



**Identyfikacja nowych modyfikatorów niekanonicznej
biosyntezy toksycznego białka poliglicynowego ze
zmutowanego mRNA *FMR1***

Katarzyna Tutak

Rozprawa doktorska

Promotor: prof. dr hab. Krzysztof Sobczak

Promotor pomocniczy: dr inż. Anna Baud

Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Uniwersytet im. Adama Mickiewicza w Poznaniu

Poznań, 2024



**Identification of novel modifiers of noncanonical
biosynthesis of toxic polyglycine protein from mutant
FMR1 mRNA**

Katarzyna Tutak

PhD thesis

Supervisor: prof. dr hab. Krzysztof Sobczak

Assistant supervisor: Anna Baud, PhD

Department of Gene Expression,
Institute of Molecular Biology and Biotechnology,
Adam Mickiewicz University Poznań

Poznań, 2024

To all the people who supported me on this journey

Content

Articles included in the dissertation	6
Funding	7
Abbreviations	8
STRESZCZENIE.....	11
ABSTRACT.....	14
1. INTRODUCTION.....	16
1.1 Expanded microsatellites in human genome can cause several neurological diseases	16
1.2 Genetic basis and phenotype of fragile X-associated syndromes.....	17
1.3 Molecular pathomechanisms of CGG repeats expansion in FMR1	19
1.3.1 Secondary structures formed within expanded CGG repeats in mutant FMR1 mRNA.....	19
1.3.2 R loop formation	20
1.3.3 Sequestration of RNA binding proteins	21
1.3.4 Repeat-associated non-ATG translation.....	21
1.3.5 RAN translation of FMR1 with CGGexp.....	22
1.4 Mechanisms and proteins implicated in the regulation of CGG-related RAN translation	23
1.5 FMRpolyG driven pathology in PM-linked conditions	27
1.6 Ribosomal heterogeneity in a nutshell	30
2. AIMS OF THE STUDY.....	33
3. THE LITERATURE REVIEW: <i>Partners in crime: Proteins implicated in RNA repeat expansion diseases</i>	35
4. THE MANUSCRIPT/PREPRINT: <i>Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions</i>	36
5. UNPUBLISHED RESULTS: <i>Characterization of other factors affecting noncanonical polyglycine synthesis from mutant FMR1 mRNA</i>	37
5.1 Results.....	39
5.1.1 Alternative preparation of MS2-based pull down samples for MS analysis revealed different proteins bound to FMR1 RNA.....	39
5.1.2 Silencing of DDX21 and DHX15 downregulated the level of FMRpolyG in two cellular models.....	40
5.1.3 The effect of ALYREF silencing on FMRpolyG.....	43

5.1.4	Technical aspects of mass spectrometry sample preparation affected identification of proteins bound to FMR1 RNA.....	44
5.2	Discussion.....	48
5.3	Materials and methods.....	52
5.3.1	MS2-based pull down followed by acetone protein precipitation and SP3 purification	52
5.3.2	Mass spectrometry (MS).....	53
5.3.3	MS data analysis.....	53
5.3.4	Cell culture and transfection.....	53
5.3.5	Quantitative real-time PCR (RT-qPCR).....	54
5.3.6	Gene Ontology	54
5.3.7	Venn diagrams.....	54
5.4	Supplementary information.....	55
5.4.1	Tables	55
5.4.2	Supplementary figures.....	56
5.5	Bibliography	58
6.	CONCLUDING REMARKS.....	71

Articles included in the dissertation

- I. Baud A*, Derbis M*, **Tutak K***, Sobczak K. (2022), Partners in crime: proteins implicated in RNA repeat expansion diseases. WIREs RNA. doi: 10.1002/wrna.1709 *these authors contributed equally
- II. **Tutak K**, Broniarek I, Zielezinski A, Niewiadomska D, Baud A, Sobczak K. (2024), Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions. Published as a preprint on bioRxiv server under the doi number: doi.org/10.1101/2024.03.27.586952

Funding

The work was supported by the following funding:

- I. The Polish National Science Centre grant (2019/35/D/NZ2/02158) to Anna Baud
- II. The Polish National Science Centre grant (UMO-2020/38/A/NZ3/00498) to Krzysztof Sobczak
- III. Adam Mickiewicz University Foundation Scholarship for best doctoral students awarded to Katarzyna Tutak for academic year 2023/2024

Abbreviations

5'UTR	5' untranslated region
40S	eukaryotic small ribosomal subunit 40S
43S	PIC 43S preinitiation complex
60S	eukaryotic large ribosomal subunit 60S
80S	eukaryotic 80S ribosome
ACG	near-cognate ACG start codon
ALYREF	THO complex subunit 4
ANKZF1	Ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1
<i>ASFMR1</i>	Antisense transcript at the <i>FMR1</i> locus
ASFMRpolyA	polyalanine-containing RAN protein translated from <i>ASFMR1</i>
ASFMRpolyP	polyproline-containing RAN protein translated from <i>ASFMR1</i>
ASFMRpolyR	polyarginine-containing RAN protein translated from <i>ASFMR1</i>
C9 ALS-FTD	amyotrophic lateral sclerosis and/or frontotemporal dementia
C9orf27	Chromosome 9 Open Reading Frame 27
CAG	CAG triplet repeat
CTG	CTG triplet repeat
CGGexp	expanded CGG repeats
CGIs	CpG islands
Ct	cycle threshold
Ctrl	control
DBA	Diamond-Blackfan Anemia
DDR	DNA damage response
DDX3X	ATP-dependent RNA helicase DDX3X
DGCR8	DiGeorge syndrome critical region 8
DHX9	ATP-dependent RNA helicase A
DHX15	ATP-dependent RNA helicase DHX15
DM1	myotonic dystrophy type 1
DMPK1	Dystrophia myotonica protein kinase 1
DOX	doxycycline
DROSHA	Drosha ribonuclease type 3
eIF	Eukaryotic initiation factor
eIF2 α	Eukaryotic translation initiation factor 2
eIF4A	Eukaryotic initiation factor 4A
eIF4E	Eukaryotic initiation factor 4E
FM	full mutation of <i>FMR1</i> gene characterized by more than 200 CGG repeats
<i>FMR1</i>	fragile X messenger ribonucleoprotein 1 gene/mRNA
FMRP	Fragile X messenger ribonucleoprotein 1 protein
FMRpolyA	polyalanine-containing RAN protein translated from mutated <i>FMR1</i>
FMRpolyG	polyglycine-containing RAN protein translated from mutated <i>FMR1</i>
FMRpolyR	polyarginine-containing RAN protein translated from mutated <i>FMR1</i>
FXAND	fragile X-associated neuropsychiatric disorders
FXPAC	fragile X-premutation-associated conditions
FXPOI	fragile X-associated primary ovarian insufficiency
FXS	fragile X syndrome
FXTAS	fragile X-associated tremor/ataxia syndrome
G4C2	GGGGCC hexanucleotide repeat expansions in C9orf72

GFP Green fluorescent protein
 GO gene ontology
 GUG near-cognate GUG start
 H3.3 Histone protein 3.3
 HD Huntington's disease
 hnRNP A2/B1 heterogeneous nuclear ribonucleoprotein A2/B1
 Hox homeobox genes
HTT huntingtin gene/mRNA
 ID uniprot unique protein identifier
 iPSCs induced pluripotent stem cells
 IRS integrated stress response
 IRES internal ribosome entry site
 L22e Large ribosomal subunit protein eL22
 LAP2 β Lamina-associated polypeptide 2 beta
 LFQ label free quantification
 LTN1 E3 ligase listerin
 LUC7L3 Luc7-like protein 3
 m7G 5' methyl-7-guanosine cap modification
 MBNL1 Muscleblind-like protein 1
 Met-tRNA initiator methionyl-tRNA
 mRNA messenger RNA
 MS mass spectrometry
 MS2 bacteriophage ms2 coat protein
 MYBBP1A Myb-binding protein 1A
 NAF1 H/ACA ribonucleoprotein complex non-core subunit NAF1
 NEMF Nuclear export mediated factor
 NOP58 Nucleolar protein 58
 NXF1 Nuclear RNA export factor 1
 NXT1 NTF2-related export factor 1
 ns non-significant
 ORF open reading frame
 PCBP2 Poly(rC)-binding protein 2
 PERK PKR-like ER kinase
 PM premutation of *FMR1* gene characterized by 55-200 CGG repeats
 POI primary ovarian insufficiency
 Pol II RNA polymerase II
 PRC2 Polycomb repressive complex 2
 Pur α Purine-rich binding protein α
 RACK1 Small ribosomal subunit protein RACK1
 RAN translation repeat-associated non-ATG translation
 RBPs RNA binding proteins
 REDs repeat expansion disorders
 R-loop RNA:DNA hybrid
 RPS6 Small ribosomal subunit protein eS6
 RPS10 Small ribosomal subunit protein eS10
 RPS15 Small ribosomal subunit protein eS15
 RPS25 Small ribosomal subunit protein eS26

RPS26 Small ribosomal subunit protein eS26
RPL10A Large ribosomal subunit protein uL1
RPL38 Large ribosomal subunit protein eL38
RQC ribosome-associated quality control pathway
Sam68 Src-Associated substrate during mitosis of 68-kDa
SCA spinocerebellar ataxia
SD standard deviation
SDS sodium dodecyl sulfate
SILAC stable isotope labeling using amino acids in cell culture
siRNA small interfering RNA
SP3 single-pot solid-phase-enhanced sample preparation
SRSF1 Serine/arginine-rich splicing factor 1
SRSF3 Serine/arginine-rich splicing factor 3
SRPK 1 Serine/arginine protein kinase 1
STR short/simple tandem repeat
Suz12 Polycomb protein Suz12 (mouse)
TDP-43 TAR DNA binding protein of 43 kDa
TSR2 Pre-rRNA-processing protein TSR2 homolog
tRNA transfer RNA
uORF upstream open reading frame
UPS ubiquitin proteasome system
WT wild type
YTDC1 YTH domain-containing protein 1

STRESZCZENIE

Zespoły łamliwego chromosomu X obejmują choroby genetyczne wywołane mutacją dynamiczną w genie *FMR1* (ang. *Fragile X messenger ribonucleoprotein 1*), który koduje białko FMRP (ang. *Fragile X messenger ribonucleoprotein 1 protein*). U zdrowego człowieka, w sekwencji rejonu 5' niepodlegającemu translacji (5'UTR) genu *FMR1* znajduje się zwykle 25-35 powtórzeń trójki nukleotydowej CGG. Jednakże, długość tej sekwencji jest wysoce polimorficzna, a powtórzenia CGG mają tendencję do wydłużania, często znacznego. Proces ten nazywany jest ekspansją powtórzeń CGG. Gdy długość sekwencji powtórzeń mieści się między 55 a 200, stan ten nazywany jest premutacją genu *FMR1* i dotyczy tzw. stanów chorobowych związanych z premutacją w chromosomie X (ang. *fragile X-premutation-associated conditions, FXPAC*). Te stany obejmują m.in. neurodegeneracyjną chorobę wieku późnego zwaną zespołem drżenia i ataksji związanym z łamliwym chromosomem X (FXTAS) oraz zespół przedwczesnego wygasania funkcji jajników (FXPOI), prowadzący do przedwczesnej menopauzy. Stan, w którym liczba powtórzeń CGG przekracza 200 określa się mianem pełnej mutacji i stanowi on podłoże genetyczne zespołu łamliwego chromosomu X (FXS). FXS jest chorobą neurorozwojową, będąca najczęstszą przyczyną wrodzonego upośledzenia umysłowego, zwłaszcza u chłopców.

Z molekularnego punktu widzenia, w chorobach z grupy FXPAC, dochodzi do prawidłowej ekspresji białka FMRP, pomimo mutacji w genie *FMR1*, podczas gdy w FXS rejon promotorowy, zawierający ekspansję CGG, ulega hipermetylacji, prowadząc do wyciszenia genu i braku ekspresji FMRP. Uważa się, że rozwój chorób z grupy FXPAC jest efektem trzech niezależnych, wzajemnie przenikających się patomechanizmów molekularnych. Pierwszy z nich związany jest z toksycznością RNA zawierającego wydłużony ciąg powtórzeń CGG. Taki toksyczny RNA, w obrębie powtórzeń CGG tworzy strukturę drugorzędową typu spinka do włosów (ang. *hairpin*), na której mogą być sekwestrowane białka, co prowadzi do upośledzenia ich prawidłowych funkcji w komórce. Drugi mechanizm dotyczy kotranskrypcyjnego powstawania hybrydowych struktur typu pętla R (hybryda DNA:RNA), które prowadzą do akumulacji pęknięć w DNA, stwarzając zagrożenie dla utrzymania integralności genomu komórki. Trzecim mechanizmem jest niekanoniczna biosynteza białka

nazwana zależną od powtórzeń translacją niewymagającą kodonu inicjatorowego ATG (ang. *repeat-associated non-ATG translation*; RAN), w skrócie translacją RAN. Proces ten prowadzi do powstawania toksycznych białek zawierających ciągi monoaminokwasowe kodowane przez sekwencję powtórzeń CGG, które mają tendencję do agregacji. Białka te składają się z powtózonego ciągu aminokwasów jednego rodzaju, np. glicyny (kodon GGC), tworząc tzw. białka poliglicynowe (FMRpolyG). Białko FMRpolyG gromadzi się w postaci złogów wewnątrzjądrowych w komórkach pacjentów, stanowiąc istotny element patogenezy FXPAC, poprzez zwiększenie śmiertelności komórek.

Ponieważ dokładny mechanizm translacji RAN nie jest jeszcze w pełni zrozumiały, celem niniejszej pracy doktorskiej było poszukiwanie nowych modyfikatorów tego procesu. W ramach projektu zastosowano system znakowania cząsteczek RNA zawierających 99 powtórzeń CGG w 5'UTR *FMR1* oraz identyfikację białek, które związały się z badanym transkrypcyjnym w komórkach, za pomocą spektrometrii mas. W wyniku przeprowadzonych badań zidentyfikowano ponad 60 białek, które w warunkach natywnych wiążą się do rejonu 5'UTR RNA *FMR1* zawierającego powtórzenia CGG. Analiza ontologii genów wykazała, że większość zidentyfikowanych białek należy do klasy białek wiążących RNA oraz takich, które uczestniczą w procesach związanych z biogenezą rybosomu, translacją oraz procesowaniem cząsteczki mRNA. Niektóre z zidentyfikowanych białek znajdują potwierdzenie w literaturze na temat interakcji mRNA *FMR1*, a zastosowana w tej pracy technologia pozwoliła na identyfikację szeregu nowych białek podlegających interakcji z badanym RNA.

Spośród zidentyfikowanych białek wybrano dziesięciu kandydatów i przetestowano ich zdolność do regulacji procesu translacji RAN, wykorzystując technikę wyciszania genów z zastosowaniem krótkich interferujących RNA. Wyciszenie kilku z tych białek obniżyło poziom toksycznego białka FMRpolyG. Na przykład, wyciszenie białka rybosomalnego eS26 małej podjednostki 40S rybosomu (ang. *small ribosomal subunit protein eS26*; RPS26), które ze względu na lokalizację w pobliżu kanału mRNA kontaktuje się podczas translacji lub skaningu z tymi RNA, spowodowało zmniejszenie ilości białka poliglicynowego w kilku niezależnych modelach

komórkowych. Dodatkowo, niedobór dwóch helikaz RNA, DHX15 (ang. *ATP-dependent RNA helicase DHX15*) oraz DDX21 (ang. *Nucleolar RNA helicase 2*), a także czynnika transportującego ALYREF (ang. *THO complex subunit 4*) negatywnie wpłynął na biosyntezę białka FMRpolyG. Co istotne, wyciszenie tych białek nie miało wpływu na poziom białka FMRP, wskazując zdolności do specyficznej regulacji otwartej ramki odczytu białka FMRpolyG. Ponadto, wyciszenie RPS26 doprowadziło do ograniczenia tworzenia się złogów białkowych, powodując częściowe zniesienie toksyczności białek poliglicynowych w modelu komórkowym. Dodatkowo, ilościowa analiza proteomu komórek linii HEK293 po wyciszeniu RPS26 ujawniła, że tylko niewielka liczba białek jest wrażliwa na niedobór RPS26. Analiza transkryptów kodujących te białka wykazała wzbogacenie nukleotydów guaninowych i cytozynowych w rejonie 5' UTR, co sugeruje podobieństwo biochemiczne tych transkryptów do mRNA *FMR1*.

W celu lepszego zrozumienia mechanizmu translacji RAN, zweryfikowano także funkcję czynnika TSR2 – białka opiekuńczego RPS26 (ang. *Pre-rRNA-processing protein TSR2 homolog*). Wykazano, że TSR2 pozytywnie wpływa na proces biosyntezy FMRpolyG. Ponadto, wykazano, że inne białko małej podjednostki rybosomu, RPS25 (ang. *Small ribosomal subunit protein eS25*) również reguluje proces translacji RAN białka FMRpolyG.

Podsumowując, przeprowadzone badanie przesiewowe w oparciu o analizę proteomiczną, pozwoliło zidentyfikować pulę białek oddziałujących ze zmutowanym *FMR1* zawierającym ciąg powtórzeń CGG, co stanowi cenne źródło wiedzy na temat biologii tej cząsteczki RNA. Głównym osiągnięciem pracy doktorskiej jest identyfikacja pięciu nowych modulatorów translacji RAN oraz propozycja koncepcji, zgodnie z którą skład małej podjednostki rybosomu odgrywa istotną rolę w regulacji niekanonicznej syntezy białka poliglicynowego – czynnika patogenetycznego w zespołach chorobowych związanych z premutacją genu *FMR1*.

ABSTRACT

Fragile X-premutation-associated conditions (FXPAC) are genetic diseases caused by dynamic mutations of the fragile X messenger ribonucleoprotein 1 gene (*FMR1*) located on the X chromosome encoding fragile X messenger ribonucleoprotein 1 protein (FMRP). The gene usually contains 25–35 CGG repeats in the 5'-untranslated region (5'UTR). However, these triplet repeats are highly polymorphic and tend to expand. The premutation (PM) state of CGG expansion (CGGexp) corresponds to 55–200 repeats and is associated with multiple FXPAC such as a late onset neurodegenerative disease called fragile X-associated tremor/ataxia syndrome (FXTAS) and fragile X-associated primary ovarian insufficiency (FXPOI). On the contrary, when CGGexp exceeds 200, it is called full mutation and underlies the pathogenesis of neurodevelopmental disease named fragile-X syndrome (FXS), the most common form of inherited intellectual disability.

In case of FXPAC, FMRP protein is produced despite the presence of CGGexp in 5'UTR of *FMR1*, whereas, in FXS, methylation of *FMR1* promoter leads to the gene silencing. It is postulated that the interplay between three major molecular pathomechanisms drives FXPAC. At first, RNA with CGGexp is toxic and forms a secondary hairpin structure that sequesters RNA binding proteins into RNA foci leading to their functional depletion. Second, co-transcriptional formation of R loops, a DNA:RNA hybrids which triggers DNA breakage and compromises genomic stability. Third mechanism involves repeat-associated non-ATG initiated (RAN) translation, which leads to the production of toxic and aggregation-prone proteins called RAN proteins, which contain long tract of a repeated monoamino acid, that can be either polyglycine, polyalanine or polyarginine depending on the open reading frame, however polyglycine-containing proteins (FMRpolyG) predominate. Toxic FMRpolyG aggregates and forms intranuclear inclusions in patient's brain, a hallmark of FXTAS pathology.

Mechanistic insights into RAN translation remain elusive, therefore we sought to identify novel RAN translation modifiers, which constitutes the main part of the PhD thesis. We applied an in cellulo RNA tagging system combined with mass spectrometry (MS) based protein identification and discovered more than 60 proteins that bind to

5'UTR of mutant *FMR1* mRNA containing CGGexp. Gene ontology analysis performed on identified proteins revealed that majority of them represent RNA binding properties, are involved in ribosome biogenesis, translation or mRNA processing. Some of identified proteins overlapped with already identified interactors of *FMR1*, however our dataset contains newly identified factors.

Among identified proteins, we selected ten candidates and we verified their RAN translation regulatory properties using small interfering RNA. As a result, we identified few proteins, which depletion affected the level of FMRpolyG. For instance, depletion of small ribosomal subunit protein eS26 (RPS26), a component of 40S subunit which contacts mRNA sequence during translation, significantly impeded RAN translation in multiple tested models. In addition, insufficiency of two RNA helicases, ATP-dependent RNA helicase DHX15 (DHX15) and Nucleolar RNA helicase 2 (DDX21) as well as THO complex subunit 4 (ALYREF) negatively affected biosynthesis of FMRpolyG. Additionally, silencing of RPS26, DDX21 and ALYREF did not affect the level of FMRP indicating specificity of regulation towards FMRpolyG frame. Importantly, we showed that depletion of RPS26 decreases the amount of aggregates formed by FMRpolyG and alleviated their toxicity in cellular model. In addition, using quantitative MS approach, we found that the number of proteins produced by RPS26-sensitive translation is limited and that the 5'UTRs of the mRNAs encoding these proteins are rich in guanosine and cytosine, similar to *FMR1* mRNA.

In order to gain mechanistic insights into RPS26-sensitive translation regulation, we verified the function of RPS26 chaperone, Pre-rRNA-processing protein TSR2 homolog (TSR2). We demonstrated that TSR2 positively regulated the production of FMRpolyG. Finally, we verified that another component of the 40S subunit, Small ribosomal subunit protein eS25 (RPS25) also regulates CGG-related RAN translation.

To sum up, performed MS-based screening provided a unique and valuable source of information about *FMR1* interacting proteins in cellulo, which may be implicated in the biology of this molecule. The main achievement of this study is the identification of five novel RAN translation modifiers and the proposal of a concept suggesting that the composition of the 40S subunit plays a pivotal role in regulating noncanonical CGG-related RAN translation in FXPAC.

1. INTRODUCTION

1.1 *Expanded microsatellites in human genome can cause several neurological diseases*

Short tandem repeats (STRs) also known as microsatellites are repeated tracts of three to seven base pairs such as CTG, CGG or GGGGCC (1). More than 1.5 million STRs have been identified in coding and non-coding regions and constitute about 3% of human genome (2, 3). When microsatellites are expanded to a critical, pathological number, they contribute to the development of almost 50, neurodevelopmental, neurodegenerative and neuromuscular diseases named repeats-expansion related disorders (REDs) (1, 4, 5). For example, expanded CTG repeats in 3' untranslated region (UTR) of *dystrophia myotonia protein kinase* gene (*DMPK*) underlies the myotonic dystrophy type 1 (DM1) (6). Pathological number of CAG repeats present in the exon 1 of *huntingtin* (*HTT*) gene contribute to the development of Huntington's disease (HD) (7). Expansion of GGGGCC (G4C2) repeats in intronic region of *c9orf72* gene underlies amyotrophic lateral sclerosis and/or frontotemporal dementia (C9 ALS/FTD) (8). Other diseases such as several spinocerebellar ataxias (SCA) and fragile X syndrome (FXS) also belong to REDs (5).

The STRs are unstable, prone to expand and to a lesser extent contract. It is thought that microsatellites cause replication stalling and are hotspots of chromosomal double strand breaks (4, 9). Repeats instability occurs predominantly in germline tissue, often leading to inheritance of longer repeats tract in progeny (10). Such phenomenon is known as genetic anticipation and inherited trait manifests in more severe disease phenotype beginning earlier in the offspring's life (11). The repeats are also unstable in somatic cells causing tissue-specific somatic mosaicism (10, 12).

Repeats expansion above a certain threshold can lead to the gene loss or gain of function phenomenon (13). The gene loss of function manifests in gene silencing. For instance, loss of fragile X messenger ribonucleoprotein 1 protein (FMRP) drives pathology in FXS (14). The gene gain of function correlates with acquiring new deleterious features by an RNA or a protein triggering disease development (13). For example, mutated Huntingtin in HD serves as an example of toxic protein gain of

function and in case of RNA gain of function, the great example is toxic RNA with expanded CUG repeats in DM1, which sequesters several cellular proteins leading to their functional depletion (15).

1.2 Genetic basis and phenotype of fragile X-associated syndromes

Fragile X messenger ribonucleoprotein 1 gene (*FMR1*) in 5'UTR contains CGG repeats that are prone to breakage and were linked to so-called fragile site on the X chromosome in the late 1960s (16). The CGG STR in this locus is conserved in mammals, and the average size of CGG repeats in humans varies mostly between 25 and 30 (17) (Figure 1.1). In 1991 the CGG expansion (CGGexp) in the *FMR1* was independently identified by four research groups and described as the causative mutation of neurodevelopmental disease known as FXS (18–21). When CGGexp exceeds 200, it induces hypermethylation of the neighboring *FMR1* promoter, leading to the transcriptional silencing and consequently lack of the primary gene product, FMRP (14). FMRP is an RNA binding protein with known implications in synaptic plasticity (22) and its insufficiency is believed to cause autism and other forms of cognitive and intellectual disability in FXS patients (23).

The premutation (PM) of *FMR1* gene corresponds to 55-200 CGG repeats expansion in 5'UTR of *FMR1* and underlies fragile X-premutation-associated conditions (FXPAC) (24). FXPAC is an umbrella term, which covers three distinct medical conditions associated with PM: fragile X-associated tremor/ ataxia syndrome (FXTAS) (25), fragile-X-associated primary ovarian insufficiency (FXPOI) (26) and fragile X-associated neuropsychiatric disorders (FXAND) (27). The estimated prevalence of PM is 1 in 150–300 women and 1 in 400–850 men. However, due to incomplete penetrance of the mutation, approximately 1 in 5,000–10,000 men will develop some form of FXPAC in their fifties or later, while random X inactivation in woman PM carrier, greatly reduces the risk of disease manifestation (28–30). On molecular level, in opposite to FXS, FMRP expression is not affected or slightly decreased in FXPAC, however the *FMR1* mRNA is elevated (31–33).

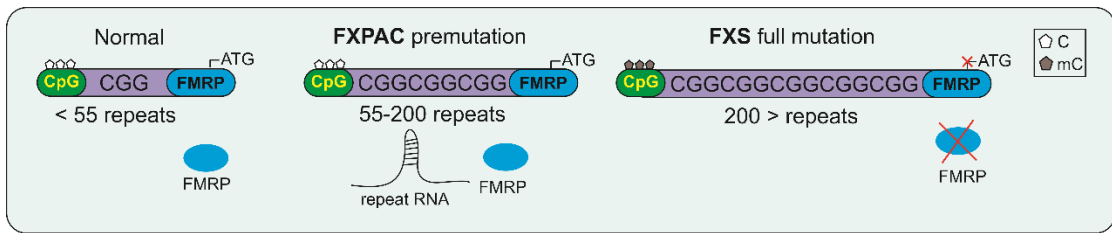


Figure 1.1. CGG repeats instability in *FMR1* locus and its molecular implications. Schematic representation of CGG repeats expansion in 5'UTR of *FMR1* and its molecular consequences in FXS and FXPAC; CpG, refers to CpG island in *FMR1* promoter, mC indicates methylation of cytosine.

Fragile X-associated tremor/ ataxia syndrome (FXTAS)

In early 2000s, FXTAS was described as a late onset, incurable neurodegenerative disease that affects mostly men (30, 34, 35). The main clinical features of FXTAS include intention tremor, cerebellar ataxia, neuropathy, balance problems, parkinsonian features, dementia as well as cognitive decline, memory problems and executive function deficits (28, 34). The pathological hallmark of the disease are ubiquitin-positive intranuclear or perinuclear inclusions in nerve cells (neurons and glia) (36–39) and neurodegeneration resulting in mild brain atrophy and white matter lesions (35, 40). FXTAS progresses with age and so far no effective cause-oriented treatment is available (28).

Fragile-X-associated primary ovarian insufficiency (FXPOI)

Primary ovarian insufficiency (POI) is described as premature cessation of ovary function leading to early menopause before the age of 40 (41). It is estimated that female PM carriers are 20% more likely to develop POI in their lifetime and suffer from infertility problems (42). Similarly to FXTAS, FMRpolyG and ubiquitin positive inclusions were found in ovaries of FXPOI patients as well as in mice model which suggests one of the disease pathomechanism (43, 44). In addition, dysfunctional hypothalamic–pituitary–gonadal-axis may contribute to FXPOI development (43). Similarly to FXTAS, only symptomatic treatment is available for FXPOI patients.

Fragile X-associated neuropsychiatric disorders (FXAND)

The neuropsychiatric disorders such as anxiety, depression, social defects, chronic fatigue, sleep disturbances but also obsessive compulsive behavior are the most common problems of psychological and/or psychiatric nature reported by almost 50% of PM patients. PM carriers who suffer from those symptoms often do not meet the criteria for FXTAS and/or FXPOI, thereby demand different treatment. Such high occurrence of neuropsychiatric symptoms led to a proposal of separate, novel term - FXAND in order to enhance recognition and promote effective research in that area (27).

1.3 Molecular pathomechanisms of CGG repeats expansion in *FMR1*

1.3.1 Secondary structures formed within expanded CGG repeats in mutant *FMR1* mRNA

High guanosine (G) and cytosine (C) content of *FMR1* RNA determinates the structural features of this molecule. Complementary binding between C and G bases is recognized as the strongest pairing between nucleotides and leads to formation of thermodynamically stable secondary structures, either hairpins and/or G-quadruplexes (structures formed by the stacking of several planar guanine quadruplets) within expanded CGG repeats in *FMR1* mRNA (45–47).

Structured, mutant GC-rich *FMR1* RNA can cause several, deleterious molecular consequences for the cell (Figure 1.2). First, high GC content triggers R loop formation, a DNA:RNA hybrid formed during transcription (48). Second, sequestration of RNA binding proteins on *FMR1* RNA containing CGGexp disrupts their functionality (49, 50). Third, mutant *FMR1* serves as template for noncanonical FMRpolyG protein synthesis *via* mechanism named repeat-associated non-ATG (RAN) translation (40). It is considered that the interplay between three main molecular pathomechanisms related to toxic RNA and protein gain of function phenomenon is involved in FXPAC development and progression (Figure 1.2) (51, 52).

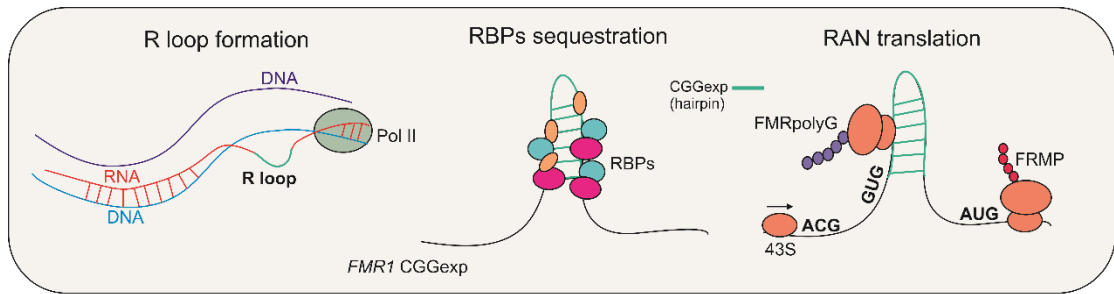


Figure 1.2. Three major molecular pathomechanisms contributing to FXPAC. Synergy between R-loop formation, RNA binding proteins sequestration and FMRpolyG biosynthesis drive FXPAC pathogenesis; Pol II – polymerase II, RBPs – RNA binding proteins, AUG – canonical start codon, ACG and GUG – noncanonical, near-cognate RAN translation start codons, 43S – preinitiation complex.

1.3.2 R loop formation

R loops are three stranded DNA:RNA hybrids formed behind elongating Polymerase II (Pol II), when the binding between nascent RNA and the template DNA (RNA:DNA) is more thermodynamically advantageous than binding between template single-stranded DNA (DNA:DNA) (53). Such situation occurs frequently on GC rich DNA template (54) and high GC content of *FMR1* gene promotes the Polymerase II (Pol II) co-transcriptional formation of R loops (48, 55). They can be formed in CpG islands (CGIs) of *FMR1* promoter region and/or within expanded CGG repeats (48, 53, 56). One hypothesis is that R loops formed in CGIs-containing promoters can relax chromatin and promote transcription, while R loops formed over CGGexp cause Pol II stalling and result in decreased transcription efficiency (53, 57, 58). With the increase of CGG repeats number, the R loop formation occurs more frequently causing DNA breakage strongly correlated with transcriptional-replication machineries collision. Accumulation of these events can be lethal to a cell by activating DNA damage response (DDR) and threatening genomic stability (48, 55, 58, 59).

Off note, in addition to impeded transcription and activated DDR, stalled Pol II can recruit Polycomb repressive complex 2 (PRC2). Methylation of histone H3 at lysine 27 by PRC2 promotes heterochromatinization of *FMR1* locus giving possible explanation for stable gene silencing in FXS (53, 60, 61).

Altogether, that activated DDR caused by R loops in *FMR1* locus can lead to cellular death constituting one of the pathomechanism underlying PM-linked disorders (48, 55).

1.3.3 Sequestration of RNA binding proteins

Toxic RNA with expanded repeats organizes within a cell in nuclear clusters known as RNA foci, which are stabilized by RNA-RNA, RNA-protein and protein-protein interactions (62, 63). Within foci, RNA containing expanded repeats sequester cellular RNA binding proteins (RBPs) disabling their functions and dysregulating many cellular processes (49, 50, 64). Depending on repeat length and sequence, multivalent base-pairing and occupancy of sequestered RBPs, RNA foci formation can promote phase transition leading to RNA gelation observed in many REDs (65–67).

Multiple RBPs were found to be sequestered on toxic *FMR1* RNA with CGGexp. These include Microprocessor complex subunit DiGeorge Syndrome Critical Region 8 (DGCR8), Ribonuclease 3 (DRISHA) (50), Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1) (68) and Transcriptional activator protein Pur-alpha (Pur- α) (69), Src-associated in mitosis 68 KDa protein (SAM68) (49) and TAR DNA binding protein of 43 kDa (TDP-43) (70). Microprocessor complex disruption results in the deregulation of microRNA maturation and SAM68 sequestration causes aberrant splicing in FXTAS models, respectively (49, 50). Importantly, overexpression of DGCR8, Pur- α and hnRNP A2/B1 in *Drosophila melanogaster* FXTAS model reduced neurodegeneration providing a proof of concept for RNA gain of function mediated toxicity (50, 68, 69).

1.3.4 Repeat-associated non-ATG translation

In 2011, Zu and colleagues described for the first time the repeat-associated non-ATG (RAN) translation, a noncanonical protein synthesis triggered by CAG repeats expansion (linked to SCA8) and CTG repeats expansion (underlying DM1). This process was shown to be initiated in multiple reading frames without canonical AUG start codon (71). Later, RAN translation induced by other repeats: G4C2 repeats in *C9orf72* gene, CAG repeats in *HTT* and CGG repeats in *FMR1* underlying ALS/FTD, HD and FXTAS,FXPOI, respectively, was reported (38, 72–74). RAN translation derived proteins (for simplicity named here RAN proteins) usually contain homopolymeric or dipeptide sequence comprised of one or two repeated amino acids. For instance, CGGexp in

FMR1 repeats encode mostly FMRpolyG, protein with long polyglycine stretch (38), and G4C2 repeats encode dipeptides such as polyglycine and polyalanine polyGA (72). RAN proteins are toxic, prone to aggregate and their accumulation in patient's tissues is believed to contribute to the development and progression of all mentioned REDs (38, 71–75).

1.3.5 RAN translation of *FMR1* with CGGexp

The open reading frame (ORF) for FMRP resides downstream to CGG hairpin of *FMR1* mRNA (Figure 1.3). FMRP translation initiates at canonical AUG codon, whereas upstream open reading frames (uORFs) for RAN proteins begin at less-favored, near-cognate start codons such as ACG or GUG (75, 76) (Figure 3). RAN translation of *FMR1* mRNA with CGGexp results in noncanonical biosynthesis of polyglycine (FMRpolyG), polyalanine (FMRpolyA) and polyarginine (FMRpolyR) containing proteins depending on the uORF. FMRpolyG predominates, whereas FMRpolyA and FMRpolyR are less abundant in cell (77, 78). RAN translation for FMRpolyG and FMRpolyR begins upstream to CGG repeats starting from ACG and GUG codons, while FMRpolyA synthesis is initiated within CGG repeats. In addition to homopolymeric tract of repeated single amino acid, RAN proteins contain amino acids encoded upstream to CGG tract as well as peptides encoded in first and partially second exon of FMRP (38, 76). Moreover, formation of hybrids of these three RAN proteins is possible as a product of frameshifting on CGGexp sequence (79). Additionally, RAN translation also occurs on antisense *ASFMR1* transcript containing CCG repeats resulting in a synthesis of polyproline (ASFMRpolyP), polyarginine (ASFMRpolyR), and polyalanine (ASFMRpolyA) proteins (80). In fact, PM of *FMR1* gene results in production of many RAN proteins, while FMRpolyG is the most abundant and the most studied one, yet biological relevance of other RAN products is not clarified and remains questionable in the field. Regardless of uORFs of RAN proteins, majority of studies indicate that FMRP translation is independent from RAN translation event. However, Rodriquez and colleagues (2020) showed that antisense oligonucleotides treatment targeting RAN start sites impeded RAN translation and enhanced FMRP synthesis in human neurons demonstrating the native regulatory function of uORFs over FMRP frame (81).

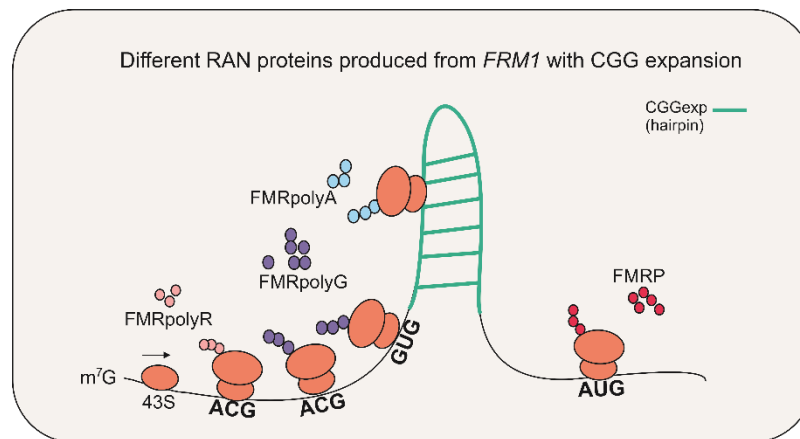


Figure 1.3. Schematic representation of canonical FMRP and noncanonical (RAN) translation of RAN proteins encoded in toxic *FMR1* RNA with CGGexp. m⁷G indicates 5′methyl-7-guanosine cap, 43S – preinitiation complex. AUG – canonical start codon, ACG and GUG – noncanonical, near-cognate RAN translation start codons.

1.4 Mechanisms and proteins implicated in the regulation of CGG-related RAN translation

Despite the important role RAN translation in multiple neurodegenerative diseases, our understanding of mechanism of this process remains incomplete. In fact, although RAN translation is 30-40% less efficient than traditional protein biosynthesis, it adapts canonical mechanism of eucaryotic initiator factors (eIF4A, eIF4E and others) recruitment to 5′ methyl-7-guanosine cap (5′ m⁷G) followed by subsequent 43S Preinitiation complex (43S PIC) scanning (76). On the other hand, in different genetic and cellular contexts RAN translation can occur in cap-independent manner utilizing tRNA initiators of translation other than methionyl-tRNA (78, 82–84). However, it is puzzling why PIC chooses near-cognate codons for translation initiation. One possible explanation refers to steric hindrance of structured CGGexp which abolishes PIC scanning and causes its stalling resulting in skewed start codon fidelity and forced translation initiation in less-favored non-AUG codons (Figure 1.4) (76). This hypothesis is supported by the increase of RAN translation rate correlated with the increase of repeats length, and the more frequent initiation at ACG codon which resides upstream to hairpin than the initiation at GUG located within hairpin (76). In addition, ribosomal stalling on structured RNA can cause translational frameshifting which results in the production of aggregation-prone chimeric proteins such as hybrid of polyarginine-polyglycine (79)

CGGexp secondary structure can be resolved by RNA helicases which positively affects RAN translation (Figure 1.4). Linsalata and colleagues (2019) showed that DEAD-box helicase DDX3X which possibly unwinds hairpin, facilitates RAN translation, as the helicase depletion resulted in the significant decline of RAN products (Figure 1.4) (85). Such results might be explained by a mechanism where DDX3X resolves hairpin paving the way for scanning 43S PIC, hence without DDX3X, near-cognate codons are hardly accessible, especially GUG codon within hairpin. In addition, Tseng and coworkers (2021) provided the evidence that another helicase, DEAH-box RNA helicase DHX36 plays a role in CGGexp G-quadruplexes and/or hairpin relaxation supporting the idea that physical hindrance formed by repeated sequences is pivotal for RAN translation initiation and efficiency (86). Altogether, one can say that RAN translation initiation is governed by fine tuning of noncanonical codons selection by 43S PIC. More complexity adds the fact that RAN translation occurs also on shorter, physiological number of repeats questioning the hypothesis that the pathological expansion of the repeats drives RAN translation (76, 81). This opens up a possibility that RAN translation may play a physiological regulatory role *in vivo* (81).

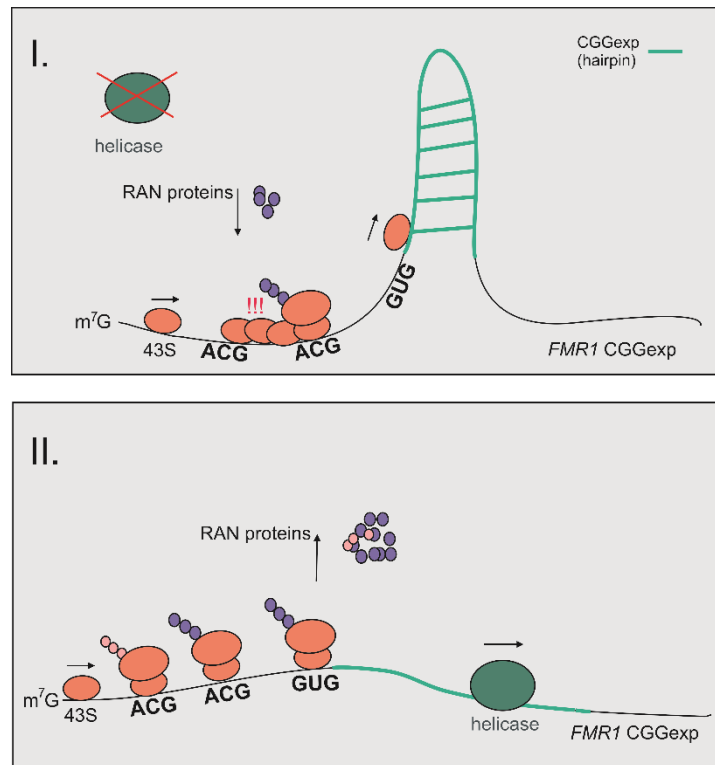


Figure 1.4. Schematic representation of two mechanisms implicated in RAN translation. I. 43S stalling while scanning (marked as !!!) caused by steric hindrance such as hairpin, induces the selection of near-cognate codons. II. Unwinding the RNA secondary structures by RNA helicases facilitates faster scanning and more frequent selection of near-cognate codons which enhances RAN translation. m⁷G indicates 5' methyl-7-guanosine cap, 43S – preinitiation complex. AUG – canonical start codon, ACG and GUG – noncanonical, near-cognate RAN translation start codons.

Expanded repeats and their secondary structures cause aberrations in translation initiation, yet the same structures may impede and abort translational elongation. This hypothesis was very recently explored pointing out that components of ribosome-associated quality control (RQC) pathway play important role in the regulation of RAN translation elongation (87). RQC pathway protects the cell from misfolded or incompletely generated polypeptides by ubiquitination and subsequent proteasomal degradation, ribosome recycling, degradation of template RNA and other processes (88). Briefly, RQC pathway consists of two sequential steps, first sensing of stalled ribosome and disassembly of 80S ribosomal subunits and second formation of RQC complex *via* initial recruitment of nuclear export mediated factor (NEMF) to 60S subunit which binds and stabilizes binding by the E3 ligase listerin (LTN1). At this stage, NEMF synthesizes carboxy-terminal alanine and threonine tails to expose lysine

residues that are inside the ribosomal exit tunnel, to be ubiquitylated by LTN1. Next, ubiquitinated, nascent peptides are released by ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1 (ANKZF1) and condemned to proteasomal degradation (89). Tseng and colleagues demonstrated that depletion of RQC complex components, NEMF, LTN1 and ANKZF1 enhanced RAN protein synthesis (87). This indicates that elongating ribosome stalls on expanded structured repeats, activates RQC pathways, which in return leads to the degradation of nascent polypeptide chain and halting RAN translation. Hence, activated RQC prevents accumulation of RAN misfolded proteins by ubiquitination and subsequent proteasomal degradation (87).

RAN protein are toxic for the cell and trigger elevated integrated stress response (ISR) (83, 90). Green and colleagues (2017) demonstrated that the presence of misfolded RAN proteins leads to the activation of endoplasmic reticulum-resident kinase (PERK), and subsequent phosphorylation of the α subunit of eukaryotic initiation factor-2 (eIF2 α) at serine 51, which is the core event of an ISR (90). Phosphorylated eIF2 α evokes cascade of events resulting in the inhibition of global, canonical protein synthesis but enhances RAN translation by alteration in start codon stringency and initiation kinetics. At the end, elevated ISR stimulates the formation of phosphorylated-eIF2 α -dependent stress granules causing neuronal death. Hence, authors proposed a deleterious feedforward loop, where RAN proteins activate ISR, leading to global translation shut down and selective enhancement of RAN translation (90).

In addition to more direct regulatory process such as start codon fidelity, RAN translation can be modulated in an indirect way, be retention of transcript with CGGexp in the nucleus. Malik and colleagues (2021) identified direct interaction between serine/arginine splicing factor 1 (SRSF1) and *FMR1* mRNA (91). They also showed that impeded activity of serine/arginine protein kinase 1 (SRPK1), which normally phosphorylates SRSF1, results in reduction of SRSF1 nuclear import, which in return halts *FMR1* transcripts in the nucleus. It was also demonstrated that pharmacological inhibition of SRPK1 prevented stress-induced enhancement of RAN translation. Hence, SRPK1 modulates translocation of toxic RNA preventing it from becoming a template for RAN translation in cytoplasm (91).

In conclusion, several regulatory pathways contribute to RAN translation regulation starting with transport, unwinding of structured RNA, start codon fidelity, RQC and stress response, however continuation of mechanistic studies underpinning noncanonical protein synthesis may broaden our understanding of this process and thus open up novel therapeutic perspectives.

Please note that other proteins implicated in RAN translation regulation not presented here are reviewed in “*Partners in crime: Proteins implicated in RNA repeat expansion diseases*”, Baud A et al., 2022, that constitutes the first part of PhD thesis (please find chapters entitled *RAN Translation, Immune response to RNA^{exp} and products of RAN translation, Targeting RBPs as potential therapeutic strategies*).

1.5 FMRpolyG driven pathology in PM-linked conditions

In FXTAS patients, FMRpolyG and ubiquitin-positive intranuclear inclusions are predominantly present throughout the brain in neurons, astroglia and Purkinje cells of cerebellum, however they can be also found in other tissues (Figure 1.5) (36–38, 75, 92–94). Except FMRpolyG and to a lesser extend FMRpolyA (38, 75, 77), the exact composition of intranuclear inclusions is not fully determined and varies in different studies, however, a pool of proteins interacting with FMRpolyG and/or present in inclusions was identified by few research groups (39, 67, 75, 93). These include Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1), Musceblind like 1 (MBNL1), Pur- α , Lamin A, Nucleolin, several heat shock proteins and other mitochondrial, cytoskeleton, proteasome and exosomal proteins. The exact sequence of inclusion formation is unresolved, however, one hypothesis is that toxic RNA acts as a scaffold for multistep RNA-RBPs complex assembly (50, 68, 95). For instance, it was shown that hnRNP A2/B1 directly interacts with CGGexp and that the CUGBP1 protein interacts with toxic RNA *via* hnRNP A2/B1 (68). Similarly, initial sequestration of DGCR8 and DORSHA to CGGexp is required for subsequent SAM68 recruitment (50). In addition, Asamitsu and colleagues (2021) have shown the direct interaction between pathogenic CGG repeat-derived RNA G-quadruplexes (CGG-G4RNA) and FMRpolyG (67) *via* the polyglycine region of FMRpolyG, promoting liquid-to-solid transition and aggregation of FMRpolyG. This suggested that FMRpolyG forms a

complex with RBPs and exosomal proteins *via* toxic RNAs (67). On the contrary, Hoem and colleagues (2019) showed that FMRpolyG itself, without co-expression of the CGGexp RNA hairpin can form aggregates and induce toxicity in several cell lines (96).

RAN translation occurs in cytoplasm, yet FMRpolyG-positive aggregated are found in nucleus. How these aggregates are transported from the cytoplasm to the nucleus is unclear. One could assume that aggregates are mobile and travel along microtubules, however this hypothesis was denied as no colocalization between aggregates and microtubule organizing center was observed (96). In addition, soluble, diffused fraction of FMRpolyG in cytoplasm and nucleus stays mobile until it attaches to aggregate, and becomes immobile, indicating possible snow-ball-like dynamic of aggregate formation (96). Altogether, the complex nature of FMRpolyG aggregates require further investigation.

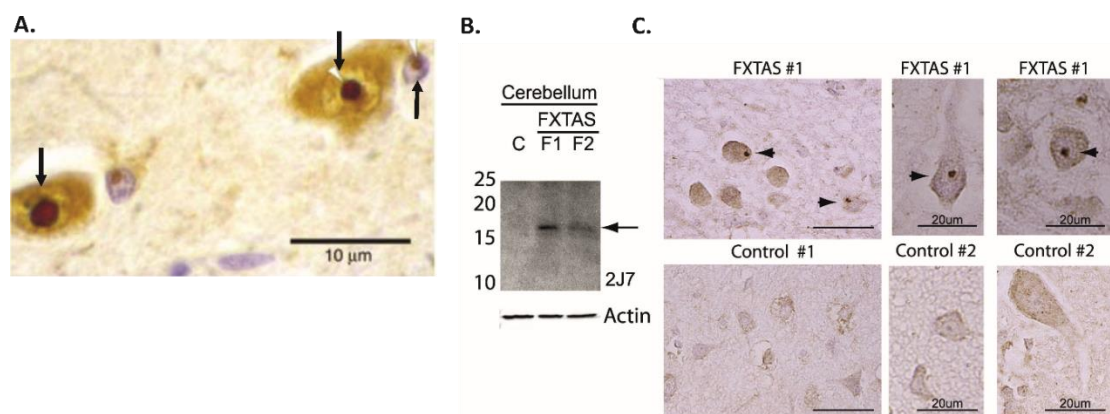


Figure 1.5. FMRpolyG detected in FXTAS patient's tissues. **A.** Intranuclear inclusions stained with anti-ubiquitin antibodies detected in FXTAS patient's neurons (indicated by arrows) **B.** Western blot with 2J7 anti-FMRpolyG antibody of cerebellar lysates from two FXTAS patients and an age-matched control. Arrow indicate FMRpolyG of expected size. **C.** Images of 2J7 immunostaining from frontal cortex and hippocampus of control and FXTAS brain. Arrows indicate FMRpolyG-positive inclusions. Adapted from Greco C et al., 2002 (A) and Todd P et al., 2013 (B & C).

Regardless of well documented presence of intranuclear inclusions in FXTAS cells, the mechanism explaining their contribution to the neurodegeneration remains uncovered. One possible explanation relates to disruption of the normal architecture of lamin A/C within the nucleus of cell expressing toxic RNA (97), which was further

supported by the discovery of interaction between FMRpolyG and Lamina-associated polypeptide 2 beta (LAP2 β), which disorganizes nuclear lamina architecture in neurons derived from FXTAS patient's induced pluripotent stem cells (iPSCs) (75). Overexpression of LAP2 β rescues neurodegeneration induced by expression of FMRpolyG indicating that destabilized lamina architecture contributes to FXTAS pathogenesis (75). In addition to lamina disruption, it was demonstrated that FMRpolyG can be sorted to exosomes, which are further propagated from cell-to-cell, negatively affecting synaptic plasticity and thus eliciting neuronal dysfunction (67). Other hypothesis regarding FMRpolyG toxicity refers to protein quality control pathways. Ubiquitin and proteasomal proteins were detected in intranuclear inclusions (93) and FMRpolyG is a stable protein that is primarily degraded *via* the ubiquitin proteasome system (UPS) (96). However, it has been shown that RAN translation and accumulation of RAN proteins leads to impairment of the UPS (98). Hence, these results suggest a mechanism in which RAN translation of FMRpolyG leads to the inability of the UPS to remove RAN proteins from the cell (96, 98).

In FXPOI, FMRpolyG-positive inclusions were found in ovary tissues and granulosa cells derived from PM women carriers and their toxicity was directly confirmed in FXPOI mice models showing significant fertility impairment in the presence of FMRpolyG (43, 44, 99). In fact, Shelly and colleagues (2021) showed that sole expression of toxic RNA was sufficient enough to impair responses to hormonal stimulation which lead to ovarian dysfunction, but expression of both RNA with CGGexp and FMRpolyG lead to premature cessation of breeding (44).

FMRpolyG-positive inclusions for many years constituted the hallmark of the disease, although, recently it was shown that the presence of RAN proteins-positive aggregates was not sufficient to induce clear behavioral disease phenotype in FXTAS mice model (100). In addition, the native FMRpolyG is hardly detectable in human tissue using MS and its quantification was possible after extensive enrichment from the analyzed patient's material (39, 78). This could be explained by the limitations of FMRpolyG detection as it was demonstrated that for successful identification a minimum 60 to 70 CGG repeats are required (75) and many patients with PM display less than 70 repeats (29). Technical limitations also play a role as RAN proteins are

small and proteolytic digestion results in very few unique peptides (39, 78). On the contrary, Sellier and colleagues showed that in FXTAS mice the sole expression of CGGexp RNA was not pathogenic, whereas FMRpolyG was required to induce locomotor deficiency (75). These findings opened a debate, if and to what extent RAN translation contributes to FXTAS pathology and what is more toxic, RNA containing CGGexp or deleterious RAN protein, or is it an interplay between those two entities that drive pathogenesis? (101) Given that FMRpolyG was initially detected by immunostaining (36, 38, 75) and other methods seem to fail or underdeliver RAN proteins identification, it is necessary to develop a reliable methodology to detect and quantify FMRpolyG to answer this question in future (39, 78).

1.6 Ribosomal heterogeneity in a nutshell

Central dogma of gene expression underlies our understanding how encoded genetic information is deciphered, leading to the biosynthesis of a final gene product. For decades, it was believed that transcribed RNAs are translated by well-defined ribosomal machinery, however, with the development of new technologies, the last two decades brought up new perspectives to this subject. It was shown that composition of the ribosome may vary and this structural rearrangements result in the translation of different pools of mRNA shifting our understanding about gene expression towards ribosome-mediated control of gene regulation. Such phenomenon named ribosomal heterogeneity or specialized ribosomes was initially described in invertebrate model organisms, although, today we know that this process is conserved among different organisms and occurs also in mammalian cells (102, 103). It is important to highlight that majority of ribosomal proteins are necessary for proper ribosomal functionality but some ribosomal proteins may be disassembled from the ribosome without harming translational efficiency, yet determining what RNAs can or cannot be translated (104–106). For instance, ribosomes containing small ribosomal subunit protein eS25 (RPS25) or Large ribosomal subunit protein uL1 (RPL10A/uL1) preferentially translate transcripts encoding proteins implicated in development, cell cycle or vitamin B12 metabolism (104).

Ribosomal heterogeneity has the significant consequences for body patterning and organism development overall (107, 108). The great example is Ribosomal Protein 38L (RPL38), which selectively facilitates translation of *homeobox (Hox)* mRNAs – crucial for formation of body plan (109, 110). Kondrashov and colleagues (2011) demonstrated that mutations in RPL38 led to abnormal body patterning in mice (109). Of similar importance is a role of RPL10A in the production of mesoderm lineage (111). Additionally, it was demonstrated that several ribosomal proteins are differentially expressed throughout the organism, for instance, ribosomal protein L22e family members in *Drosophila* (112). These RPLs can be alternatively spliced and differentially expressed in the organism. While one paralog is ubiquitous, the other has tissue-specific expression (112). Moreover, mutations of ribosomal proteins encoding genes result in the development of tissue-specific, severe congenital diseases called ribosomopathies (113). Well known example is Diamond-Blackfan Anemia (DBA), which is linked to mutations in gene encoding Small ribosomal subunit protein eS10 (RPS10) and Small ribosomal subunit protein eS26 (RPS26) (114, 115). It was shown that depletion of RPS26 from the cell abolishes erythropoiesis which may drive DBA (116). Ribosomal composition may also affect translation of RAN proteins from expanded STRs. For example, Yamada and colleagues (2019) demonstrated the role of RPS25 in RAN translation regulation. They showed that RPS25-depleted ribosomes less efficiently translate RAN peptides derived from G4C2 and CAG repeats alleviating RAN-translation related toxicity in several C9 ALS/FTD models (117). Importantly authors showed that RPS25 is non-essential for ribosome functionality as its depletion from translational machinery did not affect global translation efficiency (117).

In addition to the role of ribosomal heterogeneity in the development and disease, it was demonstrated that ribosomal composition changes upon environmental stimuli. For instance, in yeast, high-salt and high-pH stress induces the release of Rps26 from mature ribosomes by its chaperone pre-rRNA-processing protein Tsr2, enabling the translation of mRNAs engaged in stress response pathways (105, 106, 118). In eukaryotic cells, RPS26 is also involved in stress responses, although on the contrary to the yeast model, RPS26 remains associated to the ribosome under energy stress (119). It was demonstrated that cells with mutated C-terminus of RPS26

were more resistant to glucose starvation, than the wild type cells due to modulation of energy metabolism mediated by changes in the translome induced by RPS26-dependent mechanism (119). Remarkably, C-terminal domain of RPS26 was shown to specifically interact with nucleotides upstream to AUG codon near E-site of actively translating ribosome (118, 119), most likely contributing to the selection of transcripts undergoing effective translation, however the exact motif facilitating the interaction between RPS26 and RNA is not determined.

To sum up, biological relevance of specialized ribosomes is robust and extensive studies are actively conducted to better understand this phenomenon and its implications.

2. AIMS OF THE STUDY

Fragile X premutation-associated conditions (FXPAC) are linked to the pathogenic expansion of 55-200 CGG repeats in 5'UTR of *FMR1* gene. One of the pathomechanisms underlying FXPAC is noncanonical protein synthesis named repeat-associated non-ATG (RAN) translation of CGG repeats which leads to the production of toxic RNA proteins containing a tract of single, repeated amino acid, for instance, the polyglycine-containing protein named FMRpolyG. FMRpolyG is prone to aggregate and form intranuclear inclusions in the patient's tissues, which is believed to be one of the factors driving disease progression. Although RAN translation was discovered more than a decade ago, the exact mechanism and factors involved in CGG-related RAN translation regulation are yet to be uncovered, hence further studies concerning noncanonical protein synthesis in FXPAC are required. Therefore, **the main aim of the study was to identify novel modifiers of CGG-related RAN translation.**

The specific goals of the study were to:

1. Identify proteins interacting with the 5' UTR of *FMR1* containing expanded CGG repeats in cellulo, using mass spectrometry-based screening.
2. Validate the potential regulatory properties of selected candidates in modulating RAN translation by using targeted siRNAs for gene silencing.
3. Investigate the mechanism underlying RAN translation regulation facilitated by the factors that have been verified in the previous step.

The PhD thesis is structured into **three parts**. **The first part** comprises a comprehensive literature review that synthesizes current knowledge on RNA binding proteins implicated in disease-associated repeat expansion. This review was authored by Baud A, Derbis M, Tutak K & Sobczak K and was published in *Wiley Interdisciplinary Reviews: RNA* in 2022 (Chapter 3). The second and third parts of the dissertation present the results pertaining to the primary objective of the study. **Part two** is the manuscript titled "*Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*" authored by Tutak K et al., published as a preprint on bioRxiv server (Chapter 4). **Part three** titled "*Characterization of other factors affecting noncanonical polyglycine synthesis from*

mutant FMR1 mRNA” constitutes unpublished results which detail the identification and characterization of additional proteins identified through MS-based screening (Chapter 5). Furthermore, it discusses the verification of their regulatory properties in regulating RAN translation.

3. THE LITERATURE REVIEW: *Partners in crime: Proteins implicated in RNA repeat expansion diseases*

OVERVIEW

Partners in crime: Proteins implicated in RNA repeat expansion diseases

Anna Baud  | Magdalena Derbis | Katarzyna Tutak  | Krzysztof Sobczak 

Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan, Poland

Correspondence

Krzysztof Sobczak, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Uniwersytetu Poznańskiego 6, Poznan, Poland.
Email: ksobczak@amu.edu.pl

Present address

Magdalena Derbis, Department of Molecular Probes and Prodrugs, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, Poznan, Poland

Funding information

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement No. 101003385 (to Anna Baud), Polish National Science Centre No. 2019/35/D/NZ2/02158 (to Anna Baud) and No. 2020/38/A/NZ3/00498 (to Krzysztof

Abstract

Short tandem repeats are repetitive nucleotide sequences robustly distributed in the human genome. Their expansion underlies the pathogenesis of multiple neurological disorders, including Huntington's disease, amyotrophic lateral sclerosis, and frontotemporal dementia, fragile X-associated tremor/ataxia syndrome, and myotonic dystrophies, known as repeat expansion disorders (REDs). Several molecular pathomechanisms associated with toxic RNA containing expanded repeats (RNA^{exp}) are shared among REDs and contribute to disease progression, however, detailed mechanistic insight into those processes is limited. To deepen our understanding of the interplay between toxic RNA^{exp} molecules and multiple protein partners, in this review, we discuss the roles of selected RNA-binding proteins (RBPs) that interact with RNA^{exp} and thus act as “partners in crime” in the progression of REDs. We gather current findings concerning RBPs involved at different stages of the RNA^{exp} life cycle, such as transcription, splicing, transport, and AUG-independent translation of expanded repeats. We argue that the activity of selected RBPs can be unique or common among REDs depending on the expanded repeat type. We also present proteins that are functionally depleted due to sequestration on RNA^{exp} within nuclear foci and those which participate in RNA^{exp}-dependent innate immunity activation.

Abbreviations: 3'UTR/5'UTR, 3'/5' untranslated regions; AS, alternative splicing; ASO, antisense oligonucleotide; C9-ALS/FTD, *C9orf72*-linked amyotrophic lateral sclerosis and frontotemporal dementia; CAG^{exp}, expanded CAG repeats; CCG^{exp}, expanded CCG repeats; CGG^{exp}, expanded CGG repeats; CTG^{exp}, expanded CTG repeats; CUG^{exp}, expanded CUG repeats; DM1, myotonic dystrophy type 1; DM2, myotonic dystrophy type 2; DPRs, dipeptide-repeat proteins; FDA, Food and Drug Administration; FRDA, Friedreich's ataxia; FXN, frataxin; FXS, fragile X syndrome; FXTAS, fragile X-associated tremor/ataxia syndrome; G4C2^{exp}, expanded GGGGCC repeats; HD, Huntington's disease; hnRNPs, heterogeneous nuclear ribonucleoproteins; HTT, huntingtin; iPSCs, induced pluripotent stem cells; iPSNs, induced pluripotent stem cells-derived neurons; IR, intro retention; KH, K homology domain; LLPS, liquid-liquid phase separation; MBNLs, muscleblind-like proteins; MLO, membraneless organelles; NCT, nucleocytoplasmic transport; REDs, repeats expansion disorders; NSAIDs, nonsteroidal anti-inflammatory drugs; PAF1C, PAF1 complex; P-bodies, processing bodies; PBZ, phenylbutazone; PERK, protein kinase R-like ER kinase; PIC, 43S preinitiation complex; PKR, protein kinase R; Pol II, RNA polymerase II; polyG, polyglycine; polyGA, polyglycine-alanine; polyGP, polyglycine-proline; polyQ, polyglutamine; RAN translation, repeat associated non-AUG translation; RBD, RNA binding domain; RBPs, RNA binding proteins; RNA^{exp}, RNA containing expanded repeats; RNP, ribonucleoprotein complexes; RRM, RNA recognition motif; SCA, spinocerebellar ataxia; SG, stress granules; STRs, short tandem repeats.

Anna Baud, Magdalena Derbis, and Katarzyna Tutak contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. WIREs RNA published by Wiley Periodicals LLC.

Sobczak) and No. 2021/40/C/NZ3/00323 (to Magdalena Derbis). Funding for proof-reading and open access charge: Initiative of Excellence–Research University No. 05/IDUB/2019/94 at Adam Mickiewicz University, Poznan, Poland.

Edited by: Jeff Wilusz, Editor-in-Chief

Moreover, we discuss the utility of selected RBPs as targets in the development of therapeutic strategies.

This article is categorized under:

RNA Interactions with Proteins and Other Molecules > Protein-RNA Interactions: Functional Implications
RNA in Disease and Development > RNA in Disease

KEYWORDS

liquid–liquid phase separation, RAN translation, repeat expansion, RNA binding proteins, short tandem repeats

1 | INTRODUCTION

Short tandem repeats (STRs), also termed microsatellites, are repeats of 1–8 nucleotide-long sequences and occur very frequently within different parts of genomes. Some STRs are overrepresented in exons, and depending on the sequence, their frequencies in open reading frames and 5' and 3' untranslated regions (UTRs) differ (Kozłowski et al., 2010). Moreover, trinucleotide repeats are more frequent in the coding parts of genes, whereas 5- and 6-nucleotide repeats are more frequent in noncoding parts (Malik, Kelley, et al., 2021). STRs are polymorphic in length and prone to expansion or, to a lesser extent, contraction either in cells of the same individual (somatic instability) or between generations (germline instability; Ashizawa et al., 1992; Paulson, 2018; Trang et al., 2015; Trottier et al., 1994). The origin of STR instability remains incompletely understood, but the possible mechanisms include replication slippage (McMurray, 2010), replication fork stalling (Gadgil et al., 2017), mismatch repair, and bidirectional transcription (Castel et al., 2010).

To date, several possible models of pathogenesis caused by STR expansions have been proposed, and in different repeat expansion disorders (REDs), combinations of these pathomechanisms play crucial roles in disease development (extensively reviewed in Malik, Kelley, et al. (2021)). At the DNA level, STRs in noncoding parts of genes may enhance the cotranscriptional formation of RNA/DNA hybrids called R-loops, which are associated with induction of the DNA damage response and gene silencing due to methylation or blockage of transcription (Reddy et al., 2011). At the RNA level, a dominant gain-of-function mechanism in which RNA bearing expanded repeats (RNA^{exp}) sequesters RNA-binding proteins (RBPs), mostly in cell nuclei, impairing their function, has been postulated in many diseases (i.e., A. Mankodi et al., 2000; Miller et al., 2000; Sellier et al., 2010, 2013). Moreover, if RNA^{exp} is transported to the cytoplasm, its toxicity can be exerted by the translation of the repeats that do not reside in the canonical AUG-initiated open reading frame via a process called repeat-associated non-AUG (RAN) translation (Zu et al., 2011). Finally, a gain-of-function mechanism involving the protein products of genes bearing STRs in their coding sequences is associated with their tendency to form toxic aggregates or activate stress response pathways (Cook et al., 2020; Y. J. Zhang et al., 2018). In all of the mentioned pathomechanisms, processes involving RNA^{exp} molecules can be considered a causative agent from transcription to translation (Box 1).

BOX 1 Gene/RNA/protein gain- and loss-of-function in repeat expansion diseases

Gene loss-of-function: Repeat expansion in DNA might result in partial or complete lack of gene product. Altered gene expression can be caused by transcriptional gene silencing mediated by mutated allele hypermethylation or impaired transcription. For example, in fragile X-syndrome, heterochromatinization of *FMR1* locus silences the expression of fragile X mental retardation protein 1 (FMRP). Alternatively, in Friedreich's ataxia actively transcribed nascent RNA with GAA repeats can form DNA–RNA R-loop structures which initiates epigenetic silencing of *FXN* gene. **RNA gain-of-function:** Expanded repeats within (pre-)mRNAs form stable secondary structures, accumulate in the cell nuclei, attract and trap RNA binding proteins and form ribonucleoprotein complexes called RNA foci. RNA foci are pathological hallmarks of many REDs. For example, in myotonic dystrophy type 1 (DM1), muscleblind-like proteins (MBNLs) are sequestered to CUG repeat foci what leads to the global alternative splicing abnormalities due to MBNLs functional depletion. **Protein**

gain-of-function: The presence of expanded repeats in RNA, if transported to cytoplasm, may result in the production of mutated variant of the protein containing repeated amino acids, like in the case of polyQ expansion in huntingtin. Moreover, peptides derived from RNA^{exp} can aggregate and represent toxic properties. Insoluble protein aggregates are detected in many REDs, for example, polyglycine in FXTAS. On the other hand, in C9-ALS/FTD, intron retention can result in decreased synthesis of C9orf72 protein, which exemplifies protein loss-of-function mechanism due to repeat expansion.

However, it takes two to tango: in the crime of RED pathogenesis, the protein–RNA^{exp} interaction cannot be neglected. RBPs bind specific RNA sequences and/or its specific secondary/tertiary structures through RNA-binding domains (RBDs), form ribonucleoprotein (RNP) complexes, and engage in fundamental cell processes, that is, the regulation of transcription, mRNA splicing, maturation, transport, stability, cellular localization, and translation (reviewed in Jazurek et al., 2016). The most common RBDs contain either an RNA-recognition motif (RRM), a heterogeneous nuclear RNP (hnRNP) K homology domain (KH), zinc fingers (ZNFs), an S1 domain, a double-stranded RBD (dsRBD), or a combination of these domains (reviewed in Lunde et al., 2007). Eukaryotic cells contain hundreds of different RBPs with unique RNA-binding specificity, many of which have been described as involved in the development of REDs.

To date, over 50 human inherited REDs, mainly neurological or neuromuscular diseases, have been described (Figure 1a). Those typically multisystemic diseases can be inherited in an autosomal dominant manner, that is, Huntington's disease (HD; CAG^{exp} in the protein-coding region; MacDonald et al., 1993), myotonic dystrophies (DMs; CTG^{exp} in the 3'UTR; Brook et al., 1992), and C9orf72-linked amyotrophic lateral sclerosis and frontotemporal dementia (C9-ALS/FTD; G4C2^{exp} in the intron; DeJesus-Hernandez et al., 2011); an autosomal recessive manner, that is, Friedreich's ataxia (FRDA; GAA^{exp} in the intron of the frataxin *FXN* gene; Campuzano et al., 1996); or an X-linked dominant manner, that is, fragile X syndrome (FXS) and fragile X-associated tremor/ataxia syndrome (FXTAS; CGG^{exp} in the 5'UTR of the fragile X mental retardation 1 *FMR1* gene (Tassone et al., 2000). In HD, CAG^{exp} in the coding region of the huntingtin gene (*HTT*) results in the production of a mutant protein containing a polyglutamine (polyQ) stretch, which is susceptible to misfolding and aggregation and thus underlies disease pathogenesis (Martí, 2016; Persichetti et al., 1995). In FXS, CGG^{exp} exceeding 200 repeats, named full mutation, causes hypermethylation and silencing of the *FMR1* promoter (Tassone et al., 2000, 2007). The absence of the synaptic functional regulator FMR1 protein (FMRP) is linked to alterations in brain synaptic plasticity, impairing cognitive functions and resulting in intellectual disability. In FXTAS, the mutant RNA containing 55–200 CGG^{exp} can sequester some RBPs (Sellier et al., 2010, 2013) or trigger the production of toxic polyglycine (polyG) protein (Todd et al., 2013). The pathological hallmark of FXTAS is the presence of large, ubiquitin-positive inclusions in the nuclei of neurons and astrocytes (Greco et al., 2002), while the clinical symptoms include intention tremor, cerebellar ataxia, parkinsonism, and brain atrophy (Greco et al., 2006; Hagerman et al., 2001). In C9-ALS/FTD, G4C2^{exp} in the first intron of the *C9orf72* gene leads to insufficiency of the product of the mutated gene, the C9orf72 protein (DeJesus-Hernandez et al., 2011); toxic gain-of-RNA function (Xu et al., 2013); and the production of toxic dipeptide repeat proteins (DPRs; Ash et al., 2013; Gendron et al., 2013). Although DM1 is caused by CTG^{exp} in the 3'UTR of the *DMPK* gene and DM2 by CCTG^{exp} in intron 1 of the nucleic acid-binding protein-coding *CNBP* gene, these two neuromuscular disorders exhibit common pathological symptoms, that is, skeletal muscle weakness, wasting, and cognitive dysfunction (Brook et al., 1992; Liquori et al., 2001). Both CUG^{exp} and CCUG^{exp} are toxic and lead to nuclear sequestration of the Muscleblind-like (MBNL) proteins, impairing their physiological functions (Fardaei et al., 2001, 2002; Mankodi, 2001; Miller et al., 2000).

In this review, we focus on the roles of RNA^{exp} and its protein partners in the pathogenesis of selected REDs, C9-ALS/FTD, FXTAS, HD, and DM, at different stages of the RNA^{exp} life cycle (Table 1). Importantly, the pathomechanisms of these diseases are shared among many other REDs. We discuss not only the proteins that play a role in nuclear processing, that is, the transcription and splicing of RNA^{exp}, and those that are sequestered by RNA^{exp} in various disorders, but also proteins that play a role in RNA^{exp} transport, non-canonical translation, stress response, and phase separation of products of mutant genes. We also discuss these proteins in the context of the development of therapeutic strategies. The described proteins may be considered partners in crime that, together with RNA^{exp} molecules, are responsible for the development and progression of REDs.

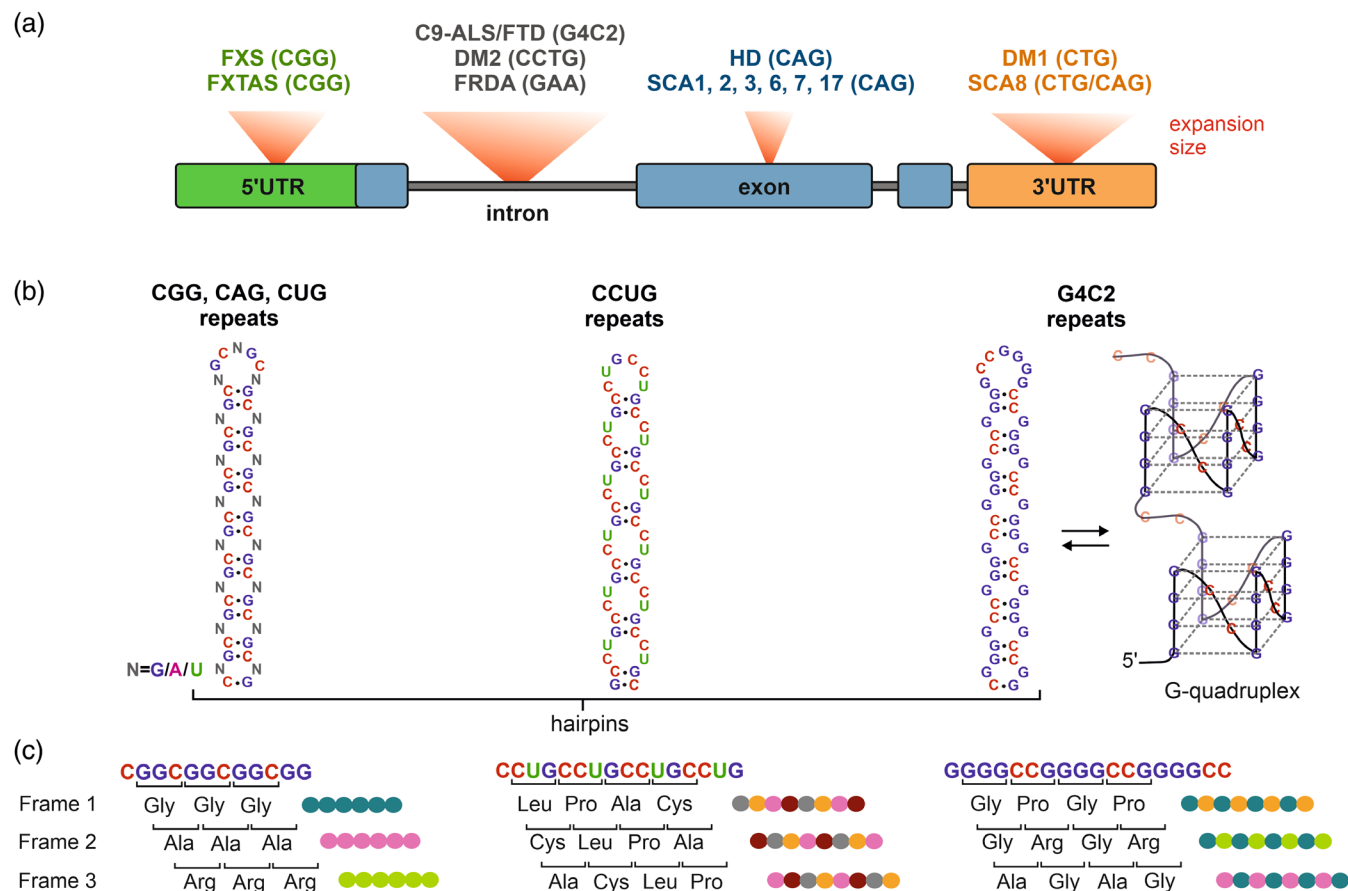


FIGURE 1 Characteristics of expanded STRs specific for different diseases. (a) *Localization and size of STRs within specific gene regions.* Expanded STRs, depending on the sequence, are located in different parts of the gene. Size of expansion of STRs necessary for the development of individual REDs differ between diseases, however, it may be roughly specified that the longest expansions are located within introns and further within 3'UTRs, middle-size expansions within 5'UTRs and the shortest within exons, and thus with protein-coding sequences. Here we show representative REDs from the larger group of diseases. (b) *Structures formed by RNA^{exp}.* Trinucleotide CNG repeats form RNA hairpin structures, all characterized by high thermodynamic stability, the highest for CGG, next CAG, and the lowest for CUG repeats. Hairpin structures are also formed by CCUG and G4C2 repeats. Moreover, G4C2 and CGG repeats are able to form G-quadruplex structure. (c) *Protein products derived from repeat-associated non-AUG (RAN) translation.* RAN translation may potentially start and produce proteins in all three reading frames. In the process of RAN translation of trinucleotide repeats, the homopolymeric proteins with tracts of single amino acids are biosynthesized. CCUG tetranucleotide repeats are RAN translated to proteins containing tracts of four amino acids, the same in all reading frames. RAN translation of G4C2 hexanucleotide repeats is the source of DPRs composed of tracts of two amino acids repeats, one of which is glycine in all reading frames

2 | TOXIC RNA^{exp} MOLECULES

In various REDs, distinct toxic RNA^{exp} molecules are generated. Interestingly, for the development of each RED, a specific “repeat load” is needed. This “repeat load” not only consists of the length of STRs within RNA^{exp} and cellular RNA^{exp} levels but also the specific location, for example, certain cellular compartments and specific backgrounds within cell types. Thus, some RNA^{exp} molecules are pathogenic when the length of inherited expanded repeats exceeds a few dozen copies, while for others, when the number of repeats exceeds hundreds or even thousands of copies (Paulson, 2018). STR length is also related to pathomechanisms in which RNA^{exp} molecules are involved; for example, RBP sequestration plays a larger role in REDs caused by very long STRs, such as DMs (up to 10,000 CTG or CCTG repeats; Malik, Kelley, et al., 2021; Rohilla & Gagnon, 2017). The toxicity of RNA^{exp} may further increase with the elongation of STRs during lifespan of patients, manifesting as increased disease severity or a younger age of onset (Paulson, 2018; Swami et al., 2009).

RNA^{exp} molecules were shown to form different types of structures with different thermodynamic stabilities, depending on the repeated sequence motif: unstructured single-stranded RNAs (e.g., AAG repeats), semistable hairpins,

TABLE 1 Proteins implicated in pathomechanisms of repeat expansion diseases (REDs)

Process	RNA ^{exp}	Protein	Protein name	UniProt ID	Effect/method	
Transcription	CAG	SUPT4H1	Transcription elongation factor SPT4	P63272 (human)	KO in mouse reduced production of mutated huntingtin	(C. R. Liu et al., 2012)
	CAG	SPT4-A	Transcription elongation factor SPT4-A	P63271 (mouse)	KD (ASO) in mouse selectively reduced mutated mRNA and protein	(H. M. Cheng et al., 2015)
	G4C2; C4G2	SUPT4H1	Transcription elongation factor SPT4	P63272 (human)	KO in yeast and in fly reduced RNA foci and RAN translation	(Kramer et al., 2016)
	G4C2	PAF1, LEO1	RNA polymerase II-associated factor 1 homolog; RNA polymerase-associated protein LEO1	Q8N7H5 (human), Q8WVC0 (human)	KD in fly suppressed toxicity, upregulated in patients	(Goodman, Prudencio, Kramer, et al., 2019)
	G4C2	AFF2/FMR2	AF4/FMR2 family member 2	P51816 (human)	KD in fly reduced RNA ^{exp} and RAN protein level, KO in iPSNs reduced level of <i>C9orf72</i> RNA, RNA foci, and RAN translation products	(Yuva-Aydemir et al., 2019)
	GGCCUG	SPT4-A, SPT5	Transcription elongation factor SPT4-A; Transcription elongation factor SPT5	P63271 (mouse), O55201 (mouse)	KD in Neuro2A cells reduced RNA foci and RAN translation	(Furuta et al., 2019)
Sequestration, co-localization	CAG	MBNL1	Muscleblind-like protein 1	Q9NR56 (human)	OE rescued splicing abnormalities in HeLa cells expressing CAG ^{exp}	(Mykowska et al., 2011)
	G4C2	Nucleolin	Nucleolin	P19338 (human)	Nucleolar stress in C9-ALS/FTD patients	(Haessler et al., 2014)
	G4C2	Pur alpha	Transcriptional activator protein Pur-alpha	Q00577 (human)	OE reduced neurodegeneration in fly, cell, and fish models.	(Swinnen et al., 2018; Xu et al., 2013)
	G4C2	hnRNP H	Heterogeneous nuclear ribonucleoprotein H	P31943 (human)	Colocalized with RNA foci in C9-ALS/FTD patients, missplicing of hnRNP H-dependent transcript	(Y. B. Lee et al., 2013)
	G4C2	hnRNP H1/F, ALYREF, and SRSF2	Heterogeneous nuclear ribonucleoprotein F, THO complex subunit 4, SRSF protein kinase 2	P52597 (human), Q86V81 (human), P78362 (human)	Colocalized with RNA foci in C9-ALS/FTD patients	(Cooper-Knock et al., 2014)

(Continues)

TABLE 1 (Continued)

Process	RNA ^{exp}	Protein	Protein name	UniProt ID	Effect/method	
	C4G2	SRSF1, hnRNP A1, hnRNP H/F, ALYREF	SRSF protein kinase 1, Heterogeneous nuclear ribonucleoprotein A1, Heterogeneous nuclear ribonucleoprotein F, THO complex subunit 4	O70551 (human), P09651 (human), P52597 (human), Q86V81 (human)	Colocalized with RNA foci in C9-ALS/FTD patients	(Cooper-Knock et al., 2015)
	G4C2	ADARB2	Double-stranded RNA- specific editase B2	Q9NS39 (human)	KD decreased RNA foci in iPSNs, colocalized with RNA foci in C9-ALS/FTD patients	(Donnelly et al., 2013)
	G4C2	ZPF106	Zinc finger protein 106	Q9H2Y7 (human)	OE suppressed neurotoxicity in C9-ALS/FTD fly model	(Celona et al., 2017)
	G4C2	Matrin-3	Matrin-3	P43243 (human)	OE suppressed neurotoxicity in C9-ALS/FTD fly model	(Ramesh et al., 2020)
	CUG	MBNL1	Muscleblind-like protein 1	Q9NR56 (human)	OE/ASO blockers complementary to RNA ^{exp} reversed DM1 phenotype in mouse model	(Kanadia et al., 2006; Wheeler et al., 2009)
	CUG	hnRNP H	Heterogeneous nuclear ribonucleoprotein H	P31943 (human)	Colocalized with RNA foci in DM1 patients	(D. H. Kim et al., 2005)
	CCUG	MBNL1	Muscleblind-like protein 1	Q9NR56 (human)	Colocalized with RNA foci in DM2 patients	(Mankodi, 2001)
	CCUG	hnRNP core protein and snRNP Sm antigen	Small nuclear ribonucleoprotein E	P62304 (human)	Colocalized with RNA foci in DM2 patients	(Perdoni et al., 2009)
	CCUG	RBFOX1	RNA binding protein fox-1 homolog	Q9NWB1 (human)	Colocalized with RNA foci in DM2 patients	(Sellier et al., 2018)
	CGG	FMRpolyG/ PolyG	FMRpolyG	–	Colocalized with RNA ^{exp} , cause neurotoxicity in primary neurons derived from FXTAS mouse	(Asamitsu et al., 2021)
	CGG	hnRNP A2/B1	Heterogeneous nuclear ribonucleoproteins A2/B1	P22626 (human)	OE suppressed neurotoxicity in FXTAS fly model	(Sofola et al., 2007)
	CGG	CELF1	CUGBP Elav-like family member 1	Q92879 (human)	OE suppressed neurotoxicity in FXTAS fly model	(Sofola et al., 2007)

TABLE 1 (Continued)

Process	RNA ^{exp}	Protein	Protein name	UniProt ID	Effect/method	
	CGG	Pur alpha	Transcriptional activator protein Pur-alpha	Q00577 (human)	OE suppressed neurotoxicity in FXTAS fly model	(Jin et al., 2007)
	CGG	TRA2A	Transformer-2 protein homolog alpha	Q13595 (human)	Colocalized with RNA ^{exp} foci in FXTAS cellular model and with inclusions in FXTAS mouse model and patients	(Cid-Samper et al., 2018)
	CGG	DROSHA-DGCR8	Ribonuclease 3 and Microprocessor complex subunit DGCR8	Q9NRR4 (human), Q8WYQ5 (human)	OE suppressed neurotoxicity in FXTAS cellular model, miRNA processing is impaired in FXTAS patients	(Sellier et al., 2013)
	CGG	SAM68	KH domain-containing, RNA-binding, signal transduction-associated protein 1	Q07666 (human)	Colocalized with RNA ^{exp} foci in FXTAS cellular model, SAM68-dependent alternative splicing is impaired in FXTAS patients	(Sellier et al., 2010)
Nucleocytoplasmic transport	CAG	U2AF65	Splicing factor U2AF 65 kDa subunit	P26368 (human)	KD in fly enhanced RNA ^{exp} toxicity and accumulation, reduction in HD mouse caused RNA ^{exp} nuclear accumulation	(Tsoi et al., 2011)
	G4C2	RanGAP1	Ran GTPase-activating protein 1	P46060 (human)	OE suppressed RNA ^{exp} toxicity in C9-ALS/FTD fly, OE rescued impairment of NCT in C9-ALS iPSNs	(K. Zhang et al., 2015)
	G4C2; CGG	SRSF1	Serine/arginine-rich splicing factor 1	O70551 (human)	KD reduced nuclear export of mutated G4C2 ^{exp} /CGG ^{exp} transcripts in cells and neurotoxicity in C9-ALS/FTD iPSNs and in FXTAS & C9-ALS/FTD fly models	(Hautbergue et al., 2017; Malik, Tseng, et al., 2021)
	G4C2	POM 121	Nuclear envelope pore membrane protein POM121	Q8TEM1 (human)	OE rescued expression of nuclear pore components and NCT impairment in C9-ALS/FTD iPSNs	(Coyne et al., 2020)

(Continues)

TABLE 1 (Continued)

Process	RNA ^{exp}	Protein	Protein name	UniProt ID	Effect/method	
	CUG	hnRNP H	Heterogeneous nuclear ribonucleoprotein H	P31943 (human)	KD reduced nuclear retention of RNA ^{exp} in cells	(D. H. Kim et al., 2005)
	CUG	Staufen1	Double-stranded RNA-binding protein Staufen homolog 1	O95793 (human)	OE reduced nuclear retention of RNA ^{exp} in cells	(Ravel-Chapuis et al., 2012)
RAN translation	G4C2; CGG	eIF4A	Eukaryotic initiation factor 4A-I	P60842 (human)	Inhibition reduced RAN translation in vitro	(Green et al., 2017; Kearse et al., 2016)
	CGG; G4C2	eIF4B and eIF4H	Eukaryotic translation initiation factor 4B and Eukaryotic translation initiation factor 4H	P23588 and Q15056 (human)	KD reduced RAN translation in C9-ALS/FTD fly model	(Goodman, Prudencio, Srinivasan, et al., 2019; Linsalata et al., 2019)
	CGG	DDX3X	ATP-dependent RNA helicase DDX3X	O00571 (human)	KD reduced RAN translation in FXTAS fly model	(Linsalata et al., 2019)
	G4C2	DDX3X	ATP-dependent RNA helicase DDX3X	O00571 (human)	KD enhanced RAN translation in C9-ALS/FTD fly model and in C9-ALS/FTD iPSNs	(W. Cheng et al., 2019)
	G4C2; CGG	DHX36	ATP-dependent DNA/RNA helicase DHX36	Q9H2U1 (human)	KD reduced RAN translation in C9-ALS/FTD iPSNs	(H. Liu, Lu, Paul, Periz, Banco, Ferré-D'Amaré, et al., 2021; Tseng et al., 2021)
	CGG	eIF1 and eIF5	Eukaryotic translation initiation factor 1 and Eukaryotic translation initiation factor 5	P41567 and P55010 (human)	Modulated RAN translation in vitro	(Linsalata et al., 2019)
	CAG; G4C2	eIF3F	Eukaryotic translation initiation factor 3 subunit F	O00303 (human)	KD reduced RAN translation in vitro	(Ayhan et al., 2018)
	CGG	5MP	eIF5-mimic protein (also known as BZW2)	E9PFD4 (human)	OE reduced RAN translation in FXTAS fly model	(Singh et al., 2021)
	G4C2	eIF2D	Eukaryotic translation initiation factor 2D	P41214 (human)	KD reduced RAN translation in <i>C. elegans</i> C9-ALS/FTD model	(Sonobe et al., 2021)
	G4C2; CAG	RPS25	40S ribosomal protein S25	P62851 (human)	KD reduced RAN translation in C9-ALS/FTD fly model and in C9-ALS/FTD iPSNs	(Yamada et al., 2019)

TABLE 1 (Continued)

Process	RNA ^{exp}	Protein	Protein name	UniProt ID	Effect/method	
Stress response	G4C2; CAG; CCTG; CAGG	PKR	Interferon-induced, double-stranded RNA-activated protein kinase	P19525 (human)	Inhibition of PKR reduced RAN translation in C9-ALS/FTD mouse model	(Zu et al., 2020)
	G4C2; CGG	eIF2 α	Eukaryotic translation initiation factor 2 subunit 1	P05198 (human)	Enhanced RAN translation when phosphorylated in stress conditions in vitro	(W. Cheng et al., 2018; Green et al., 2017)
	G4C2; CAGG; CCTG	eIF2A	Eukaryotic translation initiation factor 2A	Q9BY44 (human)	KO reduced RAN translation in vitro	(Sonobe et al., 2018; Tusi et al., 2021)
	G4C2	PERK	Eukaryotic translation initiation factor 2-alpha kinase 3	Q9NZJ5 (human)	Elevated activity in response to accumulation of RAN peptides increases RAN translation	(Zu et al., 2020)
	CUG/CAG	Dicer	Endoribonuclease Dicer	Q9UPY3 (human)	Cleaves RNA repeat regions into ~21 nt fragments	(Krol et al., 2007; Lawlor et al., 2011; Yu et al., 2011)
	CAG/CUG	ADAR1	Double-stranded RNA-specific adenosine deaminase	P55265 (human)	Co-expression of ADAR1 with (CAG/CUG)100 dsRNA rescued repeat-related pathology in <i>Drosophila</i>	(van Eyk et al., 2019)
	CAG/CUG	TLR	Toll-like receptors		KD of TLRs in <i>Drosophila</i> decreased the toxicity of CAG/CUG ₋₁₀₀ repeats	(Samaraweera et al., 2013)
MLO formation (Phase separation)	CAG	SRSF2	Serine/arginine-rich splicing factor 2	Q01130 (human)	Marker of nuclear speckles, colocalized with RNA ^{exp} foci	(Jain & Vale, 2017)
	G4C2	SRSF2	Serine/arginine-rich splicing factor 2	Q01130 (human)	Marker of nuclear speckles, colocalized with RNA ^{exp} foci	(Jain & Vale, 2017)
	G4C2	G3BP1, Caprin1, USP10, eIF3b, ELAVL1, TIAR	Ras GTPase-activating protein-binding protein 1, Caprin-1, Ubiquitin carboxyl-terminal hydrolase 10, Eukaryotic translation initiation factor 3 subunit B,	Q13283 (human), Q14444 (human), Q14694 (human), P55884 (human), Q15717 (human)	Markers of stress granules, condensed in vitro with RNA ^{exp} and lysates from cell lines and mouse brain	(Fay et al., 2017)

(Continues)

TABLE 1 (Continued)

Process	RNA ^{exp}	Protein	Protein name	UniProt ID	Effect/method
			ELAV-like protein 1, Nucleolysin TIAR	Q01085 (human)	
	G4C2	FMRP	Synaptic functional regulator FMR1	Q06787 (human)	Marker of transport granules, colocalized with RNA ^{exp} , FMRP- dependent translation regulation was impaired in C9-ALS/FTD iPSNs

Abbreviations: ASO, antisense oligonucleotides; iPSNs, induced pluripotent stem cells-derived neurons; KD, knockdown; KO, knockout; MLO, membraneless organelles; NCT, nucleocytoplasmic transport; OE, overexpression.

fairly stable hairpins, or very stable G-quadruplexes (Figure 1b). The most common RNA^{exp} molecules in REDs, CNG repeats, follow the thermodynamic stability order CGG > CAG > CUG > CCG (Sobczak et al., 2010). At physiological KCl concentration, pH, and temperature, G4C2^{exp} and CGG^{exp} form a stable parallel G-quadruplex (Asamitsu et al., 2021; Haeusler et al., 2014). These higher-order structures show extremely high thermodynamic stability and are directly linked to the pathogenic potential of RNA^{exp}, as they are recognized by different RBPs. Importantly, the structure of RNA^{exp} may be influenced by the presence of nucleotide interruptions in the sequences of repeats. In general, these protective elements against STR instability at the genomic level reduce RNA^{exp} toxicity. Interruptions including CGG, CTC, GGC, or CAG triplets in CTG^{exp} occur in ~3–5% of DM1 patients, reduce somatic instability, and may result in atypical or milder symptoms, later ages of onset, and progression (Braidia et al., 2010; Cumming et al., 2018; Wenninger et al., 2021). Similarly, AGG interruptions within CGG^{exp} in *FMR1*, increase genetic stability of alleles with 45–90 repeats (Nolin et al., 2015; Villate et al., 2020; Yrigollen et al., 2014). The CAA interruptions in CAG^{exp} in *SCA2* are linked with delayed disease onset. On the other hand, numerous CCG/CGG interruptions in CTG/CAG^{exp} of *ATXN8* related to spinocerebellar ataxia type 8 (*SCA8*), increase thermodynamic stability of RNA hairpin, change amino acid composition of RAN proteins enhancing their toxicity, and thus are associated with increased disease penetrance (Perez et al., 2021).

Despite differences concerning their sequence, the length of STRs, structure, and so forth, various RNA^{exp} molecules are engaged in similar pathogenic mechanisms, such as RBP sequestration, impairment of nucleocytoplasmic transport (NCT), or RAN translation. This is related to the natural multivalency of RNA^{exp} molecules, which makes them susceptible to interaction with each other and with multiple RBPs. It further leads to the formation of RNA foci, structures composed of RNA^{exp} and proteins, mainly in the nucleus, that varies in size, frequency, and composition but are common between different REDs (reviewed in N. Zhang & Ashizawa, 2017; Box 2).

BOX 2 Repeat load

Specific repeat load, necessary for development of individual RED, consists of a few components. First, the length of expanded repeats, which may further increase between generations or in different cells within the same organism, due to germline and more robust somatic instability. Second, the biogenesis and cellular stability of expanded repeat-bearing transcripts. Third, localization of RNA^{exp} in a defined cell type and cellular compartment where it interacts with specific partners. Finally, the sequence itself, which, if enriched in nucleotide interruptions in the repeat tract, changes the potential of RNA^{exp} molecule for relevant interactions with RBPs or toxicity of RAN protein products related to their amino acid composition.

3 | TRANSCRIPTION AND SPLICING OF RNA^{exp}

Transcription of genes containing expanded repeats plays an important role in the pathogenesis of some REDs, as RNA^{exp} can exert toxicity by sequestration of many RBPs as a function of RNA^{exp} load. The different noncanonical secondary structures formed by STR expansions need to be resolved during transcription, impeding transcriptional efficiency (Goodman & Bonini, 2020).

Slowed or stalled transcription of STRs can lead to the formation of R-loops in the gene locus. These nucleic acid structures are RNA/DNA hybrids formed during transcription when nascent RNA hybridizes to the DNA template strand behind elongating RNA polymerase II (Pol II; Thomas et al., 1976). In FXTAS, the GC-rich 5'UTR of *FMR1* is susceptible to cotranscriptional R-loop formation upon reannealing of the nascent *FMR1* pre-mRNA to the complementary DNA strand because a G-rich RNA:C-rich DNA heteroduplex is thermodynamically more stable than the corresponding DNA:DNA duplex (Loomis et al., 2014; Reddy et al., 2011). R-loops over expanded repeats may form a structural block, directly interfering with Pol II transcription elongation, and influencing transcription efficiency (Belotserkovskii et al., 2017; Crossley et al., 2019; Groh et al., 2014). Excessive R-loop formation in an individual locus can result in double-stranded DNA brakes and activation of the DNA damage response, triggering a series of signaling events that may be pathogenic (Diab et al., 2018; Loomis et al., 2014). In line with this mechanism, the phosphorylated histone variant γ H2AX, a DNA damage-response molecule, was detected in FXTAS patients (Hoem et al., 2011; Iwahashi et al., 2006). Moreover, methylation caused by the formation of atypical DNA/DNA or DNA/RNA structures in the expanded repeats at a given locus was shown to significantly modulate the expression of mutant genes. For example, in FRDA, R-loops can trigger heterochromatinization, which results in the *FXN* gene silencing (Li et al., 2016).

Expanded CTG (in DM1), G4C2 (in C9-ALS/FTD), and CGG (in FXS) repeats were shown to dysregulate the transcription process via inhibition or impairment of Pol II initiation or elongation (Brouwer et al., 2013; Colak et al., 2014; Haeusler et al., 2014). However, at least two protein complexes, the DSIF complex, and PAF1 complex (PAF1C), were shown to promote Pol II transcription at repeat expansion sites (Figure 2a).

The DSIF complex, composed of two highly conserved proteins, SUPT4H1 and SPT5H (or the yeast orthologs Spt4 and Spt5), is a transcription elongation factor that regulates Pol II processivity by reducing the efficiency of its dissociation from template DNA (Wada et al., 1998). During transcription, Spt5 interacts with the Pol II coiled-coil domain and encircles the RNA/DNA hybrid, which may prevent the dissociation of Pol II from the template (Klein et al., 2011; Martinez-Rucobo et al., 2011). Spt4 interacts indirectly with Pol II via Spt5, and the DSIF complex interacts with the DNA template outside of the transcription bubble (Klein et al., 2011; Martinez-Rucobo et al., 2011). Spt4 possesses a ZNF domain that probably plays a role in modulating interactions with DNA and stabilizes RNA polymerase/template complexes, preventing Pol II from pausing (Crickard et al., 2016; Wenzel et al., 2008).

SUPT4H1 was shown to be required for the transcription of expanded CAG (C. R. Liu et al., 2012) and G4C2 (Kramer et al., 2016) repeats in HD and C9-ALS/FTD, respectively. Deletion of SPT4-A in mouse striatal neurons expressing *Htt* containing long (>100) CAG repeats (a mouse model of HD) resulted in reduced synthesis of mutant huntingtin containing long polyQ, thus diminishing its aggregation and toxicity (C. R. Liu et al., 2012). In vivo studies showed that downregulation of SUPT4H1 by delivering antisense oligonucleotides (ASOs) that activate RNaseH-mediated target RNA degradation into the brains of HD model mice reduced the levels of mRNA and protein from the mutant but not the normal *Htt* allele (H. M. Cheng et al., 2015). Deletion of Spt4 in C9-ALS/FTD yeast, *C. elegans*, and *Drosophila* models reduced the expression of G4C2^{exp} and C4G2^{exp} transcripts, blocked accumulation of these transcripts into RNA foci, and decreased the levels of RAN-translated polyglycine-proline (polyGP; one of three dipeptide repeats (DPRs) produced from G4C2^{exp}; Figure 1c). These findings were also confirmed in fibroblasts derived from C9-ALS patients (Kramer et al., 2016).

The PAF1C is composed of the highly conserved PAF1, LEO1, CDC73, CTR9, and RTF1 proteins and plays a role in the initiation, promoter-proximal pausing, elongation, and RNA processing/termination stages of transcription (Goodman & Bonini, 2020). Recently, an RNA interference (RNAi)-based screen in a *Drosophila* model of C9-ALS/FTD revealed that the *Drosophila* orthologs of PAF1 and LEO1 (dPaf1 and dLeo1, respectively) are selectively involved in transcription of the G4C2^{exp}-containing allele, while downregulation of other PAF1C components affected transcription of both long and short G4C2 tracts. The downregulation of dPaf1 and dLeo1 suppressed the toxicity of repeat expansion in fly tissues (Goodman, Prudencio, Kramer, et al., 2019). Notably, the RNA levels of hPAF1 and hLEO1 were upregulated in post-mortem cortical tissue from C9-ALS patients, supporting the link between the PAF1C and G4C2^{exp} in C9-ALS/FTD (Goodman, Prudencio, Kramer, et al., 2019).

Recently, the AFF2/FMR2 protein, a component of superelongation complex (SEC)-like 2, was found to selectively regulate transcription of the *C9orf72* allele containing long G4C2 tracts in C9-ALS induced pluripotent stem cell (iPSC)-derived neurons (iPSNs; Yuva-Aydemir et al., 2019). Together, the emerging roles of SUPT4H1 (H. M. Cheng et al., 2015; Furuta et al., 2019; Kramer et al., 2016; C. R. Liu et al., 2012), the PAF1C (Goodman, Prudencio, Kramer, et al., 2019), and AFF2/FMR2 (Yuva-Aydemir et al., 2019) in the transcription of RNA with expanded STRs suggest that the factors implicated in transcriptional elongation are potential therapeutic targets for REDs.

In eukaryotes, splicing plays a major role in cotranscriptional gene expression, and almost 95% of protein-encoding genes undergo alternative splicing (AS). Previously, it was unclear how expanded repeats present in introns exert toxicity at the RNA and protein levels. While the majority of pre-mRNA molecules with expanded repeats undergo proper splicing and maturation, GC-rich expansion leads to intron retention (IR) (Sznajder et al., 2018; Wang et al., 2021) (Figure 2b). This interesting AS event was observed in some of the genes with repeat expansion, and the resulting RNA molecules contained an unprocessed sequence, which failed to be excised from pre-mRNA and was protected from the 5'- and 3'-ends. IR may be a consequence of spliceosome stalling or abnormal association of RBPs with *cis*-regulatory sequences, which, in the case of expanded repeats, can confer structural arrangements (Taylor & Sobczak, 2020). Although IR can be considered a physiological mechanism (Wong et al., 2013), it may also play a relevant role in REDs with intronic GC-rich sequences. For instance, *C9orf72* mRNA with an enlarged 5'UTR retaining the G4C2^{exp} intron can accumulate in the nucleus and was detected in the frontal cortex in heterozygous expansion carriers (Mori, Weng, Arzberger, May, Rentzsch, Van Broeckhoven, et al., 2013; Sznajder et al., 2018). Similarly, CCUG^{exp} associated with DM2 was shown to induce the retention of host very long intron 1 and elevated levels of mutant *CNBP* mRNA. Retention of introns containing CCUG^{exp} has been detected in many DM2 tissues, including skeletal muscle and the frontal cortex of the brain, and lymphoblastoid cell lines (Sznajder et al., 2018). It was also shown that generally, GC-rich sequences in DNA, due to their secondary structures, can slow the RNA Pol II elongation rate and cause RNA Pol II pausing over-retained introns (Velooso et al., 2014). Nevertheless, the trans-acting regulators of IR related to microsatellite repeat disorders remain to be elucidated and require further exploration.

In summary, the transcription and splicing of RNA^{exp} depend on the STR sequence and location of the repeats within the gene. Noncanonical secondary structures, including G-quadruplexes and R-loops formed by repeat expansions, impact the transcription rate and/or IR, probably by altering accessibility to trans-acting factors (Table 1), which ultimately has a significant effect on RNA^{exp} load and toxicity.

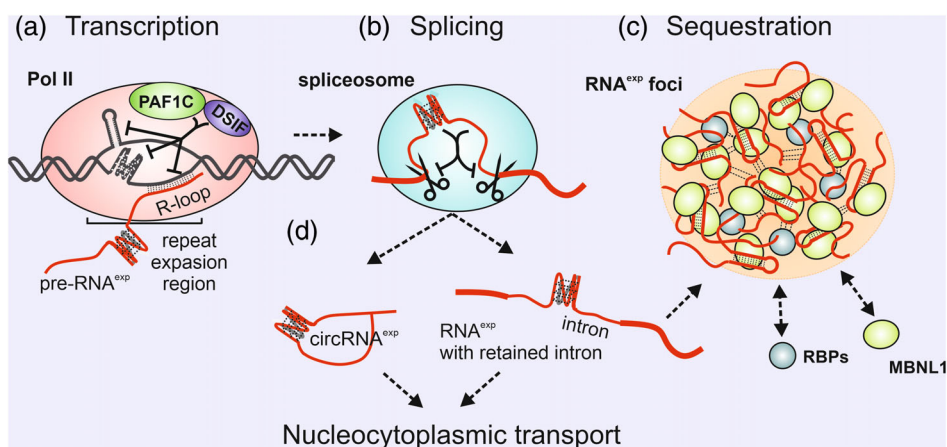


FIGURE 2 Nuclear processing and accumulation of RNA^{exp} molecules. (a) *Transcription*. DSIF and PAF1C complexes promote transcription of repeat expansion regions through inhibition of the formation of DNA secondary structures and R-loops (described for CAG^{exp} and G4C2^{exp}). (b) *Splicing*. The majority of pre-mRNAs with expanded repeats undergo correct splicing, however, in some parts of mature transcripts intron retention takes place (described for CCUG^{exp} and G4C2^{exp}). Moreover, G4C2^{exp}-containing spliced intron is stabilized in a circular form. Bold line, exon; thin line, intron. (c) *Sequestration*. RNA^{exp} molecules accumulate in nuclei where they bind multiple RBPs and sequester some of them and form RNA^{exp} foci (described for CAG^{exp}, CUG^{exp}, CCUG^{exp}, CGG^{exp}, G4C2^{exp}). (a–c) Arrow with a dotted line, change in place and/or in time; solid lines show induction or inhibition of certain processes

4 | SEQUESTRATION OF PROTEINS ON RNA^{exp} MOLECULES

As mentioned above, many REDs have been linked to pathogenic RNA gain-of-function models, also termed RNA toxicity, in which mutant RNA^{exp} molecules accumulate in the nucleus, forms aggregates, and sequesters specific RBPs within nuclear foci (Figure 2c, Table 1). This in turn leads to functional depletion of these proteins and impairment of their physiological functions (N. Zhang & Ashizawa, 2017). Depending on not only the type and load of expanded repeats but also on the tissue type and availability of RBPs, RNA^{exp} foci can present distinct morphologies and abundance (reviewed in Wojciechowska & Krzyzosiak, 2011). It should be noted that some proteins are indirectly sequestered in a process mediated by the interaction of RNA^{exp} molecules with other protein partners (Yang & Hu, 2016).

4.1 | Myotonic dystrophies

CUG^{exp} foci were reported for the first time in muscle and fibroblast biopsies from DM1 patients (Taneja et al., 1995), and they have since been observed in smooth muscle (Cardani et al., 2008), the heart (Mankodi et al., 2005), and the brain (Jiang et al., 2004) as well. Foci formation depends on the length of the CUG^{exp} repeat; in biopsies of the *vastus lateralis* muscle, sporadic inclusions were observed within the 70–100 CUG^{exp} repeat range (Mankodi, 2001). CUG^{exp} repeats fold into very stable RNA hairpin structures, attract MBNL1 and sequester it away from its normal RNA-binding sites, leading to its functional insufficiency in cells (Miller et al., 2000). MBNL1 is an RBP that recognizes multiple GCs flanked by pyrimidines, and under physiological conditions, the major role of MBNL is the regulation of tissue-specific AS and polyadenylation (Konieczny et al., 2014). Increased activity of MBNL1 and 2 during tissue development shifts the pattern of target mRNA AS from fetal- to adult-specific, while its downregulation in DM1 has the opposite effect (Konieczny et al., 2014). Missplicing of multiple mRNAs is exacerbated in the development of this progressive disease (a consequence of somatic expansion). Abnormalities in the AS of muscle-specific chloride channel (*CLCN1*; Kino et al., 2009), insulin receptor (*INSR*; Ho et al., 2004), bridging integrator 1 (*BINI*; Fugier et al., 2011), and calcium channel voltage-dependent L type alpha 1 s subunit (*CACNA1S*) cause-specific DM symptoms, including myotonia, insulin resistance, and muscle weakness, respectively. MBNL1 was also shown to interact with other RBPs (e.g., hnRNP H, H2, H3, F, A2/B1, K, and L; the probable ATP-dependent RNA helicases DDX5 and DDX17; and ATP-dependent RNA helicase A [DHX9]; Paul et al., 2011). Although only a fraction of these interactors colocalizes with CUG^{exp} foci, long CUG repeats alter the stoichiometry of MBNL1-protein complexes. Increased levels of hnRNP H, H2, H3, and F and DDX5 dysregulate the AS of many DM1-specific RNA targets in a manner similar to MBNL1 depletion, showing that the interruption of functional interactions between MBNL1 and other RBPs may contribute to DM1 pathogenesis (Paul et al., 2011).

Apart from MBNL1 sequestration, other factors also play crucial roles in DM1 pathogenesis. In DM1 myoblasts, skeletal muscle, and heart tissues, the steady-state level of CUG-binding protein (CELF1 aka CUGBP1) is augmented, leading to translation defects and misregulation of the AS of some CELF1 mRNA targets. Although the possible sequestration of CELF1 on mutant RNA containing CUG repeats was suggested two decades ago (Timchenko et al., 2001), other studies argued that CELF1 does not colocalize with CUG^{exp} foci (Fardaei et al., 2001; Jiang et al., 2004; Mankodi et al., 2005) and that its increased level is a result of hyperphosphorylation by protein kinase C, which leads to an increase in protein half-life and activity (Kuyumcu-Martinez et al., 2007).

Additionally, hnRNP H was shown to bind CUG^{exp} in vivo and colocalize to a limited extent with RNA foci in DM1 patient-derived myoblasts (D. H. Kim et al., 2005). The level of this protein increases ~2-fold, probably due to the signaling events downstream of CUG^{exp} (Paul et al., 2006). Interestingly, hnRNP H was also shown to interact with MBNL1 in normal myoblasts in an RNA-independent manner, and in DM1 myoblasts, elevated levels of MBNL1 increased the colocalization of hnRNP H with RNA foci, suggesting that hnRNP H is recruited to intranuclear DM1 foci by MBNL1 (Paul et al., 2006).

DM2 presents a clinical phenotype similar to that of DM1, even though its symptoms are generally milder, and its progression is less severe (Meola & Cardani, 2017). Ribonuclear foci containing CCUG^{exp} were shown to efficiently sequester MBNL proteins (Mankodi, 2001). Moreover, the hnRNP core protein and snRNP Sm antigen were found to colocalize with some MBNL1-CCUG^{exp} foci (Perdoni et al., 2009). Recently, in vitro studies showed that RNA-binding protein fox-1 homolog 1 (RBFOX1) directly binds the CCUG^{exp}. Although this regulator of mRNA metabolism colocalized with CCUG^{exp} foci in muscle biopsies derived from DM2 patients, its splicing regulatory functions were not impaired (Sellier et al., 2018). Interestingly, RBFOX1 seems to compete with MBNL1 for CCUG repeat-binding sites,

which results in reduced sequestration of MBNL proteins by CUG^{exp}. This phenomenon, together with the intronic localization of expanded repeats (described above), may explain the fewer splicing alterations in DM2 than in DM1 (Sellier et al., 2018). Additionally, the misregulated AS patterns in DM1 and DM2 are similar, but the extent of these changes is tissue-specific (Nakamori et al., 2013). Indeed, the muscle weakness associated with DM1 affects distal muscles, while proximal muscles are affected in DM2.

Taken together, the imbalance of multiple interacting proteins may cause aberrant splicing patterns and contribute to DM pathogenesis. The sequestration rate depends on a load of CUG^{exp} or CUG^{exp} (repeat length and expression level of RNA with expanded repeats). Due to unequal somatic instability, the efficiency of protein sequestration in distinct tissues or even myofibers differs. Moreover, MBNL proteins are mobile in RNA foci, and when saturated, they can diffuse freely between RNA foci and the nucleoplasm, which may underlie MBNL sequestration (Sznajder et al., 2016). Multiple therapeutic strategies involving limiting the sequestration or degradation of CUG^{exp} were previously developed (reviewed in Konieczny et al., 2014).

4.2 | Fragile X-associated tremor/ataxia syndrome

The pathological hallmark of FXTAS is the presence of solitary, large (2–5 μm), ubiquitin-positive inclusions in the nuclei of neurons, astrocytes (Greco et al., 2002), and Purkinje cells (Ariza et al., 2016). The number of inclusions is correlated with CGG repeat length, suggesting that toxic RNA gain-of-function plays an important role in FXTAS pathogenesis. In FXTAS, the repeat length is limited to approximately 200 repeats; longer expansions cause the development of a completely different neurodevelopmental disease, a FXS fragile X syndrome, in which complete silencing of *FMR1* occurs (Heitz et al., 1992). *FMR1* mRNA containing CGG^{exp} can sequester one or more RBPs, thus impairing their physiological function. Another possible mechanism in FXTAS is that repeat-associated non-AUG (RAN) translation leads to the production of toxic proteins containing monoamino acid tracts, mainly polyG (Figure 1c, described in the chapter on RAN translation), which colocalize with ubiquitin in FXTAS in different regions of brain sections (Todd et al., 2013) and, to a lesser extent, in other tissues (Buijsen et al., 2014). Recently, the direct interaction of CGG^{exp} and polyG was shown to promote polyG aggregation and lead to neuronal dysfunction (Asamitsu et al., 2021). PolyG colocalizes with Hsp70 (Bonapace et al., 2019; Jin et al., 2003; Todd et al., 2013) and lamina-associated polypeptide (LAP) 2β, eventually leading to abnormal nuclear envelope structure and cell death (Sellier et al., 2017). Together, these data suggest that multiple mechanisms may play a role in FXTAS and that inclusion formation can be triggered by both toxic RNA and RAN translation products.

The composition of inclusions has been investigated in a few studies. Inclusions purified from FXTAS patients were shown to be composed of multiple proteins, including lamin and hnRNP A2 (Iwahashi et al., 2006). Recent studies have identified over 200 proteins enriched within these inclusions in comparison to whole nuclei in FXTAS (Ma et al., 2019). No predominant protein was observed, but over half of the identified proteins were involved in RNA binding, protein turnover, and DNA damage repair. The most enriched proteins included small ubiquitin-related modifier 2 (SUMO 2), ubiquitin, and p62/sequestosome-1. Although the mechanism of inclusion formation is still not well understood, these results suggest that these inclusions may be a consequence of protein aggregation that exceeds the threshold of proteasomal degradation (Ma et al., 2019) and that aggregates may contain RNA^{exp}.

In accordance with the toxic RNA gain-of-function hypothesis, several proteins were found to bind CGG^{exp}. Among them, the cytoplasmic form of hnRNP A2/B1 interacts directly with CGG^{exp} (Jin et al., 2007; Sofola et al., 2007) and mediates the binding of CELF1 (Sofola et al., 2007). Importantly, overexpression of hnRNP A2/B1 or CELF1 suppressed CGG^{exp} toxicity in *Drosophila*, suggesting their involvement in FXTAS pathogenesis (Jin et al., 2007; Sofola et al., 2007). In *Drosophila*, the overexpression of ALS-associated TAR DNA-binding protein 43 (TDP-43) partially alleviated neurodegeneration, probably by preventing sequestration of the hnRNP A2/B1 protein homologs Hrb87F and Hrb98De on CGG^{exp} and thus restoring their functions (He et al., 2014). Additionally, the transcriptional activator protein Pur-alpha (Pur α) was found to bind CGG^{exp} in vitro and was also observed in the inclusions of post-mortem brain tissues from patients with FXTAS (Jin et al., 2007). The splicing regulator transformer-2 protein homolog alpha (TRA2A) binds CGG^{exp} in vitro and colocalizes with nuclear CGG^{exp} foci and polyG in FXTAS inclusions. Moreover, in cells expressing (CGG)₆₀, TRA2A-dependent splicing of genes linked to mental retardation (i.e., *ACTB*) or intellectual disabilities (i.e., *DOCK3*) was impaired (Cid-Samper et al., 2018).

CGG^{exp} also sequesters the double-stranded RBP DGCR8 and its partner, ribonuclease type 3 (DROSHA). The enzymatic complex microprocessor, in which these two proteins play a crucial role, is involved in the processing of primary

precursors of microRNA (pri-miRNA), and when impaired in FXTAS, mature miRNA levels are reduced in neuronal cells, leading to neurodegeneration (Sellier et al., 2013). Moreover, colocalization studies showed that the DROSHA-DGCR8 complex interacts with the splicing regulator KH domain-containing, RNA-binding, signal transduction-associated protein 1 (KHDRBS1, alternative name SAM68), which leads to its aggregation in CGG^{exp} foci and ultimately to the abnormal AS of several target mRNAs (Sellier et al., 2013). Importantly, overexpression of DGCR8, but not SAM68, rescued neuronal cell death induced by the expression of CGG^{exp} in cultured mouse cortical neurons, suggesting that the sequestration of DGCR8 may play a role in the neuronal degeneration observed in FXTAS and that SAM68 does not directly bind CGG^{exp} (Sellier et al., 2010, 2013).

4.3 | *C9orf72*-linked amyotrophic lateral sclerosis and frontotemporal dementia

The presence of nuclear sense G4C2 and antisense C4G2 RNA foci transcribed from G4C2^{exp} in *C9orf72*, observed in the post-mortem cerebral cortex and spinal cord tissue of C9-ALS/FTD patients, significantly augments the complexity of proposed RNA-mediated toxicity in this disease (DeJesus-Hernandez et al., 2011). Nuclear foci containing the sense transcript are more abundant in cerebellar granule neurons, while antisense RNA-containing foci are more common in Purkinje cells and motor neurons (Cooper-Knock et al., 2015). Multiple RBPs, including Pur α (Xu et al., 2013), Nucleolin (Haeusler et al., 2014), and various hnRNPs (Cooper-Knock et al., 2015; Haeusler et al., 2014; Y. B. Lee et al., 2013; Mori, Lammich, Mackenzie, Forné, Zilow, Kretschmar, et al., 2013), double-stranded RNA-specific editase B2 (ADARB2; Donnelly et al., 2013), serine/arginine-rich splicing factor 1 (SRSF1; Cooper-Knock et al., 2014), THO complex subunit 4 (ALYREF; Cooper-Knock et al., 2014), ZFP106 (Celona et al., 2017), Matrin-3 (Ramesh et al., 2020), and essential paraspeckle proteins (Česnik et al., 2019), were shown to interact/colocalize with G4C2^{exp} or C4G2^{exp} foci, yet the colocalization of a candidate protein with RNA foci does not always mirror its impact on the disease phenotype. More detailed studies to assess the role of RNA-mediated depletion/modification of the activity of these RBPs in C9-ALS/FTD should be performed. Below, we briefly describe some of the candidate proteins and their significance regarding disease phenotype.

In vitro studies showed that Nucleolin, hnRNP U, hnRNP F, and 60S ribosomal protein L7 (RPL7) bind (G4C2)₄ RNA, while hnRNP K preferentially binds antisense (C4G2)₄ RNA (Haeusler et al., 2014). Variable colocalization of Nucleolin with G4C2^{exp} foci was further confirmed in motor cortex tissue and Purkinje neurons, suggesting that Nucleolin sequestration can impair Nucleolin homeostasis and result in nucleolar stress (Cooper-Knock et al., 2015; Haeusler et al., 2014). In line with this, an increase in nucleolar volume was observed in frontal cortex neurons containing RNA foci, showing that nucleolar abnormalities are linked to RNA toxicity (Mizielinska et al., 2017).

The interaction between G4C2^{exp} and Pur α was initially shown in vitro and further confirmed in the mouse Neuro2A cell line and a *Drosophila* model carrying G4C2^{exp}; Pur α inclusions were also observed in ALS/FTD patients with G4C2^{exp}. Importantly, the overexpression of Pur α in both cell and *Drosophila* models alleviated the neurodegeneration mediated by G4C2^{exp} (Xu et al., 2013). The colocalization of Pur α with G4C2 RNA foci was also shown in a zebrafish model of C9-ALS/FTD, in which its overexpression prevented axonal abnormalities induced by toxic RNA (Swinnen et al., 2018).

hnRNP H colocalizes with G4C2^{exp} foci, changing its activity, as inclusion of TARBP2 intron 7, normally orchestrated by hnRNP H, was dysregulated in patient brain tissues. Y. B. Lee et al. (2013) suggested that sequestration of this protein may enhance the nuclear retention and aggregation of G4C2^{exp} RNA, resulting in impaired RNA processing and toxicity. Additionally, hnRNP H1/F, ALYREF, and SRSF2 were shown to colocalize in sense RNA foci in cerebellar granule cells and motor neurons derived from C9-ALS/FTD patients (Cooper-Knock et al., 2014). Later, the same group reported colocalization of SRSF1, hnRNP A1, hnRNP H/F, ALYREF, and hnRNP K with antisense foci in cerebellar Purkinje neurons derived from C9-ALS/FTD patients and confirmed the direct interaction of these proteins, with the exception of hnRNP K, with synthetic (C4G2)₅ repeats by UV cross-linking studies (Cooper-Knock et al., 2015).

The ADARB2 protein, an RNA editing regulator, was also shown to colocalize with G4C2^{exp} foci in patient tissues (Donnelly et al., 2013). The siRNA-mediated knockdown of *ADARB2* decreased the number of RNA foci in C9-ALS/FTD iPSCs, and the cells devoid of ADARB2 were more susceptible to the toxic effects of an excitatory neurotransmitter (glutamate) and showed increased cell death. The use of antisense nucleotides (ASOs) targeting G4C2^{exp} reduced the nuclear retention of ADARB2 and rescued the glutamate susceptibility phenotype (Donnelly et al., 2013).

The ZFN protein ZFP106 was recently reported to bind G4C2 repeats and other RBPs, including TDP-43 and Fused in sarcoma (FUS), suggesting its potential role in C9-ALS/FTD pathogenesis (Celona et al., 2017). In line with this,

Zfp106-knockout mice are characterized by motor neuron and muscle fiber degeneration and muscle atrophy, and this phenotype can be reversed by overexpression of ZFP106 in motor neurons. Additionally, overexpression of ZFP106 in a *Drosophila* C9-ALS/FTD model suppressed neurotoxicity, but did not reduce the expression of G4C2^{exp} RNA, suggesting that this protein may participate in the toxic RNA gain-of-function mechanism of C9-ALS/FTD (Celona et al., 2017).

Colocalization of the punctate and diffuse forms of Matrin-3 with G4C2^{exp} foci was reported in C9-ALS/FTD iPSCs and post-mortem C9-ALS/FTD motor cortex sections and validated by RNA pull-down assays (Ramesh et al., 2020). The expression of Matrin-3 in G4C2 *Drosophila* models improved eye degeneration and suppressed motor deficits but did not change G4C2 mRNA levels. Interestingly, truncation of the Matrin-3 RBD did not fully inhibit binding to G4C2 RNA, suggesting that this interaction is sometimes direct but also mediated to some extent through other protein partners (Ramesh et al., 2020).

Similar to that in FXTAS, protein sequestration in C9-ALS/FTD is not limited to RNA^{exp} foci. RNA^{exp} can be translated bidirectionally, resulting in the production of the DPRs poly-GP, poly-GA, poly-GR, poly-PR, and poly-PA (Figure 1c; Ash et al., 2013; Gendron et al., 2013; Mori, Weng, Arzberger, May, Rentzsch, Kremmer, et al., 2013; Zu et al., 2013). The positively charged, hydrophilic poly-GR repeats were found to form inclusions and colocalize with TDP-43 in an RNA-independent manner in motor neurons derived from C9-ALS/FTD patients, which suggests a role for polyGR in TDP-43 proteinopathy, a hallmark of the majority of ALS and FTD cases (Sabeti et al., 2018). NCT proteins and nucleoporins are also partially colocalized with poly-GR aggregates, suggesting that sequestration of these proteins can contribute to the cytoplasmic accumulation of TDP-43 (Cook et al., 2020). Proteomic analysis of polyGA aggregates in HEK293 cells expressing polyGA-GFP showed enrichment in the UNC119 protein, which is involved in the trafficking of lipidated cargo proteins (May et al., 2014).

In summary, various proteins can be sequestered on RNA^{exp}, and as a result, their physiological functions can be impaired in certain REDs (Table 1). Notably, some proteins do not interact with RNA^{exp} directly but rather co-aggregate with other protein partners during foci formation. To distinguish between toxic protein sequestration and colocalization, mechanistic studies showing impairment of candidate protein functions should be performed. Moreover, the role of RNA-gain of function model and protein sequestration in the pathogenesis of some REDs should be confirmed, as our current understanding is often based on either in vitro studies or simple disease models, with the exception of DM1. Confirmation of pathogenic protein sequestration in patient-derived samples would be a valuable source of knowledge on this subject (Box 3).

BOX 3 Sequestration versus co-localization

Protein sequestration: Selective recruitment and immobilization of specific proteins on RNA molecules in membraneless compartments. Physiological sequestration is used in cellular processes to temporally withdraw proteins from the available pool (e.g., nucleolar sequestration of MDM2 leads to stabilization of p53 in the nucleus, which leads to the cell growth arrest or apoptosis). Toxic sequestration is a pathogenic mechanism, where detained proteins are titrated away from their biological targets, leading to irreversible paucity in specific compartments and inhibition of their physiological roles (e.g., sequestration of MBNL1 on CUG^{exp} causes altered splicing of different pre-mRNA targets, a hallmark of DM1 pathogenesis). Toxic sequestration is often targeted in RED therapies, by using therapeutic agents which bind to toxic RNA and thus release sequestered proteins. **Colocalization:** Spatial overlap of two probes, for example, protein and RNA, which can be detected by fluorescence microscopy. It should be distinguished from sequestration, as colocalization of protein and RNA does not necessarily lead to the impairment of protein activity.

5 | NUCLEOCYTOPLASMIC TRANSPORT

Impairment of nucleocytoplasmic transport (NCT) is a common pathology in neurodegenerative diseases, including those caused by STR expansion, and a growing body of evidence indicates the great importance of this process in the development of neurodegeneration (extensively reviewed in H. J. Kim & Taylor, 2017). A number of reports have suggested the contribution of toxic proteins such as polyQ proteins or RAN translation proteins to the disruption of NCT (Boeynaems et al., 2016; Cook et al., 2020; Grima et al., 2017; Hayes et al., 2020; Jovičić et al., 2015; K. Y. Shi et al.,

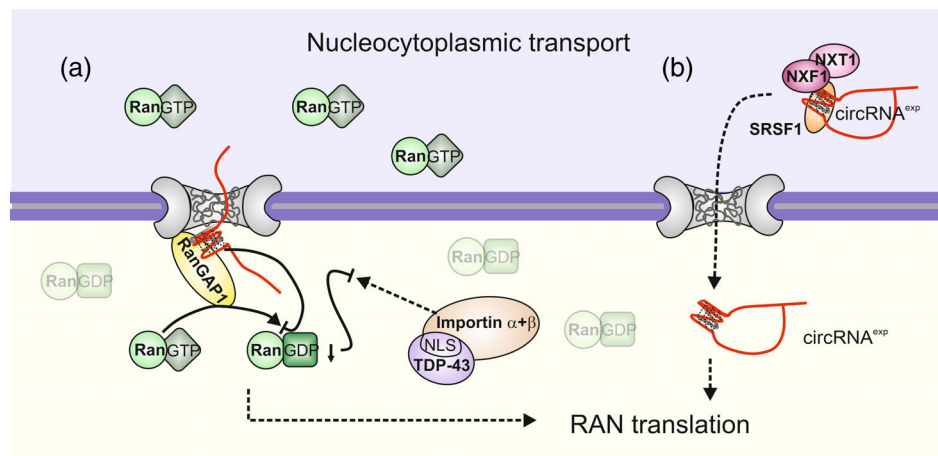


FIGURE 3 Involvement of RNA^{exp} in nucleocytoplasmic transport (NCT). (a) *Impairment of NCT by RNA^{exp}*. Gradient of RanGDP/RanGTP proteins between nucleus and cytoplasm, supported by RanGAP1, enables proper NCT. Binding of G4C2^{exp} to RanGAP1 leads to impaired import of nuclear proteins, exemplified by TDP-43. (b) *Export of RNA^{exp}*. Nuclear export adaptor SRSF1 binds to RNA with G4C2^{exp} and C4G2^{exp} and supports its transport to cytoplasm through NXF1-dependent pathway. NXF1 and its cofactor NXT1 participate also in export of circular RNA^{exp} (circRNA^{exp}) derived from G4C2^{exp}-bearing intron lariats. (a,b) Arrow with a dotted line, change in place and/or in time; solid lines show induction or inhibition of certain processes

2017; Y. J. Zhang et al., 2016). However, some of the articles presented below support the involvement of toxic RNA^{exp} due to the direct binding of proteins that participate in NCT (Figure 3a,b, Table 1).

5.1 | *C9orf72*-linked amyotrophic lateral sclerosis and frontotemporal dementia

RNA with expanded G4C2 repeats exerts toxicity in both the nucleus and cytoplasm; therefore, abnormalities in NCT may significantly contribute to the pathomechanism of C9-ALS/FTD. Genetic screens of a *Drosophila* model of this disease revealed that the pathology caused by G4C2 is related to the impairment of NCT and that Ran GTPase-activating protein (RanGAP, RanGAP1 in humans) is a potent suppressor of G4C2 repeat-associated neurodegeneration (Freibaum et al., 2015; K. Zhang et al., 2015; Figure 3a). RanGAP1 participates in the conversion of cytoplasmic RanGTP to RanGDP, a process required to maintain the Ran protein gradient between the nucleus and cytoplasm, which is necessary for the correct operation of active transport through the nuclear pore complex (Bischoff et al., 1994; Görlich et al., 1996). Binding of RanGAP to the G-quadruplex structure formed by RNA with G4C2 repeats led to disruption of the Ran protein gradient in C9-ALS iPSCs and inhibition of protein import to the nucleus in both a *Drosophila* model and iPSCs (K. Zhang et al., 2015). This is illustrated by nuclear depletion of TDP-43 over the course of ASL/FTD (Neumann et al., 2006; K. Zhang et al., 2015). On the other hand, a recent report showed that an increased concentration of TDP-43 in the cytoplasm may lead to the formation of liquid droplets, further causing mislocalization of RanGAP1 and other proteins engaged in NCT (Chou et al., 2018; Gasset-Rosa et al., 2019). G4C2^{exp} repeat-related neurodegeneration and impairment of NCT were abolished by overexpression of RanGAP or the application of molecules that inhibit its interaction with the repeats (K. Zhang et al., 2015).

Among the proteins that directly bind G4C2 repeat RNA are nuclear export adaptors, serine/arginine-rich splicing factors (SRSFs; Cooper-Knock et al., 2014; Figure 3b). Further studies showed that knockdown of SRSF1 significantly reduced repeat-related neurodegeneration in a fly model of C9-ALS/FTD (Hautbergue et al., 2017). In addition to its splicing regulatory functions, SRSF1 participates in the transport of mRNAs via a nuclear RNA export factor 1 (NXF1)-dependent pathway (Y. Huang et al., 2003; Müller-McNicoll et al., 2016). The proposed model assumes that SRSF1 directly binds RNA with expanded G4C2 or antisense C4G2 repeats and supports its export to the cytoplasm driven by NXF1, where it undergoes RAN translation to toxic DPRs (Hautbergue et al., 2017). Importantly, this process specifically concerns expansion-bearing transcripts, as SRSF1 deficiency reduced only the nuclear export of *C9orf72* mRNA that retained an intron with expanded repeats but did not influence the nuclear export of wild-type *C9orf72* mRNA in C9-ALS iPSCs (Hautbergue et al., 2017). Recently, the NXF1–NXT1 pathway was also shown to regulate the nuclear export of circular RNA containing G4C2^{exp} produced due to defective debranching of the spliced intron of *C9orf72* mRNA (Wang et al.,

2021). Here, the presence of G-rich repeats influences the stability of the circular form and its export to the cytoplasm. Interestingly, such circular RNA^{exp} undergoes RAN translation and supports DPR production (Wang et al., 2021).

Although direct binding was not shown, a recent study suggests that RNA with G4C2^{exp} contributes to the reduction of nuclear envelope pore membrane protein POM121 (POM121; Coyne et al., 2020). The observed effect was RNA^{exp}-specific and independent of DPR translation or loss of the C9orf72 protein (Coyne et al., 2020). Deficiency of POM121 resulted in further decreases in a set of nucleoporins, leading to a significant change in nuclear pore complex composition, mislocalization of RanGTPase, and reduced survival of C9-ALS iPSNs (Coyne et al., 2020).

5.2 | PolyQ diseases

Similar to the interplay between SRSF1 and G4C2 observed in ALS, RNAs with expanded CAG repeats interact with another nuclear export adapter, splicing factor U2AF 65 kDa subunit (U2AF65; Tsoi et al., 2011). U2AF65 directly binds RNA with expanded but not normal CAG repeats and then NXF1, facilitating the export of RNA^{exp} to the cytoplasm (Tsoi et al., 2011). Knockdown of the *Drosophila* ortholog of U2AF65 enhanced the nuclear accumulation of RNA^{exp} and RNA^{exp}-mediated toxicity (Tsoi et al., 2011). In a mouse model of polyQ disease expressing RNA with CAG^{exp}, symptomatic mice showed a significant reduction in U2AF65 in comparison to U2AF65 levels in asymptomatic mice, which confirmed the contribution of this protein to CAG^{exp}-related pathology (Tsoi et al., 2011). Interestingly, subsequent work suggested that nuclear retention of CAG^{exp}-bearing RNA results from its binding to MBNL1, leading to impairment of the interaction with U2AF65 and nuclear export by NXF1 (Sun et al., 2015).

5.3 | Fragile X-associated tremor/ataxia syndrome

Some RBPs that contribute to NCT may also affect this process in FXTAS due to their binding to CGG^{exp} RNA. A recent study presented the direct interaction of SRSF1 with these repeats (Malik, Tseng, et al., 2021). In line with the previously shown effects of the SRSF1/G4C2 interaction (Hautbergue et al., 2017), depletion of SRSF1 activity in a cellular model of FXTAS decreased RAN translation and led to an increased abundance of CGG^{exp}-bearing transcripts in the nucleus (Malik, Tseng, et al., 2021). SRSF1 knockdown rescued repeat-related toxicity in both C9-ALS/FTD and FXTAS *Drosophila* models (Hautbergue et al., 2017; Malik, Tseng, et al., 2021). Interestingly, articles reporting the pathological impact of RNA with G4C2 (Coyne et al., 2020) and CAG (Tsoi et al., 2011) repeats on NCT proteins showed that CGG-related toxicity did not induce mislocalization of POM121 and was not modulated by U2AF65.

5.4 | Myotonic dystrophies

DMPK transcripts containing very long expanded CUG repeats are almost fully retained in the nucleus (Brook et al., 1992; Davis et al., 1997). Nuclear export of such mRNAs is also modulated by RBPs. hnRNP H was shown to hinder this process (D. H. Kim et al., 2005). In contrast, another CUG^{exp}-binding protein, the double-stranded RBP Staufin homolog 1 (Staufen1), was shown to rescue NCT of RNA^{exp} in cellular models of DM1 (Ravel-Chapuis et al., 2012). In DM2, overexpression of MBNL1 increased RNA foci formation and reduced the production of RAN proteins, which suggests its impact on the nuclear retention of CCUG^{exp} (Zu et al., 2017).

In summary, two faces of NCT pathology relate to the interplay between RNA^{exp} and RBPs. The first concerns functional impairment of RBPs engaged in different steps of NCT due to their binding to RNA^{exp}, as exemplified by RanGAP1 and POM121. The second involves the contribution of RBPs to toxicity exerted by different RNA^{exp} molecules in either the nucleus (e.g., CUG^{exp}) or cytoplasm (e.g., G4C2^{exp}) by enabling their retention or transport to the cytoplasm (e.g., hnRNP H or SRSF1, respectively).

6 | RAN TRANSLATION

Repeat associated non-AUG (RAN) translation is a protein synthesis mechanism that, in contrast to canonical translation, does not require an AUG start codon for initiation (Zu et al., 2011). RAN translation was reported for many

diseases associated with repeat expansions, such as HD (Bañez-Coronel et al., 2015), C9-ALS/FTD (Ash et al., 2013; Mori, Weng, Arzberger, May, Rentzsch, Van Broeckhoven, et al., ; Zu et al., 2013), FXTAS (Todd et al., 2013), DM2 (Zu et al., 2017), and others (Banez-Coronel & Ranum, 2019; Zu et al., 2011). RAN proteins may be produced from sense or antisense transcripts and contain homopolymeric tracts comprised of a single amino acid, such as polyglycine (polyG) or polyalanine (polyA) tracts in the case of trinucleotide repeats (Todd et al., 2013), dipeptides repeats (DPRs), such as polyGA in case of G4C2^{exp} in C9-ALS/FTD (Ash et al., 2013; Mori, Weng, Arzberger, May, Rentzsch, Kremmer, et al., 2013; Zu et al., 2013), or tetrapeptide repeats in DM2: polyLPAC and polyQAGR translated from CCUG^{exp} and CAGG^{exp}, respectively (Zu et al., 2017; Figure 1c). RAN products have toxic properties; they mostly tend to aggregate, creating nuclear, or cytoplasmic inclusions (Mori, Weng, Arzberger, May, Rentzsch, Kremmer, et al., 2013; Sellier et al., 2017), and sequester other proteins, resulting in their functional depletion (Sellier et al., 2010, 2013). Insights describing the involvement of RAN translation products in the pathogenesis of particular REDs are beyond the scope of this review and were presented elsewhere (Banez-Coronel & Ranum, 2019). It is worth noting that so far deleterious properties of RAN peptides were mostly observed in experiments based on artificial cellular or animal models (e.g., Sellier et al., 2017; Sonobe et al., 2021; Todd et al., 2013). Indeed, RAN proteins were identified in several patient's samples (Ash et al., 2013; Banez-Coronel et al., 2015; Mori, Weng, Arzberger, May, Rentzsch, Van Broeckhoven, et al., ; Sellier et al., 2017; Todd et al., 2013; Zu et al., 2013, 2017), however, our knowledge about toxicity of those endogenously expressed peptides is limited (Freibaum & Taylor, 2017). For example, current findings concerning the presence of polyG in FXTAS patients implicate that RAN translation of CGG^{exp} may occur at relatively low level and it is not well-defined to what extent this process contributes to the pathogenesis of this disease (Ma et al., 2019; Salcedo-Arellano et al., 2020). Additionally, RAN translation can also occur when transcripts contain repeats of a normal range, indicating the possible regulatory functions of this process under physiological conditions (C. M. Rodriguez et al., 2020).

Despite the fact that RAN translation was first described nearly a decade ago, mechanistic insights into this process remain elusive (Zu et al., 2011; Figure 4a–c). In the majority of cases, initiation of RAN translation depends on a canonical cap-dependent scanning model; however, it was shown that AUG-independent translation initiates at less favored so-called near-cognate codons (such as CUG, GUG, and ACG) located upstream or within expanded, structured repeats (Green et al., 2017; Kearse et al., 2016; Tabet et al., 2018). RAN translation of DPRs from G4C2^{exp} was shown to initiate with Met-tRNA at CUG start codon in a cap-dependent manner (Green et al., 2017; Tabet et al., 2018). The mechanism of translation initiation from alternative start codons has not been fully elucidated; however, one of the proposed models of CGG RAN translation in FXTAS involves a mechanism in which the 43S preinitiation complex (PIC) scans through the 5'UTR of *FMRI* mRNA and encounters steric hindrances such as hairpins and/or G-quadruplexes formed by expanded repeats, which leads to ribosomal stalling and the initiation of translation at noncanonical codons, resulting in the production of RAN polypeptides (Kearse et al., 2016). Resolving structured RNA seems to be crucial for RAN translation, as it was shown that inhibition of the canonical DEAD box helicase eIF4A, which is needed for ribosome scanning, abolished both CGG^{exp}- and G4C2^{exp}-dependent RAN translation (Green et al., 2017; Kearse et al., 2016). It was demonstrated that the costimulatory factors of eIF4A, eIF4B, and eIF4H, also exhibit RAN translation modulatory properties (Linsalata et al., 2019). In addition, in a C9-ALS/FTD fly model, the *Drosophila* homologs of eIF4B and eIF4H were shown to modulate the production of polyGR, and eIF4H was found to be significantly down-regulated in cerebellar tissue and fibroblasts obtained from patients harboring G4C2^{exp} (Goodman, Prudencio, Srinivasan, et al., 2019). Recently, *Drosophila* Belle (a homolog of the human DEAD-box helicase DDX3X) was found to selectively affect the production of RAN peptides (Linsalata et al., 2019). The authors reported that DDX3X binds the 5'UTR of *FRMI* independent of its CGG repeats, excluding the possibility that repeat expansion determines the interaction between mRNA and the protein (Linsalata et al., 2019). Knockdown of Belle/DDX3X in an FXTAS fly model resulted in the decreased production of RAN dipeptides and ameliorated retinal degeneration (as expression of CGG^{exp} was directed to the retina using the Gmr-GAL4 driver, which manifests as a severe so-called rough-eye phenotype), confirming the regulatory role of DDX3X in vivo. In contrast, DDX3X depletion upregulated G4C2^{exp}-specific RAN translation (W. Cheng et al., 2019). This study showed that DDX3X binds G4C2^{exp} repeats and relaxes their secondary structure, thus impeding initiation of the RAN translation process (W. Cheng et al., 2019). This is not surprising, as it was previously shown that the yeast homolog of DDX3X, Ded1p, controls initiation from near-cognate codons by binding and unwinding RNA secondary structures upstream of alternative codons within 5'UTRs (Guenther et al., 2018). Recently, another DEAH-box RNA helicase, DHX36, was identified as a new modulator of G4C2 or CGG RAN translation (H. Liu et al., 2021; Tseng et al., 2021; Figure 4b). The authors postulate that similar to DDX3X, DHX36 unfolds RNA secondary structures formed within G4C2 repeats and thus modulates noncanonical protein synthesis (H. Liu

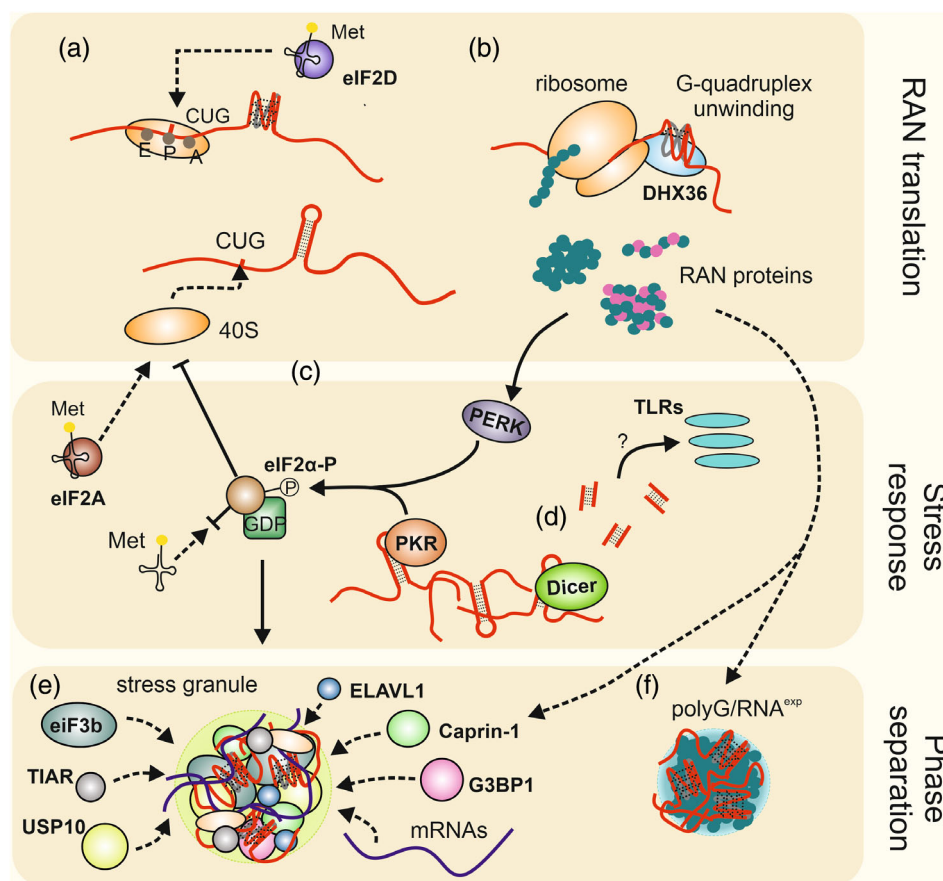


FIGURE 4 Processes involving RBPs and RNA^{exp} in cytoplasm. (a) *The role of eIF2D in RAN translation initiation.* Non-canonical translation initiation factor eIF2D delivers Met-tRNA to the P-site of 40S ribosomal subunit at CUG codon contributing to RAN translation initiation (described for G4C2^{exp}). A (aminoacyl) site, P (peptidyl) site, E (exit) site in the ribosome. (b) *Elongation of RAN translation.* DHX36 helicase unwinds G-quadruplexes formed by RNA^{exp} and thus facilitates ribosome processivity and production of toxic homopolymeric proteins or dipeptide repeat proteins (described for G4C2^{exp} and CGG^{exp}). (c) *RAN translation initiation upon stress.* Stress related to the presence of double-stranded RNA (dsRNA) formed by RNA^{exp} and RAN proteins activates PKR and PERK kinases, respectively, which catalyze phosphorylation of eIF2 α . This, in turn, inhibits eIF2 α -P binding to Met-tRNA and the formation of preinitiation complex (PIC) with 40S ribosome subunit. Under the stress, eIF2A may take over role of phosphorylated eIF2 α -P, bind Met-tRNA and participate in translation initiation at near-cognate start codons (described for G4C2^{exp} and CCUG^{exp}). (d) *Stress response caused by RNA^{exp}.* RNAi pathway component, ribonuclease Dicer, cleaves dsRNA formed by RNA^{exp} hairpin or bidirectionally transcribed CUG^{exp}/CAG^{exp} duplex into 21-mer fragments. Such RNA fragments may next activate TLRs and trigger innate immune response. (e) *Stress granules formation.* Stress stimuli such as RNA^{exp} and RAN proteins lead to phosphorylation of eIF2 α , followed by global translation suppression and formation of stress granules (SG; described for CGG^{exp} and G4C2^{exp}). Moreover, G4C2^{exp} may serve as a core component of SG and promote formation of membraneless organelles composed of different mRNAs and SG protein markers. Chronic stress may entail SG transition towards more solid-like structures. (f) *Phase separation of RNA^{exp} and RAN protein.* Homopolymeric RAN protein, polyG, binds to its own RNA^{exp} what promotes its phase transition from liquid droplets towards gel-like aggregates (described for CGG^{exp}). (a–c) arrow with a dotted line, change in place and/or in time; solid lines show induction or inhibition of certain processes

et al., 2021; Tseng et al., 2021). All of the presented evidence indicates that resolving the RNA^{exp} structure by RBPs seems to be a crucial element of RAN translation regulation. Additionally, start codon fidelity seems to substantially contribute to the regulation of RAN translation, as a recently performed screening elucidated two other translation initiation factors, eIF1 and eIF5, that influence the initiation of RAN protein synthesis in vitro (Linsalata et al., 2019), which is in line with previously published reports concerning the function of these proteins, which is contribution to start codon selection (Ivanov et al., 2010; Kozel et al., 2016; Loughran et al., 2012; Tang et al., 2017). Recent in vitro studies demonstrated that knockdown of eIF3F selectively downregulates synthesis of RAN peptides derived from expanded CAG repeats (related to SCA8) as well as G4C2 repeats (Ayhan et al., 2018). The eIF3F acts as a non-core component of eIF3 complex, which is known to regulate canonical and internal ribosome entry site (IRES)-dependent

translation but its role in RAN translation regulation is yet to be defined (Cate, 2017; Hashem et al., 2013). Further studies demonstrating the role of eIF3 in both canonical and RAN translation initiation on CGG^{exp} indicated that human oncoprotein eIF5-mimic (5MP) displaces eIF5 through eIF3c subunit within the PIC and thus affects both modes of translation via increasing the accuracy of translation initiation (Singh et al., 2021; Tang et al., 2017). In addition, it was shown that overexpression of *Drosophila* 5MP homolog in FXTAS flies reduced RAN peptides-mediated toxicity (Singh et al., 2021). Recently, Sonobe et al. (2021) developed a new *C. elegans* model of C9-ALS/FTD. The study showed that loss-of-function mutations of eIF2D lead to reduced production of polyGA (and to lesser extent polyGP), positively affecting locomotor function, and lifespan of transgenic worms (Sonobe et al., 2021). Authors presented that eIF2D is required for DPRs production and perhaps acts at initiation CUG codon for polyGA and delivers the Met-tRNA to the P-site of the 40S ribosomal subunit (Boivin et al., 2020; Dmitriev et al., 2010; Green et al., 2017; Sonobe et al., 2021; Figure 4a).

A genetic screening performed to identify RAN translation-specific modifiers in a yeast model identified 40S ribosomal protein S25-A (RPS25A), which selectively inhibited RAN translation from G4C2^{exp} (Yamada et al., 2019). Treatment with ASOs targeting RPS25 in C9-ALS iPNSs reduced DPRs production and significantly increased cell survival. Moreover, upon RNAi-mediated silencing in C9-ALS/FTD *Drosophila*, the efficiency of polyGP synthesis significantly decreased, and the lifespan of the adult male flies was extended. Additionally, in human cell lines, RPS25 silencing led to a decrease in RAN translation from another repeats, CAG^{exp} associated with the *HTT* and *ATXN2* genes (related to HD and spinocerebellar ataxia type 2 [SCA2], respectively). The exact function of RPS25 in translation remains to be determined; however, RPS25 was shown to play a crucial role in both cellular and viral noncanonical translation events, such as ribosomal shunting and IRES-mediated translation (Fuchs et al., 2015; Hertz et al., 2013; Landry et al., 2009; Y. Shi et al., 2016). As RPS25 exhibits specificity toward stable RNA secondary structures, structured G4C2 repeats might attract this protein (Nishiyama et al., 2007). Interestingly, RPS25 is not required for cap-dependent translation, which might explain why this factor plays an important role in G4C2 RAN translation of the *C9orf72* transcript, which can be translated in IRES-like cap-independent manner (W. Cheng et al., 2018; Wang et al., 2021).

Altogether, the presented reports imply that RBPs modulate RAN translation by contributing to multiple translation steps, such as PIC scanning, start codon fidelity, unfolding structured RNA^{exp}. Recent findings provide multiple lines of evidence that particular RBPs can modulate RAN translation in various ways, sometimes leading to contrary effects depending on repeat type, length and RNA sequence context (Table 1, Box 4).

BOX 4 RAN translation

Repeat-associated non-AUG (RAN) translation is a non-canonical protein synthesis initiated upstream or within expanded repeats in so-called near cognate start codons such as ACG, CUG, and GUG. RAN translated proteins may consist of single amino acids such as polyglycine (FXTAS) or repeated dipeptides (DPRs; C9-ALS/FTD). These RAN peptides tend to aggregate, interact with both RNAs and other proteins, and thus form toxic intracellular inclusions. Mechanistic insights of RAN translation are yet to be established, however, it is known that such processes as ribosomal scanning, start codon fidelity, RNA secondary structure unwinding and cellular stress contribute to RAN translation regulation. RAN translation might be also perceived as combination of RNA/protein gain-of-function phenomena as structured RNA^{exp} acts as template for production of toxic RAN peptides, affects the efficiency of RAN translation and may contribute to the formation of aggregates with RAN products.

7 | IMMUNE RESPONSE TO RNA^{exp} AND PRODUCTS OF RAN TRANSLATION

Growing evidence suggests that RNA^{exp} or double-stranded RNA (dsRNA), e.g., products of bidirectional transcription in *locus* of expanded repeats, or RAN proteins itself can induce an innate immune response. They can act as danger-associated molecular patterns (DAMP), and are recognized by pattern recognition receptors (PRRs), including Toll-like receptors (TLR), or double-stranded RNA-activated protein kinase (PKR) in various types of brain cells. These events

might correlate with pathological innate immune system activation, induction of stress response, and inflammation. The presence of structured RNAs containing different types of repeats, for example, CUG^{exp}, CAG^{exp} or G4C2^{exp}, and RAN proteins, for example, polyG or polyGP, was postulated to activate PKR, and PKR-like endoplasmic reticulum kinase (PERK) pathway, respectively (Figure 4c; W. Cheng et al., 2018; Green et al., 2017; Tian et al., 2000; Zu et al., 2020). PKR activation leads to the phosphorylation of eIF2 α , resulting in global translation shutdown, but simultaneously selective upregulation of RAN translation. Expression of expanded CGG and G4C2 repeats in primary rat cortical neurons triggered eIF2A-phosphorylation-dependent stress granule formation. Activation of integrated stress response enhanced RAN translation and halt global translation, contributing to neurodegeneration in a feed-forward loop (Green et al., 2017). The cytoplasmic dsRNA derived from G4C2 expansions was detected in brains of C9-ALS/FTD patients, and when expressed in a mouse model it triggered type-I interferon signaling and neuronal death (S. Rodriguez et al., 2021). Several strategies to inhibit eIF2 α phosphorylation and thus decrease the accumulation of RAN proteins were successfully applied in both in vitro and in vivo models (W. Cheng et al., 2018; Green et al., 2017; Westergard et al., 2019; Zu et al., 2020). When eIF2 α activity is abolished, eIF2 alternative factor, eIF2A, can participate in translation regulation (Komar & Merrick, 2020). Upon eIF2A knockout, synthesis of polyGA from G4C2^{exp} was also shown to be decreased (Sonobe et al., 2018). In addition, eIF2A can also increase the production of RAN proteins translated from both CCUG and CAGG repeats in cellular DM2 models (Tusi et al., 2021).

Interestingly, although the CGG^{exp} repeats forms hairpins/G-quadruplex, it does not activate the PKR in vitro or in vivo and is inefficiently cleaved by a major component of RNAi machinery Dicer in vitro (Handa et al., 2003). On the other hand, in patient-derived cells with mutant *DMPK* or mutant *HTT*, CUG^{exp} and CAG^{exp} are cleaved by ribonuclease Dicer into short ~21 nt RNA fragments (Krol et al., 2007; Figure 4d). In *Drosophila* model, transiently expressed transcripts containing CUG^{exp} and CAG^{exp} were cleaved by Dicer-2 and loaded on Ago2-RISC complex (Yu et al., 2011). This finding was confirmed in another fly model, where dsRNA of CAG/CUG₋₁₀₀ repeats was produced in endogenous bidirectional transcription, and was shown to be processed by Dicer-2 into CAG₇ 21-mers (Lawlor et al., 2011). Primary activity of Dicer is the cleavage dsRNAs in RNAi pathway; however, its role in antiviral activity was also reported recently (Montavon et al., 2021). Dicer requires several protein partners to cleave dsRNA in RNAi pathway, and recent studies revealed that other key components of RNAi pathway, namely R2D2 and loquacious, are not required for CAG/CUG₋₁₀₀ dsRNA toxicity in *Drosophila*, in the contrary to Dicer and Argonaute (van Eyk et al., 2019). This suggests that Dicer and Argonaute are implicated in antiviral RNA^{exp}-mediated cell death pathway rather than in RNAi pathway.

Adenosine deaminase of RNA-1 (ADAR1) is another important player in antiviral immunity, as it edits dsRNA to distinguish viral double-stranded nucleic acid (non-self) from endogenous dsRNA (i.e. Alu repeats). Depending on the number of repeats, RNA^{exp} can be conferred as “non-self” and thus triggers autoinflammatory response. Co-expression of human hADAR1c or hADAR1i with CAG/CUG₁₀₀ in *Drosophila* rescued the inflammatory phenotype (van Eyk et al., 2019).

Whole-transcriptomic analysis of *Drosophila* expressing CAG/CUG₋₁₀₀ repeats revealed alterations in several pathways of innate immunity (Samaraweera et al., 2013). Upon knockdown of TLRs in *Drosophila*, the toxicity of CAG/CUG₋₁₀₀ repeats decreased, whereas the opposite effect was observed upon knockdown of autophagy-related genes. This suggests that activation of *Toll* pathway in the presence of dsRNA contributes to neurodegeneration, and that autophagy is required for clearance of pathogenic dsRNA (Samaraweera et al., 2013).

Together, many reports show the link between innate immunity, toxic dsRNA, and RAN proteins. These aberrant molecules may induce cellular stress in a feed-forward loop, leading to global protein translation shutdown, but increase in RAN translation, thus contributing to disease progression. Nevertheless, further studies are needed to better understand the involvement of immune system in REDS.

8 | PHASE SEPARATION

In recent years, efforts to elucidate the underlying causes of neurodegenerative diseases have turned to the phenomenon of phase separation and its contribution to the formation of protein aggregates and RNA foci (comprehensively discussed in excellent reviews; Alberti & Dormann, 2019; Banani et al., 2017; Hyman et al., 2014; Nedelsky & Taylor, 2019; Polymenidou, 2018; Wolozin & Ivanov, 2019). Since RNA molecules and RBPs are involved in phase separation, alterations in the physiology of this process seem to be particularly relevant for REDS. Briefly, liquid-liquid phase separation (LLPS) is the process in which two fluids demix and one form droplets within the other, as in a mixture of oil and

vinegar. In living cells, macromolecules separate from their surroundings to perform specific functions and create assemblies called membraneless organelles (MLOs). Among MLOs in the nucleus are the nucleolus, paraspeckles, and Cajal bodies, whereas MLOs in the cytoplasm are exemplified by processing bodies (P-bodies), transport granules, or stress granules (SG). An important feature of these structures is their flexible response to changing conditions. The transition of liquid droplets to more solid-like foci influences the dynamics of their assembly and disassembly. This mechanism, widely investigated in SG, is now considered a source of the toxic foci/aggregates observed in neurodegenerative disorders, including REDs.

Proteins predisposed to undergo LLPS are characterized by multivalency and thus the ability to interact with multiple other proteins or RNA molecules or the presence of intrinsically disordered regions, which are fragments with low sequence complexity that enable noncovalent interactions with other macromolecules (reviewed in Shin & Brangwynne, 2017). Many RBPs possess these features and thus tend to undergo LLPS. The contribution of some RBPs and their mutant forms to aberrant phase separation in neurodegenerative disorders has been widely investigated (Bakthavachalu et al., 2018; Bolognesi et al., 2019; Pakravan et al., 2021; Patel et al., 2015).

Similar to proteins, RNA molecules can undergo phase separation. It was shown that different RNA^{exp} molecules (RNAs bearing CUG, CAG, and G4C2 repeats), which are multivalent by nature, form droplets in isolation in vitro (Jain & Vale, 2017). This phenomenon occurs when the number of repeats exceeds a threshold value, similar to this, which is pathogenic in particular diseases (Jain & Vale, 2017). Consistent with these observations, RNA^{exp} molecules accumulate in nuclear foci in different REDs. Interestingly, RNA^{exp} droplets in vitro are characterized as a gel because of their solid-like features (Jain & Vale, 2017).

In the cellular environment, the formation of MLOs depends on the interactions of proteins and RNAs (Langdon et al., 2018; Maharana et al., 2018; Van Treeck et al., 2018). As stated above, a number of RBPs, including those associated with LLPS, are bound to different RNA^{exp} (e.g., TDP-43, FUS, hnRNPs). However, surprisingly, the interplay between RNA^{exp} and RBPs in the context of aberrant phase separation in REDs has barely been recognized.

The aforementioned work (Jain & Vale, 2017) concerning the phase separation of RNA^{exp} in vitro also examined RNA^{exp} droplets in living cells. Induction of the expression of RNA with CAG^{exp}, but not with normal CAG repeat number resulted in the formation of liquid-like foci that colocalized with serine/arginine-rich splicing factor 2 (SRSF2 aka SC-35), a marker of nuclear speckles, MLOs involved in splicing regulation (Mintz et al., 1999). Thus, the interaction of CAG^{exp} RNA with nuclear speckle components leads to their retention in the nucleus (Jain & Vale, 2017). RNA with G4C2 repeats, but not antisense RNA with C4G2 repeats, formed nuclear foci in a length-dependent manner in vitro and in living cells (Jain & Vale, 2017). The observed difference may be associated with the secondary structures of those molecules, as the presence of G-quadruplexes promotes phase separation (Fay et al., 2017; Jain & Vale, 2017; Y. Zhang et al., 2020; Zhang et al., 2019). Moreover, foci driven by the expression of G4C2^{exp} RNA are less dynamic than CAG^{exp} RNA foci but are similarly retained in nuclei and colocalize with nuclear speckles (Jain & Vale, 2017; Y. B. Lee et al., 2013).

RNA bearing G4C2 repeats was also shown to drive the formation of SG in a length-dependent manner both in vitro and in cellulo (Fay et al., 2017; Rossi et al., 2015; Figure 4e). RBPs specific for these MLOs were identified in foci assembled in vitro with the use of G4C2 RNA and lysates derived from different cell lines or the mouse brain (Fay et al., 2017). Condensed SG markers included Ras GTPase-activating protein-binding protein 1 (G3BP1), Caprin-1, ubiquitin carboxyl-terminal hydrolase 10 (USP10), eukaryotic translation initiation factor 3 subunit B (eIF3b), ELAV-like protein 1 (ELAVL1), and nucleolysin TIAR (Fay et al., 2017). This process was promoted by the G-quadruplex structure of G4C2 repeats (Fay et al., 2017). RNA^{exp} bearing CAG and G4C2 repeats may also be components of mRNA transport granules in neurons, as they undergo active transport in neurites (Burguete et al., 2015). G4C2^{exp} RNA molecules interfere with the functions of these MLOs, possibly due to interaction with their RBP components, such as FMRP, leading to defects in neuritic branching (Burguete et al., 2015).

Interesting data concerning the interaction of RBPs and RNA^{exp} in the context of aberrant LLPS come from studies of FXTAS-related CGG^{exp} RNA repeats. As for G4C2^{exp}, overexpression of CGG^{exp} induced the formation of SG positive for the markers G3BP1 and FMRP (Green et al., 2017). Computational analyses revealed that *FMR1* mRNA carrying CGG^{exp} is a candidate scaffolding RNA component of RNP MLOs (RNP granules) exemplified by SG or P-bodies (Cid-Samper et al., 2018). Such scaffolding RNAs are characterized by the presence of secondary structure within UTRs and by a large number of interactions with MLO-related RBPs, among other characteristics (Cid-Samper et al., 2018). Indeed, the results of computational analysis, which were further validated using in vitro assays, showed that CGG^{exp} molecules interact with a considerable number of RBPs (Cid-Samper et al., 2018). A significant portion of the interactors are RBPs, which are predicted to form MLOs (Cid-Samper et al., 2018). Considering its strong CGG^{exp}-binding

and granule-forming propensities, the splicing regulator TRA2A was determined to be a potential player in CGG^{exp}-related toxicity (Cid-Samper et al., 2018). Its colocalization with polyG-positive nuclear inclusions was then reported in the brains of an FXTAS mouse model and patients (Cid-Samper et al., 2018). Surprisingly, polyG, the product of RAN translation of CGG^{exp} embedded in *FMR1* mRNA, was recently described as a protein that binds its own encoding mRNA, preferentially the expanded CGG repeats (Asamitsu et al., 2021; Figure 4f). It was shown that both RNA with CGG^{exp} and polyG undergo phase separation in vitro (Asamitsu et al., 2021). Furthermore, the addition of CGG^{exp} RNA to polyG enhanced the phase transition, leading to the formation of gel-like aggregates in vitro and possibly also in cells (Asamitsu et al., 2021). Taking the presented reports into consideration, one may conclude that the scaffolding properties of CGG^{exp}-bearing RNA and possibly other RNA^{exp} molecules lead to the formation of droplets containing RBPs and RAN translation products, which then undergo a transition to solid aggregates, the hallmark of neurodegenerative disorders. It is worth noting that DPRs, products of G4C2 repeat RAN translation, were shown to contribute to aberrant phase transition and disruption of MLO dynamics and functions (Boeynaems et al., 2017; K. H. Lee et al., 2016).

In summary, the RNA^{exp} molecules form foci by themselves and are an attractive platform for interaction with RBPs. The appearance of such unusual molecules changes cell flexibility and modulates MLO assembly and disassembly (Table 1). The presence of different types of inclusions in REDs may be a consequence of this impairment, followed by the liquid-to-solid transition of MLOs.

9 | TARGETING RBPs AS POTENTIAL THERAPEUTIC STRATEGIES

For the majority of REDs, targeting RNA^{exp} molecules by oligonucleotide therapeutics or small compounds stands as a significant strategy directed to combat repeat-associated toxicity (Scoles & Pulst, 2018). In particular, ASOs were shown to successfully target mutated transcripts harboring expanded repeats and alleviate RNA^{exp}-mediated toxicity related to HD (Lane et al., 2018; Tabrizi et al., 2019), FXTAS (Derbis et al., 2021), C9-ALS/FTD (Donnelly et al., 2013; Lagier-Tourenne et al., 2013; Sareen et al., 2013), DM1 (Jauvin et al., 2017; Wheeler et al., 2009; Wojtkowiak-Szlachcic et al., 2015), and other REDs in vivo models (Scoles & Pulst, 2018). Such an approach seems to be very promising, as some ASO-based drugs, such as nusinersen and eteplirsen, which target specific RBP-binding site, have already been approved by the Food and Drug Administration (FDA) and contribute to the treatment of neurodegenerative diseases (Goodkey et al., 2018; Lim et al., 2017). Another approach to tackle RNA^{exp}-related toxicity is the use of small molecules targeting the structural motifs of different RNA^{exp} (Angelbello et al., 2019, 2021; Khan et al., 2019). Current findings concerning this approach are detailed elsewhere (Angelbello et al., 2020; Crunkhorn, 2021), therefore, in this review, we would like to focus on strategies targeting RBPs that contribute to the pathogenesis of REDs.

One therapeutic strategy for DM1 is to restore the activity of MBNL proteins, which are sequestered on CUG^{exp} or CCUG^{exp}, to its physiological level, which seems to be a successful approach to correct AS defects and improve overall phenotype (Chamberlain & Ranum, 2012; Kanadia et al., 2006; Wheeler et al., 2009). Flow cytometry-based genetic screening allowed the identification of two histone deacetylase (HDAC) inhibitors, ISOX, and vorinostat, which increased MBNL1 levels and resulted in the partial correction of alternative splicing in DM1 patient-derived fibroblasts (F. Zhang et al., 2017). A recent screening revealed that the inhibition of cyclooxygenase 1 (COX-1) by nonsteroidal anti-inflammatory drugs (NSAIDs) caused demethylation of the MeR2 enhancer and restored *MBNL1* mRNA levels in myogenic cells (K. Huang et al., 2020). Another NSAID, phenylbutazone (PBZ) was shown to suppress methylation of an enhancer region in *Mbnl1* intron 1 and successfully upregulate MBNL1 levels in cells and a DM1 mouse model (G. Chen et al., 2016). Moreover, PBZ treatment reduced the recruitment of MBNL1 to CUG^{exp} foci in cellulo, ameliorated splicing, mitigated muscle pathology, and improved scores on mouse physical testing (G. Chen et al., 2016). Such results prove the potential value of NSAIDs as potent therapeutic candidates for DM1, especially because those drugs are already a part of the treatment regimens for other neurodegenerative diseases (G. Chen et al., 2016; H. Chen et al., 2005; Hirohata et al., 2005; Uaesoontrachoon et al., 2014). Another strategy to increase MBNL1 levels is the use of antagomiRs to silence miR-23b and miR-218, which regulate MBNL1 levels in muscle (Cerro-Herreros et al., 2018). MBNL1 derepression by antagomiRs rescued missplicing of muscle transcripts and improved functional phenotypes in DM1 mice (Cerro-Herreros et al., 2018, 2020).

Targeting pathologically increased autophagy in DM1 is another therapeutic strategy, as treatment with an anti-autophagic drug, chloroquine, in human-derived myoblasts, *Drosophila* model and DM1 mouse elevated the level of MBNL1, corrected splicing, reduced muscle degeneration, extended the lifespan of the flies, and finally improved the functionality and histopathology of murine muscle tissues (Bargiela et al., 2015, 2019). A correlation between low

miR-7 levels and increased autophagy was observed in biopsies from DM1 patients; thus, antagomiR-7 treatment was applied, which indeed improved myotube diameter and the fusion capacity in differentiating DM1 myoblasts; however, the effect turned out to be independent of MBNL1, suggesting that miR-7 acts downstream or alongside MBNL1 in the pathogenesis of DM1 (Sabater-Arcis et al., 2020).

One of the pathological phenotypic hallmarks of C9-ALS/FTD is the presence of an excessive amount of TDP-43 in SG in the patient brain, which causes toxicity in neurons (Barmada et al., 2010). As stress conditions elicit the phosphorylation of eIF2 α regulated by PERK, the small molecule GSK2606414, an inhibitor of PERK, was shown to successfully mitigate TDP-43-dependent toxicity, improve mobility in C9-ALS/FTD flies, and increase the survival rate of neurons (H. J. Kim et al., 2014; Wek et al., 2006). Small-molecule screening focused on the modulation of SG formation revealed that molecules with planar moieties, such as mitoxantrone, can decrease the accumulation of TDP-43 and other RBPs in SG (Fang et al., 2019). Another small molecule, nTRD22, was shown to bind the N-terminus of TDP-43, affecting RNA-binding properties and the degradation of TDP-43 in vitro and in vivo (Mollasalehi et al., 2020).

A study conducted in a C9-ALS/FTD mouse model indicated that inhibiting the stress-triggered PKR pathway by metformin treatment reduced the level of RAN translation. Metformin prevented the excessive accumulation of RAN peptides and improved the mouse phenotype (Zu et al., 2020). Metformin is already approved by the FDA and serves as a well-established treatment for type 2 diabetes (Sanchez-Rangel & Inzucchi, 2017) and is part of the treatment regimens for DM1 (Bassez et al., 2018), FXS (Dy et al., 2018), and HD (Hervás et al., 2017). Recent findings demonstrate that targeting another kinase, SRSF protein kinase 1 (SRPK1), the protein responsible for the phosphorylation of SRSF1 by the compounds: SRPIN340 and SPHINX31, significantly suppressed CGG^{exp} or G4C2^{exp} RAN translation in vitro, and depletion of the *Drosophila* SRPK1 homolog reduced repeat-associated toxicity and improved the rough-eye phenotype in flies expressing both types of repeats (Malik, Tseng, et al., 2021).

In summary, the presented findings underlie the value of MBNL1 rescue and regulation of autophagy in DM1 therapy by selected drugs, paving the way for further clinical validation. Targeting kinases related to stress response pathways seems to be encouraging, as the presence of structured RNA^{exp} molecules corresponds with elevated cellular stress, which is common among many REDs. However, although some already approved drugs might be applied to combat neurodegenerative diseases, new approaches to enrich existing therapies are still needed.

10 | CONCLUSION

RNA^{exp} acts as a very attractive platform that supports interactions with a number of RBPs; thus, wherever and whenever RNA^{exp} molecules appear, they distract proteins from performing their functions. To exert toxicity and drive the development of REDs, RNA^{exp} molecules require their “partners in crime” at various stages of their life cycle.

At the transcriptional level, Pol II is required for the transcription of GC-rich repeat sequences. On the one hand, RNA^{exp} impairs Pol II initiation and elongation, but in turn, the DSIF complex and PAF1C promote Pol II transcription at RNA^{exp} sites and may be potential therapeutic targets of REDs.

NCT is impaired in REDs and leads to imbalance in homeostasis between the nucleus and cytoplasm. Some evidence indicates that RNA^{exp} molecules bind RBPs involved in NCT, disturbing their functions. On the other hand, RNA^{exp}/RBPs interplay may influence the cellular localization of RNA^{exp} molecules and help them exert toxicity in specific compartments. Thus, it is important to consider potential therapeutic strategies targeting RNA^{exp} or interacting RBPs in the context of pathogenic effects triggered by these molecules in different cellular compartments.

Multiple proteins can be sequestered on RNA^{exp} via direct interaction or coaggregation with protein partners during nuclear foci formation, leading to deterioration of their physiological functions. This impairment can often be observed as symptoms of a disease, that is, in DM1, in which the sequestration of MBNL proteins on CUG^{exp} leads to aberrant AS and contributes to the disease phenotype in a CUG^{exp} load-dependent manner. Importantly, foci formation by RNA^{exp} and sequestered proteins probably changes cell flexibility and impairs LLPS. More studies are needed to decipher the relationship between RNA^{exp} and RBPs, which are MLO components, and their contribution to the liquid-to-solid transition leading to inclusion formation. Moreover, due to recent findings concerning the influence of CGG^{exp} on the phase transition of its RAN translation product, polyG, a new and exciting research field related to the convergence of RNA gain-of-function and RAN translation/protein gain-of-function mechanisms has opened up.

Toxic RAN peptides derived from non-AUG-initiated translation significantly contribute to the pathogenesis of many REDs. Recent findings provide multiple lines of evidence that RBPs play diverse roles in modulating RAN

translation starting from ribosome positioning, the unwinding of structured RNA^{exp}, and responding to stress conditions. The significance of these RBPs makes them promising therapeutic targets, especially because their modulatory properties have been confirmed by in vivo experiments.

Over the last decade, many studies have been devoted to rescue proteins trapped in RNA foci and the modulation of proteins involved in the stress response. Emerging studies suggest that RNA^{exp} and RAN proteins can be involved in the activation of innate immune system. The inflammation can account for pathogenesis in many REDs and be a possible target for therapeutic intervention. Much attention has been given to approved drugs and testing them in the context of potential REDs therapy; however, recent discoveries have pointed out a number of potential targets, requiring medical advancements. Additionally, more research on the development of biomarkers, such as RBPs involved in REDs, is desired to enable accurate diagnosis and monitoring of disease progression and effectiveness of given treatment.

CONFLICT OF INTEREST

All authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Anna Baud: Conceptualization (equal); funding acquisition (equal); writing – original draft (equal). **Magdalena Derbis:** Conceptualization (equal); visualization (equal); writing – original draft (equal). **Katarzyna Tutak:** Conceptualization (equal); writing – original draft (equal). **Krzysztof Sobczak:** Conceptualization (equal); funding acquisition (equal); supervision (equal); writing – review and editing (equal).

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Anna Baud  <https://orcid.org/0000-0003-3710-5722>

Katarzyna Tutak  <https://orcid.org/0000-0002-2801-9956>

Krzysztof Sobczak  <https://orcid.org/0000-0001-8352-9812>

RELATED WIREs ARTICLES

[Translational control in aging and neurodegeneration](#)

[Splicing and neurodegeneration: Insights and mechanisms](#)

[The role of RNA G-quadruplexes in human diseases and therapeutic strategies](#)

REFERENCES

- Alberti, S., & Dormann, D. (2019). Liquid-liquid phase separation in disease. *Annual Review of Genetics*, 53(1), 171–194. <https://doi.org/10.1146/annurev-genet-112618-043527>
- Angelbello, A. J., Benhamou, R. I., Rzuczek, S. G., Choudhary, S., Tang, Z., Chen, J. L., Roy, M., Wang, K. W., Yildirim, I., Jun, A. S., Thornton, C. A., & Disney, M. D. (2021). A small molecule that binds an RNA repeat expansion stimulates its decay via the exosome complex. *Cell Chemical Biology*, 28(1), 34–45.e6. <https://doi.org/10.1016/j.chembiol.2020.10.007>
- Angelbello, A. J., Chen, J. L., & Disney, M. D. (2020). Small molecule targeting of RNA structures in neurological disorders. *Annals of the New York Academy of Sciences*, 1471(1), 57–71. <https://doi.org/10.1111/nyas.14051>
- Angelbello, A. J., Rzuczek, S. G., Mckee, K. K., Chen, J. L., Olafson, H., Cameron, M. D., Moss, W. N., Wang, E. T., & Disney, M. D. (2019). Precise small-molecule cleavage of an r(CUG) repeat expansion in a myotonic dystrophy mouse model. *Proceedings of the National Academy of Sciences of the United States of America*, 116(16), 7799–7804. <https://doi.org/10.1073/pnas.1901484116>
- Ariza, J., Rogers, H., Monterrubio, A., Reyes-Miranda, A., Hagerman, P. J., & Martínez-Cerdeño, V. (2016). A majority of FXTAS cases present with Intranuclear inclusions within Purkinje cells. *Cerebellum*, 15(5), 546–551. <https://doi.org/10.1007/s12311-016-0776-y>
- Asamitsu, S., Yabuki, Y., Ikenoshita, S., Kawakubo, K., Kawasaki, M., Usuki, S., Nakayama, Y., Adachi, K., Kugoh, H., Ishii, K., Matsuura, T., Nanba, E., Sugiyama, H., Fukunaga, K., & Shioda, N. (2021). CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Science Advances*, 7(3), eabd9440. <https://doi.org/10.1126/sciadv.abd9440>
- Ash, P. E. A., Bieniek, K. F., Gendron, T. F., Caulfield, T., Lin, W. L., DeJesus-Hernandez, M., Van Blitterswijk, M. M., Jansen-West, K., Paul, J. W., Rademakers, R., Boylan, K. B., Dickson, D. W., & Petrucelli, L. (2013). Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, 77(4), 639–646. <https://doi.org/10.1016/J.NEURON.2013.02.004>

- Ashizawa, T., Dubel, J. R., Dunne, P. W., Fu, Y. H., Pizzuti, A., Caskey, C. T., Boerwinkle, E., Perryman, M. B., Epstein, H. F., & Hejtmancik, J. F. (1992). Anticipation in myotonic dystrophy: II. Complex relationships between clinical findings and structure of the gct repeat. *Neurology*, *42*(10), 1877–1883. <https://doi.org/10.1212/wnl.42.10.1877>
- Ayhan, F., Perez, B. A., Shorrock, H. K., Zu, T., Banez-Coronel, M., Reid, T., Furuya, H., Clark, H. B., Troncoso, J. C., Ross, C. A., Subramony, S., Ashizawa, T., Wang, E. T., Yachnis, A. T., & Ranum, L. P. (2018). SCA 8 RAN polySer protein preferentially accumulates in white matter regions and is regulated by eIF 3F. *The EMBO Journal*, *37*(19), 1–15. <https://doi.org/10.15252/embj.201899023>
- Bakthavachalu, B., Huelsmeier, J., Sudhakaran, I. P., Hillebrand, J., Singh, A., Petrauskas, A., Thiagarajan, D., Sankaranarayanan, M., Mizoue, L., Anderson, E. N., Pandey, U. B., Ross, E., VijayRaghavan, K., Parker, R., & Ramaswami, M. (2018). RNP-granule assembly via Ataxin-2 disordered domains is required for long-term memory and neurodegeneration. *Neuron*, *98*(4), 754–766.e4. <https://doi.org/10.1016/J.NEURON.2018.04.032>
- Banani, S. F., Lee, H. O., Hyman, A. A., & Rosen, M. K. (2017). Biomolecular condensates: Organizers of cellular biochemistry. *Nature Reviews Molecular Cell Biology*, *18*(5), 285–298. <https://doi.org/10.1038/nrm.2017.7>
- Bañez-Coronel, M., Ayhan, F., Tarabochia, A. D., Zu, T., Perez, B. A., Tusi, S. K., Pletnikova, O., Borchelt, D. R., Ross, C. A., Margolis, R. L., Yachnis, A. T., Troncoso, J. C., & Ranum, L. P. W. (2015). RAN translation in Huntington disease. *Neuron*, *88*(4), 667–677. <https://doi.org/10.1016/j.neuron.2015.10.038>
- Banez-Coronel, M., & Ranum, L. P. W. (2019). Repeat-associated non-AUG (RAN) translation: Insights from pathology. *Laboratory Investigation*, *99*(7), 929–942. <https://doi.org/10.1038/s41374-019-0241-x>
- Bargiela, A., Cerro-Herreros, E., Fernandez-Costa, J. M., Vilchez, J. J., Llamusi, B., & Artero, R. (2015). Increased autophagy and apoptosis contribute to muscle atrophy in a myotonic dystrophy type 1 drosophila model. *DMM Disease Models and Mechanisms*, *8*(7), 679–690. <https://doi.org/10.1242/dmm.018127>
- Bargiela, A., Sabater-Arcis, M., Espinosa-Espinosa, J., Zulaica, M., De Munain, A. L., & Artero, R. (2019). Increased Muscblind levels by chloroquine treatment improve myotonic dystrophy type 1 phenotypes in vitro and in vivo models. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(50), 25203–25213. <https://doi.org/10.1073/pnas.1820297116>
- Barmada, S. J., Skibinski, G., Korb, E., Rao, E. J., Wu, J. Y., & Finkbeiner, S. (2010). Cytoplasmic mislocalization of TDP-43 is toxic to neurons and enhanced by a mutation associated with familial amyotrophic lateral sclerosis. *Journal of Neuroscience*, *30*(2), 639–649. <https://doi.org/10.1523/JNEUROSCI.4988-09.2010>
- Bassez, G., Audureau, E., Hogrel, J. Y., Arrouasse, R., Baghdoyan, S., Bhugaloo, H., Gourlay-Chu, M. L., Le Corvoisier, P., & Peschanski, M. (2018). Improved mobility with metformin in patients with myotonic dystrophy type 1: A randomized controlled trial. *Brain*, *141*(10), 2855–2865. <https://doi.org/10.1093/brain/awy231>
- Belotserkovskii, B. P., Shin, J. H. S., & Hanawalt, P. C. (2017). Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Research*, *45*(11), 6589–6599. <https://doi.org/10.1093/nar/gkx403>
- Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A., & Ponstingl, H. (1994). RanGAP1 induces GTPase activity of nuclear Ras-related ran. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(7), 2587–2591. <https://doi.org/10.1073/PNAS.91.7.2587>
- Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., Guharoy, M., De Decker, M., Jaspers, T., Ryan, V. H., Janke, A. M., Baatsen, P., Vercruyse, T., Kolaitis, R. M., Daelemans, D., Taylor, J. P., Kedersha, N., Anderson, P., Impens, F., ... Van Den Bosch, L. (2017). Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Molecular Cell*, *65*(6), 1044–1055.e5. <https://doi.org/10.1016/j.molcel.2017.02.013>
- Boeynaems, S., Bogaert, E., Michiels, E., Gijssels, I., Sieben, A., Jovičić, A., De Baets, G., Scheveneels, W., Steyaert, J., Cuijt, I., Verstrepen, K. J., Callaerts, P., Rousseau, F., Schymkowitz, J., Cruts, M., Van Broeckhoven, C., Van Damme, P., Gitler, A. D., Robberecht, W., & Van Den Bosch, L. (2016). Drosophila screen connects nuclear transport genes to DPR pathology in c9ALS/FTD. *Scientific Reports*, *6*, 20877. <https://doi.org/10.1038/srep20877>
- Boivin, M., Pfister, V., Gaucherot, A., Ruffenach, F., Negroni, L., Sellier, C., & Charlet-Berguerand, N. (2020). Reduced autophagy upon C9ORF72 loss synergizes with dipeptide repeat protein toxicity in G4C2 repeat expansion disorders. *The EMBO Journal*, *39*(4), e100574. <https://doi.org/10.15252/embj.2018100574>
- Bolognesi, B., Faure, A. J., Seuma, M., Schmiedel, J. M., Tartaglia, G. G., & Lehner, B. (2019). The mutational landscape of a prion-like domain. *Nature Communications*, *10*(1), 1–12. <https://doi.org/10.1038/s41467-019-12101-z>
- Bonapace, G., Gullace, R., Concolino, D., Iannello, G., Procopio, R., Gagliardi, M., Arabia, G., Barbagallo, G., Lupo, A., Manfredini, L. I., Annesi, G., & Quattrone, A. (2019). Intracellular FMRpolyG-Hsp70 complex in fibroblast cells from a patient affected by fragile X tremor ataxia syndrome. *Heliyon*, *5*(6), e01954. <https://doi.org/10.1016/j.heliyon.2019.e01954>
- Braida, C., Stefanatos, R. K. A., Adam, B., Mahajan, N., Smeets, H. J. M., Niel, F., Goizet, C., Arveiler, B., Koenig, M., Lagier-Tourenne, C., Mandel, J. L., Faber, C. G., de Die-Smulders, C. E. M., Spaans, F., & Monckton, D. G. (2010). Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. *Human Molecular Genetics*, *19*(8), 1399–1412. <https://doi.org/10.1093/hmg/ddq015>
- Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H., Hunter, K., Stanton, V. P., Thirion, J. P., Hudson, T., Sohn, R., Zemelmann, B., Snell, R. G., Rundle, S. A., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., ... Housman, D. E. (1992). Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, *68*(4), 799–808. [https://doi.org/10.1016/0092-8674\(92\)90154-5](https://doi.org/10.1016/0092-8674(92)90154-5)

- Brouwer, J. R., Huguët, A., Nicole, A., Munnich, A., & Gourdon, G. (2013). Transcriptionally repressive chromatin remodelling and CpG methylation in the presence of expanded CTG-repeats at the DM1 locus. *Journal of Nucleic Acids*, 2013, 1–16. <https://doi.org/10.1155/2013/567435>
- Buijsen, R. A. M., Sellier, C., Severijnen, L. A. W. F. M., Oulad-Abdelghani, M., Verhagen, R. F. M., Berman, R. F., Charlet-Berguerand, N., Willemsen, R., & Hukema, R. K. (2014). FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathologica Communications*, 2(1), 1–5. <https://doi.org/10.1186/s40478-014-0162-2>
- Burguete, A. S., Almeida, S., Gao, F. B., Kalb, R., Akins, M. R., & Bonini, N. M. (2015). GGGGCC microsatellite RNA is neuritically localized, induces branching defects, and perturbs transport granule function. *eLife*, 4, e08881. <https://doi.org/10.7554/eLife.08881>
- Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius F., Duclos, F., Monticelli, A., Zara, F., Cañizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, A., De Frutos, R., Palau, F., Patel, P. I., Di Donato, S., Mandel, J. L., Coccozza, S., Koenig, M., & Pandolfo, M. (1996). Friedreich's ataxia: Autosomal recessive disease caused by an Intronic GAA triplet repeat expansion. *Science*, 271(5254), 1423–1427. <https://doi.org/10.1126/science.271.5254.1423>
- Cardani, R., Mancinelli, E., Saino, G., Bonavina, L., & Meola, G. (2008). A putative role of ribonuclear inclusions and MBNL1 in the impairment of gallbladder smooth muscle contractility with cholelithiasis in myotonic dystrophy type 1. *Neuromuscular Disorders*, 18(8), 641–645. <https://doi.org/10.1016/j.nmd.2008.06.366>
- Castel, A. L., Cleary, J. D., & Pearson, C. E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nature Reviews Molecular Cell Biology*, 11(3), 165–170. <https://doi.org/10.1038/nrm2854>
- Cate, J. H. D. (2017). Human eIF3: From 'blobology' to biological insight. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1716), 20160176. <https://doi.org/10.1098/rstb.2016.0176>
- Celona, B., Von Dollen, J., Vatsavayai, S. C., Kashima, R., Johnson, J. R., Tang, A. A., Hata, A., Miller, B. L., Huang, E. J., Krogan, N. J., Seeley, W. W., & Black, B. L. (2017). Suppression of c9orf72 RNA repeat-induced neurotoxicity by the ALS-associated RNA-binding protein Zfp106. *eLife*, 6, 1–17. <https://doi.org/10.7554/eLife.19032>
- Cerro-Herreros, E., González-Martínez, I., Moreno-Cervera, N., Overby, S., Pérez-Alonso, M., Llamusi, B., & Artero, R. (2020). Therapeutic potential of AntagomiR-23b for treating myotonic dystrophy. *Molecular Therapy--Nucleic Acids*, 21, 837–849. <https://doi.org/10.1016/j.omtn.2020.07.021>
- Cerro-Herreros, E., Sabater-Arcis, M., Fernandez-Costa, J. M., Moreno, N., Perez-Alonso, M., Llamusi, B., & Artero, R. (2018). MiR-23b and miR-218 silencing increase Muscleblind-like expression and alleviate myotonic dystrophy phenotypes in mammalian models. *Nature Communications*, 9(1), 2482. <https://doi.org/10.1038/s41467-018-04892-4>
- Česnik, A. B., Darovic, S., Mihevc, S. P., Štalekar, M., Malnar, M., Motaln, H., Lee, Y. B., Mazej, J., Pohleven, J., Grosch, M., Modic, M., Fonovič, M., Turk, B., Drukker, M., Shaw, C. E., & Rogelj, B. (2019). Nuclear RNA foci from C9ORF72 expansion mutation form paraspeckle-like bodies. *Journal of Cell Science*, 132(5), jcs224303. <https://doi.org/10.1242/jcs.224303>
- Chamberlain, C. M., & Ranum, L. P. W. (2012). Mouse model of muscleblind-like 1 overexpression: Skeletal muscle effects and therapeutic promise. *Human Molecular Genetics*, 21(21), 4645–4654. <https://doi.org/10.1093/hmg/dds306>
- Chen, G., Masuda, A., Konishi, H., Ohkawara, B., Ito, M., Kinoshita, M., Kiyama, H., Matsuura, T., & Ohno, K. (2016). Phenylbutazone induces expression of MBNL1 and suppresses formation of MBNL1-CUG RNA foci in a mouse model of myotonic dystrophy. *Scientific Reports*, 6, 25317. <https://doi.org/10.1038/srep25317>
- Chen, H., Jacobs, E., Schwarzschild, M. A., McCullough, M. L., Calle, E. E., Thun, M. J., & Ascherio, A. (2005). Nonsteroidal antiinflammatory drug use and the risk for Parkinson's disease. *Annals of Neurology*, 58(6), 963–967. <https://doi.org/10.1002/ana.20682>
- Cheng, H. M., Chern, Y., Chen, I. H., Liu, C. R., Li, S. H., Chun, S. J., Rigo, F., Bennett, C. F., Deng, N., Feng, Y., Lin, C. S., Yan, Y. T., Cohen, S. N., & Cheng, T. H. (2015). Effects on murine behavior and lifespan of selectively decreasing expression of mutant huntingtin allele by Supt4h knockdown. *PLoS Genetics*, 11(3), 1–17. <https://doi.org/10.1371/journal.pgen.1005043>
- Cheng, W., Wang, S., Mestre, A. A., Fu, C., Makarem, A., Xian, F., Hayes, L. R., Lopez-Gonzalez, R., Drenner, K., Jiang, J., Cleveland, D. W., & Sun, S. (2018). C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2 α phosphorylation. *Nature Communications*, 9(1), 51. <https://doi.org/10.1038/s41467-017-02495-z>
- Cheng, W., Wang, S., Zhang, Z., Morgens, D. W., Hayes, L. R., Lee, S., Portz, B., Xie, Y., Nguyen, B. V., Haney, M. S., Yan, S., Dong, D., Coyne, A. N., Yang, J., Xian, F., Cleveland, D. W., Qiu, Z., Rothstein, J. D., Shorter, J., ... Sun, S. (2019). CRISPR-Cas9 screens identify the RNA helicase DDX3X as a repressor of C9ORF72 (GGGGCC)_n repeat-associated non-AUG translation. *Neuron*, 104(5), 885–898.e8. <https://doi.org/10.1016/j.neuron.2019.09.003>
- Chou, C. C., Zhang, Y., Umoh, M. E., Vaughan, S. W., Lorenzini, I., Liu, F., Sayegh, M., Donlin-Asp, P. G., Chen, Y. H., Duong, D. M., Seyfried, N. T., Powers, M. A., Kukar, T., Hales, C. M., Gearing, M., Cairns, N. J., Boylan, K. B., Dickson, D. W., Rademakers, R., ... Rossoll, W. (2018). TDP-43 pathology disrupts nuclear pore complexes and nucleocytoplasmic transport in ALS/FTD. *Nature Neuroscience*, 21(2), 228–239. <https://doi.org/10.1038/s41593-017-0047-3>
- Cid-Samper, F., Gelabert-Baldrich, M., Lang, B., Lorenzo-Gotor, N., Ponti, R. D., Severijnen, L. A. W. F. M., Bolognesi, B., Gelpi, E., Hukema, R. K., Botta-Orfila, T., & Tartaglia, G. G. (2018). An integrative study of protein-RNA condensates identifies scaffolding RNAs and reveals players in fragile X-associated tremor/ataxia syndrome. *Cell Reports*, 25(12), 3422–3434.e7. <https://doi.org/10.1016/j.celrep.2018.11.076>

- Colak, D., Zaninovic, N., Cohen, M. S., Rosenwaks, Z., Yang, W. Y., Gerhardt, J., Disney, M. D., & Jaffrey, S. R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science*, *343*(6174), 1002–1005. <https://doi.org/10.1126/science.1245831>
- Cook, C. N., Wu, Y., Odeh, H. M., Gendron, T. F., Jansen-West, K., del Rosso, G., Yue, M., Jiang, P., Gomes, E., Tong, J., Daugherty, L. M., Avendano, N. M., Castanedes-Casey, M., Shao, W., Oskarsson, B., Tomassy, G. S., McCampbell, A., Rigo, F., Dickson, D. W., ... Petrucelli, L. (2020). C9orf72 poly(GR) aggregation induces TDP-43 proteinopathy. *Science Translational Medicine*, *12*(559), 1–12. <https://doi.org/10.1126/SCITRANSLMED.ABB3774>
- Cooper-Knock, J., Higginbottom, A., Stopford, M. J., Highley, J. R., Ince, P. G., Wharton, S. B., Pickering-Brown, S., Kirby, J., Hautbergue, G. M., & Shaw, P. J. (2015). Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathologica*, *130*(1), 63–75. <https://doi.org/10.1007/s00401-015-1429-9>
- Cooper-Knock, J., Walsh, M. J., Higginbottom, A., Highley, J. R., Dickman, M. J., Edbauer, D., Ince, P. G., Wharton, S. B., Wilson, S. A., Kirby, J., Hautbergue, G. M., & Shaw, P. J. (2014). Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain*, *137*(7), 2040–2051. <https://doi.org/10.1093/brain/awu120>
- Coyne, A. N., Zaeffel, B. L., Hayes, L., Fitchman, B., Salzberg, Y., Luo, E. C., Bowen, K., Trost, H., Aigner, S., Rigo, F., Yeo, G. W., Harel, A., Svendsen, C. N., Sareen, D., & Rothstein, J. D. (2020). G4C2 repeat RNA initiates a POM121-mediated reduction in specific nucleoporins in C9orf72 ALS/FTD. *Neuron*, *107*(6), 1124–1140.e11. <https://doi.org/10.1016/j.neuron.2020.06.027>
- Crickard, J. B., Fu, J., & Reese, J. C. (2016). Biochemical analysis of yeast suppressor of ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *Journal of Biological Chemistry*, *291*(19), 9853–9870. <https://doi.org/10.1074/jbc.M116.716001>
- Crossley, M. P., Bocek, M., & Cimprich, K. A. (2019). R-loops as cellular regulators and genomic threats. *Molecular Cell*, *73*(3), 398–411. <https://doi.org/10.1016/j.molcel.2019.01.024>
- Crunkhorn, S. (2021). Small molecule targets toxic RNA repeats. *Nature reviews. Drug discovery*, *20*(1), 20. <https://doi.org/10.1038/d41573-020-00203-z>
- Cumming, S. A., Hamilton, M. J., Robb, Y., Gregory, H., McWilliam, C., Cooper, A., Adam, B., McGhie, J., Hamilton, G., Herzyk, P., Tschannen, M. R., Worthey, E., Petty, R., Ballantyne, B., Warner, J., Farrugia, M. E., Longman, C., & Monckton, D. G. (2018). De novo repeat interruptions are associated with reduced somatic instability and mild or absent clinical features in myotonic dystrophy type 1. *European Journal of Human Genetics*, *26*(11), 1635–1647. <https://doi.org/10.1038/s41431-018-0156-9>
- Davis, B. M., Mccurrach, M. E., Taneja, K. L., Singer, R. H., & Housman, D. E. (1997). Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(14), 7388–7393. <https://doi.org/10.1073/pnas.94.14.7388>
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. C. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G. Y. R., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., ... Rademakers, R. (2011). Expanded GGGGCC Hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron*, *72*(2), 245–256. <https://doi.org/10.1016/j.neuron.2011.09.011>
- Derbis, M., Kul, E., Niewiadomska, D., Sekrecki, M., Piasecka, A., Taylor, K., Hukema, R. K., Stork, O., & Sobczak, K. (2021). Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring expanded CGG repeats. *Nature Communications*, *12*(1), 1265. <https://doi.org/10.1038/s41467-021-21021-w>
- Diab, M. A., 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*(4), 1239–1252. <https://doi.org/10.1534/genetics.118.301672>
- Dmitriev, S. E., Terenin, I. M., Andreev, D. E., Ivanov, P. A., Dunaevsky, J. E., Merrick, W. C., & Shatsky, I. N. (2010). GTP-independent tRNA delivery to the ribosomal P-site by a novel eukaryotic translation factor. *Journal of Biological Chemistry*, *285*(35), 26779–26787. <https://doi.org/10.1074/jbc.M110.119693>
- Donnelly, C. J., Zhang, P. W., Pham, J. T., Heusler, A. R., Mistry, N. A., Vidensky, S., Daley, E. L., Poth, E. M., Hoover, B., Fines, D. M., Maragakis, N., Tienari, P. J., Petrucelli, L., Traynor, B. J., Wang, J., Rigo, F., Bennett, C. F., Blackshaw, S., Sattler, R., & Rothstein, J. D. (2013). RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron*, *80*(2), 415–428. <https://doi.org/10.1016/j.neuron.2013.10.015>
- Dy, A. B. C., Tassone, F., Eldeeb, M., Salcedo-Arellano, M. J., Tartaglia, N., & Hagerman, R. (2018). Metformin as targeted treatment in fragile X syndrome. *Clinical Genetics*, *93*(2), 216–222. <https://doi.org/10.1111/cge.13039>
- Fang, M. Y., Markmiller, S., Vu, A. Q., Javaherian, A., Dowdle, W. E., Jolivet, P., Bushway, P. J., Castello, N. A., Baral, A., Chan, M. Y., Linsley, J. W., Linsley, D., Mercola, M., Finkbeiner, S., Lecuyer, E., Lewcock, J. W., & Yeo, G. W. (2019). Small-molecule modulation of TDP-43 recruitment to stress granules prevents persistent TDP-43 accumulation in ALS/FTD. *Neuron*, *103*(5), 802–819.e11. <https://doi.org/10.1016/j.neuron.2019.05.048>
- Fardaei, M., Larkin, K., Brook, J. D., & Hamshere, M. G. (2001). In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Research*, *29*(13), 2766–2771. <https://doi.org/10.1093/nar/29.13.2766>
- Fardaei, M., Rogers, M. T., Thorpe, H. M., Larkin, K., Hamshere, M. G., Harper, P. S., & Brook, J. D. (2002). Three proteins, MBNL, MBL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Human Molecular Genetics*, *11*(7), 805–814. <https://doi.org/10.1093/hmg/11.7.805>

- Fay, M. M., Anderson, P. J., & Ivanov, P. (2017). ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. *Cell Reports*, 21(12), 3573–3584. <https://doi.org/10.1016/j.celrep.2017.11.093>
- Freibaum, B. D., Lu, Y., Lopez-Gonzalez, R., Kim, N. C., Almeida, S., Lee, K.-H., Badders, N., Valentine, M., Miller, B. L., Wong, P. C., Petrucelli, L., Kim, H. J., Gao, F.-B., & Taylor, J. P. (2015). GGGGCC repeat expansion in C9ORF72 compromises nucleocytoplasmic transport HHS public access. *Nature*, 525(7567), 129–133. <https://doi.org/10.1038/nature14974>
- Freibaum, B. D., & Taylor, J. P. (2017). The role of dipeptide repeats in C9ORF72-related ALS-FTD. *Frontiers in Molecular Neuroscience*, 10, 35. <https://doi.org/10.3389/FNMOL.2017.00035>
- Fuchs, G., Petrov, A. N., Marceau, C. D., Popov, L. M., Chen, J., OLeary, S. E., Wang, R., Carette, J. E., Sarnow, P., & Puglisi, J. D. (2015). Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proceedings of the National Academy of Sciences of the United States of America*, 112(2), 319–325. <https://doi.org/10.1073/pnas.1421328111>
- Fugier, C., Klein, A. F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., Tosch, V., Vignaud, A., Ferry, A., Messaddeq, N., Kokunai, Y., Tsuburaya, R., De La Grange, P., Dembele, D., Francois, V., Precigout, G., Boulade-Ladame, C., Hummel, M. C., De Munain, A. L., ... Charlet-Berguerand, N. (2011). Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nature Medicine*, 17(6), 720–725. <https://doi.org/10.1038/nm.2374>
- Furuta, N., Tsukagoshi, S., Hirayanagi, K., & Ikeda, Y. (2019). Suppression of the yeast elongation factor Spt4 ortholog reduces expanded SCA36 GGCCUG repeat aggregation and cytotoxicity. *Brain Research*, 1711(December 2018), 29–40. <https://doi.org/10.1016/j.brainres.2018.12.045>
- Gadgil, R., Barthelemy, J., Lewis, T., & Leffak, M. (2017). Replication stalling and DNA microsatellite instability. *Biophysical Chemistry*, 225, 38–48. <https://doi.org/10.1016/j.bpc.2016.11.007>
- Gasset-Rosa, F., Lu, S., Yu, H., Chen, C., Melamed, Z., Guo, L., Shorter, J., Da Cruz, S., & Cleveland, D. W. (2019). Cytoplasmic TDP-43 Demixing independent of stress granules drives inhibition of nuclear import, loss of nuclear TDP-43, and cell death. *Neuron*, 102(2), 339–357.e7. <https://doi.org/10.1016/j.neuron.2019.02.038>
- Gendron, T. F., Bieniek, K. F., Zhang, Y. J., Jansen-West, K., Ash, P. E. A., Caulfield, T., Daugherty, L., Dunmore, J. H., Castanedes-Casey, M., Chew, J., Cosio, D. M., Van Blitterswijk, M., Lee, W. C., Rademakers, R., Boylan, K. B., Dickson, D. W., & Petrucelli, L. (2013). Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathologica*, 126(6), 829–844. <https://doi.org/10.1007/s00401-013-1192-8>
- Goodkey, K., Aslesh, T., Maruyama, R., & Yokota, T. (2018). Nusinersen in the treatment of spinal muscular atrophy. *Methods in Molecular Biology*, 1828, 69–76. https://doi.org/10.1007/978-1-4939-8651-4_4
- Goodman, L. D., & Bonini, N. M. (2020). New roles for canonical transcription factors in repeat expansion diseases. *Trends in Genetics*, 36(2), 81–92. <https://doi.org/10.1016/j.tig.2019.11.003>
- Goodman, L. D., Prudencio, M., Kramer, N. J., Martinez-Ramirez, L. F., Srinivasan, A. R., Lan, M., Parisi, M. J., Zhu, Y., Chew, J., Cook, C. N., Berson, A., Gitler, A. D., Petrucelli, L., & Bonini, N. M. (2019). Toxic expanded GGGGCC repeat transcription is mediated by the PAF1 complex in C9orf72-associated FTD. *Nature Neuroscience*, 22(6), 863–874. <https://doi.org/10.1038/s41593-019-0396-1>
- Goodman, L. D., Prudencio, M., Srinivasan, A. R., Rifai, O. M., Lee, V. M. Y., Petrucelli, L., & Bonini, N. M. (2019). eIF4B and eIF4H mediate GR production from expanded G4C2 in a drosophila model for C9orf72-associated ALS. *Acta Neuropathologica Communications*, 7(1), 62. <https://doi.org/10.1186/s40478-019-0711-9>
- Görlich, D., Panté, N., Kutay, U., Aebi, U., & Bischoff, F. R. (1996). Identification of different roles for RanGDP and RanGTP in nuclear protein import. *The EMBO Journal*, 15(20), 5584. <https://doi.org/10.1002/j.1460-2075.1996.tb00943.x>
- Greco, C. M., Berman, R. F., Martin, R. M., Tassone, F., Schwartz, P. H., Chang, A., Trapp, B. D., Iwahashi, C., Brunberg, J., Grigsby, J., Hessel, D., Becker, E. J., Papazian, J., Leehey, M. A., Hagerman, R. J., & Hagerman, P. J. (2006). Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain*, 129(1), 243–255. <https://doi.org/10.1093/brain/awh683>
- Greco, C. M., Hagerman, R. J., Tassone, F., Chudley, A. E., Del Bigio, M. R., Jacquemont, S., Leehey, M., & Hagerman, P. J. (2002). Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, 125(8), 1760–1771. <https://doi.org/10.1093/brain/awf184>
- Green, K. M., Glineburg, M. R., Kearse, M. G., Flores, B. N., Linsalata, A. E., Fedak, S. J., Goldstrohm, A. C., Barmada, S. J., & Todd, P. K. (2017). RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nature Communications*, 8(1), 2005. <https://doi.org/10.1038/s41467-017-02200-0>
- Grima, J. C., Daigle, J. G., Arbez, N., Cunningham, K. C., Zhang, K., Ochaba, J., Geater, C., Morozko, E., Stocksdale, J., Glatzer, J. C., Pham, J. T., Ahmed, I., Peng, Q., Wadhwa, H., Pletnikova, O., Troncoso, J. C., Duan, W., Snyder, S. H., Ranum, L. P. W., ... Rothstein, J. D. (2017). Mutant huntingtin disrupts the nuclear pore complex. *Neuron*, 94(1), 93–107.e6. <https://doi.org/10.1016/j.neuron.2017.03.023>
- Groh, M., Lufino, M. M. P., Wade-Martins, R., & Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genetics*, 10(5), e1004318. <https://doi.org/10.1371/journal.pgen.1004318>
- Guenther, U. P., Weinberg, D. E., Zubradt, M. M., Tedeschi, F. A., Stawicki, B. N., Zagore, L. L., Brar, G. A., Licatalosi, D. D., Bartel, D. P., Weissman, J. S., & Jankowsky, E. (2018). The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature*, 559(7712), 130–134. <https://doi.org/10.1038/s41586-018-0258-0>
- Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A. J., Shaw, P. G., Kim, M. S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., & Wang, J. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, 507(7491), 195–200. <https://doi.org/10.1038/nature13124>

- Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., & Hagerman, P. J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*, *57*(1), 127–130. <https://doi.org/10.1212/WNL.57.1.127>
- Handa, V., Saha, T., & Usdin, K. (2003). The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. *Nucleic Acids Research*, *31*(21), 6243–6248. <https://doi.org/10.1093/nar/gkg818>
- Hashem, Y., Des Georges, A., Dhote, V., Langlois, R., Liao, H. Y., Grassucci, R. A., Pestova, T. V., Hellen, C. U. T., & Frank, J. (2013). Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature*, *503*(7477), 539–543. <https://doi.org/10.1038/nature12658>
- Hautbergue, G. M., Castelli, L. M., Ferraiuolo, L., Sanchez-Martinez, A., Cooper-Knock, J., Higginbottom, A., Lin, Y. H., Bauer, C. S., Dodd, J. E., Myszczyńska, M. A., Alam, S. M., Garneret, P., Chandran, J. S., Karyka, E., Stopford, M. J., Smith, E. F., Kirby, J., Meyer, K., Kaspar, B. K., ... Shaw, P. J. (2017). SRSF1-dependent nuclear export inhibition of C9orf72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nature Communications*, *8*(May), 1–18. <https://doi.org/10.1038/ncomms16063>
- Hayes, L. R., Duan, L., Bowen, K., Kalab, P., & Rothstein, J. D. (2020). C9orf72 arginine-rich dipeptide repeat proteins disrupt karyopherin-mediated nuclear import. *eLife*, *9*, e51685. <https://doi.org/10.7554/eLife.51685>
- He, F., Krans, A., Freibaum, B. D., Paul Taylor, J., & Todd, P. K. (2014). TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. *Human Molecular Genetics*, *23*(19), 5036–5051. <https://doi.org/10.1093/hmg/ddu216>
- Heitz, D., Devys, D., Imbert, G., Kretz, C., & Mandel, J. L. (1992). Inheritance of the fragile X syndrome: Size of the fragile X premutation is a major determinant of the transition to full mutation. *Journal of Medical Genetics*, *29*(11), 794–801. <https://doi.org/10.1136/jmg.29.11.794>
- Hertz, M. I., Landry, D. M., Willis, A. E., Luo, G., & Thompson, S. R. (2013). Ribosomal protein S25 dependency reveals a common mechanism for diverse internal ribosome entry sites and ribosome shunting. *Molecular and Cellular Biology*, *33*(5), 1016–1026. <https://doi.org/10.1128/mcb.00879-12>
- Hervás, D., Fornés-Ferrer, V., Gómez-Escribano, A. P., Sequedo, M. D., Peiró, C., Millán, J. M., & Vázquez-Manrique, R. P. (2017). Metformin intake associates with better cognitive function in patients with Huntington's disease. *PLoS One*, *12*(6), e0179283. <https://doi.org/10.1371/journal.pone.0179283>
- Hirohata, M., Ono, K., Naiki, H., & Yamada, M. (2005). Non-steroidal anti-inflammatory drugs have anti-amyloidogenic effects for Alzheimer's β -amyloid fibrils in vitro. *Neuropharmacology*, *49*(7), 1088–1099. <https://doi.org/10.1016/j.neuropharm.2005.07.004>
- Ho, T. H., Charlet-B, N., Poulos, M. G., Singh, G., Swanson, M. S., & Cooper, T. A. (2004). Muscleblind proteins regulate alternative splicing. *EMBO Journal*, *23*(15), 3103–3112. <https://doi.org/10.1038/sj.emboj.7600300>
- Hoem, G., Raske, C. R., Garcia-Arocena, D., Tassone, F., Sanchez, E., Ludwig, A. L., Iwahashi, C. K., Kumar, M., Yang, J. E., & Hagerman, P. J. (2011). CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS. *Human Molecular Genetics*, *20*(11), 2161–2170. <https://doi.org/10.1093/hmg/ddr101>
- Huang, K., Masuda, A., Chen, G., Bushra, S., Kamon, M., Araki, T., Kinoshita, M., Ohkawara, B., Ito, M., & Ohno, K. (2020). Inhibition of cyclooxygenase-1 by nonsteroidal anti-inflammatory drugs demethylates MeR2 enhancer and promotes Mbn1 transcription in myogenic cells. *Scientific Reports*, *10*(1), 2558. <https://doi.org/10.1038/s41598-020-59517-y>
- Huang, Y., Gattoni, R., Stévenin, J., & Steitz, J. A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Molecular Cell*, *11*(3), 837–843. [https://doi.org/10.1016/S1097-2765\(03\)00089-3](https://doi.org/10.1016/S1097-2765(03)00089-3)
- Hyman, A. A., Weber, C. A., & Jülicher, F. (2014). Liquid-liquid phase separation in biology. *Annual Review of Cell and Developmental Biology*, *30*, 39–58. <https://doi.org/10.1146/annurev-cellbio-100913-013325>
- Ivanov, I. P., Loughran, G., Sachs, M. S., & Atkins, J. F. (2010). Initiation context modulates autoregulation of eukaryotic translation initiation factor 1 (eIF1). *Proceedings of the National Academy of Sciences of the United States of America*, *107*(42), 18056–18060. <https://doi.org/10.1073/pnas.1009269107>
- Iwahashi, C. K., Yasui, D. H., An, H. J., Greco, C. M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C. B., Hagerman, R. J., & Hagerman, P. J. (2006). Protein composition of the intranuclear inclusions of FXTAS. *Brain*, *129*(1), 256–271. <https://doi.org/10.1093/brain/awh650>
- Jain, A., & Vale, R. D. (2017). RNA phase transitions in repeat expansion disorders. *Nature*, *546*(7657), 243–247. <https://doi.org/10.1038/nature22386>
- Jauvin, D., Chrétien, J., Pandey, S. K., Martineau, L., Revillod, L., Bassez, G., Lachon, A., McLeod, A. R., Gourdon, G., Wheeler, T. M., Thornton, C. A., Bennett, C. F., & Puymirat, J. (2017). Targeting DMPK with antisense oligonucleotide improves muscle strength in myotonic dystrophy type 1 mice. *Molecular Therapy—Nucleic Acids*, *7*, 465–474. <https://doi.org/10.1016/j.omtn.2017.05.007>
- Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K., & Krzyzosiak, W. J. (2016). Identifying proteins that bind to specific RNAs—Focus on simple repeat expansion diseases. *Nucleic Acids Research*, *44*(19), 9050–9070. <https://doi.org/10.1093/nar/gkw803>
- Jiang, H., Mankodi, A., Swanson, M. S., Moxley, R. T., & Thornton, C. A. (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Human Molecular Genetics*, *13*(24), 3079–3088. <https://doi.org/10.1093/hmg/ddh327>
- Jin, P., Duan, R., Qurashi, A., Qin, Y., Tian, D., Rosser, T. C., Liu, H., Feng, Y., & Warren, S. T. (2007). Pur α binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a *Drosophila* model of fragile X tremor/ataxia syndrome. *Neuron*, *55*(4), 556–564. <https://doi.org/10.1016/j.neuron.2007.07.020>
- Jin, P., Zarnescu, D. C., Zhang, F., Pearson, C. E., Lucchesi, J. C., Moses, K., & Warren, S. T. (2003). RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in drosophila. *Neuron*, *39*, 739–747.

- Jovičić, A., Mertens, J., Boeynaems, S., Bogaert, E., Chai, N., Yamada, S. B., Iii, J. W. P., Sun, S., Herdy, J. R., Bieri, G., Kramer, N. J., Gage, F. H., Robberecht, W., & Gitler, A. D. (2015). Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nature Neuroscience*, *18*(9), 1226–1229. <https://doi.org/10.1038/nn.4085>
- Kanadia, R. N., Shin, J., Yuan, Y., Beattie, S. G., Wheeler, T. M., Thornton, C. A., & Swanson, M. S. (2006). Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(31), 11748–11753. <https://doi.org/10.1073/pnas.0604970103>
- Kearse, M. G., Green, K. M., Krans, A., Rodriguez, C. M., Linsalata, A. E., Goldstrohm, A. C., & Todd, P. K. (2016). CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Molecular Cell*, *62*(2), 314–322. <https://doi.org/10.1016/j.molcel.2016.02.034>
- Khan, E., Mishra, S. K., Mishra, R., Mishra, A., & Kumar, A. (2019). Discovery of a potent small molecule inhibiting Huntington's disease (HD) pathogenesis via targeting CAG repeats RNA and poly Q protein. *Scientific Reports*, *9*(1), 16872. <https://doi.org/10.1038/s41598-019-53410-z>
- Kim, D. H., Langlois, M. A., Lee, K. B., Riggs, A. D., Puymirat, J., & Rossi, J. J. (2005). HnRNP H inhibits nuclear export of mRNA containing expanded CUG repeats and a distal branch point sequence. *Nucleic Acids Research*, *33*(12), 3866–3874. <https://doi.org/10.1093/nar/gki698>
- Kim, H. J., Raphael, A. R., Ladow, E. S., MCGurk, L., Weber, R. A., Trojanowski, J. Q., Lee, V. M. Y., Finkbeiner, S., Gitler, A. D., & Bonini, N. M. (2014). Therapeutic modulation of eIF2 α phosphorylation rescues TDP-43 toxicity in amyotrophic lateral sclerosis disease models. *Nature Genetics*, *46*(2), 152–160. <https://doi.org/10.1038/ng.2853>
- Kim, H. J., & Taylor, J. P. (2017). Lost in transportation: Nucleocytoplasmic transport defects in ALS and other neurodegenerative diseases. *Neuron*, *96*(2), 285–297. <https://doi.org/10.1016/j.neuron.2017.07.029>
- Kino, Y., Washizu, C., Oma, Y., Onishi, H., Nezu, Y., Sasagawa, N., Nukina, N., & Ishiura, S. (2009). MBNL and CELF proteins regulate alternative splicing of the skeletal muscle chloride channel CLCN1. *Nucleic Acids Research*, *37*(19), 6477–6490. <https://doi.org/10.1093/nar/gkp681>
- Klein, B. J., Bose, D., Baker, K. J., Yusoff, Z. M., Zhang, X., & Murakami, K. S. (2011). RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(2), 546–550. <https://doi.org/10.1073/pnas.1013828108>
- Komar, A. A., & Merrick, W. C. (2020). A retrospective on EIF2A—And not the alpha subunit of EIF2. *International Journal of Molecular Sciences*, *21*(6), 2054. <https://doi.org/10.3390/ijms21062054>
- Konieczny, P., Stepniak-Konieczna, E., & Sobczak, K. (2014). MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Research*, *42*(17), 10873–10887. <https://doi.org/10.1093/nar/gku767>
- Kozel, C., Thompson, B., Hustak, S., Moore, C., Nakashima, A., Singh, C. R., Reid, M., Cox, C., Papadopoulos, E., Luna, R. E., Anderson, A., Tagami, H., Hiraishi, H., Slone, E. A., Yoshino, K. I., Asano, M., Gillaspie, S., Nietfeld, J., Perchellet, J. P., ... Asano, K. (2016). Overexpression of eIF5 or its protein mimic 5MP perturbs eIF2 function and induces ATF4 translation through delayed re-initiation. *Nucleic Acids Research*, *44*(18), 8704–8713. <https://doi.org/10.1093/nar/gkw559>
- Kozłowski, P., de Mezer, M., & Krzyżosiak, W. J. (2010). Trinucleotide repeats in human genome and exome. *Nucleic Acids Research*, *38*(12), 4027–4039. <https://doi.org/10.1093/nar/gkq127>
- Kramer, N. J., Carlomagno, Y., Zhang, Y. J., Almeida, S., Cook, C. N., Gendron, T. F., Prudencio, M., Van Blitterswijk, M., Belzil, V., Couthouis, J., Paul, J. W., Goodman, L. D., Daugherty, L., Chew, J., Garrett, A., Pregent, L., Jansen-West, K., Tabassian, L. J., Rademakers, R., ... Gitler, A. D. (2016). Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. *Science*, *353*(6300), 708–712. <https://doi.org/10.1126/science.aaf7791>
- Krol, J., Fiszer, A., Mykowska, A., Sobczak, K., de Mezer, M., & Krzyżosiak, W. J. (2007). Ribonuclease Dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. *Molecular Cell*, *25*(4), 575–586. <https://doi.org/10.1016/j.molcel.2007.01.031>
- Kuyumcu-Martinez, N. M., Wang, G. S., & Cooper, T. A. (2007). Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Molecular Cell*, *28*(1), 68–78. <https://doi.org/10.1016/j.molcel.2007.07.027>
- Lagier-Tourenne, C., Baughn, M., Rigo, F., Sun, S., Liu, P., Li, H. R., Jiang, J., Watt, A. T., Chun, S., Katz, M., Qiu, J., Sun, Y., Ling, S. C., Zhu, Q., Polymenidou, M., Drenner, K., Artates, J. W., McAlonis-Downes, M., Markmiller, S., ... Ravits, J. (2013). Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(47), E4530–E4539. <https://doi.org/10.1073/pnas.1318835110>
- Landry, D. M., Hertz, M. I., & Thompson, S. R. (2009). RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes and Development*, *23*(23), 2753–2764. <https://doi.org/10.1101/gad.1832209>
- Lane, R. M., Smith, A., Baumann, T., Gleichmann, M., Norris, D., Bennett, C. F., & Kordasiewicz, H. (2018). Translating antisense technology into a treatment for Huntington's disease. *Methods in Molecular Biology*, *1780*, 497–523. https://doi.org/10.1007/978-1-4939-7825-0_23
- Langdon, E. M., Qiu, Y., Niaki, A. G., McLaughlin, G. A., Weidmann, C. A., Gerbich, T. M., Smith, J. A., Crutchley, J. M., Termini, C. M., Weeks, K. M., Myong, S., & Gladfelter, A. S. (2018). mRNA structure determines specificity of a polyQ-driven phase separation. *Science*, *360*(6391), 922–927. <https://doi.org/10.1126/science.aar7432>

- Lawlor, K. T., O'Keefe, L. V., Samaraweera, S. E., van Eyk, C. L., McLeod, C. J., Maloney, C. A., Dang, T. H. Y., Suter, C. M., & Richards, R. I. (2011). Double-stranded RNA is pathogenic in drosophila models of expanded repeat neurodegenerative diseases. *Human Molecular Genetics*, 20(19), 3757–3768. <https://doi.org/10.1093/hmg/ddr292>
- Lee, K. H., Zhang, P., Kim, H. J., Mitrea, D. M., Sarkar, M., Freibaum, B. D., Cika, J., Coughlin, M., Messing, J., Molliex, A., Maxwell, B. A., Kim, N. C., Temirov, J., Moore, J., Kolaitis, R. M., Shaw, T. I., Bai, B., Peng, J., Kriwacki, R. W., & Taylor, J. P. (2016). C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell*, 167(3), 774–788.e17. <https://doi.org/10.1016/j.cell.2016.10.002>
- Lee, Y. B., Chen, H. J., Peres, J. N., Gomez-Deza, J., Attig, J., Štalekar, M., Troakes, C., Nishimura, A. L., Scotter, E. L., Vance, C., Adachi, Y., Sardone, V., Miller, J. W., Smith, B. N., Gallo, J. M., Ule, J., Hirth, F., Rogelj, B., Houart, C., & Shaw, C. E. (2013). Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Reports*, 5(5), 1178–1186. <https://doi.org/10.1016/j.celrep.2013.10.049>
- Li, L., Matsui, M., & Corey, D. R. (2016). Activating frataxin expression by repeat-targeted nucleic acids. *Nature Communications*, 7, 10606. <https://doi.org/10.1038/ncomms10606>
- Lim, K. R. Q., Maruyama, R., & Yokota, T. (2017). Eteplirsen in the treatment of Duchenne muscular dystrophy. *Drug Design, Development and Therapy*, 11, 533–545. <https://doi.org/10.2147/DDDT.S97635>
- Linsalata, A. E., He, F., Malik, A. M., Glineburg, M. R., Green, K. M., Natla, S., Flores, B. N., Krans, A., Archbold, H. C., Fedak, S. J., Barmada, S. J., & Todd, P. K. (2019). DDX 3X and specific initiation factors modulate FMR 1 repeat-associated non-AUG-initiated translation. *EMBO Reports*, 20, 1–18. <https://doi.org/10.15252/embr.201847498>
- Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., Day, J. W., & Ranum, L. P. W. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron I of ZNF9. *Science*, 293(5531), 864–867. <https://doi.org/10.1126/science.1062125>
- Liu, C. R., Chang, C. R., Chern, Y., Wang, T. H., Hsieh, W. C., Shen, W. C., Chang, C. Y., Chu, I. C., Deng, N., Cohen, S. N., & Cheng, T. H. (2012). Spt4 is selectively required for transcription of extended trinucleotide repeats. *Cell*, 148(4), 690–701. <https://doi.org/10.1016/j.cell.2011.12.032>
- Liu, H., Lu, Y.-N., Paul, T., Periz, G., Banco, M. T., Ferré-D'Amaré, A. R., Rothstein, J. D., Hayes, L. R., Myong, S., & Wang, J. (2021). A helicase unwinds hexanucleotide repeat RNA G-quadruplexes and facilitates repeat-associated non-AUG translation. *Journal of the American Chemical Society*, 143(19), 7368–7379. <https://doi.org/10.1021/JACS.1C00131>
- Loomis, E. W., Sanz, L. A., Chédin, F., & Hagerman, P. J. (2014). Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genetics*, 10(4), e1004294. <https://doi.org/10.1371/journal.pgen.1004294>
- Loughran, G., Sachs, M. S., Atkins, J. F., & Ivanov, I. P. (2012). Stringency of start codon selection modulates autoregulation of translation initiation factor eIF5. *Nucleic Acids Research*, 40(7), 2898–2906. <https://doi.org/10.1093/nar/gkr1192>
- Lunde, B. M., Moore, C., & Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. *Nature Reviews Molecular Cell Biology*, 8(6), 479–490. <https://doi.org/10.1038/nrm2178>
- Ma, L., Herren, A. W., Espinal, G., Randol, J., McLaughlin, B., Martinez-Cerdeño, V., Pessah, I. N., Hagerman, R. J., & Hagerman, P. J. (2019). Composition of the Intranuclear inclusions of fragile X-associated tremor/ataxia syndrome. *Acta Neuropathologica Communications*, 7(1), 143. <https://doi.org/10.1186/s40478-019-0796-1>
- MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S. A., James, M., Groot, N., MacFarlane, H., Jenkins, B., Anderson, M. A., Wexler, N. S., Gusella, J. F., Bates, G. P., Baxendale, S., Hummerich, H., Kirby, S., ... Harper, P. S. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, 72(6), 971–983. [https://doi.org/10.1016/0092-8674\(93\)90585-E](https://doi.org/10.1016/0092-8674(93)90585-E)
- Maharana, S., Wang, J., Papadopoulos, D. K., Richter, D., Pozniakovskiy, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T. M., Jahnel, M., Marrone, L., Chang, Y. T., Sternecker, J., Tomancak, P., Hyman, A. A., & Alberti, S. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science*, 360(6391), 918–921. <https://doi.org/10.1126/science.aar7366>
- Malik, I., Kelley, C. P., Wang, E. T., & Todd, P. K. (2021). Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature Reviews Molecular Cell Biology*, 2021, 1–19. <https://doi.org/10.1038/s41580-021-00382-6>
- Malik, I., Tseng, Y., Wright, S. E., Zheng, K., Ramaiyer, P., Green, K. M., & Todd, P. K. (2021). SRSF protein kinase 1 modulates RAN translation and suppresses CGG repeat toxicity. *EMBO Molecular Medicine*, 13, e14163. <https://doi.org/10.15252/emmm.202114163>
- Mankodi, A. (2001). Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Human Molecular Genetics*, 10(19), 2165–2170. <https://doi.org/10.1093/hmg/10.19.2165>
- Mankodi, A., Lin, X., Blaxall, B. C., Swanson, M. S., & Thornton, C. A. (2005). Nuclear RNA foci in the heart in myotonic dystrophy. *Circulation Research*, 97(11), 1152–1155. <https://doi.org/10.1161/01.RES.0000193598.89753.e3>
- Mankodi, A., Logigian, E., Callahan, L., McClain, C., White, R., Henderson, D., Krym, M., & Thornton, C. A. (2000). Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, 289(5485), 1769–1772. <https://doi.org/10.1126/science.289.5485.1769>
- Martí, E. (2016). RNA toxicity induced by expanded CAG repeats in Huntington's disease. *Brain Pathology*, 26(6), 779–786. <https://doi.org/10.1111/bpa.12427>
- Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M., & Cramer, P. (2011). Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *EMBO Journal*, 30(7), 1302–1310. <https://doi.org/10.1038/emboj.2011.64>
- May, S., Hornburg, D., Schludi, M. H., Arzberger, T., Rentzsch, K., Schwenk, B. M., Grässer, F. A., Mori, K., Kremmer, E., Banzhaf-Strathmann, J., Mann, M., Meissner, F., & Edbauer, D. (2014). C9orf72 FTL/ALS-associated Gly-ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathologica*, 128(4), 485–503. <https://doi.org/10.1007/s00401-014-1329-4>

- McMurray, C. T. (2010). Mechanisms of trinucleotide repeat instability during human development. *Nature Reviews Genetics*, *11*(11), 786–799. <https://doi.org/10.1038/nrg2828>
- Meola, G., & Cardani, R. (2017). Myotonic dystrophy type 2 and modifier genes: An update on clinical and pathomolecular aspects. *Neurological Sciences*, *38*(4), 535–546. <https://doi.org/10.1007/s10072-016-2805-5>
- Miller, J. W., Urbinati, C. R., Teng-umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., & Swanson, M. S. (2000). Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy. *The EMBO Journal*, *19*(17), 4439–4448. <https://doi.org/10.1093/emboj/19.17.4439>
- Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S., & Spector, D. L. (1999). Purification and biochemical characterization of interchromatin granule clusters. *The EMBO Journal*, *18*(15), 4308–4320. <https://doi.org/10.1093/EMBOJ/18.15.4308>
- Mizielinska, S., Ridler, C. E., Balendra, R., Thoeng, A., Woodling, N. S., Grässer, F. A., Plagnol, V., Lashley, T., Partridge, L., & Isaacs, A. M. (2017). Bidirectional nucleolar dysfunction in C9orf72 frontotemporal lobar degeneration. *Acta Neuropathologica Communications*, *5*(1), 29. <https://doi.org/10.1186/s40478-017-0432-x>
- Mollasalehi, N., Francois-Moutal, L., Scott, D. D., Tello, J. A., Williams, H., Mahoney, B., Carlson, J. M., Dong, Y., Li, X., Miranda, V. G., Gokhale, V., Wang, W., Barmada, S. J., & Khanna, M. (2020). An allosteric modulator of RNA binding targeting the N-terminal domain of TDP-43 yields neuroprotective properties. *ACS Chemical Biology*, *15*(11), 2854–2859. <https://doi.org/10.1021/acscchembio.0c00494>
- Montavon, T. C., Baldaccini, M., Lefèvre, M., Girardi, E., Chane-Woon-Ming, B., Messmer, M., Hammann, P., Chicher, J., & Pfeffer, S. (2021). Human DICER helicase domain recruits PKR and modulates its antiviral activity. In *PLoS Pathogens*, *17*(5), e1009549. <https://doi.org/10.1371/journal.ppat.1009549>
- Mori, K., Lammich, S., Mackenzie, I. R. A., Forné, I., Zilow, S., Kretschmar, H., Edbauer, D., Janssens, J., Kleinberger, G., Cruts, M., Herms, J., Neumann, M., Van Broeckhoven, C., Arzberger, T., & Haass, C. (2013). HnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathologica*, *125*(3), 413–423. <https://doi.org/10.1007/s00401-013-1088-7>
- Mori, K., Weng, S.-M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretschmar, H. A., Cruts, M., Van Broeckhoven, C., Haass, C., & Edbauer, D. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science*, *339*(6125), 1335–1338. <https://doi.org/10.1126/SCIENCE.1232927>
- Müller-McNicoll, M., Botti, V., de Jesus Domingues, A. M., Brandl, H., Schwich, O. D., Steiner, M. C., Curk, T., Poser, I., Zarnack, K., & Neugebauer, K. M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. *Genes and Development*, *30*(5), 553–566. <https://doi.org/10.1101/gad.276477.115>
- Mykowska, A., Sobczak, K., Wojciechowska, M., Kozłowski, P., & Krzyzosiak, W. J. (2011). CAG repeats mimic CUG repeats in the misregulation of alternative splicing. *Nucleic Acids Research*, *39*(20), 8938–8951. <https://doi.org/10.1093/nar/gkr608>
- Nakamori, M., Sobczak, K., Puwanant, A., Eichinger, K., Pandya, S., Dekdebrun, J., Heatwole, R., Mcdermott, M. P., Chen, T., Cline, M., Tawil, R., Osborne, R. J., Wheeler, T. M., Iii, R. T. M., & Thornton, C. A. (2013). Splicing biomarkers of disease severity in myotonic dystrophy. *Annals of Neurology*, *74*(6), 862–872. <https://doi.org/10.1002/ana.23992>
- Nedelsky, N. B., & Taylor, J. P. (2019). Bridging biophysics and neurology: Aberrant phase transitions in neurodegenerative disease. *Nature Reviews Neurology*, *15*(5), 272–286. <https://doi.org/10.1038/s41582-019-0157-5>
- Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretschmar, H. A., Trojanowski, J. Q., & Lee, V. M. Y. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science*, *314*(5796), 130–133. <https://doi.org/10.1126/science.1134108>
- Nishiyama, T., Yamamoto, H., Uchiumi, T., & Nakashima, N. (2007). Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistoviral intergenic internal ribosome entry site. *Nucleic Acids Research*, *35*(5), 1514–1521. <https://doi.org/10.1093/nar/gkl1121>
- Nolin, S. L., Glicksman, A., Ersalesi, N., Dobkin, C., Brown, W. T., Cao, R., Blatt, E., Sah, S., Latham, G. J., & Hadd, A. G. (2015). Fragile X full mutation expansions are inhibited by one or more AGG interruptions in premutation carriers. *Genetics in Medicine*, *17*(5), 358–364. <https://doi.org/10.1038/gim.2014.106>
- Pakravan, D., Michiels, E., Bratek-Skicki, A., De Decker, M., Van Lindt, J., Alsteens, D., Derclaye, S., Van Damme, P., Schymkowitz, J., Rousseau, F., Tompa, P., & Van Den Bosch, L. (2021). Liquid–liquid phase separation enhances TDP-43 LCD aggregation but delays seeded aggregation. *Biomolecules*, *11*(4), 548. <https://doi.org/10.3390/BIOM11040548>
- Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakowski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., & Alberti, S. (2015). A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell*, *162*(5), 1066–1077. <https://doi.org/10.1016/J.CELL.2015.07.047>
- Paul, S., Dansithong, W., Jog, S. P., Holt, I., Mittal, S., Brook, J. D., Morris, G. E., Comai, L., & Reddy, S. (2011). Expanded CUG repeats dysregulate RNA splicing by altering the stoichiometry of the Muscleblind 1 complex. *Journal of Biological Chemistry*, *286*(44), 38427–38438. <https://doi.org/10.1074/jbc.M111.255224>
- Paul, S., Dansithong, W., Kim, D., Rossi, J., Webster, N. J. G., Comai, L., & Reddy, S. (2006). Interaction of muscleblind, CUG-BP1 and hnRNP H proteins in DM1-associated aberrant IR splicing. *EMBO Journal*, *25*(18), 4271–4283. <https://doi.org/10.1038/sj.emboj.7601296>
- Paulson, H. (2018). Repeat expansion diseases. *Handbook of Clinical Neurology*, *147*, 105. <https://doi.org/10.1016/B978-0-444-63233-3.00009-9>

- Perdoni, F., Malatesta, M., Cardani, R., Giagnacovo, M., Mancinelli, E., Meola, G., & Pellicciari, C. (2009). RNA/MBNL1-containing foci in myoblast nuclei from patients affected by myotonic dystrophy type 2: An immunocytochemical study. *European Journal of Histochemistry*, 53(3), 151–158. <https://doi.org/10.4081/ejh.2009.e18>
- Perez, B. A., Shorrock, H. K., Banez-Coronel, M., Zu, T., El Romano, L., Laboissonniere, L. A., Reid, T., Ikeda, Y., Reddy, K., Gomez, C. M., Bird, T., Ashizawa, T., Schut, L. J., Brusco, A., Berglund, J. A., Hasholt, L. F., Nielsen, J. E., Subramony, S., & Ranum, L. P. (2021). CCG•CGG interruptions in high-penetrance SCA8 families increase RAN translation and protein toxicity. *EMBO Molecular Medicine*, 13, e14095. <https://doi.org/10.15252/EMMM.202114095>
- Persichetti, F., Ambrose, C. M., Ge, P., McNeil, S. M., Srinidhi, J., Anderson, M. A., Jenkins, B., Barnes, G. T., Duyao, M. P., & Kanaley, L. (1995). Normal and expanded Huntington's disease gene alleles produce distinguishable proteins due to translation across the CAG repeat. *Molecular Medicine (Cambridge, MA)*, 1(4), 374–383. <https://doi.org/10.1007/bf03401575>
- Polymenidou, M. (2018). The RNA face of phase separation. *Science*, 360(6391), 859–860. <https://doi.org/10.1126/science.aat8028>
- Ramesh, N., Daley, E. L., Gleixner, A. M., Mann, J. R., Kour, S., Mawrie, D., Anderson, E. N., Kofler, J., Donnelly, C. J., Kiskinis, E., & Pandey, U. B. (2020). RNA dependent suppression of C9orf72 ALS/FTD associated neurodegeneration by Matrin-3. *Acta Neuropathologica Communications*, 8(1), 1–20. <https://doi.org/10.1186/s40478-020-01060-y>
- Ravel-Chapuis, A., Bélanger, G., Yadava, R. S., Mahadevan, M. S., DesGroseillers, L., Côté, J., & Jasmin, B. J. (2012). The RNA-binding protein Staufen1 is increased in DM1 skeletal muscle and promotes alternative pre-mRNA splicing. *Journal of Cell Biology*, 196(6), 699–712. <https://doi.org/10.1083/jcb.201108113>
- Reddy, K., Tam, M., Bowater, R. P., Barber, M., Tomlinson, M., Nichol Edamura, K., Wang, Y. H., & Pearson, C. E. (2011). Determinants of R-loop formation at convergent bidirectionally transcribed trinucleotide repeats. *Nucleic Acids Research*, 39(5), 1749–1762. <https://doi.org/10.1093/nar/gkq935>
- Rodriguez, C. M., Wright, S. E., Kearse, M. G., Haeflner, J. M., Flores, B. N., Liu, Y., Ifrim, M. F., Glineburg, M. R., Krans, A., Jafar-Nejad, P., Sutton, M. A., Bassell, G. J., Parent, J. M., Rigo, F., Barmada, S. J., & Todd, P. K. (2020). A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nature Neuroscience*, 23(3), 386–397. <https://doi.org/10.1038/s41593-020-0590-1>
- Rodriguez, S., Sahin, A., Schrank, B. R., Al-Lawati, H., Costantino, I., Benz, E., Fard, D., Albers, A. D., Cao, L., Gomez, A. C., Evans, K., Ratti, E., Cudkowicz, M., Frosch, M. P., Talkowski, M., Sorger, P. K., Hyman, B. T., & Albers, M. W. (2021). Genome-encoded cytoplasmic double-stranded RNAs, found in C9ORF72 ALS-FTD brain, propagate neuronal loss. *Science Translational Medicine*, 13(601), 1–14. <https://doi.org/10.1126/scitranslmed.aaz4699>
- Rohilla, K. J., & Gagnon, K. T. (2017). RNA biology of disease-associated microsatellite repeat expansions. *Acta Neuropathologica Communications*, 5(1), 63. <https://doi.org/10.1186/s40478-017-0468-y>
- Rossi, S., Serrano, A., Gerbino, V., Giorgi, A., Di Francesco, L., Nencini, M., Bozzo, F., Schininà, M. E., Bagni, C., Cestra, G., Carri, M. T., Achsel, T., & Cozzolino, M. (2015). Nuclear accumulation of mRNAs underlies G4C2-repeat-induced translational repression in a cellular model of C9orf72 ALS. *Journal of Cell Science*, 128(9), 1787–1799. <https://doi.org/10.1242/jcs.165332>
- Sabater-Arcis, M., Bargiela, A., Furling, D., & Artero, R. (2020). miR-7 restores phenotypes in myotonic dystrophy muscle cells by repressing Hyperactivated autophagy. *Molecular Therapy--Nucleic Acids*, 19, 278–292. <https://doi.org/10.1016/j.omtn.2019.11.012>
- Saberi, S., Stauffer, J. E., Jiang, J., Garcia, S. D., Taylor, A. E., Schulte, D., Ohkubo, T., Schloffman, C. L., Maldonado, M., Baughn, M., Rodriguez, M. J., Pizzo, D., Cleveland, D., & Ravits, J. (2018). Sense-encoded poly-GR dipeptide repeat proteins correlate to neurodegeneration and uniquely co-localize with TDP-43 in dendrites of repeat-expanded C9orf72 amyotrophic lateral sclerosis. *Acta Neuropathologica*, 135(3), 459–474. <https://doi.org/10.1007/s00401-017-1793-8>
- Salcedo-Arellano, M. J., Dufour, B., McLennan, Y., Martinez-Cerdeno, V., & Hagerman, R. (2020). Fragile X syndrome and associated disorders: Clinical aspects and pathology. *Neurobiology of Disease*, 136, 104740. <https://doi.org/10.1016/J.NBD.2020.104740>
- Samaraweera, S. E., O'Keefe, L. V., Price, G. R., Venter, D. J., & Richards, R. I. (2013). Distinct roles for toll and autophagy pathways in double-stranded RNA toxicity in a drosophila model of expanded repeat neurodegenerative diseases. *Human Molecular Genetics*, 22(14), 2811–2819. <https://doi.org/10.1093/hmg/ddt130>
- Sanchez-Rangel, E., & Inzucchi, S. E. (2017). Metformin: Clinical use in type 2 diabetes. *Diabetologia*, 60(9), 1586–1593. <https://doi.org/10.1007/s00125-017-4336-x>
- Sareen, D., O'Rourke, J. G., Meera, P., Muhammad, A. K. M. G., Grant, S., Simpkinson, M., Bell, S., Carmona, S., Ornelas, L., Sahabian, A., Gendron, T., Petrucelli, L., Baughn, M., Ravits, J., Harms, M. B., Rigo, F., Frank Bennett, C., Otis, T. S., Svendsen, C. N., & Baloh, R. H. (2013). Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Science Translational Medicine*, 5(208), 208ra149. <https://doi.org/10.1126/scitranslmed.3007529>
- Scoles, D. R., & Pulst, S. M. (2018). Oligonucleotide therapeutics in neurodegenerative diseases. *RNA Biology*, 15(6), 1–8. <https://doi.org/10.1080/15476286.2018.1454812>
- Sellier, C., Buijsen, R. A. M., He, F., Natla, S., Jung, L., Tropel, P., Gaucherot, A., Jacobs, H., Meziante, H., Vincent, A., Champy, M.-F., Sorg, T., Pavlovic, G., Wattenhofer-Donze, M., Birling, M.-C., Oulad-Abdelghani, M., Eberling, P., Ruffenach, F., Joint, M., ... Charlet-Berguerand, N. (2017). Translation of expanded CGG repeats into FMRpolyG is pathogenic and may contribute to fragile X tremor ataxia syndrome. *Neuron*, 93(2), 331–347. <https://doi.org/10.1016/j.neuron.2016.12.016>
- Sellier, C., Cerro-herrerros, E., Blatter, M., Freyermuth, F., Gaucherot, A., Ruffenach, F., Sarkar, P., Puymirat, J., Udd, B., Day, J. W., Meola, G., Bassez, G., Fujimura, H., Takahashi, M. P., Schoser, B., & Furling, D. (2018). rFOX1/MBNL1 competition for CCUG RNA repeats binding contributes to myotonic dystrophy type 1/type 2 differences. *Nature Communications*, 9, 1–15. <https://doi.org/10.1038/s41467-018-04370-x>

- Sellier, C., Freyermuth, F., Tabet, R., Tran, T., He, F., Ruffenach, F., Alunni, V., Moine, H., Thibault, C., Page, A., Tassone, F., Willemsen, R., Disney, M. D., Hagerman, P. J., Todd, P. K., & Charlet-Berguerand, N. (2013). Sequestration of DROSHA and DGCR8 by expanded CGG RNA repeats alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Reports*, 3(3), 869–880. <https://doi.org/10.1016/j.celrep.2013.02.004>
- Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R. K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D. J., Hagerman, P. J., & Charlet-Berguerand, N. (2010). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO Journal*, 29(7), 1248–1261. <https://doi.org/10.1038/emboj.2010.21>
- Shi, K. Y., Mori, E., Nizami, Z. F., Lin, Y., Kato, M., Xiang, S., Wu, L. C., Ding, M., Yu, Y., Gall, J. G., & McKnight, S. L. (2017). Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proceedings of the National Academy of Sciences of the United States of America*, 114(7), E1111–E1117. <https://doi.org/10.1073/pnas.1620293114>
- Shi, Y., Yang, Y., Hoang, B., Bardeleben, C., Holmes, B., Gera, J., & Lichtenstein, A. (2016). Therapeutic potential of targeting IRES-dependent c-myc translation in multiple myeloma cells during ER stress. *Oncogene*, 35(8), 1015–1024. <https://doi.org/10.1038/nc.2015.156>
- Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. *Science*, 357(6357), eaaf4382. <https://doi.org/10.1126/science.aaf4382>
- Singh, C. R., Glineburg, M. R., Moore, C., Tani, N., Jaiswal, R., Zou, Y., Aube, E., Gillaspie, S., Thornton, M., Cecil, A., Hilgers, M., Takasu, A., Asano, I., Asano, M., Escalante, C. R., Nakamura, A., Todd, P. K., & Asano, K. (2021). Human oncoprotein 5MP suppresses general and repeat-associated non-AUG translation via eIF3 by a common mechanism. *Cell Reports*, 36(2), 109376. <https://doi.org/10.1016/j.celrep.2021.109376>
- Sobczak, K., Michlewski, G., De Mezer, M., Kierzek, E., Krol, J., Olejniczak, M., Kierzek, R., & Krzyzosiak, W. J. (2010). Structural diversity of triplet repeat RNAs. *Journal of Biological Chemistry*, 285(17), 12755–12764. <https://doi.org/10.1074/jbc.M109.078790>
- Sofola, O. A., Jin, P., Qin, Y., Duan, R., Liu, H., de Haro, M., Nelson, D. L., & Botas, J. (2007). RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG Premutation repeat-induced neurodegeneration in a drosophila model of FXTAS. *Neuron*, 55(4), 565–571. <https://doi.org/10.1016/j.neuron.2007.07.021>
- Sonobe, Y., Aburas, J., Krishnan, G., Fleming, A. C., Ghadge, G., Islam, P., Warren, E. C., Gu, Y., Kankel, M. W., Brown, A. E. X., Kiskinis, E., Gendron, T. F., Gao, F.-B., Roos, R. P., & Kratsios, P. (2021). A C. elegans model of C9orf72-associated ALS/FTD uncovers a conserved role for eIF2D in RAN translation. *Nature Communications*, 12(1), 6025. <https://doi.org/10.1038/s41467-021-26303-x>
- Sonobe, Y., Ghadge, G., Masaki, K., Sendoel, A., Fuchs, E., & Roos, R. P. (2018). Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiology of Disease*, 116, 155–165. <https://doi.org/10.1016/j.nbd.2018.05.009>
- Sun, X., Li, P. P., Zhu, S., Cohen, R., Marque, L. O., Ross, C. A., Pulst, S. M., Chan, H. Y. E., Margolis, R. L., & Rudnicki, D. D. (2015). Nuclear retention of full-length HTT RNA is mediated by splicing factors MBNL1 and U2AF65. *Scientific Reports*, 5, 12521. <https://doi.org/10.1038/srep12521>
- Swami, M., Hendricks, A. E., Gillis, T., Massood, T., Mysore, J., Myers, R. H., & Wheeler, V. C. (2009). Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Human Molecular Genetics*, 18(16), 3039–3047. <https://doi.org/10.1093/hmg/ddp242>
- Swinnen, B., Bento-Abreu, A., Gendron, T. F., Boeynaems, S., Bogaert, E., Nuyts, R., Timmers, M., Scheveneels, W., Hersmus, N., Wang, J., Mizielinska, S., Isaacs, A. M., Petrucelli, L., Lemmens, R., Van Damme, P., Van Den Bosch, L., & Robberecht, W. (2018). A zebrafish model for C9orf72 ALS reveals RNA toxicity as a pathogenic mechanism. *Acta Neuropathologica*, 135(3), 427–443. <https://doi.org/10.1007/s00401-017-1796-5>
- Sznajder, L. J., Michalak, M., Taylor, K., Cywoniuk, P., Kabza, M., Wojtkowiak-Szlachcic, A., Matłoka, M., Konieczny, P., & Sobczak, K. (2016). Mechanistic determinants of MBNL activity. *Nucleic Acids Research*, 44(21), 10326–10342. <https://doi.org/10.1093/nar/gkw915>
- Sznajder, L. J., Thomas, J. D., Carrell, E. M., Reid, T., McFarland, K. N., Cleary, J. D., Oliveira, R., Nutter, C. A., Bhatt, K., Sobczak, K., Ashizawa, T., Thornton, C. A., Ranum, L. P. W., & Swanson, M. S. (2018). Intron retention induced by microsatellite expansions as a disease biomarker. *Proceedings of the National Academy of Sciences of the United States of America*, 115(16), 4234–4239. <https://doi.org/10.1073/pnas.1716617115>
- Tabet, R., Schaeffer, L., Freyermuth, F., Jambeau, M., Workman, M., Lee, C. Z., Lin, C. C., Jiang, J., Jansen-West, K., Abou-Hamdan, H., Désaubry, L., Gendron, T., Petrucelli, L., Martin, F., & Lagier-Tourenne, C. (2018). CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nature Communications*, 9(1), 1–14. <https://doi.org/10.1038/s41467-017-02643-5>
- Tabrizi, S. J., Leavitt, B. R., Landwehrmeyer, G. B., Wild, E. J., Saft, C., Barker, R. A., Blair, N. F., Craufurd, D., Priller, J., Rickards, H., Rosser, A., Kordasiewicz, H. B., Czech, C., Swayze, E. E., Norris, D. A., Baumann, T., Gerlach, I., Schobel, S. A., Paz, E., ... Lane, R. M. (2019). Targeting huntingtin expression in patients with Huntington's disease. *New England Journal of Medicine*, 380(24), 2307–2316. <https://doi.org/10.1056/nejmoa1900907>
- Taneja, K. L., McCurrach, M., Schalling, M., Housman, D., & Singer, R. H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *Journal of Cell Biology*, 128(6), 995–1002. <https://doi.org/10.1083/jcb.128.6.995>
- Tang, L., Morris, J., Wan, J., Moore, C., Fujita, Y., Gillaspie, S., Aube, E., Nanda, J., Marques, M., Jangal, M., Anderson, A., Cox, C., Hiraishi, H., Dong, L., Saito, H., Singh, C. R., Witcher, M., Topisirovic, I., Qian, S. B., & Asano, K. (2017). Competition between translation initiation factor eIF5 and its mimic protein 5MP determines non-AUG initiation rate genome-wide. *Nucleic Acids Research*, 45(20), 11941–11953. <https://doi.org/10.1093/nar/gkx808>

- Tassone, F., Beilina, A., Carosi, C., Albertosi, S., Bagni, C., Li, L., Glover, K., Bentley, D., & Hagerman, P. J. (2007). Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA*, 13(4), 555–562. <https://doi.org/10.1261/rna.280807>
- Tassone, F., Hagerman, R. J., Loesch, D. Z., Lachiewicz, A., Taylor, A. K., & Hagerman, P. J. (2000). Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *American Journal of Medical Genetics*, 94(3), 232–236. [https://doi.org/10.1002/1096-8628\(20000918\)94:3<232::AID-AJMG9>3.0.CO;2-H](https://doi.org/10.1002/1096-8628(20000918)94:3<232::AID-AJMG9>3.0.CO;2-H)
- Taylor, K., & Sobczak, K. (2020). Intrinsic regulatory role of RNA structural arrangement in alternative splicing control. *International Journal of Molecular Sciences*, 21(14), 1–35. <https://doi.org/10.3390/ijms21145161>
- Thomas, M., White, R. L., & Davis, R. W. (1976). Hybridization of RNA to double stranded DNA: Formation of R loops. *Proceedings of the National Academy of Sciences of the United States of America*, 73(7), 2294–2298. <https://doi.org/10.1073/pnas.73.7.2294>
- Tian, B., White, R. J., Xia, T., Welle, S., Turner, D. H., Mathews, M. B., & Thornton, C. A. (2000). Expanded CUG repeat RNAs form hairpins that activate the double-stranded RNA-dependent protein kinase PKR. *RNA*, 6(1), 79–87. <https://doi.org/10.1017/S1355838200991544>
- Timchenko, N. A., Cai, Z. J., Welm, A. L., Reddy, S., Ashizawa, T., & Timchenko, L. T. (2001). RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *Journal of Biological Chemistry*, 276(11), 7820–7826. <https://doi.org/10.1074/jbc.M005960200>
- Todd, P. K., Oh, S. Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A. J., Chen, K., Chun, Scaglione, K. M., Basur, V., Elenitoba-Johnson, K., Vonsattel, J. P., Louis, E. D., Sutton, M. A., Taylor, J. P., Mills, R. E., Charlet-Bergerand, N., & Paulson, H. L. (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*, 78(3), 440–455. <https://doi.org/10.1016/j.neuron.2013.03.026>
- Trang, H., Stanley, S. Y., Thorner, P., Faghfoury, H., Schulze, A., Hawkins, C., Pearson, C. E., & Yoon, G. (2015). Massive CAG repeat expansion and somatic instability in maternally transmitted infantile spinocerebellar ataxia type 7. *JAMA Neurology*, 72(2), 219–223. <https://doi.org/10.1001/jamaneurol.2014.1902>
- Trottier, Y., Biancalana, V., & Mandel, J. L. (1994). Instability of CAG repeats in Huntington's disease: Relation to parental transmission and age of onset. *Journal of Medical Genetics*, 31(5), 377–382. <https://doi.org/10.1136/jmg.31.5.377>
- Tseng, Y.-J., Sandwith, S. N., Green, K. M., Chambers, A. E., Krans, A., Raimer, H. M., Sharlow, M. E., Reisinger, M. A., Richardson, A. E., Routh, E. D., Smaldino, M. A., Wang, Y.-H., Vaughn, J. P., Todd, P. K., & Smaldino, P. J. (2021). The RNA helicase DHX36/G4R1 modulates C9orf72 GGGGCC hexanucleotide repeat-associated translation. *Journal of Biological Chemistry*, 297(2), 100914. <https://doi.org/10.1016/j.jbc.2021.100914>
- Tsoi, H., Lau, C. K., Lau, K. F., & Chan, H. Y. E. (2011). Perturbation of U2AF65/NXF1-mediated RNA nuclear export enhances RNA toxicity in polyQ diseases. *Human Molecular Genetics*, 20(19), 3787–3797. <https://doi.org/10.1093/hmg/ddr297>
- Tusi, S. K., Nguyen, L., Thangaraju, K., Li, J., Cleary, J. D., Zu, T., & Ranum, L. P. W. (2021). The alternative initiation factor eIF2A plays key role in RAN translation of myotonic dystrophy type 2 CCUG•CAGG repeats. *Human Molecular Genetics*, 30(11), 1020–1029. <https://doi.org/10.1093/hmg/ddab098>
- Uaesoontrachoon, K., Quinn, J. L., Tatem, K. S., Van Der Meulen, J. H., Yu, Q., Phadke, A., Miller, B. K., Gordish-Dressman, H., Ongini, E., Miglietta, D., & Nagaraju, K. (2014). Long-term treatment with naproxen significantly improves skeletal and cardiac disease phenotype in the mdx mouse model of dystrophy. *Human Molecular Genetics*, 23(12), 3239–3249. <https://doi.org/10.1093/hmg/ddu033>
- van Eyk, C. L., Samaraweera, S. E., Scott, A., Webber, D. L., Harvey, D. P., Mecinger, O., O'Keefe, L. V., Cropley, J. E., Young, P., Ho, J., Suter, C., & Richards, R. I. (2019). Non-self mutation: Double-stranded RNA elicits antiviral pathogenic response in a drosophila model of expanded CAG repeat neurodegenerative diseases. *Human Molecular Genetics*, 28(18), 3000–3012. <https://doi.org/10.1093/hmg/ddz096>
- Van Treeck, B., Protter, D. S. W., Matheny, T., Khong, A., Link, C. D., & Parker, R. (2018). RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2734–2739. <https://doi.org/10.1073/pnas.1800038115>
- Veloso, A., Kirkconnell, K. S., Magnuson, B., Biewen, B., Paulsen, M. T., Wilson, T. E., & Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Research*, 24(6), 896–905. <https://doi.org/10.1101/gr.171405.113>
- Villate, O., Ibarluzea, N., Maortua, H., de la Hoz, A. B., Rodriguez-Revenga, L., Izquierdo-Álvarez, S., & Tejada, M. I. (2020). Effect of AGG interruptions on FMR1 maternal transmissions. *Frontiers in Molecular Biosciences*, 7(July), 1–6. <https://doi.org/10.3389/fmolb.2020.00135>
- Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., Sugimoto, S., Yano, K., Hartzog, G. A., Winston, F., Buratowski, S., & Handa, H. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes and Development*, 12(3), 343–356. <https://doi.org/10.1101/gad.12.3.343>
- Wang, S., Latallo, M. J., Zhang, Z., Huang, B., Bobrovnikov, D. G., Dong, D., Livingston, N. M., Tjoeng, W., Hayes, L. R., Rothstein, J. D., Ostrow, L. W., Wu, B., & Sun, S. (2021). Nuclear export and translation of circular repeat-containing intronic RNA in C9ORF72-ALS/FTD. *Nature Communications*, 12(1), 4908. <https://doi.org/10.1038/S41467-021-25082-9>
- Wek, R. C., Jiang, H.-Y., & Anthony, T. G. (2006). Coping with stress: eIF2 kinases and translational control. *Biochemical Society Transactions*, 34(1), 7–11. <https://doi.org/10.1042/bst20060007>
- Wenninger, S., Cumming, S. A., Gutschmidt, K., Okkersen, K., Jimenez-Moreno, A. C., Daidj, F., Lochmüller, H., Hogarth, F., Knoop, H., Bassez, G., Monckton, D. G., van Engelen, B. G. M., & Schoser, B. (2021). Associations between variant repeat interruptions and clinical outcomes in myotonic dystrophy type 1. *Neurology: Genetics*, 7(2), e572. <https://doi.org/10.1212/NXG.0000000000000572>

- Wenzel, S., Schweimer, K., Rösch, P., & Wöhrl, B. M. (2008). The small hSpt4 subunit of the human transcription elongation factor DSIF is a Zn-finger protein with α/β type topology. *Biochemical and Biophysical Research Communications*, 370(3), 414–418. <https://doi.org/10.1016/j.bbrc.2008.03.080>
- Westergard, T., McAvoy, K., Russell, K., Wen, X., Pang, Y., Morris, B., Pasinelli, P., Trotti, D., & Haeusler, A. (2019). Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Molecular Medicine*, 11(2), 9423. <https://doi.org/10.15252/EMMM.201809423>
- Wheeler, T. M., Sobczak, K., Lueck, J. D., Osborne, R. J., Lin, X., Dirksen, R. T., & Thornton, C. A. (2009). Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science*, 325(5938), 336–339. <https://doi.org/10.1126/science.1173110>
- Wojciechowska, M., & Krzyzosiak, W. J. (2011). Cellular toxicity of expanded RNA repeats: Focus on RNA foci. *Human Molecular Genetics*, 20(19), 3811–3821. <https://doi.org/10.1093/hmg/ddr299>
- Wojtkowiak-Szlachcic, A., Taylor, K., Stepniak-Konieczna, E., Sznajder, L. J., Mykowska, A., Sroka, J., Thornton, C. A., & Sobczak, K. (2015). Short antisense-locked nucleic acids (all-LNAs) correct alternative splicing abnormalities in myotonic dystrophy. *Nucleic Acids Research*, 43(6), 3318–3331. <https://doi.org/10.1093/nar/gkv163>
- Wolozin, B., & Ivanov, P. (2019). Stress granules and neurodegeneration. *Nature Reviews Neuroscience*, 20(11), 649–666. <https://doi.org/10.1038/s41583-019-0222-5>
- Wong, J. J. L., Ritchie, W., Ebner, O. A., Selbach, M., Wong, J. W. H., Huang, Y., Gao, D., Pinello, N., Gonzalez, M., Baidya, K., Thoeng, A., Khoo, T. L., Bailey, C. G., Holst, J., & Rasko, J. E. J. (2013). Orchestrated intron retention regulates normal granulocyte differentiation. *Cell*, 154(3), 583–595. <https://doi.org/10.1016/j.cell.2013.06.052>
- Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D. L., Li, H., Hales, C. M., Gearing, M., Wingo, T. S., & Jin, P. (2013). Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), 7778–7783. <https://doi.org/10.1073/pnas.1219643110>
- Yamada, S. B., Gendron, T. F., Niccoli, T., Genuth, N. R., Grosely, R., Shi, Y., Glaria, I., Kramer, N. J., Nakayama, L., Fang, S., Dinger, T. J. I., Thoeng, A., Rocha, G., Barna, M., Puglisi, J. D., Partridge, L., Ichida, J. K., Isaacs, A. M., Petrucelli, L., & Gitler, A. D. (2019). RPS25 is required for efficient RAN translation of C9orf72 and other neurodegenerative disease-associated nucleotide repeats. *Nature Neuroscience*, 22, 1383–1388. <https://doi.org/10.1038/s41593-019-0455-7>
- Yang, H., & Hu, H. Y. (2016). Sequestration of cellular interacting partners by protein aggregates: Implication in a loss-of-function pathology. *FEBS Journal*, 283(20), 3705–3717. <https://doi.org/10.1111/febs.13722>
- Yrigollen, C. M., Martorell, L., Durbin-Johnson, B., Naudo, M., Genoves, J., Murgia, A., Polli, R., Zhou, L., Barbooth, D., Rupchock, A., Finucane, B., Latham, G. J., Hadd, A., Berry-Kravis, E., & Tassone, F. (2014). AGG interruptions and maternal age affect FMR1 CGG repeat allele stability during transmission. *Journal of Neurodevelopmental Disorders*, 6(1), 24. <https://doi.org/10.1186/1866-1955-6-24>
- Yu, Z., Teng, X., & Bonini, N. M. (2011). Triplet repeat-derived siRNAs enhance RNA-mediated toxicity in a drosophila model for myotonic dystrophy. *PLoS Genetics*, 7(3), 1–11. <https://doi.org/10.1371/journal.pgen.1001340>
- Yuva-Aydemir, Y., Almeida, S., Krishnan, G., Gendron, T. F., & Gao, F. B. (2019). Transcription elongation factor AFF2/FMR2 regulates expression of expanded GGGGCC repeat-containing C9ORF72 allele in ALS/FTD. *Nature Communications*, 10(1), 5466. <https://doi.org/10.1038/s41467-019-13477-8>
- Zhang, F., Bodycombe, N. E., Haskell, K. M., Sun, Y. L., Wang, E. T., Morris, C. A., Jones, L. H., Wood, L. D., & Pletcher, M. T. (2017). A flow cytometry-based screen identifies MBNL1 modulators that rescue splicing defects in myotonic dystrophy type I. *Human Molecular Genetics*, 26(16), 3056–3068. <https://doi.org/10.1093/hmg/ddx190>
- Zhang, K., Donnelly, C. J., Haeusler, A. R., Grima, J. C., Machamer, J. B., Steinwald, P., Daley, E. L., Miller, S. J., Cunningham, K. M., Vidensky, S., Gupta, S., Thomas, M. A., Hong, I., Chiu, S. L., Haganir, R. L., Ostrow, L. W., Matunis, M. J., Wang, J., Sattler, R., ... Rothstein, J. D. (2015). The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*, 525(7567), 56–61. <https://doi.org/10.1038/nature14973>
- Zhang, N., & Ashizawa, T. (2017). RNA toxicity and foci formation in microsatellite expansion diseases. *Current Opinion in Genetics and Development*, 44, 17–29. <https://doi.org/10.1016/j.physbeh.2017.03.040>
- Zhang, Y., Glineburg, M. R., Basrur, V., Conlon, K., Hall, D. A., & Todd, P. K. (2020). Near-cognate initiation generates FMRpolyG from CGG repeats in fragile X associated tremor ataxia syndrome. *bioRxiv*, 2020, 2020.10.19.345736. <https://doi.org/10.1101/2020.10.19.345736>
- Zhang, Y., Yang, M., Duncan, S., Yang, X., Abdelhamid, M. A. S., Huang, L., Zhang, H., Benfey, P. N., Waller, Z. A. E., & Ding, Y. (2019). G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Research*, 47(22), 11746–11754. <https://doi.org/10.1093/nar/gkz978>
- Zhang, Y. J., Gendron, T. F., Ebbert, M. T. W., O'Raw, A. D., Yue, M., Jansen-West, K., Zhang, X., Prudencio, M., Chew, J., Cook, C. N., Daugherty, L. M., Tong, J., Song, Y., Pickles, S. R., Castanedes-Casey, M., Kurti, A., Rademakers, R., Oskarsson, B., Dickson, D. W., ... Petrucelli, L. (2018). Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nature Medicine*, 24(8), 1136–1142. <https://doi.org/10.1038/s41591-018-0071-1>
- Zhang, Y. J., Gendron, T. F., Grima, J. C., Sasaguri, H., Jansen-West, K., Xu, Y. F., Katzman, R. B., Gass, J., Murray, M. E., Shinohara, M., Lin, W. L., Garrett, A., Stankowski, J. N., Daugherty, L., Tong, J., Perkerson, E. A., Yue, M., Chew, J., Castanedes-Casey, M., ... Petrucelli, L. (2016). C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins. *Nature Neuroscience*, 19(5), 668–677. <https://doi.org/10.1038/nn.4272>
- Zu, T., Cleary, J. D., Liu, Y., Bañez-Coronel, M., Bubenik, J. L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H. B., Yachnis, A. T., Swanson, M. S., & Ranum, L. P. W. (2017). RAN translation regulated by Muscleblind proteins in myotonic dystrophy type 2. *Neuron*, 95(6), 1292–1305.e5. <https://doi.org/10.1016/j.neuron.2017.08.039>

- Zu, T., Gibbens, B., Doty, N. S., Gomes-Pereira, M., Huguet, A., Stone, M. D., Margolis, J., Peterson, M., Markowski, T. W., Ingram, M. A. C., Nan, Z., Forster, C., Low, W. C., Schoser, B., Somia, N. V., Clark, H. B., Schmechel, S., Bitterman, P. B., Gourdon, G., ... Ranum, L. P. W. (2011). Non-ATG-initiated translation directed by microsatellite expansions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(1), 260–265. <https://doi.org/10.1073/pnas.1013343108>
- Zu, T., Guo, S., Bardhi, O., Ryskamp, D. A., Li, J., Tusi, S. K., Engelbrecht, A., Klippel, K., Chakrabarty, P., Nguyen, L., Golde, T. E., Sonenberg, N., & Ranum, L. P. W. (2020). Metformin inhibits RAN translation through PKR pathway and mitigates disease in C9orf72 ALS/FTD mice. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(31), 18591–18599. <https://doi.org/10.1073/pnas.2005748117>
- Zu, T., Liu, Y., Bañez-Coronel, M., Reid, T., Pletnikova, O., Lewis, J., Miller, T. M., Harms, M. B., Falchook, A. E., Subramony, S. H., Ostrow, L. W., Rothstein, J. D., Troncoso, J. C., & Ranum, L. P. W. (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), E4968–E4977. <https://doi.org/10.1073/pnas.1315438110>

How to cite this article: Baud, A., Derbis, M., Tutak, K., & Sobczak, K. (2022). Partners in crime: Proteins implicated in RNA repeat expansion diseases. *Wiley Interdisciplinary Reviews: RNA*, *13*(4), e1709. <https://doi.org/10.1002/wrna.1709>


Katarzyna Tutak,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią mojej rozprawy doktorskiej, polegał na uczestniczeniu w konceptualizacji pracy, przygotowaniu rozdziałów zatytułowanych *RAN translation*, *Targeting RBPs as potential therapeutic strategies*, opracowaniu tabeli, wprowadzeniu elementów, o które postulowali recenzenci oraz edycji i korekcie finalnej wersji tekstu.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy



Podpis współautora



Podpis promotora

dr. Anna Baud,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią rozprawy doktorskiej Katarzyny Tutak, polegał na uczestniczeniu w konceptualizacji pracy, przygotowaniu rozdziałów zatytułowanych *Introduction, Transcription and splicing of RNA^{exp}, Sequestration of proteins on RNA^{exp} molecules, Immune Response to RNA^{exp} and Products of RAN translation*, wprowadzeniu elementów, o które postulowali recenzenci, edycji i korekcie finalnej wersji tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy



Podpis współautora

dr. Magdalena Derbis,
Instytut Chemii Bioorganicznej Polskiej Akademii Nauk

14.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią rozprawy doktorskiej Katarzyny Tutak, polegał na uczestniczeniu w konceptualizacji pracy, przygotowaniu rozdziałów zatytułowanych *Toxic RNA^{exp} Molecules, Nucleocytoplasmic transport, Phase Separation*, przygotowaniu rycin, wprowadzeniu elementów, o które postulowali recenzenci oraz edycji i korekcie finalnej wersji tekstu.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy



Podpis współautora

Prof. dr hab. Krzysztof Sobczak,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że mój udział w przygotowaniu artykułu: ***Partners in crime: Proteins implicated in RNA repeat expansion diseases*** autorstwa Anny Baud*, Magdaleny Derbis*, Katarzyny Tutak* oraz Krzysztofa Sobczaka opublikowanego w czasopiśmie WIREs RNA (2022), doi: 10.1002/wrna.1709, który jest częścią rozprawy doktorskiej Katarzyny Tutak, polegał na uczestniczeniu w konceptualizacji pracy, edycji i korekcie finalnej wersji tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy w równym stopniu przyczynili się do powstania tej pracy



Podpis współautora

4. **THE MANUSCRIPT/PREPRINT: *Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions***

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

Katarzyna Tutak¹, Izabela Broniarek¹, Andrzej Zielezinski², Daria Niewiadomska¹, 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 translome (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 (Addgene 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 modifying 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-Ascll-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 0.15×10^6 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% CO₂. 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, 2×10^6 HEK293T cells were co-transfected with genetic constructs: 2 μ g of MS2-HB plasmid (#35573, Tsai et al., 2011 (Tsai *et al*, 2011)) along with 8 μ g 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 μ l 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 μ l 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 μ l of solvent A (0.1% FA in H₂O, 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 μ m beads, 150 μ m 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 \pm 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 2.5x10⁶ HEK273T cells in 200 μ L of mammalian cell lysis buffer. Lysates were cleared by centrifugation and 200 μ L of supernatant was incubated for 20 min at 21°C with 5 μ g of RNA in 200 μ L 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 0.01×10^6 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 CO₂ 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' CTCGGGCATGGCGGACTTG 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 ($^{13}\text{C}_6$ $^{15}\text{N}_2$ L-lysine-2HCl and $^{13}\text{C}_6$ $^{15}\text{N}_4$ 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 x 250 mm, 2.6 μm particles; Phenomenex, USA). The injected samples were trapped at a flow rate of 5 $\mu\text{l}/\text{min}$ 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 ($\log_2[x]$), 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 \pm 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 Skłodowska-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, Choemmel 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, Ghobakhlu F, Young DJ, Visweswaraiiah 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, Martín-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^{-ΔΔC_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 FTL/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, Long 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, Albu 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

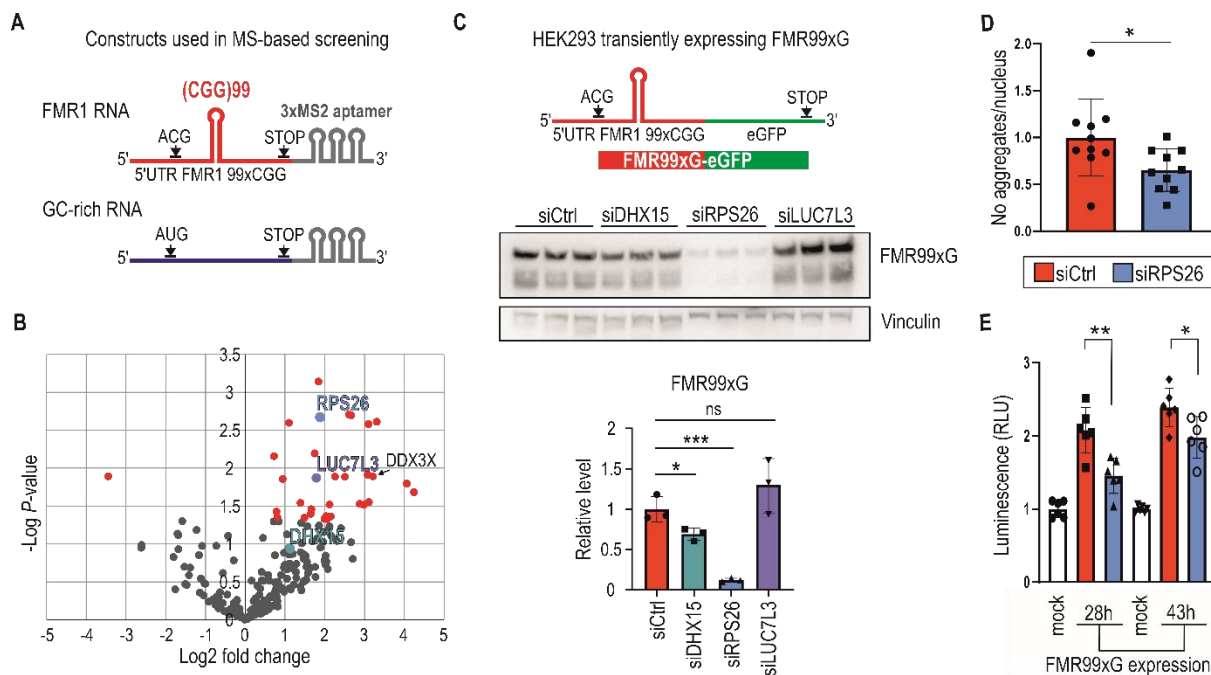


Figure 1. Mass-spectrometry (MS) -based screening revealed several proteins bound to *FMR1* RNA with expanded CGG repeats; some affect the yield of polyglycine-containing proteins, alleviating their toxicity.

(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: *, $P < 0.05$; **, $P < 0.01$.

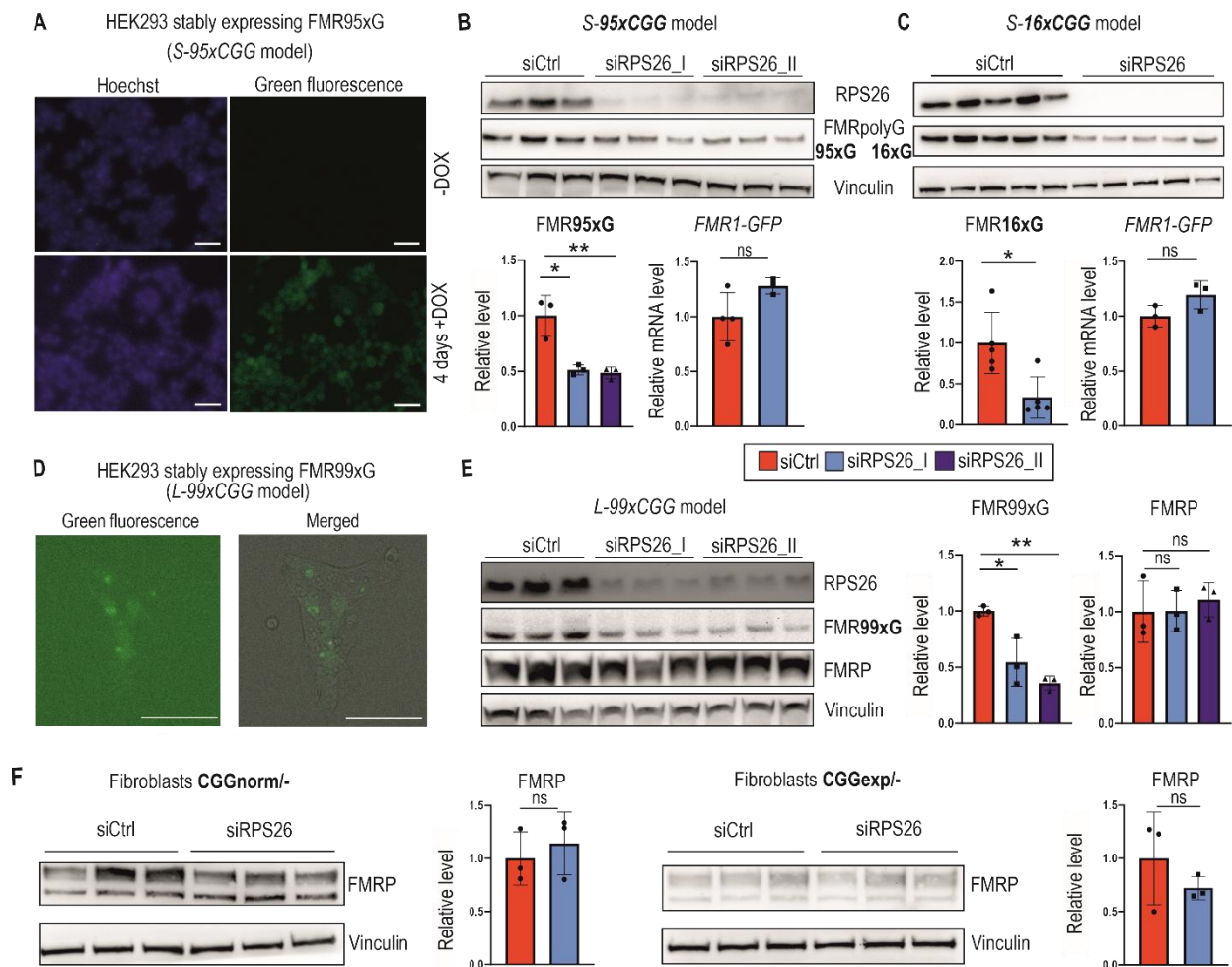


Figure 2. RPS26 insufficiency induces a lower production of polyglycine, but not FMRP, in multiple FXTAS cellular models.

(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 μ m; +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.

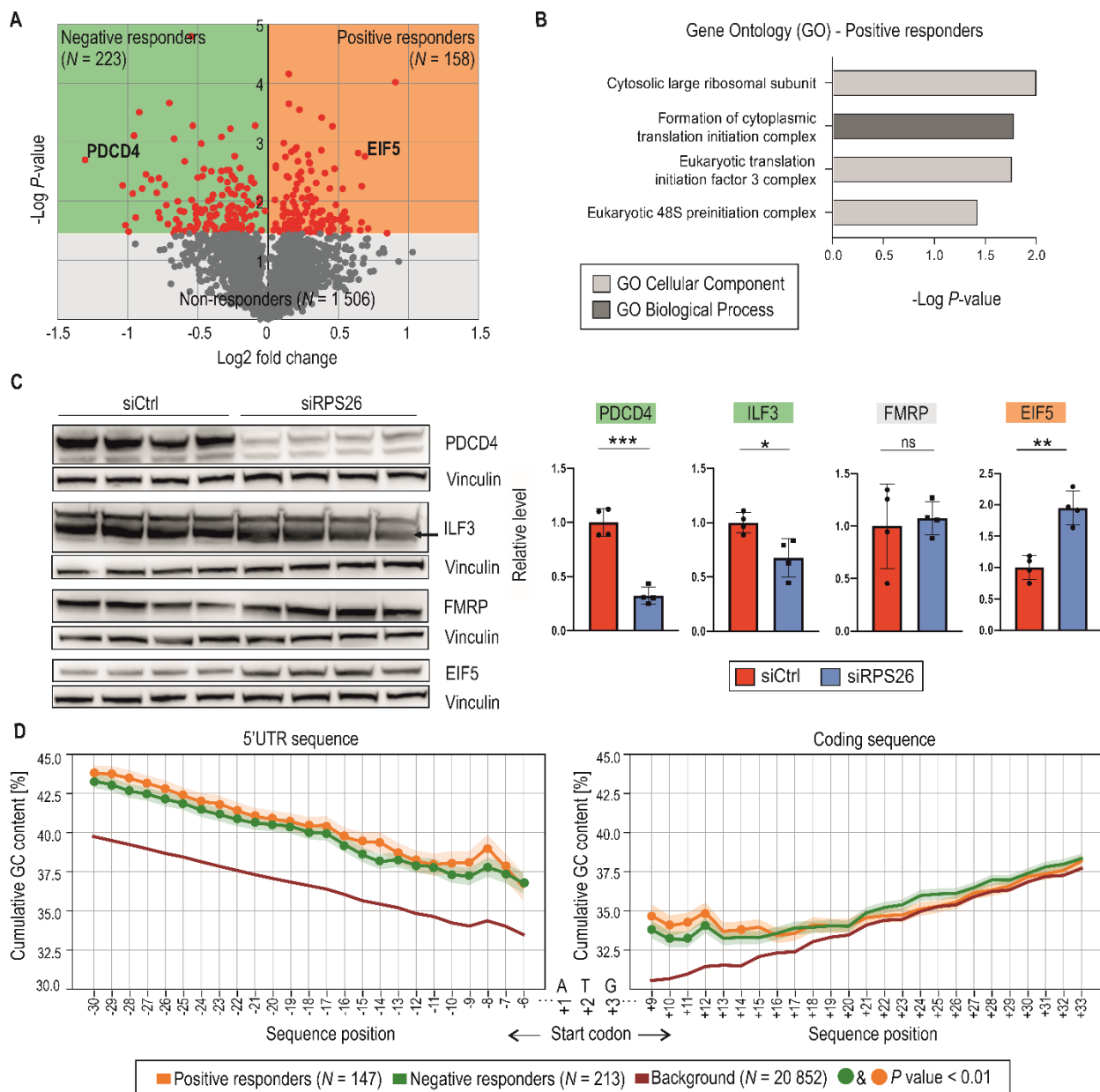


Figure 3. Changes in the proteome of RPS26-deficient cells are not robust.

(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.

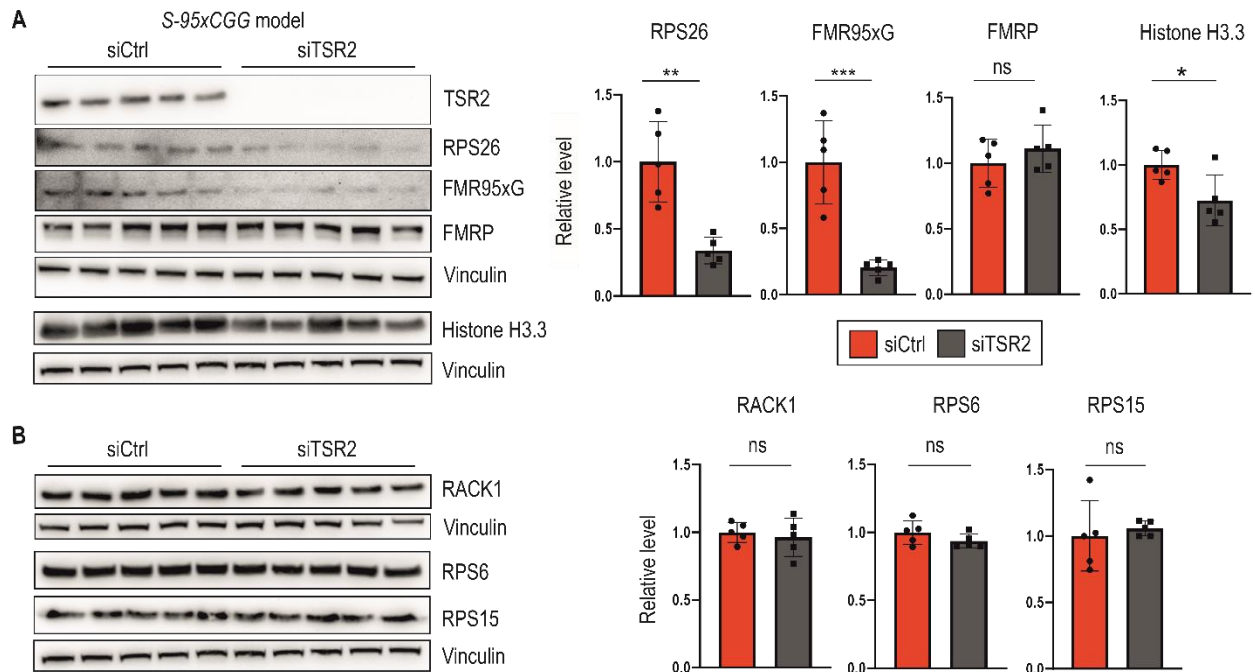


Figure 4. Insufficiency of the TSR2 chaperone protein lowers the level of RPS26 and FMRpolyG but not FMRP and selected 40S components.

(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.

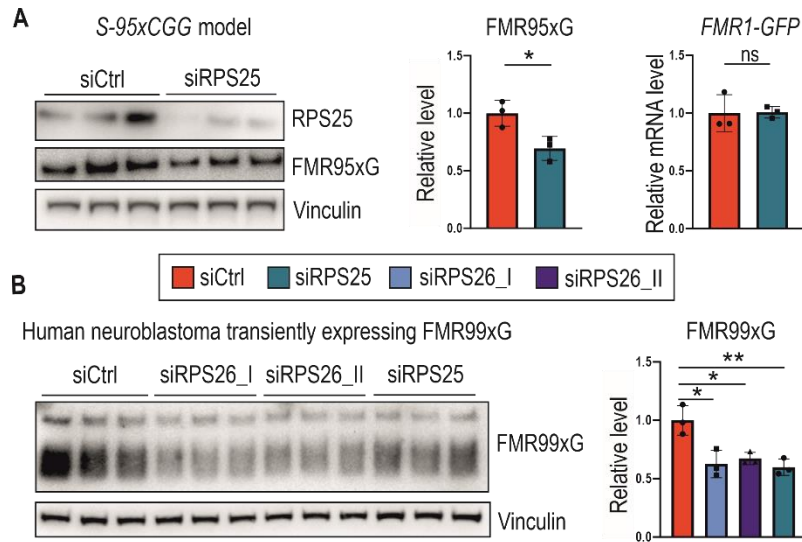


Figure 5. Silencing RPS25, the other component of 43S PIC, reduces the level of polyglycine produced from mutant *FMR1* mRNA.

(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.

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 *FMR1* 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.

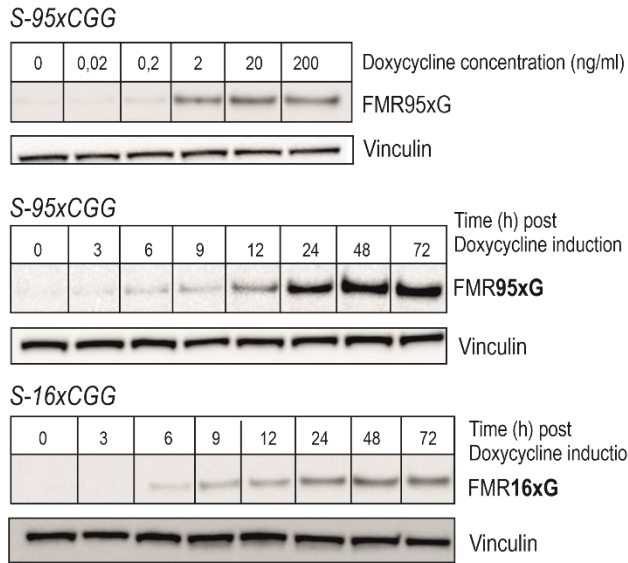
(E) RT-qPCR analysis of *FMR1-GFP* transgene expression derived from transient transfection system normalized to *GAPDH* upon RPS26 and DHX15 depletion in HEK293 cells. siRPS26_I and siRPS26_II indicate two independent siRNAs. Graphs present means from $N = 3$ biologically independent samples with SDs. 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.

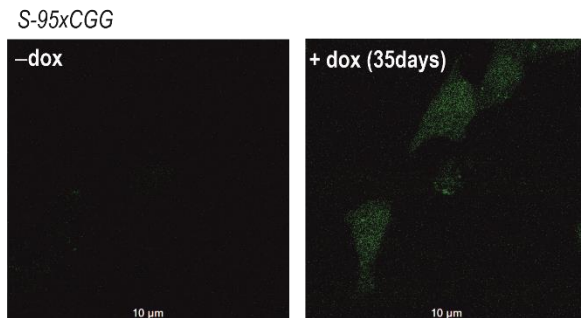
(G) Results of western blot demonstrating enrichment of SAM68 protein detected in *FMR1* RNA sample eluate after MS2-based, *in cellulo* RNA-protein pull down procedure; the percentage of following fractions were loaded on the gel: Input, 5% of total lysate; eluate, 20% of immunoprecipitated fraction.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

Supplementary Figure S2.

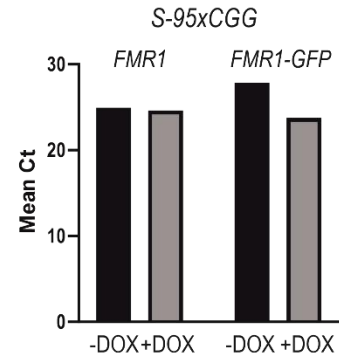
A.



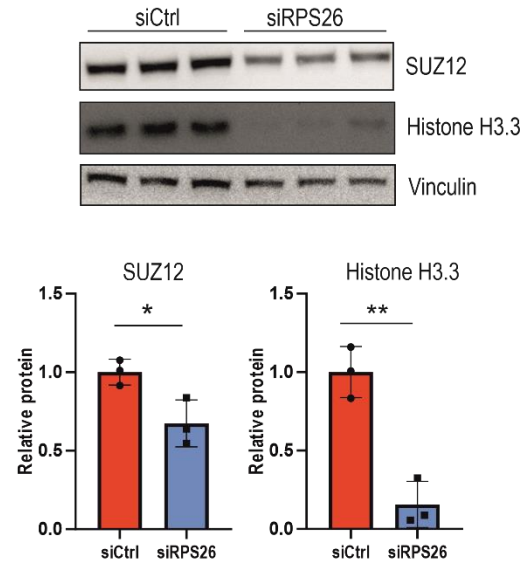
B.



C.



D.



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.

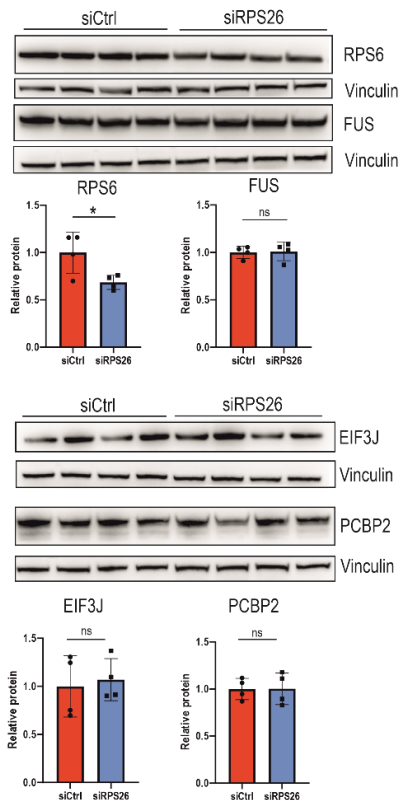
(B) The representative confocal microscopic images showing homogenous expression of FMR95xG fused with eGFP in *S-95xCGG* model after 35 days of doxycycline treatment; scale bar 10 μ m. Note that no GFP-positive aggregates of polyglycine were detected.

(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).

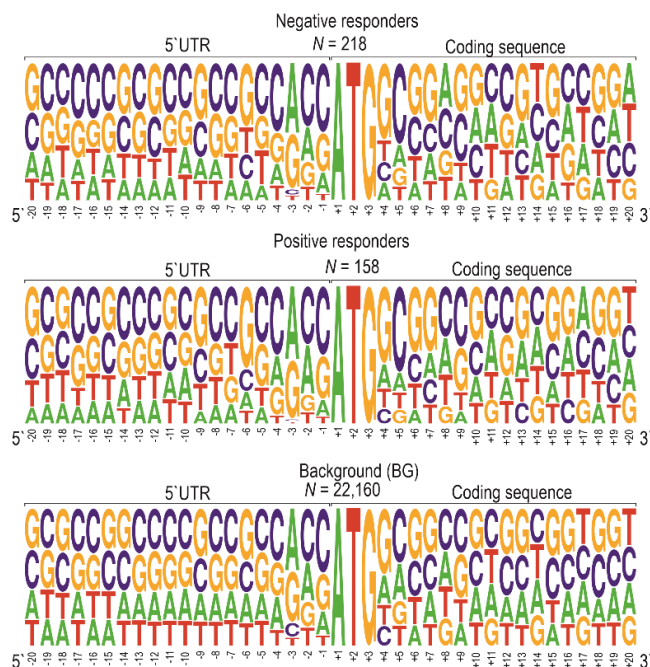
(D) Results of western blot analysis of SUZ12 and Histone H3.3 proteins normalized to Vinculin upon RPS26 silencing in COS7 cells. Graphs present means from $N = 3$ biologically independent samples with SDs. Unpaired Student's t test was used to calculate statistical significance: *, $P < 0.05$; **,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

Supplementary Figure S3.

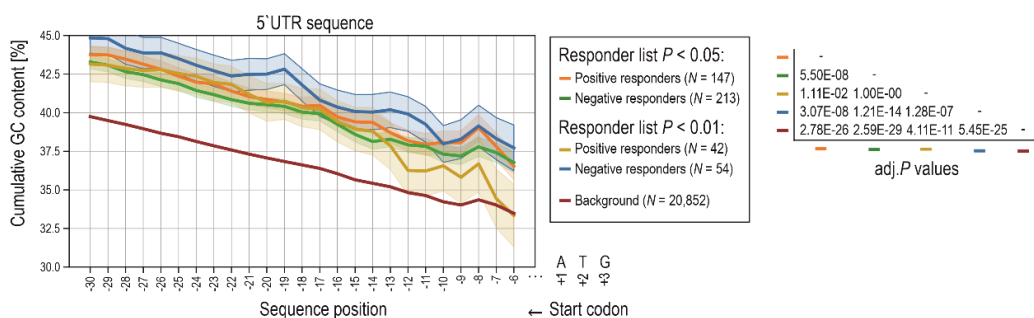
A.



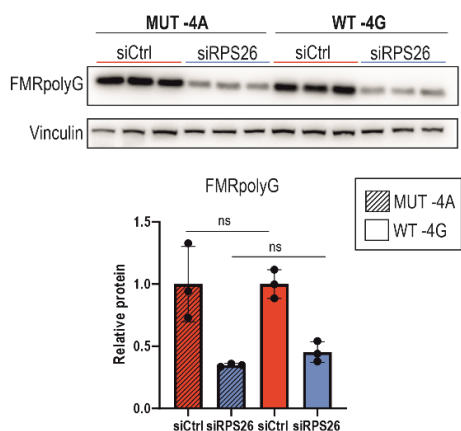
B.



C.



D.



E.

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.

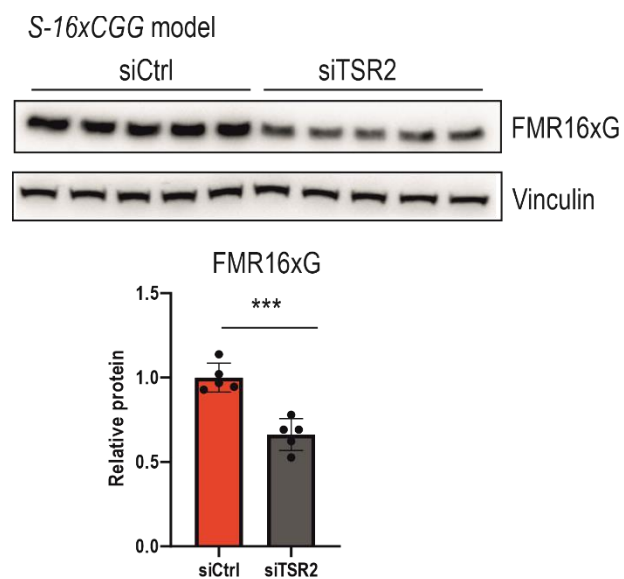
(B) Frequency of nucleotides across transcript sequence positions in the close vicinity of start codon within three groups of transcripts: negative responders ($N = 218$), positive responders ($N = 158$), and background transcripts ($N = 22,160$). The analyzed sequence positions span the last 20 nucleotides 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 \log_2 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.

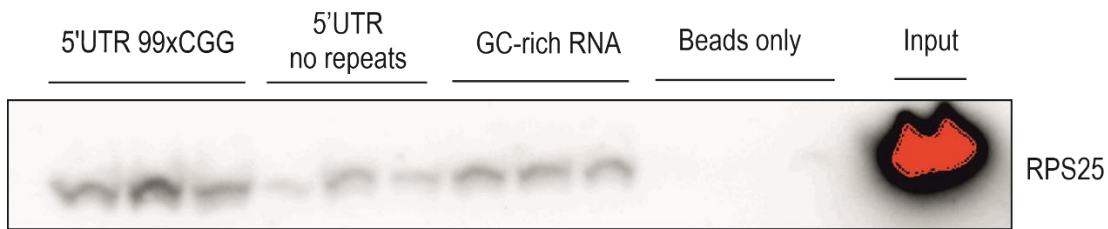
Supplementary Figure S4.



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 extent 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.

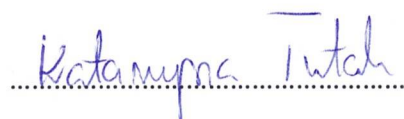
Katarzyna Tutak,
Zakład Ekspresji Genów
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią mojej rozprawy doktorskiej, mój udział polegał na uczestniczeniu w konceptualizacji pracy, wygenerowaniu modelu L-99xCGG, zaprojektowaniu i przeprowadzeniu większości eksperymentów (z wyjątkiem tych, które są wyszczególnione w oświadczeniach innych współautorów), wykonaniu analiz statystycznych, przygotowaniu rycin oraz suplementów, interpretacji wyników, a także napisaniu oraz edycji tekstu.

*Wymienieni autorzy są autorami korespondencyjnymi



Podpis współautora



Podpis promotora

dr. Izabela Broniarek,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24,Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na wygenerowaniu i charakteryzacji modeli badawczych S-95xCGG i S-16xCGG.

*Wymienieni autorzy są autorami korespondencyjnymi



Podpis współautora

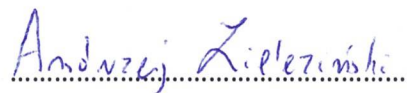
dr hab. Andrzej Zieleziński,
Zakład Biologii Obliczeniowej
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na przeprowadzeniu analizy bioinformatycznej transkryptów kodujących białka zidentyfikowane w eksperymencie SILAC-MS.

*Wymienieni autorzy są autorami korespondencyjnymi



Podpis współautora

dr. Daria Niewiadomska,
Zakład Ekspresji Genów,
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na przygotowaniu konstruktu genetycznego z mutacją -4G<A.

*Wymienieni autorzy są autorami korespondencyjnymi



Podpis współautora

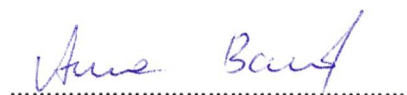
dr. Anna Baud,
Zakład Ekspresji Genów
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

14.05.24, Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na uczestniczeniu w conceptualizacji badań, nadzorowaniu prac, przeprowadzeniu analizy in vitro, wykonaniu testu na agregację białek poliglicynowych po wyciszeniu białka RPS26, pomocy w przygotowaniu próbek do analizy za pomocą spektrometrii mas (MS), analizie surowych danych MS, napisaniu i korekcie tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy są autorami korespondencyjnymi



Podpis współautora


Prof. dr hab. Krzysztof Sobczak,
Zakład Ekspresji Genów
Instytut Biologii Molekularnej i Biotechnologii,
Wydział Biologii UAM

13.05.24,Poznań

OŚWIADCZENIE WSPÓŁAUTORA ARTYKUŁU

Oświadczam, że w artykule ***Ribosomal composition affects the noncanonical translation and toxicity of polyglycine-containing proteins in fragile X-associated conditions*** autorstwa Katarzyny Tutak, Izabeli Broniarek, Andrzeja Zielezińskiego, Darii Niewiadomskiej, Anny Baud* oraz Krzysztofa Sobczaka* opublikowanym w serwisie BioRxiv jako preprint pod numerem DOI: doi.org/10.1101/2024.03.27.586952, który jest częścią rozprawy doktorskiej Katarzyny Tutak, mój udział polegał na uczestniczeniu w conceptualizacji badań, nadzorowaniu prac, napisaniu i edycji tekstu oraz pozyskaniu funduszy na badania.

*Wymienieni autorzy są autorami korespondencyjnymi



Podpis współautora

5. UNPUBLISHED RESULTS: *Characterization of other factors affecting noncanonical polyglycine synthesis from mutant FMR1 mRNA*

Given that sample preparation for MS analysis affects detectability of peptides leading to differences in the identification and the quantification of proteins (120), we repeated the biological experiment of MS2-based pull down and applied different MS sample preparation protocols including acetone precipitation of the proteins followed by single-pot solid-phase-enhanced sample preparation (SP3) procedure (described in Methodology section in details). Briefly, SP3 procedure includes immobilization of either proteins or peptides on magnetic beads followed by rinsing steps with organic solvent such as acetonitrile (121). Such intervention improves removal of common contaminants such as sodium dodecyl sulfate (SDS), which presence may negatively affect MS analysis. Results described in preprint (Tutak et al.), were obtained using another methodology, where, proteins captured on investigated bait RNA (5'UTR of *FMR1* with expanded CGG repeats) were digested on the streptavidin beads without the elution step. In another preparative approach, proteins were eluted from the beads, acetone precipitated, purified using SP3 approach and trypsin-digested prior to MS analysis (Figure 5.1). Results related to precipitation and SP3 purification variant of MS sample preparation are described below.

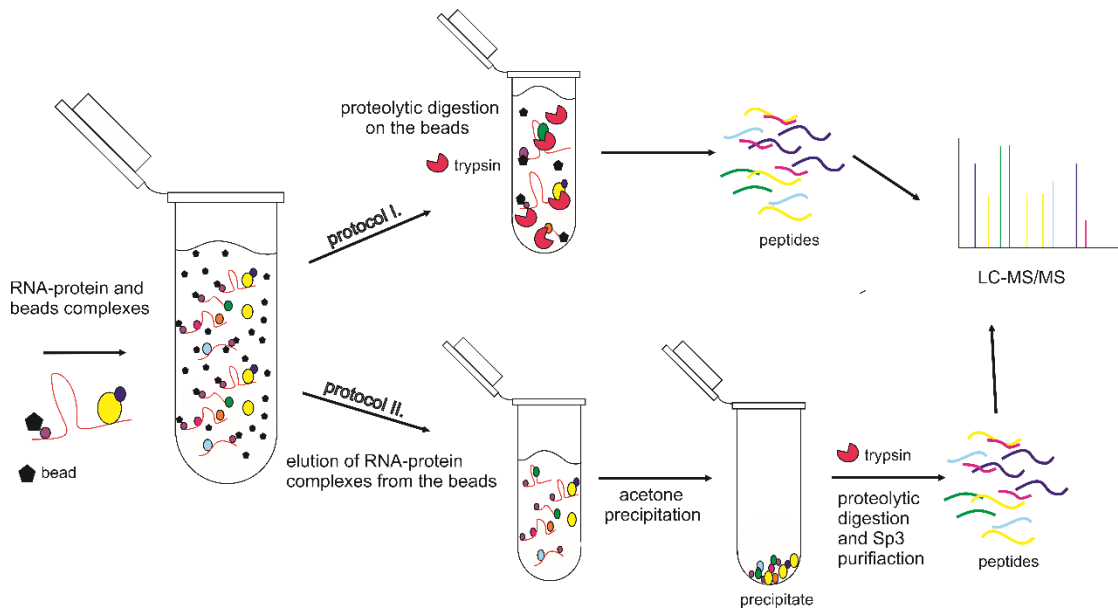
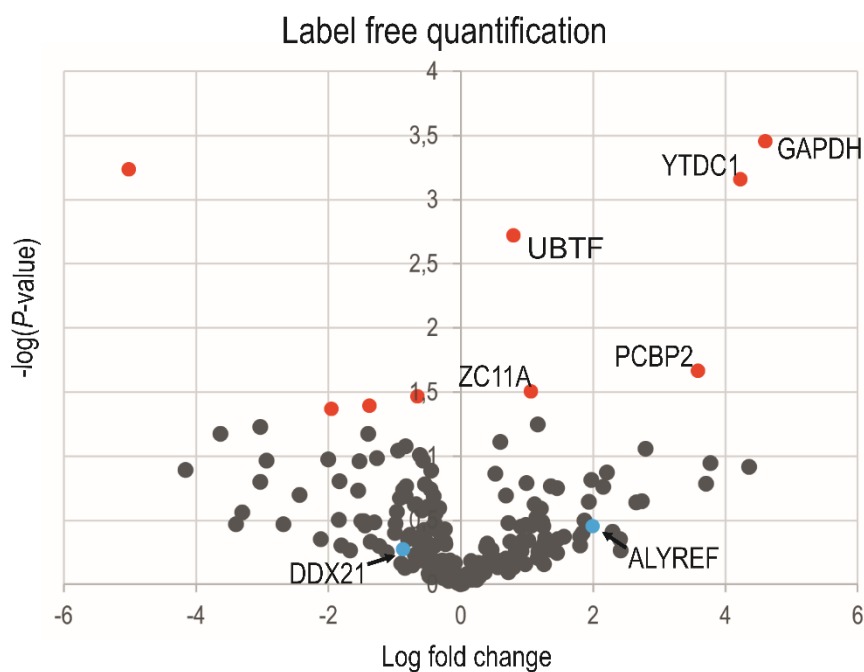


Figure 5.1. Schematic representation of two preparative approaches applied for MS sample preparation. Protocol I includes direct, on the beads digestion, protocol II includes elution of RNA-protein complexes from the beads, acetone precipitation followed by proteolytic digestion and SP3 purification. Digested peptides were analyzed by mass spectrometry.

5.1 Results

5.1.1 Alternative preparation of MS2-based pull down samples for MS analysis revealed different proteins bound to FMR1 RNA

After the RNA-protein pull down followed by MS, the bioinformatic analysis of MS RAW data was performed in an identical way as the data set described in Tutak et al., using MaxQuant software and protein identification based on Uniprot database. This analysis revealed 155 proteins common between *FMR1 RNA* replicates (proteins identified in all samples are listed in the Table 5.1). Further, we performed label free quantification (LFQ) to identify proteins enriched in *FMR1 RNA* in comparison to *GC-rich RNA*. LFQ analysis revealed five proteins significantly enriched in *FMR1 RNA* (Figure 5.2, Table 5.2).



FD	$-\log(P\text{-value})$	ID	Protein name
4.60	3.46	P04406-2	Glyceraldehyde-3-phosphate dehydrogenase GAPDH
4.22	3.16	Q96MU7-2	YTH domain-containing protein 1 YTDC1
0.80	2.72	P17480-2	Nucleolar transcription factor 1 UBTF
3.58	1.66	Q15366-7	Poly(rC)-binding protein 2 PCBP2
1.06	1.51	O75152	Zinc finger CCCH domain-containing protein 11A ZC11A

Figure 5.2. Label free quantification of proteins enriched in *FMR1 RNA* samples. 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 on the right side indicate proteins significantly enriched ($P < 0.05$) on *FMR1 RNA* compared to *GC-rich RNA*. Blue dots represent DDX21 & ALYREF analyzed in the context of potential RAN translation

regulatory properties. Proteins significantly enriched in *FMR1 RNA* in comparison to *GC-rich RNA* are listed below the graph; FD - log₂ fold change, ID - unique protein identifier from Uniprot database.

Given that proteins identified as not significantly enriched in *FMR1 RNA* samples may be also relevant to *FMR1 RNA* biology, we decided to verify RAN translation regulatory properties of Nucleolar RNA helicase 2 (DDX21) and Aly/REF Export Factor (ALYREF). We speculated that DDX21, similar to many other identified helicases may play a role in the unwinding of CGGexp hairpin structure, or thermodynamically stable RNA structure in general, and therefore modulating RAN translation efficiency (76, 85, 86). ALYREF is a component of the TRanscription EXport (TREX) protein complex that is thought to couple mRNA transcription, processing and nuclear export (122, 123). This protein represents high affinity towards 5' and 3' end of RNAs in vivo (124) and facilitate nuclear export *via* NXF1 - NTF2-related export factor 1 (NXT1) pathway (122). Components of NXF1-NXT1 pathway were described to represent RAN translation modulatory features (91, 125). Additionally, ALYREF was already identified as protein binding to mutant *FMR1* in similar screening performed by other research group (126). Moreover, both DDX21 and ALYREF were reported to be involved in guarding the genomic stability either by prevention (127) or unwinding the R loops (128) – structures involved in FXPA development (48). Altogether, it seemed reasonable to verify whether DDX21 and ALYREF modulate CGG-related RAN translation.

5.1.2 Silencing of DDX21 and DHX15 downregulated the level of FMRpolyG in two cellular models

In order to verify RAN translation regulatory properties of DDX21, we evaluated the level of FMRpolyG derived from transient transfection system upon silencing DDX21 (verified by RT-qPCR) in COS7 cell (Figure 5.3A). DDX21 depletion downregulated FMRpolyG by approximately 70%, however the level of exogenous *FMR1-GFP* mRNA was also reduced (Figure 5.3A). Despite the fact, that results were not statistically significant (perhaps due to high standard deviation (SD) in control samples), we pursued investigation of the DDX21 role in cell lines stably expressing FMRpolyG derived from expanded (95) and normal (16) CGG repeats (*S-95xCGG* & *S-16xCGG*

models, described in Tutak et al.). In addition, we included into the experiment ATP-dependent RNA helicase DHX15, which was initially identified in MS-based screening approach described in Tutak et al. The prior identification of DHX15 and verification of its RAN translation regulatory properties, encouraged us to test whether DHX15 depletion induces RAN translation suppression also in cell lines stably expressing FMRpolyG. At first, we confirmed the efficient DDX21 and DHX15 silencing by RT-qPCR (Figure 5.3A&B, lower panel) and observed that knockdown of these helicases induced lowering of FMRpolyG level regardless of CGG repeats content, as RAN translation was impeded in both *S-95xCGG* and *S-16xCGG* models (Figure 5.3B). Moreover, we checked the level of FMRP after depletion of both helicases. In *S-95xCGG* model, DHX15 depletion induced slight increase of FMRP (Figure 5.3B). On the contrary, in *S-16xCGG* model, we did not observe such effect (Figure 5.3B). Silencing DDX21 in both models, did not impact the level of FMRP (Figure 5.3B). Further, we evaluated whether observed reduction of FMRpolyG was due to change in its mRNA level. RT-qPCR analysis revealed no changes in transgenic *FMR1-GFP* mRNA level in both cell lines (Figure 5.3C).

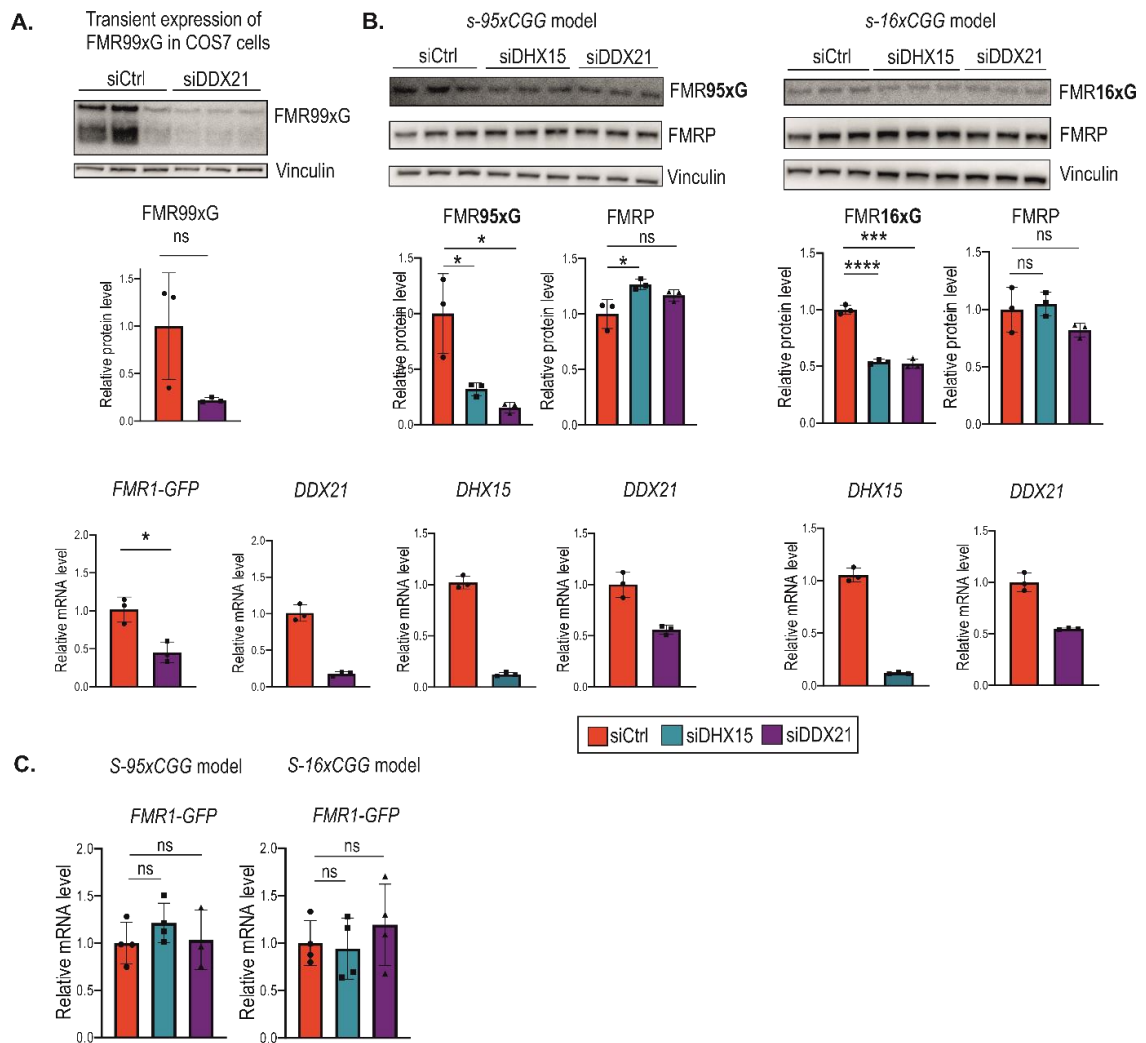


Figure 5.3. Depletion of RNA helicases, DDX21 and DHX15 downregulated FMRpolyG level in cellular models. **A.** Western blot analysis of transiently expressed FMR99xG normalized to Vinculin and RT-qPCR analysis of *FMR1-GFP* and *DDX21* normalized to *GAPDH* upon DDX21 silencing in COS7 cells. **B.** Western blot analysis of FMRP and FMRpolyG normalized to Vinculin (upper panel) and RT-qPCR knockdown verification of *DDX21* and *DHX15* normalized to *GAPDH* (lower panel) in stable *S-95xCGG* and *S-16xCGG* models upon DHX15 and DDX21 siRNA treatment. **C.** RT-qPCR analysis of transgenic *FMR1-GFP* normalized to *GAPDH* upon siDDX21 or siDHX15 treatment in *S-95xCGG* and *S-16xCGG* cell lines. The graphs present means from N = 3 (A, B; western blots, C) or N = 4 biologically independent samples (B; RT-qPCR) with SDs. An unpaired Student's t-test was used to calculate statistical significance: *, $P < 0.05$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, non-significant. (The full size images are placed in supplementary figures section).

5.1.3 The effect of ALYREF silencing on FMRpolyG

Furthermore, we evaluated the level of FMRpolyG derived from transient transfection system upon silencing ALYREF in COS7 cells. Obtained results suggested negative effect of ALYREF insufficiency on FMRpolyG, however the results were not significant (perhaps due to high SD in control samples) (Figure 5.4A). We verified if the level of *FMR1-GFP* transcript was affected by ALYREF depletion and RT-qPCR analysis revealed the increase of transgenic mRNA (Figure 5.4A). Next, we tested the effect on ALYREF silencing on FMRpolyG production in *S-95xCGG* and *S-16xCGG* models. Despite the fact that in *S-95xCGG* model, results of western blot analysis of FMR95xG were not statistically significant (perhaps due to high SD in control sample and the weak band signal hindering the quantification), similarly to transient overexpression experiment, ALYREF insufficiency seemed to evoke the downregulation of FMRpolyG level (Figure 5.4B). In case of *S-16xCGG* model, silencing of ALYREF did not change FMR16xG level. In both models, ALYREF depletion had no effect on FMRP level (Figure 5.4B). It is important to highlight, that presented results concerning ALYREF CGG-related RAN translation regulatory properties are preliminary and drawing confident conclusion require repetition of experiments.

A. Transient expression of FMR99xG in COS7 cells

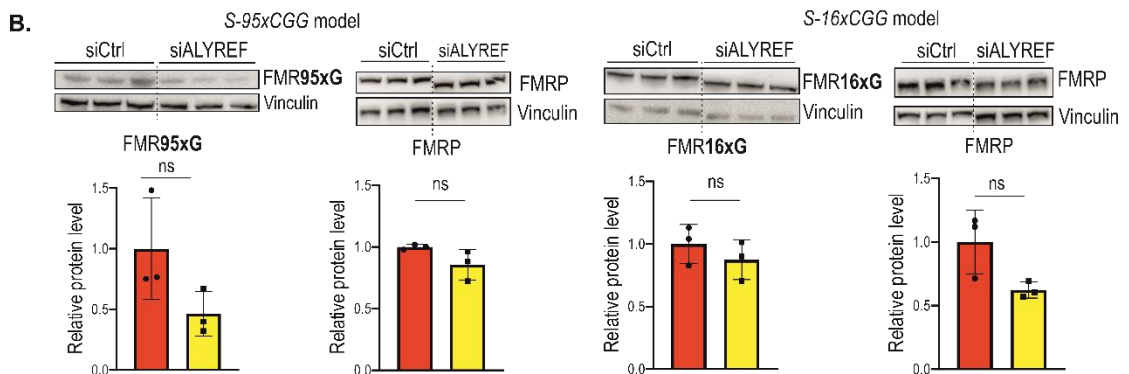
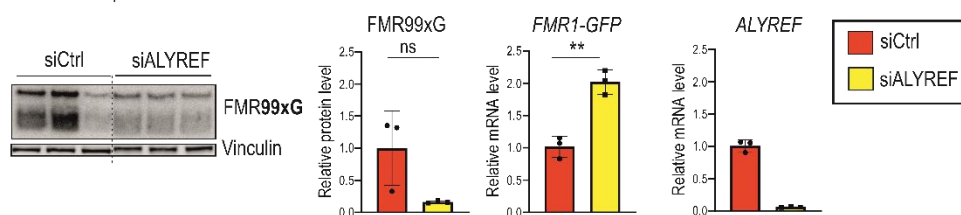


Figure 5.4. The effect of ALYREF silencing on FMRpolyG in cellular models. A. Western blot analysis of transiently expressed FMR99xG normalized to Vinculin and RT-qPCR analysis of

transgenic *FMR1-GFP* and *ALYREF* mRNA normalized to *GAPDH* upon *ALYREF* silencing in COS7 cells. **B.** Western blot analysis of FMRP and FMRpolyG normalized to Vinculin derived from expanded and normal CGG repeats in stable *S-95xCGG* and *S-16xCGG* models, respectively, after *ALYREF* depletion. Note that presented blots were cropped (the full size images are placed in supplementary figures section). The graphs present means from N = 3 biologically independent samples with SDs. An unpaired Student's t-test was used to calculate statistical significance: **, $P < 0.01$; ns, non-significant.

5.1.4 Technical aspects of mass spectrometry sample preparation affected identification of proteins bound to FMR1 RNA

In order to characterize pool of proteins bound to *FMR1 RNA* derived from the samples treated with acetone precipitation and SP3 purification, we performed a comparative analysis between technical replicates. Only proteins identified by unique peptides were considered. Analysis revealed 155 common proteins among analyzed samples (Figure 5.5A, Table 5.3). Gene ontology (GO) analysis indicated that majority of identified proteins represent RNA binding properties, are a constituent of the ribosome, participate in translation or are involved in other processes related to RNA metabolism (Figure 5.5A, Table 5.4). Given that depending on MS sample preparation identified proteins may vary (120), we compared proteins identified in two different MS sample preparation approaches, either direct, on-beads digestion or precipitation with acetone and SP3 purification. To exclude possibility that observed differences may stem from bioinformatic analysis, we applied the same bioinformatic pipeline (described in Tutak et al.) to analyze both raw data sets. Number of proteins identified in both approaches was similar, 254 and 213, for direct digestion or acetone precipitation approach respectively (while considering proteins identified in minimum 2 technical replicates), and 67 proteins (including RPS26) were overlapping between data sets (Figure 5.5B, Table 5.5). Enriched GO terms for 67 proteins were similar to the ones revealed for 155 proteins common among *FMR1 RNA* replicates (Figure 5.5A&B), however some terms were exclusive for those 67 proteins, such as rRNA metabolic process or Synapse (Figure 5.5B, Table 5.6).

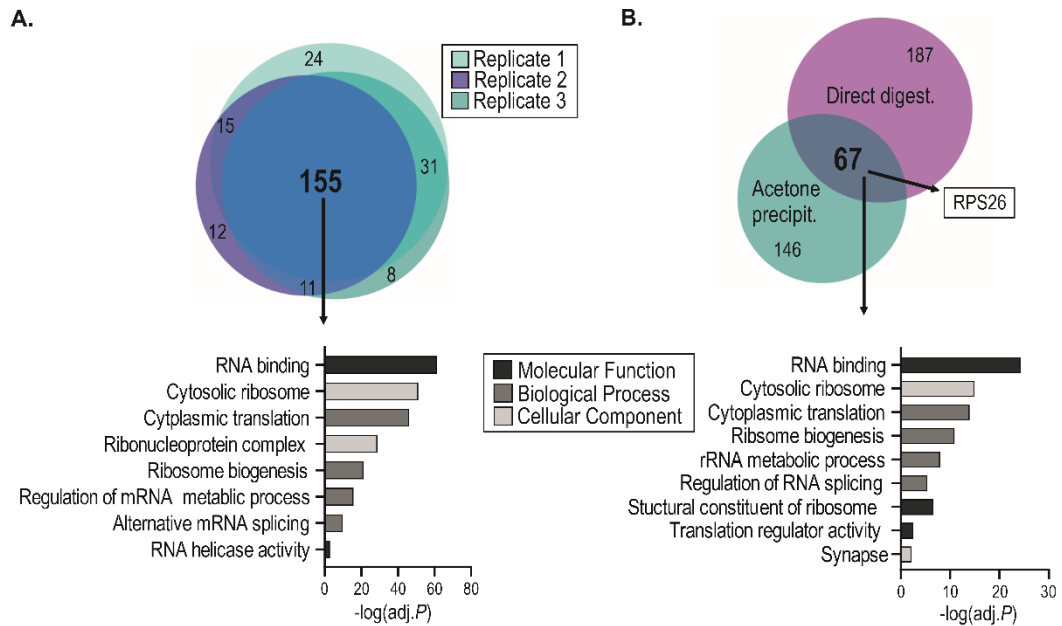


Figure 5.5. Comparison between proteins interacting with *FMR1* RNA identified in samples prepared with two different workflows for MS analysis. A. Venn diagram represents proteins common between three replicates of *FMR1* RNA samples treated with acetone and SP3 purification. Lower graph represents GO analysis performed on 155 common proteins between triplicates. **B.** Venn diagram represents proteins identified in at least 2 replicates of *FMR1* RNA samples treated by two different MS sample preparation workflows (either Direct digestion or Acetone precipitation*). Lower graph represents GO analysis performed on 67 common proteins among two sample groups. Selected, significantly enriched GO terms are presented and full lists are presented in Table 5.4 and 5.6, respectively. *Please note that the higher number of proteins analyzed in this data set (B – Acetone precipitation) is different than the one analyzed in A due to less stringent selection criteria.

Table 5.5 67 proteins identified in *FMR1* RNA samples overlapping between two MS analysis

Proteins identified in at least two technical replicates in two, separate MS data sets. Bolded proteins indicate known interactors of *FMR1* described by other research groups.

67 common proteins in two MS data sets		
Uniprot ID	Entry Name	Protein names
Q13085	ACACA	Acetyl-CoA carboxylase 1
P11498	PYC	Pyruvate carboxylase, mitochondrial
P05165	PCCA	Propionyl-CoA carboxylase alpha chain
Q96RQ3	MCCA	Methylcrotonoyl-CoA carboxylase subunit alpha
O00763-3	ACACB	Acetyl-CoA carboxylase 2
P49327	FAS	Fatty acid synthase
P05166	PCCB	Propionyl-CoA carboxylase beta chain
PODMV9	HS71B	Heat shock 70 kDa protein 1B
P26641	EF1G	Elongation factor 1-gamma
P08238	HS90B	Heat shock protein HSP 90-beta
Q9BVP2-2	GNL3	Guanine nucleotide-binding protein-like 3
O60841	IF2P	Eukaryotic translation initiation factor 5B
P39023	RL3	Large ribosomal subunit protein uL3
P04075	ALDOA	Fructose-bisphosphate aldolase A
P09874	PARP1	Poly [ADP-ribose] polymerase 1-terminus
P63261	ACTG	Actin
O75367	H2AY	Core histone macro-H2A.1
P09651-3	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1
P13639	EF2	Elongation factor 2
Q9UQ35	SRRM2	Serine/arginine repetitive matrix protein 2
P27635	RL10	Large ribosomal subunit protein uL16
Q9BZE4-3	GTPB4	GTP-binding protein 4
P62241	RS8	Small ribosomal subunit protein eS8
P63241	IF5A1	Eukaryotic translation initiation factor 5A-1
P19338	NUCL	Nucleolin
P26599-1	PTBP1	Polypyrimidine tract-binding protein 1
Q96SB4-4	SRPK1	SRSF protein kinase 1
Q99880	H2B1L	Histone H2B type 1-L
O95232	LC7L3	Luc7-like protein 3
P07437	TBB5	Tubulin beta chain
P07900	HS90A	Heat shock protein HSP 90-alpha
P11387	TOP1	DNA topoisomerase 1
P22626-2	HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1
P30414	NKTR	NK-tumor recognition protein
P51532-5	SMCA4	Transcription activator BRG1
P52272-2	HNRPM	Heterogeneous nuclear ribonucleoprotein M
P61978-3	HNRPK	Heterogeneous nuclear ribonucleoprotein K
P62244	RS15A	Small ribosomal subunit protein uS15
P62750	RL23A	Large ribosomal subunit protein uL23

P62753	RS6	Small ribosomal subunit protein eS6
P62805	H4	Histone H4
P62847-2	RS24	Small ribosomal subunit protein eS24
Q00839-2	HNRPU	Heterogeneous nuclear ribonucleoprotein U
Q07065	CKAP4	Cytoskeleton-associated protein 4
Q14241	ELOA1	Elongin-A
Q14839	CHD4	Chromodomain-helicase-DNA-binding protein 4
Q1ED39	KNOP1	Lysine-rich nucleolar protein 1
Q8NC51-4	SERB1	SERPINE1 mRNA-binding protein 1
Q9BQG0	MBB1A	Myb-binding protein 1A
Q9NX58	LYAR	Cell growth-regulating nucleolar protein
Q6FI13	H2A2A	Histone H2A type 2-A
P38919	IF4A3	Eukaryotic initiation factor 4A-III
P62854	RS26	Small ribosomal subunit protein eS26
Q99459	CDC5L	Cell division cycle 5-like protein
Q9Y2X3	NOP58	Nucleolar protein 58
P22087	FBRL	rRNA 2'-O-methyltransferase fibrillarin
P23396	RS3	Small ribosomal subunit protein uS3
P38432	COIL	Coilin
P81605	DCD	Dermcidin
Q13263-2	TIF1B	Transcription intermediary factor 1-beta
Q96T37-4	RBM15	RNA-binding protein 15
P52597	HNRPF	Heterogeneous nuclear ribonucleoprotein F
P63173	RL38	Large ribosomal subunit protein eL38
P46777	RL5	Large ribosomal subunit protein uL18
P49411	EFTU	Elongation factor Tu
P62263	RS14	Small ribosomal subunit protein uS11
P62829	RL23	Large ribosomal subunit protein uL14

5.2 Discussion

The main aim of the PhD thesis was to identify novel RAN translation modifiers specific for mutant *FMR1* harboring expanded CGG repeats. We adapted RNA-tagging technique, which allowed in cellulo RNA-protein interaction capture taking into account RNA secondary structures as well as localization of studied molecule in different cellular compartments. During experimental procedure we fixed interactions between proteins and RNAs with formaldehyde crosslinking in account for identification of weak and transient interactions, however this step may have introduced a bias for secondary interactions. Given that downstream MS sample preparation affects proteins identification (120), we applied MS2-based pull down followed by two distinct MS sample preparation protocols to minimize potential losses due to technical limitations. First technical variant relied on direct on-beads proteolytic digestion which limits potential losses of material, yet may leave more contaminants such a SDS in the sample that hinders further MS analysis. In second technical variant, eluted proteins were precipitated with acetone and digested peptides were treated with SP3 purification procedure to remove common contaminants, although precipitation step might have influenced the solubility of proteins affecting their further identification. These factors need to be taken into consideration while interpreting the obtained data.

As the result, we obtained two data sets containing *FMR1* binding proteins (Tutak et al.; Figure 1 and Figure 4A, Supplementary Table 3). Number of proteins identified in both approaches was similar and 67 proteins (Table 5.5) were overlapping between two data sets (Figure 5.5B). The differences among identified proteins may be explained by distinct downstream MS sample preparation protocols. Unsurprisingly, regardless of investigated data set, GO terms were mostly common to all analyses including following GO terms: RNA binding properties, ribosomal constituent or translation regulation (Tutak et al.; Figure 5.5A&B). However, in data containing 67 common proteins some GO terms were exclusive for this group, such as rRNA processing or Synapse. This result is especially interesting because *FMR1* encodes FMRP, a protein with well documented role in synaptic plasticity (129). Remarkably, among identified proteins some of them were already identified as *FMR1* interactors (such as serine/arginine rich proteins, for instance, SRPK1 (91)) or were

found in intranuclear inclusions in FXTAS patient's tissues (such as Nucleolin or multiple Heterogeneous nuclear ribonucleoproteins, hnRNPs and others (39, 67, 75, 93, 126, 130)). (Forementioned proteins were bolded in the Table 5.5.) Hence, this implies that our data set containing 67 interactors of *FMR1 RNA* may be a particularly valuable resource of information about mutant *FMR1* biology.

LFQ analysis elucidated proteins enriched in *FMR1 RNA* samples in comparison to *GC-rich RNA* control samples (Figure 5.2, Table 5.1). However, it is worth mentioning that precipitation steps might have incurred loss of some proteins present in samples, thus negatively affecting further quantification and skewing the data interpretation. Additionally, LFQ data should be analyzed with assumption that protein binding to two RNA molecules (*FMR1 RNA* and *GC-rich RNA*) may represent key regulatory functions specific for the one or both of the RNAs. Therefore, proteins enriched in *GC-rich RNA* samples should not be discarded by default from further analysis in the context of *FMR1* regulation. Among five proteins significantly enriched in *FMR1 RNA* samples (Figure 5.2), Poly(rC)-binding protein 2 (PCBP2) and YTH domain-containing protein 1 (YTDC1) seem to be particularly interesting in the context of potential regulation of *FMR1* mRNA. YTDC1 is a regulator of alternative splicing that specifically recognizes and binds N6-methyladenosine (m6A)-containing RNAs (131). Moreover, YTDC1 binds to Serine/arginine-rich splicing factor 3 (SRSF3) and Nuclear RNA export factor 1 (NXF1) facilitating nuclear export of RNAs harboring m6A modifications (132). Given that our preliminary experiments indicate that *FMR1* contains m6A modification (data unpublished), YTDC1 may have so far not described roles in *FMR1* processing and metabolism. PCBP2 is known to interact with structured RNAs such as internal ribosome entry site (IRES) elements (133) and play an important role in modulating cellular response to viral infection (134, 135). Binding to IRES sequences may suggest potential RAN translation regulator properties, especially that known RAN translation modifier, RPS25 binds to IRES as well (117, 136–138). To our surprise, GAPDH was identified as the most enriched protein on *FMR1 RNA*. This may be explained as a technical artifact or stem from the lack of GAPDH identification in *GC-rich RNA* samples. However, this enzyme is characterized as non-canonical RNA binding protein (139), which suggests its potential role in *FMR1* metabolism. Nevertheless, other

proteins identified in LFQ analyses (Tutak et al.; Figure 1, Supplementary table 5) regardless of statistical significance are worth further investigating in the context of RAN translation regulation of *FMR1* metabolism in general.

We verified RAN translation regulatory properties of two RNA helicases; DHX15 and DDX21. Our results are in line with current knowledge concerning roles of these enzymes in RAN translation regulation (76, 85, 86). Observed RAN translation regulation facilitated by DHX15 and DDX21 seemed not to be dependent on CGG content as both types of FMRpolyG (with normal and long glycine stretches) were sensitive to DHX15 and DDX21 depletion (Figure 5.3B). Similar regulation was observed for DDX3X, which was shown to bind to *FMR1* regardless of CGG repeats content (85). On the contrary to studied helicases, ALYREF depletion had an effect only on shorter form of FMRpolyG (Figure 5.4B).

In contrast to stable cell lines, silencing DDX21 in COS7 cells transiently expressing FMRpolyG, downregulated the level of *FMR1-GFP* (Figure 5.3A&C). Given documented role of DDX21 in resolving the R loop structure and therefore promoting transcription (128, 140), it is possible that DDX21 depletion hindered transcription from multiple copies of plasmid DNA due to insufficient R loop unwinding. Additionally, level of *DDX21* mRNA in stable cell lines indicated that silencing procedure was approximately 50% less efficient in comparison to COS7 cells. (Figure 5.3A&B). Perhaps, partial expression of DDX21 in *S-95/16xCGG* models was sufficient enough to resolve potential R loops and not affect the transcription, although this is a speculation. Surprisingly, silencing ALYREF resulted in doubling the amount of *FMR1-GFP* mRNA while downregulating the FMRpolyG (Figure 5.4A). This might be explained by transcriptional activator role of ALYREF in addition to its role in nuclear-cytoplasmic transport (141–143). Hence, the role of ALYREF in *FMR1* transcription regulation or other RNA metabolism processes requires further investigation in future.

Importantly, DDX21 and ALYREF depletion did not affect FMRP level (Figure 5.3B, 5.4B), on the contrary to DHX15 insufficiency, which induced the increase of FMRP level in *S-95xCGG* model (Figure 5.3B). This suggests that studied proteins represent some specificity towards FMRpolyG over FMRP frame. It should be noted that in humans, FMRpolyG and FMRP are translated from the same mRNA, while in

artificial models these two proteins are not synthesized from the same transcript, complicating the data interpretation.

Overall, DHX15, DDX21 and ALYREF were not previously described in the context of RAN translation regulation and we provide the evidence indicating their roles in this process. Nevertheless, conducted experiments verifying DHX15, DDX21 and ALYREF RAN translation properties were preliminary and did not cover mechanistic aspects, hence further studies to reveal the nature of discovered regulation are required. Some aspects are worth addressing in the future. For instance, verifying RNA-protein interaction (ideally by an orthogonal technique) and addressing whether CGG repeats are necessary for this interaction, or checking if overexpression of selected proteins will evoke anticipated upregulation of FMRpolyG biosynthesis. Additionally, testing multiple siRNAs targeting selected proteins by validating the level of target protein expression by western blotting may be desired to maximize the reproducibility and minimize possibility of technical artifacts affecting results interpretation.

To sum up, MS sample preparation protocol including acetone precipitation and SP3 purification steps revealed multiple proteins bound to *FMR1* RNA including DDX21 and ALYREF. Both of the proteins were tested in the context of potential RAN translation modulation and our results indicated their involvement in regulating FMRpolyG level, thus pointing novel potential therapeutic targets for FXPAC.

Finally, 67 proteins identified in both types of MS2-based pull down samples treated with different downstream protocols are particularly valuable source of information about biology of mutant *FMR1* mRNA and our dataset likely contains numerous RAN translation related proteins yet to be uncovered.

5.3 Materials and methods

The methodology of all experiments was presented in Tutak et al., except for the procedures described below.

5.3.1 MS2-based pull down followed by acetone protein precipitation and SP3 purification

Detailed experimental protocol for MS2-based pull down is described in section titled “Mass spectrometry-based proteins screening” (Tutak et al.). The protocol applied here differs from the one described in the manuscript in terms of sample preparation for MS analysis. Differences begin after pull down procedure followed by washing steps and are presented as follow; samples were transferred to the fresh Protein Lobind tubes (Eppendorf) and RNA-protein complexes were eluted from the beads by heating the samples for 10 min at 95°C in 100 µl of elution buffer (0.1% sodium dodecyl sulphate, 0.05 mM biotin in phosphate buffered saline). Next, 400 µl of ice-cold acetone was added to the sample. Samples were mixed and incubated for 60 min in 4°C followed by 10 min centrifugation at 13,000 g. Supernatant was discarded, and the protein pellet was air dried. Subsequently, protein pellets were processed in Mass Spectrometry Laboratory, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a Street, 02-106 Warsaw, Poland. Samples were suspended in dissolution buffer (100 mM NH₄HCO₃). First, the cysteines were reduced by 1 hour incubation with 20 mM tris(2-carboxyethyl)phosphine (TCEP) at 60°C followed by 10 min incubation at a RT with 50 mM methyl methanethiosulfonate (MMTS). Digestion was performed at 37°C overnight with 1 µg of trypsin (Promega). After digestion, peptides were vacuum-dried, re-suspended in 10 µl of extraction buffer (0.1% TFA 2% acetonitrile) with sonication, and processed using single-pot solid-phase-enhanced sample preparation (SP3). Magnetic beads mix were prepared by combining equal parts of Sera-Mag Carboxyl hydrophilic and hydrophobic particles (09-981-121 and 09-981-123, GE Healthcare). The bead mix was washed three times with MS-grade water and re-suspended in a working concentration of 10 µg/µl. The bead mix was then added to the samples then suspended in 100% acetonitrile, this step was repeated 3 times. Pure peptides were eluted from the beads by using 2% acetonitrile in MS-grade water. Using a magnet, the peptides solution was separated from beads. Peptide

mixture was dried in SpeedVac and re-suspended in 80 μ l extraction buffer (0.1% TFA 2% acetonitrile) with sonication.

5.3.2 Mass spectrometry (MS)

Samples were analysed using LC-MS system composed of Evosep One (Evosep Biosystems) coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific) via Flex nanoESI ion source (Thermo Fisher Scientific). 20 μ l of each sample was loaded onto disposable Evtips C18 trap columns (Evosep Biosystems, Odense, Denmark) according to the manufacturer protocol. Chromatography was carried out at a flow rate 220 nl/min using the 88 min preformed gradient on EV1106 analytical column (Dr Maisch C18 AQ, 1.9 μ m beads, 150 μ m ID, 15 cm long, Evosep Biosystems). Data was acquired in positive mode with a data-dependent method using the following parameters. MS1 resolution was set at 60 000. For MS2, resolution was set at 15 000 and top 40 precursors within an isolation window of 1.6 m/z considered for MS/MS analysis.

5.3.3 MS data analysis

MS data was analyzed as described in Tutak et al. Briefly, raw data obtained from the LC-MS/MS runs were analyzed in MaxQuant v2.0.3.0 (144) software. UniProtKB database for reviewed human canonical and isoform proteins of May 2023 was used. For comparative analysis between samples only protein identified by unique peptides were included. Label-free quantification (LFQ) using Perseus software v2.0.3.0 (145) after filtering for “reverse”, “contaminant” and “only identified by site” proteins. The LFQ intensity was logarithmized ($\log_2[x]$), 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.

5.3.4 Cell culture and transfection

COS7 and S-95/16xCGG cells were maintained as described in Tutak et al. All siRNAs were purchased from Thermo Fisher Scientific, catalog number indicated in brackets: siDHX15 (#s4029), siDDX21 (#s17564), siALYREF (#s19854) and siCtrl (#4390847). For the delivery of siRNAs with the final concentration 25 nM in culture, reverse transfection protocol was applied using jetPRIME[®] reagent (Polyplus) according to

manufacturer's instruction. In case of COS7, cells were harvested 48h post siRNA silencing and 24h post transient plasmids expression. *S-95/16xCGG* cells were harvested 48h post siRNA silencing and 72h after doxycycline treatment.

5.3.5 Quantitative real-time PCR (RT-qPCR)

RNA isolation and reverse transcription were performed as described in Tutak et al. 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. RT-qPCRs were run at 58°C annealing temperature. To amplify *DHX15*, *DDX21* and *ALYREF* following primer pairs were used:

DHX15:

Forward: 5' CGCAGATGAGGCCAAGATGA 3', Reverse: 5' CGTTCTAAATGTGCCACCTGC 3'

DDX21:

Forward: 5' GGACCCAAAGGGCAGCAGTT 3', Reverse: 5' AACGACTGGGCATCCTGCCT 3'

ALYREF:

Forward: 5' TGCCACCTCTGTTTACGCTC 3', Reverse 5' TCTGGTCGCAGCTTAGGAAC 3'

Ct values were normalized to *GAPDH* mRNA level. Fold differences in expression level were calculated according to the $2^{-\Delta\Delta Ct}$ method (146).

5.3.6 Gene Ontology

Gene ontology (GO) analysis performed was performed using an online software - g:Profiler (147) with g:SCS algorithm, where at least 95% of matches above threshold are statistically significant. As reference proteome we used total human proteome.

5.3.7 Venn diagrams

Venn diagrams were prepared with BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams (148).

5.4 Supplementary information

5.4.1 Tables

All tables (except Table 5.5) are attached to the PhD dissertation on CD in a form of the excel file titled TABLES – UNPUBLISHED RESULTS (Chapter 5) with following pages:

Table 5.1 MaxQuant output including identified proteins in MS2-screening followed by acetone precipitation and SP3 purification

Table 5.2 Perseus output with label free quantification analysis

Table 5.3 155 proteins identified in three technical replicates of *FMR1 RNA*

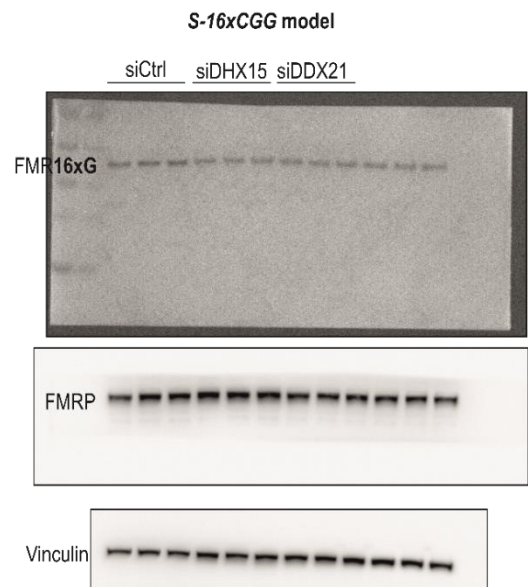
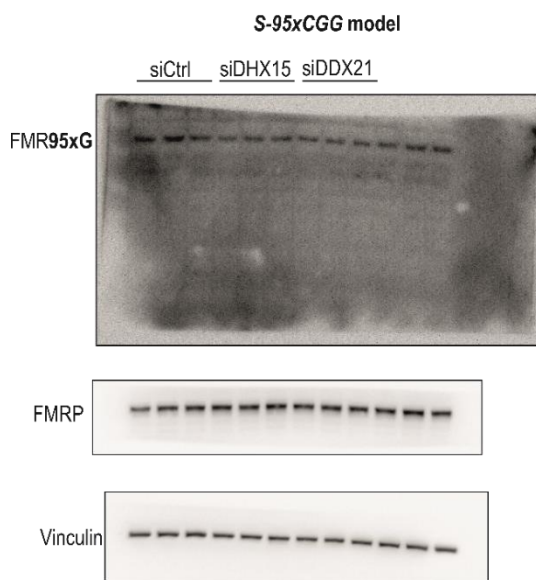
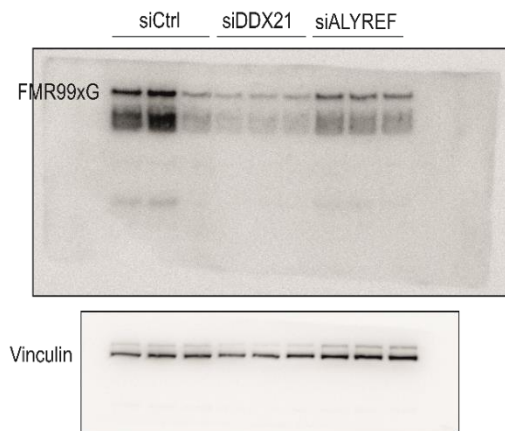
Table 5.4 Gene ontology analysis performed on 155 proteins common for three replicates of *FMR1 RNA*

Table 5.5 67 proteins identified in *FMR1 RNA* samples overlapping between two MS analysis (printed in the dissertation)

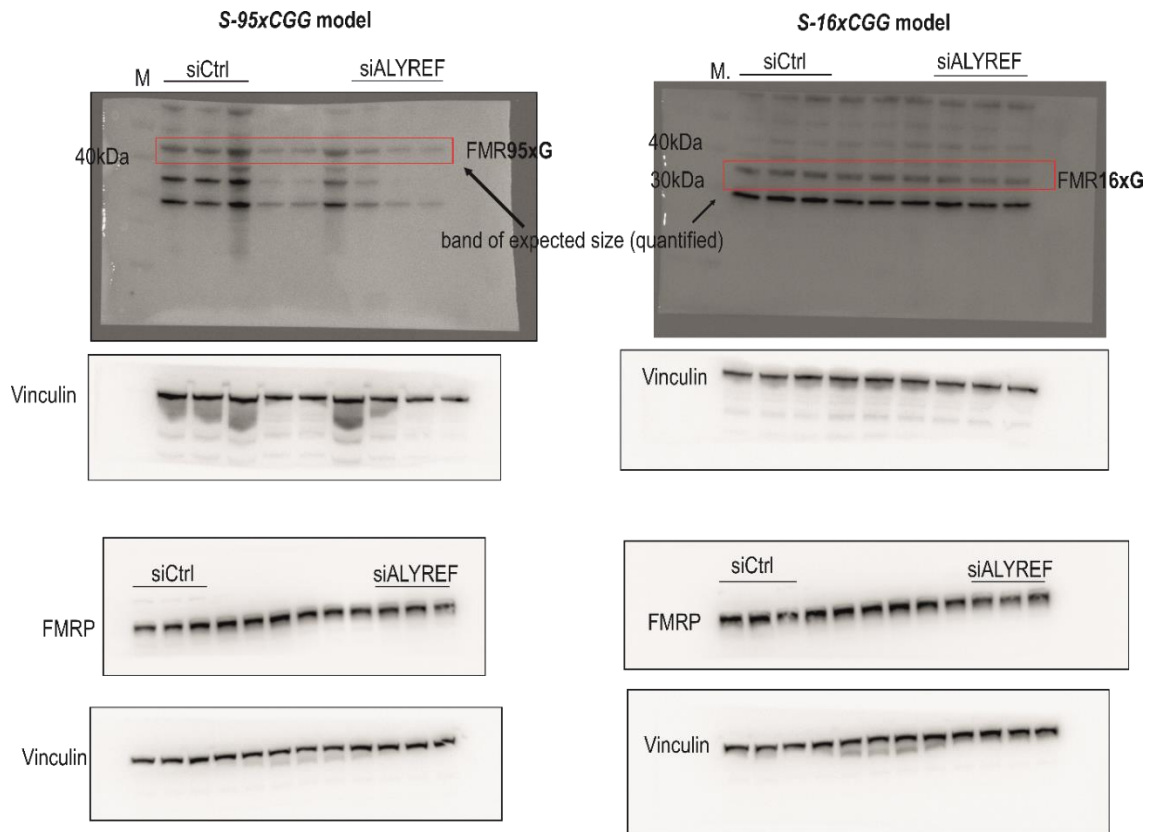
Table 5.6 Gene ontology analysis performed on 67 proteins bound to *FMR1 RNA*

5.4.2 Supplementary figures

COS7 cell line



Supplementary Figure 5.1. Full size western blot images corresponding to Figures 5.3 & 5.4.



Supplementary Figure 5.2. Full size western blot images corresponding to Figure 5.4.

5.5 Bibliography

1. Gatchel,J.R. and Zoghbi,H.Y. (2005) Diseases of Unstable Repeat Expansion: Mechanisms and Common Principles. *Nature Reviews Genetics*, **6**, 743–755.
2. Fotsing,S.F., Margoliash,J., Wang,C., Saini,S., Yanicky,R., Shleizer-Burko,S., Goren,A. and Gymrek,M. (2019) The impact of short tandem repeat variation on gene expression. *Nat Genet*, **51**, 1652–1659.
3. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., Fitzhugh,W., *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
4. Khristich,A.N. and Mirkin,S.M. (2020) On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J Biol Chem*, **295**, 4134–4170.
5. Paulson,H. (2018) Repeat expansion diseases. *Handb Clin Neurol*, **147**, 105-123.
6. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.P., Hudson,T., *et al.* (1992) Molecular basis of myotonic dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.
7. Harper,P.S. (1999) Huntington's disease: a clinical, genetic and molecular model for polyglutamine repeat disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **354**, 957-961.
8. Renton,A.E., Majounie,E., Waite,A., Simón-Sánchez,J., Rollinson,S., Gibbs,J.R., Schymick,J.C., Laaksovirta,H., van Swieten,J.C., Myllykangas,L., *et al.* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, **72**, 257–268.
9. Gadgil,R., Barthelemy,J., Lewis,T. and Leffak,M. (2017) Replication stalling and DNA microsatellite instability. *Biophys Chem*, **225**, 38–48.
10. Fortune,M.T., Vassilopoulos,C., Coolbaugh,M.I., Siciliano,M.J. and Monckton,D.G. (2000) Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. *Hum Mol Genet*, **9**, 439–445.
11. Mirkin,S.M. (2007) Expandable DNA repeats and human disease. *Nature*, **447**, 932–940.
12. Swami,M., Hendricks,A.E., Gillis,T., Massood,T., Mysore,J., Myers,R.H. and Wheeler,V.C. (2009) Somatic expansion of the Huntington's disease CAG repeat in the brain is associated with an earlier age of disease onset. *Hum Mol Genet*, **18**, 3039–3047.

13. Wenstrom,K.D. (2002) Fragile X and other trinucleotide repeat diseases. *Obstet Gynecol Clin North Am*, **29**, 367–388.
14. Sutcliffe,J.S., Nelson,D.L., Zhang,F., Pieretti,M., Caskey,C.T., Saxe,D. and Warren,S.T. (1992) DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet*, **1**, 397–400.
15. Fardaei,M., Larkin,K., Brook,J.D. and Hamshere,M.G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res*, **29**, 2766–2771.
16. Lubs,H.A. (1969) A marker X chromosome. *Am J Hum Genet*, **21**, 231-244.
17. Eichler,E.E., Kunst,C.B., Lugenbeel,K.A., Ryder,O.A., Davison,D., Warren,S.T. and Nelson,D.L. (1995) Evolution of the cryptic FMR1 CGG repeat. *Nat Genet*, **11**, 301–308.
18. Kremer,E.J., Pritchard,M., Lynch,M., Yu,S., Holman,K., Baker,E., Warren,S.T., Schlessinger,D., Sutherland,G.R. and Richards,R.I. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)_n. *Science*, **252**, 1711–1714.
19. Fu,Y.H., Kuhl,D.P.A., Pizzuti,A., Pieretti,M., Sutcliffe,J.S., Richards,S., Verkert,A.J.M.H., Holden,J.J.A., Fenwick,R.G., Warren,S.T., *et al.* (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, **67**, 1047–1058.
20. Verkerk,A.J.M.H., Pieretti,M., Sutcliffe,J.S., Fu,Y.H., Kuhl,D.P.A., Pizzuti,A., Reiner,O., Richards,S., Victoria,M.F., Zhang,F., *et al.* (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*, **65**, 905–914.
21. Oberlé,I., Rousseau,F., Heitz,D., Kretz,C., Devys,D., Hanauer,A., Boué,J., Bertheas,M.F. and Mandel,J.L. (1991) Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science*, **252**, 1097–1102.
22. Sidorov,M.S., Auerbach,B.D. and Bear,M.F. (2013) Fragile X mental retardation protein and synaptic plasticity. *Mol Brain*, **6**, 15.
23. Hagerman,R.J., Berry-Kravis,E., Hazlett,H.C., Bailey,D.B., Moine,H., Kooy,R.F., Tassone,F., Gantois,I., Sonenberg,N., Mandel,J.L., *et al.* (2017) Fragile X syndrome. *Nat Rev Dis Primers*, **3**, 17065.
24. Johnson,K., Herring,J. and Richstein,J. (2020) Fragile X Premutation Associated Conditions (FXPAC). *Front Pediatr*, **8**, 266.
25. Hagerman,P. (2013) Fragile X-associated tremor/ataxia syndrome (FXTAS): pathology and mechanisms. *Acta Neuropathol*, **126**, 1–19.

26. Buijsen,R.A.M., Visser,J.A., Kramer,P., Severijnen,E.A.W.F.M., Gearing,M., Charlet-Berguerand,N., Sherman,S.L., Berman,R.F., Willemsen,R. and Hukema,R.K. (2016) *c. Human Reproduction*, **31**, 158–168.
27. Hagerman,R.J., Protic,D., Rajaratnam,A., Salcedo-Arellano,M.J., Aydin,E.Y. and Schneider,A. (2018) Fragile X-Associated Neuropsychiatric Disorders (FXAND). *Front Psychiatry*, **9**, 564.
28. Hagerman,R.J. and Hagerman,P. (2016) Fragile X-associated tremor/ataxia syndrome-features, mechanisms and management. *Nat Rev Neurol*, **12**, 403–412.
29. Tassone,F., Long,K.P., Tong,T.H., Lo,J., Gane,L.W., Berry-Kravis,E., Nguyen,D., Mu,L.Y., Laffin,J., Bailey,D.B., *et al.* (2012) FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Med*, **4**, 100.
30. Jacquemont,S., Hagerman,R.J., Leehey,M.A., Hall,D.A., Levine,R.A., Brunberg,J.A., Zhang,L., Jardini,T., Gane,L.W., Harris,S.W., *et al.* (2004) Penetrance of the fragile X-associated tremor/ataxia syndrome in a premutation carrier population. *JAMA*, **291**, 460–469.
31. Tassone,F., Hagerman,R.J., Loesch,D.Z., Lachiewicz,A., Taylor,A.K. and Hagerman,P.J. (2000) Fragile X males with unmethylated, full mutation trinucleotide repeat expansions have elevated levels of FMR1 messenger RNA. *Am J Med Genet*, **94**, 232–236.
32. Kenneson,A., Zhang,F., Hagedorn,C.H. and Warren,S.T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum Mol Genet*, **10**, 1449–1454.
33. Tassone,F., Beilina,A., Carosi,C., Albertosi,S., Bagni,C., Li,L., Glover,K., Bentley,D. and Hagerman,P.J. (2007) Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA*, **13**, 555–562.
34. Jacquemont,S., Hagerman,R.J., Leehey,M., Grigsby,J., Zhang,L., Brunberg,J.A., Greco,C., Des Portes,V., Jardini,T., Levine,R., *et al.* (2003) Fragile X Premutation Tremor/Ataxia Syndrome: Molecular, Clinical, and Neuroimaging Correlates. *Am J Hum Genet*, **72**, 869-878.
35. Hagerman,R.J., Leehey,M., Heinrichs,W., Tassone,F., Wilson,R., Hills,J., Grigsby,J., Gage,B. and Hagerman,P.J. (2001) Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology*, **57**, 127–130.
36. Greco,C.M., Hagerman,R.J., Tassone,F., Chudley,A.E., Del Bigio,M.R., Jacquemont,S., Leehey,M. and Hagerman,P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, **125**, 1760–1771.

37. Greco,C.M., Berman,R.F., Martin,R.M., Tassone,F., Schwartz,P.H., Chang,A., Trapp,B.D., Iwahashi,C., Brunberg,J., Grigsby,J., *et al.* (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain*, **129**, 243–255.
38. Todd,P.K., Oh,S.Y., Krans,A., He,F., Sellier,C., Frazer,M., Renoux,A.J., Chen,K. chun, Scaglione,K.M., Basrur,V., *et al.* (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*, **78**, 440–455.
39. Ma,L., Herren,A.W., Espinal,G., Randol,J., McLaughlin,B., Martinez-Cerdeño,V., Pessah,I.N., Hagerman,R.J. and Hagerman,P.J. (2019) Composition of the Intranuclear Inclusions of Fragile X-associated Tremor/Ataxia Syndrome. *Acta Neuropathol Commun*, **7**, 143.
40. Glineburg,M.R., Todd,P.K., Charlet-Berguerand,N. and Sellier,C. (2018) Repeat-associated non-AUG (RAN) translation and other molecular mechanisms in Fragile X Tremor Ataxia Syndrome. *Brain Res*, **1693**, 43-54.
41. Rebar,R.W. (2009) Premature ovarian failure. *Obstetrics and Gynecology*, **113**, 1355–1363.
42. Stephanie L. Sherman (2002) Premature ovarian failure in the fragile X syndrome. *Am. J. Med. Genet.*, **97**, 189–194.
43. Buijsen,R.A.M., Visser,J.A., Kramer,P., Severijnen,E.A.W.F.M., Gearing,M., Charlet-Berguerand,N., Sherman,S.L., Berman,R.F., Willemsen,R. and Hukema,R.K. (2016) Presence of inclusions positive for polyglycine containing protein, FMRpolyG, indicates that repeat-associated non-AUG translation plays a role in fragile X-associated primary ovarian insufficiency. *Human Reproduction*, **31**, 158–168.
44. Shelly,K.E., Candelaria,N.R., Li,Z., Allen,E.G., Jin,P. and Nelson,D.L. (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.
45. Sobczak,K., de Mezer,M., Michlewski,G., Krol,J. and Krzyzosiak,W.J. (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. *Nucleic Acids Res*, **31**, 5469-5482.
46. Krzyzosiak,W.J., Sobczak,K., Wojciechowska,M., Fiszler,A., Mykowska,A. and Kozlowski,P. (2012) Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res*, **40**, 11–26.
47. Fry,M. and Loeb,L.A. (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc Natl Acad Sci U S A*, **91**, 4950.
48. Loomis,E.W., Sanz,L.A., Chédin,F. and Hagerman,P.J. (2014) Transcription-Associated R-Loop Formation across the Human FMR1 CGG-Repeat Region. *PLoS Genet*, **10**, e1004294.

49. Sellier,C., Rau,F., Liu,Y., Tassone,F., Hukema,R.K., Gattoni,R., Schneider,A., Richard,S., Willemsen,R., Elliott,D.J., *et al.* (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO Journal*, **29**, 1248–1261.
50. 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.
51. Malik,I., Kelley,C.P., Wang,E.T. and Todd,P.K. (2021) Molecular mechanisms underlying nucleotide repeat expansion disorders. *Nature Reviews Molecular Cell Biology*, **22**, 589–607.
52. Tassone,F., Protic,D., Allen,E.G., Archibald,A.D., Baud,A., Elvassore,N., Gabis,L. V, Grudzien,S.J., Hall,D.A., Hessl,D., *et al.* (2023) Insight and Recommendations for Fragile X-Premutation-Associated Conditions from the Fifth International Conference on FMR1 Premutation. *Cells*, **12**, 2330.
53. Groh,M., Lufino,M.M.P., Wade-Martins,R. and Gromak,N. (2014) R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet*, **10**, e1004318.
54. Boque-Sastre,R., Soler,M. and Guil,S. (2017) Detection and Characterization of R Loop Structures. *Methods Mol Biol*, **1543**, 231–242.
55. Abu Diab,M., Mor-Shaked,H., Cohen,E., Cohen-Hadad,Y., Ram,O., Epsztejn-Litman,S. and Eiges,R. (2018) The G-rich Repeats in FMR1 and C9orf72 Loci Are Hotspots for Local Unpairing of DNA. *Genetics*, **210**, 1239–1252.
56. Ginno,P.A., Lott,P.L., Christensen,H.C., Korf,I. and Chédin,F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*, **45**, 814–825.
57. Belotserkovskii,B.P., Shin,J.H.S. and Hanawalt,P.C. (2017) Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res*, **45**, 6589–6599.
58. Crossley,M.P., Bocek,M. and Cimprich,K.A. (2019) R-Loops as Cellular Regulators and Genomic Threats. *Mol Cell*, **73**, 398–411.
59. Gan,W., Guan,Z., Liu,J., Gui,T., Shen,K., Manley,J.L. and Li,X. (2011) R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev*, **25**, 2041–2056.
60. Kumari,D. and Usdin,K. (2014) Polycomb group complexes are recruited to reactivated FMR1 alleles in Fragile X syndrome in response to FMR1 transcription. *Hum Mol Genet*, **23**, 6575–6583.

61. Colak,D., Zaninovic,N., Cohen,M.S., Rosenwaks,Z., Yang,W.Y., Gerhardt,J., Disney,M.D. and Jaffrey,S.R. (2014) Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. *Science (1979)*, **343**, 1002–1005.
62. Wojciechowska,M. and Krzyzosiak,W.J. (2011) Cellular toxicity of expanded RNA repeats: Focus on RNA foci. *Hum Mol Genet*, **20**, 3811–3821.
63. Tassone,F., Iwahashi,C. and Hagerman,P.J. (2004) FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol*, **1**, 103–105.
64. Fardaei,M., Larkin,K., Brook,J.D. and Hamshere,M.G. (2001) In vivo co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res*, **29**, 2766–2771.
65. Jain,A. and Vale,R.D. (2017) RNA phase transitions in repeat expansion disorders. *Nature*, **546**, 243–247.
66. Lin,Y., Protter,D.S.W., Rosen,M.K. and Parker,R. (2015) Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell*, **60**, 208–219.
67. Asamitsu,S., Yabuki,Y., Ikenoshita,S., Kawakubo,K., Kawasaki,M., Usuki,S., Nakayama,Y., Adachi,K., Kugoh,H., Ishii,K., *et al.* (2021) CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related tremor/ataxia syndrome. *Sci Adv*, **7**, eabd9440.
68. Sofola,O.A., Jin,P., Qin,Y., Duan,R., Liu,H., de Haro,M., Nelson,D.L. and Botas,J. (2007) RNA-Binding Proteins hnRNP A2/B1 and CUGBP1 Suppress Fragile X CGG Premutation Repeat-Induced Neurodegeneration in a Drosophila Model of FXTAS. *Neuron*, **55**, 565–571.
69. Jin,P., Duan,R., Qurashi,A., Qin,Y., Tian,D., Rosser,T.C., Liu,H., Feng,Y. and Warren,S.T. (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.
70. He,F., Krans,A., Freibaum,B.D., Paul Taylor,J. and Todd,P.K. (2014) TDP-43 suppresses CGG repeat-induced neurotoxicity through interactions with HnRNP A2/B1. *Hum Mol Genet*, **23**, 5036–5051.
71. Zu,T., Gibbens,B., Doty,N.S., Gomes-Pereira,M., Huguet,A., Stone,M.D., Margolis,J., Peterson,M., Markowski,T.W., Ingram,M.A.C., *et al.* (2011) Non-ATG-initiated translation directed by microsatellite expansions. *Proc Natl Acad Sci U S A*, **108**, 260–265.
72. Mori,K., Weng,S.M., Arzberger,T., May,S., Rentzsch,K., Kremmer,E., Schmid,B., Kretzschmar,H.A., Cruts,M., Van Broeckhoven,C., *et al.* (2013) The C9orf72

- GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTL/ALS. *Science*, **339**, 1335–1338.
73. Bañez-Coronel, M., Ayhan, F., Tarabochia, A.D., Zu, T., Perez, B.A., Tusi, S.K., Pletnikova, O., Borchelt, D.R., Ross, C.A., Margolis, R.L., *et al.* (2015) RAN Translation in Huntington Disease. *Neuron*, **88**, 667–677.
74. Ash, P.E.A., Bieniek, K.F., Gendron, T.F., Caulfield, T., Lin, W.L., DeJesus-Hernandez, M., Van Blitterswijk, M.M., Jansen-West, K., Paul, J.W., Rademakers, R., *et al.* (2013) Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron*, **77**, 639–646.
75. Sellier, C., Buijsen, R.A.M., 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.
76. Kearse, M.G., Green, K.M., Krans, A., Rodriguez, C.M., Linsalata, A.E., Goldstrohm, A.C. and Todd, P.K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol Cell*, **62**, 314–322.
77. Krans, A., Skariah, G., Zhang, Y., Bayly, B. and Todd, P.K. (2019) Neuropathology of RAN translation proteins in fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun*, **7**, 152.
78. Zhang, Y., Glineburg, M.R., Basur, V., Conlon, K., Wright, S.E., Krans, A., Hall, D.A. and Todd, P.K. (2022) Mechanistic convergence across initiation sites for RAN translation in fragile X associated tremor ataxia syndrome. *Hum Mol Genet*, **31**, 2317–2332.
79. Wright, S.E., Rodriguez, C.M., Monroe, J., Xing, J., Krans, A., Flores, B.N., Barsur, V., Ivanova, M.I., Koutmou, K.S., Barmada, S.J., *et al.* (2022) CGG repeats trigger translational frameshifts that generate aggregation-prone chimeric proteins. *Nucleic Acids Res*, **50**, 8674–8689.
80. Krans, A., Kearse, M.G. and Todd, P.K. (2016) Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Ann Neurol*, **80**, 871–881.
81. Rodriguez, C.M., Wright, S.E., Kearse, M.G., Haenfler, J.M., Flores, B.N., Liu, Y., Ifrim, M.F., Glineburg, M.R., Krans, A., Jafar-Nejad, P., *et al.* (2020) A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat Neurosci*, **23**, 386–397.
82. Green, K.M., Linsalata, A.E. and Todd, P.K. (2016) RAN translation—What makes it run? *Brain Res*, **1647**, 30–42.

83. Cheng,W., Wang,S., Mestre,A.A., Fu,C., Makarem,A., Xian,F., Hayes,L.R., Lopez-Gonzalez,R., Drenner,K., Jiang,J., *et al.* (2018) C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2 α phosphorylation. *Nat Commun*, **9**, 51.
84. Gao,F.B., Richter,J.D., and Cleveland,D.W. (2017) Rethinking Unconventional Translation in Neurodegeneration. *Cell* , **171**, 994–1000.
85. Linsalata,A.E., He,F., Malik,A.M., Glineburg,M.R., Green,K.M., Natla,S., Flores,B.N., Krans,A., Archbold,H.C., Fedak,S.J., *et al.* (2019) DDX 3X and specific initiation factors modulate FMR 1 repeat-associated non-AUG-initiated translation. *EMBO Rep*, **20**, e47498.
86. Tseng,Y.J., Sandwith,S.N., Green,K.M., Chambers,A.E., Krans,A., Raimer,H.M., Sharlow,M.E., Reisinger,M.A., Richardson,A.E., Routh,E.D., *et al.* (2021) The RNA helicase DHX36–G4R1 modulates C9orf72 GGGGCC hexanucleotide repeat-associated translation. *Journal of Biological Chemistry*, **297**, 100914.
87. Tseng,Y.-J., Krans,A., Malik,I., Deng,X., Yildirim,E., Ovunc,S., Tank,E.M.H., Jansen-West,K., Kaufhold,R., Gomez,N.B., *et al.* (2024) Ribosomal quality control factors inhibit repeat-associated non-AUG translation from GC-rich repeats. *Nucleic Acids Res*, gkae137.
88. Brandman,O. and Hegde,R.S. (2016) Ribosome-associated protein quality control. *Nat Struct Mol Biol*, **23**, 7–15.
89. Joazeiro,C.A.P. (2019) Mechanisms and functions of ribosome-associated protein quality control. *Nat Rev Mol Cell Biol*, **20**, 368–383.
90. Green,K.M., Glineburg,M.R., Kearse,M.G., Flores,B.N., Linsalata,A.E., Fedak,S.J., Goldstrohm,A.C., Barmada,S.J. and Todd,P.K. (2017) RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat Commun*, **8**, 2005.
91. Malik,I., Tseng,Y., Wright,S.E., Zheng,K., Ramaiyer,P., Green,K.M. and Todd,P.K. (2021) SRSF protein kinase 1 modulates RAN translation and suppresses CGG repeat toxicity. *EMBO Mol Med*, **13**, e14163.
92. Ariza,J., Rogers,H., Monterrubio,A., Reyes-Miranda,A., Hagerman,P.J. and Martínez-Cerdeño,V. (2016) A Majority of FXTAS Cases Present with Intranuclear Inclusions Within Purkinje Cells. *Cerebellum*, **15**, 546–551.
93. Iwahashi,C.K., Yasui,D.H., An,H.J., Greco,C.M., Tassone,F., Nannen,K., Babineau,B., Lebrilla,C.B., Hagerman,R.J. and Hagerman,P.J. (2006) Protein composition of the intranuclear inclusions of FXTAS. *Brain*, **129**, 256–271.
94. Buijsen,R.A.M., Sellier,C., Severijnen,L.A.W.F.M., Oulad-Abdelghani,M., Verhagen,R.F.M., Berman,R.F., Charlet-Berguerand,N., Willemsen,R. and Hukema,R.K. (2014) FMRpolyG-positive inclusions in CNS and non-CNS organs of

a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol Commun*, **2**, 162.

95. Cid-Samper,F., Gelabert-Baldrich,M., Lang,B., Lorenzo-Gotor,N., Ponti,R.D., Severijnen,L.A.W.F.M., Bolognesi,B., Gelpi,E., Hukema,R.K., 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.
96. Hoem,G., Larsen,K.B., Øvervatn,A., Brech,A., Lamark,T., Sjøttem,E. and Johansen,T. (2019) The FMRpolyGlycine protein mediates aggregate formation and toxicity independent of the CGG mRNA hairpin in a cellular model for FXTAS. *Front Genet*, **10**, 249.
97. Arocena,D.G., Iwahashi,C.K., Won,N., Beilina,A., Ludwig,A.L., Tassone,F., Schwartz,P.H. and Hagerman,P.J. (2005) Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. *Hum Mol Genet*, **14**, 3661–3671.
98. Oh,S.Y., He,F., Krans,A., Frazer,M., Taylor Paul,J., Paulson,H.L. and Todd,P.K. (2015) RAN translation at CGG repeats induces ubiquitin proteasome system impairment in models of fragile X-associated tremor ataxia syndrome. *Hum Mol Genet*, **24**, 4317–4326.
99. Friedman-Gohas,M., Elizur,S.E., Dratviman-Storobinsky,O., Aizer,A., Haas,J., Raanani,H., Orvieto,R. and Cohen,Y. (2020) FMRpolyG accumulates in FMR1 premutation granulosa cells. *J Ovarian Res*, **13**, 22.
100. Haify,S.N., Mankoe,R.S.D., Boumeester,V., van der Toorn,E.C., Verhagen,R.F.M., Willemsen,R., Hukema,R.K. and Bosman,L.W.J. (2020) Lack of a Clear Behavioral Phenotype in an Inducible FXTAS Mouse Model Despite the Presence of Neuronal FMRpolyG-Positive Aggregates. *Front Mol Biosci*, **7**, 599101.
101. Boivin,M., Willemsen,R., Hukema,R.K. and Sellier,C. (2018) Potential pathogenic mechanisms underlying Fragile X Tremor Ataxia Syndrome: RAN translation and/or RNA gain-of-function? *Eur J Med Genet*, **61**, 674–679.
102. Genuth,N.R. and Barna,M. (2018) The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol Cell*, **71**, 364-374.
103. Xue,S. and Barna,M. (2012) Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nat Rev Mol Cell Biol*, **13**, 355-369.
104. Shi,Z., Fujii,K., Kovary,K.M., Genuth,N.R., Röst,H.L., Teruel,M.N. and Barna,M. (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol Cell*, **67**, 71-83.

105. Ferretti,M.B., Ghalei,H., Ward,E.A., Potts,E.L. and Karbstein,K. (2017) Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements. *Nat Struct Mol Biol*, **24**, 700–707.
106. Yang,Y.M. and 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**, 8.
107. Shi,Z. and 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.
108. Norris,K., Hopes,T. and Aspden,J.L. (2021) Ribosome heterogeneity and specialization in development. *Wiley Interdiscip Rev RNA*, **12**, e1644.
109. Kondrashov,N., Pusic,A., Stumpf,C.R., Shimizu,K., Hsieh,A.C., Xue,S., Ishijima,J., Shiroishi,T. and Barna,M. (2011) Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell*, **145**, 383–397.
110. Xue,S., Tian,S., Fujii,K., Kladwang,W., Das,R. and Barna,M. (2014) RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature*, **517**, 33–38.
111. Genuth,N.R., Shi,Z., Kunimoto,K., Hung,V., Xu,A.F., Kerr,C.H., Tiu,G.C., Oses-Prieto,J.A., Salomon-Shulman,R.E.A., Axelrod,J.D., *et al.* (2022) A stem cell roadmap of ribosome heterogeneity reveals a function for RPL10A in mesoderm production. *Nature Commun*, **13**, 5491.
112. Kearse,M.G., Chen,A.S. and Ware,V.C. (2011) Expression of ribosomal protein L22e family members in *Drosophila melanogaster*: rpL22-like is differentially expressed and alternatively spliced. *Nucleic Acids Res*, **39**, 2701–2716.
113. Mills,E.W. and Green,R. (2017) Ribosomopathies: There's strength in numbers. *Science*, **358**, 6363.
114. Doherty,L., Sheen,M.R., Vlachos,A., Choessel,V., O'Donohue,M.F., Clinton,C., Schneider,H.E., Sieff,C.A., Newburger,P.E., Ball,S.E., *et al.* (2010) Ribosomal protein genes RPS10 and RPS26 are commonly mutated in Diamond-Blackfan anemia. *Am J Hum Genet*, **86**, 222–228.
115. Li,J., Su,Y., Chen,L., Lin,Y. and 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.
116. Piantanida,N., La Vecchia,M., Sculco,M., Talmon,M., Palattella,G., Kurita,R., Nakamura,Y., Ronchi,A.E., Dianzani,I., Ellis,S.R., *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.
117. Yamada,S.B., Gendron,T.F., Niccoli,T., Genuth,N.R., Grosely,R., Shi,Y., Glaria,I., Kramer,N.J., 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.
118. Pisarev,A. V., Kolupaeva,V.G., Yusupov,M.M., Hellen,C.U.T. and Pestova,T. V. (2008) Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO Journal*, **27**, 1609–1621.
 119. Havkin-Solomon,T., Fraticelli,D., Bahat,A., Hayat,D., Reuven,N., Shaul,Y. and 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.
 120. Varnavides,G., Madern,M., Anrather,D., Hartl,N., Reiter,W. and Hartl,M. (2022) In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. *J Proteome Res*, **21**, 2397–2411.
 121. Hughes,C.S., Moggridge,S., Müller,T., Sorensen,P.H., Morin,G.B. and Krijgsveld,J. (2019) Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc*, **14**, 68–85.
 122. Viphakone,N., Hautbergue,G.M., Walsh,M., Chang,C. Te, Holland,A., Folco,E.G., Reed,R. and Wilson,S.A. (2012) TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nature Commun*, **3**, 1006.
 123. Masuda,S., Das,R., Cheng,H., Hurt,E., Dorman,N. and Reed,R. (2005) Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev*, **19**, 1512–1517.
 124. Shi,M., Zhang,H., Wu,X., He,Z., Wang,L., Yin,S., Tian,B., Li,G. and Cheng,H. (2017) ALYREF mainly binds to the 5' and the 3' regions of the mRNA in vivo. *Nucleic Acids Res*, **45**, 9640–9653.
 125. Hautbergue,G.M., Castelli,L.M., Ferraiuolo,L., Sanchez-Martinez,A., Cooper-Knock,J., Higginbottom,A., Lin,Y.H., Bauer,C.S., Dodd,J.E., Myszczyńska,M.A., *et al.* (2017) SRSF1-dependent nuclear export inhibition of C9ORF72 repeat transcripts prevents neurodegeneration and associated motor deficits. *Nat Commun*, **8**, 16063.
 126. Rosario,R., Stewart,H.L., Choudhury,N.R., Michlewski,G., Charlet-Berguerand,N. and Anderson,R.A. (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**, e22612.
 127. Argaud,D., Boulanger,M.C., Chignon,A., Mkannez,G. and Mathieu,P. (2019) Enhancer-mediated enrichment of interacting JMJD3-DDX21 to ENPP2 locus prevents R-loop formation and promotes transcription. *Nucleic Acids Res*, **47**, 8424–8438.

128. Song,C., Hotz-Wagenblatt,A., Voit,R. and Grummt,I. (2017) SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev*, **31**, 1370–1381.
129. Sidorov,M.S., Auerbach,B.D. and Bear,M.F. (2013) Fragile X mental retardation protein and synaptic plasticity. *Mol Brain*, **6**, 15.
130. Jin,P., Duan,R., Qurashi,A., Qin,Y., Tian,D., Rosser,T.C., Liu,H., Feng,Y. and Warren,S.T. (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.
131. Xiao,W., Adhikari,S., Dahal,U., Chen,Y.S., Hao,Y.J., Sun,B.F., Sun,H.Y., Li,A., Ping,X.L., Lai,W.Y., *et al.* (2016) Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell*, **61**, 507–519.
132. Roundtree,I.A., Luo,G.Z., Zhang,Z., Wang,X., Zhou,T., Cui,Y., Sha,J., Huang,X., Guerrero,L., Xie,P., *et al.* (2017) YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife*, **6**, e31311.
133. Walter,B.L., Parsley,T.B., Ehrenfeld,E. and Semler,B.L. (2002) Distinct poly(rC) binding protein KH domain determinants for poliovirus translation initiation and viral RNA replication. *J Virol*, **76**, 12008–12022.
134. Gu,H., Yang,J., Zhang,J., Song,Y., Zhang,Y., Xu,P., Zhu,Y., Wang,L., Zhang,P., Li,L., *et al.* (2022) PCBP2 maintains antiviral signaling homeostasis by regulating cGAS enzymatic activity via antagonizing its condensation. *Nat Commun*, **13**, 1564.
135. You,F., Sun,H., Zhou,X., Sun,W., Liang,S., Zhai,Z. and Jiang,Z. (2009) PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase AIP4. *Nat Immunol*, **10**, 1300–1308.
136. Nishiyama,T., Yamamoto,H., Uchiumi,T. and Nakashima,N. (2007) Eukaryotic ribosomal protein RPS25 interacts with the conserved loop region in a dicistroviral intergenic internal ribosome entry site. *Nucleic Acids Res*, **35**, 1514–1521.
137. Fuchs,G., Petrov,A.N., Marceau,C.D., Popov,L.M., Chen,J., OLeary,S.E., Wang,R., Carette,J.E., Sarnow,P. and Puglisi,J.D. (2015) Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proc Natl Acad Sci U S A*, **112**, 319–325.
138. Hertz,M.I., Landry,D.M., Willis,A.E., Luo,G. and Thompson,S.R. (2013) Ribosomal Protein S25 Dependency Reveals a Common Mechanism for Diverse Internal Ribosome Entry Sites and Ribosome Shunting. *Mol Cell Biol*, **33**, 1016–1026.
139. Garcin,E.D. (2019) GAPDH as a model non-canonical AU-rich RNA binding protein. *Semin Cell Dev Biol*, **86**, 162–173.
140. Argaud,D., Boulanger,M.C., Chignon,A., Mkannez,G. and Mathieu,P. (2019) Enhancer-mediated enrichment of interacting JMJD3-DDX21 to ENPP2 locus

- prevents R-loop formation and promotes transcription. *Nucleic Acids Res*, **47**, 8424–8438.
141. Osinalde,N., Olea,M., Mitxelena,J., Aloria,K., Rodriguez,J.A., Fullaondo,A., Arizmendi,J.M. and Zubiaga,A.M. (2013) The Nuclear Protein ALY Binds to and Modulates the Activity of Transcription Factor E2F2. *Mol Cell Proteomics*, **12**, 1087-98.
 142. Nagy,Z., Seneviratne,J.A., Kanikevich,M., Chang,W., Mayoh,C., Venkat,P., Du,Y., Jiang,C., Salib,A., Koach,J., *et al.* (2021) An ALYREF-MYCN coactivator complex drives neuroblastoma tumorigenesis through effects on USP3 and MYCN stability. *Nat Commun*, **12**, 1881.
 143. Klec,C., Knutsen,E., Schwarzenbacher,D., Jonas,K., Pasculli,B., Heitzer,E., Rinner,B., Krajina,K., Prinz,F., Gottschalk,B., *et al.* (2022) ALYREF, a novel factor involved in breast carcinogenesis, acts through transcriptional and post-transcriptional mechanisms selectively regulating the short NEAT1 isoform. *Cell Mol Life Sci*, **79**, 391.
 144. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, **26**, 1367–1372.
 145. Tyanova,S., Temu,T., Sinitcyn,P., Carlson,A., Hein,M.Y., Geiger,T., Mann,M. and Cox,J. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, **13**, 731–740.
 146. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
 147. Raudvere,U., Kolberg,L., Kuzmin,I., Arak,T., Adler,P., Peterson,H. and Vilo,J. (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res*, **47**, 191–198.
 148. Hulsen,T., de Vlieg,J. and Alkema,W. (2008) BioVenn - A web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, **9**, 488.

6. CONCLUDING REMARKS

The primary aim of the PhD dissertation was to identify novel factors of CGG repeat expansion-related RAN translation. This goal was addressed by applying proteomic screening using two different sample preparation protocols prior MS analysis. Such approach resulted in the **identification of 67 proteins binding natively to toxic *FMR1* mRNA** with expanded CGG repeats, which were present in both MS analyses. These identified proteins serve as a valuable resource for understanding the interactome of mutant *FMR1* and provide a basis for future studies underlying the biology of this molecule. Verification of RAN translation regulatory properties based on siRNA silencing of ten selected candidates resulted in the **characterization of four novel RAN translation modifiers: RPS26, DHX15, DDX21, and ALYREF**. In addition, the fifth one, **RPS25 was shown to modulate CGG-related RAN translation**, although it was not initially identified in MS screening. Notably, depleting the majority of the aforementioned factors did not affect the mRNA encoding FMRpolyG and FMRP levels, indicating a specific regulation of translation associated with the FMRpolyG open reading frame.

The newly discovered role of RPS25 and RPS26 in regulating RAN translation presents an intriguing **perspective that 40S subunit composition affects the noncanonical protein synthesis of FMRpolyG**. This view sheds light on a previously unexplored mechanistic aspect of CGG repeat-related RAN translation. Furthermore, it has been shown that the chaperon protein TSR2, associated with RPS26, was involved in RAN translation regulation, possibly through its involvement in mediating the incorporation of RPS26 into the assembling 40S subunit. Given the current knowledge about specialized ribosomes responsible for translating specific pools of mRNAs, and previous reports indicating RPS26-dependent translation, a global analysis was conducted to investigate proteins sensitive to RPS26 insufficiency, revealing a **relatively small fraction of the total proteome altered by RPS26 depletion**. Moreover, a bioinformatic study examining the features of mRNA encoding affected proteins uncovered that **they possess GC-rich 5'UTRs, similar to the one observed in *FMR1***.

Finally, it was demonstrated that **RPS26 depletion reduced toxicity caused by RAN translation in cellular model**. Given that current therapeutic strategies for fragile-X associated conditions primarily target symptoms and lack specific treatments, understanding the roles of RPS26, RPS25 DDX21, DHX15, and ALYREF in regulating RAN translation provides a foundation for future research that could yield targeted therapies aimed at alleviating symptoms and impeding the disease progression.