Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions

Katarzyna Tutak[1], Izabela Broniarek[1], Andrzej Zielezinski[2], Daria Niewiadomska[1], Anna Baud[1]*, Krzysztof Sobczak[1]*

[1] Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznan, Poland

[2] Department of Computational Biology, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, 61-614 Poznan, Poland

* To whom correspondence should be addressed:

K.S. tel: +48 61 829 57 66, email: ksobczak@amu.edu.pl

A.B. tel: +48 61 829 59 52, email: anna.baud@amu.edu.pl

Abstract

Expansion of CGG repeats (CGGexp) in the 5' untranslated region (5'UTR) of the *FMR1* gene underlies the fragile X-associated conditions including tremor/ataxia syndrome (FXTAS), a late-onset neurodegenerative disease. One pathomechanism of FXTAS is the repeat-associated non-AUG-initiated (RAN) translation of CGG repeats of mutant *FMR1* mRNA, resulting in production of FMRpolyG, a toxic protein containing long polyglycine tract. To identify novel modifiers of RAN translation we used an RNA-tagging system and mass spectrometry-based screening. It revealed proteins enriched on CGGexp-containing *FMR1* RNA in cellulo, including a ribosomal protein RPS26, a component of the 40S subunit. We demonstrated that RPS26, together with its chaperone TSR2, modulates FMRpolyG production and its toxicity. We also found that the number of proteins produced via RPS26-sensitive translation was limited, and 5'UTRs of mRNAs encoding these proteins were guanosine and cytosine-rich. Moreover, the silencing of another component of the 40S subunit, the ribosomal protein RPS25, also induced repression of FMRpolyG biosynthesis. Results of this study suggest that the composition of the 40S subunit plays important role in noncanonical CGGexp-related RAN translation.

Keywords

FMR1; unstable CGG repeats; expansion of triplet repeats; polyglycine; polyG protein; near-cognate start codon; neurotoxicity; ribosomal heterogeneity, ribosomal proteins, 40S subunit

Introduction

Fragile X chromosome-associated syndromes are rare genetic diseases caused by dynamic mutations of the *fragile X messenger ribonucleoprotein 1* (*FMR1*) gene located on the X chromosome. The gene typically contains 25–30 CGG repeats in the 5' untranslated region (5'UTR). However, these triplet repeats are highly polymorphic and prone to expand, resulting in either a full mutation (FM; over 200 CGG repeats) or premutation (PM; 55–200 CGG repeats). On the molecular level, FM causes methylation of the *FMR1* promoter, leading to transcriptional silencing, loss of *FMR1* mRNA, and a lack of the main protein product, fragile X messenger ribonucleoprotein (FMRP), which is involved in modulating synaptic plasticity. An FM causes early onset neurodevelopmental fragile X syndrome (FXS), while PM is linked to many fragile X-associated conditions including fragile X-associated tremor/ataxia syndrome (FXTAS), fragile X-associated primary ovarian insufficiency (FXPOI), and fragile X-associated neuropsychiatric disorders (FXAND). The estimated prevalence of PM is 1 in 150–300 females and 1 in 400–850 males. However, due to incomplete penetrance, approximately 1 in 5,000–10,000 men in their fifties or later will develop FXTAS. In female PM carriers, random X-inactivation lowers the risk of FXTAS development (Hagerman & Hagerman, 2016; Tassone *et al*, 2012; Jacquemont *et al*, 2004). FXTAS is a late-onset neurodegenerative disease. Its pathology includes neuropathy, white matter loss, mild brain atrophy, and ubiquitin-positive inclusions in neurons and glia (Hagerman & Hagerman, 2016; Greco *et al*, 2002, 2006). Patients suffer from cognitive decline, dementia, parkinsonism, imbalance, gait ataxia, and tremors accompanied by psychological difficulties such as anxiety or depression (Hagerman *et al*, 2018; Hagerman & Hagerman, 2016). To date, no effective treatment targeting the cause, rather than the symptoms, has been proposed for any PM-linked disorders.

Three main molecular pathomechanisms are believed to contribute to FXTAS, FXPOI, and FXAND development (Glineburg *et al*, 2018; Malik *et al*, 2021a; Hagerman *et al*, 2018). First, high-content guanosine and cytosine nucleotides in the 5'UTR of *FMR1* cause co-transcriptional DNA:RNA hybrid formations (R-loops), which trigger the DNA damage response, thus compromising genomic stability, leading to cell death (Loomis *et al*, 2014; Abu Diab *et al*, 2018).

Second, RNA gain-of-function toxicity induces nuclear foci formation by mRNA containing expanded CGG repeats (CGGexp) which form stable hairpin structures, and sequester proteins leading to their functional depletion (Sellier *et al*, 2010, 2013, 2017). Finally, mRNA containing CGGexp can act as a template for noncanonical protein synthesis called repeats-associated non-AUG initiated (RAN) translation. Protein production from expanded nucleotide repeats is initiated at different near-cognate start codons in diverse reading frames. The resultant toxic proteins contain repeated amino acid tracts, such as polyglycine (FMRpolyG), polyalanine (FMRpolyA), polyarginine (FMRpolyR), or hybrids of them produced as a result of frameshifting (Todd *et al*, 2013; Wright *et al*, 2022; Glineburg *et al*, 2018; Kearse *et al*, 2016). Notably, the open reading frame for FMRP starting from the AUG codon downstream to the repeats is canonically synthesized. In FXTAS and FXPOI, the most abundant RAN protein is the toxic FMRpolyG, which aggregates and forms characteristic intranuclear or perinuclear inclusions observed in patient cells and model systems (Greco *et al*, 2002, 2006; Ariza *et al*, 2016; Todd *et al*, 2013; Sellier *et al*, 2017; Ma *et al*, 2019; Derbis *et al*, 2018).

According to the current RAN translation model of RNA with CGGexp, the eIF4F complex and 43S pre-initiation complex (PIC) bind to the 5'-cap of *FMR1* mRNA and scan through the 5'UTR until encountering a steric hindrance, a hairpin structure formed by CGGexp, or a nearby 5'UTR sequence. This blockage increases the dwell time of PIC and lowers initiation codon fidelity. As a result, the stalled 40S ribosome initiates RAN translation at less favored, near-cognate start codons (ACG or GUG) upstream of repeats (in the FMRpolyG reading frame) or within CGG repeats (in the FMRpolyA reading frame) (Kearse *et al*, 2016; Green *et al*, 2016). Unwinding stable RNA secondary structures appears to be crucial for initiating RAN translation, as several RNA helicases such as ATP-dependent RNA helicase DDX3X, ATP-dependent DNA/RNA helicase DHX36, and eukaryotic initiation factor 4A (eIF4A) are involved in its regulation (Linsalata *et al*, 2019; Kearse *et al*, 2016; Tseng *et al*, 2021). Proteins indirectly involved in RAN translation can also contribute to RAN-mediated toxicity via pathways related to stress response and nuclear transport (Green *et al*, 2017; Zu *et al*, 2020; Malik *et al*, 2021b).

New insights into ribosome heterogeneity have explained rearrangements within ribosome components at different developmental stages or responses to environmental stimuli. This has led to a better understanding of events that shape local ribosome homeostasis and affect the translatome (Genuth & Barna, 2018; Shi & Barna, 2015). For example, although ribosomes depleted of small ribosomal subunit protein eS26 (RPS26) stays functional in the cell, they are translating preferentially selected mRNAs (Ferretti *et al*, 2017; Yang & Karbstein, 2022; Li *et al*, 2022). Moreover, ribosomes containing small ribosomal subunit protein eS25 (RPS25) and Large ribosomal subunit protein uL1 RPL10A translate different pools of mRNA including those encoding key components of cell cycle process, metabolism, and development (Shi *et al*, 2017).

Recently, different proteins involved in RAN translation regulation reviewed in Baud *et al*, 2022 were uncovered. However, mechanistic insight into this process remains unresolved. Given the toxicity of RAN translation products, identification of its regulating factors, which may serve as potential therapeutic targets to combat RAN proteins-related toxicity in fragile X-associated conditions, is essential.

In this study, we adapted an RNA-tagging technique to identify proteins natively bound to the RNA of the 5'UTR of *FMR1* with CGGexp that mimics mutated transcripts in PM carriers. Among tens of identified proteins, we focused on the RPS26 and investigated its involvement in CGGexp-related RAN translation. Previously it was shown that RPS26 can interact with mRNA associated with PIC and regulate the translation of selected transcripts depending on their sequence, 5'UTR's length, and stress conditions (Ferretti *et al*, 2017; Havkin-Solomon *et al*, 2023; Li *et al*, 2022; Yang & Karbstein, 2022; Pisarev *et al*, 2008). By regulating RPS26 at the cellular level and its incorporation into the assembling 40S subunit mediated by escortin TSR2 (Schütz *et al*, 2014; Yang & Karbstein, 2022), we found that insufficiency of this ribosomal protein has negative effect on the level and toxicity of FMRpolyG with no impact on encoding mRNA and FMRP. As FMRP is the primary protein product of the *FMR1* gene, it indicates that RPS26 selectively modulates CGG-related RAN translation. Furthermore, by using a proteomic approach, we found that the number of proteins sensitive to RPS26 insufficiency was limited.

Results

Mass-spectrometry-based screening revealed numerous proteins interacting with the 5'UTR of FMR1 mRNA

To uncover new modifiers of CGGexp-related RNA toxicity, we used an RNA-tagging method coupled with mass spectrometry (MS) that allows for the in cellulo capture and identification of native RNA-protein complexes. The RNA bait, *FMR1 RNA* (used to pull down interacting proteins) consisted of the entire sequence of the 5'UTR of *FMR1*, which contained expanded, 99-times repeated, CGG tracts tagged with three times repeated MS2 RNA stem-loop aptamers (Figure 1A, Supplementary Figure 1A). These aptamers did not affect the RAN translation process as the synthesis of FMRpolyG was initiated from the near-cognate ACG start codon located upstream to CGGexp and terminated upstream to the MS2 aptamers (Supplementary Figure 1B). Given that the sequence of the 5'UTR of *FMR1* was guanosine and cytosine (GC)-rich (ca. 90% of GC content), we used an RNA sequence of the same length with GC content higher than 70%, *GC-rich RNA* (Figure 1A, Supplementary Figure 1A), as a control. Importantly, both of these sequences contained open reading frames similar in length but with differing start codons (ACG in *FMR1 RNA* and AUG in *GC-rich RNA*). Thus, they served as templates for protein synthesis (Supplementary Figure 1A). Together with RNA baits, the MS2 protein— showing high affinity towards RNA MS2 stem-loop aptamers—was co-expressed in HEK293T cells to immunoprecipitate natively formed RNA bait-protein complexes.

The MS-based screening identified over 150 proteins potentially interacting with *FMR1 RNA* (Supplementary Table 3). The most enriched Gene ontology (GO) terms indicated that the majority of proteins binding to *FMR1 RNA* had RNA binding properties and were involved in ribosome biogenesis, translation, and regulating mRNA metabolic processes (Supplementary Figure 1C, a complete list of GO terms is in the Supplementary Table 4).

While searching for novel RAN translation modifiers, we focused on proteins enriched on *FMR1 RNA*. We applied a label-free quantification analysis, which allowed us to elucidate 32 significantly enriched interactors of *FMR1 RNA* compared to *GC-rich RNA* (Figure 1B,

Supplementary Table 5). Next, using short interfering RNA (siRNA)-based silencing, we investigated the effect of insufficiency of eight preselected proteins on the level of toxic FMRpolyG, which could indicate their role in modulating RAN translation. These experiments were conducted in HEK293 cells transiently expressing FMRpolyG containing 99 glycine residues tagged with GFP, named FMR99xG (Figure 1C). Silencing of mRNA encoding for the ATP-dependent RNA helicase DHX15 and RPS26, but not six other analyzed proteins, resulted in a significant decrease in steady-state level of FMR99xG, with no effect on its mRNA (Figure 1C, Supplementary Figure 1D&E).

We also investigated whether RPS26 depletion affected the efficiency of FMRpolyG aggregates formation and cell toxicity. In HeLa cells transiently expressing FMR99xG, the frequency of GFP-positive aggregates was reduced upon RPS26 silencing (Figure 1D). Moreover, cells with depleted RPS26 exhibited significantly lower apoptosis tendencies evoked by toxic FMRpolyG (Figure 1E). These results suggest that decreasing level of RPS26 helps to alleviate FXTAS-related phenotype in cell models.

An orthogonal biochemical assay was used to confirm coprecipitation of RPS26 with the 5'UTR of *FMR1.* We applied an in vitro RNA-protein pull-down using three biotinylated RNAs: 5'UTR of *FMR1* with 99xCGG repeats, synthetic RNA with 23xCGG repeats, and GC-rich RNA as a control. All three RNAs were incubated with an extract derived from HEK293T cells. The first two RNAs, but not GC-rich RNA, were enriched with RPS26, however, the anticipated interaction between RNAs containing CGG repeats and RPS26 was not solely dependent on the triplet number, as RPS26 was pulled down with similar efficiency by RNAs containing more or fewer repeats (Supplementary Figure 1F).

In sum, we identified 32 proteins enriched on mutant *FMR1 RNA.* RPS26 and DHX15 insufficiency hindered FMR99xG RAN translation efficiency in a transient expression system. Notably, among proteins identified in the screening there were the ATP-dependent RNA helicase (DDX3X) — the protein previously described as RAN translation modifier (Linsalata *et al*, 2019), and the Src-associated in mitosis 68 kDa protein (SAM68)—the splicing factor sequestered on RNA containing CGGexp (Sellier *et al*, 2010) (Figure 1B, Supplementary Figure 1G). Identified

interactors can be involved not only in RAN translation, but potentially also in different metabolic processes of mutant *FMR1* RNA, such as transcription, protein sequestration, RNA transport, localization, and stability. Importantly, our data showed that silencing of RPS26 alleviated the pathogenic effect of toxic FMRpolyG produced from *FMR1* mRNA containing CGGexp. These facts encouraged us to further investigate the role of RPS26 in the context of CGG-exp-related RAN translation in more natural models.

RPS26 acts as a RAN translation modulator of mRNA with short and long CGG repeats

To monitor FMRpolyG RAN translation efficiency and test the modulatory properties of preselected proteins, we generated two cell lines stably expressing a fragment of an *FMR1* gene with expanded (95xCGG) and short, normal CGG repeats (16xCGG), named *S-95xCGG* and *S-16xCGG*, respectively. These models were generated by taking advantage of the Flp-In™ T-Rex™ 293 system (Szczesny *et al*, 2018), which expresses transgenes encoding for either longer (FMR95xG) or shorter (FMR16xG) polyglycine tract-containing proteins tagged with EGFP under the control of a doxycycline-inducible promoter (Figure 2A). In these cell lines, the expression of transgenes is detectable a few hours after promoter induction (Supplementary Figure 2A); however, even in prolonged doxycycline exposure (up to 35 days), we did not observe FMR95xG positive aggregates (Supplementary Figure 2B). A lack of aggregation is advantageous for monitoring RAN translation efficiency, as it allows the entire pool of RAN proteins present in cellular lysate to be measured, which is impossible if part of the proteins is trapped in aggregates (Derbis *et al*, 2021, 2018). Moreover, single copy integration of the transgene containing a fragment of the *FMR1* gene mimics natural situation. Notably, the level of transgene expression was comparable to *FMR1* endogenous expression (Supplementary Figure 2C) and was homogeneous between cells (Figure 2A, Supplementary Figure 2B). The second model, named *L-99xCGG*, was generated using lentiviral transduction of HEK293T cells (Figure 2D). Lentiviral particles were prepared based on genetic construct used in the transient transfection system (Figure 1C). This stable cell line constantly expresses FMR99xG tagged with GFP. In the *L-99xCGG* model, FMR99xG was present in soluble and, to a lesser extent, aggregated form (Figure 2D).

The RPS26 is a component of the 40S ribosomal subunit and its inclusion or depletion from this subunit affects translation of selected mRNA but does not impact overall translation rate (Schütz *et al*, 2018; Ferretti *et al*, 2017; Havkin-Solomon *et al*, 2023; Yang & Karbstein, 2022; Gaikwad *et al*, 2021). For instance, in yeast cultured in stress conditions, Rps26 is disassociated from 40S subunit, which results in translation of different mRNA, especially those encoding proteins implicated in stress response pathway (Ferretti *et al*, 2017; Yang & Karbstein, 2022). Moreover, it was shown that RPS26 is involved in translational regulation of selected mRNA contributing to the maintenance of pluripotency in murine embryonic stem cells (Li *et al*, 2022). Its C-terminal domain interacts with mRNA sequences upstream to the E-site of an actively translating ribosome (Pisarev *et al*, 2008; Hussain *et al*, 2014; Anger *et al*, 2013). This localization may be responsible for the regulation of efficient translation initiation (Pisarev *et al*, 2008; Havkin-Solomon *et al*, 2023), especially in a context of non-canonical RAN translation. Hence, these facts encouraged us to further investigate this protein in the context of CGGexp-related RAN translation.

siRNA-induced silencing of RPS26 in the stable *S-95xCGG* cell model expressing mRNA with long 95xCGG repeats resulted in a significantly decreased level of steady-state FMR95xG, with no effect on its mRNA (Figure 2B). Moreover, the RPS26 silencing in the *S-16xCGG* model also decreased the level of RAN protein product derived from short 16xCGG repeats without affecting its mRNA level (Figure 2C). This suggests that RPS26 depletion affects RAN protein levels independently of CGG repeat content. Similarly to the *S-95/16xCGG* models, RPS26 silencing in the *L-99xCGG* model also significantly downregulated FMR99xG biosynthesis (Figure 2E).

Previously, several proteins known as RPS26 responders, e.g., murine Polycomb protein (Suz12) and Histone H3.3, were shown to be negatively affected by RPS26 depletion (Li *et al*, 2022). Here, we showed that they also negatively responded to the RPS26-specific siRNAs used in this study (Supplementary Figure 2D).

We further investigated whether the translation of FMRP might be affected by RPS26 depletion, as an open reading frame for FMRpolyG–which appears to be under the control of RPS26-sensitive translation–is derived from the same mRNA. To test this hypothesis, we treated

human fibroblasts derived from a FXTAS patient and a healthy control with RPS26-specific siRNA. Our results indicate that RPS26 depletion does not affect FMRP levels, regardless of CGG repeat length in *FMR1* mRNA (Figure 2F). Similarly, the FMRP level was not affected by RPS26 depletion in other cell models (Figure 2E).

Together, these results imply that although the presence of RPS26 in 40S subunit has a positive effect on RAN translation of FMRpolyG and other previously identified RPS26 responders, it does not affect the canonical translation of FMRP produced from the same mRNA. Moreover, a lack of significant differences in translation efficiency of long and short polyglycine-containing proteins in cells with insufficiency of RPS26 suggests that observed effect may depend on certain RNA sequences or structures within the 5'UTR of *FMR1* mRNA or other features of this mRNA, rather than the CGG repeat length.

RPS26 depletion affects only a small subset of the human proteome

To assess globally which proteins are sensitive to RPS26 insufficiency, we used stable isotope labeling with amino acids in cell culture (SILAC). The protein expression level in control and RPS26-depleted HEK293T cells was determined via quantitative mass spectrometry (SILAC-MS, Supplementary Table 6). Differential data analysis indicated that most (ca. 80%) proteins identified by SILAC-MS were not sensitive to RPS26 deficiency; we named these proteins non-responders (N = 1,506). We also identified a set of proteins that negatively (negative responders; N = 223 if the P-value was < 0.05) or positively responded (positive responders; N = 158 if the P-value < 0.05) to RPS26 deficiency (Figure 3A, Supplementary Table 7). Non-responders were used as a background list in the GO analysis. An analysis of positive responders found that the proteins in this group were mainly components of translation initiation complexes or formed large ribosomal subunits (Figure 3B, Supplementary Table 8). A similar analysis for negative responders did not reveal any significantly enriched GO terms (Supplementary Table 8), partially due to the small number of identified proteins. To further validate this data, we performed western blots for selected negative responders (PDCD4, ILF3, RPS6, and PCBP2), no-responders (FMRP and FUS) or positive responders (EIF5 and EIF3J) from independent cell samples collected 48h post siRPS26

treatment (Figure 3C, Supplementary Figure 3A). Most tested proteins (6 out of 8) aligned with the quantitative SILAC-MS data (Figure 3C, Supplementary Figure 3A).

Previously published results found that specialized ribosomes are formed to selectively translate mRNAs with specific features (Genuth & Barna, 2018; Shi *et al*, 2017). In yeast, mRNAs translated by Rps26-depleted ribosomes lack conservation of all Kozak sequence elements, while mRNAs translated by Rps26-containing ribosomes present a full Kozak consensus (Ferretti *et al*, 2017). RPS26 is localized next to the E-site of the translating ribosome and was found to interact with template mRNAs (Pisarev *et al*, 2008; Anger *et al*, 2013). For instance, if the 43S recognizes the start codon, the RPS26 contacts position −4 from the start codon in yeast (Ferretti *et al*, 2017) and from −11 up to −16 of attached mRNA in mammals (Havkin-Solomon *et al*, 2023). Given that RPS26 may be involved in start codon fidelity (Ferretti *et al*, 2017; Havkin-Solomon *et al*, 2023), we searched for specific sequence motifs in mRNAs encoding proteins responding to RPS26 deficiency. We investigated sequences containing 5'UTRs and coding sequences (CDSs) close to the start codon of mRNAs encoding positive and negative RPS26 responders. We used the total human transcriptome, named background (BG; N = 22,160), as a reference. We did not observe any significant differences in the frequency of individual nucleotide positions in the 20-nucleotide vicinity of the start codon relative to the expected distribution in the BG (Supplementary Figure 3B, Supplementary Table 9). However, we identified a significantly higher GC content in the 5'UTR sequences in the positive and negative responder groups relative to the BG at all analyzed positions from −6 to −30 (Figure 3D). Although the difference in the GC content in CDSs of responders relative to the BG was much smaller than for the 5'UTRs, we observed a significantly higher ($P < 0.05$) GC content in the close vicinity of the start AUG codon at upstream positions from +9 up to +14 (Figure 3D). Moreover, when we applied more stringent selection criteria ($P < 0.01$), the number of analyzed transcripts was significantly reduced (positive responders; N = 42, negative responders; N = 54), but the GC richness appeared to be more predominant for mRNAs encoding negative RPS26 responders (Supplementary Figure 3C).

Using bioinformatic analyses, we did not find any importance of position −4 from the start codon in any group of RPS26 responders (Supplementary Figure 3B). However, we wanted to

experimentally check whether changes in this position in human cells would affect RAN translation initiation. We substituted G to A in the –4 position from ACG near cognate codon of *FMR1*, as A has previously been shown to be an enhancer of translation initiation in yeast 43S containing Rps26 (Ferretti *et al*, 2017, 2018). We did not observe significant differences in efficiency of FMRpolyG biosynthesis between the two tested mRNAs in human HEK293 cells (Supplementary Figure 3D).

We also searched for specific, short sequence motifs, *k*-mers, which might have been enriched in the 5'UTR of mRNAs encoding responders to RPS26 deficiency. We identified a list of 6 and 14 significantly over-represented hexamers ($P < 10^{-5}$) in positive and negative responder groups, respectively. These predominantly comprised G(s) and C(s). For example, the most over-represented hexamers in the 5'UTRs of the positive responders ($P < 10^{-10}$) included GCCGCC, CCGCTG, and CCGGTC, and for negative responders, the highest over-representation ($P < 10^{-6}$) had CGCCGC, GCCGCC, and GCGGCG (Supplementary Figure 3E, Supplementary Table 10).

Altogether, these data indicate that RPS26-depleted ribosomes may demonstrate specificity towards mRNAs containing GC-rich sequences in 5'UTRs and in close proximity downstream to the AUG codon (up to 14 nucleotides). It may suggest that thermodynamic stability of RNA structure formed upstream or downstream of a scanning 43S PIC and perhaps the dynamics of translocation of this complex could be a factor modulating sensitivity to RPS26 insufficiency. We did not see any importance of A or G in –4 position from ACG initiation codon on FMRpolyG biosynthesis by ribosomes having or lacking RPS26. Our data showed also that the translation of most human proteins, including FMRP, was unaffected by the impairment of RPS26 level.

TSR2 mediates the RAN translation of FMRpolyG, perhaps via RPS26

The pre-rRNA-processing protein TSR2 (Tsr2) is a chaperone-acting protein that regulates the Rps26 cellular level and is responsible for incorporating Rps26 into 90S pre-ribosome in yeast nuclei (Schütz *et al*, 2018, 2014). RPS26 inclusion into pre-40S together with other factors facilitates final step of 40S subunit maturation (Plassart *et al*, 2021). Moreover, Tsr2 mediates the

disassembly of Rps26 from mature 40S subunit in cytoplasm in high salt and pH conditions (Yang & Karbstein, 2022) or when Rps26 is oxidized (Yang *et al*, 2023).

Assuming that depletion of TSR2 in mammalian cells may affect the level of RPS26 loaded on 40S in human cells, we hypothesized that silencing TSR2 would affect the biosynthesis of FMRpolyG. Indeed, in stable *S-95xCGG* cell line treated with siRNA against TSR2 (siTSR2) we observed lower level of both RPS26 and FMR95xG but the level of FMRP was unchanged (Figure 4A). The effect of TSR2 silencing on RAN translation was independent of CGG repeat length, as similar results were obtained for smaller FMR16xG proteins (Supplementary Figure 4). Silencing TSR2 did not affect the level of other components of the 40S subunits such as RACK1, RPS6 and RPS15, however, the level of a known responder of RPS26 insufficiency, Histone H3.3 (Li *et al*, 2022), was significantly reduced (Figure 4A&B).

This could be explained by the fact that insufficiency of TSR2 may affect incorporation of RPS26 during nuclear maturation (Schütz *et al*, 2018) or cytoplasmic regeneration of 40S (Yang *et al*, 2023), thus having an effect on FMRpolyG biosynthesis. Overall, this data strengthens our conclusion regarding the positive effect of RPS26 and TSR2 on RAN translation initiated from the near-cognate ACG or GUG codons located in the 5'UTR of *FMR1*.

The other 40S subunit component, RPS25, affects CGGexp-related RAN translation

Considering data concerning heterogenous ribosomes and their diverse roles in translation regulation (Genuth & Barna, 2018; Shi *et al*, 2017), we hypothesized that other component of the 40S subunit, the ribosomal protein RPS25, which localizes near RPS26 on 40S structure (Pisarev *et al*, 2008; Anger *et al*, 2013), may regulate CGGexp-related RAN translation. Importantly, this protein was already identified as a modifier of RAN translation of mRNAs containing other types of expanded repeats–GGGGCCexp and CAGexp (Yamada *et al*, 2019).

Upon silencing of RPS25 in the stable *S-95xCGG* model and human neuroblastoma SH-SY5Y cells with transient expression of FMRx99G, we observed a decline of FMRpolyG level with no effect on its encoding mRNAs (Figure 5A&B). The level of FMR99xG reduction was comparable after RPS25 and RPS26 depletion (Figure 5B). We also found that RPS25 coprecipitates with the

5'UTR of *FMR1* that lacks or harbors expanded 99xCGG repeats via biochemical assay using the biotinylated RNA-protein pull-down assay (Supplementary Figure 5). Altogether, these results suggest that insufficiency of RPS25, like RPS26, is a factor which negatively affects CGGexp-related RAN translation in polyglycine frame.

Discussion

RAN translation was first described in 2011 and reported for Spinocerebellar Ataxia type 8 (SCA8) linked to CAG triplet expansion (Zu *et al*, 2011) and in other repeat expansion-related disorders (REDs), such as Huntington's disease (HD) (Bañez-Coronel *et al*, 2015), C9orf72-linked amyotrophic lateral sclerosis (C9-ALS), and frontotemporal dementia (FTD) (Mori *et al*, 2013; Ash *et al*, 2013), which correlate to CAG and GGGGCC repeats expansions, respectively. RAN translation contribute to the development and progression of many REDs (Zu *et al*, 2011; Mori *et al*, 2013; Ash *et al*, 2013; Bañez-Coronel *et al*, 2015), including FXTAS (Todd *et al*, 2013; Sellier *et al*, 2017) and FXPOI (Buijsen *et al*, 2016); however, no effective therapy targets this pathomechanism. Currently, some RAN translation modifiers of *FMR1* mRNA have been described. For instance, RNA helicases such as DDX3X (Linsalata *et al*, 2019) or DHX36 (Tseng *et al*, 2021), and others extensively reviewed in Baud et al. (Baud *et al*, 2022)) have been identified as affecting CGG repeats-related RAN translation by facilitating ribosomal scanning via unwinding the structured RNA. Components of the endoplasmic reticulum ER-resident kinase (PERK) signaling pathway were shown to modulate RAN translation under stress conditions (Green *et al*, 2017), and the activity of another kinase, SRPK1, retains mutated *FMR1* mRNA containing CGGexp in the nucleus blocking its transport to the cytoplasm and subsequent translation (Malik *et al*, 2021a). Recently, it was demonstrated that RAN translation activates ribosome-associated quality control (RQC) pathway, which prevents accumulation of RAN misfolded proteins by ubiquitination and subsequent proteasomal degradation (Tseng *et al*, 2024). Nevertheless, mechanistic insights into RAN translation remain elusive.

Our research aimed to reveal proteins coprecipitating with the 5'UTR of *FMR1* mRNA containing expanded CGG repeats and discover novel modifiers of CGGexp-related RAN translation. The RNA-tagging approach allowed us to identify over a hundred and fifty proteins—

including factors with translation regulatory properties. Some identified proteins overlapped with ones already described as binding to mutant *FMR1* mRNA such as SRPK1 or TAR DNA-binding protein 43 (TDP-43) (Malik *et al*, 2021b; Rosario *et al*, 2022; Sellier *et al*, 2010; Jin *et al*, 2007; Cid-Samper *et al*, 2018; Sellier *et al*, 2017). Further, we described three new modifiers of RAN translation of this mRNA. Depletion of DHX15 helicase and two components of the 40S subunit, RPS26 and RPS25, significantly decreased FMRpolyG levels (Figure 1C & Figure 5A). Importantly, silencing RPS26 alleviated the aggregation phenotype and slowed the apoptosis process caused by FMRpolyG expression (Figure 1D&E).

In yeast, the presence or absence of Rps26 in the ribosomal structure leads to the expression of different protein pools influencing temporal protein homeostasis in response to environmental stimuli (Ferretti *et al*, 2017; Yang & Karbstein, 2022). For instance, high-salt and high-pH stress induce the release of Rps26 from mature ribosomes by its chaperone Tsr2, enabling the translation of mRNAs engaged in stress response pathways (Yang & Karbstein, 2022). A previous study indicated that RAN translation in human cells is selectively enhanced by activating stress pathways in a feed-forward loop (Green *et al*, 2017). Therefore, differences in ribosome composition may influence FMRpolyG production under certain stress conditions.

In eukaryotic cells, RPS26 is involved in stress responses, however on the contrary to yeast model, RPS26 remains associated to the ribosome under energy stress (Havkin-Solomon *et al*, 2023). It was demonstrated that cells with mutated C-terminus of RPS26 were more resistant to glucose starvation, than the wild type cells (Havkin-Solomon *et al*, 2023). Moreover, RPS26 was shown to be involved in other cellular processes such as the DNA damage response (Cui *et al*, 2013), activation of the mTOR signaling pathway (Havkin-Solomon *et al*, 2023) and cellular lineage differentiation by preferential translation of certain transcripts (Piantanida *et al*, 2022; Li *et al*, 2022). We evaluated global changes in the human cellular proteome under RPS26 depletion and found that the expression level was significantly changed in approximately 20% of human proteins, while most proteins' expression levels remained intact (Figure 3A, Supplementary List). This suggests that many proteins, including FMRP, are not negatively affected by depleting RPS26 from the ribosomal machinery, although FMRpolyG biosynthesis appears to be RPS26-sensitive.

We also verified whether silencing RPS26 would affect FMRP endogenously expressed in FXTAS patient-derived and control fibroblasts, as they differ in CGG repeats content in natural locus of *FMR1,* which potentially might influence the effect of RPS26 deficiency. We did not observe any changes in FMRP levels upon RPS26 silencing in either genetic variant with the PM or normal *FMR1* allele (Figure 2F).

Given the role of TSR2 in incorporating or replacing RPS26 into the ribosome, we investigated whether the depletion of TSR2 modulates FMRpolyG production. In line with previous studies (Schütz *et al*, 2014), we demonstrated that the RPS26 level decreases upon TSR2 silencing. We subsequently showed that TSR2 silencing incurs the same effect on RAN translation efficiency as the direct depletion of RPS26 (Figure 4A). However, it remains unclear whether this is an effect of hampering RPS26 loading to the 40S subunit by TSR2 depletion, or a consequence of decrease of RPS26 level. The silencing of this chaperon also exhibited selectivity towards FMRpolyG over the FMRP reading frame without affecting the other 40S ribosomal proteins, such as RACK1, RPS6, and RPS15 (Figure 4A&B). Previously, it was shown that Tsr2 is responsible for replacing damaged Rps26, which undergoes oxidation on mature 80S ribosomes under stress conditions (Yang *et al*, 2023). Therefore, the activity of this chaperon may play an important role in modulating RAN translation *via* Rps26 assembly or disassembly from ribosomal subunits, also during different stresses (Yang & Karbstein, 2022). Moreover, mutations in genes encoding RPS26 and TSR2 were associated with hematopoiesis impairment that underlies the genetic blood disorder Diamond-Blackfan anemia (DBA) (Piantanida *et al*, 2022; Li *et al*, 2023; Doherty *et al*, 2010). Previously, analyses of DBA patient cells and RPS26-depleted HeLa cells found alterations in ribosome biogenesis and pre-rRNA processing (Doherty *et al*, 2010). Similarly, we found that proteins implicated in the eukaryotic PIC and components of large ribosomal subunits were altered upon RPS26 silencing in HEK293 cells (Figure 3B), suggesting rearrangements in ribosomal composition. Moreover, RPS26 is highly expressed in ovarian cells and was shown to be important for female fertility (Liu *et al*, 2018). It is necessary for oocyte growth and follicle development, and its depletion causes changes in transcription and chromatin configuration, leading to premature ovarian failure (Liu *et al*, 2018). Our discovery that RPS26 regulates the level of FMRpolyG in cellular models sheds new light on the potential role of RPS26-related RAN

translation in FXPOI, where FMRpolyG aggregates are detected in ovarian cells and contribute to the development and progression of fertility issues (Buijsen *et al*, 2016; Shelly *et al*, 2021; Rosario *et al*, 2022).

Given these findings, it is likely that the depletion of RPS26 from 40S subunits contributes to its specialization and modulates the translation of specific mRNAs (Yang & Karbstein, 2022; Ferretti *et al*, 2017; Li *et al*, 2022; Gaikwad *et al*, 2021). Remarkably, our experiments identified nearly 400 proteins responding to RPS26 depletion, which is similar to 488 mRNAs, for which translation rate was altered upon Rps26 insufficiency in yeast (Gaikwad *et al*, 2021). In fact, depletion or mutations in RPS26 resulted in the reduction of 40S subunits level, (Gaikwad *et al*, 2021; Havkin-Solomon *et al*, 2023), however, overall translation rate was not impacted by RPS26 impairment (Havkin-Solomon *et al*, 2023). It has been shown that the C-terminal domain of RPS26 is essential for mRNA interaction (Havkin-Solomon *et al*, 2023), although whether RPS26 recognizes specific sequential or structural motifs within the mRNA remains unclear. Data derived from yeast models suggest that nucleotides in positions −1 to −10 upstream of the AUG codon, especially Kozak sequence elements, play an important role in Rps26 interactions and translation efficiency (Ferretti *et al*, 2017). According to the established ribosome-mRNA structure, RPS26 contacts with mRNA upstream to AUG codon and the C-terminus reaches into mRNA exit channel (Anger *et al*, 2013; Hussain *et al*, 2014). Recent findings indicate that in eukaryotic cells, positions −11 to −16 upstream of the start codon might be more significant for RNA recognition, stabilizing PIC, and translation initiation (Havkin-Solomon *et al*, 2023). Our bioinformatic analysis of proteins sensitive to RPS26 depletion revealed that their transcripts were GC-rich (especially in the 5'UTR region) and enriched with *k*-mers mainly consisting of Gs and Cs (Figure 3D, Supplementary Figure 3E), however we did not identify the importance of any specific sequence positions from AUG codon. These data and the fact that *FMR1* 5'UTR is a GC-rich sequence (~90%) suggest that RPS26-sensitive translation is selective for transcripts rich with GC nucleotides. This may point to the importance of specific RNA structural motifs localized within 5'UTRs or the speed of PIC scanning, which depends on how effectively stable secondary/tertiary structures are resolved. Other explanation would be differences in dynamics

of either scanning of PICs differing in the presence of RPS26 or assembling of 80S on the initiation codon.

The ribosomal protein RPS25, a component of 40S subunit was previously described in the context of GGGGCC and CAG repeats-related RAN translation corresponding to C9-ALS/FTD, HD, and SCA (Yamada *et al*, 2019). The depletion of RPS25 in the *Drosophila* C9orf27 model (and in induced motor neurons derived from C9-ALS/FTD patients) alleviated toxicity caused by RAN translation (Yamada *et al*, 2019). On the contrary, in FXTAS *Drosophila* model, RPS25 knockdown led to enhancement of CGG repeats-related toxicity; however, underlying mechanism was not determined (Linsalata *et al*, 2019). Here, we demonstrated that RPS25 coprecipitated with the 5'UTR of *FMR1* and its depletion negatively affected the biosynthesis of FMRpolyG (Figure 5A&B, Supplementary Figure 5), thus expanding current knowledge about RPS25 RAN translation modulatory properties in REDs.

Altogether, we have identified two ribosomal proteins, RPS26 and RPS25 as CGGexp-related RAN translation modifiers, which imply that the rearrangements within 40S subunit affects FMRpolyG biosynthesis. Importantly, we demonstrated that RPS26 depletion alleviated toxicity caused by FMRpolyG but did not affect FMRP, the main product of the *FMR1* gene. This suggests that sequence/structure elements within *FMR1* mRNA, which make this transcript sensitive to studied ribosomal proteins, may be potential therapeutic targets.

Materials and Methods

Genetic constructs

For MS-based screening *FMR1 RNA* construct was generated based on backbone described in (Sellier *et al*, 2017) (see also Addgene plasmid #63091), which contains 5'UTR of *FMR1* with expanded 99xCGG repeats. Enhanced green fluorescent protein (EGFP) sequence was removed and replaced with 4 STOP codons in FMRpolyG frame. Three times repeated MS2 stem loop aptamers (3xMS2 stem loops) were amplified by Phusion polymerase (Thermo Fisher Scientific)

with primers introducing EagI restriction site from the plasmid (Addgene, #35572, Tsai et al., 2011 (Tsai *et al*, 2011)). After gel-purification, PCR product was digested with EagI (New England Biolab) and inserted into EagI-digested and dephosphorylated backbone (CIAP; Thermo Fisher Scientific) downstream of 5'UTR of *FMR1* using T4 ligase (Thermo Fisher Scientific). To generate *GC-rich RNA* construct, GC rich sequence (corresponding to *TMEM170* mRNA, Supplementary file - sequences) was amplified by PCR introducing EagI restriction site (CloneAmp HiFi PCR Premix, TakaraBio). PCR product was digested by EagI (New England Biolabs) and ligated into *FMR1 RNA* construct backbone instead of 5'UTR of *FMR1* sequence using T4 ligase (Thermo Fisher Scientific). FMRpolyG-GFP (named here FMR99xG) transient expression was derived from 5'UTR CGG 99x FMR1-EGFP vector (Adggene plasmid #63091), a kind gift from N. Charlet-Berguerand. *FMR1 RNA* and *GC-rich RNA* constructs as well as content of CGG repeats were verified by Sanger sequencing.

Generation of cell lines stably expressing FMRpolyG

To generate cell lines with doxycycline-inducible, stable FMRpolyG expression we used the Flp-In™ T-REx™ system. Through this approach, we obtained integration of constructs containing 5'UTR FMR1 with either 95 or 16 CGG repeats fused to *EGFP* into genome of 293 Flp-In® T-Rex® cells. These cell lines were named *S-95xCGG* and *S-16xCGG,* respectively, and expressed RAN proteins referred to as FMR95xG or FMR16xG. A detailed protocol concerning experimental procedure was described previously by Szczesny et al., 2018 (Szczesny *et al*, 2018). The constructs used in the procedure were generated by modyfing pKK-RNAi-nucCHERRYmiR-TEV-EGFP (Addgene plasmid #105814). The insert containing CGG repeats within the 5'UTR of FMR1 gene fused with *EGFP* sequence was derived from 5'UTR CGG 99x FMR1-EGFP (Addgene plasmid #63091). The insert was ligated into the vector instead of *EGFP* sequence. In the second approach, we used lentiviral transduction system using constructs containing 5'UTR of *FMR1* with 99 CGG repeats fused to EGFP followed by T2A autocleavage peptide and puromycin resistance (FMRpolyG-GFP_T2A_PURO). Cloning FMRpolyG-GFP_T2A_PURO sequence into TetO-FUW-AsclI-Puro vector (Addgene plasmid #97329, Yang et al., 2017 (Yang *et al*, 2017)) as well as lentiviral particles production was performed by The Viral Core Facility, part of the Charité – Universitätsmedizin Berlin. For transduction procedure $0.15x10^6$ HEK293T cells were plated in 6-

well plates at ~60% confluency and incubated with lentiviral particles for 48h. Subsequently, in order to obtain the pool of cells with integrated transgene, selection with puromycin (in final concentration 1 μg/ml, Sigma) was initiated and lasted for 72h, until all transgene-negative cells died. After selection we obtained polyclonal cell line named *L-99xCGG* expressing FMR99xG.

Cell culture and transfection

The monkey COS7, human HEK293T, HeLa and SH-SY5Y cells were grown in a high glucose DMEM medium with L-Glutamine (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; Thermo Fisher Scientific) and 1x antibiotic/antimycotic solution (Sigma). *S-95xCGG* and *S-16xCGG* cells were grown in DMEM medium containing certified tetracycline-free FBS (Biowest). Fibroblasts derived from FXTAS male patient (1022-07) with $(CGG)_{81}$ in *FMR1* gene, and control, non-FXTAS male individual (C0603) with $(CGG)_{31}$ in *FMR1* were cultured in MEM medium (Biowest) supplemented with 10% FBS (Thermo Fisher Scientific), 1% MEM non-essential amino acids (Thermo Fisher Scientific) and 1x antibiotic/antimycotic solution (Sigma). All cells were grown at 37°C in a humidified incubator containing 5% CO_2. FXTAS 1022-07 line was a kind gift from P. Hagerman (Garcia-Arocena *et al*, 2010) while control C0603 fibroblast lines were given by A. Bhattacharyya (Rovozzo *et al*, 2016). For the delivery of siRNA with the final concentration in culture medium ranging from 15 to 25 nM, reverse transfection protocol was applied using jetPRIME® reagent (Polyplus) with the exception of fibroblasts, which were plated on the appropriate cell culture vessels the day before the transfection. To deliver plasmids, the DNA/jetPRIME® reagent (Polyplus) ratio 1:2 was applied. Cells were harvested 48h post siRNA silencing and 24h post transient plasmids expression. The list of all siRNAs used in the study is available in a Supplementary Table 1.

Mass spectrometry-based proteins screening

To capture RNA-protein complexes natively assembled within *FMR1 RNA* we adapted the MS2 *in vivo* biotin tagged RNA affinity purification (MS2-BioTRAP) technique, originally published by Tsai et al., 2011 (Tsai *et al*, 2011). The principle of this technique is to co-express bacteriophage MS2 protein fused to a HB tag which undergoes biotinylation *in vivo*, and RNAs tagged with so-called

MS2 stem loop RNA aptamers, towards which MS2-HB protein represent high affinity. Natively assembled RNA-protein complexes can be then fixed and identified with HB-tag based affinity purification using streptavidin-conjugated beads. To perform the screening, $2x10^6$ HEK293T cells were co-transfected with genetic constructs: 2 μg of MS2-HB plasmid (#35573, Tsai et al., 2011 (Tsai *et al*, 2011)) along with 8 μg of *FMR1 RNA* or *GC-rich RNA* encoding vectors (described in detail above). Three 10 cm plates were used per given mRNA with MS2 stem loop aptamers. 24h post co-transfections cells were washed 2 times with ice-cold phosphate buffered saline (PBS) and crosslinked with 0.1% formaldehyde (ChIP-grade, Pierce) for 10 min, followed by 0.5 M glycine quenching for 10 min in room temperature (RT). After 1 wash with ice-cold PBS cells were lysed for 30 min on ice in cell lysis buffer (50 mM Tris-Cl pH 7.5, 150 mM NaCl, 1% Triton X-100, 0.1% Na-deoxycholate) with Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific) and RNAsin Plus (Promega) and vortexed. Next, cells lysates were sonicated (15 cycles: 10 sec. on/10 sec. off) using sonicator (Bioruptor, Diagenode) and centrifuged at 12,000 g for 10 min at 4°C. Precleared protein extracts were transferred to Protein Lobind tubes (Eppendorf) and incubated with pre-washed 50 μl of magnetic-streptavidin beads (MyOne C1, Sigma) for 1.5h in cold room rotating. After immunoprecipitation (IP) procedure beads were washed 4-times for 5 min each on rotator: 1-time with 2% sodium dodecyl sulfate (SDS), 1-time with cell lysis buffer, 1-time with 500 mM NaCl, and last wash with 50 mM Tris-Cl pH 7.5. Finally, 20% of IP fraction was saved for western blot and remaining beads were submerged into digestion buffer (6 M Urea, 2 M Thiourea, 100 mM Tris, pH 7.8 and 2% amidosulfobetaine-14) and were shaken for 1h at RT. Then, samples were reduced using dithioerythritol for 1h at RT, and alkylated with iodoacetamide solution for 30 min in dark. Then, Trypsin/Lys-C (Promega) solution was added, and samples were incubated for 3h at 37°C, followed by adding fresh Milli-Q water to dilute Urea to ~1 M and samples were further incubated at 37°C overnight. Finally, beads were removed on a magnet, and peptides transferred to the clean tube, and desalted using C18 Isolute SPE columns (Biotage). Samples were analyzed in Mass Spectrometry Laboratory, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a Street, 02-106 Warsaw, Poland, using LC-MS system composed of Evosep One (Evosep Biosystems, Odense, Denmark) directly coupled to a Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific, USA). Peptides were loaded onto

disposable Evotips C18 trap columns (Evosep Biosystems, Odense, Denmark) according to manufacturer protocol with some modifications. Briefly, Evotips were activated with 25 µl of Evosep solvent B (0.1% formic acid in acetonitrile, Thermo Fisher Scientific, USA), followed by 2 min incubation in 2-propanol (Thermo Fisher Scientific, USA) and equilibration with 25 µl of solvent A (0.1% FA in H_2O, Thermo Fisher Scientific, Waltham, Massachusetts, USA) Chromatography was carried out at a flow rate 220 nL/min using the 88 min gradient on EV1106 analytical column (Dr Maisch C18 AQ, 1.9 µm beads, 150 µm ID, 15 cm long, Evosep Biosystems, Odense, Denmark). Data was acquired in positive mode with a data-dependent method. MS1 resolution was set at 60,000 with a normalized AGC target 300%, Auto maximum inject time and a scan range of 350 to 1,400 m/z. For MS2, resolution was set at 15,000 with a Standard normalized AGC target, Auto maximum inject time and top 40 precursors within an isolation window of 1.6 m/z considered for MS/MS analysis. Dynamic exclusion was set at 20 s with allowed mass tolerance of ±10 ppm and the precursor intensity threshold at 5e3. Precursor were fragmented in HCD mode with normalized collision energy of 30%.

Biotinylated RNA-protein pull down

Biotinylated RNA probes were produced by *in vitro* transcription using T7 RNA polymerase (Promega), and by adding 1:10 CTP-biotin analog:CTP to the reaction mix, to incorporate biotinylated cytidines in random manner. Cellular extracts were prepared by lysing $2.5x10^6$ HEK273T cells in 200 µL of mammalian cell lysis buffer. Lysates were cleared by centrifugation and 200 µL of supernatant was incubated for 20 min at 21°C with 5 µg of RNA in 200 µL of 2xTENT buffer (100 mM Tris pH 7.8, 2 mM EDTA, 500 mM NaCl, 0.1% Tween) supplemented with RNAsin (Promega). RNA-protein complexes were then incubated with MyOne Streptavidin T1 DynaBeads (Thermo Fisher Scientific) for 20 min, followed by washing steps in 1x TENT buffer. Bound proteins were released in Bolt LDS Sample Buffer (Thermo Fisher Scientific) followed by heat denaturation at 95°C for 10 min. Samples were analyzed by SDS-PAGE and western blotting.

Apoptosis assay

For luminescent-based apoptosis assay $0.01x10^6$ HeLa cells were seeded on 96-well plate and transfected with siRPS26 and siCtrl in final concentration of 15 nmol. In order to induce FMRpolyG derived toxicity, after 24h post silencing, cells well transfected with plasmid encoding FMR99xG or with jetPRIME® reagent (Polyplus) as a MOCK control. Subsequently, reagents from RealTime-Glo™ Annexin V Apoptosis Assay (Promega) were added to the culture. Measurement of luminescence signal corresponding to apoptosis progression were taken upon 28 and 43h post FMR99xG expression using SPARK microplate reader (TECAN).

Microscopic analysis

To detect aggregated form of FMRpolyG fluorescence microscopy experiments were performed as described previously (Derbis *et al*, 2018). Briefly, before analysis, HeLa cells were incubated in standard growth medium with final concentration of 5 μg/ml of Hoechst 33342 (Thermo Fisher Scientific) for 10 min. Images were taken with Axio Observer.Z1 inverted microscope equipped with A-Plan 10×/0.25 Ph1 or LD Plan-Neofluar 20×/0.4 Ph2 objective (Zeiss), Zeiss Colibri 7 excitation band 385/30 nm, emission filter 425/30 nm (Hoechst) and Zeiss Colibri 7 excitation band 469/38 nm, emission filter 514/30 nm (GFP), Zeiss AxioCam 506 camera and ZEN 2.6 pro software, 48h post siRNA transfection. Presented values were quantified from 10 images, number of cells and aggregates were calculated using ImageJ and AggreCount plugin (Klickstein *et al*, 2020). To validate FMR95xG or FMR16xG expression in *S-95xCGG* and *S-16xCGG* models, microscopic analysis was performed at two time points: 4 days and 35 days post doxycycline induction. 4 days post doxycycline induction cells were incubated in a cell culture medium with Hoechst 33342 (Thermo Fisher Scientific) at a final concentration 5 μg/ml for 30 min at 37°C. Images were taken as described above. 35 days post doxycycline induction images were captured with Leica Stellaris 8 Inverted Confocal Microscope equipped with HC PL APO CS2 63x/1.20 water objective and an onstage incubation chamber controlling temperature and CO_2 concentration. GFP was excited with 489 nm laser and detected with Power HyD S detector (spectral positions: 494 nm - 584 nm).

RNA isolation and quantitative real-time RT-PCR

Cells were harvested in TRI Reagent (Thermo Fisher Scientific) and total RNA was isolated with Total RNA Zol-Out D (A&A Biotechnology) kit according to the manufacturer's protocol. 500-1,000 ng of RNA was reversely transcribed using GoScript™ Reverse Transcriptase (Promega) and random hexamers (Promega). Quantitative real-time RT-PCRs were performed in a QuantStudio 7 Flex System (Thermo Fisher Scientific) using Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific) with 5 ng of cDNA in each reaction. Transgene *FMR1-GFP* mRNAs with expanded CGG repeats were amplified with primers: Forward: 5' GCAGCCCACCTCTCGGGG 3', Reverse: 5' CTTCGGGCATGGCGGACTTG 3' with a note that reverse primer was anchored in GFP sequence in order to distinguish endogenously expressed *FMR1* transcripts (amplified with primer pair: Forward: 5' TGTGTCCCCATTGTAAGCAA 3', Reverse: 5' CTCAACGGGAGATAAGCAG 3'). Reactions were run at 58°C annealing temperature and Ct values were normalized to *GAPDH* mRNA level (amplified with primer pair: Forward: 5' GAGTCAACGGATTTGGTCGT 3', Reverse: 5' TTGATTTTGGAGGGATCTCG 3'). Fold differences in expression level were calculated according to the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001).

SDS-PAGE and Western blot

Cells were lysed in cell lysis buffer supplemented with Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific) for 30 min on ice, vortexed, sonicated and centrifuged at 12,000 g for 10 min at 4°C. Protein lysates were heat-denatured for 10 min at 95°C with the addition of Bolt LDS buffer (Thermo Fisher Scientific) and Bolt Reducing agent (Thermo Fisher Scientific) and separated in Bolt™ 4–12% Bis-Tris Plus gels (Thermo Fisher Scientific) in Bolt™ MES SDS Running Buffer (Thermo Fisher Scientific). Next, proteins were electroblotted to PVDF membrane (0.2 µM, GE Healthcare) for 1 h at 100 V in ice-cold Bolt™Transfer Buffer (Thermo Fisher Scientific). Membranes were blocked in room temperature for 1 h in 5% skim milk (Sigma) in TBST buffer (Tris-buffered saline [TBS], 0.1% Tween 20) and subsequently incubated overnight in cold room with primary antibody solutions diluted in blocking buffer (the list of all commercially available primary antibodies used in the study including catalog numbers is presented in a Supplementary Table 2, with the exception for home-made anti-FUS antibody (Raczynska *et al*, 2015)). The

following day membranes were washed 3-times with TBST for 7 min and incubated with corresponding solutions of anti-mouse (A9044, Sigma; 1:15,000) or anti-rabbit (A9169, Sigma; 1:20,000) antibodies conjugated with horse radish peroxidase (HRP) for 1h in RT. For Vinculin and GFP detection, membranes were incubated with antibodies already conjugated with HRP (Supplementary Table 2) overnight in cold room. After final washing steps signals were developed using Immobilon Forte Western HRP substrate (Sigma) using G:Box Chemi-XR5 (Syngene) and ChemiDoc Imaging System (BioRad) and quantified using Multi Gauge 3.0 software (Fujifilm). Relative protein level was normalized to Vinculin.

Stable isotope labeling using amino acids in cell culture (SILAC) coupled with MS

To quantify changes in protein levels upon RPS26 silencing we applied SILAC in HEK293T cells grown in light and heavy amino acids ($^{13}C_6$ $^{15}N_2$ L-lysine-2HCl and $^{13}C_6$ $^{15}N_4$ L-arginine-HCl) containing media (Thermo Fisher Scientific). Cells were cultured in light or heavy media for more than 14 days to ensure maximum incorporation of labeled amino acids. Subsequently, cells were transfected with siRPS26 or siCtrl in final concentration of 15 nM and harvested 48h post silencing. Sample preparation was performed using modified filter-aided sample preparation method as described previously (Laakkonen *et al*, 2017; Wiśniewski *et al*, 2009). Briefly, 10 µg of proteins were washed 8-times with 8 M Urea, 100 mM ammonium bicarbonate in Amicon Ultra-0.5 centrifugal filters, and Lysine-C endopeptidase solution (Wako) in a ratio of 1:50 w/w was added to the protein lysates followed by incubation at room temperature overnight with shaking. The peptide digests were collected by centrifugation and trypsin solution was added in a ratio of 1:50 w/w in 50 mM ammonium bicarbonate and incubated overnight at room temperature. The peptide samples were cleaned using Pierce C18 reverse-phase tips (Thermo Fisher Scientific). Dried peptide pellets were re-suspended in 0.3% Trifluoroacetic acid and analyzed using nano-LC-MS/MS in Meilahti Clinical Proteomics Core Facility, University of Helsinki, Helsinki, Finland. Peptides were separated by Ultimate 3000 LC system (Dionex, Thermo Fisher Scientific) equipped with a reverse-phase trapping column RP-2TM C18 trap column (0.075 x 10 mm, Phenomenex, USA), followed by analytical separation on a bioZen C18 nano column (0.075 × 250 mm, 2.6 µm particles; Phenomenex, USA). The injected samples were trapped at a flow rate of 5 µl/min in

100% of solution A (0.1% formic acid). After trapping, peptides were separated with a linear gradient of 125 min. LC-MS acquisition data was performed on Thermo Q Exactive HF mass spectrometer with following settings: resolution 120,000 for MS scans, and 15,000 for the MS/MS scans. Full MS was acquired from 350 to 1400 m/z, and the 15 most abundant precursor ions were selected for fragmentation with 45 s dynamic exclusion time. Maximum IT were set as 50 and 25 ms and AGC targets were set to 3 e6 and 1 e5 counts for MS and MS/MS respectively. Secondary ions were isolated with a window of 1 m/z unit. The NCE collision energy stepped was set to 28 kJ mol–1.

MS Data analysis

Raw data obtained from the LC-MS/MS runs were analyzed in MaxQuant v2.0.3.0 (Cox & Mann, 2008) using either the label-free quantification (LFQ, for MS2 pull down samples) or stable isotope labeling-based quantification (for SILAC-MS samples) with default parameters. UniProtKB database for reviewed human canonical and isoform proteins of May 2023 was used. The false discovery rate (FDR) at the peptide spectrum matches and protein level was set to 0.01; variable peptide modifications: oxidation (M) and acetyl (N-term), fixed modification: carbamidomethyl (C), label Arg10, Lys8 (for SILAC-MS samples only), two missed cleavages were allowed. Statistical analyses were performed using Perseus software v2.0.3.0 (Tyanova *et al*, 2016) after filtering for "reverse", "contaminant" and "only identified by site" proteins. The LFQ intensity was logarithmized (log2[x]), and imputation of missing values was performed with a normal distribution (width = 0.3; shift = 1.8). Proteomes were compared using t-test statistics with a permutation-based FDR of 5% and *P*-values <0.05 were considered to be statistically significant. The data sets, the Perseus result files used for analysis, and the annotated MS/MS spectra were deposited at the ProteomeXchange Consortium (Deutsch *et al*, 2023) *via* the PRIDE partner repository (Perez-Riverol *et al*, 2022) with the dataset identifier PXD047400 - MS2 pull down data and PXD047397 – SILAC-MS data.

Gene ontology analysis

Gene ontology (GO) analysis performed on proteins that bind to *FMR1 RNA* (common between three replicates) was performed using an online tool - g:Profiler (Raudvere *et al*, 2019) with g:SCS algorithm, where at least 95% of matches above threshold are statistically significant. As reference proteome we used total human proteome. GO analysis performed on SILAC-MS data was performed with PANTHER 18.0 (Thomas *et al*, 2022). Statistical significance was calculated using Fisher's Exact test with Bonferroni correction and only GO terms with *P*-values < 0.05 were plotted. As reference proteome we used proteins named as non-responders (described in Results section).

Bioinformatic analyses

The WebLogo analysis represents the frequency of nucleotides across transcript sequence positions in the close vicinity of start codon within three groups of transcripts (positive & negative responders, background (BG) understood as total transcriptome). The analyzed sequence positions span the last 20 nucleotides of the 5'UTR sequence (from –20 to –1 downstream of the start codon) and the first 20 nucleotides of the coding sequence (from +1 to +20 upstream of the start codon). The graphs were created using WebLogo v2.8.2 (Crooks *et al*, 2004) based on unaligned 40-nucleotide sequence fragments. To calculate percentage of GC content of 5'UTRs and CDSs, gene annotations for all protein-coding human genes (GRCh38.p14), including coding and 5'UTR sequences of transcripts, were obtained from Ensembl release 110 (Nov2023) (Martin *et al*, 2023). UniProt protein accessions for positive and negative responders were mapped to corresponding Ensembl genes and transcripts using BioMart (Smedley *et al*, 2009). The reference dataset (BG) comprised all protein-coding genes excluding positive and negative responder genes, with one randomly selected transcript per gene. *P*-values reflect pairwise comparisons of GC content between transcript groups and were determined using a two-tailed paired t-test with Bonferroni correction. The frequencies of *k*-mers (hexamers) present in 5'UTR sequences of the BG, positive and negative responders datasets, were calculated using Jellyfish v2.3.1 (Marçais & Kingsford, 2011). The *P*-value associated with each hexamer in the positive and negative responder datasets determines the level of the hexamer overrepresentation in regards to BG, and

it was calculated from the binomial distribution implemented in SciPy v1.10.1 (Virtanen *et al*, 2020). All statistical analyzes related to the comparison of nucleotide composition among the BG, positive and negative responder datasets were also performed using SciPy.

Statistics

Group data are expressed as the means ± standard deviation (SD). Error bars represent SD. The statistical significance (if not indicated otherwise) was determined by an unpaired, two-tailed Student's t-test using Prism software v.8 (GraphPad): $*$, $P < 0.05$; $**$, $P < 0.01$; $***$, $P < 0.001$; ns, non-significant. All experiments presented in this work were repeated at least two times with similar results with at least three independent biological replicates ($N = 3$).

Data Availability

The data underlying this article are available in ProteomeXchange Consortium via the PRIDE partner repository and can be accessed with the dataset identifiers:

- PXD047400 - MS2 pull down data (for review: Username: reviewer_pxd047400@ebi.ac.uk Password: 5g4c0E9q)
- PXD047397 - SILAC-MS data (for review: Username: reviewer_pxd047397@ebi.ac.uk Password: KQsrwHqs).

Funding

This work was supported by the National Science Center (Poland) [2019/35/D/NZ2/02158 to A.B., 2020/38/A/NZ3/00498 to K.S.] and the European Union's Horizon 2020 Research and Innovation Program under the Marie Sklodowska-Curie grant agreement [No. 101003385 to A.B.] K.T. holds the Adam Mickiewicz University Foundation scholarship, awarded for the academic year 2023/24.

Acknowledgements

We thank A. Bhattacharyya and P. Hagerman for FXTAS and control fibroblasts. The 5'UTR CGG 99x FMR1-EGFP construct was a gift from Nicolas Charlet-Berguerand (Addgene plasmid # 63091).

We also thank Dominik Cysewski for participating in the preparation and analysis of MS samples from the MS2-based protein screening, Dorota Raczyńska for the kind gift of antibodies (anti-FUS and anti-RPS6), Wojciech Kwiatkowski and Tomasz Skrzypczak for their assistance with the microscopic analyses, and Roman Szczęsny for the pKK-RNAi-nucCHERRYmiR-TEV-EGFP genetic construct.

Author contributions

K.T., A.B., and K.S. conceptualized the study. K.T., I.B., A.Z., D.N., and A.B. performed the experiments and/or analyzed the data. A.B. performed the in vitro assay, the FMR99xG aggregation assay post RPS26 silencing, helped with the mass-spectrometry sample preparation, and performed the bioinformatics analysis of the MS-based protein screening and the SILAC-MS analysis. I.B. generated and characterized the *S-95xCGG* and *S-16xCGG* models. D.N. prepared the mutated (–4G<A) construct. A.Z. performed the bioinformatics analyses. K.T. generated the *L-99xCGG* model, performed and analyzed all other experiments, and prepared the figures. K.T., A.B., and K.S. wrote the original draft, and the other authors reviewed the manuscript.

Conflict of interest

The authors declare no competing interests.

References

Abu Diab M, Mor-Shaked H, Cohen E, Cohen-Hadad Y, Ram O, Epsztejn-Litman S & Eiges R (2018) The G-rich Repeats in FMR1 and C9orf72 Loci Are Hotspots for Local Unpairing of DNA. *Genetics* 210: 1239–1252

Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN & Beckmann R (2013) Structures of the human and Drosophila 80S ribosome. *Nature* 497: 80–85

Ariza J, Rogers H, Monterrubio A, Reyes-Miranda A, Hagerman PJ & Martínez-Cerdeño V (2016) A Majority of FXTAS Cases Present with Intranuclear Inclusions Within Purkinje Cells. *Cerebellum* 15: 546–551

Ash PEA, Bieniek KF, Gendron TF, Caulfield T, Lin WL, DeJesus-Hernandez M, Van Blitterswijk MM, Jansen-West K, Paul JW, Rademakers R, *et al* (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 77: 639–646

Bañez-Coronel M, Ayhan F, Tarabochia AD, Zu T, Perez BA, Tusi SK, Pletnikova O, Borchelt DR, Ross CA, Margolis RL, *et al* (2015) RAN Translation in Huntington Disease. *Neuron* 88: 667–677

Baud A, Derbis M, Tutak K & Sobczak K (2022) Partners in crime: Proteins implicated in RNA repeat expansion diseases. *Wiley Interdiscip Rev RNA* 13 doi:10.1002/wrna.1709

Buijsen RAM, Visser JA, Kramer P, Severijnen EAWFM, Gearing M, Charlet-Berguerand N, Sherman SL, Berman RF, Willemsen R & Hukema RK (2016) Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. *Hum Reprod* 31: 158–168

Cid-Samper F, Gelabert-Baldrich M, Lang B, Lorenzo-Gotor N, Ponti RD, Severijnen LAWFM, Bolognesi B, Gelpi E, Hukema RK, Botta-Orfila T, *et al* (2018) An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep* 25: 3422-3434.e7

Cox J & Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol 2008 2612* 26: 1367–1372

Crooks GE, Hon G, Chandonia JM & Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188–1190

Cui D, Li L, Lou H, Sun H, Ngai SM, Shao G & Tang J (2013) The ribosomal protein S26 regulates p53 activity in response to DNA damage. *Oncogene 2014 3317* 33: 2225–2235

Derbis M, Konieczny P, Walczak A, Sekrecki M & Sobczak K (2018) Quantitative evaluation of toxic polyglycine biosynthesis and aggregation in cell models expressing expanded CGG repeats. *Front Genet* 9

Derbis M, Kul E, Niewiadomska D, Sekrecki M, Piasecka A, Taylor K, Hukema RK, Stork O & Sobczak K (2021) Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring

expanded CGG repeats. *Nat Commun* 12: 1–17

Deutsch EW, Bandeira N, Perez-Riverol Y, Sharma V, Carver JJ, Mendoza L, Kundu DJ, Wang S, Bandla C, Kamatchinathan S, *et al* (2023) The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res* 51: D1539–D1548

Doherty L, Sheen MR, Vlachos A, Choesmel V, O'Donohue MF, Clinton C, Schneider HE, Sieff CA, Newburger PE, Ball SE, *et al* (2010) Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *Am J Hum Genet* 86: 222–228

Ferretti MB, Barre JL & Karbstein K (2018) Translational Reprogramming Provides a Blueprint for Cellular Adaptation. *Cell Chem Biol* 25: 1372-1379.e3

Ferretti MB, Ghalei H, Ward EA, Potts EL & Karbstein K (2017) Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements. *Nat Struct Mol Biol* 24: 700–707

Gaikwad S, Ghobakhlou F, Young DJ, Visweswaraiah J, Zhang H & Hinnebusch AG (2021) Reprogramming of translation in yeast cells impaired for ribosome recycling favors short, efficiently translated mRNAs. *Elife* 10

Garcia-Arocena D, Yang JE, Brouwer JR, Tassone F, Iwahashi C, Berry-Kravis EM, Goetz CG, Sumis AM, Zhou L, Nguyen D V., *et al* (2010) Fibroblast phenotype in male carriers of FMR1 premutation alleles. *Hum Mol Genet* 19: 299–312

Genuth NR & Barna M (2018) The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol Cell* 71: 364

Glineburg MR, Todd PK, Charlet-Berguerand N & Sellier C (2018) Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. *Brain Res* 1693: 43–54 doi:10.1016/j.brainres.2018.02.006

Greco CM, Berman RF, Martin RM, Tassone F, Schwartz PH, Chang A, Trapp BD, Iwahashi C, Brunberg J, Grigsby J, *et al* (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain* 129: 243–255

Greco CM, Hagerman RJ, Tassone F, Chudley AE, Del Bigio MR, Jacquemont S, Leehey M & Hagerman PJ

(2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain* 125: 1760–1771

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

Green KM, Linsalata AE & Todd PK (2016) RAN translation—What makes it run? *Brain Res* 1647: 30–42 doi:10.1016/j.brainres.2016.04.003

Hagerman RJ & Hagerman P (2016) Fragile X-associated tremor/ataxia syndrome-features, mechanisms and management. *Nat Rev Neurol* 12: 403–412 doi:10.1038/nrneurol.2016.82

Hagerman RJ, Protic D, Rajaratnam A, Salcedo-Arellano MJ, Aydin EY & Schneider A (2018) Fragile X-Associated Neuropsychiatric Disorders (FXAND). *Front Psychiatry* 9: 564

Havkin-Solomon T, Fraticelli D, Bahat A, Hayat D, Reuven N, Shaul Y & Dikstein R (2023) Translation regulation of specific mRNAs by RPS26 C-terminal RNA-binding tail integrates energy metabolism and AMPK-mTOR signaling. *Nucleic Acids Res* 51: 4415–4428

Hussain T, Llácer JL, Fernández IS, Munoz A, Martin-Marcos P, Savva CG, Lorsch JR, Hinnebusch AG & Ramakrishnan V (2014) Structural Changes Enable Start Codon Recognition by the Eukaryotic Translation Initiation Complex. *Cell* 159: 597

Jacquemont S, Hagerman RJ, Leehey MA, Hall DA, Levine RA, Brunberg JA, Zhang L, Jardini T, Gane LW, Harris SW, *et al* (2004) Penetrance of the Fragile X-Associated Tremor/Ataxia Syndrome in a Premutation Carrier Population. *JAMA* 291: 460–469

Jin P, Duan R, Qurashi A, Qin Y, Tian D, Rosser TC, Liu H, Feng Y & Warren ST (2007) Pur α Binds to rCGG Repeats and Modulates Repeat-Mediated Neurodegeneration in a Drosophila Model of Fragile X Tremor/Ataxia Syndrome. *Neuron* 55: 556–564

Kearse MG, Green KM, Krans A, Rodriguez CM, Linsalata AE, Goldstrohm AC & Todd PK (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell* 62: 314–322

Klickstein JA, Mukkavalli S & Raman M (2020) AggreCount: an unbiased image analysis tool for identifying and quantifying cellular aggregates in a spatially defined manner. *J Biol Chem* 295: 17672–17683

Laakkonen EK, Soliymani R, Karvinen S, Kaprio J, Kujala UM, Baumann M, Sipilä S, Kovanen V & Lalowski M (2017) Estrogenic regulation of skeletal muscle proteome: a study of premenopausal women and postmenopausal MZ cotwins discordant for hormonal therapy. *Aging Cell* 16: 1276–1287

Li D, Yang J, Huang X, Zhou H & Wang J (2022) eIF4A2 targets developmental potency and histone H3.3 transcripts for translational control of stem cell pluripotency. *Sci Adv* 8: 478

Li J, Su Y, Chen L, Lin Y & Ru K (2023) Identification of novel mutations in patients with Diamond-Blackfan anemia and literature review of RPS10 and RPS26 mutations. *Int J Lab Hematol* 45: 766–773

Linsalata AE, He F, Malik AM, Glineburg MR, Green KM, Natla S, Flores BN, Krans A, Archbold HC, Fedak SJ, *et al* (2019) DDX 3X and specific initiation factors modulate FMR 1 repeat-associated non-AUG-initiated translation. *EMBO Rep* 20

Liu XM, Yan MQ, Ji SY, Sha QQ, Huang T, Zhao H, Liu H Bin, Fan HY & Chen ZJ (2018) Loss of oocyte Rps26 in mice arrests oocyte growth and causes premature ovarian failure. *Cell Death Dis 2018 912* 9: 1–15

Livak KJ & Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402–408

Loomis EW, Sanz LA, Chédin F & Hagerman PJ (2014) Transcription-Associated R-Loop Formation across the Human FMR1 CGG-Repeat Region. *PLOS Genet* 10: e1004294

Ma L, Herren AW, Espinal G, Randol J, McLaughlin B, Martinez-Cerdeño V, Pessah IN, Hagerman RJ & Hagerman PJ (2019) Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome. *Acta Neuropathol Commun* 7: 1–26

Malik I, Kelley CP, Wang ET & Todd PK (2021a) Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nat Rev Mol Cell Biol* 22: 589–607 doi:10.1038/s41580-021-00382-6

Malik I, Tseng Y, Wright SE, Zheng K, Ramaiyer P, Green KM & Todd PK (2021b) SRSF protein kinase 1

modulates RAN translation and suppresses CGG repeat toxicity. *EMBO Mol Med* 13

Marçais G & Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27: 764–770

Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, Becker A, Bennett R, Berry A, Bhai J, *et al* (2023) Ensembl 2023. *Nucleic Acids Res* 51: D933–D941

Mori K, Weng S-M, Arzberger T, May S, Rentzsch K, Kremmer E, Schmid B, Kretzschmar HA, Cruts M, Broeckhoven C Van, *et al* (2013) The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science (80-)* 339: 1335–1338

Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Frericks-Zipper A, Eisenacher M, *et al* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 50: D543–D552

Piantanida N, La Vecchia M, Sculco M, Talmon M, Palattella G, Kurita R, Nakamura Y, Ronchi AE, Dianzani I, Ellis SR, *et al* (2022) Deficiency of ribosomal protein S26, which is mutated in a subset of patients with Diamond Blackfan anemia, impairs erythroid differentiation. *Front Genet* 13: 3573

Pisarev A V., Kolupaeva VG, Yusupov MM, Hellen CUT & Pestova T V. (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J* 27: 1609–1621

Plassart L, Shayan R, Montellese C, Rinaldi D, Larburu N, Pichereaux C, Froment C, Lebaron S, O'donohue MF, Kutay U, *et al* (2021) The final step of 40s ribosomal subunit maturation is controlled by a dual key lock. *Elife* 10

Raczynska KD, Ruepp MD, Brzek A, Reber S, Romeo V, Rindlisbacher B, Heller M, Szweykowska-Kulinska Z, Jarmolowski A & Schümperli D (2015) FUS/TLS contributes to replication-dependent histone gene expression by interaction with U7 snRNPs and histone-specific transcription factors. *Nucleic Acids Res* 43: 9711–9728

Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H & Vilo J (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* 47: W191–W198

Rosario R, Stewart HL, Choudhury NR, Michlewski G, Charlet-Berguerand N & Anderson RA (2022) Evidence for a fragile X messenger ribonucleoprotein 1 (FMR1) mRNA gain-of-function toxicity mechanism contributing to the pathogenesis of fragile X-associated premature ovarian insufficiency. *FASEB J* 36

Rovozzo R, Korza G, Baker MW, Li M, Bhattacharyya A, Barbarese E & Carson JH (2016) CGG repeats in the 5'UTR of FMR1 RNA regulate translation of other RNAs localized in the same RNA granules. *PLoS One* 11: 1–10

Schütz S, Fischer U, Altvater M, Nerurkar P, Peña C, Gerber M, Chang Y, Caesar S, Schubert OT, Schlenstedt G, *et al* (2014) A RanGTP-independent mechanism allows ribosomal protein nuclear import for ribosome assembly. *Elife* 3: e03473

Schütz S, Michel E, Damberger FF, Oplová M, Peña C, Leitner A, Aebersold R, Allain FHT & Panse VG (2018) Molecular basis for disassembly of an importin:ribosomal protein complex by the escortin Tsr2. *Nat Commun* 9

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

Sellier C, Freyermuth F, Tabet R, Tran T, He F, Ruffenach F, Alunni V, Moine H, Thibault C, Page A, *et al* (2013) Sequestration of DROSHA and DGCR8 by expanded CGG RNA Repeats Alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep* 3: 869–880

Sellier C, Rau F, Liu Y, Tassone F, Hukema RK, Gattoni R, Schneider A, Richard S, Willemsen R, Elliott DJ, *et al* (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J* 29: 1248–1261

Shelly KE, Candelaria NR, Li Z, Allen EG, Jin P & Nelson DL (2021) Ectopic expression of CGG-repeats alters ovarian response to gonadotropins and leads to infertility in a murine FMR1 premutation model. *Hum Mol Genet* 30: 923–938

Shi Z & Barna M (2015) Translating the genome in time and space: specialized ribosomes, RNA regulons, and RNA-binding proteins. *Annu Rev Cell Dev Biol* 31: 31–54

Shi Z, Fujii K, Kovary KM, Genuth NR, Rö St HL, Teruel MN & Correspondence MB (2017) Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol Cell* 67: 83

Smedley D, Haider S, Ballester B, Holland R, London D, Thorisson G & Kasprzyk A (2009) BioMart - Biological queries made easy. *BMC Genomics* 10: 1–12

Szczesny RJ, Kowalska K, Klosowska-Kosicka K, Chlebowski A, Owczarek EP, Warkocki Z, Kulinski TM, Adamska D, Affek K, Jedroszkowiak A, *et al* (2018) Versatile approach for functional analysis of human proteins and efficient stable cell line generation using FLP-mediated recombination system. *PLoS One* 13: e0194887

Tassone F, Iong KP, Tong TH, Lo J, Gane LW, Berry-Kravis E, Nguyen D, Mu LY, Laffin J, Bailey DB, *et al* (2012) FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Med* 4

Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP & Mi H (2022) PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci* 31: 8–22

Todd PK, Oh SY, Krans A, He F, Sellier C, Frazer M, Renoux AJ, Chen K chun, Scaglione KM, Basrur V, *et al* (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron* 78: 440–455

Tsai BP, Wang X, Huang L & Waterman ML (2011) Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. *Mol Cell Proteomics* 10

Tseng Y-J, Krans A, Malik I, Deng X, Yildirim E, Ovunc S, Tank EMH, Jansen-West K, Kaufhold R, Gomez NB, *et al* (2024) Ribosomal quality control factors inhibit repeat-associated non-AUG translation from GC-rich repeats. *Nucleic Acids Res*: gkae137

Tseng Y-J, Sandwith SN, Green KM, Chambers AE, Krans A, Raimer HM, Sharlow ME, Reisinger MA, Richardson AE, Routh ED, *et al* (2021) The RNA helicase DHX36/G4R1 modulates C9orf72 GGGGCC hexanucleotide repeat- associated translation. *J Biol Chem*: 100914

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M & Cox J (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods 2016 139* 13: 731–740

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, *et al* (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods 2020 173* 17: 261–272

Wiśniewski JR, Zougman A, Nagaraj N & Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods 2009 65* 6: 359–362

Wright SE, Rodriguez CM, Monroe J, Xing J, Krans A, Flores BN, Barsur V, Ivanova MI, Koutmou KS, Barmada SJ, *et al* (2022) CGG repeats trigger translational frameshifts that generate aggregation-prone chimeric proteins. *Nucleic Acids Res* 50: 8674–8689

Yamada SB, Gendron TF, Niccoli T, Genuth NR, Grosely R, Shi Y, Glaria I, Kramer NJ, Nakayama L, Fang S, *et al* (2019) RPS25 is required for efficient RAN translation of C9orf72 and other neurodegenerative disease-associated nucleotide repeats. *Nat Neurosci* 22: 1383–1388

Yang N, Chanda S, Marro S, Ng YH, Janas JA, Haag D, Ang CE, Tang Y, Flores Q, Mall M, *et al* (2017) Generation of pure GABAergic neurons by transcription factor programming. *Nat Methods* 14: 621–628

Yang YM, Jung Y, Abegg D, Adibekian A, Carroll KS & Karbstein K (2023) Chaperone-directed ribosome repair after oxidative damage. *Mol Cell* 83: 1527-1537.e5

Yang YM & Karbstein K (2022) The chaperone Tsr2 regulates Rps26 release and reincorporation from mature ribosomes to enable a reversible, ribosome-mediated response to stress. *Sci Adv* 8

Zu T, Gibbens B, Doty NS, Gomes-Pereira M, Huguet A, Stone MD, Margolis J, Peterson M, Markowski TW, Ingram MAC, *et al* (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A* 108: 260–265

Zu T, Guo S, Bardhi O, Ryskamp DA, Li J, Tusi SK, Engelbrecht A, Klippel K, Chakrabarty P, Nguyen L, *et al* (2020) Metformin inhibits RAN translation through PKR pathway and mitigates disease in C9orf72 ALS/FTD mice. *Proc Natl Acad Sci U S A* 117: 18591–18599

Tables

Supplementary **Table S1** (.xlsx file) List of siRNA used in the study

Supplementary **Table S2** (.xlsx file) List of primary antibodies used in the study

Supplementary **Table S3** (.xlsx file) Proteins identified in MS2-based screening

Supplementary **Table S4** (.xlsx file) Results of Gene ontology analysis performed on proteins bound to *FMR1 RNA*

Supplementary **Table S5** (.xlsx file) Results of Label Free Quantification of MS2-based data

Supplementary **Table S6** (.xlsx file) Proteins identified in SILAC-MS data with determined expression level

Supplementary **Table S7** (.xlsx file) Differential data analysis performed on SILAC-MS data

Supplementary **Table S8** (.xlsx file) Results of Gene ontology analysis performed on selected group of proteins (negative and positive responders) identified in SILAC-MS data

Supplementary **Table S9** (.xlsx file) Results of WebLogo analysis; frequencies of individual nucleotide at positions in the 20-nucleotide upstream or downstream to the start codon determined for transcripts groups (negative, positive responders and background)

Supplementary **Table S10** (.xlsx file) Lists of *k*-mers (hexamers) identified in 5'UTRs of negative and positive responders

Figures and supplementary figures with legends

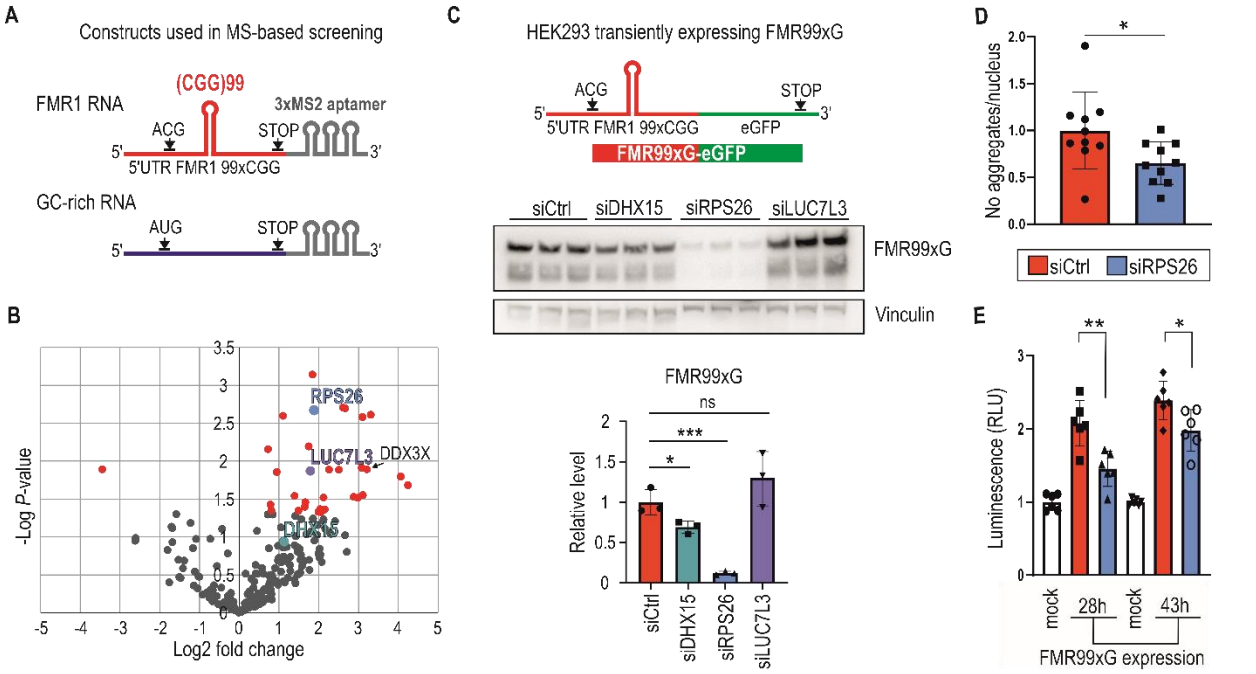


Figure 1. Mass-spectrometry (MS) -based screening revealed several proteins bound to *FMR1* RNA with expanded CGG repeats; some affect the yield of polyglycine-containing proteins, alleviating their toxicity.

(A) Scheme of two RNA molecules used to perform MS-based screening. The *FMR1 RNA* contains the entire length of the 5' untranslated region (5'UTR) of *FMR1* with expanded CGG (x99) repeats forming a hairpin structure (red). The open reading frame for the polyglycine-containing protein starts at a repeats-associated non-AUG initiated (RAN) translation-specific ACG codon. RAN translation can also be initiated from a GUG near-cognate start codon, which is not indicated in the scheme. The GC-rich RNA contains *TMEM107* mRNA enriched with G and C nucleotide residues (GC content > 70%; similar to *FMR1* RNA) with the open reading frame starting at the canonical AUG codon (blue). Both RNAs are tagged with three MS2 stem-loop aptamers (grey) interacting with an MS2 protein tagged with an in vivo biotinylating peptide used to pull down proteins interacting with the RNAs.

(B) The volcano plot representing proteins captured during MS-based screening showing the magnitude of enrichment (log2 fold change) and the statistical significance (–log *P*-value); red dots indicate proteins significantly enriched ($P < 0.05$) on *FMR1 RNA* compared to GC-rich RNA. Three proteins, RPS26, LUC7L3, and DHX15, tested in a subsequent validation experiment are marked. DDX3X is also indicated as this protein has been previously described in the context of interaction with *FMR1* mRNA.

(C) The scheme of RNA used for the transient overexpression of FMR99xG (mutant, long, 99 polyglycine tract-containing protein). The construct contains the entire length of the 5'UTR of *FMR1* with expanded CGG repeats forming a hairpin structure (red) tagged with enhanced green fluorescent protein (eGFP) (green). Western blot analysis of FMR99xG and Vinculin for HEK293 cells with insufficient DHX15, RPS26, and LUC7L3 induced by specific short interfering RNA (siRNA) treatment. To detect FMR99xG, the 9FM antibody was used. The upper bands were used for quantification. The graph presents the mean signal for FMR99xG normalized to Vinculin from $N = 3$ biologically independent samples with the standard deviation (SD). An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; ***, $P < 0.001$; ns, non-significant.

(D) Results of microscopic quantification of FMR99xG-positive aggregates in HeLa cells upon RPS26 silencing. The graph presents a normalized number of GFP-positive aggregates per nucleus in $N = 10$ biologically independent samples with the SD. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$.

(E) The influence of RPS26 silencing on apoptosis evoked in HeLa cells after 28h or 43h of FMR99xG overexpression. Apoptosis was measured as luminescence signals (relative luminescence units; RLU). The graph presents relative mean values from $N = 6$ biologically independent samples treated with either siCtrl or siRPS26 with the SD normalized to mock control (cells transfected only with the delivering reagent). An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; **, $P < 0.01$.

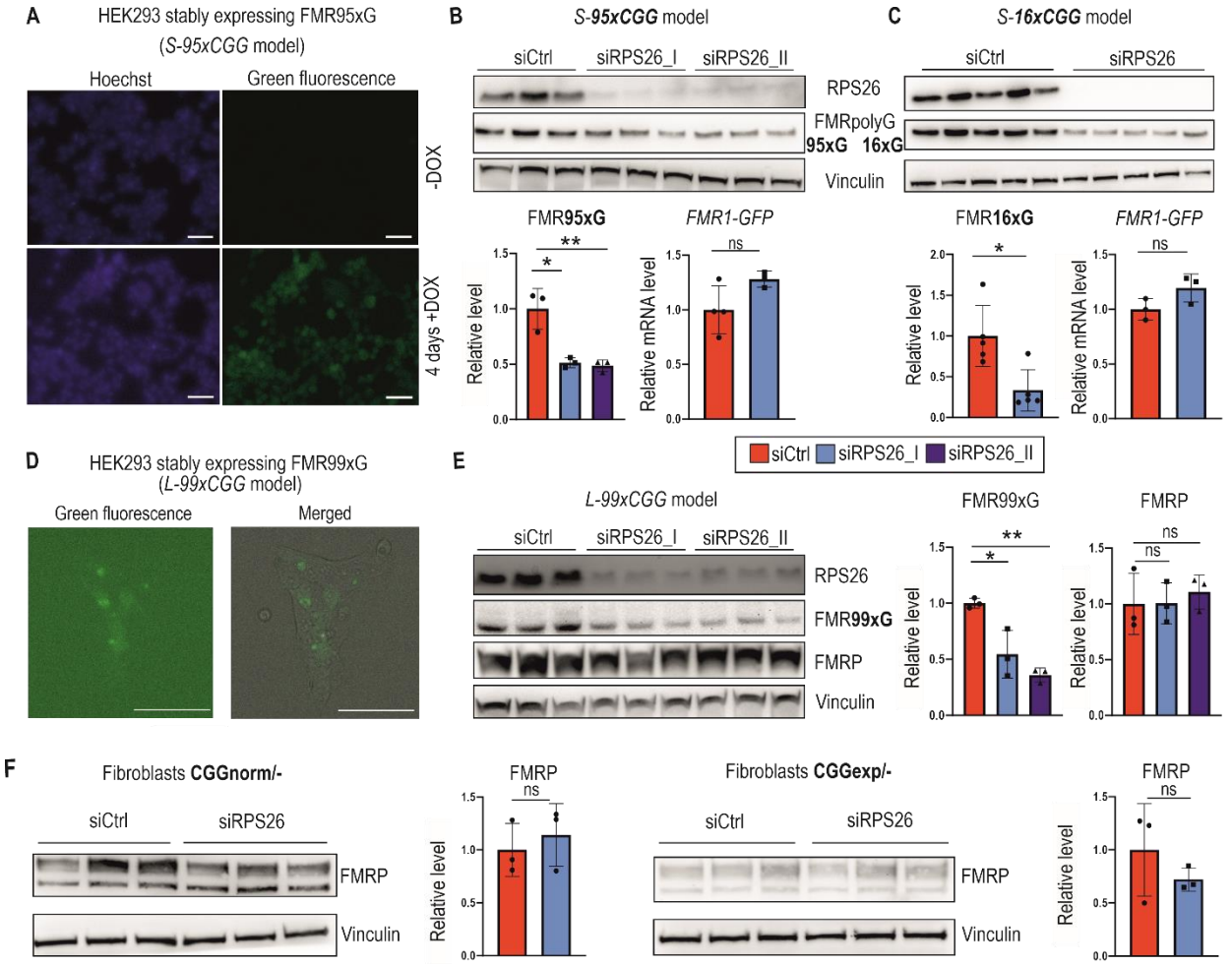


Figure 2. RPS26 insufficiency induces a lower production of polyglycine, but not FMRP, in multiple FXTAS cellular models.

(A) Representative microscopic images showing inducible expression of FMR95xG fused with eGFP in stable transgenic cell line containing a single copy of 5'UTR FMR1 99xCGG-eGFP transgene under the control of a doxycycline-inducible promoter: *S-95xCGG model*. Blue are nuclei stained with Hoechst; the green signal is derived from FMR95xG tagged with eGFP; scale bar 50 µm; +DOX and –DOX indicate cells treated (or not) with doxycycline to induce transcription of the transgene from the doxycycline-dependent promoter.

(B & C) Results of western blot analyses of FMR95xG (with long, mutant polyglycine stretches) or FMR16xG (with short, normal polyglycine stretches) normalized to Vinculin and an RT-qPCR analysis of *FMR1-GFP* transgene expression normalized to *GAPDH* upon RPS26 silencing in *S-95xCGG* and *S-16xCGG* models, respectively. siRPS26_I and siRPS26_II indicate two different siRNAs used for RPS26 silencing. The graphs present means from $N = 3$ (**B**) or $N = 5$ biologically independent samples (**C**) with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; **, $P < 0.01$; ns, non-significant.

(D) Representative microscopic images of cells stably expressing FMR99xCGG obtained after transduction with lentivirus containing the 5'UTR FMR1 99xCGG-eGFP transgene: *L-99xCGG* model. The green field image showing the signal from FMR99xG fused with eGFP was merged with the bright field image; scale bar 100 µm.

(E) Results of western blot analyses of FMR99xG and FMRP normalized to Vinculin upon RPS26 silencing in the *L-99xCGG* model. The graphs present means from $N = 3$ biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; **, $P < 0.01$; ns, non-significant.

(F) Results of western blot analyses of FMRP levels normalized to Vinculin upon RPS26 silencing in fibroblasts derived either from healthy individual (CGGnorm) or FXTAS patient (CGGexp). The graphs present means from $N = 3$ biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: ns, non-significant.

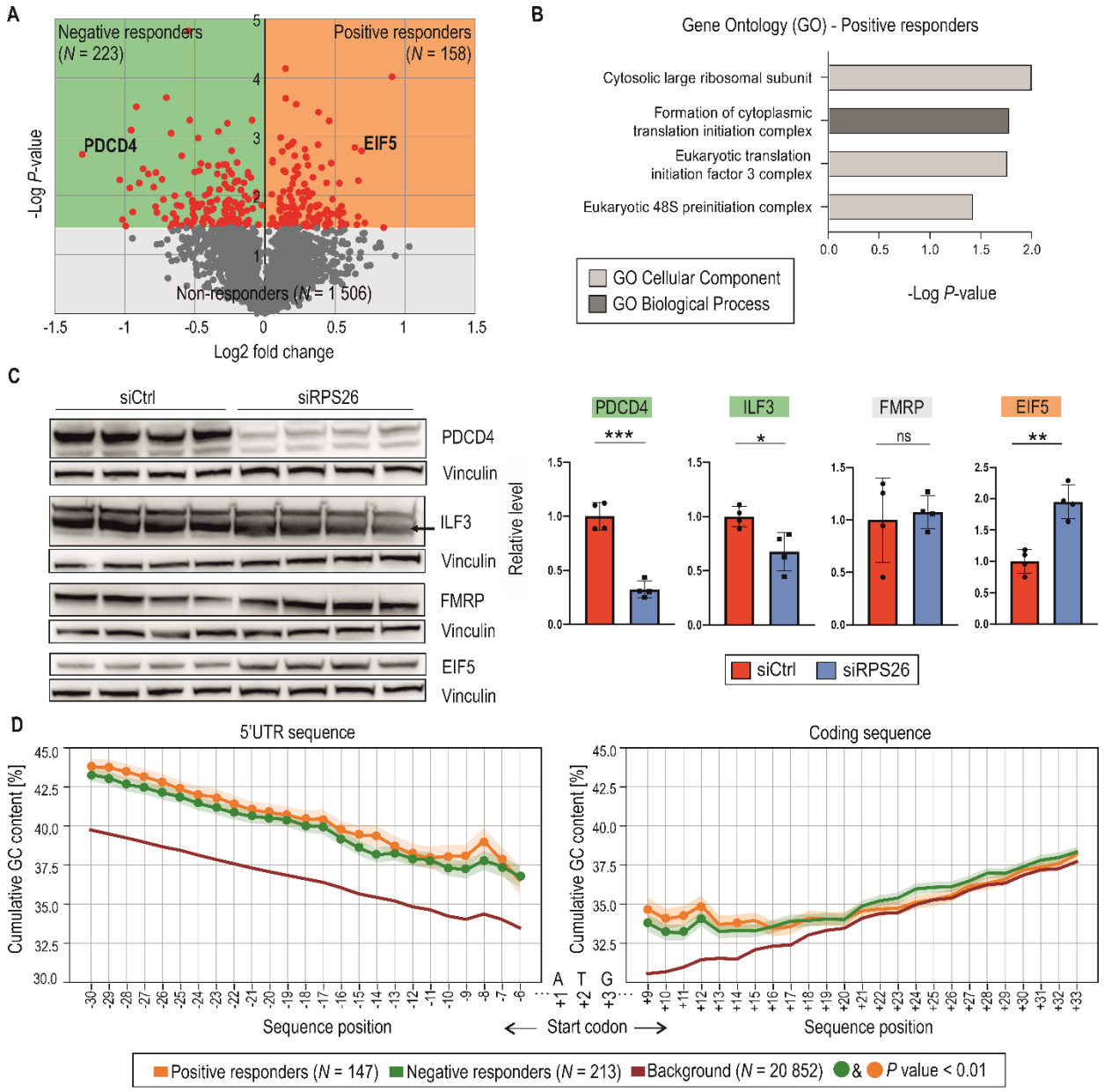


Figure 3. Changes in the proteome of RPS26-deficient cells are not robust.

(A) The volcano plot represents a stable isotope labeling using amino acids in cell culture (SILAC)-based quantitative proteomic analysis identifying proteins sensitive to RPS26 insufficiency. It shows the magnitude of protein-level changes (log2 fold change) vs. the statistical significance (–log *P*-value) 48h post siRPS26 treatment. Data were collected from three independent biological replicates for each group. Grey dots indicate proteins non-responding to RPS26 depletion ($N = 1506$). Red dots indicate proteins responding to RPS26 insufficiency ($P < 0.05$); EIF5 and PDCD4 are examples of Negative ($N = 223$; green) and Positive ($N = 158$; orange) responders, respectively. Protein groups were further analyzed in B, C, and D.

(B) Gene ontology (GO) analysis performed for positive responders of RPS26 insufficiency from the proteomic experiment shown in A. The graph presents significantly enriched GO terms ($P < 0.05$); statistical significance was calculated using Fisher's Exact test with the Bonferroni correction. Note that for negative responders, no GO terms were significantly enriched.

(C) Data validation from the proteomic experiment described in A. Western blot analyses of PDCD4, ILF3, FMRP, and EIF5 proteins normalized to Vinculin for HEK293 cells treated with siRPS26. The graphs present means from $N = 4$ biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ns, non-significant.

(D) The percentage of GC content in 5'UTRs and coding sequences across extending sequence windows initiated from the start codon within three groups of transcripts: Negative responders (green), Positive responders (orange), and Background (BG) understood as the total transcriptome (red). To avoid biases in the GC-content mean, transcripts with 5'UTR sequences shorter than 20 nucleotides were excluded from the analysis, yielding the following sample sizes: Negative responders ($N = 213$), Positive responders ($N = 147$), and BG ($N = 20{,}862$). For example, position -6 in 5' UTR sequences corresponds to a 6-nucleotide fragment (window from -6 to -1 positions upstream of ATG), while position -7 corresponds to a 7-nucleotide fragment (window from -7 to -1 positions). The solid line shows the mean GC content at a given position (i.e., within the window), and the shade indicates the standard error of the mean. P-values < 0.01 are denoted by green or yellow dots. These reflect pairwise comparisons of GC content between transcript groups (compared to BG) and were determined using a two-tailed paired t-test with Bonferroni correction.

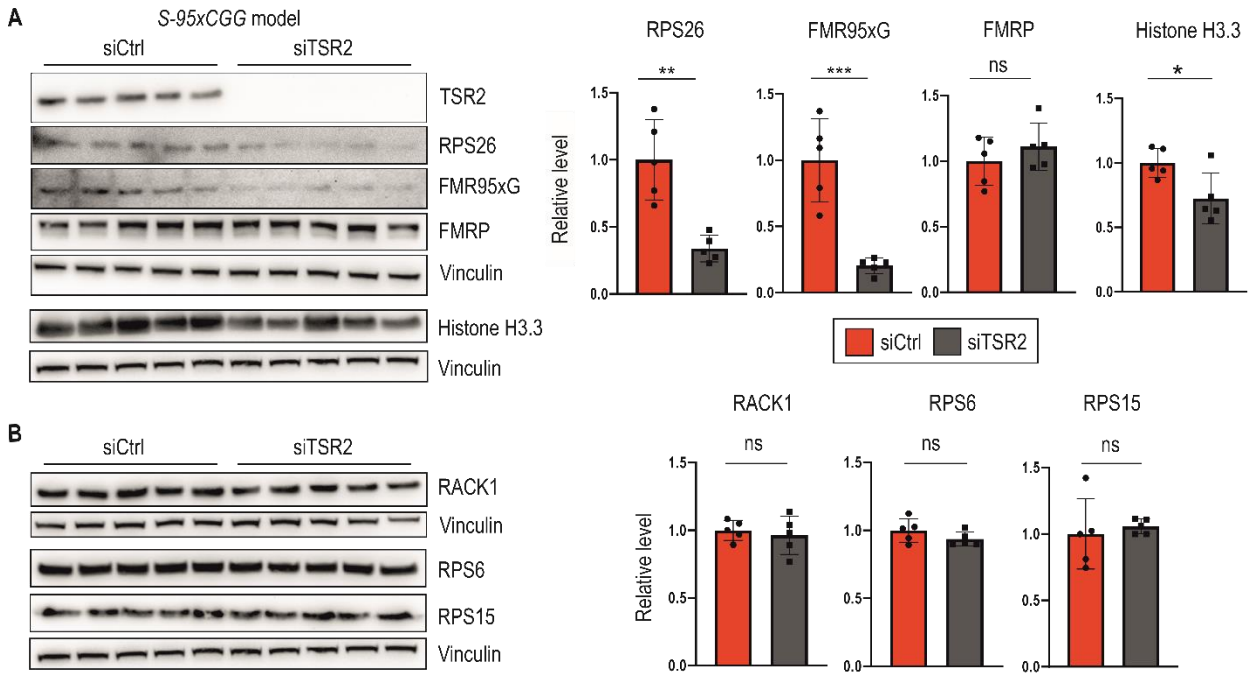


Figure 4. Insufficiency of the TSR2 chaperone protein lowers the level of RPS26 and FMRpolyG but not FMRP and selected 40S components.

(A) Western blot analyses of RPS26, FMR95xG, FMRP but also Histone H3.3, a sensor of ribosomes depleted with RPS26, normalized to Vinculin in stable *S-95xCGG* cells treated with siTSR2. Graphs represent means from $N = 5$ biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: **, $P < 0.01$; ***, $P < 0.001$; ns, non-significant.

(B) Western blot analyses of selected 40S ribosomal proteins RACK1, RPS6, and RPS15, normalized to Vinculin upon TSR2 silencing in the *S-95xCGG* model. The graphs present means from $N = 5$ biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; ns, non-significant.

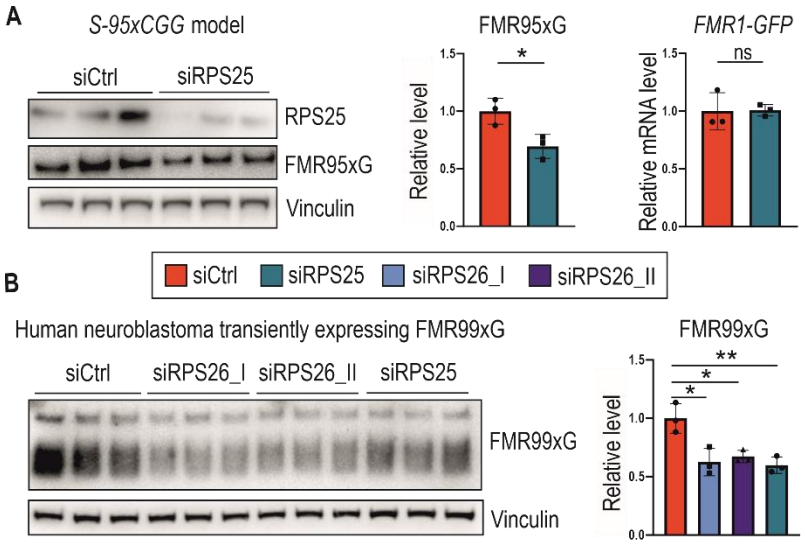


Figure 5. Silencing RPS25, the other component of 43S PIC, reduces the level of polyglycine produced from mutant *FMR1* mRNA.

(A) A western blot analysis of FMR95xG normalized to Vinculin and an RT-qPCR analysis of the *FMR1-GFP* transgene expression normalized to *GAPDH* upon RPS25 silencing in the *S-95xCGG* model. The graphs present means from $N = 3$ biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; ns, non-significant.

(B) A western blot analysis of FMR99xG normalized to Vinculin from the human neuroblastoma cell line (SH-SY5Y) upon silencing of either RPS26 or RPS25. siRPS26_I and siRPS26_II indicate two different siRNAs used for RPS26 silencing. The upper bands were used for quantification. The graph presents means from $N = 3$ biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; **, $P < 0.01$; ns, non-significant.

Supplementary Figure S1

A.

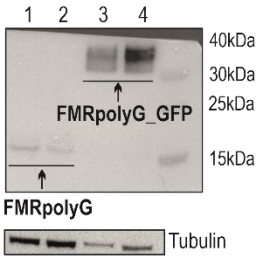
FMR1 RNA

tcagtcaggcgctcagctccgtttcggtttcacttccggtggagggccgcctctgagcgggcggcgggccgacggcgagcgcgggcggcggcggt 5'UTR *FMR1*
gacggaggcgccgctgccagggggcgtgcggcagcgcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggaggcggcggcggc
ggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc FMRpolyG
ggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc
ggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggcggc
cggggcgggctcccggcgctagcagggctgaagagaagatggaggagctggtggtggaagtgcggggctccaatggcgctttctacaaggca
cctatgataatagtgactaatcgtccGGCCgggTAATGAcggaaagatcgccgtgtaattctagatcatcgaaacatgaggatcacccatatct 4xSTOP
gcagtcgacatcgaaacatgaggatcacccatgtctgcagtcgacatcgaaacatgaggatcacccatgtctgcagtcgacatcgaaacatgagg
atcacccatgtctgcagtcgacatcgaaatcgataagcttcagatcagatcctagagcggcc
3xMS2 stem loop aptamers

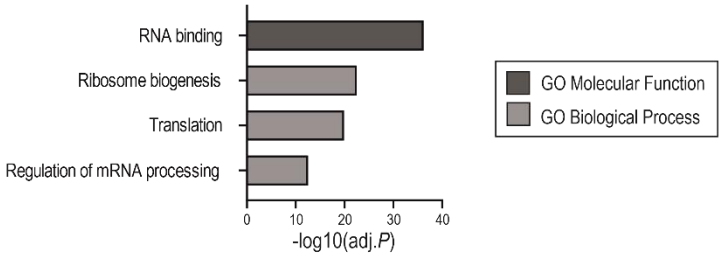
GC-rich RNA

ccgtgctgctcgcccggagcccctttccccatcacttcttcggcggtccctgaggcggagccccggagtcccgctgggcgtgagatgcagagagcg *TMEM170* ORF
gccgcaggtgggcccggggctggagggcgctcggccgccacccgagcgggtcttggccttgagcttccgagcgcctcaggttcagagctgcaccc
cacgagcccgggaggctgtggtccccgccctgcctgggttgccccacggcgcccggcctccttcgaggggtctcggagcggcccggccgcgctga
ggagtcagagctcgcgctccccttgccccggggctgcagcccggcccctccgcgcaggcgctcgttcgctgatcgcgggcgcctggggctaaatc STOP
ctccccctccgagacccgctccgctttcctaatcgtccggccgggtaaGtgacggaaagatcgccgtgtaattctagatcatcgaaacatgaggat
cacccatatctgcagtcgacatcgaaacatgaggatcacccatgtctgcagtcgacatcgaaacatgaggatcacccatgtctgcagtcgacatcg
aaacatgaggatcacccatgt
3xMS2 stem loop aptamers

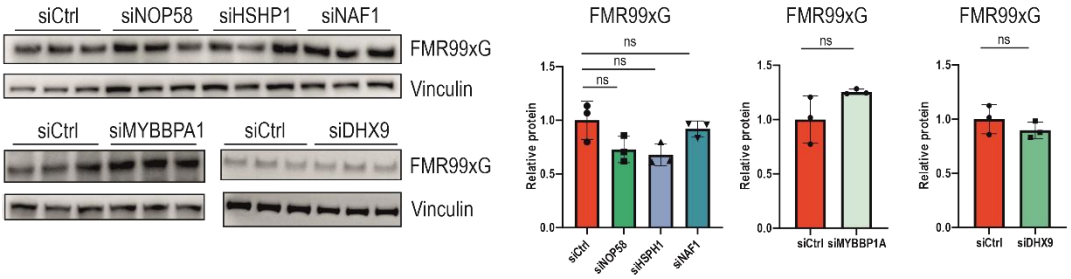
B.



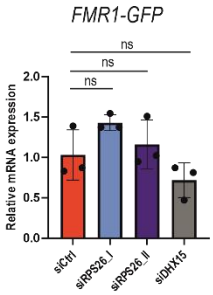
C.



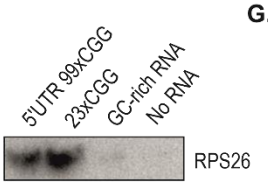
D.



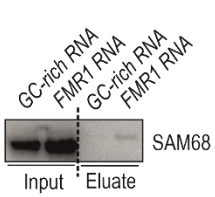
E.



F.



G.



Supplementary Figure S1.

MS-based screening revealed proteins interacting with *FMR1 RNA.*

(A) Sequences of *FMR1 RNA* and *GC-rich RNA* constructs used in MS2-based screening. Note that the sequences presented are a fragment of the plasmids used in the experiment.

(B) Results of western blot representing FMRpolyG derived from *FRM1 RNA* construct (lane 1&2) in comparison to FMRpolyG tagged with GFP (lane 3&4) derived from template construct (Addgene #63089). Tubulin was used as a loading control.

(C) Gene Ontology (GO) analysis performed on proteins detected in *FMR1 RNA* sample group (common among three technical replicates). Graph presents significantly enriched, selected GO terms.

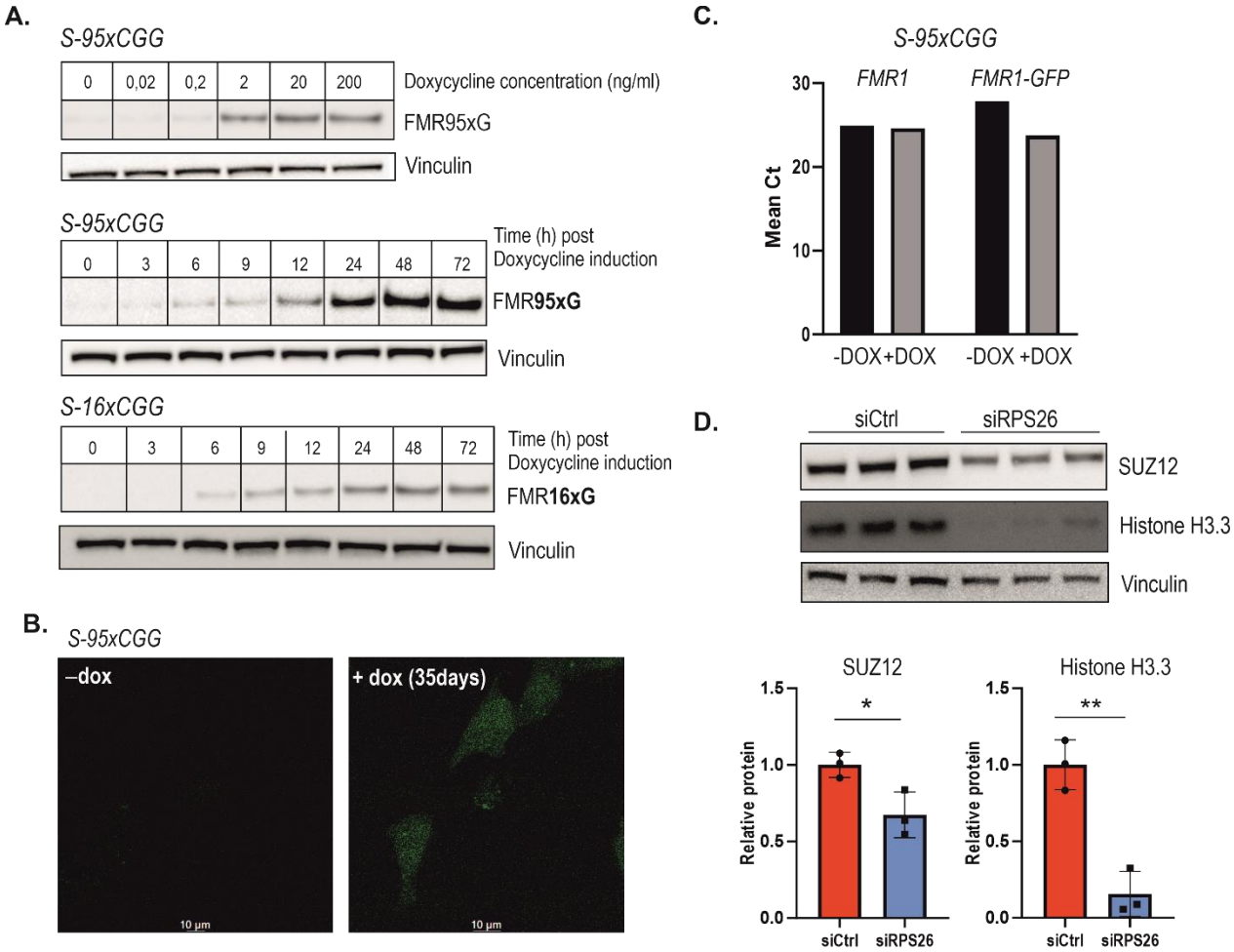
(D) Results of western blot analysis of FMR99xG normalized to Vinculin upon NOP58, HSHP1, NAF1, MYBBP1A and DHX9 siRNA-based silencing in HEK293 cells. Graphs present means from $N = 3$ biologically independent samples with SDs. Unpaired Student's t test was used to calculate statistical significance: ns, non-significant. Note that presented blots are cropped.

(E) RT-qPCR analysis of *FMR1-GFP* transgene expression derived from transient transfection system normalized to *GAPDH* upon RPS26 and DHX15 depletion in HEK293 cells. siRPS26_I and siRPS26_II indicate two independent siRNAs. Graphs present means from $N = 3$ biologically independent samples with SDs. Unpaired Student's t test was used to calculate statistical significance: ns, non-significant.

(F) Results of western blot demonstrating enrichment of RPS26 detected in 5'UTR 99xCGG and 23xCGG eluates compared to GC-rich RNA in biochemical assay. 50% of eluate fraction was loaded on the gel. All RNA molecules in the experiment were biotinylated (b) in order to perform bRNA-protein pull down procedure from HEK293 cell extract. No RNA sample served as a negative control.

(G) Results of western blot demonstrating enrichment of SAM68 protein detected in *FMR1 RNA* sample eluate after MS2-based, *in cellulo* RNA-protein pull down procedure; the percentage of following fractions were loaded on the gel: Input, 5% of total lysate; eluate, 20% of immunoprecipitated fraction. Note that presented blot is cropped.

Supplementary Figure S2.



Supplementary Figure S2.

Characterization of *S-95xCGG* and *S-16xCGG* models and RPS26-responders.

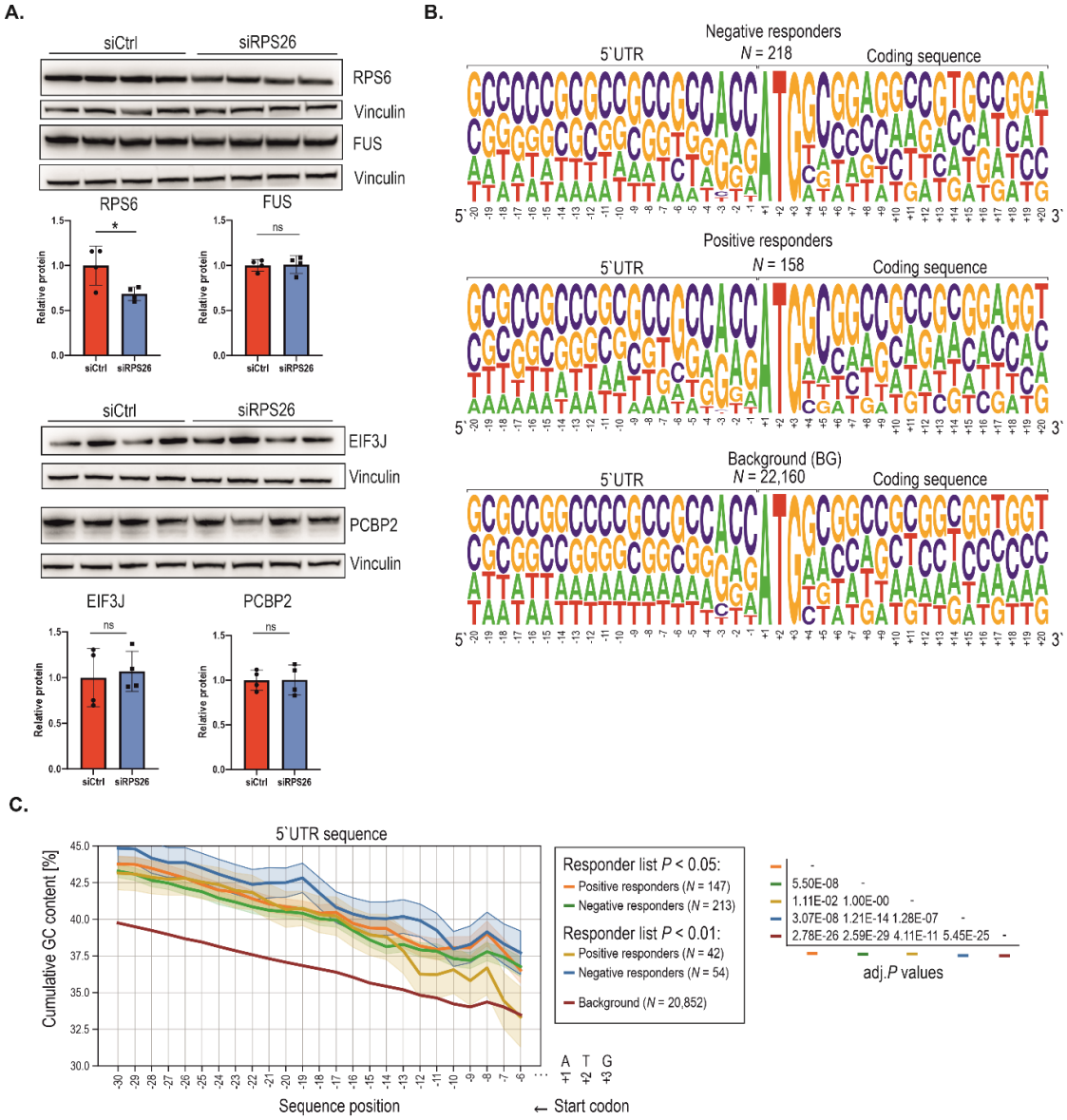
(A) Results of western blot of FMR95xG and FMR16xCGG expression in *S-95xCGG* and *S-16xCGG* models, respectively, after induction of doxycycline-inducible promoter. Upper image represents the titration of doxycycline (in ng/ml). Following images demonstrate time-dependent course of FMR95xG/FMR16xCGG expression post promoter induction. The same amount of protein sample was loaded on gels. Vinculin serves as a loading control.

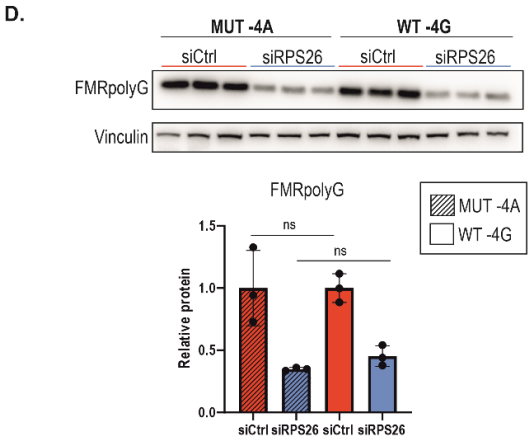
(B) The representative confocal microscopic images showing homogenous expression of FMR95xG fused with eGFP in *S-95xCGG* model after 35 days of doxycycline treatment; scale bar 10 µm. Note that no GFP-positive aggregates of polyglycine were detected.

(C) RT-qPCR analysis of endogenous *FMR1* and *FMR1-GFP* transgene expression in *S-95xCGG* model with (+DOX) or without (-DOX) doxycycline induction. Bars represent mean cycle (Ct).

(D) Results of western blot analysis of SUZ12 and Histone H3.3 proteins normalized to Vinculin upon RPS26 silencing in COS7 cells. Graphs present means from $N = 3$ biologically independent samples with SDs. Unpaired Student's t test was used to calculate statistical significance: *, $P < 0.05$; **, $P < 0.01$.

Supplementary Figure S3.





E.

5'UTRs

Negative responders			Positive responders		
hexamer	fold change (log2)	*P* value	hexamer	fold change (log2)	*P* value
CGCCGC	1,10	3,40E-13	GCCGCC	0,90	1,20E-08
GCCGCC	1,00	2,60E-12	CCGCTG	1,40	1,60E-07
GCGGCG	0,90	9,30E-11	CCGGTC	1,80	9,80E-07
CCGCCG	1,00	8,80E-10	TTCCGG	1,60	4,60E-06
GGCGGC	0,70	1,90E-07	CGCCGC	0,80	5,00E-06
CAGCGG	1,20	2,50E-07	CGGTCT	2,10	5,40E-06
GGCGCC	1,00	5,00E-07			
GCAGCG	1,10	5,40E-07			
AGCGGC	1,00	3,00E-06			
GAGCGG	1,10	3,60E-06			
CGGCGG	0,70	4,70E-06			
CGCGGC	0,90	5,40E-06			
GTGGCG	1,30	7,50E-06			
TTCCGG	1,40	7,90E-06			

Supplementary Figure S3.

Validation of SILAC-MS results and bioinformatic analysis of mRNAs encoding RPS26-sensitive proteins.

(A) Validation of SILAC-MS data. Results of western blot analysis of RPS6 & FUS (results are in line with proteomic analysis) and PCBP2 & EIF3J (results are not in line with proteomic analysis) normalized to Vinculin upon RPS26 silencing in HEK293 cells. Graphs present means from $N = 4$ biologically independent samples with SDs. Unpaired Student's t test was used to calculate statistical significance: *, $P < 0.05$; ns, non-significant.

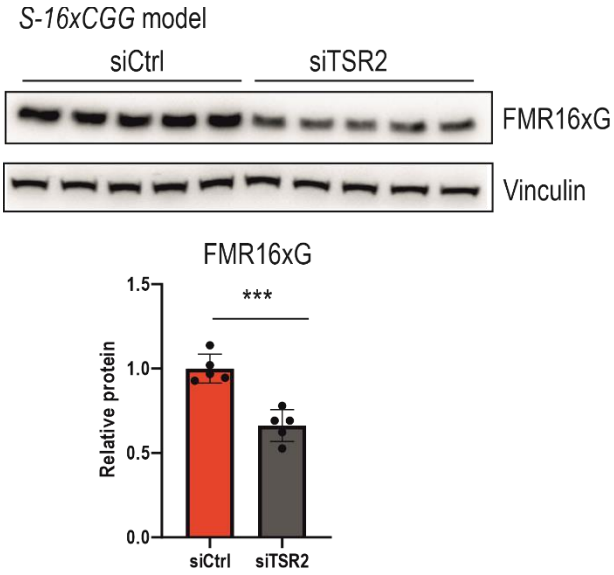
(B) Frequency of nucleotides across transcript sequence positions in the close vicinity of start codon within three groups of transcripts: negative responders ($N = 218$), positive responders ($N = 158$), and background transcripts ($N = 22{,}160$). The analyzed sequence positions span the last 20 nucleotides of the 5' UTR sequence (from −20 to −1 downstream of the start codon) and the first 20 nucleotides of the coding sequence (from +1 to +20 upstream of the start codon). Each nucleotide symbol's height on the graph is proportionate to its relative frequency at the corresponding sequence position. The cumulative height of the stack at each position sums to one, encompassing the frequencies of all four nucleotides.

(C) The percentage of GC content of 5'UTR across extending sequence windows initiated upstream from the start codon within groups of transcripts. The graph gathers analysis of transcripts determined with two *P*-value cut offs ($P < 0.05$ and $P < 0.01$) yielding different samples sizes for each group. Colors corresponding to given group as well as its size are indicated in the legend. The solid line shows the mean GC content at a given position (i.e., within the window), and the shade indicates the standard error of the mean. adj. *P* values (right panel) reflect pairwise comparisons of GC content between transcript groups and were determined using a two-tailed paired t-test with Bonferroni correction.

(D) Results of western blot analysis of FMRpolyG protein derived from wild type (WT) or mutated (−4G<A) construct normalized to Vinculin upon RPS26 silencing in HEK293 cells. Graphs present means from $N = 3$ biologically independent samples with SDs. Unpaired Student's t test was used to calculate statistical significance: ns, non-significant.

(E) Comparison of overrepresented hexamers in 5'UTR full-length sequences within two transcript groups: negative ($N = 218$) and positive responders ($N = 158$). The frequency of each hexamer is compared to the corresponding hexamer frequency in 5'UTR sequences of background transcripts (BG, $N = 22{,}160$) (indicated as log2 fold change). The *P*-value associated with each hexamer signifies the degree of overrepresentation (enrichment) and is calculated from the binomial distribution. Only hexamers with *P*-values $< 10^{-5}$ are displayed.

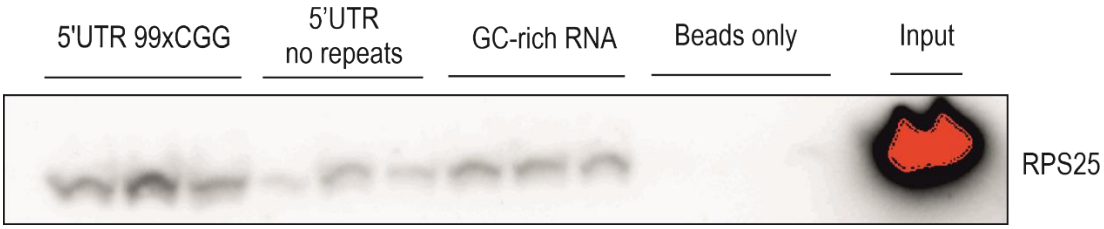
Supplementary Figure S4.



Silencing of TSR2 reduces FMR16xG level in *S-16xCGG* model.

Results of western blot analysis of FMR16xG normalized to Vinculin upon TSR2 silencing in *S-16xCGG* model. Graphs present means from $N = 5$ biologically independent samples with SDs. Unpaired Student's t test was used to calculate statistical significance: ***, $P < 0.001$.

Supplementary Figure S5.



Biotinylated RNA-protein pull down of RPS25.

Results of western blot demonstrating enrichment of RPS25 detected in samples obtained with biotinylated 5'UTR 99xCGG RNA and to a lesser extend in 5'UTR (no repeats) & GC-rich RNA eluates (ca. 20% of eluate fraction was loaded on the gel). Input, 5% of total protein lysate was loaded. All RNA molecules in the experiment were biotinylated (b) in order to perform bRNA-protein pull down procedure. Beads only was a sample (with no bRNA) which served as a negative control.