Published in final edited form as:

Brain Res. 2016 September 15; 1647: 30–42. doi:10.1016/j.brainres.2016.04.003.

RAN translation—what makes it run?

Katelyn M. Green^{1,2,*}, Alexander E. Linsalata^{1,2,*}, and Peter K. Todd^{1,2,3,#}

¹Department of Neurology, University of Michigan Medical School, Ann Arbor, MI

²Program in Cellular and Molecular Biology, University of Michigan Medical School, Ann Arbor, MI

³Veterans Affairs Medical Center, Ann Arbor, MI

Abstract

Nucleotide-repeat expansions underlie a heterogeneous group of neurodegenerative and neuromuscular disorders for which there are currently no effective therapies. Recently, it was discovered that such repetitive RNA motifs can support translation initiation in the absence of an AUG start codon across a wide variety of sequence contexts, and that the products of these atypical translation initiation events contribute to neuronal toxicity. This review examines what we currently know and don't know about repeat associated non-AUG (RAN) translation in the context of established canonical and non-canonical mechanisms of translation initiation. We highlight recent findings related to RAN translation in three repeat expansion disorders: CGG repeats in fragile X-associated tremor ataxia syndrome (FXTAS), GGGGCC repeats in *C9orf72* associated amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) and CAG repeats in Huntington disease. These studies suggest that mechanistic differences may exist for RAN translation dependent on repeat type, repeat reading frame, and the surrounding sequence context, but that for at least some repeats, RAN translation retains a dependence on some of the canonical translational initiation machinery.

Introduction

Nucleotide-repeat expansions underlie a heterogeneous group of primarily neurological diseases that in aggregate impact a large number of patients (Mason *et al.*, 2014). Repeats can cause problems through a variety of mechanisms delineated over the past 25 years. Expansion of trinucleotide repeats within protein-coding open reading frames (ORFs) cause a gain-of-function toxicity downstream of the production of polyglutamine or (less frequently) polyalanine proteins (Orr and Zoghbi, 2007). This toxicity results from both alterations in the native functions of the protein in which the repeat resides as well as toxicity independent of protein context, related to perturbations in neuronal proteostasis. Repeat expansions located outside of known protein-coding ORFs can elicit changes in the

[#]Correspondence to: Peter K. Todd, M.D., Ph.D., petertod@umich.edu, 4005 BSRB, 109 Zina Pitcher Place, Ann Arbor, MI 48109-2200.

^{*}These authors contributed equally to this work

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

expression of the gene in which they reside, leading to reduced or enhanced expression at the transcript and protein level (He and Todd, 2011). Such non-coding repeats can also elicit toxicity as RNA by binding to and sequestering specific RNA-binding proteins via presentation of a repetitive motif (Mohan *et al.*, 2014).

The discovery of repeat-associated non-AUG (RAN)-initiated translation blurs the lines that define which repeats elicit toxicity via protein gain-of-function and which act through RNA repeat-elicited gain-of-function (Cleary and Ranum, 2014; Kears and Todd, 2014). This non-canonical translational initiation process enables elongation through a repeat strand in the absence of an AUG initiation codon and in multiple reading frames, producing multiple homopolymeric or dipeptide repeat-containing proteins. Originally described in association with CAG-repeat expansions causative for spinocerebellar ataxia type 8 (SCA8), this process also occurs in association with expansions of CAG, CUG, GGGGCC, GGCCCC, and CGG repeats (Zu *et al.*, 2011; Ash *et al.*, 2013; Gendron *et al.*, 2013; Mori *et al.*, 2013a; Mori *et al.*, 2013b; Todd *et al.*, 2013; Zu *et al.*, 2013; Bañez-Coronel *et al.*, 2015). Repeats can drive RAN translation in a surprising variety of RNA contexts, including within 5' untranslated regions (UTRs), protein-coding ORFs, or introns and "non-coding" RNAs. The identification of this novel translational initiation event has led to a flurry of activity within the research community, with a significant body of work now demonstrating 1) the presence of RAN-translated peptides across a wide spectrum of neurodegenerative disorders and 2) an association between these short repeat-containing proteins and neuronal toxicity. Despite this interest, the mechanism or mechanisms by which RAN translation occurs remains largely unknown. The focus of this review will be what we know thus far about RAN translation initiation, with particular attention paid to how RAN translation compares to canonical translation and other forms of initiation described over the past 40 years. We hope that revisiting these foundational experiments will shed light on this new disease-relevant process.

Translation initiation

Translation initiation is the step-wise assembly of elongation-competent 80S ribosomes at start codons of mRNA. It is a highly complex process, entailing the concerted activity of at least nine eukaryotic initiation factors (eIFs; Jackson *et al.*, 2010). A comprehensive account of the roles of each eIF and each stage of initiation is beyond the scope of this review, but several steps are worth highlighting. In most cases, initiation begins with the recognition of the 5' methyl-7-guanosine (m⁷G) cap on mRNA by the eIF4F complex (Fig. 1, Step 1; Sonenberg *et al.*, 1978; Sonenberg *et al.*, 1979; Grifo *et al.*, 1983). The eIF4F complex is composed of eIF4E (the direct cap-binding subunit), eIF4G (a scaffolding subunit), eIF4A (a DEAD-box RNA helicase), with eIF4B and eIF4H serving as additional helicase stimulatory factors. eIF4G recognizes the poly-adenosine-binding protein (PABP; Imataka *et al.*, 1998; Kessler and Sachs, 1998), which in turn binds to the 3' polyA tail on mRNAs. This is thought to result in circularization of the mRNA and greater initiation efficiency (Borman *et al.*, 2000; Kahvejian *et al.*, 2005).

The eIF4F complex, still bound to the m⁷G cap, is joined by the 43S pre-initiation complex (PIC), composed of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, and the ternary

complex [in turn composed of methionine-conjugated tRNA (tRNA^{Met}), and eIF2-GTP; Fig. 1, Step 2]. This joining of the 43S PIC to the eIF4F complex is mediated by an eIF4G-eIF3 interaction (LeFebvre *et al.*, 2006). Successful translation of most eukaryotic mRNAs is thought to require the RNA helicase activity of eIF4A in order to resolve RNA-RNA secondary structures adjacent to the m⁷G cap and prepare a “landing pad” for the 43S PIC (Pestova and Kolupaeva, 2002). In the ribosomal scanning model of translation initiation (Kozak, 1978), the 43S PIC and components of the eIF4F complex scan through the 5' UTR in the 5' to 3' direction (Fig. 1, Step 3). This stage is also known to require eIF4A in order to resolve weaker internal secondary structures, though additional helicases assist in melting stronger structures (Jackson, 1991; Chuang *et al.*, 1997; Svitkin *et al.*, 2001; Pisareva *et al.*, 2008; Zhang *et al.*, 2015b). The 43S PIC scans until encountering an AUG codon in a good Kozak context, (A/G)NNAUGG (Fig. 1, Step 4; Kozak, 1984, 1986). At this point, base-pairing between the AUG codon and CAU anti-codon loop on tRNA^{Met} results in the ejection of eIF1, a factor which, along with eIF1A, increases the stringency of AUG start-codon selection (Yoon and Donahue, 1992; Pestova and Kolupaeva, 2002; Unbehauen *et al.*, 2004; Maag *et al.*, 2005; Passmore *et al.*, 2007). eIF2 hydrolyzes its bound GTP with the assistance of eIF5, the associated GTPase-activating protein (GAP; Fig. 1, Step 5). At this point, the 40S ribosome is committed to its selection of start codon, and forms a tighter interaction with the substrate mRNA, collectively known as the 48S PIC. In the final stages of initiation, the 40S subunit is joined by the 60S ribosomal subunit (Fig. 1, Step 6), the majority of remaining eIFs are ejected, eIF5B hydrolyzes its bound GTP (Fig. 1, Step 7) and translation elongation begins with formation of the first peptide bond (Dever and Green, 2012 for review).

The role of secondary structure in translational initiation

RNA secondary and tertiary structures contribute significantly to the dynamics and regulation of translation initiation. Structures in the 5' UTR impact translation initiation both positively and negatively, depending on the structure's location. When placed upstream of an AUG start codon, highly structured regions are known to inhibit initiation, either by blocking the eIF4E-m⁷G interaction when located adjacent to the cap, or by impeding 5'-to-3' translocation of the 43S PIC when located internally in the 5' UTR (Kozak, 1980, 1986, 1988, 1994). In contrast, secondary structures downstream of start codons facilitate initiation at imperfect start codons: those with poor Kozak context and even non-AUG codons (Kozak, 1990, 1994). Initiation at non-AUG codons occurs at reduced efficiency relative to AUG codons in *in vitro* translation systems (Peabody, 1987, 1989), but the presence of secondary structures downstream markedly increases initiation efficiency (Kozak, 1989, 1990, 1994).

Recent advances in ribosomal foot printing methodologies suggest that these *in vitro* findings may reflect common but heretofore unrecognized set of initiation events *in vivo*. Ribosome profiling combines the traditional aspects of an RNase-protection assay with next generation sequencing to identify the positions of initiating and elongating ribosomes on mRNAs on a transcriptome-wide level (Ingolia *et al.*, 2009). This technique has found evidence for thousands of unpredicted translation initiation events, many of which occur at non-AUG codons (Ingolia *et al.*, 2011; Lee *et al.*, 2012; Gao *et al.*, 2015). This is especially

true for upstream open reading frames (uORFs), which are short ORFs upstream of canonical, annotated ORFs in the same mRNA transcript. Many of these uORFs appear to play regulatory roles in translation that are dependent on metabolic conditions and cell-cycle stage (Brar *et al.*, 2012). These findings are now supported by a variety of studies utilizing mass spectroscopy to confirm the presence of these uORF-coded peptides, many of which possess functional roles (Magny *et al.*, 2013; Menschaert *et al.*, 2013; Slavoff *et al.*, 2013; Vanderperre *et al.*, 2013; Chanut-Delalande *et al.*, 2014; Pauli *et al.*, 2014; Anderson *et al.*, 2015). Thus, initiation at non-AUG codons occurs *in vivo*, may be regulated in part by *trans*-acting eIFs as well as mRNA-specific *cis* factors, and appears to play important regulatory roles in global protein translation.

While the majority of eukaryotic mRNAs are likely translated via the canonical mechanism described above, multiple atypical mechanisms exist, and like canonical initiation, atypical modes of initiation are modulated by mRNA secondary structure. Multiple viral and cellular RNAs are translated via an internal ribosome entry site (IRES)-mediated pathway (Fig. 2B; (Lozano and Martínez-Salas, 2015). IRESes are complex RNA structures that directly recruit ribosomal subunits and eIFs to internal sites within RNA transcripts. These are highly heterogeneous structurally and functionally, but the common, central feature is that translation initiation bypasses the eIF4E-m⁷G interaction. Therefore, IRES-based translation is said to be m⁷G cap-independent, allowing for initiation within bicistronic or circular RNA elements.

In addition, IRESes are heterogeneous in which eIFs are and aren't recruited or required, with some forms requiring most eIFs while others require only the 40S ribosomal subunit and an alanine-conjugated tRNA (Fig. 2C; Wilson *et al.*, 2000a; Wilson *et al.*, 2000b; Jan *et al.*, 2001). The latter, employed by the cricket paralysis virus (CrPV), intriguingly utilizes CCU for an initiation codon and codes for a protein with an N-terminal alanine. Though viral IRESes are the most thoroughly characterized, multiple eukaryotic mRNAs are also thought to contain IRESes, including c-Myc, p53, Bcl-2, and others (Komar and Hatzoglou, 2005). Translation of these proteins is maintained under various stress conditions in which canonical translation is inhibited, implying a unique mechanism that escapes this general inhibition.

In a second example of non-canonical modes of initiation, *histone 4* mRNA appears to be translated through a ribosomal tethering mechanism (Fig. 2D; Chappell *et al.*, 2006a; Chappell *et al.*, 2006b). The 5' UTR of *histone 4* is shorter than most 5' UTRs (the mouse homolog is 9 nucleotides, at the short extreme). It is efficiently translated, however, despite translation initiation generally requiring a 5' UTR of at least 20 nucleotides (Kozak, 1987, 1991). Translation of histone 4 mRNA begins with the recruitment of the eIF4F complex to a structural element within the ORF. eIF4F then binds to the m⁷G cap, which is buried within a nearby structural element. The 43S PIC is subsequently recruited to eIF4F, and is then transferred directly to the AUG start codon (Martin *et al.*, 2011). Thus, translation of *histone 4* is cap-dependent, but deviates from canonical translation in that the ribosome is initially recruited internally. And, as with IRES-mediated translation, translation of *histone 4* is mediated through interactions between eIFs and mRNA structural elements.

RAN translation: an unexpected finding at nucleotide repeats

In each of the above examples, mRNA secondary structure encodes “instructions” for how a given transcript is to be translated. Laura Ranum and colleagues’ discovery of RAN translation introduced a novel mode of translation to this mechanistic multitude (Zu *et al.*, 2011). Expansions of protein-coding CAG repeats in the gene Ataxin 8 (*ATXN8*) lead to the neurodegenerative disorder SCA8. Unexpectedly, mutation of the only AUG codon upstream of expanded CAG repeats did not abrogate protein translation (Zu *et al.*, 2011). Nevertheless, translation initiated in multiple reading frames, generating homopolymeric proteins with glutamine, serine, or alanine repeats (depending on the reading frame). RAN initiation on *ATXN8* transcripts depended on the stability of secondary structures formed from the expanded CAG repeats, as decreasing the number of CAG repeats or their GC content abrogated RAN translation (Zu *et al.*, 2011). RAN translation products from all three reading frames accumulated in cells transfected with expanded CAG reporters, occasionally even within the same transfected cell. Antisense *ATXN8* transcripts bearing expanded CUG repeats also supported RAN translation (Zu *et al.*, 2011). Antibodies generated against the predicted polyalanine product of the *ATXN8* sense transcript recognized a protein in the cerebellums of SCA8 human patients and mouse models. A similar approach provided *in vivo* evidence of a polyglutamine RAN product from *DMPK* antisense transcripts bearing expanded CAG repeats, associated with myotonic dystrophy type 1 (DM1; Zu *et al.*, 2011).

Since this initial observation, several groups have demonstrated that RAN translation occurs at a wide variety of different repeat expansions (Ash *et al.*, 2013; Gendron *et al.*, 2013; Mori *et al.*, 2013a; Mori *et al.*, 2013b; Todd *et al.*, 2013; Zu *et al.*, 2013; Bañez-Coronel *et al.*, 2015). This review focuses on RAN translation in three distinct repeat-expansion disorders that occur in different sequence contexts: CGG repeats in the 5’ UTR of the fragile X mental retardation 1 (*FMR1*) gene, as occurs in fragile X-associated tremor/ataxia syndrome (FXTAS), intronic GGGGCC repeats in *C9orf72*-associated amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD), and CAG repeats in the protein-coding sequence of Huntingtin (*HTT*) in Huntington disease. Each example offers unique insights into how RAN translation occurs, but also presents unique challenges in our effort to understand this process mechanistically. Here we review the current literature pertinent to each repeat expansion, in search of hints as to the mechanism of RAN translation.

RAN translation of CGG repeats in FXTAS

FXTAS is a late-onset neurodegenerative disorder caused by the expansion of CGG repeats in the 5’ UTR of *FMR1*. In unaffected individuals, repeats number less than 45. Individuals with FXTAS carry between 55 and 200 repeats, known as the “premutation” range (Pembrey *et al.*, 1985; Dorn *et al.*, 1994; Hagerman *et al.*, 2001). Premutation repeat expansions result in enhanced transcription of *FMR1* mRNA bearing these repeats (Tassone *et al.*, 2000; Tassone *et al.*, 2007). In contrast, expansions to greater than 200 repeats trigger transcriptional silencing of the *FMR1* locus, leading to loss of *FMR1* mRNA and the Fragile X protein, FMRP. Transcriptional silencing manifests as Fragile X syndrome, a clinically distinct neurodevelopmental disorder characterized by intellectual disability and autistic features (Pieretti *et al.*, 1991; Verkerk *et al.*, 1991). Approximately 40% of male premutation

carriers develop FXTAS (approximately 1:3000 of the total population), with increased penetrance at older ages and larger repeat sizes (Rousseau *et al.*, 1995; Dombrowski *et al.*, 2002; Jacquemont *et al.*, 2004). FXTAS is characterized clinically by action tremors, ataxia, parkinsonism, and cognitive decline, and pathologically by both neuronal and non-neuronal ubiquitinated inclusions throughout the cerebral cortex, brainstem, and cerebellum (Leehey, 2009; Leehey and Hagerman, 2012). Premutation carrier women are also at increased risk of premature ovarian insufficiency (FXPOI; Cronister *et al.*, 1991; Allingham-Hawkins *et al.*, 1999).

Our lab demonstrated that the CGG-expanded *FMR1* 5' UTR supports RAN translation initiation (Todd *et al.*, 2013). Initiation within the 5' UTR occurs in at least two reading frames in the absence of an AUG start codon: the GGC (+1) frame yields a polyglycine product (FMRpolyG), and the GCG (+2) frame yields a polyalanine product (FMRpolyA). FMRpolyG accumulates in ubiquitinated inclusions in patient tissue and cellular and animal disease models, is necessary to elicit toxicity in *Drosophila* models of disease, and induces proteasome perturbations in *Drosophila* and HeLa cells (Todd *et al.*, 2013; Buijsen *et al.*, 2014; Oh *et al.*, 2015). In an inducible mouse model of FXTAS that expresses the *FMR1* 5'UTR with 90 CGG repeats, turning off transgene expression reverses the formation of neuronal FMRpolyG-positive inclusions and repeat-elicited behavioral deficits (Hukema *et al.*, 2015). Finally, FMRpolyG-positive inclusions have been found in ovarian stromal cells in FXTAS mouse models and a FXPOI patient, suggesting FMRpolyG expression is linked to other Fragile X-related clinical phenotypes (Buijsen *et al.*, 2016).

In an effort to investigate RAN translation of CGG-expanded *FMR1* mechanistically, our lab developed several transfectable reporters for expression of FMRpolyG and FMRpolyA. First, we observed production of an FMRpolyG-green fluorescent protein (GFP) fusion reporter construct bearing 30, 50, or 88 CGG repeats, suggesting that RAN translation can occur in the absence of pathological expansions (Todd *et al.*, 2013). In contrast, FMRpolyA was expressed from reporter constructs bearing 88 repeats but not 30 repeats (Todd *et al.*, 2013). Second, insertion of a stop codon immediately upstream of the CGG repeats precluded expression of FMRpolyG reporters, indicating that RAN translation of FMRpolyG initiates upstream of the repeats (Todd *et al.*, 2013). Further mutational analysis revealed that initiation in this frame can occur at multiple upstream near-AUG codons in the human 5'UTR (Todd *et al.*, 2013). In contrast, insertion of a stop codon did not preclude expression of FMRpolyA (Todd *et al.*, 2013). This suggests that RAN translation can initiate in the GCG frame within the repeats. These results raise the intriguing possibility that RAN translation of the same sequence can differ mechanistically in different reading frames.

One important question is whether RAN translation of CGG-expanded *FMR1* is cap-dependent or utilizes an IRES-like cap-independent mechanism (Fig. 2A–C). There is some evidence to suggest that the *FMR1* 5' UTR possesses IRES activity. Insertion of the *FMR1* 5' UTR between two ORFs in a plasmid-based bicistronic reporter doesn't eliminate translation of the second ORF (Chiang *et al.*, 2001; Dobson *et al.*, 2008). Although some expression of the second ORF may be due to the presence of a cryptic promoter element within the *FMR1* 5' UTR, this same finding was observed when *in vitro* transcribed bicistronic RNA was transfected into cells (Dobson *et al.*, 2008). Further experiments

indicated that translation from monocistronic reporters with the 5' UTR of *FMR1* were less cap-dependent than reporters bearing the 5' UTR of β -globin, as would be predicted if an IRES was utilized (Dobson *et al.*, 2008). This putative IRES activity was partially dependent on the CGG repeat and required both the approximately 100 nucleotides upstream of the repeats in *FMR1* 5' UTR, as well as the region containing the repeat itself. However, these studies were done before a conceptualization of RAN translation was in place, and it is unclear whether these initiation events occurred at the AUG of the reporter ORF or within the *FMR1* 5' UTR itself. Also, some of these studies were performed with a very short repeat (9 CGGs). Follow-up work suggested that a significant fraction of FMRP is translated through a cap-dependent process, highlighting the significance of the eIF4E-m⁷G interaction both in cells and *in vitro* translation systems (Chen *et al.*, 2003; Ludwig *et al.*, 2011).

More recently, we demonstrated that RAN translation of CGG-expanded *FMR1* reporters is cap-dependent in multiple repeat reading frames (Kearse et al, 2016). Using a modified nanoLuciferase that selectively reports on RAN translation from the *FMR1* 5'UTR, we found that *in vitro* transcribed, m⁷G-capped RAN-translation reporter RNAs are efficiently translated in both transfected HeLa cells as well as rabbit Reticulocyte lysate *in vitro* translation reactions. However, when m⁷G is substituted with an A-cap which is not recognized by eIF4E, expression of these RAN reporters decreased dramatically. Furthermore, addition of excess free m⁷G cap, which binds to and sequesters eIF4E, also blocked CGG RAN translation of both FMRpolyG and FMRpolyA reporters. Neither of these manipulations affected expression of an IRES reporter, strongly suggesting that RAN translation of CGG repeats in the *FMR1* 5' UTR is a cap-dependent process akin to the first stage of canonical translation (Kearse et al, 2016).

In the next stage of canonical initiation, the 43S PIC scans through the 5' UTR, performing a base-by-base inspection for an AUG codon (Fig. 1, Step 3). If the scanning model were to hold for RAN translation of CGG-expanded *FMR1*, then the 43S PIC would need to scan through the CGG repeats to reach the AUG start codon for FMRP. *In silico* modeling predicts and *in vitro* analysis suggests that consecutive CGG repeats form a stable hairpin structure (Sobczak *et al.*, 2010; Kiliszek *et al.*, 2011), presenting a significant impedance to scanning 43S PICs. Consistent with ribosomal scanning, increasing the length of CGG repeats reduces the expression of a downstream AUG initiated reporter (Chen *et al.*, 2003; Ludwig *et al.*, 2011). In addition, Ludwig *et al.* (2011) observed initiation at a near-AUG codon in an artificial hairpin inserted in the 5' UTR in place of the CGG repeats, further supporting a scanning mechanism.

More directly, to determine whether the scanning model holds true for RAN translation of CGG-expanded *FMR1*, we treated cells with hippurastinol, an inhibitor of eIF4A, an RNA helicase that is required for 43S PIC scanning. Addition of hippurastinol effectively blocked translation from both an AUG driven reporter and CGG RAN translation reporters in both the polyglycine and polyalanine reading frames, but had no effect on IRES-mediated translation (Kearse et al, 2016). This would suggest that the initial stages of RAN translation at CGG repeats resembles canonical initiation and requires ribosomal scanning.

Marilyn Kozak demonstrated that downstream secondary structures enhance initiation at upstream AUG and non-AUG codons (Kozak, 1989, 1990, 1994). The increase in non-AUG codon usage is maximal when a hairpin falls 14 nucleotides downstream of the AUG codon. Based on the known size of ribosomes, this orientation would place the start codon within the P site of the 40S ribosome, opposite the anti-codon loop of tRNA^{Met} (Kozak, 1990). These findings led to the hypothesis that secondary structure causes scanning 43S PICs to stall, increasing initiation at optimally positioned non-AUG codons. RAN translation on CGG-expanded FMR1 mRNA to generate FMRpolyG may utilize a similar mechanism (Todd *et al.*, 2013). If mRNA secondary structures are necessary for CGG RAN translation initiation, then it creates specific testable hypotheses regarding how such structures promote initiation at non-AUG codons. For example, stalling of scanning ribosomes (Kozak, 1989, 1990, 1994) is predicted to lead to both congestion of 43S PICs on mRNAs upstream of the repeat and an increase in the dwell time of the 40S subunit over imperfect codon-anticodon matches. Stalling could also favor the dissociation of key eIFs that help determine AUG start codon fidelity (eIF1 and eIF1A) or favor alternative ribosomal conformations as occurs with IRES-mediated translation (Fernández *et al.*, 2014; Muhs *et al.*, 2015). Both of these events could presumably enhance the rate of enzymatic catalysis and 48S complex formation at upstream non-AUG codons. Consistent with this, as the size of CGG repeats increases in reporter constructs, there is an increase in the relative efficiency of RAN translation in both the polyglycine and polyalanine reading frames in RNA transfected cells (Kearse et al, 2016). This is coupled with a decrease in the need for any specific near-AUG codon upstream of the repeat in the glycine reading frame, suggesting impaired start codon selection fidelity. At large repeat sizes, there is even evidence for initiation within the CGG repeat itself in both the polyalanine and polyglycine reading frames, although these events remain inefficient compared to translation from near-cognate codons located upstream of the repeat (Kearse et al, 2016).

In consideration of the existing literature, our current working model of RAN translation at CGG-expanded FMR1 is as follows (Fig. 3A): the eIF4F complex and 43S PIC bind to the m⁷G cap on FMR1 mRNA. This complex then scans downstream through the 5' UTR until encountering secondary structure formed either by CGG repeats or the surrounding, intrinsic sequence of the 5' UTR. Ribosomal stalling results in aberrant translation initiation at non-AUG codons either upstream of or within the repeat in the +1 and +2 frames, resulting in the production of FMRpolyG and FMRpolyA. Important remaining questions include what specific protein factors or neuronal factors are important for RAN translation at CGG repeats, whether initiation also occurs in the CGG (+0, polyarginine) frame or on antisense FMR1 transcripts, and how transferable these mechanisms are to RAN translation at other repeat expansions in different sequence contexts.

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

The *C9orf72* GGGGCC/GGCCCC hexanucleotide repeat expansions was identified by two groups in 2011 as the most common known cause of ALS and FTD (DeJesus-Hernandez *et al.*, 2011; Renton *et al.*, 2011). ALS is the most frequently occurring form of motor neuron disease, affecting approximately 2–4/100,000 individuals (Johnston *et al.*, 2006), and is characterized by progressive paralysis typically leading to death within two to three years

after onset. FTD is the second most common form of presenile dementia and affects approximately 20/100,000 individuals between the ages of 45 and 65 (Onyike and Diehl-Schmid, 2013; Luukkainen *et al.*, 2015). FTD presents heterogeneously and is divided into three clinical syndromes; behavioral variant, semantic dementia, and progressive nonfluent aphasia (Neary *et al.*, 1998). The *C9orf72* hexanucleotide repeat expansions is most frequently associated with the behavioral variant (DeJesus-Hernandez *et al.*, 2011; Renton *et al.*, 2011; Gijselinck *et al.*, 2012), characterized by changes in personality and conduct. Although ALS and FTD each manifest with a unique set of symptoms and pathology, they are believed to constitute two ends of a single disease spectrum. Approximately 50% of ALS patients develop FTD-like cognitive and behavioral impairment (Lomen-Hoerth *et al.*, 2003; Ringholz *et al.*, 2005); while up to 50% of FTD patients develop motor dysfunction (Lomen-Hoerth *et al.*, 2002). Additionally, TDP-43-positive inclusions are present within the neurons and glia of a majority of ALS patients, as well as in the most common variant of FTD (FTLD-TDP; Neumann *et al.*, 2006).

In C9 ALS/FTD, the GGGGCC repeat, located within the first intron of transcript isoforms 1 and 3, and the promoter region of isoform 2, is expanded from 2–25 repeats in healthy individuals, to upwards of more than a thousand repeats in C9 ALS/FTD patients (DeJesus-Hernandez *et al.*, 2011; Renton *et al.*, 2011; Gijselinck *et al.*, 2012). Both the sense and antisense strands of *C9orf72* are transcribed in mutation carriers, resulting in the production of GGGGCC and GGCCCC-repeat containing RNAs (Gendron *et al.*, 2013; Mori *et al.*, 2013b; Zu *et al.*, 2013). These expanded repeat sequences are both predicted to form highly stable RNA secondary structures, with the sense RNA repeat generating a G-quadruplex and hairpin *in vitro* (Fratta *et al.*, 2012; Reddy *et al.*, 2013; Haeusler *et al.*, 2014; Su *et al.*, 2014) and the antisense RNA repeat recently shown to assume an A-form-like double helix *in vitro* (Dodd *et al.*, 2016).

Dipeptide Repeat Protein pathology in C9 ALS/FTD patients

In addition to TDP-43-positive inclusions within both neurons and glia (Neumann *et al.*, 2006), neuronal TDP-43-negative inclusions that co-stain for ubiquitin and ubiquitin-binding proteins are uniquely found throughout the CNS of C9-associated ALS and FTD patients (Al-Sarraj *et al.*, 2011; Boxer *et al.*, 2011). Immunohistochemical (IHC) analysis by multiple laboratories indicates that RAN-translation-derived proteins constitute these TDP-43-negative inclusions (Ash *et al.*, 2013; Gendron *et al.*, 2013; Mori *et al.*, 2013a; Mori *et al.*, 2013b; Zu *et al.*, 2013). A total of six different dipeptide repeat proteins (DRPs) are generated from the GGGGCC and GGCCCC transcripts (Fig. 3B & 3D). Specifically, glycine-alanine (GA) and glycine-arginine (GR) DRPs are generated from the sense strand, proline-alanine (PA) and proline-arginine (PR) arise from the antisense strand, and two glycine-proline (GP) containing proteins arise from RAN translation of both strands (Ash *et al.*, 2013; Gendron *et al.*, 2013; Mori *et al.*, 2013a; Mori *et al.*, 2013b; Zu *et al.*, 2013).

DRPs form both neuronal cytoplasmic and intranuclear inclusions (NCIs and NIIs) throughout the CNS. However, the distribution of DRPs throughout the brain is highly variable, with the highest burden occurring in the hippocampus, cerebellum, neocortex, and thalamus (Ash *et al.*, 2013; Zhang *et al.*, 2014; Schludi *et al.*, 2015). Additionally, although

limited by potential differences in antibody affinities, IHC studies suggest that the different DRPs are not present in equal abundance. In several brain regions assessed with multiple different antibodies for each DRP, polyGA appears to be most abundant, followed by polyGP and polyGR, while the DRPs derived exclusively from antisense transcripts (polyPA and polyPR) appear to be least abundant (Mori *et al.*, 2013a; Mori *et al.*, 2013b; Mackenzie *et al.*, 2015).

Despite the different CNS regions that exhibit marked neurodegeneration in ALS and FTD—the motor cortex and spinal cord in ALS, and the frontal and temporal lobes in FTD—quantitative IHC studies show that DRP abundance within the frontal cortex and lower motor neurons is not significantly different between C9 patients with pure ALS or FTD (Mackenzie *et al.*, 2015). Several additional studies have similarly shown that DRP burden is not well-correlated with degeneration (Mackenzie *et al.*, 2013; Davidson *et al.*, 2015; Schludi *et al.*, 2015). This is in contrast to TDP-43-positive inclusions, which are most abundant in the most severely affected brain regions (Mackenzie *et al.*, 2013; Davidson *et al.*, 2015; Mackenzie *et al.*, 2015).

This lack of correlation may suggest that RAN translation is not the driving force in disease pathogenesis. Alternatively, it may be that DRP inclusion formation is neuroprotective while the soluble DRPs oligomers drive toxicity, as has been proposed in several other neurodegenerative proteinopathies (Saudou *et al.*, 1998; Haass and Selkoe, 2007). Alternatively, Edbauer and Haas (2015) propose an “amyloid-like” mechanism of toxicity for the DRPs, in which accumulation of DRPs initiates a cascade of events that leads to TDP-43 mis-localization and aggregation in selectively vulnerable neurons (Edbauer and Haass, 2015). Work is still needed to distinguish between these possibilities, perhaps using the AAV GGGGCC₆₆ mouse model in which neuronal loss and TDP-43 pathology is detectable (Chew *et al.*, 2015).

Toxicity of Dipeptide Repeat Proteins in model systems

Although their distribution throughout the brain raises questions about their exact role in disease, it is clear from studies *in vitro* and *in vivo* that DRP expression in isolation can induce neurodegeneration. From yeast (Jovi *et al.*, 2015), to *Drosophila* (Mizielinska *et al.*, 2014; Wen *et al.*, 2014; Freibaum *et al.*, 2015; Tran *et al.*, 2015; Yang *et al.*, 2015), cultured cells (Zu *et al.*, 2013; Zhang *et al.*, 2014; Tao *et al.*, 2015; Yamakawa *et al.*, 2015), and primary mammalian neurons (May *et al.*, 2014; Wen *et al.*, 2014; Zhang *et al.*, 2014), DRP expression leads to cell death and/or reduced survival. Significantly, in many of these systems, DRP expression is sufficient to trigger toxicity, as demonstrated by the use of alternative codons in place of GGGGCC that allow for DRP production in the absence of the potentially toxic repeat-containing RNA species (May *et al.*, 2014; Mizielinska *et al.*, 2014; Wen *et al.*, 2014; Zhang *et al.*, 2014; Jovi *et al.*, 2015; Tao *et al.*, 2015; Yamakawa *et al.*, 2015; Yang *et al.*, 2015). Furthermore, transgenic flies expressing various length GGGGCC repeats with stop codons in all three reading frames form RNA foci, but only flies containing pure repeats produce polyGR and polyGP proteins and undergo significant cell death (Mizielinska *et al.*, 2014).

Multiple studies suggest that arginine-containing DRPs are the most toxic RAN species. Both polyGR and polyPR form intranuclear aggregates that disrupt nucleoli when overexpressed in model systems (Kwon *et al.*, 2014; Wen *et al.*, 2014; Tao *et al.*, 2015; Yamakawa *et al.*, 2015). However, nucleolar DRPs are not detected in patient brain tissue (Mackenzie *et al.*, 2015; Schludi *et al.*, 2015), and when co-expressed with polyGA, polyGR proteins are recruited into cytoplasmic polyGA inclusions (Yang *et al.*, 2015), suggesting that nucleolar stress may not be a significant driver of toxicity in patients. Alternatively, Wen *et al.* (2014) identified nucleolar polyPR-positive inclusions in spinal cord tissue from a C9 ALS patient, and suggest that the high toxicity of nucleolar polyPR results in increased vulnerability of neurons containing these species (Wen *et al.*, 2014). PolyGR proteins can also mediate toxicity through impairment of the Notch pathway (Yang *et al.*, 2015) and polyPR and polyGR inhibit nucleocytoplasmic transport in flies and yeast (Freibaum *et al.*, 2015; Jovi *et al.*, 2015). Importantly, GGGGCC repeats directly interact with RanGAP, a regulator of nucleocytoplasmic transport (Zhang *et al.*, 2015a), suggesting that the arginine-containing DRPs and the repeat-containing RNA both contribute to this mode of toxicity.

Although comparatively less toxic than polyGR or polyPR when each is expressed in isolation (Mizielinska *et al.*, 2014; Wen *et al.*, 2014; Freibaum *et al.*, 2015; Jovi *et al.*, 2015; Yamakawa *et al.*, 2015; Yang *et al.*, 2015), non-arginine-containing DRP proteins also appear important to neurodegeneration in model systems. Adult flies expressing exclusively GA-100 DRPs within neurons have significantly reduced survival (Mizielinska *et al.*, 2014), and expression of polyGA in primary mammalian neurons causes increased toxicity through impairment of the ubiquitin-proteasome system (Zhang *et al.*, 2014), induction of ER stress (Zhang *et al.*, 2014), and sequestration of Unc119, a trafficking protein with a GAGASA binding motif (May *et al.*, 2014).

Mechanism of RAN translation at GGGGCC/GGCCCC repeats

Despite compelling evidence that RAN translation and the resulting DRPs are involved in disease pathogenesis, little is known about the mechanism by which the expanded GGGGCC and GGCCCC repeats trigger DRP production. When placed in a 5' leader context, RAN translation at GGGGCC/GGCCCC repeats is repeat-length dependent, with more robust DRP production occurring with longer repeats (Mori *et al.*, 2013b; Zu *et al.*, 2013; Su *et al.*, 2014), consistent with observations at CAG and CGG repeats (Zu *et al.*, 2011; Kearse *et al.*, 2016). The repeat-length requirement for initiation also appears to be different for different reading frames and in different sequence contexts. When GGGGCC and GGCCCC repeats are placed downstream of a synthetic sequence, all DRPs are detected by immunocytochemistry (ICC) with as few as 30 or 40 repeats, respectively (Zu *et al.*, 2013). When GGGGCC repeats are instead placed downstream of 113 nucleotides from intron1, a partially native context (Fig. 3B), production of polyGA occurs similarly with as little as 38 repeats (Mori *et al.*, 2013b). However, within this sequence context, polyGP detection required 66 repeats in one report and 145 in another, while polyGR was not detected within cells expressing constructs containing up to 145 repeats (Mori *et al.*, 2013b; Su *et al.*, 2014). Antisense DRPs also showed different length requirements for detection when placed downstream of native sequence; while polyGP and polyPR are detected in cells expressing 66 GGCCCC repeats downstream of 99 native nucleotides, polyPA is not

(Gendron *et al.*, 2013). While these apparent differences in length requirements may reflect artifacts of detection based on antibody avidity or differences in DPR solubility, it could also indicate an inherent discrepancy in RAN translational efficiency across reading frames. For example, different RNA secondary structures might favor initiation in certain frames at shorter repeats, with an increase in promiscuity or frame-shifting at larger repeat sizes becomes more prominent.

Beyond these initial insights, however, are a series of unanswered questions related to the mechanism of RAN translation at GGGGCC and GGCCCC repeats. First, it remains unknown exactly what RNA species actually undergo RAN translation in C9 repeat expansion patients. The GGGGCC repeat expansion is located within the first intron of *C9orf72*. Therefore, in patients, RAN-translated GGGGCC repeats could conceivably derive from a retained intron, a spliced intron in a lariat, or within aberrant disease-specific transcripts generated by transcriptional stalling (Fig. 3B). There is some evidence for generation of such aberrant transcripts, at least *in vitro* (Haeusler *et al.*, 2014). The ratio of exon1a-intron1 (unspliced or abortive) RNA to exon1-exon2 (mature, spliced) RNA (Fig. 3B), however, is not altered in C9 iPSC-derived neurons and patient brain tissue relative to controls, arguing against significant production of truncated transcripts or increased intron retention (Tran *et al.*, 2015). However, a recent study suggests that intron retention occurs with some frequency in both control and C9 patient cells (Niblock M *et al.*, 2016), and this may only have pathological consequences when the expanded repeat is present. Therefore, the lack of increased retention does not rule out the possibility that such transcripts undergo RAN translation in C9 patients.

In *Drosophila*, placement of the repeat into an efficiently spliced intron dramatically reduces both RAN translation and its relative toxicity compared to repeats placed into a 5' leader sequence, suggesting that spliced lariats containing GGGGCC repeats may not be efficiently utilized to produce DRPs (Tran *et al.*, 2015). However, DRP production from the intronic repeat becomes sufficient to elicit toxicity when *Drosophila* are grown at elevated temperatures, indicating that an intronic context is able to support pathological RAN translation under certain conditions (Tran *et al.*, 2015). Whether the limited amount of DRPs observed was produced from a spliced or retained intronic repeat is unclear. However, if an intron lariat is the transcript subtype utilized, then some mechanism must exist for it to bypass normal degradation mechanisms, exit to the cytoplasm, and become engaged with translational machinery.

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

RAN translation in Huntington disease: when the message is in the middle

RAN translational events at both CGG repeats in FXTAS and GGGGCC/GGCCCC repeats in *C9orf72* occur within putatively non-coding transcripts or non-coding regions of coding transcripts. However, RAN translation can also occur efficiently at CAG repeats embedded within annotated open reading frames. This was initially suggested by work on the CAG repeat in SCA8 (Zu et al, 2011), where the AUG codon normally resides just proximal to the repeat itself, and new data demonstrates that RAN translation also occurs in Huntington disease (Bañez-Coronel *et al.*, 2015).

Huntington disease (HD) is the most common known neurodegenerative repeat expansion disorder, affecting 5.8/100,000 people worldwide (Pringsheim *et al.*, 2012). Huntington disease results from a CAG-repeat expansion in the first coding exon of the Huntingtin gene, HTT. Normal sized repeats are typically in the 20's, with a minimum threshold of disease as greater than 35 and 100% penetrance at repeat sizes of 40 or greater (Bean and Bayrak-Toydemir, 2014). The repeat begins 51 nucleotides (17 amino acids) 3' to the AUG initiation codon that demarcates the very large (3144 amino acids) annotated ORF for the full-length huntingtin protein (HTT). The CAG repeat within this annotated ORF codes for a polyglutamine stretch that can be released as a smaller peptide by either alternative splicing or protease cleavage. A large body of evidence suggests that large (usually greater than 60) CAG repeats in isolation or in the context of HTT exon 1 are sufficient to elicit toxicity in mouse and fly models of disease (Mangiarini *et al.*, 1996; Jackson *et al.*, 1998; Schilling *et al.*, 1999). However, evidence also suggests that the native functions of the Huntington protein contributes to phenotypic and molecular findings observed in Huntington Disease patients and some mouse models of HD (see review by Saudou et al, 2016).

Huntington disease shows significant genetic anticipation, with serial repeat expansions over generations leading to an earlier onset of disease and a more severe phenotype, with early dementia, dystonia and parkinsonism in addition to or in place of the cardinal features of chorea and psychosis seen in the later onset form of the disease (Hansotia *et al.*, 1968; Gonzalez-Alegre and Afifi, 2006). While long thought to represent the impact of a larger polyglutamine expansions, the qualitatively different phenotype observed in patients could reflect alternative mechanisms of pathophysiology that only occur at larger repeat sizes rather than a simple additive effect of more polyglutamine (Williams and Paulson, 2008).

Upon this background, Banez-Coronel and colleagues (2015) provided compelling evidence that RAN translation proteins are generated from the CAG repeat in Huntington disease (Banez-Coronel *et al.*, 2015). Using antibodies generated against the predicted C-terminal regions that would be produced from proteins initiating in the AGC (serine) and GCA (alanine) reading frames, they observed significant staining in both striatum and cerebellar tissues in Huntington disease patient brains and in a mouse model of Huntington disease. They also identified RAN-translated proteins arising from an antisense strand generated through the repeat in the CUG orientation, producing polyleucine and polycysteine products. In the sense strand, a cutoff for detection of the serine product by ICC was observed in cells at 35 repeats and for alanine at 45 repeats, both around the threshold for pathogenicity (Banez-Coronel *et al.*, 2015). Intriguingly, the presence of an AUG in the polyglutamine

frame appeared to have little impact on RAN translation in neighboring frames from CMV-driven, plasmid-derived transcripts (Bañez-Coronel *et al.*, 2015). No RAN translation was observed from constructs with CAA repeats or when the codon utilized to make a homopolymeric protein was varied throughout the sequence, suggesting that the repeat sequence itself and/or its secondary structure is important for RAN initiation (Bañez-Coronel *et al.*, 2015).

How might RAN translation from expanded Huntingtin transcripts be mediated mechanistically? For the most part, this question remains unaddressed. However, the findings at the *HTT* locus are internally consistent with previous work on CAG RAN translation in Spinocerebellar Ataxia type 8 (SCA8) (Zu et al, 2011). In SCA8, an AUG initiation codon in ATXN8 resides just one codon above the repeat in the normal sequence context. Mutation of that start codon to AAG failed to prevent translation of the polyglutamine containing protein (Zu *et al.*, 2011). Placement of a stop codon immediately upstream of the repeat or insertion of an AUG and V5 tag above the repeat in the glutamine reading frame failed to block RAN translation of alanine and serine containing products (Zu *et al.*, 2011). As in the HD sequence context, changing the CAG repeat to CAA in the context of the SCA8 leader sequence precluded RAN translation in all reading frames (Zu et al, 2011)

In SCA8, the exact sequence located upstream of the repeat and the cellular context influences RAN translation at CAG repeats in at least some reading frames. For example, placement of a stop codon immediately above the repeat in the polyglutamine frame of the ATXN8 sequence blocked RAN translation of a polyglutamine product below levels detectable by western blot in transfected HEK293 cells (Zu et al, 2011). Zu *et al.* (2011) also varied the 20 nucleotides immediately upstream of a CAG repeat to that of the upstream sequence other repeat expansion disorders. Lentiviral delivery of constructs with the Spinocerebellar Ataxia type 3 (SCA3) sequence placed upstream of a CAG repeat in HEK293 cells or mouse brains exhibited no detectable polyglutamine RAN production. In contrast, inclusion of the upstream sequence from another CAG repeat expansion disorder, Huntington Disease Like 2 (HDL2) supported RAN polyglutamine translation in both of these settings (Zu *et al.*, 2011). In contrast, polyalanine expression was robustly maintained in both sequence contexts (Zu *et al.*, 2011). Attempts to recapitulate CAG RAN translation in an *in vitro* rabbit reticulocyte lysate system revealed a complete loss of polyalanine production regardless of the upstream sequence (Zu *et al.*, 2011). Moreover, *in vitro* translation of polyglutamine or polyserine products from these constructs required near-AUG start codons (Zu *et al.*, 2011), akin to that reported for RAN translation at CGG repeats (Kearse et al, 2016). However, unlike the case for FMRpolyG translation from CGG repeats, this requirement for a near AUG codon was lost in CAG repeat transfected cells. Thus, it appears that for the CAG repeat in the context of a number of disease specific sequence contexts, the upstream sequence, repeat length, and cellular context impact RAN translation efficiency. While there are hints that similar sequence dependent effects may be present in Huntington's Disease, future studies using quantitative methodologies will be needed to address this question empirically.

A related question is whether methionine forms the N-terminus of RAN translation products. In canonical translation, tRNA^{Met} is already bound to the 40S ribosome throughout initiation (Fig. 1; Step 2). Initiation at non-AUG codons predominantly utilizes an N-terminal methionine *in vitro* (Peabody, 1989). If initiation of RAN translation occurs similarly, this would suggest that methionine remains the first amino acid of RAN products. The current published evidence on this question, however, is mixed. For the CAG repeat in ATXN8, data from *in vitro* rabbit reticulocyte lysate assays demonstrate incorporation of methionine at the N-terminus when a near-AUG codon is included above the repeat in the glutamine reading frame (Zu *et al.*, 2011). In contrast, in the alanine reading frame, mass spectroscopy identified a series of peptides with varying lengths of N-terminal polyalanine in transfected HEK293 cells, suggesting initiation within the repeat itself (Zu *et al.*, 2011). Whether the N-terminal amino acid utilized at initiation was alanine (which is utilized in IRES based translation by the CrPV; Fig. 2C) or methionine remains unclear, however, as N-terminal methionines are often proteolytically removed and replaced by an acetyl group (Gigliione *et al.*, 2004).

One further issue is whether frameshifting may also contribute to toxicity and inclusion burden in Huntington's disease (Wojciechowska *et al.*, 2014). Frameshifting from the polyglutamine reading frame into the polyalanine or polyserine reading frames has been suggested for a number of years as a potential mechanism for aspects of neurotoxicity in animal models of SCA3 and Huntington disease (Toulouse *et al.*, 2005, Davies and Rubinsztein, 2006; Stochmanski *et al.*, 2012; Girstmair *et al.*, 2013). It is important to note that most of this work on frameshifting was conducted prior to the discovery and description of RAN translation, which makes its interpretation difficult. However, it does occur. One recent report published after the original description of RAN translation used both mass spectroscopy and 2D electrophoresis to demonstrate that at least some of HTT polyalanine product results from -1 frameshifting (Girstmair *et al.*, 2013). In contrast, no frameshifting products were observed by Bañez-Coronel and colleagues when they used a series of Amino and Carboxy-terminally tagged AUG driven constructs to mimic exon 1 of Huntington disease in transfected cells (*et al.*, 2015). The discrepancy between these two results is potentially important, because the nature of the detection method utilized (Carboxy-terminal antibodies directed against the predicted amino-acid sequence below the repeats) for histochemical staining of sense strand derived RAN translation products in Huntington's Disease cannot distinguish between proteins generated by RAN translation or by frameshifting (Bañez-Coronel *et al.*, 2015). Thus, the relative contribution of these two processes (which are not mutually exclusive) to disease pathogenesis awaits further analysis.

Summary

RAN translation represents a new and provocative mechanism by which protein translation can occur in the setting of nucleotide repeat expansions to produce a novel set of toxic proteins. However, at this early stage in our understanding of this process, there are many more questions than answers. This review has tried to take the limited mechanistic data generated to date and place it into the context of known canonical and non-canonical translation initiation processes. These different modes of translation provide a framework for which questions are of greatest importance. By determining the cap-dependency, the

requirement for linear and continuous 5'-to-3' scanning, and the N-terminal amino acid used during RAN translation, we will be able to take advantage of previous work on processes with similar biology. Such an approach can also narrow down and prioritize which of a myriad of potential *trans* factors should be studied and guide strategies for interventions that might selectively preclude RAN translation.

One important question going forward is whether the mechanisms underlying RAN translation are the same or different across repeat types, reading frames and sequence contexts. Some discrepancies suggest that different mechanisms may be in play. For example, data on RAN translation at CGG repeats thus far is most consistent with a scanning mechanism and use of a near-AUG codon for initiation just 5' to the repeat (Todd *et al.*, 2013, Kearse *et al.*, 2016), but this would seem unlikely as a mechanism to explain initiation within an open reading frame, as occurs in Huntington disease (Bañez-Coronel *et al.*, 2015). In this fashion RAN translation may be analogous to the situation with viral IRES RNA elements, which display a significant variance in both sequence and initiation factor requirements to achieve the same goal of bypassing cap-dependent ribosomal loading. We may need to begin thinking of these processes as *RAN Translations* rather than as a single entity. However, only after careful identification of the key factors required for RAN translation can this delineation really be made across different repeat and disease contexts.

Upon these same lines, it is important to recognize that aspects of what we currently observe in RAN translation may be part of a larger set of mechanisms which allow for translation initiation in the absence of an AUG start codon- a process that may be much more common than previously thought. Data from ribosome profiling datasets suggest significant non- and near-AUG initiated translation throughout the transcriptome (Ingolia *et al.*, 2011). Thus, RAN translation may reflect aberrancy of normal non-canonical initiation processes that produces toxic proteins but which otherwise have valuable functions in other settings. Defining these normal functions and their roles in neuronal biology will be critical if RAN translation is to serve as a therapeutic target.

Lastly, it is worth noting that the novelty of RAN translation may well prove to be its greatest value, both in revealing interesting biology and in providing a particularly good target for therapy development. If this process proves important to neurodegeneration, as current data would support, then identification of factors that are selectively critical for RAN translation but not canonical translation may offer a real treatment opportunity going forward.

Acknowledgments

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

References

Al-Sarraj S, King A, Troakes C, Smith B, Maekawa S, Bodi I, *et al.* p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLN and MND/ALS. *Acta Neuropathol.* 2011; 122(6):691–702. [PubMed: 22101323]

- Allingham-Hawkins DJ, Babul-Hirji R, Chitayat D, Holden JJ, Yang KT, Lee C, et al. Fragile X premutation is a significant risk factor for premature ovarian failure: the International Collaborative POF in Fragile X study--preliminary data. *Am J Med Genet.* 1999; 83(4):322–5. [PubMed: 10208170]
- Anderson DM, Anderson KM, Chang CL, Makarewich CA, Nelson BR, McAnally JR, et al. A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. *Cell.* 2015; 160(4):595–606. [PubMed: 25640239]
- Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, DeJesus-Hernandez M, et al. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron.* 2013; 77(4):639–46. [PubMed: 23415312]
- Bañez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, et al. RAN Translation in Huntington Disease. *Neuron.* 2015; 88(4):667–77. [PubMed: 26590344]
- Bean L, Bayrak-Toydemir P. American College of Medical Genetics and Genomics Standards and Guidelines for Clinical Genetics Laboratories, 2014 edition: technical standards and guidelines for Huntington disease. *Genet Med.* 2014; 16(12):e2. [PubMed: 25356969]
- Borman AM, Michel YM, Kean KM. Biochemical characterisation of cap-poly(A) synergy in rabbit reticulocyte lysates: the eIF4G-PABP interaction increases the functional affinity of eIF4E for the capped mRNA 5'-end. *Nucleic Acids Res.* 2000; 28(21):4068–75. [PubMed: 11058101]
- Boxer AL, Mackenzie IR, Boeve BF, Baker M, Seeley WW, Crook R, et al. Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. *J Neurol Neurosurg Psychiatry.* 2011; 82(2):196–203. [PubMed: 20562461]
- Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science.* 2012; 335(6068):552–7. [PubMed: 22194413]
- Buijsen RA, Sellier C, Severijnen LA, Oulad-Abdelghani M, Verhagen RF, Berman RF, et al. FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun.* 2014; 2:162. [PubMed: 25471011]
- Buijsen RA, Visser JA, Kramer P, Severijnen EA, Gearing M, Charlet-Berguerand N, et al. Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. *Hum Reprod.* 2016; 31(1):158–68. [PubMed: 26537920]
- Chanut-Delalande H, Hashimoto Y, Pelissier-Monier A, Spokony R, Dib A, Kondo T, et al. Pri peptides are mediators of ecdysone for the temporal control of development. *Nat Cell Biol.* 2014; 16(11):1035–44. [PubMed: 25344753]
- Chappell SA, Dresios J, Edelman GM, Mauro VP. Ribosomal shunting mediated by a translational enhancer element that base pairs to 18S rRNA. *Proc Natl Acad Sci U S A.* 2006a; 103(25):9488–93. [PubMed: 16769881]
- Chappell SA, Edelman GM, Mauro VP. Ribosomal tethering and clustering as mechanisms for translation initiation. *Proc Natl Acad Sci U S A.* 2006b; 103(48):18077–82. [PubMed: 17110442]
- Chen LS, Tassone F, Sahota P, Hagerman PJ. The (CGG)_n repeat element within the 5' untranslated region of the FMR1 message provides both positive and negative cis effects on in vivo translation of a downstream reporter. *Hum Mol Genet.* 2003; 12(23):3067–74. [PubMed: 14519687]
- Chew J, Gendron TF, Prudencio M, Sasaguri H, Zhang YJ, Castanedes-Casey M, et al. Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science.* 2015; 348(6239):1151–4. [PubMed: 25977373]
- Chiang PW, Carpenter LE, Hagerman PJ. The 5'-untranslated region of the FMR1 message facilitates translation by internal ribosome entry. *J Biol Chem.* 2001; 276(41):37916–21. [PubMed: 11489899]
- Chuang RY, Weaver PL, Liu Z, Chang TH. Requirement of the DEAD-Box protein ded1p for messenger RNA translation. *Science.* 1997; 275(5305):1468–71. [PubMed: 9045610]
- Cleary JD, Ranum LP. Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders. *Curr Opin Genet Dev.* 2014; 26:6–15. [PubMed: 24852074]

- Cronister A, Schreiner R, Wittenberger M, Amiri K, Harris K, Hagerman RJ. Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. *Am J Med Genet.* 1991; 38(2–3): 269–74. [PubMed: 2018071]
- Davidson Y, Robinson AC, Liu X, Wu D, Troakes C, Rollinson S, et al. Neurodegeneration in Frontotemporal Lobar Degeneration and Motor Neurone Disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. *Neuropathol Appl Neurobiol.* 2015
- Davies JE, Rubinsztein DC. Polyalanine and polyserine frameshift products in Huntington's disease. *J Med Genet.* 2006; 43(11):893–6. [PubMed: 16801344]
- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011; 72(2):245–56. [PubMed: 21944778]
- Dever TE, Green R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol.* 2012; 4(7):a013706. [PubMed: 22751155]
- Dobson T, Kube E, Timmerman S, Krushel LA. Identifying intrinsic and extrinsic determinants that regulate internal initiation of translation mediated by the FMR1 5' leader. *BMC Mol Biol.* 2008; 9:89. [PubMed: 18922172]
- Dodd DW, Tomchick DR, Corey DR, Gagnon KT. Pathogenic C9ORF72 antisense repeat RNA forms a double helix with tandem C:C mismatches. *Biochem.* 2016; 55(9):1283–6. [PubMed: 26878348]
- Dombrowski C, Lévesque S, Morel ML, Rouillard P, Morgan K, Rousseau F. Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. *Hum Mol Genet.* 2002; 11(4):371–8. [PubMed: 11854169]
- Dorn MB, Mazzocco MM, Hagerman RJ. Behavioral and psychiatric disorders in adult male carriers of fragile X. *J Am Acad Child Adolesc Psychiatry.* 1994; 33(2):256–64. [PubMed: 8150798]
- Edbauer D, Haass C. An amyloid-like cascade hypothesis for C9orf72 ALS/FTD. *Curr Opin Neurobiol.* 2015; 36:99–106. [PubMed: 26555807]
- Fernández IS, Bai XC, Murshudov G, Scheres SH, Ramakrishnan V. Initiation of translation by cricket paralysis virus IRES requires its translocation in the ribosome. *Cell.* 2014; 157(4):823–31. [PubMed: 24792965]
- Fratta P, Mizielińska S, Nicoll AJ, Zloh M, Fisher EM, Parkinson G, et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep.* 2012; 2:1016. [PubMed: 23264878]
- Freibaum BD, Lu Y, Lopez-Gonzalez R, Kim NC, Almeida S, Lee KH, et al. GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature.* 2015; 525(7567):129–33. [PubMed: 26308899]
- Gao X, Wan J, Liu B, Ma M, Shen B, Qian SB. Quantitative profiling of initiating ribosomes in vivo. *Nat Methods.* 2015; 12(2):147–53. [PubMed: 25486063]
- Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T, et al. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.* 2013; 126(6):829–44. [PubMed: 24129584]
- Giglione C, Boularot A, Meinel T. Protein N-terminal methionine excision. *Cell Mol Life Sci.* 2004; 61(12):1455–74. [PubMed: 15197470]
- Gijselsinck I, Van Langenhove T, van der Zee J, Slegers K, Philtjens S, Kleinberger G, et al. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol.* 2012; 11(1):54–65. [PubMed: 22154785]
- Girstmair H, Saffert P, Rode S, Czech A, Holland G, Bannert N, et al. Depletion of cognate charged transfer RNA causes translational frameshifting within the expanded CAG stretch in huntingtin. *Cell Rep.* 2013; 3(1):148–59. [PubMed: 23352662]
- Gonzalez-Alegre P, Afifi AK. Clinical characteristics of childhood-onset (juvenile) Huntington disease: report of 12 patients and review of the literature. *J Child Neurol.* 2006; 21(3):223–9. [PubMed: 16901424]

- Grifo JA, Tahara SM, Morgan MA, Shatkin AJ, Merrick WC. New initiation factor activity required for globin mRNA translation. *J Biol Chem*. 1983; 258(9):5804–10. [PubMed: 6853548]
- Haass C, Selkoe DJ. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. *Nat Rev Mol Cell Biol*. 2007; 8(2):101–12. [PubMed: 17245412]
- Haeusler AR, Donnelly CJ, Periz G, Simko EA, Shaw PG, Kim MS, et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*. 2014; 507(7491):195–200. [PubMed: 24598541]
- Hagerman RJ, Leehey M, Heinrichs W, Tassone F, Wilson R, Hills J, et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*. 2001; 57(1):127–30. [PubMed: 11445641]
- Hansotia P, Cleeland CS, Chun RW. Juvenile Huntington's chorea. *Neurology*. 1968; 18(3):217–24. [PubMed: 4231060]
- He F, Todd PK. Epigenetics in nucleotide repeat expansion disorders. *Semin Neurol*. 2011; 31(5):470–83. [PubMed: 22266885]
- Hukema RK, Buijsen RA, Schonewille M, Raske C, Severijnen LA, Nieuwenhuizen-Bakker I, et al. Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. *Hum Mol Genet*. 2015; 24(17):4948–57. [PubMed: 26060190]
- Iwahashi CK, Yasui DH, An HJ, Greco CM, Tassone F, Nannan K, et al. Protein composition of the intranuclear inclusions in FXTAS. *Brain*. 2006; 129(Pt 1):256–71. [PubMed: 16246864]
- Imataka H, Gradi A, Sonenberg N. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J*. 1998; 17(24):7480–9. [PubMed: 9857202]
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324(5924):218–23. [PubMed: 19213877]
- Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011; 147(4):789–802. [PubMed: 22056041]
- Jackson GR, Salecker I, Dong X, Yao X, Arnheim N, Faber PW, et al. Polyglutamine-expanded human huntingtin transgenes induce degeneration of *Drosophila* photoreceptor neurons. *Neuron*. 1998; 21(3):633–42. [PubMed: 9768849]
- Jackson RJ. The ATP requirement for initiation of eukaryotic translation varies according to the mRNA species. *Eur J Biochem*. 1991; 200(2):285–94. [PubMed: 1889398]
- Jackson RJ, Hellen CU, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol*. 2010; 11(2):113–27. [PubMed: 20094052]
- Jacquemont S, Hagerman RJ, Leehey MA, Hall DA, Levine RA, Brunberg JA, et al. Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA*. 2004; 291(4):460–9. [PubMed: 14747503]
- Jan E, Thompson SR, Wilson JE, Pestova TV, Hellen CU, Sarnow P. Initiator Met-tRNA-independent translation mediated by an internal ribosome entry site element in cricket paralysis virus-like insect viruses. *Cold Spring Harb Symp Quant Biol*. 2001; 66:285–92. [PubMed: 12762030]
- Johnston CA, Stanton BR, Turner MR, Gray R, Blunt AH, Butt D, et al. Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. *J Neurol*. 2006; 253(12):1642–3. [PubMed: 17219036]
- Jovine A, Mertens J, Boeynaems S, Bogaert E, Chai N, Yamada SB, et al. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat Neurosci*. 2015; 18(9):1226–9. [PubMed: 26308983]
- Kahvejian A, Svitkin YV, Sukarieh R, M'Boutchou MN, Sonenberg N. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes Dev*. 2005; 19(1):104–13. [PubMed: 15630022]
- Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC, et al. CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Mol Cell*. 2016 In press.

- Kearse MG, Todd PK. Repeat-associated non-AUG translation and its impact in neurodegenerative disease. *Neurotherapeutics*. 2014; 11(4):721–31. [PubMed: 25005000]
- Kessler SH, Sachs AB. RNA recognition motif 2 of yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. *Mol Cell Biol*. 1998; 18(1):51–7. [PubMed: 9418852]
- Kiliszek A, Kierzek R, Krzyzosiak WJ, Rypniewski W. Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res*. 2011; 39(16):7308–15. [PubMed: 21596781]
- Komar AA, Hatzoglou M. Internal ribosome entry sites in cellular mRNAs: mystery of their existence. *J Biol Chem*. 2005; 280(25):23425–8. [PubMed: 15749702]
- Kozak M. How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell*. 1978; 15(4):1109–23. [PubMed: 215319]
- Kozak M. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. *Cell*. 1980; 19(1):79–90. [PubMed: 7357609]
- Kozak M. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature*. 1984; 308(5956):241–6. [PubMed: 6700727]
- Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*. 1986; 83(9):2850–4. [PubMed: 3458245]
- Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res*. 1987; 15(20):8125–48. [PubMed: 3313277]
- Kozak M. Leader length and secondary structure modulate mRNA function under conditions of stress. *Mol Cell Biol*. 1988; 8(7):2737–44. [PubMed: 3405216]
- Kozak M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol*. 1989; 9(11):5073–80. [PubMed: 2601709]
- Kozak M. Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc Natl Acad Sci U S A*. 1990; 87(21):8301–5. [PubMed: 2236042]
- Kozak M. Effects of long 5' leader sequences on initiation by eukaryotic ribosomes in vitro. *Gene Expr*. 1991; 1(2):117–25. [PubMed: 1820209]
- Kozak M. Features in the 5' non-coding sequences of rabbit alpha and beta-globin mRNAs that affect translational efficiency. *J Mol Biol*. 1994; 235(1):95–110. [PubMed: 8289269]
- Kwon I, Xiang S, Kato M, Wu L, Theodoropoulos P, Wang T, et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science*. 2014; 345(6201):1139–45. [PubMed: 25081482]
- Lee S, Liu B, Huang SX, Shen B, Qian SB. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A*. 2012; 109(37):E2424–32. [PubMed: 22927429]
- Leehey MA. Fragile X-associated tremor/ataxia syndrome: clinical phenotype, diagnosis, and treatment. *J Investig Med*. 2009; 57(8):830–6.
- Leehey MA, Hagerman PJ. Fragile X-associated tremor/ataxia syndrome. *Handb Clin Neurol*. 2012; 103:373–86. [PubMed: 21827901]
- LeFebvre AK, Korneeva NL, Trutschl M, Cvek U, Duzan RD, Bradley CA, et al. Translation initiation factor eIF4G-1 binds to eIF3 through the eIF3e subunit. *J Biol Chem*. 2006; 281(32):22917–32. [PubMed: 16766523]
- Lomen-Hoerth C, Anderson T, Miller B. The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. *Neurology*. 2002; 59(7):1077–9. [PubMed: 12370467]
- Lomen-Hoerth C, Murphy J, Langmore S, Kramer JH, Olney RK, Miller B. Are amyotrophic lateral sclerosis patients cognitively normal? *Neurology*. 2003; 60(7):1094–7. [PubMed: 12682312]
- Lozano G, Martínez-Salas E. Structural insights into viral IRES-dependent translation mechanisms. *Curr Opin Virol*. 2015; 12:113–20. [PubMed: 26004307]
- Ludwig AL, Hershey JW, Hagerman PJ. Initiation of translation of the FMR1 mRNA Occurs predominantly through 5'-end-dependent ribosomal scanning. *J Mol Biol*. 2011; 407(1):21–34. [PubMed: 21237174]

- Luukkainen L, Bloigu R, Moilanen V, Remes AM. Epidemiology of Frontotemporal Lobar Degeneration in Northern Finland. *Dement Geriatr Cogn Dis Extra*. 2015; 5(3):435–41. [PubMed: 26675285]
- Maag D, Fekete CA, Gryczynski Z, Lorsch JR. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol Cell*. 2005; 17(2):265–75. [PubMed: 15664195]
- Mackenzie IR, Arzberger T, Kremmer E, Troost D, Lorenzl S, Mori K, et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol*. 2013; 126(6):859–79. [PubMed: 24096617]
- Mackenzie IR, Frick P, Grässer FA, Gendron TF, Petrucelli L, Cashman NR, et al. Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol*. 2015; 130(6):845–61. [PubMed: 26374446]
- Magny EG, Pueyo JI, Pearl FM, Cespedes MA, Niven JE, Bishop SA, et al. Conserved regulation of cardiac calcium uptake by peptides encoded in small open reading frames. *Science*. 2013; 341(6150):1116–20. [PubMed: 23970561]
- Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, et al. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell*. 1996; 87(3):493–506. [PubMed: 8898202]
- Martin F, Barends S, Jaeger S, Schaeffer L, Prongidi-Fix L, Eriani G. Cap-assisted internal initiation of translation of histone H4. *Mol Cell*. 2011; 41(2):197–209. [PubMed: 21255730]
- Mason AR, Ziemann A, Finkbeiner S. Targeting the low-hanging fruit of neurodegeneration. *Neurology*. 2014; 83(16):1470–3. [PubMed: 25313376]
- May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM, et al. C9orf72 FTL/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol*. 2014; 128(4):485–503. [PubMed: 25120191]
- Menschaert G, Van Crielinge W, Notelaers T, Koch A, Crappé J, Gevaert K, et al. Deep proteome coverage based on ribosome profiling aids mass spectrometry-based protein and peptide discovery and provides evidence of alternative translation products and near-cognate translation initiation events. *Mol Cell Proteomics*. 2013; 12(7):1780–90. [PubMed: 23429522]
- Mizielinska S, Grönke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, et al. C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science*. 2014; 345(6201):1192–4. [PubMed: 25103406]
- Mohan A, Goodwin M, Swanson MS. RNA-protein interactions in unstable microsatellite diseases. *Brain Res*. 2014; 1584:3–14. [PubMed: 24709120]
- Mori K, Arzberger T, Grässer FA, Gijssels I, May S, Rentzsch K, et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol*. 2013a; 126(6):881–93. [PubMed: 24132570]
- Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E, et al. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science*. 2013b; 339(6125):1335–8. [PubMed: 23393093]
- Muhs M, Hilal T, Mielke T, Skabkin MA, Sanbonmatsu KY, Pestova TV, et al. Cryo-EM of ribosomal 80S complexes with termination factors reveals the translocated cricket paralysis virus IRES. *Mol Cell*. 2015; 57(3):422–32. [PubMed: 25601755]
- Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology*. 1998; 51(6):1546–54. [PubMed: 9855500]
- Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*. 2006; 314(5796):130–3. [PubMed: 17023659]
- Niblock M, Smith BN, Lee YB, Sardone V, Topp S, Troakes C, et al. Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. *Acta Neuropathol Commun*. 2016; 4(1):18–29. [PubMed: 26916632]

- Oh SY, He F, Krans A, Frazer M, Taylor JP, Paulson HL, et al. RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. *Hum Mol Genet.* 2015; 24(15):4317–26. [PubMed: 25954027]
- Onyike CU, Diehl-Schmid J. The epidemiology of frontotemporal dementia. *Int Rev Psychiatry.* 2013; 25(2):130–7. [PubMed: 23611343]
- Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci.* 2007; 30:575–621. [PubMed: 17417937]
- Passmore LA, Schmeing TM, Maag D, Applefield DJ, Acker MG, Algire MA, et al. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol Cell.* 2007; 26(1):41–50. [PubMed: 17434125]
- Pauli A, Norris ML, Valen E, Chew GL, Gagnon JA, Zimmerman S, et al. Toddler: an embryonic signal that promotes cell movement via Apelin receptors. *Science.* 2014; 343(6172):1248636. [PubMed: 24407481]
- Peabody DS. Translation initiation at an ACG triplet in mammalian cells. *J Biol Chem.* 1987; 262(24):11847–51. [PubMed: 3040720]
- Peabody DS. Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem.* 1989; 264(9):5031–5. [PubMed: 2538469]
- Pembrey ME, Winter RM, Davies KE. A premutation that generates a defect at crossing over explains the inheritance of fragile X mental retardation. *Am J Med Genet.* 1985; 21(4):709–17. [PubMed: 4040705]
- Pestova TV, Kolupaeva VG. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* 2002; 16(22):2906–22. [PubMed: 12435632]
- Pieretti M, Zhang FP, Fu YH, Warren ST, Oostra BA, Caskey CT, et al. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell.* 1991; 66(4):817–22. [PubMed: 1878973]
- Pisareva VP, Pisarev AV, Komar AA, Hellen CU, Pestova TV. Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. *Cell.* 2008; 135(7):1237–50. [PubMed: 19109895]
- Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. *Mov Disord.* 2012; 27(9):1083–91. [PubMed: 22692795]
- Reddy K, Zamiri B, Stanley SY, Macgregor RB, Pearson CE. The disease-associated r(GGGGCC)_n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J Biol Chem.* 2013; 288(14):9860–6. [PubMed: 23423380]
- Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 2011; 72(2):257–68. [PubMed: 21944779]
- Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE. Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology.* 2005; 65(4):586–90. [PubMed: 16116120]
- Rousseau F, Rouillard P, Morel ML, Khandjian EW, Morgan K. Prevalence of carriers of premutation-size alleles of the FMRI gene--and implications for the population genetics of the fragile X syndrome. *Am J Hum Genet.* 1995; 57(5):1006–18. [PubMed: 7485149]
- Saudou F, Finkbeiner S, Devys D, Greenberg ME. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell.* 1998; 95(1):55–66. [PubMed: 9778247]
- Saudou F, Humbert S. The Biology of Huntingtin. *Neuron.* 2016 Mar 2; 89(5):910–26. [PubMed: 26938440]
- Schilling G, Becher MW, Sharp AH, Jinnah HA, Duan K, Kotzuk JA, et al. Intranuclear inclusions and neuritic aggregates in transgenic mice expressing a mutant N-terminal fragment of huntingtin. *Hum Mol Genet.* 1999; 8(3):397–407. [PubMed: 9949199]
- Schludi MH, May S, Grässer FA, Rentzsch K, Kremmer E, Küpper C, et al. Distribution of dipeptide repeat proteins in cellular models and C9orf72 mutation cases suggests link to transcriptional silencing. *Acta Neuropathol.* 2015; 130(4):537–55. [PubMed: 26085200]

- Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, et al. Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol.* 2013; 9(1):59–64. [PubMed: 23160002]
- Sobczak K, Michlewski G, de Mezer M, Kierzek E, Krol J, Olejniczak M, et al. Structural diversity of triplet repeat RNAs. *J Biol Chem.* 2010; 285(17):12755–64. [PubMed: 20159983]
- Sonenberg N, Morgan MA, Merrick WC, Shatkin AJ. A polypeptide in eukaryotic initiation factors that crosslinks specifically to the 5'-terminal cap in mRNA. *Proc Natl Acad Sci U S A.* 1978; 75(10):4843–7. [PubMed: 217002]
- Sonenberg N, Rupprecht KM, Hecht SM, Shatkin AJ. Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proc Natl Acad Sci U S A.* 1979; 76(9):4345–9. [PubMed: 291969]
- Stochmanski SJ, Therrien M, Laganière J, Rochefort D, Laurent S, Karemera L, et al. Expanded ATXN3 frameshifting events are toxic in *Drosophila* and mammalian neuron models. *Hum Mol Genet.* 2012; 21(10):2211–8. [PubMed: 22337953]
- Su Z, Zhang Y, Gendron TF, Bauer PO, Chew J, Yang WY, et al. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron.* 2014; 83(5):1043–50. [PubMed: 25132468]
- Svitkin YV, Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham GJ, et al. The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA.* 2001; 7(3):382–94. [PubMed: 11333019]
- Tao Z, Wang H, Xia Q, Li K, Jiang X, Xu G, et al. Nucleolar stress and impaired stress granule formation contribute to C9orf72 RAN translation-induced cytotoxicity. *Hum Mol Genet.* 2015; 24(9):2426–41. [PubMed: 25575510]
- Tassone F, Beilina A, Carosi C, Albertosi S, Bagni C, Li L, et al. Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA.* 2007; 13(4):555–62. [PubMed: 17283214]
- Tassone F, Hagerman RJ, Loesch DZ, Lachiewicz A, Taylor AK, Hagerman PJ. Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *Am J Med Genet.* 2000; 94(3):232–6. [PubMed: 10995510]
- Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, et al. CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron.* 2013; 78(3):440–55. [PubMed: 23602499]
- Toulouse A, Au-Yeung F, Gaspar C, Roussel J, Dion P, Rouleau GA. Ribosomal frameshifting on MJD-1 transcripts with long CAG tracts. *Hum Mol Genet.* 2005; 14(18):2649–60. [PubMed: 16087686]
- Tran H, Almeida S, Moore J, Gendron TF, Chalasani U, Lu Y, et al. Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a *Drosophila* Model of C9ORF72 FTD/ALS. *Neuron.* 2015; 87(6):1207–14. [PubMed: 26402604]
- Unbehaun A, Borukhov SI, Hellen CU, Pestova TV. 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.* 2004; 18(24):3078–93. [PubMed: 15601822]
- Vanderperre B, Lucier JF, Bissonnette C, Motard J, Tremblay G, Vanderperre S, et al. Direct detection of alternative open reading frames translation products in human significantly expands the proteome. *PLoS One.* 2013; 8(8):e70698. [PubMed: 23950983]
- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell.* 1991; 65(5):905–14. [PubMed: 1710175]
- Wen X, Tan W, Westergard T, Krishnamurthy K, Markandaiah SS, Shi Y, et al. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron.* 2014; 84(6):1213–25. [PubMed: 25521377]
- Williams AJ, Paulson HL. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends Neurosci.* 2008; 31(10):521–8. [PubMed: 18778858]

- Wilson JE, Pestova TV, Hellen CU, Sarnow P. Initiation of protein synthesis from the A site of the ribosome. *Cell*. 2000a; 102(4):511–20. [PubMed: 10966112]
- Wilson JE, Powell MJ, Hoover SE, Sarnow P. Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Mol Cell Biol*. 2000b; 20(14):4990–9. [PubMed: 10866656]
- Wojciechowska M, Olejniczak M, Galka-Marciniak P, Jazurek M, Krzyzosiak WJ. RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. *Nucleic Acids Res*. 2014; 42(19):11849–64. [PubMed: 25217582]
- Yamakawa M, Ito D, Honda T, Kubo K, Noda M, Nakajima K, et al. Characterization of the dipeptide repeat protein in the molecular pathogenesis of c9FTD/ALS. *Hum Mol Genet*. 2015; 24(6):1630–45. [PubMed: 25398948]
- Yang D, Abdallah A, Li Z, Lu Y, Almeida S, Gao FB. FTD/ALS-associated poly(GR) protein impairs the Notch pathway and is recruited by poly(GA) into cytoplasmic inclusions. *Acta Neuropathol*. 2015; 130(4):525–35. [PubMed: 26031661]
- Yoon HJ, Donahue TF. The suil suppressor locus in *Saccharomyces cerevisiae* encodes a translation factor that functions during tRNA(iMet) recognition of the start codon. *Mol Cell Biol*. 1992; 12(1):248–60. [PubMed: 1729602]
- Zhang K, Donnelly CJ, Haeusler AR, Grima JC, Machamer JB, Steinwald P, et al. The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*. 2015a; 525(7567):56–61. [PubMed: 26308891]
- Zhang Y, You J, Wang X, Weber J. The DHX33 RNA Helicase Promotes mRNA Translation Initiation. *Mol Cell Biol*. 2015b; 35(17):2918–31. [PubMed: 26100019]
- Zhang YJ, Jansen-West K, Xu YF, Gendron TF, Bieniek KF, Lin WL, et al. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol*. 2014; 128(4):505–24. [PubMed: 25173361]
- Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A*. 2011; 108(1):260–5. [PubMed: 21173221]
- Zu T, Liu Y, Bañez-Coronel M, Reid T, Pletnikova O, Lewis J, et al. RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci U S A*. 2013; 110(51):E4968–77. [PubMed: 24248382]
- Zuccato C, Valenza M, Cattaneo E. Molecular mechanisms and potential therapeutic targets in Huntington's disease. *Physiological reviews*. 2010; 90(3):905–81. [PubMed: 20664076]

Highlights

- Repeat-associated non-AUG (RAN) translation produces toxic repeat proteins
- RAN translation utilizes a novel initiation mechanism to make these proteins
- Repeat content and context influence RAN translation toxicity and efficiency
- How exactly RAN translation relates to known modes of initiation remains unclear
- RAN translation is an intriguing therapeutic target in repeat expansion disorders

Figure 1. Canonical scanning model of translation initiation

Step 1 - The eIF4F complex, composed of eIF4E, eIF4G, eIF4A, binds to the 5' m⁷G cap with eIF4B and/or eIF4H. PABP associates with eIF4G to circularize the mRNA. *Step 2* - The eIF4F complex recruits the 43S Pre-initiation complex (PIC), composed of the 40S ribosomal subunit, eIF1, eIF1A, eIF3, eIF5, and the ternary complex, consisting of the initiator methionine-tRNA, eIF2, and GTP. *Step 3-4* - The PIC and components of the eIF4F complex scan through the 5' UTR in the 5' to 3' direction until encountering an AUG start codon in a good Kozak context. *Step 5* - eIF1 dissociates from the PIC, and eIF2 hydrolyzes GTP with the assistance of eIF5, committing the 40S ribosome to translation initiation at the

present AUG codon. *Step 6* - eIF5B-GTP promotes association of the 60S subunit and displacement of most eIFs. *Step 7* - eIF5B hydrolyzes its bound GTP and dissociates with eIF1A, establishing the elongation-competent 80S ribosome. eIF = eukaryotic initiation factor, m⁷G = 7 methylguanosine, PABP = poly-adenosine binding protein, PIC = preinitiation complex.

Figure 2. Canonical and alternative mechanisms of translation initiation

A) In canonical translation initiation, the 43S PIC with a full complement of eIFs binds to the 5' m⁷G cap, then linearly scans through the 5' UTR until reaching an AUG start codon in good Kozak context. B) The poliovirus IRES uses complex RNA secondary structure to recruit nearly all eIFs to an internal site within the transcript, bypassing the eIF4E-m⁷G interaction. The 43S PIC is subsequently recruited, and then scans 5' to 3' to the start codon. C) The CrPV IRES requires and recruits only the 40S and 60S ribosomal subunits and an alanine-conjugated tRNA, initiating translation at a CCU codon in the absence of any eIFs. D) Translation of *Histone 4* mRNA begins with tethering of eIF4F and the 43S PIC to internal secondary structures. The 40S subunit is then transferred to the AUG start codon upstream.

Figure 3.

Production of RAN-translated proteins across different sequence contexts. A) When located in the 5' UTR, as in *FMR1*, expanded GC-rich repeats trigger initiation of RAN translation upstream of the canonical AUG start codon, leading to the production of FMRpolyG and FMRpolyA. B) When located in an intron, as in *C9orf72*, it is unclear what RNA species is the substrate for RAN translation: a spliced lariat, an aberrantly spliced transcript in which the intron is retained, or a 3' truncated RNA resulting from stalled transcription. The relevant RNA species produces polyGA, polyGR, and polyGP RAN-translation products. C) When located within an ORF, as in *HTT*, canonical translation still initiates at the AUG codon upstream of the repeats. However, expression of polyserine (polyS) and polyalanine (polyA) proteins occurs by a combination of RAN-translation through the repeats and frameshifting out of the native (polyglutamine; polyQ) frame. D) Repeats located in antisense transcripts, as in *C9orf72* and as *HTT*, are also substrates for RAN translation, further expanding the number of dipeptide or homopolymeric RAN proteins. The post-

transcriptional modification state of these transcripts (e.g. mRNA capping and polyadenylation) are unknown.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript