PNAS

Supporting Information for:
Repeat-associated non-AUG translation causes cytoplasmic aggregation of CAG repeat-containing RNAs and induces cellular toxicity in human cells

Michael R. Das[1, 2], Yeonji Chang[1], Rachel Anderson[1,2], Reuben Saunders[3], Nan Zhang[3], Colson Tomberlin[1], Ronald D. Vale[3, 4, 5], and Ankur Jain[1, 2,*]

Corresponding author: Ankur Jain
Email: ajain@wi.mit.edu

This PDF file includes:

Supporting text
Figures S1 to S6
Table S1
Legends for Movies S1 to S5
Legend for Datasets S1

Other supporting materials for this manuscript include the following:

Movies S1 to S5
Dataset S1

Supporting Information Text

Supplemental Discussion

The MS2 system for RNA visualization requires incorporation of multiple MS2 stem loops, and the target RNA is indirectly visualized upon the binding of multiple fluorescent protein tagged MS2-coat proteins (1). Previous reports have shown that the MS2 tag can affect RNA degradation and promote its accumulation at P-bodies in yeast (2–5). These caveats need to be considered when using this system for assessing RNA localization. We previously showed that under our expression scheme, control RNAs (encoding mCherry or a non-coding sequence - reverse complement of mCherry) that were tagged with 12xMS2 hairpins and co-expressed with MS2CP-YFP did not form appreciable inclusions in the nucleus or in the cytoplasm of U-2OS cells (6).

The MS2 system detects RNA indirectly through MS2CP-YFP, and thus any localization patterns could potentially be caused by the reporter itself, rather than the RNA. In particular, the fluorescent proteins used for indirect RNA visualization, such as GFP, are prone to oligomerization (7), which may potentiate unwarranted aggregation of the MS2-tagged RNA. We used a monomerized mutant of YFP (A206K) that substantially reduces the dimerization propensity of YFP (7). Nevertheless, the high local concentration of fluorescent proteins at nuclear foci or cytoplasmic aggregates may exaggerate RNA aggregation. To further confirm that the observed sub-cellular localization of CAG repeat containing RNA is not driven by the MS2-tag, we performed control experiments using CAG repeat-containing RNAs lacking the MS2-tag and expressed these RNAs in a cell line without the co-expression of MS2CP-YFP. We observed similar localization patterns of RAN-translated CAG repeat RNA with and without the MS2-tag as assessed in fixed cells using FISH (Fig. S1D-E). We also observed comparable effects on the mislocalization of RNA binding proteins (Fig. S3E-F), RAN protein production (Fig. S2E), and cell toxicity (Fig. S4D-E) when expressing tagged or untagged RNA.

Fig. S1

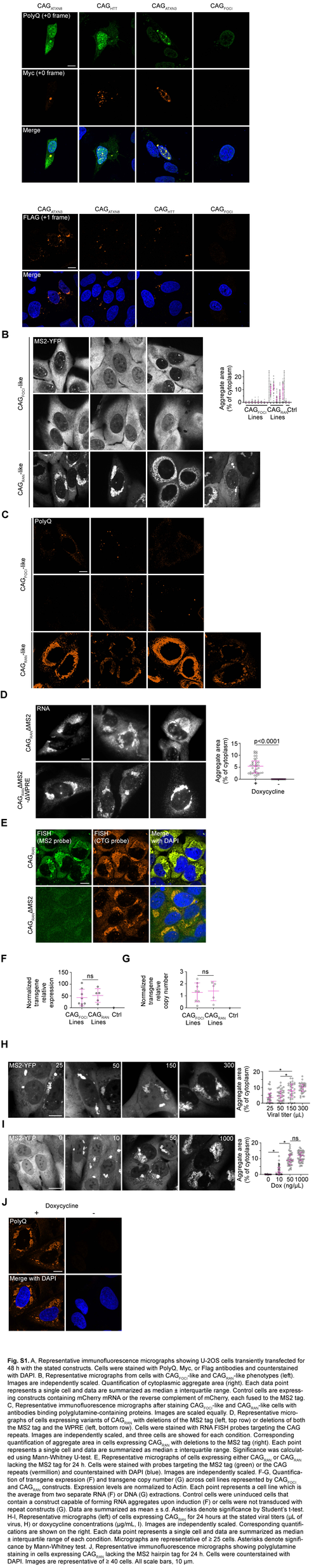

Fig. S1. A, Representative immunofluorescence micrographs showing U-2OS cells transiently transfected for 48 h with the stated constructs. Cells were stained with PolyQ, Myc, or Flag antibodies and counterstained with DAPI. B, Representative micrographs from cells with CAG_{FOCI}-like and CAG_{RAN}-like phenotypes (left). Images are independently scaled. Quantification of cytoplasmic aggregate area (right). Each data point represents a single cell and data are summarized as median ± interquartile range. Control cells are expressing constructs containing mCherry mRNA or the reverse complement of mCherry, each fused to the MS2 tag. C, Representative immunofluorescence micrographs after staining CAG_{FOCI}-like and CAG_{RAN}-like cells with antibodies binding polyglutamine-containing proteins. Images are scaled equally. D, Representative micrographs of cells expressing variants of CAG_{RAN} with deletions of the MS2 tag (left, top row) or deletions of both the MS2 tag and the WPRE (left, bottom row). Cells were stained with RNA FISH probes targeting the CAG repeats. Images are independently scaled, and three cells are showed for each condition. Corresponding quantification of aggregate area in cells expressing CAG_{RAN} with deletions to the MS2 tag (right). Each point represents a single cell and data are summarized as median ± interquartile range. Significance was calculated using Mann-Whitney U-test. E, Representative micrographs of cells expressing either CAG_{RAN} or CAG_{RAN} lacking the MS2 tag for 24 h. Cells were stained with probes targeting the MS2 tag (green) or the CAG repeats (vermillion) and counterstained with DAPI (blue). Images are independently scaled. F-G, Quantification of transgene expression (F) and transgene copy number (G) across cell lines represented by CAG_{FOCI} and CAG_{RAN} constructs. Expression levels are normalized to Actin. Each point represents a cell line which is the average from two separate RNA (F) or DNA (G) extractions. Control cells were uninduced cells that contain a construct capable of forming RNA aggregates upon induction (F) or cells were not transduced with repeat constructs (G). Data are summarized as mean ± s.d. Asterisks denote significance by Student's t-test. H-I, Representative micrographs (left) of cells expressing CAG_{RAN} for 24 hours at the stated viral titers (μL of virus, H) or doxycycline concentrations (μg/mL, I). Images are independently scaled. Corresponding quantifications are shown on the right. Each data point represents a single cell and data are summarized as median ± interquartile range of each condition. Micrographs are representative of ≥ 25 cells. Asterisks denote significance by Mann-Whitney test. J, Representative immunofluorescence micrographs showing polyglutamine staining in cells expressing CAG_{RAN} lacking the MS2 hairpin tag for 24 h. Cells were counterstained with DAPI. Images are representative of ≥ 40 cells. All scale bars, 10 μm.

Fig. S2.

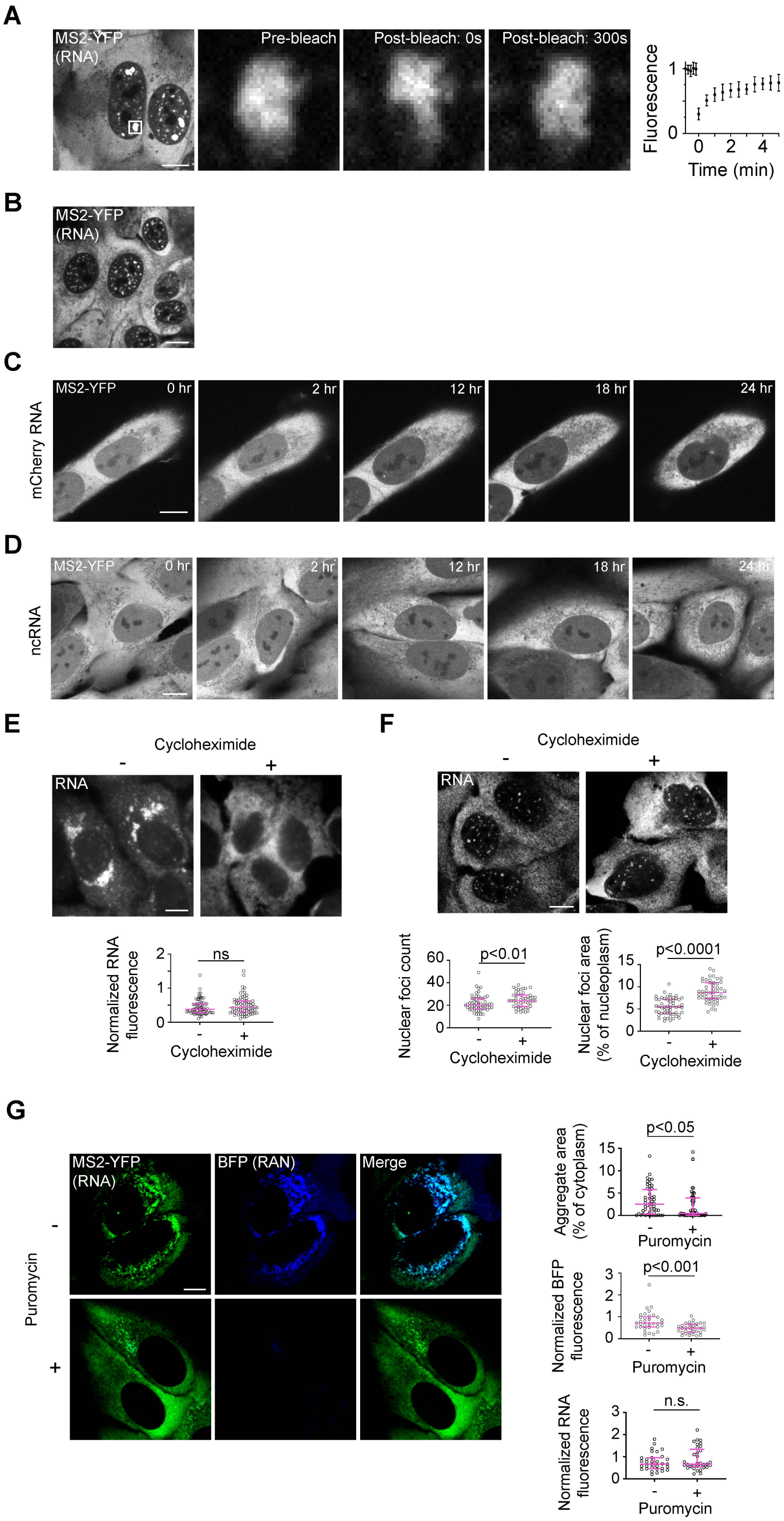

Fig. S2. A, Representative micrographs from a cell expressing CAG_{RAN} for 1 h (left). Inset shows a nuclear focus prior to and at the stated times after photobleaching. Data are representative of ≥ 10 events. Images are independently scaled. FRAP trajectory is shown with data represented as mean ± standard deviation (right). B, Representative micrograph from cells expressing CAG_{FOCI} for 6 d. Data is representative of ≥ 40 cells. C-D, Representative micrographs from cells expressing mCherry mRNA (C) and the reverse complement of mCherry (D) for the stated amounts of time. Data are representative of ≥ 4 movies. E, Representative micrographs of cells expressing CAG_{RAN} for 12 h (control) or expressing CAG_{RAN} for 12 h in the presence of cycloheximide (10 μg/mL) (top). Cells were stained with RNA FISH probes targeting the repeat RNA. Quantification of RNA levels normalized to cell area (bottom). Each point represents a single cell and data are summarized as median ± interquartile range. F, Representative micrographs of cells expressing CAG_{FOCI} for 12 h (control) or expressing CAG_{FOCI} for 12 h in the presence of cycloheximide (10 μg/mL) (left). Cells were stained with RNA FISH probes targeting the repeat tract. Quantification of RNA foci count and area (right). Each point represents a single cell and data are summarized as median ± interquartile range. G, Representative micrographs of cells expressing CAG_{RAN}BFP for 12 h (control) or expressing CAG_{RAN}BFP for 12 h in the presence of puromycin (10 μg/mL) (left). MS2-YFP images are independently scaled; BFP images are scaled equally. Quantification of cytoplasmic RNA aggregate area (right, top graph), BFP fluorescence normalized to cell area (right, middle graph), and RNA FISH fluorescence normalized to cell area (right, bottom graph). Each data point represents a single cell and data are summarized as median ± interquartile range. Significance was calculated by Mann-Whitney U-test. All scale bars, 10 μm.

Fig. S3.

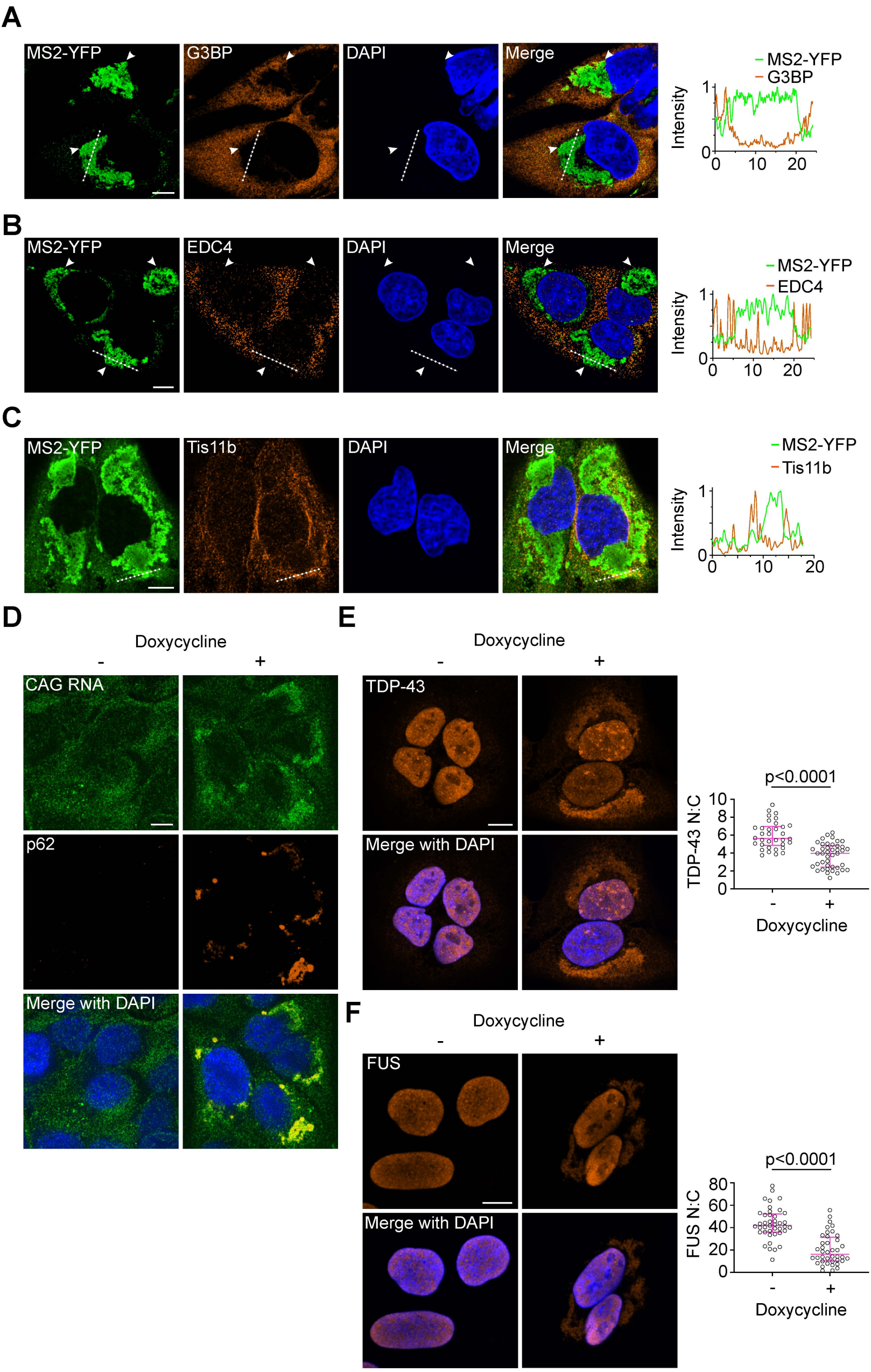

Fig. S3. A-C, Representative immunofluorescence micrographs of cells expressing CAG_{RAN} for 24 h (left). Cells were stained with antibodies against G3BP (A), EDC4 (B), and Tis11b (C) and were counterstained with DAPI. Intensity profiles of MS2-YFP and the stated epitopes along the dashed lines in the micrographs (right). D, Dual-IF FISH micrographs showing p62 protein localization and weak staining of CAG repeat RNA in cells expressing CAG_{RAN} lacking the MS2 tag for 24 h. Cells were fixed with glyoxal, stained with p62 antibodies, refixed with glyoxal, and then stained with FISH probes targeting the repeat region. Cells were counterstained with DAPI. E-F, Representative immunofluorescence micrographs of cells expressing CAG_{RAN} lacking the MS2 tag for 24 h or uninduced control cells (left). Cells were stained with antibodies against TDP-43 (E) or FUS (F) and counterstained with DAPI. Quantification of nuclear to cytoplasmic ratio of each protein (right). Each datapoint represents a single cell. Data are summarized as median ± interquartile range. Significance values were calculated using Mann-Whitney U-tests. All scale bars, 10 μm.

Fig. S4.

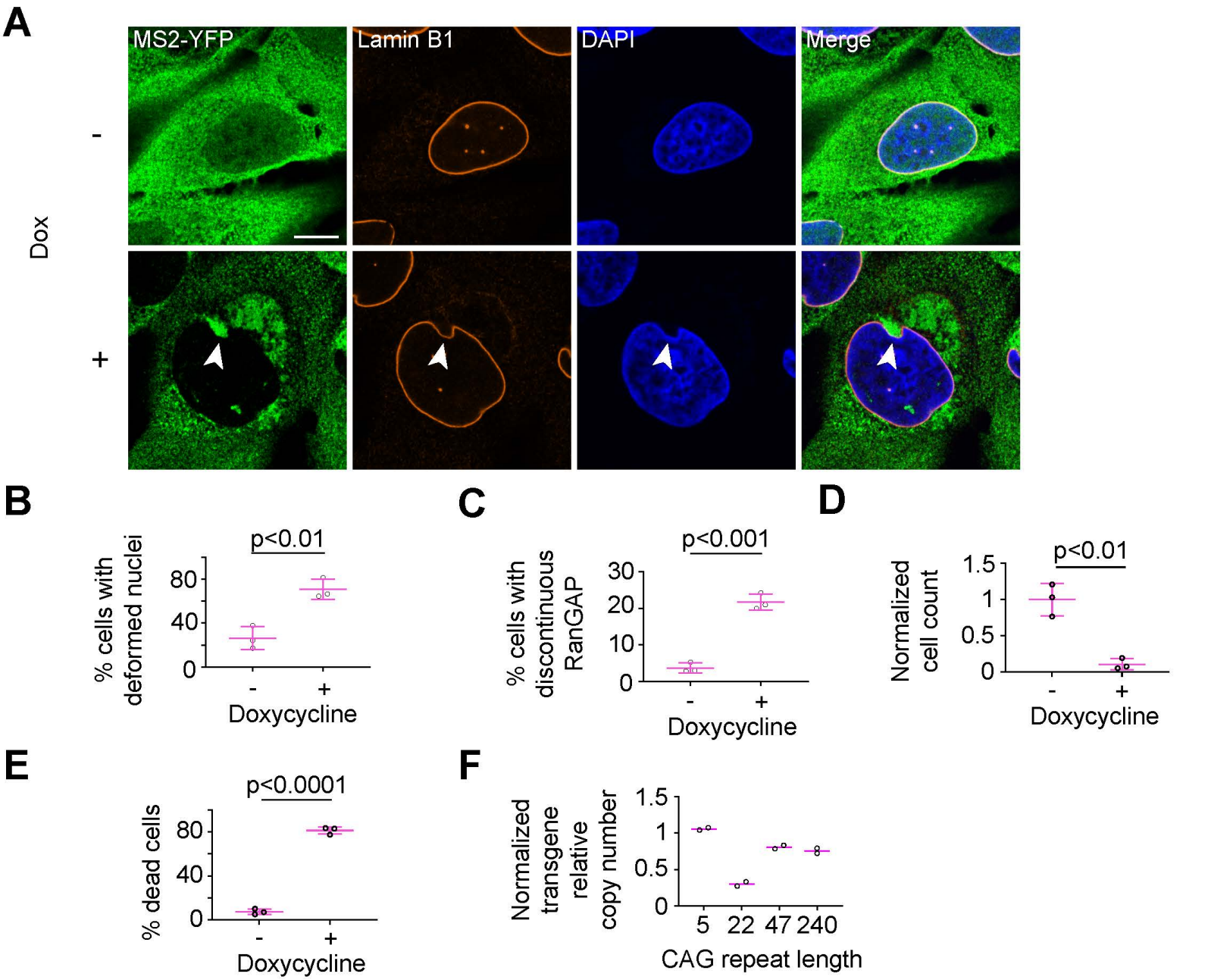

Fig. S4. A, Immunofluorescence micrographs showing localization of lamin B1 and MS2-YFP in uninduced control cells and in cells expressing CAG_{RAN} for 24 h. Arrowheads denote representative point of nuclear deformation. MS2-YFP images are independently scaled; lamin images are scaled equally. Immunofluorescence data are representative of 3 independent experiments, each of ≥ 30 cells. B-C, Quantification of cells with deformed nuclei as assessed by immunofluorescence staining for lamin (B) and discontinuous RanGAP1 staining (C) after expression of CAG_{RAN} lacking the MS2 tag for 24 h. Each datapoint represents ≥ 30 cells and data are summarized as mean ± s.d. Significance value was calculated by Student's t-test. D, Quantification of cell populations after expressing CAG_{RAN} lacking the MS2 tag for 4 days. Each point represents a separate well (biological replicate) and is the summary of at least two technical replicate measurements. Cell counts are normalized to uninduced controls. Data are summarized as mean ± s.d. Significance value was calculated by Student's t-test. E, Quantification of cell death caused by expression of CAG_{RAN} lacking the MS2 tag for 4 days, as measured by Trypan Blue staining. Each point represents a separate well and is the summary of at least two technical replicate measurements. Data are summarized as the mean of the three well replicates (biological replicates). F, Quantification of integration efficiency across cell lines expressing variations of CAG_{RAN} with varying lengths of the CAG repeat tract. Copy numbers were normalized to Actin. Each data point represents a separate DNA extraction. Data are summarized as mean. All scale bars, 10 μm.

Fig. S5.

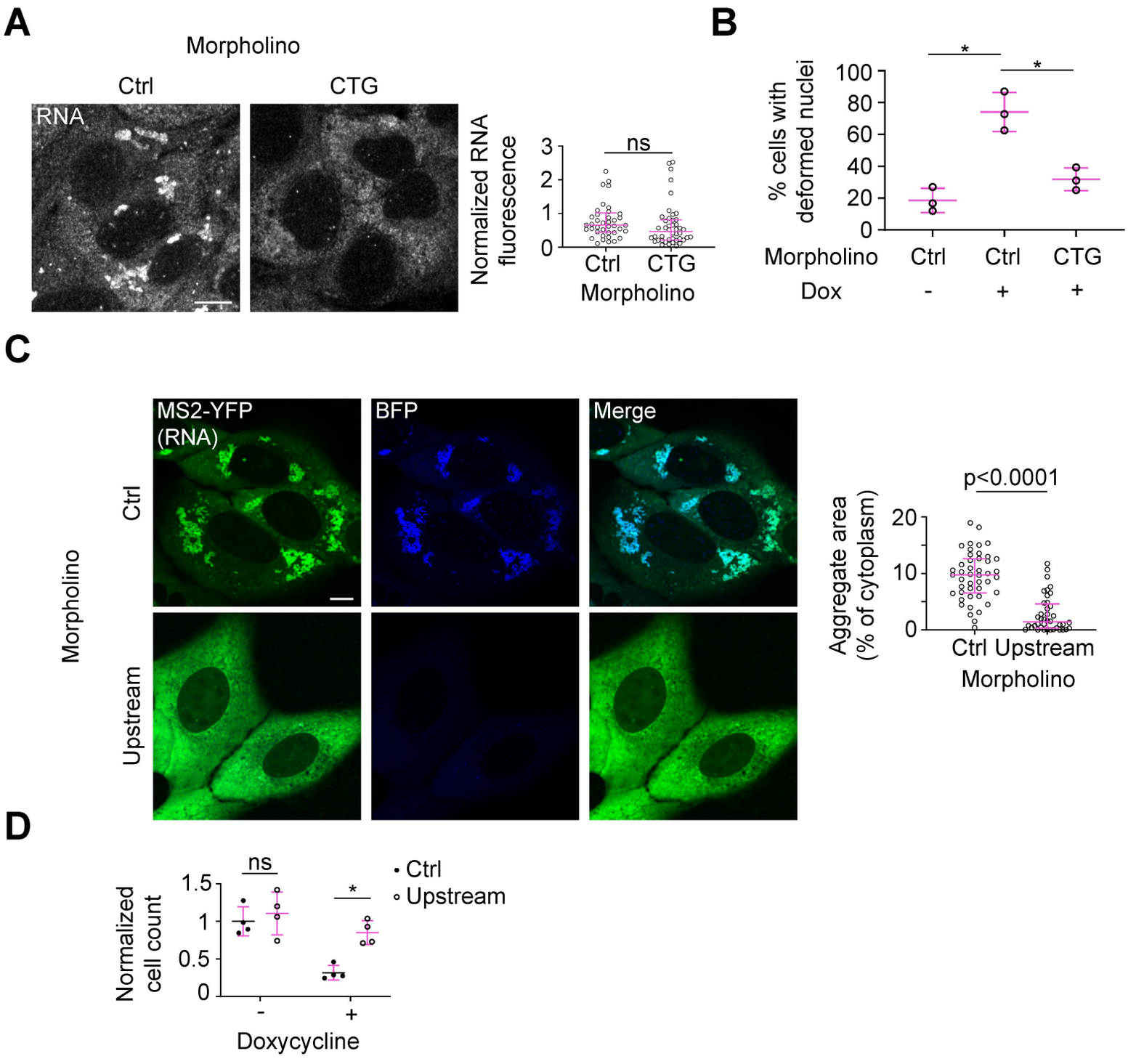

Fig. S5. A, Representative micrographs of cells expressing CAG_{RAN} for 24 h in the presence of either a control morpholino or a morpholino targeting the CAG repeat tract (morpholino concentration 50 μM) (left). Cells were stained with RNA FISH probes targeting the MS2 tag. Quantification of RNA levels normalized to cell area (right). Each point represents a single cell and data are summarized as median ± interquartile range. B, Quantification of cells with deformed nuclei expressing CAG_{RAN} for 24 h treated with morpholinos as in (A). Each data point represents ≥ 30 cells across ≥ independent experiments. Data are summarized as mean ± standard deviation. Significance was calculated by Student's t-test. C, Representative micrographs of cells expressing CAG_{RAN}BFP for 24 h and treated with either a control morpholino or a morpholino targeting the region upstream of the CAG repeat tract (morpholino concentration 50 μM) (left). MS2-YFP images are independently scaled; BFP images are scaled equally. Quantification of cytoplasmic RNA aggregate area (right). Each point represents a single cell. Data are summarized as median ± interquartile range. Significance value is calculated by Mann-Whitney U-test. D, Quantification of cell survival after four days of CAG_{RAN} expression in the presence of either a control morpholino or a morpholino targeting the region upstream of the repeat tract. Each data point represents data from one well measured with at least two technical replicates. Data are summarized as the mean ± s.d. Asterisks denote significance by Student's t-test. All scale bars, 10 μm.

Fig. S6.

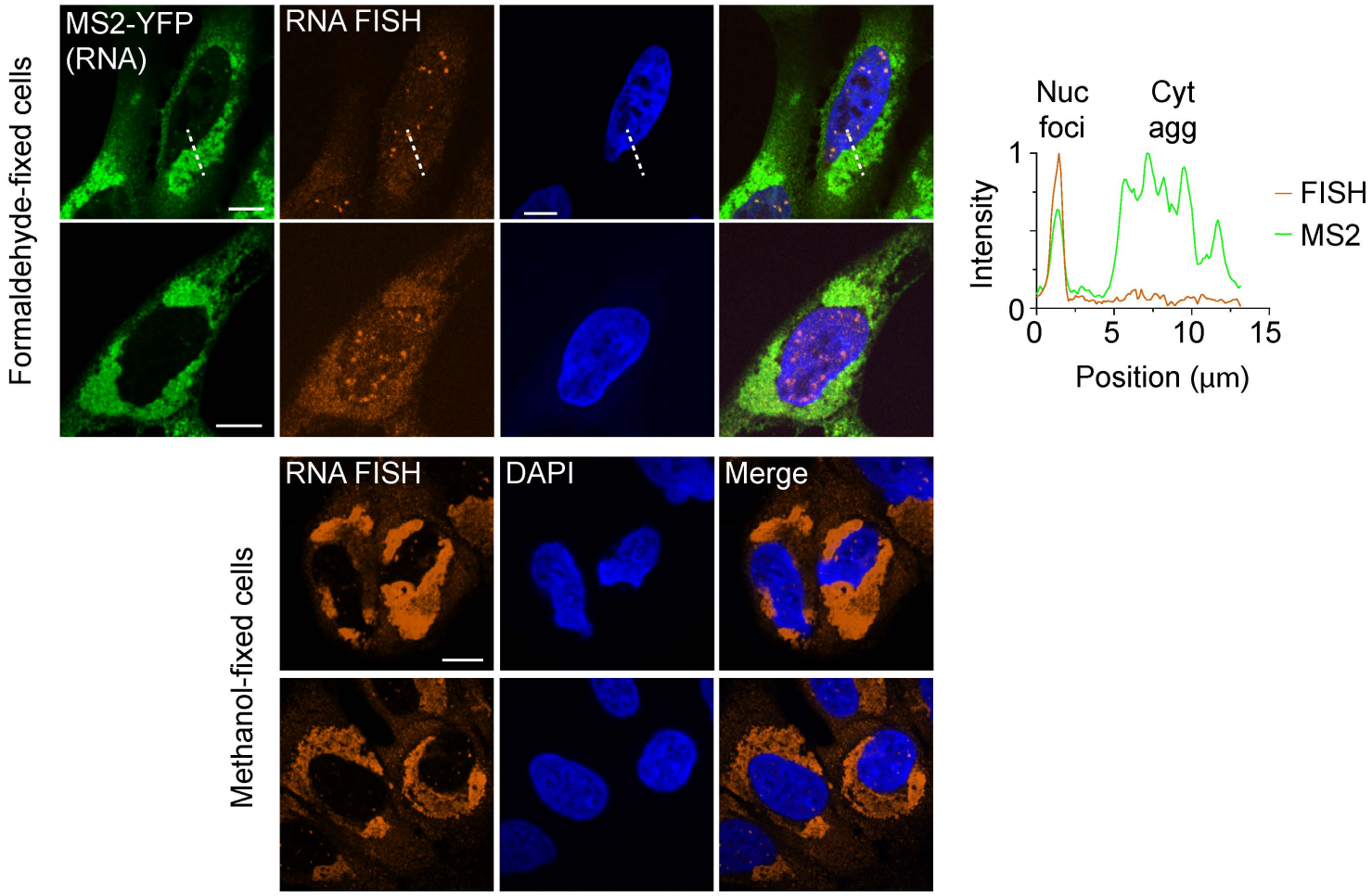

Fig. S6. Fluorescence micrographs of cells expressing CAG_{RAN} for 24 h fixed with formaldehyde and stained with RNA FISH probes (top, left). Intensity profile showing RNA intensities as detected by MS2-YFP (green) and by RNA FISH (vermillion) (top, right). Regions of the line corresponding to a nuclear focus and a cytoplasmic aggregate are labelled. Fluorescence micrographs of cells expressing CAG_{RAN} for 24 h fixed with a methanol fixation solution and stained with RNA FISH probes. MS2-YFP is not shown as methanol denatures and prevents fluorescence detection of the protein. All scale bars, 10 μm.

Table S1. List of reagents and resources used in this study

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-lamin B1	Abcam	Cat#ab16048
Mouse anti-G3BP	Abcam	Cat#ab56574
Rabbit anti-EDC4	Abcam	Cat#ab72408
Mouse anti-p62/SQSTM1	Abcam	Cat#ab56416
Rabbit anti-TDP-43	Proteintech	Cat#10782-2-AP
Rabbit anti-FUS	Invitrogen	Cat#PA552610
Rabbit anti-TIS11b/ZFP36L1	Abcam	Cat#ab230507
Goat anti-GFP	Rockland	Cat#600-101-215
Mouse anti-polyglutamine	Millipore-Sigma	Cat#MAB1574
Alexa488-conjugated donkey anti-goat	Invitrogen	Cat#A11055
Cy3-conjugated donkey anti-mouse	Jackson ImmunoResearch	Cat#715-165-151
HRP-conjugated rabbit anti-mouse	Sigma-Aldrich	Cat#A9044-2ML
Cy3-conjugated donkey anti-rabbit	Jackson ImmunoResearch	Cat#711-165-152
Bacterial and virus strains		
Stbl3 *E. coli*	Invitrogen	Cat#C7373-03
Biological samples		
Chemicals, peptides, and recombinant proteins		
EndoPorter peptide	GeneTools	SKU: OT-EP-PEG-1
Methanol	Millipore-Sigma	Cat#MX0475-1
Acetic acid	Sigma-Aldrich	Cat#695092
UltaPure 20X SSC Buffer	Invitrogen	Cat#15557-044
NP-40 substitute	Fisher Scientific	Cat#AAJ19628AP
Formamide	Sigma-Aldrich	Cat#47671
16% Formaldehyde solution (w/v) Methanol-free	Thermo Scientific	Cat#28906
Sodium citrate	Calbiochem	Cat#567446
Dextran sulfate	Sigma-Aldrich	Cat#D8906
Triton-X-100	Sigma-Aldrich	Cat#T8787
5M Sodium chloride	Promega	Cat#V4221
Sodium deoxycholate	Sigma-Aldrich	Cat#D6750
SDS	Bio-Rad	Cat#1610302
Tween-20	FisherBiotech	Cat#BP337
DTT	Thermo Scientific	Cat#R0861
Skim milk powder	BD Biosciences	Cat#232100
DAPI	Sigma-Aldrich	Cat#D9542
Doxycycline	Sigma-Aldrich	Cat#D9891
Cycloheximide	Sigma-Aldrich	Cat#C1998
Polybrene	Millipore-Sigma	Cat#TR-1003-G
Lipofectamine LTX	Invitrogen	Cat#15338-100
Viafect transfection reagent	Promega	Cat#E4981
Trypan-blue 0.4%	Invitrogen	Cat#T10282
Bovine-serum albumin	Sigma-Aldrich	Cat#2930
Glyoxal solution (~40%) in H_2O	Sigma-Aldrich	Cat#50649
Critical commercial assays		

PureLink Genomic DNA Mini kit	Invitrogen	Cat#K1820-01
PureLink RNA Mini kit	Invitrogen	Cat#12183018A
SuperScript III Reverse transcriptase	LifeTechnologies	Cat#11754050
SYBR Green PCR Master Mix	AppliedBiosystems	Cat#4309155
SuperSginal West Femto Maximum Sensitivity Substrate	Thermo Scientific	Cat#34095
Protease and phosphatase inhibitors	Thermo Scientific	Cat#78442
Experimental models: Cell lines		
Human: U-2 OS cells	ATCC	HTB-96
Human: HEK293T cells	ATCC	CRL-3216
Oligonucleotides		
qPCR primer WPRE_F: TGTCGGGGAAATCATCGTCC	IDT	N/A
qPCR primer WPRE_R: AAGGAAGGTCCGCTGGATTG	IDT	N/A
qPCR primer TetPromoter_F: AACGTATGTCGAGGTAGGCG	IDT	N/A
qPCR primer TetPromoter_R: ATTGCTCCAGGCGATCTGAC	IDT	N/A
qPCR primer Actin_F: GCTACGAGCTGCCTGACG	IDT	N/A
qPCR primer Actin_R: GGCTGGAAGAGTGCCTCA	IDT	N/A
7xCTG FISH probe: Cy3/CTGCTGCTGCTGCTGCTGCTG	IDT	N/A
MS2 FISH probe: Cy3/TTCTAGAGTCGACCTGCAG	IDT	N/A
8xCTG morpholino: CTGCTGCTGCTGCTGCTGCTG	GeneTools	N/A
Standard intron control morpholino: CCTCTTACCTCAGTTACAATTTATA	GeneTools	SKU: PCO-StandardControl-300
Upstream morpholino: CGACGGTGGCCAGGAACCTCATAT	GeneTools	N/A
Recombinant DNA		
CAG_{RAN} Lines	This study	Dataset S1
CAG_{FOCI} Lines	This study	Dataset S1
CAG_{RAN} ΔMS2	This study	Dataset S1
CAG_{RAN} ΔMS2, ΔWPRE	This study	Dataset S1
mCherry 12xMS2	(6)	N/A
mCherry-revComp 12xMS2	(6)	N/A
CAG repeats with *ATXN8* flanking sequence: pcDNA3.1 ATXN8 KKQ	(8)	N/A
CAG repeats with *HTT* flanking sequence: pcDNA3.1 HTT	(8)	N/A
CAG repeats with *ATXN3* flanking sequence: pcDNA3.1 ATXN3	(8)	N/A
MS2 coat protein – YFP fusion: pHR-tdMS2CP-YFP-WPRE	(6)	N/A
CAG_{RAN}-BFP	This study	Supp. Table 2
CAG_{RAN}-Stop-BFP	This study	Dataset S1
CAG_{RAN}-3xTag	This study	Dataset S1

Lentiviral envelope: pCMV-VSV-G	(9)	Addgene plasmid #8454
Lentiviral packaging: psPAX2	Gift from Didier Trono	Addgene plasmid #12260
NLS-tdTomato-NES	(10)	N/A
Software and algorithms		
ImageJ	(11)	https://imagej.nih.gov/ij/
Fusion	Oxford Instruments Andor	https://andor.oxinst.com/downloads/view/fusion-release-2.3
Other		
Dragonfly spinning disk confocal microscope	Oxford Instruments Andor	Model#Dragonfly505
iXon Ultra 888 EMCCD camera	Oxford Instruments Andor	Model#DU-888U3-CS0-#BV
100X Oil Immersion Objective, NA 1.45	Nikon	CAT# MRD01905
96-well glass bottom plates	Brooks	CAT# MGB096-1-2-LG-L
iBlot 2	Invitrogen	CAT# IB21001
Countess II FL automated cell counter	Invitrogen	CAT# AMQAF1000
Countess II disposable cell counting chamber slides	Invitrogen	CAT# C10283
DMEM	Gibco	CAT# 11965-126
Fetal bovine serum	Gibco	CAT# 26140-079
Penicillin-streptomycin-glutamine 100X	Gibco	CAT# 10378016
Opti-MEM	Gibco	CAT# 31985-070
IMDM	Gibco	CAT# 12440-053
DPBS	Gibco	CAT# 14190-144
Trypsin-EDTA 0.25%	Gibco	CAT# 25200-072
PBS, pH=7.2	Gibco	CAT# 20012-027
Nuclease-free water	Invitrogen	CAT# AM9932
Bovine serum albumin (BSA)	Sigma-Aldrich	CAT# A7906
4xBolt LDS sample buffer	Invitrogen	CAT# B0007
4-12% Bis-tris polyacrylamide gel	Invitrogen	CAT# NW04122

Movie S1 (separate file). FRAP of cytoplasmic RNA aggregates.
Time-lapse of cytoplasmic RNA aggregate bleached at t = 0 s. Cells transduced with construct CAG_{RAN} were induced for 24 h prior to bleaching. Arrowheads denote photobleached regions. Images are taken at 30 s intervals after bleaching. Scale bar, 10 μm.

Movie S2 (separate file). Real-time visualization cytoplasmic RNA aggregation.
Time-lapse movie of cells transduced with CAG_{RAN}. Expression of CAG_{RAN} construct is induced at time t = 0. Images show MS2-YFP. Images are taken at 30 min intervals after induction. Scale bar, 10 μm.

Movie S3 (separate file). FRAP of nuclear RNA foci.
Time-lapse of cytoplasmic RNA aggregate bleached at t = 0 s. Cells transduced with construct CAG_{RAN} were induced for 2 h prior to bleaching. Arrowheads denote photobleached regions. Images are taken at 30 s intervals after bleaching. Scale bar, 10 μm.

Movie S4 (separate file). Real-time visualization of RNA foci.
Time-lapse movie of cells transduced with CAG_{FOCI}. Expression of CAG_{FOCI} construct is induced at time t = 0. Images show MS2-YFP. Images are taken at 30 min intervals after induction. Scale bar, 10 μm.

Movie S5 (separate file). Real-time visualization of RAN translation and RNA aggregation.
Time-lapse movie of cells transduced with CAG_{RAN}-BFP translation reporter. Expression of CAG_{RAN}-BFP construct is induced at time t = 0. Images show composite (cyan) of

MS2-YFP (green) and BFP (blue) channels. Images are taken at 30 min intervals after induction. Scale bar, 10 μm.

Dataset S1 (separate file). Full sequences of plasmids generated.

SI References

1. E. Bertrand, *et al.*, Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).

2. J. F. Garcia, R. Parker, Ubiquitous accumulation of 3′ mRNA decay fragments in *Saccharomyces cerevisiae* mRNAs with chromosomally integrated MS2 arrays. *RNA* **22**, 657–659 (2016).

3. E. Tutucci, *et al.*, An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods* **15**, 81–89 (2018).

4. G. Haimovich, *et al.*, Use of the MS2 aptamer and coat protein for RNA localization in yeast: A response to "MS2 coat proteins bound to yeast mRNAs block 5′ to 3′ degradation and trap mRNA decay products: implications for the localization of mRNAs by MS2-MCP system." *RNA* **22**, 660–666 (2016).

5. J. F. Garcia, R. Parker, MS2 coat proteins bound to yeast mRNAs block 5′ to 3′ degradation and trap mRNA decay products: implications for the localization of mRNAs by MS2-MCP system. *RNA* **21**, 1393–1395 (2015).

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

7. D. A. Zacharias, J. D. Violin, A. C. Newton, R. Y. Tsien, Partitioning of Lipid-Modified Monomeric GFPs into Membrane Microdomains of Live Cells. *Science* **296**, 913–916 (2002).

8. T. Zu, *et al.*, Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci.* **108**, 260–265 (2011).

9. S. A. Stewart, *et al.*, Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA N. Y. N* **9**, 493–501 (2003).

10. H. Zhang, *et al.*, RNA Controls PolyQ Protein Phase Transitions. *Mol. Cell* **60**, 220–230 (2015).

11. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).