The Role of Upstream Open Reading Frames in Regulating Neuronal Protein Synthesis

by

Caitlin Marie Rodriguez

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Neuroscience)
in The University of Michigan
2018

Doctoral Committee:

Associate Professor Peter Todd, Chair
Associate Professor Aaron Goldstrohm, University of Minnesota
Assistant Professor Kenneth Kwan
Assistant Professor Ryan Mills
Associate Professor David Turner

Caitlin Marie Rodriguez

ctln@umich.edu

ORCID iD: 0000-0002-3827-6753

For my dad.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis mentor, Dr. Peter Todd. Having a mentor that I could fully trust and rely on is something that I do not take for granted. You took me into your lab without hesitation, put so much trust in my abilities, and always supported my curious pursuits. Your support has convinced me to accomplish things in graduate school beyond what I foresaw for myself, I will take this with me as I move forward in my career. I was fortunate have had a mentor that saw me as a collaborator and trusted my opinion throughout this entire process. I will always seek to apply the lessons you have taught me, and I will always be proud to say that I trained in your lab.

Thank you to my thesis committee for being a great support throughout my time at Michigan. Your knowledge and expertise helped to direct my focus, and provided me with training and insight on how to best approach the most important research questions.

I have also been incredibly lucky to have worked alongside some of the smartest, most talented, and hilarious people. To my lab mates, thank you for being incredible! You guys have taught me so much in our time together. I am especially thankful to Mike Kearse for always answering my questions, and for being like a second mentor to me. Thank you, Shannon, Becky, and Katelyn, for the fun times outside of lab. To all of the other lab members, I am so grateful to have had the chance to work with you. In the Mills lab, I would like to thank Tony Chun, whose expertise in bioinformatics and genomic analysis was integral to the work presented in Chapter 2.

To the Neuroscience Graduate Program, thank you for your warm support from day one. Thank you Dr. Edward Stuenkel and Dr. Audrey Seasholtz for taking genuine interest in my development as a scientist, and for providing resources and training opportunities. Valerie Smith and Rachel Harbach, you treat all of us like family. I will always have fond memories of my time in the program, and the fun times we shared as a group. I could not have picked a better program for my graduate work.

I am incredibly thankful for my boyfriend Graham and the little family we have made with our dog Penny. Graham, coming home to you every day has been the best part of these last 4 and a half years. Nightly walks with you and Penny have been some of my most cherished moments. Thank you for listening to my daily ramblings about failed experiments and other stressful situations. Thank you for hating the cold weather as much as I do, and supporting my need to move "back west". I am so excited for our next adventure!

Having the love and support of my family has kept me going through the toughest times in graduate school. Holidays and summers always provided a much-needed break from school, and I cherish the time we spent together. Dad, thank you for talking to me at all hours of the day and night, and taking genuine interest in my research. Thank you, Mom and Dad, for always supporting me, and showing your love. I am so lucky to have you both. Grammy, Auntie Di Uncle Tom, and Courtney, you guys have all always been so caring and loving, you mean the world to me.

Finally, to all my friends or "my Ann Arbor Family", thank you for the fun times and the deepest understanding of all things grad school. To my friends Brittany, Jackie, Nicole, Rob, and Scott, you made Ann Arbor truly feel like my home. I will miss having you guys nearby, but I

know we will remain close. To my cohort, I am so lucky to have had you guys from the beginning. I cannot wait to see what you all accomplish!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Spatial and temporal control of protein synthesis in response to activity is required for neuronal function and plasticity. mRNA structure and sequence provide a powerful platform for such regulation, but how such information is utilized in neurons is incompletely understood. In my thesis, I explore how functional elements within 5'leaders (traditionally termed 5'UTR or untranslated region) of mRNAs act as cis-regulatory elements to influence basal and activity-dependent translation in neurons.

First, I identified a specific role for upstream open reading frames (uORFs) in regulating mRNA translation during neuronal differentiation. uORFs are regions within the 5' leader that undergo translation. Using ribosome profiling (RP), an emerging next-generation sequencing technique which utilizes a modified RNA-sequencing library preparation to detect regions of mRNA occupied by actively translating ribosomes, I identified thousands of uORFs in human neuroblastoma cells. A portion of these uORFs demonstrated clear usage shifts with differentiation. Highly conserved uORFs exhibited increased GC content and were associated with cumulatively repressed CDSs. Importantly, changes in the translational efficiency of these conserved uORFs across differentiation were inversely correlated with CDS translation on these same transcripts. These data demonstrate uORF usage is common in neuroblastoma cells and that specific uORFs act as regulators of cell state-specific translation in neuronal differentiation.

Next, I investigated the function of CGG repeats in the 5' leader of *FMR1*. All humans have a conserved CGG-trinucleotide repeat (typically 20-45 repeats) in *FMR1* that can become

unstable and expand intergenerationally. Large expansions (>200 CGG repeats) cause Fragile X Syndrome, a common cause of intellectual disability, by silencing *FMR1*, leading to loss of the fragile X protein, FMRP. Intermediate (55-200 CGGs) expansions, in contrast, are transcribed and cause an age-related neurodegenerative condition known as Fragile-X Associated Tremor/Ataxia Syndrome (FXTAS). Our lab discovered that this repeat facilitates Repeat Associated Non-AUG translation (RANT), whereby ribosomes initiate at non-AUG codons upstream of the repeat to produce toxic homopolymeric proteins that drive pathogenesis in FXTAS. *FMR1* avidly supports RANT at normal repeat sizes, suggesting that it might serve as a regulatory uORF to control FMRP synthesis. To address this, I expressed nanoluciferase reporters in rat hippocampal neurons. Using this strategy, I found that RANT exhibits a strong negative effect on FMRP synthesis at both normal and expanded repeats. FMRP is a key synaptic protein that is rapidly synthesized in response to mGluR activity. Importantly, preventing RANT or removing the repeat itself blocked this mGluR-induced response. This suggests that *FMR1* relies on these two elements to appropriately scale synaptic FMRP synthesis.

Using non-cleaving antisense oligonucleotides (ASOs) that target the RANT initiation sites, I found that blocking RANT could decrease toxic protein production and prevent neuronal death. In a line of iPSC-derived neurons from a patient with a large CGG repeat (>200) that still generates *FMR1* mRNA but has deficits in FMRP, treatment with the ASO increased endogenous FMRP expression by 50%. These findings define a native function for RANT and CGG repeats in regulating FMRP synthesis, and delineate RANT as a therapeutic target in Fragile X-associated disorders.

CHAPTER I

The role of noncanonical mRNA translation in health and disease

Introduction

Gene expression in its simplest form relies on the process of DNA transcription to RNA, then translation of the RNA messenger into protein. This flow of information—termed the central dogma of molecular biology—provides an infrastructure for the process by which DNA is expressed as a functional protein or acts as a functional RNA element. However, it is a gross oversimplification. Every step in the pathway is highly regulated, with multiple modifiable factors that selectively influence protein generation. RNA sequence and secondary structure in particular acts as a powerful source of spatiotemporal regulation of gene expression. Moreover, every gene has the potential to reveal caveats to well-established doctrine.

The scanning mechanism of eukaryotic translation initiation

The scanning mechanism of translation initiation broadly refers to the process whereby the small ribosomal subunit with its associated factors binds to the modified guanosine cap (m^7G) on the 5' end of mRNA transcripts, then progresses in the 3' direction to a start codon where it joins with the large subunit to initiate translation (Hinnebusch 2011, Hinnebusch, Ivanov et al. 2016). This process begins when a free 40S ribosomal subunit binds the ternary complex (TC), composed of Met-$tRNA_i^{Met}$ bound to eIF2 (**Figure1.1**). eIF2 is comprised of three subunits. The γ subunit binds GTP, when GTP is hydrolyzed to GDP eIF2 cannot bind

Met-tRNA$_i^{Met}$. Phosphorylation of the α subunit of eIF2 inhibits the guanine exchange factor (GEF) eIF2B, which is responsible for adding GTP to eIF2. eIF4F is comprised of multiple initiation factors that bind to the m^{7}G cap, this interacts with the polyadenosine binding protein (PABP) present on the 3' end of the transcript, circularizing the mRNA(Hinnebusch 2011, Hinnebusch, Ivanov et al. 2016). Binding of the 40S subunit to the TC, along with eIF1, eIF1A, eIF3, and eIF5 forms the 43S preinitiation complex (PIC). The PIC is recruited to the transcript through interaction with the eIF4F complex, then begins scanning. Upon AUG start site recognition, eIF5 facilitates hydrolysis of the GTP bound to eIF2, initiation factors dissociate, and the 60S subunit binds forming the 80S ribosome with the Met-tRNA$_i^{Met}$ positioned in the P site. Not every AUG codon within the 5'leader is utilized for initiation. Robust translation initiation depends not only on the perfect complementarity of the Met-tRNA$_i^{Met}$ anticodon, but also the sequence directly surrounding (Thach, Dewey et al. 1966, Kozak 1981, Kozak 1984). The consensus or "Kozak" sequence, 5'-GCCGCC(A/G)CCAUGG-3', marks the site of initiation(Kozak 1984, Kozak 1987). Endogenously occurring suboptimal AUGs in 5' leaders are often bypassed by the PIC in a process termed "leaky scanning". However, not all upstream AUGs are bypassed, striking a balance of translation on an individual transcript. Thus, the scanning model of translation initiation has its caveats, many of these are influenced by characteristics of 5' leaders which will be further reviewed here.

Ribosome profiling illustrates global protein synthesis

Ribosome profiling (RP) in a next-generation sequencing method for locating translating ribosomes across the transcriptome (Ingolia, Ghaemmaghami et al. 2009). Cycloheximide is used to halt translating ribosomes on mRNA, RNase I is used to degrade unbound RNA, and a library

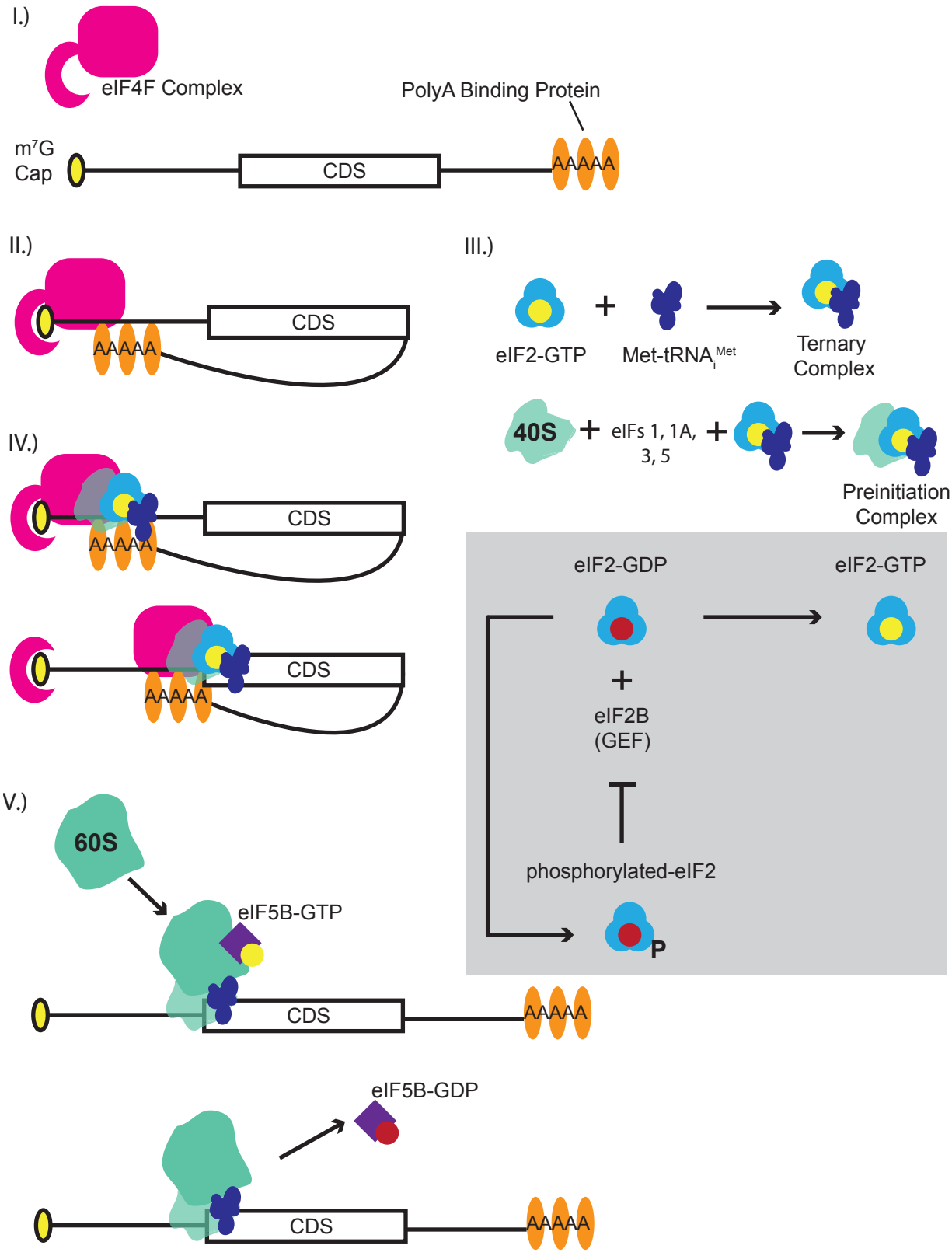

I.)
eIF4F Complex
PolyA Binding Protein
m7G Cap
CDS
AAAAA
II.)
CDS
AAAAA
III.)
eIF2-GTP
Met-tRNAiMet
Ternary Complex
40S
eIFs 1, 1A, 3, 5
Preinitiation Complex
IV.)
CDS
AAAAA
CDS
AAAAA
eIF2-GDP
eIF2-GTP
eIF2B (GEF)
phosphorylated-eIF2
P
V.)
60S
eIF5B-GTP
CDS
AAAAA
eIF5B-GDP
CDS
AAAAA

Figure 1.1:

The scanning mechanism of eukaryotic translation initiation. I) An mRNA transcript is processed to have an m^7G nucleoside at its most 5' position and a polyadenylated tail associated with Poly(A) binding proteins. II) The eIF4F complex binds to the m^7G cap. Poly(A) binding proteins interact with eIF4F which circularizes the transcript. III) eIF2 bound to GTP associates with Met-$tRNA_i^{Met}$ to form the ternary complex. The ternary complex, eIF1, eIF1A, eIF3, eIF5, and the 40S ribosome come together to form the 43S preinitiation complex. IV) The preinitiation complex is recruited to the mRNA transcript through association with eIF4F. Then, the preinitiation complex and components of eIF4F scan the mRNA in the 5' to 3' direction in search of an optimal start site. To solidify start site recognition, eIF1 dissociates, and eIF5 facilitates the hydrolysis of eIF2 bound GTP to GDP (not shown). V) The 60S subunit of the ribosome binds to the 40S subunit forming the 80S ribosome at the initiation site, facilitated by eIF5B. The eIF4F complex dissociates. When eIF5B hydrolyzes its GTP to GDP, eIF1A dissociates (not shown), and the 80S ribosome can form its first peptide bond. Inset: eIF2B stimulates GTP binding to eIF2, leaving eIF2 in its functional state to go on to form a ternary complex. eIF2 bound to GDP can be phosphorylated on the α subunit, making it a competitive inhibitor of eIF2B.

is made from the ribosome protected footprints (RPF), each of which is approximately 28 nucleotides (Ingolia 2010, Ingolia, Brar et al. 2012). Assessing global ribosome occupancy with RP has granted a new level of insight into the guiding principles of mRNA translation. Utilizing RP to investigate mRNA translation in *Saccharomyces cerevisiae* affirmed that ribosome profiling can recapitulate classic biochemical studies (Ingolia, Ghaemmaghami et al. 2009). Permutations of RP have also been highly informative. TCP-seq, or translation complex profile sequencing, utilizes formaldehyde crosslinking and a broader range of footprint sizes to select for intermediates in the initiation pathway (Shirokikh, Archer et al. 2017). This technique shows promise for identifying comprehensive principles that guide mRNA-translation machinery interactions (Archer, Shirokikh et al. 2016). At the core of the analysis of ribosome profiling studies is the calculation of translational efficiency (TE) by taking the ratio of RPF and mRNA abundance (Ingolia, Ghaemmaghami et al. 2009). RP is able to resolve active translation across transcripts with peaks in read density showing the highest three-nucleotide periodicity in the native frame(Ingolia, Ghaemmaghami et al. 2009). Interestingly, the earliest studies employing RP showed pervasive 5'leader translation and NCC initiation (NCC) (Ingolia, Ghaemmaghami et al. 2009, Ingolia, Lareau et al. 2011). RP studies will be further discussed in the following text.

The 5' Untranslated Region regulates mRNA translation

A potential role for the 5' leader in regulation of translation initiation was recognized early following establishment of the cognate start codon and its location near the 5' end of mRNA(Nirenberg and Leder 1964, Sundararajan and Thach 1966, Thach, Dewey et al. 1966, Bretscher 1968, Steitz 1969, Chang, Temple et al. 1977, Efstratiadis, Kafatos et al. 1977). The first fully sequenced human 5' leaders were present in the alpha- and beta- globin transcripts (Baralle 1977, Chang, Temple et al. 1977). The presence of secondary structure in the regions sequenced was noted, with a highly stable hairpin formed in both transcripts(Chang, Temple et al. 1977). While the exact influence of the 5' leader on peptide synthesis was unknown, the authors purported a likely interaction between this structured region and the ribosome.

In the decade that followed, there was a push to identify regions responsible for initiation complex binding (Sherman, Stewart et al. 1980, Kozak 1981, Kozak 1986). Abolishing secondary structure in the 5' leader of a reporter mRNA with denaturation or inosine monophosphate incorporation lead to further migration of the 40S ribosome past 5' proximal AUG and a decrease in translation efficiency at this site, thereby supporting the role for a structured 5' leader in influencing translation initiation(Kozak 1980). Another study in *Saccharomyces cerevisiae* hinted that the AUG sequence is not an all or nothing requirement for initiation. In-frame AUGs were cloned in various locations upstream of the main start site which was mutated to preclude initiation(Sherman, Stewart et al. 1980). In doing so expression of the protein of interest varied, but this was not fully understood until the nucleotides flanking the AUG were identified as key modulators of 80S initiation(Kozak 1981).

Upon identification of the full Kozak sequence and substantial evidence for the scanning model of translation initiation, the possible roles for the 5' leader sequence in influencing initiating ribosomes became clearer(Kozak 1981). Insertion of out-of-frame AUG codons upstream of the main initiation site AUG in preproinsulin coding sequence was an early indicator that translation in the 5' leader can influence downstream translation(Kozak 1984). More optimal start sites inserted upstream of the main AUG were shown to be more highly repressive. In addition, if the reading frame derived from the AUG insertion was overlapping the main CDS (**Figure 1.2**) the level of repression was higher than if the in-frame termination codon was in the

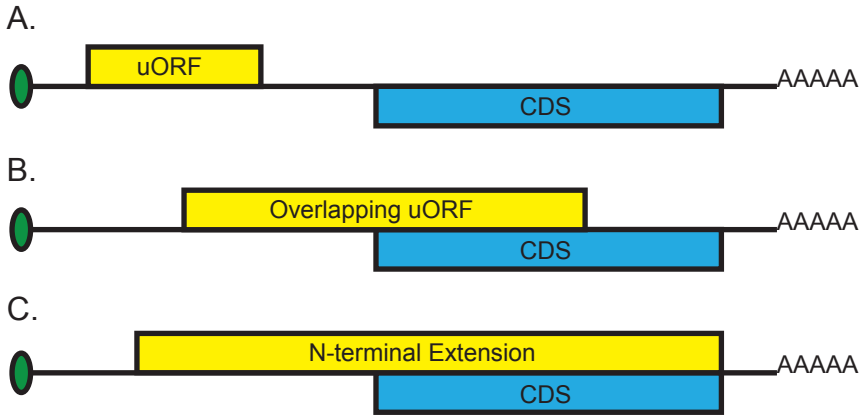


Figure 1.2:

uORF architecture is dictated by reading frame and start site. A) A uORF with an initiation site and stop codon upstream of the main CDS is fully contained in the 5' leader. B) A uORF that starts out of frame with the CDS can terminate within the CDS, forming an overlapping uORF. C) A uORF that initiates in frame with the CDS and does not encounter a termination site in the 5' leader, will produce an N-terminal extension to the CDS protein.

5' leader. It was also described that if the ORF made from the upstream AUG terminated upstream of the preproinsulin CDS, preproinsulin was still made lending evidence to the theory that ribosomes can reinitiate. Further clarification on reinitiation has suggested that 40S ribosomes can remain engaged after translation, rebind the TC, and reengage with a downstream start site(Kozak 2002). This is influenced by length of the upstream coding region and the

distance between its termination codon and the next start site (Kozak 1987, Kozak 2002). The effect of secondary structure on start site recognition was recognized when a stable hairpin structure was inserted upstream of the AUG, dramatically decreasing the overall yield of the measured protein(Kozak 1986). Leader length was also shown to affect initiation, with longer unstructured regions upstream a stem-loop structure increasing the efficiency of translation of the reporter(Kozak 1988). Importantly, positioning a hairpin structure just downstream of an NCC, increased the likelihood of initiation at that site with GUG as the most efficient initiator of the codons tested (Kozak 1989, Kozak 1990). Many of these early studies preempted a large body of research illustrating the effect that endogenously occurring ORFs residing in the 5' leader of key genes can have on translation.

eIf1 and eIF5 are key to start site choice

Two factors—eIF1 and eIF5—influence start site choice (Kozak 1986, Hinnebusch 2011). eIF1 is termed the "gatekeeper" of start site choice, as it must dissociate from the PIC where it resides so that initiation can proceed (Hinnebusch 2011). Thus, blocking the dissociation eIF1 in mammalian cells in the appropriate mRNA context can increase initiation at near-cognate start sites and suboptimal AUGs, (Valasek, Nielsen et al. 2004, Hinnebusch 2011). eIF1 interacts directly with the 40S subunit in coordination with eIF1A to assist in binding the TC, but also promotes an "open" or scanning conformation of the PIC(Passmore, Schmeing et al. 2007). Once, eIF1 dissociates, the TC is more firmly bound placing the initiator tRNA more optimally for AUG-anticodon binding(Maag, Fekete et al. 2005, Nanda, Cheung et al. 2009). In yeast, the Sui1^{-} mutation in eIF1 reduces the interaction between eIF1 and the 40S subunit and increases initiation at non-AUG start sites (Cheung, Maag et al. 2007). eIF5 is a GTPase

accelerating protein (GAP) that stimulates the hydrolysis of GTP bound to eIF2 upon start site recognition. When eIF1 is bound, it is thought to block the GAP function of eIF5(Unbehaun, Borukhov et al. 2004). eIF5 is believed to promote dissociation of eIF1 through interactions of their N-terminal domains, and thus promote initiation(Hinnebusch 2011). In yeast, eIF5 overexpression increases initiation at the near-cognate AUU codon, which is believed to primarily occur due to altered interactions with eIF1 (Nanda, Cheung et al. 2009). In human cells, eIF5 overexpression also increased initiation at suboptimal AUGs and near cognate codons(Loughran, Sachs et al. 2012). Interestingly, both eIF1 and eIF5 utilize start site context as a form of autoregulation. The AUG of eIF1 itself is in poor context, in mammals when eIF1 levels are high, eIF1 inhibits its own synthesis (Ivanov, Firth et al. 2011). When eIF1 function is compromised in yeast, such as in the Sui⁻ mutation, initiation at its suboptimal AUG is increased (Martin-Marcos, Cheung et al. 2011). eIF5 takes advantage of initiation at an inhibitory suboptimal AUG upstream of its main start site, so when eIF5 levels are high increased initiation at this upstream site inhibits further scanning to main CDS AUG of eIF5, creating a negative feedback loop(Loughran, Sachs et al. 2012). These two antagonistic factors for translation initiation highlight how important proper coordination is to start site recognition.

Upstream ORFs influence main ORF translation

A uORF can implement a regulatory role on its resident transcript through three overarching mechanisms: 1) through sequence-dependent interactions with the translation machinery, 2) through cis- and trans-interactions with uORF-derived peptides, and 3) through influencing RNA stability.

uORFs can influence CDS translation through sequence-dependent interactions

Studies on the yeast transcriptional activator Gcn4 provided the first evidence that uORFs present in the 5' leader of a eukaryotic transcript can influence main protein synthesis(Hinnebusch 1984, Thireos, Penn et al. 1984, Mueller and Hinnebusch 1986, Tzamarias, Alexandraki et al. 1986). *GCN4* bears 4 non-overlapping AUG-initiated uORFs (Figure 1.3). Initially, point mutations of these AUGs were shown to variably increase *GCN4*

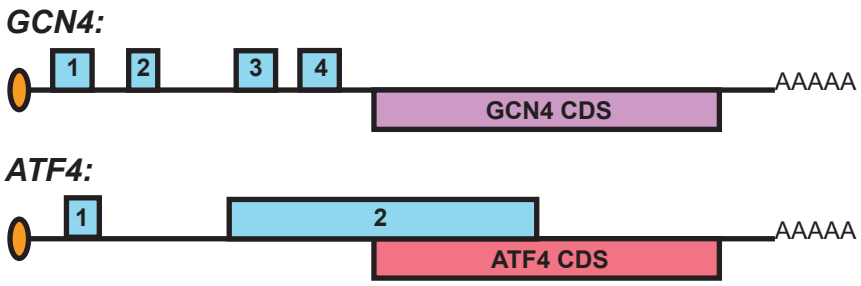


Figure 1.3:

***GCN4* and *ATF4* are transcripts regulated by uORF activity.** *GCN4* contains four uORFs in its 5' leader. In yeast, Gcn4 protein production is regulated by its 5' leader which modulates the availability of initiating ribosomes and protects the transcript from the stress-induced translational blockade. The 5' leader of *ATF4* bears one short uORF and one overlapping uORF. Like *GCN4*, these features boost the production of *ATF4* in response to stress.

production(Mueller and Hinnebusch 1986). Delving deeper into their function, the presence of these uORFs allow for the proper translational response of *GCN4* to amino acid starvation. Under normal conditions, the availability of the ternary complex is non-limiting. uORF1 is translated first, then the ribosome reinitiates at the following 3 uORFs. Removal of uORF1 actually decreases Gcn4 protein(Mueller and Hinnebusch 1986). This is likely due to increased initiation at the downstream uORFs. Following translation of uORF4, the ribosome dissociates and cannot reinitiate translation at the main initiator AUG of *GCN4* (Hinnebusch, Ivanov et al. 2016). Thus, basal levels of Gcn4 are low. This changes under conditions of stress. During amino acid starvation, the kinase Gcn2 phosphorylates serine 51 of the α subunit of eIF2.

Phosphorylation of eIF2α turns eIF2 into a competitive inhibitor of the GEF, eIF2B. eIF2B is responsible for the GDP bound to eIF2 to GTP. Thus, fully assembled ternary complexes are limiting. This changes the translation of the ORFs present on the *GCN4* transcript (Hinnebusch 1997). While uORF1 is still translated under stress, a loss of available ternary complexes is thought to preclude scanning 40S ribosomal re-initiation at the following three uORFs. By the time the scanning ribosome reaches the AUG of the main ORF, it has re-acquired bound the ternary complex and can initiate translation (Hinnebusch 1997). Gcn4 protein is thus present at increased amounts during amino acid starvation, and can act to increase transcription of genes involved in amino acid biosynthesis.

GCN4 is the premier example of antithetical protein synthesis among stress-induced translational arrest, however cellular response to stress is known to affect a number of uORF-containing transcripts(Young and Wek 2016). The Integrated Stress Response (ISR) is activated by specific stimuli, namely ER stress, viral infection, heme deprivation, and amino acid deprivation(Pakos-Zebrucka, Koryga et al. 2016). These stimuli each act on a specific eIF2α kinase leading to an increase in eIF2α phosphorylation, and a global decrease in protein synthesis(Harding, Novoa et al. 2000, Harding, Zhang et al. 2003, Pakos-Zebrucka, Koryga et al. 2016). Despite a global translational arrest, a subset of transcripts show a paradoxical increase in translation. These transcripts distinctly contain uORFs, many of the gene products are associated with cellular response to mitigate stress and maintain homeostasis. *ATF4* is the prototypical mammalian transcript modulated by uORF translation(Harding, Novoa et al. 2000, Vattem and Wek 2004). Like *GCN4*, *ATF4* is a transcription factor responsible for activating transcription of stress response factors. It has two uORFs, the second of which overlaps the CDS (**Figure 1.3**). Studies have shown that uORF1, the most 5' proximal uORF, is translated under normal

conditions followed by reinitiation at uORF2(Vattem and Wek 2004). As uORF2 is overlapping, this diverts ribosomes from translating the *ATF4* ORF. Under ER stress or amino acid starvation, when eIF2α phosphorylation is upregulated, uORF2 is bypassed after uORF1 is translated. Thus, more 43S complexes reach the AUG of *ATF4*. Overall, delayed re-initiation leads to a general increase in the main protein product. A similar mechanism has been described for *ATF5*(Zhou, Palam et al. 2008). *GADD34* and *CHOP* also encode stress response factors under similar conditions, and contain uORFs in their 5' leaders(Lee, Cevallos et al. 2009, Palam, Baird et al. 2011).

Starck et al. (2016) suggested that some uORFs exhibit enhanced translation during stress through use of alternative ternary complexes or non-canonical initiation factors such as eIF2A(Starck, Tsai et al. 2016). This was illustrated through analysis of BiP, an ER chaperone, with two NCC initiated uORFs; the first initiates at a UUG and the second initiates at a CUG, and overlaps the CDS. eIF2A knock-down decreased translation of the second uORF, and was necessary for the maintenance of BiP levels after stress induction with thapsigargin(Starck, Tsai et al. 2016). Importantly, a role for eIF2A has been explored for Leu-tRNA initiated translation at CUG codons in MHC class I molecules from mouse antigen producing cells and at a CUG initiated N-terminal extension to PTEN in human and mouse(Starck, Jiang et al. 2012, Liang, He et al. 2014). These studies highlight a role for alternative translation factors in production of uORFs with NCC start sites.

uORF-derived peptides can influence CDS translation through cis- and trans-interactions with ribosomes

The *AMD1* mRNA transcript exhibits an interesting mechanism of noncanonical translation regulation that relies on the identity of the nascent peptide. A uORF present in the transcript encodes a hexapeptide that stalls its translating ribosomes at its termination site (Hill and Morris 1993, Law, Raney et al. 2001). Interestingly, the half-life of ribosomal stalling is proportional to the concentration of polyamines present and directly affects translation of the downstream ORF in cells(Law, Raney et al. 2001). These data point to a role for translation of uORF peptides working in cis to impede ribosome function. A separate mechanism whereby the synthesized uORF peptide can act to regulate downstream translation was described in the β_2 adrenergic receptor mRNA (Parola and Kobilka 1994). In this case it is postulated that the uORF-derived peptide can interfere with mRNA and ribosome interactions after being fully synthesized. When a fusion peptide derived from this sequence is added to an *in vitro* translation reaction, production of the main ORF is impaired (Parola and Kobilka 1994). Similar mechanisms have been suggested to regulate translation on the *CHOP* and *V1bR* (vasopressin V1b receptor) mRNA transcripts (Jousse, Bruhat et al. 2001, Rabadan-Diehl, Martinez et al. 2007). Taken together, these data point to a distinct mechanism for translation regulation by uORFs.

Translation of uORFs can influence mRNA stability

Another means by which a uORF can exert a regulatory role on transcripts is through influencing mRNA stability. The principles for this mechanism are best understood in the context of nonsense mediated decay (NMD). NMD occurs in the wake of the ribosome encountering a premature termination codon, making the potential for 5' leader translation an apparent trigger for this quality control measure. If a premature termination codon (PTC) resides

within 50-55 nucleotides to the 5' end of an exon-junction site, this can trigger endonucleolytic cleavage and degradation of the transcript (Popp and Maquat 2013). This is believed to occur through interactions with the macromolecular exon-junction complex (EJC) and proteins associated with the terminating ribosome, namely UPF1 whose hyperphosphorylation as a consequence of encountering an EJC recruits decay factors(Le Hir, Izaurralde et al. 2000, Lejeune, Li et al. 2003, Popp and Maquat 2013). This process is thought to only come into effect during the first round of translation, which displaces EJCs for subsequent translation(Ishigaki, Li et al. 2001, Popp and Maquat 2013). The presence of a uORF is one of the top predictors of whether a transcript is an NMD target (Colombo, Karousis et al. 2017). Intriguingly, the transcripts of several NMD factors—*UPF2*, *SMG1*, *SMG5*, *SMG6*, *SMG7*—bear uORFs that are thought to make them substrates of this form of RNA decay themselves (Huang, Lou et al. 2011, Yepiskoposyan, Aeschimann et al. 2011). It is believed that this is one mechanism of autoregulation by which NMD maintains its balance—by a negative feedback loop to prevent over activity of RNA decay.

Ribosome Profiling reveals widespread prevalence of uORFs

As uORFs were initially described on only a case to case basis, the extent to which they are translated truly came to light through RP studies. Ingolia et al. (2009) described a significant enrichment of ribosomes in 5' leaders in *Saccharomyces cerevisiae*, with a subset showing translation comparable to coding regions(Ingolia, Ghaemmaghami et al. 2009). 153 5' leaders were deemed to contain a robustly translated AUG driven uORF, including a set of known uORFs. Interestingly, a separate set of 143 uORFs with initiation in the absence of an AUG codon were also identified. This was soon followed by research in mouse embryonic stem cells

(mESC), showing that ribosome occupancy in 5' leaders is also detected in mammalian cells (Ingolia, Lareau et al. 2011). By using harringtonine, a drug that stops ribosomes at the site of initiation, it was suggested that less than 50% of all translation initiation sites are encoded by an AUG, with near cognate codons making up the vast majority of the rest (Ingolia, Lareau et al. 2011). However, it is important to acknowledge that this number is somewhat misleading, as the most robust initiation sites- both for uORFs and for annotated start sites- were still AUG. However, the sheer number of uORFs (initial estimates were in the thousands) and the finding that they often used NCC initiation sites spurred further studies into this phenomenon (Brar and Weissman 2015, Kearse and Wilusz 2017).

As a proof of principle, ribosome profiling of HEK293T cells treated with tunicamycin, a potent activator of ER stress, delineated the translational changes associated with the ISR (Sidrauski, McGeachy et al. 2015). This study confirmed upregulation of *CHOP*, *GADD34*, *ATF4*, and *ATF5* translation upon stress, which was abrogated in response to ISRIB, an inhibitor of ISR(Sidrauski, McGeachy et al. 2015). Also identified in this analysis were several new transcripts with tunicamycin-induced upregulation of translation. Of these, *SLC35A4* was shown to contain a novel uORF enriched in RPFs which has been shown to produce a functional peptide in human proteome analysis, illustrating that well established molecular pathways can be further clarified by ribosome profiling (Kim, Pinto et al. 2014, Sidrauski, McGeachy et al. 2015).

Intriguingly, studies in multiple models have suggested that uORF translation can be conditionally regulated. uORF translation has been observed to shift during the yeast meiotic cycle and differentiation of mESCs. This shift was illustrated for both the GUG initiated uORF on *CCNB2* and the AUG initiated uORF on *ANAPC5(Ingolia, Lareau et al. 2011)*. A component of this may be attributed to expression of alternative 5' leader sequences (through alternative

splicing) at different stages, which leads to the inclusion or exclusion of potential upstream start sites(Ingolia, Lareau et al. 2011, Brar, Yassour et al. 2012). Shifts in 5' leader translation have also been described by RP studies on mitosis, neuronal differentiation of human ESCs, circadian oscillation, and organism development(Bazzini, Johnstone et al. 2014, Janich, Arpat et al. 2015, Johnstone, Bazzini et al. 2016, Blair, Hockemeyer et al. 2017, Blank, Perez et al. 2017, Sendoel, Dunn et al. 2017).

Using Ribosome Profiling to decipher the global effect of uORFs

The overall repressiveness of upstream translation on main CDS translation can be assessed through ribosome profiling. A positive correlation between 5' leader coverage and main ORF TE was reported on transcripts with RP specified uORFs in *Saccharomyces cerevisiae(Brar, Yassour et al. 2012)*. This was confirmed in human, mouse, and zebrafish datasets(Chew, Pauli et al. 2016). This purports that in order to detect these lower frequency events on transcripts, translation of the entire transcript needs to be higher overall. However, the translational efficiency of CDS coding regions in uORF-bearing transcripts are decreased relative to transcripts lacking a uORF(Chew, Pauli et al. 2016, Johnstone, Bazzini et al. 2016). Interestingly, there may be a dose-dependent effect where an increasing number of uORFs are associated with a decrease in the overall CDS TE(Chew, Pauli et al. 2016, Johnstone, Bazzini et al. 2016). The extent to which 5' leader translation is actually responsible for the decrease in CDS TE remains controversial. For example, inducing oxidative stress in yeast with hydrogen peroxide treatment lead to a significant enrichment of overall ribosome density in 5' leaders of 1,800 transcripts(Gerashchenko, Lobanov et al. 2012). Interestingly, the majority of the 5' leader regions exhibiting an increase in ribosome occupancy were devoid of AUG codons. This

conditional change in 5' leader density was then compared to the change of CDS translation of those affected transcripts, showing no real directional correlation between the two(Gerashchenko, Lobanov et al. 2012). While a subset of transcripts do show an opposing change in main ORF translation when their uORFs are increasingly translated, the global contribution of uORF translation appears to be more indirect. Andreev et al. (2015) utilized HEK293T cells to investigate the translational effects of sodium arsenite, which leads to oxidative stress and eIF2α phosphorylation (Andreev, O'Connor et al. 2015). They confirmed a decrease in global protein synthesis with *ATF4*, *ATF5*, and *SLC35A4* showing a paradoxical increase in RPF coverage. Additionally, reporter studies combined with RP demonstrated that a subset of transcripts with uORFs show only a slight decrease in translation relative to the global decrease, suggesting that uORFs can confer partial resistance to eIF2 repression(Andreev, O'Connor et al. 2015). This work was insightful because it suggested that uORF regulation of CDS translation can be subtle in its impact.

Additionally, the start codon utilized by uORFs may be important to analyzing their overarching role. Spealman et al. (2017) identified AUG and NCC-initiated uORFs in three yeast strains. While main ORFs on transcripts with AUG-initiated uORFs had significantly decreased TE, ORFs on transcripts with NCC-initiated uORFs had increased translation(Spealman, Naik et al. 2017) . Sendoel et al. (2017) investigated which translation factors can promote a shift to more upstream translation during tumor initiation (Sendoel, Dunn et al. 2017). This study showed that mRNA translation during tumor induction was independent of the canonical eIF2 ternary complex. Interestingly, most of the uORFs detected were believed to be NCC initiated, and 13 were found to produce stable peptides. Mindful of the role of eIF2A in promoting uORF translation, they determined a role for this factor in regulating protein synthesis during the

formation of tumors (Starck, Tsai et al. 2016, Sendoel, Dunn et al. 2017). Importantly, when eIF2A dependence is furthered by stress induction and the phosphorylation of eIF2α, they detected increased protein products by mass spectrometry from transcripts with eIF2A-targeted uORFs(Sendoel, Dunn et al. 2017). This highlights a positive correlation between uORF and main ORF translation, when the uORFs are mostly non-AUG initiated. The role for eIF1 was also examined globally by ribosome profiling, and revealed how translational machinery can add to the levels of uORF regulation by affecting start site stringency(Fijalkowska, Verbruggen et al. 2017). During eIF1 knockdown, increased translation at uORFs with NCC start sites was associated with a lower relative CDS TE, conversely AUG initiated uORFs exhibited a decreased TE while their associated CDS increased in translation (Fijalkowska, Verbruggen et al. 2017). Taken together, these studies purport a complex role for uORFs in CDS regulation, and suggest that cell type and cellular conditions may change the relationship of a uORF to its associated transcript.

Experiments in mouse embryonic fibroblasts revealed that ribosome profiling can be used to dissect the role of uORFs in affecting RNA stability. RP was used to identify actively translated uORFs in cells that had been treated with shRNA to knockdown UPF1 and prevent NMD (Hurt, Robertson et al. 2013). Only uORFs contained fully within the 5' leader were used in the analysis, those that overlap the CDS were excluded. Transcripts with actively translated uORFs showed a cumulative increase in abundance in the UPF1 knockdown cells verses control, suggesting that these transcripts are de-repressed when NMD is precluded(Hurt, Robertson et al. 2013). A similar result was suggested in uORF- and overlapping uORF-bearing transcripts in zebrafish, which cumulatively show decreased RNA abundance relative to transcripts without these features(Johnstone, Bazzini et al. 2016).

Transcript leaders influence uORF translation

Alternative transcript leaders impart an important level of regulation that dictates the potential role of a uORF on a transcript. The inclusion of a translation start site or the placement of the next in frame stop codon may be dictated by alternative splicing in the 5' leader as well as alternative transcription start sites. TIF-seq uses a modified library preparation and deep sequencing to identify the most 5' and 3' end of capped and polyadenylated mRNA transcripts(Pelechano, Wei et al. 2013). The implementation of TIF-seq in *Saccharomyces cerevisiae* revealed significant isoform heterogeneity with 80% of transcripts having a median of 10 major isoforms. Interestingly, of the 1,191 transcripts with uORFs identified, 703 had multiple isoforms where the uORF was excluded which may contribute to overall downstream CDS translation levels(Pelechano, Wei et al. 2013). Arribere et al. (2013) performed a modified library prep for identification of transcript leaders and deep sequencing on mRNA that sediment with different fractions of a polysome gradient(Arribere and Gilbert 2013). Gradient fractions represent a range of mRNA translation from low (unbound by ribosomes) to high (bound by polysomes. When analysis was limited to transcripts with AUG initiated uORFs, these were enriched outside of the polysome fractions suggesting lower translation(Arribere and Gilbert 2013). However, this was not observed for transcripts bearing uORFs with a NCC initiation site supporting the idea of distinct roles associated with these two categories of uORFs.

Detection of uORF-derived peptides

Active translation as characterized by ribosome profiling has revealed an abundance of noncanonical ORFs. Detection of stable products synthesized from these regions has been less

robust even for the well-defined uORF on *GCN4* (Zhang and Hinnebusch 2011). To detect proteome-wide short ORF-encoded peptides, one study inspected all peptides larger than 8 amino acids detected by LC-MS/MS in human K562 cells(Slavoff, Mitchell et al. 2013). Tryptic peptides—independent of annotated protein derivatives—were mapped to RNA-sequencing data, and counted if they could be mapped to a discrete, unannotated region of RNA with a potential ORF less than 450 nucleotides in length. 90 short ORF-encoded peptides were identified, 40% of which were predicted to reside in a 5' leader(Slavoff, Mitchell et al. 2013). Only 43% of the entire set was expected to initiate at an AUG, lending credence to the potential production of stable peptides from NCC start sites. Another study took a different approach by enriching for the N-terminus of proteins using terminal amine isotopic labeling of substrates (TAILS) with LC-MS/MS in HEK293 cells(Na, Barbhuiya et al. 2017). This method identified 90 peptides derived from 5' leaders, 69% were N-terminal extensions and 31% were derived from uORFs. Of the peptides detected, only 2% of N-terminal extensions and 11% of uORFs were expected to utilize an AUG start codon. However, nearly all of these initiated using a methionine amino acid in the first position (Na, Barbhuiya et al. 2017). These two studies highlight how proteomics can be utilized to investigate synthesis of 5' leader derived products. However, it is key to note that uORF translation occurs at lower frequency than CDS translation, and shorter peptides may not remain stable for detection by mass spectrometry.

uORFs in the nervous system

Learning and memory requires dynamic changes in synaptic structure and function in response to extracellular stimuli. Many of these changes rely on translation of mRNA transcripts localized to synapses (Santini and Klann 2014). In particular, synaptic potentiation requires

precise spatial and temporal control by key translation factors. Hippocampal dependent learning requires synapses to undergo a form of plasticity elicited by group I metabotropic receptor (mGluR) activation termed long-term depression (LTD), which is a downscaling of synaptic strength dependent on new protein synthesis (Santini and Klann 2014). Pharmacological stimulation of group I mGluRs and induction of LTD increases eIF2α phosphorylation(Di Prisco, Huang et al. 2014). LTD can also be induced by salubrinol, which pharmacologically prevents the dephosphorylation of eIF2α (Di Prisco, Huang et al. 2014). Intruigingly, uORFs are more readily translated after pharmacological mGluR activation. A key neuronal transcript, *OPHN1*, contains uORFs that regulate its expression, and *OPHN1* was required for proper LTD (REF). eIF2α phosphorylation has also been shown to play a role in a distinct form of synaptic potentiation, long term potentiation (LTP), in response to persistent cocaine injections in mice (Huang, Placzek et al. 2016). Additionally, multiple uORFs are suspected to reside in the transcript of protein kinase Mζ, a key factor in LTP maintenance (Ling, Benardo et al. 2002, Bal, Susorov et al. 2016). These data suggest that in the nervous system, a niche of uORF-containing transcripts may play a role in responding to the ever-changing microenvironment of individual neurons.

Nucleotide Repeat Disorders

Microsatellites are repetitive DNA elements that exist across the genome. These elements are inherently polymorphic in nature due to instability leading to expansion or contraction that can occur both somatically and intergenerationally (Lopez Castel, Cleary et al. 2010, McMurray 2010). The expansion of diverse sequences in coding and noncoding regions of genes is a

common cause of neurodegenerative disorders from ALS to myotonic dystrophy, all associated with a diverse set of outcomes (Paulson 2018).

Disease causing repetitive elements are traditionally categorized into three groups: dominantly inherited repeats within protein coding sequences, dominantly inherited repeat expansions outside of known protein coding sequences, and recessive or X-linked repeat expansions that impact the expression of the genes in which they reside. Most of the repeats present within protein coding exons are trinucleotide CAG repeats encoding for glutamine. These polyglutamine diseases include Huntington's Disease, spinal and bulbar muscular atrophy (SBMA), and several spinocerebellar ataxias (SCA). The mechanism thought to drive pathology in these dominantly inherited diseases is translation of the CAG repeat into an expanded polyglutamine tract. The resulting Polyglutamine-containing proteins aggregate and place a proteotoxic burden on cells (La Spada, Wilson et al. 1991, Trottier, Devys et al. 1995, Trottier, Lutz et al. 1995, Bates 1996, Paulson 2018). However, these dominantly inherited diseases are clinically diverse, suggesting a specific interaction between the polyglutamine expansion and the genes/proteinsin which these mutations arise.

Nucleotide repeat expansion mutations can also result in RNA mediated toxicity, with DM1 as the stand-out example. In DM1, a CTG repeat resides in the 3' UTR of the *DMPK* gene. This mutation does not meaningfully alter the abundance of DMPK, rather it causes RNA foci formation and sequestration of key proteins (Pettersson, Aagaard et al. 2015). Specifically, Muscleblind-like (MBNL) proteins, an important family of splicing factors, binds avidly to double-stranded CUG repeats. This sequesters these proteins into RNA foci and precludes them from performing their normal functions. In the case of MBNL sequestration, this leads to accumulation of misspliced transcripts normally regulated by MBNL (Miller, Urbinati et al.

2000, Mankodi, Urbinati et al. 2001, Jiang, Mankodi et al. 2004). Patients with DM1 experience progressive weakness and myotonia, and this can be recapitulated in mice by either overexpressing the CUG repeat out of its native gene context or by knocking out MBNL1 (Paulson 2018). Moreover, overexpression of MBNL1 is sufficient to suppress CUG repeat associated phenotypes in muscle (Kanadia, Shin et al. 2006).

Nucleotide repeat expansions can also elicit a loss of function in the genes in which they reside. Friedreich's ataxia results from a GAA repeat expansion in an intron of the *FRDA* gene which encodes frataxin. Friedreich's ataxia is the most common autosomal recessive ataxia, characterized by early onset ataxia with sensory loss, aggressive scoliosis, and cardiomyopathy(Paulson 2018). In Friedreich's ataxia, GAA repeat expansions lead to a change in the epigenetics of the *FRDA* locus that strongly suppresses the mRNA transcription (Campuzano, Montermini et al. 1996). This leads to a partial, but not complete loss of FRDA mRNA and Frataxin protein. This results from loss of function of frataxin, leading to impaired mitochondrial function and Iron homeostasis (Akbar and Ashizawa 2015). Unlike the dominant disorders described above, Friedreich's ataxia requires expansion or mutations in both *FRDA* alleles, and the size of the repeat correlates inversely with the degree of mRNA expression and the severity of the clinical phenotype (Castaldo, Pinelli et al. 2008).

The Genetic basis of Fragile X Syndrome and Fragile X-associated Disorders

The gene *FMR1* contains a polymorphic CGG repeat in its 5' leader. This repeat is conserved among mammals, and increases in size in higher order primates(Eichler, Kunst et al. 1995). In humans, this CGG element can expand to intermediate or "premutation" lengths of between 55 and 200 repeats. When this occurs, there is an enhancement in *FMR1* transcription,

but inefficient Fragile X protein (FMRP) production (Tassone, Hagerman et al. 2000). Clinically, this results in the neurodegenerative disorder Fragile X-associated Tremor/Ataxia Syndrome (FXTAS) (Hagerman 2013). FXTAS occurs in ~1:5000 men over the age of 50 and in a lower percentage of female premutation carriers. In addition, premutation sized repeat expansions also cause premature ovarian failure, or Fragile X-associated Premature Ovarian Insufficiency (FXPOI), which is the most common genetic cause of early menopause. In both cases, the repeat is thought to elicit disease primarily through a gain-of-function mechanism, as manipulating FMRP expression is unable to modulate phenotypes associated with the repeat in model systems (Arocena, Iwahashi et al. 2005, Jin, Duan et al. 2007, Hashem, Galloway et al. 2009, Lu, Lin et al. 2012).

In contrast to FXTAS and FXPOI, Fragile X Syndrome (FXS) results from expansion of CGG repeats in *FMR1* to greater than 200 CGGs. This leads to hypermethylation of the promoter region and the CGG repeats leads to heterochromatization of the FMR1 locus. Akin to the scenario that occurs in Freidriech's Ataxia, this epigenetic change leads to a partial or complete transcriptional shut off of the *FMR1* locus (Wang, Berry-Kravis et al. 2010). The end result is loss of FMRP. FXS is the most common known monogenic cause of autism and intellectual disability. Patients also have increased rates of attention deficit hyperactivity disorder (ADHD) and epilepsy. FXS affects upwards of 1:5000 males and 1:4000-1:8000 females worldwide (Hagerman, Berry-Kravis et al. 2017).

The threshold at which DNA methylation and transcriptional silencing of the *FMR1* locus occurs has recently been questioned based on studies in patient derived induced pluripotent stem cells and human embryonic stem cells (Sciascia and Usdin , Avitzour, Mor-Shaked et al. 2014). Data in these lines suggest that repeats between 400 and 200 repeats becomes incompletely

methylated at the *FMR1* locus (Sciascia and Usdin). While larger repeats are associated with silencing chromatin modifications, lines that retract in repeat size loose these modifications. This is important because the absolute cutoff of repeat size for disease phenotype is not the complete picture of patients with Fragile X disorders. The *FMR1* locus exhibits significant mosaicism in terms of both repeat length and methylation status within individuals with FXS. Mosaicism occurs in ~40% of Fragile X cases, and is associated with an increase in both *FMR1* mRNA and FMRP (Nolin, Glicksman et al. 1994, Jacquemont, Birnbaum et al. 2011, Pretto, Yrigollen et al. 2014). Generally, FMRP levels positively correlate with IQ, and percent methylation is inversely correlated with IQ (Pretto, Yrigollen et al. 2014). While this would suggest that mosaic individuals have a more favorable prognosis, case studies reveal a range of severity in these cases from unaffected to full Fragile X Syndrome (Wohrle, Salat et al. 1998, Santa Maria, Pugin et al. 2014, Basuta, Schneider et al. 2015).

FXTAS and FXS are traditionally viewed as clinically separate disorders. FXS patients begin displaying abnormalities during infancy, and by the age of two they exhibit delayed development of language, hyperactivity, and sensitivity to sensory stimulation (Hagerman, Berry-Kravis et al. 2017). Patients may have distinguishing physical features such as large ears and a long face. FXS patients may have a number of different behavioral and cognitive issues such as attention-deficit/hyperactivity disorder (ADHD), learning disabilities, intellectual disability, and/or autism.

In contrast, FXTAS affects males with an average age of onset of 60.6 years of age. Patients with FXTAS may exhibit intention tremor, parkinsonism, cerebellar ataxia, and cognitive decline (Hagerman and Hagerman 2016). Imaging studies reveal white matter abnormalities, as well as a reduction in volume of the brain stem and cerebellum (Hashimoto,

Javan et al. 2011, Hashimoto, Srivastava et al. 2011). Neurons exhibit the formation of both RNA foci and ubiquitinated inclusions in the brains of patients (Renoux and Todd 2012).

Most premutation repeat carriers have a normal neurodevelopmental trajectory in childhood and early adulthood. However, early in life, some premutation carriers experience symptoms ranging from migraines, to insomnia and anxiety and in rare cases autism and learning disabilities (Hagerman 2013). Similarly, most patients with FXS do not develop this late-onset motoric difficulties or premature ovarian failure as in FXPOI, although they do have increased rates of movement disorders and stereotypies (Wheeler, Bailey et al. 2014). The mechanisms underlying the differential phenotypes of neurodegeneration in FXTAS and neurodevelopmental abnormalities in FXS will be further discussed in the following sections.

Molecular mechanisms associated with Fragile X-associated disorders

FMRP is expressed throughout the body, however its expression is highest in brain and testis. In the mouse brain, FMRP is expressed in neurons throughout development and adulthood, and is present in glia early but is largely absent from these cells in the adult brain (Gholizadeh, Halder et al. 2015). Most studies of Fragile X Syndrome in mice have relied on an *FMR1* knock-out model, as hypermethylation and transcriptional silencing in the setting of expanded repeats is not recapitulated in mice with 300+ CGG repeats. Knock-out mice exhibit a range of phenotypes such as increased locomotion, learning difficulties, seizure sensitivity, and anxiety in behavioral assays (Hagerman, Berry-Kravis et al. 2017). There have been many avenues of research into the molecular consequences of this loss of FMRP, but the two areas of consensus are impairments in glutamatergic and GABAergic signaling.

The first evidence of glutamatergic signaling impairment came from electrophysiological studies in the Fragile X knock-out model. Hippocampal recordings from these mice showed enhanced mGluR-induce long-term depression (LTD), a form of plasticity that relies on post-synaptic protein synthesis (Huber, Gallagher et al. 2002). In addition, morphological analysis of knock-out neurons revealed dendritic spine abnormalities including increased length and density, suggesting impaired maturation (Irwin, Patel et al. 2001, Nimchinsky, Oberlander et al. 2001). This spine morphology abnormality and enhanced LTD is likely due to changes in dendritic protein synthesis in FXS model mice. FMRP is an RNA binding protein that associates with and stall ribosomes on its associated transcripts (Brown, Jin et al. 2001, Darnell, Van Driesche et al. 2011). In the absence of FMRP, it is thought that three signaling cascades converge to enhance global dendritic protein synthesis (Richter, Bassell et al. 2015). These three cascades, mediated through PI3K, ERK, and mTORC1, are all triggered by mGluR stimulation, and all converge on translation initiation machinery to promote protein synthesis. Drugs targeting any of these pathways rescues several phenotypes in the *FMR1* knock-out mice (Dolen, Osterweil et al. 2007, Michalon, Sidorov et al. 2012, Gross, Chang et al. 2015, Gross, Raj et al. 2015, Richter, Bassell et al. 2015).

FMRP is also thought to regulate its own translation. Upon mGluR stimulation, FMRP is rapidly degraded, allowing for activity dependent translation of the transcripts now freed from its repression. However, the *FMR1* mRNA is present in dendrites and mGluR activation also triggers local synthesis of FMRP (Weiler, Irwin et al. 1997, Todd, Mack et al. 2003). It is thought that newly synthesized FMRP can then rebind to transcripts to turn off translation (Hou, Antion et al. 2006, Nalavadi, Muddashetty et al. 2012). Without FMRP, transcripts undergo higher basal levels of translation, but also exhibit a loss of mGluR stimulation-dependent

translation. The end result is that, in the absence of FMRP, mGluR stimulation leads to enhanced LTD that no longer requires temporally associated protein synthesis (Huber, Gallagher et al. 2002, Bear, Huber et al. 2004).

While the downstream effects of glutamatergic signaling are upregulated, GABAergic signaling is decreased in *FMR1* knock-out mice, and GABAergic agonists rescue many behavioral phenotypes in these mice (Hagerman, Berry-Kravis et al. 2017). Based on these studies, the altered synaptic physiology described is thought to underlie the cognitive and behavioral deficits associated with FXS. Drugs targeting both glutamatergic and GABAergic signaling have been the subject of several clinical trials. However, these trials have thus far failed to meet their primary endpoints in patients, despite some subjective improvements in behaviors in FXS children (Erickson, Davenport et al. 2017). This highlights the importance of continued mechanistic research into the causes of FXS. Importantly, these disappointing outcomes purport that novel models of FXS may be necessary to ensure that findings are more translatable, and that targeting the proximal event in pathogenesis, loss of the FMRP, may be needed to achieve phenotypic correction in humans.

RAN Translation and Neurodegeneration

Microsatellites are capable of forming stable structures in RNA. While this can confer a toxic gain of function through direct binding with proteins(Todd and Paulson 2010, Renoux and Todd 2012), it was recently discovered to also support a newly recognized mechanism of translational initiation. This process is termed repeat associated non-AUG (RAN) translation, whereby toxic homopolymeric peptides are synthesized in the absence of an AUG start codon, in the presence of an expanded microsatellite (Kearse and Todd 2014, Green, Linsalata et al. 2016).

This was discovered by Zu et al. (2011) through work on the spinocerebellar ataxia 8 (SCA8) gene transcript, ATXN8, which contains a CTG repeat in in its 3' UTR that is transcribed into both sense and antisense transcripts (Zu, Gibbens et al. 2011). The antisense transcript from ATXN8 (ATXN8OS) contains a short AUG-driven ORF encompassing the CAG repeat. The surprise came when removal of the AUG or placement of stop codons just upstream of the repeat did not preclude synthesis of the product polyglutamine product. In contrast, translation persisted in all three reading frames of the repeat producing polyalanine and polyserine products in addition to a non-AUG initiated polyglutamine protein (Zu, Gibbens et al. 2011). The polyglutamine product was found to accumulate in cerebellar Purkinje cells of SCA8 patients. Central to this work was the finding that RAN translation of all three products increased with repeat size, and that replacement of the CAG repeat with CAA prevented RAN translation in all three frames (Zu, Gibbens et al. 2011). Similarly, polyglutamine, polyalanine, and polyserine products were detected in the context of the antisense DM1 transcript (in the CAG orientation), where deletion of the only in-frame AUG (for polyserine) did not remove the signal. Polyglutamine aggregates were detected in myoblasts, skeletal muscle, and blood of DM1 patients (Zu, Gibbens et al. 2011). Since this initial finding, RAN translation has been described in a variety of other nucleotide repeat expansion disorders, including genes such as C9orf72 (ALS/FTD) and HTT (Huntington's Disease), along with several (but not all) of the spinocerebellar ataxia genes (Ash, Bieniek et al. 2013, Mori, Weng et al. 2013, Banez-Coronel, Ayhan et al. 2015, Green, Linsalata et al. 2016).

The *FMR1* CGG repeat was also found to support RAN translation in all three potential reading frames on the sense strand in the absence of an AUG (Todd, Oh et al. 2013, Krans, Kearse et al. 2016), albeit with differing efficiencies. RAN translation also occurs at CCG

repeats generated by antisense transcription through the beginning of the *FMR1* gene (Krans, Kearse et al. 2016). Distinct from what was described in *ATXN8*, RAN translation in two of the three potential reading frames of the 5' leader of *FMR1* relies on initiation at NCC start codons just upstream of the repeat (Todd, Oh et al. 2013, Kearse, Green et al. 2016). On the sense strand, the 0-frame RAN initiation event uses a GUG initiation site, and the +1-frame uses both an ACG and a GUG for initiation (**Figure 1.4**) (Todd, Oh et al. 2013, Kearse, Green et al. 2016). Peaks upstream of the repeat were detected in RP datasets from mouse and man (Todd, Oh et al. 2013). Interestingly, an earlier study found that the *Drosophila* homologue to *FMR1*, *dFMR1*, encodes for a functional N-terminal extension that initiates at a CUG start site (Beerman and Jongens 2011). In CGG RAN translation, the +2-frame does not appear to rely on a start codon upstream of the repeat, making it more akin to what is observed at CAG repeat expansions with initiation occuring within the repeat (Kearse, Green et al. 2016). Interestingly, both the +1-frame and +2-frame products increase as a function of increased repeat size, and 0-frame RAN translation is attenuated at normal and expanded repeat lengths (Todd, Oh et al. 2013, Kearse, Green et al. 2016). Kearse et al. (2016) showed that translation of RAN reporter transcripts requires a functional 5' cap and interaction with the cap-binding protein eIF4E, and that blocking eIF4A, the helicase critical for PIC scanning, significantly impaired RAN translation(Kearse, Green et al. 2016). These data point to a canonical scanning mechanism of translation initiation at the CGG repeat, with a primary failure in start codon fidelity leading to RAN initiation.

Intriguingly, under ER stress and phosphorylation of eIF2α, CGG RAN translation of the +1- and +2-frames increases despite the global decrease in translation (Green, Glineburg et al. 2017). This increase was dependent on the presence of an NCC start site, and was also observed in the ALS/FTD GGGGCC repeat (Green, Glineburg et al. 2017, Cheng, Wang et al. 2018).

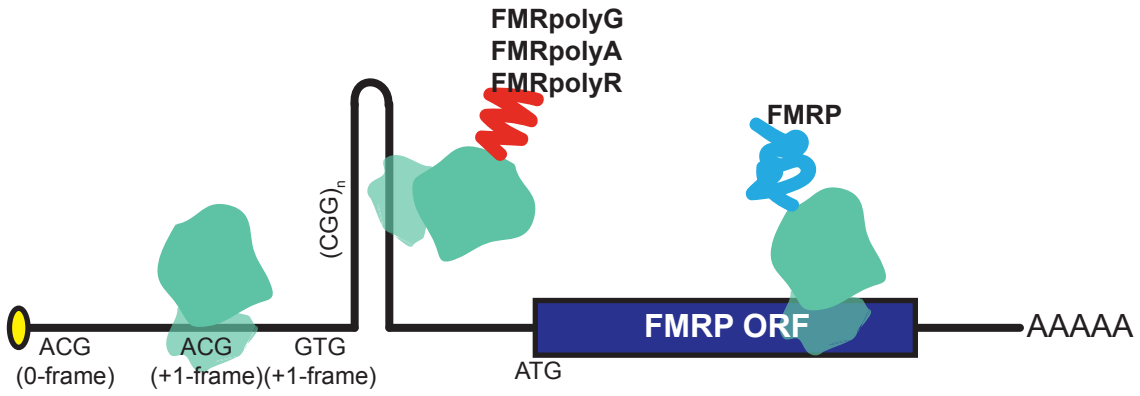


Figure 1.4:

***FMR1* RAN translation initiates at near cognate start sites, in all three frames.** The +1-frame RAN peptide, FMRpolyG, is synthesized starting with an ACG or GTG codon. The 0-frame product, FMRpolyR, is an N-terminal extension that initiates at an ACG. The +2-frame initiates within the repeat, producing FMRpolyA.

Importantly, expression of either expanded CGG or GGGGCC repeats induced stress granule formation in cells, which is dependent on eIF2α phosphorylation (Green, Glineburg et al. 2017, Cheng, Wang et al. 2018). Taken together, these data support the notion of a feed forward mechanism of toxicity whereby RAN translation is upregulated in the presence of stress, and also themselves can independently induce stress, creating a feed-forward loop.

Aggregates of the +1 CGG RAN peptide, FMRpolyG, occurs in mouse models of the disease and, importantly, in patients with FXTAS (Todd, Oh et al. 2013, Sellier, Buijsen et al. 2017). FMRpolyG aggregates are also identified in ovarian stromal cells from FXPOI patients (Buijsen, Visser et al. 2016). In FXTAS and FXPOI patients, these aggregates are ubiquitinated. Driving expression of FMRpolyG causes impaired proteasome function in model systems (Oh, He et al. 2015). Nuclear lamina integrity is also affected in FMRpolyG expressing cells, and FMRpolyG has been shown to interact directly with the inner nuclear membrane protein Lap2ß (Sellier, Buijsen et al. 2017). In mice, whole body expression of the FMRpolyG transgene impacts motor function and survival. Driving expression of the transgene in neurons decreases

survival, and alters body composition. These data provide evidence for RAN-associated toxicity in FXTAS.

Two other nucleotide repeat expansion disorders where RAN translation may contribute to overall pathology are Huntington's Disease (HD) and C9Orf72 ALS/FTD. HD is a polyglutamine disease caused by CAG repeat expansions in the coding region of *HTT,* the Huntingtin gene. Patients generally exhibit involuntary movements and cognitive decline (1993, Dickey and La Spada 2017). Individuals with larger expanded repeats are prone to earlier onset HD, linking the severity of HD to a generational effect (Andrew, Goldberg et al. 1993, Paulson 2018). While the polyglutamine containing protein defines the pathology of this disease, Banez-Coronel and colleagues (2015) showed that it is not the only homopolymeric expansion protein derived from this locus. RAN translation was shown to occur in both the sense and antisense direction through the CAG/CTG repeat, despite its residence within a protein coding open reading frame. In addition to the polyglutamine product, both poly-serine and poly-alanine products (sense) as well as poly-leucine and poly-cysteine products (antisense) were detected (Banez-Coronel, Ayhan et al. 2015). Importantly, production of the polyglutamine product remains in the absence of AUG, although in the native sequence context this is confounded by the presence of both an AUG initiated uORF and a utilized downstream AUG codon also in the Polyglutamine frame. In addition to polyglutamine, the four other potential RAN peptides accumulate in patient brains, and are increased in abundance in early onset HD (Banez-Coronel, Ayhan et al. 2015). HD RAN peptides have been linked to nucleocytoplasmic export disruption in neurons (Grima, Daigle et al. 2017).

ALS (amyotrophic lateral sclerosis) is the most common motor neuron disease, while FTD (frontotemporal lobe dementia) is a form of early dementia that generally appears in

individuals in their 50s and 60s(Renton, Majounie et al. 2011). Though FTD and ALS are two distinct diseases, there is overlap in the patient population. In 2011, expansion of the hexanucleotide repeat GGGGCC in the gene C9Orf72 was discovered to be the most common known mutation underlying both ALS and FTD (DeJesus-Hernandez, Mackenzie et al. 2011, Renton, Majounie et al. 2011). GGGGCC bearing transcripts are transcribed in the sense and antisense direction, and the repeat is present in an intron or promoter region of sense strand RNA transcripts. RAN translation through this repeat produces five potential dipeptide repeat tract proteins—poly-glycine/alanine, -glycine/arginine, -proline/alanine, -proline/arginine, and -glycine/proline (Ash, Bieniek et al. 2013, Gendron, Bieniek et al. 2013, Mori, Weng et al. 2013, Zu, Liu et al. 2013, Green, Linsalata et al. 2016). C9orf72-associated RAN proteins accumulate in the brains of patients in aggregates distinct from the TDP-43 positive aggregates classically seen in ALS/FTD. Two modes of cellular toxicity linked to the production of the various GGGGCC RAN proteins are nucleolar stress and nucleocytoplasmic transport impairment (Wen, Tan et al. 2014, Zhang, Donnelly et al. 2015). Several studies have concluded that poly-glycine/alanine is the most abundant of all GGGGCC RAN proteins followed by the other two glycine containing repeat proteins, and that each frame has a different length-dependency (Green, Linsalata et al. 2016). GGGGCC RAN translation from linear mRNA transcripts appears to mainly depend on cap-binding of eIF4E and a canonical scanning mechanism involving eIF4A, similar to what has been reported for CGG repeats (Green, Glineburg et al. 2017, Tabet, Schaeffer et al. 2018). However, GGGGCC RAN can occur to at least some degree without a cap and on spliced introns (Cheng, Wang et al. 2018). At least two of the reading frames require a near-cognate CUG for initiation, as well as an AGG closer to the repeat (Green, Glineburg et al. 2017, Tabet, Schaeffer et al. 2018). The dependency on a single start site for multiple reading

frames suggests that frame-shifting may occur; this was previously reported at CAG repeats but which was ruled out as a major contributor to the plethora of RAN translation products associated this repeat (Zu et al, 2011). Thus, while many nucleotide repeat expansions appear to support RAN, the underlying mechanism appear to have significant differences—not only between repeats but even in different frames of the same repeat. With so many permutations, there is much more research to be done into this novel mechanism of translation, and how RAN peptides may lead to toxicity.

Other potential pathogenic mechanisms in FXTAS

As in myotonic dystrophy, CGG repeat RNA is capable of forming nuclear foci with specific RNA binding proteins. Significant efforts have been made to determine if RNA-mediated toxicity is a source of dysfunction in FXTAS. Initial work focused on a disease mechanism very similar to myotonic dystrophy: RNA mediated sequestration of specific proteins. CGG repeats form highly structured and repetitive motifs *in vitro* that could serve as scaffolds for binding up RNA binding proteins at high levels, and preventing these proteins from functioning elsewhere in the cell (Renoux and Todd 2012). Initial efforts aimed at identifying which RNA binding protein is specifically associated with CGG repeats revealed two candidates: Pur-alpha and hnRNP A2/B1 (Jin, Duan et al. 2007, Sofola, Jin et al. 2007). Both proteins bound CGG repeats *in vitro* and both were found to modulate CGG repeat-associated toxicity in *Drosophila.* Moreover, mass spectrometry of isolated nuclear aggregates identified over 20 enriched proteins including the RNA binding proteins MBNL1 and hnRNP A2/B1 (Iwahashi, Yasui et al. 2006). A separate group of investigators explored this question using a CGG repeat expressed in isolation in cells (Sellier, Rau et al. 2010). This repeat was very toxic and bound to

a large group of RNA binding proteins over time, including SAM-68, hnRNP A2/B1, and MBNL1. One protein complex in particular was found to associate with CGG repeats both directly, specifically and early after expression in cells: the DGCR8-DROSHA complex, which are components of the miRNA processing machinery (Sellier, Freyermuth et al. 2013). Activity of this complex is reduced in the presence of expanded CGG repeats, and microarray profiling of postmortem FXTAS patient brains revealed a global decrease in mature miRNAs (Sellier, Freyermuth et al. 2013). Overexpression of DGCR8 increased the viability of cells derived from a mouse model of FXTAS. This lends strong evidence to the role of RNA-mediated sequestration in driving toxicity in FXTAS.

An alternative gain of function disease mechanism in FXTAS centers on transcription through the CGG repeat, and the propensity for R-loop formation during this process(Hagerman and Hagerman 2016). R-loops are nucleic acid structures formed by direct binding of newly synthesized RNA to the complimentary DNA at the site of transcription. These are vulnerable sites to DNA damage which can trigger the DNA Damage Response (DDR), a signaling pathway with a number of outcomes including apoptosis (Jackson and Bartek 2009). There are a number of neurodegenerative disorders that DDR is thought to underlie such as spinocerebellar ataxia with axonal neuropathy 1 and ataxia with oculomotor apraxia 1(Rass, Ahel et al. 2007). R-loops have been suggested to form in at least two other nucleotide repeat disorders, C9Orf72 ALS/FTD and Friedreich's ataxia (Groh, Lufino et al. 2014, Reddy, Schmidt et al. 2014). In FXTAS, R-loops form in patient cells with a modest effect of increased repeat size, and their formation is positively correlated with transcription levels (Groh, Lufino et al. 2014, Loomis, Sanz et al. 2014). More research on this subject is needed to determine if there is a link between R-loops and neuron loss in FXTAS.

Some efforts have been made to delineate which of these processes is a driver of toxicity in FXTAS. In *Drosophila*, sequence changes around the repeat that preclude RAN translation of FMRpolyG suppress toxicity whereas inclusion of an AUG codon above the repeat to enhance FMRpolyG production enhances toxicity (Todd et al, 2013). Recently, a transgenic mouse model of FXTAS was developed with 99 repeats upstream of GFP. In mice where the region encompassing the RAN start sites are deleted, survival is comparable to that of wild-type mice (Sellier, Buijsen et al. 2017). However, mice with 99 repeats and the full 5' region show significantly decreased survival as well as accumulation of RAN translation products. These data suggest that RAN translation, but not RNA mediated toxicity is the driving factor in FXTAS. It is likely that not one but all of the proposed mechanisms described above contribute in some way to the toxicity that leads to neuronal death in FXTAS.

uORFs as targets for gene therapy

This diversity of potential disease mechanisms in FXTAS and Fragile X-associated disorders coupled with the importance of the product synthesized from *FMR1* make developing a therapeutic strategy for FXTAS complicated. An ideal therapy would preclude RNA and RAN associated toxicity while sparing or enhancing FMRP synthesis. uORFs serve to regulate protein synthesis, and as such are potential targets for manipulation of the main gene product. 5' leader mutations that create or inhibit uORF function have been described in a number of diseases from Alzheimer's Disease to cancer, and many have been postulated to affect main ORF synthesis (Barbosa, Peixeiro et al. 2013).

Recently, an approach was described that utilized cleavage-resistant antisense oligonucleotides (ASO) to block uORF translation (Liang, Shen et al. 2016, Liang, Sun et al.

2017). In these studies, *RNASEH1* was mainly used as the transcript target. When the ASO was positioned to target the AUG of the uORF with its three most 5' bases, so that the ASO targeted the upstream AUG and the sequence directly upstream of that AUG, the greatest degree of increase in the main ORF was observed in cells. This is independent of mRNA changes. A lesser degree of de-repression was observed when moving the ASO so that it was targeting both regions upstream and downstream of the start site(Liang, Shen et al. 2016). Mismatches introduced into the ASO sequence also limited its function. Though the mechanism by which these ASOs function is still in its early phase of discovery, they are believed to act by interfering with start site selection of the uORF by the PIC leading to resumption of scanning downstream to eventually initiate at the main AUG. Blocking cap-dependent initiation with a small molecule that inhibits the function of eIF4F, prevents the ASO-induced increase in downstream synthesis(Liang, Shen et al. 2016, Liang, Sun et al. 2017). The helicase DHX29 may play a role in resolving the ASO-RNA interaction to allow the PIC to continue scanning downstream of the binding site, perhaps via shunting (Liang, Shen et al. 2016). A separate uORF was targeted on the mouse LRPPRC transcript and injected subcutaneously into mice. Detection of the protein in liver revealed a maximum ~82% increase in the protein, showing that this effect can be detected *in vivo (Liang, Shen et al. 2016)*.

Tiling stable stem loop regions within the 5' leader of *RNASEH1* with a similar ASO approach was also shown to increase downstream protein synthesis. One ASO, 761919, targeted a small stem loop leading to a 2-fold increase in RNase H1 protein(Liang, Sun et al. 2017). This was independent in changes in mRNA steady state levels, but increased the association with *RNASEH1* transcript with polysomes indicative of more active translation. Similarly, an ASO targeting a stem loop in the 5' leader of *ACP1* increased protein production and increased the

association of two cap-dependent initiation factors—eIF4A and eIF2α—with reporter mRNA(Liang, Sun et al. 2017). These findings show that uORFs and inhibitory elements can be selectively targeted as a means of increasing translation of key proteins, and may be critical to developing therapeutics to treat a broad range of disorders.

Perspective and Future Directions

5' leaders play important roles in regulating CDS translation across a variety of model systems and organisms. This regulation must strike a delicate balance between insufficient and overzealous inhibition of downstream initiation and is dependent on multiple factors, including uORF translation, translation initiation factor and helicase availability and RNA structural elements. More recently, a focus on non-AUG initiated translation at repeat expansions and at uORFs has highlighted the potential roles of these processes in both cellular function and human disease. My thesis addresses the relationship between RAN translation, especially at CGG repeats, and the broader field of uORFs. Specifically, I attempt to address the following questions:

1) How common are uORFs in neurons or neuron-like cells?

2) Do uORFs regulate CDS translation during neuronal differentiation? If so, how?

3) Is there a non-pathologic function for CGG repeats and RAN translation?

5) Can we use an ASOs or other methods to selectively block RAN and increase FMRP?

6) Does preventing RAN rescue neurons from RAN-driven toxicity?

In Chapter 2, I apply ribosome profiling to investigate the role of uORFs in a simplified model of neural differentiation. I performed ribosome profiling in both the wild-type and

differentiated cells. I found that uORFs are present on over 30% of transcripts, and that condition-dependent increases in uORF translation are associated with CDS repression. My data points to a mechanism for uORF-mediated inhibition which depends on the degree of translation through the uORF in order to scale its downstream repression. In Chapter 3, I investigate CGG RAN translation in the FMR1 5' leader. I hypothesized that RAN translation regulates FMRP synthesis under conditions of neuronal activity. My data show that RAN translation inhibits FMRP basally, but is needed to allow for activity dependent modulation of FMRP synthesis. In the second half of Chapter 3, I took advantage of the ASO strategy described earlier to impede uORF translation by targeting the near-cognate start sites utilized for RAN translation. I show that targeting RAN in unaffected and patient cells lead to a >50% increase in FMRP abundance, and can rescue neurons from RAN-induced cell death. These data point to an endogenous role for RAN translation in regulating FMRP, and indicate that targeting RAN initiation may be a site for future therapeutic development.

My thesis contributes to the overall knowledge on how 5' leaders can modulate protein synthesis by establishing guiding principles on which uORFs can be repressive to CDS translation, and further clarifying the point that uORFs impart their regulatory role through a mechanism directly linked to their own translation. I then expanded our understanding on RAN translation and its role in balancing FMRP translation. I show that noncanonical uORFs can be taken advantage of as strategy to increase coding sequence translation, and I show that this can be potentially beneficial to the Fragile X Related disorders.

Citations

(1993). "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. The Huntington's Disease Collaborative Research Group." Cell **72**(6): 971-983.

Akbar, U. and T. Ashizawa (2015). "Ataxia." Neurol Clin **33**(1): 225-248.

Andreev, D. E., P. B. O'Connor, C. Fahey, E. M. Kenny, I. M. Terenin, S. E. Dmitriev, P. Cormican, D. W. Morris, I. N. Shatsky and P. V. Baranov (2015). "Translation of 5' leaders is pervasive in genes resistant to eIF2 repression." Elife **4**: e03971.

Andrew, S. E., Y. P. Goldberg, B. Kremer, H. Telenius, J. Theilmann, S. Adam, E. Starr, F. Squitieri, B. Lin, M. A. Kalchman and et al. (1993). "The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease." Nat Genet **4**(4): 398-403.

Archer, S. K., N. E. Shirokikh, T. H. Beilharz and T. Preiss (2016). "Dynamics of ribosome scanning and recycling revealed by translation complex profiling." Nature **535**(7613): 570-574.

Arocena, D. G., C. K. Iwahashi, N. Won, A. Beilina, A. L. Ludwig, F. Tassone, P. H. Schwartz and P. J. Hagerman (2005). "Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells." Hum Mol Genet **14**(23): 3661-3671.

Arribere, J. A. and W. V. Gilbert (2013). "Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing." Genome Res **23**(6): 977-987.

Ash, P. E., K. F. Bieniek, T. F. Gendron, T. Caulfield, W. L. Lin, M. Dejesus-Hernandez, M. M. van Blitterswijk, K. Jansen-West, J. W. Paul, 3rd, R. Rademakers, K. B. Boylan, D. W. Dickson and L. Petrucelli (2013). "Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS." Neuron **77**(4): 639-646.

Avitzour, M., H. Mor-Shaked, S. Yanovsky-Dagan, S. Aharoni, G. Altarescu, P. Renbaum, T. Eldar-Geva, O. Schonberger, E. Levy-Lahad, S. Epsztejn-Litman and R. Eiges (2014). "FMR1 epigenetic silencing commonly occurs in undifferentiated fragile X-affected embryonic stem cells." Stem Cell Reports **3**(5): 699-706.

Bal, N. V., D. Susorov, E. Chesnokova, A. Kasianov, T. Mikhailova, E. Alkalaeva, P. M. Balaban and P. Kolosov (2016). "Upstream Open Reading Frames Located in the Leader of Protein Kinase Mzeta mRNA Regulate Its Translation." Front Mol Neurosci **9**: 103.

Banez-Coronel, M., F. Ayhan, A. D. Tarabochia, T. Zu, B. A. Perez, S. K. Tusi, O. Pletnikova, D. R. Borchelt, C. A. Ross, R. L. Margolis, A. T. Yachnis, J. C. Troncoso and L. P. Ranum (2015). "RAN Translation in Huntington Disease." Neuron **88**(4): 667-677.

Baralle, F. (1977). "Complete nucleotide sequence of the 5' noncoding region of human alpha- and beta-globin mRNA." Cell **12**(4): 1085-1095.

Barbosa, C., I. Peixeiro and L. Romao (2013). "Gene expression regulation by upstream open reading frames and human disease." PLoS Genet **9**(8): e1003529.

Basuta, K., A. Schneider, L. Gane, J. Polussa, B. Woodruff, D. Pretto, R. Hagerman and F. Tassone (2015). "High functioning male with fragile X syndrome and fragile X-associated tremor/ataxia syndrome." Am J Med Genet A **167a**(9): 2154-2161.

Bates, G. (1996). "Expanded glutamines and neurodegeneration--a gain of insight." Bioessays **18**(3): 175-178.

Bazzini, A. A., T. G. Johnstone, R. Christiano, S. D. Mackowiak, B. Obermayer, E. S. Fleming, C. E. Vejnar, M. T. Lee, N. Rajewsky, T. C. Walther and A. J. Giraldez (2014). "Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation." Embo j **33**(9): 981-993.

Bear, M. F., K. M. Huber and S. T. Warren (2004). "The mGluR theory of fragile X mental retardation." Trends Neurosci **27**(7): 370-377.

Beerman, R. W. and T. A. Jongens (2011). "A non-canonical start codon in the Drosophila fragile X gene yields two functional isoforms." Neuroscience **181**: 48-66.

Blair, J. D., D. Hockemeyer, J. A. Doudna, H. S. Bateup and S. N. Floor (2017). "Widespread Translational Remodeling during Human Neuronal Differentiation." Cell Rep **21**(7): 2005-2016.

Blank, H. M., R. Perez, C. He, N. Maitra, R. Metz, J. Hill, Y. Lin, C. D. Johnson, V. A. Bankaitis, B. K. Kennedy, R. Aramayo and M. Polymenis (2017). "Translational control of lipogenic enzymes in the cell cycle of synchronous, growing yeast cells." Embo j **36**(4): 487-502.

Brar, G. A. and J. S. Weissman (2015). "Ribosome profiling reveals the what, when, where and how of protein synthesis." Nat Rev Mol Cell Biol **16**(11): 651-664.

Brar, G. A., M. Yassour, N. Friedman, A. Regev, N. T. Ingolia and J. S. Weissman (2012). "High-resolution view of the yeast meiotic program revealed by ribosome profiling." Science **335**(6068): 552-557.

Bretscher, M. S. (1968). "Ribosomal binding of AUG by formyl-methionyl-sRNA." Nature **220**(5173): 1233-1234.

Brown, V., P. Jin, S. Ceman, J. C. Darnell, W. T. O'Donnell, S. A. Tenenbaum, X. Jin, Y. Feng, K. D. Wilkinson, J. D. Keene, R. B. Darnell and S. T. Warren (2001). "Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome." Cell **107**(4): 477-487.

Buijsen, R. A., J. A. Visser, P. Kramer, E. A. Severijnen, M. Gearing, N. Charlet-Berguerand, S. L. Sherman, R. F. Berman, R. Willemsen and R. K. Hukema (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**(1): 158-168.

Campuzano, V., L. Montermini, M. D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, A. Monticelli, F. Zara, J. Canizares, H. Koutnikova, S. I. Bidichandani, C. Gellera, A. Brice, P. Trouillas, G. De Michele, A. Filla, R. De Frutos, F. Palau, P. I. Patel, S. Di Donato, J. L. Mandel, S. Cocozza, M. Koenig and M. Pandolfo (1996). "Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion." Science **271**(5254): 1423-1427.

Castaldo, I., M. Pinelli, A. Monticelli, F. Acquaviva, M. Giacchetti, A. Filla, S. Sacchetti, S. Keller, V. E. Avvedimento, L. Chiariotti and S. Cocozza (2008). "DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients." J Med Genet **45**(12): 808-812.

Chang, J. C., G. F. Temple, R. Poon, K. H. Neumann and Y. W. Kan (1977). "The nucleotide sequences of the untranslated 5' regions of human alpha- and beta-globin mRNAs." Proc Natl Acad Sci U S A **74**(11): 5145-5149.

Cheng, W., S. Wang, A. A. Mestre and C. Fu (2018). "C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation." **9**(1): 51.

Cheung, Y. N., D. Maag, S. F. Mitchell, C. A. Fekete, M. A. Algire, J. E. Takacs, N. Shirokikh, T. Pestova, J. R. Lorsch and A. G. Hinnebusch (2007). "Dissociation of eIF1 from the 40S ribosomal subunit is a key step in start codon selection in vivo." Genes Dev **21**(10): 1217-1230.

Chew, G. L., A. Pauli and A. F. Schier (2016). "Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish." Nat Commun **7**: 11663.

Colombo, M., E. D. Karousis, J. Bourquin, R. Bruggmann and O. Muhlemann (2017). "Transcriptome-wide identification of NMD-targeted human mRNAs reveals extensive redundancy between SMG6- and SMG7-mediated degradation pathways." Rna **23**(2): 189-201.

Darnell, J. C., S. J. Van Driesche, C. Zhang, K. Y. Hung, A. Mele, C. E. Fraser, E. F. Stone, C. Chen, J. J. Fak, S. W. Chi, D. D. Licatalosi, J. D. Richter and R. B. Darnell (2011). "FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism." Cell **146**(2): 247-261.

DeJesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, A. M. Nicholson, N. A. Finch, H. Flynn, J. Adamson, N. Kouri, A. Wojtas, P. Sengdy, G. Y. Hsiung, A. Karydas, W. W. Seeley, K. A. Josephs, G. Coppola, D. H. Geschwind, Z. K. Wszolek, H. Feldman, D. S. Knopman, R. C. Petersen, B. L. Miller, D. W. Dickson, K. B. Boylan, N. R. Graff-Radford and R. Rademakers (2011). "Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS." Neuron **72**(2): 245-256.

Di Prisco, G. V., W. Huang and S. A. Buffington (2014). "Translational control of mGluR-dependent long-term depression and object-place learning by eIF2alpha." **17**(8): 1073-1082.

Dickey, A. S. and A. R. La Spada (2017). "Therapy development in Huntington disease: From current strategies to emerging opportunities."

Dolen, G., E. Osterweil, B. S. Rao, G. B. Smith, B. D. Auerbach, S. Chattarji and M. F. Bear (2007). "Correction of fragile X syndrome in mice." Neuron **56**(6): 955-962.

Efstratiadis, A., F. C. Kafatos and T. Maniatis (1977). "The primary structure of rabbit beta-globin mRNA as determined from cloned DNA." Cell **10**(4): 571-585.

Eichler, E. E., C. B. Kunst, K. A. Lugenbeel, O. A. Ryder, D. Davison, S. T. Warren and D. L. Nelson (1995). "Evolution of the cryptic FMR1 CGG repeat." Nat Genet **11**(3): 301-308.

Erickson, C. A., M. H. Davenport, T. L. Schaefer, L. K. Wink, E. V. Pedapati, J. A. Sweeney, S. E. Fitzpatrick, W. T. Brown, D. Budimirovic, R. J. Hagerman, D. Hessl, W. E. Kaufmann and E. Berry-Kravis (2017). "Fragile X targeted pharmacotherapy: lessons learned and future directions." J Neurodev Disord **9**: 7.

Fijalkowska, D., S. Verbruggen, E. Ndah, V. Jonckheere, G. Menschaert and P. Van Damme (2017). "eIF1 modulates the recognition of suboptimal translation initiation sites and steers gene expression via uORFs." Nucleic Acids Res **45**(13): 7997-8013.

Gendron, T. F., K. F. Bieniek, Y. J. Zhang, K. Jansen-West, P. E. Ash, T. Caulfield, L. Daughrity, J. H. Dunmore, M. Castanedes-Casey, J. Chew, D. M. Cosio, M. van Blitterswijk, W. C. Lee, R. Rademakers, K. B. Boylan, D. W. Dickson and L. Petrucelli (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**(6): 829-844.

Gerashchenko, M. V., A. V. Lobanov and V. N. Gladyshev (2012). "Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress." Proc Natl Acad Sci U S A **109**(43): 17394-17399.

Gholizadeh, S., S. K. Halder and D. R. Hampson (2015). "Expression of fragile X mental retardation protein in neurons and glia of the developing and adult mouse brain." Brain Res **1596**: 22-30.

Green, K. M., M. R. Glineburg, M. G. Kearse, B. N. Flores, A. E. Linsalata, S. J. Fedak, A. C. Goldstrohm, S. J. Barmada and P. K. Todd (2017). "RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response." Nat Commun **8**(1): 2005.

Green, K. M., A. E. Linsalata and P. K. Todd (2016). "RAN translation-What makes it run?" Brain Res **1647**: 30-42.

Grima, J. C., J. G. Daigle, N. Arbez, K. C. Cunningham, K. Zhang, J. Ochaba, C. Geater, E. Morozko, J. Stocksdale, J. C. Glatzer, J. T. Pham, I. Ahmed, Q. Peng, H. Wadhwa, O. Pletnikova, J. C. Troncoso, W. Duan, S. H. Snyder, L. P. W. Ranum, L. M. Thompson, T. E. Lloyd, C. A. Ross and J. D. Rothstein (2017). "Mutant Huntingtin Disrupts the Nuclear Pore Complex." Neuron **94**(1): 93-107.e106.

Groh, M., M. M. Lufino, R. Wade-Martins and N. Gromak (2014). "R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome." PLoS Genet **10**(5): e1004318.

Gross, C., C. W. Chang, S. M. Kelly, A. Bhattacharya, S. M. McBride, S. W. Danielson, M. Q. Jiang, C. B. Chan, K. Ye, J. R. Gibson, E. Klann, T. A. Jongens, K. H. Moberg, K. M. Huber and G. J. Bassell (2015). "Increased expression of the PI3K enhancer PIKE mediates deficits in synaptic plasticity and behavior in fragile X syndrome." Cell Rep **11**(5): 727-736.

Gross, C., N. Raj, G. Molinaro, A. G. Allen, A. J. Whyte, J. R. Gibson, K. M. Huber, S. L. Gourley and G. J. Bassell (2015). "Selective role of the catalytic PI3K subunit p110beta in impaired higher order cognition in fragile X syndrome." Cell Rep **11**(5): 681-688.

Hagerman, P. (2013). "Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms." Acta Neuropathol **126**(1): 1-19.

Hagerman, R. J., E. Berry-Kravis, H. C. Hazlett, D. B. Bailey, Jr., H. Moine, R. F. Kooy, F. Tassone, I. Gantois, N. Sonenberg, J. L. Mandel and P. J. Hagerman (2017). "Fragile X syndrome." Nat Rev Dis Primers **3**: 17065.

Hagerman, R. J. and P. Hagerman (2016). "Fragile X-associated tremor/ataxia syndrome - features, mechanisms and management." Nat Rev Neurol **12**(7): 403-412.
Harding, H. P., I. Novoa, Y. Zhang, H. Zeng, R. Wek, M. Schapira and D. Ron (2000). "Regulated translation initiation controls stress-induced gene expression in mammalian cells." Mol Cell **6**(5): 1099-1108.

Harding, H. P., Y. Zhang, H. Zeng, I. Novoa, P. D. Lu, M. Calfon, N. Sadri, C. Yun, B. Popko, R. Paules, D. F. Stojdl, J. C. Bell, T. Hettmann, J. M. Leiden and D. Ron (2003). "An integrated stress response regulates amino acid metabolism and resistance to oxidative stress." Mol Cell **11**(3): 619-633.

Hashem, V., J. N. Galloway, M. Mori, R. Willemsen, B. A. Oostra, R. Paylor and D. L. Nelson (2009). "Ectopic expression of CGG containing mRNA is neurotoxic in mammals." Hum Mol Genet **18**(13): 2443-2451.

Hashimoto, R., A. K. Javan, F. Tassone, R. J. Hagerman and S. M. Rivera (2011). "A voxel-based morphometry study of grey matter loss in fragile X-associated tremor/ataxia syndrome." Brain **134**(Pt 3): 863-878.

Hashimoto, R., S. Srivastava, F. Tassone, R. J. Hagerman and S. M. Rivera (2011). "Diffusion tensor imaging in male premutation carriers of the fragile X mental retardation gene." Mov Disord **26**(7): 1329-1336.

Hill, J. R. and D. R. Morris (1993). "Cell-specific translational regulation of S-adenosylmethionine decarboxylase mRNA. Dependence on translation and coding capacity of the cis-acting upstream open reading frame." J Biol Chem **268**(1): 726-731.

Hinnebusch, A. G. (1984). "Evidence for translational regulation of the activator of general amino acid control in yeast." Proc Natl Acad Sci U S A **81**(20): 6442-6446.

Hinnebusch, A. G. (1997). "Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome." J Biol Chem **272**(35): 21661-21664.

Hinnebusch, A. G. (2011). "Molecular mechanism of scanning and start codon selection in eukaryotes." Microbiol Mol Biol Rev **75**(3): 434-467, first page of table of contents.

Hinnebusch, A. G., I. P. Ivanov and N. Sonenberg (2016). "Translational control by 5'-untranslated regions of eukaryotic mRNAs." Science **352**(6292): 1413-1416.

Hou, L., M. D. Antion, D. Hu, C. M. Spencer, R. Paylor and E. Klann (2006). "Dynamic translational and proteasomal regulation of fragile X mental retardation protein controls mGluR-dependent long-term depression." Neuron **51**(4): 441-454.

Huang, L., C. H. Lou, W. Chan, E. Y. Shum, A. Shao, E. Stone, R. Karam, H. W. Song and M. F. Wilkinson (2011). "RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD." Mol Cell **43**(6): 950-961.

Huang, W., A. N. Placzek, G. Viana Di Prisco, S. Khatiwada, C. Sidrauski, K. Krnjevic, P. Walter, J. A. Dani and M. Costa-Mattioli (2016). "Translational control by eIF2alpha phosphorylation regulates vulnerability to the synaptic and behavioral effects of cocaine." **5**.

Huber, K. M., S. M. Gallagher, S. T. Warren and M. F. Bear (2002). "Altered synaptic plasticity in a mouse model of fragile X mental retardation." Proc Natl Acad Sci U S A **99**(11): 7746-7750.

Hurt, J. A., A. D. Robertson and C. B. Burge (2013). "Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay." Genome Res **23**(10): 1636-1650.

Ingolia, N. T. (2010). "Genome-wide translational profiling by ribosome footprinting." Methods Enzymol **470**: 119-142.

Ingolia, N. T., G. A. Brar, S. Rouskin, A. M. McGeachy and J. S. Weissman (2012). "The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments." Nat Protoc **7**(8): 1534-1550.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman and J. S. Weissman (2009). "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling." Science **324**(5924): 218-223.

Ingolia, N. T., L. F. Lareau and J. S. Weissman (2011). "Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes." Cell **147**(4): 789-802.

Irwin, S. A., B. Patel, M. Idupulapati, J. B. Harris, R. A. Crisostomo, B. P. Larsen, F. Kooy, P. J. Willems, P. Cras, P. B. Kozlowski, R. A. Swain, I. J. Weiler and W. T. Greenough (2001). "Abnormal dendritic spine characteristics in the temporal and visual cortices of patients with fragile-X syndrome: a quantitative examination." Am J Med Genet **98**(2): 161-167.

Ishigaki, Y., X. Li, G. Serin and L. E. Maquat (2001). "Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20." Cell **106**(5): 607-617.

Ivanov, I. P., A. E. Firth, A. M. Michel, J. F. Atkins and P. V. Baranov (2011). "Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences." Nucleic Acids Res **39**(10): 4220-4234.

Iwahashi, C. K., D. H. Yasui, H. J. An, C. M. Greco, F. Tassone, K. Nannen, B. Babineau, C. B. Lebrilla, R. J. Hagerman and P. J. Hagerman (2006). "Protein composition of the intranuclear inclusions of FXTAS." Brain **129**(Pt 1): 256-271.

Jackson, S. P. and J. Bartek (2009). "The DNA-damage response in human biology and disease." Nature **461**(7267): 1071-1078.

Jacquemont, S., S. Birnbaum, S. Redler, P. Steinbach and V. Biancalana (2011). "Clinical utility gene card for: fragile X mental retardation syndrome, fragile X-associated tremor/ataxia syndrome and fragile X-associated primary ovarian insufficiency." Eur J Hum Genet **19**(9).

Janich, P., A. B. Arpat, V. Castelo-Szekely, M. Lopes and D. Gatfield (2015). "Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames." Genome Res **25**(12): 1848-1859.

Jiang, H., A. Mankodi, M. S. Swanson, R. T. Moxley and C. A. Thornton (2004). "Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons." Hum Mol Genet **13**(24): 3079-3088.

Jin, P., R. Duan, A. Qurashi, Y. Qin, D. Tian, T. C. Rosser, H. Liu, Y. Feng and S. T. Warren (2007). "Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome." Neuron **55**(4): 556-564.

Johnstone, T. G., A. A. Bazzini and A. J. Giraldez (2016). "Upstream ORFs are prevalent translational repressors in vertebrates." Embo j **35**(7): 706-723.

Jousse, C., A. Bruhat, V. Carraro, F. Urano, M. Ferrara, D. Ron and P. Fafournoux (2001). "Inhibition of CHOP translation by a peptide encoded by an open reading frame localized in the chop 5'UTR." Nucleic Acids Res **29**(21): 4341-4351.

Kanadia, R. N., J. Shin, Y. Yuan, S. G. Beattie, T. M. Wheeler, C. A. Thornton and M. S. Swanson (2006). "Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy." Proc Natl Acad Sci U S A **103**(31): 11748-11753.

Kearse, M. G., K. M. Green, A. Krans, C. M. Rodriguez, A. E. Linsalata, A. C. Goldstrohm and P. K. Todd (2016). "CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins." Mol Cell **62**(2): 314-322.

Kearse, M. G. and P. K. Todd (2014). "Repeat-associated non-AUG translation and its impact in neurodegenerative disease." Neurotherapeutics **11**(4): 721-731.

Kearse, M. G. and J. E. Wilusz (2017). "Non-AUG translation: a new start for protein synthesis in eukaryotes." Genes Dev **31**(17): 1717-1731.

Kim, M. S., S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. A. Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T. C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D.

Leach, C. G. Drake, M. K. Halushka, T. S. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. A. Iacobuzio-Donahue, H. Gowda and A. Pandey (2014). "A draft map of the human proteome." Nature **509**(7502): 575-581.

Kozak, M. (1980). "Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits." Cell **19**(1): 79-90.

Kozak, M. (1981). "Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes." Nucleic Acids Res **9**(20): 5233-5252.

Kozak, M. (1984). "Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs." Nucleic Acids Res **12**(2): 857-872.

Kozak, M. (1984). "Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequence for preproinsulin." Nucleic Acids Res **12**(9): 3873-3893.

Kozak, M. (1986). "Influences of mRNA secondary structure on initiation by eukaryotic ribosomes." Proc Natl Acad Sci U S A **83**(9): 2850-2854.

Kozak, M. (1986). "Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes." Cell **44**(2): 283-292.

Kozak, M. (1987). "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs." Nucleic Acids Res **15**(20): 8125-8148.

Kozak, M. (1987). "Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes." Mol Cell Biol **7**(10): 3438-3445.

Kozak, M. (1988). "Leader length and secondary structure modulate mRNA function under conditions of stress." Mol Cell Biol **8**(7): 2737-2744.

Kozak, M. (1989). "Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems." Mol Cell Biol **9**(11): 5073-5080.

Kozak, M. (1990). "Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes." Proc Natl Acad Sci U S A **87**(21): 8301-8305.

Kozak, M. (2002). "Pushing the limits of the scanning mechanism for initiation of translation." Gene **299**(1-2): 1-34.

Krans, A., M. G. Kearse and P. K. Todd (2016). "Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome." Ann Neurol **80**(6): 871-881.

La Spada, A. R., E. M. Wilson, D. B. Lubahn, A. E. Harding and K. H. Fischbeck (1991). "Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy." Nature **352**(6330): 77-79.

Law, G. L., A. Raney, C. Heusner and D. R. Morris (2001). "Polyamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase." J Biol Chem **276**(41): 38036-38043.

Le Hir, H., E. Izaurralde, L. E. Maquat and M. J. Moore (2000). "The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions." Embo j **19**(24): 6860-6869.

Lee, Y. Y., R. C. Cevallos and E. Jan (2009). "An upstream open reading frame regulates translation of GADD34 during cellular stresses that induce eIF2alpha phosphorylation." J Biol Chem **284**(11): 6661-6673.

Lejeune, F., X. Li and L. E. Maquat (2003). "Nonsense-mediated mRNA decay in mammalian cells involves decapping, deadenylating, and exonucleolytic activities." Mol Cell **12**(3): 675-687.

Liang, H., S. He, J. Yang, X. Jia, P. Wang, X. Chen, Z. Zhang, X. Zou, M. A. McNutt, W. H. Shen and Y. Yin (2014). "PTENalpha, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism." Cell Metab **19**(5): 836-848.

Liang, X. H., W. Shen, H. Sun, M. T. Migawa, T. A. Vickers and S. T. Crooke (2016). "Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames." Nat Biotechnol **34**(8): 875-880.

Liang, X. H., H. Sun, W. Shen, S. Wang, J. Yao, M. T. Migawa, H. H. Bui, S. S. Damle, S. Riney, M. J. Graham, R. M. Crooke and S. T. Crooke (2017). "Antisense oligonucleotides targeting translation inhibitory elements in 5' UTRs can selectively increase protein levels." Nucleic Acids Res **45**(16): 9528-9546.

Ling, D. S., L. S. Benardo, P. A. Serrano, N. Blace, M. T. Kelly, J. F. Crary and T. C. Sacktor (2002). "Protein kinase Mzeta is necessary and sufficient for LTP maintenance." Nat Neurosci **5**(4): 295-296.

Loomis, E. W., L. A. Sanz, F. Chedin and P. J. Hagerman (2014). "Transcription-associated R-loop formation across the human FMR1 CGG-repeat region." PLoS Genet **10**(4): e1004294.

Lopez Castel, A., J. D. Cleary and C. E. Pearson (2010). "Repeat instability as the basis for human diseases and as a potential target for therapy." Nat Rev Mol Cell Biol **11**(3): 165-170.

Loughran, G., M. S. Sachs, J. F. Atkins and I. P. Ivanov (2012). "Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5." Nucleic Acids Res **40**(7): 2898-2906.

Lu, C., L. Lin, H. Tan, H. Wu, S. L. Sherman, F. Gao, P. Jin and D. Chen (2012). "Fragile X premutation RNA is sufficient to cause primary ovarian insufficiency in mice." Hum Mol Genet **21**(23): 5039-5047.

Maag, D., C. A. Fekete, Z. Gryczynski and J. R. Lorsch (2005). "A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon." Mol Cell **17**(2): 265-275.

Mankodi, A., C. R. Urbinati, Q. P. Yuan, R. T. Moxley, V. Sansone, M. Krym, D. Henderson, M. Schalling, M. S. Swanson and C. A. Thornton (2001). "Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2." Hum Mol Genet **10**(19): 2165-2170.

Martin-Marcos, P., Y. N. Cheung and A. G. Hinnebusch (2011). "Functional elements in initiation factors 1, 1A, and 2beta discriminate against poor AUG context and non-AUG start codons." Mol Cell Biol **31**(23): 4814-4831.

McMurray, C. T. (2010). "Mechanisms of trinucleotide repeat instability during human development." Nat Rev Genet **11**(11): 786-799.

Michalon, A., M. Sidorov, T. M. Ballard, L. Ozmen, W. Spooren, J. G. Wettstein, G. Jaeschke, M. F. Bear and L. Lindemann (2012). "Chronic pharmacological mGlu5 inhibition corrects fragile X in adult mice." Neuron **74**(1): 49-56.

Miller, J. W., C. R. Urbinati, P. Teng-Umnuay, M. G. Stenberg, B. J. Byrne, C. A. Thornton and M. S. Swanson (2000). "Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy." Embo j **19**(17): 4439-4448.

Mori, K., S. M. Weng, T. Arzberger, S. May, K. Rentzsch, E. Kremmer, B. Schmid, H. A. Kretzschmar, M. Cruts, C. Van Broeckhoven, C. Haass and D. Edbauer (2013). "The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS." Science **339**(6125): 1335-1338.

Mueller, P. P. and A. G. Hinnebusch (1986). "Multiple upstream AUG codons mediate translational control of GCN4." Cell **45**(2): 201-207.

Na, C. H., M. A. Barbhuiya, M. S. Kim, S. Verbruggen, S. M. Eacker, O. Pletnikova, J. C. Troncoso, M. K. Halushka, G. Menschaert, C. M. Overall and A. Pandey (2017). "Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini." Genome Res.

Nalavadi, V. C., R. S. Muddashetty, C. Gross and G. J. Bassell (2012). "Dephosphorylation-induced ubiquitination and degradation of FMRP in dendrites: a role in immediate early mGluR-stimulated translation." J Neurosci **32**(8): 2582-2587.

Nanda, J. S., Y. N. Cheung, J. E. Takacs, P. Martin-Marcos, A. K. Saini, A. G. Hinnebusch and J. R. Lorsch (2009). "eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation." J Mol Biol **394**(2): 268-285.

Nimchinsky, E. A., A. M. Oberlander and K. Svoboda (2001). "Abnormal development of dendritic spines in FMR1 knock-out mice." J Neurosci **21**(14): 5139-5146.

Nirenberg, M. and P. Leder (1964). "RNA CODEWORDS AND PROTEIN SYNTHESIS. THE EFFECT OF TRINUCLEOTIDES UPON THE BINDING OF SRNA TO RIBOSOMES." Science **145**(3639): 1399-1407.

Nolin, S. L., A. Glicksman, G. E. Houck, Jr., W. T. Brown and C. S. Dobkin (1994). "Mosaicism in fragile X affected males." Am J Med Genet **51**(4): 509-512.

Oh, S. Y., F. He, A. Krans, M. Frazer, J. P. Taylor, H. L. Paulson and P. K. Todd (2015). "RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome." Hum Mol Genet **24**(15): 4317-4326.

Pakos-Zebrucka, K., I. Koryga, K. Mnich, M. Ljujic and A. Samali (2016). "The integrated stress response." **17**(10): 1374-1395.

Palam, L. R., T. D. Baird and R. C. Wek (2011). "Phosphorylation of eIF2 facilitates ribosomal bypass of an inhibitory upstream ORF to enhance CHOP translation." J Biol Chem **286**(13): 10939-10949.

Parola, A. L. and B. K. Kobilka (1994). "The peptide product of a 5' leader cistron in the beta 2 adrenergic receptor mRNA inhibits receptor synthesis." J Biol Chem **269**(6): 4497-4505.

Passmore, L. A., T. M. Schmeing, D. Maag, D. J. Applefield, M. G. Acker, M. A. Algire, J. R. Lorsch and V. Ramakrishnan (2007). "The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome." Mol Cell **26**(1): 41-50.

Paulson, H. (2018). "Repeat expansion diseases." Handb Clin Neurol **147**: 105-123.

Pelechano, V., W. Wei and L. M. Steinmetz (2013). "Extensive transcriptional heterogeneity revealed by isoform profiling." Nature **497**(7447): 127-131.

Pettersson, O. J., L. Aagaard, T. G. Jensen and C. K. Damgaard (2015). "Molecular mechanisms in DM1 - a focus on foci." Nucleic Acids Res **43**(4): 2433-2441.

Popp, M. W. and L. E. Maquat (2013). "Organizing principles of mammalian nonsense-mediated mRNA decay." Annu Rev Genet **47**: 139-165.

Pretto, D., C. M. Yrigollen, H. T. Tang, J. Williamson, G. Espinal, C. K. Iwahashi, B. Durbin-Johnson, R. J. Hagerman, P. J. Hagerman and F. Tassone (2014). "Clinical and molecular implications of mosaicism in FMR1 full mutations." Front Genet **5**: 318.

Rabadan-Diehl, C., A. Martinez, S. Volpi, S. Subburaju and G. Aguilera (2007). "Inhibition of vasopressin V1b receptor translation by upstream open reading frames in the 5'-untranslated region." J Neuroendocrinol **19**(4): 309-319.

Rass, U., I. Ahel and S. C. West (2007). "Defective DNA repair and neurodegenerative disease." Cell **130**(6): 991-1004.

Reddy, K., M. H. Schmidt, J. M. Geist, N. P. Thakkar, G. B. Panigrahi, Y. H. Wang and C. E. Pearson (2014). "Processing of double-R-loops in (CAG).(CTG) and C9orf72 (GGGGCC).(GGCCCC) repeats causes instability." Nucleic Acids Res **42**(16): 10473-10487.

Renoux, A. J. and P. K. Todd (2012). "Neurodegeneration the RNA way." Prog Neurobiol **97**(2): 173-189.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas, H. Kalimo, A. Paetau, Y. Abramzon, A. M. Remes, A. Kaganovich, S. W. Scholz, J. Duckworth, J. Ding, D. W. Harmer, D. G. Hernandez, J. O. Johnson, K. Mok, M. Ryten, D. Trabzuni, R. J. Guerreiro, R. W. Orrell, J. Neal, A. Murray, J. Pearson, I. E. Jansen, D. Sondervan, H. Seelaar, D. Blake, K. Young, N. Halliwell, J. B. Callister, G. Toulson, A. Richardson, A. Gerhard, J. Snowden, D. Mann, D. Neary, M. A. Nalls, T. Peuralinna, L. Jansson, V. M. Isoviita, A. L. Kaivorinne, M. Holtta-Vuori, E. Ikonen, R. Sulkava, M. Benatar, J. Wuu, A. Chio, G. Restagno, G. Borghero, M. Sabatelli, D. Heckerman, E. Rogaeva, L. Zinman, J. D. Rothstein, M. Sendtner, C. Drepper, E. E. Eichler, C. Alkan, Z. Abdullaev, S. D. Pack, A. Dutra, E. Pak, J. Hardy, A. Singleton, N. M. Williams, P. Heutink, S. Pickering-Brown, H. R. Morris, P. J. Tienari and B. J. Traynor (2011). "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD." Neuron **72**(2): 257-268.

Richter, J. D., G. J. Bassell and E. Klann (2015). "Dysregulation and restoration of translational homeostasis in fragile X syndrome." Nat Rev Neurosci **16**(10): 595-605.

Santa Maria, L., A. Pugin, M. A. Alliende, S. Aliaga, B. Curotto, T. Aravena, H. T. Tang, G. Mendoza-Morales, R. Hagerman and F. Tassone (2014). "FXTAS in an unmethylated mosaic male with fragile X syndrome from Chile." Clin Genet **86**(4): 378-382.

Santini, E. and E. Klann (2014). "Reciprocal signaling between translational control pathways and synaptic proteins in autism spectrum disorders." Sci Signal **7**(349): re10.

Sciascia, N. and K. Usdin Mol Autism.

Sellier, C., R. A. M. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent, M. F. Champy, T. Sorg, G. Pavlovic, M. Wattenhofer-Donze, M. C. Birling, M. Oulad-Abdelghani, P. Eberling, F. Ruffenach, M. Joint, M. Anheim, V. Martinez-

Cerdeno, F. Tassone, R. Willemsen, R. K. Hukema, S. Viville, C. Martinat, P. K. Todd and N. Charlet-Berguerand (2017). "Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome." Neuron **93**(2): 331-347.
Sellier, C., F. Freyermuth, R. Tabet, T. Tran, F. He, F. Ruffenach, V. Alunni, H. Moine, C. Thibault, A. Page, F. Tassone, R. Willemsen, M. D. Disney, P. J. Hagerman, P. K. Todd and N. Charlet-Berguerand (2013). "Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome." Cell Rep **3**(3): 869-880.

Sellier, C., F. Rau, Y. Liu, F. Tassone, R. K. Hukema, R. Gattoni, A. Schneider, S. Richard, R. Willemsen, D. J. Elliott, P. J. Hagerman and N. Charlet-Berguerand (2010). "Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients." Embo j **29**(7): 1248-1261.

Sendoel, A., J. G. Dunn, E. H. Rodriguez, S. Naik, N. C. Gomez, B. Hurwitz, J. Levorse, B. D. Dill, D. Schramek, H. Molina, J. S. Weissman and E. Fuchs (2017). "Translation from unconventional 5' start sites drives tumour initiation." Nature **541**(7638): 494-499.

Sherman, F., J. W. Stewart and A. M. Schweingruber (1980). "Mutants of yeast initiating translation of iso-1-cytochrome c within a region spanning 37 nucleotides." Cell **20**(1): 215-222.

Shirokikh, N. E., S. K. Archer, T. H. Beilharz, D. Powell and T. Preiss (2017). "Translation complex profile sequencing to study the in vivo dynamics of mRNA-ribosome interactions during translation initiation, elongation and termination." **12**(4): 697-731.

Sidrauski, C., A. M. McGeachy, N. T. Ingolia and P. Walter (2015). "The small molecule ISRIB reverses the effects of eIF2alpha phosphorylation on translation and stress granule assembly." Elife **4**.

Slavoff, S. A., A. J. Mitchell, A. G. Schwaid, M. N. Cabili, J. Ma, J. Z. Levin, A. D. Karger, B. A. Budnik, J. L. Rinn and A. Saghatelian (2013). "Peptidomic discovery of short open reading frame-encoded peptides in human cells." Nat Chem Biol **9**(1): 59-64.

Sofola, O. A., P. Jin, Y. Qin, R. Duan, H. Liu, M. de Haro, D. L. Nelson and J. Botas (2007). "RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS." Neuron **55**(4): 565-571.

Spealman, P., A. Naik, G. May, S. Kuersten, L. Freebert, R. Murphy and J. McManus (2017). "Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data." Genome Res.

Starck, S. R., V. Jiang, M. Pavon-Eternod, S. Prasad, B. McCarthy, T. Pan and N. Shastri (2012). "Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I." Science **336**(6089): 1719-1723.

Starck, S. R., J. C. Tsai, K. Chen, M. Shodiya, L. Wang, K. Yahiro, M. Martins-Green, N. Shastri and P. Walter (2016). "Translation from the 5' untranslated region shapes the integrated stress response." Science **351**(6272): aad3867.

Steitz, J. A. (1969). "Polypeptide chain initiation: nucleotide sequences of the three ribosomal binding sites in bacteriophage R17 RNA." Nature **224**(5223): 957-964.

Sundararajan, T. A. and R. E. Thach (1966). "Role of the formylmethionine codon AUG in phasing translation of synthetic messenger RNA." J Mol Biol **19**(1): 74-90.

Tabet, R., L. Schaeffer, F. Freyermuth, M. Jambeau, M. Workman, C. Z. Lee, C. C. Lin, J. Jiang, K. Jansen-West, H. Abou-Hamdan, L. Desaubry, T. Gendron, L. Petrucelli and F. Martin (2018). "CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts." **9**(1): 152.

Tassone, F., R. J. Hagerman, A. K. Taylor, L. W. Gane, T. E. Godfrey and P. J. Hagerman (2000). "Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome." Am J Hum Genet **66**(1): 6-15.

Thach, R. E., K. F. Dewey, J. C. Brown and P. Doty (1966). "Formylmethionine codon AUG as an initiator of polypeptide synthesis." Science **153**(3734): 416-418.

Thireos, G., M. D. Penn and H. Greer (1984). "5' untranslated sequences are required for the translational control of a yeast regulatory gene." Proc Natl Acad Sci U S A **81**(16): 5096-5100.

Todd, P. K., K. J. Mack and J. S. Malter (2003). "The fragile X mental retardation protein is required for type-I metabotropic glutamate receptor-dependent translation of PSD-95." Proc Natl Acad Sci U S A **100**(24): 14374-14378.

Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur, K. Elenitoba-Johnson, J. P. Vonsattel, E. D. Louis, M. A. Sutton, J. P. Taylor, R. E. Mills, N. Charlet-Berguerand and H. L. Paulson (2013). "CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome." Neuron **78**(3): 440-455.

Todd, P. K. and H. L. Paulson (2010). "RNA-mediated neurodegeneration in repeat expansion disorders." Ann Neurol **67**(3): 291-300.

Trottier, Y., D. Devys, G. Imbert, F. Saudou, I. An, Y. Lutz, C. Weber, Y. Agid, E. C. Hirsch and J. L. Mandel (1995). "Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form." Nat Genet **10**(1): 104-110.

Trottier, Y., Y. Lutz, G. Stevanin, G. Imbert, D. Devys, G. Cancel, F. Saudou, C. Weber, G. David, L. Tora and et al. (1995). "Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias." Nature **378**(6555): 403-406.

Tzamarias, D., D. Alexandraki and G. Thireos (1986). "Multiple cis-acting elements modulate the translational efficiency of GCN4 mRNA in yeast." Proc Natl Acad Sci U S A **83**(13): 4849-4853.

Unbehaun, A., S. I. Borukhov, C. U. Hellen and T. V. Pestova (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**(24): 3078-3093.

Valasek, L., K. H. Nielsen, F. Zhang, C. A. Fekete and A. G. Hinnebusch (2004). "Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection." Mol Cell Biol **24**(21): 9437-9455.
Vattem, K. M. and R. C. Wek (2004). "Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells." Proc Natl Acad Sci U S A **101**(31): 11269-11274.

Wang, L. W., E. Berry-Kravis and R. J. Hagerman (2010). "Fragile X: leading the way for targeted treatments in autism." Neurotherapeutics **7**(3): 264-274.

Weiler, I. J., S. A. Irwin, A. Y. Klintsova, C. M. Spencer, A. D. Brazelton, K. Miyashiro, T. A. Comery, B. Patel, J. Eberwine and W. T. Greenough (1997). "Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation." Proc Natl Acad Sci U S A **94**(10): 5395-5400.

Wen, X., W. Tan, T. Westergard, K. Krishnamurthy, S. S. Markandaiah, Y. Shi, S. Lin, N. A. Shneider, J. Monaghan, U. B. Pandey, P. Pasinelli, J. K. Ichida and D. Trotti (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**(6): 1213-1225.

Wheeler, A. C., D. B. Bailey, Jr., E. Berry-Kravis, J. Greenberg, M. Losh, M. Mailick, M. Mila, J. M. Olichney, L. Rodriguez-Revenga, S. Sherman, L. Smith, S. Summers, J. C. Yang and R. Hagerman (2014). "Associated features in females with an FMR1 premutation." J Neurodev Disord **6**(1): 30.

Wohrle, D., U. Salat, D. Glaser, J. Mucke, M. Meisel-Stosiek, D. Schindler, W. Vogel and P. Steinbach (1998). "Unusual mutations in high functioning fragile X males: apparent instability of expanded unmethylated CGG repeats." J Med Genet **35**(2): 103-111.

Yepiskoposyan, H., F. Aeschimann, D. Nilsson, M. Okoniewski and O. Muhlemann (2011). "Autoregulation of the nonsense-mediated mRNA decay pathway in human cells." Rna **17**(12): 2108-2118.

Young, S. K. and R. C. Wek (2016). "Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response." J Biol Chem **291**(33): 16927-16935.

Zhang, F. and A. G. Hinnebusch (2011). "An upstream ORF with non-AUG start codon is translated in vivo but dispensable for translational control of GCN4 mRNA." Nucleic Acids Res **39**(8): 3128-3140.

Zhang, K., C. J. Donnelly, A. R. Haeusler, J. C. Grima, J. B. Machamer, P. Steinwald, E. L. Daley, S. J. Miller, K. M. Cunningham, S. Vidensky, S. Gupta, M. A. Thomas, I. Hong, S. L. Chiu, R. L. Huganir, L. W. Ostrow, M. J. Matunis, J. Wang, R. Sattler, T. E. Lloyd and J. D. Rothstein (2015). "The C9orf72 repeat expansion disrupts nucleocytoplasmic transport." Nature **525**(7567): 56-61.

Zhou, D., L. R. Palam, L. Jiang, J. Narasimhan, K. A. Staschke and R. C. Wek (2008). "Phosphorylation of eIF2 directs ATF5 translational control in response to diverse stress conditions." J Biol Chem **283**(11): 7064-7073.

Zu, T., B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski, M. A. Ingram, Z. Nan, C. Forster, W. C. Low, B. Schoser, N. V. Somia, H. B. Clark, S. Schmechel, P. B. Bitterman, G. Gourdon, M. S. Swanson, M. Moseley and L. P. Ranum (2011). "Non-ATG-initiated translation directed by microsatellite expansions." Proc Natl Acad Sci U S A **108**(1): 260-265.

Zu, T., Y. Liu, M. Banez-Coronel, T. Reid, O. Pletnikova, J. Lewis, T. M. Miller, M. B. Harms, A. E. Falchook, S. H. Subramony, L. W. Ostrow, J. D. Rothstein, J. C. Troncoso and L. P. Ranum (2013). "RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia." Proc Natl Acad Sci U S A **110**(51): E4968-4977.

CHAPTER II

Conditional translation of upstream open reading frames in a model of neuronal differentiation[1]

Abstract

Upstream open reading frames (uORFs) initiate translation within mRNA 5' leaders. uORFs can both generate unannotated proteins and impact main coding sequence (CDS) translation from the transcripts in which they reside. Ribosome profiling (RP) studies suggest that translating ribosomes are pervasive within 5' leaders across model systems. However, the significance of this observation remains unclear. To explore a role for uORF usage in neuronal differentiation, we performed RP on undifferentiated and differentiated neuroblastoma cells. Using a spectral coherence algorithm (SPECtre), we identify 3,508 uORFs in 31% of neuroblastoma transcripts. These uORFs predominantly utilize non-AUG initiation codons and exhibit translational efficiencies comparable to annotated coding regions. Among these, a highly conserved and consistently translated subset of uORFs is associated with repressed CDS translation. Intriguingly, these uORFs demonstrate directional changes in translation with differentiation that inversely correlate with cell-state shifts in CDS translation. These findings suggest that uORFs can provide a robust and state-dependent mechanism for CDS translation regulation during neuronal differentiation.

[1] This chapter will be submitted for publication as: Rodriguez, C.M., Chun, S.Y., Mills, R.E., Todd, P.K. Conditional translation of upstream open reading frames in a model of neuronal differentiation.

Introduction

Alterations in protein expression and abundance are required for successful and stable cellular differentiation (Hershey, Sonenberg et al. 2012). While changes in mRNA levels provide a partial view of networks driving such cellular changes, differences in translational efficiency act as an independent contributor to this process (Brar 2016). Determining ribosomal occupancy across the transcriptome through ribosomal profiling (RP) provides is a powerful tool for assessing the relationship between mRNA abundance and translational output (Ingolia, Ghaemmaghami et al. 2009). In particular, RP in cells and organisms has revealed a detailed picture of condition-specific changes in mRNA translation rates in multiple cellular processes from meiosis to development (Ingolia, Lareau et al. 2011, Brar, Yassour et al. 2012).

The 5' leader (traditionally referred to as the 5' UTR) of mRNAs are one well-studied source of protein synthesis regulation (Kozak 1991, Calvo, Pagliarini et al. 2009, Sonenberg and Hinnebusch 2009, Hinnebusch, Ivanov et al. 2016). 5' leaders can regulate the synthesis of the main coding sequence (CDS) product through a variety of mechanisms (Sonenberg and Hinnebusch 2009, Hinnebusch, Ivanov et al. 2016). RNA secondary structures can impede ribosomal scanning, which decreases access of assembled 40S ribosomal preinitiation complexes to CDS initiation sites. Translation can also initiate within 5' leaders at upstream open reading frames (uORFS). In the case of uORFs that terminate after the CDS initiation site (overlapping uORFs or "oORFs"), initiation in the 5' leader directly competes with CDS initiation for scanning 40S ribosomes and is thus predicted to be inhibitory on CDS translation. In contrast, uORFs that terminate within the 5' leader and before the CDS initiation site (contained uORFs or "cORFs"), ribosomes can potentially reinitiate at the CDS. Thus, cORFs sometimes bypass other 5'leader regulatory elements and can even provide stimulatory effects on CDS translation, but

may be repressive as well. uORF translation can also indirectly influence CDS translation by influencing mRNA stability (Rebbapragada and Lykke-Andersen 2009) or through interactions of newly synthesized uORF protein products with the translating ribosome (Parola and Kobilka 1994, Rabadan-Diehl, Martinez et al. 2007). As such, the relationship of each uORF to the translation of its cognate CDS can be complex, which in turn makes it difficult to define their specific functions and regulation across the transcriptome based on position alone.

Early ribosome profiling reports demonstrated the surprising finding that ribosome protected fragments (RPFs) are highly enriched within 5' leader regions of mRNAs (Ingolia, Ghaemmaghami et al. 2009, Ingolia, Lareau et al. 2011, Brar, Yassour et al. 2012). Since these first reports, there have been several studies that investigated 5' leader translation (Bazzini, Johnstone et al. 2014, Ingolia, Brar et al. 2014, Crappe, Ndah et al. 2015, Fields, Rodriguez et al. 2015, Calviello, Mukherjee et al. 2016, Chun, Rodriguez et al. 2016). These studies revealed potential roles for uORFs in circadian clock regulation, organism development, and the cell cycle (Brar, Yassour et al. 2012, Janich, Arpat et al. 2015, Johnstone, Bazzini et al. 2016, Blank, Perez et al. 2017, Fujii, Shi et al. 2017). For example, AUG initiated uORFs were detected in the transcripts of key developmental signaling proteins during murine development (Fujii, Shi et al. 2017). Homozygous deletion of an AUG initiated uORF in the 5' leader of *PTCH1*—which encodes the major receptor for SHH signaling—disrupted differentiation of mouse embryonic stem cells into neural progenitors (Fujii, Shi et al. 2017). Interestingly, ribosome profiling at various time points throughout neuronal differentiation of human embryonic stem cells revealed shifts in 5' leader coverage on a number of transcripts (Blair, Hockemeyer et al. 2017). However, these data were not systematically analyzed for active translation and characterization of uORFs, and relied solely of RPF reads as a measure translation of the whole 5' leader. Additionally, few

studies to date have included non-AUG initiated uORFs in their analysis (Brar, Yassour et al. 2012, Fijalkowska, Verbruggen et al. 2017, Spealman, Naik et al. 2017).

Treatment of human neuroblastoma cells with retinoic acid triggers their exit from the cell cycle and their differentiation into a neuron-like cell type (Sidell 1982, Pahlman, Ruusala et al. 1984). While many studies have sought to understand genetic changes underlying this process, most have focused on transcript-level changes, with evaluation of shifts in protein abundance only studied on a case-by-case basis (Hanada, Krajewski et al. 1993, Kaplan, Matsumoto et al. 1993, Korecka, van Kesteren et al. 2013, Pezzini, Bettinetti et al. 2017). Here we used RP in this simple model system to study the role of uORF activity in regulating protein translation during neuronal differentiation. Using a spectral coherence algorithm (SPECtre) and stringent dataset filtering, we find that uORFs are present in 31.5% of mRNA transcripts (Chun, Rodriguez et al. 2016). The majority of these initiate at near-cognate codons, a subset of which were validated experimentally. By filtering our dataset further, we identified a subset of uORFs which demonstrate conditional expression and an inverse relationship with the translational efficiency of their cognate CDS across cell states. Taken together, our work expands on established observations by building a set of principles which can be used to determine the translatability of a 5' leader and its potential for affecting CDS translation, while detailing the translational changes in a model of neuronal differentiation.

Results

SH-SY5Y cells are terminally differentiated into a neuronal cell population with RA

We first confirmed the efficacy of RA treatment in differentiating SH-SY5Y cells. Cells were propagated to 80% confluency prior to 10μM RA treatment for six days (**Figure 2.1A**). RA

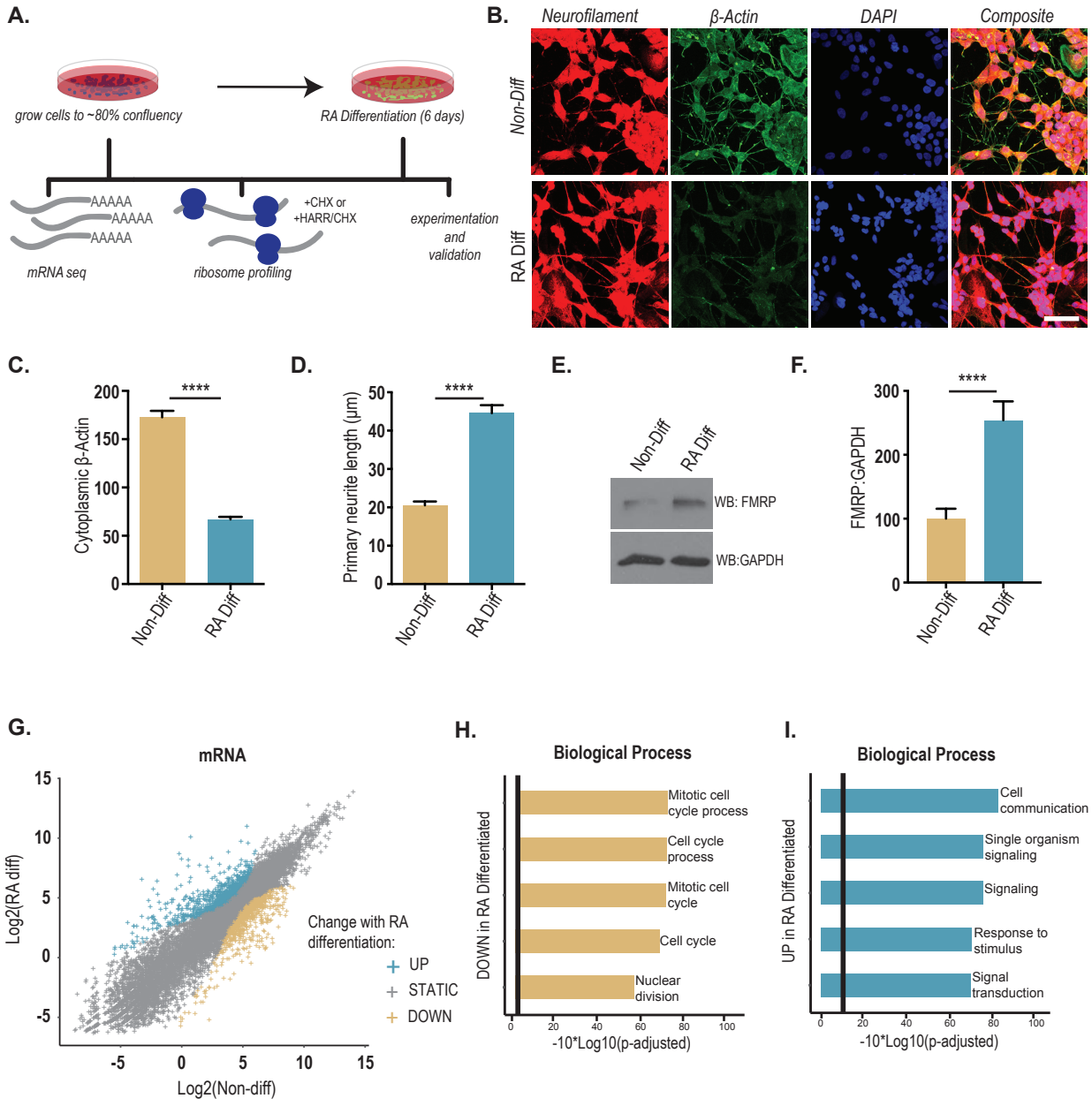


Figure 2.1:

Retinoic acid treatment induces neuronal differentiation of SH-SY5Y cells. A) Schematic of experimental design and data acquisition work-flow. B) Immunocytochemistry performed on Non-Diff and RA-Diff SH-SY5Y cells confirmed the shift to a neuronal phenotype with RA treatment. Immunocytochemistry was performed with antibodies against neurofilament (red) and β-actin (green). Nuclei were DAPI stained (blue). C) β-actin expression was decreased in RA-Diff cells. Individual cell fluorescence was quantified and represented as a corrected total cell fluorescence (CTCF) for β-actin; n=119 for Non-Diff and n=118 for RA-Diff. D) Primary neurite length measured by neurofilament staining was increased by RA treatment; n=109 for Non-Diff and n=100 for RA-Diff. E) Western blot showing an increase in FMRP after RA treatment, quantified in F); n=4 for both conditions. For panels C), D), and F) Student's t test, ****p≤0.0001. Graphs represent mean ± S.E.M. G) Differential mRNA abundance based on Non-Diff versus RA-Diff TPM. Transcripts were defined as significantly up-regulated (cyan) or down-regulated (gold) in the RA-Diff

condition based on rank-change in abundance compared to the Non-Diff condition. Significantly enriched up-regulated gene sets (biological process) in RA-Diff mRNAs (H) and down-regulated gene sets in RA-Diff mRNAs (I) based Benjamini-Hochberg multiple testing corrected p-values. Black vertical line denotes a corrected p-value cutoff of 0.05.

treatment induces an exit from the cell cycle and a change in cellular morphology. Previous studies have used a similar protocol as a model for dopaminergic neuronal differentiation, although RA treatment is thought to generate a more immature neuron-like cell than what can be achieved from a neural progenitor (Sidell 1982, Pahlman, Ruusala et al. 1984, Kaplan, Matsumoto et al. 1993). Cytoskeletal alterations confirm a shift towards a more neuron-like state after 6 days of treatment. Cytoplasmic beta-actin immunofluorescence decreased and neurofilament labeled neurites increased in length in the differentiated cells (**Figure 2.1B-D**) (Micheva, Vallee et al. 1998, Cheever and Ervasti 2013). We also observed an increase in expression of FMRP, a protein involved in neuronal function and translational control that is highly expressed in neurons relative to other cell types and tissues (**Figure 2.1E-F**) (Hinds, Ashley et al. 1993).

Global mRNA sequencing (mRNA-Seq) was performed to determine underlying transcriptional differences in undifferentiated ("Non-Diff") or RA Differentiated ("RA-Diff") cells (**Figure 2.1G**). Gene ontology analysis revealed downregulation of a network of transcripts associated with the biological processes of mitotically active cells in the RA-Diff cells (**Figure 2.1H**). In contrast, transcripts comprising biological pathways associated with cell communication and stimulus response were upregulated in this condition (**Figure 2.1I**). GO analysis of transcript groups for "cellular compartment" and "molecular function" reveal similar changes associated with both conditions (**Figure S2.1**).

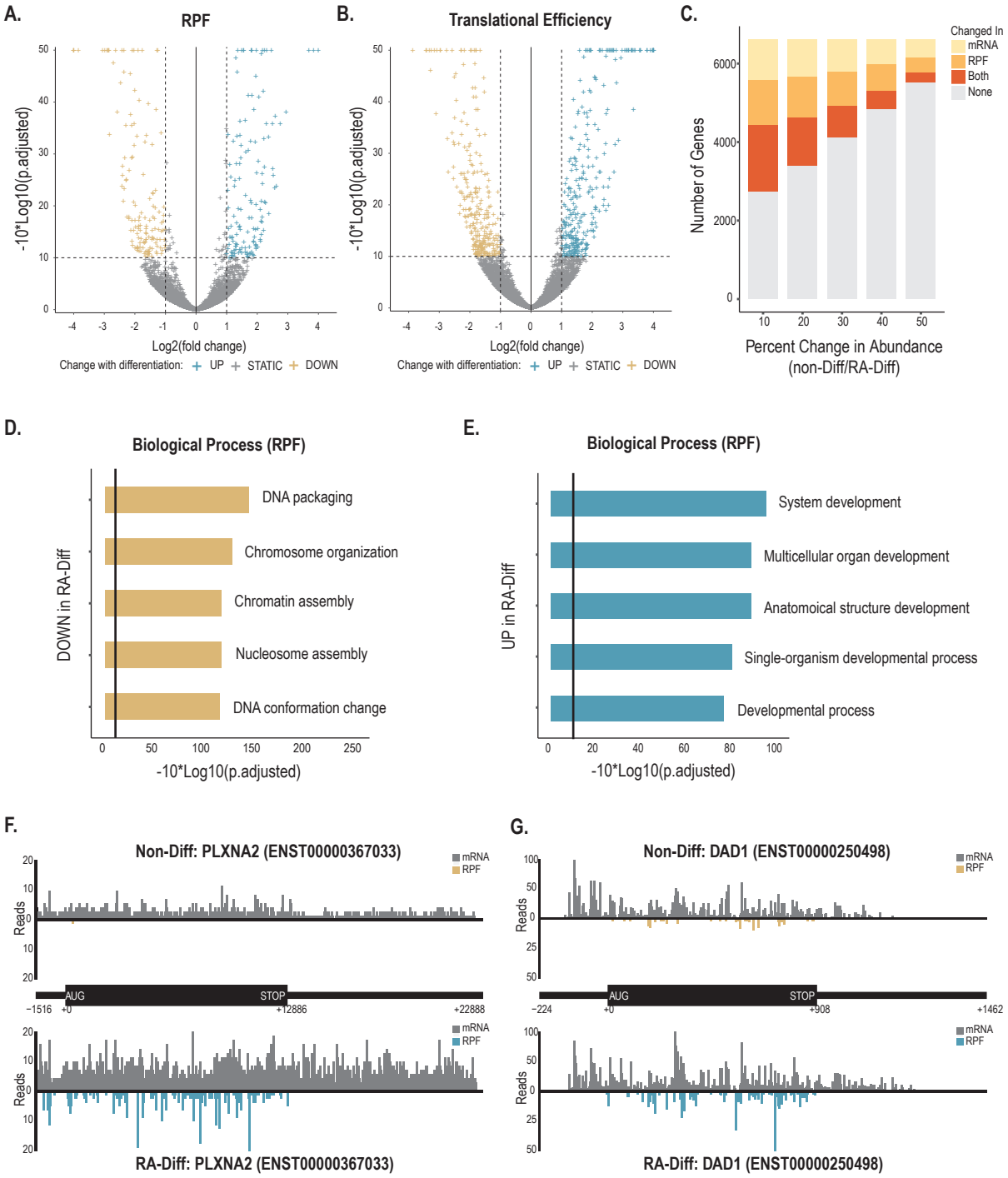


Figure 2.2:

Retinoic acid treatment induces differential translation in SH-SY5Y neuroblastoma cells. A) Volcano plot of transcripts with differential translation by RPF counts in Non-Diff and RA-Diff cells. Significantly up-regulated genes (cyan) and down-regulated genes (gold) in RA-Diff cells are defined by an absolute log2-normalized fold-change cutoff of ±1 (vertical lines), and a multiple testing corrected p-value cutoff of 0.1 (horizontal line). B) Volcano plot of transcripts with differential translational efficiency by Riborex analysis in

Non-Diff and RA-Diff cells. C) Changes in mRNA and RPF abundance across the two conditions was evaluated by pre-defined cutoffs of percent change in TPM, and the number of genes that changed as measured by one, both, or neither was counted. D) Gene sets (biological process) with significantly down-regulated RPF counts in RA-Diff cells. Genes upregulated in RA-Diff cells is shown in (E). The top five biological processes with significant change using a multiple testing corrected p-value cutoff of 0.05 (vertical line) are shown on the graph. F) Plot shows normalized mRNA reads (grey) and RPF (cyan/gold) over the 5'leader (thin line, left), CDS (thick line, middle), and 3'UTR (thin line, right). The axon guidance gene, *PLXNA2*, is representative of a transcript with higher translational efficiency and RPF in the RA-Diff condition. G) *DAD1* is an example of a transcript with consistent mRNA coverage, but an increase in RPFs in the RA-Diff condition.

Ribosome profiling can resolve the exact regions of mRNA undergoing translation at nucleotide resolution across the transcriptome within a cell population (Ingolia, Ghaemmaghami et al. 2009). By comparing ribosomal occupancy within a given transcript in Non-Diff to RA-Diff cells, we are able to gauge translational differences. This can be accomplished by directly measuring the number of ribosome protected fragments (RPFs) that map within a given transcript (**Figure 2.2A**) or RPF abundance can be normalized to mRNA expression in samples prepared in parallel as a measure of translational efficiency (TE) (**Figure 2.2B**). To discern the contribution of translation to this cell state shift, we compared the number of transcripts exhibiting a change at the level of mRNA (based on mRNA-seq), translation (based on RPF), or both across a range of percent changes. About half of all transcripts exhibit some change in one of these measures, although most exhibit less than a 50% shift. Intriguingly, the abundance of transcriptional and translational changes are comparable, suggesting that both contribute significantly to the morphological changes elicited by this shift in cell state (**Figure 2.2C**).

To determine if the same classes of genes are impacted at the translational level as those which shift transcriptionally, we performed gene ontology analysis. GO analysis of transcripts with significantly decreased RPF coverage in the RA-Diff condition were largely associated with nuclear function and aligned well with changes at the transcriptome level (**Figure 2.2D**, **Figure S2.2A-B**). In contrast, transcripts with RPF reads significantly upregulated in the RA-Diff

condition revealed changes in pathways associated with development and differed significantly from changes detected by mRNA seq (**Figure 2.2E**, **Figure S2.2C-D**). Inspection of translational efficiency (TE) changes revealed some consistent and some distinct differentiation-associated changes (**Figure S2.3**). The discrepancies between the pathways changed by RPF or TE show how these two measures can produce different results. While many of the results align, those that do not indicate how translational output is influenced by transcript abundance. As one example, the *PLXNA2* gene encodes a membrane-bound protein involved in nervous system development and axon guidance(Van Vactor and Lorenz 1999). Its mRNA coverage is upregulated upon differentiation; however, the increased expression is much greater at the RPF level (**Figure 2.2F**), consistent with higher translational output of the functional protein. Other transcripts such as *DAD1,* a factor critical for N-terminal glycosylation with roles in apoptosis and the unfolded protein response, exhibit shifts solely at the level of translation (**Figure 2.2G**) (Kelleher and Gilmore 1997).

Characterization and experimental validation of SPECtre-identified uORFs

We next explored the potential roles of uORFs in RA-induced differentiation. To annotate uORF sequences within the 5' leader of mRNA, we utilized the SPECtre algorithm for classifying active regions of translation (Chun, Rodriguez et al. 2016). SPECtre accounts for the fundamental attribute of an actively translating ribosome to shift position three nucleotides at a time as it synthesizes new peptides and the ability of ribosome profiling to resolve this behavior with peaks in read coverage. Our algorithm takes an unbiased approach to scoring all potential uORFs from start site to the next in-frame stop codon (**Figure 2.3**). For each potential uORF, the pattern of read coverage within this designated sequence is compared against the pattern of reads

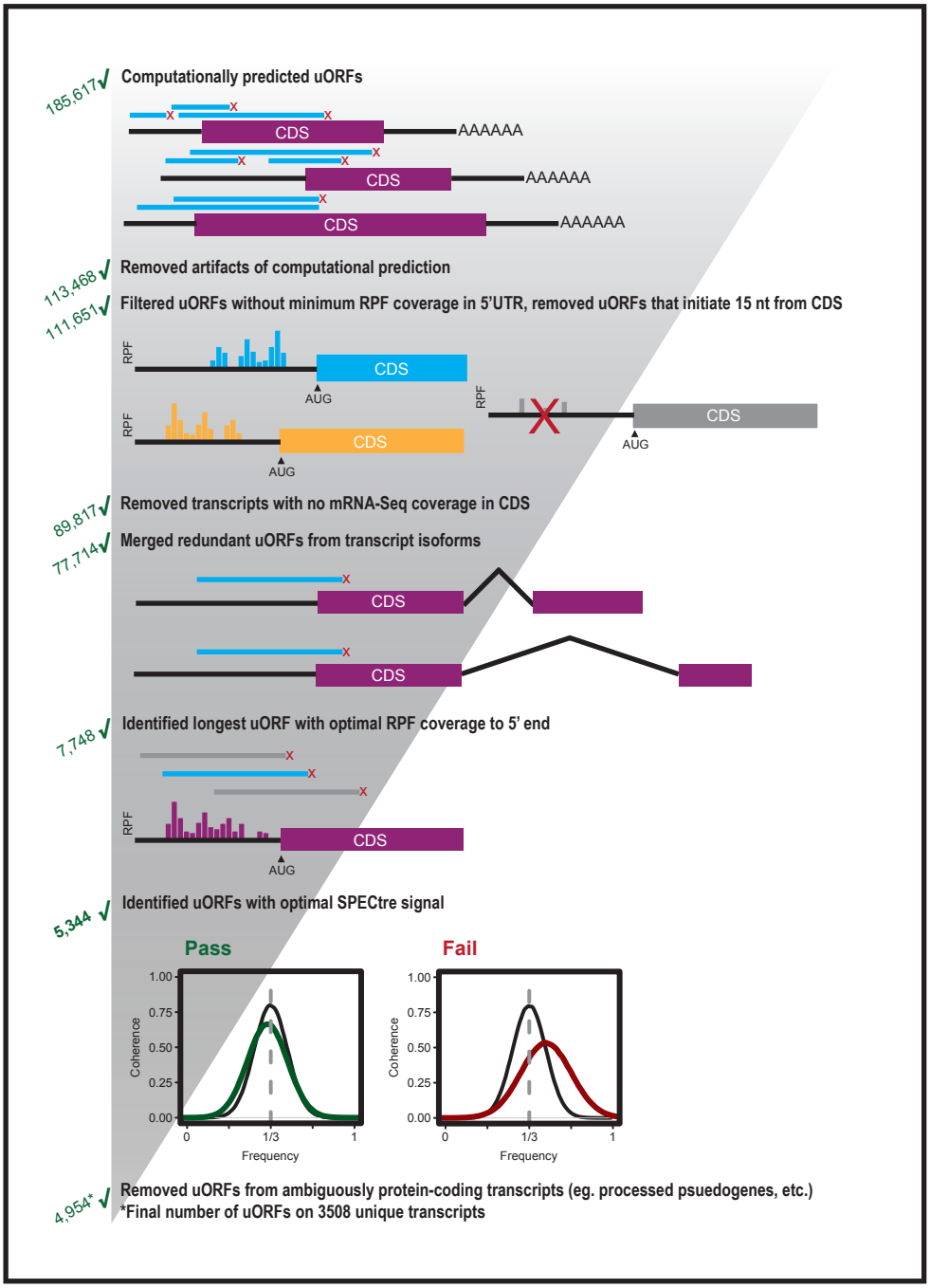


Figure 2.3:

Computational prediction and filtering of upstream-initiated open-reading frames. ORFs were predicted from AUG and non-AUG, near-cognate translation initiation sites in the 5' leader of annotated protein-coding genes, and computationally extended to the first termination site encountered in the 5' leader (upstream-terminated ORFs) or CDS (overlapping ORFs). Predicted ORFs were then filtered according to a series of heuristic filters including: 1) minimum RPF coverage in the 5' leader, 2) minimum mRNA-seq coverage in

CDS, 3) in-frame N-terminal extensions, 4) redundant isoforms, 5) minimum length with optimal RPF coverage, 6) sufficient SPECtre signal, and 7) removal of ambiguously annotated protein-coding transcripts.

across all known protein-coding regions in the experimental library. This analysis results in a set of experimentally determined scores that are then subjected to a range of transcript-level filters.

We established a translational threshold based on the distribution of scores in known coding genes to establish a minimum SPECtre score needed to classify a region as actively coding with a 5% FDR allowed. This results in a highly filtered set of 3,508 transcripts with 4,954 unique uORFs (**Figure 2.3**). Of these transcripts, 1,599 contained overlapping upstream-initiated ORFs (specified as oORFs), 1,438 uORFs fully contained in the 5' leader (cORFs), and 471 transcripts had two or more uORFs of either of these two categories (**Figure 2.4A**). The median distance of the uORF initiation site from the CDS is 99 nucleotides (**Figure 2.4B**). uORFs have a mean length of 78 nucleotides, but can span upwards of 500 nucleotides in length (**Figure 2.4C**).

Previous work using harringtonine, a drug that stalls ribosomes at initiation sites, revealed a surprising occurrence of near-AUG codons enriched in ribosome peaks (Ingolia, Lareau et al. 2011). Though near-cognate initiation had been recognized previously, this hinted that there may be a greater number of initiation events at these codons than previously expected (Zitomer, Walthall et al. 1984, Peabody 1987, Kozak 1989, Peabody 1989, Mehdi, Ono et al. 1990, Ingolia, Ghaemmaghami et al. 2009). When inspecting the translation start site of each SPECtre-identified uORF, we too see a number of non-canonical initiation events. Translation start sites were plotted to show the relative contribution of each in the final dataset (**Figure 2.4D**). AUG initiation sites were accounted for by two different methods: they were either directly identified by SPECtre, or factored into the total count if they were present within 30 nucleotides upstream or downstream of the start of the SPECtre signal without an intervening

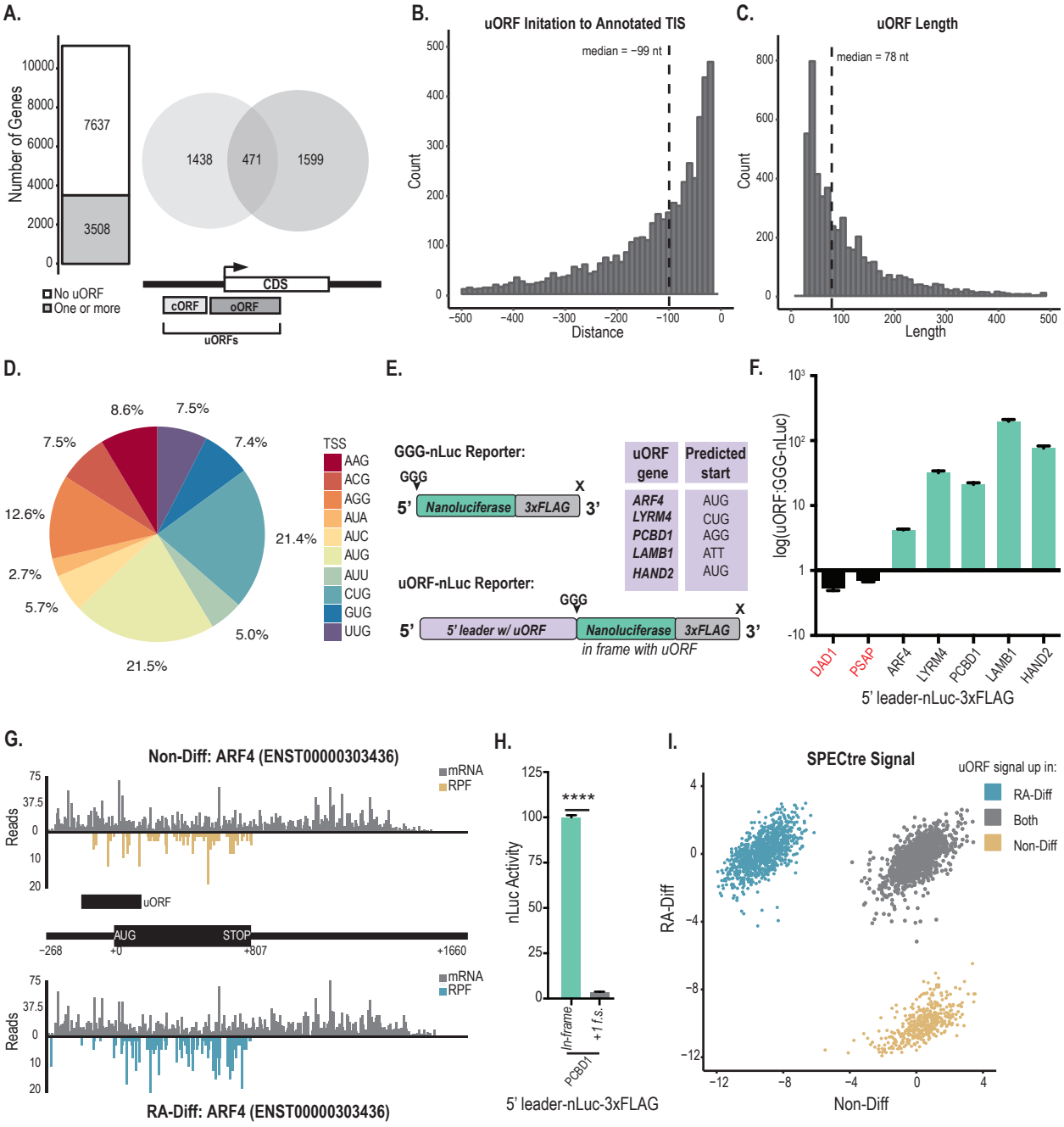


Figure 2.4:

Characterization and validation of predicted uORFs. A) The number of genes with at least one predicted ORF (bar plot) in the 5' leader of evaluated protein-coding genes. The number of genes with a predicted ORF terminated upstream in the 5' leader only (left circle), terminated in the CDS only (right circle), or with both a predicted upstream- and CDS-terminated ORF (overlap). B) Distribution of predicted ORF translation initiation position relative to the annotated protein-coding CDS start site in Non-Diff specific genes (gold), RA-Diff specific genes (cyan), and in aggregate (gray). C) Distribution of predicted ORF lengths in Non-Diff specific genes (gold), RA-Diff specific genes (cyan), and in aggregate (gray). D) Distribution of uORF translation start sites (TSS). AUG represents all AUGs predicted by SPECtre, or upstream/downstream 30-nt from the SPECtre predicted start site if no

intervening stop codon is present. Near-cognate start codons are utilized in the majority of uORFs, while AUG is the single most common start site. E) Schematic of the uORF nanoluciferase (nLuc) reporters used in this study. GGG-nLuc serves as a negative control, as its AUG initiation start codon is mutated to a GGG codon. This reporter supports very little translation. A table of the predicted start sites for each uORF reporter. F) nLuc assays performed in SH-SY5Y cells confirmed expression of these uORFs (teal). 5' leaders not included in the high-confidence dataset (black) are below the GGG-nLuc reporter activity and considered not translated. All values are normalized to the GGG-nLuc control performed in parallel during experimentation, data for individual reporters was collected in triplicate in multiple experiments. Student's t test, all teal uORFs in panel. F) have a p value $\leq$0.0001, and all black uORFs have a p value <0.05. Graph represents mean ± S.E.M. G) Plots show mRNA reads (grey) and RPF counts (cyan/gold) for *ARF4*. The annotated uORF is characterized by the presence of consistent RPF coverage in the 5' leader. H) Frameshifting the uORF relative to nLuc decreases translation of the reporter. The reporter was cloned so that the nLuc tag was frameshifted (f.s.) out of frame with the predicted uORF and the CDS start site. n=3, Student's t test, ****p$\leq$0.0001. Graph represents mean ± S.E.M. I) K-means clustering analysis of $\log_2$(uORF SPECtre Score) in Non-Diff and RA-Diff cells, reveals differentiation-associated shifts. Three clusters of CDS regulation emerge: those CDS that are up-regulated in RA-Diff cells (cyan), up-regulated in Non-Diff cells (gold), and CDS with no change in translational potential (gray).

stop site. Due to the high potential translatability of ORFs with AUG start codons, these were all annotated as AUG start sites. This constitutes 21.5% of the initiation sites used. In comparison, we detected 21.4% of uORFs use CUG as their initiation codon, consistent with previous reports (Peabody 1987, Ingolia, Lareau et al. 2011).

Detection of uORF derived peptides by mass spectrometry has proven difficult, likely due to a range of complications from sample preparation, bias in annotation algorithms, as well as intrinsic factors that make these peptides difficult to detect (Bazzini, Johnstone et al. 2014, Chugunova, Navalayeu et al. 2018). Several methods have been developed to enhance the detection of small peptides, each with variable yields (Slavoff, Mitchell et al. 2013, Bazzini, Johnstone et al. 2014, Na, Barbhuiya et al. 2018). Though techniques for detecting 5' leader-derived peptides are progressing. For example, by enriching for the N-terminus of proteins using terminal amine isotopic labeling of substrates (TAILS) with LC-MS/MS in HEK293 cells, 90 uORF peptides were detected, and about two-thirds of these were supported by ribosome profiling data (Na, Barbhuiya et al. 2018).

As many factors can hinder detection of a stable peptide, we sought to identify active translation of uORFs in our dataset through nanoluciferase (nLuc) assays. To verify that SPECtre

identified uORFs were expressed in SH-SY5Y cells, candidates were cloned for nLuc assays. The complete 5' leader upstream of the start site through the entire predicted coding region of the uORF was placed upstream of an nLuc tag without an AUG start codon mutated (**Figure 2.4E**). GGG-nLuc alone was used as a negative control, as GGG is not predicted to support translation initiation (Kearse, Green et al. 2016). SPECtre identified uORFs resided in the 5' leader of 5 genes: *ARF4*, *LYRM4*, *PCBD1*, *LAMB1*, and *HAND2*. Three of these transcripts have uORFs with near-cognate start sites, *ARF4* and *HAND2* utilize an AUG (**Figure 2.4E**). Reporters were co-transfected into SH-SY5Y cells with pGL4.13 which encodes firefly luciferase (FFluc) as a transfection control. *DAD1* and *PSAP* serve as negative controls, as their 5' leaders were filtered out early on in our analysis. All 5 of our predicted uORFs showed a significant level of translation above GGG-nLuc (**Figure 2.4F**). RPF and mRNA coverage of *ARF4* reveals significant coverage across the uORF in the Non-Diff state (**Figure 2.4G**). One key feature of SPECtre is its ability to discriminate reading frame (Chun, Rodriguez et al. 2016). While we have shown that 5' leaders can be translated, we wanted to investigate whether genes identified by SPECtre supported spurious translation or if these sequences were specific to initiation in one reading frame. To evaluate this, we mutated the reporter for *PCBD* so that the predicted uORF was out of frame. Placing nLuc out of frame resulted in a significant drop in signal (**Figure 2.4H**), revealing that the ribosome discerns coding regions based on specific sequence context, and that a 5' leader can support a discrete translation event.

We next sought to determine whether SH-SY5Y differentiation can affect uORF SPECtre score. We performed k-means clustering using the SPECtre score of uORFs in the Non-Diff and RA-Diff datasets (**Figure 2.4I**). This revealed that uORFs either group based on the cell-state (blue and gold) or show consistent scores in both states (grey). These clusters are not predictive

of a directional change in CDS translation (**Figure S2.4**). While SPECtre score is not a direct measure of translational efficiency, it is a measure of robust translation, thus this result suggests that expression of uORFs may be a regulated process. In fact, we observe a positive correlation between CDS TE and uORF TE in this dataset (**Figure S2.5**), which is consistent with a number of previous RP studies (Brar, Yassour et al. 2012, Chew, Pauli et al. 2016). This positive relationship does not preclude the potential for a uORF to act as a repressor, although it suggests that other regulatory measures can act upstream to dictate the role of uORFs such as pre-initiation complex binding. Further analysis is needed to resolve the intricacies of this relationship.

Constrained analysis of the uORF dataset reveals that conserved uORFs may be more inhibitory

To further analyze the overall oORFs and cORFs, we constrained our analysis to those that meet a higher translational threshold in at least one of the replicates from each condition. Thus, including only highly translated uORFs. This is key because we have included non-AUG initiated uORFs in our dataset, which in principle have lower potential for translation, and because of this may contribute a higher level of variability to our analysis. Additionally, after observing the positive correlation between CDS TE and uORF TE (**Figure S2.5**), we believed it was important to limit our analyses to a stricter dataset in order to parse out a relationship between these two regions. In filtering down our dataset, we obtained 158 oORFs and 137 cORFs that initiate at both AUG and near-cognate codons (**Figure 2.5A**, **Figure S2.6**). These cORFs and oORFs have comparable translational efficiencies (**Figure S2.7**). We analyzed the conservation of this set of uORFs divided into subsets based on initiation codon. uORFs were analyzed at the codon level (comprised of only the 5' leader region for oORFs) compared to all

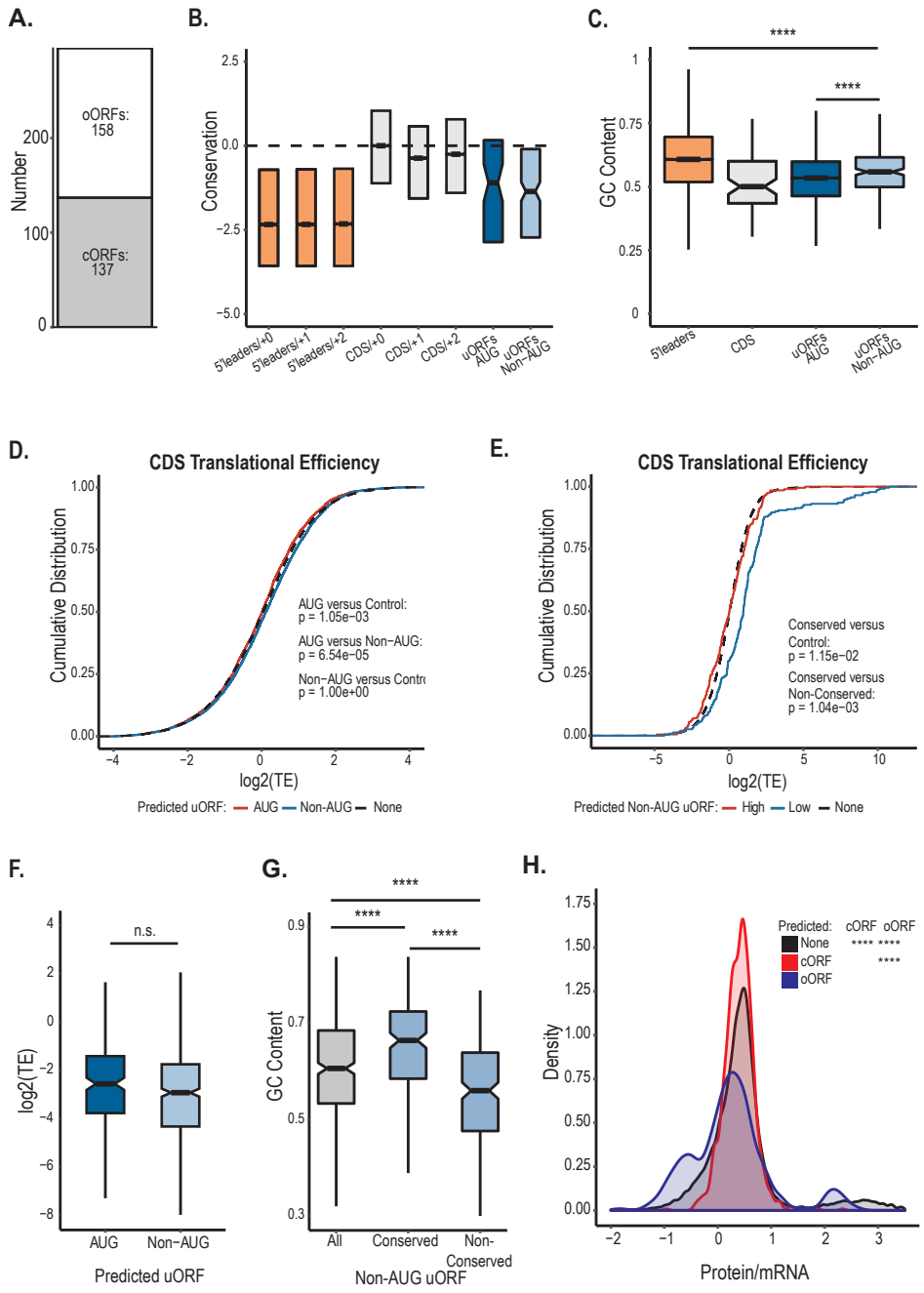


Figure 2.5:

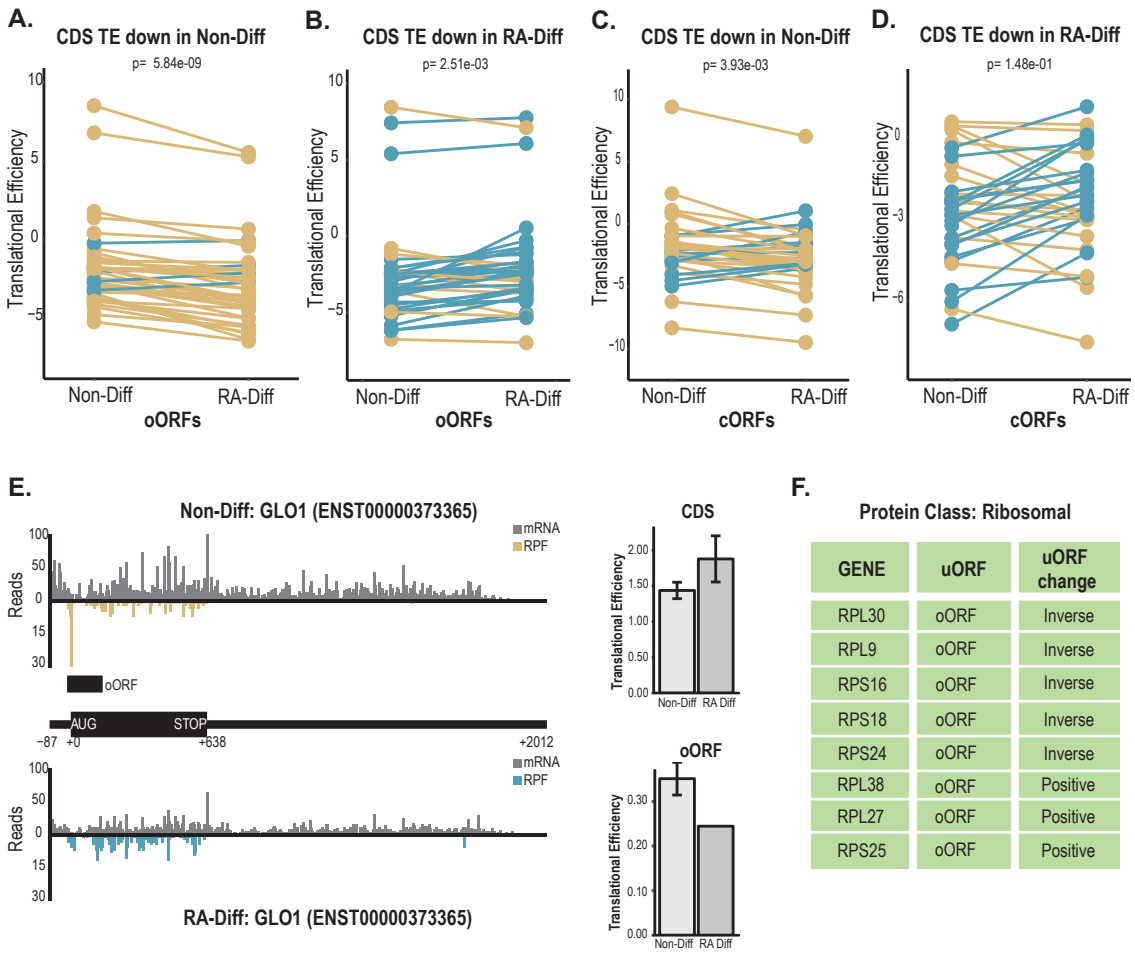
Constrained analysis of the uORF dataset reveals a repressive effect of highly conserved uORFs. SPECtre identified uORFs were further filtered to include only uORFs that have coverage in all four Non-Diff and RA-Diff libraries. A) This produces a dataset of 158 oORFs, which overlap the CDS, and 137 cORFs, which are fully contained in the 5' leader. B) Conservation analysis of annotated 5' leaders in all three reading frames (orange), annotated CDS regions over all three frames (grey), predicted AUG-initiated uORFs (dark blue), and predicted non-AUG uORFs (light blue). C) Average GC nucleotide content is shown for 5' leader regions (orange), CDSs (grey), AUG uORFs (dark blue), and non-AUG uORFs (light blue). For oORFs, only the 5' leader region of the oORF is included. 5' leaders are significantly more GC rich than both AUG uORFs and non-AUG uORFs, p= 5.72e-12 and

1.54e-07, respectively. Non-AUG uORFs are significantly more GC rich than CDSs and AUG uORFs, p=7.92e-18 and 2.16e-06. D) Empirical cumulative distribution of TE in all CDSs (black) verses CDSs from transcripts with two subsets of uORFs: those with an AUG initiation site (red) and those with a non-AUG initiation site (Blue). Distributions are significantly different with p-values annotated on graph. E) Empirical cumulative distribution of TE in all CDSs (black) verses CDSs from transcripts with two subsets of non-AUG uORFs: those in the highest quartile of conservation (Conserved, red) and those in the lowest quartile of conservation (Non-Conserved, Blue). Distributions are significantly different with p-values annotated on graph. F) Average TE for AUG uORFs and non-AUG uORFs reveals no difference between the two subtypes. G) Analysis GC content of non-AUG uORFs shows an increase in GC content of the conserved group relative to the non-conserved group. All verses conserved: p=5.89e-08, all verses non-conserved: p=2.54e-06, conserved verses non-conserved: p=1.62e-18. H) cORFs and oORFs identified in our dataset differentially impact the distribution of normalized protein abundance compared to transcripts lacking uORFs using a mass spectrometry dataset derived from human lymphoblastoid cell lines (Battle, Khan et al. 2015). All 3 groups are significantly different from one another.

5’ leader sequences, 3’UTRs, and annotated CDSs. Both non-AUG and AUG uORFs exhibit a range of conservation, with a significant portion showing conservation levels that are comparable to CDSs (**Figure 2.5B**).

Next, we investigated the sequence context for which AUG and non-AUG uORFs were derived by determining their GC content, which is generally higher in 5’ leaders than CDSs and which serves as a proxy for increased RNA structure due to the enhanced strength of GC hydrogen bonding compared to AU pairs (Banerjee 1980). Interestingly, we do observe a decrease in GC content in both AUG and non-AUG uORFs closer to what is observed for CDSs, suggesting that less stable secondary structure is necessary for robust translation of these 5’ leader regions (**Figure 2.5C**) (Kozak 1980, Kozak 1986). However, we do observe a significant increase in the GC content of non-AUG uORFs relative to AUG uORFs. This indicates that there may be a balance of secondary structure necessary to observe initiation at a near-cognate start site where too much is preclusive of translation but a slight increase relative to CDSs and AUG uORFs is necessary for near-cognate initiation. This is consistent with biochemical findings that the presence of a structured region can increase sub-optimal start site selection due to stalling of scanning ribosomes (Kozak 1990).

To investigate the role uORFs have on impacting downstream translation, we compared the cumulative distribution of CDSs without a predicted uORF to CDSs in our transcripts of interest. First, we investigated all uORFs regardless of conservation, but stratified into two groups: AUG initiated and non-AUG initiated. This analysis replicated past findings that AUG initiated uORFs are associated with less translation from their cognate CDSs, although this effect was mild (**Figure 2.5D**). In contrast, CDSs downstream of non-AUG initiated uORFs showed a comparable distribution to the control set. As the role of non-AUG uORFs has not been well elucidated, we wondered if there was an attribute of non-AUG uORFs that can delineate their role in affecting CDS translation. Recognizing the range of conservation, we next explored whether conservation carries with it any function or sequence specific features. We further divided the non-AUG uORFs into two groups: those in the top quartile of conservation (conserved) and those in the bottom quartile of conservation (non-conserved). For the top quartile of conserved transcripts, the presence of non-AUG uORFs was significantly, but modestly associated with CDS repression (**Figure 2.5E**). In contrast, non-AUG initiated uORFs in the lowest quartile of conservation were associated with highly translated CDSs, suggesting that there is a component of a uORF's regulatory role tied to sequence conservation. This is not due to TE levels of the conserved verses non-conserved, non-AUG uORFs (**Figure 2.5F**). Interestingly, the conserved group is associated with higher GC content, consistent with the finding that 5' leader secondary structure is a predictor of CDS repression and that these features are conserved (Pelletier and Sonenberg 1985, Chew, Pauli et al. 2016). These data depict the dichotomy between AUG and non-AUG uORFs, but emphasize a potential role of both for regulating downstream translation.



GENE	uORF	uORF change
RPL30	oORF	Inverse
RPL9	oORF	Inverse
RPS16	oORF	Inverse
RPS18	oORF	Inverse
RPS24	oORF	Inverse
RPL38	oORF	Positive
RPL27	oORF	Positive
RPS25	oORF	Positive

Figure 2.6:

Retinoic Acid induced differentiation elicits transcript-level reciprocal shifts in uORF and CDS translation rates. A) CDSs with oORFs that exhibited the largest increase in TE in the RA differentiated state in the Non-Diff state) exhibit a significant decrease in oORF TE in the RA-Diff state by paired analysis. B) CDSs that exhibit the largest decrease in TE with RA treatment exhibit an increase in oORF TE with RA treatment by paired analysis. C) CDSs with oORFs that exhibited the largest increase in TE in the RA differentiated state exhibit an decrease in cORF TE with RA differentiation by paired analysis. D) CDSs that exhibit the largest decrease in TE with RA treatment do not exhibit a significant reciprocal increase in cORF translation after RA treatment. For A-D, Transcripts with uORFs were separated into quartiles based on their degree and direction of change in TE ratio with differentiation. Gold lines represent a nominal decrease in uORF TE with RA treatment. Blue lines represent a nominal increase in uORF TE with RA treatment. E) Plots show mRNA reads (grey) and RPF counts (cyan/gold) for *GLO1*. Bar graphs of the average CDS TE or uORF TE across both replicates illustrate the inverse relationship of translation through these two regions. F) A sample of oORF-containing ribosomal protein transcripts show both an inverse and positive correlation between translation of the CDS and oORF, the full set is listed in Figure S2.9.

To provide further validation for our results, we cross-referenced our results with a quantitative SILAC mass spectrometry dataset derived from 62 lines of human lymphoblastoid cells (Battle, Khan et al. 2015). After restricting our analysis to transcripts with stable RNA

levels in the top 75% of expression (RPKM) as provided by RNA-sequencing data for all lines, we obtained ~1500 transcripts with mass spectrometry data (Pickrell, Marioni et al. 2010). 50 of these transcripts reside in the constrained dataset of 295 uORF-containing transcripts described above and 540 transcripts are in the full set of 4,954 transcripts with a predicted uORF, leaving ~1000 transcripts without translation in the 5' leader (None). CDS translation (SILAC:mRNA) of the 50 transcripts from the restricted set (cORF and oORF) was analyzed against CDS translation of the group of transcripts with no uORF (None). This revealed a slight decrement in normalized protein abundance of CDSs downstream of a cORF and a sizable reduction in normalized protein abundance of CDSs downstream of an oORF, relative to CDSs without 5' leader translation (**Figure 2.5H**). Though these data are derived from a cell-type other than SH-SY5Ys, this reveals a possible mechanism of translation repression dependent upon the context of uORF translation with oORFs having the potential for greater repression than a cORF. This is supported by past studies as cORFs have the ability to reinitiate if termination occurs at an optimal distance from the CDS, whereas the translation of oORFs directs ribosomes downstream of the CDS initiation site (Kozak 1987, Hinnebusch, Ivanov et al. 2016).

Conditional fluctuations in uORF translation inversely correlate with CDS translation

uORFs generally inhibit the downstream translation of their cognate CDS. In certain cases, such as cellular stress, the use of specific uORFs is regulated in response to the stimulus in a manner that changes the associated CDS translation. One well studied case is *ATF4*, a gene with two uORFs. Translation of the second uORF is inhibitory to translation of the main ATF4 protein (Harding, Zhang et al. 2003, Vattem and Wek 2004). However, under conditions where fully assembled ternary complexes are limiting (due to eIF2α phosphorylation), this second

uORF is bypassed, leading to an activation of *ATF4* translation (Vattem and Wek 2004). This balance between translation within the 5' leader and translation within the CDS results in an inverse relationship between the two.

RA treatment of SH-SY5Y cells induces potent changes in protein synthesis mediated in part at the level of translational control. We therefore evaluated the relationship between shifts in uORF translation and CDS translation with RA-induced cellular differentiation using the constrained dataset from Figure 2.5. To discern the comprehensive relationship between translation of these two regions, we first analyzed the relationship of conditional translation of oORFs between the two conditions (RA-Diff:Non-Diff) against the same ratio for CDS translation (**Figure S2.8A**). The same analysis was done for cORFs (**Figure S2.8B**). Condition-dependent uORF changes and CDS changes were positively correlated. Though this is consistent with previous RP reports on uORFs, classical biochemical studies would suggest that an inverse relationship should be observed (Hinnebusch 1997). We wondered if there was a subset of transcripts that might show an inverse relationship in response to differentiation. Thus, we limited our analysis to transcripts with the largest shifts in CDS translation. Paired analysis was performed on individual transcripts in the top quartile of differentiation-dependent changes in CDS translation. oORF-bearing transcripts that show their lowest level of CDS translation in the Non-Diff state exhibited their highest level of oORF translation in this condition (**Figure 2.6A**). The opposite was true for oORF-bearing transcripts that where repressed in the RA-Diff condition (**Figure 2.6B**). This pattern was also detected for transcripts with cORFs with CDS translation repressed in the Non-Diff state (**Figure 2.6C**). One exception to this finding are transcripts with cORFs that are repressed in the RA Diff state (**Figure 2.6D**). The reason for this is unclear, but over half of the transcripts in this particular group do show the relationship

described above. Overall, the trend determined through this analysis suggests that highly repressed CDS translation is largely accompanied by an oppositional shift in uORF translation.

This inverse relationship is further evidenced on the novel uORF-baring transcript, *GLO1* (**Figure 2.6E**). *GLO1* is an enzyme central to the glyoxalase system, which detoxifies byproducts of cellular metabolism, and is central to neuronal repair in response to injury (Pieroh, Birkenmeier et al. 2014).The oORF on *GLO1* is increased in TE in the Non-Diff condition, while the CDS is decreased in TE. The inverse is evident in the RA-Diff state. Interestingly, in transcripts with oORFs, there is an enrichment for ribosomal proteins (**Figure 2.6F, Figure S2.9**). These 19 oORF-bearing transcripts exhibit both an inverse and positive oORF to CDS relationship; whereas of the 4 ribosomal protein transcripts that contain cORFs, all show a positive correlation between translation in these two regions. However, in 20 of the 23 transcripts, the CDS is more highly translated in the Non-Diff state. This pattern may be indicative of state-dependent regulation associated with the specific molecular functions of ribosomal proteins, and oORFs may play a role in modulating their expression.

Conclusions

As ribosome profiling studies and analyses continually arise, we are still deciphering the overarching role of uORFs in regulating translation across the transcriptome. While there is a clear repressive effect of translation within the 5' leader of specific transcripts, the data derived from ribosome profiling experiments are less conclusive. Several studies have concluded a cumulatively repressed CDS TE on transcripts with an AUG-initiated uORF (Chew, Pauli et al. 2016, Johnstone, Bazzini et al. 2016). Some have suggested divergent roles for AUG-driven uORFs and uORFs with near-cognate start sites in regulating CDS translation (Fijalkowska,

Verbruggen et al. 2017, Spealman, Naik et al. 2017). Other studies have shown a positive correlation between CDS translation and uORF translation (Brar, Yassour et al. 2012, Chew, Pauli et al. 2016). The reason for this variability may be due to the differences in criteria used to identify active translation in 5' leaders. In addition, there may be a component of uORF regulation that is cell type- and condition-specific.

In this study, we show that SPECtre and stringent filtering can be used to categorize a set of uORFs translated in SH-SY5Y cells. By including near-cognate start codons in our analysis, we show that uORFs that utilize near-cognate start sites can be highly translated and are more prevalent in our highest confidence dataset than AUG-driven uORFs. Moreover, we confirmed that expression of uORFs that initiate at near-cognate start codons can be validated in cells. In aggregate, we confirm what others have described—AUG initiated uORFs are associated with a repressed CDS (Chew, Pauli et al. 2016, Johnstone, Bazzini et al. 2016, Spealman, Naik et al. 2017). However, here we find a divergent role for non-AUG uORFs based on conservation and their level of translational activity. Highly conserved uORFs are present upstream of cumulatively repressed CDSs whereas lowly conserved uORFs reside upstream of efficiently translated CDSs. This is particularly interesting because both sets of uORFs utilize predominantly near-cognate start codons. RP reports showing CDS repression by uORFs commonly restrict their analysis to AUG initiation (Bazzini, Johnstone et al. 2014, Chew, Pauli et al. 2016, Johnstone, Bazzini et al. 2016).

In yeast studies, groups have suggested that near-cognate initiation in 5' leaders is positively correlated with CDS translation, showing no apparent repressive role or repression in only a subset of transcripts (Brar, Yassour et al. 2012, Spealman, Naik et al. 2017). However, one RP study did suggest a repressive role for non-AUG uORFs under conditions of decreased

start site stringency. This study utilized a knockdown of eIF1, the "gatekeeper" of start-site selection, which favors AUG initiation (Valasek, Nielsen et al. 2004, Nanda, Cheung et al. 2009). RP of eIF1 knockdown in human epithelial cells showed an increase in translation of non-AUG uORFs relative to their CDSs, but only of uORF/CDS pairs that were regulated by the knockdown (Fijalkowska, Verbruggen et al. 2017). Our data extend on this by describing that near-cognate initiation may have a role under normal cellular conditions, where endogenous translation machinery has not been manipulated. Stratifying our analysis by conservation was key to identifying this result. We show that conserved non-AUG uORFs are mildly repressive in aggregate; we suspect that using SPECtre and strict criteria to identify robust translation helped to clarify this finding. This is consistent with findings that features of AUG-initiated uORF-mediated translation are conserved (Chew, Pauli et al. 2016).

When investigating conditional shifts in translation, an overarching role for uORFs has not been well established. A report using peroxide treatment to induce stress in *Saccharomyces cerevisiae* observed no obvious relationship between an increase in RPFs in the 5' leader and CDS translation across conditions (Gerashchenko, Lobanov et al. 2012). Several more reports have suggested a positive relationship between uORF translation and CDS translation, explaining this by the low frequency of uORF translation and a reliance on whole transcript activity to increase in order to detect such a low event (Brar, Yassour et al. 2012, Jang, Lahens et al. 2015, Chew, Pauli et al. 2016). In principle, this is likely the overarching relationship, as this is what we observe in our full dataset. It is likely that translation of 5' leaders and CDSs can increase concomitantly as a by-product of increased ribosomal-mRNA interaction. The scanning model of translation initiation would predict that an increase in suboptimal codon selection would increase as the number of preinitiation complexes increase on a transcript (Mehdi, Ono et al. 1990,

Hinnebusch, Ivanov et al. 2016). Additionally, transcription can work upstream to dictate the role of a uORF. Alternative transcription start sites have been a proposed mechanism by which a given uORF is excluded from the main transcript isoform (Pelechano, Wei et al. 2013, Spealman, Naik et al. 2017). Still, both cistrons on a transcript can increase similarly as transcription increases. Constraining our analysis to uORFs easily detected in both conditions likely eliminated some of the noise created by the mechanisms mentioned above. In our study, paired analysis of a this highly filtered set of oORFs showed robust directionality and an inverse correlation with their associated, translationally repressed CDS. This relationship was not as strong for cORFs. Regardless, these findings suggests that the pattern of repression represented in classically studied uORF-bearing transcripts—like *ATF4* and *GCN4*—appears in transcripts associated with RA induced cellular differentiation (Hinnebusch 1984, Thireos, Penn et al. 1984, Harding, Novoa et al. 2000, Vattem and Wek 2004).

Future work will be needed to understand both the relationships of these uORF regulated genes to each other and the mechanisms by which their uORF activities are altered. One set of genes that is particularly enriched in our dataset are ribosomal protein encoding transcripts that contains oORFs. While ribosomes were once thought to play a passive role in the selection of mRNA transcripts, many recent studies have suggested that specialized ribosomes can contribute to the overall proteome through sequence-specific interactions (Xue and Barna 2012). For example, RPL38 specifically interacts with *HOX* mRNAs because of structures in their 5' leaders, these encode key transcription factors required for proper development (Xue, Tian et al. 2015). Thus, the make-up of ribosomal proteins within a cell can influence the cell phenotypically. Further evidence for specialized ribosomes is illustrated in the tissue-specificity of ribosomal proteins (Kondrashov, Pusic et al. 2011, Guimaraes and Zavolan 2016). We see a

negative relationship between oORFs present on *RPL30*, *RPL9*, *RPS16*, *RPS24*, and *RPS18* transcripts and the CDSs for these genes, which may contribute to their expression level changes in differentiation.

The data presented in this study are important for demonstrating that robustly translated and conserved uORFs carry the potential to act as regulated repressors CDS translation. Importantly, many of these regulatory uORFs initiate in the absence of an AUG start codon. Scaled translation of uORFs may thus serve as an important regulatory mechanism for translation of a set of transcripts whose proteins are conditionally expressed in response to neuronal differentiation.

Methods: Cell culture and Experimentation

SH-SY5Y cell maintenance and differentiation

SH-SY5Y cells were grown in DMEM:F12 media (Invitrogen) supplemented with 10% FBS, .01mg mL^{-1} Gentamicin and .25ug mL^{-1} Amphoreticin B. Cells were plated on 150mm plates that were either coated with .1mg/mL poly-D lysine (Millipore) for differentiation or uncoated. Cells were allowed to propagate to 80% confluency for 1-2 days prior to lysing for ribosome profiling or induction of differentiation. SH-SY5Y cells were differentiated for 6 days in 10μM retinoic acid (all-trans, Sigma), with media changed every 24 hours.

Construction of the Ribosome Profiling libraries

Ribosome profiling libraries were prepared as in Ingolia et al., 2010 and Ingolia et al., 2012. Cells were washed with ice cold PBS with CHX at 100ug mL^{-1}. Plates were immediately flash frozen in liquid nitrogen, moved to dry ice, and lysed (in the presence of CHX) to prevent

ribosome loading and runoff. Additional lysates were processed in parallel for poly(A) mRNA purification and mRNA-sequencing library preparation. Polysomes were isolated from the ribosome footprinting lysates on a 1M sucrose gradient with high speed centrifugation using a 70.1Ti rotor (Beckman) at 55,000 r.p.m. for 4 hrs at 4°C. rRNA was eliminated prior to linker ligation using Ribo-Zero Gold rRNA Removal Kit (Illumina). Ribosome Profiling libraries were barcoded and multiplexed with 2-4 libraries per lane, and sequenced on a HiSeq 2000 (Illumina) using 50 cycles of single end reads. mRNA libraries were multiplexed on a single lane. All sequencing was conducted at the University of Michigan DNA Sequencing Core.

Plasmid Construction

pcDNA 3.1 plasmid was modified to encode NanoLuc and GGG-NanoLuc as previously published CITE. gBlocks® (IDT) were ordered of the 5' leader sequence to the last codon before the in-frame stop of selected genes flanked by restriction sites. These were restriction cloned upstream of GGG-nLuc using PacI and XhoI (NEB) with 12 nucleotides between the start of the 5' leader and the T7 promoter sequence to reduce spurious initiation in sequences specific to the plasmid. Frameshifts were accomplished by PCR cloning with primers that inserted one or two nucleotides between the uORF and the nanoluciferase sequence. PCR products were cloned in place of nanoluciferase in the original uORF plasmid using XbaI and SacII (NEB). Restriction digest and Sanger Sequencing were used to confirm plasmid sequence.

SH-SY5Y Transfection and Nanoluciferase Assay

SH-SY5Y cells were seeded on 6-well culture plates at 3 $x10^5$ cells per well. 24 hours post seeding, each well was transfected using 7.5µL FUGENE HD (Promega) and 1.25µg

nanoluciferase reporter plasmid along with 1.25µg pGL4.13 (internal transfection control that encodes firefly luciferase [FFluc]) in 300 µL of OptiMEM (Invitrogen). Transfections of differentiated cells were performed on day 5 in RA supplemented media. Cultures were allowed to grow for 24 hours after transfection. Cells were lysed in 250µL Glo Lysis Buffer (Promega) for 5 minutes at room temperature. 50µL lysate was mixed with 50µL prepared Nano-GLO or ONE-Glo (Promega) for 2 minutes, and bioluminescence was detected using a GloMax® 96 Microplate Luminometer. Nanoluciferase signal was normalized to FFluc signal in each sample. pcDNA 3.1 encoding nLuc the AUG start codon mutated to a GGG (GGG-nanoLuc) was run in parallel with each experimental nLuc plasmid and subjected to both conditions to serve as a control for normalization.

Immunocytochemistry and microscopy

Cells were fixed at 37°C with 4% PFA/4% sucrose in PBS with 1mM $MgCl_2$ and .1mM $CaCl_2$ (PBS-MC), permeabilized for 5 minutes in .1% Triton-X in PBS-MC, and blocked for 1 hour with 5% bovine serum albumin in PBS-MC. Cells were incubated in blocking buffer and primary antibodies against β-actin (Santa Cruz Biotechnology, cat# sc-130656, 1:1000) and neurofilament (Abcam, Ab8135, 1:1000) for 1 hour at room temperature. Following 3x10 minute washes in PBS-MC, cells were incubated in PBS-MC with Alexa Flour 488 conjugated goat anti-rabbit IgG and Alexa Flour 635 conjugated goat anti-mouse IgG to achieve secondary detection (Thermo Fisher, 1:1000). Cells were washed again, and placed in ProLong™ Gold antifade reagent with DAPI (Invitrogen).

Imaging was performed on an inverted Olympus FV1000 laser-scanning confocal microscope using a 40x objective. Acquisition parameters were identical for each condition and

optimized to eliminate signal bleed-through between channels. Images were converted to average-intensity z-projections in ImageJ. Cytoplasmic β-actin was quantified by averaging the integrated density corrected for background signal of the cells in each condition. The length of one main neurofilament-labeled primary neurite per cell was determined in ImageJ and converted from pixels to μm, and averaged for each condition.

Western Blotting

Cells were maintained as described above. Cells were washed 2X in PBS, and RIPA buffer was added to a single well of a 12-well dish either at 80% confluency or after 6 days of retinoic acid differentiation. Cells were agitated for 40 minutes at 4°C to ensure complete lysis. Lysates were clarified by centrifugation, and the supernatant was mixed with reducing SDS sample buffer and boiled for 5 minutes at 90°C. Equal amounts of lysate were loaded on an SDS-PAGE gel and subsequent western blotting was carried out with primary antibodies against FMRP (1:1000, cat# 6B8, BioLegend) or GAPDH (1:1000, cat# sc-32233, Santa Cruz Biotechnology)—in 5% (wt/vol) non-fat dry milk in TBS-T (NFDM). An HRP conjugated goat antibody to mouse IgG or to rabbit IgG was used for secondary detection (1:5000, Jackson ImmunoResearch Laboratories) in 5% NFDM.

Methods: Analysis of ribosome profiling and mRNA-Seq datasets

Alignment to the human genome and transcriptome (GRCh38 Ensembl version 78)

Ribosome profiling and mRNA-Seq reads were trimmed of adapters, and then by quality using *fasqt-mcf* from the *ea-utils* package (Aronesty, 2011). Ribosome profiling and mRNA-Seq reads in FASTQ format were trimmed based on quality if four consecutive nucleotides were

observed with Phred scores of 10 or below. The minimum read length required after trimming was 25 nucleotides.

Trimmed sequences were then aligned to a ribosomal RNA sequence index using Bowtie v1.1.2 (Langmead, Trapnell et al. 2009) to deplete them of contaminant sequences. Alignment to the rRNA sequence contaminant index was performed using the following parameters: seed alignment length of 22 nucleotides, no mismatches in the seed alignment were allowed, with the unmapped reads written to a separate FASTQ file.

```
bowtie -l 22 -n 0 -S --un /path/to/depleted_reads.fq \
      /path/to/rRNA_index \
      /path/to/trimmed_reads.fq
```

Ribosome profiling and mRNA-Seq reads depleted of rRNA contaminant sequences were aligned to the human genome and transcriptome (Ensembl, version 78) using TopHat v.2.0.10 (Trapnell, Pachter et al. 2009). The trimmed and rRNA-deplete reads were aligned to the human genome and transcriptome with the parameters specifying standard Illumina reads, with un-gapped Bowtie 1 alignment (using a seed alignment length of 22 nucleotides, with no mismatches in the seed alignment allowed), to annotated junctions only, using Solexa quality scores:

```
tophat2 -p 4 –bowtie1 \
      --no-novel-juncs \
      --library-type fr-unstranded \
      --solexa-quals \
      -G /path/to/ensemble.gtf \
      /path/to/bowtie_index \
      /path/to/depleted_reads.fq
```

Sequence alignment quality filtering and merging

Ribosome profiling and mRNA-Seq reads aligned to the human genome and transcriptome by TopHat2 were output to BAM format, and then sorted by chromosomal coordinate. Reads were then filtered by mapping quality using SAMtools (Li, Handsaker et al. 2009); read alignments were required to have minimum mapping quality of 10, or higher, to be retained for subsequent

analyses. Unique read group identifiers were assigned to each technical and biological replicate, and then the alignments were merged by technical replicates and subsequently as biological replicates using Picard (http://broadinstitute.github.io/picard/).

Metagene profile generation and alignment offset calculation

For counting reads over transcript isoforms, metagene profiles were generated from the Ensembl (version 78) transcript annotation database using Plastid (Dunn and Weissman 2016). A- and P-site offsets for harringtonine and cycloheximide ribosome profiling reads, respectively, were determined by pooling all reads that overlapped canonical AUG translation initiation start sites from annotated protein-coding genes. The most common (mode) distance from the 5' ends of reads of a given length to the position of the AUG in those reads was accepted as the A- or P-site offset distance.

Calculation of transcript abundance

Read counts over each transcript isoform, or region (5'UTR, CDS, and 3'UTR), were normalized by length, summed, and reported as transcripts per million (TPM) as described previously (Li and Dewey 2011). At the time of analysis, Cufflinks (Trapnell, Williams et al. 2010) was required for initial transcript quality control, and was run with the following parameters:

```
cufflinks -p 8 -o /path/to/output \
      -G /path/to/ensemble.gtf \
      /path/to/tophat/alignments
```

SPECtre analysis of transcripts in non-differentiated and RA-differentiated libraries

SPECtre profiling (Chun, *et. al.,* 2016) measures the strength of the tri-nucleotide periodicity inherent to the alignment of ribosome protected fragments to protein-coding genes in

a reference transcriptome. SPECtre analysis was applied in two stages: 1) to score the translational potential of annotated transcripts to build a background protein-coding reference model, and 2) to score the translational potential of predicted upstream-initiation ORFs. In this way, the translational potential of predicted upstream and overlapping ORFs are score against a background model of translation derived from annotated protein-coding transcripts. Annotated protein-coding transcripts were profiled by SPECtre using the following parameters:

```
python /path/to/SPECtre.py \
      --input /path/to/tophat/alignments \
      --output /path/to/output \
      --log /path/to/logfile \
      --gtf /path/to/ensemble.gtf \
      --fpkm /path/to/cufflinks/isoforms.fpkm_tracking \
      --len 30 \
      --fdr 0.05 \
      --min 3.0 \
      --nt 8 \
      --type mean \
      --target <chromosome_id>
```

Where the minimum FPKM required for a transcript to be considered as translated for generation of the background model was specified as 3.0, and the length of the sliding SPECtre windows was set to encompass 30 nucleotides. The SPECtre score for a transcripts was defined as the mean of the scores over these sliding windows, and a 5% false discovery rate was established to set the minimum SPECtre translational score threshold. In addition, SPECtre profiling was split by chromosome to speed computation, and the results were merged afterwards using a custom Python script. Finally, prior to analysis of predicted upstream-initiated ORFs by SPECtre profiling, the minimum SPECtre translational threshold was re-calculated using TPM instead of FPKM using a minimum cutoff of 10 transcripts per million.

Computational prediction of upstream-initiated open reading frames

Open reading frames were computationally predicted from annotated 5'UTR sequences (Ensembl, version 78) using AUG, and near-cognate non-AUG translation initiation site sequences. Open reading frame sequences were generated based on these predicted initiation site sequences and read through to the first in-frame termination codon encountered in the annotated CDS. These predicted ORFs were then used to generate coordinates over which they would be profiled and scored by SPECtre. Identical parameters to the annotated transcript SPECtre analysis were employed for consistency across analyses:

```
python /path/to/SPECtre-uORFs.py \
      --input /path/to/alignments \
      --output /path/to/output \
      --results /path/to/spectre/transcript_results \
      --log /path/to/logfile \
      --fpkm /path/to/cufflinks/isoforms.fpkm_tracking \
      --len 30 \
      --fdr 0.05 \
      --min 3.0 \
      --nt 8 \
      --type mean \
      --target <chromosome_id>
```

Three alternative inputs are required for the SPECtre analysis of predicted ORFs: 1) the annotated transcript GTF database was not required and removed, 2) the results of the annotated transcript analysis, and 3) a genomic sequence file in FASTA format. The results of the annotated transcript analysis were used to identify the set of transcripts from which to predict upstream-initiated ORFs, and the FASTA sequence file was used to generated the ORF sequences for output.

Supplemental annotation of non-AUG translation initiation sites

Upstream sequences of predicted non-AUG translation initiation sites were examined for possible in-frame AUG initiation start sites; 5'UTR sequences of predicted non-AUG sites were extracted, and then searched for the presence of in-frame AUG sites. These non-AUG sites were

then re-annotated according to the proximity of upstream AUG initiation sites: those with an in-frame AUG site within 30 nucleotides of the predicted TIS, and those with an in-frame AUG site in-frame, but beyond 30 nucleotides upstream of the predicted site.

Kernel density estimation of differential uORF translation on CDS translational efficiency

To further differentiate those uORFs with differential translation and identify those that contribute to the regulation of downstream CDS, the log-change in predicted ORF TPM was compared against the log-change in downstream CDS TPM across the conditions. The differential translational identity of each predicted ORF was retained from the SPECtre clustering analysis, and kernel density estimation was performed using R.

Calculation of translational efficiency

Ribosome profiling or mRNA-Seq reads were counted over each region (5'UTR, CDS, and 3'UTR), transcript, or upstream-initiated ORF and then normalized to length and library size as transcripts per million (Li and Dewey 2011). To calculate translational efficiency over a region, transcript or upstream-initiated ORF, ribosome profiling TPM in each biological replicate across each condition was quantile normalized (Amaratunga*, et. al.,* 2001) and then divided by the quantile normalized TPM in mRNA-Seq. Read and RPF counts from mRNA-Seq and ribosome profiling libraries does not include those that overlap the 5'UTR and 3'UTR. Furthermore, to limit the boundary effects due to translation initiation and termination, RPF and read counts do not include those whose A- or P-site adjusted position for harringtonine and cycloheximide libraries, respectively, overlap the first or last 15 nucleotides of an annotated CDS.

Differential expression analysis and gene set enrichment testing in mRNA-Seq

As described previously, the read abundance over annotated protein-coding transcripts was calculated as TPM, then quantile normalized across conditions using the preprocessCore package (Bolstad, 2016) in R (R Core Team, 2017), and then ranked. The change in rank for each gene was calculated across the non-differentiated and RA-differentiated conditions, and the significance of the up- or down-regulation of these rang-changes across conditions was classified using an extreme outlier cutoff (Tukey 1949). Functional characterization of these significantly rank-changed genes across the non-differentiated and RA-differentiated conditions was analyzed using the goseq package (Young, Wakefield et al. 2010) in R, and corrected for multiple testing using Benjami-Hochberg adjusted p-values.

Differential translation analysis and gene set enrichment testing in ribosome profiling

Ribosome profiling read fragments were A- or P-site adjusted, and then counted over annotated protein-coding CDS regions in each biological replicate using the metagene profiles generated by Plastid (Dunn and Weissman 2016). As described previously, ribosome-protected fragments with A- or P-site adjusted positions that overlapped the first or last 15 nucleotides of the boundaries defined by the annotated CDS region were masked from the analysis. DESeq2 (Love, Huber et al. 2014) was used to identify those genes with differential translation across the two states of cellular differentiation. Genes were annotated as significantly up- or down-regulated using a Benjamini-Hochberg adjusted p-value cutoff of 0.1, and fold-change in counts greater than 1, or less than 1, respectively. Functional characterization of these significantly up- and down-regulated genes was analyzed by goseq using parameters specified previously.

Differential translational efficiency and gene set enrichment testing in ribosome profiling

For each biological replicate, ribosome profiling read fragments were A- or P-site adjusted, and then counted over annotated protein-coding CDS regions using the metagene profiles generated by Plastid. As above, read counts over the first and last 15 nucleotides of protein-coding CDS regions were masked for subsequent analyses. In addition, mRNA-Seq read counts were extracted from each condition, with the proximal and terminal 15 nucleotide ends of the CDS masked for consistency with the RPF counts. The DESeq2 wrapper for differential translational efficiency analysis, Riborex (Li, Wang et al. 2017), was used to identify those genes with significant changes in translational efficiency. Genes were annotated as significantly up- or down-regulated using a Benjamini-Hochberg adjusted p-value cutoff of 0.1, and absolute fold-change of 1. Functional characterization of the sets of genes enriched in each condition by translational efficiency was analyzed by goseq using parameters described previously.

Rank-change in transcript abundance

In addition to significance cutoffs for up- and down-regulated regions or transcripts based on mRNA-Seq and ribosome profiling sequence alignments, global changes were assessed by percent change in abundance. TPM based on mRNA-Seq and RPF alignments were calculated, quantile normalized, and then compared by percent change in abundance across the non-differentiated and RA-differentiated libraries. The number of genes with a change in TPM across the two conditions was evaluated at pre-defined percent-change cutoffs of 10%, 20%, 30%, 40% and 50%.

Conservation analysis

To assess the conservation of the various regions, transcripts and ORFs, the PhyloCSF scores (Lin, Jungreis et al. 2011) over each target region was extracted. For upstream-initiated ORFs, the PhyloCSF score was extracted according to its predicted phase. In order to de-convolute the contribution of regional conservation due to overlap with annotated CDS regions, predicted ORFs that did not initiate and terminate wholly upstream of a CDS were also scored according to the subset of their coordinates defined by the 5'UTR alone. The mean PhyloCSF over each of these regions and ORFs was calculated, and then mean-shifted to canonical (+0) reading frame of the annotated CDS for comparison.

GC nucleotide content analysis

Similar to the conservation scores, the ratio of GC nucleotide content in each reading frame of 5'UTRs, CDS, and 3'UTRs was calculated. GC content over the predicted phase of each upstream-initiated was calculated, with the 5'UTR overlapping region of CDS-terminated upstream ORFs deconvoluted from the region overlapping the CDS as described above.

Cluster analysis of differential uORF translation by SPECtre score

In order to identify subsets of upstream-initiated ORFs with differential translation in one state of cell differentiation compared to the other, the SPECtre score for each predicted ORF was calculated (described in Supplemental Materials and Methods). The SPECtre score of each predicted ORF was classified by k-means clustering in R to define sets of uORFs with differential translation in one of the conditions, and those with no difference in translational potential between the two conditions.

Additional filtering of candidate uORFs

Additional replicate-based filtering was applied to the set of predicted uORFs to identify a set of highly confident candidates. As a form of internal validation, a predicted uORF was required to meet a minimal translational threshold in at least one of the technical replicate samples across both conditions. This threshold was determined on a conditional basis dependent on the 5% FDR cutoff required for translational activity according to the distribution of SPECtre scores in protein-coding genes.

Orthogonal proteomic validation of predicted uORFs

To assess their biological reproducibility, predicted uORFs from ribosomal profiling of non-differentiated and RA-differentiated SH-SY5Y cells were cross-referenced against mRNA-Seq and matched mass spectrometry data derived from the transcriptional and translational profiling of 62 AML samples (Battle, Khan et al. 2015). In order to mitigate the effects of transcriptional variation as a potential confounding variable, a common set of genes were identified that were consistently among the top 75% by transcript abundance across all 62 AML samples. The number of genes with a predicted uORF in this set of stably transcribed genes was identified. To assess the potential regulatory effect of predicted uORFs in this common set of genes, the mass spectrometry estimates of protein abundance were normalized by mRNA-Seq transcript abundance. Cumulative distribution plots and density histograms were generated from the normalized protein abundance estimates to assess the potential regulatory contribution of predicted uORFs against a background distribution of genes with no uORF predicted.

Rank-change analysis of conditional translational repression by predicted uORFs

To investigate potential conditional effects of predicted uORFs on the translational repression of annotated CDS regions, the ratio of CDS translational efficiency was calculated as the RA-differentiated replicate TE divided by the non-differentiated replicate TE. The RA-differentiated to non-differentiated CDS TE ratios were then partitioned according to the average nucleotide conservation of the predicted uORF. The top 25% and bottom 25% of predicted cORFs by conservation were extracted, and the significance of their rank-change in TE differences across conditions was calculated using a Wilcoxon Rank-Sum test. The significance in rank-change difference of the top 25% and bottom 25% of predicted oORFs by conservation was assessed separately.

Supplemental Figures

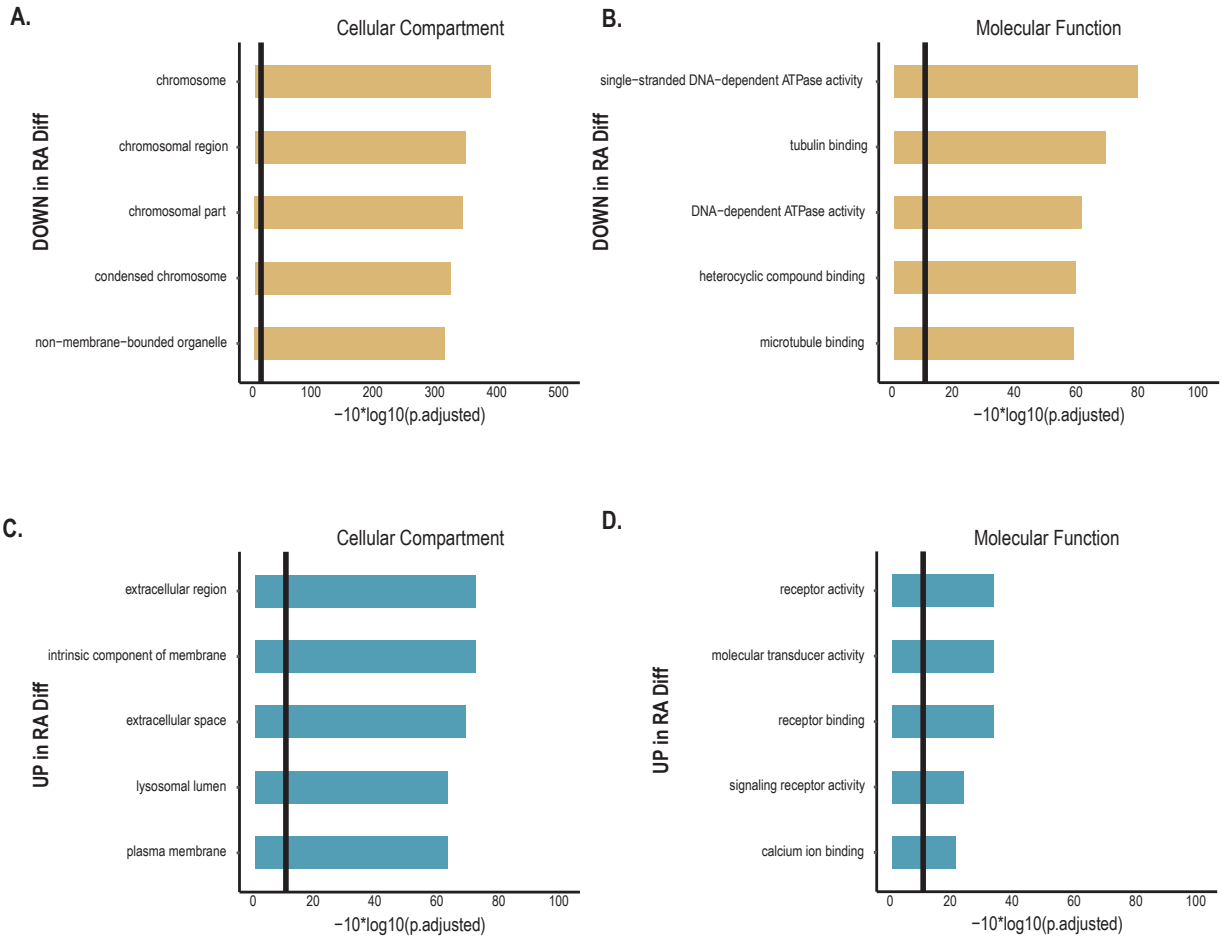


Figure S2.1:

Extended Data for Figure 2.1, #1: Gene sets with significantly down-regulated mRNA transcripts in RA-Diff cells are shown in (A) and (B). Gene sets upregulated in RA-Diff cells are shown in (C) and (D). Genes sets for Cellular Compartment are shown in (A) and (C), and sets for Biological Process are shown in (B) and (D). The top five groups with significant change using a multiple testing corrected p-value cutoff of 0.05 (vertical line) are shown on the graph.

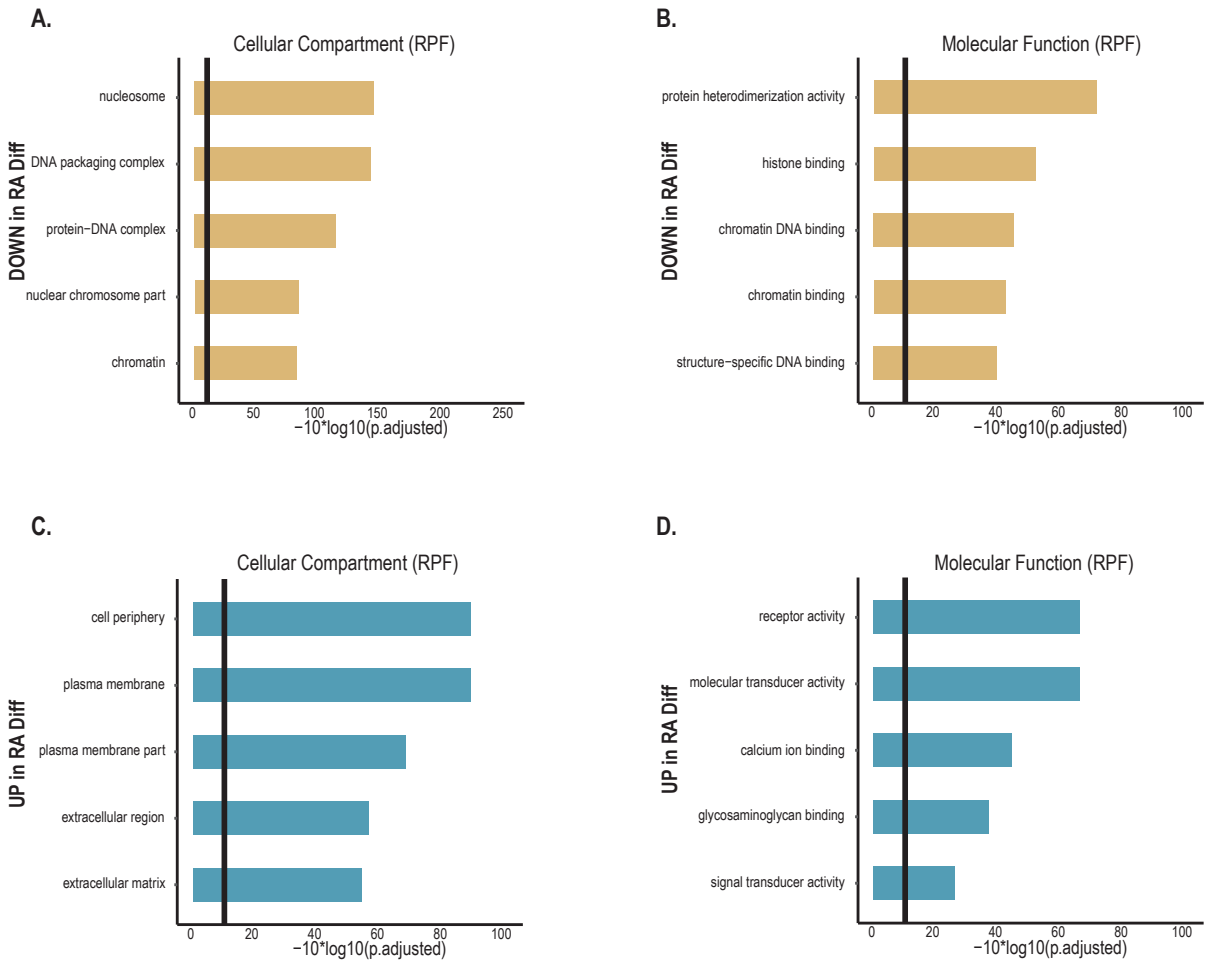


Figure S2.2:

Extended Data for Figure 2.2, #1: Gene sets with significantly down-regulated RPF counts in RA-Diff cells are shown in (A) and (B). Gene sets upregulated in RA-Diff cells are shown in (C) and (D). Genes sets for Cellular Compartment are shown in (A) and (C), and sets for Biological Process are shown in (B) and (D). The top five groups with significant change using a multiple testing corrected p-value cutoff of 0.05 (vertical line) are shown on the graph.

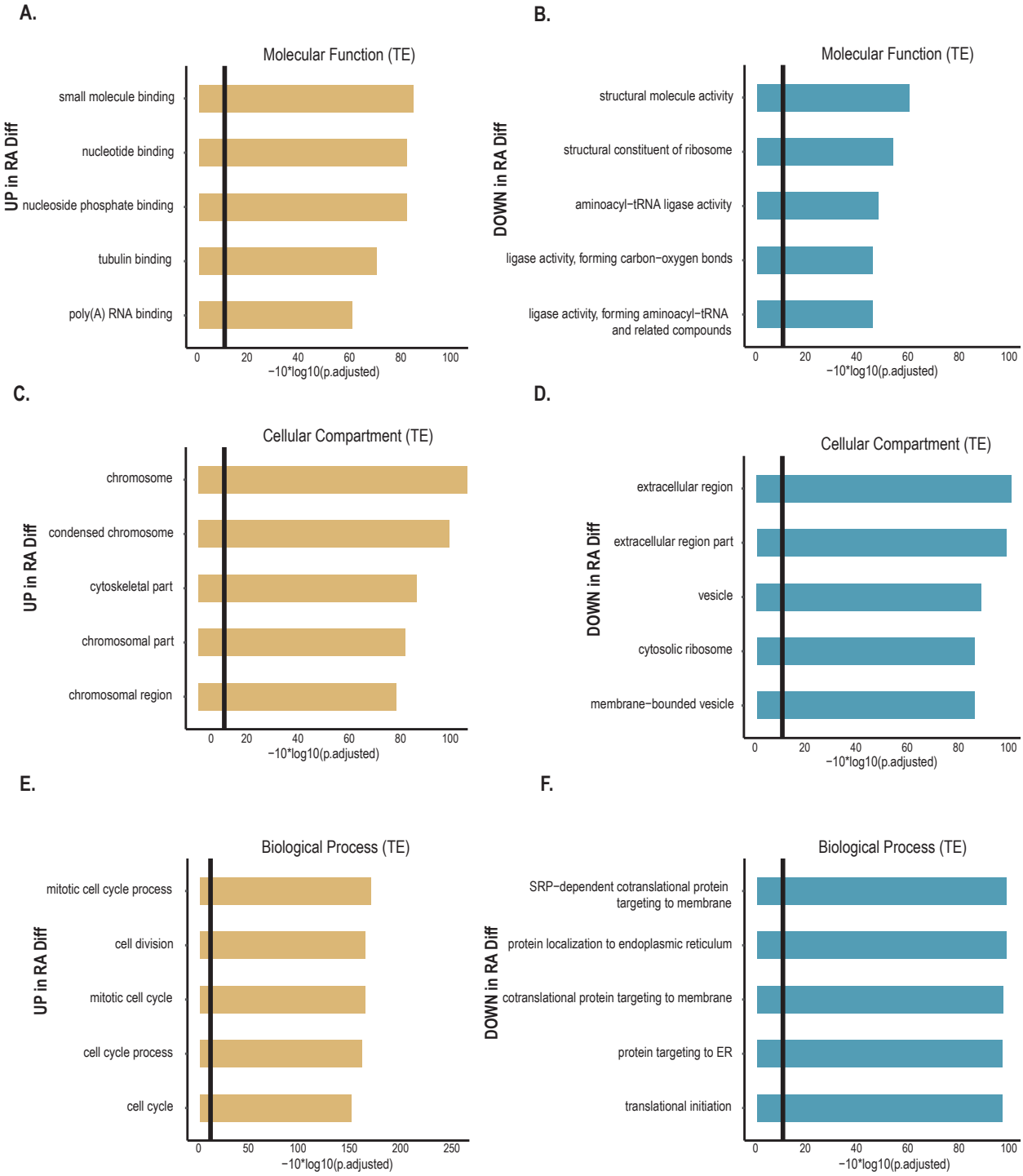


Figure S2.3:

Extended Data for Figure 2.2, #2: Gene sets with significantly down-regulated translational efficiency in RA-Diff cells are shown in (A), (C) and (E). Gene sets upregulated in RA-Diff cells are shown in (B), (D), and (F). Genes sets for Molecular Function are shown in (A) and (B), sets for Cellular Compartment are shown in (C) and (D), and sets for Biological Process are shown in (E) and (F). The top five groups with significant change using a multiple testing corrected p-value cutoff of 0.05 (vertical line) are shown on the graph.

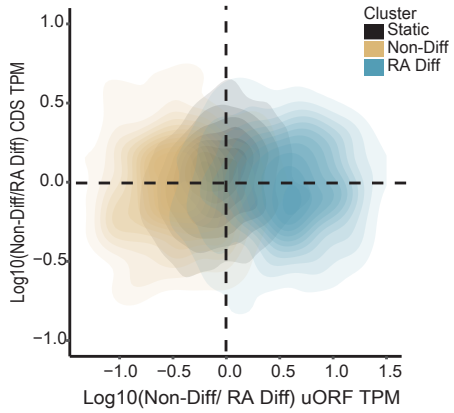


Figure S2.4:

Extended Data for Figure 2.4, #1: uORFs cluster by SPECtre score conditionally, but clustering does not correlate with directional CDS changes. Kernel density estimation analysis of changes in TPM over annotated protein-coding CDS as a function of changes in TPM over predicted upstream-initiated ORFs. Cluster identity of predicted ORF changed in translational potential as scored by SPECtre is identical to panel A): predicted ORFs enriched for translation in RA-Diff cells (cyan), predicted ORFs with enriched translation in Non-Diff cells (gold), and those with static translation across the two conditions (black) are annotated to protein-coding CDS with higher RPF abundance in Non-Diff cells (above horizontal line), and those with higher RPF abundance in RA-Diff cells (below horizontal line).

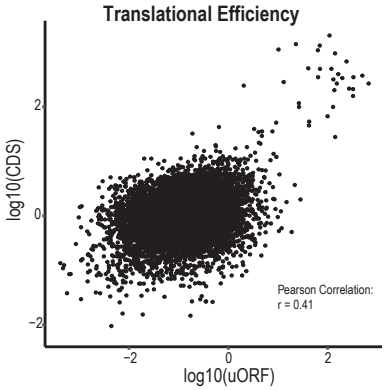


Figure S2.5:

Extended Data for Figure 2.4, #2: Analysis of the full uORF-containing transcript set reveals a positive correlation of uORF TE and CDS TE. Pearson correlation, r=0.41.

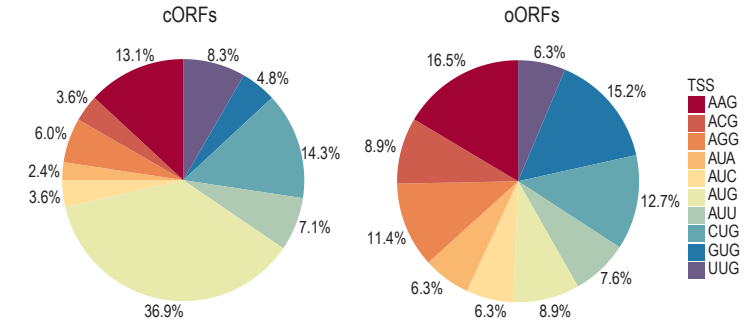


Figure S2.6:

Extended Data for Figure 2.5, #1: Distribution of cORF and oORF translation start sites (TSS) for the transcripts shown in Figure 5.

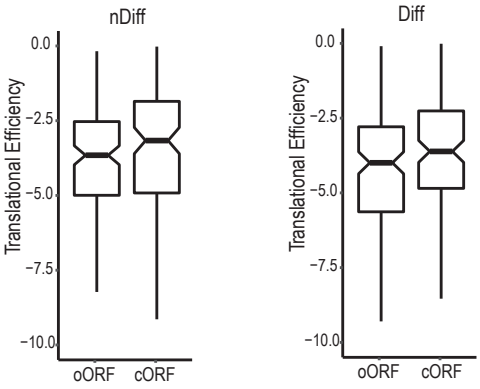


Figure S2.7:

Extended Data for Figure 2.5, #2: Average TE values for cORFs and oORFs in the Non-Diff (left) and RA-Diff (right) conditions reveals no significant difference between the two types of uORFs.

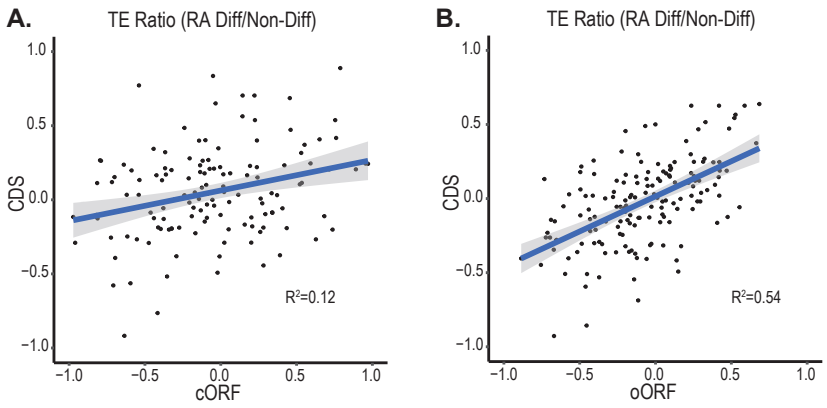


Figure S2.8:

Extended Data for Figure 2.6, #1: A) Using the filtered dataset from Figure 5, analysis of the relationship between cORF and CDS translation (log_{10}TE Non-Diff/RA-Diff) reveals that the translational efficiency of these two regions positively correlate in response to RA-Differentiation, R^2=0.12. B) This is also seen for oORFs, R^2=0.32.

Protein Class: Ribosomal

GENE	uORF	uORF change
RPL30	oORF	Inverse
RPL9	oORF	Inverse
RPS16	oORF	Inverse
RPS18	oORF	Inverse
RPS24	oORF	Inverse
RPL23	oORF	Positive
RPL27	oORF	Positive
RPL27A	oORF	Positive
RPL31	oORF	Positive
RPL34	oORF	Positive
RPL31	oORF	Positive
RPL38	oORF	Positive
RPS13	oORF	Positive
RPS15	oORF	Positive
RPS19	oORF	Positive
RPS21	oORF	Positive
RPS25	oORF	Positive
RPS27A	oORF	Positive
RPS4X	oORF	Positive
RPS8	oORF	Positive
RPL13A	**cORF**	Positive
RPL41	**cORF**	Positive
RPS14	**cORF**	Positive
RPS3	**cORF**	Positive

Figure S2.9:

Extended Data for Figure 2.6, #2: Full list of ribosomal transcripts enriched in the constrained dataset for oORFs. 4 ribosomal transcripts contain cORFs and are listed as well. While most of these transcripts show a positive CDS TE:uORF TE relationship, 20 of the 23 are repressed in the Non-Diff state.

Citations

E Aronesty. ea-utils: Command-line tools for processing biological sequencing data. https://github.com/ExpressionAnalysis/ea-utils

Banerjee, A. K. (1980). "5'-terminal cap structure in eucaryotic messenger ribonucleic acids." Microbiol Rev **44**(2): 175-205.

Battle, A., Z. Khan, S. H. Wang, A. Mitrano, M. J. Ford, J. K. Pritchard and Y. Gilad (2015). "Genomic variation. Impact of regulatory variation from RNA to protein." Science **347**(6222): 664-667.

Bazzini, A. A., T. G. Johnstone, R. Christiano, S. D. Mackowiak, B. Obermayer, E. S. Fleming, C. E. Vejnar, M. T. Lee, N. Rajewsky, T. C. Walther and A. J. Giraldez (2014). "Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation." Embo j **33**(9): 981-993.

Blair, J. D., D. Hockemeyer, J. A. Doudna, H. S. Bateup and S. N. Floor (2017). "Widespread Translational Remodeling during Human Neuronal Differentiation." Cell Rep **21**(7): 2005-2016.

Blank, H. M., R. Perez, C. He, N. Maitra, R. Metz, J. Hill, Y. Lin, C. D. Johnson, V. A. Bankaitis, B. K. Kennedy, R. Aramayo and M. Polymenis (2017). "Translational control of lipogenic enzymes in the cell cycle of synchronous, growing yeast cells." Embo j **36**(4): 487-502.

Brar, G. A. (2016). "Beyond the Triplet Code: Context Cues Transform Translation." Cell **167**(7): 1681-1692.

Brar, G. A., M. Yassour, N. Friedman, A. Regev, N. T. Ingolia and J. S. Weissman (2012). "High-resolution view of the yeast meiotic program revealed by ribosome profiling." Science **335**(6068): 552-557.

Calviello, L., N. Mukherjee, E. Wyler, H. Zauber, A. Hirsekorn, M. Selbach, M. Landthaler, B. Obermayer and U. Ohler (2016). "Detecting actively translated open reading frames in ribosome profiling data." Nat Methods **13**(2): 165-170.

Calvo, S. E., D. J. Pagliarini and V. K. Mootha (2009). "Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans." Proc Natl Acad Sci U S A **106**(18): 7507-7512.

Cheever, T. R. and J. M. Ervasti (2013). "Actin isoforms in neuronal development and function." Int Rev Cell Mol Biol **301**: 157-213.

Chew, G. L., A. Pauli and A. F. Schier (2016). "Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish." Nat Commun **7**: 11663.

Chugunova, A., T. Navalayeu, O. Dontsova and P. Sergiev (2018). "Mining for Small Translated ORFs." **17**(1): 1-11.

Chun, S. Y., C. M. Rodriguez, P. K. Todd and R. E. Mills (2016). "SPECtre: a spectral coherence--based classifier of actively translated transcripts from ribosome profiling sequence data." BMC Bioinformatics **17**(1): 482.

Crappe, J., E. Ndah, A. Koch, S. Steyaert, D. Gawron, S. De Keulenaer, E. De Meester, T. De Meyer, W. Van Criekinge, P. Van Damme and G. Menschaert (2015). "PROTEOFORMER: deep proteome coverage through ribosome profiling and MS integration." Nucleic Acids Res **43**(5): e29.

Dunn, J. G. and J. S. Weissman (2016). "Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data." BMC Genomics **17**(1): 958.

Fields, A. P., E. H. Rodriguez, M. Jovanovic, N. Stern-Ginossar, B. J. Haas, P. Mertins, R. Raychowdhury, N. Hacohen, S. A. Carr, N. T. Ingolia, A. Regev and J. S. Weissman (2015). "A Regression-Based Analysis of Ribosome-Profiling Data Reveals a Conserved Complexity to Mammalian Translation." Mol Cell **60**(5): 816-827.

Fijalkowska, D., S. Verbruggen, E. Ndah, V. Jonckheere, G. Menschaert and P. Van Damme (2017). "eIF1 modulates the recognition of suboptimal translation initiation sites and steers gene expression via uORFs." Nucleic Acids Res **45**(13): 7997-8013.

Fujii, K., Z. Shi, O. Zhulyn and N. Denans (2017). "Pervasive translational regulation of the cell signalling circuitry underlies mammalian development." **8**: 14443.

Gerashchenko, M. V., A. V. Lobanov and V. N. Gladyshev (2012). "Genome-wide ribosome profiling reveals complex translational regulation in response to oxidative stress." Proc Natl Acad Sci U S A **109**(43): 17394-17399.

Guimaraes, J. C. and M. Zavolan (2016). "Patterns of ribosomal protein expression specify normal and malignant human cells." Genome Biol **17**(1): 236.

Hanada, M., S. Krajewski, S. Tanaka, D. Cazals-Hatem, B. A. Spengler, R. A. Ross, J. L. Biedler and J. C. Reed (1993). "Regulation of Bcl-2 oncoprotein levels with differentiation of human neuroblastoma cells." Cancer Res **53**(20): 4978-4986.

Harding, H. P., I. Novoa, Y. Zhang, H. Zeng, R. Wek, M. Schapira and D. Ron (2000). "Regulated translation initiation controls stress-induced gene expression in mammalian cells." Mol Cell **6**(5): 1099-1108.

Harding, H. P., Y. Zhang, H. Zeng, I. Novoa, P. D. Lu, M. Calfon, N. Sadri, C. Yun, B. Popko, R. Paules, D. F. Stojdl, J. C. Bell, T. Hettmann, J. M. Leiden and D. Ron (2003). "An integrated stress response regulates amino acid metabolism and resistance to oxidative stress." Mol Cell **11**(3): 619-633.

Hershey, J. W., N. Sonenberg and M. B. Mathews (2012). "Principles of translational control: an overview." Cold Spring Harb Perspect Biol **4**(12).

Hinds, H. L., C. T. Ashley, J. S. Sutcliffe, D. L. Nelson, S. T. Warren, D. E. Housman and M. Schalling (1993). "Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome." Nat Genet **3**(1): 36-43.

Hinnebusch, A. G. (1984). "Evidence for translational regulation of the activator of general amino acid control in yeast." Proc Natl Acad Sci U S A **81**(20): 6442-6446.

Hinnebusch, A. G. (1997). "Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome." J Biol Chem **272**(35): 21661-21664.

Hinnebusch, A. G., I. P. Ivanov and N. Sonenberg (2016). "Translational control by 5'-untranslated regions of eukaryotic mRNAs." Science **352**(6292): 1413-1416.

Ingolia, N. T., G. A. Brar, N. Stern-Ginossar, M. S. Harris, G. J. Talhouarne, S. E. Jackson, M. R. Wills and J. S. Weissman (2014). "Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes." Cell Rep **8**(5): 1365-1379.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman and J. S. Weissman (2009). "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling." Science **324**(5924): 218-223.

Ingolia, N. T., L. F. Lareau and J. S. Weissman (2011). "Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes." Cell **147**(4): 789-802.

Jang, C., N. F. Lahens, J. B. Hogenesch and A. Sehgal (2015). "Ribosome profiling reveals an important role for translational control in circadian gene expression." Genome Res **25**(12): 1836-1847.

Janich, P., A. B. Arpat, V. Castelo-Szekely, M. Lopes and D. Gatfield (2015). "Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames." Genome Res **25**(12): 1848-1859.

Johnstone, T. G., A. A. Bazzini and A. J. Giraldez (2016). "Upstream ORFs are prevalent translational repressors in vertebrates." Embo j **35**(7): 706-723.

Kaplan, D. R., K. Matsumoto, E. Lucarelli and C. J. Thiele (1993). "Induction of TrkB by retinoic acid mediates biologic responsiveness to BDNF and differentiation of human neuroblastoma cells. Eukaryotic Signal Transduction Group." Neuron **11**(2): 321-331.

Kearse, M. G., K. M. Green, A. Krans, C. M. Rodriguez, A. E. Linsalata, A. C. Goldstrohm and P. K. Todd (2016). "CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins." Mol Cell **62**(2): 314-322.

Kelleher, D. J. and R. Gilmore (1997). "DAD1, the defender against apoptotic cell death, is a subunit of the mammalian oligosaccharyltransferase." Proc Natl Acad Sci U S A **94**(10): 4994-4999.

Kondrashov, N., A. Pusic, C. R. Stumpf, K. Shimizu, A. C. Hsieh, J. Ishijima, T. Shiroishi and M. Barna (2011). "Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning." Cell **145**(3): 383-397.

Korecka, J. A., R. E. van Kesteren, E. Blaas, S. O. Spitzer, J. H. Kamstra, A. B. Smit, D. F. Swaab, J. Verhaagen and K. Bossers (2013). "Phenotypic characterization of retinoic acid differentiated SH-SY5Y cells by transcriptional profiling." PLoS One **8**(5): e63862.

Kozak, M. (1980). "Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits." Cell **19**(1): 79-90.

Kozak, M. (1986). "Influences of mRNA secondary structure on initiation by eukaryotic ribosomes." Proc Natl Acad Sci U S A **83**(9): 2850-2854.

Kozak, M. (1987). "Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes." Mol Cell Biol **7**(10): 3438-3445.

Kozak, M. (1989). "Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems." Mol Cell Biol **9**(11): 5073-5080.

Kozak, M. (1990). "Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes." Proc Natl Acad Sci U S A **87**(21): 8301-8305.

Kozak, M. (1991). "Structural features in eukaryotic mRNAs that modulate the initiation of translation." J Biol Chem **266**(30): 19867-19870.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Li, B. and C. N. Dewey (2011). "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome." BMC Bioinformatics **12**: 323.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, W., W. Wang, P. J. Uren, L. O. F. Penalva and A. D. Smith (2017). "Riborex: fast and flexible identification of differential translation from Ribo-seq data." Bioinformatics **33**(11): 1735-1737.

Lin, M. F., I. Jungreis and M. Kellis (2011). "PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions." Bioinformatics **27**(13): i275-282.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Mehdi, H., E. Ono and K. C. Gupta (1990). "Initiation of translation at CUG, GUG, and ACG codons in mammalian cells." Gene **91**(2): 173-178.

Micheva, K. D., A. Vallee, C. Beaulieu, I. M. Herman and N. Leclerc (1998). "beta-Actin is confined to structures having high capacity of remodelling in developing and adult rat cerebellum." Eur J Neurosci **10**(12): 3785-3798.

Na, C. H., M. A. Barbhuiya, M. S. Kim, S. Verbruggen, S. M. Eacker, O. Pletnikova, J. C. Troncoso, M. K. Halushka, G. Menschaert, C. M. Overall and A. Pandey (2018). "Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini." Genome Res **28**(1): 25-36.

Nanda, J. S., Y. N. Cheung, J. E. Takacs, P. Martin-Marcos, A. K. Saini, A. G. Hinnebusch and J. R. Lorsch (2009). "eIF1 controls multiple steps in start codon recognition during eukaryotic translation initiation." J Mol Biol **394**(2): 268-285.

Pahlman, S., A. I. Ruusala, L. Abrahamsson, M. E. Mattsson and T. Esscher (1984). "Retinoic acid-induced differentiation of cultured human neuroblastoma cells: a comparison with phorbolester-induced differentiation." Cell Differ **14**(2): 135-144.

Parola, A. L. and B. K. Kobilka (1994). "The peptide product of a 5' leader cistron in the beta 2 adrenergic receptor mRNA inhibits receptor synthesis." J Biol Chem **269**(6): 4497-4505.

Peabody, D. S. (1987). "Translation initiation at an ACG triplet in mammalian cells." J Biol Chem **262**(24): 11847-11851.

Peabody, D. S. (1989). "Translation initiation at non-AUG triplets in mammalian cells." J Biol Chem **264**(9): 5031-5035.

Pelechano, V., W. Wei and L. M. Steinmetz (2013). "Extensive transcriptional heterogeneity revealed by isoform profiling." Nature **497**(7447): 127-131.

Pelletier, J. and N. Sonenberg (1985). "Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency." Cell **40**(3): 515-526.

Pezzini, F., L. Bettinetti, F. Di Leva, M. Bianchi, E. Zoratti, R. Carrozzo, F. M. Santorelli, M. Delledonne, M. Lalowski and A. Simonati (2017). "Transcriptomic Profiling Discloses Molecular and Cellular Events Related to Neuronal Differentiation in SH-SY5Y Neuroblastoma Cells." **37**(4): 665-682.

Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J. B. Veyrieras, M. Stephens, Y. Gilad and J. K. Pritchard (2010). "Understanding mechanisms underlying human gene expression variation with RNA sequencing." Nature **464**(7289): 768-772.

Pieroh, P., G. Birkenmeier and F. Dehghani (2014). "The temporal and spatial dynamics of glyoxalase I following excitoxicity and brain ischaemia." Biochem Soc Trans **42**(2): 534-537.

Rabadan-Diehl, C., A. Martinez, S. Volpi, S. Subburaju and G. Aguilera (2007). "Inhibition of vasopressin V1b receptor translation by upstream open reading frames in the 5'-untranslated region." J Neuroendocrinol **19**(4): 309-319.

Rebbapragada, I. and J. Lykke-Andersen (2009). "Execution of nonsense-mediated mRNA decay: what defines a substrate?" Curr Opin Cell Biol **21**(3): 394-402.

Sidell, N. (1982). "Retinoic acid-induced growth inhibition and morphologic differentiation of human neuroblastoma cells in vitro." J Natl Cancer Inst **68**(4): 589-596.

Slavoff, S. A., A. J. Mitchell, A. G. Schwaid, M. N. Cabili, J. Ma, J. Z. Levin, A. D. Karger, B. A. Budnik, J. L. Rinn and A. Saghatelian (2013). "Peptidomic discovery of short open reading frame-encoded peptides in human cells." Nat Chem Biol **9**(1): 59-64.

Sonenberg, N. and A. G. Hinnebusch (2009). "Regulation of translation initiation in eukaryotes: mechanisms and biological targets." Cell **136**(4): 731-745.

Spealman, P., A. Naik, G. May, S. Kuersten, L. Freebert, R. Murphy and J. McManus (2017). "Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data." Genome Res.

Thireos, G., M. D. Penn and H. Greer (1984). "5' untranslated sequences are required for the translational control of a yeast regulatory gene." Proc Natl Acad Sci U S A **81**(16): 5096-5100.

Trapnell, C., L. Pachter and S. L. Salzberg (2009). "TopHat: discovering splice junctions with RNA-Seq." Bioinformatics **25**(9): 1105-1111.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold and L. Pachter (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.

Tukey, J. W. (1949). "Comparing individual means in the analysis of variance." Biometrics **5**(2): 99-114.

Valasek, L., K. H. Nielsen, F. Zhang, C. A. Fekete and A. G. Hinnebusch (2004). "Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection." Mol Cell Biol **24**(21): 9437-9455.

Van Vactor, D. V. and L. J. Lorenz (1999). "Neural development: The semantics of axon guidance." Curr Biol **9**(6): R201-204.

Vattem, K. M. and R. C. Wek (2004). "Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells." Proc Natl Acad Sci U S A **101**(31): 11269-11274.

Xue, S. and M. Barna (2012). "Specialized ribosomes: a new frontier in gene regulation and organismal biology." Nat Rev Mol Cell Biol **13**(6): 355-369.

Xue, S., S. Tian, K. Fujii, W. Kladwang, R. Das and M. Barna (2015). "RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation." Nature **517**(7532): 33-38.

Young, M. D., M. J. Wakefield, G. K. Smyth and A. Oshlack (2010). "Gene ontology analysis for RNA-seq: accounting for selection bias." Genome Biol **11**(2): R14.

Zitomer, R. S., D. A. Walthall, B. C. Rymond and C. P. Hollenberg (1984). "Saccharomyces cerevisiae ribosomes recognize non-AUG initiation codons." Mol Cell Biol **4**(7): 1191-1197.

CHAPTER III

RAN Translation regulates FMRP synthesis and can be selectively targeted in Fragile X Disorders [2]

Abstract

CGG repeat expansions in *FMR1* impair fragile X protein (FMRP) expression, resulting in autism and intellectual disability in Fragile X Syndrome. In contrast, repeat associated non-AUG translation (RANT) from transcribed CGG repeat expansions produces toxic proteins that drive neurodegeneration in Fragile X-associated Tremor/Ataxia Syndrome (FXTAS). Here we describe fundamental roles for RANT at normal CGG repeat sizes in regulating FMRP synthesis. RANT impairs basal FMRP synthesis both *in vitro* and in neurons. This inhibition requires conserved near-AUG codons in the 5' UTR. Activation of FMRP synthesis by mGluR agonists requires both RANT and the CGG repeat. Targeting RANT initiation sites with antisense oligonucleotides (ASOs) blocks RANT and prevents neuronal death. Moreover, ASO treatment of human Fragile X neurons enhances endogenous FMRP synthesis. These findings define a native function for RANT and CGG repeats in regulating FMRP synthesis and delineate RANT as a therapeutic target across all Fragile X spectrum disorders.

[2] Most of this chapter was submitted for publication as: Rodriguez CM, Kearse MG, Haenfler JM, Flores BN, Lui Y, Wright SE, Ifrim MF, Glineburg MR, Jafar-Nejad, P, Sutton MA, Bassell GJ, Rigo F, Parent JM, Barmada SJ, Todd PK. RAN Translation regulates FMRP synthesis and can be targeted in Fragile X Disorders.

Introduction

Fragile X Syndrome (FXS) is the leading monogenic cause of autism and intellectual disability (Hagerman et al., 2017). It results from large (>200) CGG nucleotide repeat expansions in the 5' untranslated region (UTR) of the FMR1 gene present on the X chromosome FMRP plays a central role in regulating activity-dependent protein synthesis at synapses (Bhakar et al., 2012) . FMRP is an RNA-binding protein that interacts with ~4% of brain mRNAs, including an enriched fraction of synaptic transcripts from genes associated with autism (Ascano et al., 2012; Brown et al., 2001; Darnell et al., 2011). FMRP suppresses translation of target transcripts, either through direct interactions with targets or via association with translating ribosomes (Chen et al., 2014; Darnell et al., 2011; Feng et al., 1997). Upon activation of group I metabotropic glutamate receptors (mGluRs), FMRP is dephosphorylated and rapidly degraded, allowing for local translation of FMRP-associated mRNAs (Ceman et al., 2003; Hou et al., 2006; Nalavadi et al., 2012). *FMR1* mRNA itself is rapidly translated in response to mGluR activation, creating a negative feedback loop that temporally limits activity-dependent translation (Hou et al., 2006; Todd et al., 2003; Weiler et al., 1997). This activity-dependent regulation of neuronal FMRP synthesis requires specific sequences in the 3' UTR (Suhl et al., 2015). Whether the CGG repeat or *FMR1* 5' UTR contribute to this regulation is unclear.

In contrast to FXS, intermediate "premutation" sized CGG repeat expansions in the *FMR1* locus elicit a clinically distinct neurological disorder: Fragile X-associated Tremor/Ataxia Syndrome (FXTAS). Repeats in patients with FXTAS range from 55-200. FXTAS is an age-related neurodegenerative disease characterized by gait difficulties, action tremor, and dementia with executive dysfunction and, in some cases, Parkinsonism and peripheral neuropathy (Berry-Kravis et al., 2007). This un-methylated premutation repeat enhances *FMR1* transcription and

triggers formation of large ubiquitinated intranuclear neuronal inclusions in the brains of FXTAS patients (Greco et al., 2002; Tassone et al., 2007; Todd et al., 2010). Premutation repeats also cause Fragile X-associated premature ovarian insufficiency (FXPOI), the most common inherited cause of early menopause (Sullivan et al., 2011). Together with Fragile X Syndrome, these conditions and other CGG repeat-associated clinical phenotypes are collectively referred to as Fragile X spectrum disorders to capture the full range of potential pathology that results from CGG repeat expansions at this locus (Lozano et al., 2014).

Transcribed CGG repeats likely elicit neurodegeneration through multiple mechanisms, including RNA-mediated sequestration of proteins, DNA damage response activation by R-loop formation, and mitochondrial dysfunction (Hagerman and Hagerman, 2016; Kong et al., 2017). Recently, focus has been placed on the ability of CGG repeats to support repeat-associated non-AUG translation (RANT) in all three reading frames to produce toxic homopolymeric polyglycine (FMRPolyG), polyalanine (FMRpolyA), and polyarginine (FMRpolyR) proteins (Kearse et al., 2016; Todd et al., 2013). The most abundant RANT product, FMRpolyG, forms intranuclear neuronal inclusions in FXTAS patient brains and aggregates in FXPOI patient ovaries (Buijsen et al., 2016; Hukema et al., 2015; Sellier et al., 2017; Todd et al., 2013). Significant evidence suggests a direct role for FMRpolyG in disease pathogenesis. Enhancing FMRpolyG production exacerbates CGG repeat associated toxicity in cell and *Drosophila* model systems and impairs protein quality control pathways (Oh et al., 2015; Todd et al., 2013). In contrast, precluding FMRpolyG production markedly suppresses inclusion production in both mice and *Drosophila* models and is necessary for CGG repeats to elicit neurodegeneration in transgenic mice (Sellier et al., 2017; Todd et al., 2013).

RANT is a non-canonical form of translational initiation originally described at expanded CAG and CUG repeats in the context of the Spinocerebellar Ataxia 8 locus (Zu et al., 2011). RANT has now been observed in seven repeat expansion disorders, including *C9orf72*-associated amyotrophic lateral sclerosis and frontotemporal dementia and Huntington's disease (Ash et al., 2013; Banez-Coronel et al., 2015; Cleary and Ranum, 2017; Mori et al., 2013; Zu et al., 2011). At CGG repeats, initiation of this non-canonical form of translation is a cap- and scanning-dependent event that requires initiation factors eIF4E and eIF4A (Kearse et al., 2016). However, it specifically bypasses the requirement of an AUG start codon; instead, it relies on initiation predominantly at near-AUG codons that reside just 5' to the repeat or, in one reading frame (FMRpolyA), within the repeat itself.

While there is mounting evidence that RANT occurs at expanded repeats and can contribute to toxicity, potential roles for this alternative translational initiation mechanism in normal physiology are unexplored (Brar, 2016; Kearse and Wilusz, 2017). RANT of FMRpolyG occurs readily at the prominent repeat size in the human population (25-30 CGG repeats) (Kearse et al., 2016; Todd et al., 2013). Given its location within the *FMR1* 5' UTR, we tested the hypothesis that RAN translation at CGG repeats regulates FMRP synthesis. Here we describe how RANT of both normal and expanded CGG repeats regulates basal and activity-dependent FMRP synthesis. This regulation occurs in both human cell lines and within neuronal processes where *FMR1* is typically located, with specific requirements for both the CGG repeat and conserved non-AUG codons that allow for RANT initiation. Treatment with non-cleaving antisense oligonucleotides (ASOs) that sterically hinder RANT initiation suppresses neurotoxicity from repeats and to enhances endogenous FMRP synthesis in human induced pluripotent stem cell (iPSC)-derived neurons with a large unmethylated CGG repeat expansion.

Overall, these findings define a new non-pathogenic physiological role for RANT in regulating neuronal FMRP synthesis, and provide a novel therapeutic approach in multiple Fragile X spectrum disorders.

Results

RAN translation initiation at CGG repeats impairs FMRP synthesis

We previously determined the sites of RANT initiation within the 5' UTR of *FMR1*. Initiation in the 0-reading frame (relative to the AUG start codon of the FMRP open reading frame) upstream of the repeat would generate an N-terminally extended FMRP containing a polyarginine tract (FMRpolyR-FMRP). Initiation in this frame occurs at an ACG codon 60 nucleotides upstream of the first CGG repeat codon. Generation of this product is suppressed at both normal and larger repeat sizes (Kearse *et al*., 2016), and is expressed at the lowest level of the three potential reading frames as measured by luciferase assays (**Figure S3.1A**, (Kearse et al., 2016)). The most abundant RAN product, FMRpolyG, initiates in the +1 reading frame at one of two near-cognate codons: an ACG codon at 35 nucleotides upstream of the CGG repeat and a GUG codon at 11 nucleotides upstream the CGG repeat to create an out-of-frame, overlapping upstream open reading frame (uORF) that ends after the AUG start codon of FMRP. Initiation in the +2 reading frame, FMRpolyA, occurs within the repeat itself at larger repeat sizes and terminates prior to the AUG start codon of FMRP, creating a uORF with the potential for re-initiation (Hinnebusch et al., 2016; Kearse et al., 2016).

To determine whether CGG RANT might have a conserved function, we first explored the sequence conservation of the 5' UTR. The CGG repeat element in the 5' UTR of *FMR1* is highly conserved among mammals and becomes expanded in primates (Eichler et al., 1995).

Outside of the repeat itself, the 5' UTR of *FMR1* in humans is almost invariant. A large-scale analysis of novel point mutations in *FMR1* that might recapitulate FXS revealed no polymorphisms in the 5' UTR outside of the repeat (Collins et al., 2010). Within mammals, all three RANT start sites are conserved in reading frame relative to the CGG repeat and to the AUG of the FMRP main open reading frame (ORF) to armadillos (**Figure 3.1**). The mean vertebrate phyloP score of the 0-frame ACG is 0.39, and for the ACG and GUG of the +1 frame the scores are 1.60 and 1.48, respectively. These values represent slower relative evolution of all three RANT start sites and a high level of conservation of the +1 start sites amongst vertebrates, and suggest a physiological, rather than pathophysiological, role for RANT(Pollard et al., 2010).

Figure 3.1:

RANT start sites are conserved among mammals. Interspecies conservation of near-cognate codon RANT start sites in the 0/native (arginine; green) and +1 (glycine; orange) reading frames of the 5' UTR of FMR1. The CGG repeat is highlighted in blue. The AUG start codon for the FMRP open reading frame (yellow) is 69 nucleotides downstream of the repeat. All three initiation sites and their reading frames in reference to the repeat and the native frame FMRP start codon are conserved in mammals.

Assays of global protein synthesis by ribosome profiling demonstrated that translation initiation at near-AUG codons within the 5' UTR of mRNAs is quite common, with overlapping uORFs typically acting as negative regulators of translation from the main ORF (Brar et al., 2012; Ingolia et al., 2012; Ingolia et al., 2011). In published datasets from both man and mouse,

we previously detected peaks on the two near-AUG codons in the +1 reading frame associated with initiation of FMRpolyG translation (Ingolia et al., 2012; Ingolia et al., 2011; Todd et al., 2013). Re-analysis of these datasets identified a third uORF initiation site at the 0-frame ACG in mouse embryonic stem cells treated with harringtonine (a drug that stalls ribosomes at initiation sites), suggesting that this initiation event occurs in mice at normal repeat sizes (Ingolia et al., 2011).

To test the impact of these RAN events on FMRP synthesis, we generated two sets of reporters. The first set serves as a measure of FMRP synthesis with the full 5' UTR and the first coding exon of FMRP fused to a modified nanoluciferase (nLuc) with a GGG in place of its start codon (**Figure 3.2A**). The second set serves as a measure of RANT with GGG-nLuc in the 0, +1 or +2 frames of the full *FMR1* 5' UTR (Kearse et al, 2016). Both reporter sets were modified to contain a range of CGG repeats from 0 to 100. Consistent with previous results (Chen et al., 2003; Ludwig et al., 2011), increasing the number of CGG repeats in our FMRP reporter negatively impacts FMRP synthesis in *in vitro* translation assays using rabbit reticulocyte lysate (RRL) as measured by nLuc activity (**Figure 3.2B**). However, western blot analysis provides a more complete picture of how the CGG repeat impacts translation in the native frame. There are two species generated from the FMRP reporter: one at the expected molecular weight of 24.5kDa (lower band) and an upper band representing an N-terminal extension initiating upstream of the repeat. Expression of both wild-type FMRP and the N-terminal extension is inversely proportional to the CGG repeat size, with most of the detectable N-terminal extension absent at normal repeat sizes (28 CGGs) and beyond. Consistent with this, the N-terminal extension is less than 1% of the total measured FMRP reporter signal in HEK293 cells at normal and expanded repeat sizes, and is undetectable by western blot (**Figure S3.1A**).

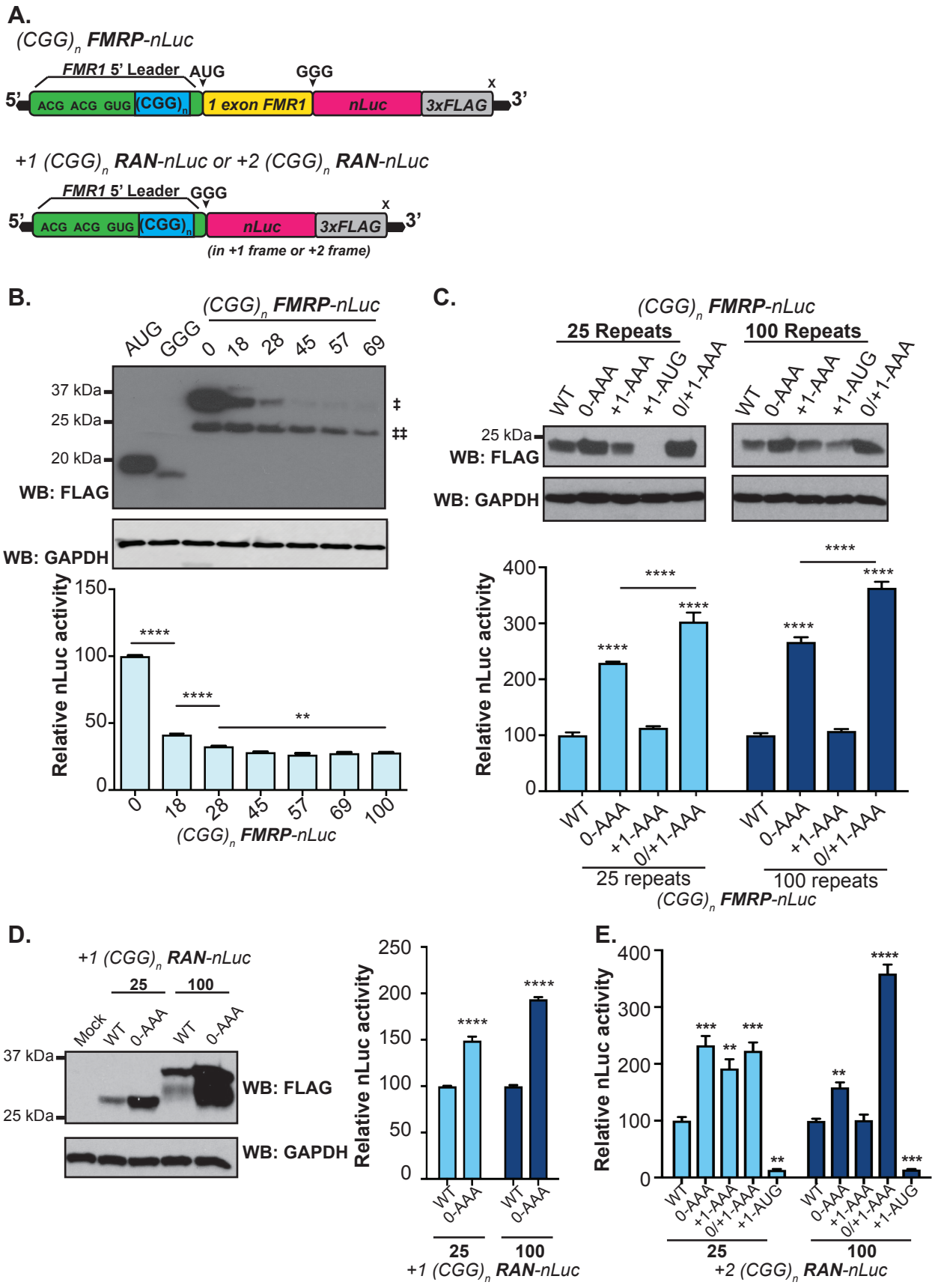

A.
(CGG)n FMRP-nLuc
FMR1 5' Leader
AUG
GGG
5'
ACG ACG GUG
(CGG)n
1 exon FMR1
nLuc
3xFLAG
x
3'
+1 (CGG)n RAN-nLuc or +2 (CGG)n RAN-nLuc
FMR1 5' Leader
GGG
5'
ACG ACG GUG
(CGG)n
nLuc
3xFLAG
x
3'
(in +1 frame or +2 frame)
B.
(CGG)n FMRP-nLuc
AUG
GGG
0
18
28
45
57
69
37 kDa
25 kDa
20 kDa
‡
‡‡
WB: FLAG
WB: GAPDH
Relative nLuc activity
150
100
50
0
****
****
**
0
18
28
45
57
69
100
(CGG)n FMRP-nLuc
C.
(CGG)n FMRP-nLuc
25 Repeats
100 Repeats
WT
0-AAA
+1-AAA
+1-AUG
0/+1-AAA
25 kDa
WB: FLAG
WB: GAPDH
Relative nLuc activity
400
300
200
100
0
****
****
****
****
****
****
WT
0-AAA
+1-AAA
0/+1-AAA
25 repeats
100 repeats
(CGG)n FMRP-nLuc
D.
+1 (CGG)n RAN-nLuc
25
100
Mock
WT
0-AAA
WT
0-AAA
37 kDa
25 kDa
WB: FLAG
WB: GAPDH
Relative nLuc activity
250
200
150
100
50
0
****
****
WT
0-AAA
WT
0-AAA
25
100
+1 (CGG)n RAN-nLuc
E.
Relative nLuc activity
400
300
200
100
0
***
**
***
**
**
****
***
WT
0-AAA
+1-AAA
0/+1-AAA
+1-AUG
WT
0-AAA
+1-AAA
0/+1-AAA
+1-AUG
25
100
+2 (CGG)n RAN-nLuc

Figure 3.2:

RANT impacts the translational efficiency of FMRP *in vitro*. A) Schematic of FMRP and RAN nanoluciferase (nLuc) reporter. The AUG of the nLuc coding sequence (pink) is mutated to GGG and placed in-frame with the first coding exon of FMRP (yellow) downstream of the full 5' UTR of *FMR1* (green) to act as a reporter for FMRP ($(CGG)_n$ FMRP-nLuc). Alternatively, GGG-nLuc was inserted into the 5' UTR of *FMR1* within different reading frames to serve as a reporter for RANT ($(CGG)_n$ RAN-nLuc). The reading frame for each RANT reporter is altered by inserting +1(FMRpolyG) or +2(FMRpolyA) nucleotide frameshifts between the repeat and nLuc. Repeat size and sites of RAN initiation codon mutations are indicated. A carboxyl-terminal 3xFLAG tag (gray) is included for antibody-based detection. B) Top: Western blot of *in vitro* translated $(CGG)_n$ FMRP-nLuc reporter mRNAs harboring different CGG repeat sizes. Bottom: Luciferase activity of *in vitro* translated FMRP-nLuc reporters with CGG repeat sizes. ‡‡: FMRP-nanoluciferase-3X FLAG protein. ‡: N-terminal extension of FMRP from initiation in the (polyarginine) 0-frame. The addition of increasing repeats leads to a significant reduction in FMRP until 45 repeats where the FMRP levels remain steady as more repeats are added. C) Top: Western blot of the $(CGG)_n$ FMRP-nLuc reporters with 25 or 100 CGG repeats with mutation of either the 0 or +1 initiation sites in isolation or in combination. +1-AUG represents insertion of AUG in place of ACG RAN initiation codon in the +1-reading frame. Bottom: nLuc Assay of 25 and 100 repeat $(CGG)_n$ FMRP-nLuc reporters with indicated mutations. A synergistic increase in FMRP reporter product is seen when all three near-AUG start sites are mutated to block RAN initiation (0/+1-AAA). D) Left: Western blot of *in vitro* translated +1 $(CGG)_n$ RAN-nLuc reporters. Mutating the +0 frame ACG initiation site (0-AAA) in +1 $(CGG)_n$ RAN-nLuc reporters increases translation in the +1 reading frame. Right: Luciferase activity of *in vitro* translated +1 $(CGG)_n$ RAN-nLuc reporters. E) 0-AAA and +1-AAA initiation codon mutations enhance RANT in the +2(FMRpolyA) reading frame by nLuc assay *in vitro*. These effects are synergistic at expanded repeat sizes. All nLuc assays were performed in triplicate in 3 or more independent experiments. One-way ANOVA was performed for panel C, and within repeat groups (25 or 100) for panels D, E, and F, **p>0.01, ****p<0.0001. All graphs represent mean +/- S.E.M.

The near-cognate start sites utilized for 0-frame (ACG) and +1 frame RAN initiation (ACG and GUG) were mutated to determine whether they influence FMRP synthesis. The single 0-frame start site was mutated to an AAA codon (0-AAA). The first +1 start site was also mutated to an AAA while the second was mutated to a TAG (+1-AAA). The 0/+1-AAA mutant has all three of these mutations to prevent RAN initiation in both reading frames. *In vitro* translation of the 0-AAA mutant FMRP reporter with no CGG repeats confirms that a large portion of the nanoluciferase signal is derived from an N-terminal extension to FMRP, and mutation of the 0-frame initiation site decreases its overall expression (**Figure S3.1B**). This suggests one reason for the CGG repeat is to direct initiation to the main, downstream AUG. Mutating the +1 frame start sites increases FMRP reporter synthesis in the absence of any repeat element. Mutating both reading frames (0/+1-AAA) also increases expression from the FMRP reporter, suggesting that removing the negative impact of the +1 reading frame is sufficient to

offset the loss of the N-terminal initiation event. This is visualized via western blot, where mutants with decreased signal in the N-terminal extension (upper band) have increase signal in the wild-type FMRP band (lower band). The continued presence of an N-terminal extension band in the absence of the 0-frame ACG codon suggests alternative non-AUG codons may be utilized in this *in vitro* system even when no repeat is present.

To evaluate the impact of these initiation events in the presence of a repeat, we added 25 or 100 CGG repeats to these mutant FMRP reporters and performed *in vitro* translation assays. Interestingly, the removal of 0-frame RAN translation increases FMRP translation in both 25 and 100 repeats (**Figure 3.2C**). In isolation, the +1 reading frame has no effect on FMRP synthesis alone. However, mutating both reading frames (0/+1-AAA) leads to a synergistic increase in FMRP translation at both normal and expanded repeat sizes. These data implicate potential coordination between RANT in the two reading frames in modulating downstream translation of the FMRP open reading frame.

As a second test of whether upstream initiation events can impair FMRP synthesis, we placed AUG codons in either the +1 or +2 reading frames. Both AUG codons severely decrease the signal from our FMRP reporters, with the +1 AUG mutations exhibiting greater inhibition of FMRP reporter synthesis than the +2 AUG and greater inhibition with larger repeats (**Figure S3.1C**). As a third test, we introduced stop codons in multiple reading frames either upstream or downstream of the CGG repeat in our FMRP reporters (**Figure S3.1D**). Stop sites were placed in all three reading frames upstream of the repeat (All STOP up) or all three reading frames downstream of the repeat (All STOP down). Placing stop codons in all three reading frames upstream of the CGG repeat significantly increased FMRP reporter expression. In contrast, placing stop sites downstream of the repeat in the three frames slightly decreased reporter protein

levels (**Figure S3.1E**). This may reflect loss of the contribution from the N-terminal extension that occurs *in vitro*. Overall, these studies are consistent with a direct role for RAN translation in negatively regulating FMRP production *in vitro*.

To understand whether there is interplay between RANT reading frames, we utilized RANT specific reporters (Kearse et al, 2016) with RAN initiation sites mutated serially across reading frames in each reporter. Mutating the 0-frame ACG to an AAA leads to a significant increase in +1 frame RAN translation in both 25 and 100 CGG repeat-containing reporters (**Figure 3.2D**). Next, we introduced the 0-AAA, +1-AAA, 0/+1-AAA, or +1-AUG mutations to the +2 frame RAN reporters. Precluding RAN initiation in the 0- and +1-reading frames increases translation in the +2-reading frame at both normal and expanded repeats (**Figures 3.2E**, **Figure S3.1F**). This effect was greatest for 0-AAA and 0/+1-AAA mutants, and was enhanced in the setting of expanded repeats. In contrast, driving translation in the +1 frame with an AUG upstream of the repeat markedly decreases production of the +2 product. In conjunction with previous data demonstrating that CGG RAN translation relies on the m^7G cap, eIF4E, and the eIF4A helicase (Kearse et al., 2016; Ludwig et al., 2011), these experiments are predominantly consistent with a scanning mode of translational initiation for both RANT and FMRP. RAN initiation in each reading frame is inhibited by previous upstream initiation events and all events negatively regulate FMRP synthesis.

RAN translation negatively regulates neuronal FMRP synthesis

Our initial experiments assessing interplay between RAN translation and FMRP synthesis were conducted in a non-neuronal *in vitro* system without direct relevance to Fragile X spectrum disorders. To determine the effect of RAN translation on FMRP synthesis in a more relevant cell

type, FMRP reporters were transfected into mature rat hippocampal neurons. Unlike the *in vitro* condition, neither the 0-AAA nor the +1-AAA mutants alone led to a significant increase in FMRP (**Figure 3.3A**). However, the combined 0/+1-AAA mutant had a greater than two-fold enhancement in the FMRP reporter signal (**Figure 3.3A**). This was true at both normal and expanded repeats, as well as in the absence of any CGG repeats (**Figure S3.2 A**). Similar to the *in vitro* studies, placing an AUG start codon in the +1 frame upstream of the repeat decreased FMRP reporter abundance (**Figure S3.2 B**).

To evaluate whether a human cell would also demonstrate this inhibitory effect of RANT on FMRP synthesis, we transfected SH-SY5Y human neuroblastoma cells with these same reporters (**Figure 3.3B**). As in neurons, mutations that precluded RANT in the 0- and +1 reading frames markedly enhanced FMRP reporter levels in SH-SY5Y cells. This was not due to altered RNA expression or stability from the different constructs, as RT-qPCR studies demonstrated no difference in mRNA abundance between control and 0/+1-AAA mutant reporters (**Figure 3.3C**). Our reporters do show an increase in mRNA abundance with a larger repeat, recapitulating past findings that larger repeats exhibit elevated transcription (**Figure S3.2C**) (Ofer et al., 2009; Tassone et al., 2007). To determine the spatial distribution and abundance of the reporter protein in transfected neurons, we performed immunocytochemistry in neurons transfected with the WT or 0/+1-AAA FMRP reporters (**Figure 3.3D-E**). As with the nLuc assay, we observed significantly increased signal in neurons with the 0/+1-AAA mutations. This effect was seen in both the soma and in dendritic processes where a portion of FMRP synthesis normally occurs (**Figure 3.3F**).

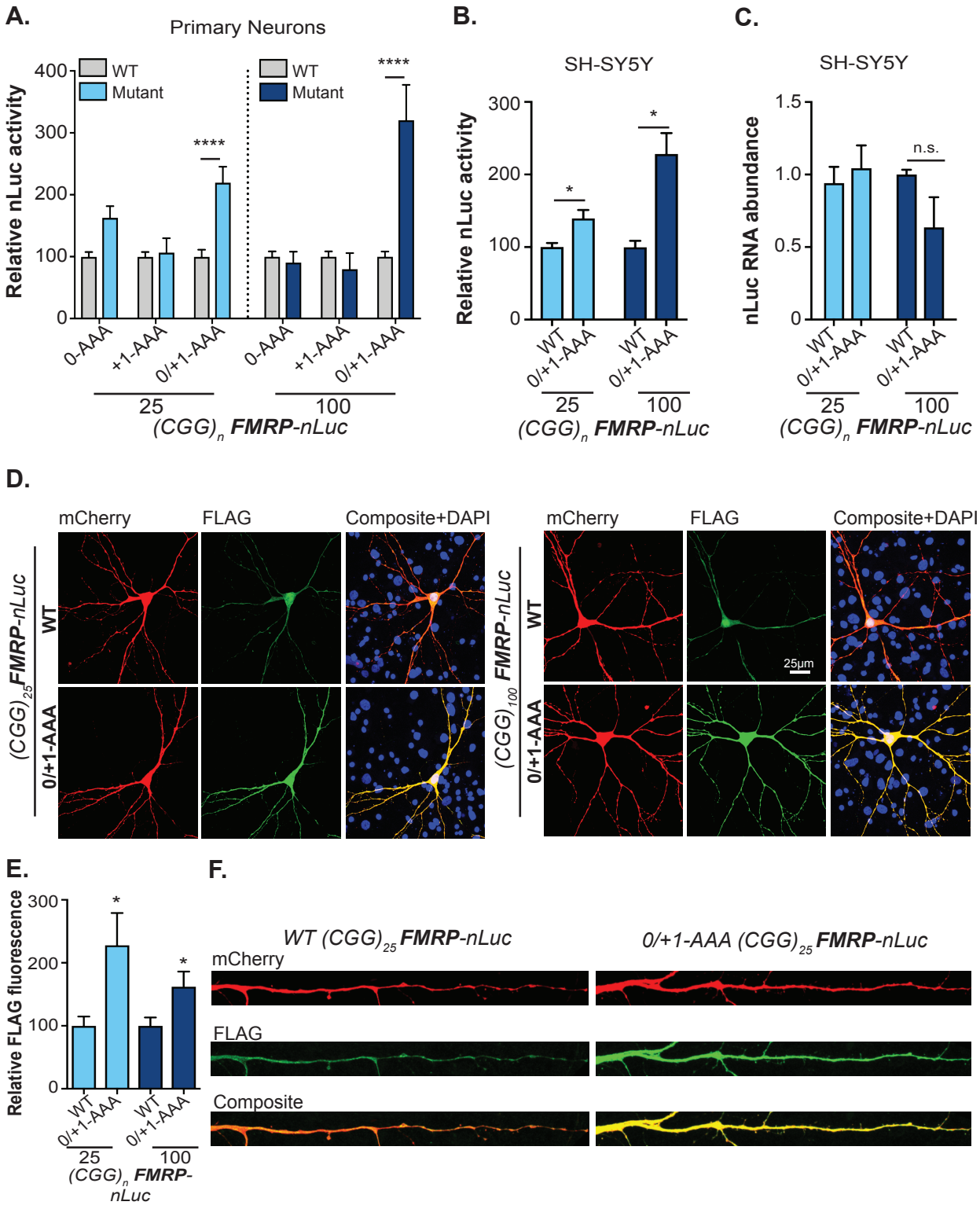

A.
Primary Neurons
Relative nLuc activity
WT
Mutant
0-AAA
+1-AAA
0/+1-AAA
25
100
(CGG)n FMRP-nLuc
B.
SH-SY5Y
Relative nLuc activity
C.
SH-SY5Y
nLuc RNA abundance
n.s.
D.
mCherry
FLAG
Composite+DAPI
(CGG)25 FMRP-nLuc
(CGG)100 FMRP-nLuc
25μm
E.
Relative FLAG fluorescence
F.
WT (CGG)25 FMRP-nLuc
0/+1-AAA (CGG)25 FMRP-nLuc
Composite

Figure 3.3:

RAN translation impedes FMRP translation in neurons. A) nLuc assay in mature (DIV 13) rat hippocampal neurons shows synergistic increase in FMRP reporters when RAN translation is precluded in the two reading frames (0/+1-AAA). n>8 for each condition. B) nLuc Assay demonstrates enhanced FMRP reporter signal in (0/+1-AAA) mutants at normal and expanded repeats in human SH-SY5Y neuroblastoma cells expressing indicated reporters. n=3 for each condition. C) RT-qPCR of SH-SY5Y cells expressing indicated FMRP reporters show no change in steady state reporter mRNA levels between WT and 0/+1-AAA mutant reporters at either 25 or 100 CGG repeats. n=3 for each condition. D) Immunocytochemistry of FLAG tagged (green) reporters transfected into neurons co-expressing mCherry to fill the cell. E) Straightened dendrites of the WT and 0/+1-AAA reporters with 25 repeats shows greater reporter signal in neuronal processes with the RAN mutant reporters. F) Reporter signal was calculated using the corrected total cell fluorescence (CTCF). n=17 for WT and 0/+1-AAA CGG_{25} reporters, n=23 for the WT CGG_{100} reporter and n=21 for the 0/+1-AAA CGG_{100} reporter. All graphs represent mean +/- S.E.M. >3 independent experiments for A-C. All individual replicates are internally normalized to a co-transfected pGL4.13 (firefly luciferase) plasmid. Panel A utilizes a two-way ANOVA with multiple comparisons within repeat groups. Panels B, C, and E utilize a Student's t-test, *p<0.05, ***p<0.001. All graphs represent mean +/- S.E.M.

CGG RAN translation occurs locally in dendrites

If RANT regulates local FMRP synthesis, then it should occur out in neuronal processes where a subset of *FMR1* mRNA is translated. To test this directly, we generated a series of RAN translation specific reporters using a Venus fluorescent protein without an AUG start codon downstream of the 5' UTR of *FMR1* in the +1 RANT reading frame. RAN reporters along with positive (AUG-Venus) and negative (ΔAUG-Venus, no AUG start codon) controls for translation were individually co-transfected alongside the cell-filling mCherry into mature rat hippocampal neurons (**Figure 3.4A**). As expected, AUG-Venus shows high translation that is expressed throughout the entire cell, while removal of the AUG (ΔAUG-Venus) erases any signal from the Venus channel. In contrast, +1 RAN protein products of both 25 and 90 CGG repeats accumulate in aggregates in neuronal soma and dendrites, with higher steady state expression of reporters with larger repeats (**Figure 3.4B**).

To decipher if RAN translation occurs in neuronal processes we performed single-molecule imaging using a protocol specifically designed to detect translation events in distal apical dendrites (Ifrim et al., 2015). Event maps were derived using the +1 RAN-Venus reporters with 25 and 90 CGG repeats (**Figure 3.4C**). Dendrites were located using the pre-bleach Venus

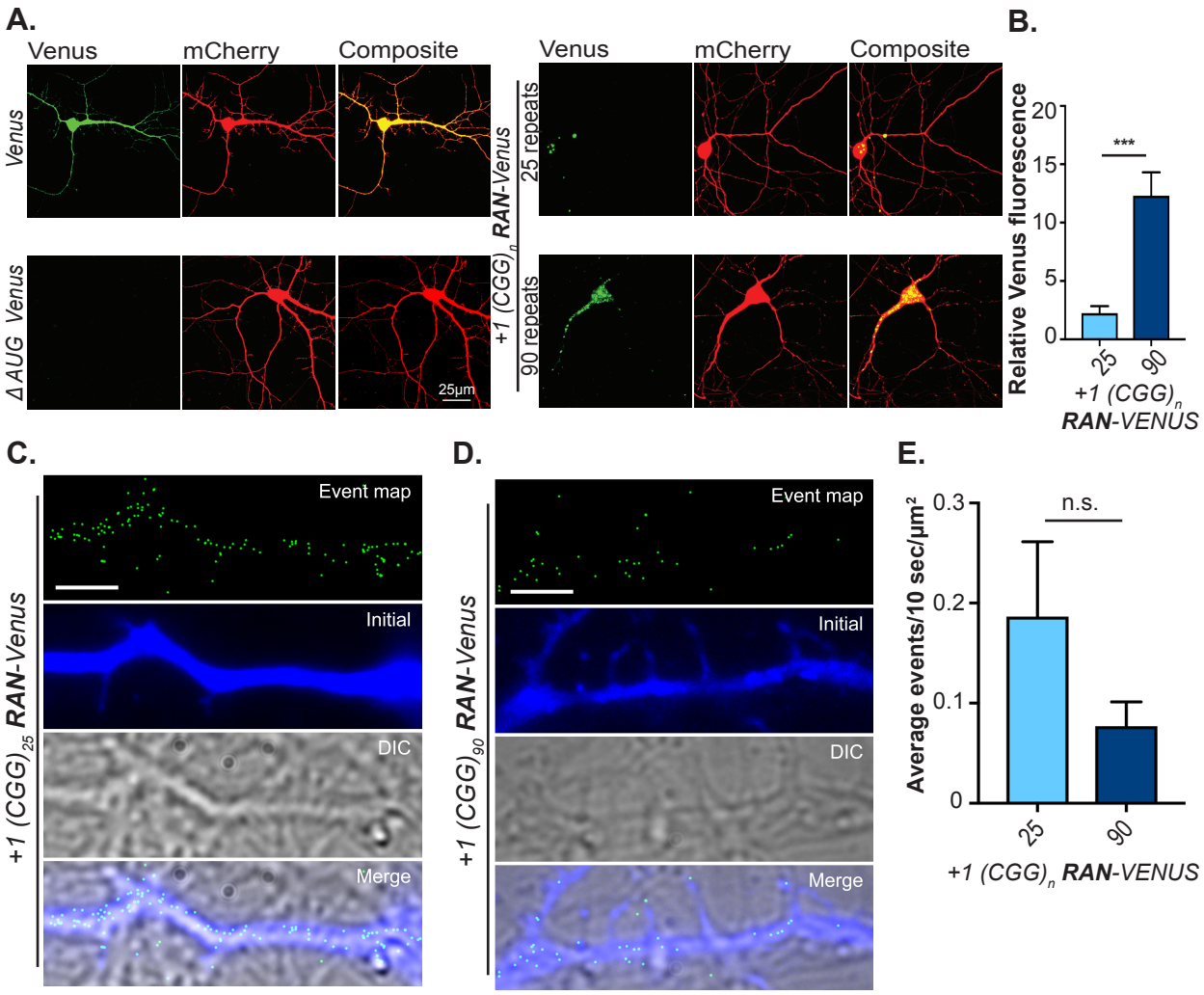


Figure 3.4:

RAN translation occurs in distal dendrites. A) Live imaging of mature rat hippocampal neurons co-transfected with RANT specific Venus reporters and mCherry. Top left: Venus alone. Bottom left: Venus without an AUG start codon (ΔAUG Venus). Top right: +1 $(CGG)_{25}$ RAN-Venus. Bottom right: +1 $(CGG)_{90}$ RAN-Venus. Live imaging revealed aggregation of FMRpolyG at large (90) and small (25) repeat sizes in soma and processes. B) +1 RAN-Venus harboring expanded CGG (90) repeats are expressed more than reporters with normal length repeats. C, D) Single molecule imaging detects translation events in distal neuronal processes of neurons transfected with +1 $(CGG)_n$ RAN-Venus reporters with 25 (C) or 90 (D) repeats. E) Quantification of RANT events in processes demonstrates no significant difference between repeat sizes. n=3 individual distal dendrites per transfection. Student's t-test, n.s.=not significant, ***p<0.001. All graphs represent mean +/- S.E.M.

signal, and single molecule events were detected post-bleaching. Importantly, we readily detected isolated RANT events within dendrites at both normal and expanded repeat sizes in the

+1 reading frame. However, we observed no significant difference in translation rates between normal and expanded repeat sizes (**Figure 3.4D**). Given that there is both more transcript overall in neurons expressing the larger repeat and higher basal +1 RAN peptide abundance at larger repeats in these same neurons, as well as in other cell-based and *in vitro* assays (Kearse et al, 2016), this finding of a similar dendritic synthesis rate suggests that RANT at normal repeat sizes may be favored in neuronal processes.

CGG repeats and RAN translation are required for mGluR dependent FMRP synthesis

Induction of lasting changes in synaptic function often requires localized activity-dependent neuronal protein synthesis(Cajigas et al., 2010; Costa-Mattioli et al., 2009; Ho et al., 2011). Activation of metabotropic glutamate receptors (mGluRs) triggers synaptic protein synthesis (Weiler and Greenough, 1993). This new protein synthesis is required for critical aspects of neurophysiology, including the maintenance of mGluR induced long-term depression (LTD) and mGLuR elicited epileptiform discharges (Huber et al., 2000; Merlin et al., 1998). FMRP plays an active role in regulating protein synthesis: its loss leads to increased basal protein synthesis in neurons and it is required for mGluR-triggered translation of FMRP target mRNAs (Osterweil et al., 2010; Qin et al., 2005; Tang et al., 2015; Todd et al., 2003). Consistent with this, loss of FMRP in rodents leads to enhanced mGluR-dependent LTD, as well as behavioral phenotypes including learning dysfunction, lowered seizure threshold, and enhanced anxiety and autistic behaviors (Dolen et al., 2007; Huber et al., 2002). These phenotypes can be largely corrected by genetic or pharmacologic attenuation of mGluR receptor signaling (Dolen et al., 2007). FMRP itself is degraded and then rapidly resynthesized in response to group I mGluR (mGluR1) receptor activation in neurons and synaptoneurosomes (Hou et al., 2006; Nalavadi et

al., 2012; Todd et al., 2003; Weiler et al., 1997). The loss and then return of FMRP coincides with a critical window for protein synthesis-dependent plasticity elicited by mGluR activation (Huber et al., 2000; Merlin et al., 1998), suggesting that it may provide negative feedback to temporally constrain local synthesis events in neurons (Bear et al., 2004; Todd and Malter, 2002).

To explore how RANT impacts activity-dependent FMRP synthesis, we evaluated the impact of mGluR activation on endogenous FMRP and our FMRP and RANT reporters. Consistent with published data(Todd et al., 2003), application of the mGluR1 agonist dihydrophenylglycine (DHPG, 50μM) to hippocampal rat neurons in culture significantly increased endogenous FMRP expression by 56% after 30 minutes of treatment (**Figure 3.5A**). FMRP and +1 RAN nanoluciferase reporters were modified with a C-terminal PEST degron sequence to stimulate reporter protein turnover, lowering the stable background of the reporter and allowing for enhanced detection of changes in expression over short timeframes (**Figure 3.5B**). In addition, the 3' UTR of *FMR1* was included as previous work demonstrated that it was required for mGluR-dependent synthesis of FMRP (Suhl et al., 2015).

First, we investigated whether our transfected reporters could recapitulate DHPG-induced changes in endogenous FMRP expression. FMRP-nLuc-3' UTR reporters bearing either 20 or 90 CGG repeats both exhibited a significant increase in activity after 30 minutes of DHPG treatment, similar to that observed with endogenous FMRP (**Figure 3.5C**). In contrast, reporters without the CGG repeat (CGG_0) showed no response to DHPG, suggesting a specific role for the repeat element in mGluR-triggered FMRP translation. To understand this further, we first replaced the CGG repeat with a GAA_{25} repeat. This modification maintains the length of the 5'

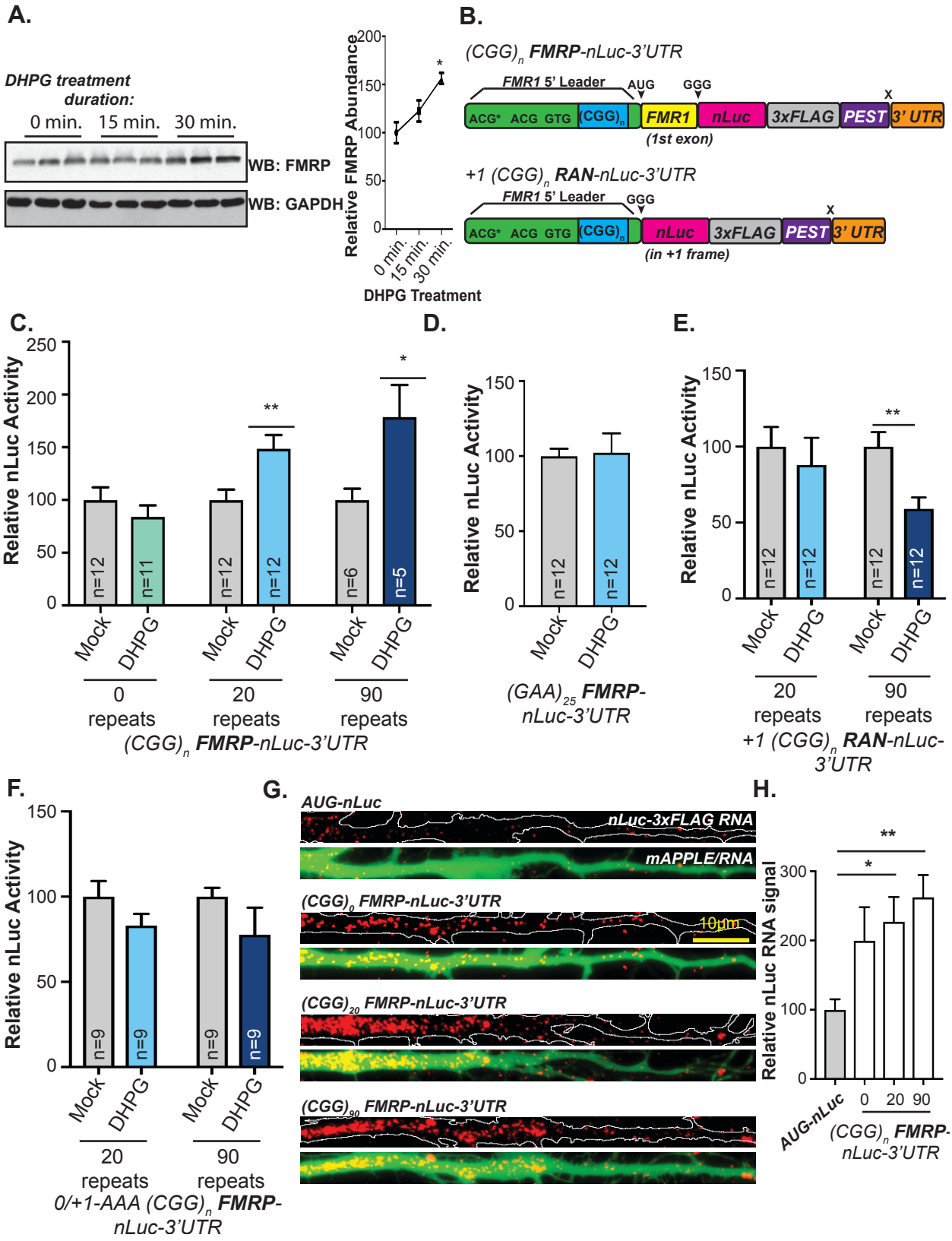

A.
DHPG treatment duration:
0 min.
15 min.
30 min.
WB: FMRP
WB: GAPDH
Relative FMRP Abundance
DHPG Treatment
B.
(CGG)n FMRP-nLuc-3'UTR
FMR1 5' Leader
AUG
GGG
ACG* ACG GTG
(CGG)n
FMR1
nLuc
3xFLAG
PEST
3' UTR
(1st exon)
+1 (CGG)n RAN-nLuc-3'UTR
(in +1 frame)
C.
Relative nLuc Activity
Mock
DHPG
0 repeats
20 repeats
90 repeats
(CGG)n FMRP-nLuc-3'UTR
D.
(GAA)25 FMRP-nLuc-3'UTR
E.
+1 (CGG)n RAN-nLuc-3'UTR
F.
0/+1-AAA (CGG)n FMRP-nLuc-3'UTR
G.
AUG-nLuc
nLuc-3xFLAG RNA
mAPPLE/RNA
(CGG)0 FMRP-nLuc-3'UTR
10μm
(CGG)20 FMRP-nLuc-3'UTR
(CGG)90 FMRP-nLuc-3'UTR
H.
Relative nLuc RNA signal
AUG-nLuc
0
20
90
(CGG)n FMRP-nLuc-3'UTR

Figure 3.5:

mGluR dependent FMRP synthesis requires the CGG repeat and RAN translation. A) Left: Western blot of cultured rat hippocampal neurons treated with 50μm DHPG for the indicated times. Right: Quantification of western blots (normalized to GAPDH) shows an increase in endogenous FMRP at 30 minutes of treatment versus sham treated cells. B) Schematic of FMRP and +1 RAN reporter constructs modified to add a PEST degron sequence and the 3' UTR of *FMR1* to allow for appropriate mRNA trafficking. (C) NLuc assays in rat hippocampal neurons 30 minutes after mock or DHPG treatment. No DHPG response was seen in constructs lacking a CGG repeat ($(CGG)_0$). Both 20 and 90 CGG repeat FMRP reporter constructs exhibited an increase in signal in response to DHPG. For all nLuc assays, each individual replicate is internally normalized to co-transfected pGL4.13 (firefly luciferase). D) Replacing the CGG repeat with a $(GAA)_{25}$ repeat prevents DHPG dependent increases in the FMRP reporter. E) RAN translation reporter (+1 $(CGG)_n$ RAN-nLuc) shows no response to DHPG at normal repeat sizes (20) and a significant decrease after DHPG treatment at expanded repeats (90). F) Mutations that prevent RANT initiation (0/+1-AAA) prevent the DHPG-induced response of FMRP reporters at both normal and expanded repeat sizes. G) Hybridization Chain Reaction (HCR) of nLuc mRNA combined with ICC to co-transfected mApple verifies dendritic export of reporter RNAs. H) Quantification of dendritic nLuc mRNA by HCR. FMRP reporters with 20 and 90 repeats exhibit enhanced dendritic mRNA abundance relative to AUG-Nanoluciferase-PEST-3xFLAG. Panels A, H: One-way ANOVA with correction for multiple comparisons, Panels C-F: Student's t-test, *p<0.05, **p<0.01, ***p<0.001. All graphs represent mean +/- S.E.M.

UTR, but removes the stable RNA secondary structure. *In vitro*, this change leads to a 4-fold increase in basal FMRP-nLuc signal (**Figure S3.3A**). Replacing the CGG repeat with GAA in neurons completely precludes increases in reporter expression induced by DHPG (**Figure 3.5D**), suggesting the structure or sequence of the CGG repeat itself is important to this regulatory process.

We next evaluated the impact of mGluR activation on CGG RANT. RANT reporters for FMRPolyG (+1 CGG RAN-nLuc-3' UTR) showed no change after DHPG treatment at normal repeat sizes. In contrast, there was a significant decrease in +1 $(CGG)_{90}$ RAN-nLuc-3' UTR expression after DHPG treatment (**Figure 3.5E**), suggesting that suppression of RANT anti-correlates with FMRP synthesis. To test whether RANT is required for mGluR-induced FMRP synthesis more directly, we introduced 0/+1-AAA mutations into our FMRP reporters to preclude RAN initiation in these reading frames. Blocking RAN initiation prevented DHPG-induced synthesis of the FMRP reporters (**Figure 3.5F**), establishing a function for RANT in regulating mGluR-dependent FMRP translation. Collectively, these data support a model

whereby the CGG repeat and RANT provide a balance of *FMR1* translation events both basally and in response to neuronal activity.

The transport of specific mRNAs into dendrites is required for local protein synthesis (Cajigas et al., 2010; Ho et al., 2011). Thus, we investigated the distribution of the reporter transcripts in transfected neurons. To accomplish this, we adapted a technique known as hybridization chain reaction (HCR) to specifically detect mRNAs containing the nLuc-3xFLAG sequence (**Figure S3.3B**) (Choi et al., 2016; Dirks and Pierce, 2004). Neurons co-expressing FMRP-nLuc-3' UTR reporters and mApple were evaluated by co-HCR/ICC with Cy5-labeled RNA probes (**Figure 3.5G**). There is significant dendritic localization of reporters that include the regulatory regions of FMR1 compared to constructs lacking this sequence, independent of the presence of any repeat element (**Figure 3.5H**), consistent with previous results (Suhl et al., 2015). There is a trending CGG repeat length-dependent increase in the abundance FMRP-nLuc-3' UTR reporter mRNAs in distal dendritic processes from 0 to 90 CGG repeats (Muslimov et al., 2011). However, this difference was largely explained by differences in relative transcription (**Figure S3.2C**). These data suggest that the absence of response to DHPG in CGG_0-FMRP-NLuc 3' UTR reporters is unlikely to result from defects in mRNA localization.

Impeding RAN initiation with an antisense oligonucleotide increases endogenous FMRP

Overlapping uORFs typically function to suppress translation from the main ORF of the transcript in which they reside (Hinnebusch et al., 2016). The translation efficiency of the main ORF in a given transcript can be enhanced by sterically hindering translation initiation at upstream AUG start sites with antisense oligonucleotides (ASOs) (Liang et al., 2016; Liang et al., 2017). To evaluate whether a similar strategy could be used to target RANT we designed a

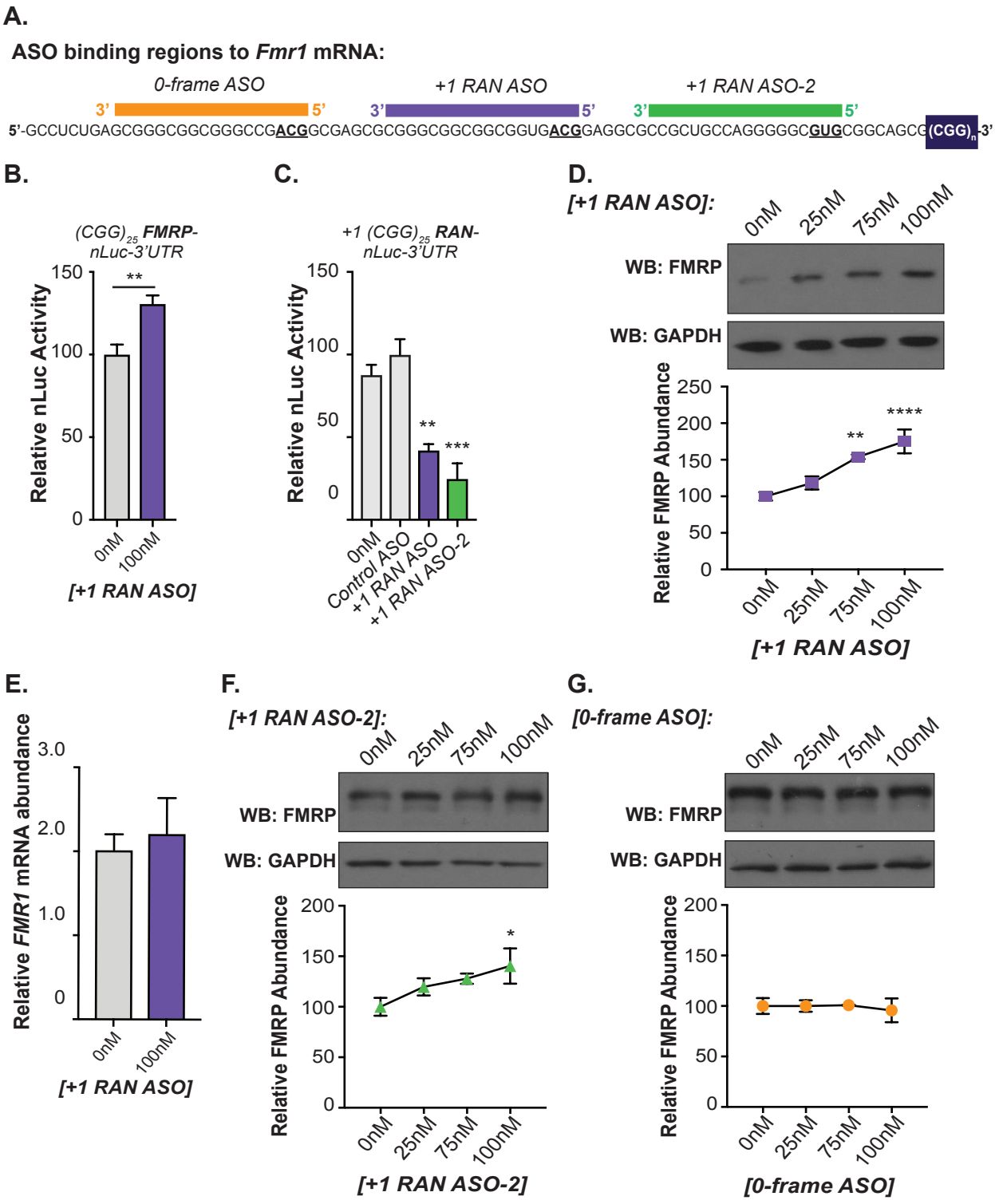


Fig 3.6:

ASO targeting of RAN translation start sites increases FMRP expression in human cells. A) Schematic of non-cleaving, RANT blocking ASOs. Colored bars overlap the corresponding *FMR1* 5' UTR sequence and

start sites; 0 frame ACG (orange), +1 frame ACG (purple), and +1 frame GUG (green). B) A significant increase in the FMRP reporter is detected by nLuc assay of HEK293 cells expressing CGG_{25} FMRP-nLuc-3'UTR mRNA that were subsequently transfected with *+1 RAN ASO* (100nM). n=6 for two independent experiments performed in triplicate. C) Cells expressing the RAN translation reporter mRNA (+1 $(CGG)_{25}$ RAN-nLuc) were subsequently transfected with the indicated ASOs. The *+1 RAN ASO* and *+1 RAN ASO-2*, but not the *Control* ASO, significantly decreased the RAN reporter compared to control. n=5 for each condition. D) Top: FMRP expression assessed by western blot in HEK293 cells transfected with increasing concentrations of *+1 RAN ASO* for 24 hours. Bottom: *+1 RAN ASO* elicits a dose-dependent increase in endogenous FMRP expression. n=3 for each condition. E) RT-qPCR of HEK293 cells transfected with the *+1 RAN ASO* shows no effect on endogenous FMR1 mRNA levels at the effective dose (100nM). n=3 for each condition. F) Treatment of HEK293 cells with *+1 RAN ASO-2*, which targets the GUG initiation site for +1 RAN, also increases endogenous FMRP. n=4 for each condition. G) *0-frame ASO* has no effect on FMRP expression in HEK293 cells. n=3 for each condition. For panel E, G, and H, One-way ANOVA with a Fisher's LSD test for dose dependency, *p<0.05, **p<0.01, ****p<0.0001. For all other panels, Student's t-test, **p<0.01. All graphs represent mean +/- S.E.M.

series of modified ASOs targeting the near-AUG sites utilized for RANT in the 5' UTR of *FMR1*. Six ASOs, harboring 2'-O-methyl modifications and phosphorothioate backbones, were designed to singly overlap with the three RAN start site targeted by the three most 5' nucleotides of the ASO (**Figures 3.6A**, **Figure S3.4A**). We first tested these ASOs to determine their impact on expression of our FMRP and RANT reporters. HEK293 cells were transfected with capped and polyadenylated *in vitro* transcribed CGG_{25} FMRP-NL reporter mRNAs and the ASOs of interest. Treatment with an 18 nucleotide ASO targeting the first +1 reading frame start site (*+1 RAN ASO*) enhanced expression of FMRP as measured by nLuc assay (**Figure 3.6B**). We next determined whether this ASO directly targeted RANT in the +1 reading frame as predicted. Cells were transfected with both +1 CGG_{25} RAN-nLuc-3' UTR reporter mRNA and *+1 RAN ASO* or an 18 nucleotide long ASO targeting the 0-frame RAN initiation site (*0-frame ASO)*. The *+1 RAN ASO* caused a significant decrease in +1 CGG_{25} RAN reporter expression, while *0-frame ASO* had no effect (**Figure 3.6C**).

To determine the impact of these ASOs on endogenous FMRP expression, we titrated the amount of *+1 RAN ASO* transfected into HEK293 cells. 24 hours after transfection, the *+1 RAN ASO* elicited a dose dependent increase in endogenous FMRP expression, peaking at 189% when

treated with 100nM ASO compared to 0nM ASO treatment (**Figure 3.6D-E**). This occurred in the absence of changes in mRNA levels (**Figure 3.6F**). A similar effect on endogenous FMRP was observed in patient-derived fibroblasts transfected with *+1 RAN ASO* (**Figure S3.4B**), while the 16-nucleotide ASO targeting this ACG start site increased FMRP at lower doses in HEK293 cells (**Figure S3.4C**).

Treatment with an ASO targeting the GUG codon used in RANT in the +1 reading frame (+1 RAN ASO-2) was also effective, showing a significant increase in endogenous FMRP at 100nM following 24-hour transfection (**Figure 3.6G**). However, neither the 16- nor 18-nucleotide ASOs targeting the 0-frame (*0 RAN ASO*) had an impact on FMRP expression (**Figures 3.6H**, **Figure S3.4D**). Co-transfecting the *+1 RAN ASO* and *0 RAN ASO* together led to no improvement above what was already seen with *+1 RAN ASO* alone (data not shown). As expected, transfection with a control "non-targeting" ASO had no impact on FMRP expression (**Figure S3.4F**). Interestingly, tiling the two +1 RANT initiation sites with the ASOs that singly increase FMRP synthesis actually decreased FMRP levels in a dose dependent manner when combined in cells (**Figure S3.4E**). This is likely due to an impact on the ability of the translation machinery to reinitiate scanning further down the mRNA transcript (Liang et al., 2017).

ASO blockade of RAN translation suppresses CGG repeat induced neuronal death

RANT in general and FMRpolyG production in particular are required for CGG repeat induced neurodegeneration in both fly and mouse models of FXTAS (Sellier et al., 2017; Todd et al., 2013). To test if the *+1 RAN ASO* also suppresses RANT in RNAs bearing a large repeat, we co-transfected HEK293 cells with both the ASOs specified and +1 $(CGG)_{90}$ nLuc-3' UTR and firefly luciferase reporters (**Figure 3.7A**). The *+1 RAN ASO* elicited a 55% reduction in

FMRpolyG production by luciferase assays, with no change elicited by a control *ASO* treatment.

Similarly, this same ASO strongly suppressed FMRpolyG reporter protein levels by western blot

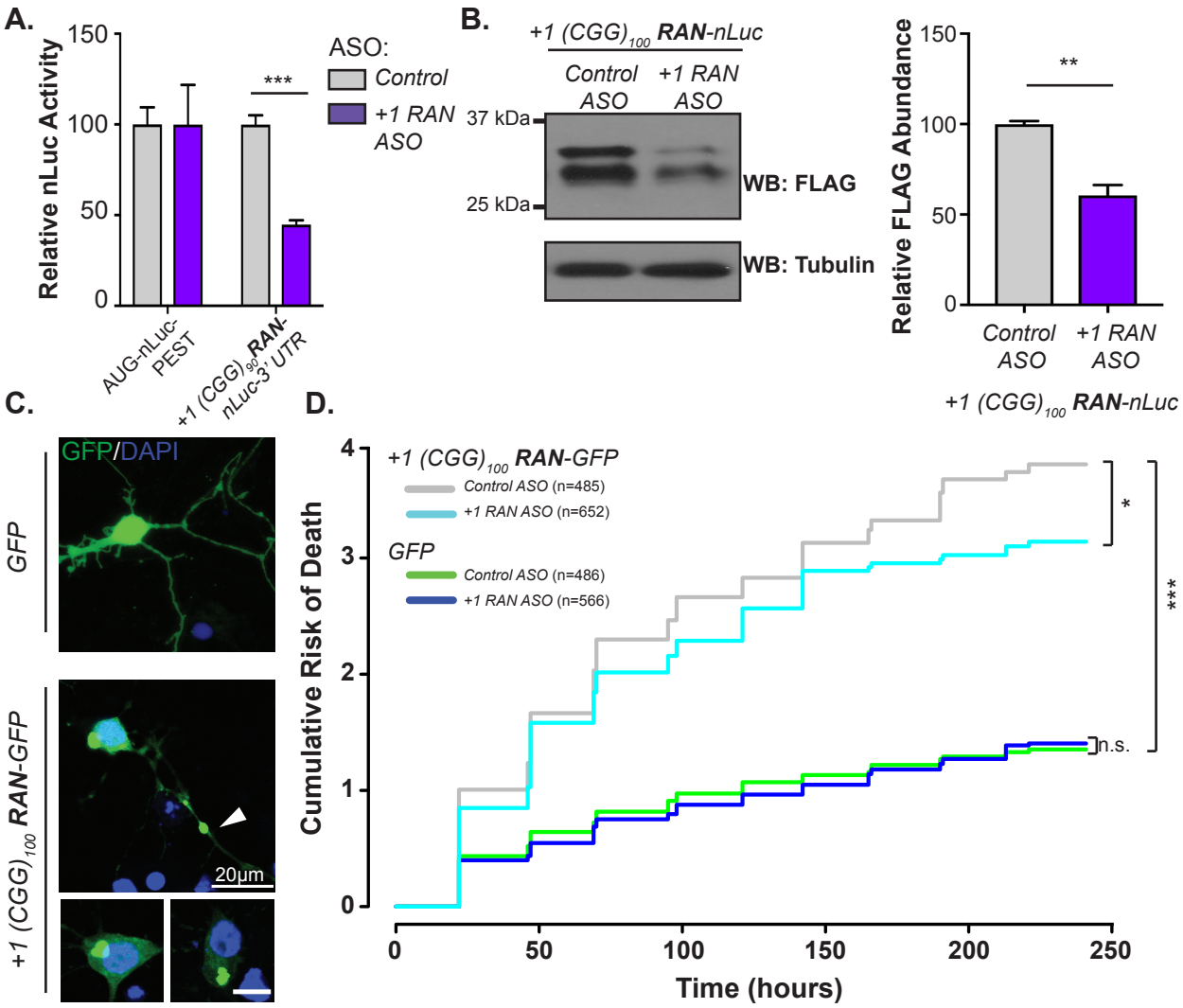


Fig 3.7:

RAN ASO suppresses CGG repeat toxicity. A) HEK293 cells expressing the indicated reporters were co-transfected with *control* or *+1 RAN ASOs*. *+1 RAN ASO* (100nM) selectively suppressed RAN reporter expression by nLuc assay. B) Western blot demonstrates selective effect of *+1 RAN ASO* on +1 $(CGG)_{100}$ RAN-nLuc expression. n=3 for each condition for experiments using nLuc reporters, Student's t-test, **p>0.01, ***p>0.001. Graphs represent mean +/- S.E.M. C) Rat neurons transfected with +1 $(CGG)_{100}$ RAN-GFP exhibit inclusions in transfected neurons. Scale bar in bottom right image represents 10µm. D) Survival analysis on rat cortical neurons transfected with the indicated constructs and ASOs tracked by longitudinal fluorescence microscopy. +1 $(CGG)_{100}$ RAN-GFP expressing neurons exhibit significant toxicity compared to GFP expressing controls. Treatment with 1µM *+1 RAN ASO*, but not the *control ASO*, reduces CGG repeat associated toxicity in cells overexpressing +1 $(CGG)_{100}$RAN-GFP. Survival is plotted as cumulative risk of death. n= number of neurons per condition. Cox proportional hazard analysis, n.s.=not significant, *p= 0.0286, ***p<2e^-16.

(**Figure 3.7B**). These data show that the ASO that increases endogenous FMRP can also selectively engage the RNA target and block RANT of FMRpolyG from expanded repeats.

To determine if this blockade of FMRpolyG synthesis modulates CGG repeat toxicity in neurons, we transfected primary neurons with +1 $(CGG)_{100}$ RAN-GFP expression constructs, which has an analogous design to the previously mentioned nLuc and Venus reporters, and tested the efficacy of the *+1 RAN ASO* to ameliorate neuronal toxicity. Expanded repeat containing constructs, elicited large somatic and dendritic aggregates of FMRpolyG fused to GFP (**Figure 3.7C**). To measure neuronal toxicity, we utilized automated longitudinal microscopy to track neuron death over 10 days on mixed primary rat cortical cultures. This system allows for accurate tracking of survival of hundreds of individual neurons over time, with differences in groups measured as a cumulative hazard of death (Flores et al., 2016). Neurons were transfected with mApple and either +1 $(CGG)_{100}$ RAN-GFP or GFP. Transfected cells were visualized by fluorescence microscopy at regular intervals spanning 10 days, and neuronal survival at each time determined on a single-cell level by using mApple as a cell marker for neuronal viability. Neurons expressing CGG repeats demonstrated significant toxicity compared to neurons expressing GFP alone across multiple experiments (**Figure 3.7D**), consistent with previous studies demonstrating that expanded CGG repeats are sufficient to cause neurodegeneration (Sellier et al., 2017; Todd et al., 2013).

To determine the efficacy of the RANT blocking-ASO to prevent cell death, transfected neurons were treated with either *+1 RAN ASO* or a *control ASO*. Parallel treatment with the RAN ASO had no toxic effect compared to the control ASO in GFP transfected neurons over the course of the experiment (**Figure 3.7D**). In contrast, treatment of +1 $(CGG)_{100}$ RAN-GFP

expressing neurons with *+1 RAN ASO* significantly reduced toxicity compared to *control ASO* treated neurons (**Figure 3.7D**). Together with previous studies(Sellier et al., 2017, Todd et al., 2013), these findings support a direct role for FMRpolyG production in CGG repeat-associated neuronal toxicity and suggest that targeting specific RANT may prevent repeat-mediated neurodegeneration.

RAN targeting ASOs increase FMRP in Fragile X full mutation patient derived neurons

The majority of FXS cases result primarily from hypermethylation of the CGG repeat and *FMR1* promoter leading to heterochromatization of the *FMR1* locus and transcriptional silencing. However, upwards of 40% of FXS patients exhibit incomplete methylation and some FMR1 transcription, often associated with mosaicism in CGG repeat size (Jacquemont et al., 2011; Nolin et al., 1994; Tassone et al., 1999). FMRP expression in such cases typically remains quite low, presumably because of translational inefficiency elicited by the repeat expansion (Feng et al., 1995). Importantly, the relative FMRP expression levels in these cases inversely correlates with the severity of symptoms, suggesting that even subtle enhancements of FMRP expression can be clinically meaningful (Lozano et al., 2014; Tassone et al., 1999).

To determine whether *+1 RAN ASO* treatment might enhance FMRP expression in the setting of a large CGG repeat expansion, we derived iPSCs from a previously established and characterized fibroblast line (TC43-97) with a fully unmethylated *FMR1* promoter and ~250-600 CGG repeats (**Figures 3.8A**, **Figure S3.5A-B**) (Burman et al., 1999). As iPSCs, the TC43-97 line maintains *FMR1* mRNA levels, but shows a significant decrease in FMRP expression relative to controls (**Figure S3.5C-D**). The iPSC TC43-97 line maintains an unmethylated repeat with a mosaic of repeat length between ~250-340 CGG repeats (not shown). Bisulfite qPCR

confirmed that the promoter remained unmethylated after iPSC derivation (**Figure S3.5E**). This new iPSC line allowed us to investigate how the *+1 RAN ASO* interacts with mRNA bearing a full FXS mutation and determine whether it can still impact FMRP expression in this more disease-relevant setting.

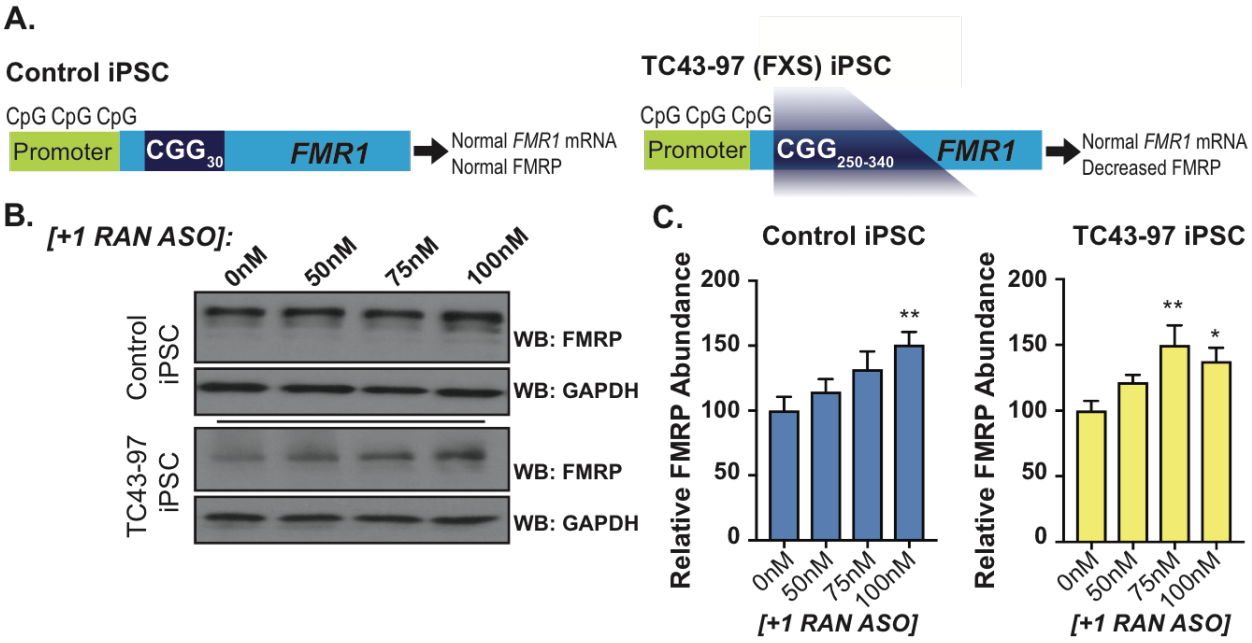


Figure 3.8:

RAN ASO increases FMRP expression in unmethylated Fragile X full mutation iPSCs. A) Schematic detailing iPSCs derived from control (30 CGG repeats) or patient (TC43-97) with unmethylated full mutation (250-340 CGG repeats). B) Western blot for FMRP in control and TC43-97 iPSCs transfected with indicated doses of *+1 RAN ASO*. C) Both iPSC lines exhibit a concentration dependent increase in FMRP by one-way ANOVA with a Fisher's LSD test. *p<0.05, **p<0.01, ***p<0.001. n=6 for each condition in both iPSC lines. All graphs represent mean +/- S.E.M.

First, we tested if the *+1 RAN ASO* could increase FMRP in control and TC43-97 iPSCs by transfecting increasing doses of the ASO. We observed a significant increase in endogenous FMRP at 100nM in both cell lines, while the TC43-97 iPSCs showed a significant increase even with 75nM of ASO (**Figure 3.8B-C**). This occurs in the absence of changes in *FMR1* mRNA expression (**Figure S3.6A**). Next, iPSCs were differentiated into neural progenitors and then

differentiated for 5 weeks to produce neurons according to an established protocol (**Figure S3.6A-B**) (Shi et al., 2012). Neurons were treated for 24 hours with *+1 RAN ASO*, followed by a media change and 5 additional days of maintenance in culture (**Figure 3.9A**). Control human iPSC derived neurons with a normal repeat size responded to 150nM of +1 RAN ASO with a significant increase in FMRP by western blot analysis (**Figure 3.9B-C**). The full mutation FXS iPSC-derived TC43-97 neurons also responded to the ASO with a 62% increase in FMRP. This effect is specific to the *+1 RAN ASO* as control neurons treated with the *control ASO* showed no change in FMRP abundance.

To confirm that this enhanced FMRP expression in the TC43-97 line was occurring in neurons, we conducted immunocytochemistry studies on TC43-97 iPSC-derived neurons to detect FMRP in cells that stain for the mature neuronal marker TUJ1 (**Figure S3.6C-D**). A more than 2-fold increase in neuronal FMRP expression is detected using this method, which represents an increase in FMRP up to the level of the control neurons (**Figures 3.9D-F**, **Figure S3.6E**). Taken together, these data demonstrate that targeting RANT in human neurons can correct FMRP deficits in the setting of a FXS full mutation.

Discussion

Simple tandem repeats make up a significant fraction of the human transcriptome and expansions in a subset of these repeats are a common cause of human disease (Gemayel et al., 2010; Nelson et al., 2013). Despite this, we know very little about the native functions of transcribed repeats within the context of normal mammalian cellular physiology. Here we define a novel, conserved and biologically relevant function of the CGG repeat in the Fragile X gene. The repeat supports non-AUG initiated translation from specific near-cognate codons at normal

repeat sizes. This translation suppresses basal FMRP production and allows for activity-dependent FMRP translation in neurons. These findings provide a framework for defining

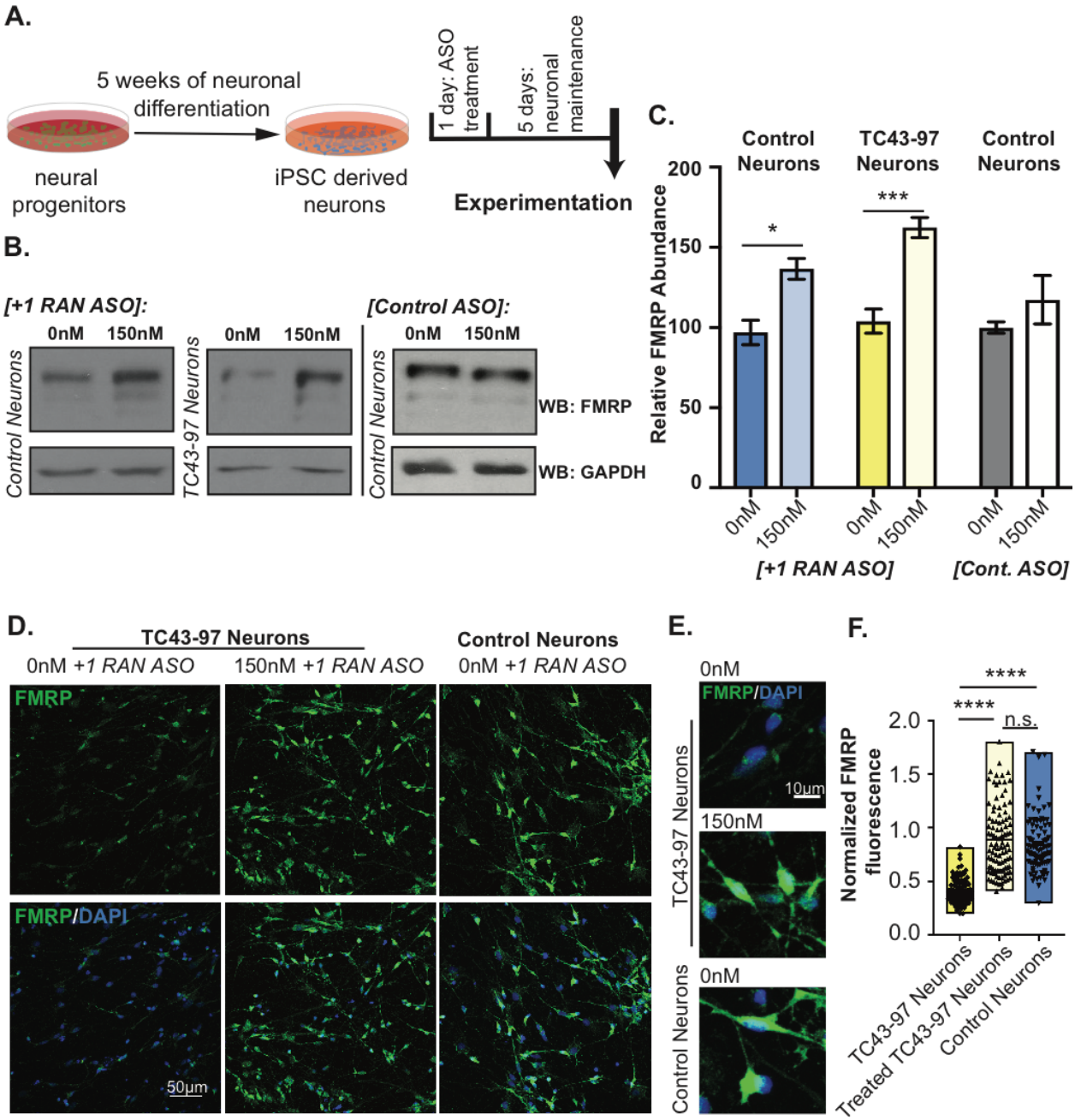


Figure 3.9

RAN ASO increases FMRP expression in unmethylated Fragile X full mutation neurons. A) Control and TC43-97 iPSC derived neurons were differentiated for 5 weeks. ASO or vehicle was added to the media for 24 hours. Cells were maintained for 5 additional days following treatment. B, C) Anti-FMRP western analysis from untreated and ASO treated neurons. *+1 RAN ASO* (150nM) treatment increases FMRP expression in control and TC43-97 neurons. Control neurons treated with the *control ASO* show no change in FMRP levels. n=3 for all control

neuron conditions, n=6 for TC43-97 neurons, **p<0.01, Student's t-test. D) Immunocytochemistry to FMRP (green) in TUJ1 positive TC43-97 neurons demonstrates increase in endogenous FMRP in *+1 RAN ASO* treated cells. Blue is DAPI. Scale bar represents 50μm. (E) Zoomed in composite images show the FMRP increase in neuronal soma and processes. F) FMRP expression normalized to TUJ1 levels in the same neurons reveals correction of FMRP expression to control levels in *+1 RAN ASO* treated neurons. n=118 for the 0nM TC43-97 condition, n=100 for the 150nM ASO TC43-97 condition, and n=95 for 0nM treated control neurons. One-way ANOVA with multiple comparisons, ****p<0.0001. Panel C represents mean +/- S.E.M. In panel F, bars represent the range of values with a line at the mean.

physiological roles for repetitive elements in both translational control and neurobiology and answers a long-standing question as to why this disease causing repeat is present and conserved in mammals (Eichler et al., 1995).

By mapping out a native function for RANT and CGG repeats, we identified a new therapeutic target in fragile X–associated disorders. To this end, we adopted a recent advance in ASO technology that allows targeting of AUG-initiated uORFs as a mechanism for enhancing translation from the main reading frame (Liang et al., 2016). Exactly how uORF ASOs function is unclear. They require RNA helicase activity and enhance recruitment of specific initiation factors (e.g. eIF2-alpha and eIF4A) to targeted transcripts (Liang et al., 2017) and based on the optimal target location for effect of these ASOs, they likely impede recognition of potential start sites during ribosomal scanning (Liang et al., 2016). This is consistent with our current knowledge of how RANT occurs at CGG repeats, where initiation is strongly cap-dependent and eIF4A-dependent (Kearse et al., 2016). However, too strong of a blockade by the ASOs may prevent continued scanning of the preinitiation complex. Consistent with this, when two ASOs that target the two potential +1 reading frame start sites are utilized, they inhibit rather than enhance FMRP synthesis despite enhancement when either is used separately (Figure S3.4E).

Treatment of neurons with an ASO targeting the +1 RAN translation initiation site significantly suppresses neuronal toxicity elicited by premutation-sized CGG repeat expansions in a cellular model of FXTAS, correcting a gain of function phenotype and supporting a role for

RANT in FXTAS disease pathogenesis. In parallel, treatment of unmethylated full mutation iPSCs and neurons with this same ASO markedly enhanced endogenous FMRP production, correcting the fundamental loss of function defect in FXS. This ASO also acts at normal *FMR1* CGG repeat sizes, further demonstrating a native function for this process in *FMR1* regulation. Thus, by selectively targeting a normal function at pathological expansions we were able to rescue multiple relevant disease states in cellular model systems.

Our work identifies RANT initiation and CGG repeats as active modulators of neuronal FMRP translation, suggesting a new function for these elements in activity-dependent translation. Previous work described how CGG repeats act to enhance FMR1 transcription and reduce FMRP translation in a length-dependent fashion in both cell based and *in vitro* studies (Chen et al., 2003; Feng et al., 1995; Ofer et al., 2009; Tassone et al., 2007; Usdin and Kumari, 2015). CGG repeats as DNA normally act as a large C_pG island near the promoter of *FMR1* to maintain an open chromatin state, and this effect is enhanced with repeat expansion (Usdin and Kumari, 2015). However, once the repeat size crosses a threshold, it triggers methylation and a closed chromatin state, perhaps through RNA-induced transcriptional silencing (Colak et al., 2014; Zhou et al., 2016). Numerous modulatory factors for this process have now been identified (Usdin and Kumari, 2015). Thus we now have a relatively nuanced understanding of how repeats function transcriptionally. In contrast, our understanding of how CGG repeats influence translation have focused largely on a model where the repeat creates a steric hindrance on ribosomal scanning that is overcome in an unregulated fashion through the activity of RNA helicases (Ofer et al., 2009). The finding that this event is activity-dependent suggests a novel mechanism by which local mRNA translation in neurons could be regulated through selective uORF usage in the setting of repetitive or structured mRNA elements in the 5' UTR.

How RANT and CGG repeats mediate this regulation is not clear, partially because how mGluR activation drives local protein synthesis is not entirely understood. mGluR1 activation stimulates a number of pathways in dendrites, including release of intracellular calcium stores, diacylglycerol and protein kinase C activation, ERK phosphorylation, mTOR signaling, and ubiquitin-proteasome activation (Bhakar et al., 2012; Hagerman et al., 2017). Neuronal

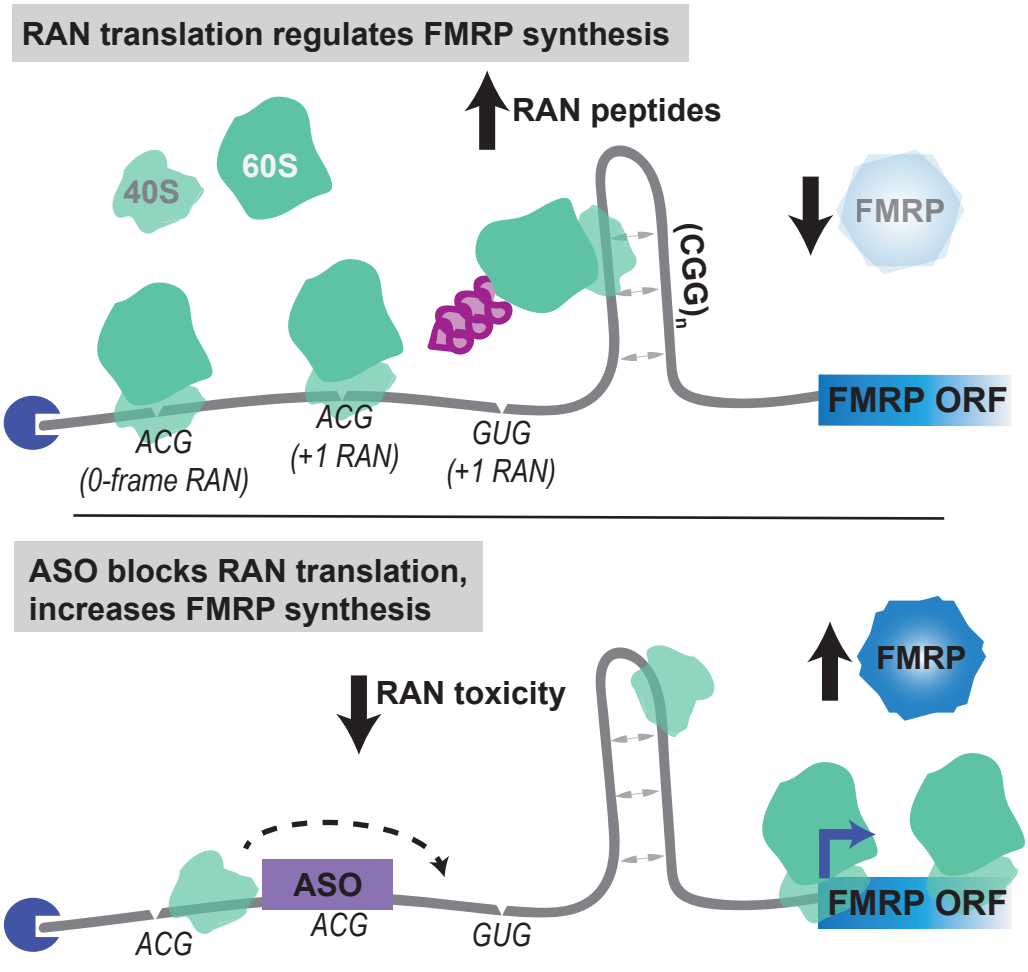


Figure 3.10:

Graphical Abstract: Expanded CGG repeats in *FMR1* underlie both a common cause of autism (Fragile X Syndrome) and neurodegeneration (Fragile X-associated Tremor/Ataxia Syndrome). Repeat associated non-AUG (RAN) translation at this repeat initiates at near-cognate codons in the 5' UTR of *FMR1* mRNA. RAN translation and the CGG repeats normally function to suppress basal FMRP production while allowing for activity-dependent synthesis of FMRP in neurons, but both functions become problematic with repeat expansions due to toxic RAN proteins and loss of FMRP. An ASO targeting RAN initiation suppresses CGG repeat toxicity while simultaneously

enhancing FMRP synthesis in Fragile X patient derived neurons, suggesting a novel therapeutic approach to both Fragile X-associated disorders.

activation by mGluRs drives synaptic alterations through cap-dependent protein synthesis (Huber et al., 2000) and additional work suggests specific roles for the cap complex protein eIF4E and the ternary complex component eIF2α in this process (Di Prisco et al., 2014; Gantois et al., 2017; Gkogkas et al., 2014). However, other studies suggest that at least the initial wave of translation that occurs in response to mGluR activation is derived from mRNAs bound to stalled polyribosomes, and cannot be precluded following incubation with a drug that binds and blocks ribosomes prior to their association with mRNA (Graber et al., 2013). Consistent with this, the eukaryotic elongation factor eEF2 appears to act as a sensor coupling synaptic activity to translational control. During particular patterns of synaptic activity, eEF2 phosphorylation is enhanced, negatively regulating protein synthesis (Sutton et al., 2007). Thus, both translation initiation and elongation regulation appear critical for controlling translation associated with synaptic plasticity.

Our data demonstrate an increase in both native FMRP and FMRP reporter synthesis after mGluR activation. This mGluR-dependent increase requires RAN translation and the CGG repeat. One explanation for this is that ribosomes stall in translation of the CGG repeat with a build-up of 40S pre-initiation complexes that follow to eventually reach the AUG of FMRP. With DHPG stimulation, there is re-activation of translation from these stalled polyribosomes, and from loaded and scanning 40S subunits that have not yet initiated. Removing the repeat or altering the secondary structure and sequence upsets this balance, preventing upscaling of FMRP translation. Given that translating ribosomes and associated proteins are a potent RNA helicase, RANT itself may have repeat structure-destabilizing effects that could alleviate secondary structure and prime *FMR1* for further translation (Todd et al., 2013). In this context, it is

interesting that RAN translation at CGG repeats is selectively and paradoxically enhanced by eIF2α phosphorylation, which typically suppresses translation initiation (Green et al, Nature Communications, *in press*). This suggests that RANT may bypass some canonical initiation pathways to serve specific regulatory roles in this synaptic context.

Translational regulation elicited by the CGG repeat and RANT on the *FMR1* transcript may be an exemplar for a broader class of regulatory events on eukaryotic mRNAs. Specific AUG-initiated regulatory uORFs in genes such as GCN2 and ATF4 are well characterized as serving to suppress translation from the main ORF normally but activate under stress conditions(Hinnebusch et al., 2016), and both GCN2 and ATF4 have been implicated as regulators of learning and memory in mouse behavioral studies (Costa-Mattioli et al., 2009). More recently, specific near cognate uORF sites were defined in transcripts encoding proteins critical for cancer and cellular responses to stress (Sendoel et al., 2017; Starck et al., 2016). Ribosome profiling studies suggest uORFs may be quite common, with some studies suggesting they are utilized on one third of all transcripts across a broad class of cell types and species (Brar, 2016; Ingolia et al., 2011). Moreover, some studies suggest that uORF usage shifts with changes in cell states, such as during mitosis in eukaryotic cells and across the life cycle of budding yeast (Brar et al., 2012). Studies in neurons suggest a select number of transcripts move to polysomes upon DHPG stimulation, and these are enriched for uORFs (Di Prisco et al., 2014). One transcript derived from a gene also implicated in X-linked intellectual disability, *Ophn1*, bears a uORF that is required for its regulation in response to DHPG. Translation of this transcript and other uORF containing transcripts is specifically regulated by phosphorylation of eIF2α, and this is in turn essential to learning in mice (Di Prisco et al., 2014). In this context,

perhaps RANT is acting as a regulatory uORF to impair basal FMRP, but in its absence proper scaling cannot be achieved in response to DHPG.

In recent years, ASO therapies have moved from the research bench to the clinic. This was initially thought of as a strategy to target solely gain-of-function disorders, where knockdown of a toxic transcript or blocked expression of a toxic protein could potentially abrogate disease pathology. However, the recent success of nusinersen, which blocks binding of hnRNPs to *SMN2* to increase production of the SMN protein in spinal muscular atrophy (Finkel et al., 2016; Rigo et al., 2012), speaks to the potential power of these molecules to treat loss-of-function disorders in the nervous system. The exact abundance of FMRP required to correct phenotypes in FXS remains unknown, but even modest enhancement of this protein may be helpful in the disease state. To this end, the approaches described here represent an exciting opportunity to potentially leverage new disease insights towards effective therapeutic development.

Methods

Plasmid Construction

Cloning of the pcDNA3.1(+) vectors expressing AUG-nLuc, +1 $(CGG)_n$ RAN-nLuc, +2 $(CGG)_n$ RAN-nLuc, and +0 $(CGG)_n$ RAN-nLuc was described in (Kearse et al., 2016). All ligations were performed using Roche Rapid DNA Dephosphorylation and Ligation Kit according to manufacturer's specification. PCR cloning was accomplished with Phusion High-Fidelity DNA Polymerase (NEB) or Platinum Taq High-Fidelity (Thermo Fisher). All site-directed mutagenesis was performed using Q5 Site-Directed Mutagenesis Kit (NEB) according

to manufacturer's protocol. All vector sequences are listed in Table 3.1, and all primers used for cloning are listed in Table 3.2.

To generate RAN translation reporters bearing out-of-frame mutants, the XhoI and XbaI fragment harboring the nLuc reporter in each particular reading frame was cloned back into the desired mutant background that contained CGG repeat (Kearse et al., 2016). +1 $(CGG)_{100}$ GGG-nLuc-3xFLAG-PEST was generated using a two-step Q5 site directed mutagenesis strategy using +1 $(CGG)_0$ GGG-nLuc-3xFLAG (Kearse et al., 2016) as a template. The PEST containing GGG-nLuc coding sequence was cloned into the original plasmid using EcoRV and XbaI to obtain a full $(CGG)_{100}$ repeat construct.

For the $(CGG)_n$ FMRP-nLuc vector, the $(CGG)_{100}$ repeat was moved from a separate pUAST vector using restriction enzymes EcoRI and XhoI. The 3' end of the *FMR1* 5' UTR and first coding exon was added by PCR cloning using XhoI and NotI. Nanoluciferase with a GGG in place of the AUG, GGG-nLuc-3xFLAG, was cloned in frame with the first FMRP exon using NotI and XbaI. The 3xSTOP (taattaattaa) was inserted 16 nucleotides downstream of the repeat by Q5 SDM using $(CGG)_0$ FMRP GGG-nLuc as a template. The mutated sequence was then cloned via XhoI and XbaI back into the original $(CGG)_{100}$ containing construct to obtain a full length repeat construct. The 3xSTOP upstream of the repeat was cloned using annealing primers via BlpI and NarI.

The 3' UTR was added to the FMRP reporters to produce the $(CGG)_n$ FMRP-nLuc-3' UTR by PCR cloning from a separate pCRII-TOPO vector bearing the sequence using restriction enzymes XmaI and XbaI. The entire nLuc-3xFLAG-PEST tag was PCR cloned out of pcDNA3.1 +1 $(CGG)_{100}$ nLuc-PEST and into the $(CGG)_n$ FMRP-nLuc-3'UTR vector using NotI

and XbaI. The +1 $(CGG)_{100}$ nLuc-PEST vector was cut with XhoI and XbaI and ligated into the $(CGG)_n$ FMRP-nLuc-3' UTR vectors to generate +1 $(CGG)_n$ nLuc-PEST-3' UTR.

The Venus open reading frame was cloned into pCRII-TOPO vector bearing the C-terminal region of +1 FMRpolyG from XhoI to the first codon before the in-frame start site. This was PCR cloned in place of GFP in the +1 $(CGG)_{100}$ GFP pcDNA3.1 plasmid (Todd et al., 2013) using XhoI and XbaI. Repeat size in all reporters was confirmed by digestion with BlpI and XhoI digestion and gel electrophoresis on a 2% (wt/vol) agarose gel, and by Sanger sequencing specific for structured DNA at the University of Michigan Sequencing Core.

In Vitro Transcription

pcDNA3.1(+) vectors were digested with PspOMI for linearization. pcDNA3.1(+) vectors with the 3' UTR of FMR1 were digested with SmaI for linearization. HiScribe T7 Quick High Yield RNA Synthesis kit (NEB) was used to make capped RNA from our linearized plasmids, mRNA was then polyadenylated as in Kearse *et al.,* 2016.

In Vitro Translation and Luciferase Assay

In vitro translation reactions using the Flexi Rabbit Reticulocyte Lysate (RRL) System (Promega), were programmed with nLuc reporter mRNAs and analyzed as in Kearse *et al.,* 2016. Briefly, luciferase assays were performed using a 25µL:25µL reaction of prepared NanoGlo reagent (Promega) and diluted reactions in Glo Lysis Buffer (Promega), and quantified using a GloMax-96 plate reader (Promega). For western blot analysis, 10µL RRL reactions were performed with saturating mRNA levels, mixed with 40µL of SDS sample buffer and heated at

70° for 15 minutes, and then resolved using SDS-PAGE. nLuc-3xFLAG reporter proteins were detected by western blot using the anti-FLAG antibody(mouse, Sigma).

Primary Hippocampal Neuron Culture and Luciferase Assay

Hippocampi were dissected from P0-P3 rat pups of both sexes. Cells were papain dissociated as in Sutton *et al.,* 2006. Hippocampal neurons were plated on poly-D-Lysine coated 12-well culture plates at a density of $6.0x10^4$ cells/plate. Hippocampal cultures were maintained for 11-12 DIV prior to transfection. On the day of transfection, cells were washed 1X with NGM supplemented with 0.0189% kynurenic acid (NGM-KY) [wt/vol], then placed in 1mL of NGM-KY. A mixture of 510µL of Opti-MEM (Fisher), 10uL Lipofectamine 2000 (Fisher), and DNA was added to the neurons for 30 minutes. 2.5µg or 5µg nanoluciferase plasmid and pGL4.13 (firefly luciferase) plasmid was added to each well for luciferase assays. 7.5µg of reporter plasmid and mCherry plasmid was transfected into cells for imaging assays. The OptiMEM/NGM-KY mixture was removed from cells followed by 2 washes in NGM-KY, and 1mL of NGM was then added. Neurons were placed back in the 37°C incubator for 24 hours.

Transfected neurons were lysed directly on plate for 10 minutes in 250µL of Glo Lysis Buffer. Luciferase assays were performed as described above in a 50µL:50µL reaction of prepared NanoGlo Reagent to lysate. For firefly luciferase quantification, 50µL of prepared ONE-Glo reagent (Promega) was used in place of NanoGlo. Nanoluciferase was normalized to firefly luciferase to control for transfection differences. All samples with firefly nanoluciferase values below 100 RLU were excluded due to viability and transfection concerns.

Single-molecule microscopy and analysis

Single-molecule imaging of Venus constructs was performed as previously described (Barbarese et al., 2013; Ifrim et al., 2015; Tatavarty et al., 2012). Single-molecule imaging was performed with a modified Nikon (Tokyo, Japan) N-SIM microscope equipped with a Nikon Apo TIRF 100X, NA 1.49 microscope objective, and an iXon3 EMCCD camera (Andor Technology, Belfast, UK). For single-molecule imaging of translation in live cells, Venus protein was excited with a 488nm laser line (1.5mW laser power) from a continuous wave (CW) solid-state laser (Coherent, Santa Clara, CA). The single-molecule imaging was done in epifluorescence mode. Following whole-cell photo-bleaching, single molecule time-lapse images of Venus protein were collected continuously at an exposure time of 150ms per frame. To minimize the effects of photo-damage on the cells an illumination area of ~25μm diameter was used.

Data analysis of time-lapse images was done as previously described (Ifrim et al., 2015; Tatavarty et al., 2012; Yu et al., 2006). Briefly, the centroid coordinates of individual molecules in each frame were determined and linked in time to construct temporal trajectories, using a standard single-particle tracking algorithm. Each Venus molecule detected was tracked from appearance (maturation) to disappearance (bleaching). Preexisting Venus fluorescent signal was used to manually define an ROI that contained neuronal dendrites, but not soma, for each dataset. From each dataset, translation events situated inside the ROI were analyzed with MATLAB (MathWorks). The analysis software is available at http://www.ccam.uchc.edu/yu.

Hybridization Chain Reaction

Mouse embryonic Fibroblasts were seeded at 2.5 x10^4 cells/well for 24 hours on 0.01% poly-D-lysine-coated 8-well chamber slides, then transfected with 250 ng nLuc reporters (AUG-

NL-3xF, +2(CGG)100-NL-3xF, and mock) and 2:1 jetPRIME® (VWR) for 24 hours. Cells were fixed according to Molecular Instruments protocol(Choi et al., 2016). Following overnight incubation with 70% ethanol, cells were rehydrated in PBS for 1 hour, permeabilized with 0.1% Triton x-100 for 6 minutes, blocked with 2% BSA for 20 minutes at room temperature, and incubated overnight with mouse anti-FLAG primary antibody in 2% BSA at 4^{o}C. Goat anti-mouse Alexa 488 secondary antibody was applied the following morning for 1 hour at room temperature, in the dark. Following ICC, reporter RNA was detected in transfected cells using DNA probes with additional sequence complimentary to Cy5 labeled self-hybridizing hairpins (Choi et al., 2016; Choi et al., 2010; Huss et al., 2015). Probes against the nanoluciferase sequence (Table 3.3) were purchased from molecularinstruments.org and applied according to manufacturer's protocol. Coverslips were then applied to slides with ProLong Gold Antifade Mountant with DAPI. 10-20 fields per condition were imaged at 20x1.6 magnification with Olympus IX71 fluorescent microscope and Slidebook 5.5 software.

Rat hippocampal neurons plated on Mattek dishes were maintained to DIV 5 prior to transfection and incubation with plasmid for 24-hours. HCR was performed as above. mAPPLE protein was detected using a rabbit anti-dsRed antibody (Clonetech). Confocal imaging was performed as outlined below.

RNA Quantification

Total RNA was isolated from cells using Quick-RNA Miniprep Kit (Zymo Research). DNA was eliminated using 1μL Turbo DNase (Thermo Fisher) according to the manufacturer's protocol. For cells transfected with plasmids, this process was repeated to ensure removal of contaminating DNA. cDNA was synthesized from equal amounts of total RNA using iScript

cDNA synthesis kit (Bio-Rad). qPCR reactions were performed using iQ SYBR® Green Supermix (Bio-Rad) and equal amounts of cDNA. qPCR was performed on a MyiQ thermocycler (Bio-Rad) using a two-step protocol. Biological triplicates were run in technical triplicate for all experiments. Standards were run alongside samples for each primer to calculate primer efficiencies. Relative abundance was calculated for each replicate using the efficiencies of each primer and C_t values from the transcript of interest relative to the housekeeping gene.

Cell Culture and Transfection

SH-SY5Y cells (ATCC) were plated on 6-well plates for nanoluciferase and RT-qPCR assays. Cells were maintained in DMEM:F12 (Fisher) and 1% Pen/Strep. Cells were transfected with 550ng of both nanoluciferase DNA and pGL4.13 using FuGENE® HD (Promega). Cells were maintained for 24 hours post-transfection before being either lysed for RNA isolation or nanoluciferase assay.

HEK293 cells (ATCC) were maintained in DMEM with high glucose (Fisher), 10% fetal bovine serum (vol/vol), and 1% Pen/Strep (vol/vol; Fisher). $1.3x10^4$ cells/well were plated on 96-well plates for nanoluciferase assays, 24 hours prior to transfection. For ASO treatments, cells were transfected with 100ng nanoluciferase reporter RNA using Lipofectamine 2000. 7 hours later media was changed and ASOs were transfected with Lipofectamine RNAiMAX (Fisher) according to manufacturer's specifications. After an additional 17 hours, cells were lysed in 200µL of Glo Lysis buffer, and reporter levels were measured by luciferase assay as detailed above. All ASO sequences are listed in Table 3.4.

For analysis of endogenous FMRP levels during ASO treatment, HEK293 cells were plated on 12-well plates, and transfected with a mixture of ASOs and 4.5µL of Lipofectamine®

RNAiMAX according to manufacturer's specifications. After 24 hours, cells were lysed in 250µL RIPA supplemented with mini cOmplete protease inhibitors (Sigma), boiled in 6X SDS sample buffer at 90°C for 5 minutes, and then were resolved on 8% polyacrylamide gels by SDS-PAGE. FMRP was detected by subsequent western blotting using anti-FMRP antibody ab17722 (Abcam). FMRP levels were quantified relative to GAPDH.

For knockdown of +1 RAN reporters bearing large repeats, HEK293 cells were plated on a 96-well plate. After 24 hours of growth, cells were co-transfected with 100nM ASO and 50µg of reporter DNA and pGL4.13 using Lipofectamine® 2000. 24 hours later, nanoluciferase assays were performed (as above). For western blotting of the knockdown, HEK293 cells were plated on 24-well plates and transfected with 250ng of reporter DNA for 3 hours. Media was change 3 hours later, and cells were re-transfected with 100nm *control* or *+1 RAN ASO*. Cells were incubated for 48 hours with a media change after 24 hours, then lysed in 200µL of RIPA buffer. Lysates were resolved on 12% SDS-PAGE gels with subsequent western blotting for detection of FLAG using anti-FLAG antibody (Sigma).

Longitudinal fluorescence microscopy of primary neurons

Mixed cortical neurons were dissected from E20 rat pups of both sexes, as previously described (Arrasate et al., 2004; Barmada et al., 2015). Cortical neurons were cultured at 0.6 x 10^6 cells/mL on 96-well plates. Cultures were maintained at 37°C in neural growth media (NGM; Neurobasal A supplemented with 2% B-27 and 1% Glutamax-1 [vol/vol] (Fisher)). On DIV4, neurons were co-transfected with 0.1µg pGW1-mApple and either 0.1µg pGW1-GFP or pGW +1 $(CGG)_{100}$ GFP DNA per well of a 96-well culture plate, using Lipofectamine 2000 (Invitrogen). ASOs targeting the CGG repeat and control ASOs were applied to neurons at 1µM

immediately following transfection, the media with ASO remained on the neurons for the entirety of the experiment. Neurons were imaged at regular 24 h intervals beginning 24 h post-transfection using an automated fluorescence microscopy platform detailed in prior studies (Arrasate et al., 2004; Barmada et al., 2015). Image processing for each timepoint and survival analysis were achieved by custom code written in Python or the ImageJ macro language, and cumulative hazard plots were generated using the survival package in R. Statistical analyses for the comparison of survival between different conditions was also accomplished in R.

iPSC Reprogramming

Fibroblasts were cultured in DMEM, 10% FBS, 1X L-glutamax (Fisher), 1 mM MEM non-essential aminoacids (Fisher), at 37°C and 5% CO_2. For episomal reprogramming, 1x 10^6 fibroblasts were collected after Trypsin treatment and mixed with a set of plasmids pCXLE-hOCT3/shP53, -hSK, -hUL (Addgene), then electroporated with Neon® device (condition:1650 Volts, 10 mm width, and 3 pulses). Induced fibroblasts were plated onto 6 well plates at density of 0.5-1x10^4cells/well and switched 1 day later to a PSC medium mTeSR1 (StemCell Technologies). IPSC colonies appeared and were manually picked and passaged onto new matrigel coated 12-well plates, and continually grown with mTeSR1. IPSCs were passaged weekly using 0.5mM EDTA and culture medium supplemented with 10μM Y-27632 ROCK inhibitor (EMD Millipore) for 24 hours. After 5-10 passages the cells were evaluated for pluripotency by immunocytochemistry (ICC) and embryoid body differentiation. For embryoid body experiments, iPSC colonies were grown in suspension for 3-4 weeks, passaged onto 0.1% gelatin (Sigma) for another week and processed for ICC. iPSC reprogramming was confirmed by

staining with antibodies against Oct-3/4, NANOG, and SSEA4 (Santa Cruz Biotech, Abcam, DSHB). IPSC lines were karyotyped by Cell Line Genetics.

iPSC Differentiation and ASO Treatment

Undifferentiated iPSCs were cultured in TeSR-E8 media (Stem Cell Technology) on MatriGel-coated plates with daily media changes and were passaged at 1:5 to 1:10 using 1 mM EDTA. Neural induction was performed using a dual-SMAD inhibition (Shi et al., 2012) protocol with modifications. In brief, two wells of a 6-well plate were grown to approximately 80% confluence, dissociated with EDTA, and plated into a single well of a MatriGel-coated 6-well plate with TeSR-E8 containing 10 μM Rock Inhibitor (Y-27632). The cells were confluent the next day and neural differentiation was induced using neural maintenance media (referred here as 3N) containing 1 μM dorsomorphin and 10 μM SB431542. The cells were cultured for 12-14 days with daily media changes. Neuroepithelial sheets were then combed into large clumps, passaged, and maintained on MatriGel-coated plates in rosette media (3N containing 20 ng/ml FGF) with daily media changes until neural rosettes appeared. Rosettes were manually picked and dissociated into single cells using Accutase. Neural progenitors were plated onto MatriGel-coated plates, grown in neural expansion media (3N containing 20 ng/ml FGF and 20 ng/ml EGF) with media changes every other day, and passaged as needed using Accutase. For differentiation into neurons, neural progenitors were plated at a density of approximately 1.5 x 10^5 cells/mL in neural expansion media on PLO-laminin coated plates or coverslips, allowed to grow for 24 hours, and switched to neural maintenance media. Neurons were maintained for up to 6 weeks with half media changes every other day and a full media change supplemented with 1 μg/ml laminin every 10 days.

For ASO treatments, undifferentiated iPSCs were plated as small colonies on MatriGel-coated plates in TeSR-E8 containing 10 uM Rock Inhibitor and grown overnight. Media was replaced with TeSR-E8 the next day. Cells were allowed to recover for at least 4 hours and media was replaced again just prior to treatment. ASOs (0-100 nM) and RNAiMax (4.5 µl per 100 µl of prepared complexes) were diluted in Opti-MEM reduced serum media, incubated together for 5 min at room temperature, and added to cells. Cells were harvested 24 hours after treatment. For iPSC-derived neurons, ASOs (0-100 nM) were diluted in neural maintenance media and added to 6-week old neurons one day after a full media change. Media was changed 24 hours after treatment. Neurons were maintained as previously stated and harvested 6 days after treatment. Westerns were performed as described above.

Immunocytochemistry

Hippocampal and iPSC derived neurons were washed 2x in PBS containing 1mM MgCl2 and 0.1 mM CaCl2 (PBS-MC). Cells were fixed for 15 minutes with a solution of 4% Paraformaldehyde/4% Sucrose in PBS-MC warmed to 37°C, washed 3x in PBS-MC, and permeabilized for 5 minutes in 0.1% Triton-X in PBS-MC. Cells were blocked in 2% BSA for an hour, then incubated in primary antibody for at least 2 hours at room temperature. Reporter protein was detected with an antibody for FLAG (mouse, Sigma) and mCherry or mApple was detected with an anti-dsRed antibody (rabbit, Clonetech). Cells were washed 3 times, followed by incubation in secondary antibody for 1 hour. Cells were subsequently washed and placed in ProLong Gold Antifade with DAPI (Thermo Fisher).

Confocal Microscopy and Live Imaging

Fixed and live imaging was performed in an inverted Olympus FV1000 laser-scanning microscope. Directly prior to live cell imaging, NGM was removed and replaced with 1mL of warmed 1x HBSS. Plates were imaged for a maximum of 1 hour to prevent imaging hyper-stressed or dying cells. For all experiments, acquisition parameters were identical between conditions within experiments. For all reporter quantification experiments, all co-transfected cells were imaged on >2 individual transfected plates. Imaging of transfected neurons was performed using a 60X objective. For human iPSC-derived neuron experiments, all cells in 3 separate 40X–imaged fields were analyzed from each plate in regions of similar confluency and TUJ1 staining. Channels were imaged sequentially and optimized to eliminate bleed-through. Neurons were imaged in a series of Z-planes to resolve the entire soma and dendritic arbor. Images were analyzed in ImageJ. Average intensity composite images were derived from raw image files. For quantification of individual soma or dendrites, cell casts were made using threshold images as a guide. The ROI was applied to the individual channels, and intensities were measured and normalized as specified.

Statistical Analysis Software and Selection of Statistical Analyses

Statistical analyses were performed in GraphPad Prism7, with the exception of the cumulative risk of death calculation, which was performed in R. An unpaired Student's t-test (two-tailed) with a 95%-CI was performed for all assays comparing two experimental conditions. A two-way ANOVA with a 95%-CI was performed for the neuron experiments comparing start site mutants to the controls run in parallel from 3-4 individual experiments run in triplicate, individually transfected wells of neurons. A one-way ANOVA with multiple comparisons was performed for all other experiments with more than two experimental conditions. A one-way

ANOVA with a Fisher's LSD test was performed on all ASO dose-response experiments. Survival measures were calculated according to Cox proportional hazard analysis. All bar graphs show the mean +/- S.E.M. unless otherwise stated.

Supplemental Figures

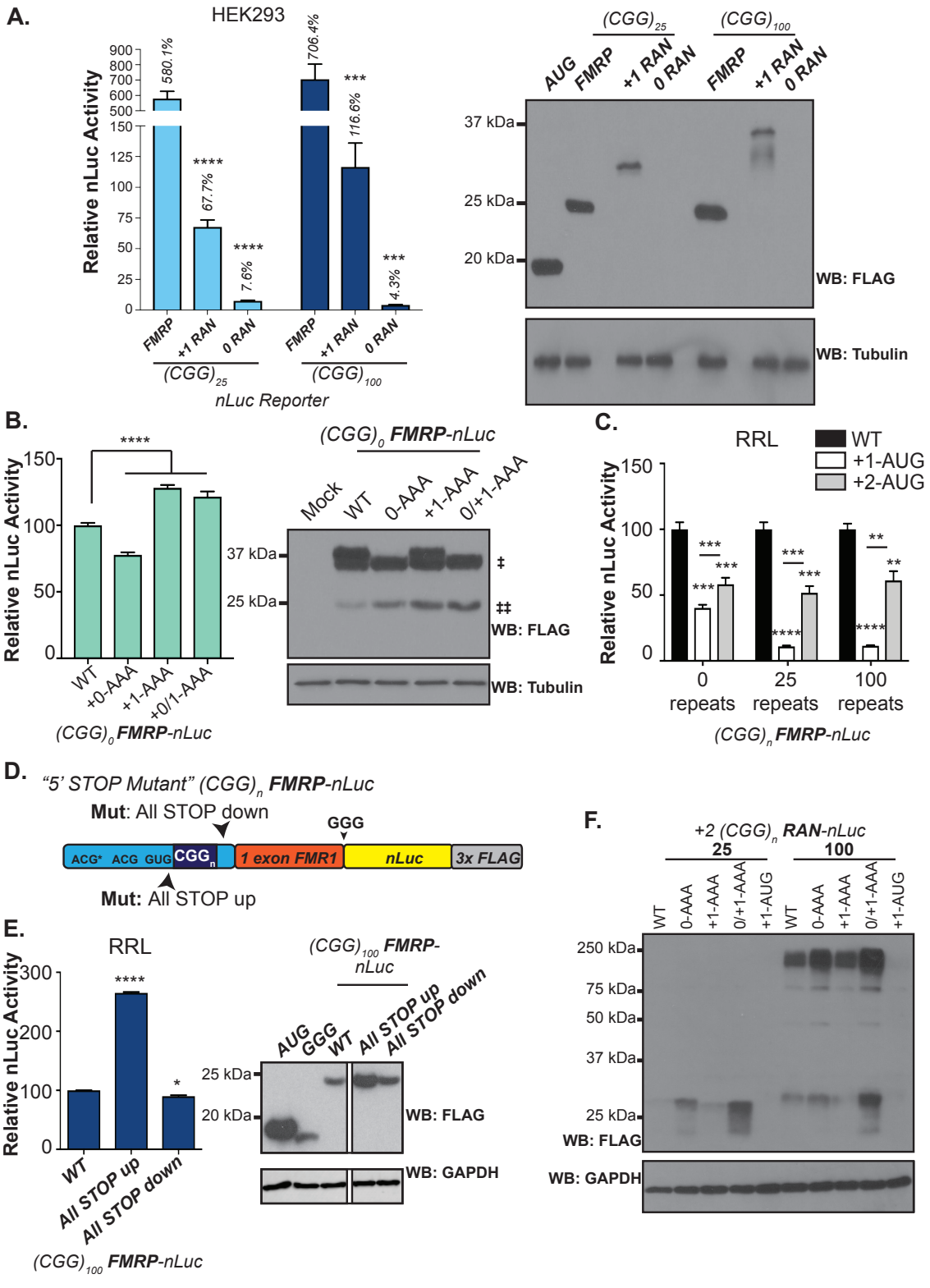

A.
HEK293
Relative nLuc Activity
580.1%
67.7%
7.6%
706.4%
116.6%
4.3%
FMRP
+1 RAN
0 RAN
(CGG)25
(CGG)100
nLuc Reporter
AUG
37 kDa
25 kDa
20 kDa
WB: FLAG
WB: Tubulin
B.
(CGG)0 FMRP-nLuc
Mock
WT
0-AAA
+1-AAA
0/+1-AAA
+0-AAA
‡
‡‡
C.
RRL
WT
+1-AUG
+2-AUG
0 repeats
25 repeats
100 repeats
D.
"5' STOP Mutant" (CGG)n FMRP-nLuc
Mut: All STOP down
GGG
ACG* ACG GUG CGGn
1 exon FMR1
nLuc
3x FLAG
Mut: All STOP up
E.
RRL
(CGG)100 FMRP-nLuc
All STOP up
All STOP down
WB: GAPDH
F.
+2 (CGG)n RAN-nLuc
25
100
250 kDa
75 kDa
50 kDa

Figure S3.1:

RAN translation modulates synthesis from FMRP reporters *in vitro.* A) Left: NLuc assay in HEK293 cells shows the relative levels of FMRP, +1 RAN, and 0-frame RAN at 25 and 100 repeats in transfected cells. Values were normalized to control nLuc vector without an AUG start codon, serving as a negative control. Right: Western blot with equivalent amounts of each reporter transfected into HEK293 cells. The decrease in abundance of the +1 product (FMRpolyG) relative to FMRP is comparable to nLuc assay derived values. The 0-frame product is almost undetectable in cells by western blot at this exposure. The greater expression of FMRP reporter at expanded repeat sizes is derived largely from enhanced reporter transcript abundance. B) Left: *In vitro* translation assay and western blot for FLAG of $(CGG)_0$ FMRP-nLuc reporters (see Figure 1 for schematic) with indicated near-cognate RAN initiation codons mutated. Right: Western blot to FLAG tag from nLuc reporters in RRL. In the absence of a repeat, there is significant signal derived from both the N-terminal extension (‡) and AUG initiation codon for FMRP (‡‡). A second N-terminal extension band derives from an alternative codon 3' to the repeat site. This is largely lost at larger repeat sizes and in transfected cells (Kearse et al, 2016, data not shown). C) NLuc assay in RRL using FMRP reporters with 0, 25, and 100 repeat lengths were mutated to include AUG codons in +1 (glycine) frame (+1-AUG) or +2(alanine) frame (+2-AUG) upstream of the repeat. Driving out of frame translation impedes FMRP reporter synthesis, with greater effects in the +1 frame and at larger repeat sizes. D) Schematic of STOP codon inserted FMRP Reporters with 100 repeats. Mutant 1 has a stop codon just 5' to the repeat in all 3 reading frames, and Mutant 2 has a stop codon inserted 3' to the repeat in all three reading frames. E) Left: *In vitro* nLuc assay on STOP mutants in RRL. Stop codons 5' of the repeat (All STOP up) enhanced FMRP reporter signal, while stop codons 3' of the repeat (All STOP down) had a modest (~10% decrease) but significant inhibitory effect. Right: Western blot analysis of *in vitro* assay. White bar in blot represents lanes removed post-hoc that were from a separate experiment not included in this analysis. All lanes shown were on the same blot at the same exposures for both FLAG and loading control. F) Western blot showing impact of 0- and +1 reading frame RAN initiation sites on +2 (FMRpolyA) RANT. All RRL nLuc assays were performed in triplicate. A one-way ANOVA was performed for each repeat group separately, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. All graphs represent mean +/- S.E.M.

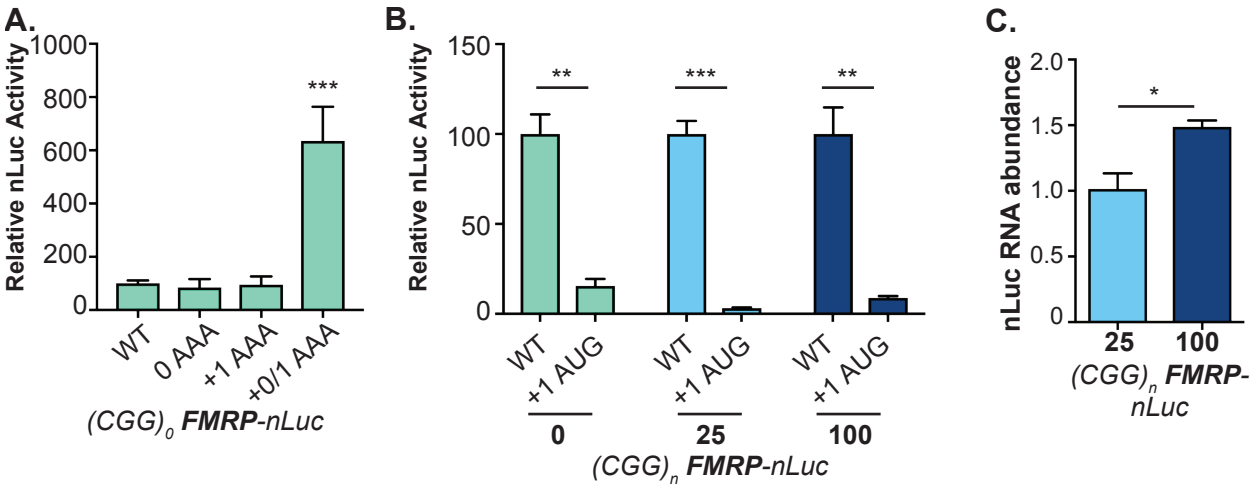


Figure S3.2:

Impact of RAN translation on FMRP reporter expression in neurons. A) NLuc assay in mature hippocampal rat neurons transfected with reporters containing no CGG repeats show that the 0/+1-AAA mutation significantly increases FMRP-nLuc signal. n=3 for each reporter, one-way ANOVA was performed with multiple comparisons, ***p<0.001. B) In neurons, FMRP nLuc reporters bearing 0, 25, and 100 repeats exhibit decreased signal when an AUG start codon replaces the ACG in the +1 frame utilized for RANT. C) RT-qPCR performed on transfected SH-SY5Y cells shows relative difference in basal mRNA expression between reporters bearing 25 and 100 repeats, consistent with previous reports. n=3 for all experiments. Student's t-test, **p<0.01, ***p<0.001. All graphs represent mean +/- S.E.M.

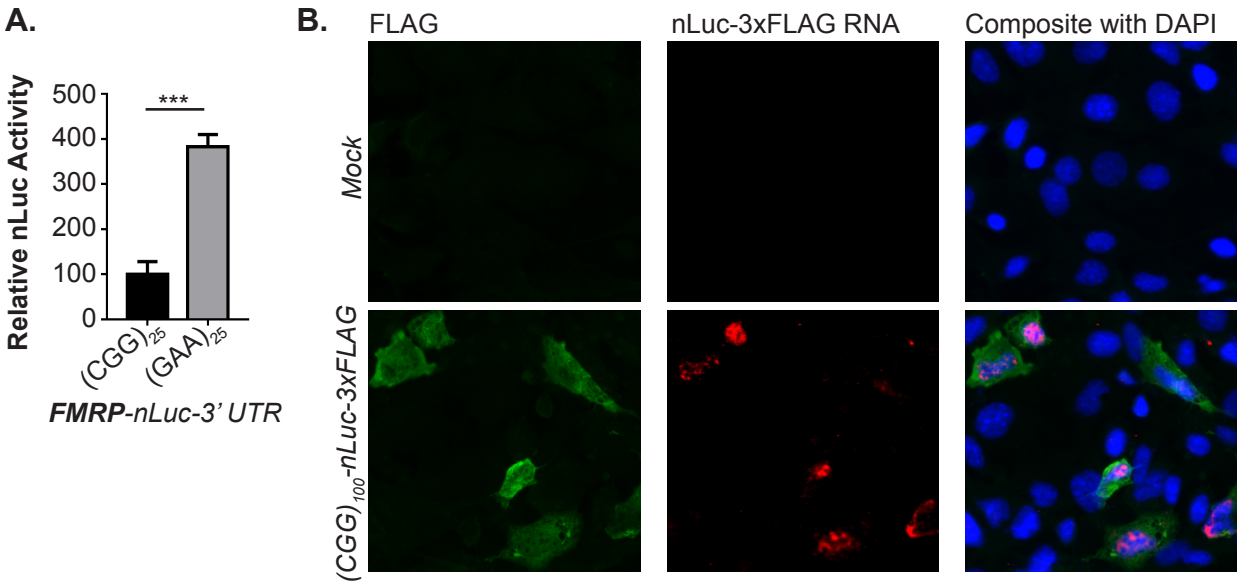


Figure S3.3:

Hybridization chain reaction to detect nanoluciferase reporter RNA. A) Replacing the $(CGG)_{25}$ repeat with a $(GAA)_{25}$ repeat *in vitro* results in an increase in basal nLuc signal from FMRP reporters. n=3. Student's t-test, ***p<0.001. Graph represents mean +/- S.E.M. B) Mouse embryonic fibroblasts were transfected with +1 CGG_{100} RAN-nLuc reporters or mock transfected were probed for nLuc RNA by HCR and co-stained for FLAG expression. The specificity of the probes for nLuc is illustrated by the presence of signal (red) in transfected cells only. Mock transfected cells do not exhibit probe binding.

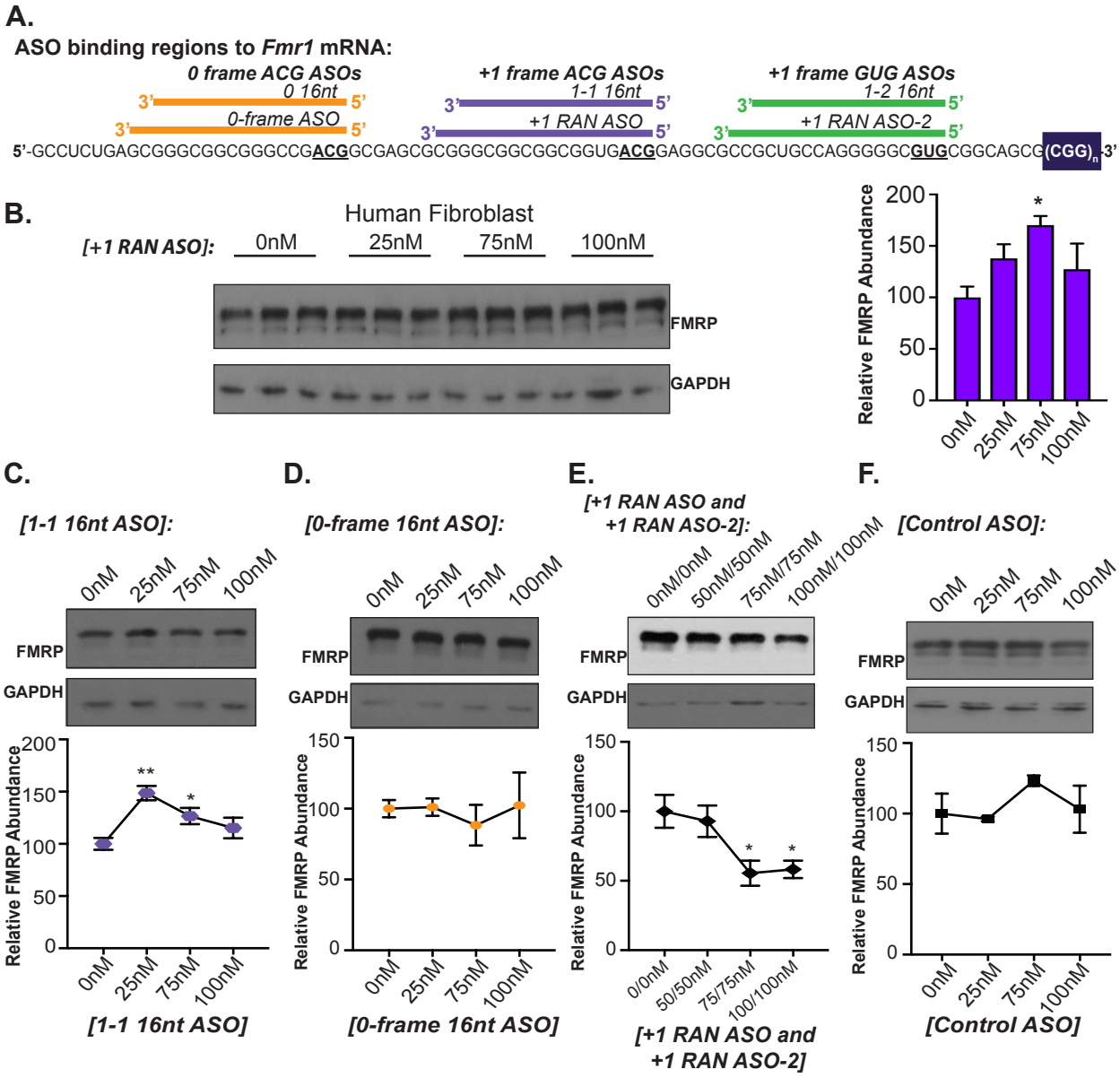


Figure S3.4:

Characterization of all tested RAN-targeting ASOs. A) Schematic of all tested non-cleaving RANT blocking ASOs. Colored bars overlap the corresponding FMR1 5' UTR sequence and start sites; 0 frame ACG(orange), +1 frame ACG (purple), and +1 frame GUG (green). B) *+1 RAN ASO* was transfected into human fibroblasts for 24 hours, producing a significant increase in FMRP at 75nM. C-F) Effect of ASOs targeting the indicated start sites as well as 18 nucleotide "non-targeting" *control ASO* in HEK293 cells at increasing doses. Combinatorial approaches using multiple ASOs either had no significant impacts beyond the single best ASO used or led to a paradoxical decrease in FMRP in transfected cells. n=3 for all conditions. One-way ANOVA with a Fisher's LSD test for dose dependency, *p<0.05, **p<0.01. All graphs represent mean +/- S.E.M.

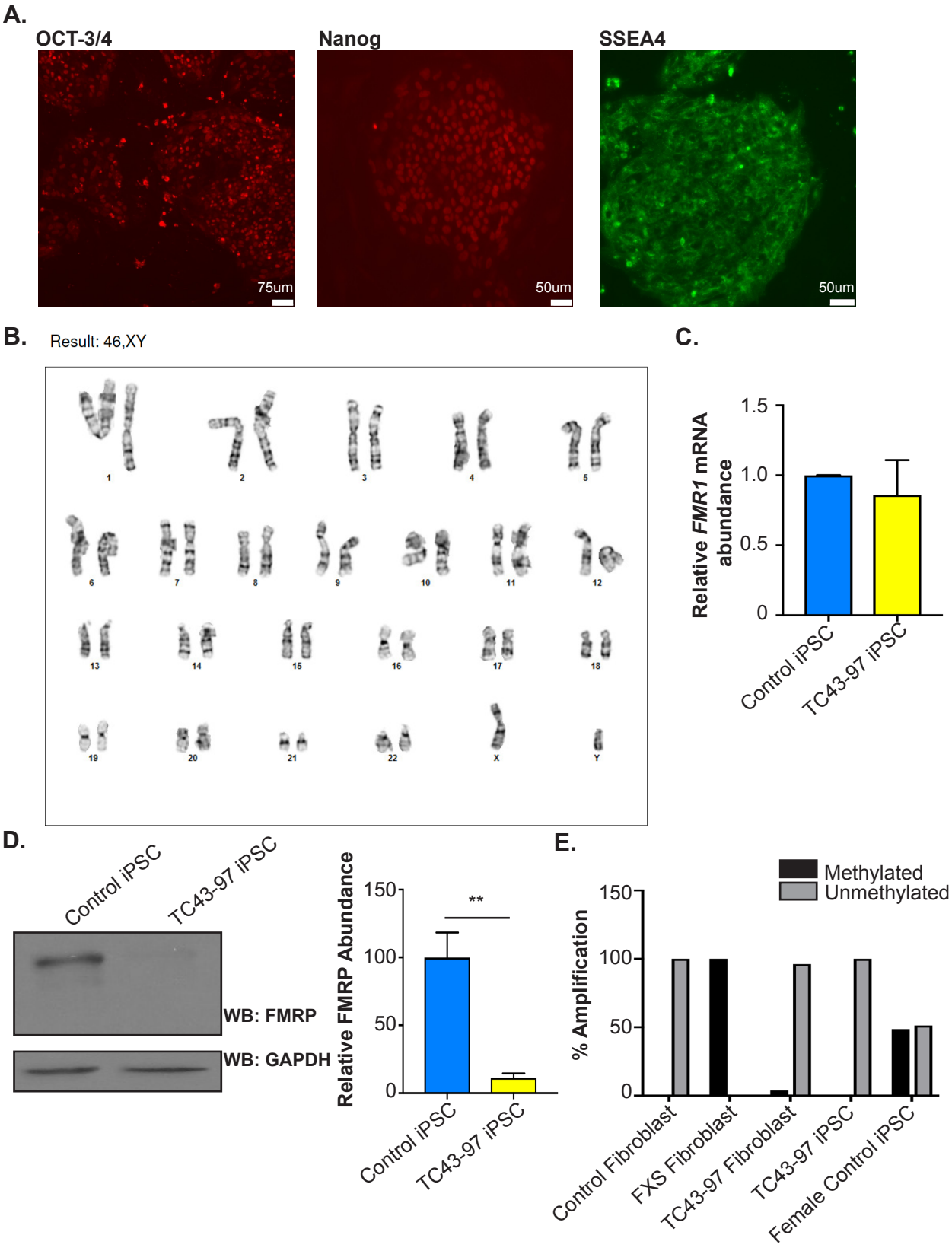

A.
OCT-3/4
Nanog
SSEA4
75um
50um
50um
B.
Result: 46,XY
C.
Relative FMR1 mRNA abundance
1.5
1.0
0.5
0
Control iPSC
TC43-97 iPSC
D.
Control iPSC
TC43-97 iPSC
WB: FMRP
WB: GAPDH
Relative FMRP Abundance
150
100
50
0
**
Control iPSC
TC43-97 iPSC
E.
Methylated
Unmethylated
% Amplification
150
100
50
0
Control Fibroblast
FXS Fibroblast
TC43-97 Fibroblast
TC43-97 iPSC
Female Control iPSC

Figure S3.5:

Characterization of unmethylated Fragile X full mutation iPSC line. A) Detection of three pluripotency markers—OCT-3/4, Nanog, and SSEA4—in the TC43-97 line confirms successful reprogramming of the fibroblast line. The three markers show proper subcellular localization. B) Karyotyping was performed by Cell Line Genetics® on the TC43-97 iPSC line. Cytogenetic analysis of cells in metaphase revealed an apparently normal male karyotype. C) TC43-97 and control iPSCs show no significant difference in *Fmr1* mRNA as determined by RT-qPCR. *Fmr1* mRNA levels are normalized to *18S* in each respective replicate. D) Western blotting of TC43-97 iPSC line shows a decrease in FMRP verses the unaffected control iPSC line. E) Methylation sensitive qPCR of *FMR1* promoter demonstrates lack of DNA methylation in TC43-97. n=3 for each condition. Student's t-test, **$p<0.01$. All graphs represent mean +/- S.E.M.

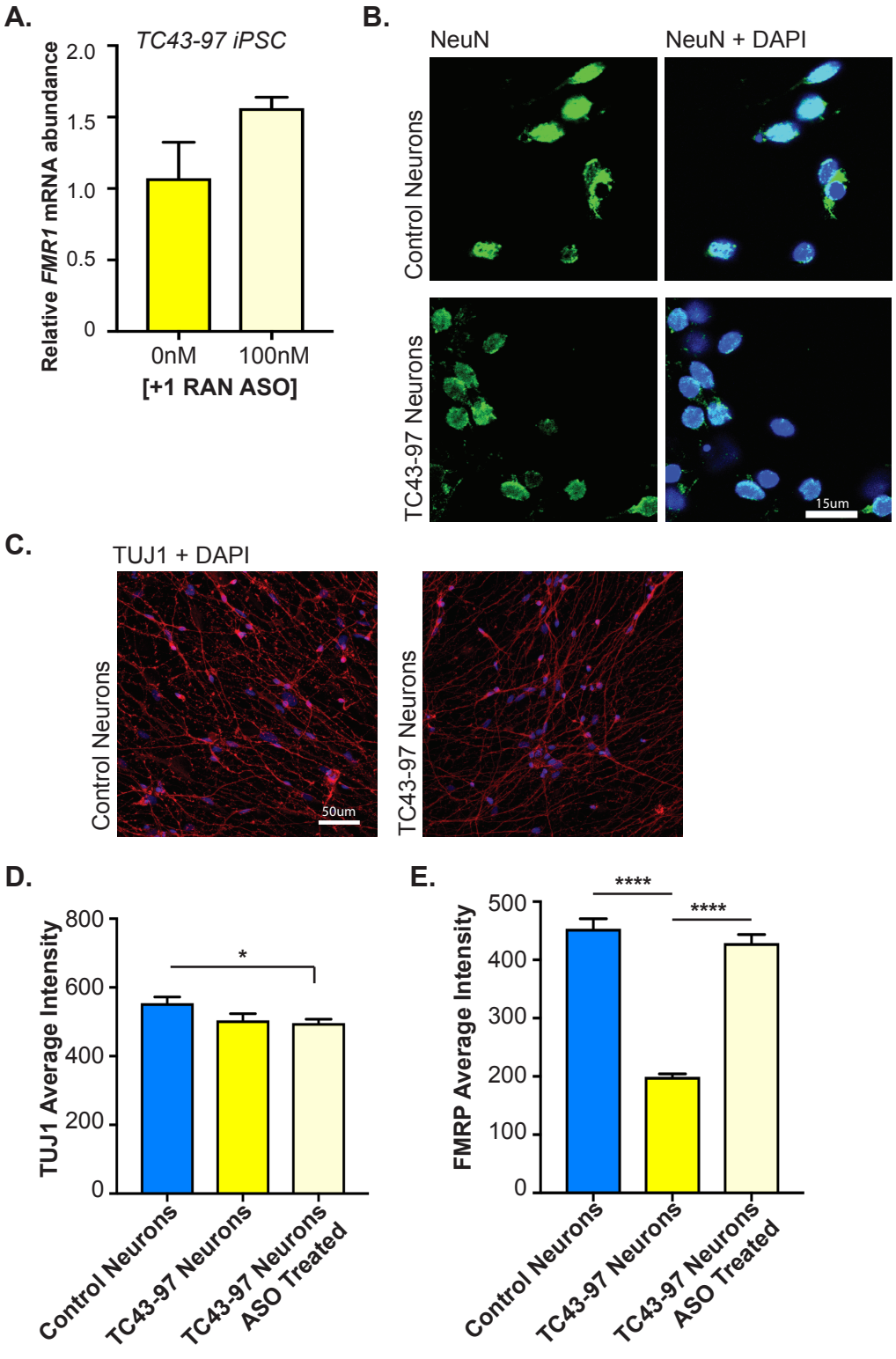


Figure S3.6:

***+1 RAN ASO* specifically increases FMRP levels in iPSC-derived neurons**. A) RT-qPCR was performed on ASO treated TC43-97 iPSCs. No significant change in *Fmr1* mRNA is detected in response to ASO treatment. n=3.

Student's t-test, n.s.=not significant. B) Immunocytochemistry performed on iPSC derived neurons confirms a neuronal cell-type with NeuN positive nuclei in the Control and TC43-97 cell line. C) Soma and processes of differentiated iPSCs are TUJ1 positive in both cell lines. D) TUJ1 immunofluorescence calculated as mean intensity for Control untreated neurons and TC43-97 ASO treated and untreated neurons. There was no difference induced by ASO treatment in TC43-97 derived neurons. There was a ~10% difference in TUJ1 immunofluorescence between control and treated TC43-97 neurons that was insufficient to explain the observed differences in FMRP expression (see Panel E, this figure). E) Average intensity of cellular FMRP levels show a significant decrease in the untreated TC43-97 neurons compared to control neurons and *+1 RAN ASO* treated TC43-97 neurons independent of normalization to TUJ1. n=118 for the untreated TC43-97 condition, n=100 for the treated TC43-97 condition, and n=95 for untreated control neurons. One-way ANOVA with multiple comparisons, ****p<0.0001. All graphs represent mean +/- S.E.M.

Supplemental Tables

Table 3.1 Vectors used in this study

5' UTR of FMR1 up to **XhoI site**	ACTTCCGGTGGAGGGCCGCCTCTGAGCGGGCGGCGGGCCGACGGCGAGCGCGGGCGG CGGCGGTGACGGAGGCGCCGCTGCCAGGGGGCGTGCGGCAGCGCGGCGGCGGCGGC GGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC GGCTGGGC**CTCGAG** **Mutations:** 0-AAA; +1 AAA or +1 AUG(at ACG); 0/+1 AAA			
Reporter Name	Full Reporter Description	Reports for:	Sequence downstream of CGG repeat, starting at XhoI restriction site	Mutations
CGG_n FMRP-nLuc	5' UTR *FMR1*-1st coding exon of FMRP-Nanoluciferase-3xFLAG	**FMRP**-nLuc-3xFlag	**CTCGAGCGCCCGCAGCCCACCTCTCGG** **GGGCGGGCTCCCGGCACTAGCAGGGC** **TGAAGAGAAG**ATGGAGGAGCTGGTGG TGGAAGTGCGGGGCTCCAATGGCGCTT TCTACAAGGCATTTGAAAGCGGCCGCG *GGGTCTTCACACTCGAAGATTTCGTTG* *GGGACTGGCGACAGACAGCCGGCTACA* *ACCTGGACCAAGTCCTTGAACAGGGAG* *GTGTGTCCAGTTTGTTTCAGAATCTCGG* *GGTGTCCGTAACTCCGATCCAAAGGATT* *GTCCTGAGCGGTGAAAATGGGCTGAAG* *ATCGACATCCATGTCATCATCCCGTATGA* *AGGTCTGAGCGGCGACCAAATGGGCCA* *GATCGAAAAAATTTTTAAGGTGGTGTAC* *CCTGTGGATGATCATCACTTTAAGGTGA* *TCCTGCACTATGGCACACTGGTAATCGA* *CGGGGTTACGCCGAACATGATCGACTAT* *TTCGGACGGCCGTATGAAGGCATCGCCG* *TGTTCGACGGCAAAAAGATCACTGTAAC* *AGGGACCCTGTGGAACGGCAACAAAATT* *ATCGACGAGCGCCTGATCAACCCCGACG* *GCTCCCTGCTGTTCCGAGTAACCATCAAC* *GGAGTGACCGGCTGGCGGCTGTGCGAA* *CGCATTCTGGCGGACTACAAAGACCATG* *ACGGTGATTATAAAGATCATGACATCGA* *TTACAAGGATGACGATGACAAGCGATCG* *TAA*	0-AAA +1 AAA +1 AUG 0/+1 AAA
+1 CGG_n RAN-	5' UTR *FMR1* in the +1 frame to XhoI site-Nanoluciferase-	**FMRpolyG**-nLuc-3xFlag	**CTCGAG**GATATCAAGATCTGGCCTCGG CGGCCAAGCTTGGCAATCCGGTACTGT TGGTAAAGCCACCG*GGGTCTTCACACT*	0-AAA

nLuc	3xFLAG		*CGAAGATTTCGTTG GGGACTGGCGACAGACAGCCGGCTACA ACCTGGACCAAGTCCTTGAACAGGGAG GTGTGTCCAGTTTGTTTCAGAATCTCGG GGTGTCCGTAACTCCGATCCAAAGGATT GTCCTGAGCGGTGAAAATGGGCTGAAG ATCGACATCCATGTCATCATCCCGTATGA AGGTCTGAGCGGCGACCAAATGGGCCA GATCGAAAAAATTTTTAAGGTGGTGTAC CCTGTGGATGATCATCACTTTAAGGTGA TCCTGCACTATGGCACACTGGTAATCGA CGGGGTTACGCCGAACATGATCGACTAT TTCGGACGGCCGTATGAAGGCATCGCCG TGTTCGACGGCAAAAAGATCACTGTAAC AGGGACCCTGTGGAACGGCAACAAAATT ATCGACGAGCGCCTGATCAACCCCGACG GCTCCCTGCTGTTCCGAGTAACCATCAAC GGAGTGACCGGCTGGCGGCTGTGCGAA CGCATTCTGGCGGACTACAAAGACCATG ACGGTGATTATAAAGATCATGACATCGA TTACAAGGATGACGATGACAAGCGATCG TAA*	
+2 CGG_n RAN-nLuc	5’ UTR *FMR1* in the +2 frame to first stop codon after the repeat-Nanoluciferase-3xFLAG	**FMRpolyA-**nLuc-3xFlag	**CTCGAGCGCCCGCAGCCCACCTCTCG GGGGCGGGCTCCCGGCGCG**GCCGCG *GGGTCTTCACACTCGAAGATTTCGTTG GGGACTGGCGACAGACAGCCGGCTAC AACCTGGACCAAGTCCTTGAACAGGG AGGTGTGTCCAGTTTGTTTCAGAATCT CGGGGTGTCCGTAACTCCGATCCAAA GGATTGTCCTGAGCGGTGAAAATGGG CTGAAGATCGACATCCATGTCATCATC CCGTATGAAGGTCTGAGCGGCGACCA AATGGGCCAGATCGAAAAAATTTTTA AGGTGGTGTACCCTGTGGATGATCAT CACTTTAAGGTGATCCTGCACTATGGC ACACTGGTAATCGACGGGGTTACGCC GAACATGATCGACTATTTCGGACGGC CGTATGAAGGCATCGCCGTGTTCGAC GGCAAAAAGATCACTGTAACAGGGA CCCTGTGGAACGGCAACAAAATTATC GACGAGCGCCTGATCAACCCCGACGG CTCCCTGCTGTTCCGAGTAACCATCAA CGGAGTGACCGGCTGGCGGCTGTGC GAACGCATTCTGGCGGACTACAAAG ACCATGACGGTGATTATAAAGATCAT GACATCGATTACAAGGATGACGATGA CAAGCGATCGTAA*	0-AAA +1 AAA 0/+1 AAA +1 AUG
+0 CGG_n RAN-nLuc	5’ UTR *FMR1* in the 0-frame to XhoI site after repeat-Nanoluciferase-3xFLAG	**FMRpolyR-**nLuc-3xFlag	**CTCGAG**G*GGGTCTTCACACTCGAAGATTT CGTTGGGGACTGGCGACAGACAGCCGGC TACAACCTGGACCAAGTCCTTGAACAGGG AGGTGTGTCCAGTTTGTTTCAGAATCTCG GGGTGTCCGTAACTCCGATCCAAAGGATT GTCCTGAGCGGTGAAAATGGGCTGAAG ATCGACATCCATGTCATCATCCCGTATGA AGGTCTGAGCGGCGACCAAATGGGCCA GATCGAAAAAATTTTTAAGGTGGTGTAC CCTGTGGATGATCATCACTTTAAGGTGA TCCTGCACTATGGCACACTGGTAATCGA CGGGGTTACGCCGAACATGATCGACTAT TTCGGACGGCCGTATGAAGGCATCGCCG*	

			TGTTCGACGGCAAAAAGATCACTGTAAC *AGGGACCCTGTGGAACGGCAACAAAATT* *ATCGACGAGCGCCTGATCAACCCCGACG* *GCTCCCTGCTGTTCCGAGTAACCATCAAC* *GGAGTGACCGGCTGGCGGCTGTGCGAA* *CGCATTCTGGCGGACTACAAAGACCATG* *ACGGTGATTATAAAGATCATGACATCGA* *TTACAAGGATGACGATGACAAGCGATCG* *TAA*	
+1 CGG_n RAN-Venus	5' UTR *FMR1*-+1 frame of first FMRP exon to first stop codon-Venus	**FMRpolyG-**Venus	**CTCGAGCGCCCGCAGCCCACCTCTCGG** **GGGCGGGCTCCCGGCGCTAGCAGGGC** **TGAAGAGAAGATGGAGGAGCTGGTG** **GTGGAAGTGCGGGGCTCCA**CCGCGG*A* *GCAAGGGCGAGGAGCTGTTCACCGGG* *GTGGTGCCCATCCTGGTCGAGCTGGAC* *GGCGACGTAAACGGCCACAAGTTCAGC* *GTGTCCGGCGAGGGCGAGGGCGATGC* *CACCTACGGCAAGCTGACCCTGAAGCT* *GATCTGCACCACCGGCAAGCTGCCCGT* *GCCCTGGCCCACCCTCGTGACCACCCTG* *GGCTACGGCCTGCAGTGCTTCGCCCGC* *TACCCCGACCACATGAAGCAGCACGAC* *TTCTTCAAGTCCGCCATGCCCGAAGGCT* *ACGTCCAGGAGCGCACCATCTTCTTCAA* *GGACGACGGCAACTACAAGACCCGCGC* *CGAGGTGAAGTTCGAGGGCGACACCCT* *GGTGAACCGCATCGAGCTGAAGGGCA* *TCGACTTCAAGGAGGACGGCAACATCC* *TGGGGCACAAGCTGGAGTACAACTAC* *AACAGCCACAACGTCTATATCACCGCC* *GACAAGCAGAAGAACGGCATCAAGGC* *CAACTTCAAGATCCGCCACAACATCGA* *GGACGGCGGCGTGCAGCTCGCCGACC* *ACTACCAGCAGAACACCCCCATCGGCG* *ACGGCCCCGTGCTGCTGCCCGACAACC* *ACTACCTGAGCTACCAGTCCGCCCTGA* *GCAAAGACCCCAACGAGAAGCGCGAT* *CACATGGTCCTGCTGGAGTTCGTGACC* *GCCGCCGGGATCACTCTCGGCATGGA* *CGAGCTGTACAAGCCCTGCAGGAACT* *GCATAATTCTGAAGTCTAGAGGGCCC* *GTTTAA*	
CGG_n FMRP-nLuc-3'UTR	5' UTR *FMR1*-1st coding exon of FMRP-Nanoluciferase-3xFLAG-Pest tag-3' UTR FMR1	**FMRP**-nLuc-Pest-3xFlag	**CTCGAGCGCCCGCAGCCC** **ACCTCTCGGGGGCGGGCTCCCGGCGC** **TAGCAGGGCTGAAGAGAAG**ATGGAG GAGCTGGTGGTGGAAGTGCGGGGCT CCAATGGCGCTTTCTACAAGGCATTT GAAAGCGGCCGCGATTTCGTTG*GGG* *ACTGGCGACAGACAGCCGGCTACAA* *CCTGGACCAAGTCCTTGAACAGGGA* *GGTGTGTCCAGTTTGTTTCAGAATCTC* *GGGGTGTCCGTAACTCCGATCCAAAG* *GATTGTCCTGAGCGGTGAAAATGGG* *CTGAAGATCGACATCCATGTCATCAT* *CCCGTATGAAGGTCTGAGCGGCGACC* *AAATGGGCCAGATCGAAAAAATTTTT* *AAGGTGGTGTACCCTGTGGATGATCA* *TCACTTTAAGGTGATCCTGCACTATG* *GCACACTGGTAATCGACGGGGTTAC* *GCCGAACATGATCGACTATTTCGGAC*	-0/+1 AAA -GAA in place of CGG

			GGCCGTATGAAGGCATCGCCGTGTTC *GACGGCAAAAAGATCACTGTAACAG* *GGACCCTGTGGAACGGCAACAAAAT* *TATCGACGAGCGCCTGATCAACCCCG* *ACGGCTCCCTGCTGTTCCGAGTAACC* *ATCAACGGAGTGACCGGCTGGCGGC* *TGTGCGAACGCATTCTGGCGGACTAC* *AAAGACGATGACGGTGATTATAAAGA* *TCATGACATCGATTACAAGGATGACG* *ATGACAAGCGATCGAATTCTCACGGC* *TTTCCGCCTGAGGTTGAAGAGCAAGC* *CGCCGGTACATTGCCTATGTCCTGCG* *CACAAGAAAGCGGTATGGACCGGCA* *CCCAGCCGCTTGTGCTTCAGCTCGCAT* *CAACGTCTTAATCTAGAA*CTGCATAA TTCTGAAGTTATATTTCCTATACCATTT CCGTAATTCTTATTCCATATTAGAAAA CTTTGTTAGGCCAAAGACAAATAGTA GGCAAGATGGCACAGGGCATGAAAT GAACACAAATTATGCTAAGAATTTTTT ATTTTTTGGTATTGGCCATAAGCAACA ATTTTCAGATTTGCACAAAAAGATACC TTAAAATTTGAAACATTGCT…(FMR1 3’ UTR)	
+1 CGG_n RAN-nLuc-3’UTR	5’ UTR *FMR1* in the +1 frame to XhoI site -Nanoluciferase-3xFLAG-Pest tag-3’ UTR FMR1	**FMRpolyG-**nLuc-Pest-3xFlag	**CTCGAG**GATATCAAGATCTGGCCT CGGCGGCCAAGCTTGGCAATCCG GTACTGTTGGTAAAGCCACCGGGGTC TTCACACTCGAAGATTTCGTTGGG GACTGGCGACAG*ACAGCCGGCTACAA* *CCTGGACCAAGTCCTTGAACAGGGA* *GGTGTGTCCAGTTTGTTTCAGAATCTC* *GGGGTGTCCGTAACTCCGATCCAAAG* *GATTGTCCTGAGCGGTGAAAATGGG* *CTGAAGATCGACATCCATGTCATCAT* *CCCGTATGAAGGTCTGAGCGGCGACC* *AAATGGGCCAGATCGAAAAAATTTTT* *AAGGTGGTGTACCCTGTGGATGATCA* *TCACTTTAAGGTGATCCTGCACTATG* *GCACACTGGTAATCGACGGGGTTAC* *GCCGAACATGATCGACTATTTCGGAC* *GGCCGTATGAAGGCATCGCCGTGTTC* *GACGGCAAAAAGATCACTGTAACAG* *GGACCCTGTGGAACGGCAACAAAAT* *TATCGACGAGCGCCTGATCAACCCCG* *ACGGCTCCCTGCTGTTCCGAGTAACC* *ATCAACGGAGTGACCGGCTGGCGGC* *TGTGCGAACGCATTCTGGCGGACTAC* *AAAGACGATGACGGTGATTATAAAGA* *TCATGACATCGATTACAAGGATGACG* *ATGACAAGCGATCGAATTCTCACGGC* *TTTCCGCCTGAGGTTGAAGAGCAAGC* *CGCCGGTACATTGCCTATGTCCTGCG* *CACAAGAAAGCGGTATGGACCGGCA* *CCCAGCCGCTTGTGCTTCAGCTCGCAT* *CAACGTCTTAATCTAGAA*CTGCATAA TTCTGAAGTTATATTTCCTATACCATTT CCGTAATTCTTATTCCATATTAGAAAA CTTTGTTAGGCCAAAGACAAATAGTA GGCAAGATGGCACAGGGCATGAAAT GAACACAAATTATGCTAAGAATTTTTT ATTTTTTGGTATTGGCCATAAGCAACA	

			ATTTTCAGATTTGCACAAAAAGATACC TTAAAATTTGAAACATTGCT…(FMR1 3'UTR)	
pGW +1 CGG_{100} GFP		**FMRpolyG-**GFP	**CTCGAGCGCCCGCAGCCCACCTCTCGG GGGCGGGCTCCCGGCGCTAGCAGGGC TGAAGAGAAGATGGAGGAGCTGGTGG TGGAAGTGCGGGGCTCCA*ATGGCGCTT TCTACAAGGCATTTGAAAG***CGGCCGC ACCGGTCGCCACC*ATGGT* *GAGCAAGGGCGAGGAGCTGTTCACC* *GGGGTGGTGCCCATCCTGGTCGAGCT* *GGACGGCGACGTAAACGGCCACAAG* *TTCAGCGTGTCCGGCGAGGGCGAGG* *GCGATGCCACCTACGGCAAGCTGACC* *CTGAAGTTCATCTGCACCACCGGCAA* *GCTGCCCGTGCCCTGGCCCACCCTCG* *TGACCACCCTGACCTACGGCGTGCAG* *TGCTTCAGCCGCTACCCCGACCACATG* *AAGCAGCACGACTTCTTCAAGTCCGCC* *ATGCCCGAAGGCTACGTCCAGGAGCG* *CACCATCTTCTTCAAGGACGACGGCA* *ACTACAAGACCCGCGCCGAGGTGAAG* *TTCGAGGGCGACACCCTGGTGAACCG* *CATCGAGCTGAAGGGCATCGACTTCA* *AGGAGGACGGCAACATCCTGGGGCA* *CAAGCTGGAGTACAACTACAACAGCC* *ACAACGTCTATATCATGGCCGACAAG* *CAGAAGAACGGCATCAAGGTGAACTT* *CAAGATCCGCCACAACATCGAGGACG* *GCAGCGTGCAGCTCGCCGACCACTAC* *CAGCAGAACACCCCCATCGGCGACGG* *CCCCGTGCTGCTGCCCGACAACCACTA* *CCTGAGCACCCAGTCCGCCCTGAGCAA* *AGACCCCAACGAGAAGCGCGATCACA* *TGGTCCTGCTGGAGTTCGTGACCGCCG* *CCGGGATCACTCTCGGCATGGACGAG* *CTGTACAAGTAA*	

Table 3.2 Primers used in this study

Mutation	**Forward Primer (5' to 3')**	**Reverse Primer (3' to 5')**
PEST tag Q5 SDM (step 1)	aagccgccggtacattgcctatgtc ctgcgcacaTAAGGCCGCGACTCTAGA	Gctcttcaacctcaggcggaaagccgtg agaattCTTGTCATCGTCATCCTTGTAATC
PEST tag Q5 SDM (step 2)	gcttgtgcttcagctcgcatcaacgt cTAAGGCCGCGACTCTAGAG	ggctgggtgccggtccataccgctttctTG TGCGCAGGACATAGGC
3X Stop downstream of repeat Q5 SDM	attaaCCGCAGCCCACCTCTCGG	taattaGCGCTCGAGGCCCAGCCG
3X Stop upstream of repeat	TCAGCTCCGTTTCGGTTTCACTTC CGGTGGAGGGCCGCCTCTGAGCG GGCGGCGGGCCGACGGCGAGCG CGGGCGGCGGCGGTGACGTAAT TAATTAAGAGG	CGCCTCTTAATTAATTACGTCAC CGCCGCCGCCCGCGCTCGCCGT CGGCCCGCCGCCCGCTCAGAGG CGGCCCTCCACCGGAAGTGAAA CCGAAACGGAGC
PCR cloning *FMR1* 3' UTR	actgcattaCCCCGGGCATCATGACTTTG AACTGAAAAACATACA	actgcattaCTCTAGAACTGCATAATT CTGAAGTTATATTTCCTATACC
Full nLuc-3xFlag-PEST XmaI/XbaI	cataattttaGCGGCCGCGATTTCGTTG GGGAC	cataattttaTCTAGATTAAGACGTTGA TGCGAGCTGAAG

Cloning Venus into CGG vector	attaattaCTCGAGCGCCCGCAGC	tttaactatTCTAGACTTCAGAATTAT GCAGTTCCTGCAGGGC
	Forward Primer (5' to 3')	**Reverse Primer (3' to 5')**
Fmr1 qPCR	CATGAAGATTCAATAACAGTTGC	CACTTTAGCTAACCACCAACA
Bisulfite qPCR for *FMR1* promoter (methylated)	GGTCGAAAGATAGACGCGC	AAACAATGCGACCTATCACCG
Bisulfite qPCR for *FMR1* promoter (universal)	TGTTGGTTTGTTGTTTGTTTAGA	AACATAATTTCAATATTTACACCC
18S qPCR	GGCCCTGTAATTGGAATGAGTC	CAAGATCCAACTACGAGCTT
NL qPCR	CAGCCGGCTACAACCTGGAC	GCCCATTTTCACCGCTCAG
FF qPCR	GCAGTACCGGATTGCCCAAG	GTCGGGGATGATCTGGTTGC

Table 3.3 Hybridization Chain Reaction Probe sequences

HCR_DNA_nLuc DNA Probe_1	GAGGAGGGCAGCAAACGGGAAGAGTCTTCCTTTACGATATT TTGTAGCCGGCTGTCTGTCGCCAGTCCCCAACGAAATCTTCGAGTGTGAA ATATA GCATTCTTTCTTGAGGAGGGCAGCAAACGGGAAGAG
HCR_DNA_nLuc DNA Probe_2	GAGGAGGGCAGCAAACGGGAAGAGTCTTCCTTTACGATATT TTGGATCGGAGTTACGGACACCCCGAGATTCTGAAACAAACTGGACACAC ATATA GCATTCTTTCTTGAGGAGGGCAGCAAACGGGAAGAG
HCR_DNA_nLuc DNA Probe_3	GAGGAGGGCAGCAAACGGGAAGAGTCTTCCTTTACGATATT AAATTTTTTCGATCTGGCCCATTTGGTCGCCGCTCAGACCTTCATACGGG ATATA GCATTCTTTCTTGAGGAGGGCAGCAAACGGGAAGAG
HCR_DNA_nLuc DNA Probe_4	GAGGAGGGCAGCAAACGGGAAGAGTCTTCCTTTACGATATT CGTAACCCCGTCGATTACCAGTGTGCCATAGTGCAGGATCACCTTAAAGT ATATA GCATTCTTTCTTGAGGAGGGCAGCAAACGGGAAGAG

Table 3.4 ASOs used in this study

Name	**Sequence**	**Length**	**Chemistry**
0-frame ASO	CGTCGGCCCGCCGCCCGC	18 nucleotides	2'-OMe, PS
0 16nt	CGTCGGCCCGCCGCCC	16 nucleotides	2'-OMe, PS
+1 RAN ASO	CGTCACCGCCGCCGCCCG	18 nucleotides	2'-OMe, PS
1-1 16nt	CGTCACCGCCGCCGCC	16 nucleotides	2'-OMe, PS
+1 RAN ASO-2	CACGCCCCCTGGCAGCGG	18 nucleotides	2'-OMe, PS
1-2 16nt	CACGCCCCCTGGCAGC	16 nucleotides	2'-OMe, PS
Control ASO	CATTGTTTTTTGTCTTCC	18 nucleotides	2'-OMe, PS

Citations

Ascano, M., Jr., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M.*, et al.* (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature *492*, 382-386.

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.

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., 3rd, Rademakers, R.*, et al.* (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. Neuron *77*, 639-646.

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

Barbarese, E., Ifrim, M.F., Hsieh, L., Guo, C., Tatavarty, V., Maggipinto, M.J., Korza, G., Tutolo, J.W., Giampetruzzi, A., Le, H.*, et al.* (2013). Conditional knockout of tumor overexpressed gene in mouse neurons affects RNA granule assembly, granule translation, LTP and short term habituation. PloS one *8*, e69989.

Barmada, S.J., Ju, S., Arjun, A., Batarse, A., Archbold, H.C., Peisach, D., Li, X., Zhang, Y., Tank, E.M., Qiu, H.*, et al.* (2015). Amelioration of toxicity in neuronal models of amyotrophic lateral sclerosis by hUPF1. Proceedings of the National Academy of Sciences of the United States of America *112*, 7821-7826.

Bear, M.F., Huber, K.M., and Warren, S.T. (2004). The mGluR theory of fragile X mental retardation. Trends in neurosciences *27*, 370-377.

Bell, M.V., Hirst, M.C., Nakahori, Y., MacKinnon, R.N., Roche, A., Flint, T.J., Jacobs, P.A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U.*, et al.* (1991). Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. Cell *64*, 861-866.

Berry-Kravis, E., Abrams, L., Coffey, S.M., Hall, D.A., Greco, C., Gane, L.W., Grigsby, J., Bourgeois, J.A., Finucane, B., Jacquemont, S.*, et al.* (2007). Fragile X-associated tremor/ataxia syndrome: clinical features, genetics, and testing guidelines. Movement disorders : official journal of the Movement Disorder Society *22*, 2018-2030, quiz 2140.

Bhakar, A.L., Dolen, G., and Bear, M.F. (2012). The pathophysiology of fragile X (and what it teaches us about synapses). Annual review of neuroscience *35*, 417-443.

Brar, G.A. (2016). Beyond the Triplet Code: Context Cues Transform Translation. Cell *167*, 1681-1692.

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.

Brown, V., Jin, P., Ceman, S., Darnell, J.C., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D.*, et al.* (2001). Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. Cell *107*, 477-487.

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. Human reproduction *31*, 158-168.

Burman, R.W., Popovich, B.W., Jacky, P.B., and Turker, M.S. (1999). Fully expanded FMR1 CGG repeats exhibit a length- and differentiation-dependent instability in cell hybrids that is independent of DNA methylation. Human molecular genetics *8*, 2293-2302.

Cajigas, I.J., Will, T., and Schuman, E.M. (2010). Protein homeostasis and synaptic plasticity. The EMBO journal *29*, 2746-2752.

Ceman, S., O'Donnell, W.T., Reed, M., Patton, S., Pohl, J., and Warren, S.T. (2003). Phosphorylation influences the translation state of FMRP-associated polyribosomes. Human molecular genetics *12*, 3295-3305.

Chen, E., Sharma, M.R., Shi, X., Agrawal, R.K., and Joseph, S. (2014). Fragile X mental retardation protein regulates translation by binding directly to the ribosome. Molecular cell *54*, 407-417.

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. Human molecular genetics *12*, 3067-3074.

Choi, H.M., Calvert, C.R., Husain, N., Huss, D., Barsi, J.C., Deverman, B.E., Hunter, R.C., Kato, M., Lee, S.M., Abelin, A.C.*, et al.* (2016). Mapping a multiplexed zoo of mRNA expression. Development *143*, 3632-3637.

Choi, H.M., Chang, J.Y., Trinh le, A., Padilla, J.E., Fraser, S.E., and Pierce, N.A. (2010). Programmable in situ amplification for multiplexed imaging of mRNA expression. Nature biotechnology *28*, 1208-1212.

Cleary, J.D., and Ranum, L.P. (2017). New developments in RAN translation: insights from multiple diseases. Current opinion in genetics & development *44*, 125-134.

Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.Y., Gerhardt, J., Disney, M.D., and Jaffrey, S.R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science *343*, 1002-1005.

Collins, S.C., Bray, S.M., Suhl, J.A., Cutler, D.J., Coffee, B., Zwick, M.E., and Warren, S.T. (2010). Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. American journal of medical genetics Part A *152A*, 2512-2520.

Costa-Mattioli, M., Sossin, W.S., Klann, E., and Sonenberg, N. (2009). Translational control of long-lasting synaptic plasticity and memory. Neuron *61*, 10-26.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W.*, et al.* (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247-261.

Di Prisco, G.V., Huang, W., Buffington, S.A., Hsu, C.C., Bonnen, P.E., Placzek, A.N., Sidrauski, C., Krnjevic, K., Kaufman, R.J., Walter, P.*, et al.* (2014). Translational control of mGluR-dependent long-term depression and object-place learning by eIF2alpha. Nature neuroscience *17*, 1073-1082.

Dirks, R.M., and Pierce, N.A. (2004). Triggered amplification by hybridization chain reaction. Proceedings of the National Academy of Sciences of the United States of America *101*, 15275-15278.

Dolen, G., Osterweil, E., Rao, B.S., Smith, G.B., Auerbach, B.D., Chattarji, S., and Bear, M.F. (2007). Correction of fragile X syndrome in mice. Neuron *56*, 955-962.

Eichler, E.E., Kunst, C.B., Lugenbeel, K.A., Ryder, O.A., Davison, D., Warren, S.T., and Nelson, D.L. (1995). Evolution of the cryptic FMR1 CGG repeat. Nature genetics *11*, 301-308.

Feng, Y., Gutekunst, C.A., Eberhart, D.E., Yi, H., Warren, S.T., and Hersch, S.M. (1997). Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. The Journal of neuroscience : the official journal of the Society for Neuroscience *17*, 1539-1547.

Feng, Y., Zhang, F., Lokey, L.K., Chastain, J.L., Lakkis, L., Eberhart, D., and Warren, S.T. (1995). Translational suppression by trinucleotide repeat expansion at FMR1. Science *268*, 731-734.

Finkel, R.S., Chiriboga, C.A., Vajsar, J., Day, J.W., Montes, J., De Vivo, D.C., Yamashita, M., Rigo, F., Hung, G., Schneider, E.*, et al.* (2016). Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. Lancet *388*, 3017-3026.

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.

Gantois, I., Khoutorsky, A., Popic, J., Aguilar-Valles, A., Freemantle, E., Cao, R., Sharma, V., Pooters, T., Nagpal, A., Skalecka, A.*, et al.* (2017). Metformin ameliorates core deficits in a mouse model of fragile X syndrome. Nature medicine *23*, 674-677.

Gemayel, R., Vinces, M.D., Legendre, M., and Verstrepen, K.J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annual review of genetics *44*, 445-477.

Gkogkas, C.G., Khoutorsky, A., Cao, R., Jafarnejad, S.M., Prager-Khoutorsky, M., Giannakas, N., Kaminari, A., Fragkouli, A., Nader, K., Price, T.J.*, et al.* (2014). Pharmacogenetic inhibition of eIF4E-dependent Mmp9 mRNA translation reverses fragile X syndrome-like phenotypes. Cell reports *9*, 1742-1755.

Graber, T.E., Hebert-Seropian, S., Khoutorsky, A., David, A., Yewdell, J.W., Lacaille, J.C., and Sossin, W.S. (2013). Reactivation of stalled polyribosomes in synaptic plasticity. Proceedings of the National Academy of Sciences of the United States of America *110*, 16205-16210.

Greco, C.M., Hagerman, R.J., Tassone, F., Chudley, A.E., Del Bigio, M.R., Jacquemont, S., Leehey, M., and Hagerman, P.J. (2002). Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. Brain : a journal of neurology *125*, 1760-1771.

Hagerman, R.J., Berry-Kravis, E., Hazlett, H.C., Bailey, D.B., Jr., Moine, H., Kooy, R.F., Tassone, F., Gantois, I., Sonenberg, N., Mandel, J.L.*, et al.* (2017). Fragile X syndrome. Nature reviews Disease primers *3*, 17065.

Hagerman, R.J., and Hagerman, P. (2016). Fragile X-associated tremor/ataxia syndrome - features, mechanisms and management. Nature reviews Neurology *12*, 403-412.

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

Ho, V.M., Lee, J.A., and Martin, K.C. (2011). The cell biology of synaptic plasticity. Science *334*, 623-628.

Hou, L., Antion, M.D., Hu, D., Spencer, C.M., Paylor, R., and Klann, E. (2006). Dynamic translational and proteasomal regulation of fragile X mental retardation protein controls mGluR-dependent long-term depression. Neuron *51*, 441-454.

Huber, K.M., Gallagher, S.M., Warren, S.T., and Bear, M.F. (2002). Altered synaptic plasticity in a mouse model of fragile X mental retardation. Proceedings of the National Academy of Sciences of the United States of America *99*, 7746-7750.

Huber, K.M., Kayser, M.S., and Bear, M.F. (2000). Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. Science *288*, 1254-1257.

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.*, et al.* (2015). Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. Human molecular genetics *24*, 4948-4957.

Huss, D., Choi, H.M., Readhead, C., Fraser, S.E., Pierce, N.A., and Lansford, R. (2015). Combinatorial analysis of mRNA expression patterns in mouse embryos using hybridization chain reaction. Cold Spring Harbor protocols *2015*, 259-268.

Ifrim, M.F., Williams, K.R., and Bassell, G.J. (2015). Single-Molecule Imaging of PSD-95 mRNA Translation in Dendrites and Its Dysregulation in a Mouse Model of Fragile X Syndrome. The Journal of neuroscience : the official journal of the Society for Neuroscience *35*, 7116-7130.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature protocols *7*, 1534-1550.

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.

Jacquemont, S., Curie, A., des Portes, V., Torrioli, M.G., Berry-Kravis, E., Hagerman, R.J., Ramos, F.J., Cornish, K., He, Y., Paulding, C.*, et al.* (2011). Epigenetic modification of the FMR1 gene in fragile X syndrome is associated with differential response to the mGluR5 antagonist AFQ056. Science translational medicine *3*, 64ra61.

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. Molecular cell *62*, 314-322.

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

Kong, H.E., Zhao, J., Xu, S., Jin, P., and Jin, Y. (2017). Fragile X-Associated Tremor/Ataxia Syndrome: From Molecular Pathogenesis to Development of Therapeutics. Frontiers in cellular neuroscience *11*, 128.

Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. (1991). Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. Science *252*, 1711-1714.

Liang, X.H., Shen, W., Sun, H., Migawa, M.T., Vickers, T.A., and Crooke, S.T. (2016). Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. Nature biotechnology *34*, 875-880.

Liang, X.H., Sun, H., Shen, W., Wang, S., Yao, J., Migawa, M.T., Bui, H.H., Damle, S.S., Riney, S., Graham, M.J.*, et al.* (2017). Antisense oligonucleotides targeting translation inhibitory elements in 5' UTRs can selectively increase protein levels. Nucleic acids research *45*, 9528-9546.

Lozano, R., Rosero, C.A., and Hagerman, R.J. (2014). Fragile X spectrum disorders. Intractable & rare diseases research *3*, 134-146.

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. Journal of molecular biology *407*, 21-34.

Merlin, L.R., Bergold, P.J., and Wong, R.K. (1998). Requirement of protein synthesis for group I mGluR-mediated induction of epileptiform discharges. Journal of neurophysiology *80*, 989-993.

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

Muslimov, I.A., Patel, M.V., Rose, A., and Tiedge, H. (2011). Spatial code recognition in neuronal RNA targeting: role of RNA-hnRNP A2 interactions. The Journal of cell biology *194*, 441-457.

Nalavadi, V.C., Muddashetty, R.S., Gross, C., and Bassell, G.J. (2012). Dephosphorylation-induced ubiquitination and degradation of FMRP in dendrites: a role in immediate early mGluR-stimulated translation. The Journal of neuroscience : the official journal of the Society for Neuroscience *32*, 2582-2587.

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

Nolin, S.L., Glicksman, A., Houck, G.E., Jr., Brown, W.T., and Dobkin, C.S. (1994). Mosaicism in fragile X affected males. American journal of medical genetics *51*, 509-512.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F., and Mandel, J.L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. Science *252*, 1097-1102.

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 research *37*, 2712-2722.

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. Human molecular genetics *24*, 4317-4326.

Osterweil, E.K., Krueger, D.D., Reinhold, K., and Bear, M.F. (2010). Hypersensitivity to mGluR5 and ERK1/2 leads to excessive protein synthesis in the hippocampus of a mouse model of fragile X syndrome. The Journal of neuroscience : the official journal of the Society for Neuroscience *30*, 15616-15627.

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.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome research *20*, 110-121.

Qin, M., Kang, J., Burlin, T.V., Jiang, C., and Smith, C.B. (2005). Postadolescent changes in regional cerebral protein synthesis: an in vivo study in the FMR1 null mouse. The Journal of neuroscience : the official journal of the Society for Neuroscience *25*, 5087-5095.

Rigo, F., Hua, Y., Chun, S.J., Prakash, T.P., Krainer, A.R., and Bennett, C.F. (2012). Synthetic oligonucleotides recruit ILF2/3 to RNA transcripts to modulate splicing. Nature chemical biology *8*, 555-561.

Sellier, C., Buijsen, R.A., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziane, H., Vincent, A.*, et al.* (2017). Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. Neuron *93*, 331-347.

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

Shi, Y., Kirwan, P., and Livesey, F.J. (2012). Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nature protocols *7*, 1836-1846.

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.

Suhl, J.A., Muddashetty, R.S., Anderson, B.R., Ifrim, M.F., Visootsak, J., Bassell, G.J., and Warren, S.T. (2015). A 3' untranslated region variant in FMR1 eliminates neuronal activity-dependent translation of FMRP by disrupting binding of the RNA-binding protein HuR.

Proceedings of the National Academy of Sciences of the United States of America *112*, E6553-6561.

Sullivan, S.D., Welt, C., and Sherman, S. (2011). FMR1 and the continuum of primary ovarian insufficiency. Seminars in reproductive medicine *29*, 299-307.

Sutton, M.A., Taylor, A.M., Ito, H.T., Pham, A., and Schuman, E.M. (2007). Postsynaptic decoding of neural activity: eEF2 as a biochemical sensor coupling miniature synaptic transmission to local protein synthesis. Neuron *55*, 648-661.

Tang, B., Wang, T., Wan, H., Han, L., Qin, X., Zhang, Y., Wang, J., Yu, C., Berton, F., Francesconi, W*., et al.* (2015). Fmr1 deficiency promotes age-dependent alterations in the cortical synaptic proteome. Proceedings of the National Academy of Sciences of the United States of America *112*, E4697-4706.

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.

Tassone, F., Hagerman, R.J., Ikle, D.N., Dyer, P.N., Lampe, M., Willemsen, R., Oostra, B.A., and Taylor, A.K. (1999). FMRP expression as a potential prognostic indicator in fragile X syndrome. American journal of medical genetics *84*, 250-261.

Tatavarty, V., Ifrim, M.F., Levin, M., Korza, G., Barbarese, E., Yu, J., and Carson, J.H. (2012). Single-molecule imaging of translational output from individual RNA granules in neurons. Molecular biology of the cell *23*, 918-929.

Todd, P.K., Mack, K.J., and Malter, J.S. (2003). The fragile X mental retardation protein is required for type-I metabotropic glutamate receptor-dependent translation of PSD-95. Proceedings of the National Academy of Sciences of the United States of America *100*, 14374-14378.

Todd, P.K., and Malter, J.S. (2002). Fragile X mental retardation protein in plasticity and disease. Journal of neuroscience research *70*, 623-630.

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

Todd, P.K., Oh, S.Y., Krans, A., Pandey, U.B., Di Prospero, N.A., Min, K.T., Taylor, J.P., and Paulson, H.L. (2010). Histone deacetylases suppress CGG repeat-induced neurodegeneration via transcriptional silencing in models of fragile X tremor ataxia syndrome. PLoS genetics *6*, e1001240.

Usdin, K., and Kumari, D. (2015). Repeat-mediated epigenetic dysregulation of the FMR1 gene in the fragile X-related disorders. Frontiers in genetics *6*, 192.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P.*, et al.* (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.

Weiler, I.J., and Greenough, W.T. (1993). Metabotropic glutamate receptors trigger postsynaptic protein synthesis. Proceedings of the National Academy of Sciences of the United States of America *90*, 7168-7171.

Weiler, I.J., Irwin, S.A., Klintsova, A.Y., Spencer, C.M., Brazelton, A.D., Miyashiro, K., Comery, T.A., Patel, B., Eberwine, J., and Greenough, W.T. (1997). Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation. Proceedings of the National Academy of Sciences of the United States of America *94*, 5395-5400.

Yu, J., Xiao, J., Ren, X., Lao, K., and Xie, X.S. (2006). Probing gene expression in live cells, one protein molecule at a time. Science *311*, 1600-1603.

Zhou, Y., Kumari, D., Sciascia, N., and Usdin, K. (2016). CGG-repeat dynamics and FMR1 gene silencing in fragile X syndrome stem cells and stem cell-derived neurons. Molecular autism *7*, 42.

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.*, et al.* (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proceedings of the National Academy of Sciences of the United States of America *108*, 260-265.

CHAPTER IV

Future Directions and General Discussion

Introduction

The process of mRNA translation is central to cellular life and its regulation plays critical roles in neuronal function, differentiation, and plasticity. This thesis explored two specific areas of translational regulation in the context of neurons: upstream open reading frames (uORFs) and translational initiation. It provides evidence that two widely accepted doctrines are principally outdated: 1) mammalian transcripts are monocistronic, with few exceptions, and 2) robust translation initiation must begin at an AUG start codon *in vivo*. Instead, I find that uORFs are—consistent with work in yeast and mice—very common in the human neuroblastoma transcriptome and that their initiation commonly occurs at a codon other than AUG. In the second half of my thesis I define how one specific type of uORF event, RAN translation at CGG repeats in *FMR1*, acts to regulate translation of the Fragile X protein in neurons—a finding with significant clinical implications. Below, I discuss the repercussions of this new understanding of post-transcriptional gene regulation

uORFs translated in neuronal differentiation may be a significant source of post-transcriptional regulation

Translation of uORFs in transcripts such as *ATF4* are known to impact the level at which the CDS is synthesized, and are generally inhibitory. How applicable this model is beyond the

handful of well characterized transcripts is an open question. My work reveals that among thousands of identified uORFs, only a relatively small subset are actively inhibitory to the synthesis of the main coding sequence (CDS) in which they arise; this regulatory group includes both uORFs fully contained in the 5' leader (cORFs) and uORFs that overlap the CDS (oORFs) and uORFs initiated at AUG codons (such as ATF4) and those initiating at non-AUG near cognate codons. These principles, illustrated globally in Chapter 2 of my thesis, both support findings from previous studies utilizing ribosome profiling and extend them in important ways.

The dichotomous effect of AUG and non-AUG uORFs

Ribosome profiling utilizes RNase protection of ribosome associated mRNA fragments to generate a cDNA library for deep sequencing. Evidence from the first reports using this technique showed enrichment of ribosome footprints in 5' leaders (Ingolia, Lareau et al. 2011, Brar, Yassour et al. 2012). The power of this technique was shown in studies looking at conditional fluctuations in protein synthesis(Ingolia, Lareau et al. 2011, Brar, Yassour et al. 2012). Based on this previous work, I hypothesized that during neuronal differentiation translation of a set of mRNA transcripts would be regulated by translation events in their 5' leaders. By tracking their 3-nucleotide periodicity as a mark of active translation (Ingolia, Ghaemmaghami et al. 2009, Chun, Rodriguez et al. 2016), I identified ~3500 uORFs in human SH-SY5Y cells, the majority of which utilize a non-AUG initiation site. The translation of these uORFs cluster based on both SPECtre score and translational efficiency (TE) across conditions, suggesting that they are regulated events. However, on a global scale their relative translation is not inversely predictive of directional CDS changes (Figure 2.5). Instead, uORF TE and CDS TE positively correlate in this large set of transcripts.

It is tempting to attribute this positive correlation to the inclusion of near-cognate initiated uORFs in our dataset. These could potentially create more noise in the analysis if they are not "real". However, a few points suggest this is not the case. First, I experimentally validated a number of uORF events based identified by RP, including multiple non-AUG initiation events, using reporter constructs. Second, this positive correlation has been reported in ribosome profiling analyses that restricted their analysis to AUG-initiated uORFs, as well (Chew, Pauli et al. 2016, Johnstone, Bazzini et al. 2016). Instead, this effect is most likely explained by the fact that both AUG and near-cognate initiated uORFs are suboptimal initiation events relative to modes of regulation that act upstream such as increased pre-initiation complex and mRNA interaction. Thus, CDS translation on a transcript with a uORF is going to change in the same direction. For this reason, we limited our analysis to uORFs above a stricter translational threshold.

In doing so, we detect that AUG initiated uORFs are generally associated with repressed CDS translation, while near-cognate initiated uORFs show a dichotomous effect based on their relative sequence conservation. When investigating non-AUG uORFs from the highest quartile of sequence conservation, we see that these are generally associated with a mildly translationally repressed CDS. However, transcripts bearing lowly conserved non-AUG uORFs show higher CDS translation. The reason for this is not clear, although I suspect that this is due to the level of noise introduced to our analysis when including near-cognate initiated uORFs. This finding suggests that highly translated transcripts support a level of uORF translation, likely due to the increased presence of initiating ribosomes. Taken together, this suggests that uORFs may impart a regulatory role dependent on sequence conservation. Interestingly, we do see a higher GC content in the conserved non-AUG uORFs relative to the lowly conserved non-AUG uORFs.

Translation through these structured regions may be inefficient, thus impairing further scanning of pre-initiation complexes, and preventing them from reaching the main CDS. Additionally, the GC content within the uORF may be an indicator of the secondary structure present across the entire 5' leader, which could be the source of repression.

Conditional CDS fluctuations are tied to oppositional uORF translation

The most interesting findings came when the dataset was constrained to uORFs that were detected above a strict translational threshold in both conditions. This filtering was key to the results that followed because it allowed us to look specifically at translation changes in uORFs with no or low input from alternative levels of regulation, like transcriptional changes, alternative transcript leaders, or alternative splicing. The most compelling evidence that the regulatory role of uORFs is largely imparted through scaled translation of these 5' elements came when we investigated the conditional translation of uORFs. To illustrate this point, we investigated transcripts bearing cORFs or oORFs in the constrained dataset that show the greatest decrease in CDS TE in either Non-Diff and RA-Diff condition (Figure 2.6). Analysis of the absolute TE changes of the cORFs and oORFs revealed significant directionality, with the majority of cORFs and oORFs showing an increased level of translation in the condition where the CDS TE is lowest. These data illustrate that conditional increases in uORF translation accompany a decrease in CDS translation, in a subset of highly expressed uORFs. This suggests an oppositional relationship between these two regions that may be important for uORFs to limit the synthesis of the main CDS product.

Future avenues of interest: uORFs in neuronal differentiation

One surprising finding from my work is the potential role of uORFs in regulating expression of ribosomal transcripts. Recent studies suggest that ribosomal heterogeneity is more widespread than previously appreciated(Xue and Barna 2012). I am intrigued by the idea that ribosomal heterogeneity might contribute to translational changes in neuronal differentiation. In this sense, RPL38 and RPS25 are two very interesting ribosomal proteins. Both are thought to exist in sub-stoichiometric amounts in ribosomes, and may be integral to the concept of heterologous ribosomes (Shi, Fujii et al. 2017). Though these two examples do not show an inverse relationship between uORF and CDS translation, they do bare oORFs, which are generally repressive and may still exert differentiation-specific inhibition of downstream translation. To investigate this further, I would specifically like to mutate these oORFs by CRISPR/Cas9 genome editing, and investigate translational changes in SH-SY5Y cells undergoing differentiation similar to what was done for *PTCH1* (Fujii, Shi et al. 2017). I would also look at the same differentiation markers, and see if the overall process is somehow affected.

Another avenue of research to follow-up this work would be to investigate alternative transcript leaders. The state dependent SPECtre clusters we observe (Figure 2.5) may largely be due to changes at the transcriptional level, which would significantly alter the landscape of uORFs translated. Lastly, because of continued concerns over the validity of RP as a tool for detecting rare translation events, I would utilize mass spectrometry with enrichment for small peptides to detect uORF-derived peptides in the cell. Recent work by other groups in comparable systems suggests that some of these uORF do produce at least transiently stable small proteins which can be detected by mass spectrometry (Bazzini, Johnstone et al. 2014, Na, Barbhuiya et al. 2018). Using the technique TAILS, terminal amine isotopic labeling of substrates, to enrich for the N-termini of peptides has yielded the highest number of 5’ leader-derived, novel peptides

(Na, Barbhuiya et al. 2018). Two-thirds of these aligned with ribosome profiling data, and over 90% of these initiated at a near-cognate codon. After identifying stable peptides by this technique, I would then tag the peptides we detect with a fluorescent marker to investigate potential cellular roles. To discern the role of two proteins generated by two distinct cistrons on a single transcript would be a very novel occurrence for a mammalian transcript.

CGG RAN translation acts as a uORF to modulate basal and activity-dependent FMRP synthesis

In Chapter 3, I investigated translation within the 5' leader of *FMR1*. *FMR1* is interesting because its 5' leader supports translation that is influenced by the presence of a repetitive CGG trinucleotide element which forms a stable hairpin structure. This form of translation is termed Repeat Associated Non-AUG translation, or RAN translation. Expansions of this repeat leads to neurodegeneration in Fragile-X Associated Tremor/Ataxia Syndrome. RAN translation of expanded CGG repeats is linked to toxicity in animal models of FXTAS, and the main protein synthesized from this region, FMRpolyG, forms aggregates in patients (Todd, Oh et al. 2013, Krans, Kearse et al. 2016, Sellier, Buijsen et al. 2017). Until now, RAN translation at CGG repeats and elsewhere has been investigated solely as a pathogenic process(Todd, Oh et al. 2013, Sellier, Buijsen et al. 2017). However, I was intrigued by the fact that FMRpolyG was highly translated at normal repeat lengths in our *in vitro* and cell assays. Upon this, I wondered how RAN translation in the 5' leader of *FMR1* affected translation of FMRP, the main protein product of this gene.

The interplay between the RAN translation reading frames and FMRP

The finding that both the 0- and +1 frame RAN events work together to impede FMRP synthesis *in vitro* and in neurons (Figure 3.2) was the first piece of evidence that CGG RAN translation can work as a uORF. This was exciting for two reasons: 1) this occurred at normal repeats, suggesting that these two reading frames may work to regulate FMRP in unaffected cells, and 2) RAN translation may have a role in decreasing the translational efficiency of FMRP in individuals with expanded repeats. In terms of the latter point, we do see more de-repression of FMRP levels when we mutate the two reading frames in the expanded repeat reporters than when it is done in reporters with a normal repeat size (Figure 3.2, 3.3). This is important because it suggests that there is an element to RAN translation acting as a uORF that may be dysfunctional in the setting of an expanded repeat. This may contribute to the overall decrease in FMRP seen in FXTAS and mosaic Fragile X patients.

RAN translation and the CGG repeat maintain the balance of FMRP translation, which is key to further scaling

The next intriguing finding pointing to a functional role for RAN translation is the data showing that removal of the RAN initiation sites prevents the DHPG-induced increase in FMRP reporter translation with normal and expanded repeats. Additionally, FMRP reporters without a CGG repeat showed the same lack of increase. These findings suggest that these two elements—RAN translation and the repeat—are required for the proper translational response of FMRP. This is interesting because it illustrates the nuances of translational regulation conferred by the 5' leader of *FMR1*. My interpretation of this data is that both elements likely limit scanning of the preinitiation complex to reach the AUG of FMRP, and in their absence translation at the main AUG is at full capacity preventing any further upscaling by DHPG-induced signaling. This is

also likely the case for the GAA reporter. This is an unstructured repeat, and likely does not impede preinitiation complex scanning like the CGG repeat. The finding that we see a selective increase in our wild-type FMRP reporters and not the RAN reporters adds to this model; it shows that RAN and FMRP translation do not function in concert in response to DHPG. Instead, RAN translation acts as a roadblock that can be overcome to allow for more downstream initiation.

This raises two questions that would inform us about how the *FMR1* transcript interacts with the translation machinery, and may provide a better understanding about how translation is regulated in response to external stimuli. First, does CGG RAN translation halt completely in response to DHPG? Or does it continue at a normal rate? Either way, translation initiation on the *FMR1* transcript is increased in response to the stimulus, so the efficiency of RAN translation as a proportion of translation on the whole transcript is decreased. Second, how is RAN translation selectively bypassed in the wake of this upscaling? One potential mechanism by which this can occur is by reactivation of stalled, initiating ribosomes. This has been suggested to occur on transcripts selectively translated in response to DHPG (Graber, Hebert-Seropian et al. 2013). It is possible that these ribosomes stall because of RAN translation and the presence of the CGG repeat, and in their absence stalling cannot occur neither can the increase in FMRP. Alternatively, RAN translation could be selectively bypassed through changes in start codon stringency. Alterations in synaptic eIF1 and eIF5 levels can potentially drive the balance of start site selection away from suboptimal start sites, and promote initiation at the AUG of FMRP. However, there is no direct evidence to support this in prior research or in our own.

An alternative mechanism for activity dependent RAN bypass is similar to what occurs in *ATF4* and *GCN4* in response to stress. In fact, some groups have suggested that DHPG induces eIF2α phosphorylation (Di Prisco, Huang et al. 2014, Trinh, Ma et al. 2014). If ternary

complexes are limiting, this may lead to the bypassing of the 0- and +1 frame start sites by scanning ribosomes, leading to more downstream initiation at the AUG of FMRP.

This proposed mechanism can also bring the +2 RAN frame into play. The +2 frame terminates 16 nucleotides from the AUG of FMRP, leaving a possibility for re-initiation. No start site has been located for the +2 frame, leading us to conclude that initiation occurs within the repeat. The best way to study the +2 frame is through manipulation of its termination site. The +2 frame does not appear to function normally to impede FMRP synthesis since when the termination site for the +2 frame is moved after the AUG of FMRP to create an overlapping uORF, FMRP translation is unaffected (Figure A.1). However, when the 0/+1 AAA mutations are introduced to an FMRP reporter with that same +2 termination site mutation, there is a slight decrease in the FMRP reporter (Figure A.1). This shows that only under conditions when these two frames are bypassed, the +2 frame affects FMRP levels. Even more, when the 0/+1-AAA mutations were introduced to the +2 frame RAN reporters, there was an increase in nanoluciferase activity (Figure 3.2) suggesting that there is indeed interplay between these events.

In summary, the final proposed mechanism may occur as follows: If eIF2α is phosphorylated during DHPG treatment, this can lead to bypassing of the start sites for the 0- and +1 frames due to limitation of available ternary complexes. Initiation may then occur in the +2 RAN reading frame, and reinitiation may then occur downstream at FMRP. One experiment left to do to clarify this point is to perform a DHPG-treated nanoluciferase assay on neurons transfected with two reporters: 1) the +2 stop site mutant, FMRP reporter and 2) the 0/+1-AAA/+2 stop site mutant, FMRP reporter. If this last proposed mechanism is correct, we should still see a DHPG-induced increase in the first condition but not in the latter condition.

Future avenues of interest: RAN translation and FMRP regulation in response to stress

It would be interesting to further investigate the role for prolonged eIF2α phosphorylation in regulating FMRP synthesis. FMRP is a key component of stress granules which are ribonucleoprotein aggregates that form following cellular stress and contain mRNA associated with ribosomes stalled during initiation (Protter and Parker 2016). Importantly, +1 and +2 CGG RAN translation of reporters with large repeats is increased in response to several hours of cell stress and eIF2α phosphorylation(Green, Glineburg et al. 2017, Cheng, Wang et al. 2018). However, this has not been investigated in reporters with normal repeat sizes. If this is seen at normal repeat sizes this may be a way by which FMRP translation is regulated in response to extended periods of eIF2α phosphorylation. At peak eIF2α phosphorylation, CGG RAN is increased and there is a global decrease in translation. This may be different for what is proposed to occur during DHPG stimulation, because eIF2α phosphorylation levels are prolonged and at their peak. Additionally, there may be an element of compartmentalization to the fluctuations that occur in eIF2α phosphorylation levels specific to neurons. Since it is unclear as to how an increase in the translation of both the +1 and +2 frames might impact FMRP, it would be interesting to look into further.

There are two potential outcomes I can see potentially happening in response to stress: 1) a decrease in FMRP and 2) maintenance of consistent FMRP levels during translational arrest. A decrease in FMRP would mirror the state of global translation in the cell. However, this is different than other uORF-bearing transcripts which usually show protection from translational arrest(Hinnebusch, Ivanov et al. 2016). *FMR1* could be different as its uORFs (RAN translation) are synthesized at a greater rate in response to stress. A decrease in FMRP may be an adaptive

response to cell stress as FMRP constitutes stress granules and is a negative regulator of translation. Decreasing FMRP levels might be a way to help cells to end the translational arrest alongside the increase in eIF2α dephosphorylation. Mutation of the RAN start sites in our FMRP reporters and investigation into whether we see a change in FMRP levels in response to stress is necessary to determine this relationship. Alternatively, the increase in +2 frame RAN translation may counter the increase in the +1 frame RAN translation, leading to a neutral effect on FMRP, and maintenance of FMRP levels. Based on my data, when translation is increased in the +2 frame, reinitiation is necessary to maintain FMRP levels. Thus, this might be a means by which FMRP escapes the translation arrest associated with cell stress, thus leading to an outcome much like other uORF-bearing transcripts. This may be necessary for the prolonged formation of stress granules. Either way, it would be interesting to investigate the role of RAN translation in regulating FMRP translation under various cell states known to impact RAN translation.

Using ASOs to impede RAN translation and increase in FMRP translation

The mechanistic studies above allowed for the rational generation of ASOs aimed at blocking RAN translation and increasing FMRP synthesis. These ASOs provide direct evidence that RAN translation acts endogenously to impede FMRP translation. Because they lack their native gene context, reporter constructs and assays always run the risk of providing misleading findings in relation to the biology they are attempting to characterize. In the absence of making mutations to the 5' leader, sterically blocking initiation at the first and second +1 start site increases endogenous FMRP. I confirmed that the ASO was acting in the expected manner by ASO treating cells transfected with the +1 RAN reporter, and performing nanoluciferase assays.

This expands on the first reports where ASOs were used with this strategy because it was never shown that the ASOs were, in fact, precluding translation of the uORFs that were targeted.

Using the ASOs in HEK cells confirmed that blocking RAN can increase FMRP derived from transcripts with normal repeat sizes. This was taken further into iPSCs and neurons derived from a patient with an expanded repeat where the effect of the ASO on FMRP was observed as well. In addition to the proof of principle that steric hindrance of RAN can increase FMRP endogenously, these data provide exciting evidence that ASOs targeting RAN start sites might be a viable therapeutic target in patients with a deficit in FMRP.

Future avenues of interest: The endogenous effect of RAN translation on FMR1

One experiment to complete moving forward is to mutate the RAN start sites endogenously by CRISPR/Cas9 genome editing. If done in iPSCs or human neural progenitors, we could further investigate the endogenous role of RAN translation in a neuronal setting. Additionally, the *+1 RAN ASO* could be applied to iPSC-derived neurons prior to DHPG stimulation. This is complicated by the developmental state of these neurons, and a potential lack of proper surface receptor expression. However, if an increase in FMRP in response to DHPG is observed in these cells, we would expect this effect to be blocked in the presence of the ASO. The would lend more confidence to the endogenous role of +1 RAN translation in modulating activity-dependent FMRP synthesis.

Additional research should apply the *+1 RAN ASO* in a mouse model of FXTAS to further elucidate the benefits of precluding RAN translation. There are caveats to this as the mouse *FMR1* transcript does not have complete sequence alignment with the human region targeted by the ASO. One way to get around this would be to virally infect reporters for RAN

translation into the brains of mice. It would be interesting to investigate the ASO's potential to knock-down this reporter. Subsequently, investigation of cell health and aggregate formation would provide insight into the cellular effects of the ASO. Behavioral assays of mice treated with the ASO would be necessary for investigating the overall benefit of an ASO treatment to knock-down RAN as a therapeutic approach to treat FXTAS.

Conclusions

My thesis both describes the universal role of uORFs in a neuronal setting and focuses on how RAN translation can act as a uORF to the key neuronal transcript, *FMR1*. The results I have gathered through my work on *FMR1* have pointed to a nuanced but important role for RAN translation in modulating the basal and activity-dependent synthesis of FMRP. While the CGG repeat and RAN translation may seem like an exceptional case, ribosome profiling of SH-SY5Y cells suggests that it is not. The CGG repeat is a significant structural barrier, but structured 5' leaders are common (Pickering and Willis 2005). Also, while upstream AUGs are largely selected against, there are thousands of transcripts that not only have one AUG but many, and near-cognate codons are even more prevalent (Chew, Pauli et al. 2016, Johnstone, Bazzini et al. 2016). Upon this, careful inspection and experimentation of the 5' leaders of individual transcripts may yield an abundance of insight about how translation is regulated. Whereas the continued use of ribosome profiling will reveal overarching themes and attributes of 5' leader regulation and its implications across diverse cell states.

Citations

Bazzini, A. A., T. G. Johnstone, R. Christiano, S. D. Mackowiak, B. Obermayer, E. S. Fleming, C. E. Vejnar, M. T. Lee, N. Rajewsky, T. C. Walther and A. J. Giraldez (2014). "Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation." Embo j 33(9): 981-993.

Brar, G. A., M. Yassour, N. Friedman, A. Regev, N. T. Ingolia and J. S. Weissman (2012). "High-resolution view of the yeast meiotic program revealed by ribosome profiling." Science 335(6068): 552-557.

Cheng, W., S. Wang, A. A. Mestre and C. Fu (2018). "C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation." 9(1): 51.

Chew, G. L., A. Pauli and A. F. Schier (2016). "Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish." Nat Commun 7: 11663.

Chun, S. Y., C. M. Rodriguez, P. K. Todd and R. E. Mills (2016). "SPECtre: a spectral coherence--based classifier of actively translated transcripts from ribosome profiling sequence data." BMC Bioinformatics 17(1): 482.

Di Prisco, G. V., W. Huang and S. A. Buffington (2014). "Translational control of mGluR-dependent long-term depression and object-place learning by eIF2alpha." 17(8): 1073-1082.

Fujii, K., Z. Shi, O. Zhulyn and N. Denans (2017). "Pervasive translational regulation of the cell signalling circuitry underlies mammalian development." 8: 14443.

Graber, T. E., S. Hebert-Seropian, A. Khoutorsky, A. David, J. W. Yewdell, J. C. Lacaille and W. S. Sossin (2013). "Reactivation of stalled polyribosomes in synaptic plasticity." Proc Natl Acad Sci U S A 110(40): 16205-16210.

Green, K. M., M. R. Glineburg, M. G. Kearse, B. N. Flores, A. E. Linsalata, S. J. Fedak, A. C. Goldstrohm, S. J. Barmada and P. K. Todd (2017). "RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response." Nat Commun 8(1): 2005.

Hinnebusch, A. G., I. P. Ivanov and N. Sonenberg (2016). "Translational control by 5'-untranslated regions of eukaryotic mRNAs." Science 352(6292): 1413-1416.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman and J. S. Weissman (2009). "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling." Science 324(5924): 218-223.

Ingolia, N. T., L. F. Lareau and J. S. Weissman (2011). "Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes." Cell 147(4): 789-802.

Johnstone, T. G., A. A. Bazzini and A. J. Giraldez (2016). "Upstream ORFs are prevalent translational repressors in vertebrates." Embo j 35(7): 706-723.

Krans, A., M. G. Kearse and P. K. Todd (2016). "Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome." Ann Neurol 80(6): 871-881.

Na, C. H., M. A. Barbhuiya, M. S. Kim, S. Verbruggen, S. M. Eacker, O. Pletnikova, J. C. Troncoso, M. K. Halushka, G. Menschaert, C. M. Overall and A. Pandey (2018). "Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini." Genome Res 28(1): 25-36.

Pickering, B. M. and A. E. Willis (2005). "The implications of structured 5' untranslated regions on translation and disease." Semin Cell Dev Biol 16(1): 39-47.

Protter, D. S. and R. Parker (2016). "Principles and Properties of Stress Granules." Trends Cell Biol 26(9): 668-679.

Sellier, C., R. A. M. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent, M. F. Champy, T. Sorg, G. Pavlovic, M. Wattenhofer-Donze, M. C. Birling, M. Oulad-Abdelghani, P. Eberling, F. Ruffenach, M. Joint, M. Anheim, V. Martinez-Cerdeno, F. Tassone, R. Willemsen, R. K. Hukema, S. Viville, C. Martinat, P. K. Todd and N. Charlet-Berguerand (2017). "Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome." Neuron 93(2): 331-347.
Shi, Z., K. Fujii, K. M. Kovary, N. R. Genuth, H. L. Rost, M. N. Teruel and M. Barna (2017). "Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide." Mol Cell 67(1): 71-83.e77.

Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur, K. Elenitoba-Johnson, J. P. Vonsattel, E. D. Louis, M. A. Sutton, J. P. Taylor, R. E. Mills, N. Charlet-Berguerand and H. L. Paulson (2013). "CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome." Neuron 78(3): 440-455.

Trinh, M. A., T. Ma, H. Kaphzan, A. Bhattacharya, M. D. Antion, D. R. Cavener, C. A. Hoeffer and E. Klann (2014). "The eIF2alpha kinase PERK limits the expression of hippocampal metabotropic glutamate receptor-dependent long-term depression." Learn Mem 21(5): 298-304.

Xue, S. and M. Barna (2012). "Specialized ribosomes: a new frontier in gene regulation and organismal biology." Nat Rev Mol Cell Biol 13(6): 355-369.

APPENDIX A

A role for reinitiation after RAN translation in regulating FMRP synthesis

In this set of experiments, I wondered if the +2 frame RAN event affected FMRP levels. It is believed that initiation in the +2 frame occurs in the repeat, thus we cannot prevent initiation in this frame. Thus, I mutated the termination site in our FMRP reporters, which is about 16 nucleotides upstream of the AUG for FMRP. This mutation (+2 Mut) removes the termination site so that the next +2 frame termination site is downstream of the start codon for FMRP, turning +2 RAN into an overlapping uORF. In doing so, we see no effect on FMRP reporter levels with a wild-type 5' leader (**Figure A.1**). However, in the combined 0/+1 AAA, +2 Mut reporter with a large repeat, we do observe a decrease in reporter levels relative to the 0/+1 AAA reporter of the same repeat size. In Figure 3.2E, we do observe an increase in +2 RAN translation when the 0/+1 AAA mutations are made. These data suggest that the +2 RAN event may not be inhibitory under wild-type conditions, until it is upregulated. Under conditions where +2 RAN is increased, reinitiation after translation may be necessary to maintain FMRP translation levels.

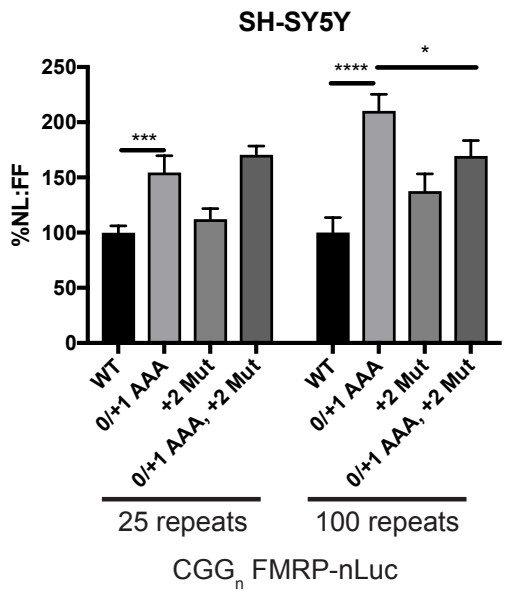


Figure A.1:

Reinitiation after +2 CGG RAN occurs in the absence of 0- and +1-frame RAN translation. SH-SY5Y cells were transfected for 24 hours with nLuc reporter. nLuc activity was normalized to FFluc, which was co-transfected. 0/+1 AAA reporters recapitulate the increase relative to WT, as previously mention. However, the +2 Mut alone does not affect FMRP reporter levels. However, in the 90 repeat reporters, we do observe a modest decrease in nLuc levels in the 0/+1 AAA, +2 Mut reporter relative to the 0/+1 AAA mutant reporter alone. n=3, one-way ANOVA with multiple comparisons within repeat size, *$p<0.05$, ***$p<0.001$, ****$p<0.0001$

APPENDIX B

Repeat length-dependent protein turnover of FMRpolyG

In this set of experiments, I sought to discern differences in turnover of FMRpolyG with varying lengths in the homopolymeric glycine stretch. To do this, I utilized Dendra2. This is a photoconvertable protein that emits green fluorescence in its unconverted state, and upon UV application the Dendra2 molecule emits red fluorescence. This allows for the tracking of proteins present in the cell at the time of photoconversion. I made a series of reporters for the +1 RAN products, FMRpolyG. These reporters had the full 5' UTR sequence for FMRpolyG (with 54 or 100 repeats) with a Dendra2 tag. The Dendra2 tag does not have the first 3 amino acids to preclude initiation and synthesis of Dendra2 without the FMRpolyG sequence upstream. To track protein degradation, photoconverted Dendra2 was imaged serially in the red channel. Green fluorescence was tracked when imaging aggregates to detect the rate at which new protein was added. Aggregate turnover was monitored in hippocampal neurons, which were transfected at DIV10 and imaged 24 hours later (**Figure B.1A**). Green and red protein was monitored directly after photobleach (0 minutes), and 20 and 40 minutes after photobleach. Interestingly, we see an increase in accumulation of green protein in the 54 repeat reporter relative to the 100 repeat reporter (**Figure B.1B**). This was unexpected. We hypothesized that larger repeat proteins would be more rapidly drawn to aggregate. We noticed that there are two populations of aggregates in the 54 repeat group: those that are dropping in green fluorescence (like the 100 repeat group) and those that are increasing in green fluorescence. We did not detect this in the 100 repeat group.

However, we believe this is due to expression differences over the 24 hours of transfection. The 100 repeat aggregates might form earlier within that span due to higher expression levels early on. Thus, we have a higher likelihood of imaging an aggregate that has been present for several hours. The 54 repeat proteins may aggregate later on within the 24 hour span, thus we are imaging aggregates as they are forming. In fact, when imaging the red channel in these forming aggregates, we also see accumulation. This is likely due to the high level of attraction for all FMRpolyG present in the cell at the time, thus even diffuse, spuriously photoconverted protein is added to the aggregate. For this reason, we only included aggregates from the 54 repeat condition that were not "forming"; these were not increasing in red fluorescence, we call this select group the "formed" group. Thus, we compared the red fluorescence of only the formed aggregates with 54 and 100 repeats. In doing so, we see no difference between the repeat sizes (**Figure B.1C**). Taken together, these data show that there is no difference between repeat sizes in turnover of the protein present in aggregates, but repeat size does appear to influence the formation dynamics of aggregates.

Next, we hypothesized that FMRpolyG turnover could be better assessed through imaging whole cells. I switched to SH-SY5Y cells so that I could capture multiple cells in one frame. Cells were transfected for 24 hours. Cells were imaged every 10 minutes post photobleaching for 60 minutes. This strategy revealed a difference in turnover of FMRpolyG with 54 and 100 repeats, with the larger repeat showing more stability as we hypothesized. The increase in stability of the larger repeat may lend some insight into what occurs in patients. In FXTAS patients we detect FMRpolyG aggregates, these are not present in control tissue. An element of this may be due to the increase in stability of FMRpolyG with a larger repeat.

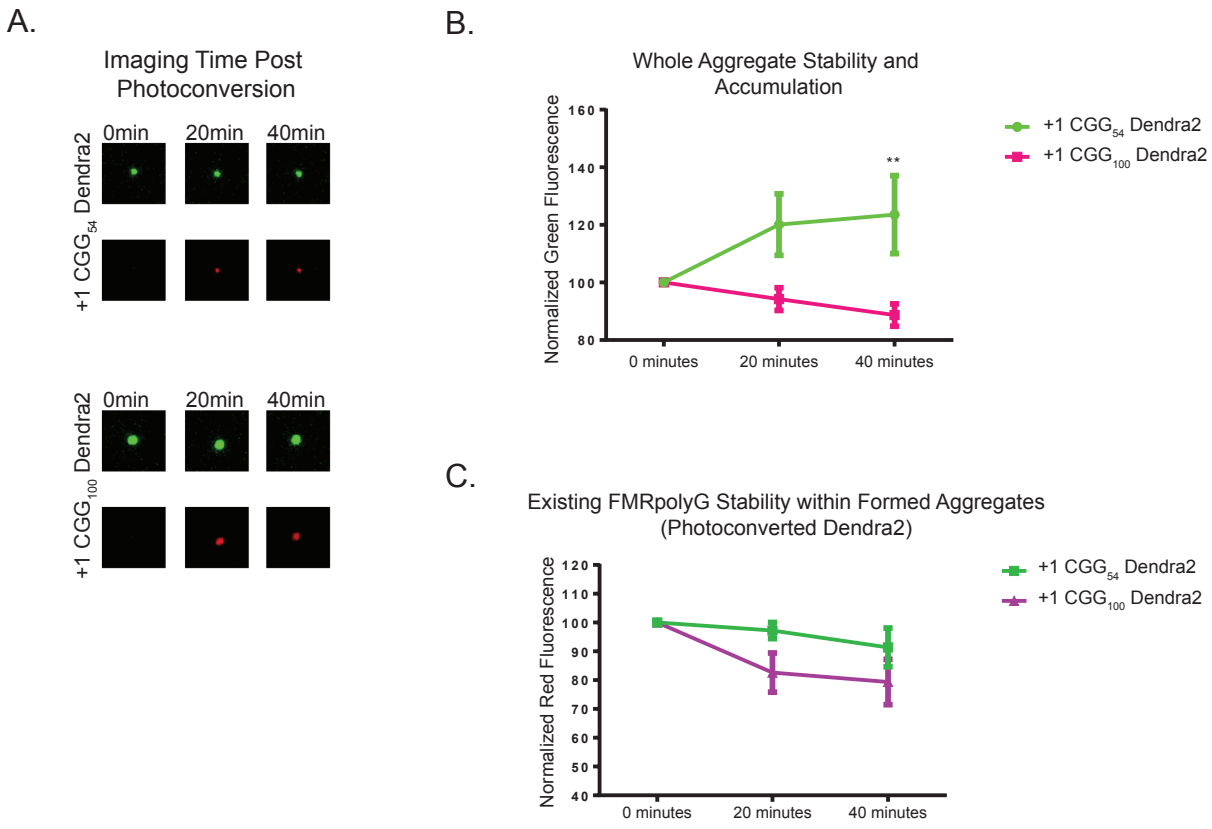


Figure B.1:

FMRpolyG aggregate formation is repeat-dependent. A) Individual aggregates were imaged at 0, 20, and 40 minutes after photobleaching. B) Aggregated green protein was tracked revealing, revealing a relative increase in active aggregation in the +1 CGG_{54} Dendra2 relative to +1 CGG_{100} Dendra2. C) Imaging already formed aggregates for photoconverted protein (red), revealed no difference in turnover based on repeat size. A) n=11, 54 repeats, n=8, 100 repeats. B) n=4, 54 repeats, n=8, 100 repeats. Two-way ANOVA with multiple comparisons, **p>0.01.

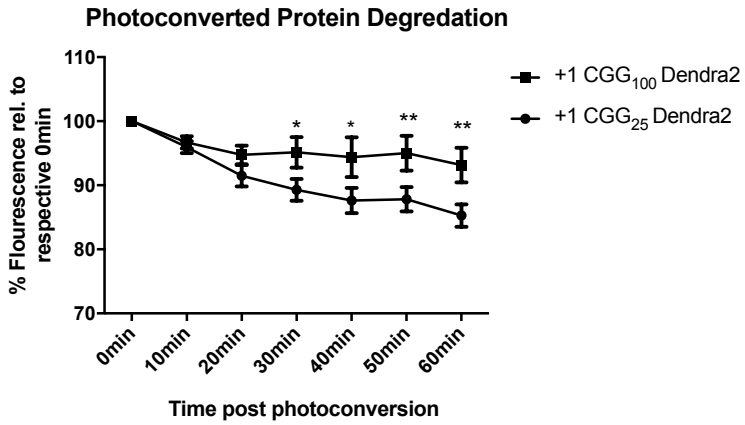


Figure B.2:

FMRpolyG turnover is repeat-dependent. Whole-cell imaging of SH-SY5Y cells transfected with +1 CGG_n Dendra2 plasmids. The larger repeat reporter is significantly more stable at 30-60minutes of imaging. n=7 cells for each condition, two-way ANOVA with multiple comparisons, *p>0.05, **p>0.01.

APPENDIX C

ASO targeting CGG RAN increases FMRP in Fragile X hESC-derived neurons

In this set of experiments, I hypothesized that the +1 RAN ASO would increase FMRP levels in human embryonic stem cell (hESC)-derived neurons. A line of control (unaffected) cells were treated, along with a line of cells with a full repeat. This line does make *FMR1* RNA at early passage numbers, so we reasoned that in these early passage cells, there should be *FMR1* mRNA for the ASO to bind. Neurons were treated at 5 weeks with ASO for 24-hours, then maintained for 6 days. Western blotting was performed (**Figure C.1**).

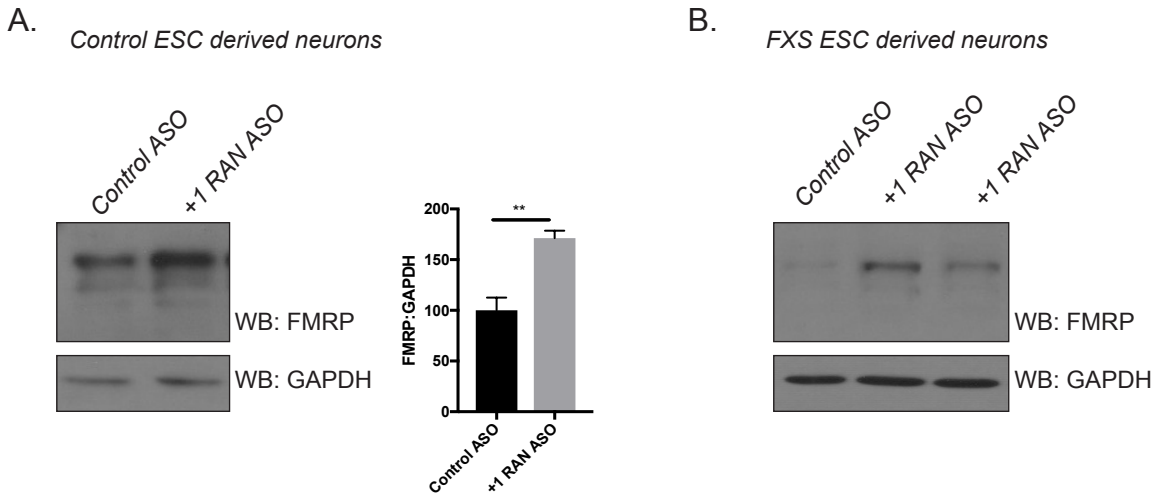


Figure C.1:

***+1 RAN ASO* increases FMRP in control and FXS hESC-derived neurons.** A) hESC-derived neurons with a normal, unmethylated repeat (~30) were treated as described above. A 71.3% increase was detected by western blot in the *+1 RAN ASO* treated cells relative to the Control ASO treated cells. n=3, Student's t-test, **p>0.01. B) The western blot is shown for the FXS neurons treated with both ASO conditions, leading ~3-fold increase. n=1, *Control ASO*, n=2, *+1 RAN ASO*.