Flanking sequences regulate the toxicity, non-ATG translation, and aggregation of RNA with tandem CAG Repeats

Michael R. Das¹,², Yeonji Chang¹, Reuben Saunders³, Nan Zhang³, Colson Tomberlin¹, Ronald D. Vale³,⁴, and Ankur Jain¹,²

¹Whitehead Institute for Biomedical Research, 455 Main Street, Cambridge, MA 02142, USA.
²Department of Biology, Massachusetts Institute of Technology, 31 Ames Street, Cambridge, MA 02139, USA.
³Department of Cellular and Molecular Pharmacology, University of California, San Francisco, 600 16th Street, San Francisco, CA 94158, USA.
⁴Howard Hughes Medical Institute, University of California, San Francisco, 600 16th Street, San Francisco, CA
Summary

Nucleotide repeat-expansions cause several neurodegenerative disorders, including Huntington's disease and spinocerebellar ataxia. The expanded repeat-containing RNA transcribed from the affected loci agglomerate in the nucleus as pathogenic foci. Here we demonstrate that depending on their surrounding sequence context, RNAs with expanded CAG repeats can also undergo nuclear export and aggregate in the cytoplasm. Cytoplasmic aggregation of repeat-containing RNA coincides with several disease hallmarks, including repeat-associated non-AUG (RAN) translation, mislocalization of RNA binding proteins, and cell toxicity. Interestingly, the repeat-containing RNA co-aggregate with RAN translation products. Inhibition of RAN translation prevents cytoplasmic RNA aggregation and also alleviates cell toxicity. Our findings provide a cogent explanation for aberrant cytoplasmic localization of RNA binding proteins and implicate cis-acting flanking sequences in mediating RAN translation and disease.
Introduction

Over 40 genetic diseases are traced to the aberrant expansion of short nucleotide repeats (Khristich and Mirkin, 2020). This family of diseases, collectively referred to as nucleotide repeat expansion disorders, includes Huntington’s disease, several spinocerebellar ataxias, myotonic dystrophy, and certain forms of amyotrophic lateral sclerosis (ALS). Disease-causing repeats are found in both protein-coding as well as non-coding regions. For example, Huntington’s disease results from a CAG trinucleotide repeat expansion in exon 1 of the huntingtin gene, whereas myotonic dystrophy is caused by a CTG repeat expansion in the 3’ untranslated region of DMPK. Disease usually manifests when the number of repeats exceeds a critical threshold (Khristich and Mirkin, 2020).

Although repeat expansion diseases are monogenic disorders, the disease mechanisms are more complex than the simple loss of function of the mutated gene (Paulson, 2018). In particular, the RNA transcripts harboring pathogenic repeat expansions are suggested to promote pathogenesis via at least two routes. First, the repeat-containing RNAs can form higher-order assemblies and accumulate as ‘foci’ in the nucleus. These nuclear RNA foci sequester various RNA-binding proteins and result in transcriptome-wide RNA processing (Jiang et al., 2004; Kanadia et al., 2006; Wojciechowska and Krzyzosiak, 2011). Second, the disease-associated repeat-containing RNA can be translated without a bona fide AUG-start codon in multiple reading frames: a process referred to as repeat-associated non-AUG or RAN translation (Zu et al., 2011). The homopolymeric peptides produced upon RAN translation disrupt cellular functions and are potentially lethal to the cell (Kwon et al., 2014; Lee et al., 2016; Zhang et al., 2014).
Retention of RNA at nuclear foci could potentially limit its access to the translation apparatus in the cytoplasm. The interplay between these two RNA gain-of-function routes and their relative contribution to cell toxicity remains to be established. In addition to RNA foci and RAN translation, a common theme in repeat expansion diseases is the aberrant cytoplasmic localization of several RNA binding proteins such as TDP-43 and FUS (DeJesus-Hernandez et al., 2011; Deng et al., 2010; Schwab et al., 2008). Mechanistically, how the repeat expansion mutation results in cytoplasmic recruitment of otherwise nuclear proteins is unknown. Another peculiar feature of repeat expansion disorders is that the sequences surrounding the repeat can also influence cellular toxicity and disease (Cleary and Pearson, 2003), but the molecular basis for this phenomenon is not clear.

Here, we examined how the surrounding sequence context influences RNA toxicity. We demonstrate that RNA with expanded CAG repeats exhibit markedly different sub-cellular localization depending on their upstream flanking sequences. In most cases, incorporation of expanded CAG-repeats results in retention of the RNA at nuclear foci. However, certain upstream flanking sequences facilitate nuclear export and RAN translation of the repeat-containing RNA. RAN translation products co-aggregate with the repeat-containing RNA and accumulate in perinuclear cytoplasmic inclusions. These RNA-RAN protein inclusions sequester TDP-43, an RNA binding protein frequently mislocalized in neurodegenerative disease, as well as p62, a marker for protein aggregation. RNA with CAG repeats in surrounding sequence contexts that result in cytoplasmic aggregates are substantially more toxic than ones that only result in nuclear foci. RAN translation is required for cytoplasmic RNA aggregation, and inhibition of RAN translation rescues the cells from repeat RNA mediated toxicity. Our findings
establish a direct regulatory role of flanking sequences in determining the localization of repeat-containing RNA, causing RAN translation, and mediating cell toxicity.

Results

Flanking sequences influence the toxicity and sub-cellular localization of CAG repeat-containing RNA.

To study the effect of the surrounding sequence context on repeat-containing RNA, we developed a minimal reporter system. We generated 12 synthetic sequences (250 bp each, GC-content 30-79%, sequences in Supp. Table 1) and introduced them upstream of CAG-repeats (Fig. 1A). Multiple stop codons were incorporated in each reading frame between the test sequences and the CAG repeats to prevent readthrough translation from potential start codons in the upstream flanking sequences. Downstream of the CAG repeats, we incorporated twelve MS2 hairpins, which allowed live-cell RNA visualization via co-expression of a YFP-tagged MS2 coat protein (MS2CP-YFP) (Bertrand et al., 1998). The repeat-containing RNAs were expressed under the control of a doxycycline-inducible promoter.

After prolonged (5 days) expression of the repeat-containing RNA, we observed two broad classes of cellular phenotypes. The majority of constructs (7 out of 12 sequences tested) did not produce a discernable difference in cellular viability as compared to uninduced controls (constructs N1-N7 for nontoxic CAG repeats, Fig. 1B). This result is similar to our previous report that the expression of RNA with expanded CAG repeat near the 5’ end of the transcript (construct N0) does not produce cell growth defects over a 2-month period (Jain and Vale, 2017). Interestingly, however, five constructs induced marked cell death (constructs T1-T5 for
toxicity-inducing CAG repeats). The 5’ flanking sequences that produced cytotoxicity had a higher GC-content compared to the flanking sequences that did not induce toxicity (Supp. Fig 1A, Supp. Table 1) but beyond this modest correlation with GC-content, we have not found other sequence motifs in the toxicity inducing flanking sequences. Cellular toxicity did not correlate with RNA expression levels (Supp. Fig. 1B-C) but was dependent on the number of CAG repeats: for a construct that induced cell death (T1), decreasing the number of CAG repeats progressively diminished the toxicity associated with RNA expression (Fig. 1C, Supp. Fig. 1D). RNA with 5×CAG repeats cloned downstream of T1 flanking sequence did not produce discernible defects in cell viability as compared to uninduced controls, indicating that cell toxicity was not trivially imparted by the flanking sequence alone. In summary, our findings reveal that the sequences upstream of CAG repeats substantially impact the cytotoxicity resulting from the expression of expanded repeat-containing RNAs.

We next examined the localization of the repeat-containing RNAs at early time points (24 h after induction) before cell toxicity was evident (Supp. Fig. 1E). In the constructs that did not induce toxicity, the repeat-containing RNA accumulated at nuclear foci (Fig. 1D, Supp. Fig. 1F-I), similar to our previous observation with transcripts harboring CAG-repeats near the 5’ end of the RNA (construct N0) (Jain and Vale, 2017). Surprisingly, we discovered that in the constructs that produced cellular toxicity, the repeat-containing RNA accumulated in peri-nuclear cytoplasmic inclusions (hereafter, called cytoplasmic RNA aggregates; Fig. 1D-F, Supp. Fig. 1J). Reducing the RNA expression level down to a few hundred copies per cell (as assessed by RNA FISH) did not affect the cytoplasmic RNA localization pattern, although the size of the inclusions decreased (Fig. 1G-H; Supp. Fig. 1K). Fluorescence in situ hybridization (FISH) using
a probe against CAG repeats or against MS2-hairpins confirmed the accumulation of cytoplasmic RNA aggregates in cells expressing constructs that induced toxicity (Supp. Fig. 1L). Control RNAs without CAG repeats, encoding a fluorescent protein or its reverse complement, that were tagged with MS2 hairpins, did not produce detectable inclusions in the nucleus or the cytoplasm (Fig. 1D-F). Deletion of the MS2-hairpin tag did not appreciably affect RNA localization (Supp. Fig. 1M), indicating that the tag was not the cause of cytoplasmic RNA accumulation and aggregation (see also, Supp. Discussion). Cytoplasmic aggregates formed whether the toxicity-inducing repeat sequence (T1) was introduced through lentiviral transductions, by piggyBac transposase-mediated integration, or via transient transfections (Supp. Fig. 1N). These results indicate that upstream flanking sequences can profoundly alter the sub-cellular localization of the expanded CAG repeat-containing RNAs, and cell toxicity correlates with the accumulation of RNA in cytoplasmic aggregates.

We examined how the native sequences that occur upstream of CAG repeats in *HTT*, *ATXN3* and *ATXN8* genes, associated with Huntington’s disease, spinocerebellar ataxia type 3, and spinocerebellar ataxia type 8 respectively, affect the localization of the CAG repeat-containing RNA. These native sequences were placed upstream of expanded CAG repeats and separated from the repeats by multiple stop codons in each frame, as previously described (Zu et al., 2011). Previous work has also shown that the expression of these repeat-containing constructs triggers cell apoptosis (Zu et al., 2011). In all three cases, we observed that a substantial fraction of cells that expressed these repeats exhibited cytoplasmic RNA aggregates whereas cells transfected with control plasmids did not (Fig 1I). Thus, disease-causing repeats occur in sequence contexts that promote nuclear export and cytoplasmic RNA aggregation.
Cytoplasmic RNA aggregates are gel-like and deform the nuclear envelope.

We previously reported that the nuclear foci formed by CAG repeat-containing RNA are liquid-like (Jain and Vale, 2017). Consistent with liquid-like behavior, RNA at the nuclear foci is mobile, and exhibited rapid fluorescence recovery after photobleaching (characteristic recovery time, $\tau = 60 \pm 28$ s, $n = 23$ foci across constructs N1 and N2; Fig. 2A-B; Supp. Movie 1). In contrast to the globular nuclear foci, cytoplasmic RNA aggregates have a dendritic morphology (Fig 1D). The RNA clustered in the cytoplasmic aggregates was immobile, and upon fluorescence photobleaching, we did not observe appreciable recovery over a 5 min period ($<15\%$ recovery, $n = 24$ aggregates across constructs T1 and T2, Fig. 2A-B, Supp. Movie 2). Cells with cytoplasmic RNA aggregates often displayed a deformed nuclear morphology (Fig. 2C-D), and in numerous cases, the nuclear envelope was indented in the regions adjoining the RNA aggregates (Fig. 2E). Thus, nuclear and cytoplasmic inclusions formed by CAG repeat-containing RNA exhibit markedly distinct biophysical properties, with the cytoplasmic aggregates being solid-like and capable of deforming the nucleus.

Cytoplasmic RNA aggregation induces mislocalization of TDP-43, FUS and p62.

We next evaluated the localization of cytoplasmic RNA aggregates in comparison to markers for known cytoplasmic RNA granules. Canonical markers for stress granules (G3BP1) (Tourrière et al., 2003), P-bodies (EDC4) (Yu et al., 2005), and TIS granules (TIS11b) (Ma and Mayr, 2018) did not appreciably enrich at the cytoplasmic RNA aggregates (Fig. 3A-B, Supp. Fig. 2A), suggesting that the aggregates of CAG repeat-containing RNA are distinct from these RNA granules.
The morphology and the localization of the RNA aggregates were similar to those of cytoplasmic inclusions that are observed in neurodegenerative diseases including Huntington’s disease and ALS (Gomez-Deza et al., 2015; Liu et al., 2015). TDP-43, normally a predominantly nuclear RNA binding protein, mislocalizes to the cytoplasm in several neurodegenerative diseases (Amador-Ortiz et al., 2007; Neumann et al., 2006; Schwab et al., 2008). We found that cytoplasmic levels of TDP-43 were significantly higher in cells with RNA aggregates than in uninduced controls (Fig. 3C; nuclear to cytoplasmic TDP-43 ratio 0.49 ± 0.2 compared to 1.4 ± 0.4 in uninduced controls, p < 0.0001 by Mann-Whitney test, n ≥ 40 cells). Further, TDP-43 was enriched at the cytoplasmic RNA aggregates (Fig. 3C; $R^2 = 0.45$, compared to $R^2 = 0.01$ for uninduced cells where $R$ is the correlation coefficient between the two images, $p < 0.0001$ by Mann-Whitney test, n ≥ 40 cells). FUS, another RNA binding protein that forms aggregates in several repeat expansion disorders (Deng et al., 2010), also mislocalized to the cytoplasm and was sequestered at the RNA aggregates (Fig. 3D; nuclear to cytoplasmic ratio 3.6 ± 3.5 compared to 19.7 ± 6.9 in uninduced controls, $p < 0.0001$ by Mann-Whitney test, n ≥ 40 cells; $R^2 = 0.29$, compared to $R^2 = 0.01$ for uninduced controls, $p < 0.0001$ by Mann-Whitney test, n ≥ 38 cells). Neither TDP-43 nor FUS mislocalized in cells with nuclear foci only, and the nuclear to cytoplasmic ratio of these proteins in cells with RNA foci was comparable to that observed in uninduced controls (Fig. 3C-D, Supp. Fig. 2B). We also found that p62, a canonical biomarker for protein aggregation and autophagy (Zatloukal et al., 2002), was substantially enriched at the RNA aggregates (Fig. 3E; $R^2 = 0.52$, compared to $R^2 = 0.01$ for uninduced controls, $p < 0.0001$ by Mann-Whitney test, n ≥ 40 cells). In summary, cytoplasmic aggregates of CAG repeat-
containing RNA co-localize with proteins that are frequently mislocalized in repeat expansion diseases.

**Cytoplasmic RNA aggregation coincides with RAN translation of the repeat-containing RNA.** RNA with expanded CAG-repeats has the potential to undergo non-canonical RAN translation (Cleary and Ranum, 2013). The repetitive proteins produced upon translation of repeat-expansion RNAs are aggregation-prone, and these aggregates recruit biomarkers for protein aggregation such as p62 (Donaldson et al., 2003; Mori et al., 2013; Nagaoka et al., 2004). We therefore tested whether cells with cytoplasmic RNA inclusions produce RAN translation products. The stop codons upstream of repeats should preclude translation readthrough from the flanking sequences and any polyglutamine production (encoded in the CAG frame) likely results from non-canonical translation of the repeats. Western blots using an antibody targeting polyglutamine revealed an ~8-fold increase in polyglutamine-containing proteins in cells with cytoplasmic aggregates compared to uninduced controls (Fig. 4A). In contrast, cells with only nuclear RNA foci did not exhibit significant polyglutamine production (Fig. 4A). Immunostaining using a polyglutamine antibody confirmed an increase in production of polyglutamine-containing proteins in cells with cytoplasmic aggregates. Interestingly, immunolabeling also revealed that polyglutamine staining co-localized with the CAG repeat-containing RNA (Fig. 4B-C, Supp. Fig. 3A; $R^2 = 0.35$ compared to $R^2 = 0.01$ for uninduced cells, $p < 0.0001$ by Mann-Whitney test, $n \geq 40$ cells). This result suggests that RAN translation products are co-aggregating with CAG repeat-containing RNA.
Large cytoplasmic inclusions of RAN-translation products may promiscuously sequester cellular macromolecules. To evaluate the progression of RAN translation with respect to RNA aggregation, we developed a live-cell reporter of RAN translation by placing a blue fluorescent protein, EBFP2, lacking an ATG-start codon, immediately downstream of the CAG-repeats (Fig 4D). We chose EBFP2 as our reporter of choice as it does not contain another in-frame methionine in the first 78 amino acids and translation initiation at this residue should not result in a fluorescent product (Heim et al., 1994). RAN translation, which likely initiates within or upstream of the repeats, will read through the downstream EBFP2, resulting in fluorescent protein production in one of the frames. Cells transduced with this construct exhibited BFP fluorescence upon induction (Fig. 4E). Incorporation of stop codons between the CAG-repeats and the EBFP2 led to a near complete loss of BFP fluorescence (Supp. Fig. 3B) validating that the translation of EBFP2 initiates within or upstream of the CAG-repeats.

Using this dual-reporter cell line, we monitored repeat RNA localization and RAN translation in real-time. Upon transcription induction, no appreciable BFP fluorescence or RNA aggregation was observed at early time points (Fig. 4E). After ~12 h, we observed an increase in BFP fluorescence (Fig. 4E, Supp. Fig. 3C, Supp. Movie 3). Intriguingly, BFP fluorescence appeared prior to any apparent RNA aggregation. Over time, the BFP-tagged RAN translation products co-aggregated with the RNA throughout the cytoplasm, and these RNA-protein co-clusters gradually coalesced into larger perinuclear aggregates. Given that the appearance of BFP fluorescence occurs ~30 min after translation (due to folding and maturation (Balleza et al., 2018)), these data suggest that RAN translation starts prior to observable RNA aggregation, and that the RAN-translation products likely aggregate co-translationally or soon after translation with the repeat-containing RNA.
Inhibition of RAN translation abrogates RNA aggregation and rescues cytotoxicity.

Finally, we investigated whether translation of CAG-repeats was required for RNA aggregation. Treatment with a translation elongation inhibitor, cycloheximide (10 µg/mL), reduced the levels of both RAN translation products as well as RNA aggregates by >90% compared to untreated controls (Fig. 5A-B, Supp. Fig. 4A; p < 0.0001 by Mann-Whitney test for each, n ≥ 40 cells). Cycloheximide did not substantially affect RNA stability (Supp. Fig. 4B). The repeat-containing RNA was still exported to the cytoplasm in cycloheximide-treated cells but did not form any observable inclusions (Supp. Fig. 4B).

Since global translation inhibition itself exerts substantial toxicity, we tested whether targeted translation inhibition of CAG-repeat-containing RNA could prevent aggregation and toxicity. We designed a phosphorodiamidate backbone 8×CTG morpholino that can hybridize with the CAG-repeats. Morpholinos sterically inhibit protein translation without affecting the stability of the target RNA (Summerton, 2007). Treatment with 8×CTG morpholino suppressed RAN translation and RNA aggregation (reduction by 43% and 71%, respectively compared to treatment with control morpholinos, p < 0.001 by Mann-Whitney test, n ≥ 40 cells) while only modestly affecting RNA levels (Fig 5C-D, Supp. Fig. 4C-D; 29% reduction in RNA compared to cells treated with control morpholino, n ≥ 40 cells). 8×CTG morpholino treatment also rescued cell viability. The expression of toxic CAG-repeat RNAs led to >85% cell death after 4 days in cells treated with a control morpholino, compared to <35% cell death upon treatment with the 8×CTG morpholino (Fig. 5E). In summary, inhibition of RAN translation is sufficient to mitigate both cytoplasmic RNA aggregation and cytotoxicity associated with CAG repeat RNA.
Discussion

Nuclear foci of repeat-containing RNA were first observed in myotonic dystrophy (Taneja et al., 1995) and have been subsequently described in nearly all repeat expansion diseases (Wojciechowska and Krzyzosiak, 2011). We demonstrate that, in addition to aggregating in the nucleus, repeat-containing RNA also aggregate in the cytoplasm. The formation of cytoplasmic RNA aggregates coincides with RAN translation, and these aggregates are the sites for the accumulation of several RNA binding proteins as well as markers for protein aggregation. We propose that this cytoplasmic aggregation of repeat-containing RNA may provide a mechanistic explanation for the aberrant cytoplasmic localization of RNA binding proteins such as TDP-43 which is commonplace in repeat expansion disorders (Goodwin and Swanson, 2014; Todd and Paulson, 2010).

Why have cytoplasmic RNA aggregates not been detected in previous RNA localization studies in repeat expansion diseases? We found that commonly used chemical cross-linking fixatives, such as formaldehyde, reduced the detection of RNA at cytoplasmic aggregates via FISH. While both nuclear foci and cytoplasmic aggregates of RNA were observed in methanol-fixed cells, only nuclear foci were stained in cells fixed with formaldehyde (Supp. Fig 5A). Even in cells where RNA detection by MS2CP-YFP revealed abundant cytoplasmic inclusions, RNA aggregates were not readily observed via FISH after formaldehyde fixation (Supp. Fig. 5A). Chemical cross-linking of proteins and RNA may prevent probes from accessing the aggregated RNA. Formaldehyde treatment can also result in the addition of mono-methylol groups to RNA bases – particularly adenine and cytosine residues which are enriched in CAG repeat RNA – and chemically cross-link RNA molecules (Masuda et al., 1999). These findings suggest that
optimization of FISH protocols may allow for better detection of cytoplasmic RNA inclusions. Consistent with this notion, a recent study reported a substantially greater number of cytoplasmic inclusions of GGGGCC- and CGG- repeat-containing RNAs when using hybridization chain reaction to amplify RNA FISH signal compared to the traditional methods (Glineburg et al., 2021).

Previous studies have described peri-nuclear cytoplasmic protein inclusions in a variety of CAG repeat expansion disorders such as Huntington’s disease and spinocerebellar ataxia 8 (Ayhan et al., 2018; Waelter et al., 2001). Our data suggest that these inclusions likely contain the repeat-containing RNA which may be revealed by using milder fixatives. Co-aggregation of RNA may also be a common feature of protein aggregation disorders that do not involve repeat-expansions. Early studies on amyloid beta plaques in Alzheimer’s disease showed that these aggregates stain with RNA-binding dyes (Ginsberg et al., 1997), and a recent study found that tau aggregates contain RNA (Lester et al., 2021). Modifications in tissue storage and fixation methods to retain integrity and advancements in RNA detection technologies may allow one to determine the true prevalence of RNA aggregates in neurodegenerative disease.

Our data directly demonstrate that RAN translation is more toxic to cells than the accumulation of repeat-containing RNAs at nuclear foci. RNAs that coalesced exclusively at nuclear foci did not induce substantial cytotoxicity in our cell culture model, and inhibiting RAN translation in cells where repeat-containing RNA was exported to the cytoplasm, was sufficient to mitigate cell death. Genetic screens have identified several RNA export-factors such as NXF1 and NXT1 as well as nuclear pore components as potential regulators of repeat expansion induced toxicity.
Previous work from the Ranum lab showed that over-expression of certain RNA binding proteins like MBNL1 may increase retention of RNA in the nucleus and consequently inhibit RAN translation and cell death (Zu et al., 2017). Our findings, based upon a tissue culture model, raise the possibility that sequestration of RNA in the nucleus at foci may be protective to cells, and if so, may provide a therapeutic intervention strategy to reduce disease burden. However, we cannot exclude the possibility that prolonged accumulation of repeat-containing RNA in nuclear foci also interferes with cell function, such as in myotonic dystrophy where foci formation is associated with various splicing defects (Jiang et al., 2004).

Our results also underscore the role of cis-acting sequences in mediating disease. Diseases caused by CAG trinucleotide repeat expansion share several pathobiological features. However, at a given repeat number, these diseases differ considerably in the age of disease onset and severity of symptoms (Gusella and MacDonald, 2009). Previous studies have shown that the surrounding sequence context may influence repeat instability (Brock et al., 1999). Likewise, when the repeats are translated, the sequence context influences the aggregation propensity of resultant polyglutamine-containing protein (Duennwald et al., 2006). Our results unequivocally show that besides affecting repeat instability and protein aggregation, flanking sequences, in combination with repeat length, modulate RNA toxicity. We found that GC-rich flanking sequences were more likely to induce RNA aggregation and RAN translation. Interestingly, recent genome-wide screens for modulators of RAN translation have identified DDX3X, a dead-box helicase that preferentially binds GC-rich UTRs as a key modulator of RAN translation (Cheng et al., 2019; Linsalata et al., 2019). The assays that we have developed in this study may provide means to screen for sequence motifs that modulate RNA export and RAN translation.
Our work provides a new lens to examine the role of repeat-containing RNA in the cellular pathology of repeat expansion diseases. Disease-associated repeat expansions in RNA provide sites for multivalent intermolecular base-pairing. Such multivalent interactions promote RNA oligomerization and its accumulation in the nucleus at foci (Jain and Vale, 2017). The sequences flanking the repeats may potentially provide binding sites for yet-unknown RNA helicases and export factors. In the cytoplasm, the repeat-containing RNA encounters the translation machinery and may undergo non-canonical RAN translation. Interestingly the cytoplasmic RNA does not appear to form aggregates until it is RAN translated, suggesting that the repeat-containing RNA in the cytoplasm is initially in a state that inhibits extensive inter-molecular RNA-RNA interactions. The process of RAN translation may relieve this repression. The RAN peptides are also aggregation-prone and another possibility is that the repeat-containing RNA may co-translationally aggregate with the RAN peptides. The smaller RNA-RAN protein aggregates are deposited, possibly by the minus-end directed motor dynein, into large perinuclear, aggresome-like inclusions (Johnston et al., 1998). These cytoplasmic inclusions sequester various RNA binding proteins, disrupt nuclear morphology, and over time, kill cells. Future work may reveal the factors involved in repeat-RNA export, RAN translation, and the mechanism of RNA-RAN-protein co-aggregation.

**Author Contributions**

M.D., Y.C., R.V., and A.J. designed research; M.D., Y.C., R.S., N.Z., C.T., and A.J. performed research; M.D. and A.J. analyzed data; M.D., R.V., and A.J. wrote the paper.
Acknowledgements

We thank members of the Jain lab for helpful discussions. Plasmids containing CAG repeats with flanking sequences from *ATXN3*, *ATXN8*, and *HTT* were a gift from Laura Ranum. This work was supported by grants from the National Institutes of Health (R00AG053434), the David and Lucile Packard Foundation, and the Smith Family Awards Program. Y.C. was supported by a fellowship from the National Research Foundation of Korea. R.V. is an investigator with the Howard Hughes Medical Institute.

Declaration of interests

The authors declare no competing interest.

STAR Methods

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ankur Jain (ajain@wi.mit.edu).

Materials availability

Plasmids generated in this study are available upon request to the lead contact.

Data and code availability
All data reported in this paper will be shared by the lead contact upon request. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**Experimental model and subject details**

Immortalized female human cell lines were used in this study. All experiments were performed in U-2 OS cells (ATCC, HTB-96) stably expressing a TetOn3G transactivator protein and an MS2 hairpin binding protein fused to YFP was previously described (Jain and Vale, 2017). HEK293T (ATCC, CRL-3216) cells were used for generating lentivirus. ATCC authenticated cell lines by STR profiling. Cells were grown in DMEM (Life Technologies, 11965-126) supplemented with 10% (v/v) fetal bovine serum (Gibco, 26140-079) and 1× penicillin, streptomycin, and glutamine (Gibco, 10378016). Transgene expression was induced by adding 1 µg mL⁻¹ doxycycline (Sigma-Aldrich, D9891) for 24 hours except where otherwise specified. Cells were maintained in a humidified incubator at 37 C with 5% CO2. Representative characterization of cytoplasmic inclusions was conducted in cells expressing construct T1 or T2; representative characterization of nuclear inclusions was conducted in cells expressing either construct N1 or N2.

**Method details**

**Cloning and plasmid preparation**

Full-length plasmid sequences are provided in Supp. Table 2. Lentiviral transfer plasmids with CAG-repeats and MS2-hairpins were previously described (Jain and Vale, 2017). The upstream flanking sequences as well as the downstream EBFP2 tag, were obtained as double stranded
DNA fragments (Quintara Biosciences or IDT). The library of 12 upstream flanking sequences were inserted via restriction digestion cloning in constructs with 47× or 240× CAG repeats (construct N1 or equivalent) between EcoRI and MluI sites. 5× or 22× CAG repeats were inserted as annealed primers/ultramers (IDT) between EcoRI and SgrDI sites downstream of the T1 flanking sequence. EBFP2 was cloned between BamHI and NotI sites. All cloning and plasmid preparations were performed in Stbl3 E. coli cells (Invitrogen, C7373-03) grown at 30°C. In constructs with 240× CAG repeats, Sanger sequencing did not provide sufficient read length to unambiguously determine the number of repeats, so repeat length was first validated via restriction digestion of the insert and gel electrophoresis. To sequence the repeat tract, we optimized a Sanger sequencing protocol (in collaboration with Quintara Biosciences) which used betaine and 7-deaza-dGTP. This optimized sequencing protocol provided ~800 base long reads from both ends. Sanger sequencing revealed eight interruptions in the CAG repeat track in constructs with 240× CAG repeats (sequences in Supp. Table 2). These sites contained a deletion of G nucleotide and the first interruption occurred at 42 bases from the start of the repeat tract. These interruptions were present in our parent plasmid (N0), and were common to the constructs with various flanking sequences (N1-N7, T1-T5). We also observed similar cellular phenotypes (e.g., toxicity, cytoplasmic RNA aggregation, and RAN translation) in constructs containing either these repeat interruptions or with pure expanded ≥47×CAG repeat tracks.

Lentiviral packaging and envelope constructs were obtained from addgene: pCMV-VSV-G was a gift from Bob Weinberg (Addgene plasmid # 8454; http://n2t.net/addgene:8454; RRID: Addgene_8454) (Stewart et al., 2003); psPAX2 was a gift from Didier Trono (Addgene plasmid # 12260; http://n2t.net/addgene:12260; RRID: Addgene_12260). Plasmids containing CAG
repeats with endogenous flanking sequences from \textit{ATXN3}, \textit{ATXN8}, and \textit{HTT} were generously provided by Laura Ranum (Zu et al., 2011).

**Transfections and lentiviral transductions**

Lentivirus was generated using second generation lentiviral packaging system in HEK293T cells. In brief, ~500,000 HEK293T cells were plated in a 6-well plate until they reached 60-80% confluency. Each well was transfected with 0.5 µg of envelope plasmid (Addgene #8454), 1 µg of packaging plasmid (Addgene #12260), and 2 µg of the transfer plasmid with 8 µL of Lipofectamine LTX (Invitrogen, 15338-100) in 500 µL of Opti-MEM reduced serum media (Gibco, 31985-070). This solution was incubated for 5 min before pipetting onto HEK293T cells in IMDM (Gibco, 12440-053) + 10% FBS + 1× PSG. After three days of incubation, virus was collected by spinning the supernatant for 5 min at 20,000 x g. U-2OS cells were transduced with varying titers of virus in 10 µg/mL polybrene (Millipore, TR-1003-G).

For transient transfections in U-2OS cells, 500 ng of plasmid was mixed with Opti-MEM reduced serum media and 1.5 µL Viafect transfection reagent (Promega, E4981) for a total volume of 50 µL. 10 µL of this mixture was added per well of U-2OS cells in a 96-well glass bottom plate in 100 µL of freshly changed DMEM +FBS + PSG media.

**Transgene copy quantification**

Genomic DNA and total RNA were collected from cells using PureLink Genomic DNA mini kits (Invitrogen, K1820-01) and PureLink RNA mini kits (Invitrogen, 12183018A), respectively, according to the manufacturer’s protocols. RNA was converted to cDNA using SuperScript III reverse transcriptase (LifeTechnologies, 11754050). Transgene genomic DNA or cDNA copies were quantified using real-time quantitative PCR (RT-qPCR) with SYBR green reagents.
(Applied Biosystems, 4309155) using primers targeting the TetPromoter for DNA and targeting the WPRE for cDNA (Key Resource Table). Transgene copies were normalized to copies of ACTB (primers listed in Key Resource Table).

**Toxicity assays**

Cell toxicity was quantified using two methods: (1) counting the number of adhered cells and (2) exclusion of trypan blue dye. (1) Cells were plated on a six well plate at 10,000 cells per well. The following day, cells were induced with 1 µg mL\(^{-1}\) doxycycline or left as uninduced controls in a total volume of 2.5 mL. Each cell line was compared to a corresponding uninduced control to account for any differences in plating density between cell lines. Five days post induction (unless otherwise specified), cells were washed five times with DPBS to remove dead cells from the plate, dissociated from the plate using trypsin, and neutralized with growth medium (DMEM + 10% (v/v) FBS) two times in order to collect all cells from the plate. The cells were then pelleted at 500 x g for 3 min and resuspended in growth medium. Cells were counted using a Countess II FL automated cell counter (Invitrogen, AMQAF1000) and disposable cell counting chamber slides (Invitrogen, C10283). At least two technical replicate counts were performed for each sample.

(2) Trypan blue assays were adapted from published protocols (Strober, 2001). Cells were plated on a six well plate and induced with doxycycline the following day. Five days after induction, the supernatant which contains any dead or detached cells was collected. Cells were washed with DPBS and this wash solution was added to the supernatant. After trypsinization, growth medium was added to neutralize trypsin and the collected cells were added to the supernatant/wash solutions. Cells were pelleted at 500 x g for 3 min and pellets were resuspended in 100 uL of DPBS. This resuspension was mixed 1:1 with 0.4% trypan blue (Invitrogen, T10282) for three
minutes and cell counts were performed using a Countess II FL automated cell counter as described above. At least two technical replicate counts were performed for each sample.

**Fluorescence microscopy**

Cells were plated in a glass-bottom 96-well plate (Brooks, MGB096-1-2-LG-L) and imaged using a Dragonfly 505 spinning-disk confocal microscope (Andor Technologies) equipped with a piezo Z-stage (ASI) and an iXon Ultra 888 EMCCD camera. Pin-hole size was kept at 40 microns. Z-stacks were acquired with a step size of 0.3 to 0.5 µm. Live cells were imaged in a humidified chamber (OKO labs) maintained at 37° C and 5% (v/v) CO$_2$ using a 100x oil immersion objective NA 1.45 (Nikon, MRD01905) (pixel size 121 nm $\times$ 121 nm). Fixed cells were imaged at room temperature. BFP and DAPI were excited with a 405 nm laser and fluorescence was collected using a 445/46 bandpass filter. YFP and Alexa488 were imaged using a 488 nm laser and corresponding 521/38 nm band pass emission filter. Cy3 or mCherry labeled samples were imaged using a 561 nm excitation and a 594/43 emission filter. Last, Cy5 or Atto647N labeled samples were imaged using 640 nm laser line and a 698/77 bandpass emission filter. At least 40 cells were imaged in at least two independent experiments.

**Image analysis and aggregate quantification**

All image analysis was performed in FIJI. Background fluorescence was determined from the average signal of ten ROIs (approximately 50 square microns in size) that did not contain any cells. This background signal was determined for each channel in each experiment and subtracted in all analyses.

Cells were manually segmented to analyze signal in the cytoplasm or in the nucleus. To identify cytoplasmic or nuclear inclusions, we empirically determined the intensity and size threshold
that adequately recapitulated the features as observed by eye. Cytoplasmic inclusions for MS2-YFP images were identified as objects with $2 \times$ the mean intensity of the cytoplasm and a size threshold of 0.1 square microns. A similar thresholding was used for identifying cytoplasmic inclusions by RNA FISH, after appropriate adjustments to the background fluorescence. We classified cells with at least 2% of their cytoplasm occupied by inclusions as having cytoplasmic aggregates. Likewise, nuclear foci were identified as having intensity 1.5-fold the median nucleoplasm intensity and a size threshold at 0.14 square microns. We classified cells with at least 1% of their nucleoplasm occupied by inclusions as having nuclear foci.

In cases where transient transfection was used, we set the size threshold for aggregates as 0.1 square microns and set an intensity threshold at $4 \times$ the mean cytoplasmic intensity after background subtraction. We classified cells with at least 0.2% of their cytoplasm occupied by inclusions as having cytoplasmic aggregates. We tested several thresholding parameters, and obtained comparable results (Supp. Fig. 5B).

To determine the nuclear or cytoplasmic immunofluorescence, we manually segmented the cells (based on DAPI and total immunofluorescence signal from the cell), and determined the total integrated fluorescence in nucleus or cytoplasm after subtracting the background. Similar trends in nuclear to cytoplasmic fluorescence were observed when we compared the mean pixel intensities in the two compartments.

For co-localization analysis, line-profiles were obtained using the ‘Plot profile’ plugin in FIJI. Background signal was subtracted from all points and the intensity of each channel was normalized to the maximum intensity along the line. For calculating correlation coefficients, we
empirically determined the intensity threshold to classify a pixel as background or foreground. The thresholded images were used to determine per-pixel intensity correlation coefficients. Thresholds were determined using the sample image, and identical settings were used for the corresponding controls.

To quantify total RNA levels by FISH, average background signal was measured in the nucleus and cytoplasm of uninduced cells. The integrated intensity of each compartment was calculated and summed to find the relative transgene RNA levels in each cell. Similar results were found in confocal and widefield images. Movies were stabilized using the Image Stabilizer plugin in FIJI (“Kang Li @ CMU - Image Stabilizer Plugin for ImageJ,” n.d.).

**FRAP imaging and analysis**

Regions approximately one square micron in size within cytoplasmic or nuclear inclusions were photobleached using a MicroPoint FRAP module (Andor Technologies) using a 405 nm wavelength photoablation laser. Cells were imaged for 5 frames prior to bleaching, immediately after the bleach (within two seconds), and then in 30 second intervals for five minutes after bleaching. After background subtraction, fluorescence of the bleached region at each timepoint was normalized to the fluorescence of the bleached region prior to photobleaching to determine fluorescence recovery. To correct for photobleaching over the course of imaging, fluorescence of the bleached region was also normalized to the fluorescence of an unbleached region of the aggregate at each timepoint. Unbleached foci were used as a control region to correct for photobleaching due to imaging. FRAP was performed at 24 hours (± 3 hours) after induction.
with doxycycline. At least 20 FRAP events from two independent experiments were performed for each case. Time constants were determined by fitting the fluorescence recovery to the equation: \( I = A - I_0 \exp(-t/\tau) \).

**RNA FISH**

Cells were fixed with a solution of 75% (v/v) methanol, 25% (v/v) acetic acid for 10 minutes at 4°C. Fixed cells were washed three times with a wash solution (PBS or nuclease-free water with 300 mM NaCl, 30 mM sodium citrate, 10% (v/v) formamide, and 0.1% (v/v) NP-40 substitute) at room temperature (23°C). Hybridization was performed at 37°C for three hours with 200 nM Cy3-conjugated probes targeting either the CAG repeats or the MS2 hairpin region (probe sequences Key Resource Table). The probes were dissolved in hybridization buffer (100 mg mL⁻¹ dextran sulfate, 10% formamide, 300 mM NaCl, and 30 mM sodium citrate in nuclease-free water). After hybridization, cells were washed for 30 minutes with wash solution and counterstained for 30 minutes with wash solution containing 0.5 µg mL⁻¹ DAPI; both this wash step and DAPI staining were performed at 37°C. Cells were then washed three times with PBS, and kept in PBS for imaging.

Where indicated, cells were fixed with 2% (w/v) formaldehyde for 10 minutes, followed by a similar protocol to that described above, except that the probes were hybridized for 16 hours. Detection of cytoplasmic inclusions was substantially impeded after formaldehyde treatment (see Discussion).

**Immunofluorescence**
Cells were fixed with 2% formaldehyde in PBS for 40 minutes, washed four times with PBS, and then permeabilized with 0.1% (v/v) Triton-X-100 in PBS for 10 minutes at room temperature. Cells were then blocked in 0.45 µm-filtered 3% (w/v) bovine serum albumin (BSA; Sigma-Aldrich, A7906) in PBS. Primary antibodies (Key Resource Table) were diluted 1:100 in 1% (w/v) BSA in PBS and incubated with cells for one hour at room temperature. Cells were washed three times with PBS and then incubated with the appropriate secondary antibody (Key Resource Table) diluted 1:2000 in 1% (w/v) BSA (in PBS) for one hour at room temperature. After three washes with PBS, cells were counterstained with DAPI solution (PBS containing 0.5 µg mL\(^{-1}\) DAPI) for three minutes, washed three times with PBS again, and then imaged as previously described. To enhance MS2CP-YFP signal after fixation, cells were incubated with an anti-GFP antibody and a subsequent Alexa488-conjugated secondary antibody. These incubations were performed along with other antibodies (i.e., cells were simultaneously incubated with anti-G3BP and anti-GFP primary antibodies). At least two independent experiments were performed for all immunofluorescence experiments.

**Western blots**

Cells were washed with PBS and then lysed with RIPA lysis buffer (25 mM Tris-HCl pH=7.5, 150 mM NaCl, 1% (v/v) NP-40, 1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS) with protease and phosphatase inhibitors (Thermo Scientific, 78442). The lysate was homogenized by passing through a 22-gauge syringe 10 times and then incubated on ice for 30 min with vortexing every 5 min. Cell debris was removed by centrifugation at 500 x g for 10 min at 4°C and lysates were mixed with 4× Bolt™ LDS sample buffer (Invitrogen, B0007) with 50 mM DTT and heated at 70°C for 10 min. Samples were loaded onto a Bolt™ 4-12% Bis-Tris polyacrylamide gel.
(Invitrogen, NW04122) and transferred to a PVDF membrane using iBlot 2 dry blotting system (Invitrogen, IB21001). Membranes were blocked in 5% (w/v) skim milk in TBST (tris-buffered saline with 0.1% (v/v) Tween-20) for one hour at room temperature. Membranes were then incubated with primary antibodies (Key Resource Table) that were diluted 1:1000 in 1% (w/v) skim milk in TBST at 4°C overnight. After three TBST washes, membranes were incubated with horseradish peroxidase-conjugated secondary antibodies at a 1:10,000 dilution in 1% (w/v) skim milk in TBST for one hour at room temperature (Key Resource Table). Membranes were washed three times with TBST and chemiluminescence signals were detected using SuperSignal West Femto Maximum sensitivity substrate (Thermo Scientific, 34095) on a ChemiDoc XRS+ imager (Bio-Rad). Band intensities were quantified using FIJI after subtracting background signal.

**Translation inhibition**

Where indicated, cells were treated with 10 µg mL\(^{-1}\) cycloheximide (Sigma-Aldrich, C1998). 1 µg mL\(^{-1}\) doxycycline was concurrently added to induce transgene expression. Cells were then imaged as previously described 14 h post treatment. Cells transduced with a high viral titer were used in these experiments so that aggregates began to form prior to substantial cell death due to the general toxicity of cycloheximide. Two independent experiments were performed.

Targeted translation inhibition was achieved through addition of 8×CTG morpholinos (GeneTools). Control experiments were performed using a standard intron control morpholino (CCTCTTACCTCAGTTACAATTTATA; GeneTools), targeting a mutated intron of the human beta-globin gene. In each case, 50 µM morpholino were delivered with 6 µM EndoPorter peptide (GeneTools, OT-EP-PEG-1) for 24 h prior to induction with doxycycline. Morpholinos were kept in the media after doxycycline addition giving a total treatment time of 48 h with
morpholinos and 24 h of treatment with doxycycline. For toxicity assays, cells were plated at 5,000 cells/well in 24-well plates in a total volume of 300 µL. Cells were treated with 50 µM 8×CTG or control morpholinos for five days and with 10 ng/mL doxycycline for four days.

**Quantification and statistical analyses**

Cell viability, qPCR, and quantifications of percent of cells exhibiting a phenotype (e.g., nuclear deformation or the presence of inclusions) were analyzed by two-tailed unpaired Student’s t-tests. Morpholino rescue of cell viability was analyzed by two-way ANOVA in addition to t-tests. FRAP experiments were analyzed using Wilcoxon signed-rank test. All other experiments were analyzed by Mann-Whitney U-tests. All statistical tests are described in figure legends including the test used, minimum values of n, definitions of center, and precision measurements.

**References**

Amador-Ortiz, C., Lin, W.-L., Ahmed, Z., Personett, D., Davies, P., Duara, R., Graff-Radford, N.R., Hutton, M.L., Dickson, D.W. (2007). TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer’s disease. Ann. Neurol. 61, 435–445. 10.1002/ana.21154

Ayhan, F., Perez, B.A., Shorrock, H.K., Zu, T., Banez-Coronel, M., Reid, T., Furuya, H., Clark, H.B., Troncoso, J.C., Ross, C.A., Subramony, S., Ashizawa, T., Wang, E.T., Yachnis, A.T., Ranum, L.P. (2018). SCA8 RAN polySer protein preferentially accumulates in white matter regions and is regulated by eIF3F. EMBO J. 37, e99023. 10.15252/embr.201899023

Balleza, E., Kim, J.M., Cluzel, P. (2018). Systematic characterization of maturation time of fluorescent proteins in living cells. Nat. Methods 15, 47–51. 10.1038/nmeth.4509
Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. Mol. Cell 2, 437–445. 10.1016/s1097-2765(00)80143-4

Brock, G.J., Anderson, N.H., Monckton, D.G. (1999). Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. Hum. Mol. Genet. 8, 1061–1067. 10.1093/hmg/8.6.1061

Cheng, W., Wang, S., Zhang, Z., Morgens, D.W., Hayes, L.R., Lee, S., Portz, B., Xie, Y., Nguyen, B.V., Haney, M.S., Yan, S., Dong, D., Coyne, A.N., Yang, J., Xian, F., Cleveland, D.W., Qiu, Z., Rothstein, J.D., Shorter, J., Gao, F.-B., Bassik, M.C., Sun, S. (2019). CRISPR-Cas9 Screens Identify the RNA Helicase DDX3X as a Repressor of C9ORF72 (GGGGCC)n Repeat-Associated Non-AUG Translation. Neuron 104, 885-898.e8. 10.1016/j.neuron.2019.09.003

Cleary, J.D., Pearson, C.E. (2003). The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. Cytogenet. Genome Res. 100, 25–55. 10.1159/000072837

Cleary, J.D., Ranum, L.P.W. (2013). Repeat-associated non-ATG (RAN) translation in neurological disease. Hum. Mol. Genet. 22, R45-51. 10.1093/hmg/ddt371

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y.R., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R., Rademakers, R. (2011). Expanded GGGGCC
hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72, 245–256. 10.1016/j.neuron.2011.09.011

Deng, H.-X., Zhai, H., Bigio, E.H., Yan, J., Fecto, F., Ajroud, K., Mishra, M., Ajroud-Driss, S., Heller, S., Sufit, R., Siddique, N., Mugnaini, E., Siddique, T. (2010). FUS-immunoreactive inclusions are a common feature in sporadic and non-SOD1 familial amyotrophic lateral sclerosis. Ann. Neurol. 67, 739–748. 10.1002/ana.22051

Donaldson, K.M., Li, W., Ching, K.A., Batalov, S., Tsai, C.-C., Joazeiro, C.A.P. (2003). Ubiquitin-mediated sequestration of normal cellular proteins into polyglutamine aggregates. Proc. Natl. Acad. Sci. 100, 8892–8897. 10.1073/pnas.1530212100

Duennwald, M.L., Jagadish, S., Muchowski, P.J., Lindquist, S. (2006). Flanking sequences profoundly alter polyglutamine toxicity in yeast. Proc. Natl. Acad. Sci. U. S. A. 103, 11045–11050. 10.1073/pnas.0604547103

Ginsberg, S.D., Crino, P.B., Lee, V.M., Eberwine, J.H., Trojanowski, J.Q. (1997). Sequestration of RNA in Alzheimer’s disease neurofibrillary tangles and senile plaques. Ann. Neurol. 41, 200–209. 10.1002/ana.410410211

Glineburg, M.R., Zhang, Y., Krans, A., Tank, E.M., Barmada, S.J., Todd, P.K. (2021). Enhanced detection of expanded repeat mRNA foci with hybridization chain reaction. Acta Neuropathol. Commun. 9, 73. 10.1186/s40478-021-01169-8

Gomez-Deza, J., Lee, Y.-B., Troakes, C., Nolan, M., Al-Sarraj, S., Gallo, J.-M., Shaw, C.E., (2015). Dipeptide repeat protein inclusions are rare in the spinal cord and almost absent from motor neurons in C9ORF72 mutant amyotrophic lateral sclerosis and are unlikely to cause their degeneration. Acta Neuropathol. Commun. 3, 38. 10.1186/s40478-015-0218-y
Goodwin, M., Swanson, M.S., 2014. RNA-binding protein misregulation in microsatellite expansion disorders. Adv. Exp. Med. Biol. 825, 353–388. 10.1007/978-1-4939-1221-6_10

Gusella, J.F., MacDonald, M.E. (2009). Huntington’s disease: the case for genetic modifiers. Genome Med. 1, 80. 10.1186/gm80

Heim, R., Prasher, D.C., Tsien, R.Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. U. S. A. 91, 12501–12504. 10.1073/pnas.91.26.12501

Jain, A., Vale, R.D. (2017). RNA phase transitions in repeat expansion disorders. Nature 546, 243–247. 10.1038/nature22386

Jiang, H., Mankodi, A., Swanson, M.S., Moxley, R.T., Thornton, C.A. (2004). Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. Hum. Mol. Genet. 13, 3079–3088. 10.1093/hmg/ddh327

Johnston, J.A., Ward, C.L., Kopito, R.R. (1998). Aggresomes: a cellular response to misfolded proteins. J. Cell Biol. 143, 1883–1898. 10.1083/jcb.143.7.1883

Kanadia, R.N., Shin, J., Yuan, Y., Beattie, S.G., Wheeler, T.M., Thornton, C.A., Swanson, M.S. (2006). Reversal of RNA missplicing and myotonia after muscleblind overexpression in a mouse poly(CUG) model for myotonic dystrophy. Proc. Natl. Acad. Sci. 103, 11748–11753. 10.1073/pnas.0604970103

Khristich, A.N., Mirkin, S.M., 2020. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. J. Biol. Chem. 295, 4134–4170. 10.1074/jbc.REV119.007678
Kwon, I., Xiang, S., Kato, M., Wu, L., Theodoropoulos, P., Wang, T., Kim, J., Yun, J., Xie, Y., McKnight, S.L. (2014). Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. Science 345, 1139–1145. 10.1126/science.1254917

Lee, K.-H., Zhang, P., Kim, H.J., Mitrea, D.M., Sarkar, M., Freibaum, B.D., Cika, J., Coughlin, M., Messing, J., Molliex, A., Maxwell, B.A., Kim, N.C., Temirov, J., Moore, J., Kolaitis, R.-M., Shaw, T.I., Bai, B., Peng, J., Kriwacki, R.W., Taylor, J.P. (2016). C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. Cell 167, 774-788.e17. 10.1016/j.cell.2016.10.002

Lester, E., Ooi, F.K., Bakkar, N., Ayers, J., Woerman, A.L., Wheeler, J., Bowser, R., Carlson, G.A., Prusiner, S.B., Parker, R. (2021). Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. Neuron 109, 1675-1691.e9. 10.1016/j.neuron.2021.03.026

Li, Kang @ CMU - Image Stabilizer Plugin for ImageJ [WWW Document], n.d. URL https://www.cs.cmu.edu/~kangli/code/Image_Stabilizer.html.

Linsalata, A.E., He, F., Malik, A.M., Glineburg, M.R., Green, K.M., Natla, S., Flores, B.N., Krans, A., Archbold, H.C., Fedak, S.J., Barmada, S.J., Todd, P.K. (2019). DDX3X and specific initiation factors modulate FMR1 repeat-associated non-AUG-initiated translation. EMBO Rep. 20, e47498. 10.15252/embr.201847498

Liu, K.-Y., Shyu, Y.-C., Barbaro, B.A., Lin, Y.-T., Chern, Y., Thompson, L.M., James Shen, C.-K., Marsh, J.L. (2015). Disruption of the nuclear membrane by perinuclear inclusions of mutant huntingtin causes cell-cycle re-entry and striatal cell death in mouse and cell models of Huntington’s disease. Hum. Mol. Genet. 24, 1602–1616. 10.1093/hmg/ddu574
Ma, W., Mayr, C. (2018). A Membraneless Organelle Associated with the Endoplasmic Reticulum Enables 3’UTR-Mediated Protein-Protein Interactions. Cell 175, 1492-1506.e19. 10.1016/j.cell.2018.10.007

Masuda, N., Ohnishi, T., Kawamoto, S., Monden, M., Okubo, K. (1999). Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. Nucleic Acids Res. 27, 4436–4443. 10.1093/nar/27.22.4436

Mori, K., Weng, S.-M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C., Haass, C., Edbauer, D. (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science 339, 1335–1338. 10.1126/science.1232927

Nagaoka, U., Kim, K., Jana, N.R., Doi, H., Maruyama, M., Mitsui, K., Oyama, F., Nukina, N., 2004. Increased expression of p62 in expanded polyglutamine-expressing cells and its association with polyglutamine inclusions. J. Neurochem. 91, 57–68. 10.1111/j.1471-4159.2004.02692.x

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., McCluskey, L.F., Miller, B.L., Masliah, E., Mackenzie, I.R., Feldman, H., Feiden, W., Kretzschmar, H.A., Trojanowski, J.Q., Lee, V.M.-Y. (2006). Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Science 314, 130–133. 10.1126/science.1134108

Paulson, H. (2018). Repeat expansion diseases, In Handbook of Clinical Neurology. (Elsevier), pp. 105–123. 10.1016/B978-0-444-63233-3.00009-9

Schwab, C., Arai, T., Hasegawa, M., Yu, S., McGeer, P.L. (2008). Colocalization of Transactivation-Responsive DNA-Binding Protein 43 and Huntingtin in Inclusions of
Huntington Disease. J. Neuropathol. Exp. Neurol. 67, 1159–1165. 10.1097/NEN.0b013e31818e8951

Strober, W. (2001). Trypan blue exclusion test of cell viability. Curr. Protoc. Immunol. Appendix 3, Appendix 3B. 10.1002/0471142735.ima03bs21

Summerton, J.E. (2007). Morpholino, siRNA, and S-DNA compared: impact of structure and mechanism of action on off-target effects and sequence specificity. Curr. Top. Med. Chem. 7, 651–660. 10.2174/156802607780487740

Taneja, K.L., McCurrach, M., Schalling, M., Housman, D., Singer, R.H. (1995). Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. J. Cell Biol. 128, 995–1002. 10.1083/jcb.128.6.995

Todd, P.K., Paulson, H.L. (2010). RNA-mediated neurodegeneration in repeat expansion disorders. Ann. Neurol. 67, 291–300. 10.1002/ana.21948

Tourrière, H., Chebli, K., Zekri, L., Courselaud, B., Blanchard, J.M., Bertrand, E., Tazi, J. (2003). The RasGAP-associated endoribonuclease G3BP assembles stress granules. J. Cell Biol. 160, 823–831. 10.1083/jcb.200212128

Waelter, S., Boeddrich, A., Lurz, R., Scherzinger, E., Lueder, G., Lehrach, H., Wanker, E.E. (2001). Accumulation of mutant huntingtin fragments in aggresome-like inclusion bodies as a result of insufficient protein degradation. Mol. Biol. Cell 12, 1393–1407. 10.1091/mbc.12.5.1393

Wojciechowska, M., Krzyzosiak, W.J. (2011). Cellular toxicity of expanded RNA repeats: focus on RNA foci. Hum. Mol. Genet. 20, 3811–3821. 10.1093/hmg/ddr299
Yu, J.H., Yang, W.-H., Gulick, T., Bloch, K.D., Bloch, D.B. (2005). Ge-1 is a central component of the mammalian cytoplasmic mRNA processing body. RNA 11, 1795–1802. 10.1261/rna.2142405

Zatloukal, K., Stumptner, C., Fuchsbichler, A., Heid, H., Schnoelzer, M., Kenner, L., Kleinert, R., Prinz, M., Aguzzi, A., Denk, H. (2002). p62 Is a Common Component of Cytoplasmic Inclusions in Protein Aggregation Diseases. Am. J. Pathol. 160, 255–263. 10.1016/S0002-9440(10)64369-6

Zhang, Y.-J., Jansen-West, K., Xu, Y.-F., Gendron, T.F., Bieniek, K.F., Lin, W.-L., Sasaguri, H., Caulfield, T., Hubbard, J., Daughrity, L., Chew, J., Belzil, V.V., Prudencio, M., Stankowski, J.N., Castanedes-Casey, M., Whitelaw, E., Ash, P.E.A., DeTure, M., Rademakers, R., Boylan, K.B., Dickson, D.W., Petrucelli, L. (2014). Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. Acta Neuropathol. (Berl.) 128, 505–524. 10.1007/s00401-014-1336-5

Zu, T., Cleary, J.D., Liu, Y., Bañez-Coronel, M., Bubenik, J.L., Ayhan, F., Ashizawa, T., Xia, G., Clark, H.B., Yachnis, A.T., Swanson, M.S., Ranum, L.P.W. (2017). RAN Translation Regulated by Muscleblind Proteins in Myotonic Dystrophy Type 2. Neuron 95, 1292-1305.e5. https://doi.org/10.1016/j.neuron.2017.08.039

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.C., Nan, Z., Forster, C., Low, W.C., Schosser, B., Somia, N.V., Clark, H.B., Schmechel, S., Bitterman, P.B., Gourdon, G., Swanson, M.S., Moseley, M., Ranum, L.P.W. (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. 108, 260–265. 10.1073/pnas.1013343108
Main figure titles and legends

Figure 1. Flanking sequences modulate the localization of CAG repeat-containing RNA

A, Schematic of repeat-containing constructs. B, C, Quantification of cell toxicity in lines expressing CAG repeats with various flanking sequences (B) or as a function of CAG repeat number for the same flanking sequence as in T1 (C). Each data point represents one experiment; toxicity was measured 5 days post induction. Data are summarized as mean ± s.d. Asterisks denote significance by Student’s t-test. D, Schematic depicting the observed localization of CAG repeat-containing RNA (top). Representative fluorescence micrographs 24 h post-induction (bottom). Arrows denote representative cytoplasmic aggregates. Micrographs are representative of ≥ 40 cells from ≥ 3 independent experiments. E, Quantification of the percent of cells containing cytoplasmic inclusions. Data are representative of at least three experiments. Control cells express mCherry mRNAs (left control on each graph) or the reverse-complement of mCherry RNAs (right control on each graph), each fused to MS2 hairpins. F, Quantification of aggregate area. Each point represents a single cell. Data are summarized as median ± interquartile range. Control cells are the same as in E. G, H, Quantification of the area of cytoplasmic aggregates (G) and nuclear foci (H) at varying doxycycline concentrations which results in varying levels of RNA expression. Each point represents a single cell; data are summarized as median ± interquartile range. Asterisks denote significance by Mann-Whitney test. I, Representative fluorescence micrographs 48 h after transfection with indicated constructs (left). Quantification of percent of cells with cytoplasmic aggregates (right). Each point represents ≥ 10 cells. Data are representative of at least 3 independent experiments. Asterisks denote significance by Student’s t-test. All scale bars, 10 µm. *: p < 0.01, n.s.: not significant (p > 0.01).
Figure 2. Cytoplasmic RNA inclusions are solid-like and deform the nucleus

A, Representative micrographs of a nuclear (top) and cytoplasmic (bottom) inclusion before (pre) and at the stated timepoints after photobleaching. B, FRAP trajectories are shown for each inclusion. Data are summarized as mean ± s.d., n ≥ 20 events. Asterisks denote significance by Wilcoxon signed-rank test. C, Representative masks of nuclei depicting nuclear envelope deformation in cells with cytoplasmic RNA aggregates. D, Quantification of the percent of cells with deformed nuclei. Each data point represents an independent experiment with each experiment consisting of ≥ 50 cells. Asterisks denote significance by Student’s t-test. E, Immunofluorescence micrographs showing localization of 240×CAG RNA (MS2-YFP) and lamin B1. Arrowheads denote representative point of nuclear deformation. Immunofluorescence data are representative of 3 independent experiments, each of ≥ 30 cells. Scale bars, 10 µm for A, E and 5 µm for C. *: p < 0.01, n.s.: not significant (p > 0.01).

Figure 3. Cytoplasmic RNA inclusions co-localize with RNA binding proteins.

A-B, Immunofluorescence micrographs (left) and representative line profiles (right) showing localization of 240×CAG RNA (MS2-YFP) and G3BP1/2 (A) and EDC4 (B) in cells with cytoplasmic inclusions. Arrowheads are used as visual guides to indicate representative RNA aggregates. Nuclei were counterstained with DAPI (blue). C, D, Same as A-B except showing TDP-43 (C) and FUS (D) in induced and uninduced controls. Ratio of nuclear:cytoplasmic TDP-43 (C) and FUS (D) as calculated from immunofluorescence micrographs (right, top). Controls are uninduced cells that contain constructs capable of forming cytoplasmic aggregates. Each point represents a single cell and data are summarized as median ± interquartile range. Asterisks
denote significance by Mann-Whitney test. Line profiles as in A-B (right, bottom). Correlation coefficient, $R^2$, for colocalization between MS2-YFP and TDP-43 (C) or FUS (D) (far right). Each point represents a single cell and data are summarized as median ± interquartile range. Asterisks denote significance by Mann-Whitney test. E, Same as A-B except showing localization of p62 in cells with cytoplasmic inclusions. $R^2$ values (far right) as in C-D. All immunofluorescence data are representative of ≥ 2 independent experiments, each of ≥ 40 cells. All scale bars, 10 µm. *: p < 0.01, n.s.: not significant (p > 0.01).

Figure 4. CAG repeat-containing RNA that accumulate at cytoplasmic inclusions undergo RAN translation and co-aggregate with their RAN translation products. A, Immunoblot with an anti-polyglutamine antibody, with actin as a loading control. B, Immunofluorescence micrographs showing localization of 240×CAG RNA (MS2-YFP) and polyglutamine-containing proteins (PolyQ). Nuclei were counterstained with DAPI. PolyQ channels are equally scaled. C, Quantification of $R^2$ values for colocalization between MS2-YFP and polyglutamine. Each point represents a single cell and data are summarized as median ± interquartile range. Asterisks denote significance by Mann-Whitney test. D, Schematic of RAN translation reporter. E, Fluorescence micrographs of RNA (MS2-YFP) and RAN translation products (RAN) at the stated timepoints after induction. All scale bars, 10 µm. Data are representative of two experiments for western blots and at least three experiments for immunofluorescence and live imaging experiments.

Figure 5. Translation is required for CAG repeat RNA aggregation and toxicity.
A, Micrographs of cells expressing the RAN translation reporter construct with doxycycline alone (Control) or with doxycycline plus cycloheximide (Cycloheximide). Micrographs depict RNA (MS2-YFP) and RAN translation products (RAN). B, Quantification of aggregate area in control and cycloheximide treated cells. Each data point represents a single cell; data are summarized as median ± interquartile range. Asterisks denote significance by Mann-Whitney test. C-D, Same as A-B except showing cells treated with doxycycline and either a control morpholino or 8×CTG morpholino. E, Quantification of cell survival after four days of induction with either a control morpholino or 8×CTG morpholino. Each data point represents one experiment, and data are summarized as the mean ± s.d. Asterisks denote significance by Student’s t-test. Data are representative of two experiments for cycloheximide treatments and three experiments for morpholino treatments, each with ≥ 40 cells. All scale bars, 10 µm. *: p < 0.01, n.s.: not significant (p > 0.01).
A. 

B. 

C. 

D. 

E.
**A**

| Dox | N1 | C1 |
|-----|----|----|
| -   | +  | -  |
| +   | C1 | +  |

**α-PolyQ**

**α-Actin**

**B**

- **MS2-YFP**
- **PolyQ**
- **DAPI**
- **Merge**

**Cyt Agg (C2)**

**Nuc foci (N2)**

**C**

- **R²**
- **Ctrl**
- **Dox**

**D**

- **Tet-On**
- **Cyt agg flank**
- **Stop codons**
- **240×CAG**
- **BFP (-ATG)**
- **MS2 hairpins**

**E**

- **MS2-YFP**
- **0 hr**
- **6 hr**
- **12 hr**
- **18 hr**
- **24 hr**

- **RAN**
- **Merge**

**DoxCtrl**

*
