

Near-cognate initiation generates FMRpolyG from CGG repeats in Fragile X associated Tremor Ataxia Syndrome

Authors: Yuan Zhang^{1,2}, M. Rebecca Glineburg^{1#}, Venkatesha Basrur³, Kevin Conlon³, Deborah A. Hall⁴, Peter K. Todd^{1,5,#}

- 1) Department of Neurology, University of Michigan, Ann Arbor, MI, USA.
- 2) Department of Respiratory Medicine, Xiangya Hospital, Central South University, Changsha, China. 410008.
- 3) Department of Pathology, University of Michigan, Ann Arbor, MI, USA.
- 4) Department of Neurological Sciences, Rush University, Chicago, IL, USA
- 5) Ann Arbor Veterans Administration Healthcare, Ann Arbor, MI, USA.

Acknowledgements: We thank the University of Michigan Brain Bank, for access to tissues. We thank the University of Michigan Mass Spectrometry-Based Proteomics Resource Facility for their assistance and use of equipment. We thank Amy Krans for cloning plasmids and acquiring brain tissue. We thank Katelyn Green for providing eIF2A KO HAP1 cells, sharing unpublished data and intellectual input.

Funding: This work was funded by grants from the NIH (R01NS099280 and R01NS086810 to PKT and T32NS00722237 to MRG) and private philanthropic support to DAH and PKT.

Abstract:

Repeat associated non-AUG (RAN) translation of *FMR1* 5' UTR CGG repeats produces toxic homopolymeric proteins that accumulate within ubiquitinated inclusions in Fragile X-associated tremor/ataxia syndrome (FXTAS) patient brains and model systems. The most abundant RAN product, FMRpolyG, initiates predominantly at an ACG codon located just 5' to the repeat. Methods to accurately measure FMRpolyG in FXTAS patients are lacking. Here we used data dependent acquisition (DDA) and parallel reaction monitoring (PRM) mass spectrometry coupled with stable isotope labeled standard peptides (SIS) to identify potential signature FMRpolyG fragments in patient cells and tissues. Following immunoprecipitation (IP) enrichment, we detected FMRpolyG signature peptides by PRM in transfected cells, FXTAS human samples and patient derived stem cells, but not in controls. Surprisingly, we identified two amino-terminal peptides: one beginning with methionine (Ac-MEAPLPGGVR) initiating at an ACG, and a second beginning with threonine (Ac-TEAPLPGGVR), initiating at a GUG. Abundance of the threonine peptide was enhanced relative to the methionine peptide upon activation of the integrated stress response. In addition, loss of the eIF2 alternative factor, eIF2A, or enhanced expression of initiation factor eIF1, preferentially suppressed GUG initiated FMRpolyG synthesis. These data demonstrate that FMRpolyG is quantifiable in human samples and that RAN translation on *FMR1* initiates at specific near cognate codons dependent on available initiation factors and cellular environment.

Key words: FXTAS; RAN; FMRpolyG; PRM

Introduction

Fragile X associated tremor ataxia syndrome (FXTAS) is an inherited late-onset neurodegenerative disorder caused by a CGG repeat expansion within the 5'UTR of *FMR1*. One of the pathological hallmarks of this disease is the presence of ubiquitinated neuronal inclusions in patient brain tissue (Greco et al. 2002; Greco et al. 2006). These inclusions are thought to occur as a result of an aberrant form of translation, termed repeat associated non-AUG (RAN) translation (Zu et al. 2011; Todd et al. 2013; Glineburg et al. 2018). RAN translation on *FMR1* transcripts is enhanced by CGG repeat expansions that form strong secondary structures in the RNA (Kearse et al. 2016; Krzyzosiak et al. 2012; Zumwalt et al. 2007; Verma et al. 2020). These repeats are thought to stall scanning ribosomes, allowing initiation just upstream of or within the repeat in the absence of an AUG start codon (Kearse et al. 2016). RAN translation at CGG repeats (CGG RAN) can occur in all three reading frames, producing three distinct toxic homopolymeric peptides, the most abundant of which is a polyglycine product, FMRpolyG (Todd et al. 2013; Kearse et al. 2016; Green et al. 2017; Linsalata et al. 2019). Overexpression of CGG repeats that support FMRpolyG expression promotes neurodegeneration in various model systems (Todd et al. 2013; Sellier et al. 2017; Oh et al. 2015; Rodriguez et al. 2020; Linsalata et al. 2019; Hukema et al. 2015; Hoem et al. 2019) and FMRpolyG is detected in human FXTAS brain tissue where it localizes to ubiquitin and P62+ inclusions (Krans et al. 2019; Buijsen et al. 2016; Buijsen et al. 2014). FMRpolyG has recently been detected via mass spectrometry (MS) in transfected cells as well as in patient tissue (Sellier et al. 2017; Ma et al. 2019); however, reliable detection of CGG RAN peptides and the ability to accurately quantify them is lacking.

In reporter assays, RAN translation of FMRpolyG predominantly initiates at two near cognate codons: an ACG, located 36 nts upstream of the repeat and a GUG located 12 nts from the repeat (Kearse et al. 2016; Sellier et al. 2017). Exactly how this initiation occurs and what factors are required is not fully understood. Recently, our lab and others uncovered a role for the integrated stress response (ISR) in upregulating RAN translation at CGG repeats and at the *C9orf72* ALS-causing G₄C₂ repeat (Green et al. 2017; Cheng et al. 2018; Sonobe et al. 2018; Westergard et al. 2019). Under normal conditions, eIF2 binds GTP and Met-tRNA_{Met} to form the ternary complex (TC) that then binds the 40S ribosome and a host of other translation initiation factors that facilitate cap binding and scanning, to form the 43S preinitiation complex (PIC). The PIC then scans along the mRNA, 5'-3', until it encounters the AUG start codon and initiation occurs (Sonenberg and Hinnebusch 2009; Hinnebusch 2011; Jackson, Hellen, and Pestova 2010; Hershey 2000). However, when any one of four ISR kinases (PERK, GCN2, HRI, and PKR) are activated, they phosphorylate canonical translation initiation factor eIF2 α , thereby inhibiting eIF2 β from exchanging GDP for GTP, and preventing a new round of TC assembly. This effectively inhibits canonical translation as no new TCs can be loaded onto 40S ribosomes to promote initiation (Jackson, Hellen, and Pestova 2010; Walter and Ron 2011).

Despite this global inhibition, some genes are able to bypass this response, and many do so by initiating at near-cognate codons, rather than the canonical AUG start codon (Starck et al. 2016; Ingolia et al. 2009; Green et al. 2017; Imataka, Olsen, and Sonenberg 1997; Ingolia, Lareau, and Weissman 2011; Kearse et al. 2019; Schwab et al. 2004). The likelihood that initiation will occur at a near-cognate codon during cellular stress is dependent on a number of factors including efficiency of TC recruitment to the 40S docked on an initiation codon, strength of that codon and its surrounding Kozak sequence, and availability of specific initiation factors (Kolitz, Takacs, and Lorsch 2009; Lind and Aqvist 2016; Tang et al. 2017; Kearse and Wilusz 2017). In particular, the eIF2 alternative factor, eIF2A, can facilitate near-cognate codon initiation, and unlike eIF2, is GTP-independent and upregulated during cellular stress (Starck et al. 2016; Starck et al. 2012; Komar et al. 2005; Kim et al. 2018; Sonobe et al. 2018). eIF2A has also been implicated in RAN translation of G₄C₂ repeats that initiates at a CUG codon (Sonobe et al. 2018).

While eIF2A is capable of promoting non-AUG initiation, the canonical initiation factor eIF1 has an opposing role of ensuring AUG start-codon fidelity. eIF1 does this by 1) keeping the 43S PIC in an opened conformation to allow for TC binding and scanning, 2) impeding eIF5 GTPase activity at non-AUGs, which slows down the conversion of eIF2-GTP to eIF2-GDP-Pi, and inhibits Pi release, and 3) dissociating from

the ribosome only after the PIC has bound to an AUG (Maag et al. 2005; Algire, Maag, and Lorsch 2005; Passmore et al. 2007; Unbehaun et al. 2004). In vitro experiments have shown that in the absence of eIF1, ribosomes can readily initiate at near cognate codons, and when eIF1 is overexpressed, use of near cognate codons is severely inhibited, specifically when the near-cognate codon differs at position 1 (Valasek et al. 2004; Pestova and Kolupaeva 2002; Fekete et al. 2007; Lind and Aqvist 2016). eIF1 overexpression was recently shown to dramatically inhibit CGG RAN translation (Linsalata et al. 2019). How eIF2A and eIF1 might interact to dictate start codon selection and whether or how they contribute to CGG RAN remain unknown.

Here, using parallel reaction monitoring (PRM) MS coupled with stable isotope labeled standard peptides (SIS), we both detected and quantified FMRpolyG peptides in overexpression systems and in FXTAS patient cells and samples. Furthermore, we identified an additional initiation codon, GUG, immediately 5' of the ACG, that is preferentially used in response to cellular stress. Initiation at this codon was selectively inhibited in the absence of eIF2 alternative factor, eIF2A, and during overexpression of canonical translation initiation factor, eIF1. Together these data provide a promising new technique for quantifying low abundant RAN peptides, and provide mechanistic insights into non-AUG initiation in a neurodegenerative disease.

Results

Development of PRM-SIS for FMRpolyG quantification

RAN translation of FMRpolyG primarily initiates at two near cognate codons, ACG and GUG, located in the *FMR1* 5' UTR, upstream of the CGG repeat (Kearse et al. 2016); however, while initiation at the ACG with a methionine has been previously reported in transfected samples, and recently found at extremely low levels endogenously, whether other species exist, and whether they can be detected at quantifiable levels has not been determined (Sellier et al. 2017; Ma et al. 2019). In order to detect and quantify endogenous FMRpolyG, we first identified potential FMRpolyG signature peptides. We predicted four different FMRpolyG sequences based on known initiation start sites for FMRpolyG: two longer isoforms beginning at ACG and initiating with either a methionine or a threonine, and two shorter isoforms beginning at GUG and initiating with either a methionine or a valine (Figure 1A). From these sequences, we identified four potential FMRpolyG signature peptides (two N-terminal and two C-terminal) to use for quantification (Figure 1A underlined sequences). We then applied data dependent acquisition (DDA) MS to identify these signature peptides in an FMRpolyG overexpression system. After comparing with the latest SwissProt human protein database (42054 sequences) (<https://www.uniprot.org/statistics/Swiss-Prot>), we found three of the four signature FMRpolyG peptides produced in HEK293Ts expressing exogenous FMRpolyG: the two C-terminal fragments—a carbamidomethyl modified peptide, CAM-CGAPMALSTR, and an unmodified peptide, SPPLGGGLPALAGLK—and an acetyl-modified methionine-initiated N-terminal fragment, Ac-MEAPLPGGVR (Figure 1B-D). This N-terminal fragment and the SPPLGGGLPALAGLK fragment have previously been identified using similar methods (Sellier et al. 2017; Ma et al. 2019).

The carbamidomethyl modification on an NH₂-terminal cysteine can cause the peptide to spontaneously undergo cyclization during MS sample preparation in a pH- and time-dependent manner, making CAM-CGAPMALSTR a poor peptide for quantification (Geoghegan et al. 2002). Thus, we chose Ac-MEAPLPGGVR and SPPLGGGLPALAGLK as signature peptides for FMRpolyG quantification by parallel reaction monitoring combined with stable isotope-labeled signature peptides (PRM-SIS). PRM-SIS is one of the most accurate approaches for directly quantifying the absolute abundance of known target proteins in a complex mixture (Peterson et al. 2012). By monitoring the response and determining the area under the curve of specific fragment ions of both endogenous and SIS peptides (spiked in at known concentrations during sample preparation), we can calculate the absolute abundance of our protein in a given sample (Figure 2A). Light and heavy peptides for Ac-MEAPLPGGVR and

SPPLGGGLPAGLK were synthesized and used to determine the limit of detection and limit of quantification in the background of Fragile X Syndrome patient derived fibroblasts which do not produce any *FMR1* mRNA, as well as in control CSF (Supplementary Figure 1A-D, Supplementary Table 1). Both conditions had values in the sub-nanomolar range (data not shown).

Optimizing FMRpolyG detection in transfected HEK293Ts by PRM-SIS

We next quantified FMRpolyG abundance in HEK293Ts transfected with or without untagged FMRpolyG₁₀₀. Despite being able to detect FMRpolyG in overexpression systems quite readily by western blot (Supplemental Figure 2C) (Todd et al. 2013; Kearse et al. 2016; Green et al. 2017), measurement of the two signature peptides Ac-MEAPLPGGVR and SPPLGGGLPAGLK by PRM-SIS were near the limit of quantification even with overexpression (Figure 2B). This discrepancy was not due to FMRpolyG aggregation, as FMRpolyG was primarily in the soluble fraction following ultracentrifugation, with only a small fraction found in the urea soluble pellet (Supplemental Figure 2A, B). Furthermore, FMRpolyG was efficiently digested by trypsin (Supplemental Figure 2C), suggesting the low quantification observed was not due to insufficient digestion of FMRpolyG but due to low abundance. Consistent with these observations, digesting lysates with the pentaglycine endopeptidase, lysostaphin (LS), which cleaves the polyglycine stretch in FMRpolyG (Todd et al. 2013), had no effect on the levels of Ac-MEAPLPGGVR, and only modestly enhanced detection of SPPLGGGLPAGLK (Supplemental Figure 2D, E). To overcome this low level of detection, we enriched for FMRpolyG through immunoprecipitation (IP). Following IP of FMRpolyG using a specific FMRpolyG N-terminal antibody (NTF1) (Krans et al. 2019), FMRpolyG was detected >100 fold more than in whole cell lysate (Figure 2B). This fold increase correlated with amounts of starting material (50 µg/cell lysate vs 4mg/IP) and was maintained when normalized to spiked-in SIS Ac-MEAPLPGGVR and SPPLGGGLPAGLK.

Detection and Quantification of endogenous FMRpolyG in FXTAS patient derived samples

We next sought to detect FMRpolyG in IPed human derived samples using PRM-SIS. Only one of the monitored FMRpolyG peptides, Ac-MEAPLPGGVR, was detectable at low levels in a CGG premutation iPSC line (FX 11-9U) and an un-methylated full mutation line (Rodriguez et al. 2020) but not in a control iPSC line with 23 repeats (2E) (Supplemental Figure 4A-B and Supplemental Table 2). We reasoned that the above are not ideal cell types to detect FMRpolyG, since *FMR1* is highly expressed in the brain and most FMRpolyG pathology reported to date is seen in neurons and glia (Tassone et al. 2004; Greco et al. 2006; Iwahashi et al. 2006; Hagerman et al. 2007; Todd et al. 2013). Using PRM-SIS, we detected Ac-MEAPLPGGVR in FXTAS cortex samples from 3 separate cases and 1 FXTAS hippocampus sample, but not in 2 age-matched control cortices (Figure 2C). The absolute abundance for Ac-MEAPLPGGVR ranged widely from 2.96 to 45.36 fmol/mg (Figure 2C, and Supplemental Table 2). We observed more FMRpolyG in cortex vs hippocampus for FXTAS-3, which may reflect the increased diffuse staining (rather than accumulation in ubiquitinylated inclusions) observed for cortical FMRpolyG by immunohistochemical techniques (Krans et al. 2019).

Discovery of an additional RAN initiation codon

DDA MS only detected one N-terminal fragment, Ac-MEAPLPGGVR. However, using the more sensitive PRM method, we identified a second N-terminal peptide—Ac-TEAPLPGGVR (Figure 3A). To measure this peptide, we synthesized light and heavy SIS for Ac-TEAPLPGGVR, and determined the LOD (0.9766 fmol/µl) and LOQ (2.9298 fmol/µl) in Fragile-X Syndrome fibroblasts to be in the sub-nanomolar range similar to that of Ac-MEAPLPGGVR (Supplementary Figure 3A-B). When we quantified Ac-TEAPLPGGVR from HEK293Ts transfected with untagged FMRpolyG, we found the abundance for the new Ac-TEAPLPGGVR fragment of FMRpolyG was lower than the Ac-MEAPLPGGVR fragment (Figure 3B, see Figure 2B), but still made up a significant proportion (15.69%) of detected FMRpolyG initiation

peptides. We saw a similar pattern in HEK293Ts transfected with FMRpolyG₁₀₀-3xFlag, using lysates directly (17.42%) and following IP (18.37%) (Figure 3D).

As Ac-TEAPLPGGVR is acetylated (a common posttranslational modification of N-terminal amino acids) and the codon 5' to the ACG is a GUG, coding for valine, we can conclude that this is an N-terminal initiation fragment and not a trypsin cleavage product of a spurious upstream initiation event (Verkerk et al. 1993; Lange et al. 2014; Ree, Varland, and Arnesen 2018). It is possible that Ac-TEAPLPGGVR is derived from the same near cognate codon ACG, which is supported by a similar observation in another gene, *CLK2* (Na et al. 2018). However, Ac-TEAPLPGGVR could also arise from Met initiation at the GUG followed by N-terminal Met (iMet) excision, an event that is particularly common when Thr is the second amino acid (Varland, Osberg, and Arnesen 2015; Martinez et al. 2008; Meinnel, Peynot, and Giglione 2005). To determine if Ac-TEAPLPGGVR was due to initiation at the 5' GUG with Met-tRNA_{iMet} followed by iMet excision, we designed two new reporters. In the first, we mutated the GUG upstream of the ACG to GGG. Importantly, this mutation did not alter the strength of the Kozak sequence for the ACG codon. In the second, we mutated the GUG upstream of the ACG to an AUG. We hypothesized that if Ac-TEAPLPGGVR was derived from the ACG, the GGG mutation (mut1) would have no effect on Ac-TEAPLPGGVR abundance, while the AUG mutation (mut2) would eliminate Ac-TEAPLPGGVR. Conversely, if Ac-TEAPLPGGVR was derived from initiation at the GUG upstream of the ACG, the GGG mutation would eliminate Ac-TEAPLPGGVR while the AUG mutation would enhance Ac-TEAPLPGGVR. We observed almost complete loss of Ac-TEAPLPGGVR from our mut1 GGGACG reporter, and a significant increase in Ac-TEAPLPGGVR from our mut2 AUGACG reporter. (Figure 3C-E). We also detected Ac-MTEAPLPGGVR from our mut2 AUGACG (Supplemental Figure 5). These results strongly suggest Ac-TEAPLPGGVR is generated by GUG mediated Met-tRNA_{iMet} initiation followed by iMet excision, and identifies a previously overlooked near cognate codon utilized for FmrpolyG synthesis.

Impact of integrated stress response activation on initiation codon usage for FMRpolyG

During the integrated stress response (ISR), canonical translation initiation factor eIF2 α is phosphorylated, preventing new TC formation, and consequently inhibiting global canonical translation (Walter and Ron 2011; Jackson, Hellen, and Pestova 2010). However, CGG RAN translation is selectively enhanced by the ISR in a codon specific manner (Green et al. 2017). As our data supports the role for two different initiation codons, GUG and ACG, we explored whether the ISR was preferentially promoting initiation at one or both of these codons. Similar to what we see with RAN translation-specific nanoluciferase reporters, we observed an ~2 fold increase in FMRpolyG-3xFlag upon treatment with the ER stress inducer, thapsigargin (TG) (Green, 2017, Figure 4A-B and Supplemental Figure 7A). Using PRM-SIS quantification, both methionine and threonine mediated RAN translational initiation were significantly enhanced by TG (Figure 4C). However, the relative ratio of Ac-TEAPLPGGVR to Ac-MEAPLPGGVR was enhanced by Thapsigargin (Figure 4D). This suggests that GUG mediated initiation in particular may occur independent of the eIF2 ternary complex.

FMRpolyG synthesis via GUG initiation is partially dependent on eIF2A and inhibited by eIF1 overexpression.

Canonical translation initiation is dependent on eIF2 forming a ternary complex with GTP and the initiator Met-tRNA_{iMet}, which binds the 40S ribosome along with other translation initiation factors to form the PIC. The PIC then binds the 5' cap of the mRNA and scans along in a 5' to 3' direction until it reaches an AUG start codon (Sonenberg and Hinnebusch 2009; Hinnebusch 2011; Jackson, Hellen, and Pestova 2010; Hershey 2000). However, during ISR activation, functional eIF2 is depleted, suggesting that initiation at these near cognate codons likely requires an eIF2 alternative factor. One particular factor, eIF2A, is known to be upregulated during cellular stress (Kim et al. 2018; Ventoso et al. 2006; Starck et al. 2016) and was recently suggested to play a role in RAN translational initiation at G₄C₂ repeats in C9orf72 (Sonobe et al. 2018). Thus we investigated whether eIF2A was required for ACG or GUG initiated RAN translation.

We performed PRM-SIS on WT and eIF2A KO HAP1 cell lines transfected with FMRpolyG₁₀₀-3xFlag. We found loss of the alternative eIF2 factor, eIF2A, reduced total FMRpolyG, but significantly suppressed GUG initiation by ~40%, relative to ACG mediated initiation (Figure 5A-C, Supplemental Figure 6A, and Supplemental Figure 7B). This suggests that GUG initiation is at least partially mediated by eIF2A.

Noncanonical translation initiation can also be facilitated by altered levels of canonical translation initiation factors. In particular, eIF1 and eIF5 have known roles in regulating start codon fidelity, where eIF1 enhances codon stringency, and eIF5 promotes dissociation of eIF1, reducing codon stringency (Pestova and Kolupaeva 2002; Pisarev et al. 2006; Luna et al. 2012). Our lab has previously shown that CGG RAN translation decreases with eIF1 overexpression and increases with eIF5 overexpression (Linsalata et al. 2019). To determine whether eIF1 and/or eIF5 differentially affect ACG or GUG initiation, we co-expressed FMRpolyG₁₀₀-3xFlag with either empty, eIF1, or eIF5 expression vectors, and performed PRM-SIS. While there was no effect on either Ac-MEAPLPGGVR or Ac-TEAPLPGGVR product abundance with eIF5 overexpression, we found overexpression of eIF1 significantly decreased both ACG and GUG initiation, with the impact being greater on GUG initiation, decreasing ~50% relative to ACG initiation (Figure 5D-F, Supplemental Figure 6B-C and Supplemental Figure 7C). These data are consistent with eIF1 having a preferential effect on GUG initiation.

Discussion

RAN translated proteins have emerged as important potential participants in the pathogenesis of multiple neurodegenerative disorders. Accurate quantification of these proteins and a deep understanding of the mechanisms underlying their production are both key steps in biomarker and therapy development for these untreatable diseases. Here, using a highly sensitive MS method, PRM-SIS, we identified N-terminal and C-terminal peptides of FMRpolyG in transfected samples and in patient tissue and cells. By comparing endogenous values to their respective SIS, we were able to quantify the absolute amount of FMRpolyG in patient brain tissue, uncovering a large range of FMRpolyG detection in patients. Moreover, this MS based methodology allowed us to identify a novel GUG initiation site for FMRpolyG production and discern its relative usage in conditions of stress and in response to specific perturbations in initiation factor abundance. Together, these data provide new insights into RAN translation and the pathogenesis of FXTAS and other repeat expansion disorders.

Both IRES and ribosomal scanning-dependent initiation mechanisms have been suggested for RAN translation at different repeat expansions (Zu et al. 2011; Zu et al. 2020; Cheng et al. 2018; Sellier et al. 2017; Kearse et al. 2016; Green et al. 2017). In our own work, we previously defined contributions from all near-cognate codons 5' to the repeat in *FMR1* by mutational analysis, which identified 2 key codons as likely drivers of FMRpolyG production (Kearse et al. 2016). Here we identify a third near-cognate codon as a meaningful contributor to FMRpolyG production. This particular GUG codon was originally dismissed as a contributor because its mutation placed the neighboring ACG codon into a stronger Kozak sequence (Kearse et al. 2016). How initiation site choice is made at these near-cognate codons on *FMR1* is not fully understood, but our data here provide some insight. We observe that loss of eIF2A dramatically reduces RAN translation from both ACG and GUG initiation, with the effect being particularly dramatic for GUG initiation, reducing it 2-fold relative to ACG initiation (Figure 5C). To date, there is no documentation of the effects of eIF2A on near cognate codons ACG and GUG, and thus, the mechanism by which eIF2A makes this distinction is not clear. eIF2A has previously been implicated in Met-tRNA_{iMet} mediated initiation at CUG and UUG codons (Starck et al. 2016; Starck et al. 2012; Liang et al. 2014), which share more similarity to GUG than ACG, and could explain the preferential decrease in GUG initiation in our case.

We also observe that GUG initiation is enhanced more so than ACG initiation upon activation of the integrated stress response. There are a number of likely reasons for this. First, eIF2A expression has been shown to increase in response to cellular stress (Starck et al. 2016; Kim et al. 2018; Reid et al. 2014). Given that both GUG and ACG mediated initiation are enhanced during thapsigargin-induced cellular

stress, this favors a role for eIF2A in stress-induced CGG RAN initiation. However, both GUG and ACG initiation still occur to some extent in eIF2A knockout cells, suggesting additional alternative initiation factors are at play. Alternatively, this differential stress response may reflect differential recruitment of the TC following 40S ribosomal scanning. ACG has the second lowest free energy of any near cognate codon (after CUG), and thus in general is more efficient at initiating under non-stressed conditions (Lind and Aqvist 2016). However, binding of the TC to a ribosome already sitting on an initiation codon is much stronger and more efficient when the ribosome is on a GUG vs an ACG (Kolitz, Takacs, and Lorsch 2009). Therefore, under non-stressed conditions, when the TC is bound to the PIC, ACG is the more favorable codon; however, under stress, when TC is limited, and recruitment of TC following PIC loading and scanning is necessary, GUG initiation becomes more favorable. Similarly, reduced GUG initiation in the setting of eIF1 overexpression is consistent with previous *in vitro* work showing eIF1 primarily acts by discriminating codons based on the first position (Lind and Aqvist 2016). In this system, eIF1 levels had no influence on ACG initiation while dramatically altering the free energy state of PIC binding to GUG. eIF1 overexpression has also been shown to strongly repress translation from a GUG but not an ACG codon, particularly when eIF1 expression is not auto regulated, which is the case in our system (Ivanov et al. 2010; Linsalata et al. 2019; Stewart et al. 2015).

A current paradox in the field is the ability to readily detect FMRpolyG by IHC in FXTAS cases and model systems while observing comparably low levels by MS (Ma et al. 2019; Krans et al. 2019; Green et al. 2017; Todd et al. 2013; Kearse et al. 2016). Overall, our PRM-SIS methodology exhibited excellent quantification characteristics in both CSF and cell lysates, with limits of detection for specific FMRpolyG peptides in the sub-nanomolar range. Moreover, immunoprecipitation led to enrichment of FMRpolyG peptides in a linear fashion from transfected cells (Figure 2B and Figure 3B), suggesting that the biochemistry of the peptides is not altered at higher concentrations. This strategy led to reliable detection and quantification of FMRpolyG in both FXTAS patient brains and cells, confirming the presence of FMRpolyG in FXTAS patient tissue and demonstrating definitively that initiation at near-AUG codons above the repeat contributes to its production in people. However, despite its sensitivity, we were not able to measure FMRpolyG using this methodology in CSF and its total abundance in FXTAS patients remained quite low. The exact reasons that this strategy failed to provide a reliable biomarker are not yet clear. Efforts to further increase FMRpolyG enrichment in PRM-SIS MS through enzymatic digests of the polyglycine tract with lysostaphin or 8M urea extraction methods were unsuccessful. It is likely that our immunoprecipitation strategy failed to capture all of the endogenous FMRpolyG protein, given that additional near-AUG codons (i.e. at the GUG closest to the repeat (Figure 1A)) or initiation within the repeat itself can be used for production of isoforms of FMRpolyG whose N-terminus is too short for MS detection and/or lacks the full antibody epitope. To date, C-terminally targeted antibodies to FMRpolyG have not been sufficiently robust to allow for effective immunoprecipitation of this endogenous protein from brain tissue (Todd et al. 2013, Sellier et al. 2017, Krans et al. 2019, and data not shown). The total abundance of FMRpolyG may also be lower than that observed by prior ICC and IHC measurements and reporter assays. Regardless, while our data demonstrate that PRM-MS is capable of reliably quantifying FMRpolyG at high abundance, the low level of detection in endogenous samples indicates that alternative approaches will be needed if this protein is to serve as a viable biomarker in FXTAS.

In sum, the studies presented here describe a new tool for quantifying the non-AUG initiated CGG repeat product, FMRpolyG and provides novel insights into the mechanisms underlying CGG RAN translation, including a specific role for a previously overlooked initiation codon. Furthermore, it describes differential effects on initiation at these codons, highlighting the importance of targeting multiple factors to successfully downregulate RAN translation. Further work into whether these alternative N-terminal ends influence the physical properties of FmrpolyG and its cellular toxicity are warranted, as are efforts to measure RAN translation products as a biomarker of FXTAS and other repeat disorder pathogenesis.

Methods

1. Cell lines and culture

HEK293Ts were obtained from ATCC and grown in high glucose DMEM with 10% fetal bovine serum (FBS), 1% penicillin/streptomycin (P/S). WT and eIF2A KO Hap1 cells were obtained from Horizon Discovery and grown in IMDM with 10%FBS, 1% P/S. All cell lines were maintained in 10 cm dishes at 37°C.

2. Reporters

FMRpolyG₁₀₀-3XFlag, and FMRpolyG₁₀₀ RAN translation reporters were made by cloning in their respective sequences into pcDNA3.1+ between BamHI and PspOMI (Supplemental Table 3). eIF1 and eIF5 expression vectors are previously published (Linsalata et al. 2019; Stewart et al. 2015).

3. Transfection and treatment

HEK293Ts were seeded at 4.0 X 10⁶ cells/10 cm dish and transfected at ~60-70% confluence for 24 hrs with 10 µg of either pcDNA3.1+ vector, FMRpolyG₁₀₀ or FMRpolyG₁₀₀-3XFlag and 3:1 Fugene HD to DNA. For stress induced experiments, the cells were transfected as above for 19 hours followed by 5 hrs with specified drugs (Tg, DMSO). For eIF1 and eIF5 experiments, HEK293Ts were seeded as above and cotransfected for 24 hrs with 2.5 µg FMRpolyG₁₀₀-3XFlag and 25 µg of either pcDNA3.1+ vector, eIF1 or eIF5 and 3:1 Fugene HD to DNA.

Hap1 WT and eIF2A KO cells were seeded at 7.0 X 10⁶ cells/10 cm dish and transfected at ~60-70% confluence for 24 hrs with 10 µg of FMRpolyG₁₀₀-3XFlag and 3:1 Genjet to DNA.

4. Clinical Specimens

FXS iPSCs were obtained from Wicell (Doers et al. 2014). Control 2E iPSC line was obtained from the Parent lab at the University of Michigan and have been previously published (Rodriguez et al. 2020). FXTAS peripheral lymphocytes (LCL1) were kindly gifted to us by Stephanie Sherman, and FXTAS skin fibroblasts (F3), and FXS skin fibroblasts (4026) were kindly gifted by Paul Hagerman. All have been previously reported (Garcia-Arocena et al. 2010; Todd et al. 2010). The 3 FXTAS patient frozen brain tissues (FXTAS1-1906, FXTAS2-345, FXTAS3-1558), 2 control frozen brain tissues (control1-382, control2-8662) were obtained from the University of Michigan Brain Bank. FXTAS patient CSF samples were obtained from Rush University Children's Hospital, and control CSF pool was purchased from Biochemed Services (3057BF). The information for these human derived samples are presented in Supplementary Table 2.

Lymphocyte cells were grow in 1640 RPMI containing 10% FBS and 1% P/S. Skin fibroblasts were grown in high glucose DMEM with 10% FBS, 1% non-essential amino acid (NEAA) and 1% P/S. The CSF and brain tissues were kept in frozen in -80 until use.

5. Immunoprecipitation (IP) and silver staining gel

Samples were immunoprecipitated according to manufacturer's protocol (LSKMAGAG02, Millipore). In brief, cells were washed with 1XPBS 2x then lysed with RIPA buffer (PI89900, Fisher) with Complete™, Mini Protease Inhibitor Cocktail (11836153001, Sigma). Protein concentration for each sample was determined by Pierce™ BCA protein assay kit (23225, Thermo Scientific™). 1.5-4.0 mg of lysate from FMRpolyG₁₀₀ transfected cells or 1.5-5mg of human derived samples were incubated with NTF1 antibody overnight at 4°C followed by adding PureProteome protein A/G mix magnetic beads for another 2 hours. 1.5-4.0 mg of lysate from FMRpolyG₁₀₀-3XFlag transfected cells were incubated with anti-flag M2 magnetic

beads (M8823, Sigma) overnight at 4°C. Beads were washed 3x with 1XTBS for 5 minutes at 4°C. IPed proteins were isolated by directly digesting on the beads themselves for mass spectrometry.

6. Ultracentrifuge

Samples were centrifuged at 70,000g for 1 hour. The pellet was resuspended in 8M urea, and both the supernatant and the pellet were run on a 15% SDS-PAGE, and blotted with the indicated antibodies.

7. Enzymatic digests

For trypsin experiments, lysates from 293Ts transfected with FMRpolyG₁₀₀ or FMRpolyG₁₀₀-3XFlag (dilute to 1 µg/µl) were digested with indicated amounts of trypsin (1 µg/µl) at 37°C for 1 hour. For lysostaphin (LS) experiments, lysates from 293Ts transfected with FMRpolyG₁₀₀ were digested with indicated amounts of LS at room temperature for 1 hour.

8. Western blot (WB)

Western blots were performed according to standard protocol (Krans, 2019). The following antibodies were used for western blots where indicated: NTF1 (1:1000, rabbit (Krans et al. 2019)), Flag M2 (1:1000, mouse, F1804, Sigma), eIF2A (1:8000, rabbit, 11233-1-AP, ProteinTech), eIF1 (1:1000, rabbit, 12496S, CST), eIF5 (1:1000, rabbit, 13894S, CST), tubulin-β (mouse, 1:1000, E7, DSHB), and GAPDH (mouse, 1:2000, sc-32233, SCBT) in 5% milk. HRP-conjugated secondary antibodies, goat anti-mouse HRP (111-035-146, Jackson ImmunoResearch), goat anti-rabbit HRP (111-035-144, Jackson Immuno Research) were used at 1:5000 in 5% milk.

9. Sample digestion for MS

The silver stained gel slice was destained according to the manufacturer's protocol (PROTSIL2-1KT, Sigma). The digestion for gel slice, whole lysates and IPed beads were done according to previously published protocols (Singh et al. 2018).

10. Liquid Chromatography Mass Spectrometry (LC-MS/MS)

A hybrid quadrupole-orbitrap mass spectrometer (Q Exactive HF, Thermo Scientific) coupled to nano-UHPLC (Ultimate 3000 RSLC Nano, ThermoScientific) was used for acquiring all mass spectrometry data. For Ultra High-Performance Liquid Chromatography (UHPLC), the procedures were according to protocol (Casado et al. 2018).

Resolved peptides were directly sprayed onto the mass spectrometer (MS) operating in either Data-Dependent Acquisition (DDA) or Parallel Reaction Monitoring (PRM, targeted) mode.

For DDA mode, the MS was set to collect one high-resolution MS1 scan (60,000 FWHM @200 m/z; scan range 400-1500 m/z) followed by data-dependent High-energy C-trap Dissociation MS/MS (HCD) spectra on the 20 most abundant ions observed in MS1. Other MS parameters were as previously published (Singh et al. 2018).

For PRM mode: Synthetic peptides with heavy isotope labeled arginine or lysine at the C-termini were obtained from ThermoScientific (AQUA Ultimate grade, 99% pure, ±5% quantitation precision). The digested samples were spiked with 25 fmol/µl of heavy synthetic peptides prior to UHPLC separation. In PRM mode, the quadrupole was set to isolate only the desired precursor ions (see Supplementary Table 1) followed by HCD-MS/MS on the isolated ions.

11. Protein identification

Proteome Discoverer (v2.1; Thermo Fisher) was used for data analysis. MS2 spectra were searched against latest SwissProt human protein database (42054 sequences) using the following search parameters: MS1 and MS2 tolerance were set to 10 ppm and 0.1 Da, respectively; carbamidomethylation of cysteines (57.02146 Da) was considered a static modification; acetylation of protein N-terminus, oxidation of methionine (15.9949 Da) and deamidation of asparagine and glutamine (0.98401 Da) were considered as variable modifications. Percolator algorithm (PD2.1) was used to determine the false discovery rate (FDR) and proteins/peptides with $\leq 1\%$ FDR were retained for further analysis. MS/MS spectra assigned to the peptides of interest were manually examined.

12. Protein quantification by PRM

The raw data was processed using Skyline 4.1.0.18169. A spectral library was created in Skyline for some of the target peptides using several DDA of FMRpolyG samples as well as blanks spiked with AQUA Ultimate endogenous (light) and isotopically labeled (heavy) peptides obtained from ThermoFisher Scientific (Supplementary Table 1). Product ions (minimum of 3) as determined by DDA analysis or Skyline prediction were monitored for each peptide (see the Condensed Transition List.xlsx attachment). Samples analyzed by PRM were spiked with 25 fmol/ μ l of heavy peptides and the concentration of the endogenous peptide was determined by comparison of the area ratios.

13. Statistical analysis

Blots were quantified by Image J software from NIH. Data plots and statistical analyses were done using GraphPad Prism 7. The error bars represent standard deviation (SD) of the mean. Unpaired t-test was used for WB blots intensity. Paired t-test was applied for all PRM-SIS quantification value analysis. We set $p < 0.05$ as the cutoff for significance.

References

Algire, M. A., D. Maag, and J. R. Lorsch. 2005. 'Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation', *Mol Cell*, 20: 251-62.

Buijsen, R. A., C. Sellier, L. A. Severijnen, M. Oulad-Abdelghani, R. F. Verhagen, R. F. Berman, N. Charlet-Berguerand, R. Willemse, and R. K. Hukema. 2014. 'FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome', *Acta Neuropathol Commun*, 2: 162.

Buijsen, R. A., J. A. Visser, P. Kramer, E. A. Severijnen, M. Gearing, N. Charlet-Berguerand, S. L. Sherman, R. F. Berman, R. Willemse, and R. K. Hukema. 2016. 'Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency', *Hum Reprod*, 31: 158-68.

Casado, M., C. Sierra, M. Batllori, R. Artuch, and A. Ormazabal. 2018. 'A targeted metabolomic procedure for amino acid analysis in different biological specimens by ultra-high-performance liquid chromatography-tandem mass spectrometry', *Metabolomics*, 14: 76.

Cheng, W., S. Wang, A. A. Mestre, C. Fu, A. Makarem, F. Xian, L. R. Hayes, R. Lopez-Gonzalez, K. Drenner, J. Jiang, D. W. Cleveland, and S. Sun. 2018. 'C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation', *Nat Commun*, 9: 51.

Doers, M. E., M. T. Musser, R. Nichol, E. R. Berndt, M. Baker, T. M. Gomez, S. C. Zhang, L. Abbeduto, and A. Bhattacharyya. 2014. 'iPSC-derived forebrain neurons from FXS individuals show defects in initial neurite outgrowth', *Stem Cells Dev*, 23: 1777-87.

Fekete, C. A., S. F. Mitchell, V. A. Cherkasova, D. Applefield, M. A. Algire, D. Maag, A. K. Saini, J. R. Lorsch, and A. G. Hinnebusch. 2007. 'N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection', *EMBO J*, 26: 1602-14.

Garcia-Arocena, D., J. E. Yang, J. R. Brouwer, F. Tassone, C. Iwahashi, E. M. Berry-Kravis, C. G. Goetz, A. M. Sumis, L. Zhou, D. V. Nguyen, L. Campos, E. Howell, A. Ludwig, C. Greco, R. Willemse, R. J. Hagerman, and P. J. Hagerman. 2010. 'Fibroblast phenotype in male carriers of FMR1 premutation alleles', *Hum Mol Genet*, 19: 299-312.

Geoghegan, K. F., L. R. Hoth, D. H. Tan, K. A. Borzilleri, J. M. Withka, and J. G. Boyd. 2002. 'Cyclization of N-terminal S-carbamoylmethylcysteine causing loss of 17 Da from peptides and extra peaks in peptide maps', *J Proteome Res*, 1: 181-7.

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

Greco, C. M., R. F. Berman, R. M. Martin, F. Tassone, P. H. Schwartz, A. Chang, B. D. Trapp, C. Iwahashi, J. Brunberg, J. Grigsby, D. Hessl, E. J. Becker, J. Papazian, M. A. Leehey, R. J. Hagerman, and P. J. Hagerman. 2006. 'Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS)', *Brain*, 129: 243-55.

Greco, C. M., R. J. Hagerman, F. Tassone, A. E. Chudley, M. R. Del Bigio, S. Jacquemont, M. Leehey, and P. J. Hagerman. 2002. 'Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers', *Brain*, 125: 1760-71.

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

Hagerman, R. J., S. M. Coffey, R. Maselli, K. Soontarapornchai, J. A. Brunberg, M. A. Leehey, L. Zhang, L. W. Gane, G. Fenton-Farrell, F. Tassone, and P. J. Hagerman. 2007. 'Neuropathy as a presenting feature in fragile X-associated tremor/ataxia syndrome', *Am J Med Genet A*, 143A: 2256-60.

Hershey, J.W.B., Merrick, W.C. 2000. 'Translational Control of Gene Expression', *Cold Spring Harbor Laboratory Press*: 33-88.

Hinnebusch, A. G. 2011. 'Molecular mechanism of scanning and start codon selection in eukaryotes', *Microbiol Mol Biol Rev*, 75: 434-67, first page of table of contents.

Hoem, G., K. Bowitz Larsen, A. Overvatn, A. Brech, T. Lamark, E. Sjottem, and T. Johansen. 2019. 'The FMRpolyGlycine Protein Mediates Aggregate Formation and Toxicity Independent of the CGG mRNA Hairpin in a Cellular Model for FXTAS', *Front Genet*, 10: 249.

Hukema, R. K., R. A. Buijsen, M. Schoneville, C. Raske, L. A. Severijnen, I. Nieuwenhuizen-Bakker, R. F. Verhagen, L. van Dessel, A. Maas, N. Charlet-Berguerand, C. I. De Zeeuw, P. J. Hagerman, R. F. Berman, and R. Willemse. 2015. 'Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS', *Hum Mol Genet*, 24: 4948-57.

Imataka, H., H. S. Olsen, and N. Sonenberg. 1997. 'A new translational regulator with homology to eukaryotic translation initiation factor 4G', *EMBO J*, 16: 817-25.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman, and J. S. Weissman. 2009. 'Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling', *Science*, 324: 218-23.

Ingolia, N. T., L. F. Lareau, and J. S. Weissman. 2011. 'Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes', *Cell*, 147: 789-802.

Ivanov, I. P., G. Loughran, M. S. Sachs, and J. F. Atkins. 2010. 'Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1)', *Proc Natl Acad Sci U S A*, 107: 18056-60.

Iwahashi, C. K., D. H. Yasui, H. J. An, C. M. Greco, F. Tassone, K. Nannen, B. Babineau, C. B. Lebrilla, R. J. Hagerman, and P. J. Hagerman. 2006. 'Protein composition of the intranuclear inclusions of FXTAS', *Brain*, 129: 256-71.

Jackson, R. J., C. U. Hellen, and T. V. Pestova. 2010. 'The mechanism of eukaryotic translation initiation and principles of its regulation', *Nat Rev Mol Cell Biol*, 11: 113-27.

Kearse, M. G., D. H. Goldman, J. Choi, C. Nwaezeapu, D. Liang, K. M. Green, A. C. Goldstrohm, P. K. Todd, R. Green, and J. E. Wilusz. 2019. 'Ribosome queuing enables non-AUG translation to be resistant to multiple protein synthesis inhibitors', *Genes Dev*, 33: 871-85.

Kearse, M. G., K. M. Green, A. Krans, C. M. Rodriguez, A. E. Linsalata, A. C. Goldstrohm, and P. K. Todd. 2016. 'CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins', *Mol Cell*, 62: 314-22.

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

Kim, E., J. H. Kim, K. Seo, K. Y. Hong, S. W. A. An, J. Kwon, S. V. Lee, and S. K. Jang. 2018. 'eIF2A, an initiator tRNA carrier refractory to eIF2alpha kinases, functions synergistically with eIF5B', *Cell Mol Life Sci*, 75: 4287-300.

Kolitz, S. E., J. E. Takacs, and J. R. Lorsch. 2009. 'Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation', *RNA*, 15: 138-52.

Komar, A. A., S. R. Gross, D. Barth-Baus, R. Strachan, J. O. Hensold, T. Goss Kinzy, and W. C. Merrick. 2005. 'Novel characteristics of the biological properties of the yeast *Saccharomyces cerevisiae* eukaryotic initiation factor 2A', *J Biol Chem*, 280: 15601-11.

Krans, A., G. Skariah, Y. Zhang, B. Bayly, and P. K. Todd. 2019. 'Neuropathology of RAN translation proteins in fragile X-associated tremor/ataxia syndrome', *Acta Neuropathol Commun*, 7: 152.

Krzymosiak, W. J., K. Sobczak, M. Wojciechowska, A. Fiszer, A. Mykowska, and P. Kozlowski. 2012. 'Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target', *Nucleic Acids Res*, 40: 11-26.

Lange, P. F., P. F. Huesgen, K. Nguyen, and C. M. Overall. 2014. 'Annotating N termini for the human proteome project: N termini and Nalpha-acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome', *J Proteome Res*, 13: 2028-44.

Liang, H., S. He, J. Yang, X. Jia, P. Wang, X. Chen, Z. Zhang, X. Zou, M. A. McNutt, W. H. Shen, and Y. Yin. 2014. 'PTENalpha, a PTEN isoform translated through alternative initiation, regulates mitochondrial function and energy metabolism', *Cell Metab*, 19: 836-48.

Lind, C., and J. Aqvist. 2016. 'Principles of start codon recognition in eukaryotic translation initiation', *Nucleic Acids Res*, 44: 8425-32.

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

Luna, R. E., H. Arthanari, H. Hiraishi, J. Nanda, P. Martin-Marcos, M. A. Markus, B. Akabayov, A. G. Milbradt, L. E. Luna, H. C. Seo, S. G. Hyberts, A. Fahmy, M. Reibarkh, D. Miles, P. R. Hagner, E. M. O'Day, T. Yi, A. Marintchev,

A. G. Hinnebusch, J. R. Lorsch, K. Asano, and G. Wagner. 2012. 'The C-terminal domain of eukaryotic initiation factor 5 promotes start codon recognition by its dynamic interplay with eIF1 and eIF2beta', *Cell Rep*, 1: 689-702.

Ma, L., A. W. Herren, G. Espinal, J. Randol, B. McLaughlin, V. Martinez-Cerdeno, I. N. Pessah, R. J. Hagerman, and P. J. Hagerman. 2019. 'Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome', *Acta Neuropathol Commun*, 7: 143.

Maag, D., C. A. Fekete, Z. Gryczynski, and J. R. Lorsch. 2005. 'A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon', *Mol Cell*, 17: 265-75.

Martinez, A., J. A. Traverso, B. Valot, M. Ferro, C. Espagne, G. Ephritikhine, M. Zivy, C. Giglione, and T. Meinnel. 2008. 'Extent of N-terminal modifications in cytosolic proteins from eukaryotes', *Proteomics*, 8: 2809-31.

Meinnel, T., P. Peynot, and C. Giglione. 2005. 'Processed N-termini of mature proteins in higher eukaryotes and their major contribution to dynamic proteomics', *Biochimie*, 87: 701-12.

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

Oh, S. Y., F. He, A. Krans, M. Frazer, J. P. Taylor, H. L. Paulson, and P. K. Todd. 2015. 'RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome', *Hum Mol Genet*, 24: 4317-26.

Passmore, L. A., T. M. Schmeing, D. Maag, D. J. Applefield, M. G. Acker, M. A. Algire, J. R. Lorsch, and V. Ramakrishnan. 2007. 'The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome', *Mol Cell*, 26: 41-50.

Pestova, T. V., and V. G. Kolupaeva. 2002. 'The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection', *Genes Dev*, 16: 2906-22.

Peterson, A. C., J. D. Russell, D. J. Bailey, M. S. Westphall, and J. J. Coon. 2012. 'Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics', *Mol Cell Proteomics*, 11: 1475-88.

Pisarev, A. V., V. G. Kolupaeva, V. P. Pisareva, W. C. Merrick, C. U. Hellen, and T. V. Pestova. 2006. 'Specific functional interactions of nucleotides at key -3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex', *Genes Dev*, 20: 624-36.

Ree, R., S. Varland, and T. Arnesen. 2018. 'Spotlight on protein N-terminal acetylation', *Exp Mol Med*, 50: 1-13.

Reid, D. W., Q. Chen, A. S. Tay, S. Shenolikar, and C. V. Nicchitta. 2014. 'The unfolded protein response triggers selective mRNA release from the endoplasmic reticulum', *Cell*, 158: 1362-74.

Rodriguez, C. M., S. E. Wright, M. G. Kearse, J. M. Haenfler, B. N. Flores, Y. Liu, M. F. Ifrim, M. R. Glineburg, A. Krans, P. Jafar-Nejad, M. A. Sutton, G. J. Bassell, J. M. Parent, F. Rigo, S. J. Barmada, and P. K. Todd. 2020. 'A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis', *Nat Neurosci*, 23: 386-97.

Schwab, S. R., J. A. Shugart, T. Horng, S. Malarkannan, and N. Shastri. 2004. 'Unanticipated antigens: translation initiation at CUG with leucine', *PLoS Biol*, 2: e366.

Sellier, C., R. A. M. Buijsen, F. He, S. Natla, L. Jung, P. Tropel, A. Gaucherot, H. Jacobs, H. Meziane, A. Vincent, M. F. Champy, T. Sorg, G. Pavlovic, M. Wattenhofer-Donze, M. C. Birling, M. Oulad-Abdelghani, P. Eberling, F. Ruffenach, M. Joint, M. Anheim, V. Martinez-Cerdeno, F. Tassone, R. Willemsen, R. K. Hukema, S. Viville, C. Martinat, P. K. Todd, and N. Charlet-Berguerand. 2017. 'Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome', *Neuron*, 93: 331-47.

Singh, A. K., P. Khare, A. Obaid, K. P. Conlon, V. Basrur, R. A. DePinho, and K. Venuprasad. 2018. 'SUMOylation of ROR-gammat inhibits IL-17 expression and inflammation via HDAC2', *Nat Commun*, 9: 4515.

Sonenberg, N., and A. G. Hinnebusch. 2009. 'Regulation of translation initiation in eukaryotes: mechanisms and biological targets', *Cell*, 136: 731-45.

Sonobe, Y., G. Ghadge, K. Masaki, A. Sendoel, E. Fuchs, and R. P. Roos. 2018. 'Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress', *Neurobiol Dis*, 116: 155-65.

Starck, S. R., V. Jiang, M. Pavon-Eternod, S. Prasad, B. McCarthy, T. Pan, and N. Shastri. 2012. 'Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I', *Science*, 336: 1719-23.

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

Stewart, J. D., J. L. Cowan, L. S. Perry, M. J. Coldwell, and C. G. Proud. 2015. 'ABC50 mutants modify translation start codon selection', *Biochem J*, 467: 217-29.

Tang, L., J. Morris, J. Wan, C. Moore, Y. Fujita, S. Gillaspie, E. Aube, J. Nanda, M. Marques, M. Jangal, A. Anderson, C. Cox, H. Hiraishi, L. Dong, H. Saito, C. R. Singh, M. Witcher, I. Topisirovic, S. B. Qian, and K. Asano. 2017. 'Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide', *Nucleic Acids Res*, 45: 11941-53.

Tassone, F., R. J. Hagerman, D. Garcia-Arocena, E. W. Khandjian, C. M. Greco, and P. J. Hagerman. 2004. 'Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome', *J Med Genet*, 41: e43.

Todd, P. K., S. Y. Oh, A. Krans, F. He, C. Sellier, M. Frazer, A. J. Renoux, K. C. Chen, K. M. Scaglione, V. Basrur, K. Elenitoba-Johnson, J. P. Vonsattel, E. D. Louis, M. A. Sutton, J. P. Taylor, R. E. Mills, N. Charlet-Berguerand, and H. L. Paulson. 2013. 'CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome', *Neuron*, 78: 440-55.

Todd, P. K., S. Y. Oh, A. Krans, U. B. Pandey, N. A. Di Prospero, K. T. Min, J. P. Taylor, and H. L. Paulson. 2010. 'Histone deacetylases suppress CGG repeat-induced neurodegeneration via transcriptional silencing in models of fragile X tremor ataxia syndrome', *PLoS Genet*, 6: e1001240.

Unbehaun, A., S. I. Borukhov, C. U. Hellen, and T. V. Pestova. 2004. 'Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP', *Genes Dev*, 18: 3078-93.

Valasek, L., K. H. Nielsen, F. Zhang, C. A. Fekete, and A. G. Hinnebusch. 2004. 'Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection', *Mol Cell Biol*, 24: 9437-55.

Varland, S., C. Osberg, and T. Arnesen. 2015. 'N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects', *Proteomics*, 15: 2385-401.

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

Verkerk, A. J., E. de Graaff, K. De Boulle, E. E. Eichler, D. S. Konecki, E. Reyniers, A. Manca, A. Poustka, P. J. Willems, D. L. Nelson, and et al. 1993. 'Alternative splicing in the fragile X gene FMR1', *Hum Mol Genet*, 2: 1348.

Verma, A. K., E. Khan, S. K. Mishra, A. Mishra, N. Charlet-Berguerand, and A. Kumar. 2020. 'Curcumin Regulates the r(CGG)(exp) RNA Hairpin Structure and Ameliorate Defects in Fragile X-Associated Tremor Ataxia Syndrome', *Front Neurosci*, 14: 295.

Walter, P., and D. Ron. 2011. 'The unfolded protein response: from stress pathway to homeostatic regulation', *Science*, 334: 1081-6.

Westergard, T., K. McAvoy, K. Russell, X. Wen, Y. Pang, B. Morris, P. Pasinelli, D. Trott, and A. Haeusler. 2019. 'Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress', *EMBO Mol Med*, 11.

Zu, T., B. Gibbens, N. S. Doty, M. Gomes-Pereira, A. Huguet, M. D. Stone, J. Margolis, M. Peterson, T. W. Markowski, M. A. Ingram, Z. Nan, C. Forster, W. C. Low, B. Schoser, N. V. Somia, H. B. Clark, S. Schmechel, P. B. Bitterman, G. Gourdon, M. S. Swanson, M. Moseley, and L. P. Ranum. 2011. 'Non-ATG-initiated translation directed by microsatellite expansions', *Proc Natl Acad Sci U S A*, 108: 260-5.

Zu, T., S. Guo, O. Bardhi, D. A. Ryskamp, J. Li, S. Khoramian Tusi, A. Engelbrecht, K. Klippe, P. Chakrabarty, L. Nguyen, T. E. Golde, N. Sonenberg, and L. P. W. Ranum. 2020. 'Metformin inhibits RAN translation through PKR pathway and mitigates disease in C9orf72 ALS/FTD mice', *Proc Natl Acad Sci U S A*, 117: 18591-99.

Zumwalt, M., A. Ludwig, P. J. Hagerman, and T. Dieckmann. 2007. 'Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the fragile X mental retardation 1 (FMR1) gene', *RNA Biol*, 4: 93-100.

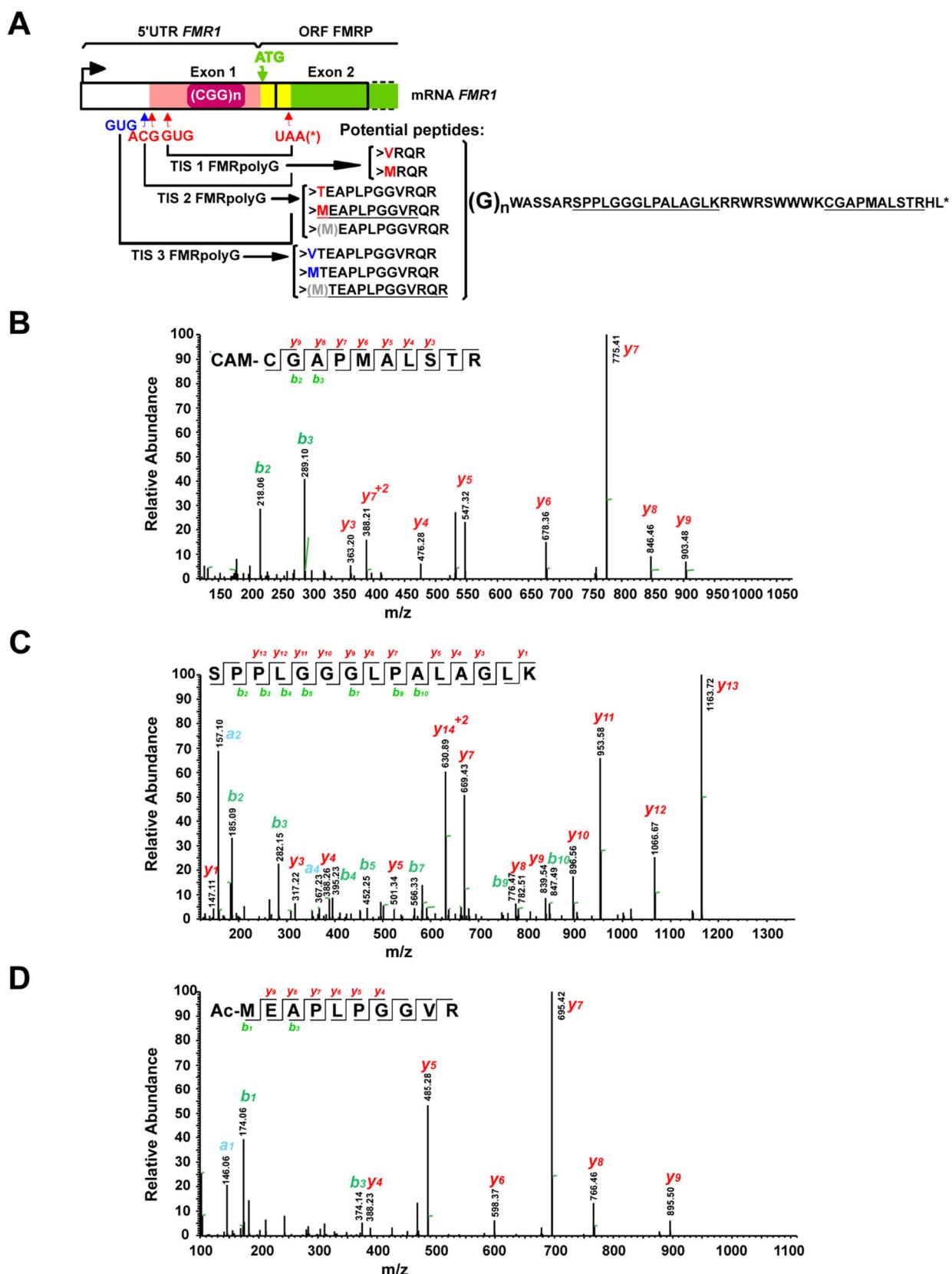


Figure 1. Predicted potential peptide sequences for RAN translation product FMRpolyG.

A. Schematic of the peptides produced via alternative translation initiation on *FMR1*. Top) *FMR1* schematic showing location of the FMRpolyG uORF (pink), FMRP ORF (green), and sequence corresponding to

overlap of these two ORFs (yellow). Bottom) The predicted FMRpolyG peptide sequences from three potential translation initiation sites (TIS) on *FMR1*, two previously identified initiation sites (red), and one novel initiation site (blue). Predicted trypsin-digested signature peptides for FMRpolyG that are monitored in this study are underlined. The grey M in brackets indicates a potential product of iMet excision.

B-D. LC-MS/MS spectra of FMRpolyG signature peptides CAM-CGAPMALSTRHL (B), SPPLGGGLPAGLK (C) and Ac-MEAPLPGGVR (D) from HEK293T cells transfected with FMRpolyG₁₀₀-3XFlag. Observed *b*- and *y*-ions are indicated.

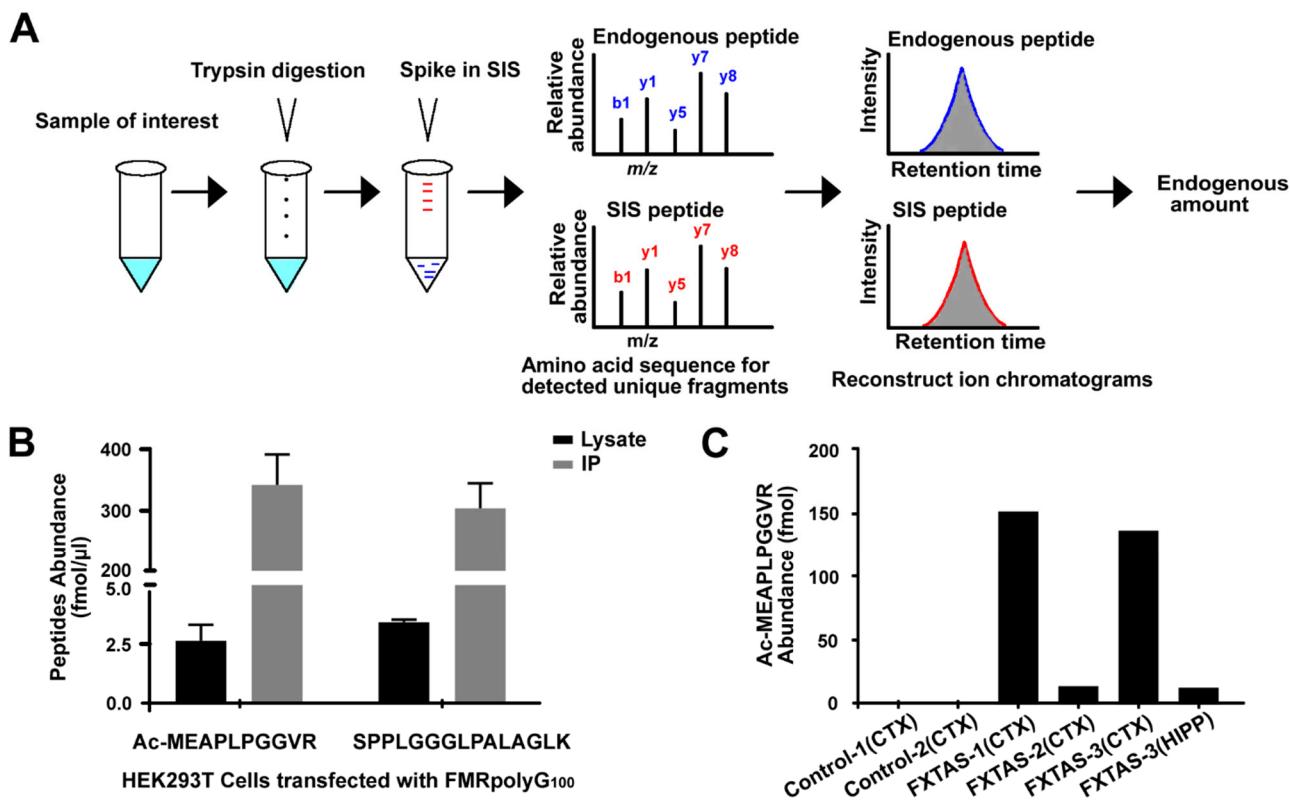


Figure 2. Parallel Reaction Monitoring based quantification of FMRpolyG.

A. Schematic for PRM-SIS in quantification. Samples were digested with trypsin and then spiked with known concentrations of a signature stable isotope-labeled (SIS) peptide. The amino acid sequence for our signature peptide of interest is detected by MS, and the endogenous peptide concentration is quantified relative to the SIS peptide.

B. Quantification of signature peptides for FMRpolyG, Ac-MEAPLPGGVR and SPPLGGGLPALAGLK, in HEK293T cell lysates transfected with untagged FMRpolyG₁₀₀ for 24 hrs and either processed directly or following immunoprecipitation with NTF1 antibody. Bars represent mean \pm SD. N=2.

C. Quantification of the absolute abundance of endogenous FMRpolyG peptide Ac-MEAPLPGGVR from human FXTAS and control brains by PRM-SIS.

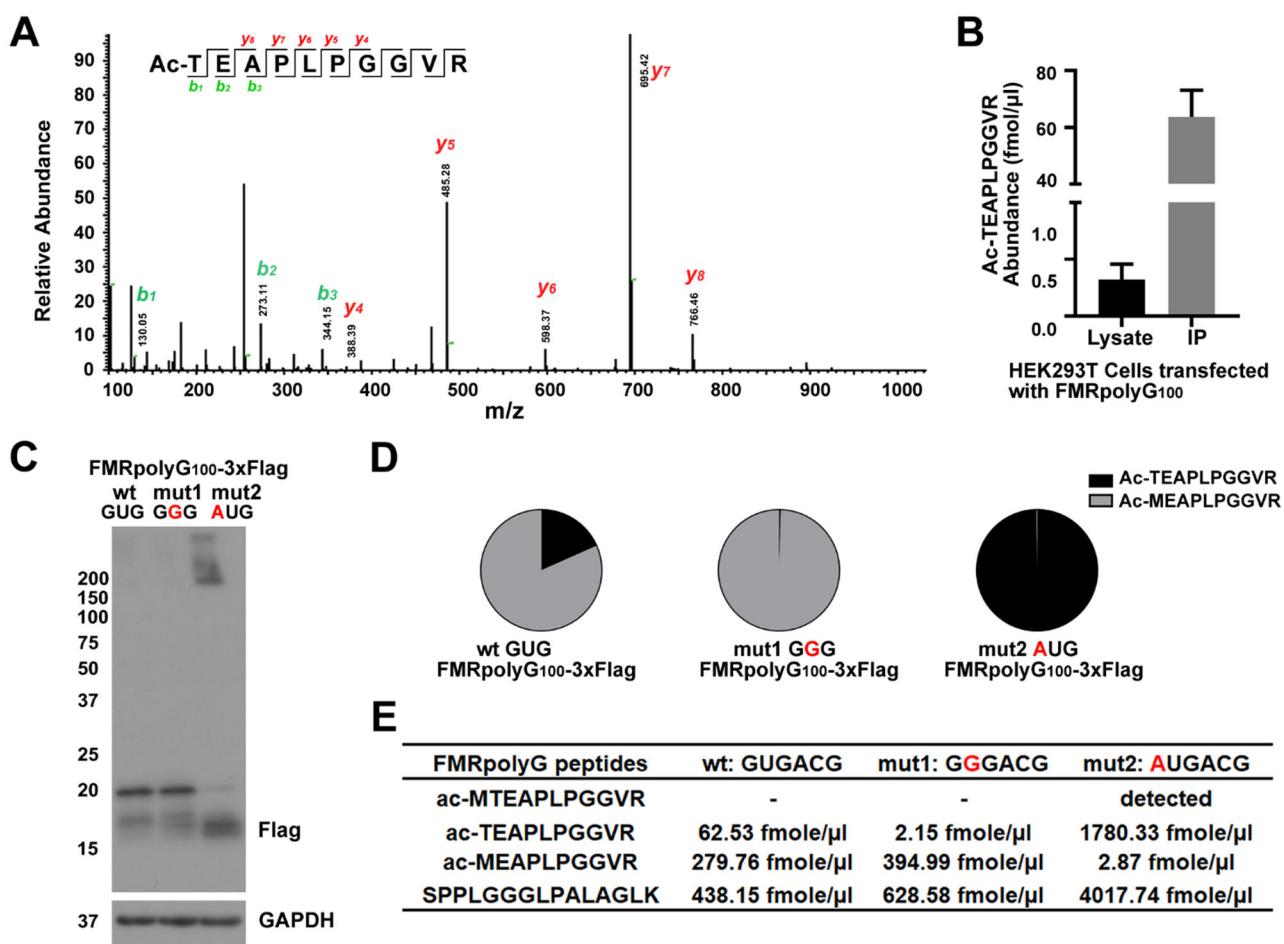


Figure 3. CGG RAN translation initiation can occur at a 5' GUG codon

A. Representative LC-MS/MS chromatogram of FMRpolyG peptide, Ac-TEAPLPGGVR from NTF1 immunoprecipitated HEK293T cells transfected with FMRpolyG₁₀₀.

B. Quantification of Ac-TEAPLPGGVR absolute abundance in HEK293T cells transfected with FMRpolyG₁₀₀. Bars represent mean ± SD. N=2.

C. Western blot of FMRpolyG from HEK293T cells transfected with the indicated constructs. GAPDH serves as a loading control.

D. The distribution of two different initiation fragments quantified via PRM-SIS from IPed lysates of 293Ts transfected with the indicated constructs.

E. Quantification of the absolute amount of signature peptides PRM-SIS from IPed lysates of 293Ts transfected with the indicated reporters. Detected indicates observed, but not quantified (See supplemental Figure 5)..

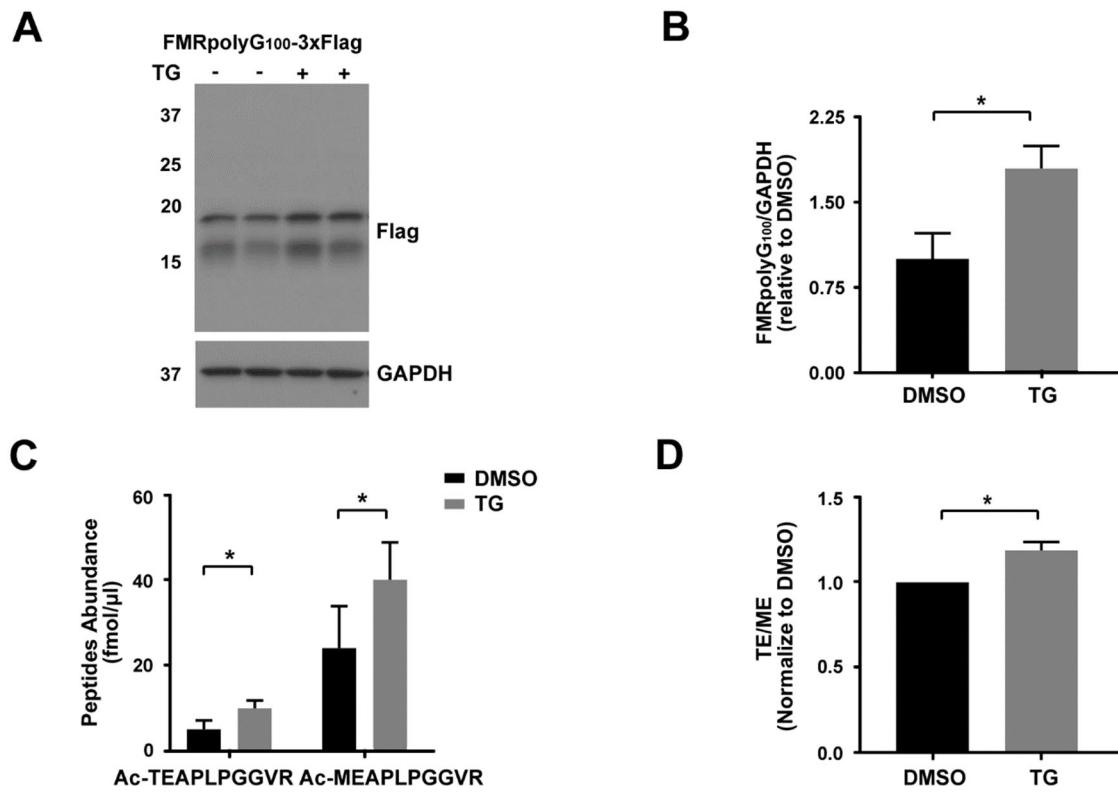


Figure 4. GUG initiation preferentially increases in response to thapsigargin.

A. Representative western blot of FMRpolyG from HEK293T cells transfected with FMRpolyG₁₀₀-3XFlag followed by DMSO or TG treatment. GAPDH serves as a loading control. N=3.

B. Quantification FMRpolyG from the above western blot. Bars represent mean \pm SD, N=3. *p < 0.05.

C. The absolute abundance of Ac-TEAPLPGGVR and Ac-MEAPLPGGVR by PRM-SIS in lysates of above samples. Bars represent mean \pm SD, N=3. *p < 0.05.

D. Ratios of the two N-terminal peptides in TG vs DMSO treated 293Ts. Bars represent mean \pm SD, N=3.

*p < 0.05.

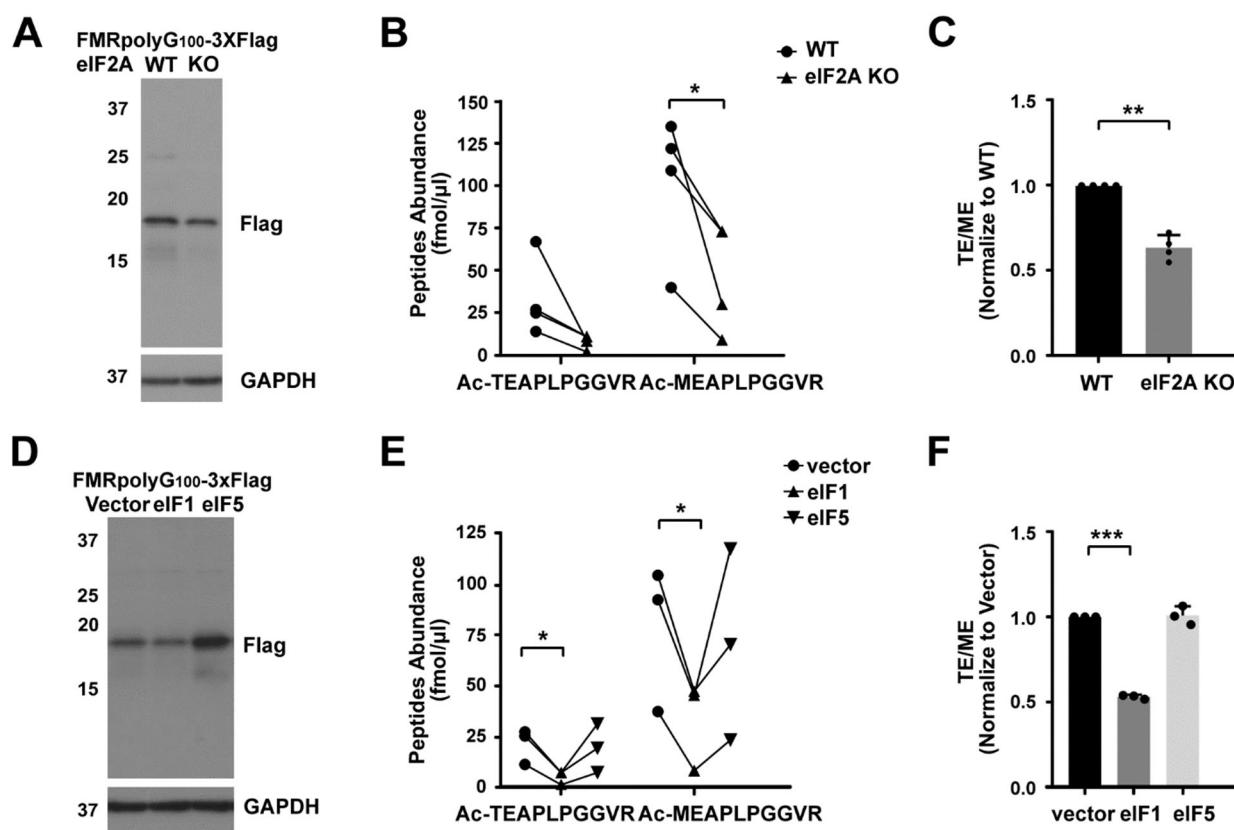


Figure 5. FMRpolyG initiation codon choice is influenced by eIF1 and eIF2A availability

A. Representative western blot of FMRpolyG from HAP1 WT and eIF2A KO cells transfected with FMRpolyG₁₀₀-3XFlag. GAPDH serves as a loading control. N=4.

B. Quantification of Ac-TEAPLPGGVR and Ac-MEAPLPGGVR absolute abundance by PRM-SIS in Flag-IPed lysates of HAP1 cells transfected with FMRpolyG₁₀₀-3XFlag. N=4. *p < 0.05.

C. Ratio of the two N-terminal peptides in HAP1 WT and eIF2A KO cells transfected with FMRpolyG₁₀₀-3XFlag. Bars represent mean ± SD, N=4. **p < 0.01.

D. Representative western blot of FMRpolyG from HEK293T cells co-transfected with FMRpolyG₁₀₀-3XFlag and control, eIF1 or eIF5 vectors. GAPDH serves as a loading control N=3.

E. Quantification of Ac-TEAPLPGGVR and Ac-MEAPLPGGVR absolute abundance by PRM-SIS in Flag-IPed lysates of HEK293T cells co-transfected with FMRpolyG₁₀₀-3XFlag and control, eIF1, or eIF5 vectors. N=3. *p < 0.05.

F. Ratio of the two N-terminal peptides in HEK293T cells co-transfected with FMRpolyG₁₀₀-3XFlag and control, eIF1 or eIF5 vectors. Bars represent mean ± SD, N=3. ***p < 0.001.