

## **RNA Gelation in Repeat Expansion Disorders**

Ankur Jain and Ronald D. Vale

Department of Cellular and Molecular Pharmacology and Howard Hughes Medical Institute,  
University of California, San Francisco, CA 94158, USA

Howard Hughes Medical Institute Summer Institute, Marine Biological Laboratory, Woods  
Hole, MA 02543, USA

Address correspondence to Ronald D. Vale: [ron.vale@ucsf.edu](mailto:ron.vale@ucsf.edu)

**Expansions of short nucleotide repeats in the protein coding and non-coding regions of >30 genes produce a variety of neurological and neuromuscular disorders including Huntington’s disease (CAG repeats), muscular dystrophy (CTG repeats) and amyotrophic lateral sclerosis (GGGGCC repeats) [1-3]. Expression of expanded repeats alone is sufficient to recapitulate disease pathology in animal models [4-6]. Repeat-containing transcripts accumulate in the nucleus as aberrant “RNA foci” [7-10] and sequester numerous RNA binding proteins [11,12], leading to a disruption of cellular homeostasis [13,14]. Interestingly, RNA foci, as well as the disease symptoms, only manifest at a critical threshold of nucleotide repeats: >30 for CAG/CTG expansions [1] and >7 for the GGGGCC expansion [15]. However, the reason for this characteristic threshold, as well as the molecular mechanism of foci formation, remain unresolved [16]. Here, we show that nucleotide repeat expansions in RNA create templates for multivalent Watson-Crick (CAG/CUG expansions) or Hoogsteen (GGGGCC expansion) base-pairing. These multivalent interactions cause purified RNAs containing repeat expansions to undergo a sol-gel transition and form micron-sized clusters. Reflecting an increase in the valency for intermolecular hybridization, the gelation of purified RNA only occurs above a critical number of trinucleotide or hexanucleotide repeats. These thresholds for *in vitro* RNA gelation are similar to those associated with manifestation of disease. By visualizing RNA in live cells, we show that nuclear foci form as a result of phase separation of the repeat-containing RNA and that these foci 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.**

The RNAs associated with most repeat expansion diseases contain a regular pattern of Gs and Cs and can form intramolecular hairpins [17]. Because of their repeating nature, these sequences can also, in principle, template multivalent intermolecular interactions (Fig. 1a). Such multivalent interactions could potentially lead to the formation of large clusters via liquid-liquid phase separation [18,19] or a sol-gel transition [20-22]. To examine whether repeat-containing RNAs assemble into large clusters, we synthesized fluorescently-labeled RNAs containing 47 triplet

repeats of CAG (47xCAG), which is the expansion underlying Huntington's disease and spinocerebellar ataxias or CUG (47xCUG), which produces myotonic dystrophy and HDL2 [1]. As controls, we used RNAs of equivalent length but with arbitrary sequences with 30-75% GC content and RNAs with identical base composition as 47xCAG and 47xCUG RNAs but with scrambled sequences. Upon annealing (see Methods), the 47xCAG and 47xCUG repeat-containing RNAs formed micron-sized spherical clusters, while the control RNAs remained soluble (~5  $\mu$ M for each RNA) (Fig. 1b, Supplementary Fig. 1a, 1b). The spherical clusters were >100-fold enriched in the RNA as compared to the solution phase and contained nearly one half of the RNA in the reaction (Supplementary Fig. 1c). As expected, these clusters were dissolved by RNaseA, but not proteinase K or DNase I (Supplementary Fig. 1d), confirming that the RNA clustering is not mediated by protein or DNA contaminants.

Molecules that form multivalent interactions show abrupt phase transitions with increasing valency of interaction [18-20]. An increase in the number of triplet repeats results in an increased valency for intermolecular hybridization. Consistent with this valency dependence, we find that the formation of spherical CUG and CAG RNA clusters occurred only with 30 or more triplet repeats (Fig. 1c). The intermolecular interactions between the transcripts could potentially be competed by shorter complementary antisense oligonucleotides. Indeed, addition of a 6xCTG antisense oligonucleotide prevented clustering of 47xCAG RNA, while control oligonucleotides that either cannot hybridize with the RNA or bind outside the repeat region did not (Fig. 1d). Collectively, these experiments indicate that intermolecular base-pairing interactions in the CAG/CUG-repeat region can lead to the assembly of these disease-associated RNAs in to micron sized clusters.

We next analyzed the physical properties of the CAG/CUG RNA clusters. Their spherical shapes (aspect ratio  $1.05 \pm 0.1$ , mean  $\pm$  s.d.,  $n = 214$ ) are characteristic of polymers undergoing liquid-liquid phase separation [23]. Molecules within the liquid phase are mobile and undergo fast internal rearrangement. However, fluorescence photobleaching of the RNA in the clusters revealed little or no fluorescence recovery over  $\sim 10$  min, indicating that the RNA in the clusters is immobile (Fig. 1e, Supplementary Fig. 2a, b, Supplementary Movie 1) and thus in a solid-like state. Because of their solid-like behavior, we refer to these CAG/CUG RNA clusters as “RNA gels” [21,22,24]. We hypothesize that these RNAs initially phase separate into spherical liquid-like droplets but then rapidly become cross-linked into a gel due to increasing intermolecular base-pairing. Consistent with this idea, we occasionally (in  $\sim 1\%$  of the clusters) find evidence of incomplete fusion events where two droplets solidified prior to complete relaxation to a single spherical geometry (Supplementary Fig. 2c). As further evidence for this model, we found that single stranded DNA (polyT) in the presence of polyvalent cations formed liquid-like droplets but that the incorporation of multivalent base-pairing sites imparted solid-like properties to the DNA droplets (Supplementary Fig. 3, Supplementary Movie 2). Similar liquid-to-solid phase transitions have also been observed in proteins with increasing strength of intermolecular interactions [25,26]. In summary, multivalent base-pairing interactions lead to the gelation of CAG/CUG triplet repeat containing RNA *in vitro* at a similar critical number of repeats as observed for the manifestation of clinical symptoms.

Using *in situ* hybridization on tissues from patient biopsies, several groups have described the accumulation of trinucleotide repeat-containing RNA in aberrant nuclear foci [8]. We wished to

establish a live cell reporter assay to follow the repeat-containing RNAs after transcription and determine if they assemble into nuclear foci. To enable live cell RNA visualization, we tagged the repeat-containing RNAs with 12x MS2-hairpin loops [27] and co-expressed YFP-tagged MS2-coat binding protein (Fig. 2a). Multiple stop codons upstream of the repeats were included to minimize translation of repeat containing proteins [28] (Fig. 2a). After induction of 47xCAG (Fig. 2b, Supplementary Fig. 4a) or 120xCAG (Supplementary Fig. 4b) RNA transcription in U-2OS cells, numerous nuclear foci appeared as early as 1 h post induction (Supplementary Fig. 4c, Supplementary Fig. 5, Supplementary Movie 3). The RNA foci recruited endogenous Muscleblind-like-1 protein (Supplementary Fig. 4d), sequestration of which has been previously implicated in CAG/CUG RNA toxicity [5,11]. The number of foci per nucleus increased with higher levels of RNA induction (Supplementary Fig. 4e) and with increasing repeat number (Fig. 2c). In contrast, 5xCAG (Fig. 2b, c) or control RNAs with coding or non-coding sequences (Supplementary Fig. 4f) did not form nuclear puncta.

Real-time microscopy revealed that the 47xCAG and 120xCAG RNA foci exhibit liquid-like properties. For example, two or more foci that came in close contact could fuse with one another (Fig. 2d, 2e, Supplementary Movies 3, 4), a hallmark of liquid-like behavior as has been previously shown for various physiological membraneless organelles [29,30]. Upon photobleaching, nuclear puncta exhibited near-complete fluorescence recovery ( $83 \pm 13\%$  recovery,  $\tau_{\text{FRAP}} = 81 \pm 24$  s; mean  $\pm$  s.d.,  $n = 5$  foci across 3 independent experiments), indicating that the RNA is dynamic and can move into and out of the nuclear foci (Fig. 2f, g, Supplementary Movie 5). Upon photobleaching a portion of a 47xCAG RNA punctum, the fluorescence also recovered rapidly ( $\tau_{\text{FRAP}} = 18 \pm 5$  s, mean  $\pm$  s.d.,  $n = 5$  foci) (Fig. 2h, i,

Supplementary Movie 6), suggesting that RNA can undergo internal rearrangement within the foci, which is further indicative of liquid-like behavior [29]. Thus, unlike the solid-like behavior of clusters prepared from pure CAG RNA *in vitro* (Fig. 1e), CAG RNA nuclear puncta in cells display liquid-like properties. We hypothesized that the increased dynamicity might arise from specialized proteins (e.g. helicases) in the nucleoplasm that remodel RNA base-pairing. Consistent with this hypothesis, depletion of cellular ATP significantly reduced fluorescence recovery of the RNA foci upon photobleaching ( $23 \pm 7\%$  recovery, mean  $\pm$  s.d.,  $n = 7$ , Fig. 2g, Supplementary Movie 7).

Next, we asked if perturbations that prevent RNA gelation *in vitro* also affect the stability of RNA foci in cells. We found that doxorubicin, a nucleic acid intercalator that disrupts base pairing [31], blocked the formation of CAG RNA gels *in vitro* (Fig. 3a). Similarly, addition of doxorubicin (2.5  $\mu$ M) potently dissolved the 47xCAG nuclear foci in cells (Fig. 3b). Second, phase separation of anionic polymers, like RNAs, is promoted by multivalent cations, which can bridge between polymers, and is inhibited by monovalent cations that compete for occupancy with the multivalent cations [32,33]. We found that the gelation of the CAG repeat-containing RNAs was inhibited by monovalent cations such as  $\text{NH}_4^+$  (Fig. 3c). To test the effect of monovalent cations in cells, we used ammonium acetate, which readily permeates into cells and does not perturb intracellular pH [34]. Strikingly, the addition of 100 mM ammonium acetate led to the disappearance of 47xCAG RNA foci within minutes (Fig. 3d, Supplementary Movie 8). To preclude the possibility that ammonium acetate or doxorubicin cause the dissociation of MS2-coat binding protein from the RNA, we performed RNA fluorescence in situ hybridization experiments and observed a similar reduction in the number of nuclear foci (Supplementary Fig.

6). Thus, agents that inhibit the gelation of purified repeat-containing RNAs *in vitro* also disrupt nuclear foci in cells.

Besides the canonical Watson-Crick base pairing, nucleic acids may also form Hoogsteen base-pairs such as in G-quadruplexes. Interestingly, RNA with the ALS/FTD-associated GGGGCC repeat expansion found in the *C9orf72* locus, the most common genetic cause of ALS [35], is shown to form G-quadruplexes *in vitro* and *in vivo* [36,37]. A single G-quartet can bring up to four RNA strands together, and the repeating G-quadruplex as found in *C9orf72* could potentially give rise to multimolecular RNA complexes [36] (Fig. 4a). Additionally, G-quadruplexes form stronger interactions than Watson-Crick base pairing [38]. Indeed, we find that an RNA with 3xGGGGCC is largely soluble but as few as 5x repeats of GGGGCC forms spherical clusters (Fig. 4b). Longer repeats (10x and 23x) resulted in the formation of an interconnected mesh-like network of aggregated RNA (Fig. 4b). In contrast, 23xCCCCGG RNA, which can form multivalent Watson-Crick base pairing but not G-quadruplexes, remained soluble under the same conditions in which 23xGGGGCC formed aggregates (Fig. 4c). As described for the CAG/CUG RNA gels, the GGGGCC RNA clusters are also solid-like and did not exhibit fluorescence recovery after photobleaching (Supplementary Fig. 7). In summary, repetitive GGGGCC RNA assemble in to supramolecular aggregates *in vitro* by forming multimolecular G-quadruplexes at a lower repeat number than repetitive CAG/CUG RNA.

Expression of 29xGGGGCC RNA in U-2OS cells also led to the formation of nuclear puncta (Fig. 4d). However, the GGGGCC RNA foci exhibit incomplete recovery after fluorescence photobleaching ( $36 \pm 16\%$  recovery, mean  $\pm$  s.d.,  $n = 5$ , Fig. 4e, f, Supplementary Movie 9),

indicating that the GGGGCC RNA foci are less dynamic than the CAG RNA foci. This result is consistent with a stronger inter-molecular interaction between the GGGGCC quadruplex (Fig. 4a). Lastly, treatment with doxorubicin or ammonium acetate disrupted GGGGCC RNA gels *in vitro* as well as the RNA foci in cells (Fig. 4g-j) indicating that both base-pairing and electrostatic interactions are essential for the assembly of GGGGCC RNA foci.

In summary, we have shown that the propensity of an RNA to form multivalent canonical Watson-Crick (in CAG/CUG expansions) or Hoogsteen (in GGGGCC expansions) 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 nucleolus, are phase separated compartments [24]. Numerous recent studies have characterized proteins with regard to their ability to phase separate and mediate the assembly of these RNP granules [26,39,40]. Our results demonstrate that sequence-specific base-pairing properties of RNAs can lead 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 diseases produced by nucleotide repeat expansions, our data suggest that intermolecular RNA base-pairing can result in the aggregation and sequestration of RNA into nuclear foci (Fig. 4k). RNA gelation, which occurs at a boundary condition of increasing valency, might explain why a critical threshold of nucleotide repeats must be reached to produce disease (Fig. 4k), a puzzling phenomenon that has lacked a mechanistic explanation thus far [1,16]. RNA gelation and foci formation likely occur upstream of known drivers of cellular toxicity such as sequestration of RNA binding proteins [11,12], nucleolar stress [41] and

disruption of nucleocytoplasmic transport [42,43]. In addition to these direct consequences of nuclear RNA aggregation, aberrant protein production due to canonical [44] or non-ATG translation [45] from repeat expansions also likely contributes to toxicity and neurological disease. Our results may explain why placement of distinct repeat expansions in seemingly unrelated genes can result in similar clinical syndromes [46,47]. In addition, our findings explain how RNA duplex [48,49] or quadruplex [50] destabilizing compounds disrupt nuclear foci and suggest that new strategies for disrupting RNA-RNA base pairing could lead to novel therapeutics for treating repeat expansion diseases.

## References:

1. La Spada AR, Taylor JP. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat. Rev. Genet.* 11(4), 247–258 (2010).
2. Renton AE, Majounie E, Waite A, *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 72(2), 257–268 (2011).
3. Gatchel JR, Zoghbi HY. Diseases of unstable repeat expansion: mechanisms and common principles. *Nat. Rev. Genet.* 6(10), 743–755 (2005).
4. Mankodi A, Logigian E, Callahan L, *et al.* Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science.* 289(5485), 1769–1773 (2000).
5. Li L-B, Yu Z, Teng X, Bonini NM. RNA toxicity is a component of ataxin-3 degeneration in *Drosophila*. *Nature.* 453(7198), 1107–1111 (2008).
6. Chew J, Gendron TF, Prudencio M, *et al.* C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science.* 348(6239), 1151–1154 (2015).
7. Taneja KL, McCurrach M, Schalling M, Housman D, Singer RH. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.* 128(6), 995–1002 (1995).
8. Wojciechowska M, Krzyzosiak WJ. Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum. Mol. Genet.* 20(19), 3811–3821 (2011).
9. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 72(2), 245–256 (2011).
10. Du J, Aleff RA, Soragni E, *et al.* RNA toxicity and missplicing in the common eye disease fuchs endothelial corneal dystrophy. *J. Biol. Chem.* 290(10), 5979–5990 (2015).
11. Miller JW, Urbinati CR, Teng-Umnuay P, *et al.* Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.* 19(17), 4439–4448 (2000).
12. Cooper-Knock J, Walsh MJ, Higginbottom A, *et al.* Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions. *Brain.* 137(Pt 7), 2040–2051 (2014).
13. Belzil VV, Gendron TF, Petrucelli L. RNA-mediated toxicity in neurodegenerative disease. *Mol. Cell. Neurosci.* 56, 406–419 (2013).
14. Todd PK, Paulson HL. RNA-mediated neurodegeneration in repeat expansion disorders. *Ann. Neurol.* 67(3), 291–300 (2010).

15. van der Zee J, Gijselinck I, Theuns J, *et al.* A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats. *Hum. Mutat.* 34(2), 363–373 (2013).
16. Lee D-Y, McMurray CT. Trinucleotide expansion in disease: why is there a length threshold? *Curr. Opin. Genet. Dev.* 26, 131–140 (2014).
17. Krzyzosiak WJ, Sobczak K, Wojciechowska M, Fiszler A, Mykowska A, Kozlowski P. Triplet repeat RNA structure and its role as pathogenic agent and therapeutic target. *Nucleic Acids Res.* 40(1), 11–26 (2012).
18. Flory PJ. Principles of Polymer Chemistry. Cornell University Press.
19. Li P, Banjade S, Cheng H-C, *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature.* 483(7389), 336–340 (2012).
20. Stockmayer WH. Theory of Molecular Size Distribution and Gel Formation in Branched Polymers II. General Cross Linking. *The Journal of Chemical Physics.* 12(4), 125–131 (1944).
21. Semenov AN, Rubinstein M. Thermoreversible gelation in solutions of associative polymers. 1. Statics. *Macromolecules.* 31(4), 1373–1385 (1998).
22. Falkenberg CV, Blinov ML, Loew LM. Pleomorphic ensembles: formation of large clusters composed of weakly interacting multivalent molecules. *Biophys. J.* 105(11), 2451–2460 (2013).
23. Larson RG. The Structure and Rheology of Complex Fluids. Oxford University Press.
24. Hyman AA, Weber CA, Jülicher F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* 30(1), 39–58 (2014).
25. Patel A, Lee HO, Jawerth L, *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell.* 162(5), 1066–1077 (2015).
26. Molliex A, Temirov J, Lee J, *et al.* Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell.* 163(1), 123–133 (2015).
27. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA particles in living yeast. *Molecular Cell.* 2(4), 437–445 (1998).
28. Kearse MG, Green KM, Krans A, *et al.* CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Molecular Cell.* 62(2), 314–322 (2016).
29. Brangwynne CP, Eckmann CR, Courson DS, *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science.* 324(5935), 1729–

- 1732 (2009).
30. Brangwynne CP, Mitchison TJ, Hyman AA. Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes. *Proc. Natl. Acad. Sci. U.S.A.* 108(11), 4334–4339 (2011).
  31. Frederick CA, Williams LD, Ughetto G, *et al.* Structural comparison of anticancer drug-DNA complexes: adriamycin and daunomycin. *Biochemistry.* 29(10), 2538–2549 (1990).
  32. Jia TZ, Fahrenbach AC, Kamat NP, Adamala KP, Szostak JW. Oligoarginine peptides slow strand annealing and assist non-enzymatic RNA replication. *Nat Chem.* 8(10), 915–921 (2016).
  33. Aumiller WM, Keating CD. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat Chem.* 8(2), 129–137 (2016).
  34. Hamaguchi MS, Watanabe K, Hamaguchi Y. Regulation of intracellular pH in sea urchin eggs by medium containing both weak acid and base. *Cell Struct. Funct.* 22(4), 387–398 (1997).
  35. Taylor JP, Brown RH, Cleveland DW. Decoding ALS: from genes to mechanism. *Nature.* 539(7628), 197–206 (2016).
  36. Reddy K, Zamiri B, Stanley SYR, Macgregor RB, Pearson CE. The disease-associated r(GGGGCC)<sub>n</sub> repeat from the C9orf72 gene forms tract length-dependent uni- and multimolecular RNA G-quadruplex structures. *J. Biol. Chem.* 288(14), 9860–9866 (2013).
  37. Conlon EG, Lu L, Sharma A, *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).
  38. Lane AN, Chaires JB, Gray RD, Trent JO. Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* 36(17), 5482–5515 (2008).
  39. Lin Y, Protter DSW, Rosen MK, Parker R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell.* 60(2), 208–219 (2015).
  40. Feric M, Vaidya N, Harmon TS, *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell.* 165(7), 1686–1697 (2016).
  41. Tsoi H, Lau TC-K, Tsang S-Y, Lau K-F, Chan HYE. CAG expansion induces nucleolar stress in polyglutamine diseases. *Proc. Natl. Acad. Sci. U.S.A.* 109(33), 13428–13433 (2012).
  42. Freibaum BD, Lu Y, Lopez-Gonzalez R, *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature.* 525(7567), 129–133 (2015).

43. Zhang K, Donnelly CJ, Haeusler AR, *et al.* The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature*. 525(7567), 56–61 (2015).
44. Paulson HL, Bonini NM, Roth KA. Polyglutamine disease and neuronal cell death. *Proc. Natl. Acad. Sci. U.S.A.* 97(24), 12957–12958 (2000).
45. Zu T, Gibbens B, Doty NS, *et al.* Non-ATG-initiated translation directed by microsatellite expansions. *Proc. Natl. Acad. Sci. U.S.A.* 108(1), 260–265 (2011).
46. Holmes SE, O'Hearn E, Rosenblatt A, *et al.* A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2. *Nat. Genet.* 29(4), 377–378 (2001).
47. Elden AC, Kim H-J, Hart MP, *et al.* Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*. 466(7310), 1069–1075 (2010).
48. Kumar A, Parkesh R, Sznajder LJ, Childs-Disney JL, Sobczak K, Disney MD. Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. *ACS Chem. Biol.* 7(3), 496–505 (2012).
49. García-López A, Llamusi B, Orzáez M, Pérez-Payá E, Artero RD. In vivo discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models. *Proc. Natl. Acad. Sci. U.S.A.* 108(29), 11866–11871 (2011).
50. Zamiri B, Reddy K, Macgregor RB, Pearson CE. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)<sub>n</sub> repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J. Biol. Chem.* 289(8), 4653–4659 (2014).

### **Acknowledgments:**

We thank M. K. Rosen, W. W. Seeley and the members of Vale lab for helpful discussions. A.J. is supported by the Damon Runyon Cancer Research Foundation postdoctoral fellowship. R.V. is an investigator with the Howard Hughes Medical Institute.

## Figure captions

**Figure 1. Triplet repeat containing RNAs undergo gelation *in vitro* at a critical number of repeats.** (a) Schematic for possible multivalent base-pairing and multimerization of CAG repeat RNA. (b) Representative fluorescence micrographs for a control RNA with 75% GC content (left), 47xCUG (center) and 47xCAG (right) RNA. Quantitation of the inhomogeneity in the sample as normalized variance ( $\sigma^2/\mu$ ); higher values reflect higher inhomogeneity (see Methods); each data point represents an independent imaging area ( $1800 \mu\text{m}^2$ ). (c) Representative micrographs of RNA with different CAG- and CUG- repeat number (200 ng/ $\mu\text{l}$  RNA concentration). Data are quantified as in (b). (d) DNA oligonucleotides are designed against the 47xCAG RNA that can hybridize with the CAG-repeats (6xCTG), with the 5' linker sequence (ASO-2) or without sequence complementarity (ASO-1). 6xCTG blocks clustering of 47xCAG RNA, whereas the control oligonucleotides, ASO-1 and ASO-2 do not. (ASO: antisense oligonucleotide) (e) Fluorescence time-trajectories for 47xCUG and 47xCAG RNA clusters indicating that the clusters do not exhibit fluorescence recovery upon photobleaching. Sample images for 47xCAG RNA (top) at indicated time points before and after photobleaching. Dashed line depicts the time point of photobleaching. Scale bars are  $5 \mu\text{m}$ . Error bars depict median and interquartile range. Data are representative of  $n = 3$  or more independent experiments.

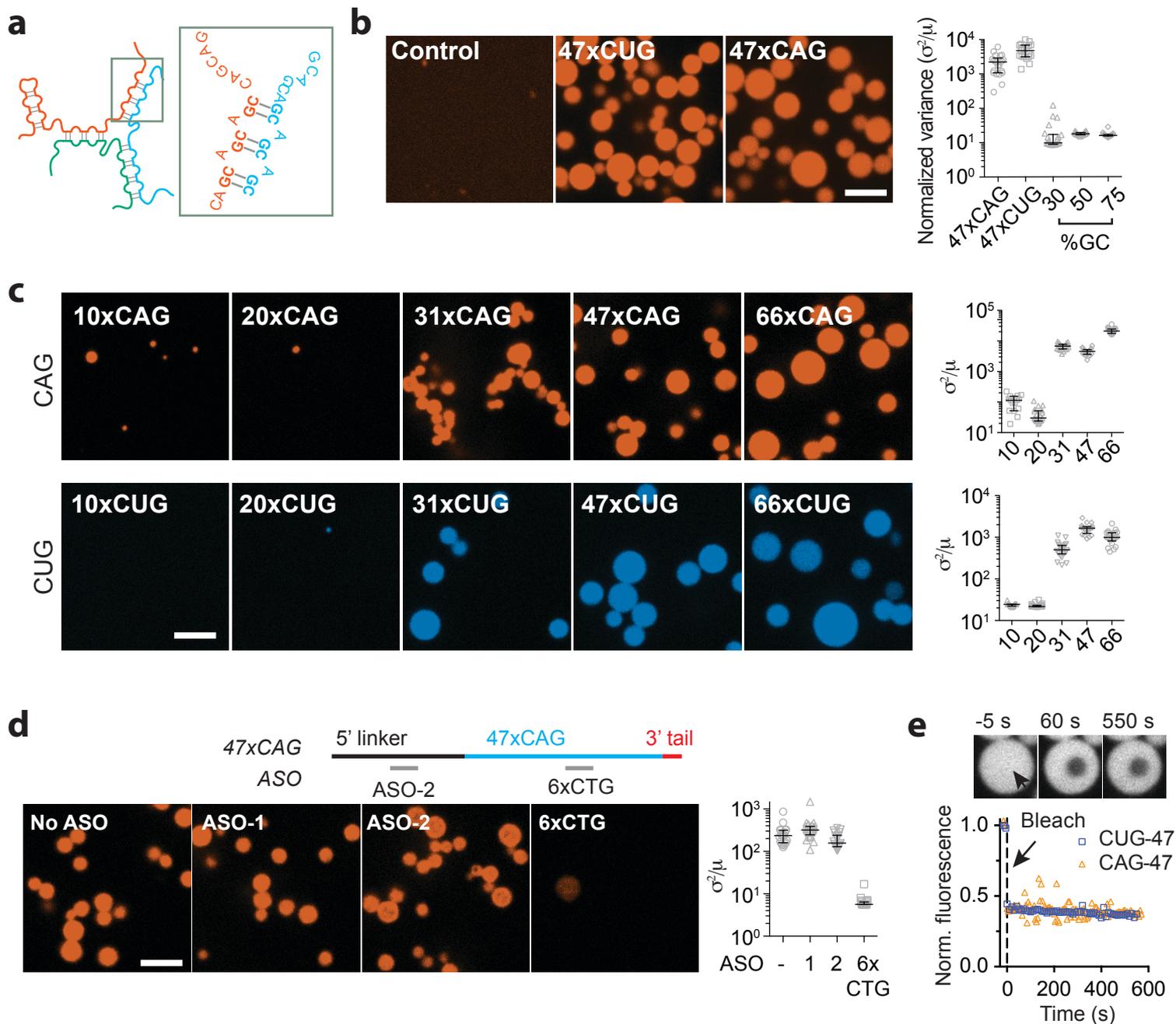
**Figure 2. CAG repeat-containing RNAs coalesce in to liquid-like nuclear foci.** (a) Schematic for RNA visualization in cultured U-2OS cells. (b) MS2-YFP fluorescence images showing accumulation of RNA as nuclear foci in cells expressing 47xCAG (right) but not 5xCAG (left). (c) Number of foci per cell as a function of CAG repeat number. Each data point represents one cell. Error bars depict median and interquartile range. (d, e) Fusion event where two adjacent

RNA foci merge in to one. Images at indicated time-points after induction (d) and a corresponding kymograph (e). (f) Fluorescence micrographs for 47xCAG RNA foci before and after photobleaching at indicated time points after photobleaching. Arrow depicts site of bleaching. (g). Fluorescence recovery plots after photobleaching for 47xCAG RNA punctum before (47xCAG) and after ATP depletion (- ATP). (h, i) Partial photobleaching of a 47xCAG RNA punctum. Fluorescence micrographs at indicated time points before and after photobleaching (h) and a corresponding kymograph (i). Scale bars are 5  $\mu\text{m}$ . Error bars depict median and interquartile range. Data are representative of  $n = 3$  or more independent experiments.

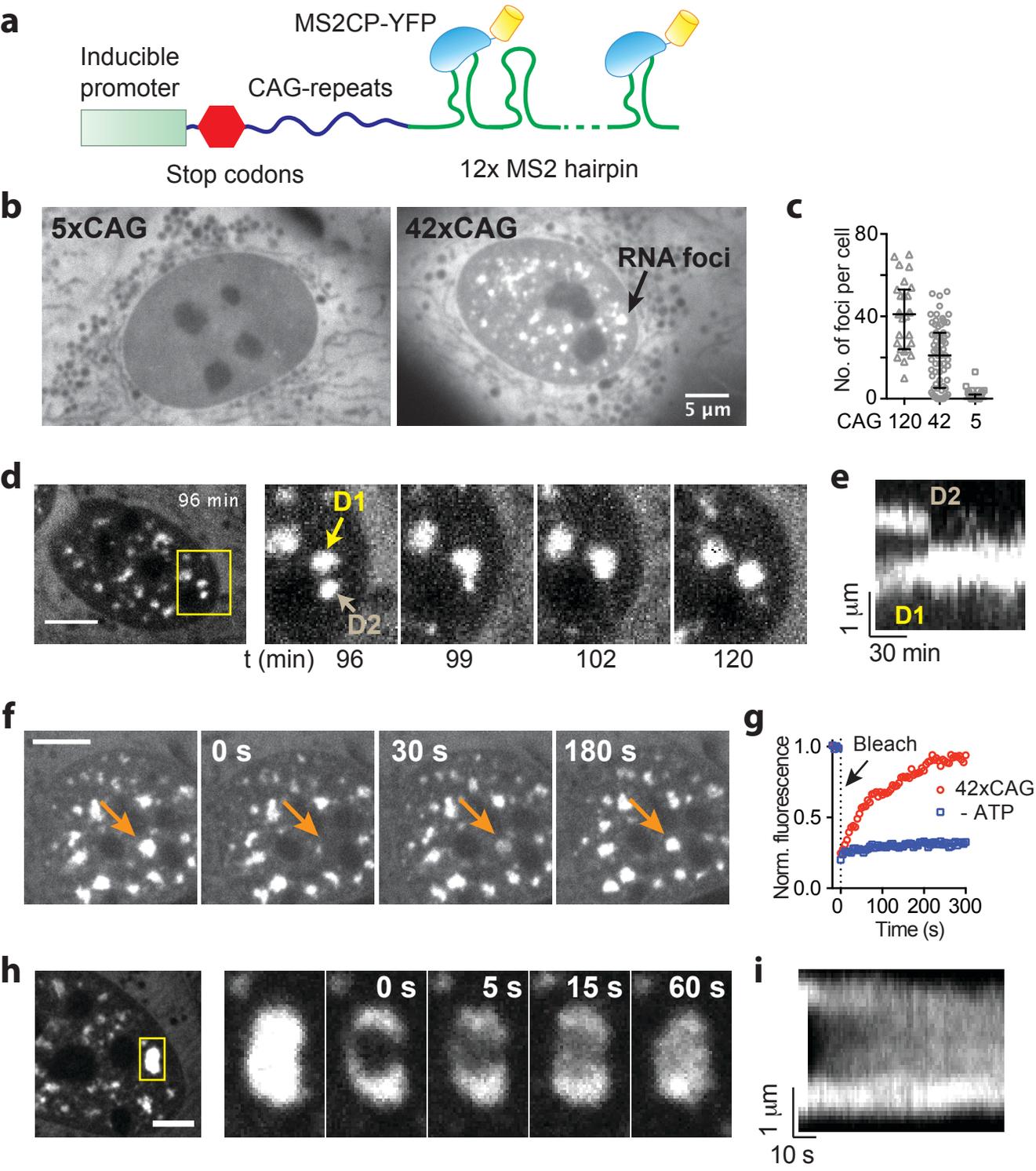
**Figure 3. Nuclear foci are reversed by agents that disrupt RNA gelation *in vitro*.** (a) Representative fluorescence micrographs and corresponding quantification of inhomogeneity for 47xCAG RNA *in vitro* with or without 1 mM doxorubicin. (b) Fluorescence micrographs of 47xCAG expressing cells with or without treatment with 2.5  $\mu\text{M}$  doxorubicin. Quantification of the number of foci per cell 2 h after treatment with doxorubicin at indicated doses. (c) Same as (a) for 47xCAG RNA with or without 100 mM  $\text{NH}_4\text{OAc}$ . (d) Treatment of 47xCAG expressing cells with 100 mM  $\text{NH}_4\text{OAc}$  disrupted RNA foci. Images of a cell before and 300 s after treatment with  $\text{NH}_4\text{OAc}$ . Quantification of the number of foci per cell before and 10 min after treatment with 100 mM  $\text{NH}_4\text{OAc}$ . Scale bars are 5  $\mu\text{m}$ . Error bars depict median and interquartile range.

**Figure 4. Gelation of GGGGCC repeat containing RNA and model for RNA foci formation.** (a) Schematic for GGGGCC RNA multimerization (GQ: G-quadruplex, WC: Watson-Crick). (b) Representative micrographs for GGGGCC RNA at indicated number of repeats. (c) 23xCCCCGG RNA is soluble under similar conditions. (d) Expression of 29xGGGGCC RNA leads to formation of nuclear foci in U-2OS cells. (e, f) 29xGGGGCC RNA foci exhibit incomplete fluorescence recovery upon photobleaching. Micrographs at indicated time points (e) and recovery plot (f). (g) Representative fluorescence micrographs and corresponding quantification of inhomogeneity for 23xGGGGCC RNA *in vitro* with or without 1 mM doxorubicin. (h) Fluorescence micrographs of 29xGGGGCC expressing cells treated with 2.5  $\mu$ M doxorubicin for 2 h (right) or DMSO only (left). Quantification of the number of foci per cell with or without doxorubicin treatment. (i) Same as (g) with or without 100 mM  $\text{NH}_4\text{OAc}$  (j) Treatment of 29xGGGGCC RNA expressing cells with 100 mM  $\text{NH}_4\text{OAc}$  disrupted RNA foci. Images of a cell before and 300 s after treatment with  $\text{NH}_4\text{OAc}$ . Quantification of the number of foci per cell before and 10 min after treatment with 100 mM  $\text{NH}_4\text{OAc}$ . Scale bars are 5  $\mu$ m. Error bars depict median and interquartile range. (k) Model for RNA foci formation in repeat expansion disorders. The repeat expansion sequences form templates for intermolecular base pairing. An increase in the repeat number leads to an increased valency of intermolecular interaction, and thus an increased propensity to undergo gelation. In cells, the RNA clusters are retained in the nucleus and may disrupt cellular homeostasis.

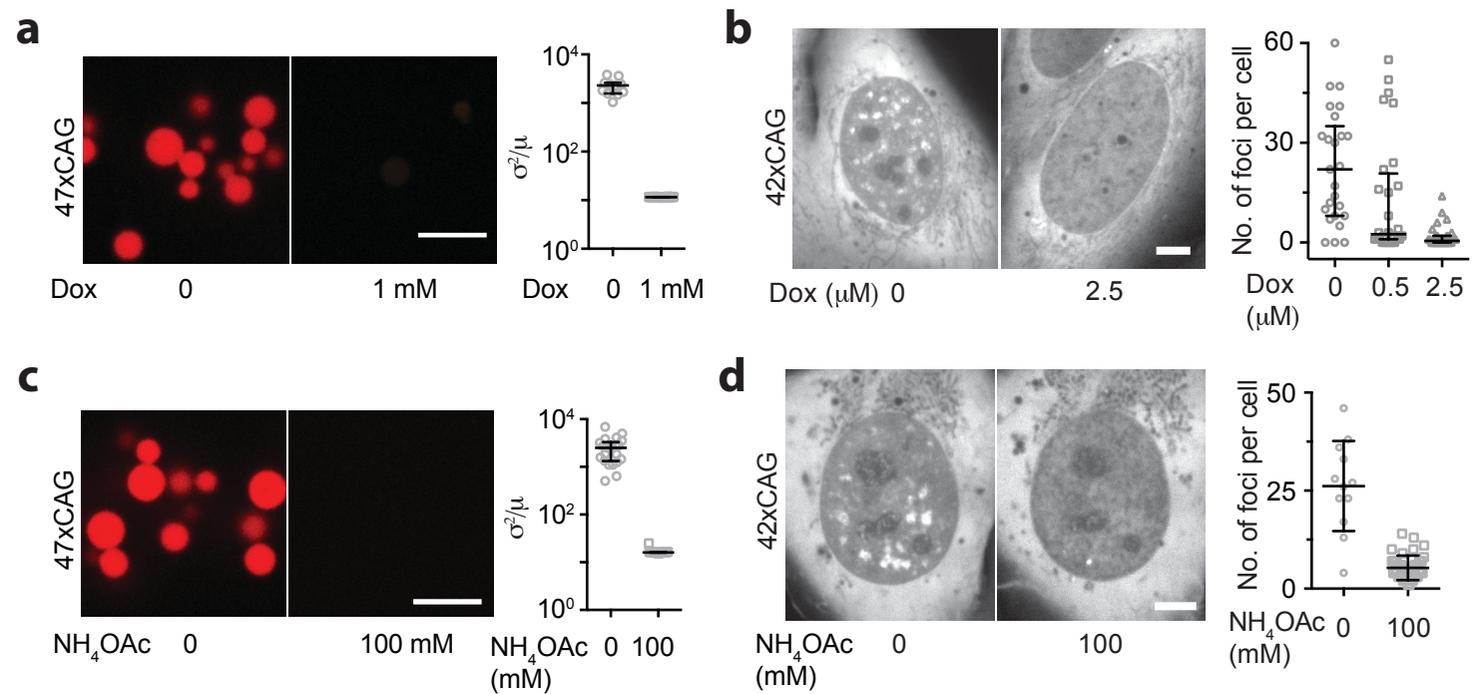
**Figure 1. Triplet repeat containing RNAs undergo gelation *in vitro* at a critical number of repeats.**



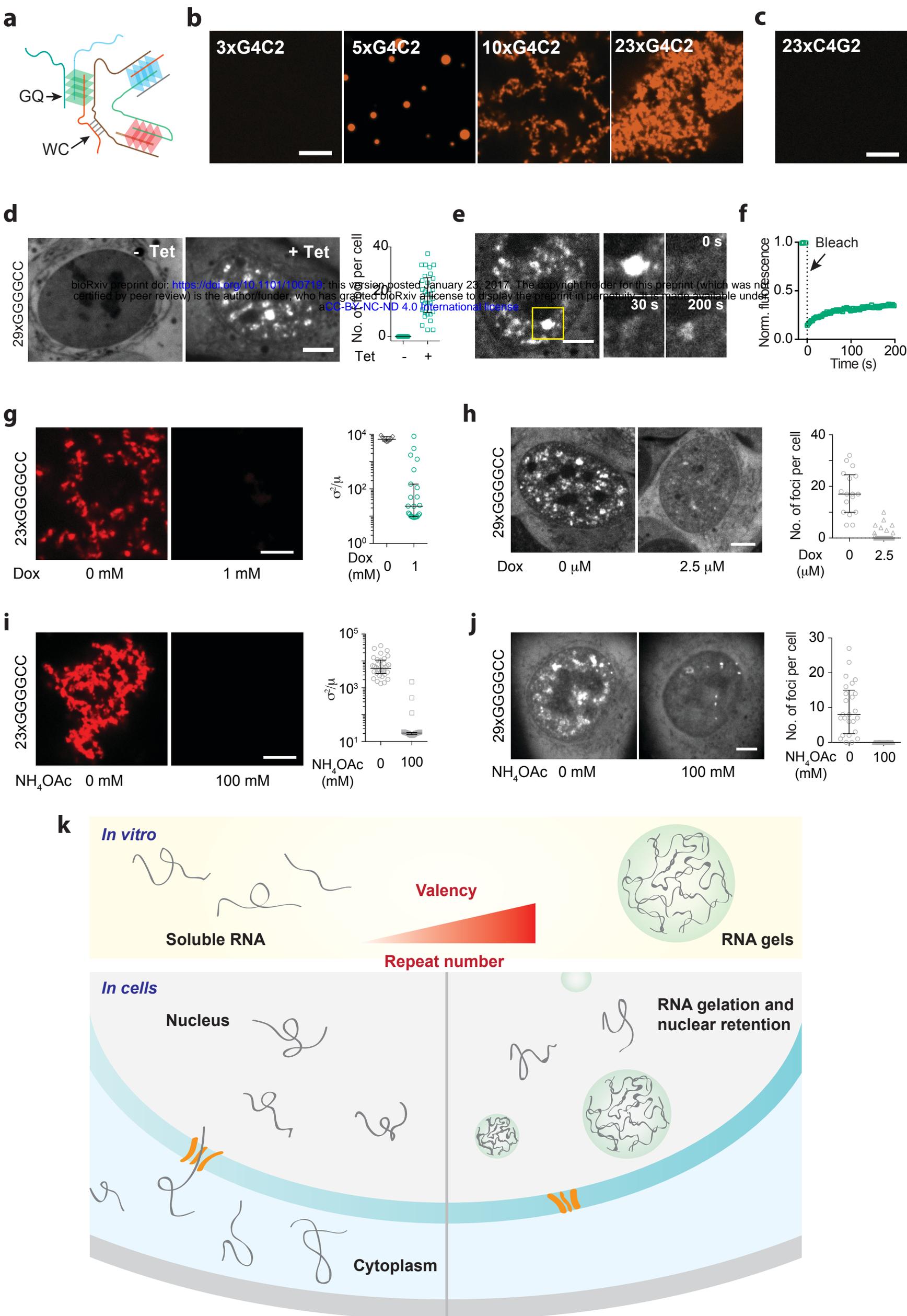
**Figure 2. CAG repeat-containing RNAs coalesce in to liquid-like nuclear foci.**



**Figure 3. Nuclear foci are reversed by agents that disrupt RNA gelation *in vitro*.**



**Figure 4. Gelation of GGGGCC repeat containing RNA and model for RNA foci formation.**



bioRxiv preprint doi: <https://doi.org/10.1101/100719>; this version posted January 23, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.