Nucleation landscape of biomolecular condensates

Shunsuke Shimobayashi, Pierre Ronceray, David Sanders, Mikko Haataja, Clifford Brangwynne

▶ **To cite this version:**

Shunsuke Shimobayashi, Pierre Ronceray, David Sanders, Mikko Haataja, Clifford Brangwynne. Nucleation landscape of biomolecular condensates. Nature, 2021, 599 (7885), pp.503-506. 10.1038/s41586-021-03905-5 . hal-03540908

HAL Id: hal-03540908
https://hal.science/hal-03540908

Submitted on 17 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Nucleation landscape of biomolecular condensates

Shunsuke F. Shimobayashi[1], Pierre Ronceray[2,3], David W. Sanders[1], Mikko P. Haataja,[4,5], and Clifford P. Brangwynne[1,5,6,*]

[1]Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ 08544, USA
[2]Center for the Physics of Biological Function, Princeton University, Princeton, NJ 08544, USA
[3]Aix Marseille Univ, Université de Toulon, CNRS, CPT, Turing Center for Living Systems, Marseille, France
[4]Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, USA
[5]Princeton Institute for the Science and Technology of Materials
[6]Howard Hughes Medical Institute, Princeton University, Princeton, NJ 08544, USA
[*]cbrangwy@princeton.edu

Forming at the right place and time is important for all structures within living cells. This includes condensates such as the nucleolus, Cajal bodies, and stress granules, which form via liquid-liquid phase separation (LLPS) of biomolecules, particularly proteins enriched in intrinsically-disordered regions (IDRs)[1,2] (Fig. 1a). In non-living systems, the initial stages of nucleated phase separation arise when thermal fluctuations overcome an energy barrier due to surface tension. This phenomenon can be modeled by classical nucleation theory (CNT), which describes how the rate of droplet nucleation depends on the degree of supersaturation, while the location at which droplets appear is controlled by interfacial heterogeneities[3,4]. In living cells, however, it is unknown whether this framework applies, due to the multicomponent and highly complex nature of the intracellular environment, including the presence of diverse IDRs, whose specificity of biomolecular interactions is unclear[5–8]. Here we show that despite this complexity, nucleation within living cells occurs through a physical process not unlike that within inanimate materials, but where the efficacy of nucleation sites can be tuned by their biomolecular features. By quantitatively characterizing the nucleation kinetics of endogenous and biomimetic condensates within living cells, we found that key features of condensate nucleation can be quantitatively understood through a CNT-like theoretical framework. Nucleation rates can be significantly enhanced by compatible biomolecular (IDR) seeds, while the kinetics of cellular processes can impact condensate nucleation rates and location specificity. This quantitative framework sheds light on the intracellular nucleation landscape, and paves the way for engineering synthetic condensates precisely positioned in space and time.

To quantify condensate nucleation in living cells, we measure their nucleation rate J, defined as the number of condensates formed per unit volume and per unit time. (Extended Data Fig. 1). Specifically, $J = d\rho(t)/dt|_{t=t_0}$, where $\rho(t)$ and t_0 denote the observed number of visible droplets per unit volume and a delay time after which droplets first emerge, respectively. For various endogenous condensates, we find nucleation rates are in the range of 10^{-4} to 10^{-6} $[\mu m^{-3}s^{-1}]$ (Extended Data Fig. 1b). To probe the molecular biophysical determinants of intracellular condensate nucleation in a tractable system, we take advantage of a recently developed optogenetic tool, Corelets, which allows the spatial and temporal control of intracellular LLPS using light[9]. Corelets are comprised of a two-module optogenetic system built around a light-activatable core, that mimics the endogenous oligomerization of phase separating proteins, often enriched in IDRs (Fig. 1b,c). Consistent with recent work[9–11], we find that light-activated oligomerization drives phase separation of constructs including FUS or HNRNPA1 IDRs, model aromatic-rich polar sequences[7,8,12], forming liquid condensates in human osteosarcoma (U2OS) cells; the nucleation probability is near zero within nucleoli, due to their intrinsic exclusion of Corelets (Extended data Fig. 2a)[13].

Two different phase separation modes are observed: nucleation and growth (NG), and connected network-like growth and coarsening akin to spinodal decomposition (SD) (Fig. 1d, Extended Data Fig. 3 and Supplementary Videos 1-3)[9]. These distinct behaviors reflect the location of the cell within the phase diagram, as can be seen in our experimentally measured phase diagram of FUS_N-Corelets as functions of mean Core concentration and Core-to-IDR ratio, f, (Fig. 1g); the core-to-IDR ratio represents the degree of IDR oligomerization, which impacts the entropic contribution to the free energy[14], and is thus akin to temperature. Weakly supersaturated cells lying close to the binodal only display visible condensates after ∼80 seconds of activation, while cells close to the spinodal region show visible condensates in ∼5 seconds or less, as seen in Fig. 1d,e, Extended Data Fig. 4, and Supplementary Video 1.

To further quantitatively examine this dependence, we plot J vs. the initial supersaturation S, defined as $S = \log(C_{\mathrm{dil}}/C_{\mathrm{sat}})$[3], where C_{dil} is the initial Core concentration in the dilute phase of the nucleus (or cytoplasm) and C_{sat} is the saturating Core concentration (supplemental discussion). Remarkably, we find that S controls J over three orders of magnitude, with a similar dependence for both FUS_N- and $HNRNPA1_C$-Corelets (Fig. 1h). Despite the complexity of the intracellular milieu, this data

can be fit to a functional form predicted by CNT[3] (see also Extended Data Fig. 5):

$$J(S) = \kappa \cdot \exp\left[-\left(\frac{S^*}{S}\right)^2\right], \quad (1)$$

where κ is a kinetic factor proportional to the condensate-specific density of possible nucleation sites, while S^* is a dimensionless parameter that describes the crossover supersaturation between activated and diffusion-limited regimes of nucleation. At low supersaturation $S < S^*$, nucleation is an activated process which is strongly dependent on the supersaturation. By contrast, for $S \gg S^*$, $J \simeq \kappa$, as the nucleation rate is controlled by microscopic transport processes and the limited density of condensate nucleation sites[3] (supplemental discussion). Finally, we quantified the dependence of the delay time t_0 on supersaturation and found $t_0 \sim S^{-\xi}$, with $\xi \approx 1.2$ for FUS_N- and $\xi \approx 1.7$ for $HNRNPA1_C$-Corelets, likely reflecting surface reaction limited growth of the critical nucleus ($\xi = 2$)[15] (Extended Data Fig. 6).

To understand the microscopic origin of S^*, we note that for inanimate systems, CNT predicts $S^* \sim v(\gamma/k_BT)^{3/2}$, where v, γ, and k_BT denote molecular volume, surface tension between the droplet and the surrounding fluid, and thermal energy (with k_B the Boltzmann constant), respectively. Furthermore, surface tension in inanimate systems strongly depends on the position of the system in the phase diagram: close to a critical point, $\gamma \to 0$[16,17], implying that $S^* \to 0$. Consistent with S^* being strongly affected by surface tension, we indeed find that S^* decreases monotonically as a function of the Core-to-IDR ratio for FUS_N and $HNRNPA1_C$-Corelets (Fig. 1f, Extended Data Fig. 7,5). Assuming that the Core-to-IDR ratio f acts as an effective temperature around the critical point, theory predicts[17] that $S^* \propto |f - f_c|^{0.84}$[17]. The best fit yields a critical Core-to-IDR ratio ($f_c^{FUS_N} = 0.18 \pm 0.02$ and $f_c^{HNRNPA1_C} = 0.12 \pm 0.01$), consistent with the phase diagrams in Fig. 1g and Extended Data Fig. 4.

The dependence of J on S in Eq. 1 corresponds to a homogeneous system where all possible nucleation sites are equivalent – i.e., spatially nonspecific droplet formation, consistent with the roughly flat nucleation probability outside of nucleoli (Extended Data Fig. 2). However, a broad variety of condensates, such as those involved in transcription[18–21] or DNA repair[22], assemble at a specific locus of protein, DNA and/or RNA, suggesting that biomolecules can function as "seeds"[23,24]. To parse the potential ability of the seeds to spatially modulate S^* – and thus to impart a more complex nucleation landscape – we use a convenient and engineerable chromatin locus, the telomere, which can be modified by fusing telomeric repeat-binding protein TRF1 to various proteins, including IDRs (Fig. 2a). When FUS_N-TRF1 was co-expressed with FUS_N-Corelets, light activated FUS_N condensates nucleated and grew at telomeres, as expected for IDR-IDR interaction between the condensates and telomere associated seeds (Fig. 2b, Extended data Fig. 8 and Supplementary Video 4).

To examine the effects of seeds on S^*, we again measured J as a function of S for FUS_N-Corelets with FUS_N (WT)-seeds. Interestingly, seeding results in a lower value of κ (indicating fewer nucleation sites) and a lower value of S^* (indicating facilitated nucleation), compared to FUS_N-Corelets without seeds (Fig. 2c). Quantitatively, we find $S^*_{\text{FUS, seed(WT)}}$ ($= 0.44 \pm 0.04$), compared to S^*_{FUS} ($= 0.97 \pm 0.08$) (see also Extended Data Fig. 5). Thus, favorable molecular interactions at specific intracellular seeds can catalyze the nucleation process, by locally reducing S^*. Note that at high supersaturation, the driving force for phase separation is so large that the nucleation landscape is effectively "blurred out", seeding becomes unimportant, and the $J(S)$ curve overlaps with the unseeded one (Fig. 2c, Extended Data Fig. 2b, Supplementary Video 5).

We quantify seeding efficacy by measuring the degree of co-localization of condensates and seeds using the Pearson correlation coefficient (PCC) that measures linear correlation between two images[25], which reveals a monotonic decrease of nucleation specificity as the supersaturation S increases (Fig. 2d). This effect can be understood by again examining Eq. 1, now considering a nucleation landscape consisting of two populations of loci. Nucleation at specific (s) sites is characterized by nucleation parameters $S^{*;s}$ and κ^s, while non-specific (ns) nucleation corresponds to κ^{ns} and $S^{*;ns}$, with $S^{*;s} < S^{*;ns}$. This allows us to define a specificity index $\Delta(S) = \frac{J_{\text{specific}}(S)}{J_{\text{total}}(S)}$, which indicates the fraction of droplets that form at specific seeds (see supplemental discussion). Facilitated nucleation, with $\Delta(S) > 0.5$ occurs for supersaturations $S < S^\dagger$, where $S^\dagger = \sqrt{\frac{(S^{*;ns})^2 - (S^{*;s})^2}{\log(\kappa^{ns}/\kappa^s)}}$ indicates the supersaturation at which the specific and non-specific nucleation curves first overlap (Fig. 2c, $S^\dagger \approx 0.69$ for FUS_N-Corelets); we note that in order for nucleation to proceed with both rapidity and specificity, $S^{*;s} < S^\dagger$ (Fig. 2c, "Strong Seeding"). Using fitted values for the nucleation parameters, we obtain a predicted curve for the specificity index whose variations are consistent with that of the PCC (Fig. 2d). Thus, in order to nucleate with high specificity ($\Delta \to 1$), cells must not only provide energetically favorable seeding sites (i.e., low values of $S^{*;s}$), but also maintain a low degree of supersaturation.

To gain further insight into the biomolecular determinants of $S^{*,s}$, we mutated the telomere-bound FUS_N sequence, changing the number of its tyrosine residues, a key amino acid mediating IDR-IDR interactions[7–9,11,12]. Specifically, we mutated n (= 5, 15 or 27) out of 27 tyrosine (Y) into serine (S) in FUS_N-TRF1, referred to as $FUS_N^{-n\text{YS}}$-seeds, and co-expressed with FUS_N-Corelets. Upon blue light activation, FUS_N condensates still nucleate at the seeds in cells which express FUS_N-Corelets plus FUS_N-5YS seeds (Extended Data Fig. 9), similar to the behavior in unmutated FUS_N seeds. However, with either 15 or 27

tyrosines mutated, the FUS_N seeds no longer nucleated FUS_N-Corelets (Fig. 2b and Extended Data Fig. 9), as can be seen in the near-zero PCC values for all supersaturation levels (Fig. 2d). Consistent with this sensitive dependence of the detailed IDR sequence in nucleation sites, telomere-fused HNRNPA1 IDR, which is known to promote phase separation and can interact with FUS[12], was similarly incapable of nucleating FUS condensates (Extended Data Fig. 9 and Supplementary Video 6). By plotting J against S for FUS_N^{-nYS}-seeds for $n = 15$ or 27 or $HNRNPA1_C$-seeds, all yield data roughly consistent with Eq. 1 with the best-fit S^* and κ for FUS_N-Corelets without seeds (Extended Data Fig. 9d). Collectively, these results suggest that $S^{*,s}$ exhibits a dependence on the biomolecular features of the seed.

Given the biophysical picture that emerges from examining these model condensates, we sought to evaluate its applicability in a system that more closely mimics an endogenous condensate. To this end, we utilized the Corelet system with the stress granule protein G3BP1, whose native oligomerization domain is deleted and replaced with the light activatable sspB module [6]. When expressed in G3BP knock-out cells, this leads to light-activated synthetic stress granules ("opto-SGs"), which recapitulate essentially all features of endogenous stress granules, including recruitment of other key SG proteins, and a strong dependence of their phase behavior on RNA concentration[6]. Consistent with this previous work, upon flooding the cell with exposed RNA by treatment with Sodium arsenite (+As), opto-SGs form more readily than in untreated cells (Fig. 3a). Quantifying the nucleation rate as a function of supersaturation, we find that both +As and -As cells appear to fall on the same curve, which is well-described by Eq. 1, with a lower value of S^* and prefactor compared with FUS_N-Corelets (Fig. 3b, Extended Data Fig. 5). The lower S^* likely reflects the impact of heterogeneous SG nucleation with endogenous binding partners (Extended Data Fig. 10a), again underscoring the biomolecular specificity of heterogenous condensate nucleation. To probe time-dependent changes in nucleation caused by Arsenite treatment, we sequentially activated and then deactivated cells immediately after Arsenite addition. Consistent with the degree of supersaturation increasing with each successive quench, due to the steady increase in exposed RNA concentration, the nucleation rate increases over the course of roughly 10-60 min (Fig. 3c,d). This corresponds to the system moving along the J vs. S curve in time (Fig. 3d), reflecting an interplay of nonequilibrium biological processes with our CNT-like framework.

An intriguing difference between the opto-SGs and endogenous SGs is that the nucleation rate of endogenous SGs is 10^{-4} $[\mu m^{-3} s^{-1}]$, which is only comparable to that of the opto-SGs at the earliest times after Arsenite treatment (Fig. 3d), when the system is only modestly supersaturated upon light-stimulated quenching. Indeed, the nucleation rates for endogenous SGs remain low across a range of larger nominal supersaturation $S^{\mathrm{nom}} = \log(C_{\mathrm{tot}}/C_{\mathrm{sat}})$ values (Fig. 3b). Nonetheless, opto-SGs recapitulate most other aspects of endogenous SGs[6], leading us to speculate that the endogenous condensates form at unexpectedly low values of effective supersaturation; the apparent supersaturation for SG nucleation can be determined by the intersection of the J-S curve for opto-SGs and the dashed horizontal line for endogenous stress granules (Fig. 3b), resulting in $S_{\mathrm{nuc}}^{SG} \sim 0.1$. We reasoned that this low apparent supersaturation could potentially reflect a competition between the kinetics of nucleation and the kinetics of Arsenite-induced RNA-release. To test this picture, we examined the nucleation rates of our opto-SG system under conditions of constant illumination, more closely mimicking endogenous SGs, where G3BP is oligomerized throughout the duration of Arsenite-treatment. Remarkably, under these conditions, the opto-SG nucleation rate upon Arsenite treatment becomes nearly identical to that of endogenous SGs (Fig. 3b).

We can understand this kinetic effect by noting that any real physical system does not become instantaneously supersaturated, but instead is driven into a supersaturated state at a finite "quench" rate α, i.e., $S^{\mathrm{nom}}(t) = \alpha t$. As the supersaturation steadily increases, nucleation commences at a certain supersaturation $S = S_{\mathrm{nuc}}$, which is related to α via $\alpha \approx V_{\mathrm{cell}} \cdot S_{\mathrm{nuc}} \cdot J(S_{\mathrm{nuc}})$ (Fig. 3e and supplemental discussion). This causes a decrease in the dilute phase concentration (and hence a decrease in the supersaturation), thereby suppressing further nucleation. Consistent with this picture, in addition to endogenous SGs, we find that Cajal bodies, DNA repair condensates, and nucleoli all exhibit nucleation rates that are independent of the nominal supersaturation (S^{nom})(Extended Data Fig. 10b), suggesting that the relatively slow kinetics of the governing biological processes control nucleation rates, by maintaining cells in states of low supersaturation. Interestingly, this effect is intimately coupled to the specificity of the nucleation process: under slow quench rates, such that $\alpha \ll V_{\mathrm{cell}} \cdot S^{\dagger} \cdot J(S^{\dagger})$, nucleation will only occur at specific nucleation sites, while fast quench rates, for which $\alpha \gg V_{\mathrm{cell}} \cdot S^{\dagger} \cdot J(S^{\dagger})$, will lead to rapid and non-specific nucleation. Thus, the timescales of biological processes governing quench rates (e.g., protein phosphorylation, RNA transcription/release, protein transport etc.) are intimately coupled to the subcellular locations at which condensates form.

In conclusion, we have proposed a general framework to quantify and interpret the rate and specificity of droplet nucleation for synthetic and endogenous biomolecular condensates. Despite the complexity of cells, we have shown that the scaling form predicted by classical nucleation theory (CNT, Eq. 1) can be fruitfully employed to interpret the nucleation rate and its dependence on supersaturation. Biomolecular and mechanical heterogeneities within the cell result in a spatially varying nucleation landscape (i.e. S^*) governing condensate formation. We have here focused on simple nucleation landscapes characterized by only one or two values of S^* – the minimum required to control the specificity of nucleation – but in full generality the spatial distribution of S^* is dictated by a nucleation landscape that is continuous (Fig. 3e, (supplemental discussion). Nonetheless, this simplification facilitates a mathematical criterion that cells must meet in order to ensure specific

nucleation. Such specific nucleation is mediated by favorable interactions between seeds and condensates, which we have shown are strongly dependent on molecular features, including amino acid patterning of IDRs. Nucleation specificity is likely crucial for processes such as DNA repair and gene expression, wherein condensates must assemble at particular locations. Finally, the framework we have introduced will help guide efforts to design intracellular condensates for various bioengineering applications, where targeting their nucleation to specific spatial location within the cell may strongly impact their engineered functions.

References

1. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaar4382, DOI: 10.1126/science.aaf4382 (2017).
2. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298, DOI: 10.1038/nrm.2017.7 (2017).
3. Kashchiev, D. *Nucleation* (Elsevier, 2000).
4. Auer, S. & Frenkel, D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature* **409**, 1020–1023, DOI: https://doi.org/10.1038/35059035 (2001).
5. Mitrea, D. M. *et al.* Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying r-rich linear motifs and rrna. *eLife* **5**, 1–33, DOI: 10.7554/elife.13571 (2016).
6. Sanders, D. W. *et al.* Competing protein-RNA interaction networks control multiphase intracellular organization. *Cell* **181**, 306–324.e28, DOI: 10.1016/j.cell.2020.03.050 (2020).
7. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **367**, 694–699, DOI: 10.1126/science.aaw8653 (2020).
8. Wang, J. *et al.* A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688–699.e16, DOI: 10.1016/j.cell.2018.06.006 (2018).
9. Bracha, D. *et al.* Mapping local and global liquid phase behavior in living cells using photo-oligomerizable seeds. *Cell* **175**, 1467–1480, DOI: 10.1016/j.cell.2018.10.048 (2018).
10. Shin, Y. *et al.* Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* **175**, 1481–1491.e13, DOI: 10.1016/j.cell.2018.10.057 (2018).
11. Wei, M. T. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol.* **22**, 1187–1196, DOI: 10.1038/s41556-020-00578-6 (2020).
12. Kato, M. *et al.* Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767, DOI: 10.1016/j.cell.2012.04.017 (2012).
13. Riback, J. A. *et al.* Composition-dependent thermodynamics of intracellular phase separation. *Nature* **581**, 209–214, DOI: 10.1038/s41586-020-2256-2 (2020).
14. Rubinstein, M. & Colby, R. H. *Polymer Physics* (Oxford Univ. Press, 2003).
15. Clouet, E. *Modeling of Nucleation Processes* (ASM Handbook, 2009).
16. Honerkamp-Smith, A. R., Veatch, S. L. & Keller, S. L. An introduction to critical points for biophysicists; observations of compositional heterogeneity in lipid membranes. *Biochimica et Biophys. Acta - Biomembr.* **1788**, 53–63, DOI: 10.1016/j.bbamem.2008.09.010 (2009).
17. Mayoral, E. & Goicochea, A. G. Hyperscaling relationship between the interfacial tension of liquids and their correlation length near the critical point. *Soft Matter* **10**, 9054–9058, DOI: 10.1039/c4sm01981d (2014).
18. Lafontaine, D. L., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* DOI: 10.1038/s41580-020-0272-6 (2020).
19. Cho, W. K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415, DOI: 10.1126/science.aar4199 (2018).
20. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555, DOI: 10.1126/science.aar2555 (2018).
21. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958, DOI: 10.1126/science.aar3958 (2018).

22. Kilic, S. *et al.* Phase separation of 53 BP1 determines liquid-like behavior of DNA repair compartments. *The EMBO J.* **38**, 1–17, DOI: 10.15252/embj.2018101379 (2019).

23. Shevtsov, S. P. & Dundr, M. Nucleation of nuclear bodies by RNA. *Nat. Cell Biol.* **13**, 167–173, DOI: 10.1038/ncb2157 (2011).

24. Berry, J., Weber, S. C., Vaidya, N., Haataja, M. & Brangwynne, C. P. RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci.* **112**, E5237 – E5245, DOI: 10.1073/pnas.1509317112 (2015).

25. Aaron, J. S., Taylor, A. B. & Chew, T. L. Image co-localization - co-occurrence versus correlation. *J. Cell Sci.* **131**, 1–10, DOI: 10.1242/jcs.211847 (2018).

Methods

Cell culture

Human cell lines used in this study include U2OS (Sex: Female) and Lenti-X 293T (Takara Bio USA, Sex: Female). Lenti-X 293T cells were only used for virus production, while U2OS cells were used for experiments. Cells were cultured in growth medium (DMEM, GIBCO) supplemented with 10% FBS (Atlanta Biological) and penicillin and streptomycin (GIBCO) at 37°C with 5% CO2 in a humidified incubator. All cell lines were validated by STR profiling (ATCC) with 100% match between submitted samples and database profiles and were tested negative for mycoplasma (method: Universal Mycoplasma Detection Kit, ATCC® 30-1012K).

Plasmid construction

DNA constructs used for tissue culture were cloned using In-Fusion HD cloning kit (Clonetech) in a standard reaction mixture containing 50 ng of PCR-amplified inserts and 20 ng of linearized pHR-SFFV backbone in a 5 µl reaction set to 50°C for 15 min. PCR fragments were produced using a standard PCR reaction using Phusion High-Fidelity DNA Polymerase (NEB). PCR products were verified on an agarose gel and purified using PCR extraction kit (QIAGEN). Plasmids were transformed into Stellar cells (Clontech), from which single colonies were picked, grown in LB supplemented by Ampicillin for 16 hours, and miniprepped (QIAGEN) following manufacturer instruction. All cloning products were confirmed by sequencing (GENEWIZ). pHR-FUS_N (1-214, WT)-miRFP-TRF1 was first generated by inserting DNA fragments encoding FUS_N and miRFP670-TRF1[10] into a pHR-based vector backbone (pHR- FUS_N-mCh-sspb, Addgene 122148) linearized by MluI-HF (NEB) and SbfI-HF (NEB) using following manufacturer's instruction. For other FUS_N-nYS-miRFP-TRF1 (n=5, 15, 27) and $HNRNPA1_C$-miRFP-TRF1 constructs, FUS_N in FUS_N-miRFP-TRF1 was swapped out for the DNA sequence encoding FUS_N-5YS, FUS_N-15YS, FUS_N-27YS, $HNRNPA1_C$[9]. FUS_N-5YS harbored tyrosine to serine mutations at Y17, Y75, Y81, Y143, Y149. FUS_N-15YS harbored tyrosine to serine mutations at Y14, Y17, Y33, Y38, Y41, Y55, Y58, Y91, Y97, Y100, Y130, Y143, Y149, Y155, Y161. FUS_N-27YS harbored tyrosine to serine mutations at Y6, Y14, Y17, Y25, Y33, Y38, Y41, Y50, Y55, Y58, Y66, Y75, Y81, Y91, Y97, Y100, Y113, Y122, Y130, Y136, Y143, Y149, Y155, Y161, Y177, Y194. The other constructs (pHR-FUS_N-mCherry-sspb, pHR-$HNRNPA1_C$-mCherry-sspb, pHR-NLS-iLID-EGFP-FTH1, pHR-iLID-EGFP-FTH1, FM5-NPM1-GFP, FM5-Coilin-EYFP, FM5-mCherry-G3BP1, pHR-H2B-miRFP670, pHR-mGFP-P2A-mCherry, FM5-sspB-mCherry-G3BP1 DelNTF2, FM5-FXR1-miRFP, FM5-UBAP2L-miRFP, FM5-miRFP-LSM14A, FM5-iLID-mGFP-FTH1) were generated in our previous studies[6,9,10,13].

Lentiviral transduction

Lentiviruses were produced by cotransfecting Lenti-X 293T cells grown to approximately 70% confluency in 6-well plates with the transfer plasmids (1.5 µg), pCMV-dR8.91 (1.33 µg) and pMD2.G (0.17 µg) using FuGENE HD Transfection Reagent (Promega) following manufacturer's instruction. After 2 days, 2 mL of supernatant containing viral particles was harvested and filtered with 0.45 µm filter (VWR). Supernatant was immediately used for transduction or stored at -80°C in aliquots.

Construction of stable cell lines

U2OS cells were grown to 10-20% confluency on 35-mm glass-bottom dishes (MatTek) and 10-500 µl of filtered viral supernatant was added to the cells. Virus-containing medium was replaced with fresh growth medium 48 hours post-infection. Cells infected were typically imaged no earlier than 72 hr after infection.

Arsenite treatment on G3BP1 expressing cells

G3BP1/G3BP2 knockout U2OS cells[6], expressing FM5-mCherry-G3BP1 or G3BP1 Corelet (FM5-iLID-mGFP-FTH1 and FM5-sspB-mCherry-G3BP1 DelNTF2) via lentiviral transfection, were treated to a total concentration of 400 µM sodium Arsenite (Sigma-Aldrich).

Bleomycin treatment on 53BP1 expressing cells

Bleomycin (Sigma-Aldrich) was dissolved in PBS at 10 mg/mL and further diluted in growth medium to a final concentration of 10 μg/ml before being applied to U2OS cells expressing FM5-miRFP-53BP1 via lentiviral transfection.

Corelets (Core scaffolds to promote droplets)

Corelets have a two-module optogenetic system that mimics the endogenous oligomerization of IDR-rich proteins to drive endogenous nucleation of membrane-less condensates, using a light-activatable high valency core[9]. The core is comprised of 24 human ferritin heavy chain (FTH1) protein subunits, which self-assemble to form a spherical particle of 12 nm diameter (referred to as "Core"), which is fused to a nuclear localization signal (NLS) and an engineered protein iLID. The second module is comprised of a self-interacting IDR fused to SspB, which upon blue light activation strongly heterodimerize ($K_d \sim$ 130 nM) with iLID to form self-interacting particles.

Microscopy

Fluorescence images were taken using a spinning-disk confocal microscope (Yokokgawa CSU-X1) with a Nikon 100x oil immersion objective (CFI Apo TRIF, NA 1.49) and an Andor iXon Ultra DU-897 EMCCD camera on a Nikon Eclipse Ti-E body. Samples were maintained at 37°C and 5% CO2 with a stage top incubator (Okolab). A laser combiner (Nikon LU-NV) was used to excite mGFP and EYFP and for activating iLID with 488 nm laser light at an excitation power of approximately 0.1 W/cm2 as measured with a microscope slide photodiode power sensor (PM100D, Thorlabs). 561 and 640 nm lasers were used for imaging mCherry and miRFP, respectively. Phase separation dynamics for the Corelet system were captured by imaging in the mCherry channel with global activation performed by the 488 nm laser line. Telomere-targeted seeds were captured in the miRFP channel. In the case where a fast frame rate was desirable, a frame interval of 60 ms was used with a 256x256 pixel ROI. All image acquisition was performed using Nikon NIS-Elements AR software. All image analysis was performed using ImageJ or custom-built MATLAB scripts. A laser scanning confocal microscope (Nikon A1) was used only for fluorescence correlation spectroscopy (FCS) experiments.

Absolute concentrations for Corelets components

Absolute concentrations for Corelets components, i.e., Core and IDR, were estimated using FCS, as reported previously[9]. Data for concentration of proteins were obtained using 30 s FCS measurement time. The measurements were performed on HEK293 cells expressing pHR- FUS_N-mCh-sspb using a Nikon A1 with an oil immersion objective (Plan Apo 60X/1.4). All measurements and data analysis were performed using the SymPhoTime Software (PicoQuant). The autocorrelation function for simple diffusion is:

$$G(\lambda) = G(0)\left(1+\frac{\lambda}{\lambda_D}\right)^{-1}\left(1+\frac{\lambda}{\eta^2\lambda_D}\right)^{-1/2}, \tag{2}$$

here $G(0)$ is the magnitude at short timescales, λ is the lag time, λ_D is the half decay time, and η is the ratio of axial to radial of measurement volume ($\eta = \omega_z/\omega_{xy}$). Here, $\omega_{xy} = 2.3$ and $\eta = 3.1$ which is determined by fluophore dye Atto 550 in water. The parameters $G(0)$ and λ_D are optimized in the fit and are used to determine the molecule concentration (C= ($\pi^{3/2}$ ω_{xy}^2 $\omega_z G(0)$)$^{-1}$). mCherry fluorescence was converted to absolute concentration using FCS. GFP fluorescence conversion was done by determining the mCherry-to-GFP fluorescence ratio, which was determined by equimolar expression of mCherry and GFP monomers in HEK293 cell using auto-catalytic P2A containing construct mCherry-P2A-EGFP. Then, further fluorescence calibration was done between the Nikon A1 microscope and spinning-disk confocal microscope on a Nikon Eclipse Ti-E body (Yokokgawa CSU-X1), using WT U2OS cells expressing mCherry-P2A-EGFP.

Phase diagram construction

The nucleus or cytoplasm boundary in each cell was determined by applying an automated image segmentation MATLAB code for the Gaussian-filtered image of Core component (EGFP channel) before phase separation. Then, mean EGFP and mCherry fluorescence within the segmented region were background-subtracted and translated to absolute concentration via the FCS-based concentration estimation, thus allowing to estimate mean Core absolute concentration and Core-to-IDR ratio. The mean values are shown in Fig. 1g and Extended Data Fig. 4b. The binodal boundary is defined as the boundary to separate if any visible droplets are observed within our experimental time-scale ($\sim$ 5 min) after light-activation. The spinodal region is defined as the region in which the interconnected network-like growth is observed, as shown in Extended Data 3.

Quantification of nucleation rate

To quantify the nucleation rate J, the number of droplets per unit volume, i.e., droplet density ρ, was first measured as a function of time, t. The nucleus or cytoplasm boundary in each cell was determined by applying an automated image segmentation

MATLAB code for the Gaussian-filtered image of Core component (EGFP channel) before nucleation. Then, the droplets were automatically segmented by a single intensity threshold based on the bimodal fluorescent histogram of IDR component (mCherry channel) within the segmented nucleus or cytoplasm. The droplet number was then counted. The depth of focus ($\sim$1 μm) of confocal images was used to estimate the analyzed volume. For Fig. 1e and Extended Data 3b, the measures of centre for the shaded error bars are the mean and the error bars show standard deviation. For Fig.3c(ii), the data points are the mean. The error bars are the standard deviation. The nucleation rate was obtained as

$$J = \frac{d\rho}{dt}, \tag{3}$$

at $t = t_0$ from fitting the data with the equation

$$\rho(t) = \rho_0(1 - e^{-\frac{t-t_0}{\tau}}). \tag{4}$$

The data points in Fig. 1h, 2c, 3b and Extended Data Fig. 1b(ii), 7, 9d, and 10a are the best fit parameters. The error bars are the standard errors with 95% confidence intervals of the fits. For Fig. 3d, the data points and error bars are the mean and standard deviation.

Quantification of supersaturation

Supersaturation is defined as $S = \log(C_{\mathrm{dil}}/C_{\mathrm{sat}})$, where C_{dil} is the initial Core concentration in the dilute phase of the nucleus (or cytoplasm) and C_{sat} is the saturating Core concentration, determined from the dilute phase after nucleation ends, since $C_{\mathrm{sat}} \approx C_{\mathrm{dil}}$ in the steady state. For the quantification of supersaturation, the nucleus or cytoplasm boundary in each cell was first determined by applying an automated image segmentation MATLAB code for the Gaussian-filtered image of Core component (EGFP channel) before nucleation. The initial Core concentration C_{dil} was estimated from the average fluorescence intensity of the background-subtracted image in the segmented region using the FCS-based calibration. For estimating the Core concentration in the diluted phase after nucleation in the steady state, the droplets were automatically segmented by a single intensity threshold based on the bimodal fluorescent histogram of Core component within the segmented nucleus or cytoplasm. To accurately determine the concentration within the dilute phase, morphological erosion was performed for the segmented droplets, such that three pixels ($\approx 0.48\ \mu m$) near the droplet interface were excluded from the analysis. The Core concentration in the diluted phase was estimated from the average fluorescence intensity of the background-subtracted image in the segmented diluted region using the FCS-based calibration. The nominal supersaturation $S^{\mathrm{nom}} = \log(C_{\mathrm{tot}}/C_{\mathrm{sat}})$ in Fig. 3b and Extended Data 10b was estimated from the concentrations of the total composition and diluted phase of a reporter protein (ex. NPM1-mGFP for nucleoli) after the nucleation process ends. The error bars in Fig. 2h, 3c and Extended Data 7, 10b are the standard deviation.

Model fit

The data of J against S were fit to a functional form predicted by CNT (Eq. 1) by minimizing chi-squared χ^2. For Fig. 1h, 2c, and 4b, the model fit parameters (i.e., S^* and κ) and reduced chi-squared χ^2_{red} values are shown in Extended Data Fig. 5.

Quantification of relative nucleation probability

To quantify the relative nucleation probability $p^{\mathrm{rel}}_{\mathrm{nuc}}(\mathbf{x})$ in Extended Data 2, defined by the number of formed droplets per unit volume at position $\mathbf{x}$, U2OS cells expressing $\mathrm{FUS_N}$ Corelets w/wo $\mathrm{FUS_N}$ (WT)-miRFP-TRF1 were sequentially light-activated five times. The image where the nucleation is saturated after 9.9 sec in (a), 58.7 sec (b-(i))), and 32.1 sec (b-(ii)) of light-activation was selected for each activation, and the five images were spatially aligned based on the center of the nucleus. We then counted the total number of formed condensates in a volume of $\sim 5\ [\mu m^3]$ with a focus on a position $\mathbf{x}$ by iterative activation and averaged it in space to calculate $p^{\mathrm{rel}}_{\mathrm{nuc}}(\mathbf{x})$.

Colocalization analysis

The colocalization coefficient between two images was estimated by calculating Pearson's correlation coefficient (PCC) defined as;

$$PCC = \frac{\Sigma_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\Sigma_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{5}$$

where x_i and $\bar{x}$ are the i^{th} pixel intensity and mean values in the segmented first image (i.e. color 1), respectively. Likewise, y_i and $\bar{y}$ are the i^{th} pixel intensity and mean values in the second image (i.e. color 2), respectively. The perfect-correlation, no-correlation, and anti-correlation give the values of PCC = 1, 0, -1, respectively. The correlation in the whole nucleus was

calculated. The data points in Fig. 2d show the mean. The error bars show standard deviation.

Reproducibility
A representative of at least three independent experiments is shown in Fig. 1a,c 3a,b, c(i), Extended Data 1a, 2, 3, 8, 4a, 9, and 10a.

Data availability

Source data for Figs. 1-3 are provided with the paper. All other data are available from the corresponding authors upon reasonable request.

Code availability

Custom code used to process and analyse the images, as detailed in the Methods, are available from the corresponding authors upon reasonable request.

Acknowledgements

We thank members of the Brangwynne laboratory for helpful discussions and comments on this manuscript. The authors are grateful to lab members Yoonji Kim, Mack Walls, and Dan Bracha for providing plasmids. This work was supported by the Howard Hughes Medical Institute, and grants from the NIH 4D Nucleome Program (U01 DA040601), and the Princeton Center for Complex Materials, an MRSEC (NSF DMR-2011750). We thank Princeton's School of Engineering and Applied Sciences for supporting this work through a Focused Research Team award. S.F.S is supported by JSPS Postdoctoral Fellowships for Research Abroad. P.R. is supported by the NSF through the Center for the Physics of Biological Function (PHY-1734030), by the "Investissements d'Avenir" French Government program managed by the French National Research Agency (ANR-16-CONV-0001) and by the Excellence Initiative of Aix-Marseille University - A*MIDEX.

Author contributions statement

S.F.S., P.R., D.W.S., M.P.H. and C.P.B. designed the research; S.F.S. and D.W.S. performed the experiments; S.F.S analyzed the data; P.R. and M.P.H took the lead in the theoretical formalism; S.F.S., P.R., M.P.H. and C.P.B. wrote, and all authors reviewed and edited the paper.

Competing interests

C.P.B. is a founder and consultant of Nereid Therapeutics. S.F.S., P.R., D.W.S., and M.P.H. declare that they have no conflict of interest.

Legends for Extended Data 1-10

Extended Data 1. Nucleation process of various endogenous and biomimietic condensates. (a) Time-lapse images of U2OS cells show the nucleation of Cajal bodies (coilin-EYFP) and nuclear speckles (SRRM1-EYFP) in mitosis, DNA repair condensates (miRFP-53BP1) upon 10 μg/ml bleomycin treatment, and engineered FUS condensates. Scale bars, 10 μm. (b) (i) Number density of condensates, ρ, as a function of time, t for nucleoli and Cajal bodies in mitosis, stress granules (SGs) upon 400 μM As treatment, and 53BP1 condensates upon 10 μg/ml bleomycin treatment. The nucleation rate, J, is quantified by the slope. Here, $t = 0$ is $t_0 - 5$ (min). (ii) Mean nucleation rate, J, and its standard deviation for nucleoli ($n = 6$), Cajal bodies ($n = 9$), SGs ($n = 24$) and DNA repair condensates ($n = 9$); $n =$ number of cells. (c) Schematic diagram of the nucleation process of intracellular condensates.

Extended Data 2. Quasi-1D nucleation probability. (a) (i) Fluorescence images of U2OS cells expressing $\mathrm{HNRNPA1_C}$-Corelets before and after blue light activation. Scale bar, 10 μm. (ii) Droplet number density as a function of time upon iterative blue-light activation and deactivation. (iii) Quasi-1D nucleation probability $p_{\mathrm{nuc}}^{\mathrm{rel}}$ in the indicated region which include nucleoli (A) or not (B). (b) Quasi-1D nucleation probability $p_{\mathrm{nuc}}^{\mathrm{rel}}$ of U2OS cells expressing $\mathrm{FUS_N}$-Corelets and $\mathrm{FUS_N}$-miRFP-TRF1 in the indicated region ((i) low and (ii) high supersaturation), calculated from five successive activation cycles. Scale bar, 10 μm.

Extended Data 3. Photo-activated phase separation in nucleus and cytoplasm. (a) Photo-activated phase separation in nucleus. (a-i) Confocal images of U2OS cells with different expression levels of $\mathrm{FUS_N}$ Corelets ($\mathrm{FUS_N}$ IDRs (red) and Cores (green)) after light-activation. The cells display nucleation and growth (NG) regimes between the binodal boundary and spinodal region, and spinodal decomposition (SD). Scale bars, 10 μm. (a-ii) Connected network-like growth and coarsening akin to spinodal decomposition. Scale bar, 10 μm. (b) Photo-activated phase separation in cytoplasm. (b-i) Time-lapse confocal images of photo-activated U2OS cells expressing $\mathrm{FUS_N}$-Corelets composed of Core without NLS and $\mathrm{FUS_N}$-mCh-sspb. Scale bars, 10 μm. (b-ii) Time change of light-induced droplet number density, ρ. The shaded error bars show standard deviation.

Extended Data 4. Phase separation behaviors of $\mathrm{HNRNPA1_C}$-Corelets. (a) Time-lapse confocal images of photo-activated U2OS cells with different expression levels of $\mathrm{HNRNPA1_C}$-Corelets. Nucleation growth (NG) regime near the binodal boundary (top) and spinodal region (bottom). Scale bars, 10 μm. (b) Phase diagram of $\mathrm{HNRNPA1_C}$-Corelets as functions of Core concentration and Core-to-IDR ratio ($n = 161$). Solid circles exhibit cells where nucleation growth is observed, while empty triangles and squares show cells where no phase separation and spinodal decomposition are observed, respectively. The colors of solid circles indicate the observed nucleation rate, J. n: number of cells.

Extended Data 5. Model fit parameters and chi-square values. (a,b) Model fit parameters of S^*, κ, χ^2_{red} for various biomimetic condensates. The errors show the standard errors with 68% confidence intervals of the fits by Eq. 1. The parameter $\chi^2_{\mathrm{red}} \gg 1$ indicates a poor model fit while $\chi^2_{\mathrm{red}} \lesssim 1$ indicates a good model fit. (c) S^* (top) and κ (bottom) obtained from fits to Eq. 1 for $\mathrm{FUS_N}$ Corelets with ($n = 23$) or without ($n = 76$) $\mathrm{FUS_N}$ (WT) seeds and G3BP1 corelets +/- 400 μM As (n=45). Error bars represent standard fit errors.

Extended Data 6. Delay time against supersaturation. Delay time, t_0, against supersaturation, S, for $\mathrm{FUS_N}$-Corelets ($n = 70$), $\mathrm{HNRNPA1_C}$-Corelets ($n = 60$), and $\mathrm{FUS_N}$ Corelets plus $\mathrm{FUS_N}$ (WT) seeds ($n = 23$) with $1/24 < f < 1/10$. The dashed lines show the best power-law fits. Error bars for t_0 and S show the standard errors of the fits and standard deviation, respectively. n: number of cells.

Extended Data 7. J against S with various ranges of Core-to-IDR ratio. Nucleation rate, J, plotted against the degree of supersaturation, S, for (i) $\mathrm{FUS_N}$-Corelets ($n = 26$ for $1/24 < f < 1/18$, $n = 33$ for $1/18 \leq f < 1/12$ and $n = 56$ for $1/12 \leq f < 1/$ 1/6) and (ii) $\mathrm{HNRNPA1_C}$-Corelets ($n = 12$ for $1/24 < f < 1/18$, $n = 38$ for $1/18 \leq f < 1/12$ and $n = 18$ for $1/12 \leq f < 1/$ 1/6). Error bars for J and S show the standard errors of the fits and standard deviation, respectively. See Extended Data Fig. 5 for model fit parameters and χ^2_{red}. n: number of cells.

Extended Data 8. Seeded nucleation near the binodal boundary and spinodal region. Time-lapse confocal images of U2OS cells expressing $\mathrm{FUS_N}$-Corelets and $\mathrm{FUS_N}$-miRFP-TRF1 near the (a) binodal boundary and (c) spinodal region. The insets show the regions indicated by the arrows. Scale bars, 10 μm. (b) The seeded condensates, which are indicated in (a), are enlarged. Scale bar, 1 μm.

Extended Data 9. Nucleation of $\mathrm{FUS_N}$-Corelets with tyrosine-mutated $\mathrm{FUS_N}$-seeds. (a-c) Fluorescent images of U2OS cells expressing $\mathrm{FUS_N}$-Corelets and $\mathrm{FUS_N}$ (-5YS, -27YS)-miRFP-TRF1 or $\mathrm{HNRNPA1_C}$-miRFP-TRF1 before and after blue

light activation. The insets and enlarged images show the regions indicated by the squares. Scale bars, 10 μm. (d) Nucleation rate, J, plotted against the degree of supersaturation, S, for FUS_N-Corelets ($1/24 < f < 10$) and FUS_N-Corelets ($1/30 < f < 10$) with FUS_N (-15YS, -27YS)-miRFP-TRF1 ($n_{15YS}, n_{27YS} = 12, 5$) or $HNRNPA1_C$-miRFP-TRF1 ($n = 8$). n: number of cells.

Extended Data 10. Specificity of the nucleation of endogenous condensates. (a) Confocal images of G3BP KO cells expressing G3BP1 corelets and iRFP-tagged FXR1/UBAP2L/LSM14 before and after blue light activation in the absence of As. Scale bars, 1 μm. (b) Nucleation rate of endogenous condensates against nominal supersaturation. Nucleation rate, J, plotted against the nominal supersaturation, S^{nom}, for nucleoli ($n = 6$) and Cajal bodies ($n = 9$) in mitosis and 53BP1 condensates ($n = 9$) upon 10 μg/ml bleomycin treatment. n: number of cells.

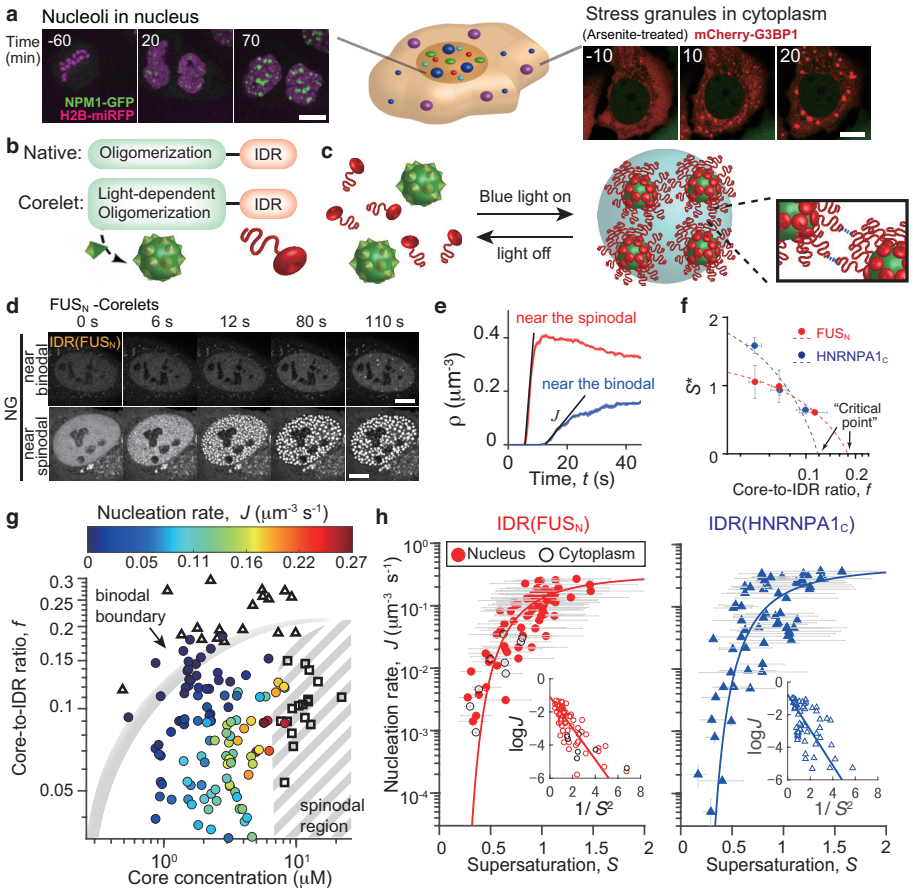


Figure 1. Quantifying the nucleation landscape with synthetic condensates. (a) Time-lapse images of U2OS cells show the nucleation of nucleoli (NPM1-mGFP) and stress granules (SGs, mCherry-G3BP1) upon 400 μM arsenite (As) treatment. Scale bars, 10 μm. (b) Schematic diagram of the Corelet system, which mimics the oligomerization of native condensate forming proteins. (c) Schematic of Corelet phase separation. Upon blue-light illumination, up to 24 IDR modules are captured by each Core, which may subsequently phase separate by multivalent IDR interactions. (d) Time lapse confocal images of photo-activated U2OS cells with different expression levels of FUS_N Corelets. The cells display nucleation and growth (NG) between the binodal boundary and spinodal region. Scale bars, 10 μm. (e) Time change of light-induced droplet density, ρ, for FUS_N Corelets. The nucleation rate, denoted by J, is quantified by the slope as indicated. The maximum droplet density is 0.33 ± 0.16 (s.d.) [μm^{-3}] ($n = 124$). (f) S^* vs. Core-to-IDR ratio, f for FUS_N ($n = 115$) and $HNRNPA1_C$ ($n = 68$) Corelets. The solid circles represent the mean and error bars for S^* and f show standard errors of the fits with Eq. 1 and standard deviation, respectively. Dashed lines show best fit to $S^* \propto |f - f_c|^{0.84}$, where f_c is the critical Core-to-IDR ratio. (g) Phase diagram of FUS_N Corelets as functions of Core concentration and Core-to-IDR ratio ($n = 175$). Solid circles indicate cells where NG is observed, while empty triangles and squares indicate cells where no phase separation and SD are observed, respectively. Circle colors correspond to observed nucleation rate, J. [See Extended Data 4 for phase diagram of $HNRNPA1_C$ Corelets.] (h) J vs. supersaturation, S, for FUS_N ($n = 85$) and $HNRNPA1_C$ ($n = 64$) Corelets with $1/24 < f < 1/9$. Solid lines show the best fit to Eq. 1. n: number of cells.

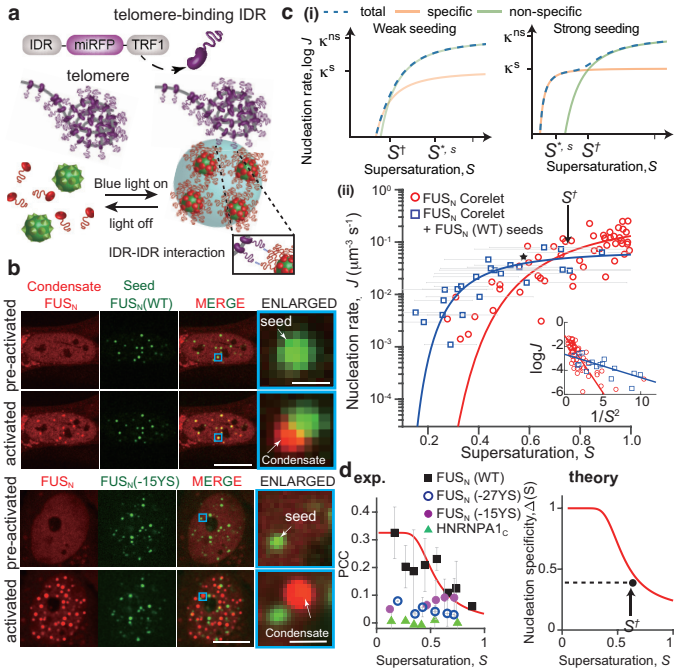


Figure 2. Engineering the condensate nucleation landscape. (a) Schematic diagram of a synthetic seed, consisting of miRFP-tagged telomere binding protein, TRF1, fused to IDRs, promoting localized Corelet phase separation. (b) Fluorescence images of U2OS cells expressing FUS_N Corelets plus FUS_N (WT, -15YS) seeds, before and after blue light activation. Enlarged images of indicated regions are shown on the right. Scale bars, 10 μm and 1 μm for enlarged images. (c) (i) Example $J-S$ curves for the cases of weak ($S^{*,s} > S^{\dagger}$) vs strong ($S^{*,s} < S^{\dagger}$) seeding. (ii) Nucleation rate, J, plotted against the supersaturation, S, for FUS_N Corelets with ($n = 23$) or without ($n = 76$) FUS_N (WT) seeds with $1/24 < f < 1/9$. The solid lines show the best fit to Eq. 1. (d) (left) Pearson correlation coefficient (PCC) between mCherry (Corelets) and miRFP (seeds) signals for FUS_N Corelets and FUS_N (WT ($n = 57$), -15YS ($n = 23$), -27YS ($n = 18$)) or $HNRNPA1_C$ ($n = 19$) seeds, against supersaturation, S. (right) Theoretical prediction of nucleation specificity $\Delta(S)$ vs. S. n: number of cells.

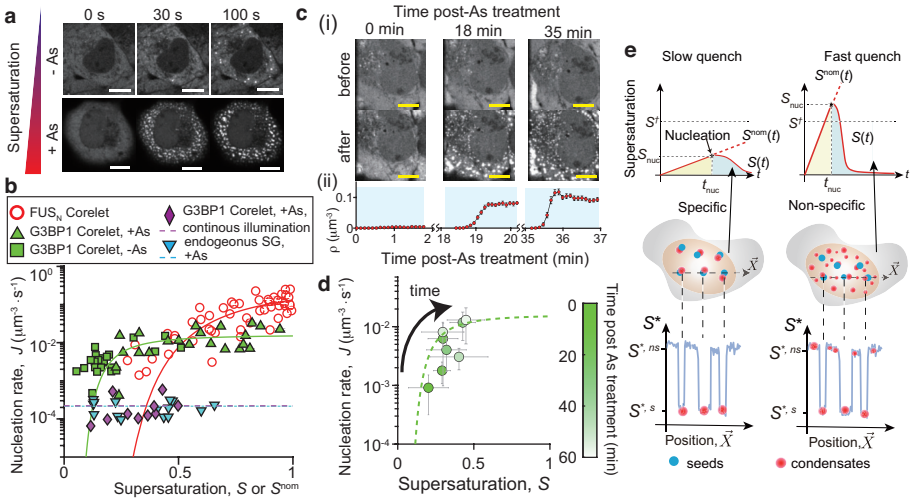
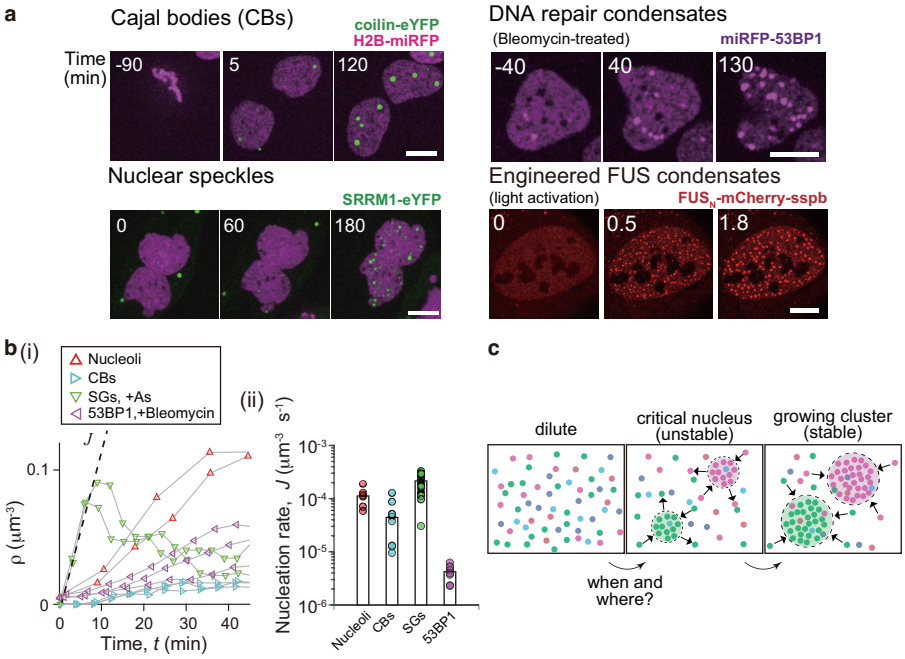
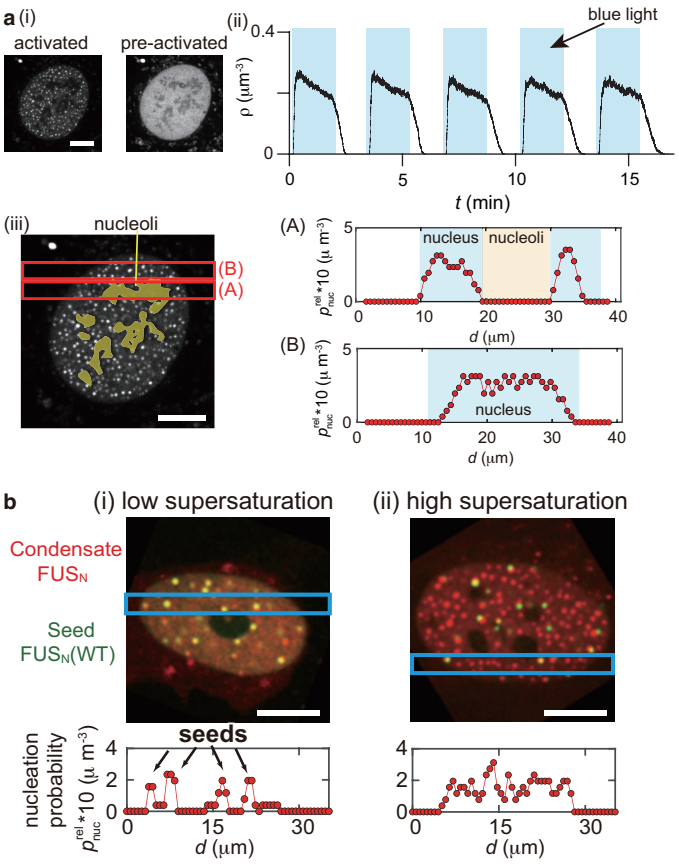


Figure 3. Temporal dynamics and the specificity of condensate nucleation. (a) Time lapse confocal images of photo-activated G3BP KO U2OS cells expressing G3BP1 corelets +/- one hour of As treatment (400 μM). Scale bars, 10 μm. (b) Nucleation rate, J, plotted against the supersaturation, S, for $\mathrm{FUS_N}$ Corelets ($n = 76$) and G3BP1 corelets +/- 400 μM As ($n = 45$) with $1/24 < f < 1/9$, and plotted against nominal supersaturation S^{nom} for G3BP1 corelets +As under conditions of constant illumination ($n = 10$) and endogenous SGs +As ($n = 15$). Dashed lines show mean J values for the latter two conditions, and solid lines show best fit to Eq. 1. (c) (i) Fluorescent images of G3BP KO cells expressing G3BP1 corelets before and after blue light activation at different time points post As (400 μM) treatment. Scale bars, 10 μm. (ii) Light-induced droplet number density, ρ, at different time points post As (400 μM) treatment. (d) Time change of J and S post As (400 μM) treatment ($n = 56$). (e) Schematic diagram for nucleation landscape of intracellular condensates. n: number of cells.

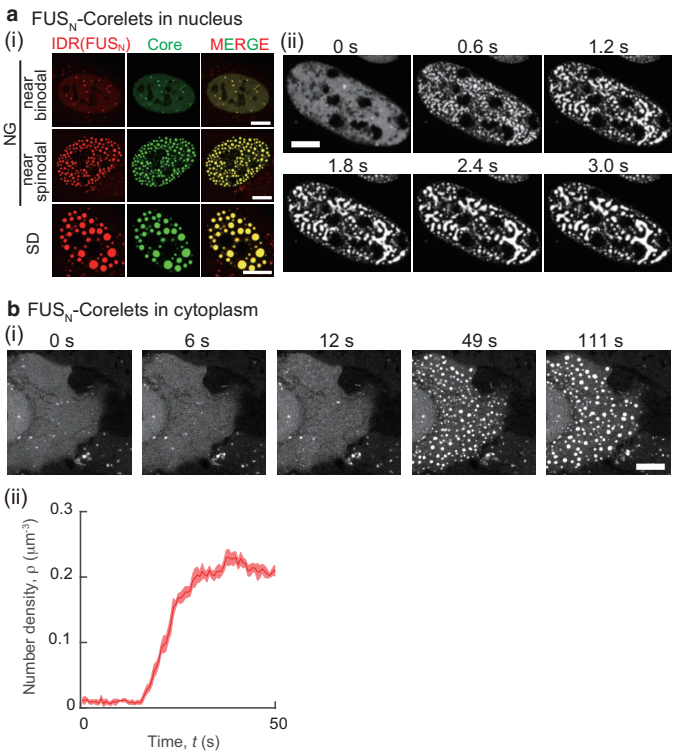
Extended Data



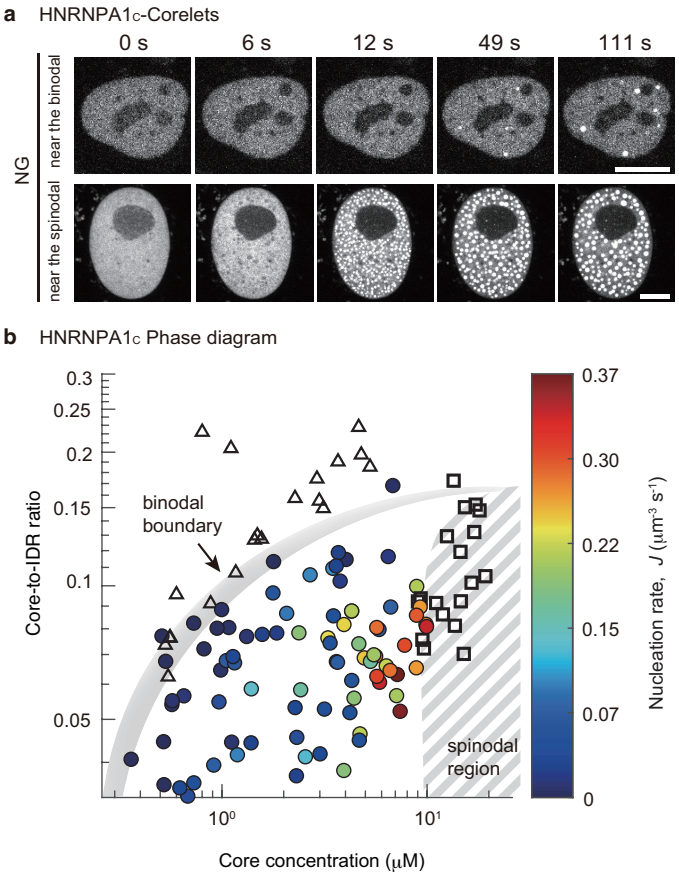
Extended Data Fig. 1. Nucleation process of various endogenous and biomimietic condensates. (a) Time-lapse images of U2OS cells show the nucleation of Cajal bodies (coilin-EYFP) and nuclear speckles (SRRM1-EYFP) in mitosis, DNA repair condensates (miRFP-53BP1) upon 10 μg/ml bleomycin treatment, and engineered FUS condensates. Scale bars, 10 μm. (b) (i) Number density of condensates, ρ, as a function of time, t for nucleoli and Cajal bodies in mitosis, stress granules (SGs) upon 400 μM As treatment, and 53BP1 condensates upon 10 μg/ml bleomycin treatment. The nucleation rate, J, is quantified by the slope. Here, $t = 0$ is $t_0 - 5$ (min). (ii) Nucleation rate, J, for nucleoli ($n = 6$), Cajal bodies ($n = 9$), SGs ($n = 24$) and DNA repair condensates ($n = 9$); $n =$ number of cells. (c) Schematic diagram of the nucleation process of intracellular condensates.



Extended Data Fig. 2. Quasi-1D nucleation probability. (a) (i) Fluorescence images of U2OS cells expressing HNRNPA1$_C$ -Corelets before and after blue light activation. Scale bar, 10 μm. (ii) Droplet number density as a function of time upon iterative blue-light activation and deactivation. (iii) Quasi-1D nucleation probability $p_{\mathrm{nuc}}^{\mathrm{rel}}$ in the indicated region which include nucleoli (A) or not (B). (b) Quasi-1D nucleation probability $p_{\mathrm{nuc}}^{\mathrm{rel}}$ of U2OS cells expressing FUS$_N$-Corelets and FUS$_N$-miRFP-TRF1 in the indicated region ((i) low and (ii) high supersaturation), calculated from five successive activation cycles. Scale bar, 10 μm.



Extended Data Fig. 3. Photo-activated phase separation in nucleus and cytoplasm. (a) Photo-activated phase separation in nucleus. (a-i) Confocal images of U2OS cells with different expression levels of FUS_N Corelets (FUS_N IDRs (red) and Cores (green)) after light-activation. The cells display nucleation and growth (NG) regimes between the binodal boundary and spinodal region, and spinodal decomposition (SD). Scale bars, 10 μm. (a-ii) Connected network-like growth and coarsening akin to spinodal decomposition. Scale bar, 10 μm. (b) Photo-activated phase separation in cytoplasm. (b-i) Time-lapse confocal images of photo-activated U2OS cells expressing FUS_N-Corelets composed of Core without NLS and FUS_N-mCh-sspb. Scale bars, 10 μm. (b-ii) Time change of light-induced droplet number density, ρ. The measure of centre for the shaded error bars is the mean and the error bars show standard deviation.



Extended Data Fig. 4. Phase separation behaviors of HNRNPA1$_C$-Corelets. (a) Time-lapse confocal images of photo-activated U2OS cells with different expression levels of HNRNPA1$_C$-Corelets. Nucleation growth (NG) regime near the binodal boundary (top) and spinodal region (bottom). Scale bars, 10 μm. (b) Phase diagram of HNRNPA1$_C$-Corelets as functions of Core concentration and Core-to-IDR ratio ($n = 161$). Solid circles exhibit cells where nucleation growth is observed, while empty triangles and squares show cells where no phase separation and spinodal decomposition are observed, respectively. The colors of solid circles indicate the observed nucleation rate, J. n: number of cells.

a

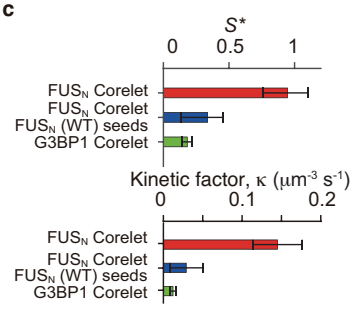
	FUS_N-Corelets	FUS_N Corelets plus FUS_N (WT) seeds	$HNRNPA1_C$-Corelets	G3BP1 corelets
$\kappa\ [\mu m^{-3} \cdot s^{-1}]$	0.34 ± 0.03	0.07 ± 0.01	0.47 ± 0.04	0.02 ± 0.00
S^*	0.97 ± 0.08	0.44 ± 0.04	1.04 ± 0.02	0.26 ± 0.01
χ^2_{red}	0.55	1.62	2.70	39.16

b

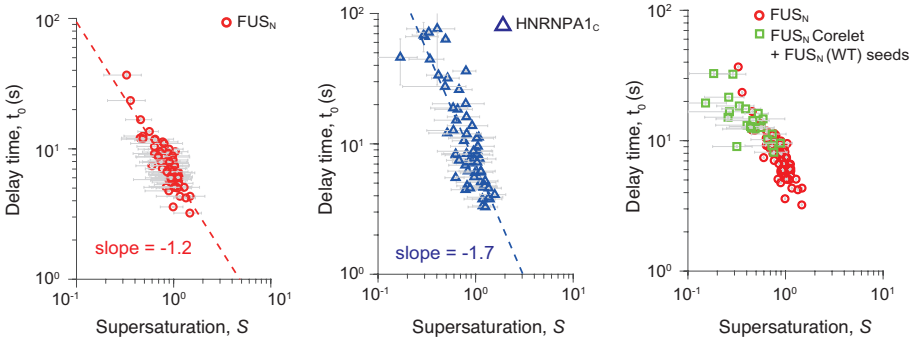
	FUS_N-Corelets, $1/24 < f < 1/18$	FUS_N-Corelets, $1/18 \leq f < 1/12$	FUS_N-Corelets, $1/12 \leq f < 1/6$
$\kappa\ [\mu m^{-3} \cdot s^{-1}]$	0.28 ± 0.07	0.33 ± 0.01	0.26 ± 0.01
S^*	1.05 ± 0.24	0.99 ± 0.24	0.61 ± 0.01
χ^2_{red}	0.34	0.29	6.83

	$HNRNPA1_C$-Corelets, $1/24 < f < 1/18$	$HNRNPA1_C$-Corelets, $1/18 \leq f < 1/12$	$HNRNPA1_C$, $1/12 \leq f < 1/6$
$\kappa\ [\mu m^{-3} \cdot s^{-1}]$	0.76 ± 0.11	0.46 ± 0.03	0.29 ± 0.03
S^*	1.59 ± 0.12	0.93 ± 0.08	0.64 ± 0.06
χ^2_{red}	1.39	1.25	2.85

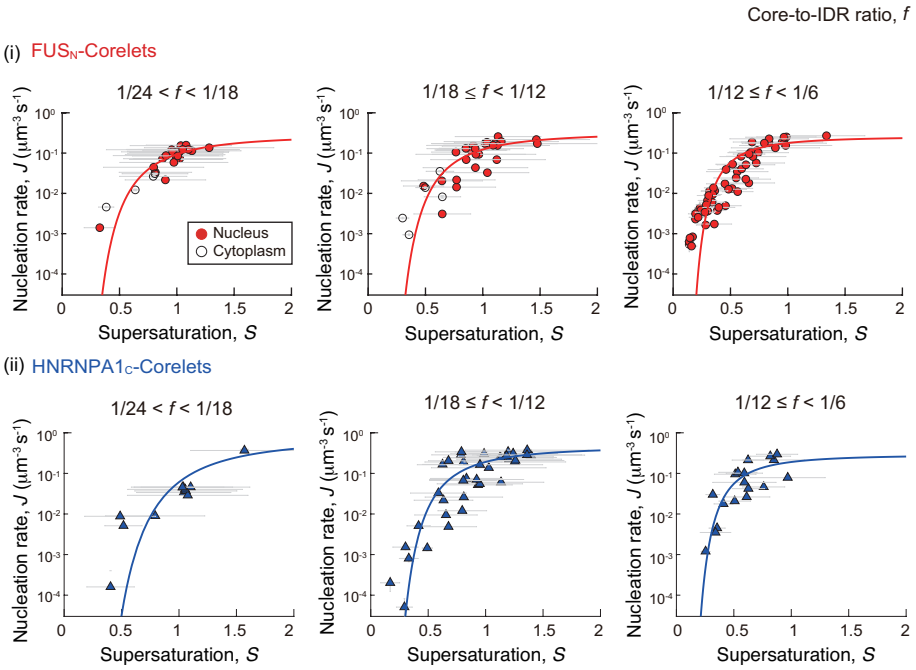
c



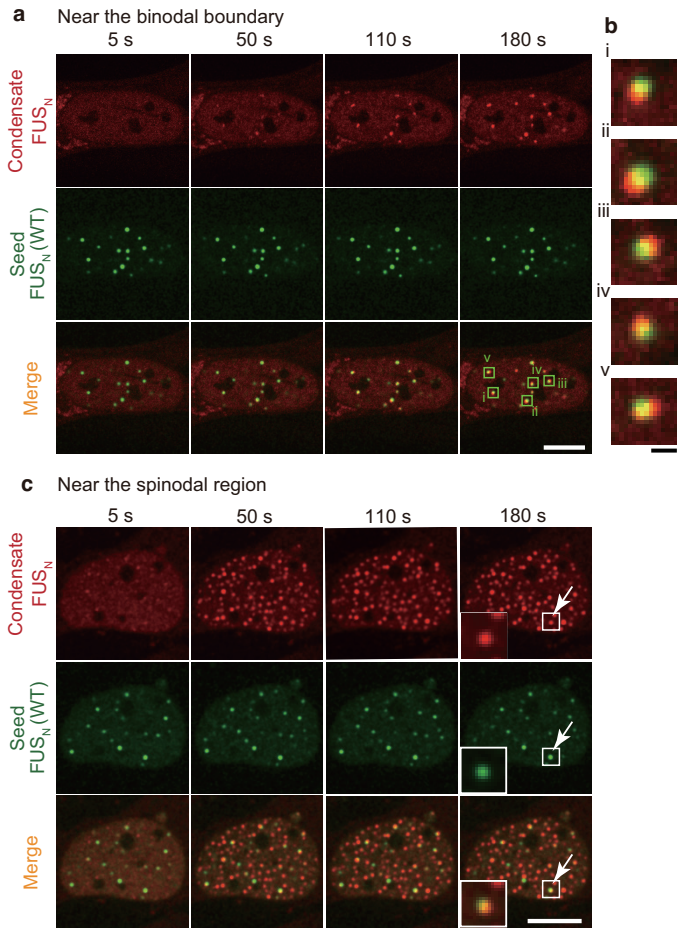
Extended Data Fig. 5. Model fit parameters and chi-square values. (a,b) Model fit parameters of S^*, κ, and χ^2_{red} for various biomimetic condensates. The errors show the standard errors with 68% confidence intervals of the fits by Eq. 1. The parameter $\chi^2_{red} \gg 1$ indicates a poor model fit while $\chi^2_{red} \lesssim 1$ indicates a good model fit. (c) S^* (top) and κ (bottom) obtained from fits to Eq. 1 for FUS_N Corelets with ($n = 23$) or without ($n = 76$) FUS_N (WT) seeds and G3BP1 corelets +/- 400 μM As (n=45). Error bars represent standard fit errors.



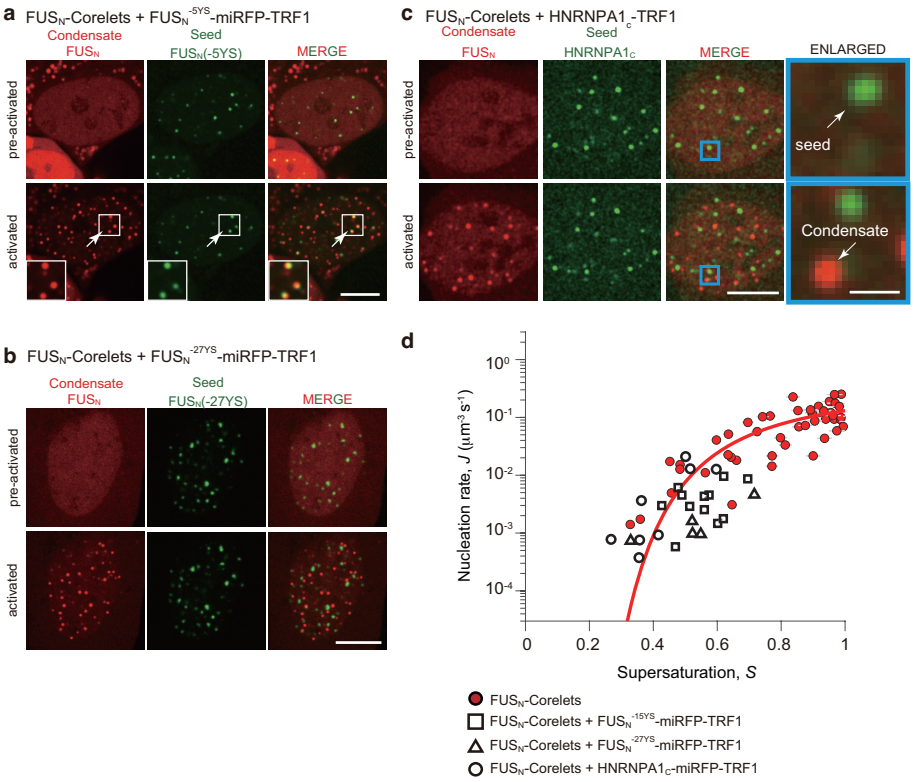
Extended Data Fig. 6. Delay time against supersaturation. Delay time, t_0, against supersaturation, S, for FUS_N-Corelets ($n = 70$), $HNRNPA1_C$-Corelets ($n = 60$), and FUS_N Corelets plus FUS_N (WT) seeds ($n = 23$) with $1/24 < f < 1/10$. The dashed lines show the best power-law fits. The data points for t_0 and S represent the best fit parameters and the mean, respectively. Error bars for t_0 and S show the standard errors of the fits and standard deviation, respectively. n: number of cells.



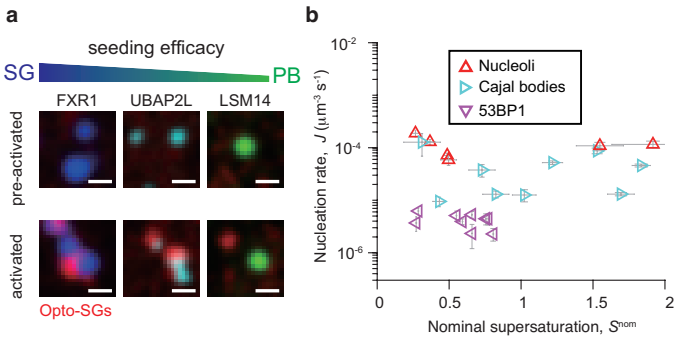
Extended Data Fig. 7. J against S with various ranges of Core-to-IDR ratio. Nucleation rate, J, plotted against the degree of supersaturation, S, for (i) FUS_N-Corelets ($n = 26$ for $1/24 < f < 1/18$, $n = 33$ for $1/18 \leq f < 1/12$ and $n = 56$ for $1/12 \leq f <$ 1/ 1/6) and (ii) $HNRNPA1_C$-Corelets ($n = 12$ for $1/24 < f < 1/18$, $n = 38$ for $1/18 \leq f < 1/12$ and $n = 18$ for $1/12 \leq f <$ 1/ 1/6). The data points J and S represent the best fit parameters and the mean. Error bars for J and S show the standard errors of the fits and standard deviation, respectively. See Extended Data Fig. 5 for model fit parameters and χ^2_{red}. n: number of cells.



Extended Data Fig. 8. Seeded nucleation near the binodal boundary and spinodal region. Time-lapse confocal images of U2OS cells expressing FUS_N-Corelets and FUS_N-miRFP-TRF1 near the (a) binodal boundary and (c) spinodal region. The insets show the regions indicated by the arrows. Scale bars, 10 μm. (b) The seeded condensates, which are indicated in (a), are enlarged. Scale bar, 1 μm.



Extended Data Fig. 9. Nucleation of FUS$_N$-Corelets with tyrosine-mutated FUS$_N$-seeds. (a-c) Fluorescent images of U2OS cells expressing FUS$_N$-Corelets and FUS$_N$ (-5YS, -27YS)-miRFP-TRF1 or HNRNPA1$_C$-miRFP-TRF1 before and after blue light activation. The insets and enlarged images show the regions indicated by the squares. Scale bars, 10 μm. (d) Nucleation rate, J, plotted against the degree of supersaturation, S, for FUS$_N$-Corelets ($1/24 < f < 10$) and FUS$_N$-Corelets ($1/30 < f < 10$) with FUS$_N$ (-15YS, -27YS)-miRFP-TRF1 ($n_{15YS}, n_{27YS} = 12, 5$) or HNRNPA1$_C$-miRFP-TRF1 ($n = 8$). n: number of cells.



Extended Data Fig. 10. Specificity of the nucleation of endogenous condensates. (a) Confocal images of G3BP KO cells expressing G3BP1 corelets and iRFP-tagged FXR1/UBAP2L/LSM14 before and after blue light activation in the absence of As. Scale bars, 1 μm. (b) Nucleation rate of endogenous condensates against nominal supersaturation. Nucleation rate, J, plotted against the nominal supersaturation, S^{nom}, for nucleoli ($n = 6$) and Cajal bodies ($n = 9$) in mitosis and 53BP1 condensates ($n = 9$) upon 10 μg/ml bleomycin treatment. n: number of cells.