FFLuc) reporter (**Fig. 4-4, G**). This

suggests that loss of DENR may reduce global translation levels. No reduction of FFLuc was observed with eIF2A or eIF2D KD (**Fig. 4-4, E & F**).

eIF2A, eIF2D, and DENR knockdown do not prevent increased C9 and CGG RAN translation upon ISR induction

In previous studies, eIF2A was most critical for translation initiation under conditions of cellular stress, when functional eIF2 levels were limited by ISR-mediated phosphorylation of eIF2α at serine 51 (6,27,28). We thus specifically assessed the ability of eIF2A, eIF2D, and DENR/MCTS1 KD to regulate C9 and CGG RAN translation following induction of ER stress. Consistent with previous findings (14-17), 2 μM thapsigargin (TG) treatment for five hours selectively increased C9 and CGG RAN translation expression in HEK293 cells, while reducing expression of AUG-initiated control reporters (**Fig. 4-5, A to C**). However, KD of eIF2A, eIF2D, and DENR/MCTS1 did not consistently prevent the increase in RAN levels (**Fig. 4-5, A to C**).

These results suggest that despite limited levels of functional eIF2·GTP·Met-$tRNA_i^{Met}$ TC, RAN translation may still utilize this factor for initiation during the ISR. Alternatively, stress-resistant RAN translation may receive an initiator tRNA through a mechanism independent of eIF2 and the additional factors assessed here. Future experiments monitoring the behavior of RAN reporters in HEK293 cells co-treated with TG and NSC119893, to inhibit TC formation, could distinguish between these possibilities.

GGGGCC repeat-mediated toxicity in Drosophila is modulated by knockdown of eIF2 alternatives

As an *in vivo* model of repeat toxicity, we developed two transgenic *Drosophila* lines conditionally expressing 28 GGGGCC repeats under the UAS promoter, one with the repeat inserted into the second chromosome and the other with the repeat inserted into the third chromosome. When expressed with a non-targeting control shRNA in the fly eye, using the eye-specific GMR-Gal4 driver, the repeat causes a severe rough eye phenotype with noticeable eye shrinkage (**Fig 4-6, A & B**). We next co-expressed the GGGGCC repeat with an shRNA targeting the predicted *Drosophila* homolog of eIF2A (CG7414), using the GMR-Gal4 driver, and imaged the fly eyes 1-3 days post eclosion (**Fig 4-6, B**). Using ImageJ, we measured the width of repeat-expressing fly eyes in the presence of the control versus CG7414 ("eIF2A") shRNAs, and found that eIF2A KD modestly, but significantly, increased the width of the GGGGCCx28 expressing fly eye, relative to a control shRNA (**Fig 4-6, B & C**).

We next expressed the GGGGCCx28 repeat, along with a non-targeting control shRNA, ubiquitously throughout the fly using the RU-486 inducible tubulin driver,Tub5-GS Gal4. This results in a substantial decrease in the lifespan of both male and female flies (**Fig 4-6, D & E**). However, in this assay, co-expression of the repeat with an shRNA targeting CG7414 ("eIF2A") did not significantly improve survival of male or female flies (**Fig 4-6, F & G**).

We then assessed the ability of shRNAs targeting the fly homologs of eIF2D and DENR to modulate the repeat-associated toxicity observed through the rough eye phenotype and reduced survival. Two of three shRNAs against eIF2D modestly, but

significantly increased the fly eye width (**Fig 4-7, A to C**). However, none improved fly survival (**Fig 4-7, D & E**).

Both shRNAs targeting the fly homology of DENR again modesty, yet significantly, increased the width of GGGGCCx28 expressing eyes (**Fig 4-8, A & B**). Additionally, “DENR shRNA #2” significantly enhanced male fly survival in the presence of the repeat (**Fig 4-8, C**). However, in female flies, neither shRNA improved survival (**Fig 4-8, D**).

It is important to note that these GGGGCCx28 flies contain significant non-native C9orf72 sequence upstream of the repeat. Therefore, any non-repeat sequence-specific regulation these factors have on C9RAN translation may not be captured in this model.

GGGGCC repeat expression and toxicity in eIF2A KO mice

Despite inconsistency in the eIF2A’s effect on RAN translation across systems, the lysate results encouraged us to look at C9RAN translation in a different eIF2A KO model. Mice lacking eIF2A are born at a normal frequency and survive to adulthood (24). We are thus performing intracerebroventricular (ICV) injections on P0 eIF2A KO and WT mice with AAV9 vectors expressing 2 or 66 GGGGCC repeats (29). This C9ALS/FTD mouse model has been reported to support robust levels of RAN translation and develop motor deficits in a repeat-dependent manner (29). In preliminary studies, we have confirmed that expression of 66, but not 2 GGGGCC repeats, leads to accumulation of polyGA in mice brains by three months of age (**Fig. 4-9, B**). We have also found that WT mice injected with the 66 repeat AAV take longer to cross a square 5 mm balance beam than WT mice injected with 2 repeats (**Fig. 4-9, C**). Preliminary

studies suggest that this motor deficit may be reduced in eIF2A KO mice injected with the 66 repeat AAV, however additional eIF2A KO animals are needed before any conclusions can be drawn (**Fig. 4-9, D**).

In our hands, the eIF2A KO mice are more prone than WT mice to developing hydrocephalus between 4 and 8 weeks of age, and this condition appears to be exacerbated by ICV injection, leading to greater than anticipated dropout of eIF2A animals in our studies. Additionally, eIF2A mice perform better in multiple motor tests than WT mice at 6 months of age (**Fig. 4-9, E & F**), though not on the balance beam (**Fig. 4-9, G**). These phenotypes may confound interpretation of behavioral test results on WT and eIF2A KO mice injected with the GGGGCC repeats.

4.4 Discussion

eIF2A and RAN translation

When first discovered, the ability of eIF2A to deliver initiator tRNAs to the 40S ribosome led many to suspect that it was the main factor involved in this essential translation initiation step (8), but subsequent experiments established the ability of eIF2 to out-perform eIF2A in stimulating eukaryotic protein synthesis (23). However, in these initial studies, eIF2A was able to function in ways eIF2 could not. First, it was able to initiate translation of a poly(U) transcript by promoting phenylalanine tRNA binding to the ribosome (8). Second, it was able to bind the ribosome independent of GTP, but it required the presence of RNA to do so (8). In current models, these characteristics allow eIF2A to function in place of eIF2 during the ISR, when functional eIF2 levels are limiting due to eIF2α phosphorylation, and to promote stress resistant non-AUG

translation initiation with non-methional tRNAs (6,9,27,28). These characteristics, along with the fact that it is non-essential in yeast and mice (22,24), also make eIF2A an appealing factor in regulating RAN translation.

To date, eIF2A has been implicated in four cellular translation events; translation of the 26S alphavirus RNA (28), the hepatitis C virus (HCV) RNA (27), two BIP upstream reading frames (6), and C9RAN translation in the polyGA reading frame (16). However, subsequent experiments have failed to replicate a role for eIF2A in alphavirus or HCV translation (25,26,30). This may suggest unknown nuances in eIF2A's function in translation initiation that differ between cell types and experimental conditions, and is consistent with data reported here. While eIF2A KO has a strong inhibitory effect on RAN translation *in vitro*, eIF2A KD had little to no effect in HEK293 cells. This also highlights the importance of assessing the role of RAN translation targets in disease-relevant contexts, such as patient cells. Through ongoing experiments, we are knocking down eIF2A in patient-derived iPSCs, and using a previously published ELISA to determine if this has an effect on C9RAN translation in the polyGP reading frame (31).

Awaiting results from these ongoing studies, including mouse studies, a role for eIF2A in C9 and CGG RAN translation is inconclusive, based on inconsistent effects of its KD or deletion, in reporter-based and fly experiments.

DENR/MCTS1, eIF2D and RAN translation

MCTS1 and DENR are homologous to eIF2D's N and C-terminal domains, respectively, and both factors participate in the same translation processes, including initiator tRNA delivery, ribosome reinitiation, and ribosome recycling (7). Despite these

similarities, in our hands in HEK293 cells, eIF2D KD had no inhibitory effect on RAN translation, while DENR KD caused a robust and specific reduction in C9 and CGG RAN translation. Additionally, DENR was the only factor in these studies that both improved the rough eye and survival phenotypes of the GGGGCCx28 fly. This suggests that in these models, DENR/MCTS1 and eIF2D do not function in an entirely redundant manner.

Because DENR is involved in multiple translation initiation processes, it is unclear in what way it supports RAN translation. That DENR KD reduces basal RAN translation but does not prevent ISR-induced increases in C9RAN and CGG RAN translation, suggests that in this context, it is not functioning in place of eIF2 to deliver initiator tRNAs. Further, DENR's role in reinitiation is most critical for sustained translation of main coding ORFs, following translation termination of uORFs. Accordingly, knockdown of DENR shifts ribosome occupancy from coding ORFs to uORFs (32). As RAN translation embeds CGG repeats in *FMR1* (33), and GGGGCC repeats in the reporters used in these studies, within uORFs, ribosome reinitiation, if it occurred at all, would occur *after* the repeats are translated, and thus not be a requirement for their translation. Lastly, DENR KD did not affect translation of the expanded repeats when initiation was driven with an AUG start codon. Therefore, it is likely not involved in helping resolve ribosomes that may stall during elongation through the repetitive elements, but specifically plays a role in the non-AUG translation initiation event.

DENR and the C-terminus of eIF2D have homology to eIF1, and all three factors adopt a similar SUI1-like fold that interact with the 40S ribosome at the same location

(34). eIF1 is an essential initiation factor that is involved in regulating start codon fidelity; reduced levels of eIF1 leads to increased non-AUG initiation (35). Consistent with RAN translation exclusively occurring at non-AUG codons, previous work from our group found that overexpression of eIF1, to enhance start codon fidelity, decreased CGG RAN translation in HEK293 cells (36). Therefore, one model for how DENR KD reduces RAN translation is that it may increase the occupancy of repeat-associated ribosomes with eIF1, resulting in increased start codon stringency. How this could specifically target repeat-containing reporters, and not the no repeat near-AUG reporters used in this study, will require further investigation.

RAN translation during the ISR

Our interest in each of these factors was sparked by their ability to act in place of eIF2. Therefore, it was unexpected that none of their KDs prevented increased RAN translation during cellular stress conditions, and thus leaves the important question of how RAN translation evades ISR inhibition unanswered. Several alternative explanations exist. First, eIF2A has been shown to associate and work synergistically with eIF5B, as eIF5B increases eIF2A tRNA binding and both factors have increased ribosome association during cellular stress (37). Additionally, while loss of either factor is well-tolerated in yeast, loss of both is lethal (22). Consequently, if eIF2A is indeed involved in stress-resistant translation, knocking down eIF5B in concert may be necessary in some systems to observe effects.

Alternatively, it is possible the RAN translation initiation still uses eIF2 during the ISR, despite its low functional levels. This could be achieved through preferential

loading of 43S ribosomes with eIF2 TCs onto repeat-containing RNAs, or through pre-loading of translation competent ribosomes prior to stress induction. Consistent with the second model is evidence that inefficient translation initiation events at near-AUG start codons, including events in CGG RAN translation, cause ribosome queuing that allows for continued translation following treatment with translation elongation inhibitors like cycloheximide (5). Lastly, ISR-resistant RAN translation could proceed through a previously uncharacterized mechanism. As inhibiting this event is one potential avenue to allow for selective targeting of RAN translation in patients, unraveling its mechanism is of high importance.

In conclusion, the data in this chapter implicate DENR in regulating C9 and CGG RAN translation and suggest a potential role for eIF2A in specific conditions. Further work is required to determine if the effects observed here with reporter systems and model organisms translate to regulating RAN translation in patient cells, and if these factors could serve as novel targets in disease treatments.

4.5 Experimental methods

RNA synthesis

RNAs were *in vitro* transcribed from PspOMI-linearized pcDNA3.1(+) reporter plasmids, using HiScribe T7 High Yield RNA Synthesis Kit (NEB) with 3´-O-Me-m^7GpppG anti-reverse cap analog (ARCA) or ApppG cap (NEB) added at 8:1 to GTP, for a capping efficiency of ~90%, as previously described (14,18). After RNA synthesis, DNA templates were removed with RNase-free DNaseI (NEB), and RNAs were poly-adenylated with *E. coli* Poly-A Polymerase (NEB), as previously described (14,18).

mRNAs were then clean and concentrated with RNA Clean and Concentrator-25 Kit from Zymo Research and run on a denaturing formaldehyde RNA gel to verify mRNA size and integrity. Reporter sequence have been previously published (14,18).

Rabbit reticulocyte lysate in vitro translation

mRNAs were *in vitro* translated with Flexi Rabbit Reticulocyte Lysate System (Promega), as previously described (14,18). 10 µL reactions for luminescence assays were programmed with 3 nM mRNA and contained 30% RRL, 10 µM amino-acid mix minus methionine, 10 µM amino acid mix minus leucine, 0.5 mM MgOAc, 100 mM KCl, and 0.8 U/uL Murine RNAse Inhibitor (NEB), and incubated at 30^{o}C for 30 minutes before termination by incubation at 4^{o}C. Reactions were diluted 1:7 in Glo Lysis Buffer (Promega), and incubated 1:1 for 5 minutes in the dark in opaque 96 well plates with NanoGlo Substrate freshly diluted 1:50 in NanoGlo Buffer (Promega). Luminescence was measured on a GloMax 96 Microplate Luminometer.

Cell line information and maintenance

HEK293 cells were purchased from American Type Culture Collection (ATCC CRL-1573). They were maintained within 37^{o}C incubators at 5% CO_2 in DMEM + high glucose (GE Healthcare Bio-Science, SH30022FS) supplemented with 9.09% fetal bovine serum (50 mL added to 500 mL DMEM; Bio-Techne, S11150).

eIF2A KO (Cat # HZGHC002650c001) and isogenic control HAP1 cells were purchased from Horizon Discovery (Cambridge, UK) (25,26). They were maintained

within 37ºC incubators at 5% CO_2 in IMDM (Invitrogen, 124400-61) supplemented with 9.09% fetal bovine serum (50 mL added to 500 mL IMDM; Bio-Techne, S11150).

HAP1 cell lysate in vitro translation

Lysates were prepared using a protocol previously developed within our lab (14). HAP1 cells were trypsinized, centrifuged at 200xg for 5 minutes, with cell pellet washed once with 1X PBS. Cell pellets were then weighed and resuspended in RNAse-free hypotonic buffer containing 10 mM HEPES-KOH (pH 7.6), 10 mM potassium acetate, 0.5 mM magnesium acetate, 5 mM DTT, and EDTA-free protease inhibitor cocktail (Roche), 250 µL buffer added per 200 mg cells (14). Resuspended cells were incubated on ice for 20 minutes, passed 10X through a 27G syringe, incubated for another 20 minutes on ice, and centrifuged at 10,000xg for 10 minutes at 4°C to pellet cell debris. The supernatant was recovered, and total protein quantified with a BCA assay. Lysates were diluted to 8 µg/µl protein in the hypotonic lysis buffer, and stored in single use aliquots at -80ºC.

For *in vitro* translation reactions, 8 µg lysate was supplemented to final concentrations of 20 mM HEPES-KOH (pH 7.6), 44 mM potassium acetate, 2.2 mM magnesium acetate, 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 (14). *In vitro* transcribed reporter mRNAs were added to 4 nM. Translation assays and luminescence measurements were then performed as with RRL reactions.

Cell transfection, drug treatments, and analysis

For luminescence assays, HEK293 cells were seeded in 96-well plates and transfected 24 hours later at 40-50% confluency with siRNAs using Lipofectamine RNAiMax. The following amount of stealth siRNA (Thermo) was added per well, to achieve maximum knockdown efficiency: 0.1 pmol eIF2A siRNA, 1 pmol DENR siRNA, 4 pmol eIF2D siRNA. The control non-targeting siRNA was added at same concentration as targeting siRNA it was compared to. siRNA sequences are listed in **Table 4-1**.

Reporters were transfected into cells ~24 hours post siRNA transfections, when cells were 80-90% confluent. DNA transfections were performed in triplicate with FuGene HD at a 3:1 ratio to DNA, with 50 ng NLuc reporter DNA and 50 ng pGL4.13 FLuc reporter added per well. RNA transfections were performed with *Trans*IT-mRNA Transfection Kit from Mirus Bio, per manufacturer's recommended protocol, with 90 ng reporter mRNA and 200 ng pGL4.13 FLuc control DNA added to each well in triplicate.

For TG experiments, 2 µM TG or equal volume DMSO was added to HEK293 cells 19 hours post transfection, for 5 hours.

Cells were lysed 24 hours post transfection with 60 µL Glo Lysis Buffer for 5 minutes at room temperature. 25 µL of lysate was mixed with NanoGlo Substrate prepared as for RRL reactions, and 25 µL of ONE-Glo Luciferase Assay System (Promega), for 5 minutes in the dark, in opaque 96-well plates. Luminescence measurements were obtained as with RRL reactions.

For western blots, HEK293 cells were seeded in 12-well plates and transfected 24 hours later, at 40-50% confluency, with siRNAs using RNAiMax. Where indicated, 24 hours post transfections, at 80-90% confluency, cells were transfected with 500 ng

NLuc reporter DNAs and 4:1 FuGene HD. 48 hours post siRNA transfection, cells were lysed in 300 µL RIPA buffer with protease inhibitor for 30 minutes at 4^{o}C. Lysates were homogenized by passing through a 28G syringe, mixed with 6X sample buffer, and stored at -20^{o}C.

Western blots

All samples for western blot were run on 12% SDS-polyacrylamide gels at 150V for ~90 minutes. For western blot analysis of HAP1 lysates, 60 µg protein was loaded per well. For HEK293 western blots, 30 µL lysate for each sample was loaded. Gels were transferred to PVDF membranes either overnight at 30V and 4^{o}C, or for 2.5 hours at 320 mAmps and 4^{o}C. Membranes were blocked with 5% non-fat dry milk, and all antibodies were diluted in 5% non-fat dry milk. Primary antibodies information and probing conditions are listed in **Table 4-2**. For eIF2A, eIF2D, DENR, MCTS1, and FLAG western blots, HRP secondary antibodies were applied at 1:10000, with 1-hour incubations at room temperature. Bands were then visualized on film. For GAPDH loading controls, LiCor IRDye secondary antibodies were applied at 1:10000, with 1-hour incubations at room temperature, and bands visualized with LiCor Odyssey CLx Imaging Systems.

Drosophila work

The GGGGCCx28 repeat, along with 30 nt on both ends in the first intron of gene C9ORF72 were PCR cloned from the genomic DNA from fibroblast of an ALS patient of Central Biorepository of University of Michigan, and placed to the 5′ upstream in the +1 reading frame (GP) relative to the GFP gene in the vector

PGFPN1 (Clonetech). The repeat and GFP were then subcloned into the NotI site of vector pUAST and sequence verified for the repeat length and reading frame. This vector was used to generate transgenic flies by standard p-element insertion (Best Gene, CA).

Information the GAL4 and shRNA lines interrogated in this chapter is included in Table **4-3**. For eye shrinkage experiments, male flies containing the UAS-GGGGCCx28 and a UAS-shRNA transgene were crossed to GMR-GAL4 virgin females at 29^{o}C. One eye of each resulting progeny was imaged 0-2 days post eclosion using Leica M125 stereomicroscope and a Leice DFC425 digital camera. Eye widths were then measured with ImageJ, and normalized to values obtained from flies expressing control shRNAs.

For survival experiments, male flies containing the UAS-GGGGCCx28 and a UAS-shRNA transgene were crossed to Tub5-GAL4 GeneSwitch (GS) virgin females at 25^{o}C. 0-2 days post eclosion, resulting progeny were placed on SY10 food containing 200 µM RU486 at 29^{o}C. Male and female progeny were housed separately, with no more than 28 flies per tube. RU486 food was changed every 2 days, with number of dead flies counted during each flip.

Mouse work

Mouse work was performed in accordance with ethical standards and regulations from the National Institute of Health and University of Michigan's Institutional Animal Care and Use Committee. eIF2A KO mice, on a C57-B6N background, were received as a kind gift from the Komar Lab at Cleveland State University, and housed in a specific pathogen free, environmentally controlled vivarium with 12-hour light/dark

cycles. Mouse cages contained corncob bedding and mice were fed standard 5LOD chow, except for breeders, whom received the high fat 5008 chow. eIF2A genotyping was performed on tail biopsy DNA, as previously described (24).

ICV injections were performed as previously described, using previously developed AAVs (29). P0 mice were cold anesthetized on an aluminum plate placed on ice. They were bilaterally injected with 2 µL 3:1 AAV:trypan blue for a virus titer of $1.1x10^5$ genomes/µl.

All behavioral assays were performed blinded to mouse genotype and injection type. Balance beam experiments were performed as previously described (38). Mice were placed on a clear 20cm x 20cm platform raised 53cm above the bench top, with a 44cm long, 5mm wide clear square rod spanning the distance to a 20cm x 20cm x 20cm opaque black box. The amount of time it took the mice to cross from the clear platform to the block box was measured with a stopwatch. The maximum time allowed to cross the balance beam was 20 seconds. If a mouse fell off the balance beam, a maximum time of 20 seconds was assigned. Mice were trained for three consecutive days before experimental trial was performed on the fourth day. Each mouse crossed the balance beam two times each day, with the average time for both crosses presented.

For rotarod experiments, mice were placed on spinning bar with constant acceleration from 4-40 rpm over a course of 300 seconds. The time at which each mouse fell was recorded. Mice were trained for three consecutive days before the experimental trial was performed on the fourth day. Each mouse performed the rotarod test two times each day, with a 30 minute break between trials, and the average time presented.

Open field experiments were performed as previously described (38). Mice were placed in the photobeam activity system open-field apparatus (San Diego Instruments) for 30 minutes, on a day with no rotarod or balance beam testing. Total distance traveled during the 30 minutes was measured by the total number of beam breaks during that time.

Mouse brain immunohistochemistry

AAV-injected eIF2A+/+ mice brains were harvested at 24 weeks of age, following PBS perfusion. For immunohistochemistry (IHC), one hemisphere of each brain was post-fixed in 4% paraformaldehyde at 4^{o}C for 3 days. It was then stored in 30% sucrose in PBS at 4^{o}C until sectioning.

20-25 µM sagittal brain sections were generated with a microtome (SM200R; Leica Biosystems), and stored in cryopreserve media (in 1L: 300 g sucrose, 300 mL ethylene glycol, 500 mL 0.1M PB pH7.4) at -20^{o}C until staining. Staining and imaging was performed blinded to injection type. Sections from the central region of each hemisphere were selected. Acidic antigen retrieval was performed with incubation in 10 mM sodium citrate, pH 6.0 at 80^{o}C for 30 minutes. Tissue was blocked in 5% normal goat serum (NGS), then incubated in anti-HA antibody in 5% NGS overnight at 4^{o}C. The antibody was detected using the Vector Lab's ABC Kit; tissue was incubated with biotinylated anti-mouse antibody followed by avidin-biotin coupled HRP, briefly incubated with DAB (ImmPACT DAB, Vector Labs), and stained with hematoxylin. Sections were mounted on slides and dehydrated (70% ethanol - 1 min, 95% ethanol - 2

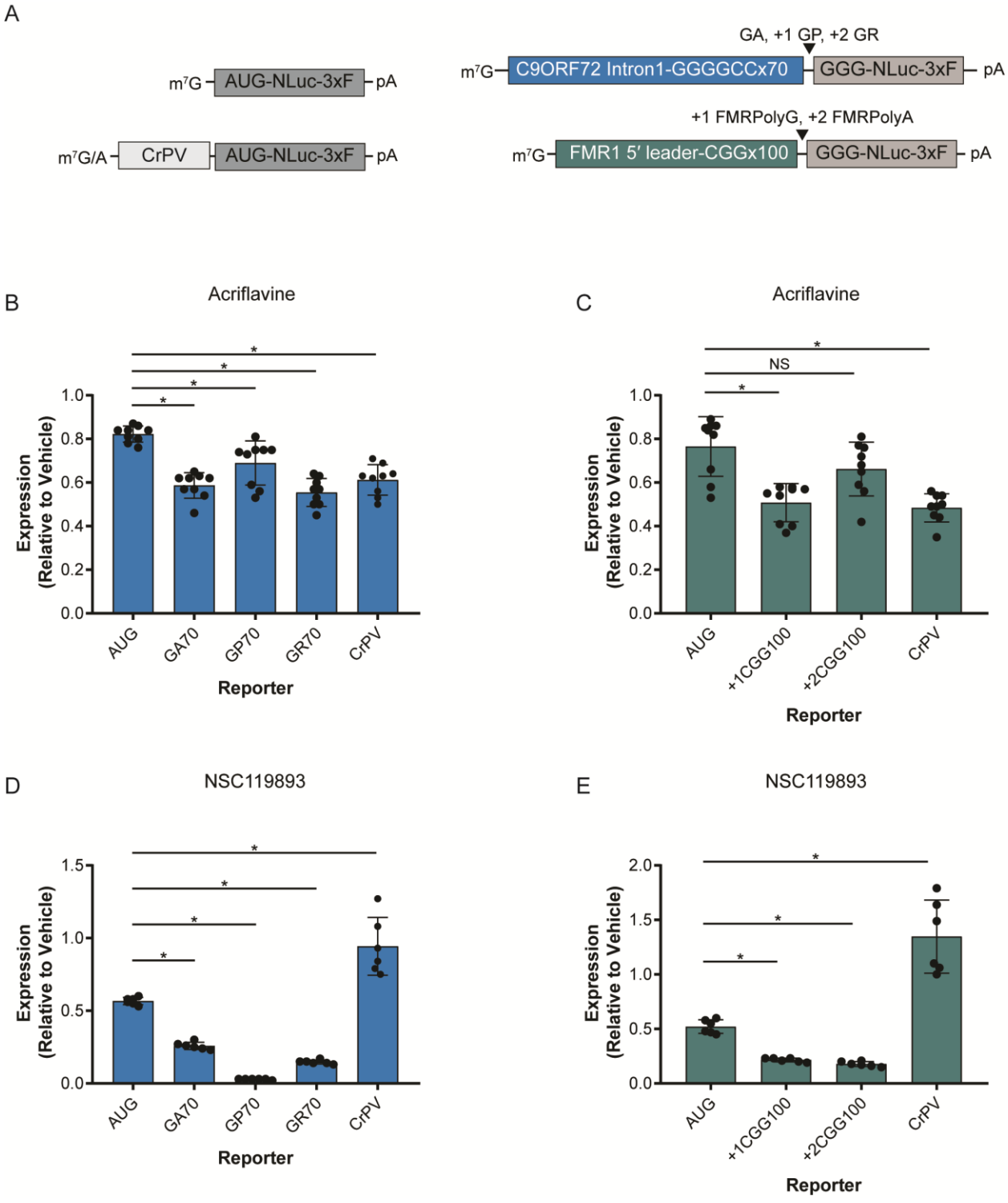
min, 100% ethanol – 2 x 2 min, xylenes - 2 x 3 min). Coverslips were applied with DPX mountant. Brightfield images were acquired with an Olympus BX51 microscope.

4.6 Chapter-specific acknowledgement

We thank the Petrucelli Lab at the Mayo Clinic in Jacksonville, FL for sharing their AAV vectors. We also thank the Komar Lab at Cleveland State University in Cleveland, OH for sharing their eIF2A$^{-/-}$ mice.

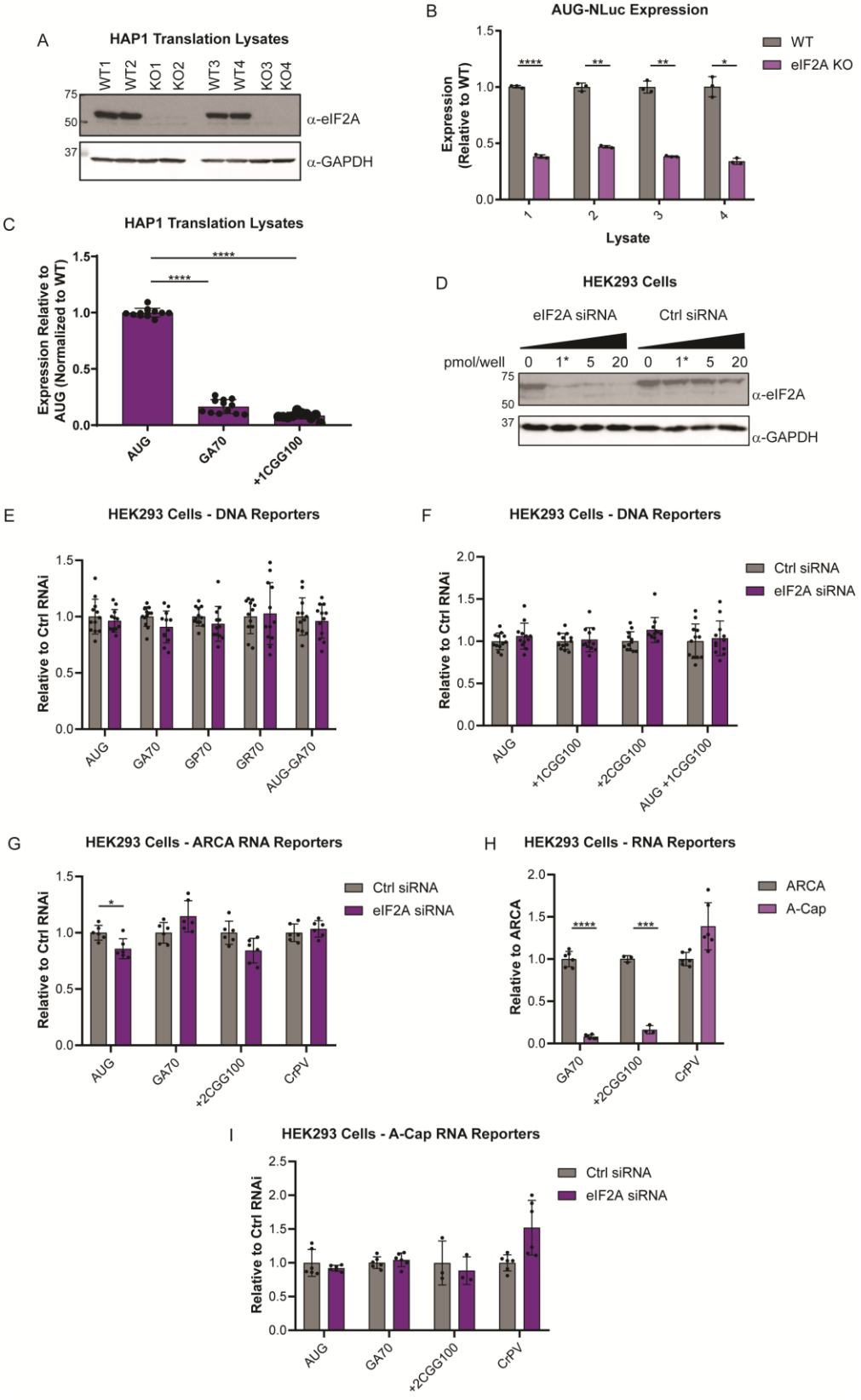
4.7 Figures

Figure 4-1. Effect on non-AUG or AUG translation initiation inhibiting small molecules on RAN translation



(A) Schematic of Nanoluciferase (NLuc) reporter mRNAs used in rabbit reticulocyte lysate (RRL) *in vitro* translation assays. 3xF = 3x Flag tag, CrPV = Cricket Paralysis Virus IRES, GA = glycine-alanine, GP = glycine-proline, GR = glycine-arginine. (B-C) NLuc expression for indicated reporter mRNAs expressed in RRL treated with 50 µM acriflavine, relative to vehicle treated controls, n=9. (D-E) NLuc expression for indicated reporter mRNAs expressed in RRL treated with 10 µM NSC119893, relative to vehicle treated controls, n=6. All graphs represent mean with errors +/- standard deviation. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$, **** $p< 0.0001$, One-way ANOVA with Dunnett's multiple comparison test.

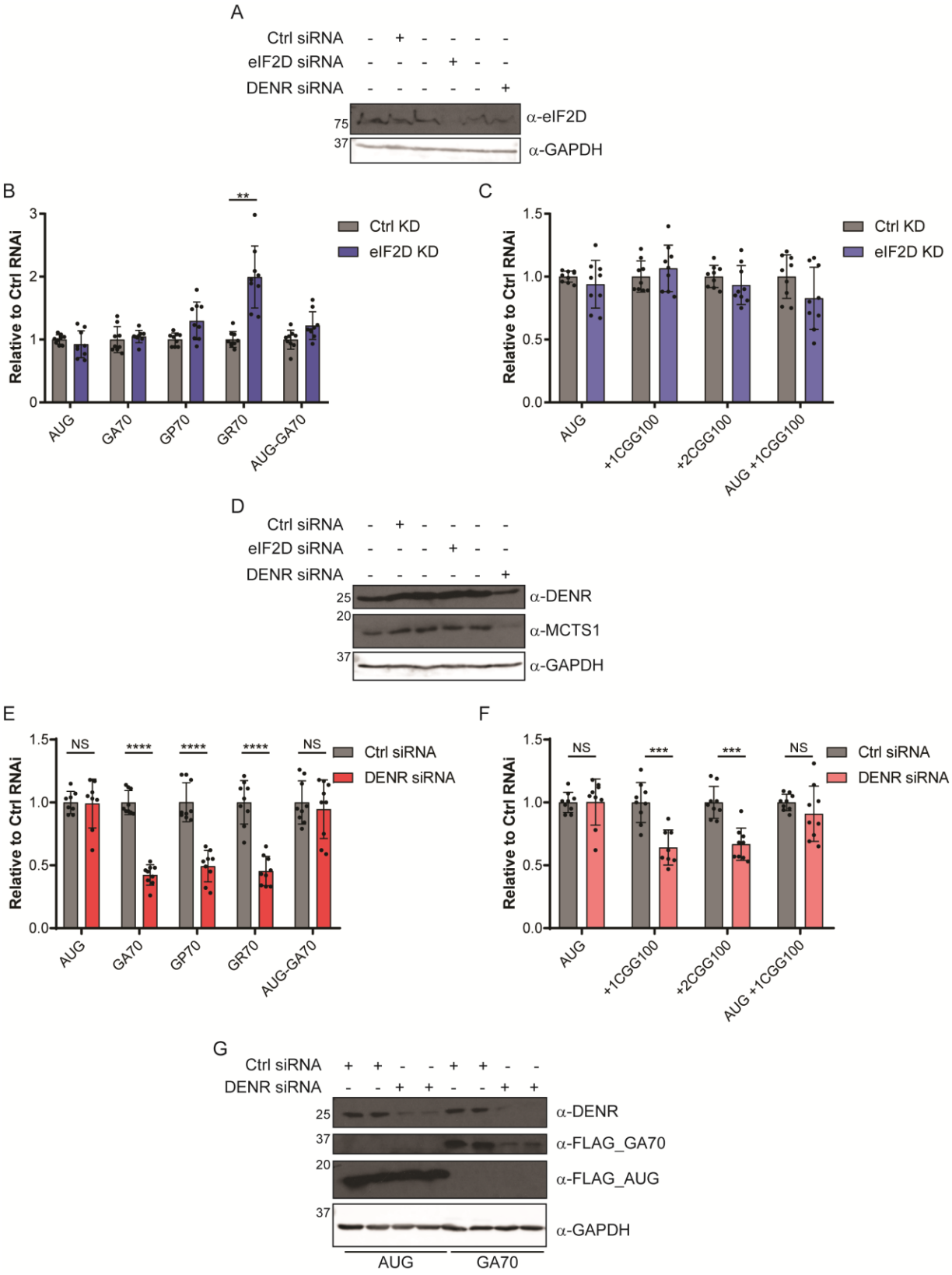
Figure 4-2. eIF2A deletion reduces RAN translation *in vitro* in translation lysates



(A) Western blot of wildtype (WT) and eIF2A knockout (KO) lysates, probed with an anti-eIF2A antibody, to confirm loss of eIF2A expression. GAPDH serves as a loading control. (B) NLuc expression of AUG-NLuc reporter mRNA in four independently generated sets of WT and KO translation lysates, with expression normalized to the WT lysate in each set, n=3. (C) Expression of AUG-NLuc, GA70-NLuc, and +1CGG100-NLuc reporter mRNAs in eIF2A KO lysates, relative to WT lysates, after controlling for the difference in AUG-NLuc expression between paired lysates, n=12. (D) Western blot showing efficiency of eIF2A KD in HEK293 cells following increasing concentrations of eIF2A siRNA. Starred lanes indicate siRNA concentration used in subsequent experiments. GAPDH was used as a loading control. (E-F) NLuc expression of indicated reporters expressed from DNA plasmids in HEK293 cells 24 hours post transfection with non-targeting or eIF2A siRNAs. NLuc levels are expressed relative to levels in cells transfected with the non-targeting siRNA, n=12. (G) NLuc expression of indicated reporters expressed from ARCA-capped mRNAs transfected into HEK293 cells 24 hours post transfection with non-targeting or eIF2A siRNAs. NLuc levels are expressed relative to levels in cells transfected with the non-targeting siRNA, n=6. (H) Comparison of NLuc levels from ARCA vs. A-capped RNAs transfected into HEK293 cells previously transfected with a non-targeting siRNA. CrPV were used a cap-independent control. NLuc levels are expressed relative to ARCA-capped mRNAs, n=3-6. (I) NLuc expression of indicated reporters expressed from A-capped RNAs transfected into HEK293 cells 24 hours post transfection with non-targeting or eIF2A siRNAs. NLuc levels are expressed relative to levels in cells transfected with the non-targeting siRNA, n=3-6. All graphs represent mean with errors +/- standard deviation. * $p < 0.05$, ** $p <$

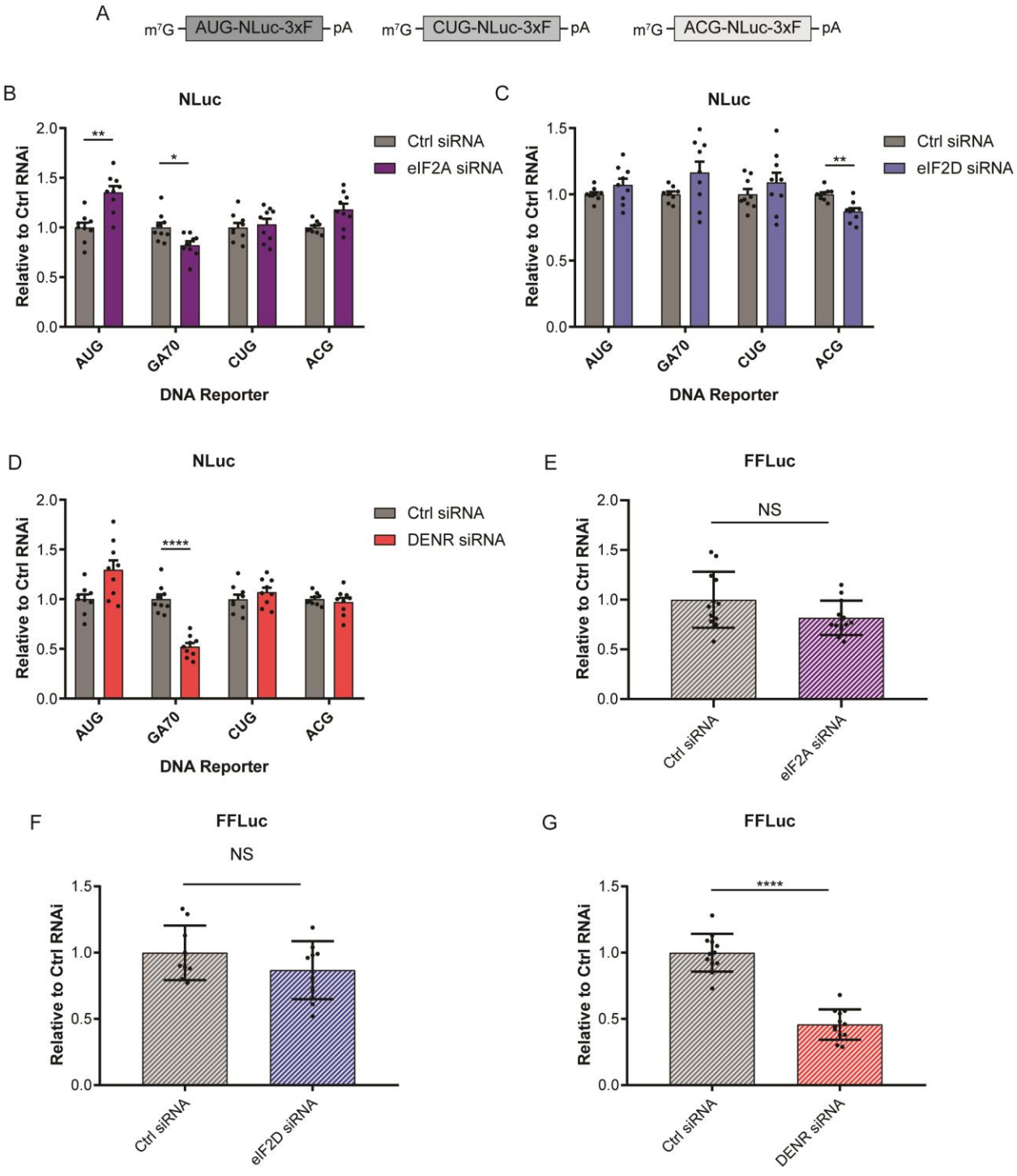
0.01, *** p< 0.001, **** p< 0.0001, (B, E-I) Two-way ANOVA with Sidak's multiple comparison test. (C) One-way ANOVA with Dunnett's multiple comparison test.

Figure 4-3. DENR, but not eIF2D KD, reduces RAN translation in HEK293 cells



(A) Western blot showing efficiency of eIF2D KD in HEK293 cells 48 hours post transfection with an eIF2D targeting siRNA. GAPDH was used as a loading control. (B-C) NLuc expression of indicated reporters expressed from DNA plasmids transfected into HEK293 cells 24 hours post transfection with non-targeting or eIF2D siRNAs. NLuc levels are expressed relative to levels in cells transfected with the non-targeting siRNA, n=9. (D) Western blot showing efficiency of DENR and MCTS1 KD in HEK293 cells 48 hours post transfection with an DENR targeting siRNA. GAPDH was used as a loading control. (E-F) NLuc expression of indicated reporters expressed from DNA plasmids transfected into HEK293 cells 24 hours post transfection with non-targeting or DENR siRNAs. NLuc levels are expressed relative to levels in cells transfected with the non-targeting siRNA, n=9. (G) Western blots showing effects of DENR KD on expression of AUG-NLuc and GA70-NLuc reporters expressed in HEK293 cells, 48 hours post KD and 24 hours post reporter transfection. NLuc reporters were probed with an antibody targeting their c-terminal FLAG tag. GAPDH was used as a loading control. All graphs represent mean with errors +/- standard deviation. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$, **** $p< 0.0001$, Two-way ANOVA with Sidak's multiple comparison test.

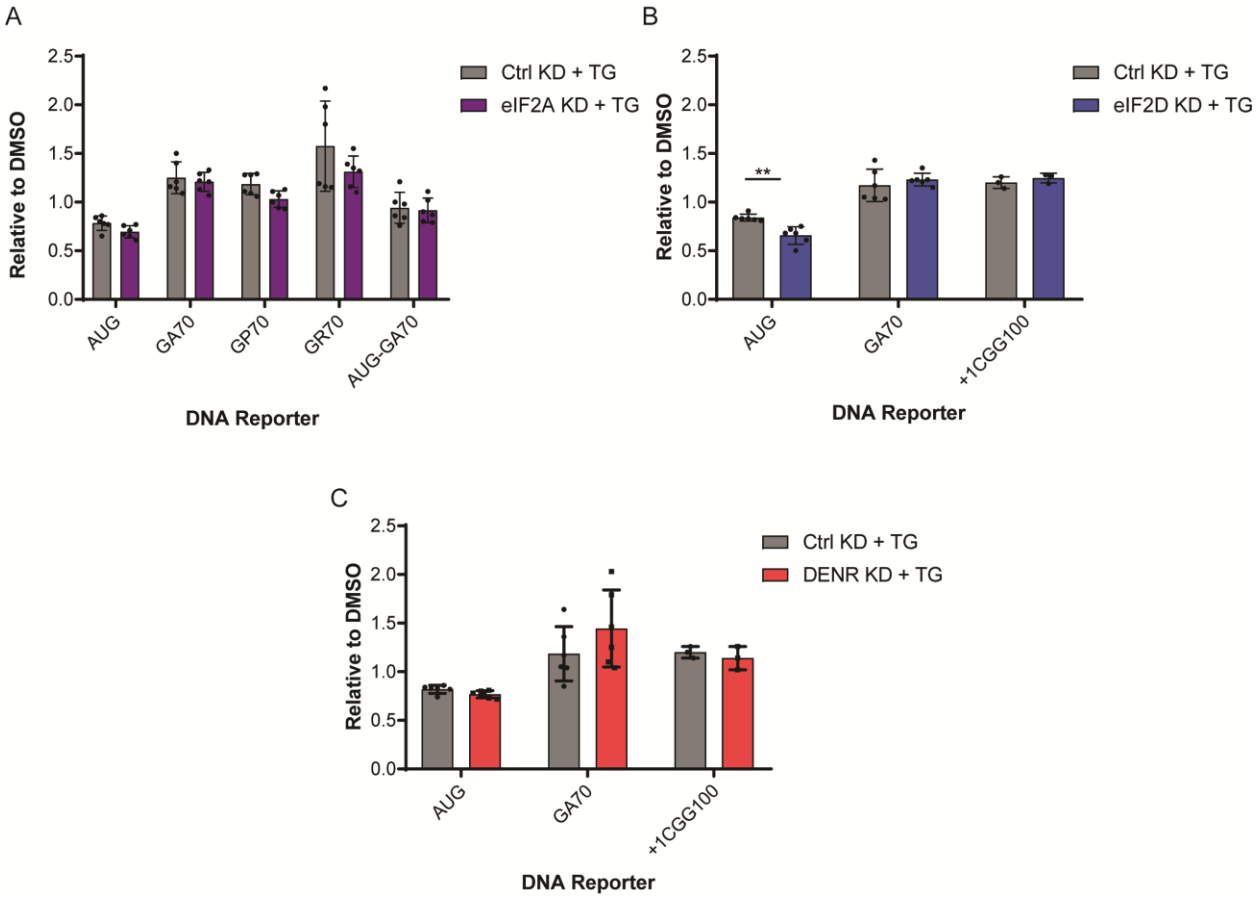
Figure 4-4. KD of eIF2 alternatives does not reduces near-AUG translation from reporters lacking expanded repeats



(A) Schematic of CUG and ACG-NLuc reporter mRNAs. (B-D) NLuc expression of indicated reporters expressed from DNA plasmids transfected into HEK293 cells 24

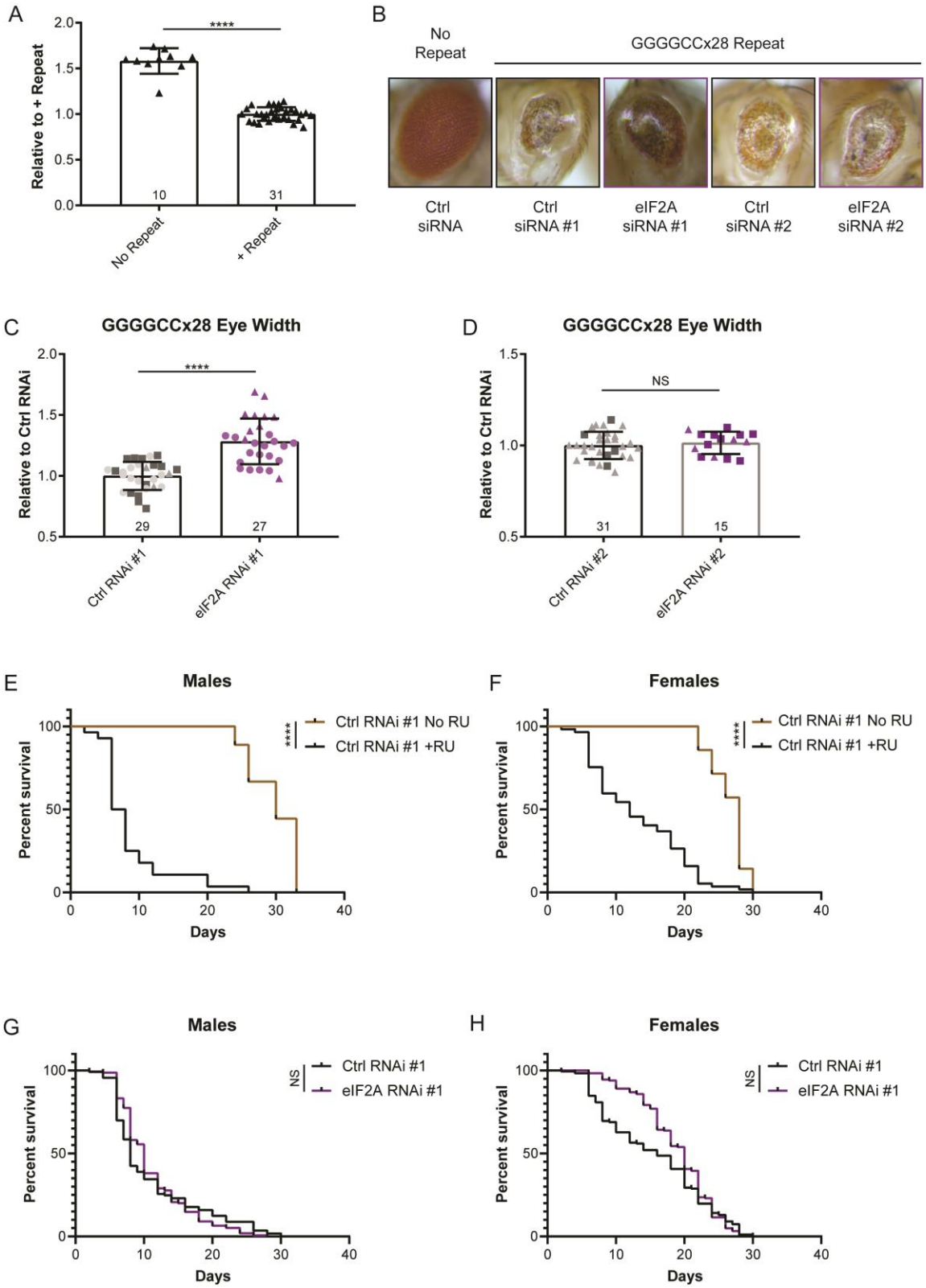
hours post transfection with non-targeting or eIFA, eIF2D, or DENR siRNAs, respectively. NLuc levels are expressed relative to levels in cells transfected with the non-targeting siRNA, n=9. (E-G) Expression of FFLuc reporter in wells co-transfected with AUG-NLuc, 24 hours post transfection with non-targeting or eIFA, eIF2D, or DENR siRNAs. FFLuc levels are expressed relative to levels in cells transfected with the non-targeting siRNA, n=9-12. All graphs represent mean with errors +/- standard deviation. * $p< 0.05$, ** $p< 0.01$, *** $p< 0.001$, **** $p< 0.0001$, (B-D) Two-way ANOVA with Sidak's multiple comparison test. (E-G) Unpaired t-test with Welch's correction.

Figure 4-5. KD of eIF2 alternatives does not prevent ISR-induced increases in RAN translation



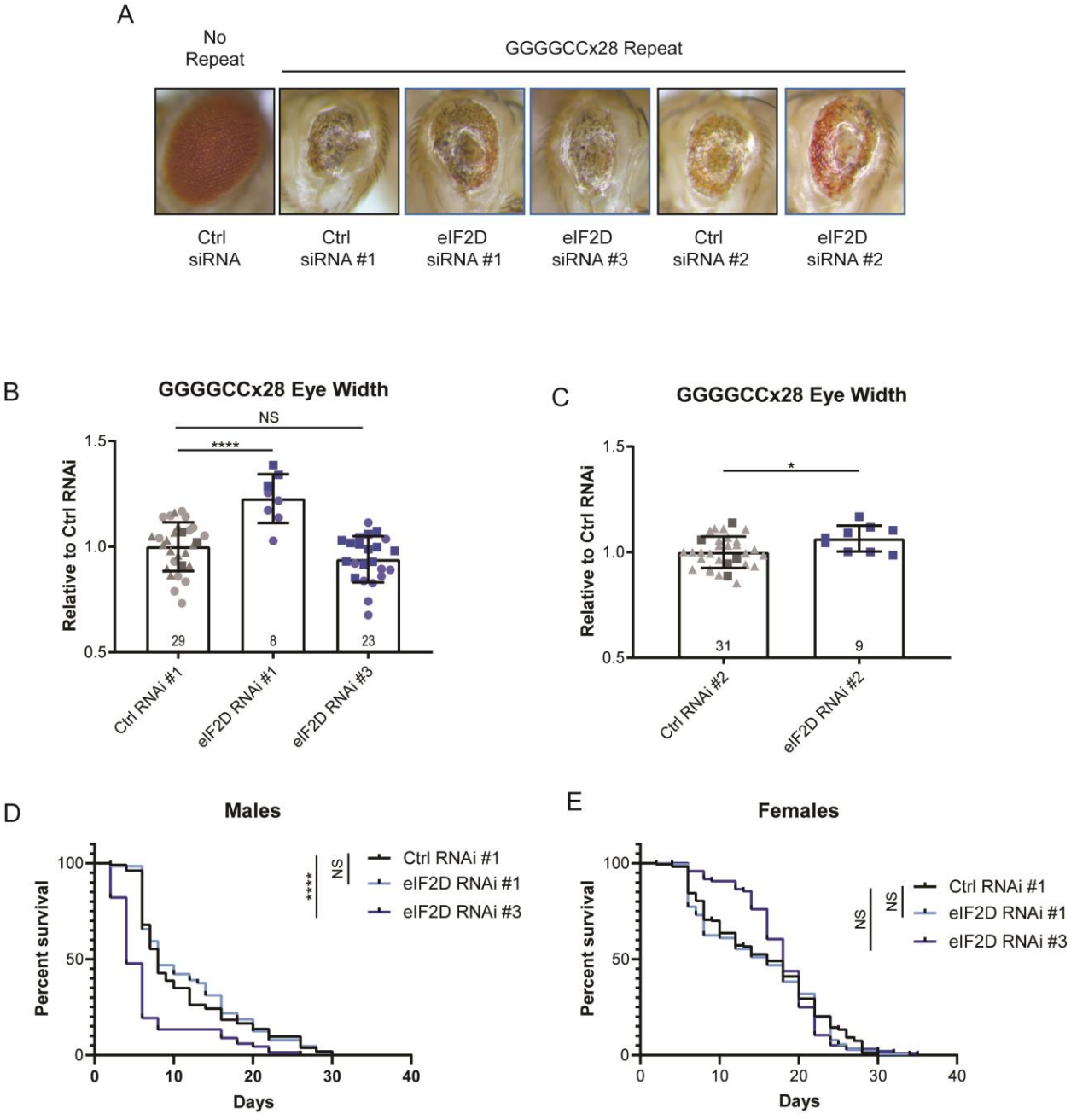
(A-C) HEK293 cells were transfected with eIF2A, eIF2D, or DENR siRNAs, respectively, as well as a non-targeting control siRNA for 24 hours before transfection with NLuc reporter plasmids. 19 hours post reporter transfection, HEK293 cells were treated with 2 µM thapsigargin (TG) for 5 hours, before NLuc levels were measured. NLuc levels are expressed relative to vehicle (DMSO) treated cells. All graphs represent mean with errors +/- standard deviation. ** $p < 0.01$, Two-way ANOVA with Sidak's multiple comparison test.

Figure 4-6. KD of *Drosophila* eIF2A homolog modestly improves GGGGCC-repeat-mediated eye toxicity



(A) Quantification of eye width in flies expressing control (ctrl) shRNA #1 in the absence or presence of the GGGGCCx28 repeat, under the GMR-GAL4 driver at 29^{o}C. Experimental numbers are indicated within bars, and each fly is represented by a single data point. (B) Representative images of fly eyes expressing control or eIF2A shRNA in the presence of absence of the GGGGCCx28 repeat, under the GMR-GAL4 driver at 29^{o}C. (C-D) Comparison of eye width in flies expressing indicated ctrl or eIF2A shRNA in the presence of the GGGGCCx28 repeat, under the GMR-GAL4 driver at 29^{o}C. Experimental numbers are indicated within bars, and each fly is represented by a single data point. Shape of data points represent experimental trials that data points were collected from. (E-F) Survival curve of control male and female Tub5-GAL4 GeneSwitch flies, respectively. Survival of flies with induced expression of the GGGGCCx28 repeat and control shRNA through RU486 treatment (+RU), is compared to survival of flies without induced expression (No RU). (G-H) Survival curve of RU486-treated GGGGCCx28 male and female Tub5-GAL4 GeneSwitch flies, respectively. Survival of flies expressing control shRNA #1 compared to flies expressing eIF2A shRNA. Bar graphs represent mean with errors +/- standard deviation. **** $p < 0.0001$, unpaired t-test with Welch's correction. For survival curves, **** $p < 0.0001$, mantel-cox test.

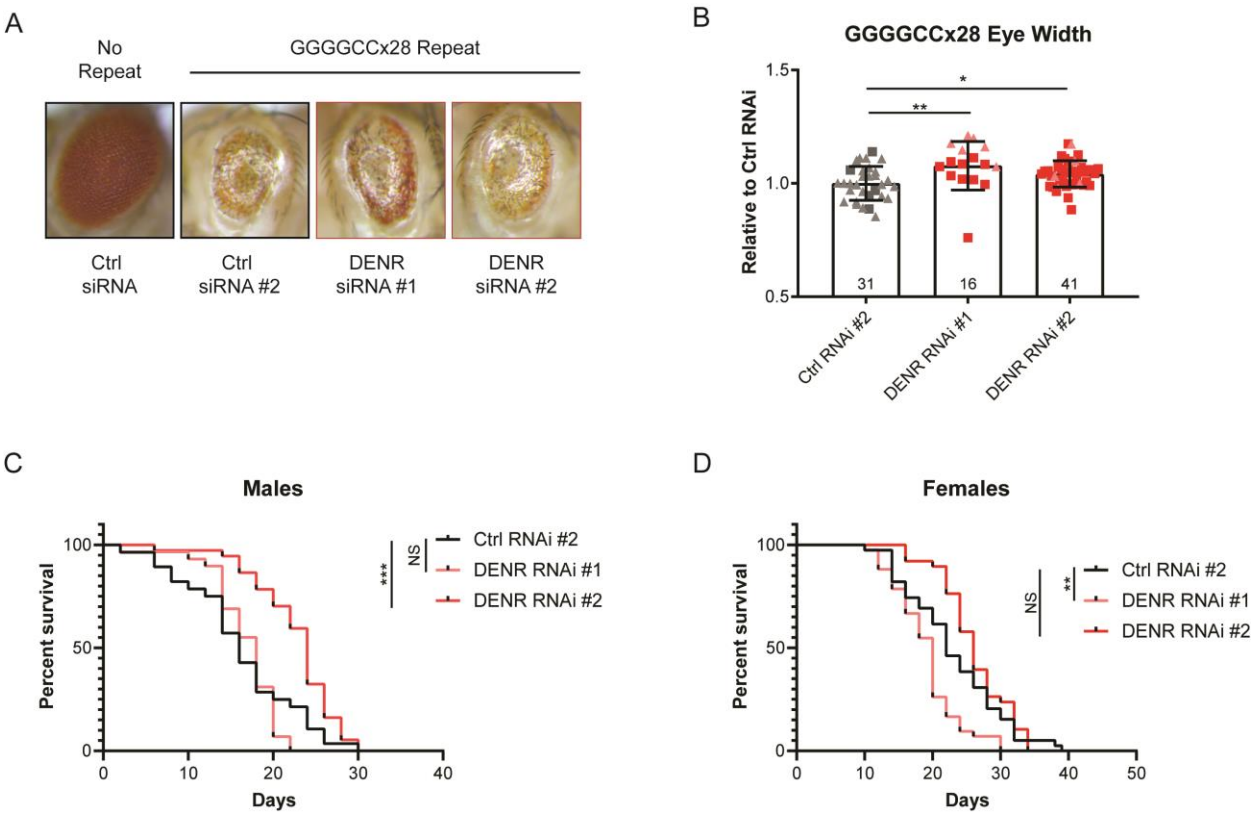
Figure 4-7. KD of *Drosophila* eIF2D homolog modestly improves GGGGCC-repeat-mediated eye toxicity



(A) Representative images of fly eyes expressing control or eIF2D shRNAs in the presence of absence of the GGGGCCx28 repeat, under GMR-GAL4 driver at 29^{o}C. (B-C) Comparison of eye width in flies expressing indicated ctrl or eIF2D shRNAs in the

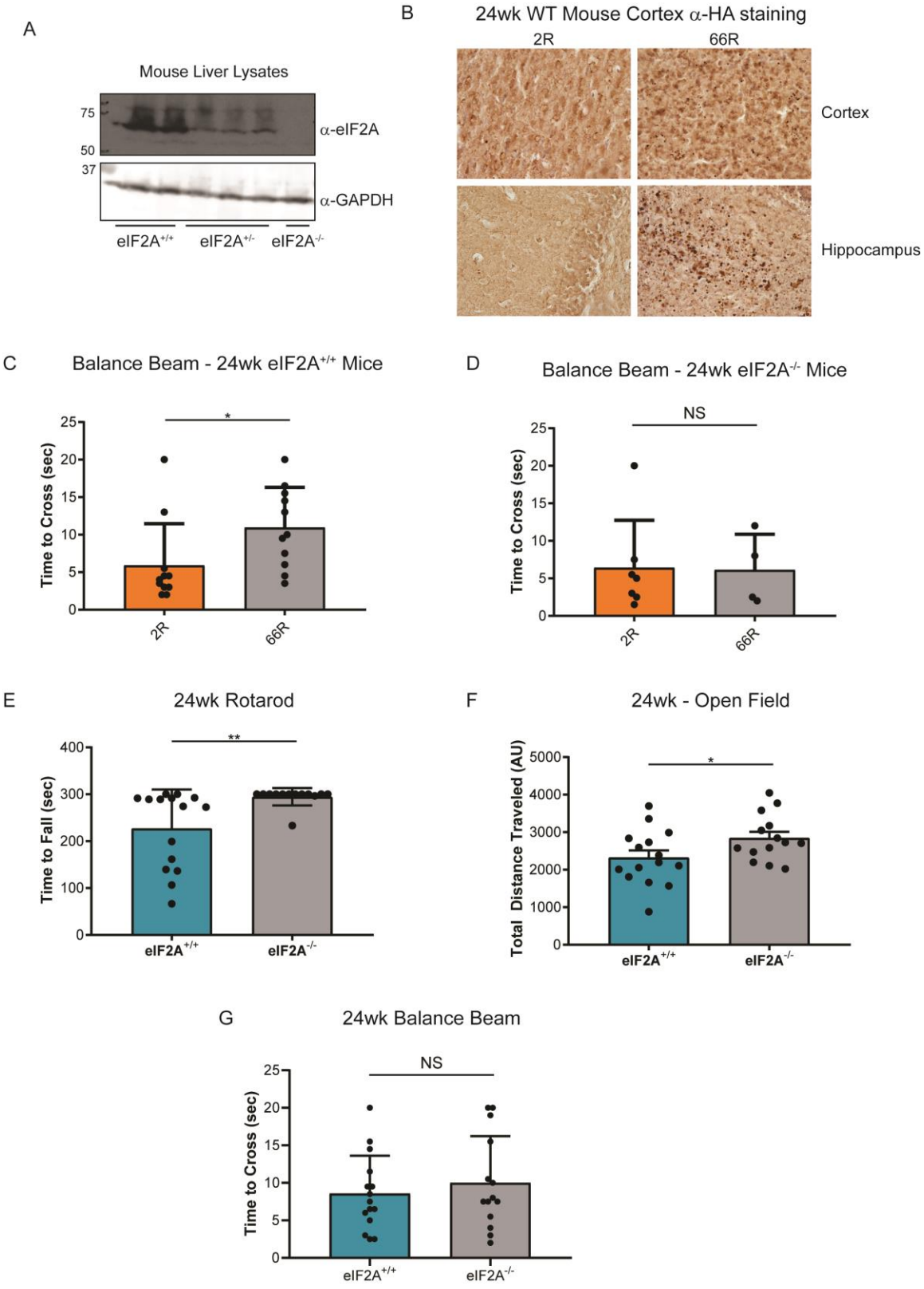
presence of the GGGGCCx28 repeat, under GMR-GAL4 driver at 29^{o}C. Experimental numbers are indicated within bars, and each fly is represented by a single data point. Shape of data points represent experimental trials that data points were collected from. (G-H) Survival curve of RU486-treated GGGGCCx28 male and female Tub5- GAL4 GeneSwitch flies, respectively. Survival of flies expressing control shRNA #1 compared to flies expressing eIF2D shRNAs #1 and #3. Bar graphs represent mean with errors +/- standard deviation. * $p < 0.05$, **** $p < 0.0001$, One-way ANOVA with Dunnett's multiple comparison test or unpaired t-test with Welch's correction. For survival curves, **** $p < 0.0001$, mantel-cox test.

Figure 4-8. KD of Drosophila DENR homolog modestly improves GGGGCC-repeat-mediated eye toxicity and prolongs survival



(A) Representative images of fly eyes expressing control or DENR shRNAs in the presence of absence of the GGGGCCx28 repeat, under GMR-GAL4 driver at 29ºC. (B) Comparison of eye width in flies expressing indicated ctrl or DENR shRNAs in the presence of the GGGGCCx28 repeat, under GMR-GAL4 driver at 29ºC. Experimental numbers are indicated within bars, and each fly is represented by a single data point. Shape of data points represent experimental trials that data points were collected from. (C-D) Survival curve of RU486-treated GGGGCCx28 male and female Tub5-GAL4 GeneSwitch flies, respectively. Survival of flies expressing control shRNA #2 compared to flies expressing DENR shRNAs #1 and #2. Bar graphs represent mean with errors +/- standard deviation. * p< 0.05, **** p< 0.0001, One-way ANOVA with Dunnett's multiple comparison test. For survival curves, ** p< 0.01, **** p< 0.0001, mantel-cox test.

Figure 4-9. Preliminary toxicity of GGGGCC repeats in eIF2A$^{+/+}$ verses eIF2A$^{-/-}$ mice



(A) Western blot of liver lysates harvested from eIF2A+/+, eIF2A+/-, and eIF2A-/- mice, stained with anti-eIF2A antibody to confirm protein deletion. GAPDH was used as loading control. (B) Representative images of mouse cortex and hippocampus harvested from 24-week-old WT mice injected with GGGGCCx2 (2R) or GGGGCCx66 (66R) expressing AAVs at P0. DAB staining with anti-HA antibody that detects polyGA product, and hematoxylin staining to color nuclei. 60X magnification. (C-D) PRELIMINARY DATA: Comparison between the amount of time 24-week-old eIF2A+/+ and eIF2A-/- mice, injected with 2R or 66R AAVs take to cross 44cm long, 5mm wide square balance beam. (E) Comparison of the latency to fall for uninjected 24-week-old eIF2A+/+ and eIF2A-/- mice on the rotarod, with constant 4-40 rpm ramping over 300 seconds. (F) Comparison of total distance traveled by uninjected 24-week-old eIF2A+/+ and eIF2A-/- mice in an open field, expressed as arbitrary units. (G) Comparison between the amount of time uninjected 24-week-old eIF2A+/+ and eIF2A-/- mice take to cross 44cm long, 5mm wide square balance beam. Bar graphs represent mean with errors +/- standard deviation. All data points represent the average of two trials for each individual animal, except in (F) where data point represent a single trial for each individual animal. NS $p > 0.05$, * $p < 0.05$, ** $p < 0.01$, unpaired student's t-test with Welch's correction.

4.8 Tables

Table 4-1. siRNA information

siRNA Target	Product Information	Sequence Sense/Antisense
eIF2A	Thermo stealth siRNA – HSS130478	ACGAAACACUGUCUCUCAGUCAAUU/ AAUUGACUGAGAGACAGUGUUUCGU
eIF2D	Thermo stealth siRNA – HSS103085	GGACAGGAGAAAGCUUCGAGCUGAU/ AUCAGCUCGAAGCUUUCUCCUGUCC
DENR	Thermo stealth siRNA – HSS112532	ACCAACAGAGUACUGUGAAUAUAUG/ CAUAUAUUCACAGUACUCUGUUGGU
EGFP (non-targeting control)	Thermo stealth siRNA	CACAUGAAGCAGCACGACUUCUUCA/ UGAAGAAGUCGUGCUGCUUCAUGUG

Table 4-2. Primary antibody information

Antibody Target	Product Information	Species	Probing Conditions
eIF2A	Protein Tech, 11233-1-AP	Rabbit	1:8000 - 1 hour at room temperature
eIF2A	Abcam; ab169528	Rabbit	1:1000 – overnight at 4C
eIF2D	Protein Tech, 12840-1-AP	Rabbit	1:1000 – overnight at 4C
DENR	Sigma, Clone 1H3, WH0008562M1	Mouse	1:1000 – overnight at 4C
MCTS1	Sigma, SAB2701331 (discontinued)	Rabbit	1:1000 – overnight at 4C
FLAG	Sigma, M2, F1804	Mouse	1:1000 – overnight at 4C
GAPDH	SCBT, sc-32233	Mouse	1:1000 – 1 hour at room temperature
HA	Biolegend/Fisher; HA 11 16B12, 50-103-0101	Mouse	IHC: 1:100 – overnight at 4C

Table 4-3. *Drosophila* line information

Name used within chapter	Gene target	Vendor	Stock #
GMR-Gal4	-	BDSC	8605
Tub5-GS-Gal4	-	Internal	-
Ctrl shRNA #1	Luciferase	BDSC	31603
Ctrl shRNA #2	LexA	BDSC	67947
eIF2A shRNA #1	CG7414	BDSC	50649
eIF2A shRNA #2	CG7414	VDRC	107955/KK
eIF2D shRNA #1	eIF2D	BDSC	33995
eIF2D shRNA #2	eIF2D	VDRC	100304
eIF2D shRNA #3	eIF2D	VDRC	21340
DENR shRNA #1	DENR	VDRC	101746
DENR shRNA #2	DENR	VDRC	49895

4.9 References

1. Ingolia, N. T., Lareau, L. F., and Weissman, J. S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789-802
2. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., and Weissman, J. S. (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223
3. Kearse, M. G., and Wilusz, J. E. (2017) Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev* **31**, 1717-1731
4. Imataka, H., Olsen, H. S., and Sonenberg, N. (1997) A new translational regulator with homology to eukaryotic translation initiation factor 4G. *EMBO J* **16**, 817-825
5. Kearse, M. G., Goldman, D. H., Choi, J., Nwaezeapu, C., Liang, D., Green, K. M., Goldstrohm, A. C., Todd, P. K., Green, R., and Wilusz, J. E. (2019)

Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors. *Genes Dev* **33**, 871-885

6. Starck, S. R., Tsai, J. C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., Martins-Green, M., Shastri, N., and Walter, P. (2016) Translation from the 5' untranslated region shapes the integrated stress response. *Science* **351**, aad3867
7. Skabkin, M. A., Skabkina, O. V., Dhote, V., Komar, A. A., Hellen, C. U., and Pestova, T. V. (2010) Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev* **24**, 1787-1801
8. Merrick, W. C., and Anderson, W. F. (1975) Purification and characterization of homogeneous protein synthesis initiation factor M1 from rabbit reticulocytes. *J Biol Chem* **250**, 1197-1206
9. Starck, S. R., Jiang, V., Pavon-Eternod, M., Prasad, S., McCarthy, B., Pan, T., and Shastri, N. (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719-1723
10. Sendoel, A., Dunn, J. G., Rodriguez, E. H., Naik, S., Gomez, N. C., Hurwitz, B., Levorse, J., Dill, B. D., Schramek, D., Molina, H., Weissman, J. S., and Fuchs, E. (2017) Translation from unconventional 5' start sites drives tumour initiation. *Nature* **541**, 494-499
11. Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., Nan, Z., Forster, C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. (2011) Non-

ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* **108**, 260-265

12. Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., Pletnikova, O., Borchelt, D. R., Ross, C. A., Margolis, R. L., Yachnis, A. T., Troncoso, J. C., and Ranum, L. P. (2015) RAN Translation in Huntington Disease. *Neuron* **88**, 667-677
13. Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. C., Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., and Paulson, H. L. (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440-455
14. Green, K. M., Glineburg, M. R., Kearse, M. G., Flores, B. N., Linsalata, A. E., Fedak, S. J., Goldstrohm, A. C., Barmada, S. J., and Todd, P. K. (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications* **8**, 2005
15. Cheng, W., Wang, S., Mestre, A. A., Fu, C., Makarem, A., Xian, F., Hayes, L. R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., Cleveland, D. W., and Sun, S. (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat Commun* **9**, 51
16. Sonobe, Y., Ghadge, G., Masaki, K., Sendoel, A., Fuchs, E., and Roos, R. P. (2018) Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol Dis* **116**, 155-165

17. Westergard, T., McAvoy, K., Russell, K., Wen, X., Pang, Y., Morris, B., Pasinelli, P., Trotti, D., and Haeusler, A. (2019) Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med* **11**

18. Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., and Todd, P. K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**, 314-322

19. Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C. Z., Lin, C. C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., Désaubry, L., Gendron, T., Petrucelli, L., Martin, F., and Lagier-Tourenne, C. (2018) CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* **9**, 152

20. Malina, A., Khan, S., Carlson, C. B., Svitkin, Y., Harvey, I., Sonenberg, N., Beal, P. A., and Pelletier, J. (2005) Inhibitory properties of nucleic acid-binding ligands on protein synthesis. *FEBS Lett* **579**, 79-89

21. Robert, F., Kapp, L. D., Khan, S. N., Acker, M. G., Kolitz, S., Kazemi, S., Kaufman, R. J., Merrick, W. C., Koromilas, A. E., Lorsch, J. R., and Pelletier, J. (2006) Initiation of protein synthesis by hepatitis C virus is refractory to reduced eIF2.GTP.Met-tRNA(i)(Met) ternary complex availability. *Mol Biol Cell* **17**, 4632-4644

22. Zoll, W. L., Horton, L. E., Komar, A. A., Hensold, J. O., and Merrick, W. C. (2002) Characterization of mammalian eIF2A and identification of the yeast homolog. *J Biol Chem* **277**, 37079-37087

23. Adams, S. L., Safer, B., Anderson, W. F., and Merrick, W. C. (1975) Eukaryotic initiation complex formation. Evidence for two distinct pathways. *J Biol Chem* **250**, 9083-9089

24. Golovko, A., Kojukhov, A., Guan, B. J., Morpurgo, B., Merrick, W. C., Mazumder, B., Hatzoglou, M., and Komar, A. A. (2016) The eIF2A knockout mouse. *Cell Cycle* **15**, 3115-3120

25. González-Almela, E., Williams, H., Sanz, M. A., and Carrasco, L. (2018) The Initiation Factors eIF2, eIF2A, eIF2D, eIF4A, and eIF4G Are Not Involved in Translation Driven by Hepatitis C Virus IRES in Human Cells. *Front Microbiol* **9**, 207

26. Sanz, M. A., González Almela, E., and Carrasco, L. (2017) Translation of Sindbis Subgenomic mRNA is Independent of eIF2, eIF2A and eIF2D. *Sci Rep* **7**, 43876

27. Kim, J. H., Park, S. M., Park, J. H., Keum, S. J., and Jang, S. K. (2011) eIF2A mediates translation of hepatitis C viral mRNA under stress conditions. *EMBO J* **30**, 2454-2464

28. Ventoso, I., Sanz, M. A., Molina, S., Berlanga, J. J., Carrasco, L., and Esteban, M. (2006) Translational resistance of late alphavirus mRNA to eIF2alpha phosphorylation: a strategy to overcome the antiviral effect of protein kinase PKR. *Genes Dev* **20**, 87-100

29. Chew, J., Gendron, T. F., Prudencio, M., Sasaguri, H., Zhang, Y. J., Castanedes-Casey, M., Lee, C. W., Jansen-West, K., Kurti, A., Murray, M. E., Bieniek, K. F., Bauer, P. O., Whitelaw, E. C., Rousseau, L., Stankowski, J. N., Stetler, C., Daughrity, L. M., Perkerson, E. A., Desaro, P., Johnston, A., Overstreet, K.,

Edbauer, D., Rademakers, R., Boylan, K. B., Dickson, D. W., Fryer, J. D., and Petrucelli, L. (2015) Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* **348**, 1151-1154

30. Jaafar, Z. A., Oguro, A., Nakamura, Y., and Kieft, J. S. (2016) Translation initiation by the hepatitis C virus IRES requires eIF1A and ribosomal complex remodeling. *Elife* **5**

31. Su, Z., Zhang, Y., Gendron, T. F., Bauer, P. O., Chew, J., Yang, W. Y., Fostvedt, E., Jansen-West, K., Belzil, V. V., Desaro, P., Johnston, A., Overstreet, K., Oh, S. Y., Todd, P. K., Berry, J. D., Cudkowicz, M. E., Boeve, B. F., Dickson, D., Floeter, M. K., Traynor, B. J., Morelli, C., Ratti, A., Silani, V., Rademakers, R., Brown, R. H., Rothstein, J. D., Boylan, K. B., Petrucelli, L., and Disney, M. D. (2014) Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**, 1043-1050

32. Castelo-Szekely, V., De Matos, M., Tusup, M., Pascolo, S., Ule, J., and Gatfield, D. (2019) Charting DENR-dependent translation reinitiation uncovers predictive uORF features and links to circadian timekeeping via Clock. *Nucleic Acids Res* **47**, 5193-5209

33. Rodriguez, C. M., Wright, S. E., Kearse, M. G., Haenfler, J. M., Flores, B. N., Liu, Y., Ifrim, M. F., Glineburg, M. R., Krans, A., Jafar-Nejad, P., Sutton, M. A., Bassell, G. J., Parent, J. M., Rigo, F., Barmada, S. J., and Todd, P. K. (2020) A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat Neurosci*

34. Weisser, M., Schäfer, T., Leibundgut, M., Böhringer, D., Aylett, C. H. S., and Ban, N. (2017) Structural and Functional Insights into Human Re-initiation Complexes. *Mol Cell* **67**, 447-456.e447

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

36. Linsalata, A. E., He, F., Malik, A. M., Glineburg, M. R., Green, K. M., Natla, S., Flores, B. N., Krans, A., Archbold, H. C., Fedak, S. J., Barmada, S. J., and Todd, P. K. (2019) DDX3X and specific initiation factors modulate FMR1 repeat-associated non-AUG-initiated translation. *EMBO Rep* **20**, e47498

37. Kim, E., Kim, J. H., Seo, K., Hong, K. Y., An, S. W. A., Kwon, J., Lee, S. V., and Jang, S. K. (2018) eIF2A, an initiator tRNA carrier refractory to eIF2α kinases, functions synergistically with eIF5B. *Cell Mol Life Sci* **75**, 4287-4300

38. Costa, M. o. C., Luna-Cancalon, K., Fischer, S., Ashraf, N. S., Ouyang, M., Dharia, R. M., Martin-Fishman, L., Yang, Y., Shakkottai, V. G., Davidson, B. L., Rodríguez-Lebrón, E., and Paulson, H. L. (2013) Toward RNAi therapy for the polyglutamine disease Machado-Joseph disease. *Mol Ther* **21**, 1898-1908

Chapter 5: RAN-Translation-Competent GGGGCC Repeats Are Present in Capped C9orf72 Transcripts with Novel 5′ Start Sites

5.1 Statement of other's contribution to data presented in this chapter

iNeurons used in Fig. 5-3 were generated by Elizabeth Tank, who initially inserted the neurogenin expression cassette into patient iPSCs, and performed the neuronal differentiations and maturations. I generated all other reagents and data presented in this chapter. I also analyzed all the data, generated each figure, and wrote the entire chapter, with the exception of the methods section related to iNeuron work, which was written by Elizabeth Tank.

5.2 Introduction

In 2011, a GGGGCC hexanucleotide repeat expansion (HRE) mutation in the gene C9orf72 was identified as the most common genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) (1,2), accounting for approximately 37% of familial ALS, 25% of familial FTD, and approximately 6% of sporadic cases for both diseases (3). Most healthy individuals contain only 2 GGGGCC repeats in C9orf72, while those with C9ALS/FTD typically have several hundreds to thousands (1,2).

Shortly after its discovery, the HRE was shown to undergo a non-canonical form of translation initiation, termed repeat-associated non-AUG (RAN) translation (4-7). As a

result, six different dipeptide-repeat-containing proteins (DPRs), generated through RAN translation of the sense and antisense repeat, accumulate within inclusions in patient neurons (4,5,7). The DPRs produced from the sense GGGGCC strand include a poly-glycine-alanine (GA), a poly-glycine proline (GP), and a poly-glycine-arginine (GR) containing protein, while the antisense CCCCGG strand produces a poly-proline-arginine (PR), poly-proline-alanine (PA), and a second poly-GP containing protein. Across model systems, overexpression of DPRs, even in the absence of the repeating mRNA sequence, causes severe toxicity (5,8-15).

To understand the mechanism by which the HRE undergoes RAN translation, multiple groups have developed C9RAN translation reporters for the expanded sense GGGGCC repeat. Despite differences in the design of these reporters, several common mechanistic features have emerged across studies. First, C9RAN translation occurs most robustly in the poly-GA reading frame (16-20). Second, while C9RAN translation can occur from the second cistron of bicistronic reporters, and in a cap-independent manner (19,20), C9RAN translation occurs most efficiently from RNAs containing a functional 5′ m^7G cap (16,17,19,20). Third, a CUG start codon located in good Kozak sequence context 24 nucleotides (nts) upstream of the GGGGCC repeat is used for RAN translation initiation in the poly-GA frame, as loss of this codon dramatically reduces poly-GA expression across systems, including patient-derived motor neurons, and from both monocistronic and bicistronic reporters (16,17,19,21). Lastly, C9RAN translation is upregulated following induction of the integrative stress response and phosphorylation of eIF2α (16,18-20).

While these reporter systems have provided critical insight into the mechanism of C9RAN translation, more work is needed to fully understand how this process occurs on endogenous C9orf72 RNAs in patient cells. In particular, as reflected by the various designs of C9RAN translation reporters developed by different groups, the actual sequence context of the endogenously translated HRE-containing C9orf72 RNA is unknown. This is because, under normal circumstances, the GGGGCC repeat is believed to be located either within the first intron of C9orf72 or within its promoter. Therefore, proper splicing and degradation of this intron within the nucleus should preclude the repeat from interacting with cytoplasmic ribosomes to undergo RAN translation. However, the presence of DPRs in patient neurons indicates that this is not the case.

There are multiple models by which the HRE could evade nuclear degradation. First, impaired splicing of the HRE would allow it to be exported to the cytoplasm within otherwise mature C9orf72 transcripts (22-24). Alternatively, the spliced intron itself could evade degradation, and undergo nuclear export in either a linearized or lariat form (20,25-27). Lastly, it is possible that the HRE interferes with C9orf72 transcription, generating novel transcripts lacking the 5′ or 3′ splice site, either through altered transcription initiation (28,29) or premature transcription termination (30), respectively, so that the HRE is no longer in an intronic context.

To address this question, here we performed 5′ RNA-ligation mediated rapid amplification of cDNA ends (RLM-RACE) on C9orf72 transcripts from control and C9ALS/FTD patient fibroblasts and iNeurons using GGGGCC repeat targeting primers. In doing so, we identified multiple functionally capped C9orf72, repeat-containing

transcripts with novel 5′ ends occurring in close proximity to the GGGGCC repeat. Interestingly, several of these species contain the CUG start codon previously identified as important for poly-GA production. Using polysome profiling and C9RAN translation reporters with the newly identified 5′ ends, we show that these short C9orf72 transcripts are associated with translating ribosomes in patient neurons and are capable of supporting RAN translation, and are thus likely contribute to DPR burden in patients.

5.3 Results

Upstream sequence context impacts the efficiency of C9RAN translation

To determine how upstream sequence context affects C9RAN translation, we generated nanoluciferase (NLuc) reporters containing either 162nts of intron1 found upstream of the GGGGCC repeat in C9orf72, or this sequence along with the additional 158nts of the upstream exon1a, inserted above 70 GGGGCC repeats (**Fig. 5-1, A**). Consistent with previous findings (17), when expressed as *in vitro* transcribed mRNAs in a rabbit reticulocyte lysate (RRL) *in vitro* translation system, the exonic sequence significantly reduced expression of the RAN reporters for all three reading frames relative to reporters containing only the intronic sequence (**Fig. 5-1, B**). This inhibition remained even when an AUG start codon was inserted upstream of the repeat to drive initiation in the GA reading frame (**Fig. 5-1, C**). This finding was unexpected and prompted us to look more closely at the exonic sequence. Within it, we found an AUG start codon that creates a short upstream open reading frame (uORF) terminating halfway into the intronic sequence (**Fig. 5-1, A**). Mutating this exonic AUG to AAA,

restored or increased C9RAN translation for all three reading frames in RRL, suggesting that in this context it acts to support production of an inhibitory uORF (**Fig. 5-1, B & C**).

To evaluate whether these same findings were observed in cells, we expressed reporter plasmids in HEK293T. As observed in RRL, the exonic sequence inhibited C9RAN translation in all three reading frames relative to the intron-only sequence (**Fig. 5-1, D**). However, in these cells, the AUG to AAA mutation only partially restored GA70 and GP70 expression, and did not increase GR70 expression (**Fig. 5-1, D**). These data demonstrate that upstream sequence context can affect C9RAN translation levels, and suggest that in patients, C9orf72 transcripts in which the HRE is downstream of only intronic sequence would support more RAN translation than transcripts resulting from intron retention, due to the presence of this inhibitory uORF.

Repeat-primed 5′ RACE in patient fibroblasts reveals novel C9orf72 5′ ends

To identify the sequence 5′ of the GGGGCC repeat in C9orf72 expressed from the endogenous locus in cells, we isolated total RNA from one control (Ctrl #1) and two C9orf72 derived patient fibroblasts (C9 #1 and C9 #2) and performed 5′ RLM-RACE (**Fig. 5-2, A**). Through sequential treatment of RNA with calf intestine phosphatase (CIP) to remove 5′ monophosphates resulting from RNA degradation, followed by tobacco acid pyrophosphatase (TAP) treatment to remove 5′ caps and expose new 5′ monophosphates, a 5′ RNA oligo was ligated only to the ends of previously capped, full-length mRNAs (**Fig. 5-2, A**).

cDNA was then synthesized using a pool of random primers and CCCCGGx4 primers, to enrich for RNAs that contain a GGGGCC repeat and increase the likelihood

of cDNA synthesis making it through the HRE, which is known to impair PCR (1) (**Fig. 5-2, A**). Following RNA degradation with RNAse H, we performed a control RACE reaction using primers specific for beta-actin and the 5′ ligated RNA oligo. The presence of a single band of the expected size on an agarose gel indicates that all steps of the reactions were successful for both the control and C9ALS/FTD sample (**Fig. 5-2, B**); CIP treatment prevented amplification of any degraded/truncated transcripts, TAP treatment properly exposed full-length RNAs for ligation, and RNA ligation successfully added the RNA oligo used to prime the RACE reaction.

We next used a series of repeat/C9orf72-specific primers (**Fig. 5-2, C to F**). The first primer (P1) is complimentary to only the GGGGCC repeat and the first three nts located upstream of the repeat in C9orf72 (**Fig. 5-2, D)**. We choose such a non-specific primer to allow us to detect any possible capped RNA species with 5′ ends located immediately upstream of the repeat. Sequencing ten clones from Ctrl #1 revealed amplification of the 5′ ends of nine different genes containing GGGGCC or GGGGCC-like sequences, none of which were C9orf72 (**Fig. 5-2, D and Table 5-1**). However, five of the six clones sequenced from C9 #1 contained no sequence between the GGGGCC and the RNA oligo (**Fig. 5-2, D and Table 5-1**).

The presence of capped RNAs with 5′ ends within a GGGGCC repeat, in the C9ALS/FTD but not control line, represents an intriguing possibility that the HRE promotes transcription initiation within itself. However, the lack of unique sequence prevents conclusive mapping of these 5′ ends to C9orf72. We therefore also used two additional primers; one complimentary to two repeats + 12nts of C9orf72 upstream of

the repeat (P2), and the other entirely complimentary C9orf72 sequence upstream of the repeat (P3; **Fig. 5-2, E and F**).

5′ RACE with P2 amplified two C9orf72-specific products in the C9 #1 line. The first 5′ end is immediately adjacent to the end of the P2 sequence (**Fig. 5-2, E**). The second occurs 20nts upstream of P2, encompassing 31nt of C9orf72 sequence upstream of the repeat (**Fig. 5-2, E**). Interestingly, the 5′ end of the 31nt sequence begins with a +1 A following a -1 T, a short motif enriched at transcription start sites (TSSs) (31). Of note, this longer species begins 8nts upstream of the CUG codon previously identified as being important for polyGA synthesis (16,17,19,21).

5′ RACE on C9 #1 using P3 also amplified the same C9orf72 species beginning 31nts upstream of the repeat in C9 #1 (**Fig. 5-2, F**). However, P3 would not be able to amplify the shorter transcripts identified by P1 and P2, due to its priming location upstream of or within these sites (**Fig. 5-2, F**). Both C9orf72 5′ ends were also detected following 5′ RLM-RACE of a second C9orf72-patient fibroblast line (C9 #2, **Fig. 5-2, E**). They were also amplified within the Ctrl #1 line, but at a lower frequency than either C9ALS/FTD line (**Fig. 5-2, E and Table 5-1**). Together, these data provide evidence suggesting that capped, but shortened isoforms of C9orf72 with 5′ ends in close proximity to the GGGGCC repeat are enriched in C9ALS/FTD patient cells.

Short C9orf72 isoforms are actively translated in patient-derived iNeurons

To determine whether C9orf72 transcripts with these novel 5′ ends are capable of undergoing RAN translation in disease-relevant cells, we next generated lysates from DIV10 control and C9orf72-patient derived iNeurons (**Fig. 5-3, A**). We then fractionated

these lysates on 10-50% sucrose gradients using ultracentrifugation, and performed polysome profiling (**Fig. 5-3, A to C**). Actively translated fractions, including the 80S fraction and all polysome fractions, were pooled for RNA isolation (**Fig. 5-3, A to C**). The actively translated iNeuron RNA was then processed for 5′ RLM-RACE using the same protocol described above.

As before, RACE with P2 identified only short C9orf72 isoforms, including the same species identified within total fibroblast RNA, that begins immediately upstream of the P2 sequence (**Fig. 5-3, D**). Interestingly, this species was equally enriched in the control and C9ALS/FTD line (**Fig. 5-3, D**). As Ctrl #2 contains a larger than average GGGGCC repeat length (9 vs. 2), this may suggest that differences in even non-pathological repeat lengths can promote the use of different C9orf72 RNA 5′ ends. Alternatively, these shorter C9orf72 species may be enriched under normal conditions within actively translated RNA populations or are more abundantly generated in neurons compared to fibroblasts. Further experiments using additional lines would help discern between these possibilities.

RACE with P2 also amplified a rare, longer C9orf72 isoform that contained 71nt of intronic sequence upstream of the GGGGCC repeat, that was only present in a single clone from C9 #3 (**Fig. 5-3, D**). As with the 5′ end beginning 31nt upstream of the repeat, this species initiated at a -1T, +1A motif. RACE with P3 also detected the same C9orf72 transcript beginning 31nt upstream of the GGGGCC repeat, previously detected in total fibroblast RNA (**Fig. 5-3, D**). This species was similarly enriched within Ctrl #2 and C9 #3 clones (**Fig. 5-3, D**).

Consequently, short, functionally-capped C9orf72 isoforms are present within the actively translated fraction of iNeuron RNA. This suggests that these RNA species could contribute to DPR production in patient brains.

Reporters containing short C9orf72 5' ends undergo RAN translation and utilize the CUG start codon

To further understand the capabilities of short C9orf72 transcripts to support RAN translation, we developed additional NLuc reporters with either the 31nt or 71nt C9orf72 sequence inserted upstream of 70 GGGGCC repeats (**Fig. 5-4, A**). The C9orf72 sequence was inserted immediately after the T7 promoter, so no additional vector sequence was present in RNAs following *in vitro* transcription. We also generated new "whole intron" reporters in the same manner, to compare expression of the shorter C9orf72 RNAs to (**Fig. 5-4, A**). Of note, all of these sequences contained the CUG start codon (**Fig. 5-4, A**).

As with all other previously developed C9RAN translation reporters (16-20), when expressed as *in vitro* transcribed RNA in RRL or HEK293 cells, polyGA was produced at higher levels than polyGP or polyGR (**Fig. 5-3, B & F**). Additionally, despite being located only 8nts from the 5' cap, the CUG codon was again found to be important for polyGA production from the 31nt reporters, as its mutation to CCC decreased polyGA levels in RRL and HEK293 cells (**Fig. 5-3, C & G**). As seen before (16,17), loss of the CUG codon had less of an inhibitory effect on polyGP production in both systems (**Fig. 5-3, C & G**). However, in comparison to the whole intron reporters, expression of the C9RAN reporters was significantly reduced from the 31nt reporters in

RRL and HEK293 cells (**Fig. 5-3, C & G**). In contrast, relative to a reporter containing the entire upstream intron1 sequence, polyGR expression was significantly enhanced from the 71nt reporter in RRL (**Fig. 5-3, E**).

These results again highlight the role of upstream sequence context in regulating C9RAN translation levels. They demonstrate that short isoforms of C9orf72, present in monosome or polysome fractions of iNeurons, can undergo RAN translation. In future studies, it will be interesting to determine if RAN translation occurs in each reading frame from reporters that start even closer to the repeat sequence and lack the CUG start codon.

5.4 Discussion

Since discovering that the GGGGCC HRE undergoes RAN translation in C9ALS/FTD patient neurons, the question of how a normally intronic sequence becomes accessible to ribosomes has remained unanswered. While reporter-based studies have provided insight into how different sequence contexts affect levels of GGGGCC RAN translation (16-20), which transcripts actually undergo RAN translation *in vivo* has not been previously assessed. Knowledge of whether the endogenous RAN-translated HRE resides within a retained intron, a spliced circularized or linearized lariat, or an RNA species in which the repeat is no longer intronic due to altered transcription initiation or termination, has important implications for developing strategies that effectively target the HRE and its ability to undergo RAN translation in patients. For instance, if DPR production primarily comes from an expanded repeat imbedded within an intron lariat, strategies that promote lariat debranching or inhibit cap-independent

translation could be effective at reducing C9RAN translation. However, these same strategies would have little benefit if DPR production arises from linear and capped RNA species.

Here, using a repeat-primed 5′ RLM-RACE approach, we identify novel, actively translated C9orf72 transcripts with 5′ ends that occur in close proximity to the expanded repeat. Successful amplification of these RNA species through 5′ RLM-RACE indicates a high likelihood that they contain a functional 5′ cap. Additionally, the CUG start codon present in two of the three identified transcripts is critical for polyGA production. Together, these data suggest that at least some DPR production in patient neurons is generated from capped, linear transcripts through use of a near-AUG start codon.

While these findings support one potential explanation for how the HRE undergoes RAN translation, they do not rule out contributions from other RNA species. In particular, the sequential treatment of RNAs with CIP and TAP prior to RNA oligo ligation allows only for the amplification of functionally capped RNAs. Consequently, any uncapped species, such as spliced introns, would be not detected through these experiments, nor would circularized RNAs lacking available 5′ ends for RNA oligo ligation. Therefore, additional studies to determine if repeat-containing lariat introns are present within the polysome fractions of iNeurons would provide valuable data in support of or against the contribution of these RNA species to DPR production.

Additionally, current experiments do not control for the possibility that the GGGGCC repeat structure impedes CIP activity, allowing uncapped repeat-containing RNAs to maintain their 5′ monophosphates and serve as suitable substrates for RNA

oligo ligation. Future experiments, in which the 5′ RLM-RACE reactions are repeated without the de-capping step would test this possibility.

An unexpected result from these experiments is that none of our GGGGCC or C9orf72-intronic primers amplified 5′ ends that encompass exon1 and begin at the TSS most commonly associated with expression of the C9orf72 isoforms predicted to contain the GGGGCC repeat, despite PCR cycling conditions set to detect these longer transcripts. This may suggest that in cDNA pools enriched for GGGGCC-repeat containing species, this TSS is less abundant than the shorter 5′ ends identified here. Additionally, consistent with our findings, cap analysis of gene express (CAGE) data (32), available through the UCSC genome browser, indicates use of many more C9orf72 TSSs in control individuals than generally recognized, including several within what is normally considered the first intron. Furthermore, our findings are consistent with another 5′ RACE study showing that the HRE promotes more variability in C9orf72 TSS locations than control length repeats (28).

The -1T, +1A motif present at multiple of the newly identified 5′ ends is consistent with the possibility that the shortened C9orf72 transcripts are generated through use of previously unmapped TSSs (31). This is also consistent with reports that GC-rich repeats exclude nucleosomes (33-35). However, it is also possible that they arise from cytoplasmic RNA recapping (36). Indeed, it has been estimated that approximately 30% of 5′ ends mapped through cap-capture techniques like CAGE are the result of recapping, as these sites do not align with chromosomal marks of active transcription (37). Future experiments investigating RNA-pol II occupancy and chromosome accessibility in control and C9ALS/FTD patient derived cells would help discern whether

transcription initiation is occurring at the novel 5′ C9orf72 ends identified here, and if the HRE increases the use of these TSSs, or if HRE-containing C9orf72 transcripts are substrates for cytoplasmic recapping. Identifying the mechanism through which these RAN-translated RNAs are generated could reveal new targets for preventing this event in patients.

5.5 Experimental methods

RNA synthesis, rabbit reticulocyte lysate in vitro translation, and HEK293 or HEK293T cell reporter expression

These experiments were performed as previously described (16,38). See method section in Chapter 2 or 4 for details. The sequences of unpublished reporters are included in **Table 5-2**.

Patient fibroblasts

Fibroblasts were maintained in 10 cm plates at 37^{o}C and 5% CO_2, in DMEM + high glucose (GE Healthcare Bio-Science, SH30022FS) supplemented with 9.09% fetal bovine serum (50 mL added to 500 mL DMEM; Bio-Techne, S11150), 1% MEM non-essential amino acids (Corning, MT25025CI), and 1% penicillin/streptomycin (Gibco, 15070-063). When confluent, cells were trypsinized and pelleted at 200xg and room temperature for 5 minutes. Media was aspirated, cell pellets flash frozen in liquid nitrogen, and stored at -80^{o}C until use.

For each 5′ RLM-RACE reaction, two-three pellets per fibroblast line were combined. Total RNA was then isolated from combined pellets using Qiagen RNeasy

Mini Kit, according to manufacturer's protocol. It was then treated with Turbo DNAse (Thermo) to ensure RNA purity, and clean and concentrated with Zymo Research's RNA Clean and Concentrator Kit. Isolated RNAs were stored at -80^{o}C until use. Information on fibroblast GGGGCC repeat lengths is included in **Table 5-1**.

iNeuron generation

iPSCs were split using EDTA and plated into a vitronectin-coated 6 well plate and incubated overnight in E8 media with ROCK inhibitor (Fisher BDB562822). Thirty minutes prior to transfection, cells were changed into mTESR-1 media (Cell Technologies 85850) and then transfected with 2.5 µg of donor DNA containing a dox inducible Ngn1/2 eIF1alpha-mCherry plasmid and 1.25 µg of each CLYBL targeting locus using Lipofectamine Stem (Invitrogen STEM00003) according to the manufacturer's instructions. The following morning, cells were changed into fresh E8 media. Media was changed daily, and cells were screened for red fluorescence. When the partially positive colonies reached 100-500 cells, they were carefully scraped/aspirated using a P200 pipet tip and transferred to a new vitronectin-coated dish. This process was repeated, enriching the fluorescent cells until a 100% fluorescent colony was identified. This was then relocated to a new dish, and expanded for future use. The Ngn1/2 integration cassette and accompanying targeting constructs were a gift from M. Ward (NIH).

iNeuron differentiation

Day 0. Induced pluripotent stem cells were washed in PBS and incubated in

prewarmed accutase (Sigma A6964) at 37°C for 5min. Four volumes of E8 media were added to the plate, and the cells were collected and pelleted at 200xg for 5min. The media was aspirated, and the pellet was resuspended in 1ml of fresh E8 media. Cells were counted using a hemocytometer and plated at a density of 300,000 cells per well in a 6 well dish coated with Matrigel in E8 media with ROCK inhibitor and incubated at 37°C overnight. **Day 1.** Media was changed to N2 media (1x N2 Supplement (Gibco 17502-048), 1x NEAA Supplement (Gibco 11140-050), 10 ng/ml BDNF (Peprotech 450-02), 10 ng/ml NT3 (Peprotech 450-03), 0.2 µg/ml laminin (Sigma L2020), 2 mg/ml doxycycline (Sigma D3447) in E8 media). **Day 2.** Media was changed to transition media ((1x N2 Supplement, 1x NEAA Supplement, 10 ng/ml BDNF, 10 ng/ml NT3, 0.2 µg/ml laminin, 2 mg/ml doxycycline in half E8 media, half DMEM F12 (Gibco 11320-033)). **Day 3.** Media was changed into B27 media (1x B27 Supplement (Gibco 17504-044), 1x Glutamax Supplement (Gibco 35050-061), 10 ng/ml BDNF, 10 ng/ml NT3, 0.2 µg/ml laminin, 2 mg/ml doxycycline, and 1x Culture One (Gibco A33202-01) in Neurobasal-A (Gibco 12349-015)). **Day 6.** An equal volume of B27 media without Culture One was added to each well. **Day 10.** Half conditioned media was removed and replaced with fresh B27 media without Culture One was added to each well. iNeurons were used at D10 of differentiation.

Polysome profiling

DIV10 iNeurons were treated with 100 µg/ml cycloheximide (CHX) for 5 min at 37°C. They were then harvested as previously described (39). They were transferred to ice and washed with 0.5 ml ice-cold PBS containing 100 µg/ml CHX, collected by

scraping in cold PBS + CHX, and pelleted at 234xg and 4°C for 5 min. 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, 1× EDTA-free protease inhibitor cocktail, 20 U/ml Turbo DNase, 1% Triton X-100) (40). Lysates were passed through a 20G needle 10x and incubated on ice for 5 min. Cellular debris was pelleted at 14,000xg and 4°C for 5 min, and supernatant transferred to a fresh tube. Total lysate RNA was estimated by NanoDrop. Lysates were flash-frozen in liquid nitrogen 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 (40). Prior to use, gradients were allowed to thaw and linearize overnight at 4°C. For fractionation, approximately 50-100 µg total RNA was applied to the top of the sucrose gradient. Gradients were spun at 151,263xg 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 baselined with 60% sucrose chase solution, and its sensitivity set to 0.2. For fractionation, 60% sucrose was pumped at a rate of 1.5 ml/min. Brandel's PeakChart software was used to collect profile data. 0.5 mL fractions were collected and stored at -80^{o}C, until use. RNA was this extracted using TRIzol, according to manufacturer’s protocol.

5′ RLM-RACE

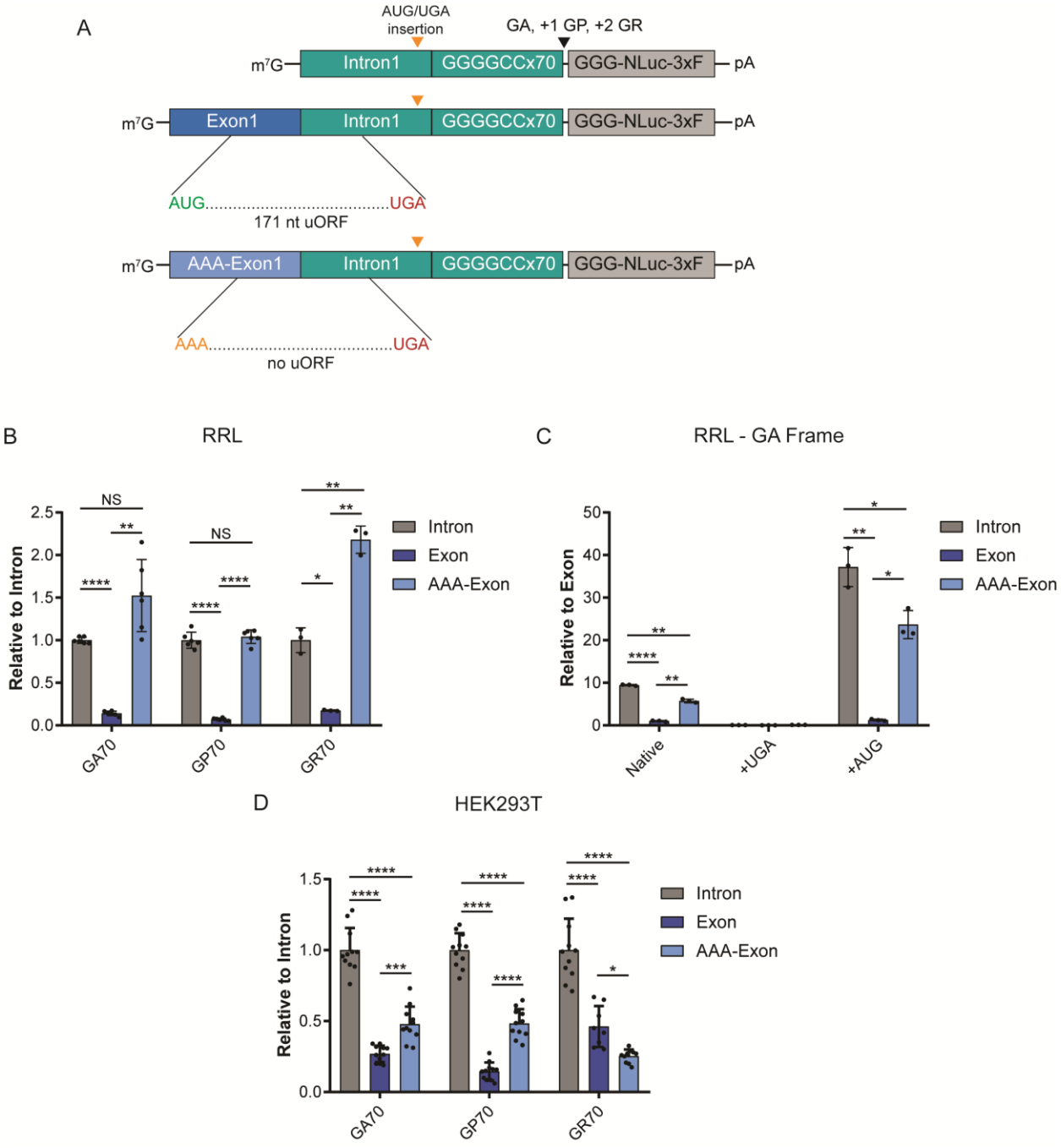
5′ RLM-RACE was performed using GeneRacer kit (Thermo Fisher Scientific), according to kit protocol. For fibroblast experiments, ~2.5-4.5 µg DNase-treated total RNA was used in the initial CIP reaction. For monosome/polysome isolated iNeuron experiments, ~0.2 µg RNA was used. cDNA synthesis was performed with SuperScript III, 50 ng random primers and 1 pmol CCCCGGx4 primer. Reaction was heated at 25^{o}C for 5 minutes prior to 60-minute incubation at 55^{o}C.

5′ RACE reactions were performed with Platinum PCR Supermix High Fidelity (Thermo), using 1 µL of cDNA and the touchdown protocol recommended in the GeneRacer Kit. Primer sequences are included in main figures. Cycling conditions for each primer: P1 and P2, annealing temperature 68^{o}C, 25 second extension; P3, annealing temperature 66^{o}C, 25 second extension.

RACE products were run on a 1% agarose gel, all bands excised, and DNA extracted with S.N.A.P. columns according to GeneRacer kit protocol (Thermo). Extracted DNA was then TOPO-cloned into pCR4-TOPO vector, and miniprepped DNA sanger sequenced with a T3 primer.

5.6 Figures

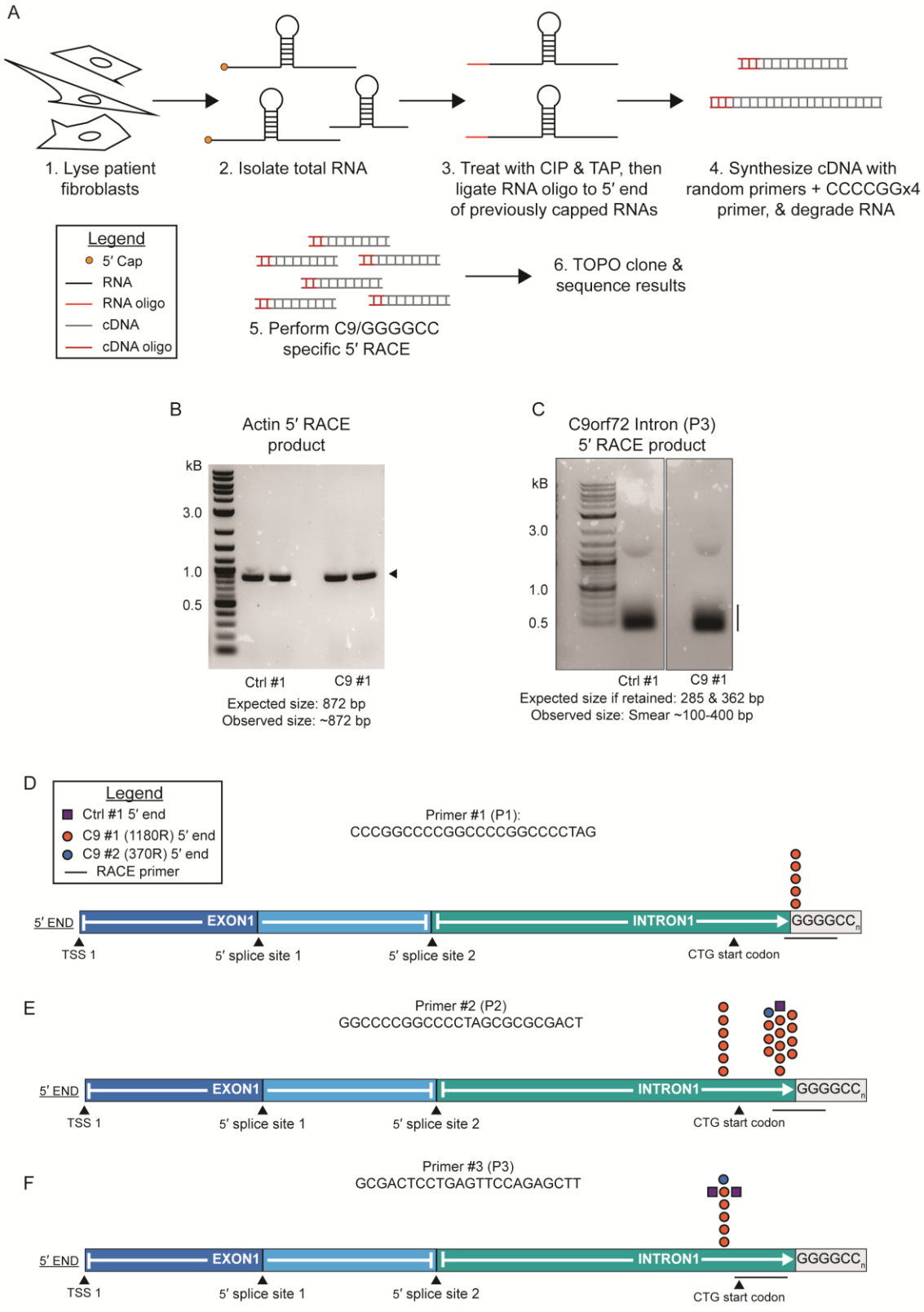
Figure 5-1. Upstream sequence context modulates C9RAN translation levels



(A) Schematic of GGGGCCx70 reporters for the GA, GP and GR reading frames, with different upstream sequence context. Spanning exon1 and intron1 is a short upstream open reading frame (uORF). For the “AAA-Exon” reporters, the AUG start codon of this

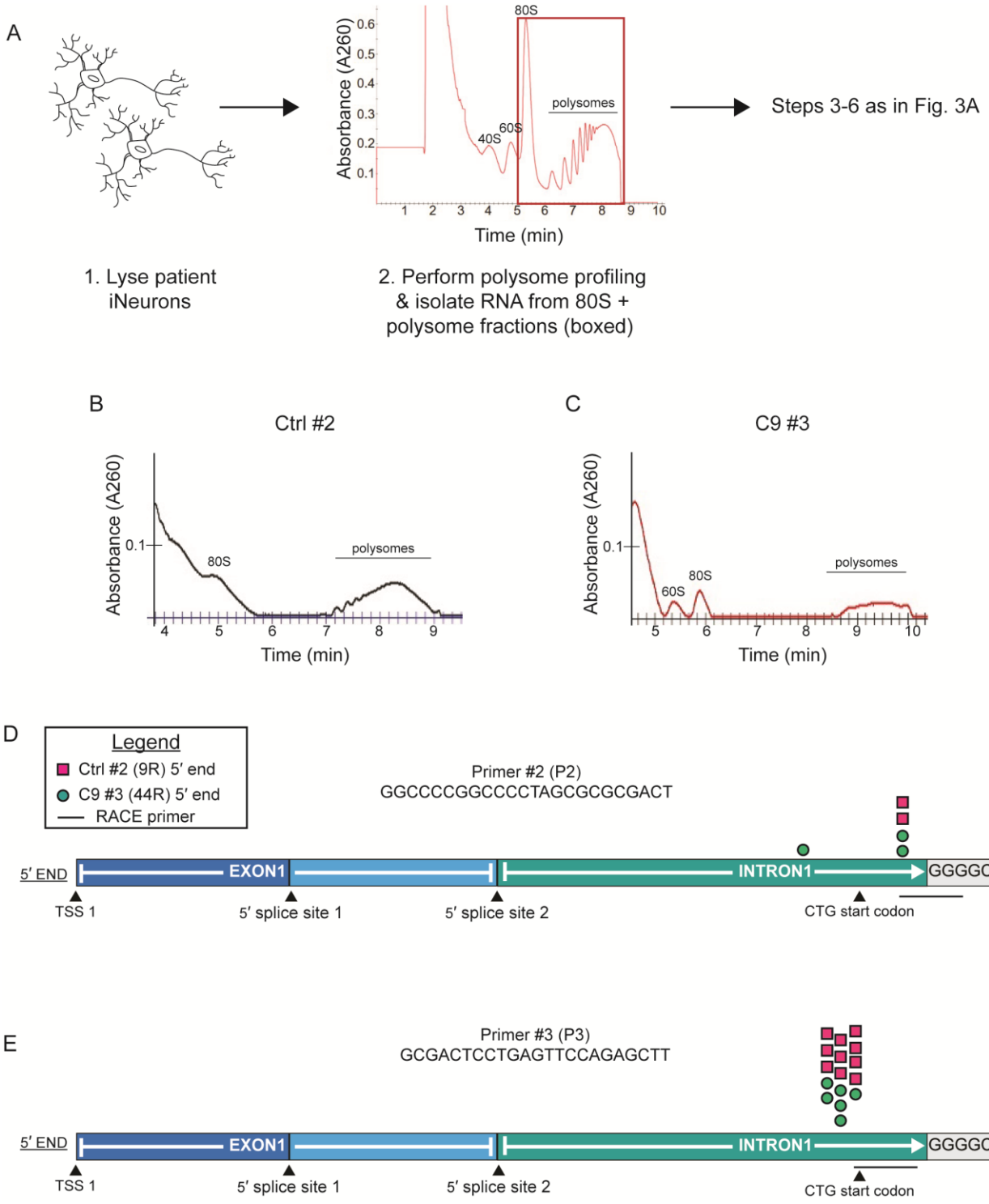
uORF is mutated to AAA. (B) Reporter NLuc signal in rabbit reticulocyte lysate (RRL), with expression of all reporters for each reading frame normalized to the “intron” reporter for that frame. n=3 or 6. (C) GA70 reporter NLuc signal in RRL, with expression of all reporters normalized to the “native exon” reporter, n=3. (D) NLuc signal 24 hours post transfection of reporter plasmids into HEK293T cells. Expression normalized to the “intron” reporter for each reading frame. n=9. Graphs represent mean and error bars +/- standard deviation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, 2-way ANOVA with Sidak’s multiple comparison test.

Figure 5-2. Repeat-primed 5′ RLM-RACE reveals novel C9orf72 5′ ends in control and patient fibroblasts



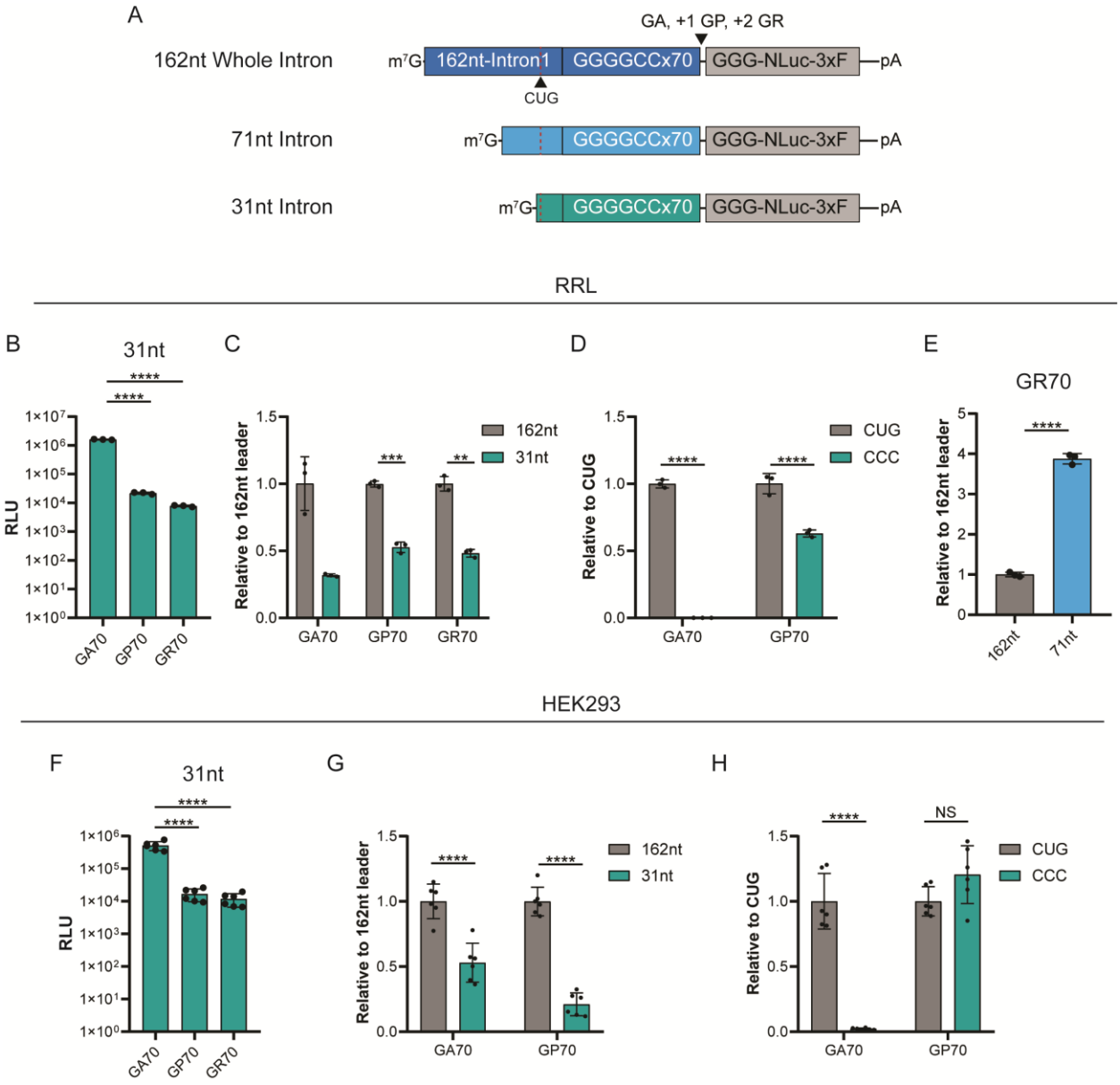
(A) Schematic of 5′ RLM-RACE experimental flow for total RNA isolated from control and C9orf72 patient fibroblasts. (B) Representative agarose gel of full-length 5′ RLM-RACE product following amplification of beta-actin in Ctrl #1 and C9 #1 lines, demonstrating success of technique. (C) Representative agarose gel of 5′ RLM-RACE product following amplification of C9orf72 intron using primer #3, on cDNA generated from Ctrl #1 and C9 #1 lines. (D-F) Schematic of C9orf72 gene structure 5′ of GGGGCC repeat, with intron and exonic sequences depicted to scale. Circle and square markers represent individual 5′ RNA ends identified following sequencing of colonies produced from RACE with C9orf72 primers 1, 2, and 3, on cDNA generated from C9 #1 and #2, and Ctrl #1 fibroblasts, respectively. Location of each primer depicted to scale by black line below complimentary C9orf72 sequence.

Figure 5-3. Repeat-primed 5′ RLM-RACE reveals novel C9orf72 5′ ends in actively translated RNA within patient iNeurons



(A) Schematic of 5′ RLM-RACE experimental flow for RNA isolated from control and C9orf72 patient-derived iNeuron 80S and polysome fractions. Polysome profile in step 2 obtained from ~300 µg total RNA isolated from HEK293 cells, to clearly show 40S, 60S, 80S and polysome fractions. (B) Polysome profile obtained from ~90 µg total RNA isolated from Ctrl #2 iNeurons, with 80S and polysome fractions labeled. (C) Polysome profile obtained from ~90 µg total RNA isolated from C9 #3 iNeurons, with 60S, 80S and polysome fractions labeled. (D-E) Schematic of C9orf72 gene structure 5′ of GGGGCC repeat, with intron and exonic sequences depicted to scale. Circle and square markers represent individual 5′ RNA ends identified following sequencing of colonies produced from RACE with C9orf72 primers 2 and 3, on cDNA isolated from C9 #3 and Ctrl #2 lines, respectively. Location of each primer depicted to scale by black line below complimentary C9orf72 sequence.

Figure 5-4. GGGGCCx70 reporters containing novel 5′ ends support RAN translation in multiple systems, and utilize the CUG start codon.



(A) Schematic of GGGGCCx70 reporters for the GA, GP and GR reading frames, with different upstream sequence context beginning at different 5′ ends identified through 5′ RLM-RACE. Intron segments are depicted to scale, as is the location of the CUG start codon within each. (B & F) Comparison of NLuc expression from the “31nt intron” GA, GP, and GR RAN reporters in rabbit reticulocyte lysate (RRL) or 24 hours post RNA

transfection in HEK293 cells, respectively. (C & G) Comparison of NLuc signal between “31nt intron” and “whole intron” reporters, following expression in RRL or 24 hours post RNA transfection in HEK293 cells, respectively. Data graphed relative to “whole intron” expression. For each n=6, from 2 independent experiments with n=3 each. (D & H) Comparison of NLuc signal between “31nt intron” reporters with CUG start codon, or with CUG codon mutated to CCC, for GA and GP reading frames. Data graphed relative to “CUG” expression. Expression in RRL or 24 hours post RNA transfection in HEK293 cells, respectively. (E) NLuc expression in RRL of GR70 reporters containing “whole intron” or “71 nt intron,” expressed relative to “whole intron.” All graphs represent mean and error bars +/- standard deviation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. For RRL, n=3; for HEK293, n=6. (B and F) Dunnett’s multiple comparison test. (E) Student’s t-test. (C, D, G, and H) 2-way ANOVA with Sidak’s multiple comparison test.

Tables

Table 5-1. Patient cell lines information and sequencing results

Cell Line	GGGGCC Repeat Length	Total P2 + P3 clones sequenced	P2 + P3 Clones containing C9orf72 sequence	% C9orf72+ Clones
Control #1	NA	45	3	6.7%
C9orf72 #1	1180	54	22	40.7%
C9orf72 #2	370	15	2	13.3%
Control #2	9	40	12	30%
C9orf72 #3	44	36	10	27.8%

Table 5-2. NLuc reporter mRNA sequences

Reporter Name	**DNA sequence to 3′ PspOMI cut site, excluding addition of poly-A tail**
Intron-GA70-PSP-GGG-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCGTGTGTGTTTTTGTTTTTCCC ACCCTCTCTCCCCACTACTTGCTCTCACAGTACTCGCTGAGG GTGAACAAGAAAAGACCTGATAAAGATTAACCAGAAGAAAACA AGGAGGGAAACAACCGCAGCCTGTAGCAAGCTCTGGAACTCA GGAGTCGCGCGCTAGCGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCAGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGTCGTGGAAGGGTGGGCGCGCCCACCGG TCTCGAGGTCCTCTTCCAGGGACCCGATGGGGTCTTCACACT CGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCTACAA CCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTT TCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGT CCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATGTCAT CATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCCAGAT CGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCATCAC TTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGACGGG GTTACGCCGAACATGATCGACTATTTCGGACGGCCCTATGAA GGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGG ACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTGATC

	AACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGA GTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGACTA CAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTAC AAGGATGACGATGACAAGTAAGGCCGCGACTCGAGAG
Exon-GR70-PSP-GGG-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCACGTAACCTACGGTGTCCCG CTAGGAAAGAGAGGTGCGTCAAACAGCGACAAGTTCCGCCCA CGTAAAAGATGACGCTTGGTGTGTCAGCCGTCCCTGCTGCCC GGTTGCTTCTCTTTTGGGGGCGGGGTCTAGCAAGAGCAGGTG TGGGTTTAGGAGGTGTGTGTTTTTGTTTTTCCCACCCTCTCTC CCCACTACTTGCTCTCACAGTACTCGCTGAGGGTGAACAAGA AAAGACCTGATAAAGATTAACCAGAAGAAAACAAGGAGGGAA ACAACCGCAGCCTGTAGCAAGCTCTGGAACTCAGGAGTCGCG CGCTAGCGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCAGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGTCGTGGAAGGGTGGGCGCGCCCACCGGTCCCTCGAG GTCCTCTTCCAGGGACCCGATGGGGTCTTCACACTCGAAGAT TTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGA CCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAAT CTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGC GGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCG TATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAA ATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGG TGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGC CGAACATGATCGACTATTTCGGACGGCCCTATGAAGGCATCG CCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGT GGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCG ACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCG GCTGGCGGCTGTGCGAACGCATTCTGGCGGACTACAAAGAC CATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATG ACGATGACAAGTAAGGCCGCGACTCGAGAG
AAA-Exon-GR70-PSP-GGG-NLuc-3xF	GGGAGACCCAAGCTGGCTAGCACGTAACCTACGGTGTCCCG CTAGGAAAGAGAGGTGCGTCAAACAGCGACAAGTTCCGCCCA CGTAAAAGAAAACGCTTGGTGTGTCAGCCGTCCCTGCTGCCC GGTTGCTTCTCTTTTGGGGGCGGGGTCTAGCAAGAGCAGGTG TGGGTTTAGGAGGTGTGTGTTTTTGTTTTTCCCACCCTCTCTC CCCACTACTTGCTCTCACAGTACTCGCTGAGGGTGAACAAGA AAAGACCTGATAAAGATTAACCAGAAGAAAACAAGGAGGGAA

	ACAACCGCAGCCTGTAGCAAGCTCTGGAACTCAGGAGTCGCG CGCTAGCGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCAGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCG GGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGG CCGGTCGTGGAAGGGTGGGCGCGCCCACCGGTCCCTCGAG GTCCTCTTCCAGGGACCCGATGGGGTCTTCACACTCGAAGAT TTCGTTGGGGACTGGCGACAGACAGCCGGCTACAACCTGGA CCAAGTCCTTGAACAGGGAGGTGTGTCCAGTTTGTTTCAGAAT CTCGGGGTGTCCGTAACTCCGATCCAAAGGATTGTCCTGAGC GGTGAAAATGGGCTGAAGATCGACATCCATGTCATCATCCCG TATGAAGGTCTGAGCGGCGACCAAATGGGCCAGATCGAAAAA ATTTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGG TGATCCTGCACTATGGCACACTGGTAATCGACGGGGTTACGC CGAACATGATCGACTATTTCGGACGGCCCTATGAAGGCATCG CCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGT GGAACGGCAACAAAATTATCGACGAGCGCCTGATCAACCCCG ACGGCTCCCTGCTGTTCCGAGTAACCATCAACGGAGTGACCG GCTGGCGGCTGTGCGAACGCATTCTGGCGGACTACAAAGAC CATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATG ACGATGACAAGTAAGGCCGCGACTCGAGAG
Whole Intron-GA70-GGG-NLuc-3xF	GGTGTGTGTTTTTGTTTTTCCCACCCTCTCTCCCCACTACTTG CTCTCACAGTACTCGCTGAGGGTGAACAAGAAAAGACCTGAT AAAGATTAACCAGAAGAAAACAAGGAGGGAAACAACCGCAGC CTGTAGCAAGCTCTGGAACTCAGGAGTCGCGCGCTAGCGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC AGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGTGGGGT CTTCACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGC CGGCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTC CAGTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAA

	AGGATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATC CATGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATG GGCCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGAT GATCATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAA TCGACGGGGTTACGCCGAACATGATCGACTATTTCGGACGGC CCTATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTG TAACAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGC GCCTGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCA TCAACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTG GCGGACTACAAAGACCATGACGGTGATTATAAAGATCATGAC ATCGATTACAAGGATGACGATGACAAGTAAGGCCGCGACTCG AGAG
71 nt Intron-GR70-GGG-NLuc-3xF	GAACCAGAAGAAAACAAGGAGGGAAACAACCGCAGCCTGTAG CAAGCTCTGGAACTCAGGAGTCGCGCGCTAGCGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCAGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGTTTGGGGTCTTCA CACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCGGCT ACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCAGTT TGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAGGA TTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCATG TCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGGCC AGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGATCA TCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCGAC GGGGTTACGCCGAACATGATCGACTATTTCGGACGGCCCTAT GAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACA GGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCCTG ATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCAAC GGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCGGA CTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGAT TACAAGGATGACGATGACAAGTAAGGCCGCGACTCGAGAG
31 nt Intron-GA70-GGG-NLuc-3xF	GAGCAAGCTCTGGAACTCAGGAGTCGCGCGCTAGCGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCAGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG

	GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGTGGGGTCTT CACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCG GCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCA GTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAG GATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCA TGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGG CCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGAT CATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCG ACGGGGTTACGCCGAACATGATCGACTATTTCGGACGGCCCT ATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAA CAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCC TGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCA ACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCG GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCG ATTACAAGGATGACGATGACAAGTAAGGCCGCGACTCGAGAG
31 nt CCC Intron-GA70-GGG-NLuc-3xF	GAGCAAGCTCCCGAACTCAGGAGTCGCGCGCTAGCGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCAGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGC CGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGG GGCCGGGGCCGGGGCCGGGGCCGGGGCCGGTGGGGTCTT CACACTCGAAGATTTCGTTGGGGACTGGCGACAGACAGCCG GCTACAACCTGGACCAAGTCCTTGAACAGGGAGGTGTGTCCA GTTTGTTTCAGAATCTCGGGGTGTCCGTAACTCCGATCCAAAG GATTGTCCTGAGCGGTGAAAATGGGCTGAAGATCGACATCCA TGTCATCATCCCGTATGAAGGTCTGAGCGGCGACCAAATGGG CCAGATCGAAAAAATTTTTAAGGTGGTGTACCCTGTGGATGAT CATCACTTTAAGGTGATCCTGCACTATGGCACACTGGTAATCG ACGGGGTTACGCCGAACATGATCGACTATTTCGGACGGCCCT ATGAAGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAA CAGGGACCCTGTGGAACGGCAACAAAATTATCGACGAGCGCC TGATCAACCCCGACGGCTCCCTGCTGTTCCGAGTAACCATCA ACGGAGTGACCGGCTGGCGGCTGTGCGAACGCATTCTGGCG GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCG ATTACAAGGATGACGATGACAAGTAAGGCCGCGACTCGAGAG

5.7 References

1. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256
2. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V. M., Kaivorinne, A. L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C., Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J., Traynor, B.

J., and Consortium, I. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268

3. Majounie, E., Renton, A. E., Mok, K., Dopper, E. G., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., van Swieten, J. C., Abramzon, Y., Johnson, J. O., Sendtner, M., Pamphlett, R., Orrell, R. W., Mead, S., Sidle, K. C., Houlden, H., Rohrer, J. D., Morrison, K. E., Pall, H., Talbot, K., Ansorge, O., Hernandez, D. G., Arepalli, S., Sabatelli, M., Mora, G., Corbo, M., Giannini, F., Calvo, A., Englund, E., Borghero, G., Floris, G. L., Remes, A. M., Laaksovirta, H., McCluskey, L., Trojanowski, J. Q., Van Deerlin, V. M., Schellenberg, G. D., Nalls, M. A., Drory, V. E., Lu, C. S., Yeh, T. H., Ishiura, H., Takahashi, Y., Tsuji, S., Le Ber, I., Brice, A., Drepper, C., Williams, N., Kirby, J., Shaw, P., Hardy, J., Tienari, P. J., Heutink, P., Morris, H. R., Pickering-Brown, S., Traynor, B. J., Consortium, C.-A. F., FTLD/FTLD/ALS, F. r. n. o., and Consortium, I. (2012) 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

4. Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C., and Edbauer, D. (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335-1338

5. Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., Ostrow, L. W., Rothstein, J. D., Troncoso, J. C., and Ranum, L. P. (2013) RAN proteins and RNA foci from

antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-4977

6. Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646
7. Mori, K., Arzberger, T., Grässer, F. A., Gijselinck, I., May, S., Rentzsch, K., Weng, S. M., Schludi, M. H., van der Zee, J., Cruts, M., Van Broeckhoven, C., Kremmer, E., Kretzschmar, H. A., Haass, C., and Edbauer, D. (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* **126**, 881-893
8. Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., Moens, T., Norona, F. E., Woollacott, I. O., Pietrzyk, J., Cleverley, K., Nicoll, A. J., Pickering-Brown, S., Dols, J., Cabecinha, M., Hendrich, O., Fratta, P., Fisher, E. M., Partridge, L., and Isaacs, A. M. (2014) C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192-1194
9. Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., Lin, S., Shneider, N. A., Monaghan, J., Pandey, U. B., Pasinelli, P., Ichida, J. K., and Trotti, D. (2014) 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

10. Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., Paul, J. W., Sun, S., Herdy, J. R., Bieri, G., Kramer, N. J., Gage, F. H., Van Den Bosch, L., Robberecht, W., and Gitler, A. D. (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18**, 1226-1229
11. Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., Kim, N. C., Almeida, S., Lee, K. H., Badders, N., Valentine, M., Miller, B. L., Wong, P. C., Petrucelli, L., Kim, H. J., Gao, F. B., and Taylor, J. P. (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129-133
12. Zhang, Y. J., Gendron, T. F., Ebbert, M. T. W., O'Raw, A. D., Yue, M., Jansen-West, K., Zhang, X., Prudencio, M., Chew, J., Cook, C. N., Daughrity, L. M., Tong, J., Song, Y., Pickles, S. R., Castanedes-Casey, M., Kurti, A., Rademakers, R., Oskarsson, B., Dickson, D. W., Hu, W., Gitler, A. D., Fryer, J. D., and Petrucelli, L. (2018) Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat Med* **24**, 1136-1142
13. Hao, Z., Liu, L., Tao, Z., Wang, R., Ren, H., Sun, H., Lin, Z., Zhang, Z., Mu, C., Zhou, J., and Wang, G. (2019) Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR. *Nat Commun* **10**, 2906
14. Zhang, Y. J., Jansen-West, K., Xu, Y. F., Gendron, T. F., Bieniek, K. F., Lin, W. L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., Chew, J., Belzil, V. V.,

Prudencio, M., Stankowski, J. N., Castanedes-Casey, M., Whitelaw, E., Ash, P. E., DeTure, M., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2014) Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* **128**, 505-524

15. May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grässer, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., Mann, M., Meissner, F., and Edbauer, D. (2014) C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* **128**, 485-503
16. Green, K. M., Glineburg, M. R., Kearse, M. G., Flores, B. N., Linsalata, A. E., Fedak, S. J., Goldstrohm, A. C., Barmada, S. J., and Todd, P. K. (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications* **8**, 2005
17. Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C. Z., Lin, C. C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., Désaubry, L., Gendron, T., Petrucelli, L., Martin, F., and Lagier-Tourenne, C. (2018) CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* **9**, 152
18. Westergard, T., McAvoy, K., Russell, K., Wen, X., Pang, Y., Morris, B., Pasinelli, P., Trotti, D., and Haeusler, A. (2019) Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med* **11**

19. Sonobe, Y., Ghadge, G., Masaki, K., Sendoel, A., Fuchs, E., and Roos, R. P. (2018) Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol Dis* **116**, 155-165

20. Cheng, W., Wang, S., Mestre, A. A., Fu, C., Makarem, A., Xian, F., Hayes, L. R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., Cleveland, D. W., and Sun, S. (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat Commun* **9**, 51

21. Almeida, S., Krishnan, G., Rushe, M., Gu, Y., Kankel, M. W., and Gao, F. B. (2019) Production of poly(GA) in C9ORF72 patient motor neurons derived from induced pluripotent stem cells. *Acta Neuropathol* **138**, 1099-1101

22. Niblock, M., Smith, B. N., Lee, Y. B., Sardone, V., Topp, S., Troakes, C., Al-Sarraj, S., Leblond, C. S., Dion, P. A., Rouleau, G. A., Shaw, C. E., and Gallo, J. M. (2016) Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. *Acta Neuropathol Commun* **4**, 18

23. Lian, Y., and Garner, H. R. (2005) Evidence for the regulation of alternative splicing via complementary DNA sequence repeats. *Bioinformatics* **21**, 1358-1364

24. Sznajder, Ł., Thomas, J. D., Carrell, E. M., Reid, T., McFarland, K. N., Cleary, J. D., Oliveira, R., Nutter, C. A., Bhatt, K., Sobczak, K., Ashizawa, T., Thornton, C. A., Ranum, L. P. W., and Swanson, M. S. (2018) Intron retention induced by microsatellite expansions as a disease biomarker. *Proc Natl Acad Sci U S A* **115**, 4234-4239

25. Qian, L., Vu, M. N., Carter, M., and Wilkinson, M. F. (1992) A spliced intron accumulates as a lariat in the nucleus of T cells. *Nucleic Acids Res* **20**, 5345-5350

26. Armakola, M., Higgins, M. J., Figley, M. D., Barmada, S. J., Scarborough, E. A., Diaz, Z., Fang, X., Shorter, J., Krogan, N. J., Finkbeiner, S., Farese, R. V., and Gitler, A. D. (2012) Inhibition of RNA lariat debranching enzyme suppresses TDP-43 toxicity in ALS disease models. *Nat Genet* **44**, 1302-1309

27. Talhouarne, G. J. S., and Gall, J. G. (2018) Lariat intronic RNAs in the cytoplasm of vertebrate cells. *Proc Natl Acad Sci U S A* **115**, E7970-E7977

28. Sareen, D., O'Rourke, J. G., Meera, P., Muhammad, A. K., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., Gendron, T., Petrucelli, L., Baughn, M., Ravits, J., Harms, M. B., Rigo, F., Bennett, C. F., Otis, T. S., Svendsen, C. N., and Baloh, R. H. (2013) Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med* **5**, 208ra149

29. Beilina, A., Tassone, F., Schwartz, P. H., Sahota, P., and Hagerman, P. J. (2004) Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum Mol Genet* **13**, 543-549

30. Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., and Wang, J. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195-200

31. Zhang, Z., and Dietrich, F. S. (2005) Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. *Nucleic Acids Res* **33**, 2838-2851

32. Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P., and Hayashizaki, Y. (2003) Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A* **100**, 15776-15781

33. Fotsing, S. F., Margoliash, J., Wang, C., Saini, S., Yanicky, R., Shleizer-Burko, S., Goren, A., and Gymrek, M. (2019) The impact of short tandem repeat variation on gene expression. *Nat Genet* **51**, 1652-1659

34. Datta, S., Alam, M. P., Majumdar, S. S., Mehta, A. K., Maiti, S., Wadhwa, N., and Brahmachari, V. (2011) Nucleosomal occupancy and CGG repeat expansion: a comparative analysis of triplet repeat region from mouse and human fragile X mental retardation gene 1. *Chromosome Res* **19**, 445-455

35. Wang, Y. H., Gellibolian, R., Shimizu, M., Wells, R. D., and Griffith, J. (1996) Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. *J Mol Biol* **263**, 511-516

36. Trotman, J. B., and Schoenberg, D. R. (2019) A recap of RNA recapping. *Wiley Interdiscip Rev RNA* **10**, e1504

37. Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F.,

Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Bar, N. S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O. J., Park, E., Persaud, K., Preall, J. B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L. H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S. E., Hannon, G., Giddings, M. C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., and Gingeras, T. R. (2012) Landscape of transcription in human cells. *Nature* **489**, 101-108

38. Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., and Todd, P. K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**, 314-322

39. Linsalata, A. E., He, F., Malik, A. M., Glineburg, M. R., Green, K. M., Natla, S., Flores, B. N., Krans, A., Archbold, H. C., Fedak, S. J., Barmada, S. J., and Todd, P. K. (2019) DDX3X and specific initiation factors modulate FMR1 repeat-associated non-AUG-initiated translation. *EMBO Rep* **20**, e47498

40. Simsek, D., Tiu, G. C., Flynn, R. A., Byeon, G. W., Leppek, K., Xu, A. F., Chang, H. Y., and Barna, M. (2017) The Mammalian Ribo-interactome Reveals Ribosome Functional Diversity and Heterogeneity. *Cell* **169**, 1051-1065.e1018

Chapter 6: Discussion and Future Directions

6.1 Introduction

In 2011, repeat-associated non-AUG (RAN) translation initiation was first described at expanded CTG and CAG repeats causative of myotonic dystrophy type 1 (DM1) and spinocerebellar ataxia type 8 (SCA8), respectively (1). Over the course of the last nine years, RAN translation has been described in seven additional diseases, the majority of which are neurodegenerative (2-8), and it is expected that this list will continue to grow as advances in long-read sequencing technology improve our ability to more readily detect and attribute repeat expansion mutations to human disease (9) (**Table 6-1**).

For each of these diseases, our understanding of how the proteins generated through RAN translation contribute to pathogenesis is still incomplete, and an area of research that warrants continued attention. However, perhaps the most extensive progress along these lines has been made with RAN translation in C9orf72-associated amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), that occurs at a GGGGCC/CCCCGG hexanucleotide repeat expansion mutation (HRE) (2,4,10-13).

Early disease models, in which dipeptide repeat-containing proteins (DPRs) were overexpressed absent the repeat-RNA, convincingly established that several C9RAN proteins synthesized from the GGGGCC/CCCCGG repeat, namely the poly-glycine-

arginine (GR), poly-glycine-proline (PR), and poly-glycine-alanine (GA), are toxic and likely contributors to C9ALS/FTD pathogenesis (14-22). However, early evidence also supported possible C9orf72 loss-of-function (LOF) (12,13) and repeat-RNA gain-of-function (GOF) mechanisms of toxicity (13). Additionally, areas of highest DPR burden in patients' brains do not correlate well with the areas of greatest neurodegeneration (23-26), and multiple *C9orf72* HRE BAC-transgenic mouse models support neuronal RAN translation, but two of four lack any neurological phenotypes (27-30), raising questions about the impact of RAN translation in disease.

However, additional C9ALS/FTD models developed in the last five years further support a role of DPRs in disease. First, new mouse models allow for expression of polyGA, polyGR or polyPR DPRs absent the repeat RNA (31-35). In all these mice, overexpression of the individual DPRs is sufficient to cause motor deficits and neurodegeneration (31-35). Additionally, new antibody-based therapies that increase DPR turnover in mouse brains, reduce repeat-associated motor deficits and neurodegeneration, including in a BAC line with unchanged levels of repeat RNA expression (36,37).

Lastly, a rapidly emerging area of research suggests that *C9orf72* haploinsufficiency, due to reduced gene expression from the HRE-containing allele (12,13), enhances DPR GOF toxicity in disease models. Recent research has established that the C9orf72 protein functions in protein homeostasis by promoting autophagosome formation (38-43). While *C9orf72* knockout (KO) mice do not have neurodegenerative phenotypes (27,44-46), heterozygous or homozygous deletion of mouse C9orf72 worsens motor deficits of *C9orf72* BAC mice, in a dose-dependent

manner (47). Furthermore, patient-derived and C9orf72 KO motor neurons (iMNs) die more rapidly than control motor neurons upon polyGR or polyPR overexpression in the absence of the repeat RNA (48).

Consequently, in current models, DPR toxicity remains a central driver of C9ALS/FTD disease pathogenesis. For this reason, understanding the mechanism of C9RAN translation is crucial to developing strategies for inhibiting toxic DPR production in patients, as this has the potential to serve as a fruitful avenue for therapy development.

6.2 Contribution of this dissertation to the broader field

The work presented in this dissertation advances the field's understanding of the mechanism of C9RAN translation in several important ways that are discussed in-depth below.

C9RAN cap and scanning-dependency, and near-AUG start codon use

As the first group to develop a luciferase-based reporter system to quantitatively measure levels of RAN translation in each reading frame of the HRE, we were able to describe several key features of C9RAN translation (49).

First, C9RAN translation occurs most efficiently in the GA reading frame, followed by the glycine-proline (GP) and GR reading frames (49). This expression pattern is seen across multiple systems, from *in vitro* lysates to primary rodent neurons, and has be replicated in subsequent reporter-based experiments by other groups (49-53), suggesting that the greater abundance of polyGA+ inclusions in patient neurons,

compared to polyGP+ and ployGR+ inclusions (24), is at least in part due to polyGA's greater rate of synthesis.

Second, a CUG codon positioned 24 nucleotides upstream of the GGGGCC repeat, in good Kozak sequence context in the GA reading frame, is important for polyGA production, as mutating it to CCC significantly reduces polyGA synthesis in multiple models (49). This finding has also been replicated in other reporter-based experiments by other groups (50,53), and a recent report showed that CRISPR-mediated deletion of a region in *C9orf72* containing the CUG codon greatly reduces polyGA production in patient-derived iNeurons (54). Thus, this dissertation work has established a sequence that can be targeted to reduce polyGA production in patient cells (50,54).

Third, C9RAN translation occurs most efficiently through a canonical scanning-model of translation initiation, that begins with 40S ribosome recruitment to a functional 5′ m^7G cap, and utilizes the helicase eIF4A to promote ribosomal scanning in a 5′ to 3′ direction along the RNA in search of start codons (49). This same base finding was also replicated in subsequent reporter-based experiments by other groups (50,51,53). However, two of these reports showed that C9RAN translation can also occur, albeit at lower rates, when the HRE is placed within the second cistron of a bicistronic reporter (51,53). This finding has been used to propose an IRES-based mechanism of C9RAN translation initiation (51,53). However, unlike bonafide IRESs, C9RAN translation in all reading frames occurs much less efficiently from RNA reporters with non-functional caps, compared to reporters with m^7G caps (49-51). Regardless, that some level of reporter-based C9RAN translation may occur through a cap-independent mechanism

suggests that this event, along with cap-dependent C9RAN translation, could be important for DPR production in patients (49-51,53).

Additionally, in this dissertation, we have replicated another report that an upstream open reading frame in exon1a inhibits C9RAN translation (50). That an upstream sequence can inhibit downstream translation initiation at the repeat is further evidence that C9RAN translation reporters utilize 5′ to 3′ ribosomal scanning to identify start codons.

Together, the reporter-based work presented in this dissertation, in consideration of subsequent reporter-based findings, suggests that if there is a heterogenous pool of C9orf72 transcripts in patient neurons, generated through use of multiple transcription start sites (TSSs) (55) or abortive transcription through the repeat (56), then species that lack exon1a, contain a function 5′ cap, and have the CUG start codon upstream of the repeat, would contribute the most to DPR production. These species would therefore be the first priority to target in therapeutic development. However, other transcripts and initiation mechanisms likely also contribute to RAN translation in patients and could be relevant to patient health. This is especially true when considering that these diseases take decades to manifest symptoms.

C9RAN translation and the integrative stress response

The work done through this dissertation was the first of four reports to show that C9RAN translation levels increase following induction of the integrative stress response (ISR) (49,51-53). We showed that this increase is a direct result of eIF2α phosphorylation at serine 51, and that it depends on the non-AUG initiation event used

by C9RAN translation (49). Other studies indicate that cap-independent forms of C9RAN translation are also increased by the ISR, though the dependence on the non-AUG initiation event in this context has not been directly assessed (51,53).

As eIF2α phosphorylation inhibits global cellular translation, this work established a clear way in which C9RAN translation is mechanistically distinct from canonical translation and thus may provide the groundwork for therapies that selectively reduce C9RAN translation without affecting global translation.

Along these lines, within this dissertation, I investigated the ability of factors that can function in place of eIF2 and deliver initiator tRNA to ribosomes in a stress-resistant manner, to support C9RAN translation. Through this work, I found that under certain conditions, eIF2A and DENR can selectively support RAN translation from reporters (**Table 6-2**). However, these factors do not appear to be important specifically for C9RAN translation during cellular stress, as initially hypothesized. In the case of DENR, this work has identified a novel factor that is more important for C9RAN translation than canonical translation, providing additional mechanistic insight into how RAN translation initiation occurs.

Small molecule inhibition of RAN translation

The high-throughput small molecule screen performed through this dissertation is the first unbiased attempt to identify small molecules that selectively inhibit RAN translation, while sparing canonical translation, through any potential mechanism (57). Through this screen, we identified five novel small molecules that inhibit RAN translation *in vitro* and can be used as tools in future mechanistic studies. Additionally, because of

the screen's unique design using RAN translation as its readout, we identified two classes of RAN-selective inhibitors; those that bind repeat RNAs and those that do not appear to with any biologically relevant affinity. This has thus provided evidence that, in future attempts to design small molecule inhibitors, multiple strategies can be used to selectively target RAN translation, beyond just repeat binding.

Furthermore, the same five small molecules inhibit RAN translation across multiple reading frames of multiple disease-causing repeats, establishing that future therapies for one repeat-expansion disease may be applicable to other repeat-expansion diseases that also undergo RAN translation (57).

Identification of novel C9orf72 5′ ends

Lastly, work in this dissertation provides evidence that the HRE promotes production of functionally-capped C9orf72 transcripts with novel, shortened 5′ ends located near the repeat sequence. These novel RNAs are found in patient iNeuron monosomes/polysomes, and reporters based on several of the newly identified sequences support C9RAN translation, suggesting that they contribute to DPR production in patient neurons.

The presence of these species is of importance to the C9ALS/FTD field. First, it suggests a mechanism by which the intronic repeat becomes accessible to ribosomes for RAN translation. Additionally, it will inform ongoing efforts in developing C9orf72-targeting ASOs and siRNAs that selectively and effectively reduce C9RAN translation in patient cells. Further work establishing how these RNA species are generated, through

either altered transcription initiation or RNA re-capping, may also reveal new C9RAN translation targets.

6.3 Other recent advance, from across the field, in inhibiting C9RAN translation

Beyond the data presented in this dissertation, work by other groups has made significant advances to our understanding of C9RAN translation. Several C9RAN translation modifiers have been identified in diverse model systems using both genetic screens and candidate-based approaches (**Table 6-2**). Additionally, several small molecule inhibitors of C9RAN translation have been identified.

The first of the genetic modifier screens was conducted in *Drosophila*, using a reporter with 114nts of C9orf72 above ~40 GGGGCC repeats, tagged in the GR reading frame with a no-AUG GFP (58). When expressed in the fly eye, the repeat underwent RAN translation and resulted in loss of retinal tissue (58). This fly was then crossed to 48 RNAi and LOF lines targeting canonical translation factors, several of which were found to reduce polyGR-GFP levels and lessen eye degeneration (58). Among the most robust rescuers were eIF4B and eIF4H, which promote eIF4A-mediated ribosomal scanning (58). Interestingly, a recent *Drosophila* RNAi screen performed in our lab, for inhibitors of CGG RAN translation, also identified eIF4B and eIF4H as suppressors (61). However, manipulation of these factors in HEK293 cells affected canonical translation to a similar degree as CGG RAN translation (61). Regardless, these findings support those reported within this dissertation that reducing ribosomal scanning reduces C9RAN translation (49).

A subsequent screen was conducted by transforming yeast with a galactose-inducible vector expressing 66 GGGGCC repeats downstream of intronic C9orf72 sequence, with no c-terminal tag (59). C9RAN levels were monitored through a GP ELISA following repeat expression in 275 yeast lines with mutations in translation-related genes (59). Again, many modifiers were identified, with the RPS25A mutant being among the most robust suppressors of polyGP expression (59). Subsequent experiments established that RPS25 KD also reduces C9RAN translation in a GGGGCCx36 fly model, as well as in C9ALS/FTD patient-derived iPSCs and iMNs, which improved iMN survival and provided further evidence of the role of DPRs in disease pathogenesis (59). RPS25 is a non-essential translation factor involved in non-canonical IRES-mediated translation and ribosomal shunting (59). The mechanism by which it regulates C9RAN translation in patient cells was not investigated but could be instrumental in understanding whether cap-independent RAN translation is important in disease.

Next, a genome-wide CRISPR-Cas9 screen was performed on RPE-1 cells stably expressing a GGGGCCx70 reporter with upstream C9orf72 intronic sequence, fused to GFP in the GA reading frame (60). Flow cytometry was used to measure polyGA-GFP expression in CRISPR KO cells (60). DDX3X deletion was found to enhance polyGA expression, which was replicated in multiple cell lines, *Drosophila*, and patient iPSCs and iNeurons (60). DDX3X is a helicase that was shown to bind the GGGGCC repeat, and its helicase activity was necessary for suppressing C9RAN translation (60). This led to a model whereby the stable secondary structure of the HRE

is necessary to support C9RAN translation, and enhancing its unwinding inhibits DPR production (60).

However, a candidate-based *Drosophila* screen from our group identified DDX3X KD as a suppressor of CGG RAN translation (61). As previous work from our group, including work within this dissertation, suggests mechanistic similarities between C9 and CGG RAN translation (49,62), it is not immediately clear how the same helicase has opposing effects on RAN translation of these two repeats, but it could be related to difference in the secondary structure of the surrounding sequence (61).

Lastly, as a candidate-based approach, eIF2A was identified as being important for polyGA production (53). Using a NLuc reporter with 75 GGGGCC repeats embedded within the second cistron of a bicistronic construct, deletion of eIF2A in HEK293 cells and eIF2A KD in chick embryo spinal cord, significantly reduced polyGA expression (53). These results are consistent with ours showing eIF2A KO lysates allow less polyGA production than WT lysates under basal conditions and support the potential of C9RAN translation's near-AUG start codon usage to serve as a molecular target. However, as eIF2A KD in HEK293 cells in our hands had little to no effect on RAN translation, further work is needed to assess the value of this target in patient cells.

Other's work related to small molecule inhibitors of C9RAN translation has focused on GGGGCC repeat binding. Consistent with results from our unbiased small molecule screen, this strategy has proven effective. The first of these studies took advantage of the fact that the hairpin formed by the GGGGCC repeat is structurally similar to the hairpin formed by the CGG repeat (63). Thus, 132 compounds structurally similar to compound 1a, which binds the CGG repeat hairpin, were screened for binding

to the GGGGCC repeat (63,64). This led to the identification of compounds 1a, 2, and 3. Each compound binds the repeat in cells, and compound 1a reduces C9RAN in patient iNeurons (63). A subsequent experiment used displacement of 1a from the GGGGCC repeat RNA to identify four additional compounds that bound the hairpin structure with even greater affinity (65). All four small molecules inhibited RAN translation from reporters, and compound 4 prevented the reporter repeat RNA from associating with polysomes (65). These studies provide strong evidence that small molecule binding to the hairpin structure of the GGGGCC repeat RNA is effective at inhibiting C9RAN translation. However, another recent screen of 138 g-quadruplex binding compounds found two (DB1246 and DB1273) that reduced polyGP levels in patient iNeurons (66). Therefore, while repeat-RNA binding in cells can reduce C9RAN translation, it is unclear if targeting the hairpin or the g-quadruplex structure is more effective, and therefore if these different RNA structures play a role in regulating C9RAN translation levels in patients.

6.4 Remaining questions and future directions

While this dissertation has provided many insights into the mechanism of C9RAN translation, many more questions remain. These questions, outlined below, highlight areas worthy of future research attention.

Where does RAN translation initiate in the polyGP and polyGR reading frames of the GGGGCC repeat?

Knowing how and where initiation for the different C9RAN DPRs occurs is important for identifying specific sequences or processes that can be targeted to reduce DPR production in patients.

My initial finding that a CUG codon is important for RAN translation in the polyGA reading frame (49) has since been replicated by three groups (50,53,54). However, it remains less clear where initiation in the polyGP and polyGR frames occurs. While mutating the CUG codon also decreases polyGR expression, in our hands, this inhibition is less complete than in the GA frame (49). Additionally, through this dissertation work and another report, loss of the CUG codon had varying effects on C9RAN translation in the GP reading frame depending on the model system in which the reporters were (49,50).

Furthermore, while CRISPR-mediated deletion of an 86bp region upstream of the repeat that contains the CUG codon nearly entirely prevents polyGA production in C9ALS/FTD iNeurons, it does not decrease polyGP or polyGR levels (54). However, this large deletion shifts the reading frame of the sequence upstream of the repeat, which places new near-AUG codons in the GP and GR frames that could conceivable be used for initiation.

Therefore, it remains a possibility that the same CUG codon is used for initiation in multiple reading frames, and the polyGP and/or polyGR products result from ribosome frameshifting during translation elongation (49,50). The possibility of frameshifting in patient cells is supported by a recent study showing that when polyGA peptides are pulled-down from C9ALS/FTD patient brain lysates, they are detected by a polyGP antibody, suggesting that both DPRs exist in a single chimeric species (67). More

targeted CRISPR approaches in C9ALS/FTD patient cells, where only the CUG codon is mutated or deleted, could further shed light on the importance of this codon for all three sense reading frames. If the CUG codon is not used for the polyGP and polyGR frames, then the question remains as to where initiation in these frames occurs. Is it within the repeat itself, as the stop codon immediately upstream of the repeat in the GP frame would require for polyGP production in the absence of frameshifting? Are there other specific non-AUG codons used? If so, is the same non-AUG codon used each time, or is initiation site selection promiscuous?

How is C9RAN translation increased during the ISR?

In addition to this dissertation work, multiple groups have observed that C9RAN translation is increased following ISR induction and eIF2α phosphorylation (49,51-53). As this aspect of C9RAN translation represents a way in which its behavior diverges from global canonical translation, understanding the mechanism through which it occurs could yield new strategies for selective targeting.

The data generated through this dissertation indicate that stress-resistant C9RAN translation depends on its use of a non-AUG start codon (49). This led us to investigate the role of eIF2A, eIF2D, and DENR/MCTS1 in promoting C9RAN translation following eIF2α phosphorylation, as these factors can deliver initiator methionine and non-methionine tRNAs to the ribosomes to promote non-AUG initiation, and are not regulated by the ISR (68-70).

However, knockdown of none of these factors prevented increased C9RAN translation during cellular stress. Therefore, future studies are needed to determine if

other targetable factors are involved in this process. Possibilities include other non-essential translation factors that can promote initiator tRNA delivery, factors that enhance start-codon stringency, or factors that destabilize the repeat-RNA secondary structure if ribosome queuing proves important in this phenomenon.

What is the mechanism by which small molecule inhibitors impair RAN translation?

Three of the five small molecules identified through the small molecule screen in Chapter 3 show evidence of interacting with repeat RNAs, while two do not (57). However, unexpectedly, the repeat-interacting compounds, but not the non-RNA-interactors, also selectively inhibit *in vitro* translation of near-AUG reporters that lack any repeat element (57). Therefore, the mechanism by which either class of small molecules more significantly inhibits RAN translation than canonical translation is unclear. Some possibilities are that while the RNA-interactors can bind the repeat RNA, they also interact with other GC-rich RNAs, perhaps in the ribosome, and it is through these interactions that RAN translation is selectively inhibited. Additionally, the non-RNA-interactors may inhibit a non-RNA based factor important for RAN translation.

Experiments in which chemical modifications are made to the small molecule inhibitors, to permit proximity labeling or pulldown (63,71), could identify their targets in *in vitro* lysates, which would reveal important insights into the mechanism of RAN translation and how to selectively inhibit it.

How does C9RAN translation occur in patient neurons and how are the novel C9orf72 5' ends generated?

In Chapter 5 of this dissertation, using 5′ RNA ligation-mediated rapid amplification of cDNA ends, I provided evidence that the HRE promotes generation of novel C9orf72 5′ ends located just upstream of the repeat. While these are important initial findings, many questions remain for future experiments. First among these is whether the novel 5′ ends are generated by altered transcription initiation or RNA re-capping, which can be addressed by assessing RNA Pol II occupancy at *C9orf72* in patient and control cells, as well as other chromatin features associated with active transcription. Additionally, 3′ RACE experiments are needed to determine what sequence is present downstream of the repeat, as this will provide additional insight into how the short C9orf72 transcripts are generated, and how they can be targeted. Lastly, work is needed to determine if other uncapped HRE-containing RNA species, and potentially repeat-containing lariat RNAs, that the 5′ RLM-RACE technique cannot detect, are also present in patient monosome/polysome fractions.

Answers to each of the questions outlined in this section will inform the field's long-term goal of developing therapies for inhibiting RAN translation, not only in C9ALS/FTD, but across the ever-growing list of human diseases that support this non-canonical form of translation initiation.

6.5 Table

Table 6-1: RAN-translated repeats in human diseases

Disease	Repeat	RAN peptides detected in patient tissue
Myotonic dystrophy type 1	CTG·CAG	AS: poly-glutamine (1)
Spinocerebellar ataxia type 8	CAG	S: poly-alanine (1)
Fragile X-associated tremor ataxia syndrome	CGG·CCG	S: poly-glycine (3) S: poly-alanine (72) AS: poly-proline (73) AS: poly-alanine (73)
Amyotrophic lateral sclerosis and frontotemporal dementia	GGGGCC·CCCCGG	S: poly-glycine-alanine (2) S: poly-glycine-proline (2,4) S: poly-glycine-arginine (2) AS: poly-proline-alanine (10,11) AS: poly-proline-arginine (10,11) AS: poly-glycine-proline (10,11)
Myotonic dystrophy type 2	CCTG·CAGG	S: poly-leucine-proline-alanine-cysteine (8) AS: poly-glutamine-alanine-glycine-arginine (8)
Huntington disease	CAG·CTG	S: Poly-alanine (6) S: poly-serine (6) AS: poly-leucine (6) AS: poly-cysteine (6)
Fragile x-associated primary ovarian insufficiency	CGG·CCG	S: poly-glycine (5)
Fuch's endothelial corneal dystrophy	CTG·CAG	S: poly-cysteine (7)

S indicates products generated from the sense strand of the expanded repeat, AS from the antisense strand

Table 6-2: RAN translation modifiers

Gene	Methods of C9RAN detection	Model systems effective in
RPS25	GP ELISA,GR ELISA western blot, immunocytochemistry	Yeast (59), HAP1 cells (59), *Drosophila* (59), patient iPSCs (59), patient iMNs (59)
eIF4B	polyGR-GFP fluorescence	*Drosophila* (58)
eIF4H	polyGR-GFP fluorescence	*Drosophila* (58)
DDX3X	polyGA-GFP fluorescence, nanoluciferase expression, western blot, GP ELISA, GR ELISA	RPE-1 cells (60), HeLa cells (60), *Drosophila* (60) patient iPSCs (60), patient iNeurons (60)
eIF2A	Nanoluciferase expression	HEK293 cells (53), chick embryo (53), *in vitro* lysates*
DENR	Nanoluciferase expression, western blot	HEK293 cells*

*From work described in this dissertation

6.6 References

1. Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A., Nan, Z., Forster, C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., Swanson, M. S., Moseley, M., and Ranum, L. P. (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* **108**, 260-265

2. Mori, K., Weng, S. M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C., and Edbauer, D. (2013) The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* **339**, 1335-1338

3. Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K. C., Scaglione, K. M., Basrur, V., Elenitoba-Johnson, K., Vonsattel, J. P.,

Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Berguerand, N., and Paulson, H. L. (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* **78**, 440-455

4. Ash, P. E., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., Dejesus-Hernandez, M., van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639-646
5. Buijsen, R. A., Visser, J. A., Kramer, P., Severijnen, E. A., Gearing, M., Charlet-Berguerand, N., Sherman, S. L., Berman, R. F., Willemsen, R., and Hukema, R. K. (2016) 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
6. Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., Pletnikova, O., Borchelt, D. R., Ross, C. A., Margolis, R. L., Yachnis, A. T., Troncoso, J. C., and Ranum, L. P. (2015) RAN Translation in Huntington Disease. *Neuron* **88**, 667-677
7. Soragni, E., Petrosyan, L., Rinkoski, T. A., Wieben, E. D., Baratz, K. H., Fautsch, M. P., and Gottesfeld, J. M. (2018) Repeat-Associated Non-ATG (RAN) Translation in Fuchs' Endothelial Corneal Dystrophy. *Invest Ophthalmol Vis Sci* **59**, 1888-1896
8. Zu, T., Cleary, J. D., Liu, Y., Bañez-Coronel, M., Bubenik, J. L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H. B., Yachnis, A. T., Swanson, M. S., and Ranum,

L. P. W. (2017) RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. *Neuron* **95**, 1292-1305.e1295

9. Mantere, T., Kersten, S., and Hoischen, A. (2019) Long-Read Sequencing Emerging in Medical Genetics. *Front Genet* **10**, 426

10. Mori, K., Arzberger, T., Grässer, F. A., Gijselinck, I., May, S., Rentzsch, K., Weng, S. M., Schludi, M. H., van der Zee, J., Cruts, M., Van Broeckhoven, C., Kremmer, E., Kretzschmar, H. A., Haass, C., and Edbauer, D. (2013) Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* **126**, 881-893

11. Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., Ostrow, L. W., Rothstein, J. D., Troncoso, J. C., and Ranum, L. P. (2013) RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A* **110**, E4968-4977

12. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., Schymick, J. C., Laaksovirta, H., van Swieten, J. C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A. M., Kaganovich, A., Scholz, S. W., Duckworth, J., Ding, J., Harmer, D. W., Hernandez, D. G., Johnson, J. O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R. J., Orrell, R. W., Neal, J., Murray, A., Pearson, J., Jansen, I. E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J. B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M. A., Peuralinna, T., Jansson, L., Isoviita, V. M.,

Kaivorinne, A. L., Hölttä-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wuu, J., Chiò, A., Restagno, G., Borghero, G., Sabatelli, M., Heckerman, D., Rogaeva, E., Zinman, L., Rothstein, J. D., Sendtner, M., Drepper, C., Eichler, E. E., Alkan, C., Abdullaev, Z., Pack, S. D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N. M., Heutink, P., Pickering-Brown, S., Morris, H. R., Tienari, P. J., Traynor, B. J., and Consortium, I. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257-268

13. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., Wszolek, Z. K., Feldman, H., Knopman, D. S., Petersen, R. C., Miller, B. L., Dickson, D. W., Boylan, K. B., Graff-Radford, N. R., and Rademakers, R. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245-256

14. Mizielinska, S., Grönke, S., Niccoli, T., Ridler, C. E., Clayton, E. L., Devoy, A., Moens, T., Norona, F. E., Woollacott, I. O., Pietrzyk, J., Cleverley, K., Nicoll, A. J., Pickering-Brown, S., Dols, J., Cabecinha, M., Hendrich, O., Fratta, P., Fisher, E. M., Partridge, L., and Isaacs, A. M. (2014) C9orf72 repeat expansions cause neurodegeneration in Drosophila through arginine-rich proteins. *Science* **345**, 1192-1194

15. Flores, B. N., Dulchavsky, M. E., Krans, A., Sawaya, M. R., Paulson, H. L., Todd, P. K., Barmada, S. J., and Ivanova, M. I. (2016) Distinct C9orf72-Associated

Dipeptide Repeat Structures Correlate with Neuronal Toxicity. *PLoS One* **11**, e0165084

16. May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grässer, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., Mann, M., Meissner, F., and Edbauer, D. (2014) C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* **128**, 485-503
17. Wen, X., Tan, W., Westergard, T., Krishnamurthy, K., Markandaiah, S. S., Shi, Y., Lin, S., Shneider, N. A., Monaghan, J., Pandey, U. B., Pasinelli, P., Ichida, J. K., and Trotti, D. (2014) 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
18. Zhang, Y. J., Jansen-West, K., Xu, Y. F., Gendron, T. F., Bieniek, K. F., Lin, W. L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., Chew, J., Belzil, V. V., Prudencio, M., Stankowski, J. N., Castanedes-Casey, M., Whitelaw, E., Ash, P. E., DeTure, M., Rademakers, R., Boylan, K. B., Dickson, D. W., and Petrucelli, L. (2014) Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* **128**, 505-524
19. Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., Paul, J. W., Sun, S., Herdy, J. R., Bieri, G., Kramer, N. J., Gage, F. H., Van Den Bosch, L., Robberecht, W., and Gitler, A. D. (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci* **18**, 1226-1229

20. Tao, Z., Wang, H., Xia, Q., Li, K., Jiang, X., Xu, G., Wang, G., and Ying, Z. (2015) Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet* **24**, 2426-2441

21. Yamakawa, M., Ito, D., Honda, T., Kubo, K., Noda, M., Nakajima, K., and Suzuki, N. (2015) Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum Mol Genet* **24**, 1630-1645

22. Yang, W. Y., Wilson, H. D., Velagapudi, S. P., and Disney, M. D. (2015) Inhibition of Non-ATG Translational Events in Cells via Covalent Small Molecules Targeting RNA. *J Am Chem Soc* **137**, 5336-5345

23. Mackenzie, I. R., Arzberger, T., Kremmer, E., Troost, D., Lorenzl, S., Mori, K., Weng, S. M., Haass, C., Kretzschmar, H. A., Edbauer, D., and Neumann, M. (2013) Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* **126**, 859-879

24. Mackenzie, I. R., Frick, P., Grässer, F. A., Gendron, T. F., Petrucelli, L., Cashman, N. R., Edbauer, D., Kremmer, E., Prudlo, J., Troost, D., and Neumann, M. (2015) Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol* **130**, 845-861

25. Davidson, Y., Robinson, A. C., Liu, X., Wu, D., Troakes, C., Rollinson, S., Masuda-Suzukake, M., Suzuki, G., Nonaka, T., Shi, J., Tian, J., Hamdalla, H., Ealing, J., Richardson, A., Jones, M., Pickering-Brown, S., Snowden, J. S., Hasegawa, M., and Mann, D. M. (2015) Neurodegeneration in Frontotemporal Lobar Degeneration and Motor Neurone Disease associated with expansions in

C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. *Neuropathol Appl Neurobiol*

26. Schludi, M. H., May, S., Grässer, F. A., Rentzsch, K., Kremmer, E., Küpper, C., Klopstock, T., Arzberger, T., Edbauer, D., Degeneration, G. C. f. F. L., and Alliance, B. B. B. (2015) Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathol* **130**, 537-555

27. Jiang, J., Zhu, Q., Gendron, T. F., Saberi, S., McAlonis-Downes, M., Seelman, A., Stauffer, J. E., Jafar-Nejad, P., Drenner, K., Schulte, D., Chun, S., Sun, S., Ling, S. C., Myers, B., Engelhardt, J., Katz, M., Baughn, M., Platoshyn, O., Marsala, M., Watt, A., Heyser, C. J., Ard, M. C., De Muynck, L., Daughrity, L. M., Swing, D. A., Tessarollo, L., Jung, C. J., Delpoux, A., Utzschneider, D. T., Hedrick, S. M., de Jong, P. J., Edbauer, D., Van Damme, P., Petrucelli, L., Shaw, C. E., Bennett, C. F., Da Cruz, S., Ravits, J., Rigo, F., Cleveland, D. W., and Lagier-Tourenne, C. (2016) Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs. *Neuron* **90**, 535-550

28. Peters, O. M., Cabrera, G. T., Tran, H., Gendron, T. F., McKeon, J. E., Metterville, J., Weiss, A., Wightman, N., Salameh, J., Kim, J., Sun, H., Boylan, K. B., Dickson, D., Kennedy, Z., Lin, Z., Zhang, Y. J., Daughrity, L., Jung, C., Gao, F. B., Sapp, P. C., Horvitz, H. R., Bosco, D. A., Brown, S. P., de Jong, P., Petrucelli, L., Mueller, C., and Brown, R. H. (2015) Human C9ORF72

Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice. *Neuron* **88**, 902-909

29. O'Rourke, J. G., Bogdanik, L., Muhammad, A. K. M. G., Gendron, T. F., Kim, K. J., Austin, A., Cady, J., Liu, E. Y., Zarrow, J., Grant, S., Ho, R., Bell, S., Carmona, S., Simpkinson, M., Lall, D., Wu, K., Daughrity, L., Dickson, D. W., Harms, M. B., Petrucelli, L., Lee, E. B., Lutz, C. M., and Baloh, R. H. (2015) C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD. *Neuron* **88**, 892-901
30. Liu, Y., Pattamatta, A., Zu, T., Reid, T., Bardhi, O., Borchelt, D. R., Yachnis, A. T., and Ranum, L. P. (2016) C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD. *Neuron* **90**, 521-534
31. Zhang, Y. J., Gendron, T. F., Ebbert, M. T. W., O'Raw, A. D., Yue, M., Jansen-West, K., Zhang, X., Prudencio, M., Chew, J., Cook, C. N., Daughrity, L. M., Tong, J., Song, Y., Pickles, S. R., Castanedes-Casey, M., Kurti, A., Rademakers, R., Oskarsson, B., Dickson, D. W., Hu, W., Gitler, A. D., Fryer, J. D., and Petrucelli, L. (2018) Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat Med* **24**, 1136-1142
32. Hao, Z., Liu, L., Tao, Z., Wang, R., Ren, H., Sun, H., Lin, Z., Zhang, Z., Mu, C., Zhou, J., and Wang, G. (2019) Motor dysfunction and neurodegeneration in a C9orf72 mouse line expressing poly-PR. *Nat Commun* **10**, 2906
33. Zhang, Y. J., Gendron, T. F., Grima, J. C., Sasaguri, H., Jansen-West, K., Xu, Y. F., Katzman, R. B., Gass, J., Murray, M. E., Shinohara, M., Lin, W. L., Garrett, A.,

Stankowski, J. N., Daughrity, L., Tong, J., Perkerson, E. A., Yue, M., Chew, J., Castanedes-Casey, M., Kurti, A., Wang, Z. S., Liesinger, A. M., Baker, J. D., Jiang, J., Lagier-Tourenne, C., Edbauer, D., Cleveland, D. W., Rademakers, R., Boylan, K. B., Bu, G., Link, C. D., Dickey, C. A., Rothstein, J. D., Dickson, D. W., Fryer, J. D., and Petrucelli, L. (2016) C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nat Neurosci* **19**, 668-677

34. Schludi, M. H., Becker, L., Garrett, L., Gendron, T. F., Zhou, Q., Schreiber, F., Popper, B., Dimou, L., Strom, T. M., Winkelmann, J., von Thaden, A., Rentzsch, K., May, S., Michaelsen, M., Schwenk, B. M., Tan, J., Schoser, B., Dieterich, M., Petrucelli, L., Hölter, S. M., Wurst, W., Fuchs, H., Gailus-Durner, V., de Angelis, M. H., Klopstock, T., Arzberger, T., and Edbauer, D. (2017) Spinal poly-GA inclusions in a C9orf72 mouse model trigger motor deficits and inflammation without neuron loss. *Acta Neuropathol* **134**, 241-254

35. Herranz-Martin, S., Chandran, J., Lewis, K., Mulcahy, P., Higginbottom, A., Walker, C., Valenzuela, I. M. Y., Jones, R. A., Coldicott, I., Iannitti, T., Akaaboune, M., El-Khamisy, S. F., Gillingwater, T. H., Shaw, P. J., and Azzouz, M. (2017) Viral delivery of. *Dis Model Mech* **10**, 859-868

36. Nguyen, L., Montrasio, F., Pattamatta, A., Tusi, S. K., Bardhi, O., Meyer, K. D., Hayes, L., Nakamura, K., Banez-Coronel, M., Coyne, A., Guo, S., Laboissonniere, L. A., Gu, Y., Narayanan, S., Smith, B., Nitsch, R. M., Kankel, M. W., Rushe, M., Rothstein, J., Zu, T., Grimm, J., and Ranum, L. P. W. (2020)

Antibody Therapy Targeting RAN Proteins Rescues C9 ALS/FTD Phenotypes in C9orf72 Mouse Model. *Neuron* **105**, 645-662.e611

37. Zhou, Q., Mareljic, N., Michaelsen, M., Parhizkar, S., Heindl, S., Nuscher, B., Farny, D., Czuppa, M., Schludi, C., Graf, A., Krebs, S., Blum, H., Feederle, R., Roth, S., Haass, C., Arzberger, T., Liesz, A., and Edbauer, D. (2020) Active poly-GA vaccination prevents microglia activation and motor deficits in a C9orf72 mouse model. *EMBO Mol Med* **12**, e10919

38. Sellier, C., Campanari, M. L., Julie Corbier, C., Gaucherot, A., Kolb-Cheynel, I., Oulad-Abdelghani, M., Ruffenach, F., Page, A., Ciura, S., Kabashi, E., and Charlet-Berguerand, N. (2016) Loss of C9ORF72 impairs autophagy and synergizes with polyQ Ataxin-2 to induce motor neuron dysfunction and cell death. *EMBO J* **35**, 1276-1297

39. Yang, M., Liang, C., Swaminathan, K., Herrlinger, S., Lai, F., Shiekhattar, R., and Chen, J. F. (2016) A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy. *Sci Adv* **2**, e1601167

40. Amick, J., Roczniak-Ferguson, A., and Ferguson, S. M. (2016) C9orf72 binds SMCR8, localizes to lysosomes, and regulates mTORC1 signaling. *Mol Biol Cell* **27**, 3040-3051

41. Boivin, M., Pfister, V., Gaucherot, A., Ruffenach, F., Negroni, L., Sellier, C., and Charlet-Berguerand, N. (2020) Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders. *EMBO J* **39**, e100574

42. Wang, M., Wang, H., Tao, Z., Xia, Q., Hao, Z., Prehn, J. H. M., Zhen, X., Wang, G., and Ying, Z. (2020) C9orf72 associates with inactive Rag GTPases and regulates mTORC1-mediated autophagosomal and lysosomal biogenesis. *Aging Cell*, e13126

43. Webster, C. P., Smith, E. F., Bauer, C. S., Moller, A., Hautbergue, G. M., Ferraiuolo, L., Myszczynska, M. A., Higginbottom, A., Walsh, M. J., Whitworth, A. J., Kaspar, B. K., Meyer, K., Shaw, P. J., Grierson, A. J., and De Vos, K. J. (2016) The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy. *EMBO J* **35**, 1656-1676

44. O'Rourke, J. G., Bogdanik, L., Yáñez, A., Lall, D., Wolf, A. J., Muhammad, A. K., Ho, R., Carmona, S., Vit, J. P., Zarrow, J., Kim, K. J., Bell, S., Harms, M. B., Miller, T. M., Dangler, C. A., Underhill, D. M., Goodridge, H. S., Lutz, C. M., and Baloh, R. H. (2016) C9orf72 is required for proper macrophage and microglial function in mice. *Science* **351**, 1324-1329

45. Burberry, A., Suzuki, N., Wang, J. Y., Moccia, R., Mordes, D. A., Stewart, M. H., Suzuki-Uematsu, S., Ghosh, S., Singh, A., Merkle, F. T., Koszka, K., Li, Q. Z., Zon, L., Rossi, D. J., Trowbridge, J. J., Notarangelo, L. D., and Eggan, K. (2016) Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease. *Sci Transl Med* **8**, 347ra393

46. Sudria-Lopez, E., Koppers, M., de Wit, M., van der Meer, C., Westeneng, H. J., Zundel, C. A., Youssef, S. A., Harkema, L., de Bruin, A., Veldink, J. H., van den Berg, L. H., and Pasterkamp, R. J. (2016) Full ablation of C9orf72 in mice causes

immune system-related pathology and neoplastic events but no motor neuron defects. *Acta Neuropathol* **132**, 145-147

47. Shao, Q., Liang, C., Chang, Q., Zhang, W., Yang, M., and Chen, J. F. (2019) C9orf72 deficiency promotes motor deficits of a C9ALS/FTD mouse model in a dose-dependent manner. *Acta Neuropathol Commun* **7**, 32

48. Shi, Y., Lin, S., Staats, K. A., Li, Y., Chang, W. H., Hung, S. T., Hendricks, E., Linares, G. R., Wang, Y., Son, E. Y., Wen, X., Kisler, K., Wilkinson, B., Menendez, L., Sugawara, T., Woolwine, P., Huang, M., Cowan, M. J., Ge, B., Koutsodendris, N., Sandor, K. P., Komberg, J., Vangoor, V. R., Senthilkumar, K., Hennes, V., Seah, C., Nelson, A. R., Cheng, T. Y., Lee, S. J., August, P. R., Chen, J. A., Wisniewski, N., Hanson-Smith, V., Belgard, T. G., Zhang, A., Coba, M., Grunseich, C., Ward, M. E., van den Berg, L. H., Pasterkamp, R. J., Trotti, D., Zlokovic, B. V., and Ichida, J. K. (2018) Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. *Nat Med* **24**, 313-325

49. Green, K. M., Glineburg, M. R., Kearse, M. G., Flores, B. N., Linsalata, A. E., Fedak, S. J., Goldstrohm, A. C., Barmada, S. J., and Todd, P. K. (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications* **8**, 2005

50. Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C. Z., Lin, C. C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., Désaubry, L., Gendron, T., Petrucelli, L., Martin, F., and Lagier-Tourenne, C. (2018) CUG initiation and

frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat Commun* **9**, 152

51. Cheng, W., Wang, S., Mestre, A. A., Fu, C., Makarem, A., Xian, F., Hayes, L. R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., Cleveland, D. W., and Sun, S. (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2α phosphorylation. *Nat Commun* **9**, 51
52. Westergard, T., McAvoy, K., Russell, K., Wen, X., Pang, Y., Morris, B., Pasinelli, P., Trotti, D., and Haeusler, A. (2019) Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol Med* **11**
53. Sonobe, Y., Ghadge, G., Masaki, K., Sendoel, A., Fuchs, E., and Roos, R. P. (2018) Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol Dis* **116**, 155-165
54. Almeida, S., Krishnan, G., Rushe, M., Gu, Y., Kankel, M. W., and Gao, F. B. (2019) Production of poly(GA) in C9ORF72 patient motor neurons derived from induced pluripotent stem cells. *Acta Neuropathol* **138**, 1099-1101
55. Sareen, D., O'Rourke, J. G., Meera, P., Muhammad, A. K., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., Gendron, T., Petrucelli, L., Baughn, M., Ravits, J., Harms, M. B., Rigo, F., Bennett, C. F., Otis, T. S., Svendsen, C. N., and Baloh, R. H. (2013) Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med* **5**, 208ra149
56. Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A., Shaw, P. G., Kim, M. S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., and

Wang, J. (2014) C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195-200

57. Green, K. M., Sheth, U. J., Flores, B. N., Wright, S. E., Sutter, A. B., Kearse, M. G., Barmada, S. J., Ivanova, M. I., and Todd, P. K. (2019) High-throughput screening yields several small-molecule inhibitors of repeat-associated non-AUG translation. *J Biol Chem* **294**, 18624-18638

58. Goodman, L. D., Prudencio, M., Srinivasan, A. R., Rifai, O. M., Lee, V. M., Petrucelli, L., and Bonini, N. M. (2019) eIF4B and eIF4H mediate GR production from expanded G4C2 in a Drosophila model for C9orf72-associated ALS. *Acta Neuropathol Commun* **7**, 62

59. Yamada, S. B., Gendron, T. F., Niccoli, T., Genuth, N. R., Grosely, R., Shi, Y., Glaria, I., Kramer, N. J., Nakayama, L., Fang, S., Dinger, T. J. I., Thoeng, A., Rocha, G., Barna, M., Puglisi, J. D., Partridge, L., Ichida, J. K., Isaacs, A. M., Petrucelli, L., and Gitler, A. D. (2019) RPS25 is required for efficient RAN translation of C9orf72 and other neurodegenerative disease-associated nucleotide repeats. *Nat Neurosci* **22**, 1383-1388

60. Cheng, W., Wang, S., Zhang, Z., Morgens, D. W., Hayes, L. R., Lee, S., Portz, B., Xie, Y., Nguyen, B. V., Haney, M. S., Yan, S., Dong, D., Coyne, A. N., Yang, J., Xian, F., Cleveland, D. W., Qiu, Z., Rothstein, J. D., Shorter, J., Gao, F. B., Bassik, M. C., and Sun, S. (2019) 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

61. Linsalata, A. E., He, F., Malik, A. M., Glineburg, M. R., Green, K. M., Natla, S., Flores, B. N., Krans, A., Archbold, H. C., Fedak, S. J., Barmada, S. J., and Todd, P. K. (2019) DDX3X and specific initiation factors modulate FMR1 repeat-associated non-AUG-initiated translation. *EMBO Rep* **20**, e47498

62. Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., and Todd, P. K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* **62**, 314-322

63. Su, Z., Zhang, Y., Gendron, T. F., Bauer, P. O., Chew, J., Yang, W. Y., Fostvedt, E., Jansen-West, K., Belzil, V. V., Desaro, P., Johnston, A., Overstreet, K., Oh, S. Y., Todd, P. K., Berry, J. D., Cudkowicz, M. E., Boeve, B. F., Dickson, D., Floeter, M. K., Traynor, B. J., Morelli, C., Ratti, A., Silani, V., Rademakers, R., Brown, R. H., Rothstein, J. D., Boylan, K. B., Petrucelli, L., and Disney, M. D. (2014) Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**, 1043-1050

64. Disney, M. D., Liu, B., Yang, W. Y., Sellier, C., Tran, T., Charlet-Berguerand, N., and Childs-Disney, J. L. (2012) A small molecule that targets r(CGG)(exp) and improves defects in fragile X-associated tremor ataxia syndrome. *ACS Chem Biol* **7**, 1711-1718

65. Wang, Z. F., Ursu, A., Childs-Disney, J. L., Guertler, R., Yang, W. Y., Bernat, V., Rzuczek, S. G., Fuerst, R., Zhang, Y. J., Gendron, T. F., Yildirim, I., Dwyer, B. G., Rice, J. E., Petrucelli, L., and Disney, M. D. (2019) The Hairpin Form of r(G. *Cell Chem Biol* **26**, 179-190.e112

66. Simone, R., Balendra, R., Moens, T. G., Preza, E., Wilson, K. M., Heslegrave, A., Woodling, N. S., Niccoli, T., Gilbert-Jaramillo, J., Abdelkarim, S., Clayton, E. L., Clarke, M., Konrad, M. T., Nicoll, A. J., Mitchell, J. S., Calvo, A., Chio, A., Houlden, H., Polke, J. M., Ismail, M. A., Stephens, C. E., Vo, T., Farahat, A. A., Wilson, W. D., Boykin, D. W., Zetterberg, H., Partridge, L., Wray, S., Parkinson, G., Neidle, S., Patani, R., Fratta, P., and Isaacs, A. M. (2018) G-quadruplex-binding small molecules ameliorate C9orf72 FTD/ALS pathology in vitro and in vivo. *EMBO Mol Med* **10**, 22-31

67. McEachin, Z. T., Gendron, T. F., Raj, N., García-Murias, M., Banerjee, A., Purcell, R. H., Ward, P. J., Todd, T. W., Merritt-Garza, M. E., Jansen-West, K., Hales, C. M., García-Sobrino, T., Quintáns, B., Holler, C. J., Taylor, G., San Millán, B., Teijeira, S., Yamashita, T., Ohkubo, R., Boulis, N. M., Xu, C., Wen, Z., Streichenberger, N., Fogel, B. L., Kukar, T., Abe, K., Dickson, D. W., Arias, M., Glass, J. D., Jiang, J., Tansey, M. G., Sobrido, M. J., Petrucelli, L., Rossoll, W., Bassell, G. J., and Network, N. C. N. (2020) Chimeric Peptide Species Contribute to Divergent Dipeptide Repeat Pathology in c9ALS/FTD and SCA36. *Neuron* **107**, 292-305.e296

68. Skabkin, M. A., Skabkina, O. V., Dhote, V., Komar, A. A., Hellen, C. U., and Pestova, T. V. (2010) Activities of Ligatin and MCT-1/DENR in eukaryotic translation initiation and ribosomal recycling. *Genes Dev* **24**, 1787-1801

69. Starck, S. R., Tsai, J. C., Chen, K., Shodiya, M., Wang, L., Yahiro, K., Martins-Green, M., Shastri, N., and Walter, P. (2016) Translation from the 5' untranslated region shapes the integrated stress response. *Science* **351**, aad3867

70. Starck, S. R., Jiang, V., Pavon-Eternod, M., Prasad, S., McCarthy, B., Pan, T., and Shastri, N. (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719-1723

71. Hill, Z. B., Pollock, S. B., Zhuang, M., and Wells, J. A. (2016) Direct Proximity Tagging of Small Molecule Protein Targets Using an Engineered NEDD8 Ligase. *J Am Chem Soc* **138**, 13123-13126

72. Krans, A., Skariah, G., Zhang, Y., Bayly, B., and Todd, P. K. (2019) Neuropathology of RAN translation proteins in fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun* **7**, 152

73. Krans, A., Kearse, M. G., and Todd, P. K. (2016) Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Ann Neurol* **80**, 871-881