

# Single-molecule imaging reveals distinct elongation and frameshifting dynamics between frames of expanded RNA repeats in C9ORF72-ALS/FTD

Received: 13 November 2022

Accepted: 30 August 2023

Published online: 11 September 2023

 Check for updates

Malgorzata J. Latallo<sup>1,2,9</sup>, Shaopeng Wang<sup>1,2,3,4,9</sup>, Daoyuan Dong<sup>3,4,9</sup>, Blake Nelson<sup>1,2</sup>, Nathan M. Livingston<sup>1,2</sup>, Rong Wu<sup>3,4</sup>, Ning Zhao<sup>5</sup>, Timothy J. Stasevich<sup>5</sup>, Michael C. Bassik<sup>6</sup>, Shuying Sun<sup>2,3,4,7,8</sup> ✉ & Bin Wu<sup>1,2,7</sup> ✉

*C9ORF72* hexanucleotide repeat expansion is the most common genetic cause of both amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). One pathogenic mechanism is the accumulation of toxic dipeptide repeat (DPR) proteins like poly-GA, GP and GR, produced by the noncanonical translation of the expanded RNA repeats. However, how different DPRs are synthesized remains elusive. Here, we use single-molecule imaging techniques to directly measure the translation dynamics of different DPRs. Besides initiation, translation elongation rates vary drastically between different frames, with GP slower than GA and GR the slowest. We directly visualize frameshift events using a two-color single-molecule translation assay. The repeat expansion enhances frameshifting, but the overall frequency is low. There is a higher chance of GR-to-GA shift than in the reversed direction. Finally, the ribosome-associated protein quality control (RQC) factors ZNF598 and Pelota modulate the translation dynamics, and the repeat RNA sequence is important for invoking the RQC pathway. This study reveals that multiple translation steps modulate the final DPR production. Understanding repeat RNA translation is critically important to decipher the DPR-mediated pathogenesis and identify potential therapeutic targets in *C9ORF72*-ALS/FTD.

The most prevalent genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) is the hexanucleotide repeat expansion of GGGGCC in the non-coding region of the first intron of the *C9ORF72* gene<sup>1–3</sup>. The repeat is typically fewer than ten in healthy individuals but can expand up to thousands in *C9ORF72*-ALS/FTD patients<sup>4,5</sup>. The *C9ORF72* repeat expansion may cause RNA-mediated gain-of-toxicity: formation of RNA granules disrupting the RNA processing, as well as production of toxic dipeptide repeat (DPR) proteins<sup>6–9</sup>. The DPRs are synthesized through noncanonical repeat-

associated non-AUG (RAN) translation<sup>10,11</sup>. Due to bidirectional transcription in the *C9ORF72* locus, both sense and antisense RNA repeats exist, which are translated to five different DPRs: GA, GP, GR, PR and PA<sup>10,12,13</sup>. Numerous studies have revealed the biophysical properties and toxicity pathways of different DPRs<sup>14,15</sup>. The arginine-containing DPRs (poly-GR and poly-PR) are the most toxic<sup>14,16–27</sup>, however, poly-GA and poly-GP aggregates are more widely observed in patient post-mortem tissues<sup>13,28</sup>. Chimeric DPR species consisting of fused GA and GP/GR were detected in patients, but it is hard to quantify the relative

A full list of affiliations appears at the end of the paper. ✉ e-mail: [shuying.sun@jhmi.edu](mailto:shuying.sun@jhmi.edu); [bwu20@jhmi.edu](mailto:bwu20@jhmi.edu)

abundance. They can be produced either by frameshift during translation or due to genetic interruptions in the expanded repeats<sup>29–31</sup>. To better understand the relative contribution of the different DPRs to the disease, it is imperative to thoroughly study the DPR biosynthesis and translation dynamics of different reading frames<sup>14</sup>.

Protein synthesis is an elaborate process, subject to tight regulation at each step. Translation is divided into four stages: initiation, elongation, termination, and ribosome recycling<sup>32</sup>. Any one of the steps may contribute to the final protein production. For RAN translation, most studies focused on the initiation, commonly believed to be the rate-limiting step in determining translation efficiency. Several translation initiation factors have been shown to regulate GGGGCC-associated RAN translation, including eIF2A<sup>33</sup>, eIF2D<sup>34</sup>, eIF4B and eIF4H<sup>34,35</sup>. The phosphorylation of eIF2 $\alpha$ , the key factor in the integrated stress response (ISR) pathway that generally reduces the global translation, can significantly enhance RAN translation<sup>36,37</sup>. For the cis-acting elements that regulate RAN translation, it was shown that the sequence preceding the repeats plays an important role in initiation. Particularly, the CUG codon, 24 nucleotides (nt) upstream of the repeats, is a near-cognate start codon used for the initiation in GA frame<sup>37,38</sup>. Therefore, the translation efficiency from the GA frame is higher than GP and GR frames, which could account for the different DPR abundance in patients. However, it remains elusive whether initiation alone can explain different DPR expressions and how translation elongation and frameshift contribute.

Translation efficiency is often measured experimentally as the ribosome occupancy density<sup>39,40</sup>, which is proportional to the initiation rate and inversely related to the elongation rate<sup>41</sup>. Recently, it has been increasingly recognized that translation elongation also plays a regulatory role<sup>42</sup>. On the one hand, optimal codon usage, which presumably promotes faster codon adaptation and elongation, is known to increase protein yield<sup>43–45</sup> and associate with mRNA stability<sup>46</sup>. On the other hand, stalling sequences, such as strong RNA secondary structures or stretches of polybasic amino acids, inhibit ribosome elongation, induce ribosome collisions and invoke protein/RNA quality control pathways to degrade nascent peptides and/or mRNAs<sup>47–49</sup>. This is known as the ribosome-associated protein quality control (RQC)<sup>50</sup>. The GGGGCC repeat is prone to form G-quadruplexes or strong hairpin structures<sup>51,52</sup>, which could potentially inhibit ribosome elongation. Poly-GR is highly positively charged, and poly-GP is proline-rich and may induce steric hindrance. These DPRs are strong candidates for ribosome stalling<sup>53</sup> and targets for RQC<sup>54,55</sup>. How they influence translation elongation dynamics is largely unknown.

Traditionally, DPR biosynthesis is studied through ensemble methods to measure the final protein level, which depends on the RNA transcription, processing, export, translation, degradation, and protein stability. The final protein amount may not precisely represent the translation efficiency. Here, we employed the SunTag-based single-molecule translation imaging approach developed by us and other laboratories<sup>56–60</sup> to address the mechanistic questions of DPR translation that are difficult to address with ensemble methods. The single-molecule imaging technique directly measures the temporal translation dynamics of single mRNAs in live cells. It can reveal the molecular heterogeneity hidden in the ensemble measurements. We constructed reporters to measure initiation, elongation, and frameshifts for different DPR frames. We mutated the CUG near-cognate start codon to examine its influence on the initiation efficiency in each frame. We generated a two-color translation reporter to observe frameshifting events on single RNAs. We measured the elongation speed in different frames and characterized its dependence on the amino acid sequence and codon usage. Lastly, we characterized how RQC factors were involved in modulating the translation dynamics through repeats. The insights gained from this single-molecule study are instrumental in understanding the biogenesis of DPR and their

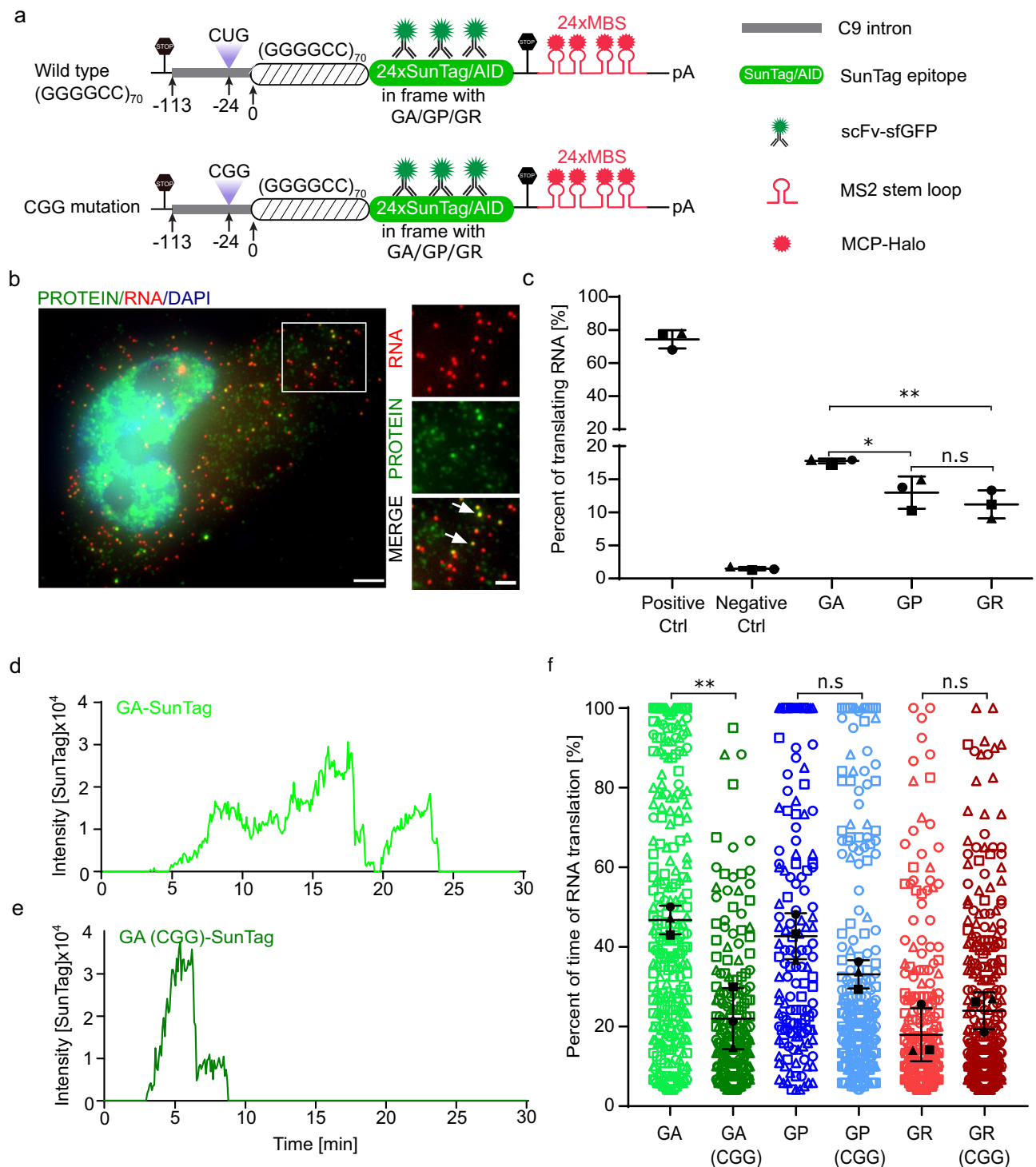
contribution to the disease. Modulation of translation is a promising therapeutic approach to prevent the accumulation of toxic DPRs.

## Results

### The CUG near-cognate start codon contributes to the initiation of the GA frame, but not GP or GR frame

To compare RAN translation initiation efficiency among different reading frames, we constructed single-molecule imaging reporters for each frame (Fig. 1a): 70 $\times$  (GGGGCC) repeats, including the 113-nucleotide upstream intron sequences, were placed in front of 24 $\times$  SunTag without the AUG start codon. Zero, one or two nucleotides were inserted between the (GGGGCC) expansion and SunTag to make SunTag in frame with a particular frame (GA, GP or GR). SunTag epitope is followed by an auxin-inducible degron (AID), which degrades mature proteins upon the addition of auxin to reduce the background fluorescence<sup>61,62</sup>. 24 $\times$  MS2-binding sites (MBS) were inserted in the 3' untranslated region (UTR) to visualize the RNA<sup>63</sup>. Next, each of these constructs was stably integrated into the U-2 Osteosarcoma (U-2 OS) cell line expressing stdMCP-HaloTag (synonymized tandem MS2 coat protein fused with HaloTag) for RNA labeling<sup>63,64</sup>, and scFv-sfGFP (single-chain antibody fused with superfolder GFP) for protein visualization<sup>65</sup>. We performed single-molecule Fluorescence In Situ Hybridization (smFISH), with probes targeting the MBS and SunTag regions, and immunofluorescence (IF) with antibodies against GFP<sup>68</sup> in these cells to quantify the nascent peptides on mRNAs. Without RNA reporters, scFv-sfGFP itself does not form detectable green puncta (Supplementary Fig. 1a, b). With RNA reporters, both red (RNA) and green puncta (protein) were observed in the cytoplasm. Dim green puncta in the cytoplasm represented the fully synthesized proteins since they could bind up to 24 scFv-sfGFPs and were readily detectable. Bright green puncta colocalized with RNA were translation sites (TLS, arrow) as one or more ribosomes were translating on the mRNA to produce multiple nascent peptides (Fig. 1b). To compare the RAN translation initiation efficiency in different frames, we quantified the percentage of translating mRNA in each reporter cell line. The repeats increased the fraction of translating mRNA (RAN translation, 10–20%) significantly compared with negative control (without insertion of repeats, <2%) (Fig. 1c, Supplementary Fig. 1c). The RAN translation was significantly less efficient than regular mRNAs (~80%)<sup>68</sup> (Fig. 1c), as expected. Interestingly, the fraction of translating mRNA in the GA frame was higher than that of GP and GR frames (Fig. 1c), supporting the higher initiation rate of poly-GA.

It has been reported that a near-cognate start codon CUG 24 nt upstream of GGGGCC repeat expansions is the initiation site of poly-GA<sup>37,38</sup>. To measure the influence of CUG on translation initiation, we made reporters in which the CUG was mutated to CGG (Fig. 1a). We performed live-cell imaging to quantify the translation dynamics of single mRNAs. To facilitate long-term tracking, mRNAs were tethered to the cell plasma membrane to limit their mobilities (by fusing stdMCP to a CAAX motif)<sup>56</sup> and imaged with total internal reflection fluorescence (TIRF) microscopy to reduce background noise. Cells were imaged to track translation of single mRNAs for 30 min (Fig. 1d, e, Supplementary Movies 1 and 2). The fraction of time that an individual mRNA molecule was being translated was quantified. In the GA frame, the RNA was translated around 45% of the time, similar to the GP frame (~40%) but significantly higher than the GR frame (~17%) (Fig. 1f). Mutation of CUG to CGG caused a 2-fold reduction in the fraction of translation time in GA frame; but did not significantly change in the GP and GR frame (Fig. 1f). The influence of CUG on different reading frames can also be validated using bicistronic luciferase reporters with the splicing context of C9ORF72<sup>66</sup>. The NanoLuc luciferase was placed in the first intron after the repeats in different reading frames to measure the corresponding DPR levels, while Firefly luciferase was put in exon 2 to normalize the expression level (Supplementary Fig. 1d). GA



frame has significantly higher expression than GP and GR. When CUG was mutated to CGG in the intron, the GA-NLuc was significantly reduced, while there was no decrease in GP and GR (Supplementary Fig. 1e).

Taken together, both the fixed, live cell and bulk experiments suggest that RAN translation initiation efficiency in poly-GA is higher than poly-GP and poly-GR. This is due to the presence of the near-cognate CUG start codon in the GA frame. The poly-GP and poly-GR translation likely initiate independently from the CUG codon (otherwise, we should expect poly-GP and poly-GR to decrease in CGG mutant as well). Surprisingly, the difference between frames, especially between GA and GR, measured by the

luciferase assay is higher than that in the single-molecule measurements (Fig. 1c, Supplementary Fig. 1e), prompting us to look at other aspects of translation that may contribute to the variation in DPR levels.

### Frameshift occurs, but infrequently during the translation of C9ORF72 repeats

It was suggested that the chimeric DPR species existed in C9ORF72-ALS/FTD patients, either by translation frameshift or genetic interruptions of repeats<sup>29</sup>. Although the translation initiation reporter data suggest that the majority of poly-GP and poly-GR translation is not due to frameshifting from the GA frame, it is intriguing to directly assess

**Fig. 1 | Single-molecule imaging of RAN translation showed higher initiation rate of poly-GA than poly-GP and poly-GR.** **a** Schematic of constructs for single-molecule imaging of RAN translation. The wild-type constructs contain 113 bp endogenous 5' intron sequence upstream of the GGGGCC repeats. A CUG near-cognate start codon (−24 bp) is in frame with GA. Multiple stop codons were placed at the beginning of the reporters to prevent any leakage from canonical translation. SunTag epitopes were fused with GGGGCC repeats in each of the three frames separately. In CGG mutants, the CUG codon was mutated to CGG. MBS MS2-binding sites, MCP MS2 coat proteins, scFv-sfGFP single-chain variable fragment against SunTag epitope fused with superfolder GFP, AID auxin-inducible degen. **b** Representative smFISH-IF image of RAN translation reporters. The protein signals colocalized with RNA indicated the active translating RNA. Red: RNAs; Green: protein. Scale bar: 5 μm. **c** The percentage of actively translating RNAs measured from smFISH-IF experiments. Data are mean ± SD from three biological replicates

(Ctrl: 86, 39, 40 cells; Negative Ctrl: 66, 60, 58 cells; GA: 82, 67, 29 cells; GP: 51, 33, 26 cells; GR: 61, 53, 25 cells). Each symbol represents the average of one biological replicate. Statistical analysis using two-tailed, equal variance, t-test: GA vs GP,  $*P = 0.028$ ; GA vs GR,  $**P = 0.0061$ ; GP vs GR,  $P = 0.39$ . **d, e** Representative traces of translation site intensity for a single mRNA in GA (**d**) and GA-CGG mutant (**e**) reporters. Also, see Supplementary Movies 1 and 2. **f** The fraction of time that RNAs were undergoing translation during the 30 min. Each symbol represented a single translating RNA, and the shapes represented biological replicates (numbers of translating RNA traces in replicates: GA: 43, 150, 100; GA-CGG: 81, 91, 49; GP: 60, 30, 52; GP-CGG: 77, 35, 116; GR: 103, 98, 99; GR-CGG: 170, 33, 119). The mean of each replicate (black shapes) was used to calculate the mean (horizontal bar) and standard deviation (error bars) in each group. Statistical analysis using two-tailed, equal variance t-test: GA vs GA(CGG),  $**P = 0.0071$ ; GP vs GP(CGG),  $P = 0.070$ ; GR vs GR(CGG),  $P = 0.27$ . Data are mean ± SD.

whether the frameshifting can occur on the repeats during translation and, if it does, how frequently it occurs.

Ribosomal frameshifting is an evolutionarily conserved biological process found in many organisms<sup>67</sup>. Frameshift can be accidental as well as programmed. For example, ribosome stalling due to secondary RNA structure may induce a ribosome to shift its reading frame<sup>68</sup>. To directly visualize the frameshifting event, we utilized a dual-color translation imaging assay to measure the translation in two reading frames simultaneously on single RNAs in live cells<sup>56,69</sup>. We made constructs in which 70× (GGGGCC) repeats were placed between 24× synonymized HA-tag (Human influenza hemagglutinin) and 24× SunTag. The HA-tag was initiated with AUG start codon and placed before the GGGGCC repeats in one reading frame (DPR1), followed by the stop codon. The SunTag lacking AUG start codon was fused after the repeats in +1 or +2 reading frame (DPR2) with respect to DPR1 and followed by an in-frame stop codon (Fig. 2a). The mRNAs were labeled with 24× MBS in 3'UTR and tethered to cell membrane for long-term imaging by 12× PP7 binding sites (PBS) with membrane-localized PP7 coat protein<sup>70</sup>. This construct allows us to quantify the frameshift events in live cells. Normally, when a ribosome initiates at the AUG start codon, translates HA-tag and DPR1 in GGGGCC repeats, and terminates before SunTag, the mRNA will contain only the HA-tag TLS signal. However, if the ribosome shifts into the DPR2 frame during the translation of GGGGCC repeats, the SunTag will be translated. In such cases, both the HA-tag and SunTag translation signal will colocalize with the mRNA (Fig. 2b).

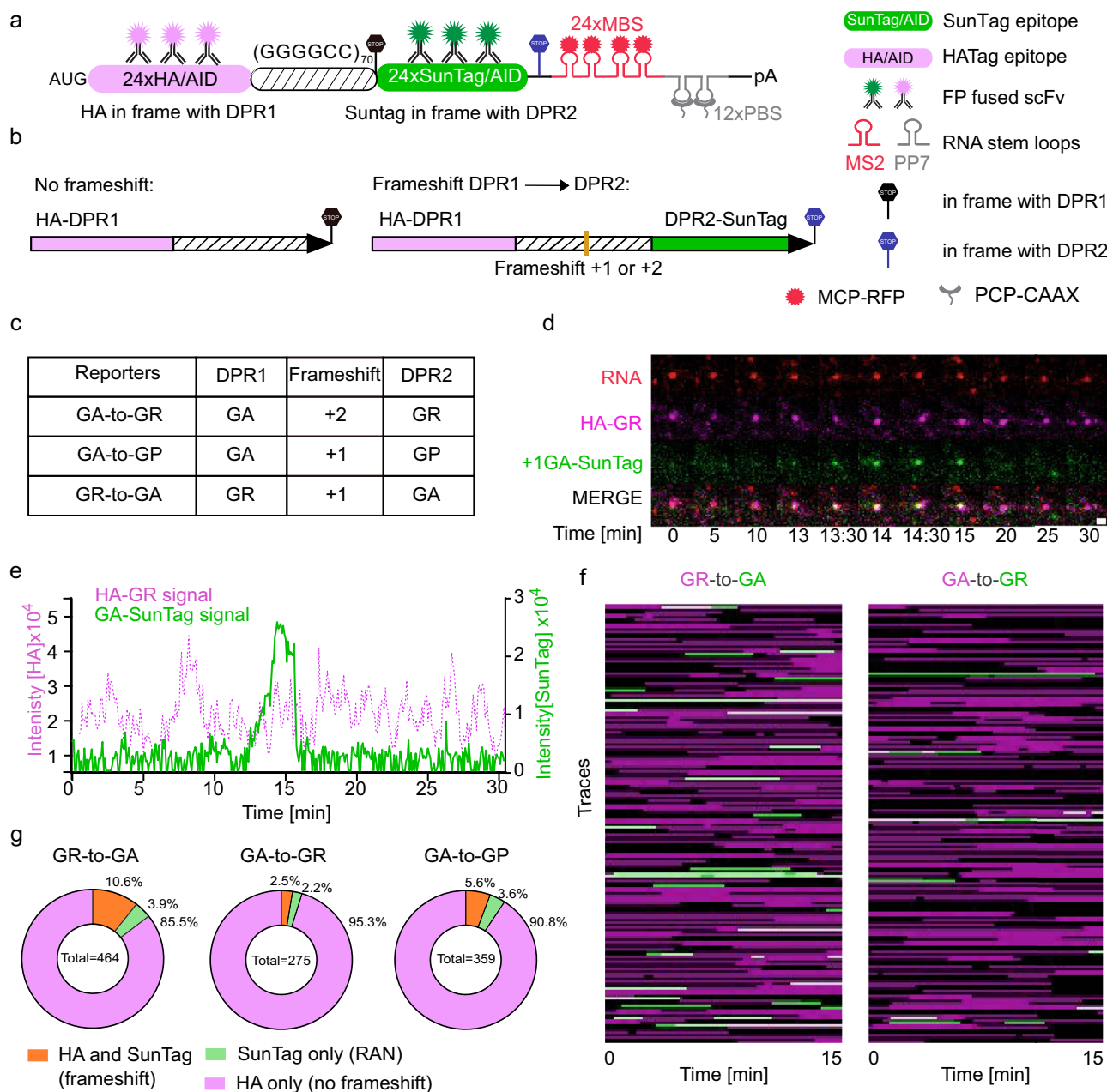
To validate that the two tags can faithfully represent the translation of two reading frames, we first constructed a negative control reporter with no repeats between HA-tag and the +1 SunTag (Supplementary Fig. 2a) and a readthrough reporter in which HA-tag and SunTag sandwiching the 70× (GGGGCC) repeats were both placed in the GA frame (Supplementary Fig. 2a, b). We stably expressed these reporters in the U-2 OS cell line containing scFv-sfGFP and Fb-HaloTag (anti-HA Frankenbody: an scFv against the HA-tag fused with HaloTag<sup>71</sup>). In the negative control reporter, there is no SunTag signal colocalized with RNA and HA-tag TLS (Supplementary Movie 3). For the readthrough reporter, the vast majority (85%) of translating mRNAs (red) colocalized with both HA-tag (magenta) and SunTag (green) signals (Supplementary Fig. 2c, d, Supplementary Movie 4). These experiments demonstrate that the two epitopes reliably report translation dynamics under our experimental conditions.

To systematically compare frameshift efficiency, we made a series of reporters to quantify frameshift events from GA to GR, GA to GP, or from GR to GA (Fig. 2c). As expected, most of the mRNAs (red) colocalized with HA-tag TLS (magenta) (Fig. 2d, Supplementary Movie 5), indicating the high translation efficiency of AUG initiated translation. In all three reporters, we observed occasional and transient SunTag fluorescence signals (green) colocalized with both mRNA and HA-tag signals during the 30-min live-cell imaging, indicating the occurrence of translational frameshift (Fig. 2d, e,

Supplementary Movie 5). This supports that the expanded repeats enhance the frameshift probability compared to the negative control. We analyzed individual translating mRNAs and quantified the frequencies of frameshift by calculating the percentage of mRNAs showing both SunTag and HA-tag signals within the imaging window (Fig. 2f, Supplementary Fig. 3a–c). First, we noticed that frameshift is rare: more than 85% of translating mRNAs in all three reporters did not have any frameshift events in 30 min (Fig. 2g). Second, the percentage of mRNAs undergoing frameshift from GR to GA (10.6%) was significantly higher than from GA to GR (2.5%) and from GA to GP (5.6%) (Fig. 2g). Similarly, we quantified the amount of time spent in particular translation state. About 90% of the time, mRNAs were translated in the original ORF without frameshifting. However, when frameshift did happen, it occurred more often for GR to GA (6.4%), followed by GA to GP (4.3%) and lastly GA to GR (2.3%) (Supplementary Fig. 3d).

We observed low numbers of mRNAs that only had SunTag signals colocalized, which likely represented cap-independent RAN translation events as reported previously<sup>66,72</sup>. To quantify the contribution of RAN translation, we constructed a bicistronic reporter. We placed stop codons in all three reading frames between HA-tag and repeats, such that the SunTag signal arose only from the cap-independent RAN translation in the GA frame (Supplementary Fig. 4a). We observed rare RAN SunTag TLS (Supplementary Fig. 4b, c, Supplementary Movie 6), with a frequency similar to the SunTag-only events in the frameshift reporters (about 2%, Fig. 2g). This supports that the quantification of the frameshift events is not significantly influenced by the cap-independent RAN translation and represent the relative frequency between different frames.

To further quantify the relative amount of frameshift products, we made dual-luciferase constructs in which the AUG initiated Firefly Luciferase (FLuc) was placed before the 70× (GGGGCC) repeats in the GR frame followed by the in-frame stop codon (Fig. 3a). Nanoluc Luciferase (NLuc) without an AUG start codon was placed after the repeats in either GR (0), GA (+1) or GP (+2) frame to compare frameshift efficiency. Identical constructs without (GGGGCC) repeats were used as negative controls (Fig. 3a). We transiently expressed these constructs in three different cell lines and quantified the frameshift products. Without GGGGCC repeats, no NLuc activity was detected in the +1 or +2 reading frame (Fig. 3b–d). 70× (GGGGCC) repeats significantly increased frameshift products in both +1 (GR-to-GA) and +2 (GR-to-GP) frames (Fig. 3b–d). The GR-to-GA shift was more efficient (2.3–5.3%) than the GR-to-GP shift (0.5–1.6%), despite the overall low frameshift efficiency. Taken together, both single-molecule imaging experiment and bulk luciferase data indicate that frameshift could occur during the translation of GGGGCC repeat expansion and the chimeric DPRs can be produced. However, considering the overall low frameshift frequency in multiple cell lines, chimeric DPRs likely account only for a small proportion of total DPRs.



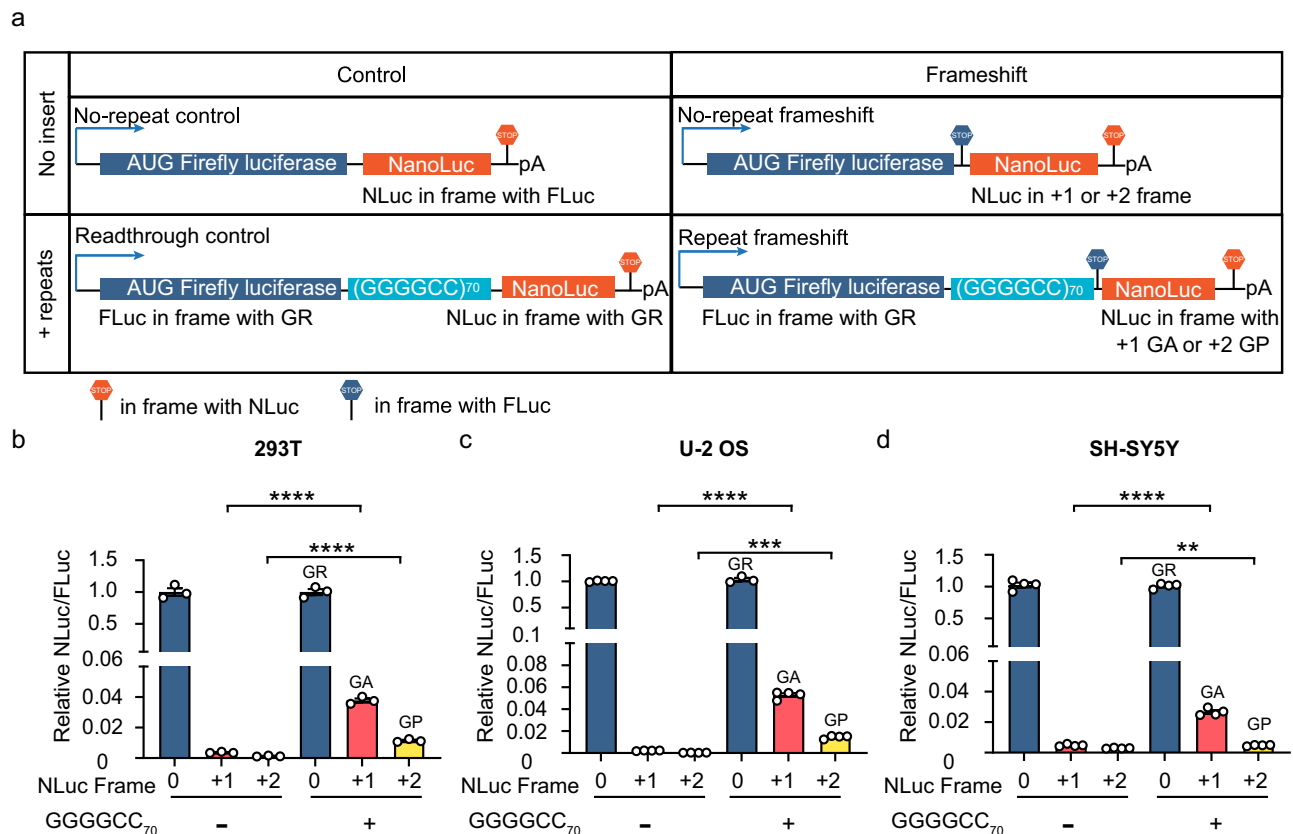
**Fig. 2 | Two-color translation assay revealed translation frameshift frequencies between DPR frames of GGGGCC repeats.** **a** Schematic of the dual-color frameshift constructs. 24× HA-tag were placed before 70× (GGGGCC) repeats in frame with DPR1. The translation was initiated by a canonical AUG start codon before HA and terminated with a stop codon after GGGGCC repeats. 24× SunTag lacking AUG start codon was placed after 70× (GGGGCC) repeats in the frame of DPR2. 24× MBS was used for RNA visualization and 12× PBS was used to tether mRNA on the cell membrane for long-term tracking. **b** Expected protein products without or with translation frameshift. **c** Frameshift reporters used in this study. **d** Time-lapse images illustrating translation frameshift events for GR-to-GA reporter. Red (RNA): stdMCP-RFP; magenta (GR): HA-Fb-HaloTag; green (GA): SunTag-scFv-sfGFP. See

also Supplementary Movie 5. **e** Fluorescence intensity traces of both HA-tag and SunTag on a single mRNA showing frameshift events (Corresponding to **d**). **f** Combined TLS tracks in two translation channels for GR-to-GA and GA-to-GR frameshift reporters. Magenta: HA-Fb-HaloTag indicating the translation of AUG-HA-DPR1; green: SunTag-scFv-sfGFP indicating the translation of DPR2-SunTag. **g** Percentage of translating mRNAs undergoing normal translation (only in HA-DPR1 frame, magenta), frameshift (both frames, orange), or RAN translation (only DPR2-SunTag frame, green). The number of mRNAs under specific state was divided by the total number of tracks to calculate the percentage: GR-to-GA (5 cells, 464 translation events); GA-to-GR (6 cells, 275 of translation events); GA-to-GP (6 cells, 359 of translation events).

### The translation elongation speeds in GP and GR frames are slower than in GA frame

The GGGGCC repeat sequence involves highly repetitive codon usage and can form strong secondary structures, such as hairpin or G-quadruplex. It is not well understood how they influence the elongation of ribosomes through the repeats. Earlier, we demonstrated ribosomes translating poly-GR were more likely to frameshift. A plausible explanation for the observation is that the slower elongation

speed of poly-GR leads to ribosomal collision and subsequent sliding to a different frame. We therefore examined the elongation kinetics of ribosomes through different reading frames. We constructed a series of reporters with SunTag placed before the repeats in different reading frames (Fig. 4a). After the SunTag sequence is translated, the fluorescence signal remains on the nascent peptide while the ribosome translates the repeats, therefore reporting on the elongation kinetics through the specific reading frames of the repeats. The control



**Fig. 3 | Luciferase assay to measure proteins produced from frameshifting.** **a** Schematic of dual-luciferase constructs used to compare frameshift products. In the no-repeat control, NLuc was fused in frame with FLuc as a single open reading frame. In the readthrough control, NLuc was placed in the same frame between FLuc and GR-(GGGGCC)<sub>70</sub>. In the no-repeat frameshift control, NLuc was placed in the +1 or in the +2 frame, with stop codon placed in the FLuc frame before NLuc. In the repeat frameshift construct, the GR frame of (GGGGCC)<sub>70</sub> was placed in frame with FLuc. **b–d** Normalized frameshift products of different reporters in 293T (b),

U-2 OS (c) and SH-SY5Y (d) cells, respectively. All constructs were transiently transfected into cells for 24 h. NLuc signals were normalized to FLuc in each sample. The no-repeat frameshift is normalized by no-repeat control, and the repeat frameshift is normalized by the readthrough control. Data were reported as mean ± SEM from three or four biological replicates. Statistical analysis using two-tailed t-test. 293T: frame+1, \*\*\*\**P* = 0.000016; frame+2, \*\*\*\**P* = 0.00007. U-2 OS: frame+1, \*\*\*\**P* = 0.000021; frame+2, \*\*\**P* = 0.00014. SH-SY5Y: frame+1, \*\*\*\**P* = 0.000019; frame+2, \*\**P* = 0.0059.

reporter contains an NLuc sequence in place of repeats, which has a similar length and indicates the elongation through a regular protein sequence.

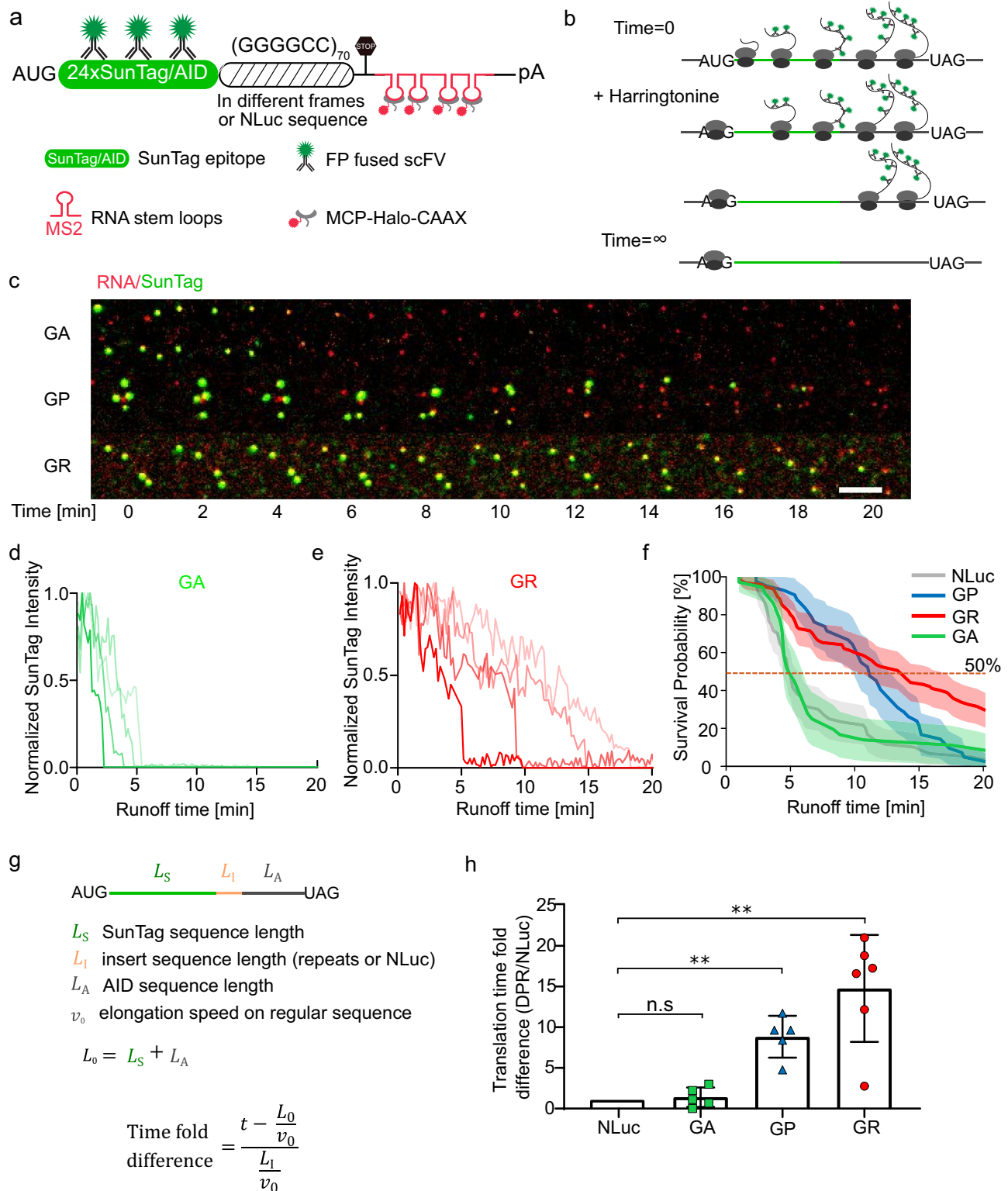
To measure the elongation speed, we performed a ribosome runoff experiment in live cells<sup>56,73</sup>. Harringtonine, a translation inhibitor that prevents the first peptide bond formation, effectively stops translation initiation but does not interfere with existing 80S ribosomes on mRNAs<sup>74</sup>. Application of harringtonine synchronously stops initiation while allowing initiated ribosomes to run off the mRNA, the duration of which depends on the elongation speed and the length of the open reading frame (ORF)<sup>75</sup> (Fig. 4b). To facilitate long-term tracking of single mRNAs, the reporter mRNAs were tethered to the plasma membrane via membrane-localized MCP<sup>73</sup>. As expected, harringtonine results in a steady loss of translation intensity in the GA frame but with a slower speed for GP and GR (Fig. 4c–e, Supplementary Movie 7). We calculated the survival probability of the translation signal: the percentage of translating mRNAs as a function of time after adding Harringtonine (“Methods”, only translating mRNAs at time zero were considered). The survival curve depends on the elongation of ribosome through the ORF: the curve shifts to the right if elongation is lower. Interestingly, it took much longer for the translation signal to disappear in the GP and GR frames than in the GA frame, which resembled the control NLuc reporter (Fig. 4f). The median runoff time for the GA reporter was 5.1 min, close to NLuc (4.9 min), whereas it was 11.2 min for GP and 13.5 min for GR (Fig. 4f), representing 2.3 and 2.8

folds increase respectively. Note that the GGGGCC repeat sequence (420 bp) only occupied a small fraction of the whole ORF (3009 bp) (Fig. 4g). The actual elongation speed discrepancy between DPRs was much larger than the difference in runoff time. To estimate the elongation speed through the repeat sequence, we calculated the ribosome residence time per amino acid in the repeat region, which was inversely proportional to the elongation speed (Fig. 4g, “Methods”). We estimated that the elongation speeds in the GP and GR frames were 10 and 15 folds slower compared with NLuc, while that of GA was statistically identical to NLuc (Fig. 4h).

We used fluorescence recovery after photobleaching (FRAP) of single translation sites as an independent method to measure elongation speed<sup>58</sup>. In agreement with the ribosome runoff experiment, we observed the TLS fluorescence in the control NLuc and GA reporters recovered rapidly, whereas the recovery was significantly slower in GP and GR reporters (Supplementary Fig. 5a, b, Supplementary Movie 8). Altogether, these results demonstrate that the GP and GR frames have slower elongation speed than the GA frame for the GGGGCC repeats.

### The amino acid composition of DPRs affects the translation elongation speed

The potential secondary structures formed by GGGGCC repeat RNA cannot explain the slow elongation through GP and GR since the GA frame shares the same secondary structure but elongates similarly with regular control sequences. We therefore investigated the

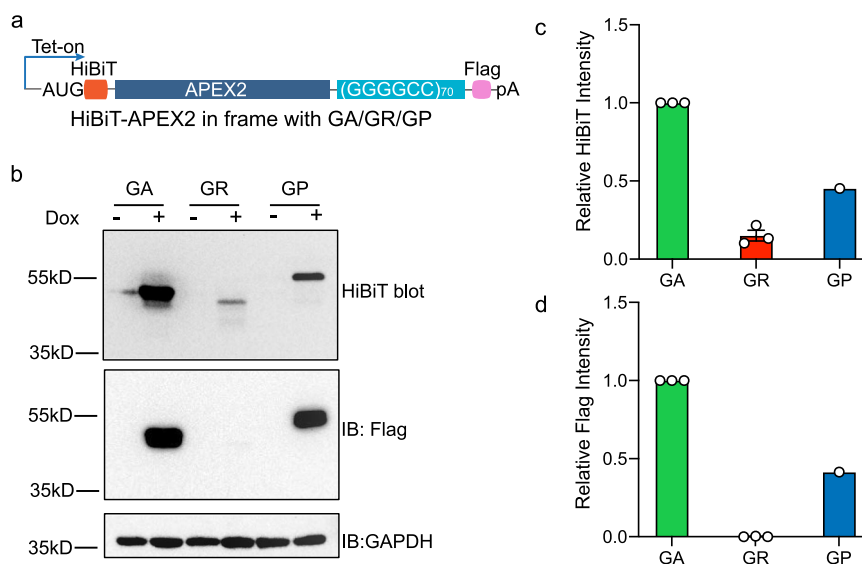


influence of the encoded amino acids. Proline introduces steric hindrance in the peptide and is known to cause ribosome pausing<sup>76,77</sup>. The antisense CCCCCG repeats also code for GP in one of the frames. Similarly, we observed a slower ribosome runoff in GP frame encoded by the antisense repeats (Supplementary Fig. 5a, c). The FRAP experiment also showed consistent results (Supplementary Fig. 5b). The amino acid arginine is positively charged. A stretch of arginine is a strong ribosome stalling sequence and induces ribosome quality

control in yeast<sup>78</sup>. Poly-GR has been shown to block the peptide exit tunnel of the ribosome in vitro<sup>79</sup>. We used synonymously randomized codons to encode poly-GR sequences. Similarly, we observed that the ribosome runoff was significantly slower than the NLuc control and similar to the repeat-encoded poly-GR (Supplementary Fig. 5c). Altogether, these results support that the amino acid composition of DPRs affects the translation elongation speed through the expanded repeat RNA.

**Fig. 4 | The translation elongation speeds in the GP and GR frames are slower than in the GA frame.** **a** Diagram of the single-molecule DPR elongation reporter construct. The (GGGGCC)<sub>70</sub> repeats were inserted after AUG initiated SunTag in different reading frames separately. The N-terminal SunTag fluorescence stayed on the RNA while the ribosome translated the repeats, therefore reporting the elongation speed. 24× MBS were used for RNA visualization and tethering to the plasma membrane through membrane-localized MCP (stdMCP-HaloTag-CAAX). **b** Schematics of ribosome runoff assay. Harringtonine effectively stops translation initiation but does not affect elongating ribosomes, which continue translating until “runoff”. This assay is used to determine the elongation speed. **c** Snapshots from videos of ribosome runoff experiment for GA, GP and GR reporter cells, respectively. The time reported was after adding harringtonine (Supplementary Movie 7). Scale bar: 5 μm. **d, e** Example translation intensity traces for GA (**d**) and GR (**e**) reporters after harringtonine treatment. Traces were normalized to the respective maximum intensities. **f** The survival curves of translation sites as a

function of runoff time (“Methods”): the percent of mRNAs still having translation signals at given times was calculated in each cell and then averaged. Dashed line represents the 50% of mRNAs that finished runoff. The shadow represents the 95% confidence bounds (Greenwood’s formula). NLuc: 6 cells, 60 TLS; GP: 5 cells, 42 TLS; GR: 6 cells, 95 TLS; GA: 5 cells, 36 TLS. **g** Estimation of elongation speed through the inserted sequence using median runoff time measured in (**f**) (“Methods”). The time spent translating the DPR is compared with translating NLuc. L<sub>S</sub>: length of SunTag; L<sub>A</sub>: length of AID; L<sub>0</sub> = L<sub>S</sub> + L<sub>A</sub> is identical for all reporters; L<sub>I</sub>: length of inserted sequence (NLuc or repeats); v<sub>0</sub>: regular elongation speed (through NLuc or the rest reporter); t: runoff time. **h** Elongation time (per amino acid) through different repeat frames compared with through NLuc (median). Each symbol represents the average measurement in one cell. Data are mean ± SD from six (NLuc, GR) or five (GP, GA) independent experiments. Statistical analysis using one-sample t-test: GA, *P* = 0.50; GP, *\*\*P* = 0.0024; GR, *\*\*P* = 0.0037.



by HiBiT blot. Data are presented as mean ± SEM from three (GA, GR) or one (GP) independent experiments. **d** Densitometry quantification of the relative reporter expressions by Flag blot. Data are presented as mean ± SEM from three (GA, GR) or one (GP) independent experiments.

by HiBiT blot. Data are presented as mean ± SEM from three (GA, GR) or one (GP) independent experiments. **d** Densitometry quantification of the relative reporter expressions by Flag blot. Data are presented as mean ± SEM from three (GA, GR) or one (GP) independent experiments.

To assess the effect of elongation on the final protein product, we designed a split luciferase reporter to measure DPR production. The AUG-driven 11-amino acid fragment of NanoLuc (HiBiT) was fused upstream of the GGGGCC repeat in different frames to measure the N-terminal product. A Flag tag was placed at the C-terminus in frame with AUG-HiBiT to report the full-length proteins (Fig. 5a). The construct was controlled by an inducible Tet-On promoter and stably integrated through the Flp-In system in HeLa cells. The protein level was measured through HiBiT blotting or immunoblotting against Flag. When the GGGGCC repeat was placed in the GA frame, the protein level was significantly higher than that in the GP and GR frames (Fig. 5b–d) after the RNA level was normalized (Supplementary Fig. 6a). As all the DPR translation was initiated by AUG start codon in the same reporter backbone, the discrepancy cannot be explained by the initiation differences. We next tested whether the protein stability contributes to the different steady-state DPR levels. We measured the DPR stability and found that poly-GA was, in fact, less stable than poly-GP or poly-GR (Supplementary Fig. 6b–f). Therefore, the higher poly-GA expression level was not due to its increased stability either. Taken together, we conclude that the

protein amount correlates with the translation elongation rate in different DPR frames.

To understand the potential mechanism of how elongation may influence the protein yield, we performed in vitro translation of the elongation reporters (Supplementary Fig. 6g). The in vitro transcribed reporter mRNAs (Supplementary Fig. 6h) were capped and translated in rabbit reticulocyte lysate. We included the K20 (K<sup>AAA</sup>)<sub>20</sub> reporter as a positive control, which is known to induce translation stalling<sup>80</sup>. The no repeat construct (K0) was used as negative control for stalling. As expected, there was barely any detectable translation product from the K20 reporter RNA (Supplementary Fig. 6i). The translation yield from the GA reporter was comparable to the no repeat control (Supplementary Fig. 6i), consistent with the in vivo translation measurement that the elongation in GA frame was similar to the NLuc. Interestingly, we observed less full-length protein from the (GGGGCC)<sub>40</sub>-GR reporter. There were multiple shorter protein bands, likely representing truncated translation products (Supplementary Fig. 6i). This indicates that there is ribosome stalling and truncation of the nascent peptide during translation elongation in the GR



frame, which could contribute to reduced poly-GR production compared to poly-GA.

### RQC factors modulate the translation dynamics of DPRs

To identify genetic modifiers for poly-GR production, we performed a CRISPR-Cas9 knockout screening. We constructed an AUG-driven translation reporter of GGGGCC repeat in the GR frame fused with a GFP at the C-terminus. We stably expressed it in the HeLa Flp-In cells under the Tet-On promoter (Fig. 6a). The reporter cells were infected with a lentiviral guide RNA (gRNA) library targeting genes encoding 1906 RNA-binding proteins and ribosome proteins. After drug selection for gRNA expression, GR-GFP was induced by doxycycline. The GFP-high and GFP-low cell populations were collected by fluorescence-activated cell sorting (FACS) (Supplementary Fig. 7a, b)<sup>81</sup>. We sequenced the gRNA fragments from genomic DNA in each cell population and analyzed their distribution<sup>81</sup>. Many protein/RNA quality control genes, such as *ZNF598* and *Pelota*, were top hits in the screen whose gRNAs were enriched in the GFP-high population, suggesting knockdown of these genes can increase GR-GFP level (Fig. 6b, Supplementary Data 1).

We further used the single-molecule imaging platform to directly evaluate the function of RQC factors on DPR translation. The slowdown or stalling of ribosomes may lead to the collision of trailing ribosomes. The collided ribosomes signal aberrant translation status, resulting in the splitting of ribosomes and subsequent degradation of nascent peptides and mRNAs, known as the RQC pathway<sup>82,83</sup>. *ZNF598* is an E3 ubiquitin ligase that detects the collided ribosomes, ubiquitinates the small ribosomal subunit and triggers RQC response<sup>84,85</sup>. In the polyA-induced ribosome collision, *ZNF598* accelerated the clearance of the collided ribosomal queue<sup>73</sup>.

As the elongation speed through poly-GR is significantly slower than a general protein sequence, it may cause trailing ribosomes to collide. That might be the reason why many RQC factors were enriched in the CRISPR screen. To test the hypothesis, we evaluated whether *ZNF598* knockdown influenced the translation kinetics of the elongation reporters (Fig. 4a). We performed ribosome runoff assays as described previously and compared the *ZNF598* siRNA transfected cells with scrambled siRNA control (Supplementary Fig. 7c, d). Knocking down *ZNF598* further extended the ribosome runoff time of the (GGGGCC)<sub>n</sub>-encoded GR (Fig. 6c, d, Supplementary Movie 9), but interestingly had no effect on the GR reporter encoded by randomly synonymized codons (Fig. 6e). More surprisingly, the *ZNF598* knockdown also slowed down the ribosome runoff on the (GGGGCC)<sub>n</sub>-encoded GA repeat (Fig. 6f), even though its elongation speed was similar to the control NLuc sequences (Fig. 4f). Likewise, *ZNF598* knockdown had no effect on randomized GA (Fig. 6g). *Pelota* is the paralogue of translation release factor *eRF1*<sup>86</sup>. The yeast homolog of *Pelota* is implicated in recycling ribosomes at the truncated message<sup>49</sup>. Knocking down *Pelota* recapitulated the observation from the *ZNF598* knockdown experiment (Fig. 6d–g, Supplementary Fig. 7e). Altogether, our data support that the translation of C9ORF72 repeats can be modulated by the ribosome-associated quality control pathway. The repeat RNA sequence, rather than the DPR composition, is important to be recognized as the RQC substrate.

To evaluate whether RQC factors were involved in diseases, we further analyzed the proteomics data of C9ORF72-ALS/FTD patient-derived induced pluripotent stem cell (iPSC)-differentiated neurons (iPSN) from AnswerALS<sup>87</sup>. C9ORF72-ALS patient neurons showed a significantly decreased level of *Pelota* protein compared to non-neurological controls (Fig. 6h). This could influence the DPR production as well as other protein homeostasis in neurons, potentially contributing to the disease pathogenesis.

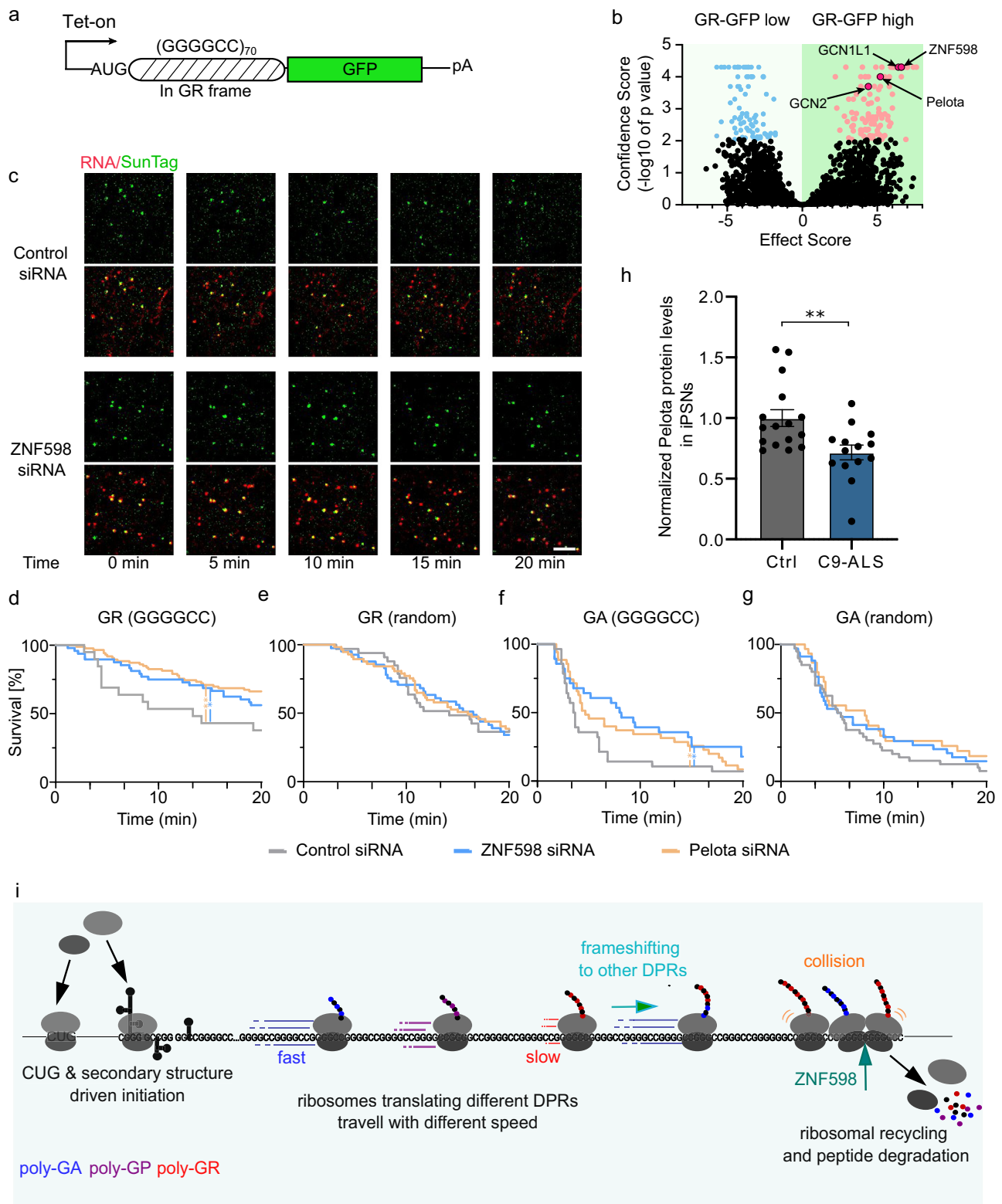
## Discussion

Despite recent progress in understanding the RAN translation initiation mechanisms in C9ORF72-ALS/FTD, little is known about translation dynamics through the expanded repeats. Previous work mostly relied on ensemble approaches to measure the final DPR level, which was an indirect readout of translation activity. In this study, we used single-molecule imaging in live cells to directly examine the translation kinetics and investigated various aspects of repeat-associated translation: initiation, elongation, frameshift, and quality control. We recently identified that the spliced intron of *C9ORF72* was stabilized by the GGGGCC repeats in a circular form, which was exported to the cytoplasm and subjected to cap-independent translation<sup>72</sup>. In this paper, the repeat was placed in a capped mRNA reporter for translation analysis, as this allowed a higher expression level and more accurate quantification of translation properties. We used the canonical AUG start codon for translation initiation when assessing the elongation, frameshift, and quality control response in different DPR frames, which was independent of the initiation step.

To study RAN translation, we and others have placed reporters after the GGGGCC repeats in different frames to report the in-frame DPR production separately<sup>36,37</sup>. With the limitation in mind, we found that in agreement with previous work<sup>37,38</sup>, translation initiation in the GA frame had higher efficiency than GP and GR frames, and the CUG near-cognate start codon preceding the repeats contributed to the initiation in the GA frame. However, the difference in the initiation rate was insufficient to explain the measured difference in the final DPR amounts. Instead, we found that the elongation speeds of different reading frames varied significantly, with GP slower than GA and GR the slowest. The discrepancy of translation elongation may contribute to the different amounts of the final DPR products. It is noted that this is not direct proof that the specific DPR is produced from the initiation of the labeled frame because frameshift and abortive translation can occur. Nevertheless, we have shown that frameshift is rare. On the first-order approximation, we believe that the initiation reporters can represent the RAN translation efficiency.

We directly observed frameshift between different frames with a two-color translation assay. The frameshift frequency on the repeats was higher than the regular RNA sequences. Therefore, chimeric DPRs likely exist in patients. However, the frequency was overall low and unlikely to be the major determinant for the DPR abundance. Contrary to the previous hypothesis, we found that the frameshift from GR to GA had a higher probability than the reverse. This was likely due to the slow elongation speed of GR, which could increase the chance of frameshift. Furthermore, mutating the near-cognate CUG codon in the GA frame did not reduce the translation efficiency of GP and GR frames. Taken together, our study supports that the majority of poly-GP and poly-GR should arise from independent initiation in the corresponding frames rather than frameshift from poly-GA. It is possible that longer expanded repeats in patients can slightly increase the chance of frameshifting, which cannot be fully recapitulated by the 70× repeat reporter used in this study. Nevertheless, the relative frequency among different frameshifting directions (such as higher GR-to-GA than GA-to-GR) should remain similar. Consistent with our finding, a recent study demonstrated that homozygous deletion of an intron fragment preceding the endogenous repeats, which contains the CUG codon, in C9ORF72-ALS/FTD patient-derived iPSNs only reduced the level of endogenous poly-GA but not poly-GP and poly-GR<sup>88</sup>. This evidence strongly suggests that frameshifting is not the main mechanism for poly-GP or poly-GR production from the GGGGCC repeats. Overall, our work indicates that there are chimeric DPRs produced via translation frameshift in C9ORF72-ALS/FTD patients, but the proportion of chimeric peptides could be low.

The GGGGCC repeat is prone to form strong secondary structures and may slow down elongation. However, we showed that the coded



amino acid, rather than the secondary structure of the RNA molecule, was the main determinant for the different elongation speeds between frames. The GA frame had a similar elongation speed as a general protein, but the GP and GR frames had much slower elongation, even though the RNA template was the same. The slow elongation of GP and GR frames was unlikely due to the rare codon effect because the antisense (CCCCGG)<sub>n</sub>- or the randomly synonymized poly-GP or poly-GR had similar elongation speeds compared to the (GGGGCC)<sub>n</sub>-

encoded ones. Therefore, the amino acid composition of the DPRs, rather than RNA or codon sequences, determines the elongation speed. Recently it has been shown that the positively charged poly-GR interacted with the peptide exit tunnel<sup>79</sup>, a possible mechanism for its slow elongation. Additionally, the bulky proline residue may cause a steric hindrance and slow down the elongating ribosomes<sup>76,79,89</sup>.

Aberrant mRNAs can lead to ribosomal stalling and degradation of nascent proteins and/or mRNAs<sup>90,91</sup>. Cells can readily resolve the

**Fig. 6 | Ribosome quality control factors influence elongation dynamics through repeat RNAs.** **a** Diagram of doxycycline-inducible AUG-driven GR-GFP translation reporter for CRISPR screening. The reporter was stable integrated in HeLa cells. **b** Volcano plot depicts gene knockout enrichment in cells with different GR levels. The top GFP-high and GFP-low cell populations were collected by FACS and subjected to deep sequencing and statistical analysis for gRNA distribution. Red: genes contributing to high GR-GFP levels when knocked out (10% FDR), Blue: genes contributing to low GR-GFP levels when knocked out (10% FDR). **c** Representative time-lapse images of runoff experiment for scrambled control siRNA (top) and ZNF598 siRNA transfected GR (GGGGCC)<sub>70</sub> elongation reporter cells (Fig. 4a). Scale bar: 2.5 μm. Also see Supplementary Movie 9. **d–g** The translation site survival curves of ribosome runoff experiment for control (gray), ZNF598 knockdown (blue), and Pelota knockdown (orange) in elongation reporter cells: GR-(GGGGCC)<sub>70</sub> (**d**), GR random (**e**), GA (GGGGCC)<sub>70</sub> (**f**) and GA random (**g**). For DPR random reporters, the repeat sequence was replaced with randomly

synonymized sequences coding for the same DPR of interest. The graphs were generated from 4–6 cells per condition, with 48 TLS in GR(GGGGCC), 41 TLS in GR(random), 28 TLS in GA(GGGGCC), 34 TLS in GA (random), for ZNF598 siRNA KD; 86 TLS in GR(GGGGCC), 57 TLS in GR(random), 35 TLS in GA(GGGGCC), 30 TLS in GA(random) for Pelota siRNA KD and with 19 TLS in GR(GGGGCC), 33 TLS in GR(random), 28 TLS in GA(GGGGCC), 40 TLS in GA(random) for Control siRNA KD. The survival curve comparison between Control and ZNF598 and Pelota groups was done using Log-rank (Mantel-Cox) test. GR(GGGGCC): ZNF598 siRNA, \**P* = 0.042; Pelota siRNA, \*\**P* = 0.0051. GA(GGGGCC): ZNF598 siRNA, \**P* = 0.016; Pelota siRNA, \**P* = 0.042. Data are median survival. **h** Normalized Pelota protein levels in iPSN derived from control (*n* = 16) and C9-ALS patients (*n* = 14). Graph shows proteomics measurements from AnswerALS dataset<sup>87</sup>. Data are mean ± SEM. Statistical test using two-tailed nonparametric Wilcoxon test: \*\**P* = 0.0045. **i** A working model showing the influence of translation initiation, elongation, frameshift and quality control factors on the DPR productions.

transient ribosome pause, but prolonged stalling will invoke RQC<sup>73</sup>. Slow elongation may induce ribosome collision and associated quality control. We found that the RQC factors ZNF598 and Pelota modulate translation dynamics through repeats. Surprisingly, both poly-GA and poly-GR encoded by GGGGCC repeats were modulated, but not the DPR encoded by synonymously randomized codons. This indicates that the repeat RNA sequence or structure, instead of the DPR amino acid composition or elongation rate, is important for invoking the quality control mechanism. There is precedence for this phenomenon: poly-lysine encoded by AAA was more efficient to induce RQC than the same peptide encoded by AAG<sup>92</sup>. It is worth further study on how GGGGCC structurally interacts with collided ribosomes to recruit RQC factors. In C9ALS patient-derived iPSN lines, Pelota protein level is lower, which could lead to higher DPR expression and exacerbate the disease. Mutation in ZNF598 (R69C) has been linked to ALS from large-scale genome-wide association study<sup>93</sup>. The dysregulation in RQC pathways will be a risk factor not only for increasing DPR expression but also for generating global cellular stress<sup>94</sup>.

The ribosome collision and the queue formed by stalling sequence depend on the initiation rate and repeat length. A higher initiation rate may lead to more frequent collisions and cause a stronger RQC effect. The elongation reporters in this study were driven by canonical AUG, with an initiation rate much higher than RAN translation. The low ribosomal flux for RAN translation may reduce the collision rate or allow the collision to be naturally resolved and bypass detection. On the other hand, longer repeats in patients may have more translating ribosomes or take a longer time to translate through. In addition, the RAN translation may occur in any frame of the repeat. If a ribosome stalls in one frame (such as GR), it will stop ribosomes in any frame, including GA, which has a higher initiation and elongation rate. In this work, we only used 70x GGGGCC repeats due to technical difficulty to construct longer repeats. Longer repeat in patient may lead to higher levels of RQC.

The RNA-mediated gain-of-toxicity in C9ORF72-ALS/FTD may result from the formation of RNA granules that sequester RNA-binding proteins (RBP) and disrupt RNA metabolism<sup>95</sup>. The RBPs play essential roles in regulating RNA splicing, export, stability as well as translation. It is plausible that some RBPs may influence translation initiation, elongation, termination or frameshift, for example, by influencing the repeat RNA structures. However, it is unlikely that RBPs may exert differential influences on different DPR frames. In addition, the sub-cellular localization of the RBPs should be considered for their pathophysiological functions. The repeat RNA granules are predominantly in the nucleus, while the translation occurs in the cytoplasm. The single-molecule imaging systems reported here may provide a valuable platform to study how specific RBPs can affect the different properties of repeat RNA translation.

DPR level is regulated by diverse cellular signaling pathways. C9ORF72 DPR production is upregulated by the integrated stress

response through eIF2α phosphorylation<sup>36,37</sup>. Increased neuronal activity can elevate the DPR levels<sup>96</sup>. The protein kinase A (PKA) pathway has been shown to positively influence RAN translation<sup>97</sup>. It is generally believed that the repeat translation initiation is modulated by these signaling pathways. It will be interesting to explore whether they also influence the translation elongation or frameshifting of the expanded repeats.

Taken together, this work studied various aspects of translation associated with GGGGCC repeat in C9ORF72-ALS/FTD. The single-molecule approach points out the importance of direct measurements of multiple translation steps to achieve a better understanding of the repeat RNA translation. This platform can be applied to other repeat expansion diseases. Our study reveals that the final DPR level depends on a multifaceted mechanism (Fig. 6i), including initiation, elongation, frameshifting, and quality control pathways. This work also sheds light on other repeat expansion diseases that it is important to examine different translation steps and properties besides the initiation. Different RNA or amino acid repeat sequences probably have different influences on translation, which can be explored using similar single-molecule imaging approaches. Understanding the mechanism of the repeat RNA translation and DPR production is of critical importance to identify the genetic modifiers and create a therapeutic intervention.

## Methods

### Plasmid

**Reporters to measure translation initiation.** To compare the RAN translation efficiency between different frames, pCAG-(GGGGCC)<sub>70</sub>-24×SunTag-24×MBS mRNA expression plasmids were constructed. 24×MBSV5 was PCR amplified from puc57-24×MBSV5 template<sup>63</sup> and inserted downstream of (GGGGCC)<sub>70</sub> via NsiI and SacI in the pCAG-(GGGGCC)<sub>70</sub> vector<sup>81</sup>. The FLAG-24×SuntagV4 was cut out from pUbC-FLAG-24×SuntagV4-oxEBFP-AID-baUTRI-24×MBSV5-Wpre<sup>88</sup> (Addgene Plasmid #84561) by NotI and FspI and inserted into pCAG-(GGGGCC)<sub>70</sub>-24×MBS via NotI and NsiI (Blunt) between (GGGGCC)<sub>70</sub> and 24×MBS. The ATG start codon was removed, and a frameshift was introduced by replacing the region between NotI and AflIII with a linker sequence AGATTACAAGGACGACGACGATAAGGGCGGACCGGGTGGATCTGGAGGTGGAGTTCTGGAGGAGAAGAAGTCTTTGAGCAAGAATTA TCATCTTGAGAACGAAGTGGCTCGT for GA frame, GATTACAAGGAC GACGACGATAAGGGCGGACCGGGTGGATCTGGAGGTGGAGTTCTG GAGGAGAAGAAGTCTTTGAGCAAGAATTATCATCTTGAGAACGAAGT GCTCGT for GR frame, and TCGATTACAAGGACGACGACGAT AAGGGCGGACCGGGTGGATCTGGAGGTGGAGTTCTGGAGGAGAAG AACTTTTGAGCAAGAATTATCATCTTGAGAACGAAGTGGCTCGT for GP frame. For reporters with CUG to CGG mutation, the preceding intron region was replaced by (GTGAACAAGAAAAGACCTGATAAAG ATTAACCGAAGAAAACAAGGAGGGAACAACCGCAGCCTGTAGCA AGCTCgGGAAGTCAAGGAGTC). For the pCAG-stop×3-24×SunTag-

24×MBS no repeat control plasmid, the (GGGGCC)<sub>70</sub> repeat sequence between HindIII and NotI (blunt) was replaced by the intron sequence cut from the previous bicistronic plasmid (Cheng 2018, Nat Comm) by HindIII and BssHII (blunt).

To make bicistronic splicing luciferase reporters with CGG mutation, the preceding intron region between HindIII and BSSHII of our previous bicistronic splicing reporter<sup>66</sup> was replaced by DNA sequences containing CGG mutation.

**Reporters to measure translation elongation.** For elongation reporters, ATG-SunTag was placed in front of repeat sequences. To make pCAG-ATG-24×SunTag-MCS-24×MBS plasmids, BFP sequence after SunTag in pCAG-ATG-24×SunTag-24×MBS was replaced with fragments containing multiple cloning sites (MCS) using PstI and BsrGI: MCS1 (AAGCTTGCAGGTACCGGAGCAGATATCGCTGGAGCGG CCGC) for (GGGGCC)<sub>70</sub> GA frame and (CCCCGG)<sub>70</sub> GP frame; MCS2 (AAAGCTTGCAGGTACCGGAGCAGATATCGCTGGAGCGGCCGCAT) for (GGGGCC)<sub>70</sub> GR frame. MCS3 (AAAAGCTTGCAGGTACCGGAGC AGATATCGCTGGAGCGGCCGCT) for (GGGGCC)<sub>70</sub> GP frame. Next, the (GGGGCC)<sub>70</sub> repeat was cut from PBS-(GGGGCC)<sub>70</sub> plasmid<sup>66</sup> by KpnI and XhoI (DNA blunting at the XhoI end) and cloned into pCAG-ATG-24×SunTag-MCS-24×MBS plasmids at the KpnI + EcoRV sites. To make antisense GP frame reporter, pBS-(GGGGCC)<sub>70</sub> was cut by HindIII and Acc65I (blunt) and inserted into pCAG-ATG-24×SunTag-MCS1-24×MBS by HindIII and EcoRV. To make control plasmid, NLuc sequence (513 bp) was PCR amplified and inserted into pCAG-ATG-24×SunTag-MCS1-24×MBS by HindIII and EcoRV. For elongation reporters with randomized codon, the DNA fragments encoding GA50, GP50 and GR50 were synthesized (Genewiz) and ligated to produce GA100, GP80 (shortened during cloning) and GR100. GP80 was cloned into pCAG-ATG-24×SunTag-MCS2-24×MBS by EcoRV and NotI. GA100 and GR100 cloned into pCAG-ATG-24×SunTag-MCS1-24×MBS by HindIII and EcoRV.

**Reporters to measure translation frameshift.** For frameshift reporters, a fragment containing multiple cloning sites (CCTGCAGGGAA TTCTTAATTAATGCCTCGAGACTAGTTGTACATAGAAGCTTGCCCCG GCGCGCGGATCCGATATC) was first placed between AflII and XbaI in pcDNA5-TO plasmid. 24×rHA-sNluc-smAID was synthesized (Genewiz) and cloned into pcDNA5-TO-MCS by XhoI (blunt) and BsrGI. Next, (GGGGCC)<sub>70</sub>-24×SunTag-24×MBS was cut from pCAG-Δintron-(GGGGCC)<sub>70</sub>-24×SunTag-24×MBS (GA frame) by NotI (blunt) and BglII and inserted into the plasmid at the SrfI and BamHI sites. The 24×MBS was replaced with 24×MBS-12×PBS cut from pcDNA5-ATF4-24×SunTag-24×MBS-12×PBS<sup>73</sup> by NheI and NsiI (blunt) to make the final pcDNA5-TO-24×rHA-(GGGGCC)<sub>70</sub>-24×SunTag-24×MBS-12×PBS reporter (**GR to GA reporter**). For the **no repeat control reporter** (pcDNA5-TO-24×rHA-stop×3-24×SunTag-24×MBS-12×PBS), the (GGGGCC)<sub>70</sub>-24×SunTag in the **GR to GA reporter** was replaced with 24×SunTag from pCAG-stop×3-24×SunTag-24×MBS by HindIII and NheI. For the **dual-color GA RAN reporter** (pcDNA5-TO-24×rHA-stop×3-(GGGGCC)<sub>70</sub>-24×SunTag-24×MBS-12×PBS), the (GGGGCC)<sub>70</sub>-24×SunTag was cut from pCAG-Δintron-(GGGGCC)<sub>70</sub>-24×SunTag-24×MBS (GA frame) and inserted into the above **no repeat control** plasmid by NotI and NheI. The stop codons in all three reading frames were subsequently placed between HindIII and NotI. To make **HA-GA-SunTag readthrough reporter**, the stop codon in GA frame between HindIII and NotI was removed to make HA in frame with GA reading frame of (GGGGCC)<sub>70</sub>. For frameshift **GA to GR reporter**, (GGGGCC)<sub>70</sub>-24×SunTag (GR frame) was cut from pCAG-Δintron-(GGGGCC)<sub>70</sub>-24×SunTag-24×MBS (GR frame) and inserted into **HA-GA-SunTag readthrough reporter** by NotI and NheI. Similarly, to make **GA to GP reporter**, (GGGGCC)<sub>70</sub>-24×SunTag (GP frame) was cut from pCAG-Δintron-(GGGGCC)<sub>70</sub>-24×SunTag-24×MBS (GP frame) and inserted into **HA-GA-SunTag readthrough reporter** by NotI and NheI.

**Luciferase reporters for frameshift.** For dual-luciferase frameshift reporters, FLuc and NLuc were PCR amplified from template. To keep NLuc in frame, in +1 frame and in +2 frame with FLuc, one or two nucleotides were introduced into the linker sequence GCTTGGCGCCGCTAGGTAGC, GCTTGGCGCCGCTAGGTAGC, GC TTGGCGCCGCTTAGGTAGC, respectively. pcDNA5-FRT-TO vector was digested by HindIII/NotI. The FLuc, NLuc (0, +1, +2 frame) and the digested vector were ligated together via Gibson assembly. Then the (GGGGCC)<sub>70</sub> repeat was inserted into the vector via BssHII and NotI.

**Reporters to measure the stability of different DPRs.** To study the stability of different DPRs, DNA fragments encoding GA50, GP50 and GR50 with randomized codon were synthesized (Genewiz). The HiBiT tag with ATG start codon and 2× Flag were introduced to the N-terminal and C-terminal of those DNA fragments, respectively. Next, the whole fragment was cloned into DOX inducible lentiviral vector by NheI and AgeI to generate Lenti-TRE3G-HiBiT-GA50/GP50/GR50-2×Flag.

**AUG-GGGGCC<sub>70</sub> (GR)-EGFP reporter.** To construct GR-EGFP reporter for CRISPR screening, HiBiT-HA tag with ATG start codon (AAGCTTACCATTGGTGGAGCGGCTGGCGGCTGTTCAGAA-GATTAGCGGTTCAAGTGGATACCCATACGACGTCCTCCCA-GACTACGCTGGGTAC) was introduced into the 5' end of EGFP by PCR amplification. The fragment was then inserted into pcDNA5-FRT-TO vector via HindIII and NsiI. Next, (GGGGCC)<sub>70</sub> repeats were placed between HA tag and EGFP by KpnI and XhoI.

**In vitro translation reporter.** The K0 and K20 reporters were constructed by PCR amplification from pmGFP-P2A-K0-P2A-RFP and pmGFP-P2A-K20-P2A-RFP (Addgene plasmid # 105686 and # 105688)<sup>80</sup> with N-terminal HiBiT-HA tag and a C-terminal 2× Flag tag. The amplified sequence was inserted into pBluescript II SK(+) by HindIII and NotI. The GGGGCC<sub>40</sub> fragment was cut from the pBS-GGGGCC<sub>40</sub> plasmid<sup>81</sup> via KpnI and EcoRV and cloned into the same sites of the above K0 reporter to produce the GGGGCC<sub>40</sub>(GA) reporter. The GGGGCC<sub>40</sub> fragment was cloned into the K0 reporter at KpnI and Sall (blunt) sites to generate the GGGGCC<sub>40</sub>(GR) reporter.

### Cell culture and transfection

U-2 OS (American Type Culture Collection HTB-96), HEK293T (American Type Culture Collection CRL-1573), HeLa (American Type Culture Collection CCL-2), HeLa Flp-In (Thermo Fisher), and SH-SY5Y (American Type Culture Collection CRL-2266) cells were cultured in DMEM supplemented with 10% (v/v) FBS, 100 U/mL penicillin, and 100 μg/mL streptomycin.

i<sup>3</sup>N iPSC<sup>98</sup> were grown in E8 medium (Thermo Fisher, A15169-01) on Matrigel (Corning, 354277) coated plates. For neuron differentiation, iPSC cells were first seeded and cultured in Neuronal Induction Medium (DMEM/F12 containing N-2 supplement, Nonessential amino acids, Gluta-MAX, 10 μM Y-27623 and 2 μg/mL doxycycline) for 3 days. Then cells were dissociated with Accutase (MilliporeSigma, A6964) and seeded on a 24-well plate pre-coated with 0.1 mg/mL poly-L-ornithine (Sigma, P3655) at a density of 3 × 10<sup>5</sup> per well. Cells were maintained in BrainPhys Neuronal Medium (STEMCELL Technologies), containing B-27 supplement, 10 ng/mL BDNF, 10 ng/mL neurotrophin-3, and 1 μg/mL laminin, with half-media changes every other day until harvest.

To generate stable cell lines, U-2 OS cells, U2PA stably expressing stdMCP-HaloTag and scFv-sfGFP for initiation reporters, U-2 OS stably expressing stdMCP-HaloTag-CAAX and scFv-sfGFP for initiation and elongation reporters (live cell imaging), U-2 OS stably expressing stdMCP-tagRFPT, stdPCP-CAAX, scFv-sfGFP (SunTag targeting), and Fb-HA-HaloTag (HA targeting for frameshift reporters)

were transfected with linearized reporter constructs by nucleofection (Lonza nucleofector 2b, VACA-1003), and followed by drug selection (200 µg/mL hygromycin or 1 µg/mL puromycin) for 3 weeks. HeLa Flp-In stable cell lines expressing bicistronic splicing reporters with CGG mutation were generated as described before<sup>66</sup>. For transient transfection experiments, TransIT<sup>®</sup>-LT1 Transfection Reagent (Mirus Bio) or X-tremeGene HP transfection reagent (Roche, 6366236001) was used for plasmid transfection, and Lipofectamine<sup>®</sup> RNAiMAX (Invitrogen) was used for siRNAs transfection. ON-TARGETplus pooled siRNAs of *ZNF598*, *PELO*, and nontargeting control (GE Dharmacon) were transfected twice at 0.25 nM, and cells were imaged 24 h later after the second transfection.

### Dual-luciferase assay

Nano-Glo Dual Luciferase Assay system (Promega, N1620) was used to measure the NLuc and FLuc luciferase activities in cells transfected with frameshift reporters according to the manufacturer's instructions. NLuc levels were normalized to FLuc in each sample for comparison. Three biological replicates were included at each condition.

### DPR stability measurement

To study the stability of different DPRs, Dox inducible lentiviral plasmids expressing HiBiT tagged GAS0/GP50/GR50 were packaged into lentivirus. Neurons were infected on day 4 post-differentiation with 100% MOI. On day 6, DPR expression was induced by adding doxycycline (2 µg/mL). Doxycycline was removed from day 8, and cells were collected every day (including day 8) till day 14. Cells without doxycycline induction on day 6 were also collected on day 8 as a non-induction control. Then the HiBiT activity, which represents the levels of DPRs, was measured by the Nano-Glo<sup>®</sup> HiBiT Lytic Detection System (Promega, N3030) according to the manufacturer's instructions. To compare the relative protein level among different samples, the HiBiT activity was normalized to the total protein level measured by BCA assay (Thermo Fisher Scientific).

### In vitro transcription and translation

RNA was in vitro transcribed from NotI linearized vector using MEGAscript T7 Transcription Kit (Thermo Fisher Scientific, AM1333) in the presence of m<sup>7</sup>GpppG cap analog (NEB, S1404). Reactions were carried out at 37 °C for 3 h. Synthesized RNAs were purified by TRIzol reagent. RNA size and quality were verified by agarose gel. Then, 500 ng mRNAs were in vitro translated with Flexi Rabbit Reticulocyte Lysate System (Promega, L4540), following the manufacturer's protocol. The reactions were incubated at 30 °C for 30 min before termination by incubation at 4 °C. Next, 10 µL reaction was mixed with 40 µL 2x Laemmli sample buffer and heated at 70 °C for 15 min, then 20 µL was loaded on 15% polyacrylamide gel. Expression was detected by Nano-Glo<sup>®</sup> HiBiT Blotting detection kit (Promega, PRN2410).

### Immunoblotting

For immunoblotting, goat anti-mouse or anti-rabbit IgG HRP-conjugated (1:5000, GE healthcare, #NA934V and #NA934V) was used along with chemiluminescent detection reagents (Thermo Scientific). The primary antibody included GAPDH (1:1000, Cell Signaling Technology, #2118), FLAG (1:1000, Sigma, #F1804), ZNF598 (1:1000, GeneTex, #GTX119246), PELO (1:1000, proteintech, #10582-1-AP). For HiBiT tag detecting, Nano-Glo<sup>®</sup> HiBiT Blotting System (Promega, N2410) was used according to the manufacturer's instructions. Uncropped blots for Fig. 5b and Supplementary Fig. 7d, e can be found in the Source Data file.

### CRISPR screening

Guide RNAs (gRNAs) targeting genes encoding 1906 potential RNA-binding proteins (RBPs), ribosome proteins and nontargeting negative control gRNAs were chosen from the genome-wide guide RNA

library to make RBP sub-library as described previously<sup>99</sup>. The RBP sub-library was packaged into lentivirus using 293T cells. About 30 million HeLa Flp-In cells expressing AUG-(GGGGCC)<sub>70</sub>-EGFP (GR in frame with AUG start codon and EGFP) were infected with the library at an MOI of 0.4. After infection, cells were placed under puromycin selection at 1 µg/mL for 3 days. After drug selection, 2 µg/mL doxycycline was applied to cells to induce the reporter expression. Cells were split every 3 days or as needed. Throughout the screen, the cells were maintained at ≥1000× gRNA coverage to keep the library complexity. After 14 days of culture, 20 million cells were analyzed by fluorescence-activated cell sorting (FACS), and the top 5% GFP-high and 5% GFP-low cell populations were collected<sup>81</sup>. Genomic DNA was extracted from the individual cell population using the QIAamp Blood Mini Kit (Qiagen). The DNA fragment libraries containing gRNAs were generated by PCR and subjected to high-throughput sequencing on the Illumina Nextseq platform. The enrichment of individual gRNAs in GFP-high vs GFP-low population from biological duplicates was analyzed by castLE<sup>100</sup>.

### Single-molecule fluorescence in situ hybridization with immunofluorescence (smFISH-IF)

The detailed protocol for smFISH IF assay was described in detail<sup>101</sup>. Briefly, 20mer single-stranded DNA oligoes were ordered from Integrated DNA Technology in a 96-well plate format. The oligoes were pooled together and labeled with Cy3-NHS ester (Lumiprobe 11020) according to the described protocol<sup>102</sup>. Cells were seeded on 18 mm cover glass and treated with 500 µM indole-3-acetic acid (Sigma, I2886) to degrade AID-labeled reporters on the day before the experiment. During the experiment, the cells were fixed with 4% PFA (Electron Microscopy Sciences, 50-980-492) for 10 min, blocked in 50 mM Glycine (Thermo Fisher Scientific, BP 381-1), in phosphate buffer saline (PBS, Corning, 46-013-CM) buffer for 10 min, then permeabilized with 0.1% Triton-X (Sigma-Aldrich, T8787), and blocked with BSA 5 mg/mL (VWR, VWRV033) for 10 min, all at room temperature. The cells were pre-equilibrated with hybridization buffer (10% formamide, 2×SSC (Roche, 11666681001), BSA 5 mg/mL), and incubated with 100 nM FISH probes targeting transcript of interest and primary antibodies targeting the sf-GFP (Aves Labs, GFP-1010) in hybridization buffer (10% formamide (Acros Organics, AC205821000), 2×SSC, BSA 2 mg/mL) for 3 h at 37 °C. The excess probes and antibodies were washed away with wash buffer (10% formamide, 2×SSC), and the samples were incubated for 20 min at 37 °C with secondary antibodies (Goat anti-Chicken IgY (H+L) Alexa Fluor 647, Thermo Fisher Scientific, A21449). Finally, the samples were washed with PBS and mounted on a microscope slide using ProLong Diamond Antifade Mountant containing DAPI for nucleus staining (Thermo Fisher Scientific, P36962).

### Microscopy

The fixed smFISH-IF samples were imaged on an automated inverted Nikon Ti-2 wide-field microscope equipped with a 60× 1.4NA oil immersion objective lens (Nikon), Spectra X LED light engine (Lumencor), and Orca 4.0 v2 sCMOS camera (Hamamatsu).

The live cell imaging was conducted on a customized wide-field microscope built around a Nikon Ti-E stand. The microscope was equipped with three separate EMCCD cameras (Andor iXon Ultra 897) with ultra-flat 2 mm thick imaging splitting dichroic mirrors (T565LPXR-UF2, T640LPXR-UF2). Each of the cameras was equipped with band pass emission filter (ET525/50m, ET595/50m, and ET655lp). The light source of the microscope is an LU-n4 laser launch (Nikon) with four wavelengths: 405, 488, 561, and 640 nm. The excitation was done through H-TIRF with 100× 1.49NA oil immersion TIRF objective (Nikon). The excitation laser and fluorescence emission were separated by a quad band dichroic mirror (Chroma, ET-405/488/561/640 nm). The microscope was also equipped with an

automated linear encoded XY-stage with Piezo-Z (Applied Scientific Instrumentation).

### Live-cell imaging

**Imaging steady state translation.** Cells stably expressing reporters were seeded in 35 mm cover glass (Cellvis, D35-20-1.5-N) and grown overnight. Then, 500  $\mu$ M 3-Indoleacetic acid (Sigma, I2886) was added to the culture medium throughout the experiment. One hour before imaging, cells were incubated with 100 nM Janelia Fluor Halo dye (JF646, provided by Lavis lab) for 30 min<sup>103</sup> and washed with warm DMEM for 30 min to remove excessive dyes. Prior to imaging, the medium was changed to Leibovitz's L-15 media without phenol red (Gibco, 21083-027) supplemented with 10% FBS. The sample was kept at 37 °C with humidity controlled by the Tokai Hit stage top incubator during the imaging session. For transient transfection of translation reporters, 400 ng plasmids were transfected 12 h before imaging with X-tremeGene HP transfection reagent (Roche, 6366236001). The rest of the protocol is identical to the imaging of stably integrated reporters.

**Ribosome runoff experiment.** Cells with translation sites were identified and imaged. Half of the warm imaging media was mixed with harringtonine (Cayman Chemical, #15361) and added back to the imaging dish to a final concentration of 3  $\mu$ g/mL. Live imaging was started exactly 1 min after adding the drug. The RNA and protein channel were acquired every 10 s for 30 min.

**Fluorescence recovery after photobleaching for single translation sites.** After identifying the target translation sites, cells were imaged for 30 s prior to bleaching the translation site with a focused 488 nm laser controlled by Bruker microscanner. After bleaching the site, the cell was imaged every 5 s for 10 min.

### Data analysis

**smFISH-IF data analysis.** All fixed cell imaging data analysis was performed using existing and custom Matlab programs as described<sup>101</sup>. In brief, the coordinates and amplitudes of smFISH and IF signals were identified independently using FISH-Quant<sup>104</sup>. FISH-Quant uses Gaussian fitting to determine the spot localization and its intensity amplitude. After the identification of single mRNA signals, a bounding box of 11  $\times$  11 pixels was drawn around the cytoplasmic mRNA and checked as a potential translation site. The transcription site fitting algorithm of FISH-Quant was used to assess the translation site intensity of the corresponding mRNA. To determine the number of nascent peptides per transcript, the single mature SunTag proteins in the cytoplasm were identified in a similar fashion as RNA transcripts, and the median of single protein intensities was used to normalize the integrated intensities of translation sites.

**Tracking translation of single mRNAs in live cells.** All live cell imaging analysis was performed using existing software packages written in Matlab (MathWorks) supplemented with custom script. Single particles of RNA or translation sites were identified with Airlocalize<sup>105</sup> and tracked by u-track<sup>106</sup>. Only tracks longer than five frames were further analyzed. Translation sites and mRNAs that overlap in space and time are linked together. A temporal overlap of at least five frames was required to link RNA and protein tracks. All tracks were inspected, and mRNAs with merging or splitting events were discarded.

For the ribosome runoff experiment, we measured the time for each translating mRNA to completely lose its translation intensity, starting from the moment of adding harringtonine. This is defined as ribosome runoff time:  $t$  (Fig. 4d, e). Practically, if the translation site intensity fell below 10% of the maximum intensity of the trace, it was deemed runoff. If TLS and the corresponding mRNA disappear together, the molecules were not included in the analysis because

these events may be loss of the tracked mRNA. The RNA with low initial protein signal in the first four frames, below 10% of maximum protein intensity, were discarded, as they were not visibly translating at the beginning of the assay. We compiled all runoff times into a translation site survival probability curve: the percentage of translating mRNAs was plotted as a function of time after adding Harringtonine. We used GraphPad Prism software to plot the results of runoff assays as Kaplan–Meier curves at the same time. We used Greenwood's formula to calculate the 95% confidence bounds in Fig. 4. To test the statistical significance between the survival curves, we performed Cox proportional hazards analysis at 15 min using GraphPad Prism (v8.2.1)<sup>107</sup>.

**Calculating elongation time through repeats.** The repeats are just a small fraction of the reporter, but the measured ribosome runoff time depends on elongation through the whole reporter. We therefore estimated the elongation time through the GGGGCC repeats, which is inversely proportional to the elongation speed, using the measured ribosome runoff time. Theoretically, the runoff time is the time it takes the last ribosome to translate through the entire construct. We used the median of the runoff time to estimate the total elongation time  $t$ .

Assume that the lengths of the inserted sequence (repeat or NLuc) and the rest of the construct are  $L_I$  and  $L_O$ , respectively (Fig. 4g). The elongation speed through the insert is  $v_I$  and through the rest of the protein is  $v_O$ . The time spent on the insert and the rest of the protein is  $t_O$  and  $t_I$ , respectively. The total runoff time is equal to the sum of elongation times through the two regions.

$$t = t_I + t_O = \frac{L_I}{v_I} + \frac{L_O}{v_O} \quad (S1)$$

Rearrange the equation, we can derive

$$\frac{v_O}{v_I} = \frac{t - \frac{L_O}{v_O}}{\frac{L_I}{v_O}} \quad (S2)$$

which is the ratio between elongation speeds of a general protein sequence and of repeats. Note the variables at the right-hand side of the equation are all experimentally measurable quantities. For NLuc, the median runoff time is  $t = 4$  min 55 s, from which we can estimate the elongation speed of  $v_O = 3.644$  /s, consistent with previously published data<sup>58,59,73</sup>. The elongation speeds through different frames can be estimated according to Eq. S2.

### Proteomic data analysis

The analysis of proteomic data followed the previous pipeline<sup>108</sup>. Briefly, the data matrix was extracted from the Answer ALS database, and the iPSN lines used are listed in Supplementary Data 2. The acquisition of AALS data followed the protocols approved by the Johns Hopkins Institutional Review Board. The protection of health information was in accordance with HIPAA and the HITECH Act. Proteins that had more than four missing values across samples were removed. The protein levels were normalized to the AEs iPSN batch control line for each batch. \* $P < 0.05$ , Wilcox test.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. The source data underlying Figs. 1–6, Supplementary Figs. 1–7 are provided as a Source Data file. The CRISPR screen sequencing data (Fig. 6a) is deposited in

the NCBI BioProject database under the accession number [PRJNA905075](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA905075). Source data are provided with this paper.

## Code availability

The custom code used in the analysis is freely available at <https://github.com/binwulab/>.

## References

- DeJesus-Hernandez, M. et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
- Renton, A. E. et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
- Majounie, E. et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol.* **11**, 323–330 (2012).
- Van Mossevelde, S., van der Zee, J., Cruts, M. & Van Broeckhoven, C. Relationship between C9orf72 repeat size and clinical phenotype. *Curr. Opin. Genet. Dev.* **44**, 117–124 (2017).
- Gitler, A. D. & Tsuiji, H. There has been an awakening: emerging mechanisms of C9orf72 mutations in FTD/ALS. *Brain Res.* **1647**, 19–29 (2016).
- Kumar, V., Hasan, G. M. & Hassan, M. I. Unraveling the role of RNA mediated toxicity of C9orf72 repeats in C9-FTD/ALS. *Front. Neurosci.* **11**, 711 (2017).
- Gendron, T. F. & Petrucelli, L. Disease mechanisms of C9ORF72 repeat expansions. *Cold Spring Harb. Perspect. Med.* **8**, a024224 (2018).
- Taylor, J. P., Brown, R. H. & Cleveland, D. W. Decoding ALS: from genes to mechanism. *Nature* **539**, 197–206 (2016).
- Balendra, R. & Isaacs, A. M. C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat. Rev. Neurol.* **14**, 544–558 (2018).
- Mori, K. et al. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol.* **126**, 881–893 (2013).
- Zu, T. et al. Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl Acad. Sci. USA* **108**, 260–265 (2011).
- Gendron, T. F. et al. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.* **126**, 829–844 (2013).
- Mackenzie, I. R. et al. Quantitative analysis and clinicopathological correlations of different dipeptide repeat protein pathologies in C9ORF72 mutation carriers. *Acta Neuropathol.* **130**, 845–861 (2015).
- Freibaum, B. D. & Taylor, J. P. The role of dipeptide repeats in C9ORF72-related ALS-FTD. *Front. Mol. Neurosci.* **10**, 35 (2017).
- Schmitz, A., Marques, J. P., Oertig, I., Maharjan, N. & Saxena, S. Emerging perspectives on dipeptide repeat proteins in C9ORF72 ALS/FTD. *Front. Cell. Neurosci.* **15**, 637548 (2021).
- Jovicic, A. et al. Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci.* **18**, 1226–1229 (2015).
- Kwon, I. et al. Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145 (2014).
- Wen, X. et al. Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death. *Neuron* **84**, 1213–1225 (2014).
- Lopez-Gonzalez, R. et al. Poly(GR) in C9ORF72-related ALS/FTD compromises mitochondrial function and increases oxidative stress and DNA damage in iPSC-derived motor neurons. *Neuron* **92**, 383–391 (2016).
- May, S. et al. C9orf72 FTL/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol.* **128**, 485–503 (2014).
- Shi, K. Y. et al. Toxic PRn poly-dipeptides encoded by the C9orf72 repeat expansion block nuclear import and export. *Proc. Natl Acad. Sci. USA* **114**, E1111–E1117 (2017).
- Bennion Callister, J., Ryan, S., Sim, J., Rollinson, S. & Pickering-Brown, S. M. Modelling C9orf72 dipeptide repeat proteins of a physiologically relevant size. *Hum. Mol. Genet.* **25**, 5069–5082 (2016).
- Boeynaems, S. et al. Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell* **65**, 1044–1055.e1045 (2017).
- Lee, K. H. et al. C9orf72 dipeptide repeats impair the assembly, dynamics, and function of membrane-less organelles. *Cell* **167**, 774–788.e717 (2016).
- Lin, Y. et al. Toxic PR Poly-dipeptides encoded by the C9orf72 repeat expansion target LC domain polymers. *Cell* **167**, 789–802.e712 (2016).
- Mizielinska, S. et al. C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* **345**, 1192–1194 (2014).
- Zhang, Y. J. et al. Poly(GR) impairs protein translation and stress granule dynamics in C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis. *Nat. Med.* **24**, 1136–1142 (2018).
- Davidson, Y. et al. Neurodegeneration in frontotemporal lobar degeneration and motor neurone disease associated with expansions in C9orf72 is linked to TDP-43 pathology and not associated with aggregated forms of dipeptide repeat proteins. *Neuropathol. Appl. Neurobiol.* **42**, 242–254 (2016).
- McEachin, Z. T. et al. Chimeric peptide repeat species contribute to divergent dipeptide repeat pathology in c9ALS/FTD and SCA36. *Neuron* **107**, 292–305.e296 (2020).
- Gao, F. B., Richter, J. D. & Cleveland, D. W. Rethinking unconventional translation in neurodegeneration. *Cell* **171**, 994–1000 (2017).
- Ebbert, M. T. W. et al. Long-read sequencing across the C9orf72 ‘GGGGCC’ repeat expansion: implications for clinical use and genetic discovery efforts in human disease. *Mol. Neurodegener.* **13**, 46 (2018).
- Dever, T. E. & Green, R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **4**, a013706 (2012).
- Sonobe, Y. et al. Translation of dipeptide repeat proteins from the C9ORF72 expanded repeat is associated with cellular stress. *Neurobiol. Dis.* **116**, 155–165 (2018).
- Sonobe, Y. et al. A *C. elegans* model of C9orf72-associated ALS/FTD uncovers a conserved role for eIF2D in RAN translation. *Nat. Commun.* **12**, 6025 (2021).
- Goodman, L. D. et al. eIF4B and eIF4H mediate GR production from expanded G4C2 in a *Drosophila* model for C9orf72-associated ALS. *Acta Neuropathol. Commun.* **7**, 62 (2019).
- Cheng, W. et al. C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2alpha phosphorylation. *Nat. Commun.* **9**, 51 (2018).
- Green, K. M. et al. RAN translation at C9orf72-associated repeat expansions is selectively enhanced by the integrated stress response. *Nat. Commun.* **8**, 1–13 (2017).
- Tabet, R. et al. CUG initiation and frameshifting enable production of dipeptide repeat proteins from ALS/FTD C9ORF72 transcripts. *Nat. Commun.* **9**, 1–14 (2018).

39. Ingolia, N. T., Ghaemmghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
40. Eastman, G., Smircich, P. & Sotelo-Silveira, J. R. Following ribosome footprints to understand translation at a genome wide level. *Comput. Struct. Biotechnol. J.* **16**, 167–176 (2018).
41. Gorochowski, T. E., Avciilar-Kucukgoze, I., Bovenberg, R. A., Roubos, J. A. & Ignatova, Z. A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. *ACS Synth. Biol.* **5**, 710–720 (2016).
42. Riba, A. et al. Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates. *Proc. Natl Acad. Sci. USA* **116**, 15023–15032 (2019).
43. Quax, T. E., Claassens, N. J., Söll, D. & van der Oost, J. Codon bias as a means to fine-tune gene expression. *Mol. Cell* **59**, 149–161 (2015).
44. Chaney, J. L. & Clark, P. L. Roles for synonymous codon usage in protein biogenesis. *Annu. Rev. Biophys.* **44**, 143–166 (2015).
45. Liu, Y., Yang, Q. & Zhao, F. Synonymous but not silent: the codon usage code for gene expression and protein folding. *Annu. Rev. Biochem.* **90**, 375 (2021).
46. Presnyak, V. et al. Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
47. Brandman, O. & Hegde, R. S. Ribosome-associated protein quality control. *Nat. Struct. Mol. Biol.* **23**, 7–15 (2016).
48. Joazeiro, C. A. Mechanisms and functions of ribosome-associated protein quality control. *Nat. Rev. Mol. Cell Biol.* **20**, 368–383 (2019).
49. Filbeck, S., Cerullo, F., Pfeffer, S. & Joazeiro, C. A. Ribosome-associated quality-control mechanisms from bacteria to humans. *Mol. Cell* **82**, 1451–1466 (2022).
50. Brandman, O. et al. A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* **151**, 1042–1054 (2012).
51. Conlon, E. G. et al. The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains. *Elife* **5**, e17820 (2016).
52. Su, Z. et al. Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS. *Neuron* **83**, 1043–1050 (2014).
53. Requião, R. D., de Souza, H. J. A., Rossetto, S., Domitrovic, T. & Palhano, F. L. Increased ribosome density associated to positively charged residues is evident in ribosome profiling experiments performed in the absence of translation inhibitors. *RNA Biol.* **13**, 561–568 (2016).
54. Li, S. et al. Quality-control mechanisms targeting translationally stalled and C-terminally extended poly (GR) associated with ALS/FTD. *Proc. Natl Acad. Sci. USA* **117**, 25104–25115 (2020).
55. Park, J. et al. ZNF598 co-translationally titrates poly (GR) protein implicated in the pathogenesis of C9ORF72-associated ALS/FTD. *Nucleic Acids Res.* **49**, 11294–11311 (2021).
56. Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of translation of single mRNA molecules in vivo. *Cell* **165**, 976–989 (2016).
57. Morisaki, T. et al. Real-time quantification of single RNA translation dynamics in living cells. *Science* **352**, 1425–1429 (2016).
58. Wu, B., Eliscovich, C., Yoon, Y. J. & Singer, R. H. Translation dynamics of single mRNAs in live cells and neurons. *Science* **352**, 1430–1435 (2016).
59. Wang, C., Han, B., Zhou, R. & Zhuang, X. Real-time imaging of translation on single mRNA transcripts in live cells. *Cell* **165**, 990–1001 (2016).
60. Pichon, X. et al. Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells. *J. Cell Biol.* **214**, 769–781 (2016).
61. Holland, A. J., Fachinetti, D., Han, J. S. & Cleveland, D. W. Inducible, reversible system for the rapid and complete degradation of proteins in mammalian cells. *Proc. Natl Acad. Sci. USA* **109**, E3350–E3357 (2012).
62. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kane-maki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* **6**, 917–922 (2009).
63. Wu, B. et al. Synonymous modification results in high-fidelity gene expression of repetitive protein and nucleotide sequences. *Genes Dev.* **29**, 876–886 (2015).
64. Wu, B., Chao, J. A. & Singer, R. H. Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophys. J.* **102**, 2936–2944 (2012).
65. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**, 79–88 (2006).
66. Cheng, W. et al. C9ORF72 GGGGCC repeat-associated non-AUG translation is upregulated by stress through eIF2 $\alpha$  phosphorylation. *Nat. Commun.* **9**, 1–12 (2018).
67. Jacobs, J. L., Belew, A. T., Rakauskaitė, R. & Dinman, J. D. Identification of functional, endogenous programmed-1 ribosomal frameshift signals in the genome of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **35**, 165–174 (2007).
68. Yan, S., Wen, J.-D., Bustamante, C. & Tinoco, I. Jr Ribosome excursions during mRNA translocation mediate broad branching of frameshift pathways. *Cell* **160**, 870–881 (2015).
69. Lyon, K., Aguilera, L. U., Morisaki, T., Munsky, B. & Stasevich, T. J. Live-cell single RNA imaging reveals bursts of translational frameshifting. *Mol. Cell* **75**, 172–183.e179 (2019).
70. Livingston, N. M. et al. Bursting translation on single mRNAs in live cells. *Mol. Cell* **83**, 2276–2289.e2211 (2023).
71. Zhao, N. et al. A genetically encoded probe for imaging nascent and mature HA-tagged proteins in vivo. *Nat. Commun.* **10**, 1–16 (2019).
72. Wang, S. et al. Nuclear export and translation of circular repeat-containing intronic RNA in C9ORF72-ALS/FTD. *Nat. Commun.* **12**, 1–14 (2021).
73. Goldman, D. H., Livingston, N. M., Movsik, J., Wu, B. & Green, R. Live-cell imaging reveals kinetic determinants of quality control triggered by ribosome stalling. *Mol. Cell* **81**, 1830–1840.e1838 (2021).
74. FRESNO, M., JIMÉNEZ, A. & VÁZQUEZ, D. Inhibition of translation in eukaryotic systems by harringtonine. *Eur. J. Biochem.* **72**, 323–330 (1977).
75. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147**, 789–802 (2011).
76. Pavlov, M. Y. et al. Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc. Natl Acad. Sci. USA* **106**, 50–54 (2009).
77. Buskirk, A. R. & Green, R. Ribosome pausing, arrest and rescue in bacteria and eukaryotes. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, 20160183 (2017).
78. Matsuo, Y. et al. Ubiquitination of stalled ribosome triggers ribosome-associated quality control. *Nat. Commun.* **8**, 1–14 (2017).
79. Loveland, A. B. et al. Ribosome inhibition by C9ORF72-ALS/FTD-associated poly-PR and poly-GR proteins revealed by cryo-EM. *Nat. Commun.* **13**, 1–13 (2022).
80. Juskiewicz, S. & Hegde, R. S. Initiation of quality control during poly (A) translation requires site-specific ribosome ubiquitination. *Mol. Cell* **65**, 743–750.e744 (2017).
81. Cheng, W. et al. CRISPR-Cas9 screens identify the RNA helicase DDX3X as a repressor of C9ORF72 (GGGGCC) $_n$  repeat-associated non-AUG translation. *Neuron* **104**, 885–898.e888 (2019).
82. Doma, M. K. & Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561–564 (2006).



83. Letzring, D. P., Wolf, A. S., Brule, C. E. & Grayhack, E. J. Translation of CGA codon repeats in yeast involves quality control components and ribosomal protein L1. *RNA* **19**, 1208–1217 (2013).
84. Garzia, A. et al. The E3 ubiquitin ligase and RNA-binding protein ZNF598 orchestrates ribosome quality control of premature polyadenylated mRNAs. *Nat. Commun.* **8**, 1–10 (2017).
85. Juszkiwicz, S. et al. ZNF598 is a quality control sensor of collided ribosomes. *Mol. Cell* **72**, 469–481.e467 (2018).
86. Pisareva, V. P., Skabkin, M. A., Hellen, C. U., Pestova, T. V. & Pisarev, A. V. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J.* **30**, 1804–1817 (2011).
87. Baxi, E. G. et al. Answer ALS, a large-scale resource for sporadic and familial ALS combining clinical and multi-omics data from induced pluripotent cell lines. *Nat. Neurosci.* **25**, 226–237 (2022).
88. Almeida, S. et al. Production of poly (GA) in C9ORF72 patient motor neurons derived from induced pluripotent stem cells. *Acta Neuropathol.* **138**, 1099–1101 (2019).
89. Doerfel, L. K. et al. Entropic contribution of elongation factor P to proline positioning at the catalytic center of the ribosome. *J. Am. Chem. Soc.* **137**, 12997–13006 (2015).
90. Shoemaker, C. J. & Green, R. Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* **19**, 594–601 (2012).
91. Inada, T. Quality controls induced by aberrant translation. *Nucleic Acids Res.* **48**, 1084–1096 (2020).
92. Arthur, L. L. et al. Translational control by lysine-encoding A-rich sequences. *Sci. Adv.* **1**, e1500154 (2015).
93. Park, J. et al. ZNF598 co-translationally titrates poly(GR) protein implicated in the pathogenesis of C9ORF72-associated ALS/FTD. *Nucleic Acids Res.* **49**, 11294–11311 (2021).
94. Viera Ortiz, A. P. et al. Impaired ribosome-associated quality control of C9orf72 arginine-rich dipeptide-repeat proteins. *Brain* **146**, 2897–2912 (2023).
95. Mori, K. et al. 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 Neuropathol.* **125**, 413–423 (2013).
96. Westergard, T. et al. Repeat-associated non-AUG translation in C9orf72-ALS/FTD is driven by neuronal excitation and stress. *EMBO Mol. Med.* **11**, e9423 (2019).
97. Licata, N. V. et al. C9orf72 ALS/FTD dipeptide repeat protein levels are reduced by small molecules that inhibit PKA or enhance protein degradation. *EMBO J.* **41**, e105026 (2022).
98. Tian, R. et al. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* **104**, 239–255.e212 (2019).
99. Morgens, D. W. et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 1–8 (2017).
100. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 (2016).
101. Latallo, M. J., Livingston, N. M. & Wu, B. Translation imaging of single mRNAs in established cell lines and primary cultured neurons. *Methods* **162**, 12–22 (2019).
102. Gaspar, I., Wippich, F. & Ephrussi, A. Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* **23**, 1582–1591 (2017).
103. Grimm, J. B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250 (2015).
104. Mueller, F. et al. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nat. Methods* **10**, 277–278 (2013).
105. Lionnet, T. et al. A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nat. Methods* **8**, 165–170 (2011).
106. Jaqaman, K. et al. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).
107. Machin, D., Cheung, Y. B. & Parmar, M. *Survival Analysis: A Practical Approach* (Wiley, 2006).
108. Li, Y. et al. Globally reduced N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) in C9ORF72-ALS/FTD dysregulates RNA metabolism and contributes to neurodegeneration. *Nat. Neurosci.* **26**, 1328–1338 (2023).

## Acknowledgements

We thank the Sun and Wu lab members for their helpful discussion. This work is supported by Johns Hopkins University School of Medicine Discovery Fund Synergy Award, National Institute of Health (RFINS113820 to S.S. and B.W.; RO1GM136897 to B.W.; RO1NS107347, R21AG072078 and RFINS127925 to S.S.), National Science Foundation (MCB 1817447 to B.W.), Pew Charitable Trust (00030601 to B.W.), and the Packard Center for ALS Research. M.J.L. was supported by the NIH Training Grant (T32 GM008403), and N.M.L. and B.L. were supported by the NIH Training Grant (T32 GM007445).

## Author contributions

M.J.L., S.W., D.D., S.S. and B.W. contributed to the experimental design and interpretation of the study and wrote the manuscript with input from other authors. M.J.L., S.W., D.D., B.N., N.M.L. and R.W. conducted all experiments under the mentorship of S.S. and B.W. N.Z., T.J.S. and M.C.B. edited the manuscript and provided key reagents.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-41339-x>.

**Correspondence** and requests for materials should be addressed to Shuying Sun or Bin Wu.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. <sup>2</sup>Center for Cell Dynamics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. <sup>3</sup>Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. <sup>4</sup>Brain Science Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. <sup>5</sup>Department of Biochemistry and Molecular Genetics, University of Colorado-Anschutz Medical Campus, Aurora, CO 80045, USA. <sup>6</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>7</sup>Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. <sup>8</sup>Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. <sup>9</sup>These authors contributed equally: Malgorzata J. Latallo, Shaopeng Wang, Daoyuan Dong. ✉ e-mail: [shuying.sun@jhmi.edu](mailto:shuying.sun@jhmi.edu); [bwu20@jhmi.edu](mailto:bwu20@jhmi.edu)