

RNA phase transitions in repeat expansion disorders

Ankur Jain^{1,2} & Ronald D. Vale^{1,2}

Expansions of short nucleotide repeats produce several neurological and neuromuscular disorders including Huntington disease, muscular dystrophy, and amyotrophic lateral sclerosis. A common pathological feature of these diseases is the accumulation of the repeat-containing transcripts into aberrant foci in the nucleus. RNA foci, as well as the disease symptoms, only manifest above a critical number of nucleotide repeats, but the molecular mechanism governing foci formation above this characteristic threshold remains unresolved. Here we show that repeat expansions create templates for multivalent base-pairing, which causes purified RNA to undergo a sol–gel transition *in vitro* at a similar critical repeat number as observed in the diseases. In human cells, RNA foci form by phase separation of the repeat-containing RNA and can be dissolved by agents that disrupt RNA gelation *in vitro*. Analogous to protein aggregation disorders, our results suggest that the sequence-specific gelation of RNAs could be a contributing factor to neurological disease.

Nucleotide repeat expansion disorders constitute some of the most common inherited diseases^{1,2}. Several of the disease-associated repeat expansions comprise a nucleotide triplet of high G/C content, such as CAG in Huntington disease and spinocerebellar ataxias, and CTG in myotonic dystrophy^{2,3}. Likewise, the expansion of the hexanucleotide GGGGCC in the *C9orf72* gene is the most common mutation associated with familial amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)^{4,5}. Disease-causing repeats can occur in the coding or non-coding regions of RNA transcripts². Intriguingly, in all the nucleotide repeat expansion disorders identified so far, disease only manifests beyond a critical number of repeats^{2,4–6}.

Three non-exclusive hypotheses have been put forward to explain how repeat expansions give rise to neurotoxicity. First, expansions could result in a gain or loss of function of the associated gene¹. However, mutations or deletion of the gene associated with the repeats do not always elicit the disease phenotype, while expression of the expanded repeats alone is sufficient to replicate the disease pathology in animal models^{7–9}. Second, repeat-containing RNAs can be non-canonically translated¹⁰ to produce short peptides that can give rise to neurotoxicity¹¹. Third, the repeat-containing RNA accumulate in the nuclei as aberrant foci^{4,8,12,13} that sequester various RNA binding proteins^{14–16}, resulting in a potential disruption of cellular homeostasis^{17,18}. The molecular mechanism by which a threshold number of nucleotide repeats results in formation of nuclear foci is not known. Here, we tested the hypothesis that sequence-specific physicochemical properties of repeat-containing RNAs may lead to their coalescence into nuclear foci.

Repeat-containing RNAs form gels *in vitro*

RNAs associated with many repeat expansion diseases contain a regular pattern of G and C nucleotides and can form intramolecular hairpins³. Because of their repeating nature, these sequences, in principle, can also template multivalent intermolecular interactions (Fig. 1a). Such multivalent interactions could potentially lead to the formation of large clusters via liquid–liquid phase separation^{19,20} or a sol–gel transition^{21–23}. To examine whether repeat-containing RNAs assemble into large clusters, we synthesized fluorescently labelled RNAs containing 47 triplet repeats of CAG (47×CAG) or CUG (47×CUG).

As controls, we used RNAs of equivalent length (~250 bases), but with arbitrary sequences with 30–75% GC content, and RNAs with scrambled sequences but with identical base composition as 47×CAG and 47×CUG. Upon annealing (see Methods), the 47×CAG and 47×CUG RNAs formed micrometre-sized spherical clusters, while the control RNAs remained soluble (Fig. 1b and Extended Data Fig. 1a, b). Clusters were observable at RNA concentrations as low as 25 nM (Extended Data Fig. 1c). These clusters were enriched more than 100-fold in the RNA compared with the solution phase and contained nearly half of the RNA in the reaction (Extended Data Fig. 1d). The clusters were dissolved by RNaseA, but not proteinase K or DNase I (Extended Data Fig. 1e, f), confirming that clustering was not mediated by protein or DNA contaminants. RNA clustering required Mg²⁺ and was inhibited by monovalent cations such as Na⁺ (Extended Data Fig. 1g, h), suggesting that, besides multivalent base-pairing, electrostatic interactions also play a prominent role in RNA clustering, as has been described for homopolymeric RNAs²⁴.

Molecules that form multivalent interactions show abrupt phase transitions with increasing valency of interaction^{19–21}. Consistent with this valency dependence, we found that the formation of CAG/CUG RNA clusters occurred only with more than 30 triplet repeats (Fig. 1c). The intermolecular interactions between the transcripts potentially could be competed out by shorter complementary antisense oligonucleotides (ASO). Indeed, a 6×CTG ASO prevented clustering of 47×CAG RNA, while control oligonucleotides did not (Fig. 1d). Collectively, these experiments indicate that intermolecular base-pairing interactions in the CAG/CUG-repeat region can lead to the coalescence of RNAs into micrometre-sized clusters.

We next analysed the physical properties of the CAG/CUG RNA clusters. Their spherical shapes (aspect ratio 1.05 ± 0.1 , mean \pm s.d., $n = 214$) are characteristic of polymers undergoing liquid–liquid phase separation²⁵. Molecules within the liquid phase are mobile and undergo fast internal rearrangement. However, fluorescence recovery after photobleaching (FRAP) experiments revealed little or no fluorescence recovery over ~10 min, indicating that RNA in the clusters is immobile (Fig. 1e, Extended Data Fig. 1i and Supplementary Video 1). Because of their solid-like behaviour, we refer to these clusters as ‘RNA gels’^{22,23}. We hypothesize that these RNAs initially phase-separate

¹Department of Cellular and Molecular Pharmacology and Howard Hughes Medical Institute, University of California, San Francisco, California 94158, USA. ²Howard Hughes Medical Institute Summer Institute, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA.

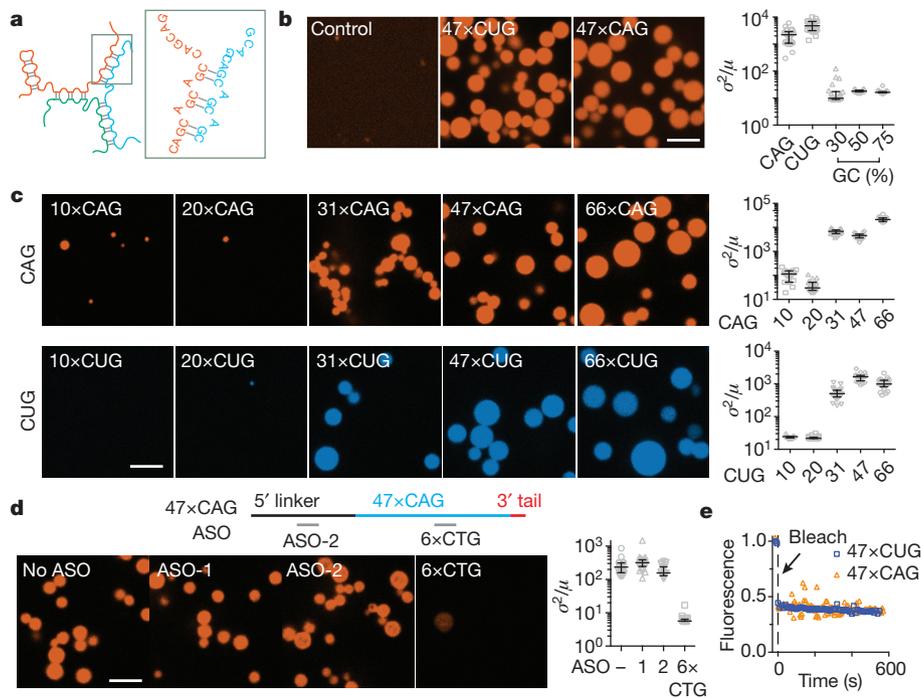


Figure 1 | Triplet repeat-containing RNAs undergo gelation *in vitro* at a critical number of repeats. **a**, Schematic for CAG RNA multimerization. **b**, Fluorescence micrographs for indicated RNA and quantitation of inhomogeneity as normalized variance (σ^2/μ). **c**, Effects of CAG- or CUG-repeat number on RNA gelation. **d**, Effects of control ASO (ASO-2,

and non-hybridizing oligonucleotide, ASO-1) or 6×CTG on 47×CAG RNA clustering. **e**, FRAP trajectories for RNA clusters. Scale bars, 5 μ m. Data are median \pm interquartile range. Data are representative of at least three independent experiments.

into spherical liquid-like droplets but then rapidly become cross-linked into gels because of increasing intermolecular base-pairing. Consistent with this idea, we occasionally ($\sim 1\%$ of clusters) observed incomplete fusion events where two droplets solidified before relaxation to a single spherical geometry (Extended Data Fig. 1j). As further evidence for this model, we found that single-stranded DNA, in the presence of polyvalent cations, formed liquid-like droplets and that incorporation of multivalent base-pairing sites progressively imparted solid-like properties to the DNA droplets (Extended Data Fig. 2 and Supplementary Video 2). In summary, multivalent base-pairing interactions lead to the gelation of CAG/CUG-repeat-containing RNAs *in vitro* at a similar critical number of repeats as observed in diseases.

CAG-repeat RNAs phase-separate in cells

We next sought to establish a live-cell reporter assay in U-2OS cells to visualize repeat-containing RNAs and determine whether they form aberrant nuclear foci, as described in patient cells¹³. For this purpose, we tagged the RNA with 12×MS2-hairpin loops²⁶ and co-expressed yellow fluorescent protein (YFP)-tagged MS2-coat binding protein (MS2CP-YFP) (Fig. 2a). Multiple stop codons were incorporated upstream of the repeats to minimize translation of repeat-containing proteins²⁷ (Fig. 2a). Upon induction of 47×CAG or 120×CAG RNA transcription, numerous nuclear foci appeared as early as 1 h after induction (Fig. 2b, Extended Data Fig. 3a–c and Supplementary Video 3). The number of foci per nucleus increased with higher levels of RNA induction (Extended Data Fig. 3d). In contrast, 5×CAG (Fig. 2b, c and Extended Data Fig. 4) or control RNAs with coding or non-coding sequences did not form nuclear puncta (Extended Data Fig. 3e, f) when expressed at similar levels ($\sim 10,000$ copies per cell; Extended Data Fig. 3g, h). The formation of foci did not induce discernible cell death or impede cell growth over a 7-day period after induction (Extended Data Fig. 3i). Since repeat expansion disorders take years to manifest in patients², short-term toxicity in cells is not necessarily expected.

The CAG RNA nuclear foci exhibited liquid-like properties. For example, two or more foci could fuse with one another (Fig. 2d, e and

Supplementary Videos 3 and 4), a hallmark of liquid-like behaviour²⁸. Upon photobleaching, nuclear foci also exhibited near-complete fluorescence recovery ($83 \pm 13\%$ recovery, time constant $\tau_{\text{FRAP}} = 81 \pm 24$ s, mean \pm s.d., $n = 5$ foci), indicating that the RNA can move into

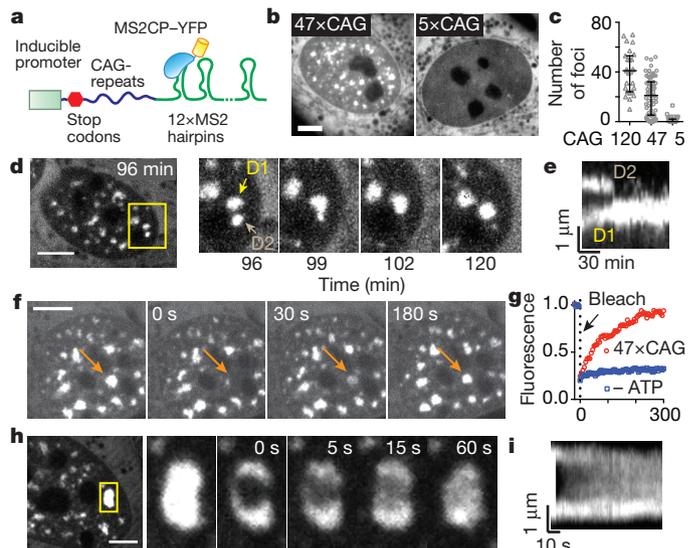


Figure 2 | CAG-repeat-containing RNAs coalesce into liquid-like nuclear foci. **a**, Schematic for RNA visualization. **b**, Micrographs of cells expressing 5×CAG or 47×CAG RNA. **c**, Quantification of foci as a function of CAG-repeat number. Each data point represents one cell. Data are median \pm interquartile range. **d**, Typical fusion event between RNA foci (time after induction indicated). **e**, Kymograph corresponding to **d**. **f**, Images for 47×CAG RNA foci before and after photobleaching (arrow, bleach site). **g**, FRAP trajectories for 47×CAG RNA punctum before (47×CAG) and after ATP depletion (–ATP). **h**, Similar to **f**, after partial photobleaching. **i**, Kymograph corresponding to **h**. Scale bars, 5 μ m. Data are representative of at least three independent experiments.

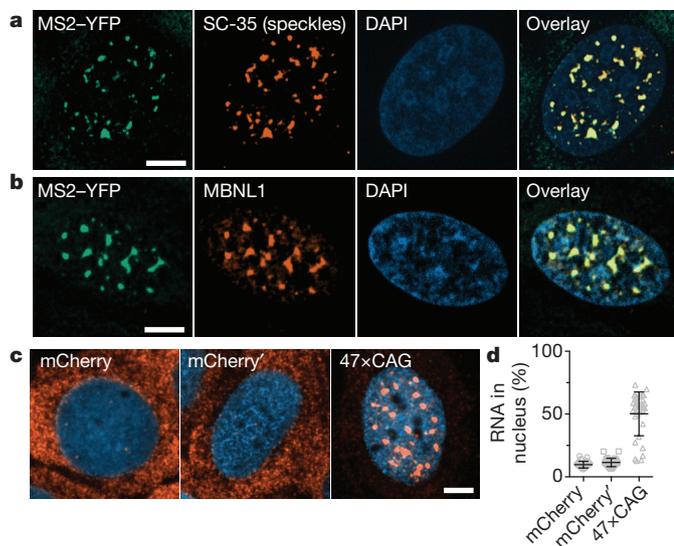


Figure 3 | Triplet repeat 47×CAG RNA is retained within the nucleus and co-localizes with nuclear speckles. **a**, Immunofluorescence images depicting co-localization of 47×CAG RNA foci (MS2-YFP) with nuclear speckles (SC-35). Nuclei are counterstained with 4',6-diamidino-2-phenylindole (DAPI). **b**, Similar to **a**, staining for MBNL1. **c**, **d**, FISH images (**c**) and relative RNA abundance per nucleus (**d**) in cells expressing MS2-tagged 47×CAG, a control coding sequence (mCherry), or its reverse complement (mCherry'). Scale bars, 5 μm. Data are mean ± s.d. Data are representative of at least three independent experiments.

and out of the foci (Fig. 2f, g and Supplementary Video 5). Upon photobleaching a portion of a 47×CAG RNA punctum, the fluorescence recovered rapidly ($\tau_{FRAP} = 18 \pm 5$ s, mean ± s.d., $n = 5$ foci) (Fig. 2h, i and Supplementary Video 6), suggesting that RNA within the foci can undergo internal rearrangement. Thus, unlike their solid-like behaviour *in vitro* (Fig. 1e), CAG RNA foci in cells display liquid-like properties²⁸. We hypothesized that the increased dynamicity might arise from specialized proteins (for example, helicases) in the nucleoplasm that remodel RNA base-pairing. Consistent with this hypothesis, depletion of cellular ATP substantially reduced fluorescence recovery of the RNA foci after photobleaching ($23 \pm 7\%$ recovery, mean ± s.d., $n = 7$; Fig. 2g and Supplementary Video 7).

Similar to the endogenous foci in patient-derived fibroblasts²⁹, the induced RNA foci co-localized with the SC-35 marker for nuclear speckles (Fig. 3a and Extended Data Fig. 5), non-membranous bodies that are enriched in pre-messenger RNA (pre-mRNA) splicing

factors³⁰. The foci also recruited endogenous muscleblind-like-1 (MBNL1) protein (Fig. 3b), sequestration of which has been implicated in CAG/CUG-repeat-containing RNA-mediated pathogenicity^{14,31}. Using fluorescence *in situ* hybridization (FISH)³², we found that ~50% of the 47×CAG RNA was retained in the nucleus, compared with <10% of control RNAs (Fig. 3c, d and Extended Data Fig. 3f–h). Thus, CAG repeats cause the RNA to be retained in the nucleus within liquid-like bodies that sequester splicing factors.

Inhibitors of RNA gelation disrupt foci

Next, we asked whether perturbations that prevent RNA gelation *in vitro* also affect the stability of RNA foci in cells. *In vitro*, RNA gelation is inhibited by monovalent cations (Fig. 4a and Extended Data Fig. 1g, h). To test the effect of monovalent cations in cells, we used ammonium acetate, which readily permeates into cells and does not perturb intracellular pH³³. Strikingly, addition of 100 mM ammonium acetate led to the disappearance of 47×CAG RNA foci within minutes (Fig. 4b, Extended Data Fig. 6a and Supplementary Video 8). Interestingly, nuclear speckles also were disrupted by ammonium acetate, suggesting that the assembly of these ribonucleoprotein bodies depends upon ionic interactions as well (Extended Data Fig. 7). Consistent with previous studies²⁹, we find that the disruption of nuclear speckles with tautomycin also disrupts the 47×CAG RNA foci (Extended Data Fig. 7).

Next, we tested agents that might specifically disrupt the base-pairing in RNA foci without dissolving nuclear speckles. Transfection of an 8×CTG ASO reduced the number and size of 47×CAG foci compared against control oligonucleotides (Fig. 4c and Extended Data Fig. 6b). ASO may disrupt cellular RNA foci either by inhibiting intermolecular base-pairing or by degrading RNA via the RNase H machinery³⁴. To specifically perturb base-pairing interactions, we used doxorubicin, a nucleic acid intercalator³⁵. Doxorubicin blocked the formation of CAG RNA gels *in vitro* and potently dissolved the 47×CAG nuclear foci in cells (Fig. 4d, e and Extended Data Fig. 6c, d) without disrupting nuclear speckles (Extended Data Fig. 7). The concentration of doxorubicin required to disrupt RNA foci in cells (2.5 μM) was lower than that needed for the *in vitro* experiments (~1 mM), potentially because of the aid of cellular proteins that unwind RNA base-pairing and facilitate doxorubicin intercalation. Next, we tested whether doxorubicin can alleviate RNA foci derived from an endogenous locus. To this end, we used fibroblasts derived from patients with myotonic dystrophy type 1 (DM1) with a CTG expansion in the *DMPK* gene, which exhibit RNA foci containing multiple *DMPK* transcripts¹² (Extended Data Fig. 8a–c). Treatment with doxorubicin (2 μM for 24 h) substantially reduced the number as well as total volume of foci per cell (~65% and ~85%

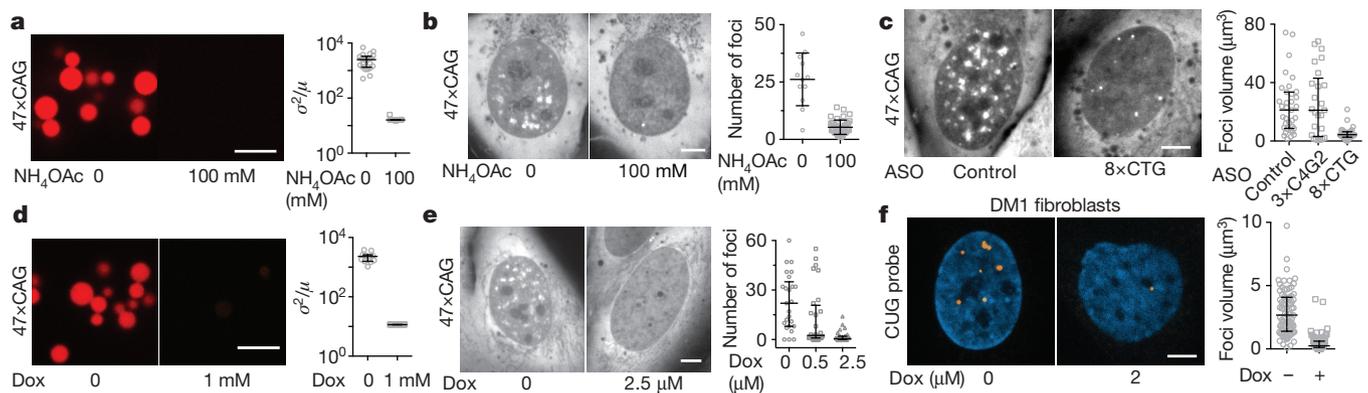


Figure 4 | Nuclear foci are reversed by agents that disrupt RNA gelation *in vitro*. **a**, **d**, NH_4OAc (**a**) and doxorubicin (Dox) (**d**) prevent 47×CAG RNA clustering. **b**, Images of the same cell before and 300 s after treatment with NH_4OAc , and corresponding quantification. **c**, **e**, **f**, Representative images and quantification of RNA foci upon transfection with indicated

ASO (**c**) and upon Dox treatment in U-2OS cells expressing exogenous 47×CAG RNA (**e**) and fibroblasts from patients with DM1 (**f**). Scale bars, 5 μm. Data are median ± interquartile range. Data are representative of at least three independent experiments.

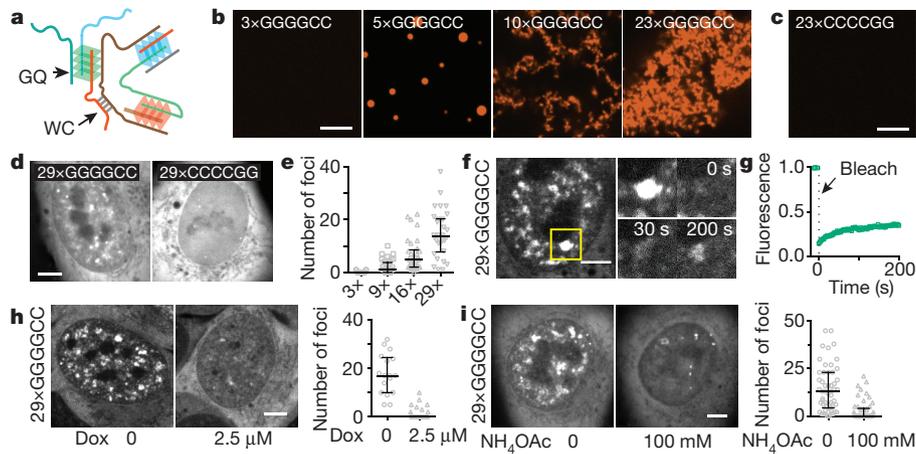


Figure 5 | RNA with ALS/FTD-associated GGGGCC repeats forms gels *in vitro* and in cells. **a**, Schematic for GGGGCC RNA multimerization. GQ, G-quadruplex; WC, Watson–Crick. **b**, RNA clusters at indicated number of GGGGCC repeats. **c**, Repeat 23×CCCCGG is soluble under similar conditions. **d**, Images of cells expressing 29×GGGGCC or 29×CCCCGG RNA. **e**, Quantification of RNA foci at indicated number of

decrease, respectively; Fig. 4f and Extended Data Fig. 8d). In summary, agents that inhibit gelation of purified repeat-containing RNAs *in vitro* also disrupt RNA foci in cells.

ALS/FTD-linked GGGGCC repeats form gels

Besides the canonical Watson–Crick base-pairing, nucleic acids can also form Hoogsteen base pairs such as in G-quadruplexes. The GGGGCC repeat in the *C9orf72* locus associated with ALS/FTD was found to form G-quadruplexes *in vitro* and *in vivo*^{15,36}. A single G-quadruplex can bring up to four RNA strands together, but a GGGGCC repeat expansion could potentially give rise to multimolecular RNA complexes³⁶ (Fig. 5a). Indeed, while the 3×GGGGCC RNA is largely soluble, we found that 5×GGGGCC RNA formed spherical clusters *in vitro* (Fig. 5b). Longer repeats (10× and 23×) formed an interconnected mesh-like network of aggregated RNA (Fig. 5b). In contrast, 23×CCCCGG RNA, which can form multivalent Watson–Crick base-pairing but not G-quadruplexes, was soluble (Fig. 5c). Similar to the CAG/CUG RNA, GGGGCC RNA clusters exhibited solid-like properties, and clustering was inhibited by monovalent cations (Extended Data Fig. 9a, b).

Cellular expression of 29×GGGGCC, but not 29×CCCCGG, RNA resulted in the formation of nuclear puncta in a dose-dependent manner (Fig. 5d and Extended Data Fig. 9c, d). The number of foci per cell increased with the number of repeats (Fig. 5e and Extended Data Fig. 9e). The threshold number of repeats for disease onset (>23×GGGGCC⁴) is similar to the repeat length in which most cells exhibit foci (16–29×GGGGCC). This GGGGCC repeat number is higher than that required for RNA gelation *in vitro*, possibly because of cellular proteins that may unfold G-quadruplexes³⁷. Interestingly, the 29×GGGGCC RNA foci, as well as those of shorter length, exhibited incomplete FRAP recovery ($37 \pm 20\%$ recovery, mean \pm s.d., $n = 10$; Fig. 5f, g, Extended Data Fig. 9f, g and Supplementary Video 9), indicating that they are less dynamic than CAG RNA foci. This result suggests a stronger intermolecular interaction between the GGGGCC repeats, which is consistent with intracellular G-quadruplex formation (Fig. 5a).

The GGGGCC RNA foci recruited hnRNP H, as previously shown¹⁵, as well as MBNL1, and co-localized with nuclear speckles (Extended Data Fig. 10). Like the 47×CAG RNA, most GGGGCC RNA was retained in the nucleus ($\sim 60\%$; Extended Data Fig. 9h). The *c9orf72* GGGGCC expansion is located in an intron with ~ 150 bases upstream and ~ 6 kb region downstream of the repeats. We found that the incorporation of the endogenous (~ 150 bases) upstream or a long (~ 1.7 kb) downstream flanking sequence did not affect the formation of RNA

GGGGCC repeats. **f**, GGGGCC RNA punctum at indicated time points after photobleaching. **g**, Recovery plot corresponding to **f**. **h**, **i**, Doxorubicin (**h**) and NH_4OAc (**i**) disrupt 29×GGGGCC RNA foci. Images in **i** of the same cell before and 300 s after treatment. Scale bars, 5 μm . Data are median \pm interquartile range. Data are representative of at least three independent experiments.

foci (Extended Data Fig. 9i). Intriguingly, incorporation of a longer sequence (~ 1 kb) upstream of the repeats abolished the formation of nuclear puncta (Extended Data Fig. 9i), suggesting that sequences flanking the repeats may influence their assembly into RNA foci. Similar to the CAG RNA foci, the GGGGCC RNA foci were disrupted by antisense oligonucleotides, doxorubicin, or ammonium acetate (Fig. 5h, i and Extended Data Fig. 9 j–l), indicating that both base-pairing and electrostatic interactions are essential for GGGGCC RNA foci formation.

Conclusion

In summary, we have shown that the propensity of an RNA to form multivalent base-pairing can lead to its gelation without requiring protein components. An emerging body of research has shown that RNP granules, such as stress granules, P-granules, and nucleoli, are phase-separated liquid-like compartments³⁸. Numerous recent studies have characterized proteins for their ability to phase-separate and mediate the assembly of these RNP granules³⁸. Our results demonstrate that sequence-specific base-pairing properties of RNAs can lead

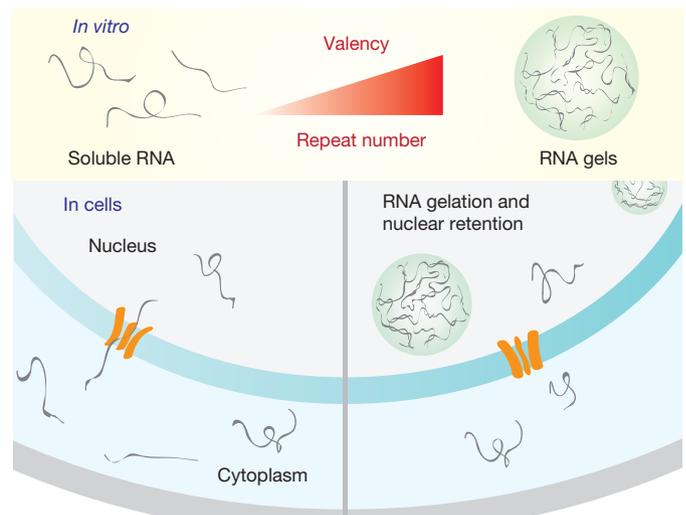


Figure 6 | Model for RNA foci formation in repeat expansion disorders. The repeat expansion sequences form templates for multivalent intermolecular base-pairing, which leads to the formation of RNA foci and the retention of the RNA in the nucleus. See text for details.

to their phase separation and gelation, and raise the possibility that such phenomena could contribute to physiological granule assembly as well.

In the case of repeat expansions diseases, our data suggest that intermolecular base-pairing can result in the aggregation and sequestration of RNA into nuclear foci (Fig. 6). RNA gelation, which occurs at a boundary condition of increasing valency, might explain why disease appears to be triggered after an expansion of a nucleotide repeat has reached a threshold number (Fig. 6), a puzzling phenomenon that has lacked a mechanistic explanation thus far^{2,6}. Our results may also explain why placement of distinct repeat expansions in seemingly unrelated genes can result in similar clinical syndromes^{39,40}. RNA gelation and foci formation probably occur upstream of known drivers of cellular toxicity such as sequestration of RNA binding proteins¹⁴, nucleolar stress⁴¹, and disruption of nucleocytoplasmic transport^{42,43}, although a direct link between foci formation and cellular toxicity remains to be established. Our findings raise the possibility that new strategies for disrupting RNA–RNA base-pairing could be effective for treating repeat expansion diseases.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 January; accepted 24 April 2017.

Published online 31 May 2017.

- Gatchel, J. R. & Zoghbi, H. Y. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* **6**, 743–755 (2005).
- La Spada, A. R. & Taylor, J. P. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* **11**, 247–258 (2010).
- Krzyzosiak, W. J. *et al.* Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* **40**, 11–26 (2012).
- 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).
- Lee, D.-Y. & McMurray, C. T. Trinucleotide expansion in disease: why is there a length threshold? *Curr. Opin. Genet. Dev.* **26**, 131–140 (2014).
- Chew, J. *et al.* Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* **348**, 1151–1154 (2015).
- Mankodi, A. *et al.* Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science* **289**, 1769–1773 (2000).
- Jin, P. *et al.* RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in *Drosophila*. *Neuron* **39**, 739–747 (2003).
- Zu, T. *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl Acad. Sci. USA* **108**, 260–265 (2011).
- Cleary, J. D. & Ranum, L. P. W. Repeat-associated non-ATG (RAN) translation in neurological disease. *Hum. Mol. Genet.* **22** (R1), R45–R51 (2013).
- Taneja, K. L., McCurrach, M., Schalling, M., Housman, D. & Singer, R. H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.* **128**, 995–1002 (1995).
- Wojciechowska, M. & Krzyzosiak, W. J. Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum. Mol. Genet.* **20**, 3811–3821 (2011).
- Miller, J. W. *et al.* Recruitment of human muscleblind proteins to (CUG)_n expansions associated with myotonic dystrophy. *EMBO J.* **19**, 4439–4448 (2000).
- 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**, 345 (2016).
- Timchenko, N. A. *et al.* RNA CUG repeats sequester CUGBP1 and alter protein levels and activity of CUGBP1. *J. Biol. Chem.* **276**, 7820–7826 (2001).
- Todd, P. K. & Paulson, H. L. RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.* **67**, 291–300 (2010).
- Belzil, V. V., Gendron, T. F. & Petrucelli, L. RNA-mediated toxicity in neurodegenerative disease. *Mol. Cell. Neurosci.* **56**, 406–419 (2013).
- Flory, P. J. *Principles of Polymer Chemistry* (Cornell Univ. Press, 1953).
- Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
- Stockmayer, W. H. Theory of molecular size distribution and gel formation in branched polymers II. General cross linking. *J. Chem. Phys.* **12**, 125–131 (1944).
- Semenov, A. N. & Rubinstein, M. Thermoreversible gelation in solutions of associative polymers. 1. *Statics. Macromolecules* **31**, 1373–1385 (1998).
- Falkenberg, C. V., Blinov, M. L. & Loew, L. M. Pleomorphic ensembles: formation of large clusters composed of weakly interacting multivalent molecules. *Biophys. J.* **105**, 2451–2460 (2013).
- Aumiller, W. M. Jr & Keating, C. D. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat. Chem.* **8**, 129–137 (2016).
- Larson, R. G. *The Structure and Rheology of Complex Fluids* (Oxford Univ. Press, 1999).
- Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).
- Kearse, M. G. *et al.* CGG repeat-associated non-AUG translation utilizes a cap-dependent scanning mechanism of initiation to produce toxic proteins. *Mol. Cell* **62**, 314–322 (2016).
- Brangwynne, C. P. *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
- Urbaneck, M. O., Jazurek, M., Switonski, P. M., Figura, G. & Krzyzosiak, W. J. Nuclear speckles are detention centers for transcripts containing expanded CAG repeats. *Biochim. Biophys. Acta* **1862**, 1513–1520 (2016).
- Spector, D. L. & Lamond, A. I. Nuclear speckles. *Cold Spring Harb. Perspect. Biol.* **3**, a000646 (2011).
- Li, L.-B., Yu, Z., Teng, X. & Bonini, N. M. RNA toxicity is a component of ataxin-3 degeneration in *Drosophila*. *Nature* **453**, 1107–1111 (2008).
- Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
- Hamaguchi, M. S., Watanabe, K. & Hamaguchi, Y. Regulation of intracellular pH in sea urchin eggs by medium containing both weak acid and base. *Cell Struct. Funct.* **22**, 387–398 (1997).
- Bennett, C. F. & Swayze, E. E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* **50**, 259–293 (2010).
- Frederick, C. A. *et al.* Structural comparison of anticancer drug–DNA complexes: adriamycin and daunomycin. *Biochemistry* **29**, 2538–2549 (1990).
- Reddy, K., Zamiri, B., Stanley, S. Y. R., Macgregor, R. B. Jr & Pearson, C. E. The disease-associated r(GGGGCC)_n repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.* **288**, 9860–9866 (2013).
- Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science* **353**, aaf5371 (2016).
- Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid–liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
- Holmes, S. E. *et al.* A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat. Genet.* **29**, 377–378 (2001).
- Elden, A. C. *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* **466**, 1069–1075 (2010).
- Tsoi, H., Lau, T. C.-K., Tsang, S.-Y., Lau, K.-F. & Chan, H. Y. E. CAG expansion induces nucleolar stress in polyglutamine diseases. *Proc. Natl Acad. Sci. USA* **109**, 13428–13433 (2012).
- Freibaum, B. D. *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129–133 (2015).
- Zhang, K. *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56–61 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. K. Rosen, W. W. Seeley, and the members of the Vale laboratory for discussions. A.J. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2181-14).

Author Contributions A.J. and R.V. designed research and wrote the paper. A.J. performed experiments and analysed data.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.V. (ron.vale@ucsf.edu).

Reviewer Information *Nature* thanks C. P. Brangwynne, M. Carmo-Fonseca, J. Shorter and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cloning. CAG and GGGGCC repeats were cloned via sequential repeat directed elongation as previously described⁴⁴ in a modified pBluescript vector. Inserts were verified via sequencing from both ends (up to 700 bp read length for CAG/CTG repeats and up to 200 bp read length for GGGGCC repeats) and by verifying the insert size by restriction digestion. All cloning and amplification were performed in *Escherichia coli* Stbl3 cells (Invitrogen) grown at 30 °C. For synthesizing the mammalian expression constructs, repeats were cut directly from the cloning plasmids and ligated at the compatible restriction sites in a modified lentiviral expression vector with tetracycline-inducible expression promoter. Plasmid sequences are provided in Supplementary Table 1. We observed that the purified plasmids formed higher-order complexes when stored for prolonged periods (>1 month) at 4 °C or –20 °C. Stored plasmids when re-transformed in bacteria often resulted in substantial truncations in the repeat region. To avoid such retransformation-associated repeat truncations, we maintained a bacterial stock of each plasmid. The plasmid DNA was freshly purified for each cloning/transfection or RNA transcription experiment, and the sequence was verified as described above.

RNA transcription and gelation. Repeat-containing RNA was transcribed using a T7 or T3 MegaScript kit (Ambion) according to the manufacturer's recommendation. Template DNA up to 200 bases long (for 10×CAG, 20×CAG, 10×CUG, 20×CUG, 3×GGGGCC, 5×GGGGCC, and 23×GGGGCC) were purchased from Integrated DNA Technologies as single-stranded DNA oligonucleotides. Complementary strand was synthesized by using a single complementary primer and a standard polymerase chain reaction kit (Advantage GC 2 PCR Kit, Clontech). The double-stranded DNA thus generated was gel purified and used as a template for transcription reactions. Longer templates were obtained by either PCR amplification from plasmids (31×CAG, 47×CAG, 31×CUG, 47×CUG) or by restriction digestion of the repeat-containing vectors (66×CAG, 66×CUG) upstream and downstream of the repeat region. Template sequences for transcription reactions are provided in Supplementary Table 2. For synthesis of fluorescently labelled RNA, transcription reactions were doped with Cy3-UTP or Cy5-UTP (Enzo Lifesciences). Free nucleoside 5'-triphosphates were removed using lithium chloride precipitation or an RNA purification kit (Zymo Research). Similar results were obtained in both purification schemes. The size of the RNA products was verified using denaturing agarose gel electrophoresis. RNAs were resuspended in water, and either used immediately or aliquoted and flash-frozen in liquid nitrogen and stored at –80 °C for up to 1 month.

For phase separation/gelation assays, RNAs were diluted to concentrations of 0.5 ng μl^{-1} to 0.5 $\mu\text{g} \mu\text{l}^{-1}$ in 10 mM Tris pH 7.0, 10 mM MgCl_2 , 25 mM NaCl buffer, unless indicated otherwise. Nuclease-free buffer stocks were purchased from Ambion. RNA was denatured at 95 °C for 3 min and cooled down at 1–4 °C per min to 37 °C final temperature in a thermocycler and imaged immediately. Samples were visualized using a custom spinning disk confocal microscope (Nikon Ti-Eclipse equipped with a Yokogawa CSU-X spinning disk module) using a $\times 100$, 1.49 numerical aperture oil immersion objective and an air-cooled EM-CCD (electron multiplying charge-coupled device). The extent of phase separation/gelation was quantified by the index of dispersion (σ^2/μ) of fluorescence intensity per pixel (pixel size 83 nm \times 83 nm). In brief, variance in the fluorescence intensity per image was determined, and normalized to the mean fluorescence intensity in the solution phase of the RNA. For dilute solution (<10% of imaging area occupied with clusters), this parameter reports the extent of inhomogeneity in the sample. At least 20 independent imaging areas ($\sim 1,800 \mu\text{m}^2$ each) were analysed for each condition to achieve a representative measure across the sample. Each data point in the bar graphs represents one imaging area. Data shown are representative of three or more independent replicates, across two or more independent RNA preparations.

For antisense DNA-mediated repression of RNA phase separation, 47×CAG RNA (200 ng μl^{-1} or 2.4 μM) was incubated with the ASO at 20 μM final concentration, followed by heat denaturation and annealing as described above. Doxorubicin was purchased from Cell Signaling Technology (catalogue number 5927). For *in vitro* experiments, doxorubicin was added to pre-formed RNA clusters, and samples were incubated at 37 °C for 1 h. Alternatively, doxorubicin and RNA were pre-mixed at indicated concentrations before annealing. Similar results were obtained in both cases.

For FRAP experiments, RNA clusters were prepared as described above. RNA clusters were allowed to settle on to the glass surface for ~ 15 min. A region of $\sim 1 \mu\text{m}^2$ was photobleached using a 405 nm laser modulated by a Rapp UGA-40 photo targeting unit and the fluorescence recovery was monitored over time. The

fluorescence recovery was fitted to the equation $I = A - I_0 \exp(-t/\tau_{\text{FRAP}})$, and the time constant, τ_{FRAP} was determined.

DNA phase separation. DNA oligonucleotides were purchased from Integrated DNA Technologies. Spermine hydrochloride (Sigma) was resuspended in water and pH was adjusted to 7.5. DNA was heat denatured at 90 °C for 2 min to melt secondary structure, incubated on ice for 2 min, and used immediately for phase separation assays. Phase separation was triggered by adding spermine to the DNA solution. All phase separation assays were performed in 10 mM Tris pH 7.0 buffer with the indicated amounts of DNA and salts. DNA clusters were visualized using standard bright-field or confocal microscopy as described above. To prevent DNA droplets from fusing onto the glass surface, coverslips were passivated with polyethylene glycol as previously described⁴⁵. FRAP experiments and analysis were performed as described above.

Cell culture and imaging. U-2OS cells, authenticated by STR profiling, were purchased from the University of California, San Francisco, Cell Culture Facility. We generated a monoclonal U-2OS cell line stably expressing Tet-On 3G trans-activator protein (Clontech) and a tandem-dimeric MS2 hairpin binding protein tagged with enhanced YFP (MS2CP-YFP) via sequential lentiviral infection and selection. This stable cell line was transduced with repeat-containing plasmids under tetracycline-inducible promoter. Cells were maintained in DMEM with 10% (v/v) tetracycline-free fetal bovine serum (Clontech) and 1 \times penicillin-streptomycin-glutamine cocktail (Gibco). Cell lines were tested for mycoplasma contamination using a standard PCR kit (LookOut Mycoplasma PCR Detection Kit, Sigma) and verified routinely by live-cell DNA staining.

RNA expression was induced by adding 1,000 ng ml^{-1} doxycycline for 12–48 h, or as indicated. Before imaging, the culture medium was replaced with DMEM with 25 mM HEPES pH 7.5 or FluoroBrite DMEM (Invitrogen) with serum and antibiotics as listed above. For long-term imaging (>2h), cells were placed in a live-cell imaging chamber supplemented with 5% CO_2 . Cells were imaged using a spinning disk confocal microscope (Nikon Ti-Eclipse equipped with a Yokogawa CSU-X spinning disk module) using a $\times 100$, 1.49 numerical aperture oil immersion objective and an air-cooled EM-CCD. For each experimental condition, at least 30 randomly chosen cells were imaged and analysed. Each data point in the bar graphs represents one cell. Data shown are representative of three or more independent replicates.

ATP depletion was achieved by rinsing cells twice in DMEM without glucose (Gibco), followed by incubation for 10 min in the ATP depletion medium (DMEM without glucose with 1% (v/v) dialysed FBS (Gibco), 10 mM sodium azide and 6 mM 2-deoxy-D-glucose). Doxorubicin (stock, 10 mM in dimethylsulfoxide (DMSO)) was diluted to the desired concentration in cell culture medium and added to cells pre-induced with doxycycline for 24 h. Cells were incubated with doxorubicin or an equivalent dilution of DMSO only as control, for 2 h, and imaged as described above. Ammonium acetate (stock, 5 N) was diluted to 200 mM in cell culture medium. This intermediate dilution (2 \times) was added to cells pre-induced with doxycycline for 24 h to achieve a final concentration of 100 mM of ammonium acetate. Cells were incubated in this medium for 10 min at 37 °C. Normal cell culture medium was replaced after treatment, and cells were imaged immediately or 1 h after medium replacement. For treatment with ASO, cells pre-induced with doxycycline for 48 h were transfected with 100 nM final concentration of ASO using either Lipofectamine RNAiMAX (Invitrogen) or TransIT-Oligo Transfection Reagent (Mirus Bio) according to the manufacturers' recommended protocols. Similar results were obtained with both transfection reagents. Cells were imaged 12 h after transfection. ASO sequences are provided in Supplementary Table 2.

Analysis of RNA foci. We used a fluorescence-intensity and size-based threshold to identify RNA foci. In brief, U-2OS cells expressing the RNA of interest together with MS2CP-YFP were imaged using a spinning disk confocal microscope, and 0.3 μm Z-stacks were acquired. To account for variability in MS2CP-YFP expression levels, we used a cell-intrinsic intensity threshold for foci identification. We manually segmented the nuclei, and determined the mean YFP fluorescence intensity in the nucleus. RNA foci were identified using the FIJI 3D Objects Counter plugin, with an intensity threshold as 1.6 \times the mean fluorescence intensity in the nucleus of the cell, and a size cut-off of more than 50 adjoining pixels (pixel size, 83 nm \times 83 nm). This algorithm accurately identified the foci, as depicted in Extended Data Fig. 4. This method was used to determine the number, volume, surface area, and the fluorescence intensity of the foci. Various metrics such as total number of foci per cell, total volume of foci per cell, coefficient of dispersion (σ^2/μ), and integrated intensity of foci were compared and yielded similar results. The number of foci per cell, and the total volume occupied by the foci per cell, were chosen as the parameters of choice to quantify the extent of foci formation. Statistical significance was analysed using unpaired, two-tailed Mann-Whitney *U*-tests. For this analysis, the numbers of foci per cell in each experiment were assumed to be symmetrically distributed about the median.

Quantification of RNA copy number. To determine the copy number of RNA in cells, we used two alternative approaches. First, we used NanoString, a proprietary PCR-free RNA quantitation platform, and determined that, under our highest induction conditions, the copy number of 47×CAG RNA is about ten times that of GAPDH or β-actin RNA, or about 8,800 ± 1,500 copies per cell ($n = 3$ independent experiments). Second, we used single-molecule FISH to obtain quantitative RNA localization information. We designed fluorescent probes against the MS2 hairpin loop region, such that the 12×MS2 tag could accommodate a maximum of 32 fluorescently labelled probes (probe sequences in Supplementary Table 2). For cells expressing low levels of MS2-tagged control RNAs such as mCherry or 5×CAG RNA, we observed isolated fluorescent spots that exhibited a uniform distribution of intensities, probably arising from single RNA molecules. Similarly, in the cytoplasm of cells expressing 47×CAG or 29×GGGGCC RNA, we observed isolated RNA spots with a similar uniform distribution of fluorescence intensities (Extended Data Fig. 3f, g). We ascribed this intensity value as corresponding to that of a single RNA. Fluorescence spots and corresponding intensities were quantified using the ImageJ Spot Counter plugin. The approximate RNA copy number in each cell was then calculated by dividing the total fluorescence intensity of the cell by the fluorescence intensity of a single RNA, after background subtraction. By this method, we determined that, under maximal induction conditions leading to RNA foci formation, the copy number of 47×CAG was 13,000 ± 7,000 copies per cell (mean ± s.d., $n = 24$ cells), for 29×GGGGCC RNA 2,500 ± 1,800 copies per cell ($n = 30$ cells), and for the control cells expressing mCherry was 20,000 ± 7,000 copies per cell ($n = 21$ cells). The fraction of RNA retained in the nucleus was determined by dividing the fluorescence intensity in the nucleus of the cell by the total fluorescence of the cell.

FRAP experiments and data analysis. To assess the dynamicity of RNA foci, we performed FRAP experiments by bleaching MS2CP–YFP protein. Previous studies have shown that the dimeric MS2CP–YFP is attached with high affinity to the MS2 hairpin sequence and does not dissociate during the observation timescales of a few minutes⁴⁶, and that the fluorescence recovery of MS2CP–YFP can be used to report on the RNA dynamics. To monitor exchange of RNA between foci and the nucleoplasm, an entire punctum, typically a few micrometres in size, was photobleached and the fluorescence recovery was monitored by time-lapse imaging. To examine internal turnover, relatively large puncta were manually selected and a region ~1 μm in diameter was photobleached. The fluorescence intensity of the bleached region was normalized and corrected for photobleaching using previously described methods⁴⁷. To determine fluorescence relaxation time, the recovery curves were fitted to the equation $I = A - I_0 \exp(-t/\tau_{\text{FRAP}})$, where A and I_0 are also fit parameters.

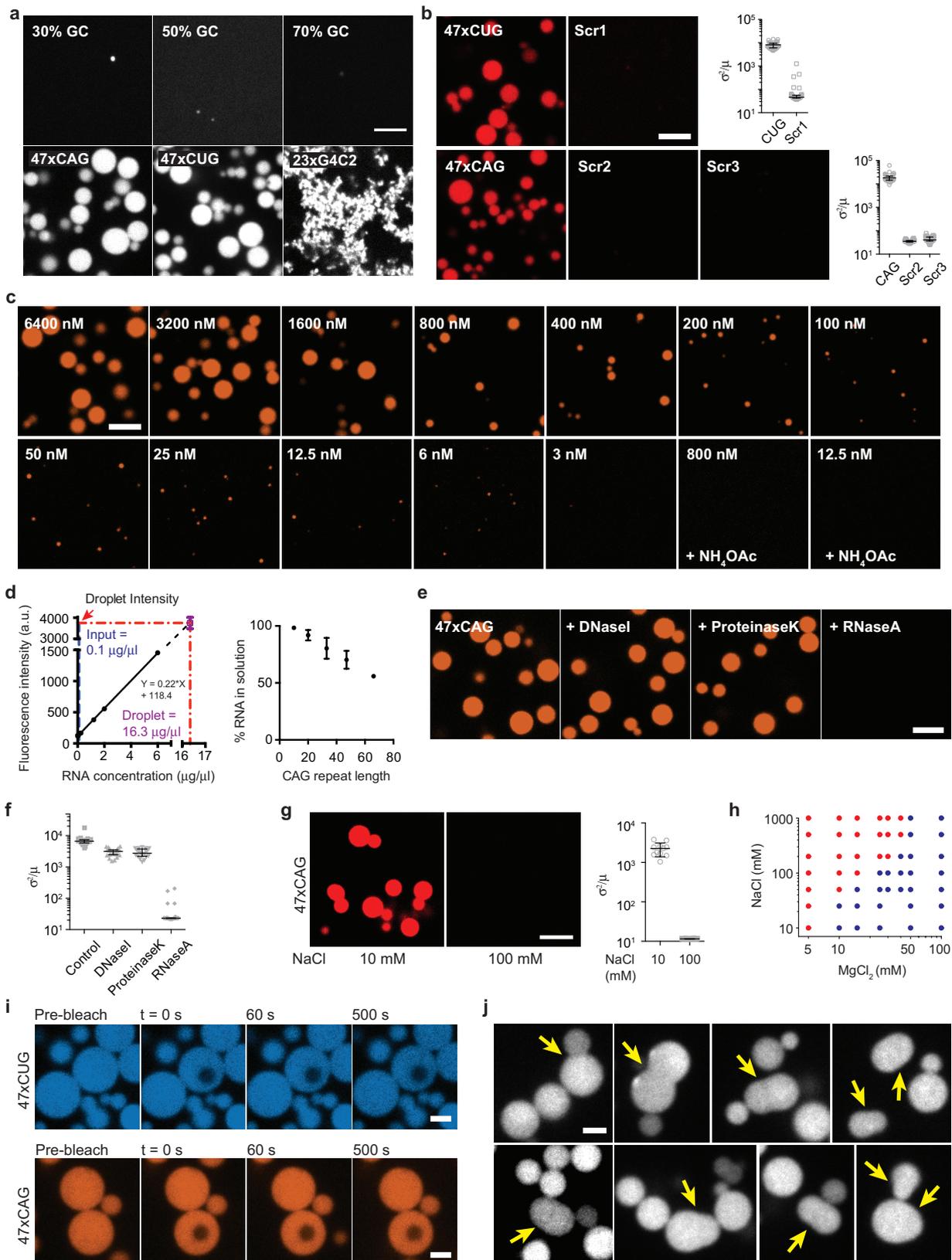
RNA FISH and immunofluorescence. For RNA FISH in U-2OS cells, cells expressing the desired RNA (induced for 24 h) were fixed with 2% paraformaldehyde

for 10 min at room temperature and permeabilized by overnight incubation in 70% ethanol at 4 °C. Alternatively, cells were fixed and permeabilized by incubation for 10 min in methanol with 10% (v/v) acetic acid. Similar results were obtained with both fixation protocols. Fixed and permeabilized cells were either used immediately, or stored in the permeabilization medium at –20 °C until needed. RNA was detected using Cy3-labelled DNA oligonucleotides designed against the MS2-hairpin sequence. Probe sequences are provided in Supplementary Table 2. Hybridization and wash buffers were purchased from Biosearch Technologies and used according to the manufacturer's protocol. For immunofluorescence detection of proteins, methanol-fixed cells were stained using antibodies against muscleblind-like-1 (MBNL1, Abcam, ab45899), hnRNP H (Abcam, ab10374), SC-35 (Abcam, ab11826), coilin (Abcam, ab87913), fibrillarin (Abcam, ab5821), PML (Abcam, ab179466), and a corresponding Alexa Fluor 647-labelled secondary antibody (Invitrogen A-21236 or Invitrogen A-21244). Samples were co-stained with an anti-green fluorescent protein (GFP) booster antibody (GBA488, Bulldog Bio) to visualize RNA foci. After labelling, samples were mounted in Prolong Gold antifade medium (Thermo Scientific) and imaged using confocal microscopy as described above.

Fibroblasts from patients with DM1. Fibroblasts from patients with DM1 were obtained from the Coriell Institute (catalogue numbers GM03132 and GM03987). Control fibroblasts (Hs27) were obtained from the University of California, San Francisco, Cell Culture Facility. These cell lines were used without further validation. Cells were maintained in DMEM with 10% (v/v) fetal bovine serum (Clontech) and 1× penicillin–streptomycin–glutamine cocktail (Gibco). To detect RNA foci, RNA FISH was performed as described above using an 8×CAG oligonucleotide labelled with Atto647N (sequence in Supplementary Table 2) or using a pool of 48 oligonucleotide probes designed against the wild-type *DMPK* allele obtained as a pooled library from Biosearch Technologies. To disrupt RNA foci, cells were incubated for 24 h with 2 μM doxorubicin or an equivalent dose of DMSO-only control. Total volume and the number of RNA foci were quantified using the ImageJ 3D Objects Counter plugin, with an empirically determined fluorescence threshold.

Data availability. The raw data that support the findings of this study are available from the corresponding author upon reasonable request.

44. Scior, A., Preissler, S., Koch, M. & Deuerling, E. Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnol.* **11**, 87 (2011).
45. Jain, A. *et al.* Probing cellular protein complexes using single-molecule pull-down. *Nature* **473**, 484–488 (2011).
46. Shav-Tal, Y. *et al.* Dynamics of single mRNPs in nuclei of living cells. *Science* **304**, 1797–1800 (2004).
47. Phair, R. D., Gorski, S. A. & Misteli, T. Measurement of dynamic protein binding to chromatin *in vivo*, using photobleaching microscopy. *Methods Enzymol.* **375**, 393–414 (2004).

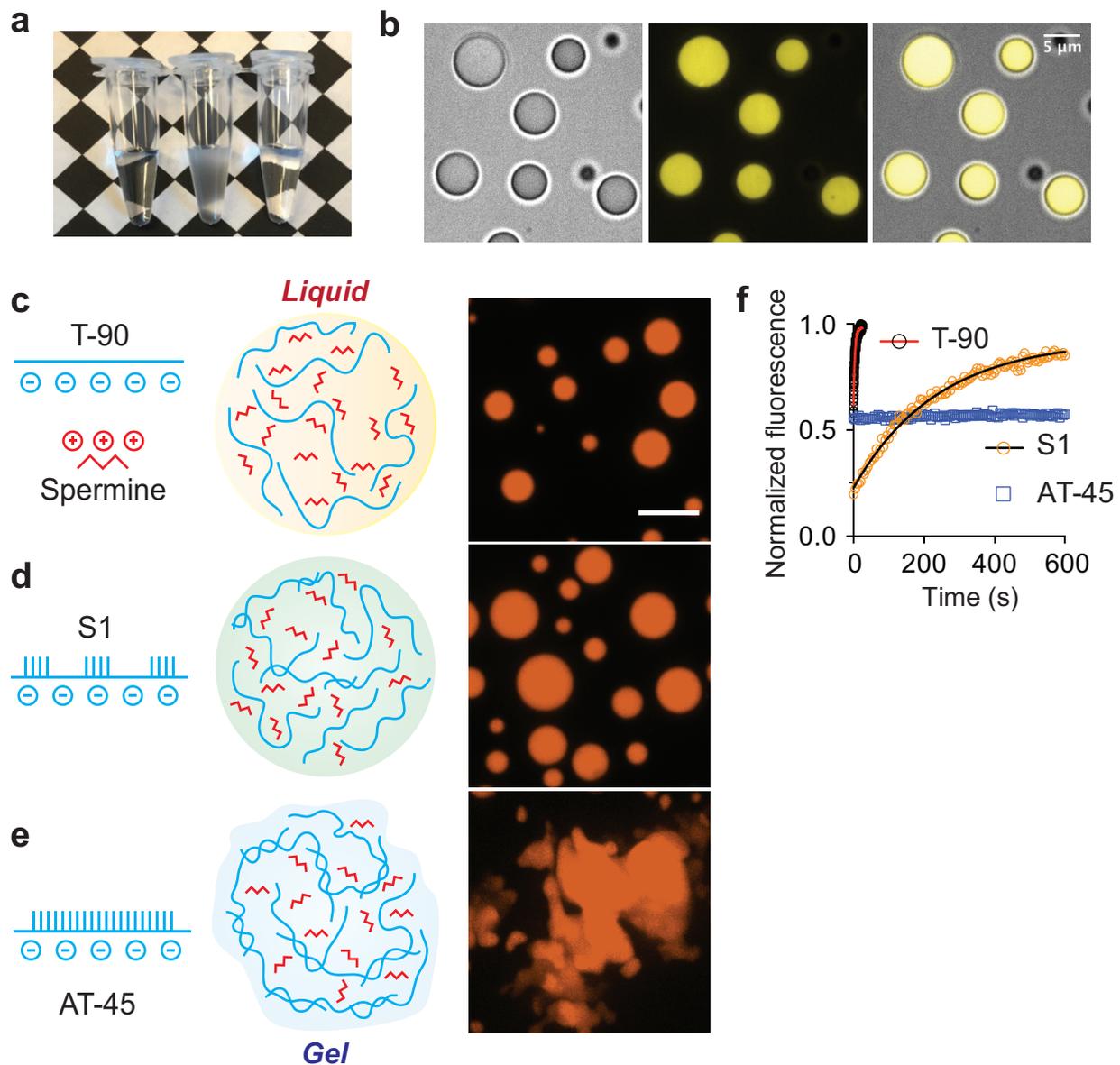


Extended Data Figure 1 | See next page for caption.

Extended Data Figure 1 | Disease-associated repeat-containing RNAs form clusters *in vitro*.

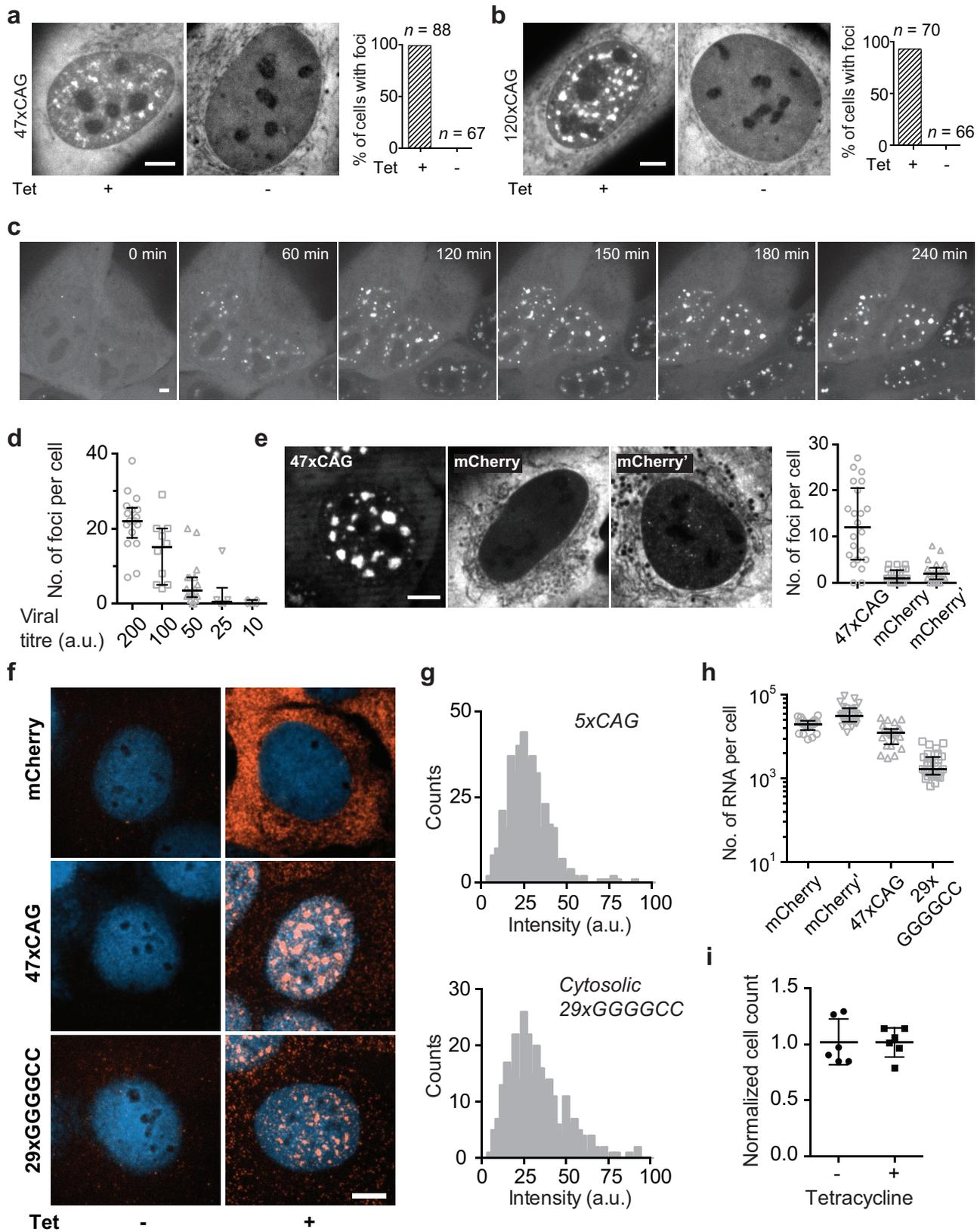
a, Fluorescence micrographs for RNA with various GC content compared against RNA with disease-associated repeat expansions. Sequences of the DNA templates used for transcription are provided in Supplementary Table 2. **b**, Fluorescence micrographs comparing 47×CUG RNA and a corresponding control RNA (Scr1) with identical base composition but with a scrambled sequence. Similarly, 47×CAG RNA was compared with two different RNAs (Scr2, Scr3) having the same base composition as 47×CAG but whose sequences were scrambled. The extent of inhomogeneity was quantified by the index of dispersion (σ^2/μ) across more than 20 independent imaging areas (1,800 μm^2 each). Each data point represents an independent imaging area. **c**, Representative micrographs of 47×CAG RNA clusters at indicated concentrations. Spherical RNA clusters are observable up to 25 nM RNA concentration. In this concentration regime, the reaction is reactant limited and the cluster size is below the diffraction limit. RNA clustering at all concentrations was not observed in the presence of 100 mM ammonium acetate. Representative images at the indicated RNA concentration in the presence of 100 mM NH_4OAc (+ NH_4OAc). **d**, RNA enrichment in the clusters. Left, Cy3-labelled 66×CAG RNA was serially diluted in conditions preventing RNA clustering (10 mM Tris pH 7.0, 10 mM MgCl_2 , 100 mM NaCl), and the bulk solution fluorescence was calibrated against the RNA concentration. The enrichment of RNA in the clusters was determined by comparing their fluorescence intensity against this calibration. When the input RNA concentration was

100 $\text{ng } \mu\text{l}^{-1}$, the concentration in the clusters corresponded to $\sim 16.3 \mu\text{g } \mu\text{l}^{-1}$, or an enrichment of 163-fold. a.u., arbitrary units. Right, RNA clusters were precipitated by centrifugation at 16,000g for 10 min at room temperature. The concentration of the soluble RNA after centrifugation was determined by measuring absorbance at 260 nm. The concentration of the RNA in the solution phase decreased with the increasing CAG-repeat number. **e, f**, The 47×CAG RNA clusters were treated with proteinase K (60 U ml^{-1}), DNase I (200 U ml^{-1}), or RNaseA (0.7 U ml^{-1}) for 10 min at room temperature. Representative micrographs (**e**) and quantification (**f**). **g**, Clustering of 47×CAG RNA is inhibited by NaCl. **h**, Binary phase diagram for 1.25 μM 47×CAG RNA as a function of MgCl_2 and NaCl concentrations. Blue dots represent two-phase regime while the red dots indicate a homogenous single-phase regime. **i**, The RNA clusters were in a solid-like state and did not exhibit fluorescence recovery upon photobleaching, as indicated by the representative micrographs for 47×CUG (top) and 47×CAG (bottom) RNA at the indicated time points. **j**, Sample images showing aborted fusion events between 47×CAG RNA clusters, suggesting that the clusters were liquid-like initially and later underwent a liquid-to-solid transition. Fusion events were probably aborted as the clusters solidified before relaxation to a spherical geometry. Sites of aborted fusion are marked by arrows. Scale bars in **a–c, e, g**, 5 μm , in **i, j**, 1 μm . Data are median \pm interquartile range. Data are representative of at least three independent experiments across at least two independent RNA preparations.



Extended Data Figure 2 | Base-pairing interactions impart solid-like properties to DNA-spermine complexes. Electrostatic interactions between polymeric anions (such as nucleic acids) and multivalent cations can lead to phase separation via formation of polyelectrolyte complexes, a phenomenon known as complex coacervation. We found that spermine, a tetravalent cation at pH 7, could induce phase separation of single-stranded DNA oligonucleotides. **a**, Mixing 10 mM spermine pH 7 (left tube) with 10 μ M T-90 DNA (90-mer polyT DNA oligonucleotide) (right tube) immediately resulted in a turbid solution (centre tube). **b**, Examination by bright-field microscopy revealed numerous spherical droplets. Using a fluorescently labelled T-90, we confirmed that the droplet phase was enriched in DNA. Representative bright-field (left), fluorescence (centre), and overlay (right) images for the DNA-spermine complexes. **c-f**, We investigated the effect of base-pairing interactions on DNA-spermine complexes. The T-90 DNA-spermine complexes were

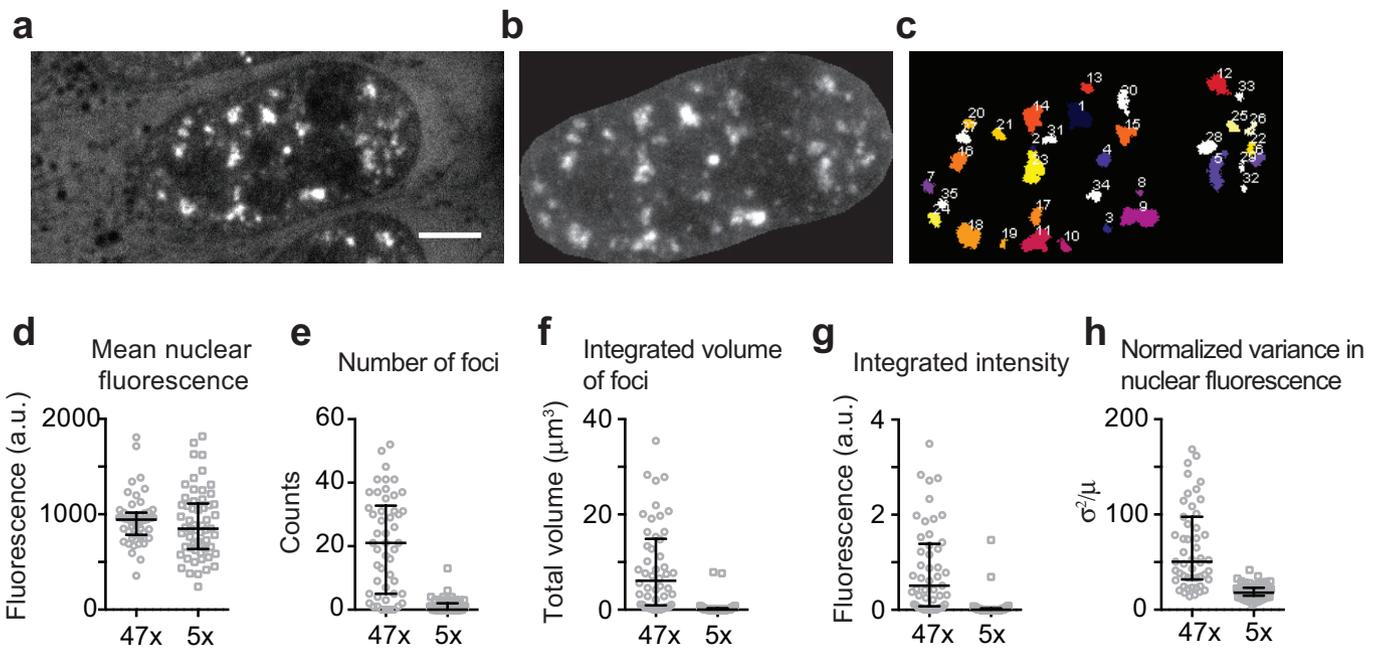
liquid-like, as evidenced by their spherical geometry (**c**) and a rapid FRAP (**f**, $99 \pm 1\%$ recovery, $\tau_{\text{FRAP-T90}} = 5 \pm 2$ s, mean \pm s.d., $n = 3$ droplets). Next, we designed a 90-base-long DNA with five 8-bp palindromic hybridization sites separated by poly-dT spacers (sequence S1, sequences in Supplementary Table 2). S1 DNA also phase-separated and formed spherical liquid-like droplets in the presence of spermine (**d**). However, the S1-spermine droplets exhibited reduced fluidity, as evidenced by a slower recovery upon photobleaching (**f**, $90 \pm 4\%$ recovery, $\tau_{\text{FRAP-S1}} = 335 \pm 41$ s, mean \pm s.d., $n = 5$). **e**, We performed similar spermine-mediated coacervation experiments with a (dAdT)₄₅ oligonucleotide (AT-45) that could form multivalent A:T base-pairing interactions. AT-45 DNA formed interconnected network-like structures spanning hundreds of micrometres or gels (**e**). These AT-45 DNA gels were in a solid-like state, as evidenced by lack of FRAP (**f**, $14 \pm 5\%$ recovery, mean \pm s.d., $n = 4$ clusters). Scale bars, 5 μ m.



Extended Data Figure 3 | See next page for caption.

Extended Data Figure 3 | Disease-associated, repeat-containing RNAs coalesce into nuclear foci in cells. **a, b**, Expression of $47\times\text{CAG}$ (**a**) and $120\times\text{CAG}$ (**b**) RNA leads to the formation of nuclear puncta. Representative images (left) and quantification of the percentage of cells showing RNA foci (right) with and without induction; n , number of cells analysed. **c**, Time-lapse images of $120\times\text{CAG}$ RNA accumulation in the nuclei of U-2OS cells. Cells were induced with $1\ \mu\text{g ml}^{-1}$ of doxycycline at $t=0$. See also Supplementary Video 3. **d**, Number of foci per cell increased with increasing $47\times\text{CAG}$ RNA expression levels. The expression levels were controlled by increasing the virus titre. **e**, The $47\times\text{CAG}$ RNA accumulated in the nuclei as puncta, while control RNA with coding (mCherry) or a non-coding sequence (mCherry', reverse complement of mCherry sequence) did not form nuclear inclusions, as shown in the representative MS2-YFP fluorescence micrographs (left) and quantification of the number of foci per cell (right). **f**, U-2OS cells were transduced with the indicated constructs tagged with $12\times\text{MS2}$ hairpins under a tetracycline-inducible promoter. RNA was visualized by FISH using Cy3-labelled oligonucleotide probes against MS2-hairpins. Representative micrographs showing the localization of mCherry (top),

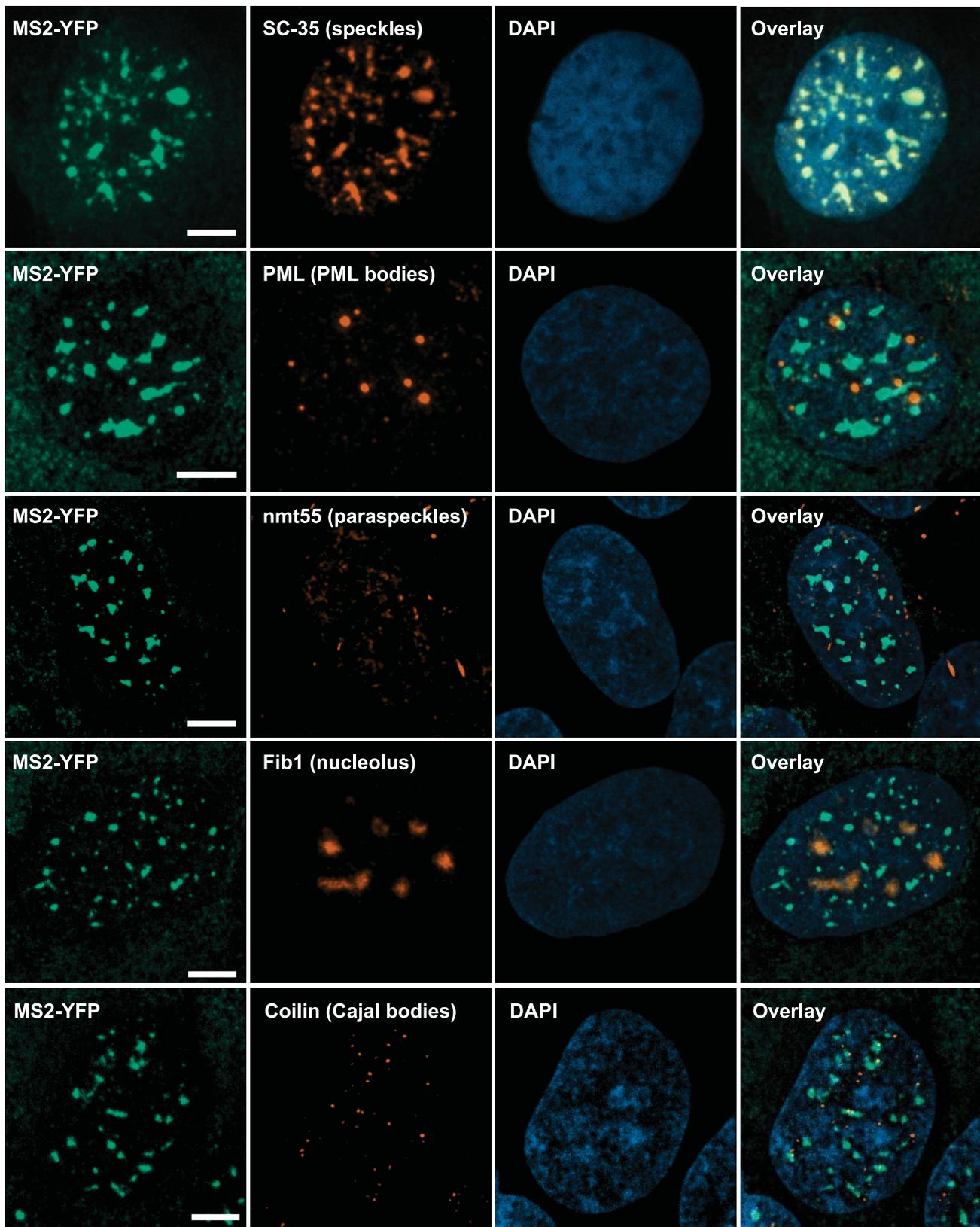
$47\times\text{CAG}$ (middle), and $29\times\text{GGGGCC}$ (bottom) RNA with (+ Tet) or without (- Tet) doxycycline induction. The probes did not bind in the absence of induction. Nuclei are counterstained with DAPI (depicted in blue). **g**, Intensity distribution for single RNA spots in cells expressing $5\times\text{CAG}$ (top) and in the cytoplasm of cells expressing $29\times\text{GGGGCC}$ (bottom). **h**, RNA copy number was determined by dividing the total Cy3 fluorescence intensity in a cell by that of a single RNA, as determined in **g**. The $47\times\text{CAG}$ RNA copy number corresponds to the highest viral titre used in **d**. Similar results were obtained using NanoString ($8,800 \pm 1,500$ copies per cell for $47\times\text{CAG}$ RNA, mean \pm s.d., $n=3$ independent experiments). **i**, Induction of $47\times\text{CAG}$ RNA foci did not cause overt toxicity or a reduction in cell division rates over 7 days. Normalized cell counts in $47\times\text{CAG}$ -transduced cells with or without doxycycline induction. Cell counts were normalized to control cells (without $47\times\text{CAG}$ transduction), grown under corresponding induction conditions. Each data point in **d**, **e**, **h** represents one cell, and data are shown as median \pm interquartile range. Data points in **i** represent technical replicates, and are shown as mean \pm s.d. Scale bars, $5\ \mu\text{m}$.



Extended Data Figure 4 | Identification of RNA foci in live cells.

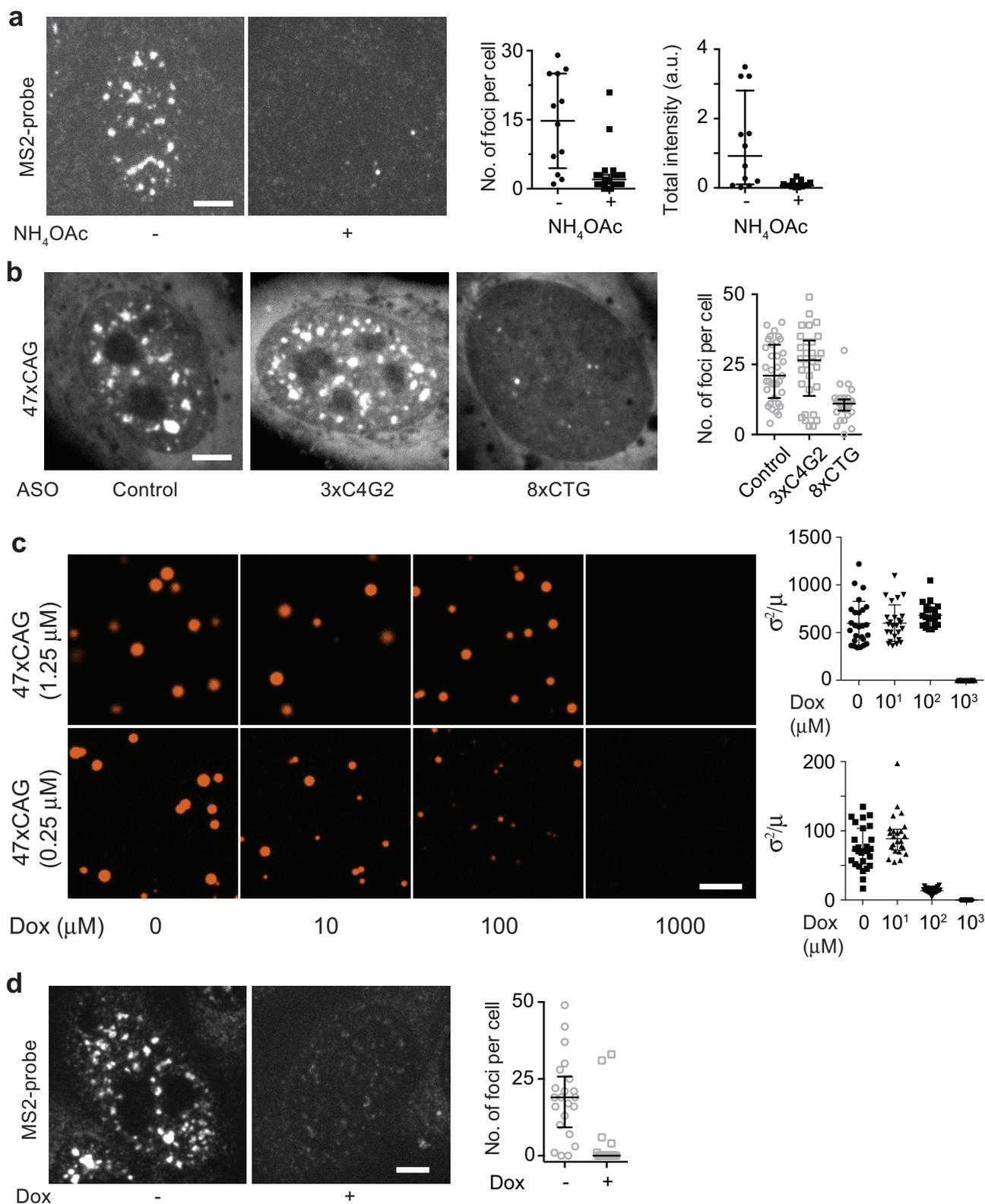
a–c, We used a fluorescence-intensity and size-based threshold to identify RNA foci. In brief, U-2OS cells expressing the RNA of interest together with MS2CP-YFP were imaged using a spinning disk confocal microscope, and $0.3\ \mu\text{m}$ Z-stacks were acquired (**a**). To account for variability in MS2CP-YFP expression levels, we used a cell-intrinsic intensity threshold for foci identification. We manually segmented the nuclei (**b**) and determined the mean fluorescence intensity in the nucleus. RNA foci were identified using the FIJI 3D Objects Counter plugin, with an intensity threshold as $1.6\times$ the mean fluorescence intensity in the

nucleus of the cell, and a size cut-off of more than 50 adjoining pixels (pixel size, $83\ \text{nm} \times 83\ \text{nm}$). This algorithm accurately identified the foci as depicted in **c**. **d–h**, We compared the extent of foci formation in $47\times\text{CAG}$ and $5\times\text{CAG}$ expressing cells. **d**, The mean nuclear fluorescence intensity was similar between the $47\times\text{CAG}$ and $5\times\text{CAG}$ expressing cells. The cells were compared via various metrics: **e**, number of foci per cell; **f**, total volume of foci per cell; **g**, integrated fluorescence intensity of the foci per cell; and **h**, normalized variance in the fluorescence intensity in the nucleus per cell. Scale bar, $5\ \mu\text{m}$. Data are median \pm interquartile range.



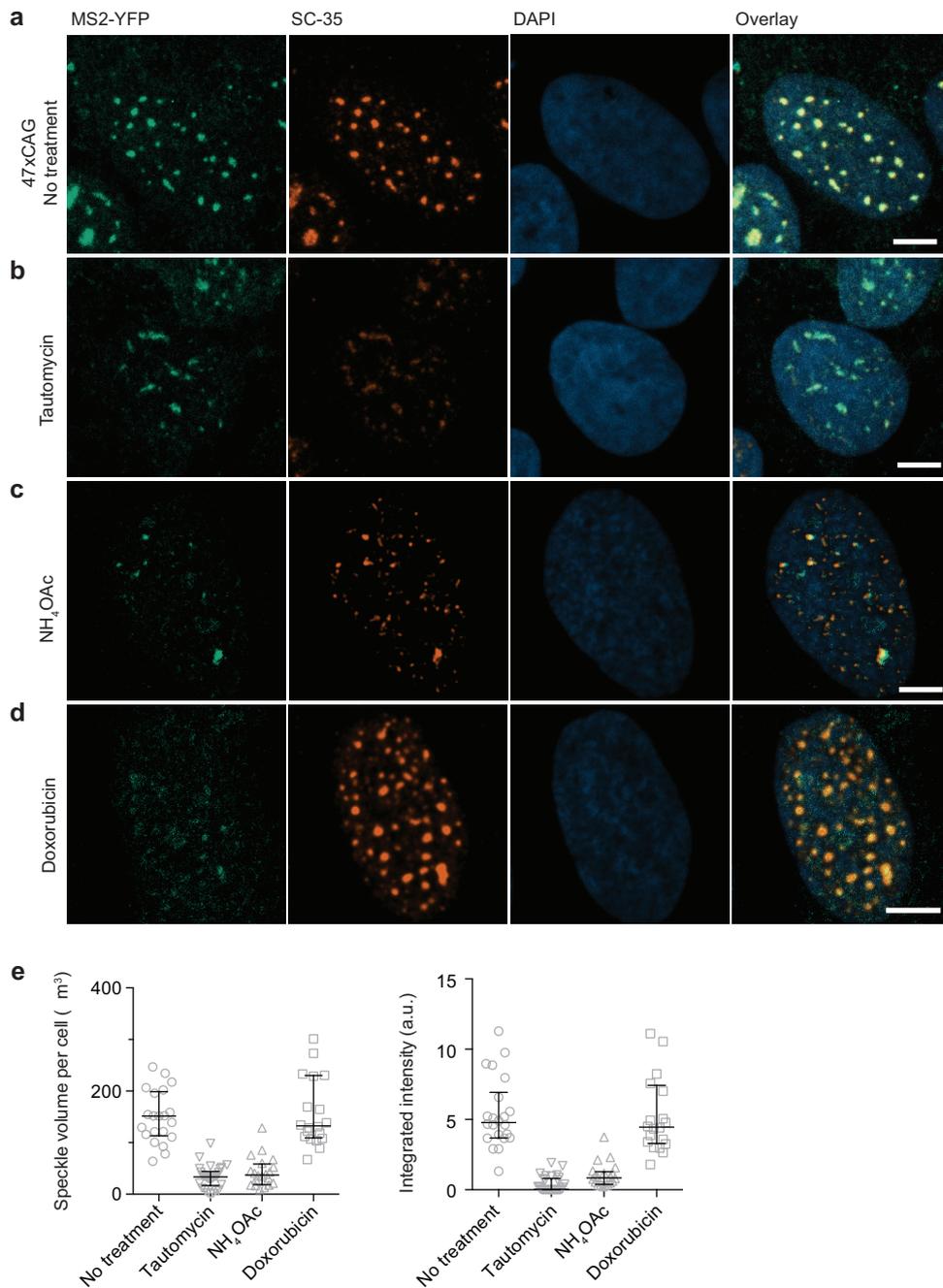
Extended Data Figure 5 | CAG RNA foci co-localize with nuclear speckles. Representative immunofluorescence micrographs depicting that the $47\times$ CAG RNA foci co-localized with the marker for nuclear speckles (SC-35) but not with other nuclear bodies such as PML bodies (PML),

paraspeckles (nmt55), nucleoli (Fib1), or Cajal bodies (coilin). RNA foci were stained using an antibody against GFP. Nuclei were stained with DAPI. Data are representative of three or more independent experiments. Scale bars, 5 μm.



Extended Data Figure 6 | RNA foci are disrupted by treatments that prevent RNA gelation *in vitro*. **a**, RNA FISH using a probe directed against MS2 hairpin loops confirmed that 47×CAG RNA foci were disrupted by treatment with 100 mM NH₄OAc, thus precluding the possibility that the observed disruption of RNA foci in live cells was due to dissociation of MS2CP–YFP from the MS2 hairpins. Representative images and corresponding quantification. **b**, Transfection of an 8×CTG oligonucleotide disrupted 47×CAG RNA foci while control oligonucleotides (3×C4G2 or Control) did not. Representative images and quantification of the number of RNA foci per cell. Sequences of the

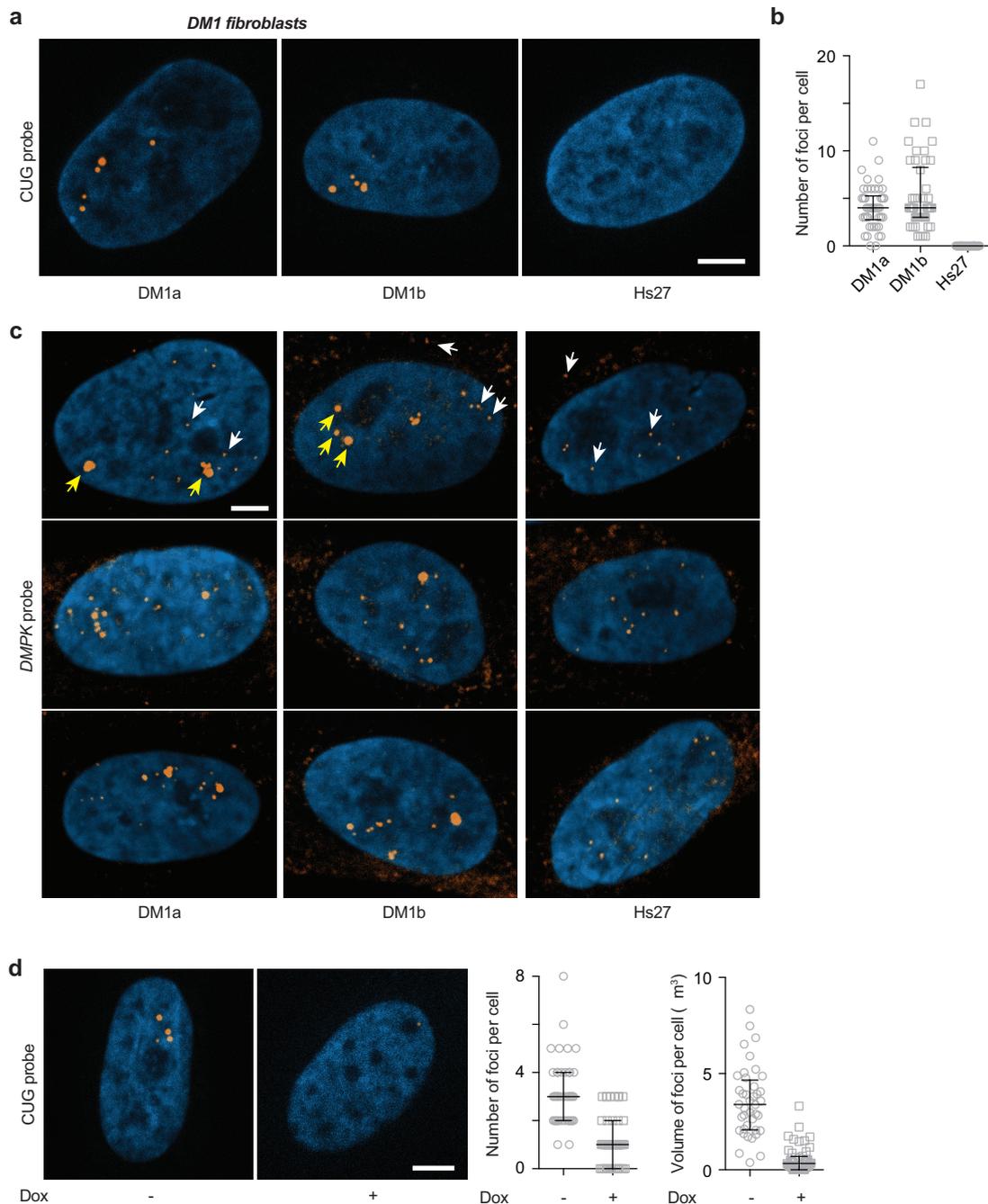
oligonucleotides are provided in Supplementary Table 2. **c**, Doxorubicin (Dox) disrupted 47×CAG RNA clustering *in vitro* in a dose-dependent manner. Representative micrographs and the quantification of the inhomogeneity in the solution at indicated RNA and doxorubicin concentrations. **d**, RNA FISH using a probe directed against MS2 hairpin loops confirmed that 47×CAG RNA foci were disrupted by treatment with 2.5 μM doxorubicin, suggesting that the observed disruption of RNA foci in live cells was probably not an artefact of MS2CP–YFP dissociation from MS2 hairpins. Scale bars, 5 μm. Data are median ± interquartile range. Data are representative of three or more independent experiments.



Extended Data Figure 7 | Doxorubicin disrupts RNA foci but not nuclear speckles. **a**, Representative immunofluorescence micrographs of U-2OS cells expressing 47×CAG stained with antibodies against GFP (MS2-YFP) and SC-35, as a marker for nuclear speckles.

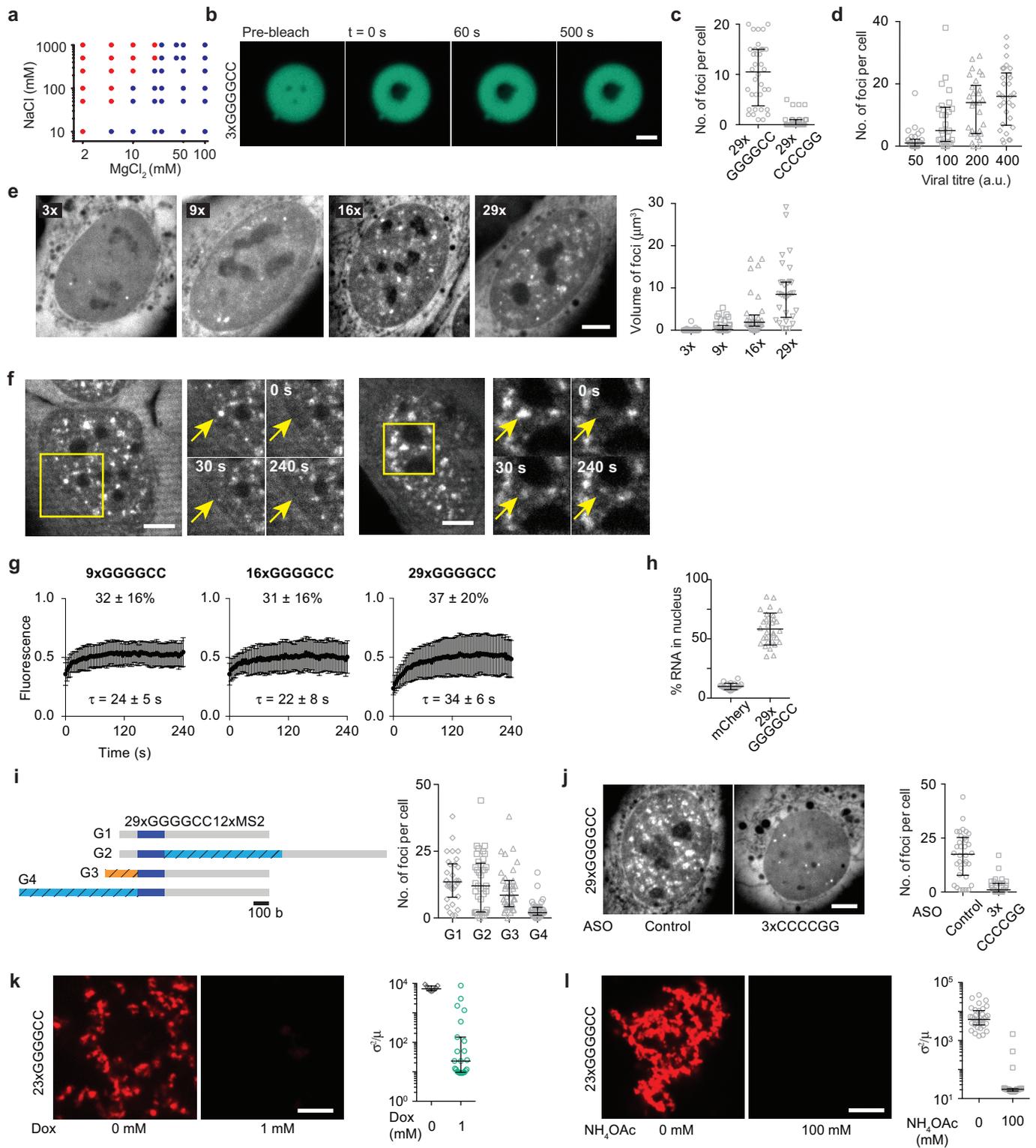
b, **c**, Treatment with 2 μM tautomycin for 4 h (**b**) or 100 mM NH₄OAc for 10 min (**c**) disrupted both the 47×CAG RNA foci as well as the nuclear speckles. **d**, Treatment with 2.5 μM doxorubicin for 2 h specifically

abrogated the 47×CAG RNA foci but the nuclear speckles were not disrupted. Nuclei were counterstained with DAPI (blue). Scale bars, 5 μm. **e**, Quantification of the total volume occupied by nuclear speckles (left) and the integrated intensity of the SC-35 immunofluorescence (right) per cell under various treatments. Data are median ± interquartile range. Data are representative of three or more independent experiments.



Extended Data Figure 8 | Doxorubicin disrupts RNA foci in fibroblasts from patients with DM1. **a**, Fibroblasts derived from patients with DM1 (DM1a, DM1b) or control fibroblasts (Hs27) were stained using a FISH probe directed against expanded CUG repeats ($8\times\text{CAG}$ labelled with Atto647N). Representative fluorescence images showing RNA foci in DM1 cell lines but not in control. Nuclei were counterstained with DAPI (blue). **b**, Quantification of the number of RNA foci per cell across various cell lines. **c**, Single-molecule FISH using probes designed against the wild-type *DMPK* allele showed isolated diffraction-limited spots in control fibroblasts (Hs27), indicated by white arrows, probably arising because of single mRNA. In the patient-derived fibroblasts (DM1a, DM1b), isolated spots (white arrows) as well as several bright puncta (yellow arrows)

were observed. Since both the wild-type and the mutant transcript with expanded CUG repeats could each accommodate the same number of fluorescent probes (48 probes per transcript), the higher brightness indicates that each punctum in cells derived from patients with DM1 (yellow arrows) contained multiple *DMPK* mRNAs. **d**, Treatment with $2\mu\text{M}$ doxorubicin for 24 h reduced the average number of RNA foci per cell by 66% and the total volume of foci per cell by 87%. Representative images with or without doxorubicin treatment and corresponding quantification. Data are aggregated from two independent experiments, and are representative of four or more independent experiments. Scale bars, $5\mu\text{m}$. Data are median \pm interquartile range.

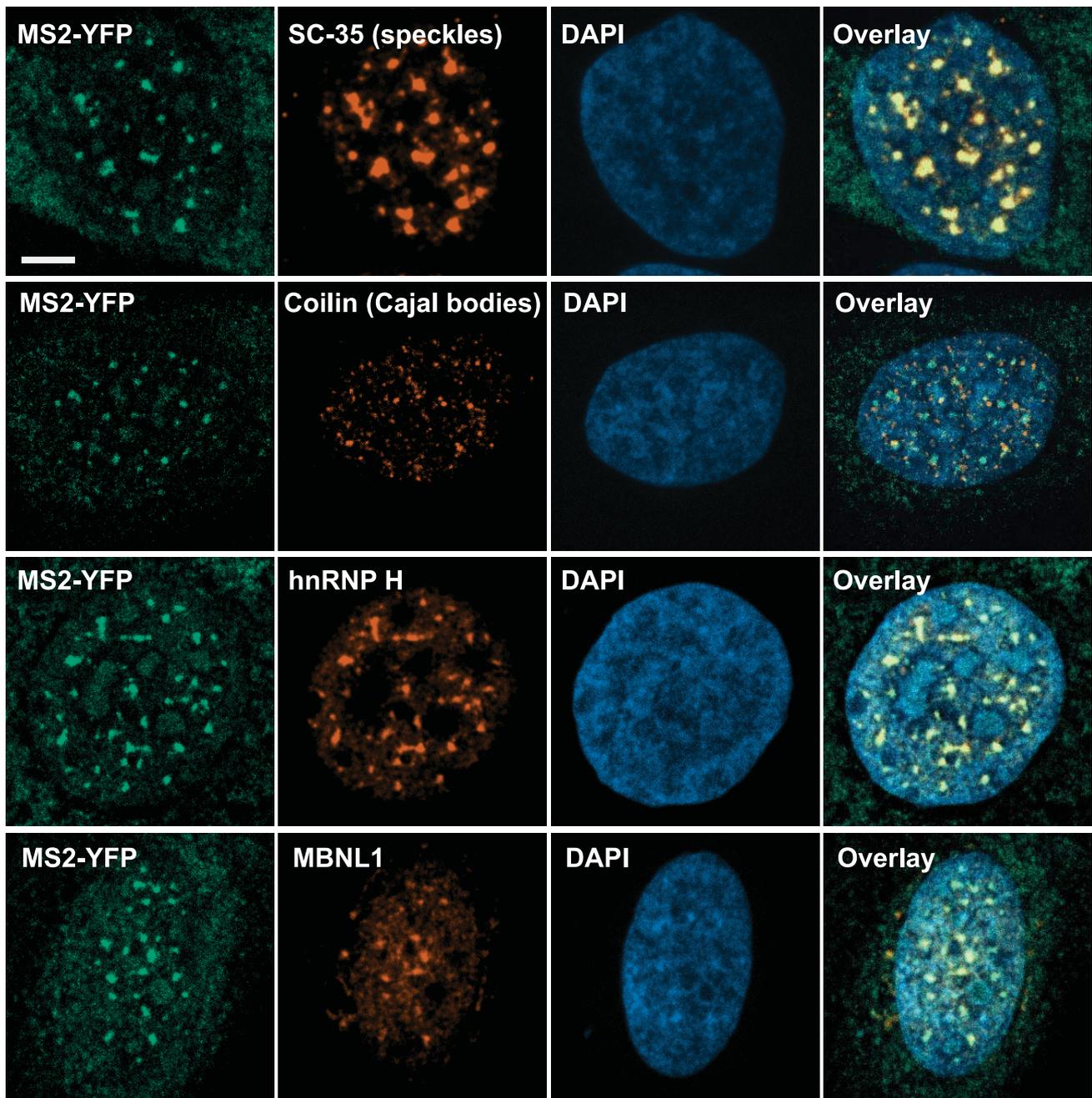


Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | GGGGCC repeat-containing RNAs

form clusters *in vitro* and foci in cells. **a**, Binary phase diagram for 23×GGGGCC RNA clustering *in vitro* as a function of NaCl and MgCl₂ concentrations. RNA concentration was 1.5 μM. Blue dots represent two-phase regime while the red dots indicate a homogenous single-phase regime. **b**, Representative fluorescence micrographs for 3×GGGGCC RNA clusters before and after photobleaching at indicated time points. The lack of fluorescence recovery indicated that the RNA in the clusters was immobile or in a solid-like state. Scale bar, 1 μm. **c**, Quantification of the number of RNA foci per cell for U-2OS cells expressing 29×GGGGCC or 29×CCCCGG RNA. **d**, Number of 29×GGGGCC RNA foci increased with the increasing level of RNA expression. The expression levels were controlled by increasing the virus titre. **e**, Representative fluorescence micrographs and corresponding quantification of the total volume of foci per cell in U-2OS cells transduced with 12×MS2 tagged RNA with the indicated number of GGGGCC repeats. **f**, GGGGCC RNA foci exhibited incomplete recovery upon fluorescence photobleaching. Representative fluorescence micrographs for 29×GGGGCC RNA foci at indicated time points. **g**, Fluorescence recovery plots for GGGGCC RNA foci with indicated number of repeats. Data are

average of $n = 10$ foci at each repeat number. **h**, Percentage of 29×GGGGCC RNA retained in the nucleus compared against a control RNA encoding for mCherry. **i**, Effect of flanking sequences on the formation of GGGGCC RNA foci. Construct G1 had 29×GGGGCC repeats with 12×MS2 hairpins (~0.7 kb) downstream of the repeats for RNA visualization. Incorporation of a ~1 kb long sequence (G2, sequence in Supplementary Table 1) between the 29×GGGGCC repeats and 12×MS2 repeats did not inhibit foci formation. Similarly, RNA foci were observed in construct G3, which had the same 5' flanking sequence as found in the endogenous locus in intron 1 of *c9orf72*. However, incorporation of a longer ~1 kb 5' flanking sequence (G4) inhibited the formation of RNA foci. **j**, Transfection of U-2OS cells with a 3×CCCCGG ASO disrupted the 29×GGGGCC RNA foci while a control ASO did not. Representative micrographs and the quantification of the number of RNA foci per cell. **k**, Representative fluorescence micrographs and corresponding quantification of inhomogeneity for 23×GGGGCC RNA *in vitro* with or without 1 mM doxorubicin. **l**, Same as **k** with or without 100 mM NH₄OAc. Scale bars in **e**, **f**, **j-l**, 5 μm. Data are shown as median ± interquartile range (**c-e**, **i-l**) or mean ± s.d. (**g**, **h**). Data are representative of three or more independent experiments.



Extended Data Figure 10 | GGGGCC RNA foci co-localize with nuclear speckles. Representative immunofluorescence images illustrating that the 29×GGGGCC RNA foci co-localized with the marker for nuclear speckles

(SC-35) but not for Cajal bodies (coilin). The GGGGCC RNA foci also recruited endogenous hnRNP H and MBNL1. Scale bars, 5 μ m. Data are representative of three or more independent experiments.