CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts

Ricardos Tabet1,2, Laure Schaeffer3, Fernande Freyermuth1,2, Melanie Jambeau1,2, Michael Workman1, Chao-Zong Lee1, Chun-Chia Lin1, Jie Jiang4, Karen Jansen-West5, Hussein Abou-Hamdan6, Laurent Désaubry6, Tania Gendron5, Leonard Petrucelli5, Franck Martin3 & Clotilde Lagier-Tourenne1,2

Expansion of G4C2 repeats in the C9ORF72 gene is the most prevalent inherited form of amyotrophic lateral sclerosis and frontotemporal dementia. Expanded transcripts undergo repeat-associated non-AUG (RAN) translation producing dipeptide repeat proteins from all reading frames. We determined cis-factors and trans-factors influencing translation of the human C9ORF72 transcripts. G4C2 translation operates through a 5′–3′ cap-dependent scanning mechanism, requiring a CUG codon located upstream of the repeats and an initiator Met-tRNAMet. Production of poly-GA, poly-GP, and poly-GR proteins from the three frames is influenced by mutation of the same CUG start codon supporting a frameshifting mechanism. RAN translation is also regulated by an upstream open reading frame (uORF) present in mis-spliced C9ORF72 transcripts. Inhibitors of the pre-initiation ribosomal complex and RNA antisense oligonucleotides selectively targeting the 5′-flanking G4C2 sequence block ribosomal scanning and prevent translation. Finally, we identified an unexpected affinity of expanded transcripts for the ribosomal subunits independently from translation.
myotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are devastating neurodegenerative disorders with a considerable clinical and pathological overlap, which is further substantiated by the discovery of C9ORF72 repeat expansions as the most frequent genetic cause for both diseases. Indeed, expansion of a G₄C₂ hexanucleotide repeat in the first intron of the C9ORF72 gene is identified in ~40% and ~25% of familial ALS and FTD, respectively, as well as 5% of sporadic patients. The number of G₄C₂ repeats is normally lower than 30 and can extend to several hundred repeats in patients. As in other microsatellite diseases, C9ORF72 expansions are transcribed from both sense and antisense strands (reviewed in ref. 4). Bidirectional transcription of the C9ORF72 locus results in the production of transcripts containing either G₄C₂ or G₂C₄ repeats that accumulate into RNA foci. The G₄C₂-containing RNAs were proposed to form G-quadruplex secondary structures and sequester several RNA-binding proteins (RBPs) including hnRNP H1/F, ALYREF, SRSF2, hnRNPA1, hnRNPA3, ADARB2, Pur-α, and Nucleolin (reviewed in ref. 4). In addition, C9ORF72 expanded transcripts are Attilated into dipeptide repeat (DRP) proteins through unconventional translation, known as repeat-associated non-AUG (RAN) translation. RAN translation occurs in absence of an AUG start codon, in multiple reading frames of the same repeat-containing transcript, and within coding as well as non-coding regions. This mechanism has now been described in several microsatellite expansion diseases, including spinocerebellar ataxia type 8 (SCA8), myotonic dystrophy (DM1 and DM2), Huntington’s disease (HD), fragile X-associated tremor/ataxia syndrome (FXTAS), spinocerebellar ataxia type 31, and C9ORF72 ALS/FTD. Both G₄C₂ sense and G₂C₄ antisense transcripts are translated from the three coding frames into five DPR proteins, which aggregate in C9ORF72 ALS/FTD patients. Poly-Glycine-Alanine (poly-GA) and poly-Glycine-Arginine (poly-GR) are translated from the sense strand G₂C₄ transcripts, while poly-Proline-Alanine (Poly-PA) and poly-Proline-Arginine (poly-PR) are produced from the antisense strand G₄C₂ RNA. Poly-Glycine-Proline (poly-GP) may be produced from both RNA strands. These DPR proteins are the main components of cytoplasmic poly-GA and poly-GR, and toxic and play a central role in neurodegeneration due to C9ORF72 expansions (reviewed in ref. 24).

However, how RAN translation of C9ORF72 expanded transcripts occurs and which factors are required is unknown. Translation initiation of canonical mRNAs is a complex process that requires numerous eukaryotic initiation factors (eIFs) and is crucial for regulation of gene expression. The 40S ribosomal subunit binds to the 5' cap and then scans along the mRNA until encountering an initiation codon. Most of the regulation is exerted at the first stage, where the AUG start codon is identified and decoded by the methionyl-tRNA specialized for initiation (Met-tRNA⁰⁰). The efficiency of start codon selection is strongly influenced by surrounding sequences and the recruitment of eIFs. Certain viral and cellular messenger RNAs escape the canonical translation pathway to attract the ribosomes in a cap-independent scanning mechanism. These RNAs contain highly structured sequence, called internal ribosome entry site (IRES), mimicking initiation factors to directly recruit the ribosome at the start codon. Repeat-containing RNAs may also adopt stable structures, such as stem loops or G-quadruplexes and an IRES-like mechanism could be at the origin of RAN translation in microsatellite expansion diseases. Against this hypothesis, RAN translation of CGG repeats associated with FXTAS was recently shown to involve a canonical cap-dependent scanning mechanism. The cis-factors and trans-factors influencing the translation of the human C9ORF72 expansion transcripts are not yet identified. Determining whether hexanucleotide G₄C₂ transcripts recruit the ribosome following the canonical translation initiation or using an IRES mechanism is a crucial step for the development of therapeutic approaches targeting RAN translation in C9ORF72 ALS/FTD patients.

Herein, we provide mechanistic insights delineating the different steps needed to recruit the ribosome and initiate RAN translation from G₄C₂ expansions to produce poly-GA, GA, and GR proteins. Similar to a canonical mechanism of translation, the production of DPR proteins from expanded transcripts requires a 5' cap insertion, involves the initiator methionine and strongly relies on sequences upstream of the repeat. G₂C₄ RAN translation proceeds by a 5'-3' canonical scanning mechanism to start translation at a near-cognate CUG codon and produce DPR proteins by frameshifting. Consistent with this mechanism, we also demonstrate that G₂C₄ RAN translation is downregulated by upstream open reading frame (uORF) present in abnormally spliced C9ORF72 transcripts. Inhibitors of the pre-initiation ribosomal complex and RNA anticodon oligonucleotides (ASOs) targeting the sequence upstream of the repeats inhibit G₂C₄ RAN translation, confirming a scanning-dependent mechanism that may be targeted for therapeutic intervention. Finally, G₄C₂-containing RNAs are found to be associated with ribosomal subunits in a translation independent manner supporting a new RNA gain of function mechanism in C9ORF72 disease.

Results
Translation efficiency of G₄C₂ RAN translation. To identify cis-factors and trans-factors influencing the translation of G₂C₄ repeats in the context of the C9ORF72 gene, we used a construct containing 66 repeats that was shown to undergo RAN translation in all three frames when expressed in cultured cells and in the mouse central nervous system. This construct was modified to generate a series of vectors with different sequences flanking the repeat at the human C9ORF72 locus (Supplementary Fig. 1 and Table 1). Sequences encoding for a specific tag in each of the three reading frames were inserted downstream of the repeat to monitor the production of poly-GA (HA in the +1 frame), poly-GR (His in the +1 frame), and poly-PR (FLAG in the +1 frame). RAN translation from all three reading frames is therefore examined from the same G₂C₄ construct.

A cell-free translation assay based on rabbit reticulocyte lysates (RRL) was developed to monitor RAN translation efficiency from C9ORF72 transcripts. In vitro RAN translation was observed in all three frames from capped RNAs with 66 repeats (Fig. 1a–c, Supplementary Fig. 2a). To accurately compare the translation efficiency of the repeat in each frame, we used as reference Renilla luciferases with either HA, His, or FLAG tags under the control of the intergenic region (IGR) IRESs from cricket paralysis virus (CrPV). IRES are structural RNA elements that allow translation in a translation-independent manner supporting a new RNA gain of function mechanism in C9ORF72 disease.

A cell-free translation assay based on rabbit reticulocyte lysates (RRL) was developed to monitor RAN translation efficiency from C9ORF72 transcripts. In vitro RAN translation was observed in all three frames from capped RNAs with 66 repeats (Fig. 1a–c, Supplementary Fig. 2a). To accurately compare the translation efficiency of the repeat in each frame, we used as reference Renilla luciferases with either HA, His, or FLAG tags under the control of the intergenic region (IGR) IRESs from cricket paralysis virus (CrPV). IRES are structural RNA elements that allow translation in a translation-independent manner supporting a new RNA gain of function mechanism in C9ORF72 disease.

A cell-free translation assay based on rabbit reticulocyte lysates (RRL) was developed to monitor RAN translation efficiency from C9ORF72 transcripts. In vitro RAN translation was observed in all three frames from capped RNAs with 66 repeats (Fig. 1a–c, Supplementary Fig. 2a). To accurately compare the translation efficiency of the repeat in each frame, we used as reference Renilla luciferases with either HA, His, or FLAG tags under the control of the intergenic region (IGR) IRESs from cricket paralysis virus (CrPV). IRES are structural RNA elements that allow translation in a translation-independent manner supporting a new RNA gain of function mechanism in C9ORF72 disease.
GA frame was 17 times more efficient than the IGR-luciferase reporter (Fig. 1a, d). In contrast, translation efficiency from poly-GP in the +2 frame and poly-GR in the +3 frame was equivalent to the translation of IGR-luciferase (Fig. 1b–d). Notably, poly-GA aggregates are the most prevalent DPR proteins accumulated in post-mortem brain samples from C9ORF72 ALS/FTD patients (Supplementary Fig. 3)\(^1\),\(^7\),\(^4\) supporting that translation of the C9ORF72 repeat is most efficient in the +1 frame both in vitro and in vivo.

We also uncovered that the size of the expansion does not equally influence translation of the different frames. Production of poly-GP in the +2 frame was strongly influenced by the size of the repeat when comparing 30 and 66 repeats (Fig. 1b, f, Supplementary Fig. 1; constructs #3 vs. #4). In contrast, no significant difference was observed for poly-GA or poly-GR, which were equally expressed from both 30 and 66 G4C2 repeats (Fig. 1a, c, e, g).

**Cap-dependent G4C2 translation initiates with methionine.**

Our in vitro assay provided the opportunity to determine whether RAN translation of the C9ORF72 repeat depends on the presence of a 5’ m7G cap. Levels of poly-GA produced from 66 repeats increased more than five times when transcripts were capped (Fig. 1a, e) and poly-GP/GR syntheses were strongly repressed in absence of the cap, supporting a canonical cap-dependent mechanism of translation for all three DPR proteins (Fig. 1a, b, c, e, f, g).

---

**Fig. 1** G4C2 RAN translation is length dependent and displays different efficiencies across reading frames. RNA transcripts with (G4C2)\(^{30}\) or (G4C2)\(^{66}\) repeats were transcribed in vitro with T7 RNA polymerase, capped or not capped and subjected to translation in rabbit reticulocyte lysate (RRL) system. Increasing RNA concentrations (100 and 200 nM) were used for translation in RRL. RAN translation was probed on immunoblot with antibodies to (a) HA tag in the +1 poly-GA frame, (b) His tag in the +2 poly-GP frame, and (c) FLAG tag in the + 3 poly-GR frame. Schematics of constructs with 30 repeats (#3) and 66 repeats (#4) are shown in Figure S1. (d) Efficiencies of RAN translation in the different frames were measured relatively to Renilla Luciferase with the corresponding tags driven by the intergenic region (IGR) IRES from the cricket paralysis virus. The efficiencies of RAN translation from capped RNAs were compared to uncapped RNAs at 100 nM for (e) poly-GA, (f) poly-GP, (g) poly-GR with 30 or 66 G4C2 repeats, relatively to the capped 66 repeats. Graphs represent mean ± SEM, n = 3. Student’s t-test, *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001.
Canonical translational initiation consists of base-pairing between the initiator Met-tRNA\textsuperscript{Met} anticodon and the AUG start codon. The incorporation of [\textsuperscript{35}S]-methionine during the translation of transcripts expressing 30 repeats (#3) or 66 repeats (#4) was measured to determine whether RAN translation requires Met-tRNA\textsuperscript{Met} for the production of DPR proteins (Fig. 2a, b). Notably, the sequence of the transcripts #3 and #4 do not contain any AUG codon and the presence of [\textsuperscript{35}S]-methionine in RAN products cannot derive from the incorporation of an internal methionine (Supplementary Fig. 1 and Table 1). A specific [\textsuperscript{35}S]-methionine band was detected at the expected 14.5 and 20.5 kDa molecular weight from constructs #3 and #4, respectively (Fig. 2b). The level of [\textsuperscript{35}S]-methionine labeled polypeptide(s) was proportional to RNA concentration indicating that RAN translation is observed in sub-saturating conditions. Immunoprecipitation of poly-GA products with a HA-specific antibody confirmed that RAN translation initiates with the incorporation of a methionine residue (Fig. 2c).

To further demonstrate that G\textsubscript{4}C\textsubscript{2} RAN translation starts with a methionine, we inhibited the activity of the methionylated initiator tRNA\textsuperscript{Met} carrier eIF2 by inducing the phosphorylation of its α subunit with poly(I:C)/salubrinal treatment as previously described\textsuperscript{41} (Supplementary Fig. 4a). While this treatment did not have any impact on the non-canonical translation of IGR-renilla luciferase, it inhibited the translation of a capped-dependent renilla luciferase reporter and the incorporation of [\textsuperscript{35}S]-methionine in DPR products (Supplementary Fig. 4b).

To gain further insights in the mechanism of RAN translation, we investigated the requirement of eukaryotic initiation factor eIF4E. In canonical translation the cap binding protein eIF4E is part of a larger complex called eIF4F, which also contains the platform protein eIF4G and the RNA helicase eIF4A\textsuperscript{34} (Fig. 2d). To test whether eIF4E is involved in the RAN translation of G\textsubscript{4}C\textsubscript{2} repeats we monitored the translation efficiency in the presence of an excess of cap analog (m\textsuperscript{7}GpppG) or its non-functional ApppG counterpart. The competition assay was performed in RRL (Fig. 2e, f) and wheat germ extract (WGE) (Supplementary Fig. 4c, d), a highly cap-dependent system\textsuperscript{42}. Increasing...
5 G4C2 RAN translation of all reading frames initiates at the same near-cognate CUG start codon in RRL. (a) Scheme of the pre-initiation complex loaded at the 5′ cap and the 4OS ribosomal subunit ready to scan toward the start codon. RAN translation occurs in absence of an AUG codon. (b) Schemes of the transcripts #4 to #11 showing mutations in the 5′ flanking sequence of (G4C2)66 used in (c-f). Construct #4 contains the native sequence of 113 nucleotides upstream of the G4C2 repeat. Construct #5 has a CUG → CGG mutation (blue nucleotide) in a near-cognate start codon located in a perfect Kozak sequence 24 nucleotides upstream of the repeat. Construct #6 has CUG → CGG and GAG → GGG (blue nucleotides) mutations in two potential start codons located 24 and 13 nucleotides upstream of the repeat, respectively. Construct #7 contains a GAG → GGG mutation in a potential near-cognate start codon located 13 nucleotides upstream of the repeat. Constructs #8 and #9 harbor a deletion leaving 33 nucleotides (including CUG and GAG codons) and eight nucleotides (deleting both CUG and GAG codons) upstream of the repeat, respectively. Construct #10 has a CUG → AUG mutation in the near-cognate start codon. Construct #11 has GCCUGG > UCUCUG mutations in the Kozak sequence. (c-f) Translation was performed in presence of [35S]-methionine using each RNA variant separately (#4 to #11). (c) and (e) [35S]-methionine RAN translation products were detected by autoradiography. (d and f) Poly-GA, poly-GP, and poly-GR were detected by immunoblot using antibodies against HA-tag, His-tag, and FLAG-tag, respectively. Asterisk indicates unspecific proteins translated in the RRL system.

G4C2 translation initiation at a near-cognate CUG start codon.

We next sought to identify the candidate(s) used to initiate translation of C9ORF72 transcripts (Fig. 3a). The presence of a single band on SDS-PAGE for the different DPR products (Fig. 1) and poly-GA accumulation (Fig. 2f; Supplementary Fig. 4d) were reduced by increased concentrations of cap analog, demonstrating the role of the canonical initiation factor eIF4E in C9ORF72 RAN translation.

Concentrations of cap analog, but not ApppG, lead to elf4E titration thereby affecting the efficiency of elf4F-dependent translation. The levels of [35S]-methionine-DPRs (Fig. 2c, Supplementary Fig. 4c) were reduced by increased concentrations of cap analog, demonstrating the role of the canonical initiation factor elf4E in C9ORF72 RAN translation.

Identification of translation initiation codon.

A candidate start site is a near-cognate CUG codon located 24 nucleotides upstream of the repeats in the +1 frame and embedded in a perfect Kozak sequence (G/A in −3 and G in +4) (Supplementary Fig. 1, Fig. 3b, and Supplementary Table 1). Site-directed mutagenesis of this codon from CUG to CCG was sufficient to abolish the production of [35S]-methionine labeled DPR proteins in RRL, demonstrating that this CUG is used as the start codon to translate C9ORF72 G4C2 repeats (Fig. 3b, c; construct #4 vs. #5). In contrast, a point mutation from GAG to GGG in another putative start site located 13 nucleotides upstream of the repeats in the +2 frame only slightly reduced the level of [35S]-methionine DPR proteins (Fig. 3b, c; construct #4 vs. #7). Transcripts containing mutations at both putative start codons confirmed the necessity of the CUG codon to initiate RAN translation of the C9ORF72 repeat (Fig. 3b, c; construct #4 vs. #6). This was further corroborated by using constructs with 5′ truncations either preserving the near cognate CUG codon (#8) or deleting the entire region (#9) (Fig. 3b, c; construct #8 vs. #9). Importantly, immunoblot analyses revealed that syntheses of all three DPRs, poly-GA, poly-GP, and poly-GR, are equally disabled by the CUG mutation located 24 nucleotides upstream of the repeats (Fig. 3d; construct #4 vs. #5). In contrast, mutation of the GAG codon located 13 nucleotides upstream of the repeats reduced the levels of the three DPRs without abolishing their production (Fig. 3d; construct #4 vs. #7). Site-directed mutagenesis of the near cognate CUG codon to a canonical start codon AUG increases the incorporation of [35S]-methionine in DPR products (Fig. 3b, e; construct #10 vs. #4) and concomitantly the level of DPRs from all three frames, poly-GA, poly-GP, and poly-GR (Fig. 3b, f; construct #10 vs. #4). Interestingly, mutating the Kozak sequence inhibits the production of DPR proteins detected by [35S]-methionine autoradiography (Fig. 3b, e; construct #11 vs. #4).
An uORF represses G\textsubscript{4}C\textsubscript{2} translation. Recently, Niblock et al. identified poly-adenylated C9ORF72 RNA species that retain the repeat-containing intron 1 and in which downstream exons are correctly spliced\textsuperscript{33}. This finding opens the possibility that G\textsubscript{4}C\textsubscript{2} RAN translation occurs from a C9ORF72 mRNA variant with an enlarged 5′-untranslated region containing the G\textsubscript{4}C\textsubscript{2} repeat sequences. Notably, retention of intron 1 creates a potential uORF with 55 codons flanked by an AUG start codon and two consecutive stop codons (UGA and UAA) (Supplementary Fig. 1 and Table 1). Emerging evidence suggests that the presence of uORF may regulate the expression of downstream ORF\textsuperscript{25,45,46}. Indeed, translation of uORFs located in the 5′UTRs of transcripts often inhibits translation of the downstream ORF likely by reducing its accessibility to the preinitiation complex\textsuperscript{47,48}. Hence, we tested whether the uORF created by the retention of intron 1 in C9ORF72 transcripts may influence RAN translation of DPR proteins (Fig. 5a). We generated a construct with 66 repeats and the entire 5′ end sequence of C9ORF72 starting with exon 1A (Fig. 5b, Supplementary Fig. 1; construct #1). The uORF was found to strongly repress RAN translation in all frames of C9ORF72 repeat. Indeed, [\textsuperscript{35}S]-methionine labeled DPR proteins were not detected in the presence of the uORF (Fig. 5c, construct #1 vs. #4). Immuno blot analysis confirmed the influence of the uORF with a severe reduction of poly-GA (+1 frame) levels and non-detectable poly-GP (+2) and poly-GR (+3) products (Fig. 5d; construct #1 vs. #4). Mutation of the uORF AUG start codon into CGG (Fig. 5b, construct #2) restored G\textsubscript{4}C\textsubscript{2} RAN translation from all reading frames confirming its role in repressing RAN translation (Fig. 5e; construct #2 vs. #1). Overall, these findings strongly support that RAN translation operates through a 5′–3′ scanning mechanism and is regulated by an uORF in C9ORF72 transcripts that retain intron 1.

5′–3′ scanning-dependent mechanism for G\textsubscript{4}C\textsubscript{2} translation. To further demonstrate that RAN translation uses a canonical 5′–3′ scanning mechanism we investigated whether the elf4A, an RNA helicase required for efficient scanning during translation initiation, is involved in G\textsubscript{4}C\textsubscript{2} RAN translation (Fig. 6a–e). FL3, an elf4A-specific inhibitor\textsuperscript{49}, was found to reduce RAN translation in RRL as demonstrated by the levels of [\textsuperscript{35}S]-methionine labeled DPR proteins (Supplementary Fig. 6a; construct #4) and poly-GA (Fig. 6b, Supplementary Fig. 6c; construct #4) generated from two different concentrations of expanded RNAs. Consistently, FL3 treatment significantly reduced the levels of poly-GA, poly-GP, and poly-GR in HEK293T (Fig. 6c, d) without affecting the level of the repeat-containing transcripts (Fig. 6e). To confirm the role of elf4A in 5′–3′ scanning mechanism in RAN translation, we used a longer transcript that includes exon 1a and the entire intronic region upstream of the C9ORF72 repeat with a AUG > CGG mutation in the uORF start codon (Supplementary Fig. 6a; construct #2). Consistent with our previous results, production of DPR proteins was partially restored in presence of the mutated uORF, but was strongly inhibited after treatment with FL3 (Supplementary Fig. 6a–c; construct #2). Another important component for scanning is the platform elf4E, which links the cap binding factor elf4E with the small ribosomal subunit (Fig. 6f). To investigate whether elf4E is required for G\textsubscript{4}C\textsubscript{2} RAN translation we used 4EIRCat, an inhibitor that prevents the direct interaction between elf4E and elf4F\textsuperscript{50}. Consistently, synthesis of poly-GA from two different RNA concentrations was also reduced by 4EIRCat (Fig. 6b). Finally, we found that both edine and cicloheximide completely inhibited the RAN translation from all three reading frames (Fig. 6b, Supplementary Fig. 6b–d). Edine is a translation inhibitor that prevents the interaction of Met-tRNA\textsubscript{Met}\textsuperscript{51} anticodon with the start codon in the P site of the ribosome (Fig. 6a). Cicloheximide binds between the E site and P site of the ribosome and thereby blocks translational elongation to the next codon (Fig. 6a)\textsuperscript{51}. The profound effect of these inhibitors on RAN translation is consistent with our previous results showing that G\textsubscript{4}C\textsubscript{2} RAN translation uses a canonical translation mechanism and initiates at a CUG codon with Met-tRNA\textsubscript{Met}\textsuperscript{51} anticodon interaction in the P site of the ribosome (Figs. 2–4). Overall, the effect of specific translation inhibitors on the production of DPR proteins demonstrate that G\textsubscript{4}C\textsubscript{2} RAN translation requires elf4F components (4E, 4G and 4A) to promote efficient cap-dependent 5′–3′ scanning.

Inhibition of G\textsubscript{4}C\textsubscript{2} translation by RNA ASOs. We previously showed that DNA ASOs targeting sense strand G\textsubscript{4}C\textsubscript{2}-containing transcripts mediate their cleavage through action of the primarily
nuclear enzyme RNase H, reducing the level of RNA foci and DPR proteins in a C9ORF72 transgenic mouse model and patient fibroblasts. To determine whether RNA ASOs targeting the 5’ flanking G₄C₂ sequence can block the scanning of ribosomes and inhibit RAN translation without inducing RNAse-H-dependent degradation, we generated ASOs selectively targeting the region upstream of the repeats and tested their potency in inhibiting G₄C₂ RAN translation in RRL system (Fig. 6f–h). One RNA C9ORF72 ASO (RNA-ASO1) was complementary to a sequence that overlaps the near-cognate CUG codon, and two ASOs (RNA-ASO2, RNA-ASO3) were chosen to cover sequences located at 41 and 82 nucleotides distal from the repeats, respectively (Fig. 6f). Corresponding RNA sense oligonucleotides (RNA-SOs) were used as controls (RNA-SO1, RNA-SO2, and RNA-SO3, Fig. 6f). All three RNA-ASOs induced a dose-dependent reduction of DPR proteins produced from the capped G₄C₂ 66 repeats RNAs.

**Fig. 4** Poly-GA, poly-GP, and poly-GR RAN translation initiate at the near-cognate CUG start codon in cells. (a) Schematic representations of constructs #4 and #5 containing the near cognate start codon CUG or mutant CCG upstream of (G₄C₂)₆₆ repeats. These constructs are driven by a CMV early enhancer/chicken β-actin (CAG) promoter. Human neural progenitor cells (b–c), mouse motor neuron like cells (NSC-34) (d–e) and human HEK293T cells (f–g) were co-transfected with the constructs #4 or #5 along with a GFP plasmid reporter. GFP, Hsp90 or β-Actin proteins were analyzed by immunoblot to control for the transfection efficiency and protein loading. Poly-GA, poly-GP, and poly-GR proteins were identified by immunoblot using antibodies raised against poly-GA, poly-GP, and poly-GR. Levels of the different DPR proteins were quantified and normalized to GFP, HSP90 or β-Actin. Error bars indicate SEM of three independent transfections. Student’s t-test, ** and *** indicate p < 0.01 and p < 0.001, respectively.
**Fig. 5** RAN translation of G\(_4\)C\(_2\) repeats is down-regulated by a short upstream open reading frame (uORF). (a) Retention of intron 1 in C9ORF72 repeat-containing transcripts creates an uORF located 226 nucleotides upstream of the start CUG codon. This uORF may inhibit or enhance G\(_4\)C\(_2\) RAN translation. (b) To interrogate the regulation of RAN translation by this uORF, RNAs harboring the 5′ full-length sequence including C9ORF72 exon 1A (#1) and a AUG > CCG mutation in the uORF start codon (#2) were compared to RAN translation from RNA without the uORF (#4). Black boxes represent exons 1a and 1b and the gray box represents the uORF overlapping exon 1a and intron 1. (c) Translation in RRL system was performed in presence of \(^{35}\)S-methionine and capped RNA (#1, #2, or #4) followed by detection of \(^{35}\)S-methionine proteins by autoradiography. (d) Poly-GA, poly-GP, and poly-GR were detected by immunoblot using antibodies against HA-tag, His-tag, and FLAG-tag, respectively.

as measured by \(^{35}\)S-methionine-labeling (Fig. 6g) and immunoblot (Fig. 6h). In contrast, SO controls did not affect the levels of DPR proteins. These results demonstrate that RNA ASOs targeting the 5′ flanking G\(_4\)C\(_2\) sequence are sufficient to block RAN translation independently of C9ORF72 RNAs degradation and identify the 5′–3′ scanning of ribosomes as a potential therapeutic target in C9ORF72 ALS/FTD.

G\(_4\)C\(_2\) RNAs bind ribosomes independently from translation.

To assess ribosome loading onto (G\(_4\)C\(_2\))\(_{66}\) RNAs, we performed sucrose gradient analysis with radiolabeled capped RNAs containing either 30 or 66 repeats. As a control for canonical translation we used radiolabelled capped human \(\beta\)-globin mRNA. Radiolabeled capped RNAs with 66 antisense G\(_4\)C\(_2\) repeats were also used as control for RAN translation (Fig. 7a). Sucrose gradient analysis with 30 and 66 G\(_4\)C\(_2\) repeat transcripts showed that RNA-containing repeats are mainly associated with polysomes (Fig. 7a, b green graph, Supplementary Fig. 6e orange graph; heavy fractions 0–20). Only a small proportion of RNAs was free (RNP; ribonucleoproteins), associated with the ribosomal small subunit in complex with initiation factors (48S) or with monosomes (80S), which is consistent with active RAN translation (Fig. 1). Since transcripts containing expanded repeats, including G\(_4\)C\(_2\) RNAs, were recently shown to undergo abnormal phase transition and form gel-like structures in vitro\(^{35}\), we determined whether the presence of radiolabeled G\(_4\)C\(_2\) RNAs in the heavy fractions could be due to self-aggregation rather than association with polyribosomes. Against this possibility, G\(_4\)C\(_2\)-free RNAs remained in the light fractions of sucrose gradients strongly supporting that expanded RNAs associate with polyribosomes in RRL. Contrary to the sense (G\(_4\)C\(_2\))\(_{66}\) RNAs, transcripts containing the antisense (C\(_4\)G\(_2\))\(_{66}\) repeat sediments mainly in the light fractions or were associated to monosomes, consistent with a low translation efficiency of the antisense transcripts (Fig. 7a, b; blue graph)\(^{40}\). Unexpectedly, treatment with edeine, that blocks the translation (Fig. 6b) and lead to the accumulation of \(\beta\)-globin mRNA in the light fractions (Fig. 7c, Supplementary Fig. 6f; light fractions 20–40, red graphs), did not prevent loading of polysomes on transcripts with 66 or 30 G\(_4\)C\(_2\) repeats (Fig. 7c, Supplementary Fig. 6f; heavy fractions 0–20, green and orange graphs). The same abnormal sedimentation of G\(_4\)C\(_2\) transcripts in heavy fractions was observed after treatment with GMP-PNP, a non-hydrolysable GTP analog that normally leads to the accumulation of the transcripts in the fraction corresponding to the 48S particles, showing that G\(_4\)C\(_2\) RNAs can recruit ribosomes in a translation-independent manner (Supplementary Fig. 6g).

As expected, blocking ribosomal translocation with cycloheximide induced the accumulation of the control \(\beta\)-globin mRNAs in the fraction corresponding to monosomes 80S that are prevented from translocating after assembly (Fig. 7d, Supplementary Fig. 6h; red graphs). In contrast, inhibiting RAN translation with cycloheximide (Fig. 6b, Supplementary Fig. 6b–d) did not prevent ribosomal loading on expanded transcripts with 30 or 66 repeats (Fig. 7d, Supplementary Fig. 6b; heavy fractions 0–20, green and orange graphs). As expected the 80S peak was slightly increased consistent with a small proportion of expanded G\(_4\)C\(_2\) RNAs being associated with monosomes after cycloheximide treatment, but most transcripts remained present in the heavy fractions despite cycloheximide blockade of translation. Notably, radiolabeled (G\(_4\)C\(_2\))\(_{66}\) transcripts were more abundant in heavy fractions when they were folded in presence of K\(^{+}\) ions that stabilize G-quadruplex structures, comparatively to Na\(^{+}\) and Li\(^{+}\) ions (Supplementary Fig. 6f). Finally, to confirm that G\(_4\)C\(_2\) RNAs recruit
the ribosome independently from DPR translation, we performed sucrose gradient analysis with purified ribosomal 40S and 60S. Expanded transcripts with 30 repeats were able to recruit and load several 40S and 60S ribosomal subunits without the need of 5′-cap and any other initiation factors (Fig. 7e). Overall, we demonstrate here that G_4C_2 repeat-containing transcripts associate with ribosomal subunits independently of translational factors.

**Discussion**

G_4C_2 hexanucleotide expansions in the C9ORF72 gene were recently discovered as the major genetic cause of ALS and FTD, two fatal neurodegenerative disorders. Emerging evidence supports pathogenic RNA gain-of-function mechanisms, where expanded G_4C_2 transcripts form RNA foci sequestering RNA-binding proteins in the nuclei or undergo RAN translation to produce toxic DPR proteins. We developed robust assays to...
study RAN translation and determine specific cis-requirements and trans-requirements for expanded G\textsubscript{4}C\textsubscript{2} translation. G\textsubscript{4}C\textsubscript{2} RAN translation was found to share many aspects with canonical translation initiation, including the requirement of a 5' cap structure, methionylated initiator tRNA\textsubscript{Met}, and the recruitment of the 40S subunit by the eIF4F complex (eIF4A, E, and G) to begin scanning toward the start codon (Fig. 8a, b). These findings are consistent with mechanisms involved in RAN translation of CGG triplet repeats in the fragile X FMR1 gene which also depends on a cap-dependent scanning mechanism\textsuperscript{15,33,34}. Since eIF4F's functions were shown to be critical in dysregulation of the translational machinery in cancers, major efforts have been undertaken to develop specific compounds directed against its components for therapeutic purposes\textsuperscript{35}. Our work highlights the importance of eIF4F in ALS/FTD pathogenesis, thereby opening the potential for new therapeutic strategies using existing eIF4F inhibitors to mitigate the effects of this neurodegenerative disease.

Ribosome profiling on higher eukaryotes showed that translation occurs on numerous ORFs without an AUG-initiator but operates with near-cognate start codons (CUG > GUG > UUG > ACG > others)\textsuperscript{36,37}. We discovered that the CUG codon located 24 nucleotides upstream of the G\textsubscript{4}C\textsubscript{2} repeat, in the +1 (GA) frame and in an optimal Kozak sequence, is utilized as a start codon to produce DPR proteins. Mutations of this CUG codon or the others\textsuperscript{36,37} led by polyribosome fractionation in RRL comparatively to capped β-globin mRNA and (C\textsubscript{4}G\textsubscript{2})\textsubscript{66} antisense RNAs. Fractionation on sucrose gradients was performed without inhibitor (b), in presence of Edelene (c) or CHX (d). (e) Sucrose gradient fractionation of radiolabeled uncapped (G\textsubscript{4}C\textsubscript{2})\textsubscript{30} transcripts (#3) was performed in presence of purified 40S or 60S ribosomal subunits.
GP translation was prevented by mutation of the CUG repeat in RRL and the two neuronal models, poly-GP levels were slightly increased in HEK293T cells supporting a context-dependent regulation that differs between the three frames when the CUG is absent. The presence of specific RNA helicases might explain the differences on poly-GP translation between the different cell types, such as DDX21 recently shown to unfold RNA G-quadruplex structures in HEK293T58. Notably, an UAG stop codon in phase with the poly-GP frame is present at the beginning of the G4C2 repeats (UAG GGG CC sequence, Supplementary Fig. 1), indicating that the ribosome must initiate in another reading frame and then frameshift to produce poly-GP or directly initiate within the repeat. As we observed a single band on SDS-PAGE with comparable molecular weight between all reading frames and systems used (Figs. 3 and 4), initiation further downstream inside the repeats is less likely to occur from (G4C2)66 transcripts. When comparing translation efficiencies for the three reading frames, poly-GA (+1) is predominant, followed by poly-GP (+2) and poly-GR (+3), which is in agreement with a frameshifting model. This is also consistent with staining and immunoassay from human post-mortem tissues, where poly-GA accumulates at higher levels than poly-GP and poly-GR (Supplementary Fig. 3)17,40,59,60.

G4C2 RAN translation initiation is influenced by repeat length, with different sensitivity among the three reading frames. While RAN translation efficiency is only reduced in the +1 poly-GA and +3 poly-GR frames with shorter repeat length, it is completely abolished for poly-GP at 30 comparatively to 66 repeats (Fig. 1). This repeat length dependence could reflect secondary structures, which differentially affect ribosomal scanning, translation...

---

**Fig. 8** Model of translation mechanisms associated with G4C2 expansions in C9ORF72 ALS/FTD. (a) Pre-Initiation ribosomal complex (PIC) assembles on the 5′ cap of mRNA by interacting with eIF4F complex formed by the cap binding factor eIF4E, the platform eIF4G and the RNA helicase eIF4A. The PIC complex scans the 5′ end for an appropriate AUG start codon, where the 60S ribosomal subunit joins the 40S to form a functional 80S ribosome ready to translate the coding sequence. (b) G4C2 RAN translation initiation shares the same pathway as the canonical one to translate poly-GA dipeptides, including the need of 5′ cap, eIF4E, eIF4G, eIF4A, initiator methionyl-tRNA, and the scanning mechanism. However, it initiates on a near-cognate CUG codon embedded in a perfect Kozak sequence, in frame with poly-GA, instead of a canonical AUG start codon. The ability of G4C2 expansions to form stable G-quadruplex structures forces the ribosome to occasionally undergo frameshifting to translate poly-GP and poly-GR in the +2 and +3 frames, respectively. (c) When G4C2 repeats are expanded, a subset of C9ORF72 mRNA is mis-spliced retaining intron 1 with the repeats35. RAN translation from these RNAs is inhibited by a uORF that is translated canonically. (d) G4C2 expanded transcripts associate with ribosomal subunits independently from their translation.
elongation or force the ribosome to undergo a frameshifting. Indeed, G\textsubscript{C}G\textsubscript{C} expansions can adopt RNA G-quadruplexes\textsuperscript{29–32}, a structure that was recently demonstrated to induce frameshifting during translation\textsuperscript{31,62}. These RNA secondary structures are stable in presence of monovalent cations, in the order of K\textsuperscript{+} > Na\textsuperscript{+} > Li\textsuperscript{+}. Thus, variations of ion concentration in the cell or specific RNA binding proteins\textsuperscript{38} may modulate the presence of G-quadruplex structure in G\textsubscript{C}G\textsubscript{C}-containing transcripts and could influence frameshifting or initiation at non-AUG start codon.

Another major finding is the down-regulation of G\textsubscript{C}G\textsubscript{C} RAN translation by a short uORF. Indeed, in mis-spliced C9ORF72 transcripts that retain intron 1, an uORF is present with an AUG and in two in-frame stop codons located 76 nucleotides upstream of the G\textsubscript{C}G\textsubscript{C} repeats. Notably, the AUG codon in exon 1A is the only AUG identified in the 5′ end of the mis-spliced RNA. Upstream ORF is cis-acting elements that regulate the expression of downstream protein coding sequences\textsuperscript{35,46}. We demonstrated that mutating the AUG start codon of the uORF is sufficient to increase G\textsubscript{C}G\textsubscript{C} RAN translation in all three reading frames, confirming that this uORF is efficiently used by the ribosome during 5′–3′ scanning and is therefore inhibiting the translation of the downstream G\textsubscript{C}G\textsubscript{C} repeat (Figs. 5 and 8c). It is noteworthy that translation of synaptic mRNA(s) was shown to be downregulated by uORFs located in their 5′UTR, but upregulated upon metabotropic glutamate receptor activation\textsuperscript{64–66}. Thus, it will be important to determine whether the uORF in mis-spliced C9ORF72 transcripts influences G\textsubscript{C}G\textsubscript{C} RAN translation level upon synaptic activation or external stimuli in neurons.

Notably, ASOs targeting RNase-H-dependent degradation of C9ORF72 transcripts are under therapeutic development\textsuperscript{5,7,52}. The identification of sequences upstream of the repeat that influence RAN translation (CUG near-cognate start codon and uORF) opens the possibility of using alternative strategies based on ASOs that modulate translation without reducing transcript levels\textsuperscript{67,68}. In agreement, we demonstrated that several RNA ASOs specifically targeting the region immediately upstream of the repeats block ribosomal scanning and efficiently reduce the level of RAN translation products (Fig. 6h).

Finally, we show that G\textsubscript{C}G\textsubscript{C} repeat transcripts unexpectedly associate with ribosomal subunits in a translation independent manner (Fig. 8d). Indeed, blocking cap initiation factors, codon–anticodon interaction, 80S ribosome assembly and ribosomal elongation did not avert the sedimentation of radiolabeled G\textsubscript{C}G\textsubscript{C} RNAs in the heavy fractions of sucrose gradients (Fig. 7). In addition, removing the 5′cap, shortening the repeat size, or using purified ribosomal subunits did not prevent the assembly of the transcript to multiple ribosomal subunits. On the contrary, antisense transcripts with G\textsubscript{C}G\textsubscript{C} repeat did not associate with the ribosome. This striking finding supports a RNA gain-of-function mechanism, independent from RAN translation and DPR proteins accumulation. Ribosomal subunits are assembled in the nucleolus and exported to the cytoplasm by multiple export receptors\textsuperscript{69}. It will be important to determine whether sequestration of ribosomal subunits by expanded repeats and disruption of nucleocytoplasmic transport recently identified in C9ORF72 disease\textsuperscript{4} negatively impact overall translation in cells with C9ORF72 expansions.

Overall, we provide new insights into RAN translation of C9ORF72 G\textsubscript{C}G\textsubscript{C} repeat which uses a cap-dependent mechanism initiating at a near-cognate CUG codon. A novel mechanism of toxicity associated to C9ORF72 expansion is supported by the association of G\textsubscript{C}G\textsubscript{C} transcripts with ribosomal subunits independently of their translation. Importantly, this work identifies sequences upstream of the G\textsubscript{C}G\textsubscript{C} repeats and specific initiation factors as possible therapeutic targets to inhibit RAN translation in C9ORF72 ALS/FTD patients.

**Methods**

**Generation of C9ORF72 constructs with G\textsubscript{C}G\textsubscript{C} repeats.** To generate the different constructs listed in Supplementary Fig. 1 and Table 1, a plasmid pAG2 containing 66 repeat\textsuperscript{20,36} was first digested with restriction sites BshHII and SacI to isolate the intrinsic region of human C9ORF72 with (G\textsubscript{C}G\textsubscript{C})\textsubscript{66}, including 8 bp of 5′, 99 bp of 3′ flanking sequences and three tags in frame with DPR proteins. BshHII is a restriction site naturally present in the human C9ORF72 gene located two nucleotides upstream of the repeats. Second, pUC18 (ThermoFisher, # SD0051) was modified to contain the three HindIII, BshHII, and SacI restriction sites, enabling the insertion of the digested BshHII/SacI C9ORF72 insert and the addition of any 5′ end sequence between the HindIII and BshHII sites. After cloning the C9ORF72 insert in modified pUC18 with BshHII and SacI, primers listed in Supplementary Table 2 were used to generate the different constructs listed in Supplementary Fig. 1. Primers were designed to add the 5′ Promoter for in vitro transcription (construct #9), followed immediately by 113 bp of 5′ Banking G\textsubscript{C}G\textsubscript{C} sequence with CUG > CGG mutation (construct #5), GAG > G\textsubscript{C}G mutation (construct #6) and double mutations CUG > CGG + GAG > G\textsubscript{C}G (construct #7). tRNA conjugates were designed to add 17 pr tRNA Untranslated RNAs by 320 bp of 5′ sequence (construct #1), 320 bp with AUG > CGG mutations (construct #2) and to generate a short 5′ end by adding T7 promoter with 33 bp (construct #8). All primers were designed to harbor the HindIII restriction site at the 5′ end and BshHII site at the 5′ end. After phosphorylation with T4 Polynucleotide Kinase (ThermoFisher, #E0031) of the primers at the 5′ end and hybridization of corresponding forward and reverse primers, the generated inserts were cloned in HindIII/BshHII pUC18 with (G\textsubscript{C}G\textsubscript{C})\textsubscript{66} repeats. The original plasmid was modified to contain T7 promoter by cloning using the HindIII restriction site (construct #4). Construct #3 with 30 G\textsubscript{C}G\textsubscript{C} repeats was generated by expansion retraction during amplification of the primer that contains 5′ digest construct #5. Construct #5 was used to generate T7 promoter, CUG > CGG mutation was digested with HindIII and NotI to be cloned in pAG3 downstream of the CMV′ early enhancer/chicken β-actin (CAG) promoter for human cell transfection.

**In vitro transcription.** The different variants of C9ORF72 (G\textsubscript{C}G\textsubscript{C})\textsubscript{66} constructs were cloned downstream of T7 promoter in pUC18 as described in Supplementary Fig. 1 and Table 1. Vectors were digested by Xhol and used for run-off in vitro transcription with T7 RNA polymerase capped with 5′ terminal cap analog. The experiments were performed in the presence of MgCl\textsubscript{2} at a constant [MgCl\textsubscript{2}]/[cap analog] ratio of 0.8. For transcription in presence of RNA sense (RNA- SO) and antisense (RNA-ASO) oligonucleotides (Supplementary Table 3) were annealed to 100 nM capped 66 repeat RNA (construct #4) in 20 mM Hepes-K (pH 7.6) and 100 mM KC\textsubscript{1} for 5 min at 65 °C and 20 min at room temperature with a 10 or 50 fold molar excess of oligonucleotides over construct #4. This annealing mixture was kept on ice before addition to the translation reaction. RRL were incubated 5 min at 30 °C in presence of the different translational inhibitors at the following concentrations: 150 ng mL\textsuperscript{-1} for the polyLC, 15 μM for salubrinal, 4.5 mg mL\textsuperscript{-1} cycloheximide, 10 μM edeine, 15 μM FL3, and 5 μM 4EIRCat.

**Sucrose-gradient analysis.** For sucrose-gradient analysis, 5′-32P-labeled or 3′-32P-labeled mRNA were incubated in RRL or with purified 40S and 60S ribosomal subunits, in the presence of specific inhibitors (Edeine leads to 43S accumulation, GMP-PNP leads to 48S formation, cycloheximide blocks translation and leads to 80S accumulation) or without inhibitor to assemble polysomes. Translational inhibitors were incubated with RRL 5 min prior to addition of radiolabeled RNAs. The translation initiation complexes were separated on a 7–45% linear sucrose gradient in buffer (25 mM Tris–HCl [pH 7.4], 50 mM KC\textsubscript{1}, 5 mM MgCl\textsubscript{2}, 1 mM DTT). The reactions were loaded on the gradients and spun (23,411xg for 2.5 h at 4 °C) in a SW41 rotor. mRNA sedimentation on sucrose gradients was monitored by Cerenkov counting after fractionation. In Supplementary Fig. 5, uncapped RNAs were folded in presence of KCl or LiCL at 195 mM, by denaturing 1 min at 95 °C, followed by 5 min at 20 and 4 °C until adding the RRL (75 mM final ion concentrations).
Cell culture and plasmid transfection. The HEK293T cells were cultured in DMEM (10% v/v) FBS and penicillin/streptomycin. ReNcell VM human neural progenitors (Millipore; Catalog number SCOC008) were maintained in high-glucose DMEM/F12 (Thermo Fisher Scientific) media supplemented with 2 μM Mg²⁺-2H₂O (Sigma-Aldrich, #M6644), 20 μM β-D-glycerolphosphate, 20 μg/mL heparin (StemCell Technologies, #07980), 2% (v/v) B27 neural supplement (Thermo Fisher Scientific, #17500-044), 20 μg/mL hEGF (Sigma-Aldrich, #E5644), 20 μg/mL β-D-Glucuronidase (Stemgent, #03-0002) and 1% penicillin/streptomycin (Thermo Fisher Scientific) and were plated on BD Matrigel (BD Biosciences)-coated cell culture flasks with B27, EGF, FGF, and heparin on precoated Matrigel dishes. The NSC-34 cells (CellLUTION Biosystems Inc; Catalog number—CLU140) were grown in DMEM supplemented with 10% FBS, 100 U·mL⁻¹ penicillin, and 100 μg/mL streptomycin at 37 °C in a humified atmosphere of 5% CO₂. HEK293T were plated 24 h prior transfection with different C9ORF72 (G₄C₃G)₉ expansion constructs (Supplementary Fig.1 and Table 1) and a reporter pGFPmax (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, 2000 reagent was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-44 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005) was used according to manufacturer instruction (Invitrogen) for HEK293T and NSC-34 transfections. Nucleofection using Nucleofector kit (Lonza, #VPG 1005)...
31. Haeuwer, A. R. et al. C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature 507, 195–200 (2014).

32. Condron, E. G. et al. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNPH to disrupt splicing in ALS brains. Elife 5, e17820 (2016).

33. Kearse, M. G. et al. CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. Mol. Cell 63, 314–322 (2016).

34. Hinnebusch, A. G. The scanning mechanism of eukaryotic translation initiation. Annu. Rev. Biochem. 83, 779–812 (2014).

35. Niblock, M. et al. Retention of hexanucleotide repeat-containing intron in C9orf72 mRNA: implications for the pathogenesis of ALS/FTD. Acta Neuropathol. Commun. 4, 18 (2016).

36. Chou, J. et al. Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. Science 348, 1151–1154 (2015).

37. Murray, J. et al. Structural characterization of ribosome recruitment and translocation by type IV IRES. Elife 5, e15367 (2016).

38. Pestova, T. V. & Hellen, C. U. Translation elongation after assembly of ribosomes on the Cricket paralysis virus internal ribosomal entry site without initiation factors or initiator tRNA. Genes Dev. 17, 181–186 (2003).

39. MacKenzie, I. R. et al. Quantitative analysis and clinico-pathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. Acta Neuropathol. 130, 845–861 (2015).

40. Namer, L. S. et al. An ancient pseudoknot in 5TF-a Pre-mRNA activates PKR, inducing eIF2alpha phosphorylation that potently enhances splicing. Cell Rep. 20, 188–200 (2017).

41. Martin, F. et al. Cap-assisted internal initiation of translation of histone H4. Mol. Cell 41, 197–209 (2011).

42. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44, 283–292 (1986).

43. Edwards, J. E. et al. A three-dimensional human neural cell culture model of Alzheimer’s disease. Nature 513, 274–278 (2014).

44. Chev, G. L., Pauli, A. & Schier, A. F. Conservation of uORF repression and sequence features in mouse, human and zebrafish. Nat. Commun. 7, 11663 (2016).

45. Liang, X. H. et al. Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. Nat. Biotechnol. 34, 873–880 (2016).

46. Barbosa, C., Peixiro, I. & Romao, L. Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 9, e1003529 (2013).

47. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc. Natl Acad. Sci. USA 106, 7507–7512 (2009).

48. Boussouar, J. et al. elf4F is a nexus of resistance to anti-3’-BAP and anti-MEK cancer therapies. Nature 513, 105–109 (2014).

49. Robert, F. et al. Translation initiation factor elf4F modifies the dexamethasone response in multiple myeloma. Proc. Natl Acad. Sci. USA 111, 13421–13426 (2014).

50. Garreau de Loubresse, N. et al. Structural basis for the inhibition of the eukaryotic ribosome. Nature 513, 517–522 (2014).

51. Jiang, J. et al. Gain of toxicity from ALS/FTD-linked repeat expansions in huntingtin gene. Nucleic Acids Res. 42, 307–324 (2014).

52. Johansson, H. E., Belsham, G. J., Sproat, B. S. & Bentzen, M. W. Target-specific arrest of mRNA translation by antisense 2′-O-alkylguoribonucleotides. Nucleic Acids Res. 22, 4591–4598 (1994).

53. Kohler, A. & Hurt, E. Exporting RNA from the nucleus to the cytoplasm. Nat. Rev. Mol. Cell Biol. 8, 761–773 (2007).

54. Soto Rojo, R., Ricci, E. P., Decimo, D., Moncorge, O. & Ohlmanna, T. Back to basics: the untreated rabbit reticulocyte lysate as a competitive system to recapitulate cap/poly(A) synergy and the selective advantage of IRES-driven translation. Nucleic Acids Res. 35, e121 (2007).

Acknowledgements

We thank Dr Doo Yeon Kim for his support in cultivating ReNeCl VM. Amelie Lauget, Michael Baughn, Dr Anna-Claire Devlin, and Dr Ying Sun for technical assistance, Dr Gilbert Elman, members of Dr Brian J. Wajing and Dr Mark W. Albers laboratories, Dr. Merit Cudkowicz, Dr Shaying Sun, Dr Raymond Kaemper, Dr Frank Rigo, and Dr Don Cleveland for helpful discussions and continuous support. R.T. was supported by a grant from the Prince Philip Foundation. This work was supported by CNRS, Université de Strasbourg, a grant from the ANR to F.M. (ANR-11-SVS802501), C.L.T was supported by the Department of Neurology at the Massachusetts General Hospital and grants from Target ALS (13-04827) and from NINDS/NIMH (R01NS087827).

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-02643-5.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.